Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Attention based multi-scale nested network for biomedical image segmentation

Dapeng Cheng [a,b,*], Jia Deng [a], Jinjie Xiao [a], Mao Yanyan [a], Jialong Kang [c], Jiale Gai [a], Baosheng Zhang [a], Feng Zhao [a,b]

[a] *School of Computer Science and Technology, Shandong Business and Technology University, No. 191 Binhai Middle Road, Yantai, 264000, Shandong Province, China*

[b] *Shandong Co-Innovation Center of Future Intelligent Computing, No. 191 Binhai Middle Road, Yantai, 264000, Shandong Province, China*

[c] *School of Information and Electronic Engineering, Shandong Business and Technology University, No. 191 Binhai Middle Road, Yantai, 264000, Shandong Province, China*

ARTICLE INFO

Dataset link: https://doi.org/10.1016/j.compmedimag.2015.02.007

Dataset link: https://doi.org/10.1016/j.patcog.2012.03.002

Dataset link: https://doi.org/10.48550/arXiv.1902.03368

Dataset link: https://doi.org/10.1038/sdata.2018.161

Dataset link: https://doi.org/10.1016/j.dib.2019.104863

Dataset link: https://doi.org/10.1016/j.media.2016.08.008

Keywords:
Convolutional neural network
Medical image segmentation
ReSidual U-CBAM module
CBAM

ABSTRACT

Convolutional neural network-based methods have significantly enhanced the segmentation performance of biomedical images in recent years. Nevertheless, medical image segmentation presents a challenge marked by layout specificity, with limited variation between samples in medical datasets but significant variation within each individual sample. This aspect has been often overlooked by many models. Consequently, we propose a novel architecture called Attention based multi-scale nested network (AMNNet), specifically designed for efficient biomedical image segmentation. AMNNet comprises four components: early ReSidual U-CBAM (RSUC) modules and convolutional stages, a MLP stage in latent stage, and Convolutional Block Attention Modules (CBAM) integrated into the decoder stage. We introduce a lightweight CBAM to concentrate on regions proximate to the target and suppress extraneous features without substantial parameter increments. The RSUC module is proposed to combine receptive fields of different sizes, capturing comprehensive contextual information across various scales in medical samples. Extensive experiments conducted on the AMNNet reveal its outperformance compared to prevailing medical image segmentation methods across the ISIC2018, CVC-ClinicDB, CVC-ColonDB, BUSI, and GlaS datasets. Notably, AMNNet achieves Dice Similarity Coefficients (DSC) of 91.35%, 90.01%, 90.80%, 81.61%, and 94.31%, respectively.

1. Introduction

Presently, cancer remains one of the most formidable threats to human health [1]. Its incidence and mortality rates consistently ranking among the highest of all diseases. Medical image segmentation plays a pivotal role in diverse clinical cancer diagnoses and has

[*] Corresponding author at: School of Computer Science and Technology, Shandong Business and Technology University, No.191 Binhai Middle Road, Yantai, 264000, Shandong Province, China.

E-mail address: chengdapeng@sdtbu.edu.cn (D. Cheng).

https://doi.org/10.1016/j.heliyon.2024.e33892
Received 17 November 2023; Received in revised form 27 June 2024; Accepted 28 June 2024

Available online 2 July 2024

garnered considerable research focus within the domain of medical image analysis [2–4]. Semantic segmentation outcomes aid in the detection of lesions within critical areas, including colorectal polyp examination, melanoma screening, colorectal adenocarcinoma, and breast cancer diagnosis. Consequently, these segmentation results enable doctors to accurately pinpoint the precise locations of cancerous lesions. This is a critical factor in the planning and guidance of surgical procedures, radiotherapy, and other therapeutic interventions. Achieving this necessitates access to an extensive repository of high-quality medical image datasets.

The manual annotation of medical image data is a laborious process. It necessitates collaboration with highly specialized medical experts and incurring substantial costs. An alternative to manual image segmentation is the creation of diagnostic support systems. These systems rely on automated computer-assisted segmentation, providing swifter, more precise, and more dependable solutions. This revolutionizes clinical procedures and elevates patient care. Computer-assisted diagnosis will relieve the burden on experts and lead to a reduction in overall treatment expenses.

Furthermore, the diversity and intricacy of medical image data require medical image segmentation networks to exhibit a requisite level of robustness [5]. With the widespread adoption of deep Convolutional Neural Networks (CNNs) in the field of computer vision, CNNs have quickly been utilized in medical image segmentation. This has significantly improved the efficiency of medical image segmentation. Additionally, CNN-based methodologies have surmounted the constraints associated with conventional segmentation techniques [6] across a range of medical imaging modalities, such as dermoscopy, endoscopy, ultrasound imaging, stained histopathological slides, computed tomography (CT), and magnetic resonance imaging (MRI). UNet [7], 3D U-Net [8], ResUNet [9], Dense-UNet [10], U-Net++ [11], and U-Net3+ [12] have been explicitly utilized for image segmentation across diverse medical imaging modalities. These networks have yielded commendable segmentation outcomes on numerous demanding datasets, underscoring the efficacy of CNNs in acquiring discriminative features for organ and lesion segmentation in medical scans.

CNNs constitute a foundational element in the majority of contemporary image segmentation techniques. Nevertheless, within CNNs, individual convolutional kernels exclusively attend to local information across the entire image. Their primary design aims at local feature extraction at each layer, rendering them less proficient in yielding superior outcomes for regions with diverse shapes, intricate structures, and blurred boundaries.

Additionally, when tasked with segmenting a multitude of organs, tissues, vessels, tumors, and other structures in medical images, the presence of noise and low contrast in certain images frequently leads to diminished object-background distinguishability. Achieving precise segmentation of distinct organs, tissues, vessels, and tumors through conventional deep learning-based segmentation networks can pose challenges, especially in intricate medical image contexts.

Consequently, the task of medical image segmentation has consistently presented significant challenges. To tackle the mentioned challenges, we have designed an innovative medical image segmentation architecture known as AMNNet, meticulously crafted to enhance segmentation performance.

This paper's primary contributions can be summarized as follows:

(1)We introduced CBAM [13], an attention mechanism module that fuses channel and spatial information. CBAM enhances feature representation, suppresses noise, adapts to diverse image conditions, improves segmentation accuracy, and mitigates overfitting concerns, thereby facilitating more precise identification and segmentation of medical structures and lesions.

(2)To capture comprehensive global contextual information across the stages, we devised the RSUC module and integrated it into the initial stages of the encoder. This is realized through a nested U-like architecture. At the lower level, we introduced the innovative RSUC module for extracting multiscale features within the stages. Meanwhile, at the top level, a structure resembling UNet is employed, with each early encoder stage housing an RSUC block. This dual-level configuration creates a nested U-like structure for the AMNNet network.

(3)We have validated the effectiveness of AMNNet using five publicly available datasets: GlaS [14], CVC-ClinicDB [15], CVC-ColonDB [16], BUSI [17], and ISIC2018 [18,19]. The experimental results, along with a comparative analysis against existing computer vision methods, establish the superior performance and broader applicability of our approach.

2. Related work

2.1. Medical image segmentation

Long et al. [20] pioneered a fully convolutional network (FCN) comprising exclusively of convolutional layers tailored for semantic segmentation. Following this, Ronneberger et al. [7] enhanced the FCN by employing an encoder-decoder architecture, UNet, to segment neural structures within Hela cells and electron microscopy stacks. In UNet [7], low-level and high-level feature maps are merged via skip connections. The high-level feature maps traverse deeper layers of the encoder network and continue through the decoder, while the low-level features originate from the initial network layers. This configuration can result in a semantic disparity between high-level and low-level features. Ibtehaz et al. [21] subsequently extended UNet and introduced the MultiResUNet architecture, which incorporates convolutional layers with residual connections within the skip connections to mitigate disparities between encoder-decoder features. Valanarasu et al. [22] introduced a lightweight segmentation model known as UNext, constructed through the fusion of MLP [23] and UNet. UNext stands as the pioneering model that employs MLP-based convolutional neural networks for image segmentation. It substantially diminishes the parameter count and inference time, all the while upholding segmentation performance.

2.2. Attention mechanism

Empirical evidence suggests that transformer-based models typically excel when trained on extensive datasets [24]. In contrast to datasets employed in visual applications, medical imaging data typically features a relatively smaller sample size. Valanarasu et al. [25] introduced a gated position-sensitive axial attention mechanism that integrates four gates to regulate the extent of positional embedding assigned to keys, queries, and values. These gates are adaptable learnable parameters, rendering the proposed mechanism viable for datasets of varying sizes.

Oktay et al. [26] introduced an attention U-Net structure that employs attention mechanisms for pancreas segmentation. Attention blocks are utilized to adapt feature maps in skip connections. Xiao et al. [27] developed a convolutional network with a weighted attention mechanism and integrated skip connections for high-resolution retinal vessel segmentation. Guo et al. [28] introduced a lightweight SA-UNet that excels on small datasets and effectively mitigates overfitting using the DropBlock mechanism. Hu et al. [29] created SeNet, utilizing the network's attention mechanism to emphasize channel relationships, enabling automatic learning of the significance of different channel features. A limitation is its focus solely on feature channel relationships without incorporating contextual information linkage. Woo et al. [13] expanded upon SeNet's concept and introduced CBAM, extending SeNet's channel attention module (CAM) with a spatial attention module (SAM). CBAM encompasses both CAM and SAM sub-modules, catering to channel and spatial attention, respectively. CAM distinguishes itself from SeNet by incorporating a parallel max-pooling layer, enhancing comprehensive high-level feature extraction.

2.3. Multi-scale fusion

Successful medical image segmentation necessitates both local and global information integration. A 3x3 filter is adept at capturing local features within each layer. Nonetheless, enlarging convolutional kernel sizes to encompass global information would substantially augment the network's parameter count. Consequently, numerous studies have concentrated on extracting global features. Cheng et al. [30] used Channel Attention Module to integrate features of different stages at channel level. Zhao et al. [31] utilized pyramid pooling to extract global contextual information. Chen et al. [32] applied Atrous Spatial Pyramid Pooling (ASPP) with varying sampling rates and multiscale dilated convolutions to achieve multiscale fusion. To enhance the handling of decreased model detection performance attributed to small target sizes, Wu et al. [33] reconceptualized infrared small target detection. They reframed it as a semantic segmentation challenge rather than a typical object detection issue, thereby introducing UIU-Net. UIU-Net introduced Resolution-Maintaining Deep Supervision (RM-DS) networks [34] to acquire profound multiscale features and elevate global contextual representation.

Inspired by the above works, we design a new network model, AMNNet, for medical image segmentation. We introduced CBAM into AMNNet to better focus on the edge features of cancer sites or organ tissues in medical images. In addition, we also design the RSUC module and embed it into the network to better extract the global context information through multi-scale feature fusion, and achieve improved performance.

3. Proposed method

This section provides a detailed exposition of the AMNNet network model. We thoroughly expound upon its primary components, namely CBAM, the RSUC module, the tokenized MLP module, and the convolution module. Towards the conclusion of this section, we present the loss function employed in our approach.

3.1. The AMNNet architecture

AMNNet retains the foundational five-layer deep encoder-decoder structure of UNet [7], preserving the skip connections. Yet, we have reconfigured the design of each module, as depicted in Fig. 1. To enhance information acquisition, we introduce CBAM into the decoder stage, facilitating the model's focus on pertinent data while discarding irrelevant information. In the initial encoder stages, we incorporate RSUC modules with varying depths, facilitating the capture of multi-scale features at different image processing stages. Additionally, we integrate CBAM modules within the RSUC modules, further amplifying the model's performance. Moreover, we employ tokenized MLP modules in the latent stage of AMNNet and implement convolution blocks with reduced filters in the network's final block (The number of filters for the convolution phase encoder that we use in the l1, l2, and l3 stages of the final block in AMNNet is (16,32,128) and the number of filters for the decoder is (16,16,32)), achieving the objective of sustaining superior segmentation performance while managing parameter growth.

3.2. Convolutional block attention module

Within the realm of medical imaging, the target structures requiring segmentation encompass a wide range, comprising organs, tissues, blood vessels, tumors, and beyond. These structures often manifest intricate shapes, sizes, and textural variations, thus intensifying the complexity of image segmentation.

We present CBAM as a solution to tackle the aforementioned challenges. The attention mechanism operates by adaptively assigning weights and filtering information, enabling the extraction of valuable insights from a vast feature pool to enhance network training, and subsequently passing this enriched information to the convolution stack. The CBAM stands as a straightforward yet

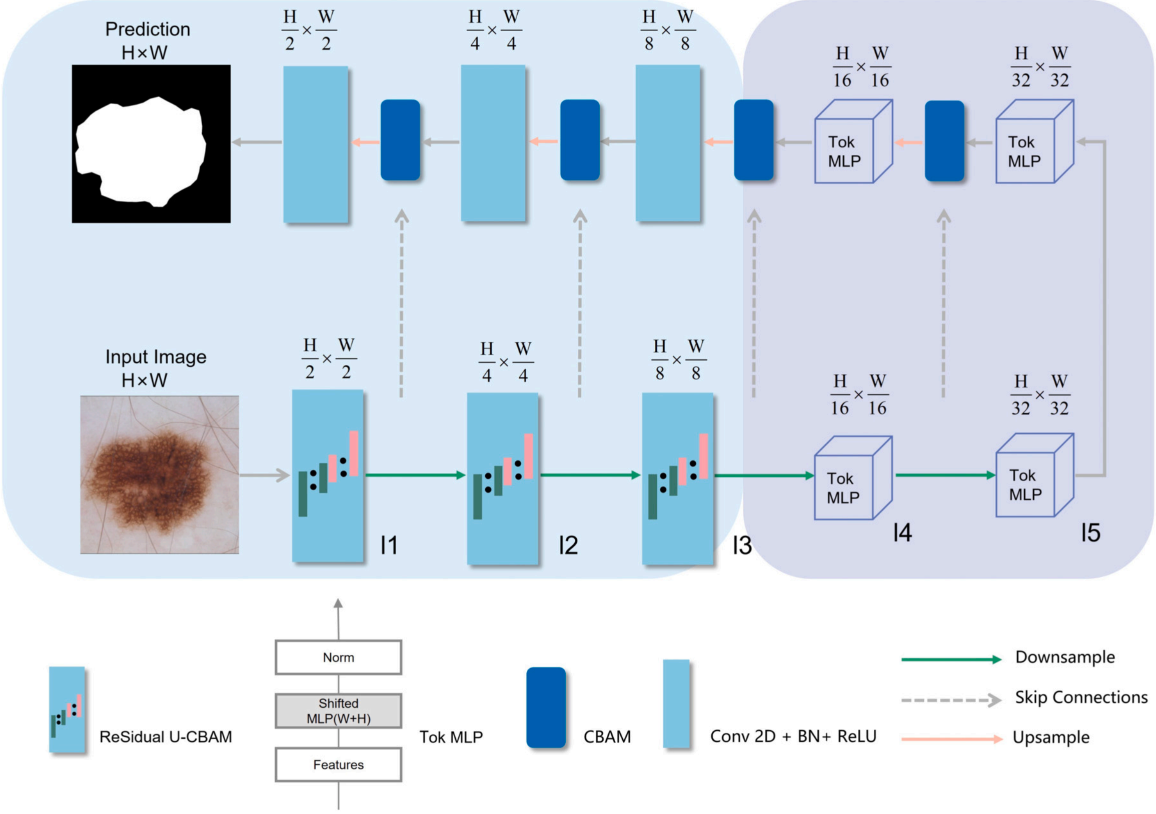
Prediction
H×W
$\frac{H}{2}\times\frac{W}{2}$
$\frac{H}{4}\times\frac{W}{4}$
$\frac{H}{8}\times\frac{W}{8}$
$\frac{H}{16}\times\frac{W}{16}$
$\frac{H}{32}\times\frac{W}{32}$
Tok MLP
Input Image
H×W
l1
l2
l3
l4
l5
Norm
Shifted MLP(W+H)
Features
ReSidual U-CBAM
Tok MLP
CBAM
Conv 2D + BN+ ReLU
Downsample
Skip Connections
Upsample

Fig. 1. Illustration of our proposed AMNNet architecture.

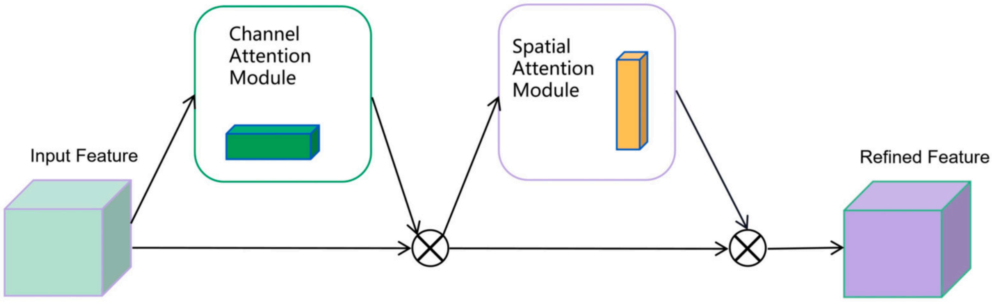

Fig. 2. Convolutional block attention module.

highly efficient feedforward convolutional neural network attention module. CBAM consists of two distinct sub-modules: the channel attention module and the spatial attention module. As a lightweight and adaptable module, CBAM can be seamlessly integrated into any CNN architecture with minimal computational overhead.

When provided with an intermediate feature map as input, the initial step involves the calculation of a one-dimensional channel attention map, which is then multiplied with the said intermediate feature map. Subsequently, the 2D spatial attention map is computed and subjected to multiplication with the feature map from the preceding layer. In the process of multiplication, attention values are duplicated as necessary; channel attention values are disseminated across the spatial dimension, and vice versa. Fig. 2 illustrates the CBAM's construction. The calculation method of CBAM is shown in formulas (1) and (2) below:

$$F' = M_c(F) \otimes F, \tag{1}$$

$$F'' = M_s\left(F'\right) \otimes F', \tag{2}$$

In the above equations, $F \in R^{C\times H\times W}$ represents the intermediate feature map, $M_c \in R^{C\times 1\times 1}$ is the one-dimensional channel attention map, $M_s \in R^{1\times H\times W}$ is the two-dimensional spatial attention map, $\otimes$ denotes element-wise multiplication, and F'' is the refined feature map.

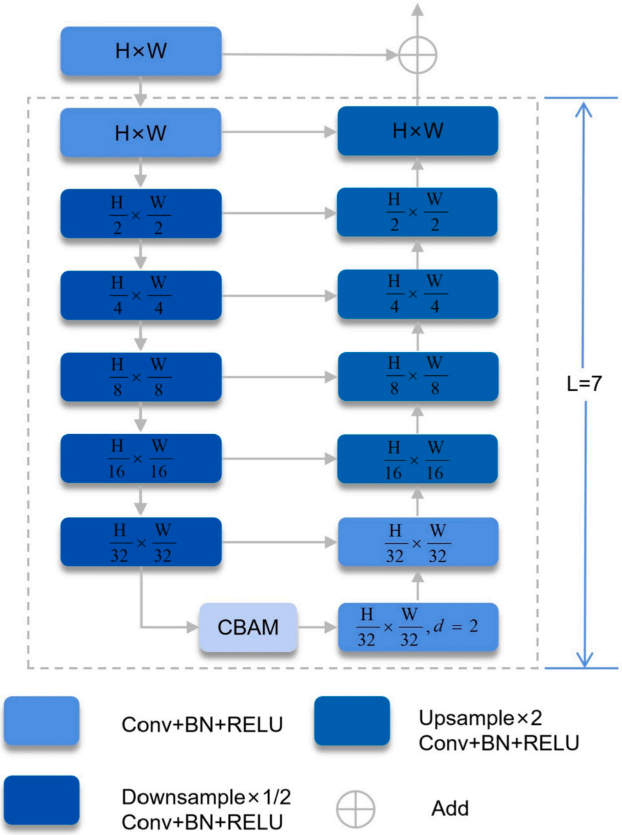

Fig. 3. ReSidual U-CBAM module.

3.3. *ReSidual U-CBAM module*

In CNN design, such as VGG [35], ResNet [36], and DenseNet [37], etc., small convolution filters with a size of 1×1 or 3×3 are the most commonly used components for feature extraction. They require less storage space and are computationally efficient. Nonetheless, 1x1 or 3x3 filters possess a restricted receptive field and cannot effectively encompass global information. This limitation results in shallow-level output features that primarily capture local characteristics. Drawing inspiration from U2-Net and UIU-Net in the realm of object detection, we have innovated the Residual U-CBAM module, denoted as RSUC, to capture multiscale features within each stage, thus augmenting the segmentation performance. Our RSUC mainly consists of three components. The structure of RSUC is visually depicted in Fig. 3.

Intermediate feature alignment part: An initial convolutional layer that aligns the channel count of the input feature map with that of the intermediate feature map. This layer operates using the intermediate feature map as its input for the purpose of learning and encoding deep multiscale features, as opposed to using the input image. This convolutional layer adheres to standard design principles for local feature extraction.

Layer variations and enhanced feature integration part: The architectural framework closely resembles UNet, but in AMNNet's l1, l2, and l3, we introduce variations in the number of layers within the RSUC blocks of the encoder-decoder, and we also incorporate CBAM. With increasing network depth, the RSUC block incorporates a higher number of pooling and upsampling operations, which leads to an expanded receptive field and a more extensive array of local and global features. Thus, in cases of feature maps with larger dimensions, we deploy deeper networks to capture larger-scale information. In l1, we employ a 7-layer encoder-decoder. Conversely, in l2 and l3, where the feature map resolutions decrease, we opt for shallower networks to safeguard against the loss of valuable information, utilizing a 6-layer encoder in l2 and 5 layers in l3. Within each RSUC block, CBAM is introduced following the final downsampling phase, enabling the effective capture of Region of Interest (ROI) features within medical images while suppressing non-ROI features.

Subsequent to CBAM application, a dilated convolution layer is employed to enlarge the receptive field and acquire a greater amount of contextual information. In order to illustrate the influence of attention placement on performance, we performed four sets of comparative experiments on both the ISIC2018 dataset and the CVC-ClinicDB dataset. The outcomes of these experiments are presented in Table 1 and Table 2, respectively.

Local and multiscale feature fusion part: Ultimately, we integrate the local features with the multiscale feature residuals. The operations in RSUC can be summarized as shown in formula (3) below:

Table 1
Comparison results of 5 different models on ISIC2018 dataset. We designate the complete network formed by removing CBAM between the encoder and decoder in the early stages of the AMNNet encoder as AMNNet-C1. The full network, where CBAM is excluded from the early stages of the AMNNet encoder between the encoder and decoder but added to both the encoder and decoder, is denoted as AMNNet-C1+C2. The comprehensive network, where CBAM is omitted from the early stages of the AMNNet encoder between the encoder and decoder but added to the encoder, is labeled as AMNNet-C1+C3. The overall network, created by removing CBAM between the encoder and decoder in the early stages of the AMNNet encoder and introducing it to the decoder, is identified as AMNNet-C1+C4.

Model	Encoder CBAM	Encoder-CBAM-Decoder	Decoder CBAM	DSC	IoU	TPR	PPV	Parameters
AMNNet-C1				90.89	83.78	89.75	92.51	1.64M
AMNNet-C1+C2	√		√	91.00	83.90	90.51	91.98	1.64M
AMNNet-C1+C3	√			90.92	83.70	90.32	92.00	1.64M
AMNNet-C1+C4			√	90.83	83.54	89.69	**92.59**	1.64M
AMNNet(Ours)		√		**91.35**	**84.36**	**90.96**	92.14	1.64M

Table 2
Comparison results of 5 different models on CVC-ClinicDB dataset.

Model	Encoder CBAM	Encoder-CBAM-Decoder	Decoder CBAM	DSC	IoU	TPR	PPV	Parameters
AMNNet-C1				88.79	80.21	88.63	89.31	1.64M
AMNNet-C1+C2	√		√	89.16	80.86	87.48	**91.39**	1.64M
AMNNet-C1+C3	√			86.83	77.37	87.35	87.08	1.64M
AMNNet-C1+C4			√	88.45	79.82	86.60	90.99	1.64M
AMNNet(Ours)		√		**90.01**	**82.23**	**89.18**	91.17	1.64M

$$F_{RSUC}(x) = U(f(x)) + f(x), \tag{3}$$

where $f(x)$ represents the input intermediate feature map, and U stands for the RSUC block that resembles a UNet.

3.4. Tokenized MLP module and convolutional module

Within the tokenized MLP module, the initial step involves feature relocation and projection into tokens. Subsequently, these tokens are directed to the shifted MLP (along the width), with the hidden dimension of the MLP determined by the hyperparameter H. Lastly, layer normalization (LN) is applied, and the resulting features are passed to the next module. In the final module of the AMNNet network, convolutional modules with a reduced number of filters are utilized. Each convolutional module comprises a convolutional layer, a batch normalization layer, and a ReLU activation layer.

3.5. Dice loss

The Dice Coefficient loss (DL) addresses the issue of imbalance between background and foreground pixels by adjusting the segmentation evaluation metric, the Dice Similarity Coefficient (DSC), between the predicted samples and the ground truth annotations. It has demonstrated superior performance in segmentation tasks. The formula (4) is shown as follows:

$$DL(p,g) = 1 - \frac{2\sum_{i=1}^{N} p_i g_i + \alpha}{\sum_{i=1}^{N} p_i^2 + \sum_{i=1}^{N} g_i^2 + \alpha}, \tag{4}$$

Here, α, which falls within the range [0, 1], represents a tunable parameter crafted to prevent zero errors and enhance gradient propagation for negative samples [38].

4. Experiments and results

4.1. Dataset

To assess the efficacy of AMNNet, we utilized five publicly available biomedical imaging datasets: GlaS, CVC-ClinicDB, CVC-ColonDB, BUSI, and ISIC2018. Notably, BUSI encompasses ultrasound images of breast cancer cases, encompassing normal, benign, and malignant instances, accompanied by their respective segmentation masks. For experimental purposes, we exclusively selected the malignant images, amounting to a total of 210. Table 3 lists detailed information about the datasets. These datasets all comprise images and their associated ground truth masks. These chosen datasets are frequently employed in biomedical image segmentation tasks. The selection of datasets from diverse imaging modalities and distinct anatomical regions aims to evaluate the performance and robustness of the proposed approach.

Table 3
Medical datasets used in our experiments.

Dataset	Modality	Images	Original Resolution	Input Resolution	Train	Test
CVC-ClinicDB [15]	Endoscopy	612	384×288	384×384	490	122
CVC-ColonDB [16]	Colonoscopy	380	574×500	384×384	304	76
ISIC2018 [18,19]	Dermoscopy	2594	Variable	512×512	2075	519
BUSI [17]	Ultrasound	210	Variable	384×384	168	42
GlaS [14]	H&E Stained Histological Sections	165	Variable	384×384	132	33

4.2. Evaluation metrics

We chose several performance evaluation metrics for the network models, including the dice similarity coefficient (DSC), intersection over union (IoU), true positive rate (TPR), positive predictive value (PPV), and the number of parameters. DSC and IOU quantify the similarity between segmented regions and ground truth regions. TPR assesses the segmentation model's performance, specifically its capability to accurately detect true positives. A higher TPR value signifies a model's improved capacity to accurately detect true positive cases, indicating heightened sensitivity. PPV is another metric for evaluating the segmentation model's performance, specifically measuring the proportion of identified positives that are true positives. A higher PPV value signifies the model's enhanced accuracy in identifying positives, indicating higher precision. The parameter count reflects the model's complexity and capacity, where smaller parameter counts indicate reduced model size and complexity. The definitions of DSC, IOU, TPR, and PPV are shown in formulas (5), (6), (7), and (8) below:

$$DSC = \frac{2\left|I_{gt} \bigcap I_{pred}\right|}{\left|I_{gt}\right| + \left|I_{pred}\right|} = \frac{2TP}{2TP + FP + FN} \tag{5}$$

$$IoU = \frac{\left|I_{gt} \bigcap I_{pred}\right|}{\left|I_{gt} \bigcup I_{pred}\right|} = \frac{TP}{TP + FP + FN} \tag{6}$$

$$TPR = \frac{I_{gt} \bigcap I_{prep}}{I_{gt}} = \frac{TP}{TP + FN} \tag{7}$$

$$PPV = \frac{I_{gt} \bigcap I_{prep}}{I_{prep}} = \frac{TP}{TP + FP} \tag{8}$$

In this context, TP, FP, and FN represent True Positives, False Positives, and False Negatives, respectively. I_{gt} and I_{prep} refer to the ground truth segmentation and predicted segmentation, respectively.

4.3. Implementation details

We trained and evaluated the proposed model on a workstation equipped with an NVIDIA RTX 3060 GPU. We used the Adam optimizer with a learning rate of 0.0001 for the ISIC2018 dataset and a learning rate of 0.001 for the BUSI, GlaS, CVC-ClinicDB and CVC-ColonDB datasets. The momentum was set to 0.9. Additionally, we employed a cosine annealing learning rate scheduler with a minimum learning rate of 0.00001. The batch size was set to 8. AMNNet was trained for a total of 400 epochs. 80% of the dataset was used for training, while the remaining 20% was reserved for testing. For ISIC2018, the image size was adjusted to 512×512, whereas images from CVC-ClinicDB, CVC-ColonDB, BUSI, and GlaS were resized to a resolution of 384×384.

4.4. Experimental results

4.4.1. Method comparisons

In this section, we will present a comparison of AMNNet with other state-of-the-art methods.

(1) Comparison on ISIC-2018 Skin Lesion Segmentation Challenge: Melanoma is a common form of cancer, and automated diagnostic tools for skin lesions can help in accurately detecting melanoma, potentially saving up to 99% of cases [39]. The quantitative results for ISIC2018 are summarized in Table 4. Our method achieved remarkable metrics: a DSC of 91.35%, an IoU of 84.36%, a TPR of 90.96%, and a PPV of 92.14%. Notably, our results closely resemble those of Attention UNet [26] in terms of DSC, IoU, TPR, and PPV, while significantly reducing the parameter count by 33.24M compared to Attention UNet. We also outperform Attention UNet with improvements of 0.63% in DSC and 0.9% in IoU. Additionally, our TPR and PPV exhibit enhancements of 1.13% and 0.9%, respectively, compared to UNext. The network's melanoma segmentation capabilities are visually evident through the comparison of ground truth and predicted masks (Fig. 4).

(2) Comparison on CVC-ClinicDB: Early detection of polyps, which can potentially progress into colorectal cancer, significantly improves survival rates [42]. In our experiments, we selected the widely-used colonoscopy dataset, CVC-ClinicDB. The quantitative evaluation of AMNNet is presented in Table 5, while the qualitative results are depicted in Fig. 5. From the quantitative results, our

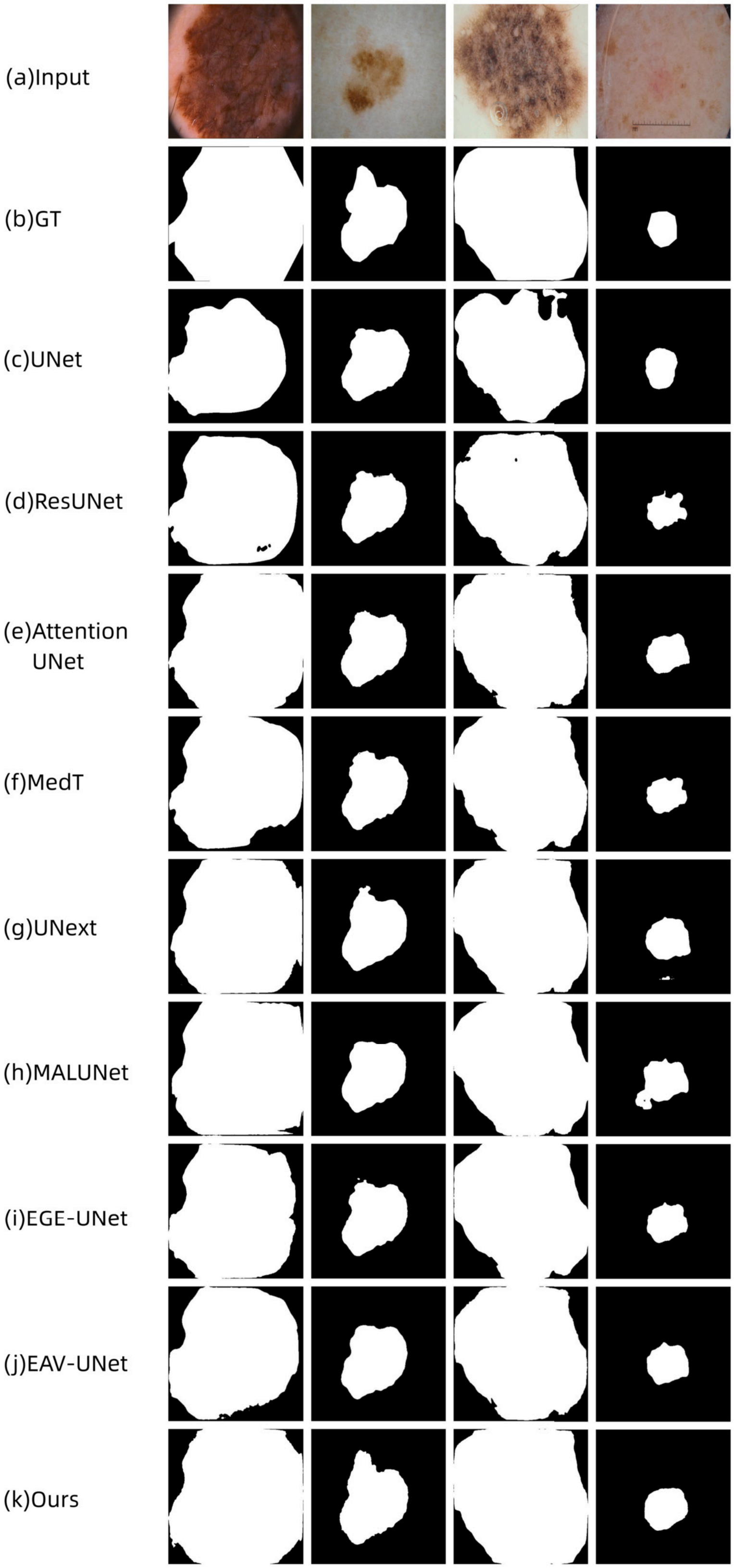

Fig. 4. Figure displays the qualitative results of UNet, ResUNet, Attention UNet, MedT, UNext, MALUNet, EGE-UNet, EAV-UNet, and Our method on the ISIC2018 dataset.

Table 4
Comparison results of UNet, ResUNet, Attention UNet, MedT, UNext, MALUNet, EGE-UNet, EAV-UNet, and Our method on the ISIC2018 dataset.

Method	DSC	IoU	TPR	PPV	Parameters
UNet [7]	85.14	75.01	80.71	91.55	31.04M
ResUNet [9]	86.48	76.91	82.86	91.45	31.56M
Attention UNet [26]	90.72	83.46	90.68	91.13	34.88M
MedT [25]	88.27	79.70	85.72	91.66	1.56M
UNext [22]	90.21	82.84	89.83	91.24	1.47M
MALUNet [30]	88.83	80.57	90.33	88.09	0.175M
EGE-UNet [40]	89.22	81.21	87.45	91.76	**0.053M**
EAV-UNet [41]	86.93	77.46	84.96	90.02	32.67M
Ours	**91.35**	**84.36**	**90.96**	**92.14**	1.64M

Table 5
Comparison results of UNet, ResUNet, Attention UNet, MedT, UNext, MALUNet, EGE-UNet, EAV-UNet, and Our method on the CVC-ClinicDB dataset.

Method	DSC	IoU	TPR	PPV	Parameters
UNet [7]	82.71	71.17	79.44	87.25	31.04M
ResUNet [9]	86.19	76.08	81.98	91.91	31.56M
Attention UNet [26]	83.59	72.58	79.07	89.76	34.88M
MedT [25]	67.38	58.88	60.39	82.67	1.56M
UNext [22]	86.71	77.04	83.58	90.82	1.47M
MALUNet [30]	71.78	56.45	69.29	76.10	0.175M
EGE-UNet [40]	63.95	47.35	65.22	64.58	**0.053M**
EAV-UNet [41]	85.39	74.95	80.21	**92.25**	32.67M
Ours	**90.01**	**82.23**	**89.18**	91.17	1.64M

Table 6
Comparison results of UNet, ResUNet, Attention UNet, MedT, UNext, MALUNet, EGE-UNet, EAV-UNet, and Our method on the CVC-ColonDB dataset.

Method	DSC	IoU	TPR	PPV	Parameters
UNet [7]	78.35	65.41	74.62	85.57	31.04M
ResUNet [9]	86.27	76.31	84.47	88.70	31.56M
Attention UNet [26]	61.26	45.36	59.73	68.41	34.88M
MedT [25]	50.51	34.50	48.86	57.97	1.56M
UNext [22]	84.98	74.51	82.95	87.99	1.47M
MALUNet [30]	63.36	48.09	66.84	63.53	0.175M
EGE-UNet [40]	53.17	38.04	70.76	48.06	**0.053M**
EAV-UNet [41]	86.93	77.46	84.96	90.02	32.67M
Ours	**90.80**	**83.41**	**90.10**	**91.58**	1.64M

method achieved a DSC of 90.01%, an IoU of 82.23%, a TPR of 89.18%, and a PPV of 91.17%. Compared to the best-performing method, UNext, our approach exhibited a 3.3% increase in DSC, a 5.19% increase in IoU, a 5.6% increase in TPR, and a 0.35% increase in PPV. Despite EGE-UNet having 1.59M less parameters than AMNNet, AMNNet outperformed EGE-UNet by 26.06% in DSC and 34.88% in IoU. In comparison to MedT, our method demonstrated a 28.79% increase in TPR and an 8.5% increase in PPV. Although ResUNet had a 0.74% higher PPV than AMNNet, AMNNet outperformed ResUNet by 7.2% in TPR.

(3) Comparison on CVC-ColonDB: CVC-ColonDB is the second polyp dataset used in our experiments. The quantitative results are shown in Table 6. In terms of segmentation performance, our method outperforms other state-of-the-art methods. Specifically, our method achieves a DSC of 90.80% and an IoU of 83.41%. Compared to EAV-UNet, our method improves by 3.87% and 5.95%, respectively. Compared to ResUNet, our method increases DSC and IoU by 4.53% and 7.1%, respectively. In terms of TPR and PPV, our method achieves 90.10% and 91.58%, respectively. Compared to the method with the fewest parameters, EGE-UNet, our method increases TPR by 19.34% and PPV by 43.52%. Compared to UNext, our method improves TPR by 7.15% and PPV by 3.59%. Qualitative results are shown in Fig. 6.

(4) Comparison on BUSI: Breast cancer stands as a significant cause of female mortality worldwide, underscoring the importance of early detection in reducing premature deaths. In our research, we utilized a breast ultrasound dataset, BUSI, to assess the performance of AMNNet. The quantitative results presented in Table 7 reveal that our method achieved a DSC of 81.61% and an IoU of 69.04%. When compared to the state-of-the-art UNext, our method demonstrated a noteworthy improvement, enhancing DSC by 3.84% and IoU by 5.16%. AMNNet improved DSC by 13.84% and IoU by 17.2% compared to MALUNet. AMNNet exhibited a TPR of 82.09% and a PPV of 81.78%, marking substantial enhancements of 11.68% and 17.05%, respectively, in comparison to EGE-UNet, despite the latter having only 0.053M parameters. In comparison to MedT, our approach showcased remarkable progress with a 27.14% increase

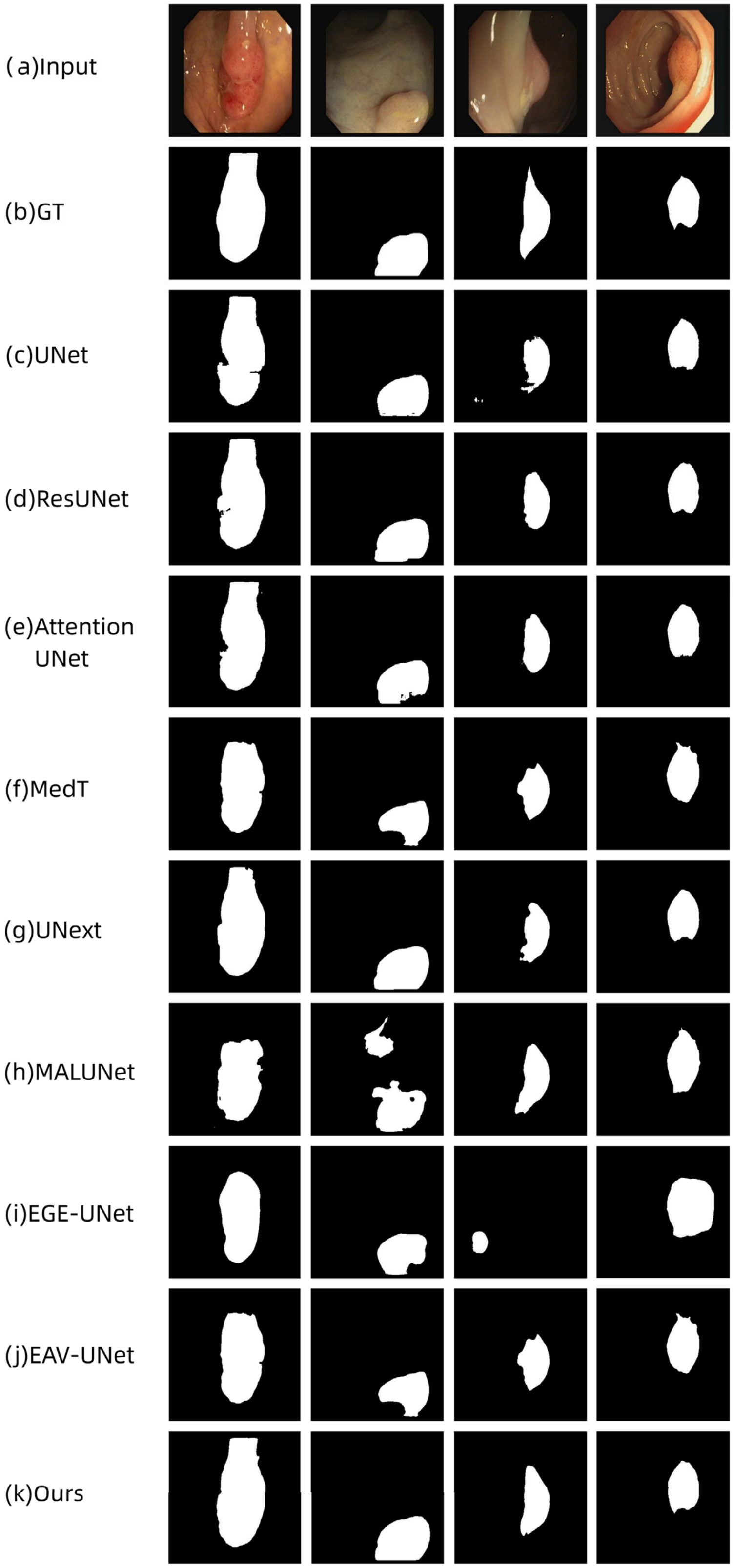

Fig. 5. Figure displays the qualitative results of UNet, ResUNet, Attention UNet, MedT, UNext, MALUNet, EGE-UNet, EAV-UNet, and Our method on the CVC-ClinicDB dataset.

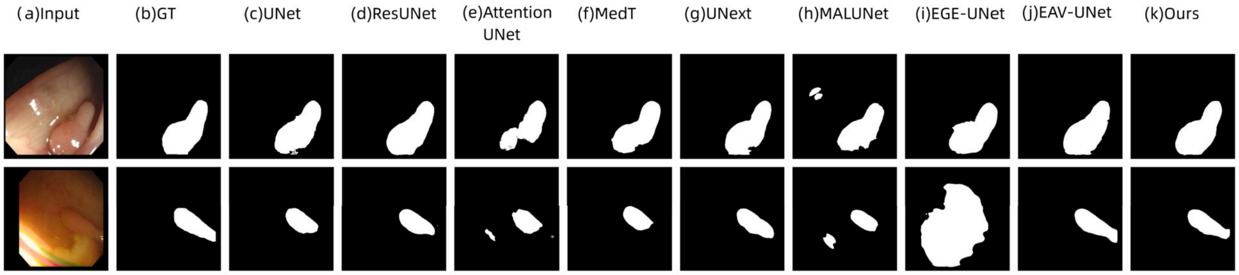

Fig. 6. Figure displays the qualitative results of UNet, ResUNet, Attention UNet, MedT, UNext, MALUNet, EGE-UNet, EAV-UNet, and Our method on the CVC-ColonDB datasets.

Table 7
Comparison results of UNet, ResUNet, Attention UNet, MedT, UNext, MALUNet, EGE-UNet, EAV-UNet, and Our method on the BUSI dataset.

Method	DSC	IoU	TPR	PPV	Parameters
UNet [7]	71.96	56.39	69.92	75.60	31.04M
ResUNet [9]	74.37	59.48	70.93	78.75	31.56M
Attention UNet [26]	74.10	59.26	70.49	79.39	34.88M
MedT [25]	53.12	39.45	54.95	61.96	1.56M
UNext [22]	77.77	63.88	76.74	80.24	1.47M
MALUNet [30]	67.77	51.84	66.18	72.00	0.175M
EGE-UNet [40]	65.86	49.41	70.41	64.73	**0.053M**
EAV-UNet [41]	73.82	59.23	78.51	71.59	32.67M
Ours	**81.61**	**69.04**	**82.09**	**81.78**	1.64M

Table 8
Comparison results of UNet, ResUNet, Attention UNet, MedT, UNext, MALUNet, EGE-UNet, EAV-UNet, and Our method on the GlaS dataset.

Method	DSC	IoU	TPR	PPV	Parameters
UNet [7]	93.40	87.69	93.52	93.34	31.04M
ResUNet [9]	93.52	87.91	93.63	93.50	31.56M
Attention UNet [26]	93.27	87.49	93.06	93.52	34.88M
MedT [25]	78.91	66.80	89.63	73.34	1.56M
UNext [22]	93.24	87.47	93.81	92.68	1.47M
MALUNet [30]	84.28	73.48	79.32	90.76	0.175M
EGE-UNet [40]	87.85	78.53	91.16	84.85	**0.053M**
EAV-UNet [41]	93.30	87.54	**95.65**	91.09	32.67M
Ours	**94.31**	**89.33**	94.21	**94.45**	1.64M

in TPR and a 19.82% surge in PPV. The segmentation capabilities of our method, accurately delineating lesions of varying sizes, are visually evident in Fig. 7.

(5) Comparison on GlaS: Colorectal adenocarcinoma, the most prevalent form of colon cancer, originates from glandular structures in the colon [14]. In clinical practice, quantifying gland morphology enables pathologists to assess patient prognosis and tailor personalized treatments based on the morphological characteristics of colonic glands, encompassing their structural appearance and glandular formation. Our experiments were conducted using the GlaS dataset. The quantitative results presented in Table 8 demonstrate that our method achieved a DSC of 94.31% and an IoU of 89.33%, indicating an improvement of 0.79% and 1.42% compared to the high-performing ResUNet. In comparison to Attention UNet, our approach showcased a 1.04% increase in DSC and a 1.84% increase in IoU. Compared with EAV-UNet, DSC and IoU of AMNNet increased by 1.01% and 1.79% respectively. Concerning TPR and PPV, our method exhibited 94.21% and 94.45%, showcasing improvements of 4.58% and 21.11% compared to MedT. These enhancements amounted to 3.05% and 9.6% in comparison to EGE-UNet. As seen in the results in Fig. 7, it is evident that AMNNet's predicted masks most closely resemble the ground truth masks.

4.5. Ablation study

To assess the efficacy of the various modules developed in our experiments and study their effectiveness in melanoma, polyp, breast cancer, and gland segmentation, we employed a controlled variable approach to evaluate the network's performance across the ISIC2018, CVC-ClinicDB, CVC-ColonDB, BUSI, and GlaS datasets. Table 9, Table 10, Table 11, Table 12, and Table 13 present the influence of each module on the precision of melanoma, polyp, breast cancer, and gland segmentation. These tables reveal that the CBAM and RSUC modules effectively identify and delineate lesion regions within the images, resulting in enhanced segmentation accuracy.

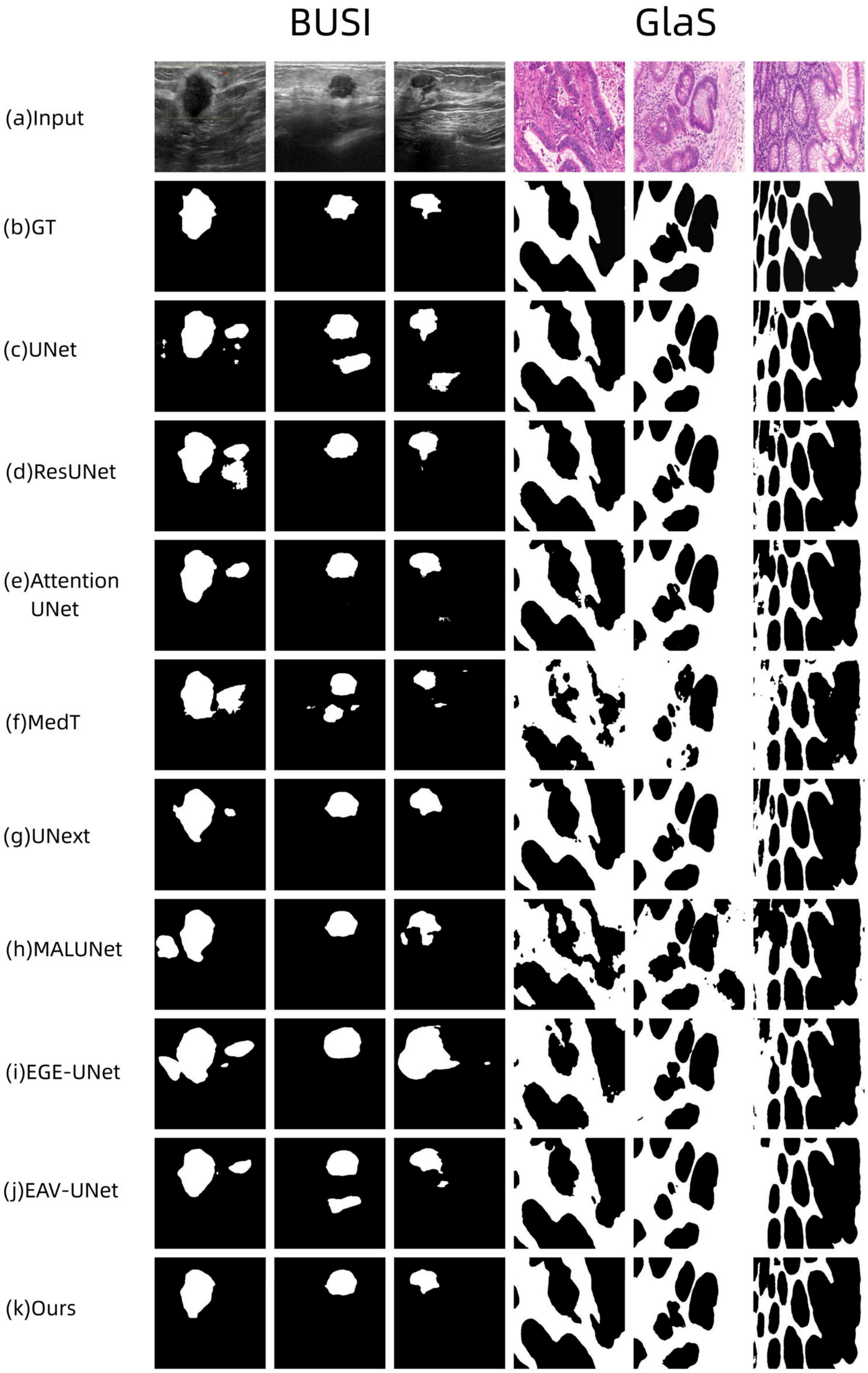

Fig. 7. Figure displays the qualitative results of UNet, ResUNet, Attention UNet, MedT, UNext, MALUNet, EGE-UNet, EAV-UNet, and Our method on the BUSI and GlaS datasets.

Table 9
Ablation studies of each module on the ISIC2018 dataset.

Method	DSC	IoU	TPR	PPV	Parameters
MLP+Conv stage(L+R)(baseline)	90.21	82.84	89.83	91.24	**1.47M**
MLP+Conv stage(L+R)+CBAM	90.91	83.62	90.09	**92.16**	**1.47M**
MLP+RSUC+Conv stage(R)+CBAM(Ours)	**91.35**	**84.36**	**90.96**	92.14	1.64M

Table 10
Ablation studies of each module on the CVC-ClinicDB dataset.

Method	DSC	IoU	TPR	PPV	Parameters
MLP+Conv stage(L+R)(baseline)	86.71	77.04	83.58	90.82	**1.47M**
MLP+Conv stage(L+R)+CBAM	87.90	78.98	88.03	88.06	**1.47M**
MLP+RSUC+Conv stage(R)+CBAM(Ours)	**90.01**	**82.23**	**89.18**	**91.17**	1.64M

Table 11
Ablation studies of each module on the CVC-ColonDB dataset.

Method	DSC	IoU	TPR	PPV	Parameters
MLP+Conv stage(L+R)(baseline)	84.98	74.51	82.95	87.99	**1.47M**
MLP+Conv stage(L+R)+CBAM	88.85	80.34	87.47	90.58	**1.47M**
MLP+RSUC+Conv stage(R)+CBAM(Ours)	**90.80**	**83.41**	**90.10**	**91.58**	1.64M

Table 12
Ablation studies of each module on the BUSI dataset.

Method	DSC	IoU	TPR	PPV	Parameters
MLP+Conv stage(L+R)(baseline)	77.77	63.88	76.74	80.24	**1.47M**
MLP+Conv stage(L+R)+CBAM	80.10	66.86	78.98	**81.99**	**1.47M**
MLP+RSUC+Conv stage(R)+CBAM(Ours)	**81.61**	**69.04**	**82.09**	81.78	1.64M

Table 13
Ablation studies of each module on the GlaS dataset.

Method	DSC	IoU	TPR	PPV	Parameters
MLP+Conv stage(L+R)(baseline)	93.24	87.47	93.81	92.68	**1.47M**
MLP+Conv stage(L+R)+CBAM	93.59	88.07	93.74	93.47	**1.47M**
MLP+RSUC+Conv stage(R)+CBAM(Ours)	**94.31**	**89.33**	**94.21**	**94.45**	1.64M

4.6. Cross validation

In order to further evaluate the performance and verify the robustness of the model, we conducted five 5-fold cross-validations on the BUSI and GlaS datasets, and the results were shown in Table 14 and 15. We divided the data set into five equal parts, using four of them as the training set and the remaining one as the test set. So, do five training and testing sessions. The experimental results in Table 14 and 15 indicate the robustness of AMNNet.

5. Discussion

Accurate medical image segmentation results serve as crucial clinical indicators, assisting healthcare professionals in cancer diagnosis and enhancing patient care. However, medical images often exhibit noise, artifacts, low contrast, and unclear boundaries between pathological and normal tissues. Furthermore, some images originate from different modalities or scanning devices, complicating the precise segmentation of tumors, organs, and tissues. To address these challenges, we introduced the CBAM to the decoder of the AMNNet. This strategic integration aligns the network's focus on the extraction of essential features while suppressing irrelevant information, including noise, thereby enhancing segmentation performance. While traditional medical image segmentation methods demonstrate commendable segmentation capabilities, they possess certain limitations. For instance, UNet, ResNet, DenseNet, and UNet++ primarily rely on relatively uniform convolution kernels, making it challenging to extract features across various scales. To better capture phased multi-scale features within the images, we developed the RSUC module, which is seamlessly integrated into the early stages of the AMNNet encoder, resulting in improved segmentation performance. In the initial encoder stage (l1), characterized by larger-sized feature maps in shallow stages, we employ the RSUC module with a parameter L = 7 to capture information on a larger scale. Conversely, as we progress to the l2 and l3 stages, where feature map sizes decrease, we employ RSUC modules with L = 6 and L = 5, respectively, to prevent the loss of vital information. In the RSUC module, we also incorporate the CBAM module to

Table 14
Method 1,2,3,4,5 is the result of our 5 cross-validations on the BUSI dataset, Mean value is the mean of our 5 results, and the last line is the result of our original training set 80% and test set 20% random splitting on BUSI.

Method	DSC	IoU	TPR	PPV
1	81.97	70.03	77.94	88.03
2	81.60	69.37	84.79	79.80
3	80.43	68.05	78.47	84.19
4	81.05	68.77	79.76	83.85
5	80.59	68.47	78.72	85.02
Mean value	81.13	68.94	79.94	84.18
Ours	81.61	69.04	82.09	81.78

Table 15
Method 1,2,3,4,5 is the result of our 5 cross-validations on the GlaS dataset, Mean value is the mean of our 5 results, and the last line is the result of our original training set 80% and test set 20% random splitting on GlaS.

Method	DSC	IoU	TPR	PPV
1	94.27	89.17	95.46	93.15
2	94.55	89.66	93.60	95.52
3	94.21	89.06	94.77	93.67
4	94.65	89.84	92.97	96.41
5	93.98	88.68	92.21	95.91
Mean value	94.33	89.28	93.80	94.93
Ours	94.31	89.33	94.21	94.45

optimize segmentation performance. Analysis of Table 1 and Table 2 consistently reveals that the positioning of the CBAM module within the RSUC module yields varying results. Notably, the highest model performance is achieved when CBAM is placed between the encoder and decoder. This strategic placement balances feature extraction and integration, thereby enhancing model robustness and performance. However, incorporating CBAM solely in the encoder presents limitations since the spatial feature maps at the onset of the encoder stage are too extensive with inadequate channel capacity, resulting in an insufficient ability to capture specific features and, consequently, limited performance improvements.

In practical clinical settings, the performance of deep learning-based segmentation methods often deteriorates due to variations in imaging protocols and patient characteristics. Models capable of generalizing across multiple medical center datasets are highly desirable in clinical applications [43]. As evidenced by the segmentation results presented in Figs. 4, 5, 6, and 7, and the experimental data provided in Tables 4, 5, 6, 7, and 8, AMNNet outperforms UNet, ResUNet, Attention UNet, MedT, UNext, MALUNet, EGE-UNet and EAV-UNet on ISIC2018, CVC-ClinicDB, CVC-ColonDB, BUSI, and GlaS datasets, respectively, in terms of segmentation performance. The experimental data in Tables 9, 10, 11, 12, and 13 clearly demonstrate that the inclusion of the CBAM module and the integration of the RSUC module within the early stages of the AMNNet encoder significantly enhance the model's segmentation performance across the ISIC2018, CVC-ClinicDB, CVC-ColonDB, BUSI, and GlaS datasets. These results underscore the versatility of our proposed AMNNet, attributable to the RSUC module's capacity to capture phased multi-scale features while retaining representative features within different categories.

While AMNNet demonstrates effective model performance optimization, it is important to note that it does not possess the fewest parameters, as indicated in Tables 4, 5, 6 7, 8, 9, 10, 11, 12, and 13. When compared to UNet, ResUNet, Attention UNet, and EAV-UNet our model boasts a reduction in parameters by 29.4M, 29.92M, 33.24M and 31.03M, respectively. However, it is essential to acknowledge that it exhibits a marginal increase of 0.08M, 0.17M, 1.465M, and 1.587M in parameters when contrasted with MedT, UNext, MALUNet, and EGE-UNet. Therefore, future research endeavors will be directed towards the development of more lightweight models tailored for practical applications.

6. Conclusion

Cancer is a significant health concern with life-threatening implications. Various cancer types can impact multiple organs and systems, resulting in functional disruptions. Inaccurate segmentation of organs, tissues, and tumors can lead to medical misdiagnoses with severe repercussions. In this paper, we design the AMNNet architecture for the segmentation of medical images. The AMNNet model incorporates CBAM, which dynamically adjusts feature map weights using channel and spatial attention mechanisms to emphasize specific regions and structures in medical images while suppressing irrelevant information. Furthermore, we propose the RSUC module, integrated into the initial stages of the AMNNet encoder, to capture more comprehensive local and global information from shallow and deep layers, thus enhancing segmentation performance. Our experiments provide evidence that AMNNet surpasses several state-of-the-art computer vision methods on five distinct biomedical datasets, showcasing its superior segmentation performance and its potential to enhance automated cancer disease analysis and intelligent diagnosis.

CRediT authorship contribution statement

Dapeng Cheng: Resources, Methodology, Data curation. **Jia Deng:** Writing – review & editing, Writing – original draft, Visualization, Validation, Resources, Methodology, Investigation, Data curation. **Jinjie Xiao:** Visualization, Data curation. **Mao Yanyan:** Visualization, Data curation. **Jialong Kang:** Visualization, Data curation. **Jiale Gai:** Visualization, Validation. **Baosheng Zhang:** Visualization, Validation. **Feng Zhao:** Visualization, Validation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that support the findings of this study are openly available in CVC-ClinicDB dataset at https://doi.org/10.1016/j.compmedimag.2015.02.007, CVC-ColonDB dataset at https://doi.org/10.1016/j.patcog.2012.03.002, ISIC2018 dataset at https://doi.org/10.48550/arXiv.1902.03368, https://doi.org/10.1038/sdata.2018.161, BUSI dataset at https://doi.org/10.1016/j.dib.2019.104863, GlaS dataset at https://doi.org/10.1016/j.media.2016.08.008 reference number [15,16,18,19,17,14].

Funding

This work was supported in part by the National Natural Science Foundation of China (62176140).

General Special Project of Shandong Provincial Education “14th Five-Year Plan” + Application Research of Artificial Intelligence in Financial Management in Higher Education + Project Number: 2021CYB012.

References

[1] J. Fan, Z. Liu, X. Mao, X. Tong, T. Zhang, C. Suo, X. Chen, Global trends in the incidence and mortality of esophageal cancer from 1990 to 2017, Cancer Med. 9 (2020) 6875–6887, https://doi.org/10.1002/cam4.3338.

[2] H.-Y. Zhou, J. Guo, Y. Zhang, X. Han, L. Yu, L. Wang, Y. Yu, nnformer: volumetric medical image segmentation via a 3d transformer, IEEE Trans. Image Process. (2023), https://doi.org/10.1109/TIP.2023.3293771.

[3] H. Bao, Y. Zhu, Q. Li, Hybrid-scale contextual fusion network for medical image segmentation, Comput. Biol. Med. 152 (2023) 106439, https://doi.org/10.1016/j.compbiomed.2022.106439.

[4] B. Chen, Y. Liu, Z. Zhang, G. Lu, A.W.K. Kong, Transattunet: multi-level attention-guided u-net with transformer for medical image segmentation, IEEE Trans. Emerg. Top. Comput. Intell. (2023), https://doi.org/10.1109/TETCI.2023.3309626.

[5] D. Jha, S. Ali, N.K. Tomar, H.D. Johansen, D. Johansen, J. Rittscher, M.A. Riegler, P. Halvorsen, Real-time polyp detection, localization and segmentation in colonoscopy using deep learning, IEEE Access 9 (2021) 40496–40510, https://doi.org/10.1109/ACCESS.2021.3063716.

[6] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.A. Van Der Laak, B. Van Ginneken, C.I. Sánchez, A survey on deep learning in medical image analysis, Med. Image Anal. 42 (2017) 60–88, https://doi.org/10.1016/j.media.2017.07.005.

[7] O. Ronneberger, P. Fischer, T. Brox, 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, Springer, 2015, pp. 234–241.

[8] 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 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19, Springer, 2016, pp. 424–432.

[9] Z. Zhang, Q. Liu, Y. Wang, Road extraction by deep residual u-net, IEEE Geosci. Remote Sens. Lett. 15 (2018) 749–753, https://doi.org/10.1109/LGRS.2018.2802944.

[10] S. Cai, Y. Tian, H. Lui, H. Zeng, Y. Wu, G. Chen, Dense-unet: a novel multiphoton in vivo cellular image segmentation model based on a convolutional neural network, Quant. Imag. Med. Surg. 10 (2020) 1275, https://doi.org/10.21037/qims-19-1090.

[11] Z. Zhou, M.M. Rahman Siddiquee, N. Tajbakhsh, J. Liang, Unet++: a nested u-net architecture for medical image segmentation, in: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4, Springer, pp. 3–11, 2018.

[12] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Y.-W. Chen, J. Wu, Unet 3+: a full-scale connected unet for medical image segmentation, in: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2020, pp. 1055–1059.

[13] 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.

[14] K. Sirinukunwattana, J.P. Pluim, H. Chen, X. Qi, P.-A. Heng, Y.B. Guo, L.Y. Wang, B.J. Matuszewski, E. Bruni, U. Sanchez, et al., Gland segmentation in colon histology images: the glas challenge contest, Med. Image Anal. 35 (2017) 489–502, https://doi.org/10.1016/j.media.2016.08.008.

[15] J. Bernal, F.J. Sánchez, G. Fernández-Esparrach, D. Gil, C. Rodríguez, F. Vilariño, Wm-dova maps for accurate polyp highlighting in colonoscopy: validation vs. saliency maps from physicians, Comput. Med. Imaging Graph. 43 (2015) 99–111, https://doi.org/10.1016/j.compmedimag.2015.02.007.

[16] J. Bernal, F.J. Sánchez, F. Vilariño, Towards automatic polyp detection with a polyp appearance model, Pattern Recognit. 45 (2012) 3166–3182, https://doi.org/10.1016/j.patcog.2012.03.002.

[17] W. Al-Dhabyani, M. Gomaa, H. Khaled, A. Fahmy, Dataset of breast ultrasound images, Data Brief 28 (2020) 104863, https://doi.org/10.1016/j.dib.2019.104863.

[18] N. Codella, V. Rotemberg, P. Tschandl, M.E. Celebi, S. Dusza, D. Gutman, B. Helba, A. Kalloo, K. Liopyris, M. Marchetti, et al., Skin lesion analysis toward melanoma detection 2018: a challenge hosted by the international skin imaging collaboration (isic), arXiv preprint arXiv:1902.03368, 2019, https://doi.org/10.48550/arXiv.1902.03368.

[19] P. Tschandl, C. Rosendahl, H. Kittler, The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions, Sci. Data 5 (2018) 1–9, https://doi.org/10.1038/sdata.2018.161.

[20] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431–3440.

[21] N. Ibtehaz, M.S. Rahman, Multiresunet: rethinking the u-net architecture for multimodal biomedical image segmentation, Neural Netw. 121 (2020) 74–87, https://doi.org/10.1016/j.neunet.2019.08.025.

[22] J.M.J. Valanarasu, V.M. Patel, Unext: Mlp-based rapid medical image segmentation network, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2022, pp. 23–33.

[23] I.O. Tolstikhin, N. Houlsby, A. Kolesnikov, L. Beyer, X. Zhai, T. Unterthiner, J. Yung, A. Steiner, D. Keysers, J. Uszkoreit, et al., Mlp-mixer: an all-mlp architecture for vision, Adv. Neural Inf. Process. Syst. 34 (2021) 24261–24272, https://doi.org/10.48550/arXiv.2105.01601.

[24] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al., An image is worth 16x16 words: transformers for image recognition at scale, arXiv preprint arXiv:2010.11929, 2020, https://doi.org/10.48550/arXiv.2010.11929.

[25] J.M.J. Valanarasu, P. Oza, I. Hacihaliloglu, V.M. Patel, Medical Transformer: Gated Axial-Attention for Medical Image Segmentation, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24, Springer, 2021, pp. 36–46.

[26] O. Oktay, J. Schlemper, L.L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N.Y. Hammerla, B. Kainz, et al., Attention u-net: learning where to look for the pancreas, arXiv preprint arXiv:1804.03999, 2018, https://doi.org/10.48550/arXiv.1804.03999.

[27] X. Xiao, S. Lian, Z. Luo, S. Li, Weighted res-unet for high-quality retina vessel segmentation, in: 2018 9th International Conference on Information Technology in Medicine and Education (ITME), IEEE, 2018, pp. 327–331.

[28] C. Guo, M. Szemenyei, Y. Yi, W. Wang, B. Chen, C. Fan, Sa-unet: spatial attention u-net for retinal vessel segmentation, in: 2020 25th International Conference on Pattern Recognition (ICPR), IEEE, 2021, pp. 1236–1242.

[29] 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.

[30] J. Ruan, S. Xiang, M. Xie, T. Liu, Y. Fu, Malunet: a multi-attention and light-weight unet for skin lesion segmentation, CoRR, arXiv:2211.01784, 2022, https://doi.org/10.48550/arXiv.2211.01784.

[31] H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia, Pyramid scene parsing network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2881–2890.

[32] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille, Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs, IEEE Trans. Pattern Anal. Mach. Intell. 40 (2017) 834–848, https://doi.org/10.1109/TPAMI.2017.2699184.

[33] X. Wu, D. Hong, J. Chanussot, Uiu-net: U-net in u-net for infrared small object detection, IEEE Trans. Image Process. 32 (2022) 364–376, https://doi.org/10.1109/TIP.2022.3228497.

[34] X. Qin, Z. Zhang, C. Huang, M. Dehghan, O.R. Zaiane, M. Jagersand, U2-net: going deeper with nested u-structure for salient object detection, Pattern Recognit. 106 (2020) 107404, https://doi.org/10.1016/j.patcog.2020.107404.

[35] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556, 2014, https://doi.org/10.48550/arXiv.1409.1556.

[36] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[37] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4700–4708.

[38] F. Milletari, N. Navab, S.-A. Ahmadi, V-net: fully convolutional neural networks for volumetric medical image segmentation, in: 2016 Fourth International Conference on 3D Vision (3DV), Ieee, 2016, pp. 565–571.

[39] S. Ac, Cancer facts & figures 2018, Cancer Facts Fig. 2018 (2018) 1–71.

[40] J. Ruan, M. Xie, J. Gao, T. Liu, Y. Fu, Ege-unet: an efficient group enhanced unet for skin lesion segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2023, pp. 481–490.

[41] D. Cheng, X. Gao, Y. Mao, B. Xiao, P. You, J. Gai, M. Zhu, J. Kang, F. Zhao, N. Mao, Brain tumor feature extraction and edge enhancement algorithm based on u-net network, Heliyon 9 (2023).

[42] B. Levin, D.A. Lieberman, B. McFarland, K.S. Andrews, D. Brooks, J. Bond, C. Dash, F.M. Giardiello, S. Glick, D. Johnson, et al., Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, Gastroenterology 134 (2008) 1570–1595, https://doi.org/10.1053/j.gastro.2008.02.002.

[43] S. Ali, D. Jha, N. Ghatwary, S. Realdon, R. Cannizzaro, O.E. Salem, D. Lamarque, C. Daul, M.A. Riegler, K.V. Anonsen, et al., Polypgen: a multi-center polyp detection and segmentation dataset for generalisability assessment, arXiv preprint arXiv:2106.04463, 2021, https://doi.org/10.48550/arXiv.2106.04463.