Computers in Biology and Medicine 158 (2023) 106838

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

RMAU-Net: Residual Multi-Scale Attention U-Net For liver and tumor segmentation in CT images

Linfeng Jiang [a], Jiajie Ou [a], Ruihua Liu [a,*], Yangyang Zou [a], Ting Xie [a], Hanguang Xiao [a], Ting Bai [b]

[a] School of Artificial Intelligence, Chongqing University of Technology, Chongqing, China
[b] School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Stockholm, Sweden

ARTICLE INFO

Keywords:
Liver and tumor segmentation
Multi-scale feature
Attention mechanism
Deep learning
Medical imaging

ABSTRACT

Liver cancer is one of the leading causes of cancer-related deaths worldwide. Automatic liver and tumor segmentation are of great value in clinical practice as they can reduce surgeons' workload and increase the probability of success in surgery. Liver and tumor segmentation is a challenging task because of the different sizes, shapes, blurred boundaries of livers and lesions, and low-intensity contrast between organs within patients. To address the problem of fuzzy livers and small tumors, we propose a novel Residual Multi-scale Attention U-Net (RMAU-Net) for liver and tumor segmentation by introducing two modules, *i.e.*, Res-SE-Block and MAB. The Res-SE-Block can mitigate the problem of gradient disappearance by residual connection and enhance the quality of representations by explicitly modeling the interdependencies and feature recalibration between the channels of features. The MAB can exploit rich multi-scale feature information and capture inter-channel and inter-spatial relationships of features simultaneously. In addition, a hybrid loss function, that combines focal loss and dice loss, is designed to improve segmentation accuracy and speed up convergence. We evaluated the proposed method on two publicly available datasets, *i.e.*, LiTS and 3D-IRCADb. Our proposed method achieved better performance than the other state-of-the-art methods, with dice scores of 0.9552 and 0.9697 for LiTS and 3D-IRCABb liver segmentation, and dice scores of 0.7616 and 0.8307 for LiTS and 3D-IRCABb liver tumor segmentation.

1. Introduction

The liver is one of the most important organs in the human body due to its detoxifying and digestive functions [1]. As the fourth-highest death rate of all malignancies [2], liver cancer has become a serious hazard to human health. Computer tomography (CT) is one of the most common imaging modalities, and is typically used by radiologists and oncologists to evaluate and analyze liver and lesions. Radiologists and oncologists can find areas of the lesion and thus develop a diagnosis and treatment plan by analyzing computed tomography (CT) or magnetic resonance images (MRI). Currently, most of the segmentation of the liver and tumor is performed manually, which is labor-intensive, time-consuming, and depends on the experience of the surgeons. Computer-assisted liver and tumor segmentation can reduce the surgeon's workload and can increase the success rate of surgery, which is of great clinical value. However, the shape, location, and volume of livers and tumors vary from patients to patients, the boundaries between lesions and surrounding normal liver tissues are blurred, and differences in imaging equipment and settings can lead to significant differences in tumor color and contrast, making automated computer-assisted liver and tumor segmentation a challenging study.

Researchers have explored some traditional methods [3–5] for biomedical image segmentation tasks. Traditional methods are significantly more efficient compared to manual segmentation methods. Traditional methods, including region growing [6], level-set [7] and edge-based [8] method. However, traditional methods require the manual design of features and the manual setting of important parameters. Deep learning methods have superiority in weight learning and model generalization without the manual design of features and setting of parameters. With the development of GPU hardware and the open-source availability of large amounts of medical datasets, deep-learning methods can perform even better in liver and tumor segmentation tasks.

Convolutional neural networks (CNN) have achieved great success in the field of computer vision over the last few years [9]. In the field of image semantic segmentation, a new type of convolutional neural network, the fully convolution network (FCN) [10], has been proposed with the advantage that the input and output images of the network are of the same size and that the input images are all full image of arbitrary resolution. FCN has rapidly gained attention for its outstanding advantages in feature extraction. Compared to natural

* Corresponding author.
E-mail address: cvlab.cqut@hotmail.com (R. Liu).

https://doi.org/10.1016/j.compbiomed.2023.106838
Received 26 September 2022; Received in revised form 8 March 2023; Accepted 26 March 2023
Available online 28 March 2023

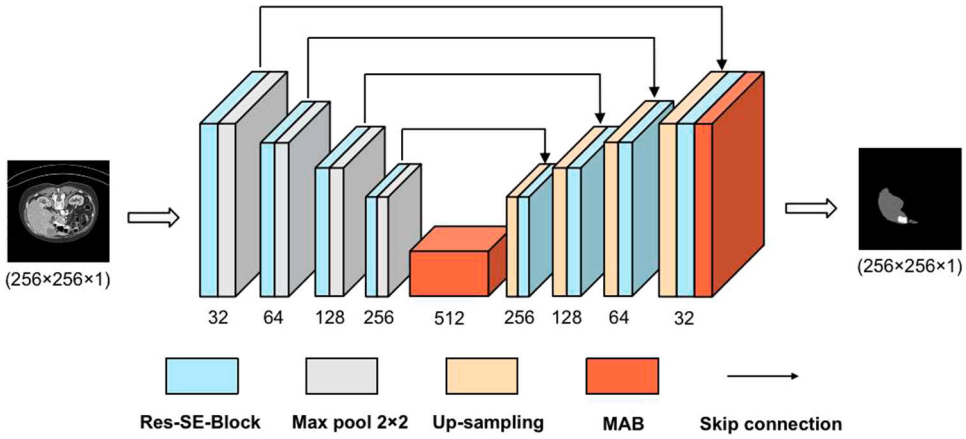

Fig. 1. The architecture of RMAU-Net. RMAU-Net mainly consists of Res-SE-Block, MAB, max-pooling, upsampling and skip connections. The network receives input from CT slices of size (256 × 256 × 1) and directly outputs the mask with the same size. The numbers under the module represent the numbers of channels of feature maps.

image segmentation, the accuracy of medical image segmentation is often affected by factors such as different sizes, shapes, and locations of lesioned regions and the low-intensity contrast between organs within the patient. In addition, accurate segmentation is regarded as an extremely complex task due to the blurred boundaries of the lesions. Consequently, some deep learning methods for medical image segmentation have been developed to overcome the aforementioned factors. U-Net [11] is one of the most widely used encoder–decoder networks in the field of medical image segmentation. U-Net employs skip connections from the lower to the higher layer to exploit multi-scale feature information and make up for information lost through down-sampling. The addition of skip connections significantly improves the utilization of information and the accuracy of segmentation. Inspired by U-Net, a number of variants based on the U-Net have been developed, including Attention U-Net [12], 3D U-Net [13], and U-Net++ [14]. Attention U-Net presents an attention mechanism that focuses on the goal and suppresses background features and eliminates irrelevant information and noise by inserting the attention module before the splicing of encoder and decoder features. To recover 3D information, 3D U-Net converts all 3D convolution operations in U-Net to 3D convolution operations. As stated in [9,15], 3D methods include a more complicated network in comparison with 2D methods, resulting in a higher computation load and lower computation efficiency. Therefore, it is practical to apply 2D segmentation methods in scenarios where the computation resource is limited and a high computation efficiency is required U-Net++, which is based on improved U-Net, enhances performance by combining four structures of different depths and dense skip connections of different lengths. In addition, novel skip connections, such as residual connection [16] and dense connection [17], have been introduced to the network architecture. While the variances of the proposed skip connections are helpful to capture the rich semantic feature information of different levels and reduce the semantic gap, it fails to describe the channel-wise dependencies and spatial-wise relationships between the pixels of images that are crucial for medical image segmentation.

Conventional CNNs perform poorly in global information modeling and multi-scale feature extraction. In liver and tumor segmentation tasks, the lesions have various sizes, shapes, locations, and numbers between patients, and even the same patient, which can cause a significant challenge to automatic segmentation. In addition, edge segmentation is usually performed unsatisfactorily because of the lack of clear boundaries of some lesions, combined with severe noise. To capture multi-scale features in the network, some state-of-the-art methods [18–20] introduce atrous convolutions and pooling operations with different sampling rates. But, the pooling operations and atrous convolutions are unable to take advantage of the channel-wise and spatial-wise dependencies in the global information. Additionally, pooling operations induce to loss of detailed information on the feature map.

To deal with the above issues, we provide Redial Multi-scale Attention U-Net (RMAU-Net) for liver and tumor segmentation, as seen in Fig. 1. The RMAU-Net includes two main modules, Res-SE-Block and Multi-scale Attention Block (MAB). The Res-SE-Block instead of the origin two convolution layer is used to mitigate the problem of gradient disappearance by residual connection and enhance the quality of representations by squeeze-and-excitation operations. The MAB can capture both multi-scale feature information and channel-wise and spatial-wise dependencies. In addition, we design a hybrid loss function that combined Focal loss and dice loss to solve the imbalance of class and poor segmentation of difficult samples.

In summary, this study makes the following contributions.

- We propose a new network architecture named RMAU-Net to augment the ability of feature representation and improve performance on liver and tumor segmentation tasks.
- We develop a multi-scale attention mechanism, which can exploit global spatial information and channel dependencies and solve the multi-scale problem efficiently and effectively.
- We redesign a novel hybrid loss function based on a combination of dice loss and focal loss in order to solve the imbalance of class and poor segmentation of difficult samples.

The remainder of the paper is organized as follows. Section 2 describes the previous research and related work. Section 3 discusses the proposed method in detail. Section 4 describes the experiment, and Section 5 analyzes the results. Finally, Section 6 provides conclusions of this study.

2. Related work

2.1. Liver and tumor segmentation

In recent years, many researchers have proposed many approaches based on convolutional neural networks for liver and tumor segmentation. These approaches are mainly classified as 2D networks and 3D networks. Sun et al. [21] proposed a multi-channel fully convolutional network (MC-FCNs), which is used to automatically segment liver and tumors in CT scans. The MC-FCNs has three channels, which can be trained separately and independently for different stages of CT images, and feature fusion is performed at the higher levels of the network. In order to obtain multi-scale feature maps for live and tumor segmentation, Song et al. [22] proposed a bottleneck supervised U-Net (BS U-Net) with convolutional kernels of different sizes. The

BS U-Net adds a dense module, an inception module, and a dilated convolution to the encoder of the network based on the original U-Net. Liu et al. [23] proposed a GIU-Net that integrates the graph cut method to the improved U-Net. Kaur et al. [24] proposed a GA-UNet for 2D and 3D image segmentation, respectively. To improve the resolution of the output image, the 3D GA-UNet eliminates the effect of the network's shrinkage path on resolution through successive layers and replaces the pooling operation with an upsampling operation. Jin et al. [25] proposed an RA-UNet for liver and tumor segmentation, replacing the convolutional blocks of the traditional U-Net with residual blocks, and proposing an attention mechanism and using it to combine low-level feature maps with high-level feature maps to extract contextual information by skip connections. Gao et al. [26] proposed an ASU-Net++ to improve gradient flow and feature retention by integrating dense skip connections. In addition, ASU-Net++ based U-Net++ modified the original Atrous Spatial Pyramidal Pooling (ASPP) to an adaptive pooling structure for better performance and compatibility. Zhang et al. [27] proposed the Scale Attention mechanism, which is effective for multi-scale problems in liver and tumor segmentation. Kushnure et al. [15] proposed HFRU-Net which modifies skip connections by using a feature fusion mechanism and local feature reconstruction.

2.2. Multi-scale feature extraction

In the past few years, many approaches [28–35] have been proposed to enhance context aggregation by applying multi-scale feature information of images due to that multi-scale feature information can provide rich semantic features for medical image segmentation. We review several approaches about multi-scale feature extraction.

It is well known that FPN [28] was the first work to address the problem of multi-scale feature extraction. Many studies explored multi-scale feature extraction in the field of biomedical image segmentation, such as CE-Net [29], U-Net++ [30], MDAN-UNet [31] and MS-UNet [32]. Gu et al. [29] designed a pooling strategy with pool kernels of different sizes for medical image segmentation. Zhou et al. [30] proposed a network called U-Net++, which applies dense and nested skip connections to connect encoders and decoders before fusing them with the corresponding semantic from different layers of the encoder network progressively enriched before fusing with rich multi-scale feature mappings. Liu et al. [31] proposed an improved nested U-Net (MDAN-UNet) for the automatic segmentation of OCT images, which takes advantage of a dual attention mechanism, and multi-scale feature extraction. Kushnure et al. [32] introduced the MS-UNet which utilized the multi-scale approach to improve the receptive field of CNN and extract global and local features. Huang et al. [33] developed a novel variant of U-Net named Unet3+ that utilizes deep supervision and full skip connections for organ segmentation. Khan et al. [35] presented the RMS-Unet with a multi-scale approach, which has been utilized to explore novel inter-slice features with multi-channel input images.

2.3. Attention mechanism

Attention mechanisms are inspired by the biological systems of humans that tend to focus on the area of interest when processing large amounts of information. At the same time, attention mechanisms can capture long-range dependencies. Attention mechanisms were first applied to natural language processing tasks. [36] is regarded as the first work to use attention mechanisms to capture the global dependence of inputs. In recent years, attention mechanisms have been widely used in computer vision tasks, as researchers have discovered that attention mechanisms also perform well in computer vision tasks. There are various attention mechanisms have been proposed for computer vision tasks. Hu et al. [37] firstly presented the concept of channel attention that focus on the channel dependencies between feature maps, and proposed a novel architecture unit named Squeeze-and-Excitation block (SE) that can adaptively recalibrate channel dependencies responses of feature maps by explicitly modeling the interdependencies between channels. Woo et al. [38] proposed a novel attention module (CBAM) in convolutional neural networks, which can calculate the attention graph sequentially along channel-wise and spatial-wise. Meanwhile, Park et al. [39] proposed the bottleneck attention module (BAM) that can be integrated with a convolutional neural network that can infer the attention graph along two different paths, channel, and space. Both CBAM [38] and BAM [39] use convolutional operations with large kernels that operations to compute local dependencies. Wang et al. [40] proposed channel attention (CA) block and inserted it into the skip connections between encoder and decoder to enhance performance on medical image segmentation. Zhang et al. [41] introduced the Scale Attention and the Axis Attention mechanisms, which are efficient to capture essential information in global pooling.

To date, extensive segmentation methods based on 2D data have been developed to improve the segmentation accuracy of liver and tumors. Notwithstanding, most of the existing methods ignore the multi-scale feature information and spatial channel information, which results in relatively low segmentation accuracy. In this paper, we design multi-scale attention units for accurate segmentation of liver edges by capturing global feature information, and combine the ideas of spatial attention mechanism and channel attention mechanism to enhance the segmentation of small tumors.

3. Methodology

In this section, we describe the details of RMAU-Net, including network architecture, Res-SE-Block, MAB module, and loss function.

3.1. RMAU-Net

The RMAU-Net architecture is shown in Fig. 1, which is improved on the encoder–decoder architecture of U-Net. The original U-Net based on encoder–decoder architecture uses four downsampling layers to obtain high-dimensional information and utilizes four upsampling layers to restore the feature map to the original size, then fuse the high-level features and low-level features by four skip connections. But U-Net can poorly capture multi-scale feature information and focus on more important features. To solve these issue of U-Net, we modify its basic deep learning neural network with PyTorch as the backend. Different from two convolutional layers in U-Net, we develop Res-SE-Block which consists of two 3×3 convolutional blocks, residual block, and squeeze-excitation block to extract high-dimensional feature information. In addition, we insert the MAB module at the end of the encoder and the decoder respectively to increase the receptive fields of the network and compensate for the loss of information by max pooling, and the MAB module can capture multi-scale feature information and spatial-wise and channel-wise dependencies of feature maps (see Fig. 5).

3.2. Res-SE-Block

Our proposed Res-SE-Block is built upon residual block and squeeze-excitation block, as shown in Fig. 2. With the increase in the number of network layers, CNNs will suffer from gradient disappearance and gradient explosion, which cause CNNs to fail to converge. [16] designed a novel skip connection termed residual connection to deal with the problem. Inspired by residual connections, we incorporate residual modules into our network. In the residual connection path, we use 1×1 Conv to control the number of output channels. Inspired by SE-Net [37], we consider focusing on the importance of each channel in the feature map through squeeze-excitation operation to solve the problem of channel dependence.

In addition, to further prevent the gradient from vanishing, we use Leaky ReLU as the activation function instead of ReLU. Because of the zero slopes of the ReLU activation, the ReLU may stay in a state where it can only output zero. We avoid this problem by using a leaky ReLU

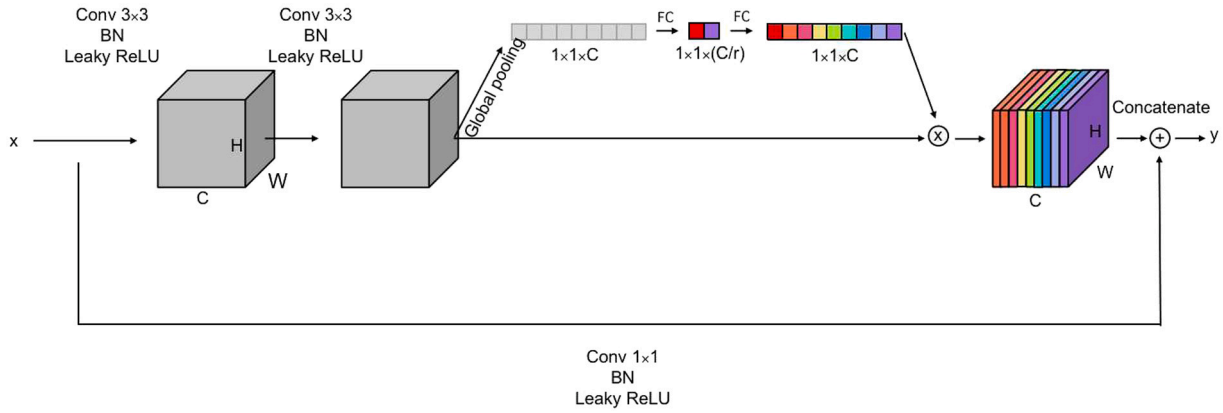

Fig. 2. The diagram of Res-SE-Block. Our proposed Res-SE-Block mainly consists of squeeze-excitation block, residual connection, convolution and Leaky ReLU.

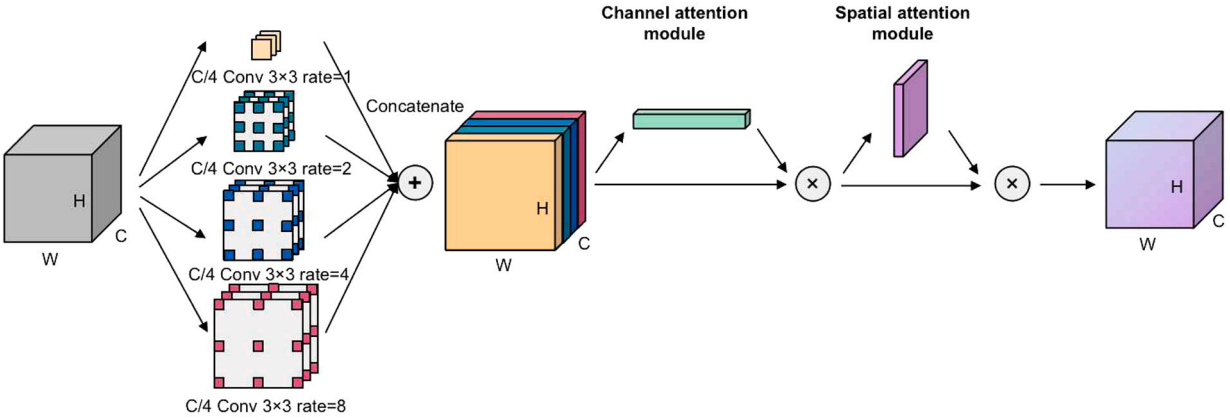

Fig. 3. The overview of MAB. Our proposed MAB mainly consists of ASPP module, Channel Attention module, and Spatial Attention module. ASSP module can integrate multi-scale features by atrous convolution with different dilated rates. The Channel module and Spatial module generate channel attention vectors and spatial attention vectors respectively.

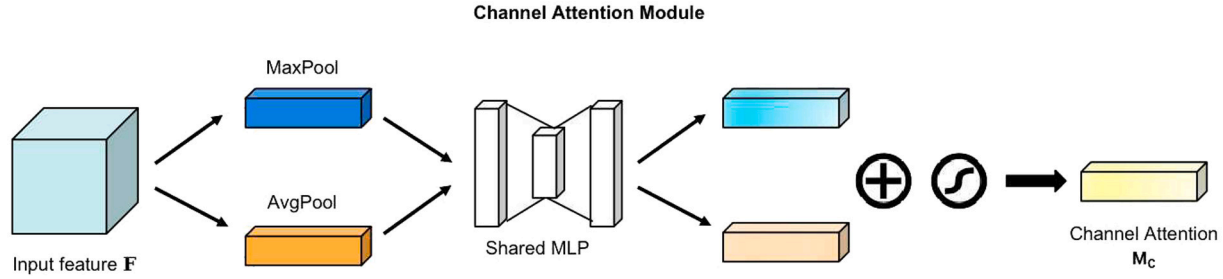

Fig. 4. The diagram of the Channel module. The Channel module utilizes average-pooling and max-pooling to aggregate spatial information of a feature map and apply a shared network with one hidden layer. Then, we get the channel attention map by merging the output feature vectors.

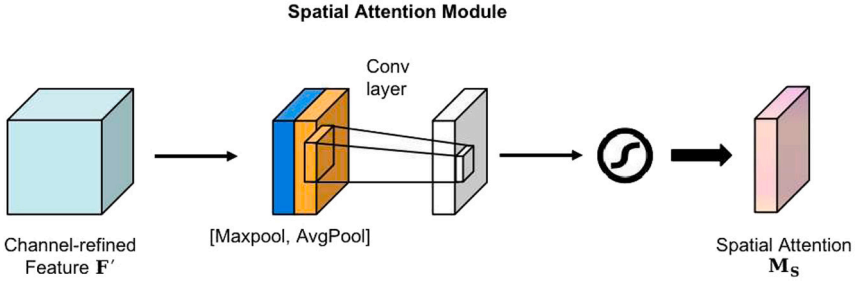

Fig. 5. The diagram of the Spatial module. The Spatial module aggregates channel information by using average-pooling and max-pooling. The outputs of two pooling operations are concatenated and computed by a convolutional layer, producing a spatial attention map.

rather than a ReLU. Batch normalization is also added to each block to prevent over-fitting and improve the gradient flow and thus facilitate the convergence of the network.

The squeeze operation is achieved by using global average pooling to squeeze global spatial information into a channel descriptor. Mathematically, the squeeze operation can be written as:

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_c(i,j), \tag{1}$$

where z_c refers to the data after squeeze operation, F_{sq} denotes the squeeze function, u_c refers to the input, H refers to the height of the feature map, W refers to the width of the feature map.

The purpose of the excitation operation is to exploit the aggregated information in the squeeze operation and fully capture channel-wise dependencies through two fully connected (FC) layers. The first FC layer reduces the channel dimension with the reduction ratio of r = 6 to limit the complexity of the network. The second FC layer restores the dimension to the original dimension after applying the ReLU activation function. Mathematically, the excitation operation can be written as:

$$s = F_{ex}(z, W) = \sigma\left(W_2 \delta\left(W_1 z\right)\right), \tag{2}$$

where s refers to the result after excitation operation, F_{ex} refers to the excitation function, σ refers to the sigmoid function, δ refers to the ReLU function, $W_1 \in R^{\frac{C}{r} \times C}$ and $W_2 \in R^{C \times \frac{C}{r}}$

The squeeze operation and excitation operation can enhance the quality of representations by explicitly modeling the interdependencies between the channels of its convolutional features, it also allows the network to perform feature recalibration through which it can learn to use global information to selectively emphasize informative features and suppress less useful ones. The Res-SE-Block module can significantly improve the accuracy of the liver and tumor segmentation while slightly increasing computation time and the complexity of the model. In addition, it is easier to integrate into other networks compared with other attention mechanisms. Finally, the function of the Res-SE-Block is represented as the following equation:

$$y = x + u \times s, \tag{3}$$

where x refers to the input of Res-SE-Block, y refers to the output of Res-SE-Block.

3.3. MAB

The MAB module takes advantage of ASPP [42] combined with the attention mechanism module and is designed to exploit multi-scale feature information and capture channel dependencies and spatial information between pixels, as shown in Fig. 3. The idea for ASPP comes from the spatial pyramid pooling [43], which successfully resamples features at multiple scales. ASPP module incorporates many parallel atrous convolutions with different dilated rates to capture contextual information at different scales from feature maps. In addition, atrous convolution can control the receptive field of CNNs to accurately capture multi-scale feature information [44]. Thanks to its multi-scale extraction function, the ASPP model shows satisfactory results on various segmentation tasks. Therefore, we use ASPP to capture useful multi-scale information in the liver and tumor segmentation task.

The attention mechanism module includes two attention modules, the channel attention module and spatial attention module, which focus on 'what' and 'where' respectively by computing complementary attention, as shown in 4.

The channel attention module exploits the inter-channel relationship of features and focuses on 'what' is meaningful. The first step is to aggregate spatial information of the feature map, average-pooling and max-pooling operations are used to generate two different spatial context features: F^c_{avg} and F^c_{max}. The two features are forwarded to a shared network which is composed of a multi-layer perceptron (MLP) to generate a channel attention map $M_c \in R^{C \times 1 \times 1}$. After applying the shared network to each feature, the output feature vectors are merged by using element-wise summation. In brief, the channel attention is computed as:

$$\begin{aligned} M_c(F) = \sigma\,(&MLP(AvgPool(F)) \\ &+ MLP(MaxPool(F))), \end{aligned} \tag{4}$$

where σ denotes the sigmoid function, and F is the input.

The spatial attention module utilizes the inter-spatial relationship of features and focuses on 'where' is an informative part. As the same as the channel attention, the first step is applying average-pooling and max-pooling operations to generate a highlighting feature map efficiently. And then we apply a convolution layer to produce a spatial attention map $M_s \in R^{C \times 1 \times 1}$. In brief, the spatial attention is computed as:

$$M_s(F) = \sigma\left(f^{7\times7}([AvgPool(F); MaxPool(F)])\right), \tag{5}$$

where σ denotes the sigmoid function and $f^{7\times7}$ represents a convolution layer with the kernel size of 7×7.

3.4. Loss function

The cross-entropy (CE) loss is one of the most widely used loss functions for deep learning models. The loss function based on the dice coefficient can alleviate the problem of imbalance between background and foreground pixels. Therefore, most of the previous work optimizes the network by combining the cross-entropy loss function and the dice coefficient loss function in the model to obtain a weighted loss function. Inspired by [45], considering the imbalance with the numbers of liver and tumor in the task of liver and tumor segmentation, and the liver and tumors are more complex and diverse, with the tumors having a smaller and more blurred shape, we propose a novel loss function combined with dice coefficient loss and the Focal loss [46] which deal with the issue of class imbalance by reducing the weight the contribution of easy examples and more focusing on harder examples.

The cross-entropy (CE) loss is defined as the following:

$$\mathrm{CE}(p, y) = \left\{ \begin{array}{ll} -\log(p), & \text{if } y = 1 \\ -\log(1-p), & \text{if } y = 0 \end{array} \right\}, \tag{6}$$

where $y \in 0, 1$ denotes the ground-truth class and $p \in [0, 1]$ refers to the predict probability for the class with label $y = 1$. Then, we define p_t as:

$$p_t = \left\{ \begin{array}{ll} p, & \text{if } y = 1 \\ 1-p, & \text{if } y = 0 \end{array} \right\}, \tag{7}$$

and rewrite $\mathrm{CE}(p, y) = \mathrm{CE}(p_t) = -\log(p_t)$.

We define the Focal loss($\mathscr{L}_{\mathrm{F}}$) as:

$$\mathscr{L}_{\mathrm{F}p_t} = \alpha_t \left(1 - p_t\right)^{\gamma} \cdot \mathscr{L}_{\mathrm{CE}(p,y)}, \tag{8}$$

where α_t controls the weights of the class, γ denotes the weight of easy samples.

The dice loss function can be written as:

$$\mathscr{L}_{dice} = 1 - \frac{2 \times \sum_{i=1}^{N} p_i y_i}{\sum_{i=1}^{N} {p_i}^2 + \sum_{i=1}^{N} {y_i}^2}, \tag{9}$$

where N indicates the number of all predicted voxels. p_i represents the predict probability that the voxel i, y_i denotes the voxel i in the ground truth.

In this study, the final loss function of the proposed method is shown as:

$$\mathscr{L}_{total} = \alpha \mathscr{L}_F + \beta \mathscr{L}_{dice}, \tag{10}$$

where α and β are to control the weight of Focal loss and Dice loss. In this study, $\alpha = 0.5$ and $\beta = 1$.

Table 1
The quantitative comparison of preprocessing based on LiTS dataset. (Without data augmentation).

Model	Liver			Tumor		
	DSC	VOE	RVD	DSC	VOE	RVD
U-Net (Without Pre)	0.8950 ± 0.026	0.1569 ± 0.039	0.1313 ± 0.368	0.5553 ± 0.068	0.5822 ± 0.058	0.8517 ± 0.417
U-Net (With Pre)	**0.9168 ± 0.010**	**0.1371 ± 0.019**	**−0.1028 ± 0.021**	**0.6063 ± 0.060**	**0.5348 ± 0.055**	**−0.1631 ± 0.593**
Ours (Without Pre)	0.9497 ± 0.012	0.0874 ± 0.008	−0.0083 ± 0.012	0.7547 ± 0.053	0.3940 ± 0.055	−0.0449 ± 0.764
Ours (With Pre)	**0.9521 ± 0.010**	**0.0826 ± 0.007**	**−0.0057 ± 0.073**	**0.7594 ± 0.075**	**0.3837 ± 0.072**	**−0.0235 ± 0.393**

Note. Pre means preprocessing.

Table 2
The quantitative comparison of data augmentation based on LiTS dataset. (With preprocessing).

Model	Liver			Tumor		
	DSC	VOE	RVD	DSC	VOE	RVD
U-Net (Without Aug)	0.9168 ± 0.010	0.1371 ± 0.019	−0.1028 ± 0.021	0.6063 ± 0.060	0.5348 ± 0.055	−0.1631 ± 0.593
U-Net (With Aug)	**0.9221 ± 0.027**	**0.1270 ± 0.047**	**−0.0725 ± 0.047**	**0.6252 ± 0.148**	**0.5209 ± 0.159**	**0.0946 ± 0.437**
Ours (Without Aug)	0.9521 ± 0.010	0.0826 ± 0.007	−0.0057 ± 0.073	0.7594 ± 0.075	0.3837 ± 0.072	−0.0235 ± 0.393
Ours (With Aug)	**0.9552 ± 0.012**	**0.0792 ± 0.023**	**−0.0042 ± 0.026**	**0.7616 ± 0.118**	**0.3709 ± 0.135**	**0.0118 ± 0.291**

Note. Aug means Augmentation.

Table 3
Experimental hardware and software configuration.

Environment	Configuration Information
GPU	RTX3090
Memory	64G
Operating system	Ubuntu 18.04
Hard disk	4TB
Programming Software	PyTorch 1.7; Python 3.7

Table 4
Training hyperparameters of the proposed method.

Hyperparameters	Setting
Learning rate	0.0001
Batch_size	8
Epoch	250
Optimizer	Adam

4. Experiments AND results

4.1. Datasets

4.1.1. LiTS

The LiTS dataset is the public dataset that comes from the Liver tumor segmentation challenge held by ISBI 2017 and MICCAI 2017, and is currently the most commonly used dataset for liver and tumor segmentation studies. The LiTS dataset consists of a training set of 131 CT scans and a test set of 70 CT scans. The number of CT slices included in each scan ranged from 42 to 1026, with an axial plane resolution of 512 × 512 pixels and slice spacing ranging from 0.45 mm to 6.0 mm. The training data set was manually labeled by four radiologists at six clinical sites around the world, while the test set was not labeled. The uniqueness of the data source makes it difficult to segment tumor using this dataset due to the significant variation in reconstructed layer thickness, slice thickness, storage orientation of scanned images, image quality, and spatial resolution. However, due to its relatively large number and high image quality, it is still the most widely used dataset for liver and tumor segmentation.

4.1.2. 3D-IRCADb

The 3D Image Reconstruction for Comparison of Algorithm Database (3D-IRCADb) is a publicly available dataset that has been used extensively in related research. This dataset provides more complex data on the liver and its tumors, including medical images of anonymous patients and images of the region of interest manually segmented by clinical experts. The 3D-IRCADb-01 consists of enhanced CT scans of 10 women and 10 men, 75% of women had liver tumors, while the 3D-IRCADb-02 consists of 2 anonymous enhanced 3D CT scans of the chest and abdomen. The resolution of this dataset is also 512 × 512 pixels, but some of the livers and tumors in the dataset have low contrast, and the liver and tumor areas almost overlap, which may affect the model training to some extent, and thus the segmentation results.

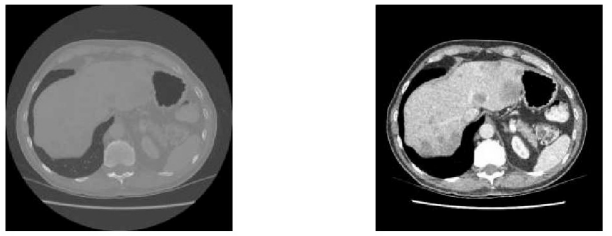

Fig. 6. The comparison of the original and processed images.

4.2. Preprocessing and augmentation

In this paper, we train, test our proposed model on the LiTS dataset, and evaluate its generalization ability on the 3D-IRCADb dataset. As for the LiTS dataset, in general, data splitting is randomly performed based on the training set, since image labels of the test set are not publicly available. In this work, we used volumes 0–26 of the training set as the test subset and volumes 27–130 as the training subset, which can facilitate readers to reproduce our method and experimental results. And we obtained 19,080 2D images by slicing 131 volumes, which includes 11,893 normal images and 7187 cancer images. In order to make the CT slices beneficial for network training, the raw liver CT images need to be preprocessed by using windowing techniques [9,17]. For preprocessing of the dataset, windows were opened in the range of Hounsfield's unit value [−200, 200] to remove other irrelevant tissues and enhance the contrast between the liver and other tissues, then the voxel values were normalized to [−1, 1]. Finally, the images are normalized before being processed. Fig. 6 shows a comparison between the original CT slice and the preprocessed CT slice. It can be seen that the area of the liver is more visible and has a clearer texture and contour after preprocessing. In addition, we accomplish a comparative experiment on the effectiveness of the preprocessing methods. The result is shown in Table 1.

The original CT images are 512 × 512 pixels in size and have been cropped to 256 × 256 pixels to accelerate the training of the network

Table 5
Ablation analysis for the proposed method on LiTS dataset.

Model	Liver			Tumor		
	DSC	VOE	RVD	DSC	VOE	RVD
U-Net(backbone)	0.9221 ± 0.027	0.1270 ± 0.047	−0.0725 ± 0.047	0.6252 ± 0.148	0.5209 ± 0.159	0.0946 ± 0.437
U-Net+Res-SE-Block	0.9394 ± 0.016	0.1116 ± 0.029	−0.0444 ± 0.035	0.6906 ± 0.155	0.4578 ± 0.158	−0.0277 ± 0.240
U-Net+MAB	0.9472 ± 0.032	0.0981 ± 0.048	0.0164 ± 0.048	0.7257 ± 0.139	0.4140 ± 0.154	−0.1126 ± 0.102
RMAU-Net	**0.9552 ± 0.012**	**0.0792 ± 0.023**	**−0.0042 ± 0.026**	**0.7616 ± 0.118**	**0.3709 ± 0.135**	**0.0118 ± 0.291**

Table 6
Ablation analysis for Loss function based on RMAU-Net on LiTS dataset.

LOSS FUNCTION	Liver			Tumor		
	DSC	VOE	RVD	DSC	VOE	RVD
CE loss	0.9471 ± 0.011	0.0872 ± 0.023	0.0075 ± 0.038	0.7415 ± 0.149	0.4008 ± 0.154	−0.1081 ± 0.216
CE loss + dice loss	0.9526 ± 0.014	0.0864 ± 0.025	−0.0058 ± 0.032	0.7494 ± 0.178	0.3892 ± 0.177	−0.0996 ± 0.239
Proposed loss	**0.9552 ± 0.012**	**0.0792 ± 0.023**	**−0.0042 ± 0.026**	**0.7616 ± 0.118**	**0.3709 ± 0.135**	**0.0118 ± 0.291**

and reduce the region of the background. The role of data augmentation is to increase the generalization ability and robustness of the model and to avoid model overfitting. We augmented data by using the tool named Transforms of the PyTorch deep learning framework following the other methods [26,27]. The data augmentation methods used include: (1) scaling the image between 0.8 and 1.2 with a 50% probability, (2) rotating the image between 0 degrees and 30 degrees with a 30% probability, (3) horizontal and vertical flipping with a 30% probability. In addition, we test the effectiveness of the data augmentation methods on the backbone(U-Net) and our proposed model. The results are shown in Table 2. The experimental results demonstrate that the data augmentation method can improve the performance of the model by increasing the diversity of samples.

4.3. Experimental environment and parameters

The experimental hardware and software configuration for this study are shown in Table 3. The settings of training hyperparameters are shown in Table 4.

4.4. Evaluation metrics

The common evaluation metrics for liver and tumor segmentation include Dice similarity coefficient (DSC), volume overlap error (VOE), and relative volume difference (RVD). If A denotes the ground truth(GT) and B denotes the predicted results(PR), the relevant evaluation metrics are as follows:

(1) DSC: the most common evaluation metric for image segmentation and represents the overlapping similarity between GT and PR. The calculation formula is as follows:

$$\mathrm{DSC}(A,B)=\frac{2|A\cap B|}{|A|+|B|}. \tag{11}$$

(2) VOE: similar to the DSC and is the ratio between intersection and union between segmentation results and markers. The VOE is the error rate of the segmentation, which is in the range of 0 to 1. The calculation formula is as follows:

$$\mathrm{VOE}(A,B)=1-\frac{|A\cap B|}{|A\cup B|}. \tag{12}$$

(3) RCD: used to express the relative difference in volume between GT and PR. The formula is as follows:

$$\mathrm{RVD}(A,B)=\frac{|B|-|A|}{|A|}. \tag{13}$$

4.5. Ablation analysis

To evaluate the effectiveness of our proposed RMAU-Net for liver and tumor segmentation, we conduct comprehensive experiments with ablation analysis. We present the Res-SE-Block instead of all the original two convolution layers based on U-Net and add the MAB module after the encoder and decoder respectively. We mainly evaluate the effectiveness of the Res-SE-Block and the MAB module. The result of ablation analysis for RMAU-Net is shown in Table 5. It demonstrates that the Res-SE-Block and the MAB module are effective to improve the segmentation performance, and the MAB gets a greater segmentation effect than Res-SE-Block. Then to evaluate the effectiveness of the proposed loss function, we train the RMAU-Net with CE loss, CE loss + dice loss, and the proposed loss respectively. The result of ablation analysis for loss function is shown in Table 6. It demonstrates that the proposed loss function is also beneficial to enhance the performance of the liver and tumor segmentation.

4.6. Comparison of models

In this section, we compare the proposed RMAU-Net with five state-of-the-art approaches to evaluate the effectiveness and robustness of RMAU-Net on the LiTS dataset and 3D-IRCADb dataset. In the medical image segmentation tasks, U-Net [11] is the most classic network of encoder–decoder architecture with skip connections. U-Net++ [14] adds a series of nested, dense skip connections based on U-net to reduce the semantic gap between the feature maps. RA-UNet [25] introduces the attention residual mechanism to improve the performance of U-Net. ASU-Net++ [26] introduces the adaptive feature extractions for liver and tumor segmentation. SAA-Net [27] proposes the Scale Attention mechanism based on U-Net. HFRU-Net [15] modifies skip paths by applying feature fusion mechanism and local feature reconstruction to improve U-Net. The results of the comparison of six methods on the LiTS dataset are shown in Table 7. The result demonstrates that the proposed RMAU-Net achieved better performance than the other methods. The proposed method achieves 0.9522 (DSC) and 0.7616 (DSC) on liver and tumor segmentation by measuring dice values respectively. In addition, we show six CT slices that contain the liver and tumor to visualize the segmentation results. The visual comparison of the output of different models is shown in Fig. 7. We can see that U-Net does not work well in liver and tumor segmentation, and our proposed RMAU-Net obtains better performance on both liver segmentation and tumor segmentation. To evaluate the validity and robustness of our proposed RMAU-Net, we also conducted experiments on the 3D-IRCADb dataset. The result of comparison and visual comparison on the 3D-IRCADb dataset is shown in Table 8 and Fig. 8 respectively. We can see that

Table 7
The quantitative comparison of different methods on LiTS dataset (27 test volumes).

Model	Liver			Tumor		
	DSC	VOE	RVD	DSC	VOE	RVD
U-Net [11]	0.9221 ± 0.027	0.1270 ± 0.047	−0.0725 ± 0.047	0.6252 ± 0.148	0.5209 ± 0.159	0.0946 ± 0.437
U-Net++[14]	0.9399 ± 0.026	0.1092 ± 0.045	0.0220 ± 0.058	0.6903 ± 0.137	0.4629 ± 0.150	0.2334 ± 0.661
RA-UNet [25]	0.9450 ± 0.018	0.0992 ± 0.037	0.0152 ± 0.037	0.7021 ± 0.135	0.4427 ± 0.145	0.1370 ± 0.431
ASU-Net [26]	0.9494 ± 0.012	0.0915 ± 0.021	0.0129 ± 0.029	0.7198 ± 0.133	0.4311 ± 0.142	−0.2661 ± 0.179
SAA-Net [27]	0.9538 ± 0.014	0.0861 ± 0.025	0.0136 ± 0.029	0.7312 ± 0.133	0.4088 ± 0.142	−0.2218 ± 0.179
HFRU-Net [15]	0.9503 ± 0.014	0.0911 ± 0.025	−0.0141 ± 0.032	0.7494 ± 0.107	0.3802 ± 0.128	−0.2181 ± 0.152
RMAU-Net	**0.9552 ± 0.012**	**0.0792 ± 0.023**	**−0.0042 ± 0.026**	**0.7616 ± 0.118**	**0.3709 ± 0.135**	**0.0118 ± 0.291**

Table 8
The quantitative comparison of different methods on 3D-IRCADb dataset.

Model	Liver			Tumor		
	DSC	VOE	RVD	DSC	VOE	RVD
U-Net [11]	0.9458 ± 0.013	0.0743 ± 0.023	−0.0081 ± 0.024	0.5063 ± 0.153	0.6201 ± 0.139	1.6562 ± 1.246
U-Net++[14]	0.9647 ± 0.022	0.0627 ± 0.039	0.0102 ± 0.052	0.6145 ± 0.140	0.5145 ± 0.153	1.0971 ± 0.864
RA-UNet [25]	0.9527 ± 0.011	0.0823 ± 0.030	0.0063 ± 0.027	0.7027 ± 0.131	0.4125 ± 0.161	−0.0461 ± 0.135
ASU-Net [26]	0.9535 ± 0.022	0.0840 ± 0.038	0.0035 ± 0.036	0.7354 ± 0.142	0.3914 ± 0.169	−0.0440 ± 0.255
SAA-Net [27]	0.9552 ± 0.016	0.0814 ± 0.028	0.0293 ± 0.035	0.6782 ± 0.149	0.4595 ± 0.164	0.1145 ± 0.338
HFRU-Net [15]	0.9594 ± 0.013	0.0742 ± 0.023	−0.0029 ± 0.031	0.7894 ± 0.111	0.3257 ± 0.142	0.0327 ± 0.170
RMAU-Net	**0.9697 ± 0.008**	**0.0531 ± 0.014**	**0.0011 ± 0.019**	**0.8307 ± 0.095**	**0.2751 ± 0.125**	**0.1258 ± 0.186**

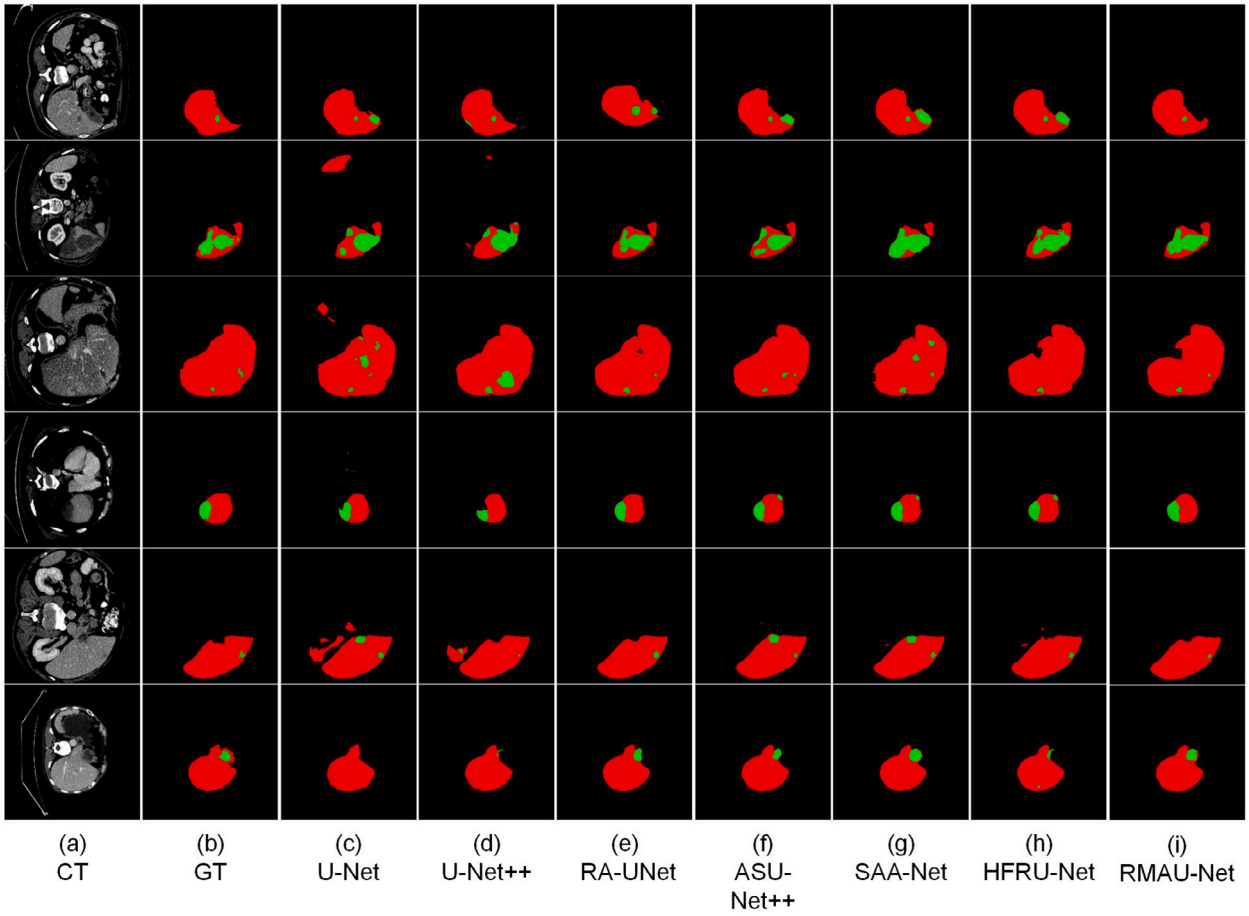

Fig. 7. The visual comparison of the output of different models on LiTS dataset. The red region refers to the livers and the green region refers to the lesions.

our proposed method still outperforms the other methods on the 3D-IRCADb dataset. In addition, We used the t-test method to calculate the p-value on DSC to evaluate the statistical significance. The results in Tables 9 and 10 prove that RMAU-Net has a statistically significant improvement in DSC (all p-value < 0.01). The experimental comparison validated the superiority of our proposed method in comparison with other methods.

In addition, we conducted a 5-fold cross-validation on both LiTS and 3D-IRCADb to test if there is any bias toward the training data resulting from the fixed test data. Specifically, in line with the 5-fold cross-validation rules, each dataset is evenly divided into five groups where at each time, one group of the data is selected exclusively and used as the test data, and the remaining four groups of data are used for training the model. This procedure is repeated five times until each

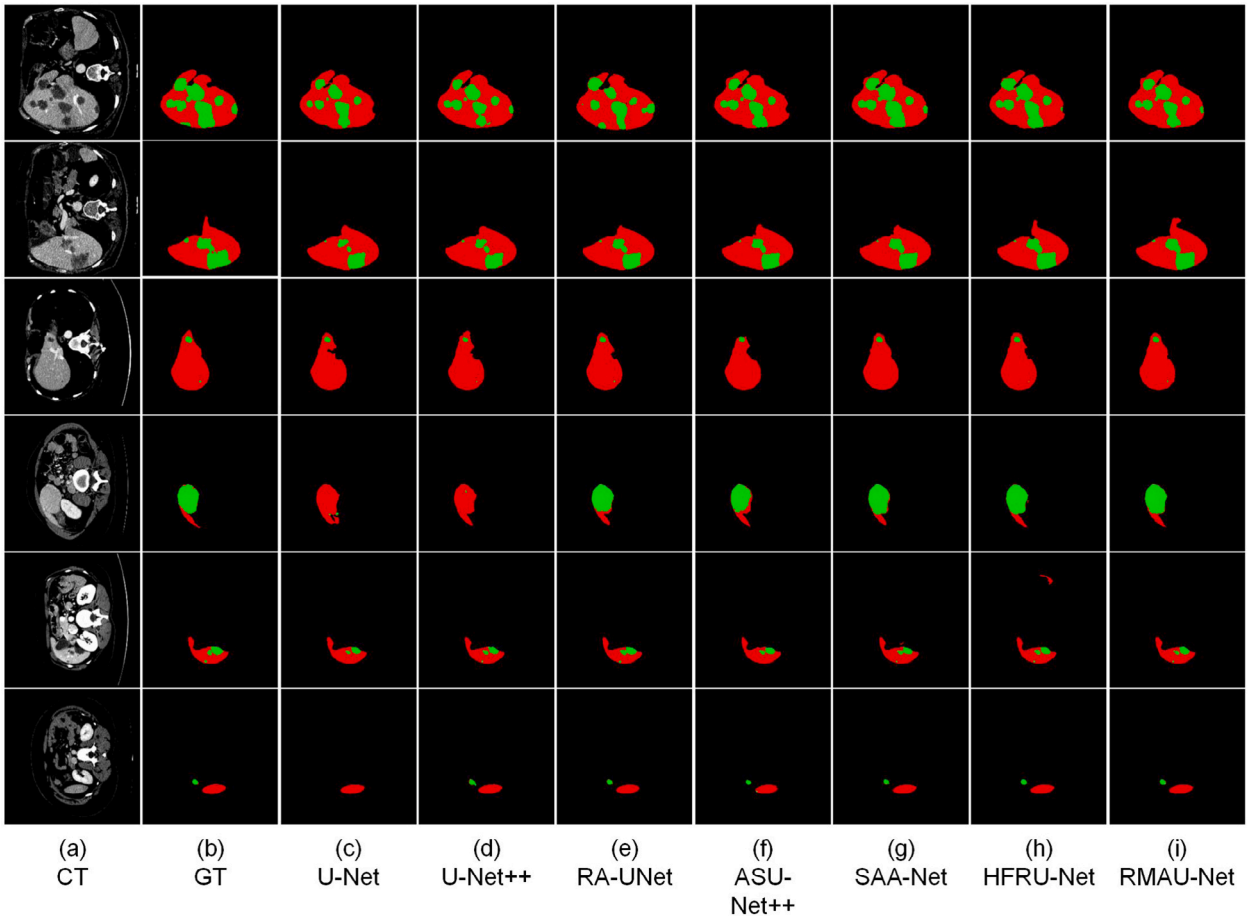

Fig. 8. The visual comparison of the output of different models on 3D-IRCADb dataset. The red region refers to the livers and the green region refers to the lesions.

Table 9
Statistical analysis for t-test method on LiTS dataset.

Model	Liver		Tumor	
	DSC	p-value (DSC)	DSC	p-value (DSC)
U-Net [11]	0.9221 ± 0.027	1.16×10^{-20}	0.6252 ± 0.148	5.54×10^{-16}
U-Net++[14]	0.9399 ± 0.026	7.15×10^{-12}	0.6903 ± 0.137	9.21×10^{-12}
RA-UNet [25]	0.9450 ± 0.018	2.16×10^{-8}	0.7021 ± 0.135	3.52×10^{-10}
ASU-Net [26]	0.9494 ± 0.012	1.23×10^{-6}	0.7198 ± 0.133	7.81×10^{-9}
SAA-Net [27]	0.9538 ± 0.014	1.78×10^{-3}	0.7312 ± 0.133	6.93×10^{-7}
HFRU-Net [15]	0.9503 ± 0.014	2.52×10^{-4}	0.7494 ± 0.107	3.25×10^{-6}
RMAU-Net	**0.9552 ± 0.012**		**0.7616 ± 0.118**	

Table 10
Statistical analysis for t-test method on 3D-IRCADb dataset.

Model	Liver		Tumor	
	DSC	p-value (DSC)	DSC	p-value (DSC)
U-Net [11]	0.9458 ± 0.013	8.41×10^{-16}	0.5063 ± 0.153	4.94×10^{-26}
U-Net++ [14]	0.9647 ± 0.022	5.07×10^{-3}	0.6145 ± 0.140	1.56×10^{-20}
RA-UNet [25]	0.9527 ± 0.011	1.21×10^{-8}	0.7027 ± 0.131	3.51×10^{-12}
ASU-Net [26]	0.9535 ± 0.022	2.13×10^{-8}	0.7354 ± 0.142	9.21×10^{-8}
SAA-Net [27]	0.9552 ± 0.016	9.04×10^{-6}	0.6782 ± 0.149	3.03×10^{-16}
HFRU-Net [15]	0.9594 ± 0.013	2.31×10^{-4}	0.7894 ± 0.111	7.69×10^{-5}
RMAU-Net	**0.9697 ± 0.008**		**0.8307 ± 0.095**	

group has been used as the test data. Finally, the mean value obtained from the five iterations is employed to evaluate the proposed model.

The experimental results of the 5-fold cross-validation with RMAU-net on the LiTS and 3D-IRCADb datasets are shown in Table 11. Moreover, the cross-validation results performed on the datasets are shown in Table 12 and Table 13, respectively. The experimental results of the cross-validation demonstrate good stability of the developed model and do not show significant bias toward the training set.

5. Discussion

Automatic liver and tumor segmentation assist radiologists and diagnosis in clinical practice. In this paper, we propose RMAU-Net for liver and tumor segmentation, which is based on improved U-Net. We introduce multi-scale feature information extraction and an attention mechanism in the proposed method, which contains Res-SE-Block and MAB modules. The Res-SE-Block can enhance the quality of representations and perform feature recalibration by combining residual connection and squeeze-and-extraction operation. The MAB module can infuse multi-scale feature information and exploit the inter-channel and inter-spatial relationship of features that both benefit liver and tumor segmentation.

To evaluate the effectiveness and robustness of our proposed approaches, we conduct comprehensive experiments with ablation analysis. According to the results of the ablation analysis, the Res-SE-Block, the MAB module, and the proposed loss function are effective for liver and tumor segmentation.

To further demonstrate the superiority of our proposed approaches, we compare RMAU-Net with other state-of-the-art methods. Table 2 shows the results compared to the other methods. We can see that our proposed method outperforms other methods on segmentation performance. The proposed method achieves 0.9522 (DSC) and 0.7616 (DSC) on liver and tumor segmentation by measuring dice values respectively. In addition, we list some visual presentations of the liver and tumor segmentation results, as shown in Fig. 7.

Table 11
The results of five-fold cross-validation experiments with RMAU-Net on LiTS and 3D-IRCADb dataset.

Model	Liver			Tumor		
	DSC	VOE	RVD	DSC	VOE	RVD
LiTS	0.9563	0.0788	−0.0037	0.7623	0.3700	0.0109
	0.9556	0.0784	−0.0030	0.7629	0.3691	0.0113
	0.9566	0.0788	−0.0038	0.7621	0.3702	0.0103
	0.9564	0.0790	−0.0037	0.7622	0.3709	0.0111
	0.9564	0.0789	−0.0037	0.7620	0.3694	0.0104
mean	**0.9563 ± 0.001**	**0.0788 ± 0.001**	**−0.0036 ± 0.001**	**0.7623 ± 0.001**	**0.3699 ± 0.001**	**0.0108 ± 0.001**
3D-IRCADb	0.9703	0.0510	0.0014	0.8304	0.2743	0.1253
	0.9702	0.0523	0.0012	0.8305	0.2742	0.1255
	0.9703	0.0529	0.0014	0.8307	0.2744	0.1254
	0.9707	0.0520	0.0011	0.8304	0.2746	0.1258
	0.9701	0.0504	0.0012	0.8306	0.2747	0.1253
mean	**0.9703 ± 0.001**	**0.0517 ± 0.001**	**0.0013 ± 0.001**	**0.8305 ± 0.001**	**0.2744 ± 0.001**	**0.1255 ± 0.001**

Table 12
The results of cross-validation experiments on LiTS dataset.

Model	Liver			Tumor		
	DSC	VOE	RVD	DSC	VOE	RVD
U-Net [11]	0.9308 ±0.005	0.1219 ±0.002	−0.0417 ± 0.015	0.6627 ± 0.015	0.4734 ± 0.009	0.1305 ± 0.029
U-Net++ [14]	0.9419 ± 0.001	0.1061 ± 0.002	0.0194 ± 0.001	0.6948 ± 0.004	0.4525 ± 0.003	0.1462 ± 0.007
RA-UNet [25]	0.9477 ± 0.001	0.0962 ± 0.001	0.0141 ± 0.001	0.7138 ± 0.002	0.4374 ± 0.003	0.2178 ± 0.038
ASU-Net [26]	0.9513 ± 0.001	0.0881 ± 0.001	0.0132 ± 0.001	0.7265 ± 0.004	0.4236 ± 0.005	−0.2404 ± 0.013
SAA-Net [27]	0.9518 ± 0.001	0.0879 ± 0.001	0.0139 ± 0.001	0.7411 ± 0.007	0.3909 ± 0.012	−0.2202 ± 0.001
HFRU-Net [15]	0.9525 ± 0.001	0.0876 ± 0.003	−0.0083 ± 0.003	0.7556 ± 0.004	0.3738 ± 0.001	−0.1374 ± 0.049
RMAU-Net	**0.9563 ± 0.001**	**0.0788 ± 0.001**	**−0.0036 ± 0.001**	**0.7623 ± 0.001**	**0.3699 ± 0.001**	**0.0108 ± 0.001**

Table 13
The results of cross-validation experiments on 3D-IRCADb dataset.

Model	Liver			Tumor		
	DSC	VOE	RVD	DSC	VOE	RVD
U-Net [11]	0.9567 ± 0.006	0.0704 ± 0.002	−0.0090 ± 0.001	0.5625 ± 0.033	0.5899 ± 0.017	1.4120 ± 0.138
U-Net++[14]	0.9596 ± 0.004	0.0742 ± 0.006	0.0095 ± 0.001	0.6665 ± 0.028	0.4635 ± 0.033	0.4605 ± 0.191
RA-UNet [25]	0.9532 ± 0.001	0.0828 ± 0.001	0.0045 ± 0.001	0.7228 ± 0.010	0.4021 ± 0.007	−0.0446 ± 0.001
ASU-Net [26]	0.9542 ± 0.001	0.0822 ± 0.001	0.0174 ± 0.010	0.7026 ± 0.012	0.4359 ± 0.011	−0.0828 ± 0.011
SAA-Net [27]	0.9575 ± 0.001	0.0776 ± 0.001	0.0138 ± 0.004	0.7299 ± 0.023	0.3831 ± 0.039	0.0906 ± 0.013
HFRU-Net [15]	0.9653 ± 0.003	0.0646 ± 0.007	−0.0016 ± 0.001	0.8134 ± 0.007	0.3040 ± 0.009	0.0884 ± 0.032
RMAU-Net	**0.9703 ± 0.001**	**0.0517 ± 0.001**	**0.0013 ± 0.001**	**0.8305 ± 0.001**	**0.2744 ± 0.001**	**0.1255 ± 0.001**

To demonstrate the generalization ability of our proposed RAMU-Net in clinical practice, we test RMAU-Net training on the LiTS dataset on the 3D-IRCADb dataset and achieve state-of-the-art results for liver and tumor segmentation, with 0.9697 and 0.8307 on DSC respectively. The results on the 3D-IRACDb dataset also demonstrate that our approach is not simply overtrained, but can be effectively generalized to different datasets under different data collection conditions.

Our proposed method mainly uses a multi-scale fusion attention mechanism to segment small tumors and fuzzy livers. Although our proposed RMAU-Net framework yielded encouraging results, it is limited when dealing with discontinuous livers and marginal tumors, as contextual semantic information has not been taken into account. In the future work, we would like to explore solutions to capture contextual semantic information in CT images for segmenting discontinuous livers and marginal tumors.

6. Conclusion and future works

In this study, we proposed RMAU-Net for liver and tumor segmentation. We introduce an effective module named Res-SE-Block to capture important feature information of images. Specially, we design a novel module named MAB which can capture multi-scale feature information and exploit inter-channel and inter-spatial dependencies simultaneously. In addition, we propose a loss function that combines the Focal loss and dice loss. To evaluate the performance of the proposed method, we conducted experiments on LiTS and 3D-IRCADb datasets. The results show that RMAU-Net achieves good performance on DSC (0.9552) and DSC (0.9697) for liver segmentation and DSC (0.7616) and DSC (0.8307) for liver tumor segmentation on LiTS and 3D-IRCADb dataset, which surpassed the performance of other state-of-the-art methods for liver and tumor segmentation.

In this study, we mainly focus on the problem of the multi-scale problem and attention mechanism in the liver and tumor segmentation task, but do not consider about the 3D information of CT images, which is also critical for medical image segmentation. We will consider improving the performance of RMAU-Net by adding 3D information on CT images in future studies.

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.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (Grant Nos. 61501070, 61971078), the Natural Science Foundation of Chongqing, China (Grant No: cstc2019jcyj-msxmX0240, cstc2020jscx-msxmX0086, cstc2021jcyj-msxmX0605), and Science and Technology Foundation of Chongqing Education Commission, China (Grant Nos. CQUT20181124, KJQN202001137, KJQN202001129, KJQN202101104).

References

[1] Marina Galicia-Moreno, Jorge A Silva-Gomez, Silvia Lucano-Landeros, Arturo Santos, Hugo C Monroy-Ramirez, Juan Armendariz-Borunda, Liver cancer: Therapeutic challenges and the importance of experimental models, Can. J. Gastroenterol. Hepatol. 2021 (2021).

[2] Sang Hee Ahn, Adam Unjin Yeo, Kwang Hyeon Kim, Chankyu Kim, Youngmoon Goh, Shinhaeng Cho, Se Byeong Lee, Young Kyung Lim, Haksoo Kim, Dongho Shin, et al., Comparative clinical evaluation of atlas and deep-learning-based auto-segmentation of organ structures in liver cancer, Radiat. Oncol. 14 (1) (2019) 1–13.

[3] Hyunseok Seo, Masoud Badiei Khuzani, Varun Vasudevan, Charles Huang, Hongyi Ren, Ruoxiu Xiao, Xiao Jia, Lei Xing, Machine learning techniques for biomedical image segmentation: An overview of technical aspects and introduction to state-of-art applications, Med. Phys. 47 (5) (2020) e148–e167.

[4] Pablo Mesejo, Andrea Valsecchi, Linda Marrakchi-Kacem, Stefano Cagnoni, Sergio Damas, Biomedical image segmentation using geometric deformable models and metaheuristics, Comput. Med. Imaging Graph. 43 (2015) 167–178.

[5] J.M. Pardo, Diego Cabello, J. Heras, A snake for model-based segmentation of biomedical images, Pattern Recognit. Lett. 18 (14) (1997) 1529–1538.

[6] A. Baâzaoui, W. Barhoumi, A. Ahmed, E. Zagrouba, Semi-automated segmentation of single and multiple tumors in liver CT images using entropy-based fuzzy region growing, IRBM 38 (2) (2017) 98–108.

[7] Changyang Li, Xiuying Wang, Stefan Eberl, Michael Fulham, Yong Yin, Jinhu Chen, David Dagan Feng, A likelihood and local constraint level set model for liver tumor segmentation from CT volumes, IEEE Trans. Biomed. Eng. 60 (10) (2013) 2967–2977.

[8] Yuanzhi Cheng, Xin Hu, Ji Wang, Yadong Wang, Shinichi Tamura, Accurate vessel segmentation with constrained B-snake, IEEE Trans. Image Process. 24 (8) (2015) 2440–2455.

[9] Sidra Gul, Muhammad Salman Khan, Asima Bibi, Amith Khandakar, Mohamed Arselene Ayari, Muhammad E.H. Chowdhury, Deep learning techniques for liver and liver tumor segmentation: A review, Comput. Biol. Med. (ISSN: 0010-4825) 147 (2022) 105620, http://dx.doi.org/10.1016/j.compbiomed.2022.105620.

[10] Jonathan Long, Evan Shelhamer, Trevor Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431–3440.

[11] Olaf Ronneberger, Philipp Fischer, Thomas Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241.

[12] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al., Attention u-net: Learning where to look for the pancreas, 2018, arXiv preprint arXiv:1804.03999.

[13] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, Olaf Ronneberger, 3D U-net: Learning dense volumetric segmentation from sparse annotation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2016, pp. 424–432.

[14] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, Jianming 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, Springer, 2018, pp. 3–11.

[15] Devidas T. Kushnure, Sanjay N. Talbar, HFRU-net: High-level feature fusion and recalibration UNet for automatic liver and tumor segmentation in CT images, Comput. Methods Programs Biomed. 213 (2022) 106501.

[16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[17] Xiaomeng Li, Hao Chen, Xiaojuan Qi, Qi Dou, Chi-Wing Fu, Pheng-Ann Heng, H-DenseUNet: Hybrid densely connected UNet for liver and tumor segmentation from CT volumes, IEEE Trans. Med. Imaging 37 (12) (2018) 2663–2674.

[18] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, Jiaya Jia, Pyramid scene parsing network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2881–2890.

[19] Xi Fang, Pingkun Yan, Multi-organ segmentation over partially labeled datasets with multi-scale feature abstraction, IEEE Trans. Med. Imaging 39 (11) (2020) 3619–3629.

[20] Pingping Zhang, Wei Liu, Yinjie Lei, Huchuan Lu, Xiaoyun Yang, Cascaded context pyramid for full-resolution 3D semantic scene completion, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 7801–7810.

[21] Changjian Sun, Shuxu Guo, Huimao Zhang, Jing Li, Meimei Chen, Shuzhi Ma, Lanyi Jin, Xiaoming Liu, Xueyan Li, Xiaohua Qian, Automatic segmentation of liver tumors from multiphase contrast-enhanced CT images based on FCNs, Artif. Intell. Med. 83 (2017) 58–66.

[22] L.I. Song, K.F. Geoffrey, H.E. Kaijian, Bottleneck feature supervised U-net for pixel-wise liver and tumor segmentation, Expert Syst. Appl. 145 (2020) 113131.

[23] Zhe Liu, Yu-Qing Song, Victor S Sheng, Liangmin Wang, Rui Jiang, Xiaolin Zhang, Deqi Yuan, Liver CT sequence segmentation based with improved U-net and graph cut, Expert Syst. Appl. 126 (2019) 54–63.

[24] Amrita Kaur, Lakhwinder Kaur, Ashima Singh, GA-UNet: Unet-based framework for segmentation of 2D and 3D medical images applicable on heterogeneous datasets, Neural Comput. Appl. 33 (21) (2021) 14991–15025.

[25] Qiangguo Jin, Zhaopeng Meng, Changming Sun, Hui Cui, Ran Su, RA-UNet: A hybrid deep attention-aware network to extract liver and tumor in CT scans, Front. Bioeng. Biotechnol. (2020) 1471.

[26] Qinhan Gao, Mohamed Almekkawy, ASU-net++: A nested U-net with adaptive feature extractions for liver tumor segmentation, Comput. Biol. Med. 136 (2021) 104688.

[27] Chi Zhang, Jingben Lu, Qianqian Hua, Chunguo Li, Pengwei Wang, SAA-Net: U-shaped network with Scale-Axis-Attention for liver tumor segmentation, Biomed. Signal Process. Control 73 (2022) 103460.

[28] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, Serge Belongie, Feature pyramid networks for object detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2117–2125.

[29] Zaiwang Gu, Jun Cheng, Huazhu Fu, Kang Zhou, Huaying Hao, Yitian Zhao, Tianyang Zhang, Shenghua Gao, Jiang Liu, Ce-net: Context encoder network for 2D medical image segmentation, IEEE Trans. Med. Imaging 38 (10) (2019) 2281–2292.

[30] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, Jianming Liang, Unet++: Redesigning skip connections to exploit multiscale features in image segmentation, IEEE Trans. Med. Imaging 39 (6) (2019) 1856–1867.

[31] Wen Liu, Yankui Sun, Qingge Ji, Mdan-unet: Multi-scale and dual attention enhanced nested u-net architecture for segmentation of optical coherence tomography images, Algorithms 13 (3) (2020) 60.

[32] Devidas T. Kushnure, Sanjay N. Talbar, MS-UNet: A multi-scale UNet with feature recalibration approach for automatic liver and tumor segmentation in CT images, Comput. Med. Imaging Graph. 89 (2021) 101885.

[33] Huimin Huang, Lanfen Lin, Ruofeng Tong, Hongjie Hu, Qiaowei Zhang, Yutaro Iwamoto, Xianhua Han, Yen-Wei Chen, Jian 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.

[34] Ying Chen, Cheng Zheng, Fei Hu, Taohui Zhou, Longfeng Feng, Guohui Xu, Zhen Yi, Xiang Zhang, Efficient two-step liver and tumour segmentation on abdominal CT via deep learning and a conditional random field, Comput. Biol. Med. (ISSN: 0010-4825) 150 (2022) 106076, http://dx.doi.org/10.1016/j.compbiomed.2022.106076.

[35] Rayyan Azam Khan, Yigang Luo, Fang-Xiang Wu, RMS-UNet: Residual multi-scale UNet for liver and lesion segmentation, Artif. Intell. Med. (2022) 102231.

[36] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, Illia Polosukhin, Attention is all you need, Adv. Neural Inf. Process. Syst. 30 (2017).

[37] Jie Hu, Li Shen, Gang Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132–7141.

[38] Sanghyun Woo, Jongchan Park, Joon-Young Lee, In So Kweon, Cbam: Convolutional block attention module, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 3–19.

[39] Jongchan Park, Sanghyun Woo, Joon-Young Lee, In So Kweon, Bam: Bottleneck attention module, 2018, arXiv preprint arXiv:1807.06514.

[40] Zekun Wang, Yanni Zou, Peter X. Liu, Hybrid dilation and attention residual U-net for medical image segmentation, Comput. Biol. Med. 134 (2021) 104449.

[41] Chi Zhang, Jingben Lu, Qianqian Hua, Chunguo Li, Pengwei Wang, SAA-net: U-shaped network with scale-axis-attention for liver tumor segmentation, Biomed. Signal Process. Control 73 (2022) 103460.

[42] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, Alan L Yuille, Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs, IEEE Trans. Pattern Anal. Mach. Intell. 40 (4) (2017) 834–848.

[43] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, Spatial pyramid pooling in deep convolutional networks for visual recognition, IEEE Trans. Pattern Anal. Mach. Intell. 37 (9) (2015) 1904–1916.

[44] Liang-Chieh Chen, George Papandreou, Florian Schroff, Hartwig Adam, Rethinking atrous convolution for semantic image segmentation, 2017, arXiv preprint arXiv:1706.05587.

[45] Michael Yeung, Evis Sala, Carola-Bibiane Schönlieb, Leonardo Rundo, Unified focal loss: Generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation, Comput. Med. Imaging Graph. 95 (2022) 102026.

[46] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, Piotr Dollár, Focal loss for dense object detection, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2980–2988.