Electronic Research Archive

ERA, 31(8): 4324–4345.
DOI: 10.3934/era.2023221
Received: 09 May 2023
Revised: 28 May 2023
Accepted: 30 May 2023
Published: 07 June 2023

http://www.aimspress.com/journal/ERA

Research article

2.5D cascaded context-based network for liver and tumor segmentation from CT images

Rongrong Bi[1,*], Liang Guo[2], Botao Yang[1], Jinke Wang[1,2] and Changfa Shi[3]

[1] Department of Software Engineering, Harbin University of Science and Technology, Rongcheng 264300, China
[2] School of Automation, Harbin University of Science and Technology, Harbin 150080, China
[3] Mobile E-business Collaborative Innovation Center of Hunan Province, Hunan University of Technology and Business, Changsha 410205, China

* **Correspondence:** Email: haligong_b@163.com.

Abstract: The existing 2D/3D strategies still have limitations in human liver and tumor segmentation efficiency. Therefore, this paper proposes a 2.5D network combing cascaded context module (CCM) and Ladder Atrous Spatial Pyramid Pooling (L-ASPP), named CCLNet, for automatic liver and tumor segmentation from CT. First, we utilize the 2.5D mode to improve the training efficiency; Second, we employ the ResNet-34 as the encoder to enhance the segmentation accuracy. Third, the L-ASPP module is used to enlarge the receptive field. Finally, the CCM captures more local and global feature information. We experimented on the LiTS17 and 3DIRCADb datasets. Experimental results prove that the method skillfully balances accuracy and cost, thus having good prospects in liver and liver segmentation in clinical assistance.

Keywords: liver and tumor; segmentation; cascaded context; ASPP; deep learning

1. Introduction

The fourth leading cause of cancer-related death is liver cancer, which is also the fifth most common cancer in the world [1]. For the early diagnosis and evaluation of both primary and secondary hepatic tumors, abnormalities in the liver are crucial. For thorough tumor staging, the liver and its lesions are regularly examined. Additionally, the greatest target lesion's diameter must be measured

according to the Response Evaluation Criteria in Solid Tumor (RECIST) or modified RECIST procedures. Therefore, accurate segmentation of liver tumors is necessary for the diagnosis of cancer, planning of the course of therapy and monitoring of the effectiveness of that treatment [2]. Although only around one-third of patients are eligible for curative procedures like percutaneous ablation, surgical resection or liver transplantation, clinical results continue to be poor [3].

Clinically, the segmentation of the human liver and tumors is usually done manually by experts with rich professional knowledge and experience. However, this work is time-consuming, labor-intensive and susceptible to experts' subjective judgment. Therefore, there is an urgent need to develop an automated segmentation technology based on medical images to assist doctors in diagnosing liver disease. Among them, CT imaging has attracted the attention of scholars thanks to its high quality and low price.

Currently, some challenges exist in automatically segmenting liver and tumor from CT, as shown in Figure 1. (i) Tumors located at the edge of the liver have different gray scales from the liver, which are easily mistaken for other organs or tissues (Figure 1a)). (ii) Liver tumors' size, number and location information are challenging to determine (Figure 1b)). (iii) Once the liver has adjacent organs, the boundaries between the liver and these organs are blurred, which may lead to under-/over-segmentation (Figure 1c)). (iv) Differences in CT scanning instruments and imaging acquisition schemes lead to considerable differences in the intensity and density of the liver and tumor regions. Besides, there will also be different noise levels, blurred borders, etc.

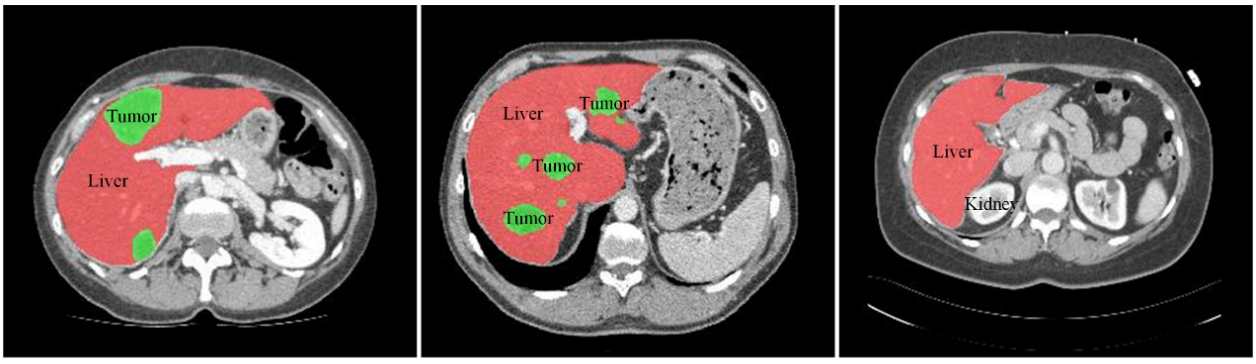

Figure 1. Illustrations of segmentation challenges. a) Liver tumor located at the liver edge; b) The size and number of tumors are variable; c) Liver with adjacent organ.

In response to these challenges, scholars have conducted extensive research based on deep learning technology in recent years. As a result, many 2D/3D-based methods have emerged, significantly improving segmentation performance. However, these approaches still have certain limitations. For example, 2D-based methods are usually trained by 2D slices, making full use of 2D plane information, but ignoring the information between pieces, causing low accuracy. On the other hand, the 3D-based methods need fully use 3D convolution to learn the information within and between layers thoroughly. However, a large amount of memory consumption and computational cost is a significant burden. Therefore, considering the above limitations, this paper proposes a 2.5D network, which combines CCM and L-ASPP modules, called CCLNet, for liver and tumor segmentation. The main contributions of this work are summarized as follows:

1) Leverage 2.5D mode to overcome the insufficient learning in 2D and the high training cost of 3D.
2) Capture more local and global feature information by CCM modules.
3) Employ L-ASPP to increase the receptive field and improve the ability to learn contextual features;
4) Use ResNet-34 as the encoder to improve the segmentation accuracy through the scSE module.

The rest of this paper is organized as follows: Section 2 presents the related work; Section 3 gives a detailed description of the proposed method; Section 4 provides the experimental configuration, results and analysis; In the final Section 5, we summarize our research and provide a future outlook.

2. Related work

Traditional machine learning requires manual feature extraction. Instead, the deep learning-based network performs all processes automatically and has shown superior image classification, detection and segmentation performance. Currently, the liver and tumor segmentation approaches can be divided into three categories: 1) fully convolutional neural network (FCN)-based methods; 2) U-Net-based methods; 3) generative adversarial networks (GAN)-based methods.

2.1. FCN-based methods

The FCN-based approaches allow for arbitrary input sizes and generate outputs of corresponding sizes with effective inference and learning. It was first proposed by Long et al. [4] in 2015. It plays an essential role in medical image segmentation because it can pixel-wise predict and predict different categories simultaneously. Ben-Cohen et al. [5] presented an FCN for the first time for liver segmentation and metastases detection. According to their tests, the FCN with data augmentation, neighbor slices and the right class weights produced the greatest outcomes. Zhang et al. [6] first employed an FCN for coarse liver segmentation. Then they performed fine liver segmentation through post-processing methods such as level sets and conditional random fields (CRF). Their experimental results demonstrated that the enhancement of FCN could achieve superior segmentation performance. Jiang et al. [7] designed an attention hybrid connection network structure that combines soft and hard attention and long and short skip connections. This structure can use attention to learn more liver information and achieve a significant segmentation effect. The suggested approach can hasten network convergence. The precision of tumor segmentation, however, still has to be improved. Christ et al. [8] designed a cascaded FCN for automatic localization and joint volume segmentation of the liver and its lesions and post-processed the resulting segmentation results with 3D dense CRF. Although their computation times are below 100 s per volume, the liver and lesion segmentation Dice score is not high enough.

2.2. U-Net-based methods

U-Net was first proposed by Ronneberger et al. [9]. Most researchers now use this network because of its exceptional performance in tasks involving the segmentation of medical images. Seo et al. [10] added a residual path in the skip connection of U-Net to avoid duplication of low-resolution information of features. It can produce improved segmentation outcomes for the liver and tumor regions when dealing with indistinct segmentation boundaries and small objects. But it also has a common drawback of deep learning, i.e., less generalizability. The attention module Wang et al. [11]

added to U-Net allows it to extract picture information and adaptively suppress unimportant regions. At the same time, the advantages of residual learning and atrous convolution are used to extract liver image information at multiple scales. Their suggested approach is susceptible to missing crucial context information on the z-axis because it is built on a 2D network. To extract contextual information, Jin et al. [12] presented a 3D hybrid deep attention-aware network that combines low-level and high-level feature maps. The limitation of the proposed method is the training time because the 3D convolutions require larger parameters than the 2D convolutions. H-DenseUNet was proposed by Li et al. [13] and is based on 2D and 3D networks. By parallelizing the gradient computation for a mini-batch across mini-batch elements, they trained the network using parallel data training, a useful technique for accelerating gradient descent. The GPU memory does, however, have a limit on the model complexity. Lv et al.'s [14] novel 2.5D lightweight RIU-Net network for quick and precise liver and tumor segmentation from CT uses methods from the residual and inception theories. However, the suggested technique makes mistakes when dealing with low-contrast boundaries between the liver and surrounding organs because of poor feature integration. Meng et al. [15] created a two-stage densely connected UNet (DCUNet) and achieved good results for segmenting liver and liver tumors. However, the Dice score for liver segmentation is still a little lower compared to manual detection. The Aim-Unet, a fusion of Unet and Inception v3 architectures for liver and tumor segmentation, was proposed by Özcan et al. [16]. At appropriate accuracy levels, it is roughly 100 times faster than hand segmentation.

2.3. GAN-based methods

A generative model based on game theory is called a GAN. In terms of producing realistic data, particularly photographs, they have achieved a great deal of practical success. It was proposed by Goodfellow et al. [17] to generate new copies of data by learning the data distribution. The idea of this network is to implement a generator through a neural network so that it models a transformation function that takes random variables as input and follows a target distribution when trained. While another network is trained as a discriminator simultaneously, it distinguishes between real and fake generated data. Enokiya et al. [18] proposed a liver segmentation method based on U-Net and Wasserstein GAN for the problem of difficult training of small data sets, and the Dice value increased by 3%–5%. However, in the presence of discontinuous livers in the same slice, the detection rate of small livers is low. An automatic and effective technique for liver segmentation from 3D CT was proposed by Yang et al. [19]. The method uses a deep image-to-image network (DI2IN) to create liver segmentations. Then, it uses an adversarial network to distinguish DI2IN output from actual data during training, further improving the performance of DI2IN. Their research demonstrates that for 3D volumetric datasets, a framework that combines the encoder-decoder structure, skip connections and deep supervision scheme has a better structural design. Demir et al. [20] used the transformer's self-attention mechanism to allow the network to aggregate high-dimensional features, provide the advantages of global information modeling and combine the transformer and GAN to segment the liver. The proposed hybrid architecture (i.e., a combination of GAN and Transformers) can potentially be applied to various medical image segmentation tasks. However, they did not offer the method's time performance. An adversarial learning model (ALM) and a boundary mining model (BMM) make up the semi-supervised model that Xu et al. [21] suggested. They employed the dataset from a scanner even though they demonstrated the usefulness of the proposed approach. To segment liver cancers from CT, Chen et al. [22] created a cascaded adversarial training method. Utilizing the cascade

framework has the advantage of making the network less complex and simpler to train. However, the fact that this technique is not always end-to-end, could make training take longer.

3. Methods

3.1. The overall framework of the proposed CCLNet

This paper presents a CCLNet for liver and tumor segmentation. Its overall structure is shown in Figure 2.

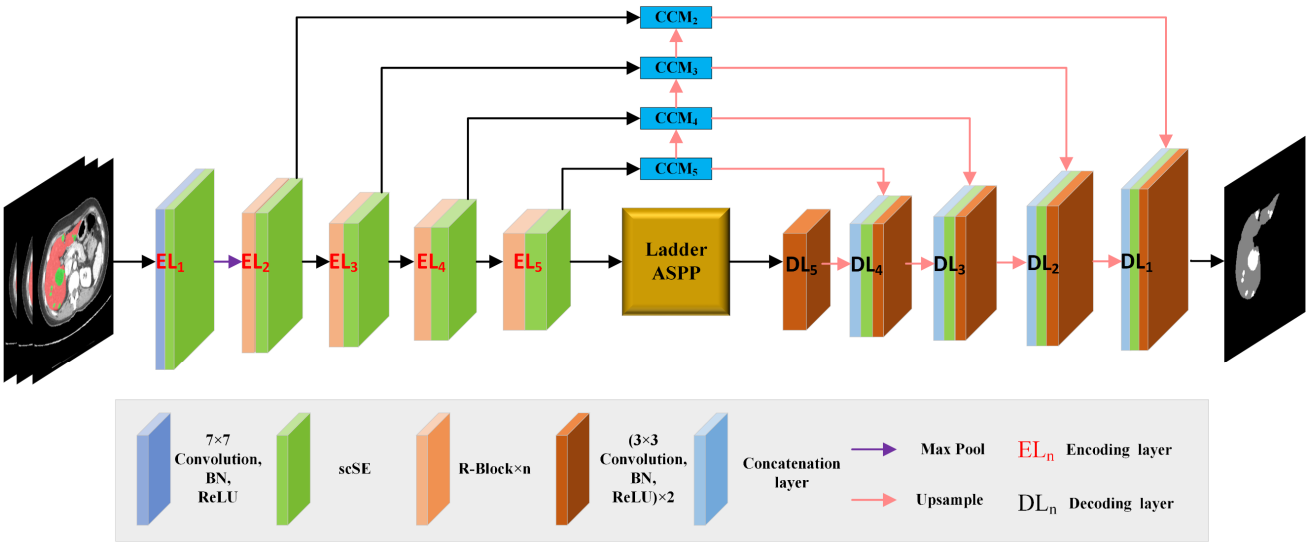

Figure 2. The framework of the proposed CCLNet.

The input of the proposed model is a set of adjacent slices, and the output is the segmentation result of the central portion. The model contains an encoder and a decoder. In the encoder, the R-Block output of each layer passes through a scSE module (where R-Blocks at different levels correspond to varying levels of the ResNet-34 network). It is used to recalibrate the response of features through spatial and channel weighting mechanisms, resulting in a finer segmentation effect. Between codecs, richer feature information is extracted through L-ASPP. The decoder stage consists of a series of convolutions and four upsampling operations. Additionally, the model incorporates a CCM, which takes the R-Block's output information and extracts regional context data, which is then coupled with the earlier context data from lower layers. Thus, this method can extract global and local information to the greatest extent.

We employed scSE modules in both encoding and decoding layers. In the encoding layer, denoted as EL_n, the output feature map from the R-Block of EL_n is fed into the scSE module, which recalibrates the input both spatially and channel-wise, and outputs the feature map as input for the R-Block of the next encoding layer, EL_{n+1}. In the decoding layer, denoted as DL_n, the output from CCM_n is concatenated with the output from DL_n and fed as input to the scSE module of DL_{n-1}, whose output is then fed to the convolutional unit of the same decoding layer.

Ladder-ASPP takes the output feature map of EL_5 as input and outputs a feature map concatenated from two cascaded parts. The global pooling output makes up the first part, while the second is the

dense connection with a ladder-like structure. The DL_5 decoding layer then accepts the output of Ladder-ASPP as input.

CCM_n takes the output from the EL_n encoding layer and employs four branches to extract contextual information. To extract the final representative feature of this layer, the features from the four branches are first concatenated. It is then upsampled and added pixel-by-pixel to the CCM_{n+1} output. Then, it is upsampled and added pixel-wise to the output of CCM_{n+1}, producing the final output of CCM_n, which is then used as input to the DL_{n-1} decoding layer. The specific input and output relationships are shown in Table 1.

Table 1. Input and output of the CCM module.

CCM	Input	Output
CCM_2	output of the EL_2 add output 3	output 2
CCM_3	output of the EL_3 add output 4	output 3
CCM_4	output of the EL_4 add output 5	output 4
CCM_5	output of the EL_5	output 5

3.2. Spatial and channel squeeze and excitation (scSE)

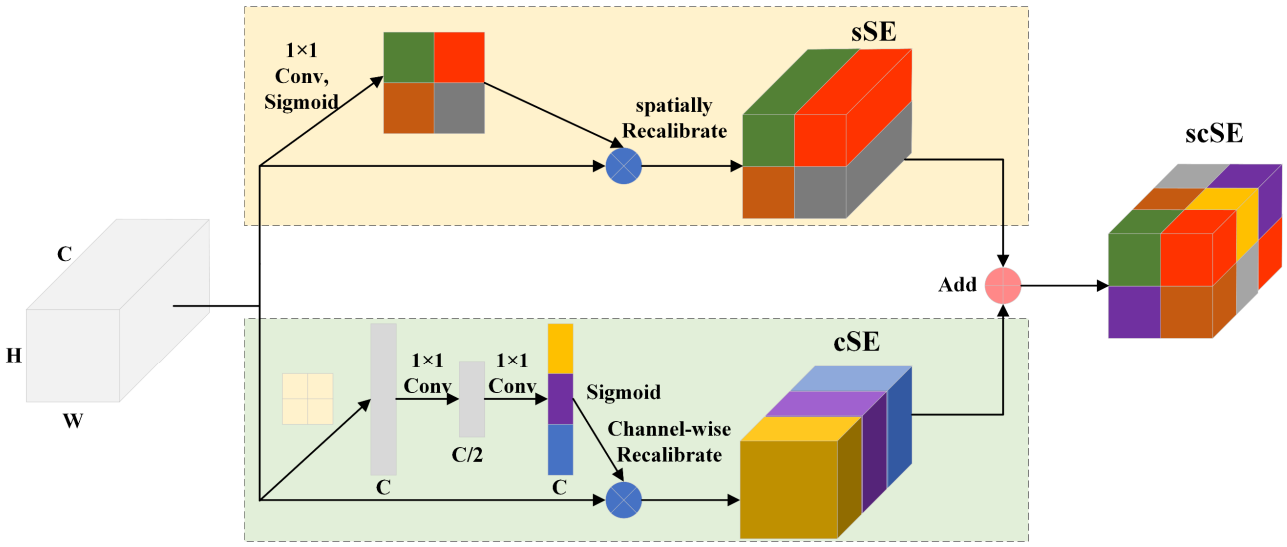

Figure 3. The structure of the scSE module.

The squeeze & excitation (SE) module belongs to the channel attention, including two operations of squeezing and stimulating. Three SE versions, including cSE (channel SE), sSE (spatial SE) and scSE, were proposed by Roy et al. [23] to move the SE module from the image classification network to the image segmentation network. (i) The cSE module is used to reweight channels, which can be adapted to ignore less critical channels and emphasize more important ones. (ii) the sSE module reweights space to deliver more accurate spatial localization by eliminating unimportant spatial locations. (iii) The scSE module is made up of the cSE module and the sSE module, which can integrate the channel and space feature maps into the output layer after recalibrating their respective feature maps. Figure 3 depicts the scSE module's structural layout.

scSE is obtained by adding cSE and sSE. For the sSE module, the feature map is initially obtained

using a 1 × 1 convolution, the weights for the spatial attention are then obtained using the Sigmoid activation function, and finally, the calibration of the spatial attention is completed by multiplying with the original feature map. The initial step in the cSE module is to deconstruct the spatial characteristics of the feature map using the global average pooling technique. Second, it performs SE operations on each channel and utilizes the ReLU activation function and Sigmoid to obtain the corresponding channel attention weight. Finally, the channel attention calibration is finished by doing the channel multiplication of the weights and the original feature map.

3.3. Ladder Atrous Spatial Pyramid Pooling (L-ASPP)

Because they have a larger receptive field than ordinary pooling, Pyramid Pooling (PP) and Atrous Spatial Pyramid Pooling (ASPP) are two prominent methods for encoding contextual information. However, since PP uses multi-scale pooling kernels to accomplish pooling directly, it frequently leads to irreversible information loss and subpar segmentation of small objects. Besides, ASPP does not use multi-scale pooling kernels to perform hole convolution but uses multiple dilated convolution kernels to serve. Therefore, compared with PP, hole convolution achieve superior performance than pooling operations in preserving details; thus, ASPP can provide better context information.

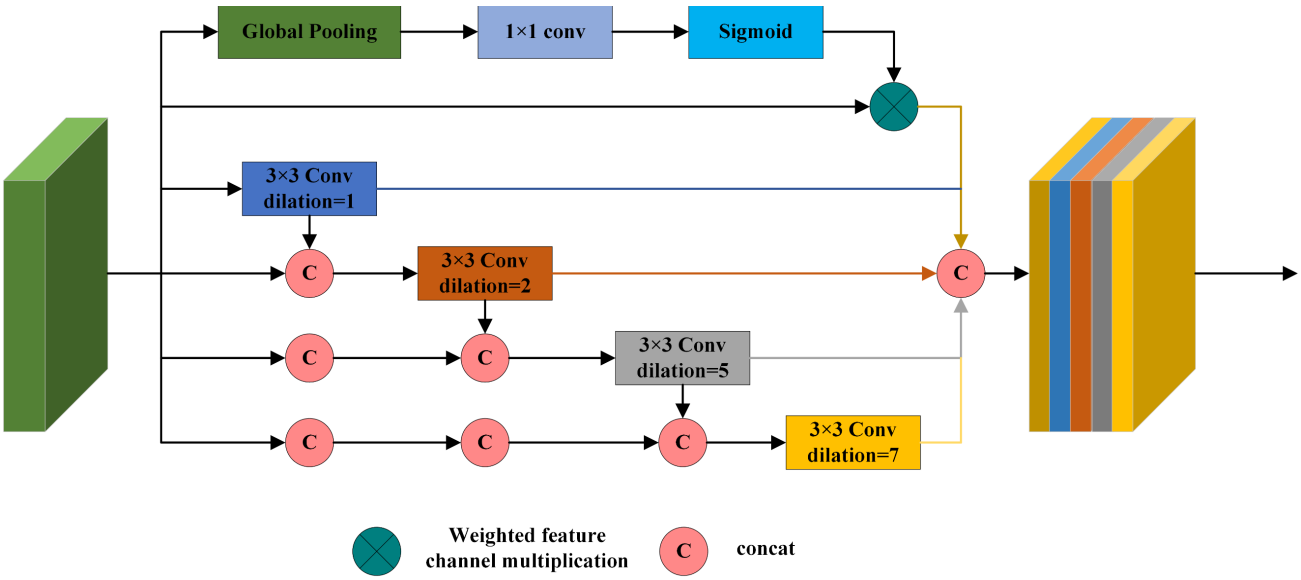

Figure 4. Structure of the L-ASPP module.

Nevertheless, there are still two difficulties with ASPP in real-world usage. One of them is that a constant dilation rate will produce a meshing effect, and some pixels that fall into the receptive field cannot participate in the convolution operation. The global context data is ignored by ASPP, which is the second problem. To solve this problem, Lei et al. [24] designed the L-ASPP, whose structure is shown in Figure 4.

The L-ASPP output feature maps are concatenated into two sections. The global pooling output is the first, and the feature-fusion liking ladder is the second. In Figure 4, L-ASPP improves context encoding by employing variable dilation rates for atrous convolutions and densely connected ladder connections for an improved feature fusion.

Dense networks, however, quickly result in an increase in the number of parameters and large memory needs. It incorporates depthwise separable convolution (DSC) with L-ASPP to decrease the number of parameters. DSC can accomplish decoupled computation between spatial and channel characteristics, as opposed to normal convolutions, which often combine spatial and channel features together. This lowers the number of parameters needed.

Finally, we integrated it into L-ASPP to enhance the feature representation of ASPP, considering that global pooling can contain the priority of channels with more important information. Figure 4 shows how the information buried in the spatial and channel dimensions is used simultaneously, and how the combined global and local information results in the final feature map.

3.4. Cascading context modules (CCM)

This paper adopts a cascaded context module (CCM) for more contextual information, which can collect contextual data obtained from R-Block on a global and regional level, thus realizing multi-scale receptive fields [25]. In Figure 5, the structure of CCM is depicted.

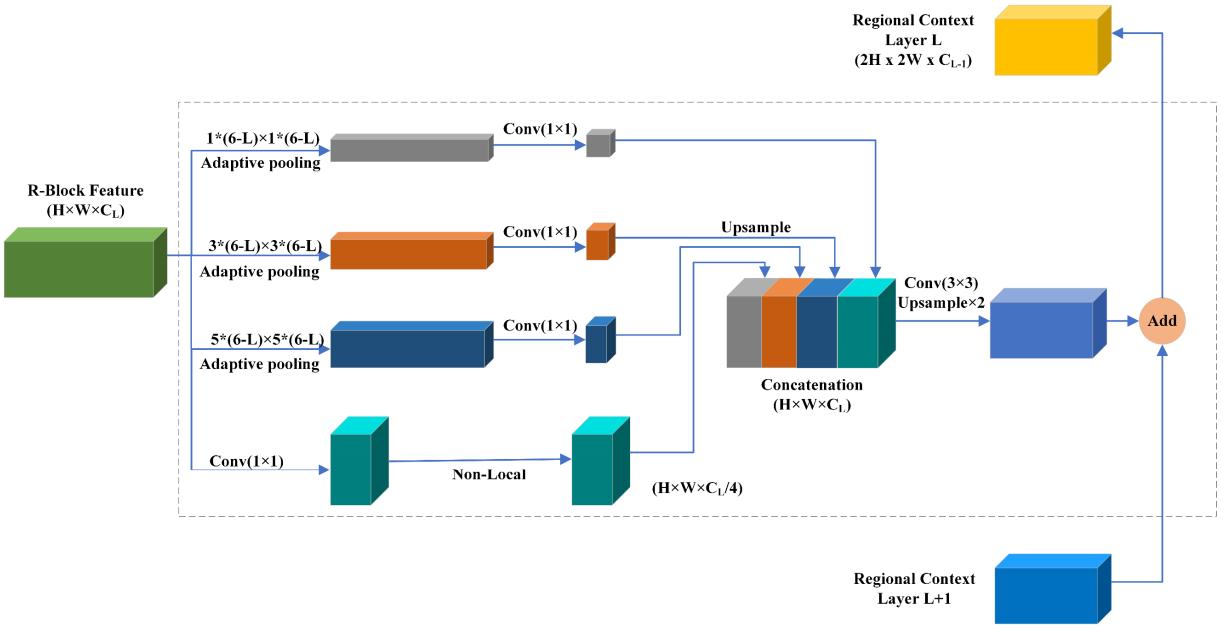

Figure 5. The structure of the cascading context module.

CCM receives features from R-Block and processes extracting contextual information with four branches. The last branch employs non-local operations to capture long-distance dependencies. The other three units first use adaptive pooling operations with kernel sizes A, B and C, respectively. Subsequently, a 1 × 1 convolution is used to downscale the channels to a quarter of the original size and the features are upsampled to the same spatial size using this technique. To actualize the regional context of several distinct layer combinations, the features from the four branches are then concatenated to get this layer's final representative features. Upsampling is then used to add pixels that match the context features of the previous CCM layer. Additionally, the deepest CCM at the top of the encoder branch can record global context, which can be included in the regional context through the cascade connection of CCMs.

3.5. Evaluation metrics and statistical tests

In this paper, six metrics are used to evaluate the segmentation results of liver and tumor, including Dice coefficient, volume overlap error (VOE), relative volume error (RVD), average symmetric surface distance (ASD), maximum average surface distance (MSD) and root mean square symmetric surface distance (RMSD) [26]. We use A and B to represent two segmentation results.

(i) Dice coefficient: the larger the Dice value, the better the segmentation effect. The formula is as follows: The gold standard and the predicted result is represented by A and B, respectively.

$$Dice(A,B) = \frac{2|A \cap B|}{|A|+|B|} \tag{1}$$

(ii) VOE: represents the volumetric overlap error between two sets of voxels and its formula is defined as:

$$VOE(A,B) = 1 - \frac{|A \cap B|}{|A \cup B|} \times 100(\%) \tag{2}$$

(iii) RVD: This metric is an asymmetric metric whose formula is as follows, cannot be the only indicator of segmentation quality.

$$RVD(A,B) = \frac{|B|-|A|}{|A|} \times 100(\%) \tag{3}$$

(iv) ASSD: The shortest Euclidean distance from the voxel v to the surface voxel of the segmentation result X is represented by $d(v, S(X))$, and it indicates the average distance between the surfaces of the two segmentation results A and B. Its formula is defined as:

$$ASSD(A,B) = \frac{1}{|S(A)|+|S(B)|}\left(\sum_{p \in S(A)} d^2\,(p, S(B)) + \sum_{q \in S(B)} d^2\,(q, S(A))\right) \tag{4}$$

(v) MSD: Similar to ASSD, the operation of calculating the average value is changed to the operation of calculating the maximum value. Its formula is defined as:

$$MSD(A,B) = max\left\{\max_{p \in S(A)} d(p, S(B)), \max_{q \in S(B)} d(q, S(A))\right\} \tag{5}$$

(vi) RMSD: This indicator is one of the crucial criteria for evaluating segmentation accuracy. The closer the value is to 0, the better the segmentation effect. Its formula is defined as:

$$RMSD(A,B) = \sqrt{\frac{1}{|S(A)|+|S(B)|}} \times \sqrt{\sum_{p \in S(A)} d\,(p, S(B)) + \sum_{q \in S(B)} d\,(q, S(A))} \tag{6}$$

We used the paired t-test [27] with a significance level of $p < 0.05$ on all measures to examine whether the differences in segmentation accuracy between our proposed method and the compared methodologies were statistically significant. The comparison of techniques' mean values for the same evaluation metric is the null hypothesis.

4. Experiments and results

4.1. Experimental setup

Two publicly available datasets with liver and tumor data, LiTS17 and 3DIRCADb, are used to train and test the proposed technique. (The two datasets are publicly available at https://competitions.codalab.org/competitions/17094 and https://www.ircad.fr/research/3d-ircadb-01/). LiTS17 datasets are utilized for training and verification in 116 sets. The training set and verification sets are randomly allocated according to the ratio of 3:1, and 15 sets are used for testing. Twelve sets of 3DIRCADb data sets are used for training and verification. The training set and verification sets are randomly allocated according to the ratio of 3:1, and eight groups are used for testing.

We conduct the experiments using the Ubuntu 18.04 operating system, Python 3.7, Pytorch 1.8 and CUDA 11.2 platform. The Binary Cross Entropy (BCE) + Dice hybrid loss function is utilized during training. Adam is the optimizer and the initial learning rate is 0.003. We set the maximum number of iterations to 200, the number of input data channels to 3 and the batch size to 4 within the constraints of the computer resources.

4.2. Slice arrangement

To verify the appropriate number of slices, we conducted a comparative experiment. Under the premise of keeping other parameters the same, five different experiments were conducted with the number of consecutive slices 1, 3, 5, 7 and 9 on the LiTS17 dataset (shown in Figure 6). It can be seen in Table 2 that inputting three consecutive slices achieved the best results under the same parameter settings.

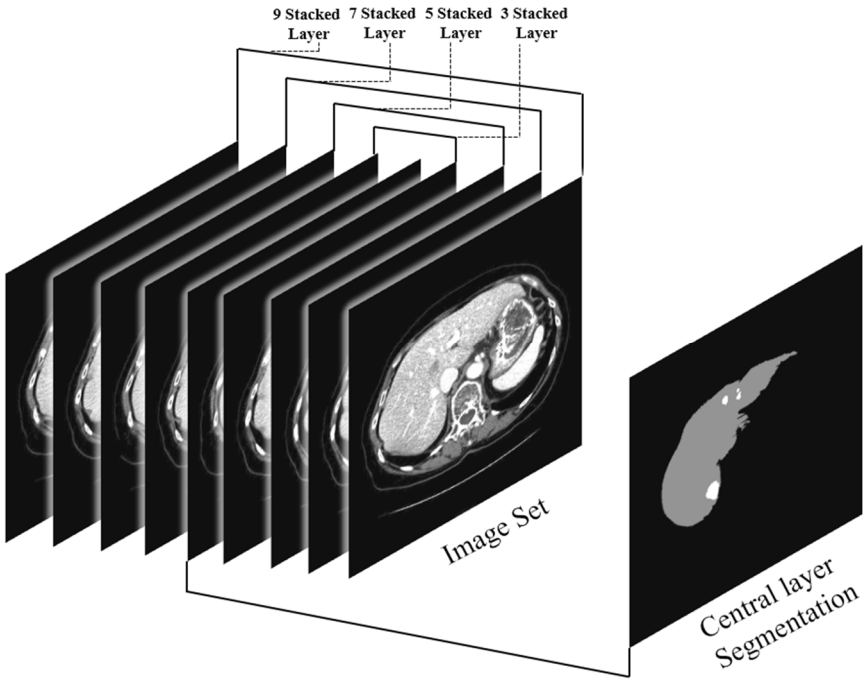

Figure 6. Slice arrangement test.

Table 2. The result of slice arrangement test on LiTS17 validation dataset.

Model	Slice-1	Slice-3	Slice-5	Slice-7	Slice-9
Dice score					
Liver	93.32 ± 0	**97.58 ± 0.28**%	95.17 ± 0.21%	95.26 ± 0.23%	96.21 ± 0.37%
Tumor	70.63 ± 7.67%	**76.78 ± 9.40**%	74.27 ± 8.37%	71.29 ± 8.96%	73.11 ± 0.31%
Hausdorff distance					
Liver	17.54 ± 10.57%	**13.73 ± 8.26%**	14.87 ± 9.11%	15.77 ± 9.29%	13.97 ± 9.73%
Tumor	37.71 ± 15.43%	**28.73 ± 12.27%**	29.73 ± 14.81%	30.42 ± 13.53%	32.47 ± 14.83%

Note: bold values indicate the best result in that row.

4.3. Image preprocessing

The preprocessing steps for liver and tumor model training include three stages (as shown in Figure 7): 1) Windowing: adjust the CT value range to [-200, 200]. 2) Resolution adjustment: find the slices at the beginning and end of the liver and tumor area, expand 20 pieces each outward as training samples, and crop the 512 × 512 size image to 448 × 448 size. 3) Histogram equalization: normalize the image pixels to [0, 1]. These three steps are integrated into one preprocessing program, which takes 5 m 17 s for the LiTS17 dataset and 3 m 11 s for the 3Dircadb dataset. After preprocessing, the data volume of the LiTS17 and 3DIRCADb datasets for training changed to 16,745 and 1154, respectively.

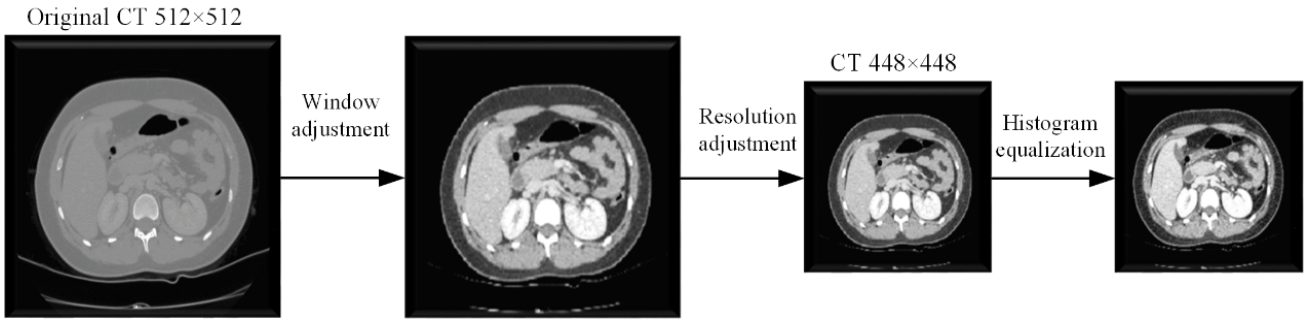

Figure 7. The illustration of image preprocessing.

4.4. Ablation on LiTS17 dataset

4.4.1. Training in ablation

Our ablation experiment consists of four modules using ResNet34 as the basic model: 1) Base, 2) Base+CCM, 3) Base+L-ASPP, 4) Base+CCM+L-ASPP (Proposed CCLNet).

The Dice curve is depicted in Figure 8a),b) during the liver and liver tumor training processes, respectively. The liver Dice curve shows that CCLNet has the fastest convergence speed and locates at the highest position. In contrast, it can be seen from the tumor Dice curve that the training of tumor segmentation is relatively more complicated, the Dice value is not high during the training process and the curve fluctuates greatly.

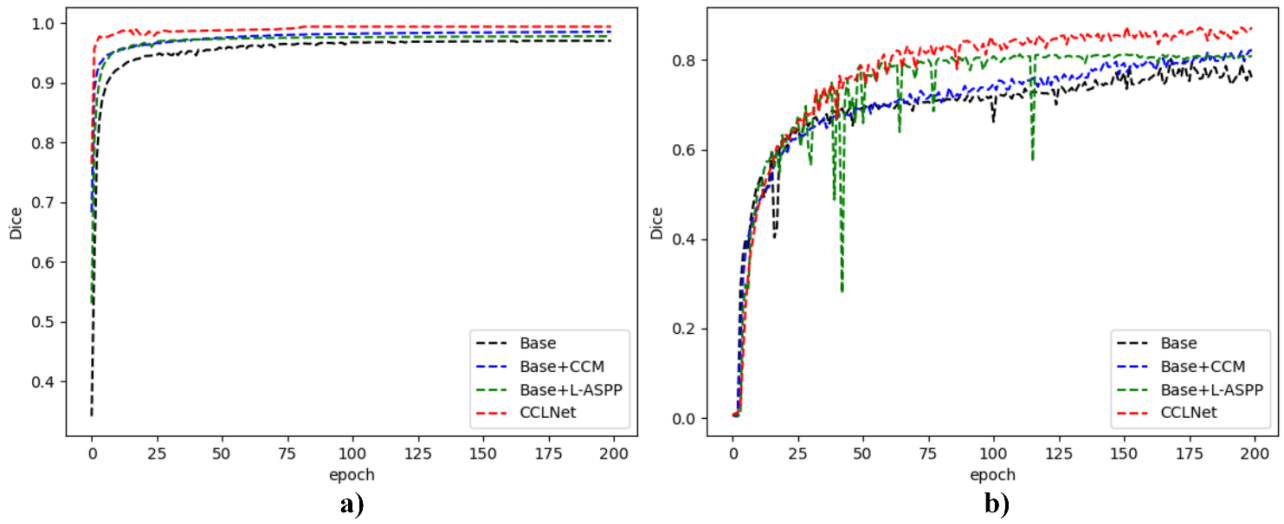

Figure 8. Dice curve in ablation experiment. a) Dice curve of live; b) Dice curve of liver tumor.

4.4.2. Results of ablation

The liver and tumor segmentation findings are shown in Tables 3 and 4, respectively. The tables demonstrate the following:

Table 3. Liver segmentation results in ablation experiment.

Method	Dice (%)	VOE (%)	RVD (%)	ASSD (mm)	RMSD (mm)
Base	96.84 ± 1.78*	6.01 ± 3.23	0.29 ± 0.96	1.16 ± 1.90	4.99 ± 4.82
Base+CCM	97.22 ± 0.33*	4.88 ± 0.62	**0.17 ± 1.03**	1.15 ± 0.30	2.69 ± 1.36
Base+L-ASPP	97.18 ± 0.71*	5.67 ± 1.32	0.67 ± 1.41	1.18 ± 0.41	3.03 ± 1.45
CCLNet	**97.58 ± 0.28**	**4.01 ± 0.68**	0.38 ± 0.60	**1.03 ± 0.25**	**2.20 ± 1.25**

Note: Bold value indicates the best outcome for each statistic. At a significance level of 0.05, * denotes a statistically significant difference between the marked Dice score and our technique.

Table 4. Tumor segmentation results in ablation experiment.

Method	Dice (%)	VOE (%)	RVD (%)	ASSD (mm)	RMSD (mm)
Base	69.73 ± 12.19*	45.39 ± 13.01	0.51 ± 0.96	11.33 ± 11.44	21.09 ± 17.97
Base+CCM	73.51 ± 14.06*	38.87 ± 16.88	-0.27 ± 0.35	5.07 ± 0.30	9.04 ± 8.91
Base+L-ASPP	71.19 ± 19.67*	39.32 ± 21.16	0.67 ± 0.91	8.24 ± 12.43	15.39 ± 13.91
CCLNet	**76.78 ± 9.40**	**24.34 ± 3.78**	**0.15 ± 0.75**	**2.88 ± 2.97**	**7.31 ± 5.49**

(i) Compared with the Base model, the Base+CCM improved the Dice score of the liver segmentation from 96.84% to 97.22% (increased by 0.39%). In addition, the Dice value of tumor segmentation was increased from 69.73% to 73.51% (increased by 5.42%), thus verifying that the cascaded context module can improve segmentation performance by capturing more spatial context information.

(ii) Compared with the Base model, the Base+L-ASPP improved the Dice value of liver segmentation from 96.84 to 97.18% (increased by 0.35%). Besides, the Dice value of tumor

segmentation rose from 69.73% to 71.19% (increased by 2.09%). Thus, the verification shows that the L-ASPP module obtains more information to improve the segmentation accuracy.

(iii) Compared with Base+CCM and Base+ASPP, the Base+CCM+L-ASPP (CCLNet) module is only slightly lower than Base+CCM on the RVD in liver segmentation; the other four metrics have all improved. Therefore, it can be proved that the CCM and the L-ASPP module benefit the overall framework.

4.4.3. Visualization of ablation

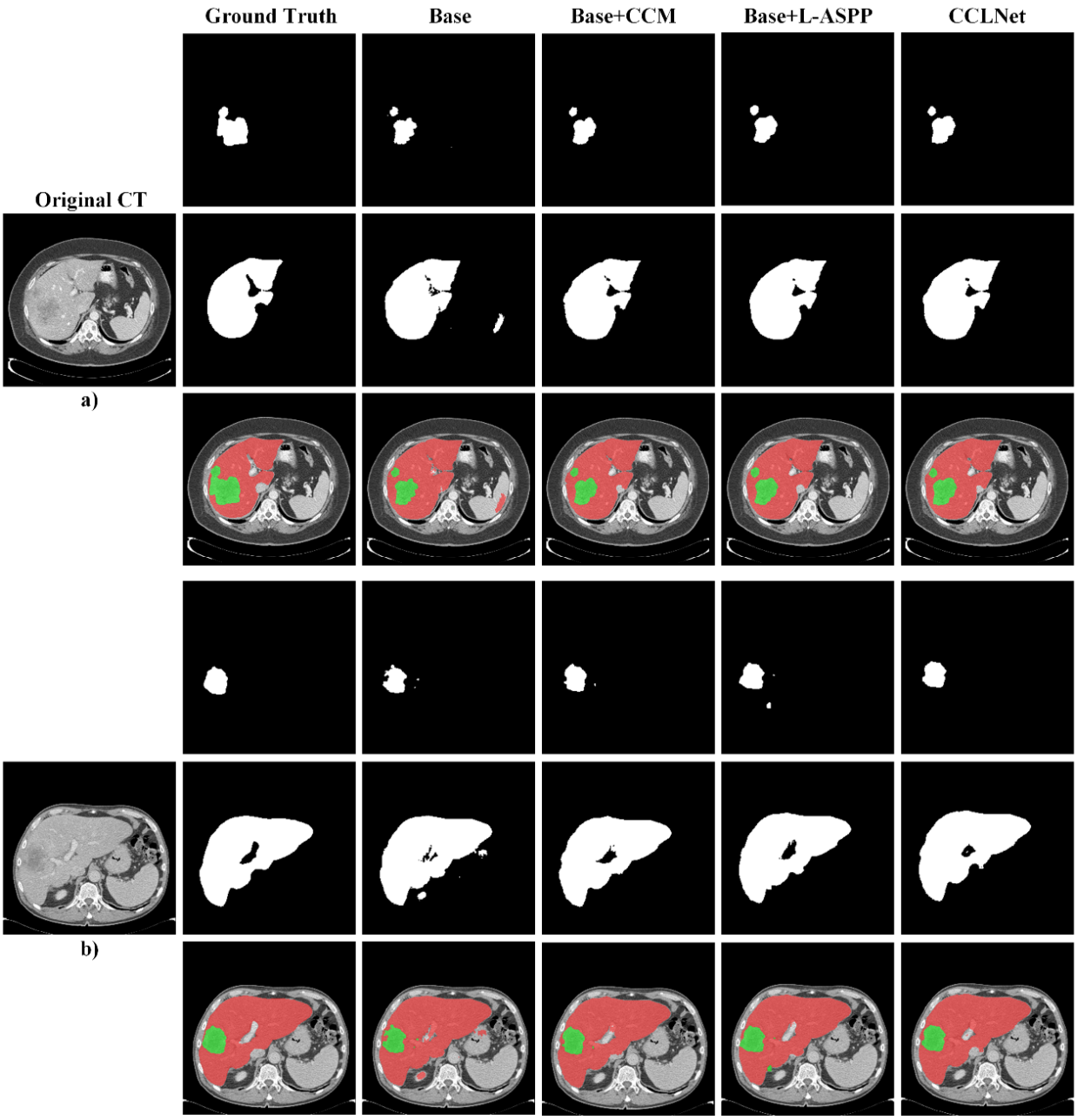

Figure 9. Liver and tumor segmentation visualization during ablation. The first and fourth rows demonstrate the results of tumor segmentation, the second and fifth rows are the results of liver segmentation and the third and last rows provide the segmentation effect on the original CT. (The red/green part indicates the liver/tumor segmentation result).

We compare the segmentation of two cases of liver and tumor using the LiTS17 dataset in Figure 9 to more clearly illustrate the segmentation effect of various combination models in the ablation experiment.

Figure 9a) shows that the Base model clearly under-segmented the liver during segmentation, and the spleen was incorrectly segmented out. In contrast, the other three networks did not show such a problem and were almost close to the ground truth. As the model changes, the segmentation effect is gradually optimized in liver tumor segmentation.

Figure 9b) shows that the Base model causes obvious under-segmentation for the liver segmentation, while margin errors of the other four methods are minor. Besides, all the models produced errors when processing the interlobar fissure region. On the other hand, the suggested CCLNet model is more accurate in segmenting liver tumors when compared to the other three models. In contrast, the other models suffer from significant under-segmentation and poor edge segmentation.

4.5. Comparisons with SOTA methods on the LiTS17 dataset

4.5.1. Quantitative comparison

On the LiTS17 dataset, we compared the performance of CCLNet with four other state-of-the-art (SOTA) methods (FCN [6], UNet [9], MiniSeg [28], Attention UNet [29], HFRU-Net [30] and FRA-UNet [31]) for the segmentation of tumor and liver. Tables 5 and 6 lists the evaluation results.

As can be seen from Table 5, CCLNet achieved 97.58% of Dice score in the liver segmentation, which is 2.06%, 1.27%, 1.21%, 0.41%, 2.58% and 0.45% higher than FCN, UNet, MiniSeg, Attention UNet, HFRU-Net and FRA-Net, respectively, and it also received the highest ratings on the other four parameters.

Table 5. Liver segmentation results in model comparison experiment on the LiTS17 dataset.

Method	Dice (%)	VOE (%)	RVD (%)	ASSD (mm)	RMSD (mm)
FCN	95.52 ± 3.64*	7.94 ± 6.19	0.87 ± 1.81	2.05 ± 1.71	6.23 ± 3.86
UNet	96.31 ± 2.12*	7.04 ± 3.79	0.45 ± 1.62	2.39 ± 1.83	7.11 ± 5.56
MiniSeg	96.37 ± 1.52*	7.22 ± 4.68	0.59 ± 0.87	2.21 ± 1.68	6.01 ± 2.54
Attention UNet	97.17 ± 0.83*	6.33 ± 4.11	0.57 ± 1.32	1.26 ± 1.47	4.61 ± 1.86
HFRU-Net	95.00*	10.50	6.60	3.02	--
FRA-UNet	97.13*	5.50	1.60	1.10	3.67
CCLNet	**97.58 ± 0.28**	**4.01 ± 0.68**	**0.38 ± 0.60**	**1.03 ± 0.25**	**2.20 ± 1.25**

CCLNet also produced the best segmentation outcomes on the five assessment measures for the tumor segmentation displayed in Table 6. Specifically, Dice reached 76.78%, 23.64%, 12.56%, 10.86%, 5.72%, 15.18% and 5.00% higher than FCN, UNet, MiniSeg, Attention UNet, HFRU-Net and FRA-Net, respectively, achieving an undeniable segmentation advantage. According to experimental findings, the suggested CCLNet performs liver and tumor segmentation better than other SOTA networks.

Table 6. Tumor segmentation results in model comparison experiment on the LiTS17 dataset.

Method	Dice (%)	VOE (%)	RVD (%)	ASSD (mm)	RMSD (mm)
FCN	53.14 ± 24.06*	60.91 ± 20.54	-1.48 ± 2.11	22.51 ± 28.67	30.00 ± 29.67
UNet	64.22 ± 29.89*	47.44 ± 27.09	1.63 ± 0.89	9.45 ± 11.97	14.44 ± 14.39
MiniSeg	65.92 ± 22.73*	47.64 ± 20.77	-0.71 ± 0.72	9.40 ± 8.12	17.97 ± 29.08
Attention UNet	71.06 ± 13.96*	36.82 ± 17.03	0.44 ± 0.61	8.04 ± 9.72	15.45 ± 16.09
RU-Net	61.60*	38.40	22.30	**1.24**	--
FRA-UNet	71.78*	35.00	0.27	1.25	**4.55**
CCLNet	**76.78 ± 9.40**	**24.34 ± 3.78**	**0.15 ± 0.75**	**2.88 ± 2.97**	**7.31 ± 5.49**

The parameters of several network models and the training and testing times on the LiTS17 dataset are listed in Table 7. It is clear that the model suggested in this research has a complex configuration of parameters. Besides, it is slightly higher than FCN, UNet and MiniSeg regarding training time and the typical test duration for each set of CT. However, although the proposed model's parameters are comparable to the Attention UNet's, our training time is much lower.

Table 7. Parameters and efficiency for five models on the LiTS17 dataset.

Method	Parameters	Training time	Test time
FCN	18,643,746	**41 h 13 m 18 s**	331 s
UNet	7,765,442	42 h 21 m 6 s	330 s
Miniseg	**82,666**	45 h 54 m 57 s	**327 s**
Attention UNet	34,878,638	102 h 18 m 23 s	369 s
CCLNet	37,706,674	52 h 8 m 16 s	342 s

Note: Bold values represent the best outcomes for each metric in that column.

4.5.2. Qualitative comparison

We chose two special cases from the LiTS17 dataset to test in order to confirm the proposed CCLNet's higher segmentation performance: one contains multiple tumors, and the other reports a large marginal tumor. In Figure 10, the results of the suggested approach and the other four methods are displayed.

In the liver segmentation task of Figure 10a), the FCN mis-segmented part of the kidney as the liver, and an obvious under-segmentation error occurred in the lower part of the image. In contrast, the other four networks accurately extract the liver. For the tumor segmentation, FCN, UNet and MiniSeg resulted in local over-/under-segmentation, while CCLNet and Attention UNet greatly improved the segmentation effect.

For the liver segmentation shown in Figure 10b), FCN and UNet create under-segmentation errors outside the liver region. In the tumor segmentation task, the FCN segmentation results can be found to be poor, and a complete tumor is segmented into multiple discontinuous small tumor blocks. Furthermore, the accuracy of UNet, MiniSeg and Attention UNe around the upper edge of the tumor is not high, and extra under-segmentation errors occur outside the tumor region. In contrast, the results of the proposed CCLNet segmentation are reasonably similar to the ground truth.

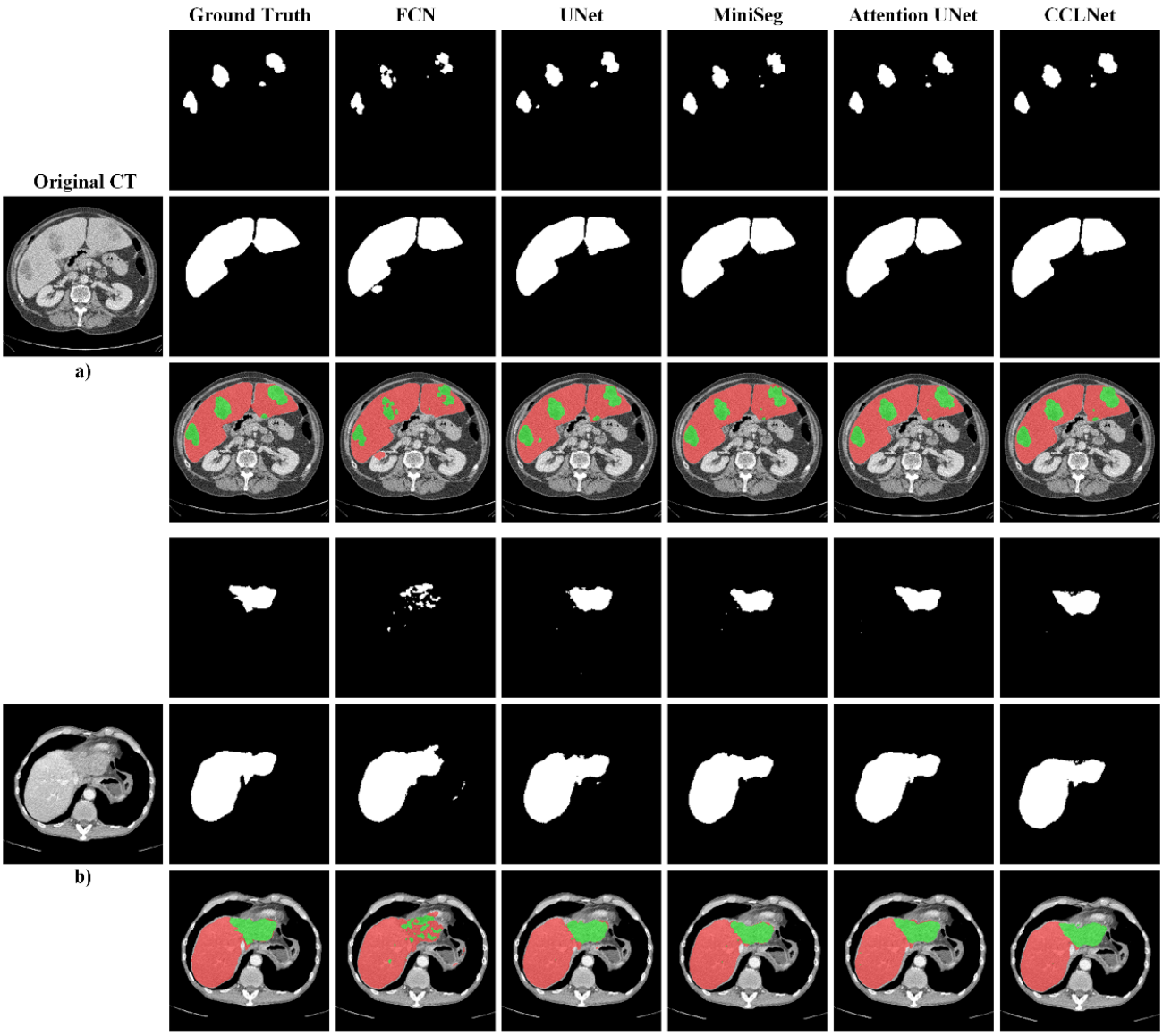

Figure 10. Visualization of liver and tumor segmentation results on the LiTS17 dataset. a) Liver with variable number and size of tumors; b) Liver with large tumors on the edge of parenchyma.

4.6. Comparisons on the 3DIRCADb dataset

4.6.1. Quantitative comparison

To verify the generalization ability of the proposed CCLNet network, we conduct experiments on the 3DIRCADb dataset. Under the same experimental environment configuration, we compared the performance of the proposed model with four networks of FCN, UNet, MiniSeg and Attention UNet. Tables 8 and 9 show the comparative results. As can be observed, the CCLNet network surpasses the other four approaches in every metric except RVD for the liver and tumor segmentation performance.

Table 8. Liver segmentation results in model comparisons on the 3DIRCADb dataset.

Method	Dice (%)	VOE (%)	RVD (%)	ASSD (mm)	RMSD (mm)
FCN	94.82 ± 2.53*	9.75 ± 4.50	-0.69 ± 2.38	2.39 ± 1.42	6.94 ± 4.29
UNet	95.88 ± 1.42*	7.89 ± 2.61	1.30 ± 1.78	1.78 ± 0.83	5.51 ± 2.66
MiniSeg	95.83 ± 1.73*	8.57 ± 3.14	**0.25 ± 1.03**	1.39 ± 0.82	4.66 ± 2.29
Attention UNet	96.46 ± 1.52*	6.82 ± 2.68	-0.47 ± 0.97	1.26 ± 1.68	2.70 ± 2.54
CCLNet	**96.84 ± 0.52**	**6.52 ± 0.98**	0.28 ± 1.26	**1.15 ± 0.54**	**2.56 ± 2.19**

Table 9. Tumor segmentation results in model comparisons on the 3DIRCADb dataset.

Method	Dice (%)	VOE (%)	RVD (%)	ASSD (mm)	RMSD (mm)
FCN	50.57 ± 37.39*	59.81 ± 31.82	-2.35 ± 5.69	19.77 ± 23.18	31.13 ± 34.51
UNet	61.78 ± 12.36*	47.43 ± 14.14	0.74 ± 3.53	7.98 ± 8.73	30.99 ± 18.52
MiniSeg	63.72 ± 18.92*	45.83 ± 16.27	**0.12 ± 2.84**	8.63 ± 8.11	19.05 ± 25.73
Attention UNet	69.47 ± 15.37*	32.12 ± 17.43	0.90 ± 1.07	12.42 ± 9.36	15.57 ± 14.20
CCLNet	**74.53 ± 10.51**	**29.04 ± 8.15**	-0.23 ± 0.94	**5.72 ± 6.33**	**7.87 ± 2.49**

On the 3DIRCADb dataset, Table 10 compares the performance of the SOTA algorithms in terms of parameters and training/test times. We can see that the proposed CCLNet needs to configure the most network parameters; besides, CCLNet takes somewhat longer to train and test than the other four models, at 2 h 42 m 27 s and 208 s, respectively. Such a cost is still reasonable, though, given the increase in model accuracy.

Table 10. Parameters and efficiency for five models on the 3DIRCADb dataset.

Model	Parameters	Training time	Test time
FCN	18,643,746	**2 h 7 m 56 s**	**197 s**
UNet	7,765,442	2 h 13 m 19 s	198 s
Miniseg	**82,666**	2 h 16 m 24 s	214 s
Attention UNet	34,878,638	4 h 31 m 48 s	221 s
CCLNet	37,706,674	2 h 42 m 27 s	208 s

4.6.2. Qualitative comparison

To compare the segmentation ability of the proposed CCLNet in complex situations (one with irregular tumors and the other with small tumors), In Figure 11, we contrast the suggested method with the other four methods.

In the liver segmentation task of Figure 11a), FCN, UNet and MiniSeg showed obvious under-segmentation errors. In the tumor segmentation task, all five methods incorrectly bisected the tumor. In Figure 11b), in the task of accurately segmenting the liver and tumor, all five networks demonstrated strong performance. However, compared with the other four methods, the edge processing by CCLNet is smoother.

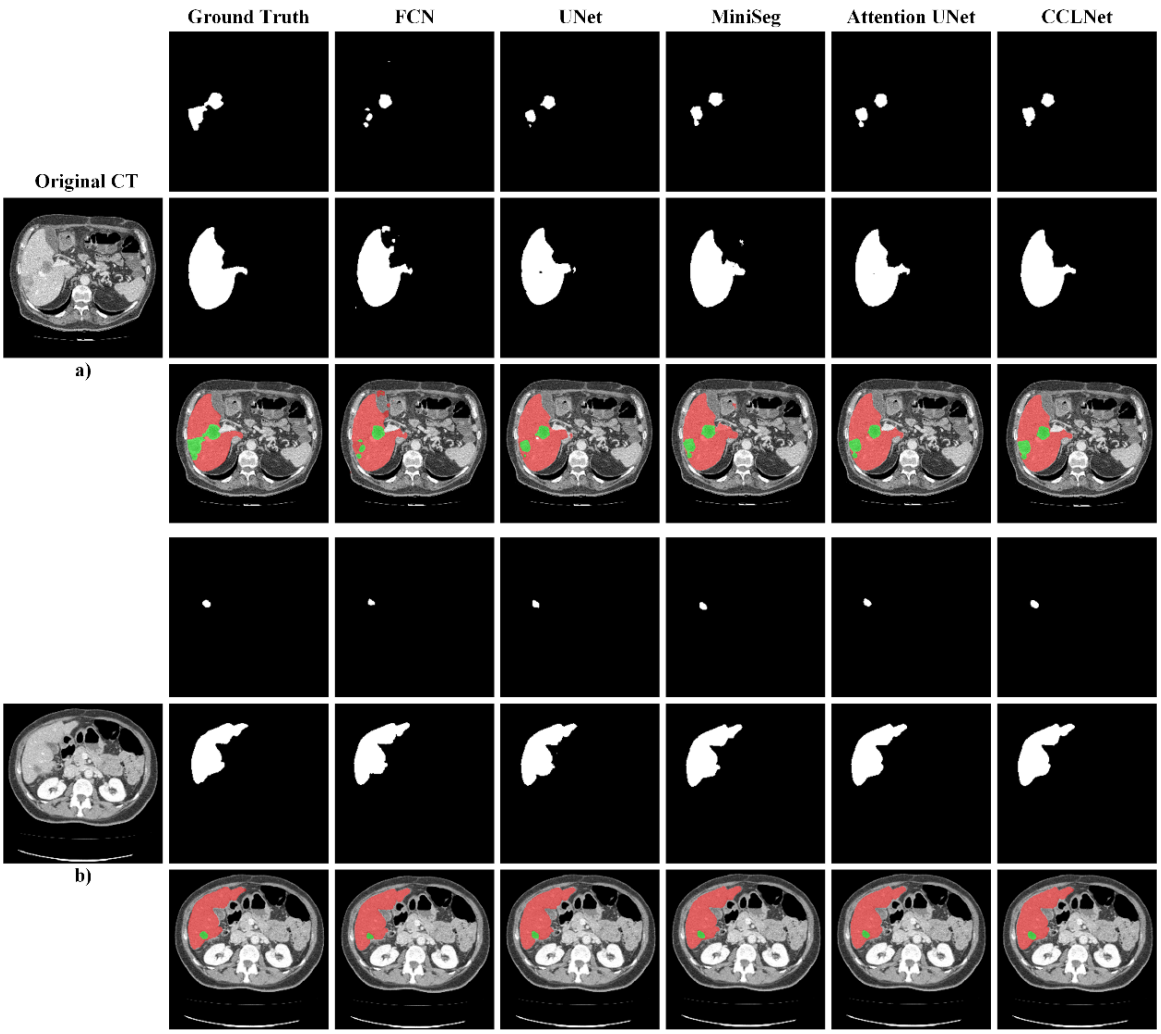

Figure 11. Visualization of liver and tumor segmentation outcomes of various techniques on the 3DIRCADb dataset. a) liver has an irregular size tumor; b) liver has a tiny tumor.

5. Conclusions

Considering the low accuracy of 2D network segmentation and the high computational demand of the 3D network, we proposed a 2.5D network, called CCLNet, for liver and tumor segmentation to balance between accuracy and cost. Specifically, we adopted a 2.5D mode to balance between accuracy and cost. In addition, we used ResNet-34 to enhance the segmentation accuracy. Furthermore, the receptive field is expanded by using the L-ASPP. To finally gather additional local and global feature data, we added the CCM module.

Our experiments were carried out on the LiTS17 and 3DIRCADb datasets. The proposed network can learn enough spatial information, according to experimental findings with higher accuracy, meanwhile saving computing resources and improving training efficiency with closer segmentation accuracy. Furthermore, the suggested method has advantages over existing methods in tumor segmentation and can accurately segment the liver. Nevertheless, this paper only conducts experiments in liver CT. The proposed hybrid architecture is essentially general, which may be used for a variety of medical picture segmentation tasks, including optical coherence tomography (OCT) which has wide

application in imaging different parts of human body [32–36]. In future work, we will conduct robustness tests on medical images of other morphologies and organs.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (No. 61701178), the Scientific Research Fund of Hunan Provincial Education Department (No. 22B0653).

Conflict of interest

There are no competing interests according to the authors.

References

1. J. Ferlay, M. Colombet, I. Soerjomataram, D. M. Parkin, M. Piñeros, A. Znaor, et al., Cancer statistics for the year 2020: An overview, *Int. J. Cancer*, **149** (2021), 778–789. https://doi.org/10.1002/ijc.33588
2. P. Bilic, P. Christ, B. H. Li, E. Vorontsov, A. Ben-Cohen, G. Kaissis, et al., The liver tumor segmentation benchmark (LiTs), *Med. Image Anal.*, **84** (2023), 102680. https://doi.org/10.1016/j.media.2022.102680
3. J. Calderaro, M. Ziol, V. Paradis, J. Zucman-Rossi, Molecular and histological correlations in liver cancer, *J. Hepatol.*, **71** (2019), 616–630. https://doi.org/10.1016/j.jhep.2019.06.001
4. J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Boston, USA, (2015), 3431–3440. https://doi.org/10.1109/CVPR.2015.7298965
5. A. Ben-Cohen, I. Diamant, E. Klang, M. Amitai, H. Greenspan, Fully convolutional network for liver segmentation and lesions detection, in *Deep Learning and Data Labeling for Medical Applications*, Springer, Athens, Greece, (2016), 77–85. https://doi.org/10.1007/978-3-319-46976-8_9
6. Y. Zhang, Z. He, C. Zhong, Y. Zhang, Z. Shi, Fully convolutional neural network with post-processing methods for automatic liver segmentation from CT, in 2017 *Chinese Automation Congress (CAC)*, IEEE, Jinan, China, (2017), 3864–3869. https://doi.org/10.1109/CAC.2017.8243454
7. H. Jiang, T. Shi, Z. Bai, L. Huang, Ahcnet: an application of attention mechanism and hybrid connection for liver tumor segmentation in CT volumes, *IEEE Access*, **7** (2019), 24898–24909. https://doi.org/10.1109/access.2019.2899608

8. F. P. Christ, A. E. M. Elshaer, F. Ettlinger, S. Tatavarty, M. Bickel, P. Bilic, et al., Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, Athens, Greece, (2016), 415–423. https://doi.org/10.1007/978-3-319-46723-8_48
9. O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, Munich, Germany, (2015), 234–241. https://doi.org/10.1007/978-3-319-24574-4_28
10. H. Seo, C. Huang, M. Bassenne, R. Xiao, L. Xing, Modified U-Net (mU-Net) with incorporation of object-dependent high level features for improved liver and liver-tumor segmentation in CT images, *IEEE Trans. Med. Imaging*, **39** (2019), 1316–1325. https://doi.org/10.1109/TMI.2019.2948320
11. J. Wang, P. Lv, H. Wang, C. Shi, SAR-U-Net: squeeze-and-excitation block and atrous spatial pyramid pooling based residual U-Net for automatic liver segmentation in computed tomography, *Comput. Methods Programs Biomed.*, **208** (2021), 106268. https://doi.org/10.1016/j.cmpb.2021.106268
12. Q. Jin, Z. Meng, C. Sun, H. Cui, R. Su, RA-UNet: a hybrid deep attention-aware network to extract liver and tumor in CT scans, *Front. Bioeng. Biotechnol.*, **8** (2020), 1471. https://doi.org/10.3389/fbioe.2020.605132
13. X. Li, H. Chen, X. Qi, Q. Dou, C. W. Fu, P. A. Heng, H-DenseUNet: hybrid densely connected Unet for liver and tumor segmentation from CT volumes, *IEEE Trans. Med. Imaging*, **37** (2018), 2663–2674. https://doi.org/10.1109/TMI.2018.2845918
14. P. Lv, J. Wang, H. Wang, 2.5D lightweight RIU-Net for automatic liver and tumor segmentation from CT, *Biomed. Signal Process. Control*, **75** (2022), 103567. https://doi.org/10.1016/j.bspc.2022.103567
15. L. Meng, Q. Zhang, S. Bu. Two-stage liver and tumor segmentation algorithm based on convolutional neural network, *Diagnostics*, **11** (2021), 1806. https://doi.org/10.3390/diagnostics11101806
16. F. Özcan, N. O. Uçan, S. Karaçam, D. Tunçman, Fully automatic liver and tumor segmentation from CT image using an AIM-Unet, *Bioengineering*, **10** (2023), 215. https://doi.org/10.3390/bioengineering10020215
17. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, et al., Generative adversarial networks, *Commun. ACM*, **63** (2020), 139–144. https://doi.org/10.1007/978-3-030-50017-7_16
18. Y. Enokiya, Y. Iwamoto, W. Y. Chen, X. H. Han, Automatic liver segmentation using U-net with Wasserstein GANs, *Int. J. Image Graphics*, **6** (2018), 152–159. https://doi.org/10.18178/joig.7.3.94-101
19. D. Yang, D. Xu, S. K. Zhou, B. Georgescu, M. Chen, S. Grbic, et al., Automatic liver segmentation using an adversarial image-to-image network, in *20th International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, Quebe, Canada, (2017), 507–515. https://doi.org/10.1007/978-3-319-66179-7_58

20. U. Demir, Z. Zhang, B. Wang, M. Antalek, E. Keles, D. Jha, et al., Transformer based generative adversarial network for liver segmentation, *arXiv preprint*, (2022), arXiv:2205.10663. https://doi.org/10.48550/arXiv.2205.10663
21. C. Xu, Y. Wang, D. Zhang, L. Han, Y. Zhang, J. Chen, et al., BMAnet: boundary mining with adversarial learning for semi-supervised 2D myocardial infarction segmentation, *IEEE J. Biomed. Health. Inf.*, **27** (2022), 87–96. https://doi.org/10.1109/JBHI.2022.3215536
22. L. Chen, H. Song, C. Wang, Y. Cui, J. Yang, X. Hu, et al., Liver tumor segmentation in CT volumes using an adversarial densely connected network, *BMC Bioinf.*, **20** (2019), 1–13. https://doi.org/10.1186/s12859-019-3069-x
23. A. G Roy, N. Navab, C. Wachinger, Concurrent spatial and channel 'squeeze & excitation' in fully convolutional networks, in *21th International Conference on Medical Image Computin and Computer-Assisted Intervention*, Springer, Granada, Spain, (2018), 421–429. https://doi.org/10.1007/978-3-030-00928-1_48
24. T. Lei, R. Wang, Y. Zhang, Y. Wan, C. Liu, A. K. Nandi, DefED-Net: deformable encoder-decoder network for liver and liver tumor segmentation, *IEEE Trans. Radiat. Plasma Med. Sci.*, **6** (2021), 68–78. https://doi.org/10.1109/TRPMS.2021.3059780
25. T. C. Nguyen, T. P. Nguyen, G. H. Diep, A. H. Tran-Dinh, T. V. Nguyen, M. T. Tran, Ccbanet: cascading context and balancing attention for polyp segmentation, in *24th International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, Strasbourg, France, (2021), 633–643. https://doi.org/10.1007/978-3-030-87193-2_60
26. T. Heimann, B. Van Ginneken, M. Styner, Y. Arzhaeva, V. Aurich, C. Bauer, et al., Comparison and evaluation of methods for liver segmentation from CT datasets, *IEEE Trans. Med. Imaging*, **28** (2009), 1251–1265. https://doi.org/10.1109/TMI.2009.2013851
27. M. W. Li, D. Y. Xu, J. Geng, W. C. Hong, A hybrid approach for forecasting ship motion using CNN–GRU–AM and GCWOA, *Appl. Soft Comput.*, **114** (2022), 108084. https://doi.org/10.1016/j.asoc.2021.108084
28. Y. Qiu, Y. Liu, S. Li, J. Xu, Miniseg: An extremely minimum network for efficient covid-19 segmentation, in *Proceedings of the AAAI Conference on Artificial Intelligence*, AAAI, (2021), 4846–4854. https://doi.org/10.1609/aaai.v35i6.16617
29. O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, et al., Attention U-Net: learning where to look for the pancreas, *Arxiv preprint*, (2018), arXiv:1804.03999. https://doi.org/10.48550/arXiv.1804.03999
30. D. T. Kushnure, S. 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. https://doi.org/10.1016/j.cmpb.2021.106501
31. Y. Chen, C. Zheng, F. Hu, T. Zhou, L. Feng, G. Xu, et al., Efficient two-step liver and tumour segmentation on abdominal CT via deep learning and a conditional random field, *Comput. Biol. Med.*, **150** (2022), 106076. https://doi.org/10.1016/j.compbiomed.2022.106076
32. R. K. Meleppat, C. R. Fortenbach, Y. Jian, E. S. Martinez, K. Wagner, B. S. Modjtahedi, et al., In Vivo imaging of retinal and choroidal morphology and vascular plexuses of vertebrates using swept-source optical coherence tomography, *Transl. Vision Sci. Technol.*, **11** (2022), 11. https://doi.org/10.1167/tvst.11.8.11

33. K. M. Ratheesh, L. K. Seah, V. M. Murukeshan, Spectral phase-based automatic calibration scheme for swept source-based optical coherence tomography systems, *Phys. Med. Biol.*, **61** (2016), 7652. https://doi.org/10.1088/0031-9155/61/21/7652
34. R. K. Meleppat, K. E. Ronning, S. J. Karlen, K. K. Kothandat, M. E. Burns, E. N. Pugh, et al., In situ morphologic and spectral characterization of retinal pigment epithelium organelles in mice using multicolor confocal fluorescence imaging, *Invest. Ophthalmol. Visual Sci.*, **61** (2020), 1. https://doi.org/10.1167/iovs.61.13.1
35. R. K. Meleppat, C. Shearwood, S. L. Keey, M. V. Matham, Quantitative optical coherence microscopy for the in situ investigation of the biofilm, *J. Biomed. Opt.*, **21** (2016), 127002–127002. https://doi.org/10.1117/1.JBO.21.12.127002
36. V. M. Murukeshan, L. K. Seah, C. Shearwood, Quantification of biofilm thickness using a swept source based optical coherence tomography system, in *International Conference on Optical and Photonic Engineering* (icOPEN 2015), SPIE, Singapore, (2015), 683–688. https://doi.org/10.1117/12.2190106

AIMS Press