Journal of Imaging Informatics in Medicine (2024) 37:1874–1888
https://doi.org/10.1007/s10278-024-01070-5

GA-UNet: A Lightweight Ghost and Attention U-Net for Medical Image Segmentation

Bo Pang[1] · Lianghong Chen[1] · Qingchuan Tao[1] · Enhui Wang[1] · Yanmei Yu[1]

Received: 7 December 2023 / Revised: 13 February 2024 / Accepted: 22 February 2024 / Published online: 13 March 2024

Abstract

U-Net has demonstrated strong performance in the field of medical image segmentation and has been adapted into various variants to cater to a wide range of applications. However, these variants primarily focus on enhancing the model's feature extraction capabilities, often resulting in increased parameters and floating point operations (Flops). In this paper, we propose GA-UNet (Ghost and Attention U-Net), a lightweight U-Net for medical image segmentation. GA-UNet consists mainly of lightweight GhostV2 bottlenecks that reduce redundant information and Convolutional Block Attention Modules that capture key features. We evaluate our model on four datasets, including CVC-ClinicDB, 2018 Data Science Bowl, ISIC-2018, and BraTS 2018 low-grade gliomas (LGG). Experimental results show that GA-UNet outperforms other state-of-the-art (SOTA) models, achieving an F1-score of 0.934 and a mean Intersection over Union (mIoU) of 0.882 on CVC-ClinicDB, an F1-score of 0.922 and a mIoU of 0.860 on the 2018 Data Science Bowl, an F1-score of 0.896 and a mIoU of 0.825 on ISIC-2018, and an F1-score of 0.896 and a mIoU of 0.853 on BraTS 2018 LGG. Additionally, GA-UNet has fewer parameters (2.18M) and lower Flops (4.45G) than other SOTA models, which further demonstrates the superiority of our model.

Keywords Medical image segmentation · GhostV2 bottleneck · Convolutional Block Attention Module · Computer-aided diagnosis

Introduction

Currently, medical image analysis plays a vital role in the diagnosis and treatment of diseases [1]. For example, doctors using microscopes to analyze images of cells can determine the stage and type of disease. These images allow doctors to provide valuable insights into disease diagnosis by closely observing diseased areas of the cells. Traditional medical image analysis is typically performed by doctors, and the diagnostic results can be influenced by the subjective judgment of doctors. Furthermore, diagnosing medical images is time-consuming and labor-intensive. And working for long periods of time can make doctors feel tired, which can affect the accuracy of diagnostic results. With the development of computers, computer-aided diagnosis (CAD) has received widespread attention from pathology researchers [2]. The results of the CAD system are more objective. CAD can handle a large amount of work in a short time, improving doctors' work efficiency and reducing their burden. Medical image segmentation is to separate the target area from the original image, determine the location that requires detailed analysis, and provide a more targeted medical diagnosis.

Traditional medical image segmentation algorithms include edge detection and threshold segmentation, such as the Canny operator [3] and the Otsu threshold method [4]. However, these traditional algorithms require adjusting parameters for different applications. These parameters are highly sensitive to both image quality and noise, which makes it challenging to process complex images and changing tasks [5].

With the advancement of deep learning technology, U-Net [6] and its derived models have been widely utilized in medical image segmentation tasks. The success of U-Net is mainly attributed to the skip connection strategy of the encoder and decoder framework. This strategy combines low-level and high-level semantic information from the encoder to form more complex and effective features. Attention-UNet [7] improves the skip connection by incorporating coarse-scale information for gating,

✉ Yanmei Yu
yuyanmei@scu.edu.cn

[1] College of Electronics and Information Engineering, Sichuan University, Chengdu 610065, China

aiming to mitigate the noise response caused by the skip connection. However, both methods rely on traditional convolutional blocks for feature extraction, leading to the generation of redundant information. In order to fully exploit voxel information in computed tomography (CT) and magnetic resonance imaging (MRI), some 3D models have been proposed [8, 9]. However, 3D models have many parameters. Recently, Vision Transformer (ViT) [10] has shown promising potential in medical image segmentation tasks. LeViT-UNet [11] integrates a LeViT Transformer module [12] into the U-Net architecture, striking a balance between model accuracy and efficiency. However, the Transformer module imposes a significant computational burden primarily due to self-attention operations, resulting in an increase in model size in terms of parameters and floating point operations (Flops). The models mentioned above require a large amount of training data to avoid overfitting. This can be a challenge when working with medical image datasets, which are often small in size. Therefore, a key challenge is to effectively use fewer parameters to achieve better performance in medical image segmentation.

In this paper, we propose a lightweight Ghost and Attention U-Net for medical image segmentation, called GA-UNet. We use the GhostV2 bottleneck [13] to extract features in the encoder and decoder parts, and depthwise separable convolution [14] for downsampling to achieve a lightweight model. In order to further improve the segmentation capability of the lightweight model, we introduce the Convolutional Block Attention Module (CBAM) [15]. The main contributions of this work can be summarized as follows:

1. A lightweight and efficient model for medical image segmentation is proposed, namely GA-UNet. Our proposed model achieves higher accuracy with only 2.18M parameters and 4.45G Flops, demonstrating faster convergence speed and stronger generalization performance.
2. To address the issue of feature redundancy caused by traditional convolution modules, we employ the GhostV2 bottleneck [13] for feature extraction. This effectively reduces model complexity.
3. To further enhance the model's accuracy, we incorporate CBAM [15] in the first three decoders. By combining channel and spatial information into attention maps, our model better captures the lesion locations.
4. Four datasets are used to evaluate our model, including CVC-ClinicDB [16], 2018 Data Science Bowl [17], ISIC-2018 Task 1 Lesion Boundary Segmentation [18, 19] and BraTS 2018 low-grade gliomas (LGG) [20]. Experiments show that our GA-UNet performs better than other state-of-the-art (SOTA) methods in terms of mean Intersection over Union (mIoU) and F1-score indicators.

Related Work

Medical Image Segmentation

Deep learning methods have demonstrated remarkable efficiency in the field of medical image segmentation. The U-Net [6] comprises two components: the encoder, responsible for feature extraction and image size compression, and the decoder, responsible for image resolution restoration. To achieve more accurate segmentation results, skip connections are employed between the encoder and decoder. Inspired by U-Net, Jha et al. [21] proposed the DoubleU-Net architecture which connects two U-Nets and incorporates atrous spatial pyramid pooling (ASPP) [22] between each U-Net's encoder and decoder, achieving effective segmentation in various tasks such as intestinal polyp segmentation, lesion boundary segmentation, and cell nucleus segmentation. Additionally, Jha et al. [23] integrated the residual structure [24], ASPP [22] and Squeeze-and-Excitation (SE) [25] attention mechanism into U-Net to design the ResUNet++ network for accurate intestinal polyp segmentation. To fully leverage multi-scale features and reduce the false positives in non-organ images, Huang et al. [26] presented UNet3+, which achieved efficient segmentation on liver and spleen datasets. Lama et al. [27] incorporated EfficientNet [28] along with SE module into U-Net, achieving satisfactory results on the ISIC2017 lesion segmentation dataset. By utilizing LinkNet [29] with EfficientNet [28] variants, Singh et al. [30] yielded promising results in glomerular detection tasks. The proposal of a residual deformable split depthwise separable U-Net (RDSDSU-NET) [31] achieved precise liver and liver tumor segmentation results. Furthermore, the success of Transformer [32] in the field of natural language processing has attracted widespread attention from researchers. For example, the Vision Transformer structure proposed by Dosovitskiy et al. [10] is widely adopted within computer vision research. Xu et al. [11] constructed LeViT-UNet to achieve good segmentation performance on the Synapse multi-organ segmentation dataset and Automatic Cardiac Diagnosis Challenge dataset. Feng et al. [33] developed a framework for brain tumor segmentation based on a deep convolutional neural network (DCNN). This framework incorporates a novel sequence dropout technique, exhibiting better robustness and performance than before. To extract comprehensive contextual information effectively and achieve fast and accurate segmentation of medical images, Tang et al. [34] designed CMUNeXt, a fully convolutional lightweight medical image segmentation network based on U-Net. This approach demonstrated effective results in breast cancer and thyroid nodule segmentation tasks. In order to mitigate the loss of U-Net encoder header information and enhance the network's ability for expressing multi-scale features, Xu et al. [35] devised distinct modules, namely

the primary feature conservation module and compact split attention module, to achieve efficient segmentation in tasks involving intestinal polyps, lesion boundaries, cell nuclei, and myeloma plasma cells.

Lightweight Modules

Convolutional neural networks have significantly enhanced the performance of various computer vision tasks. However, a large amount of computational cost and parameters are required. Howard et al. [14] proposed depthwise separable convolution (DSC) as a method to construct MobileNets, a lightweight neural network that strikes a balance between resource utilization and accuracy. This approach has been extensively employed in detection and classification tasks. To address the issue of poor performance of DSC in low-channel features, Sandler et al. [36] introduced convolution before DSC and proposed the MobileNetV2 network, which maintains lightweight while mitigating performance degradation. Han et al. [37] designed GhostNet to reduce the computational costs of deep neural networks and achieve a better balance between efficiency and accuracy. In order to tackle the problem of weak long-distance modeling ability of lightweight convolutional networks, Tang et al. [13] stacked the Ghost modules in the GhostNetV2 model and introduced the decoupled fully connected (DFC) attention module in the Ghost module to enhance feature representation capabilities of the middle layer, thereby achieving a better trade-off between accuracy and inference speed.

Attention Modules

In recent years, attention modules have been widely used in the field of deep learning. Hu et al. [25] proposed the SE module, which compresses global spatial information and then learns features in the channel dimension, assigning different weights to channels. Jaderberg et al. [38] proposed a spatial attention module to transform various deformation data in space and capture effective features. Woo et al. [15] introduced CBAM, which generates two independent dimensions of attention maps (channel and space), then fuses these attention maps with the input features for adaptive feature refinement, which can improve the network's ability to extract effective features.

Method

A Lightweight Model: GA-UNet

We propose GA-UNet, a lightweight U-shaped Model based on the GhostV2 bottleneck [13] and CBAM [15] for medical image segmentation, as shown in Fig. 1. Our model not only achieves higher segmentation accuracy, but also has fewer parameters and Flops. The encoder, consisting mainly of the GhostV2 bottleneck and the DSC downsample module, extracts information and downsamples the input features. Different from U-Net, our DSC downsample module uses DSC [14] instead of max pooling. The decoder, consisting mainly of the GhostV2 bottleneck, upsample module, and CBAM, is used to reconstruct features. The encoders and decoders perform features interaction through skip connections.

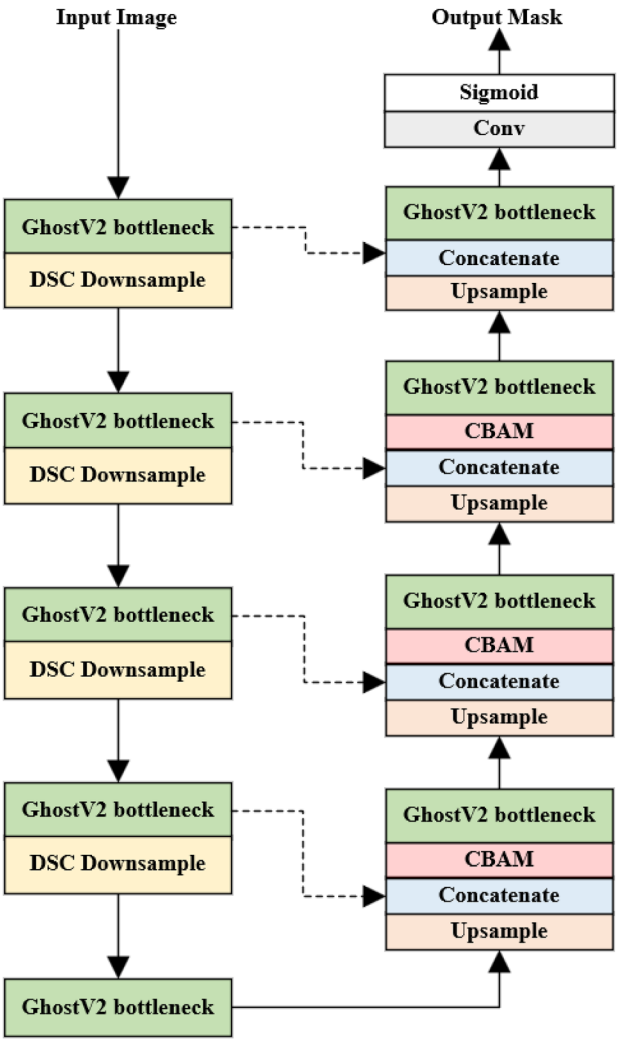

Fig. 1 The architecture of GA-UNet. GA-UNet consists mainly of the GhostV2 bottleneck, the DSC downsample module and CBAM

GhostV2 Bottleneck

To address the issue of redundant features and excessive model parameters caused by the convolution module, we introduce the GhostV2 bottleneck, as shown in Fig. 2. The GhostV2 bottleneck consists of two Ghost modules, a DFC attention module and a residual connection. The

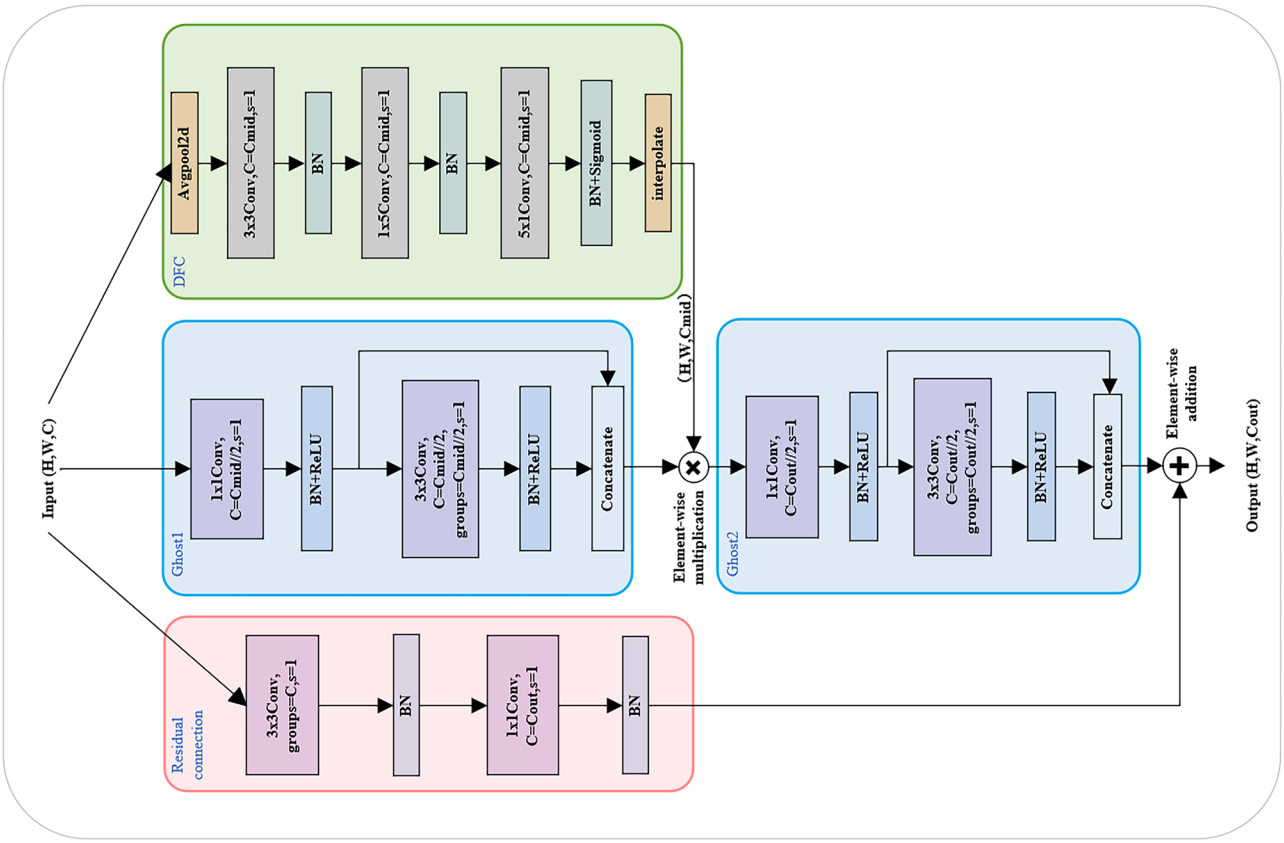

Fig. 2 The diagram of the GhostV2 bottleneck. It includes a residual connection (red box), two Ghost modules (blue boxes), and a DFC attention module (green box)

implementation of each Ghost module is divided into two steps: the first step involves a regular convolution with a kernel size of 1×1 to generate the intermediate feature $Y^{'}$, and the second step involves a cheap operation (depthwise convolution) to generate additional feature maps, which are then concatenated with the intermediate feature $Y^{'}$ to form the output feature Y. The two steps for the implementation of the Ghost module mentioned above are calculated as Eqs. (1) and (2):

$$Y^{'} = X * Conv^{1\times1} \tag{1}$$

$$Y = Concat([Y^{'}, Y^{'} * Conv_{dp}^{3\times3}]) \tag{2}$$

where $X \in \mathbb{R}^{H\times W\times C}$ is the input features with channel number C, height H, and width W, $*$ stands for convolution operation, $Conv^{1\times1}$ is a convolution with a kernel size of 1×1, $Conv_{dp}^{3\times3}$ is the depthwise convolution with a kernel size of 3×3, and the output feature is $Y \in \mathbb{R}^{H\times W\times C_{out}}$.

To better capture spatial information, the DFC attention module is merged with the first Ghost module of GhostV2 bottleneck. The DFC attention module generates a global attention map by stacking both horizontal and vertical fully connected layers. This module allows the GhostV2 bottleneck to achieve a better balance between accuracy and inference speed. The specific calculation of the DFC attention module can be written as Eqs. (3) and (4):

$$a_{hw}^{'} = \sum_{h^{'}=1}^{H} L_{h,h^{'}w}^{H} \odot x_{h^{'}w} \tag{3}$$

$$a_{hw} = \sum_{w^{'}=1}^{W} L_{w,hw^{'}}^{W} \odot a_{hw^{'}}^{'} \tag{4}$$

where $\odot$ denotes element-wise multiplication, L is the learnable weights in the fully connected layers, $X = \{x_{11}, x_{12}, ...x_{HW}\}$ is the input feature that can be seen as HW tokens $x_i \in \mathbb{R}^C$, $A = \{a_{11}, a_{12}, ...a_{HW}\}$ is the attention map, and $h = 1, 2, ..., H, w = 1, 2, ..., W$.

Finally, a residual connection is utilized to combine the output features of the second Ghost module with the input features X to obtain the final output features.

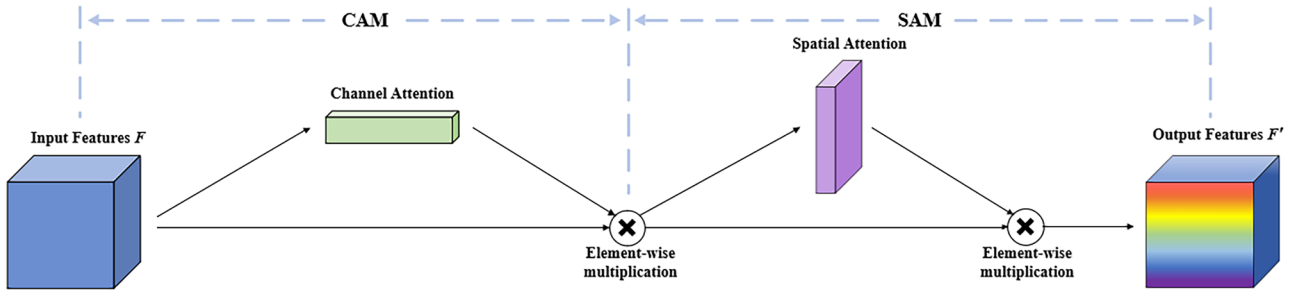

Fig. 3 The overview of CBAM. CBAM has two sequential sub-modules: CAM and SAM

Convolutional Block Attention Module

The Convolutional Block Attention Module (CBAM) [15] combines both channel and spatial attention to generate a more comprehensive attention map, as shown in Fig. 3.

The channel attention module (CAM) provides adaptive weights to the input features in the channel dimension, as shown in Fig. 4. In CAM, the spatial information of input features is first compressed using max pooling and average pooling, and then two linear layers are applied to extract channel information. Finally, the channel attention map $M_c(F)$ is generated by element-wise addition and a sigmoid layer. The channel attention map $M_c(F)$ is calculated according to Eq. (5):

$$M_c(F) = \sigma(Linearlayers(Avgpool(F)) + Linearlayers(Maxpool(F))) \tag{5}$$

where σ is the sigmoid activation function.

In Eq. (6), $M_c(F)$ is then multiplied with the input features F to generate the channel attention features F_c.

$$F_c = M_c(F) \otimes F \tag{6}$$

where $\otimes$ denotes element-wise multiplication.

The spatial attention module (SAM) provides spatially adaptive feature weights for input features, as shown in Fig. 5. In SAM, the channel attention features F_c first perform the global max pooling and average pooling operations, respectively. The results are then concatenated and input into a convolutional layer with a kernel size of 7×7 for feature extraction. Finally, the spatial attention map $M_s(F_c)$ is generated by sigmoid activation function as shown in Eq. (7):

$$M_s(F_c) = \sigma(Conv^{7\times7}([Avgpool(F_c); Maxpool(F_c)]) \tag{7}$$

where σ is the sigmoid activation function, and $Conv^{7\times7}$ is a convolution layer with a kernel size of 7×7.

$M_s(F_c)$ is then multiplied with the channel attention features F_c to generate the final output features F' in Eq. (8):

$$F' = M_s(F_c) \otimes F_c \tag{8}$$

where $\otimes$ denotes element-wise multiplication.

We introduce CBAM in the first three decoders. This module enhances the ability of GA-UNet to extract effective information from high-dimensional abstract features, thereby improving accuracy without significantly increasing computational effort.

Experiments and Results

Datasets

To evaluate the effectiveness of GA-UNet, we test it on four public medical datasets:

1. CVC-ClinicDB [16]: this is the training database for the MICCAI 2015 Polyp Detection Challenge, a common

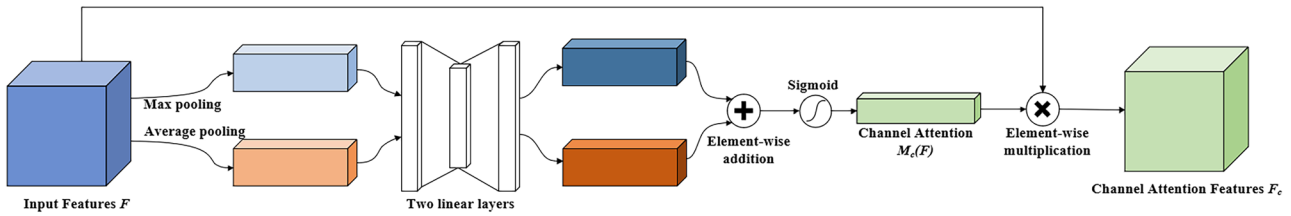

Fig. 4 The detail of CAM. CAM consists mainly of max pooling, average pooling and two linear layers

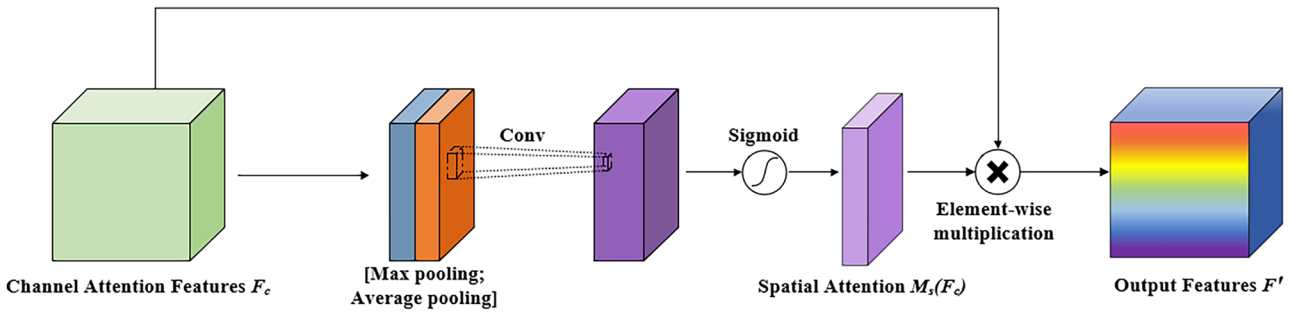

Fig. 5 The detail of SAM. SAM consists mainly of max pooling, average pooling and a convolutional layer

dataset for the polyp segmentation task. This dataset includes 612 images with a resolution of 384 × 288 from 29 colonoscopy sequences.

2. 2018 Data Science Bowl [17]: this dataset was provided for the Data Science Bowl competition held in 2018. It is used for the nuclei segmentation task, containing 670 images of cell nuclei at various resolutions. Before feeding the images into our model, we resize them to 256 × 256.
3. ISIC-2018 [18, 19]: this dataset is the Task 1 dataset released by the International Skin Imaging Collaboration (ISIC) in 2018. It is used for skin lesion area segmentation task and contains 2594 dermoscopic images of various resolutions.
4. BraTS 2018 LGG [20]: this dataset comprises preoperative MRI scans of 75 low-grade gliomas (LGG) from the Brain Tumor Segmentation (BraTS) Challenge 2018. It is used for the entire tumor area segmentation task. There are four kinds of multimodal scans for each patient: T1-weighted, T1-weighted with contrast, T2-weighted, and Fluid-Attenuated Inversion Recovery (FLAIR). We extract 4845 2D slides from the FLAIR MRI scans of LGG.

More details about the data split are shown in Table 1.

Evaluation Metrics

The main evaluation indicators of this paper are as follows: Accuracy, Precision, Recall, F1-score, and mIoU. They can be calculated by Eqs. (9–13):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

$$Recall = \frac{TP}{TP + FN} \tag{11}$$

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{12}$$

$$mIoU = \frac{1}{k} \sum_{i=1}^{k} \frac{Prediction \cap GroundTruth}{Prediction \cup GroundTruth} \tag{13}$$

where True Positive (TP) is the number of correctly predicted positive samples, True Negative (TN) is the number of correctly predicted negative samples, False Positive (FP) and False Negative (FN) indicate the number of incorrectly predicted positive and negative samples. And mIoU refers to the mean Intersection over Union between the prediction and ground truth.

As shown in Fig. 6, we visualize TP, TN, FP, and FN. Figure 6(d) is the comparison between the ground truth and the prediction results, which is generated according to the following rules:

1. If a pixel is white in Fig. 6(b) and white in Fig. 6(c), then it is white in Fig. 6(d), representing TP, as shown in Fig. 6(e).
2. If a pixel is black in Fig. 6(b) and black in Fig. 6(c), then it is black in Fig. 6(d), representing TN, as shown in Fig. 6(f).

Table 1 Details of the medical segmentation datasets used in our experiments

Datasets	Images	Input size	Train	Valid	Test
CVC-ClinicDB	612	384 × 288	440	111	61
2018 Data Science Bowl	670	Variable	482	121	67
ISIC-2018	2594	Variable	1868	467	259
BraTS 2018 LGG	4845	240 × 240	3488	872	485

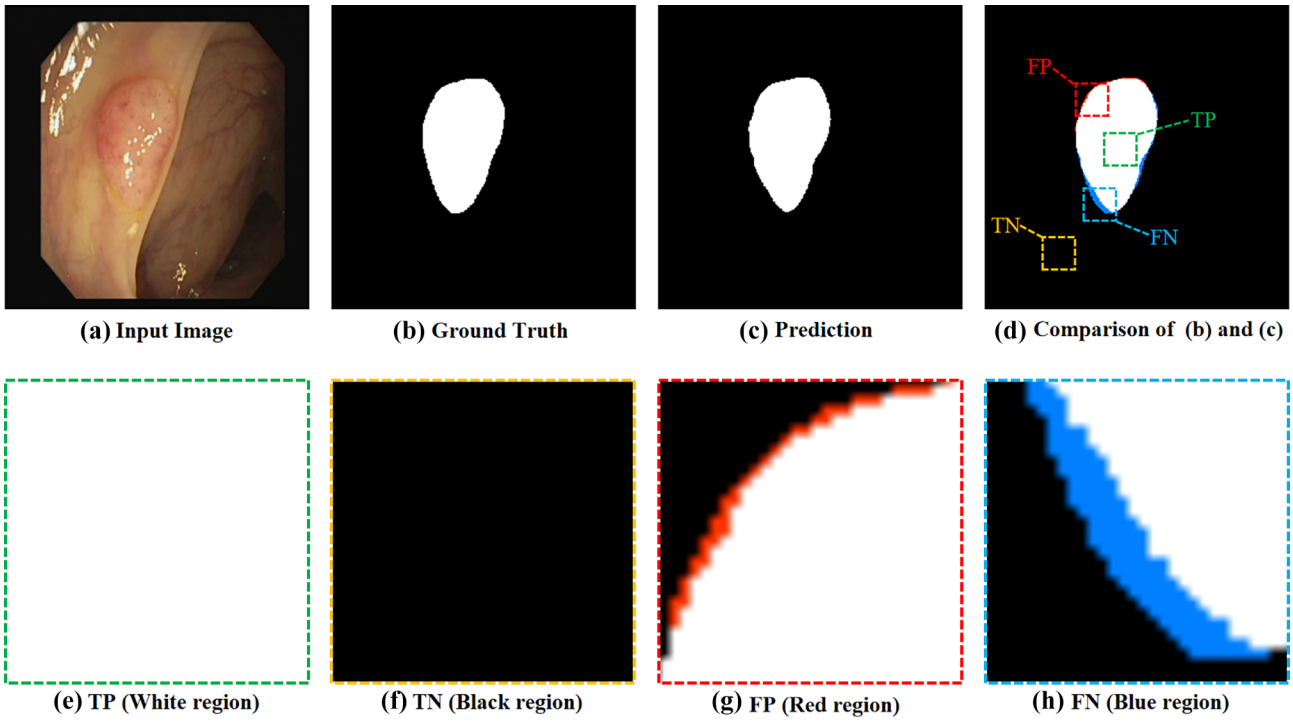

Fig. 6 Visualization of TP, TN, FP, FN

3. If a pixel is black in Fig. 6(b) and white in Fig. 6(c), then it is red in Fig. 6(d), representing FP, as indicated by the red region in Fig. 6(g).
4. If a pixel is white in Fig. 6(b) and black in Fig. 6(c), then it is blue in Fig. 6(d), representing FN, as indicated by the blue region in Fig. 6(h).

In addition, the indicators parameters and Flops are used to evaluate the effectiveness of the model.

Data Augmentation

The sizes of medical image datasets are usually small due to the expensive and time-consuming process of obtaining and annotating these images. This limitation can result in model overfitting. To address this issue, we incorporate data augmentation techniques during the training phase to increase sample diversity and enhance the model's generalization ability. Specifically, we apply horizontal flip, cutout, and rotation data augmentation methods with a probability of 0.25 on the training set, as depicted in Fig. 7.

Implementation Details

All experiments are implemented using PyTorch 1.10.0 framework on a RTX 3090 (24GB) and 12 vCPU Intel (R) Xeon (R) Platinum 8255C CPU @ 2.50GHz.

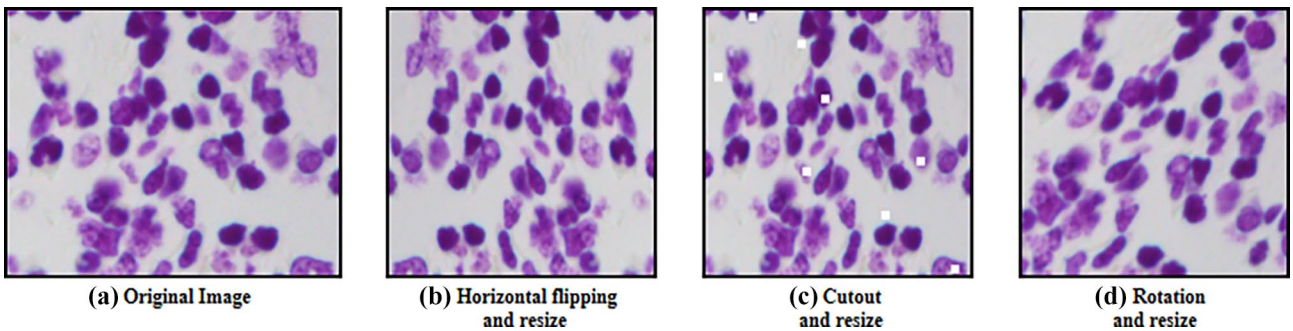

Fig. 7 Examples of data augmentation

During the training phase, we use the Dice loss function [39] and the Adam optimizer [40]. The batch size is set to 16. The initial learning rate is 1e-3 and then gradually reduced to 0 using the cosine annealing scheduler [41] over the entire 150 epochs.

Images in all four datasets are resized to 256 × 256 before being input into the model. The experiments on the four datasets all use the same split method of training sets, validation sets and test sets as shown in Table 1. Those SOTA models used for comparison are trained from scratch with the same parameters as our model. The batch size of UNet3+ in the training phase is set to 8.

Analysis of Results

In this section, our GA-UNet is compared with seven previous SOTA methods on four medical image segmentation datasets.

Comparison on CVC-ClinicDB Dataset

Early diagnosis and treatment of polyps are crucial for colon cancer prevention. We compare the performance of GA-UNet and other seven SOTA models on the polyp segmentation task (CVC-ClinicDB dataset), as shown in Table 2. Our GA-UNet's F1-score and mIoU index are improved by 0.9% and 1.6%, respectively, compared to DCSAU-Net, and by 3.1% and 4.2%, respectively, compared to DoubleU-Net. Meanwhile, compared with DCSAU-Net, our GA-UNet reduces parameters and Flops by 16.2% (0.42M) and 35.7% (2.47G), respectively. In Fig. 8(a), GA-UNet (red line) achieves the highest scores in all five assessment metrics, demonstrating the potential of our model for the clinical task of intestinal polyp segmentation.

Comparison on 2018 Data Science Bowl

The results and visualization on the 2018 Data Science Bowl dataset are shown in Table 3 and Fig. 8(b), our GA-UNet's F1-score and mIoU index of GA-UNet are 0.4% and 0.6% higher than those of DCSAU-Net, respectively, 2.0% and 2.8% higher than those of LeViT-UNet, respectively, and 0.5% and 0.7% higher than those of DoubleU-Net, respectively. Although GA-UNet shows similar performance in evaluation metrics to UNet3+, the advantage of our model lies in its significantly fewer parameters and lower Flops about 1/40 and 1/10 of UNet3+, respectively.

Comparison on ISIC-2018

Accurately delineating the area of a lesion is crucial for diagnosing dermatologic diseases. We assess the performance of GA-UNet and other SOTA models using the ISIC-2018 dataset, as shown in Table 4 and Fig. 8(c), GA-UNet improves the F1-score and mIoU by 1.0% and 1.0%, respectively, compared to DCSAU-Net. And it achieved 0.963, 0.910, and 0.911 in Accuracy, Precision, and Recall, respectively, which are better than other models.

Comparison on BraTS 2018 LGG

Brain tumor segmentation is indispensable in MRI analysis. The performance of our model and other SOTA models in the BraTS 2018 LGG dataset is shown in Table 5 and Fig. 8(d). GA-UNet achieves the mIoU of 0.853 and the F1-score of 0.896, which are 3% and 2.7% higher than DCSAU-Net, respectively. And it outperforms other SOTA models in Accuracy, Precision, and Recall metrics, as shown in Fig. 8(d). GA-UNet demonstrates strong performance, suggesting that it is an efficient model for medical image segmentation.

Ablation Study

In this section, we conduct an ablation study on GA-UNet to verify the effectiveness of our improvements, as shown in Table 6.

In Table 6, E_1 represents baseline model U-Net, E_2 represents U-Net+GhostV2 bottleneck and DSC downsample, and E_3 represents U-Net+CBAM. Lastly, E_4 represents our GA-UNet.

Table 2 Results on the CVC-ClinicDB. The best results are in bold

Method	Accuracy	Precision	Recall	F1-score	mIoU	Flops	Parameters
U-Net (2015) [6]	0.977 ± 0.031	0.898 ± 0.152	0.874 ± 0.182	0.875 ± 0.158	0.802 ± 0.181	48.33G	28.95M
ResUNet++ (2019) [23]	0.976 ± 0.029	0.878 ± 0.185	0.846 ± 0.202	0.854 ± 0.185	0.776 ± 0.196	70.99G	14.48M
DoubleU-Net (2020) [21]	0.981 ± 0.028	0.923 ± 0.086	0.903 ± 0.146	0.903 ± 0.120	0.840 ± 0.154	44.37G	18.84M
UNet3+ (2020) [26]	0.981 ± 0.027	0.920 ± 0.105	0.909 ± 0.143	0.902 ± 0.110	0.837 ± 0.152	199.74G	26.97M
LeViT-UNet (2021) [11]	0.972 ± 0.031	0.848 ± 0.226	0.817 ± 0.232	0.816 ± 0.224	0.734 ± 0.244	33.21G	52.14M
CMUNeXt (2023) [34]	0.982 ± 0.025	0.931 ± 0.083	0.905 ± 0.109	0.911 ± 0.076	0.846 ± 0.118	7.42G	3.15M
DCSAU-Net (2023) [35]	0.985 ± 0.018	0.935 ± 0.064	0.923 ± 0.087	0.925 ± 0.061	0.866 ± 0.095	6.92G	2.60M
GA-UNet (ours)	**0.987 ± 0.019**	**0.941 ± 0.066**	**0.933 ± 0.084**	**0.934 ± 0.062**	**0.882 ± 0.098**	**4.45G**	**2.18M**

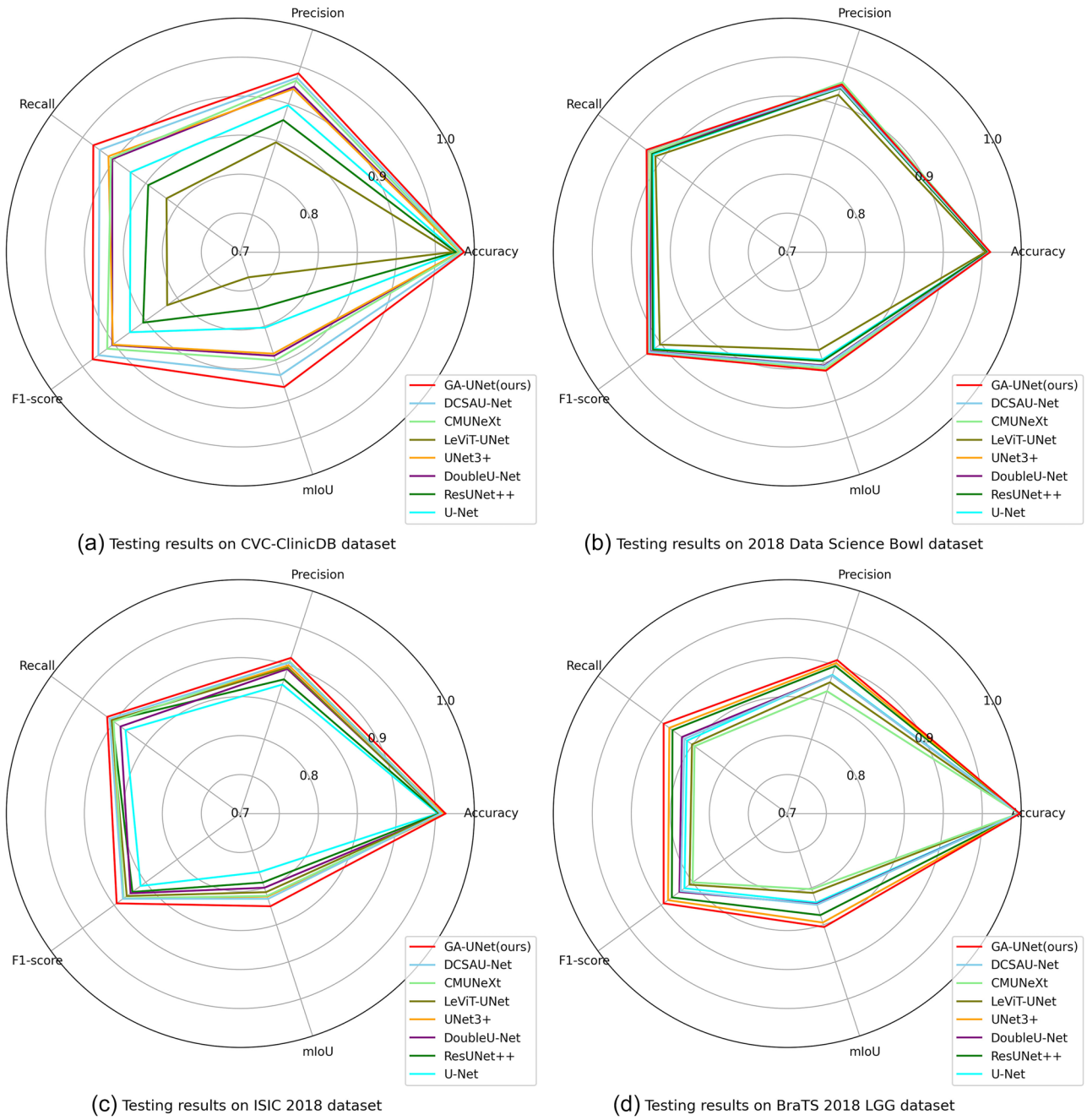

Fig. 8 Comparison of visualization performance across four datasets. GA-UNet achieves the highest performance in terms of Accuracy, F1-score, and mIoU

Effectiveness of GhostV2 Bottleneck and DSC Downsample Module

According to E_1 and E_2 in Table 6, the GhostV2 bottleneck and the DSC downsample module (our feature extraction module) increase the F1-score on the four datasets by 3.4%, 0.9%, 2.8%, and 2.7%, respectively, and mIoU by 5.2%, 1.1%, 3.5%, and 2.9%, respectively. The number of parameters and Flops decrease from 28.95M and 48.33G of U-Net to 2.01M and 4.44G, respectively.

In addition, we also utilize other feature extraction modules (Residual block [24] and Mobilenetv2 block [36]) to compare with our feature extraction module, as illustrated in Table 7. The experimental results on the CVC-ClinicDB dataset demonstrate that the performance of our feature extraction module in terms of mIoU has been improved by

Table 3 Results on the 2018 Data Science Bowl Dataset. The best results are in bold

Method	Accuracy	Precision	Recall	F1-score	mIoU	Flops	Parameters
U-Net (2015) [6]	0.957 ± 0.041	0.921 ± 0.094	0.913 ± 0.084	0.911 ± 0.081	0.845 ± 0.110	48.33G	28.95M
ResUNet++ (2019) [23]	0.957 ± 0.041	0.921 ± 0.091	0.915 ± 0.0781	0.913 ± 0.077	0.847 ± 0.106	70.99G	14.48M
DoubleU-Net (2020) [21]	0.959 ± 0.040	0.923 ± 0.081	0.917 ± 0.083	0.917 ± 0.071	0.853 ± 0.103	44.37G	18.84M
UNet3+ (2020) [26]	0.960 ± 0.039	0.927 ± 0.073	0.921 ± 0.070	**0.922 ± 0.059**	0.859 ± 0.089	199.74G	26.97M
LeViT-UNet (2021) [11]	0.954 ± 0.046	0.912 ± 0.111	0.909 ± 0.080	0.902 ± 0.097	0.832 ± 0.118	33.21G	52.14M
CMUNeXt (2023) [34]	0.960 ± 0.039	**0.929 ± 0.056**	0.917 ± 0.067	0.921 ± 0.051	0.857 ± 0.081	7.42G	3.15M
DCSAU-Net (2023) [35]	0.959 ± 0.040	0.922 ± 0.086	0.922 ± 0.067	0.918 ± 0.066	0.854 ± 0.095	6.92G	2.60M
GA-UNet (ours)	**0.960 ± 0.039**	0.926 ± 0.078	**0.923 ± 0.064**	0.922 ± 0.061	**0.860 ± 0.091**	**4.45G**	**2.18M**

4.9% and 1.6%, respectively, compared to the Residual block and Mobilenetv2 block. Additionally, our feature extraction module has significantly reduced the number of model parameters. Consequently, the GhostV2 bottleneck and the DSC downsample module not only improve performance but also reduce the number of parameters and Flops.

Effectiveness of CBAM

According to E_1 and E_3 in Table 6, CBAM improves the F1-score on the first three datasets by 4.4%, 0.5%, and 2.0%, respectively, and mIoU by 5.5%, 0.6%, and 2.3%, respectively. On the fourth dataset, E_1 and E_3 have similar performance. Overall, introducing CBAM in the first three decoders can improve the model's performance.

At the same time, by comparing the results of E_2 and E_4 as well as E_3 and E_4 in Table 6, it can be observed that our GA-UNet (E_4) can further improve the F1-score by 0.2–2.9% and mIoU by 0.4–3.3% compared to either E2 or E3. Additionally, the parameters and Flops of GA-UNet model are only 2.18M and 4.45G, respectively, which are much lower than those of U-Net.

Discussion

In order to evaluate GA-UNet more comprehensively, we conduct a statistical analysis. For the sake of discussion, we divide the seven SOTA models into two categories according to the number of parameters in the models: the non-lightweight models (U-Net, ResUNet++, DoubleU-Net, UNet3+ and LeViT-UNet) and the lightweight models (CMUNeXt and DCSAU-Net). We choose the best performing model based on the results presented in Tables 2, 3, 4, and 5 from each category, specifically UNet3+ from the non-lightweight models and DCSAU-Net from the lightweight models. Subsequently, pair t-tests are conducted to compare each selected model and GA-UNet. As shown in Table 8, most of the p-values are less than 0.05, which means the performance improvement of GA-UNet is statistically significant.

We also assess the convergence speed of our model and the other seven SOTA models within the first 20 epochs, as illustrated in Fig. 9. The experimental results indicate that GA-UNet exhibits faster convergence speed compared to the other SOTA models during the early stage. It suggests that GA-UNet has the potential to achieve superior performance with fewer epochs.

To further demonstrate the superiority of GA-UNet, we visualize segmentation results of challenging images from four datasets in Fig. 10. Qualitative analysis reveals that GA-UNet effectively captures information and produces smoother edges with reduced burrs, even in low-quality images. These edges are more consistent with the shape of lesions (indicated by red, blue, and green circles). In contrast to SOTA models, GA-UNet can better segment diseased cells instead of treating them as a whole entity (indicated by yellow circles). Moreover,

Table 4 Results on the ISIC-2018 Dataset. The best results are in bold

Method	Accuracy	Precision	Recall	F1-score	mIoU	Flops	Parameters
U-Net (2015) [6]	0.953 ± 0.072	0.874 ± 0.187	0.882 ± 0.172	0.858 ± 0.164	0.779 ± 0.192	48.33G	28.95M
ResUNet++ (2019) [23]	0.955 ± 0.069	0.881 ± 0.174	0.903 ± 0.144	0.871 ± 0.143	0.793 ± 0.175	70.99G	14.48M
DoubleU-Net (2020) [21]	0.958 ± 0.067	0.895 ± 0.164	0.890 ± 0.166	0.874 ± 0.152	0.800 ± 0.180	44.37G	18.84M
UNet3+ (2020) [26]	0.960 ± 0.062	0.900 ± 0.142	0.903 ± 0.131	0.886 ± 0.119	0.812 ± 0.156	199.74G	26.97M
LeViT-UNet (2021) [11]	0.956 ± 0.077	0.897 ± 0.154	0.904 ± 0.146	0.880 ± 0.138	0.806 ± 0.171	33.21G	52.14M
CMUNeXt (2023) [34]	0.958 ± 0.067	0.905 ± 0.148	0.901 ± 0.139	0.885 ± 0.126	0.811 ± 0.162	7.42G	3.15M
DCSAU-Net (2023) [35]	0.957 ± 0.082	0.904 ± 0.153	0.907 ± 0.135	0.886 ± 0.136	0.815 ± 0.167	6.92G	2.60M
GA-UNet (ours)	**0.963 ± 0.055**	**0.910 ± 0.142**	**0.911 ± 0.117**	**0.896 ± 0.108**	**0.825 ± 0.145**	**4.45G**	**2.18M**

Table 5 Results on the BraTS 2018 LGG Dataset. The best results are in bold

Method	Accuracy	Precision	Recall	F1-score	mIoU	Flops	Parameters
U-Net (2015) [6]	0.997 ± 0.003	0.887 ± 0.240	0.859 ± 0.263	0.863 ± 0.257	0.820 ± 0.255	48.33G	28.95M
ResUNet++ (2019) [23]	0.998 ± 0.002	0.899 ± 0.211	0.882 ± 0.226	0.883 ± 0.221	0.837 ± 0.226	70.99G	14.48M
DoubleU-Net (2020) [21]	0.997 ± 0.003	0.886 ± 0.227	0.867 ± 0.242	0.871 ± 0.233	0.822 ± 0.235	44.37G	18.84M
UNet3+ (2020) [26]	0.998 ± 0.002	0.903 ± 0.215	0.887 ± 0.227	0.889 ± 0.222	0.847 ± 0.225	199.74G	26.97M
LeViT-UNet (2021) [11]	0.997 ± 0.003	0.877 ± 0.244	0.851 ± 0.264	0.855 ± 0.255	0.807 ± 0.253	33.21G	52.14M
CMUNeXt (2023) [34]	0.997 ± 0.003	0.865 ± 0.261	0.847 ±0.271	0.850 ± 0.264	0.802 ± 0.259	7.42G	3.15M
DCSAU-Net (2023) [35]	0.997 ± 0.003	0.886 ± 0.237	0.863 ± 0.250	0.869 ± 0.242	0.823 ± 0.241	6.92G	2.60M
GA-UNet (ours)	**0.998 ± 0.002**	**0.907 ± 0.196**	**0.896 ± 0.213**	**0.896 ± 0.205**	**0.853 ± 0.211**	**4.45G**	**2.18M**

Table 6 Detailed ablation study of the GA-UNet architecture. The best results are in bold

Datasets	Method	Accuracy	Precision	Recall	F1-score	mIoU	Flops	Parameters
CVC-ClinicDB	E_1(U-Net [6])	0.977 ± 0.031	0.898 ± 0.152	0.874 ± 0.182	0.875 ± 0.158	0.802 ± 0.181	48.33G	28.95M
	E_2(U-Net+GhostV2 bottleneck and DSC downsample)	0.983 ± 0.027	0.926 ± 0.140	0.904 ± 0.161	0.909 ± 0.146	0.854 ± 0.163	**4.44G**	**2.01M**
	E_3(U-Net+CBAM)	0.982 ± 0.025	0.939 ± 0.052	0.909 ± 0.103	0.919 ± 0.066	0.857 ± 0.102	48.34G	29.12M
	E_4(GA-UNet (ours))	**0.987 ± 0.019**	**0.941 ± 0.066**	**0.933 ± 0.084**	**0.934 ± 0.062**	**0.882 ± 0.098**	4.45G	2.18M
2018 Data Science Bowl	E_1(U-Net [6])	0.957 ± 0.041	0.921 ± 0.094	0.913 ± 0.084	0.911 ± 0.081	0.845 ± 0.110	48.33G	28.95M
	E_2(U-Net+GhostV2 bottleneck and DSC downsample)	0.960 ± 0.039	**0.926 ± 0.065**	0.917 ± 0.065	0.920 ± 0.053	0.856 ± 0.083	**4.44G**	**2.01M**
	E_3(U-Net+CBAM)	0.958 ± 0.040	0.922 ± 0.079	0.916 ± 0.081	0.916 ± 0.070	0.851 ± 0.100	48.34G	29.12M
	E_4(GA-UNet (ours))	**0.960 ± 0.039**	0.926 ± 0.078	**0.923 ± 0.064**	**0.922 ± 0.061**	**0.860 ± 0.091**	4.45G	2.18M
ISIC-2018	E_1(U-Net [6])	0.953 ± 0.072	0.874 ± 0.187	0.882 ± 0.172	0.858 ± 0.164	0.779 ± 0.192	48.33G	28.95M
	E_2(U-Net+GhostV2 bottleneck and DSC downsample)	0.960 ± 0.062	0.908 ± 0.146	0.900 ± 0.135	0.886 ± 0.125	0.814 ± 0.160	**4.44G**	**2.01M**
	E_3(U-Net+CBAM)	0.959 ± 0.063	0.899 ± 0.153	0.892 ± 0.152	0.878 ± 0.135	0.802 ± 0.168	48.34G	29.12M
	E_4(GA-UNet (ours))	**0.963 ± 0.055**	**0.910 ± 0.142**	**0.911 ± 0.117**	**0.896 ± 0.108**	**0.825 ± 0.145**	4.45G	2.18M
BraTS 2018 LGG	E_1(U-Net [6])	0.997 ± 0.003	0.887 ± 0.240	0.859 ± 0.263	0.863 ± 0.257	0.820 ± 0.255	48.33G	28.95M
	E_2(U-Net+GhostV2 bottleneck and DSC downsample)	0.998 ± 0.002	0.905 ± 0.210	0.890 ± 0.225	0.890 ± 0.219	0.849 ± 0.224	**4.44G**	**2.01M**
	E_3(U-Net+CBAM)	0.997 ± 0.003	0.885 ± 0.231	0.866 ± 0.247	0.867 ± 0.239	0.820 ± 0.241	48.34G	29.12M
	E_4(GA-UNet (ours))	**0.998 ± 0.002**	**0.907 ± 0.196**	**0.896 ± 0.213**	**0.896 ± 0.205**	**0.853 ± 0.211**	4.45G	2.18M

Table 7 Ablation study of different feature extraction modules on the CVC-ClinicDB dataset. The best results are in bold

Method	Accuracy	Precision	Recall	F1-score	mIoU	Flops	Parameters
U-Net [6]	0.977 ± 0.031	0.898 ± 0.152	0.874 ± 0.182	0.875 ± 0.158	0.802 ± 0.181	48.33G	28.95M
U-Net + Residual block [23]	0.978 ± 0.031	0.907 ± 0.154	0.870 ± 0.211	0.870 ± 0.187	0.805 ± 0.211	51.09G	30.35M
U-Net + Mobilenetv2 block [36]	0.980 ± 0.027	0.920 ± 0.077	**0.904 ± 0.119**	0.906 ± 0.086	0.838 ± 0.127	20.14G	8.46M
U-Net + GhostV2 bottleneck and DSC downsample	**0.983 ± 0.027**	**0.926 ± 0.140**	0.904 ± 0.161	**0.909 ± 0.146**	**0.854 ± 0.163**	**4.44G**	**2.01M**

Table 8 The p-values on the four datasets. P_{UG}denotes the p-values of comparing UNet3+ with GA-UNet. P_{DG}denotes the p-values of comparing DCSAU-Net with GA-UNet. A p-value < 0.05 indicates the performance improvement is statistically significant

P-values \ Datasets	CVC-ClinicDB	2018 Data Science Bowl	ISIC-2018	BraTS 2018 LGG
P_{UG}	1.495e-3	8.075e-3	1.153e-2	3.598e-5
P_{DG}	2.160e-2	1.465e-1	3.440e-2	2.061e-2

for the isolated regions (pink and orange circles) that are difficult to detect in the image, GA-UNet makes a more complete delineation of lesion areas, which is crucial for medical diagnosis and treatment. These findings suggest that GA-UNet excels at capturing intricate details from medical images and represents an effective model for medical image segmentation.

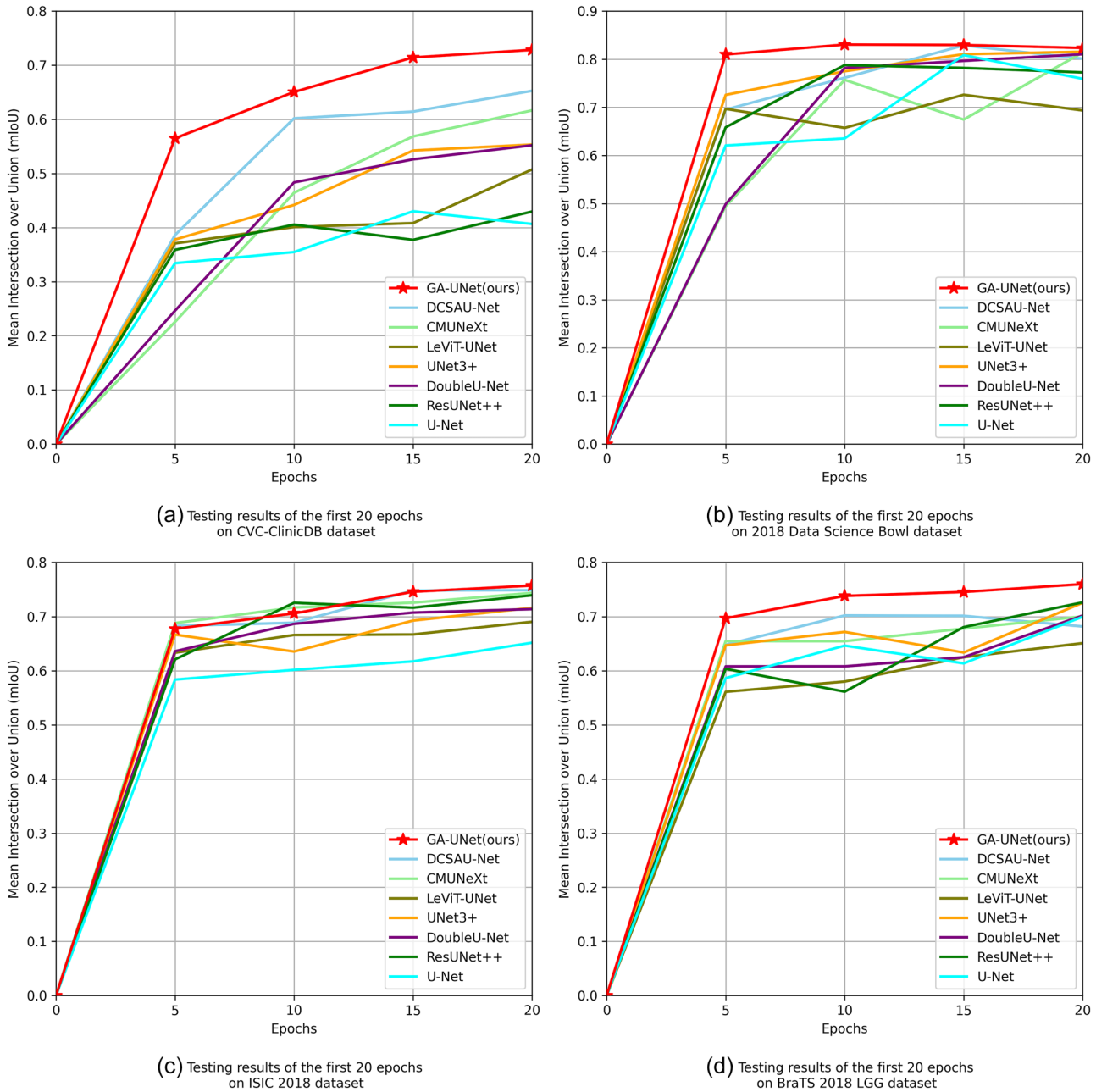

Fig. 9 Testing results of the first 20 epochs on four medical image segmentation tasks. Our GA-UNet exhibits the fastest convergence speed

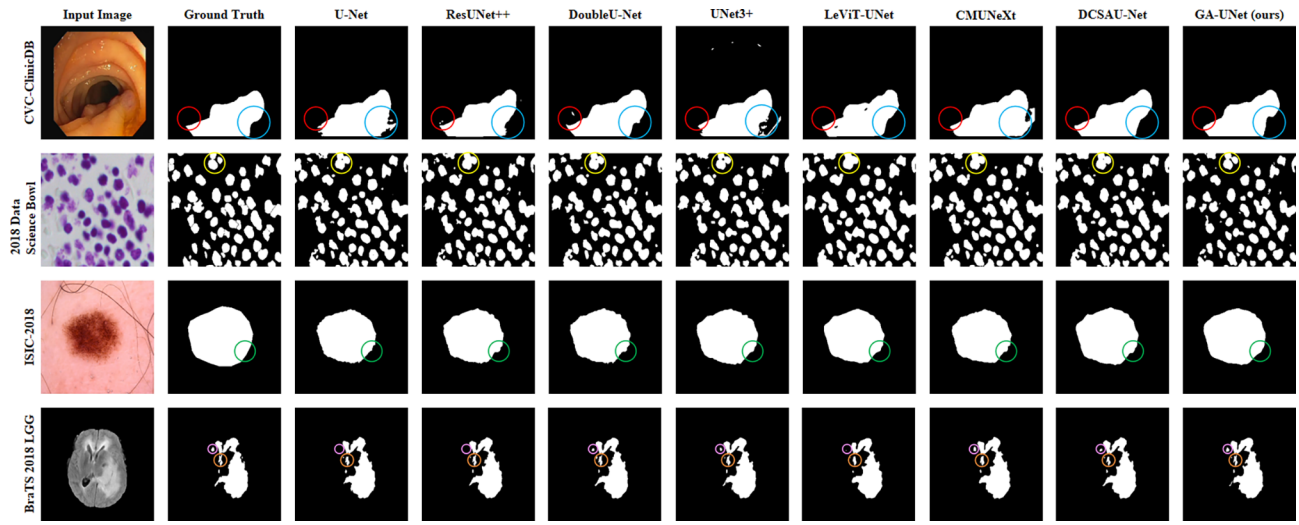

Fig. 10 Qualitative comparison results between GA-UNet and seven SOTA models on four medical segmentation datasets

However, there is still room to improve our model. Red circles in Fig. 11 show that our GA-UNet model has limitations in segmenting similar foreground and background, delineating complex edges, and identifying tiny lesion regions. Addressing these challenges will be a key focus of our future research.

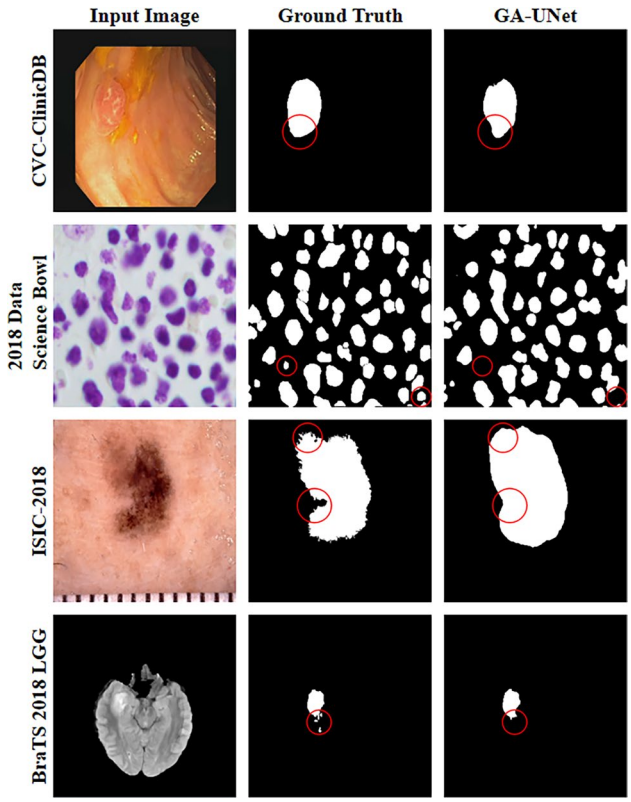

Fig. 11 Example of GA-UNet's failed segmentation on four datasets

Conclusion

This work proposes GA-UNet, a novel lightweight encoder-decoder model for medical image segmentation which consists mainly of the GhostV2 bottleneck, the DSC downsample module and CBAM. To assess the performance, we evaluate our model on four medical image segmentation datasets (i.e., Polyp, Nuclei, Cellular lesions and Brain tumors). The experimental results show that our model is better than other seven SOTA models in terms of F1-score and mIoU. It is worth mentioning that our GA-UNet has much fewer parameters and Flops. The segmentation results of GA-UNet are more consistent with ground truth than other SOTA models, demonstrating its potential as an aid to diagnostic tools. In the future, our work will focus on improving the model's ability to segment the challenging regions mentioned above and moving towards a lightweight architecture for 3D medical image segmentation.

Author Contributions All authors contributed to the study conception and design.

Funding The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Data Availability The code that support our study will be available from the corresponding author upon reasonable request.

Declarations

Ethics Approval Not applicable.

Consent to Participate Not applicable.

Consent to Publish Not applicable.

Conflict of Interest The authors declare no competing interests.

References

1. Guan H and Liu M. Domain adaptation for medical image analysis: a survey. *IEEE Transactions on Biomedical Engineering*, 69(3):1173–1185, 2021. https://doi.org/10.1109/TBME.2021.3117407
2. Yanase J and Triantaphyllou E. A systematic survey of computer-aided diagnosis in medicine: Past and present developments. *Expert Systems with Applications*, 138:112821, 2019. https://doi.org/10.1016/j.eswa.2019.112821
3. Canny J. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, (6):679–698, 1986. https://doi.org/10.1109/TPAMI.1986.4767851
4. Otsu N. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1):62–66, 1979. https://doi.org/10.1109/TSMC.1979.4310076
5. Ramesh K, Kumar G, Swapna K, Datta D, and Rajest S. A review of medical image segmentation algorithms. *EAI Endorsed Transactions on Pervasive Health and Technology*, 7(27):e6–e6, 2021. https://doi.org/10.4108/eai.12-4-2021.169184
6. Ronneberger O, Fischer P, and Brox T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015. https://doi.org/10.1007/978-3-319-24574-4_28.
7. Oktay O, Schlemper J, Folgoc L, Lee M, Heinrich M, Misawa K, Mori K, McDonagh S, Hammerla N, Kainz B, Glocker B, and Rueckert D. Attention U-Net: Learning where to look for the pancreas. 04 2018. https://doi.org/10.48550/arXiv.1804.03999.
8. Çiçek Ö, Abdulkadir A, Lienkamp S, Brox T, and Ronneberger O. 3d U-Net: Learning dense volumetric segmentation from sparse annotation. 2016. https://doi.org/10.48550/arXiv.1606.06650.
9. Milletari F, Navab N, and Ahmadi S. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. 06 2016. https://doi.org/10.48550/arXiv.1606.04797.
10. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint https://doi.org/10.48550/arXiv.2010.11929, 2020.
11. Xu G, Zhang X, Fang Y, Cao X, Liao W, He X, and Wu X. LeViT-UNet: Make faster encoders with transformer for biomedical image segmentation. https://doi.org/10.48550/arXiv.2107.08623.
12. Graham B, El-Nouby A, Touvron H, Stock P, Joulin A, Jégou H, and Douze M. LeViT: a vision transformer in Convet's clothing for faster inference. pages 12239–12249, 2021. https://doi.org/10.1109/ICCV48922.2021.01204.
13. Tang Y, Han K, Guo J, Xu C, Xu C, and Wang Y. Ghostnetv2: enhance cheap operation with long-range attention. *Advances in Neural Information Processing Systems*, 35:9969–9982, 2022. https://doi.org/10.48550/arXiv.2211.12905.
14. Howard A, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, and Adam H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint https://doi.org/10.48550/arXiv.1704.04861, 2017.
15. Woo S, Park J, Lee J, and Kweon I. Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)*, pages 3–19, 2018. https://doi.org/10.48550/arXiv.1807.06521.
16. Bernal J, Sánchez F, Fernández-Esparrach G, Gil D, Rodríguez C, and Vilariño F. WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. *Computerized medical imaging and graphics*, 43:99–111, 2015. https://doi.org/10.1016/j.compmedimag.2015.02.007
17. Caicedo J, Goodman A, Karhohs K, Cimini B, Ackerman J, Haghighi M, Heng C, Becker T, Doan M, McQuin C, et al. Nucleus segmentation across imaging experiments: the 2018 data science bowl. *Nature methods*, 16(12):1247–1253, 2019. https://doi.org/10.1038/s41592-019-0612-7
18. Codella N, Gutman D, Celebi M, Helba B, Marchetti M, Dusza S, Kalloo A, Liopyris K, Mishra N, Kittler H, et al. Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC). In *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*, pages 168–172. IEEE, 2018. https://doi.org/10.1109/ISBI.2018.8363547
19. Tschandl P, Rosendahl C, and Kittler H. The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific Data*, 5(1):1–9, 2018. https://doi.org/10.1038/sdata.2018.161
20. Bakas S, Reyes M, Jakab A, Bauer S, Rempfler M, Crimi A, Shinohara R, Berger C, Ha S, Rozycki M, Prastawa M, Alberts E, Lipkova J, Freymann J, Kirby J, Bilello M, Fathallah-Shaykh H, Wiest R, Kirschke J, and Menze B. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge. 11 2018. https://doi.org/10.48550/arXiv.1811.02629
21. Jha D, Riegler M, Johansen D, Halvorsen P, and Johansen H. Doubleu-net: A deep convolutional neural network for medical image segmentation. In *2020 IEEE 33rd International symposium on computer-based medical systems (CBMS)*, pages 558–564. IEEE, 2020. https://doi.org/10.1109/CBMS49503.2020.00111.
22. He K, Zhang X, Ren S, and Sun J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE transactions on pattern analysis and machine intelligence*, 37(9):1904–1916, 2015. https://doi.org/10.1109/TPAMI.2015.2389824
23. Jha D, Smedsrud P, Riegler M, Johansen D, De Lange T, Halvorsen P, and Johansen H. Resunet++: An advanced architecture for medical image segmentation. In *2019 IEEE international symposium on multimedia (ISM)*, pages 225–2255. IEEE, 2019. https://doi.org/10.1109/ISM46123.2019.00049
24. He K, Zhang X, Ren S, and Sun J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. https://doi.org/10.48550/arXiv.1512.03385
25. Hu J, Shen L, and Sun G. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018. https://doi.org/10.48550/arXiv.1709.01507
26. Huang H, Lin L, Tong R, Hu H, Zhang Q, Iwamoto Y, Han X, Chen Y, and Wu J. Unet 3+: A full-scale connected unet for medical image segmentation. pages 1055–1059, 2020. https://doi.org/10.1109/ICASSP40776.2020.9053405.
27. Lama N, Hagerty J, Nambisan A, Stanley R, and Van Stoecker W. Skin lesion segmentation in dermoscopic images with noisy data. *Journal of Digital Imaging*, pages 1–11, 2023. https://doi.org/10.1007/s10278-023-00819-8.
28. Tan M and Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks. 05 2019. https://doi.org/10.48550/arXiv.1905.11946.
29. Chaurasia A and Culurciello E. Linknet: Exploiting encoder representations for efficient semantic segmentation. pages 1–4, 2017. https://doi.org/10.48550/arXiv.1707.03718.
30. Singh Samant S, Chauhan A, Dn J, and Singh V. Glomerulus detection using segmentation neural networks. *Journal of Digital Imaging*, pages 1–10, 2023. https://doi.org/10.1007/s10278-022-00764-y.

31. Saumiya S and Franklin S. Residual deformable split channel and spatial u-net for automated liver and liver tumour segmentation. *Journal of Digital Imaging*, 36(5):2164–2178, 2023. https://doi.org/10.1007/s10278-023-00874-1
32. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A, Kaiser Ł, and Polosukhin I. Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 2017. https://doi.org/10.48550/arXiv.1706.03762.
33. Feng X, Ghimire K, Kim D, Chandra R, Zhang H, Peng J, Han B, Huang G, Chen Q, Patel S, Bettagowda C, Sair H, Jones C, Jiao Z, Yang l, and Bai H. Brain tumor segmentation for multi-modal mri with missing information. *Journal of Digital Imaging*, 36, 06 2023. https://doi.org/10.1007/s10278-023-00860-7.
34. Tang F, Ding J, Wang L, Ning C, and Zhou S. Cmunext: An efficient medical image segmentation network based on large kernel and skip fusion. *ArXiv*, abs/2308.01239, 2023. https://doi.org/10.48550/arXiv.2308.01239.
35. Xu Q, Ma Z, Na H, and Duan W. Dcsau-net: A deeper and more compact split-attention u-net for medical image segmentation. *Computers in Biology and Medicine*, 154:106626, 2023. https://doi.org/10.48550/arXiv.2202.00972.
36. Sandler M, Howard A, Zhu M, Zhmoginov A, and Chen L. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4510–4520, 2018. https://doi.org/10.48550/arXiv.1801.04381.
37. Han K, Wang Y, Tian Q, Guo J, Xu C, and Xu C. Ghostnet: More features from cheap operations. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1580–1589, 2020. https://doi.org/10.48550/arXiv.1911.11907.
38. Jaderberg M, Simonyan K, Zisserman A, et al. Spatial transformer networks. *Advances in neural information processing systems*, 28, 2015. https://doi.org/10.48550/arXiv.1506.02025.
39. Li X, Sun X, Meng Y, Liang J, Wu F, and Li J. Dice loss for data-imbalanced nlp tasks. arXiv preprint https://doi.org/10.48550/arXiv.1911.02855, 2019.
40. Kingma D and Ba J. Adam: A method for stochastic optimization. arXiv preprint https://doi.org/10.48550/arXiv.1412.6980, 2014.
41. Loshchilov I and Hutter F. Sgdr: Stochastic gradient descent with warm restarts. arXiv preprint https://doi.org/10.48550/arXiv.1608.03983, 2016.