Mathematical Biosciences and Engineering

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

MBE, 19(5): 4703–4718.
DOI: 10.3934/mbe.2022219
Received: 12 January 2022
Revised: 20 February 2022
Accepted: 24 February 2022
Published: 09 March 2022

Research article

Residual based attention-Unet combing DAC and RMP modules for automatic liver tumor segmentation in CT

Rongrong Bi[1], Chunlei Ji[1], Zhipeng Yang[2], Meixia Qiao[1], Peiqing Lv[2,*] and Haiying Wang[2]

[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

* **Correspondence:** Email: peiqinglv@163.com; Tel: +8618769171796; Fax: +86045186390850.

Abstract: *Purpose*: Due to the complex distribution of liver tumors in the abdomen, the accuracy of liver tumor segmentation cannot meet the needs of clinical assistance yet. This paper aims to propose a new end-to-end network to improve the segmentation accuracy of liver tumors from CT. *Method*: We proposed a hybrid network, leveraging the residual block, the context encoder (CE), and the Attention-Unet, called ResCEAttUnet. The CE comprises a dense atrous convolution (DAC) module and a residual multi-kernel pooling (RMP) module. The DAC module ensures the network derives high-level semantic information and minimizes detailed information loss. The RMP module improves the ability of the network to extract multi-scale features. Moreover, a hybrid loss function based on cross-entropy and Tversky loss function is employed to distribute the weights of the two-loss parts through training iterations. *Results*: We evaluated the proposed method in LiTS17 and 3DIRCADb databases. It significantly improved the segmentation accuracy compared to state-of-the-art methods. *Conclusions*: Experimental results demonstrate the satisfying effects of the proposed method through both quantitative and qualitative analyses, thus proving a promising tool in liver tumor segmentation.

Keywords: liver tumor; segmentation; CT; residual; attention

1. Introduction

Among all organ tumors, the mortality of liver tumors is much higher than that of others [1]. Many treatments are available for liver tumors, in which the directional resection of the focus area is one of the fundamental approaches for liver tumors. In addition, the rapid development of various imaging techniques (e.g., Ultrasound, Computed Tomography (CT), Optical coherence tomography (OCT), Magnetic Resonance (MRI), etc.) also provides the possibility of clinical computer-aided diagnosis [2–4]**.** However, the tumor's location, size, and contrast affect the difficulty of resection and postoperative recovery (e.g., two challenges are shown in Figure 1). Therefore, automatic and accurate segmentation of liver tumors is extremely valuable in the clinical environment.

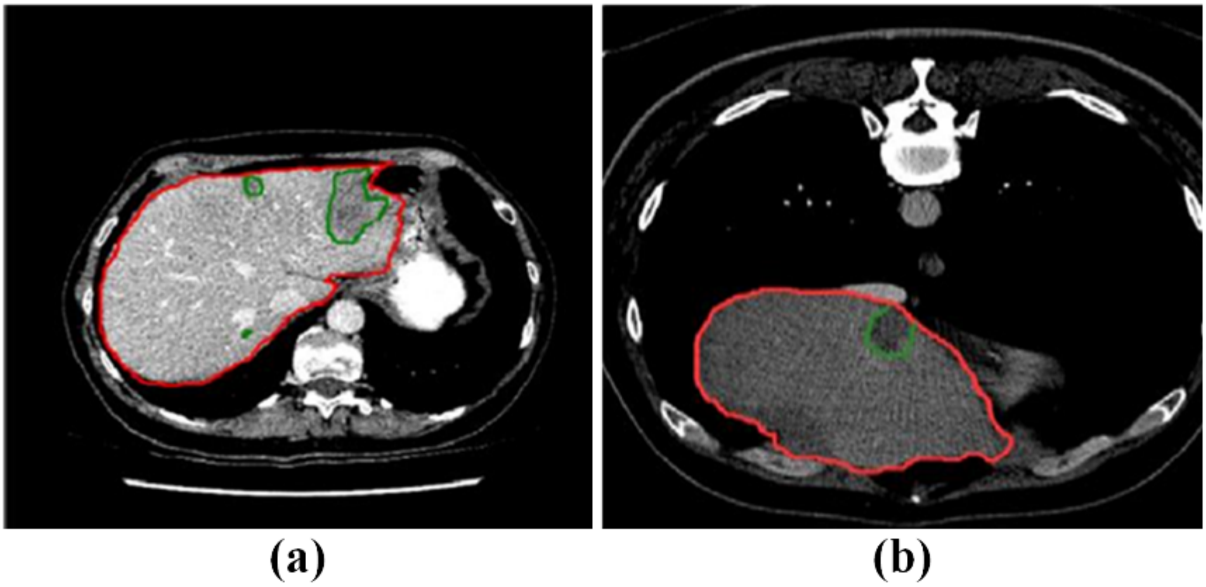


Figure 1. Challenges for liver tumor segmentation. (a) Blurred contour of liver tumor. (b) Low contrast between liver tumor and parenchyma.

With the rapid development of computer hardware for the past few decades, scholars have proposed many sophisticated deep learning-based segmentation methods. For example, Mu et al. [5] proposed a progressive global perception and local polishing (PCPLP) network to automatically segment COVID-19 infections in CT images, integrating multi-scale features with multi-level features. Experimental evaluations show that the proposed PCPLP effectively improves the learning ability to identify well-defined lung infection regions accurately. Zhou et al. [6] suggested a 3D semantic V-net (SV-net) for extraocular muscle (EOM) and optic nerve (ON) extractions in orbital CT images. Qualitative and Quantitative evaluation showed that the method has good performance in automatically extracting total EOM and ON in orbital CT images and measuring their volumes. Finally, Zhao et al. [7] proposed a combination of a 3D network and a shape deformation module implemented by a spatial transformation network (STN) to achieve automatic segmentation of lung parenchyma, and the experimental results showed that the method is helpful for clinical diagnosis of COVID-19 infection in CT images. Ronneberger et al. [8] proposed the U-Net network, which adopts the skip connection to minimize the loss of detail caused by upsampling. Thanks to its excellent segmentation ability, the U-Net network has become the benchmark network of many medical segmentation models.

Many scholars have made extensive efforts on U-Net. For example, Oktay et al. [9] applied the attention mechanism to the U-Net and proposed the attention-Unet. The network can adjust the weight of some feature map elements to make the network focus on easily misclassified targets. Chen et al. [10]

proposed DeepLab series networks. First, they added void convolution to the convolution module of FCN [11]. Then, the authors [12] also proposed an atrous spatial pyramid pooling (ASPP) to extract context information of different scales and applied this module to a DeepLab network. Alom et al. [13] applied cyclic residual convolution to U-Net and proposed the R2UNet network with the highest sensitivity on the ISIC dataset. Then, they added the attention mechanism threshold and proposed the attention-R2UNet. In addition, Wang et al. [14] introduced Squeeze-and-Excitation (SE) block, ASPP, and residual block into U-Net for robust liver segmentation. Zongwei et al. [15] improved the hop connection of U-Net and proposed a mesh UNet++ network, which connects different levels of UNet networks through the cascade method.

Moreover, there are also variants of U-type networks, such as V-type and M-type networks. Arnab et al. [16] proposed the V-Net network, which covers the scope of application of U-Net to 3D medical images for the first time. Li et al. [17] proposed a 2.5D semantic segmentation model H-DenseUNet based on V-Net. They present a transformation processing function to transform 3D CT volumes into 2D adjacent slices. Finally, Jun et al. [18] proposed the M-Net network, which uses multi-scale input to extract the features of four-scale input images. To improve the network's ability on multi-scale targets, the author proposed a loss function based on multi-label and used polar coordinates to convert to different scale layers to generate local prediction maps. As a result, the network achieves the state-of-the-art effect on the dataset of optic cups and discs.

Some scholars also employed a non-coding-decoding structure, such as bilateral networks proposed by Chang et al. [19]. They abandoned the traditional encoding-decoding design but adopted a novel bilateral network instead. The bilateral network comprises spatial path, semantic path, and feature fusion module connecting the two, in which the spatial path mainly extracts the detailed features of the image. Meanwhile, the semantic path extracts the abstract semantic information by increasing the range of receptive fields through continuous downsampling.

Nevertheless, the end-to-end semantic segmentation model still has some limitations, which can be outlined as follows:

(i) The conventional U-Net is suitable for large targets but not low contrast tumors and small ones.

(ii) Although the segmentation effect of attention-Unet has been improved, the model's speed is limited since there is no downsampling process in the spatial branch.

This paper added the CE structure with the DAC and RMP modules to attention-UNet and infused it with the Residual module to propose the ResCEAttUNet network. In summary, the main contributions of this paper are as follows:

- Use a 2D end-to-end model, which converges faster and has fewer parameters than 2.5D and 3D networks.
- Employ the CE structure via extracting spatial details and multi-scale semantic information to improve tumor edges' segmentation effect and reduce the misclassification rate of small tumors.
- Integrate the residual and batch normalization (BN) modules to the network to improve convergence speed.

We organize the rest of the paper as follows: Section 2 describes the proposed method in detail. Section 3 provided the experiments and results. Finally, Section 4 summarizes the work.

2. Methods

2.1. Proposed ResCEAttUnet network

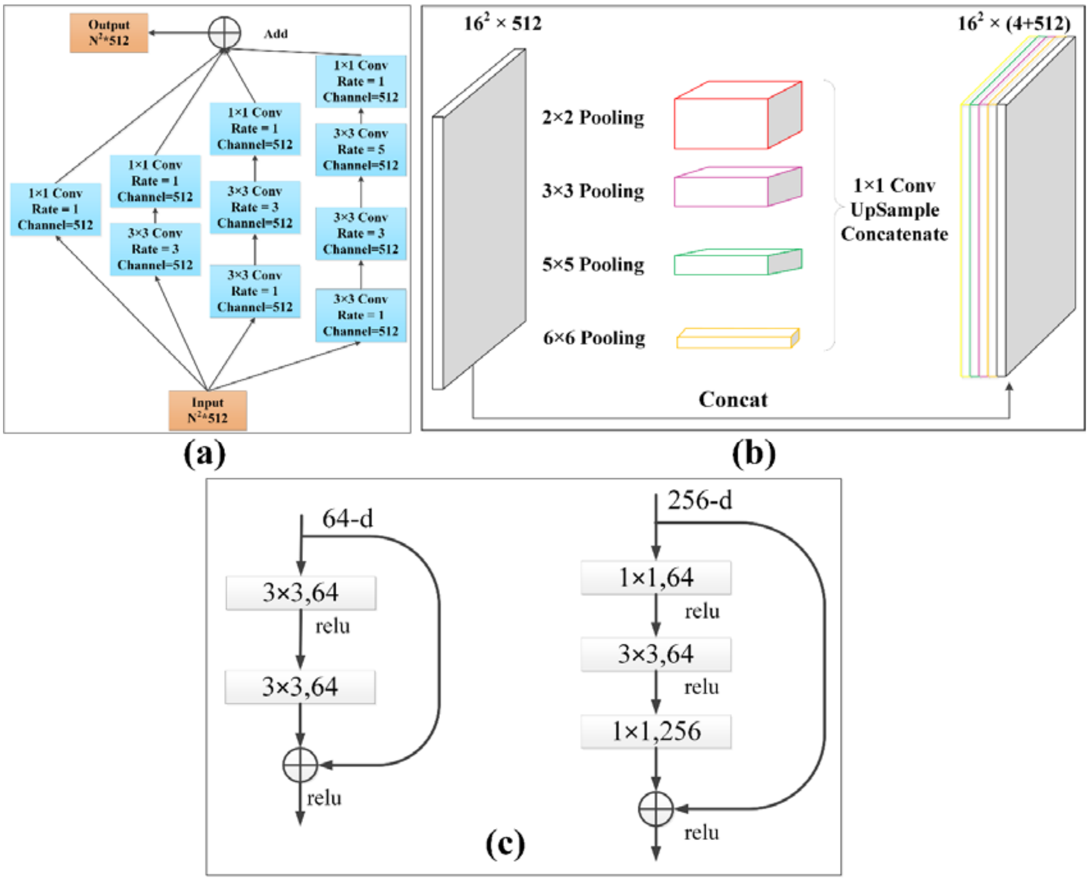


Figure 2. Primary modules of the proposed network. (a) DAC. (b) RMP. (c) Residual Block.

To improve the segmentation ability of the network for tumors with different scales, we employed the improved CE structure proposed by Gu et al. [20]. It comprises DAC and RMP blocks. We directly apply this CE module to the attention-Unet. To ensure fair comparisons, we use the official source code provided by Attention-Unet and CE-Net. All the experimental models in this paper are based on Attention-Unet, using the same parameters.

Figure 2(a) shows the structure of the DAC module. The dimension of the output feature map of the structure is consistent with the input one, which minimizes the loss of spatial information while extracting semantic information. We connect the DAC module to the end of the shrinkage path of attention U-Net for multi-scale features.

Figure 2(b) shows the structure of RMP. We adjusted some designs in PSPNET [21] to adapt to liver tumor segmentation. Specifically, the global pooling is removed, and the pool sizes of the four branches are set to 2, 3, 5, and 6, respectively. In addition, to reduce the model parameters, the number of channels of the feature map is reduced to 1 by using 1 × 1 convolution kernel. Finally, we link the DAC and RMP modules to form the CE net. Then, we connect this structure with the shrinkage path of attention-Unet and get the CE-Att-Unet.

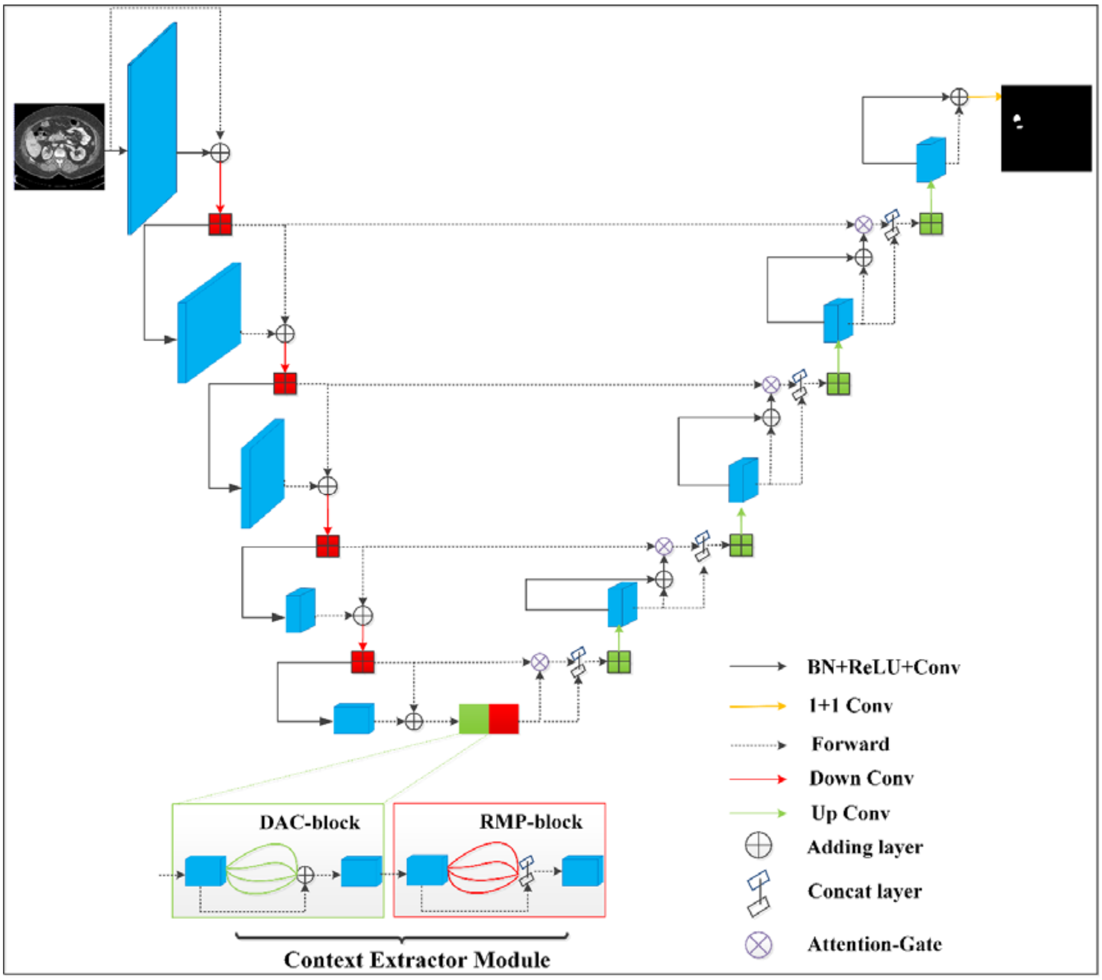


Figure 3. The structure of the proposed ResCEAttUnet network.

Furthermore, to improve the model's generalization ability, we add the BN layer to all the convolution modules in the CE-Att-UNet to reduce the model's sensitivity to parameter initialization. Moreover, to avoid the problem of network degradation, we employ the residual module based on bottleneck convolution (as shown in Figure 2(c)). Finally, the residual and BN form a new convolution module, ResNet-BN. Then, we applied the ResNet-BN to CE-Att-UNet and obtained the proposed network ResCEAttUnet (shown in Figure 3).

Compared with CE-NET, the proposed ResCEAttUnet is improved in the following aspects: First, we use attention-Unet as the basic framework. Then, we use DAC and RMP in CE-Net as the connection point between encoding and decoding regions in attention-Unet. Finally, each convolutional sequence in attention-Unet was replaced with residual connection and BN layer.

2.2. Loss function

For semantics segmentation, the loss function based on cross-entropy treats each pixel independently, which results in the loss function lacking global information attention. Meanwhile, the Tversky function based on cross-ratio can evaluate the segmentation result as a whole. Its definition is shown in Eq (1).

$$L_{TL}(\alpha,\beta)=1-\frac{\sum_{i=0}^{l}\sum_{j=0}^{N}p_{ij}g_{ij}}{\sum_{i=0}^{l}\sum_{j=0}^{N}p_{ij}g_{ij}+\alpha\sum_{i=0}^{l}\sum_{j=0}^{N}(p_{ij}g_{\tilde{ij}})^2+\beta\sum_{i=0}^{l}\sum_{j=0}^{N}(p_{\tilde{ij}}g_{ij})^2} \tag{1}$$

where i is the index of the channel in the real image, j is the index of the pixel in the predicted image, l is the total number of classes in the segmentation task, N is the total number of pixels in the image, p is the probability of the output of the classification model, and g is true. We use p_{ij} to denote the probability that pixel j belongs to class i at the time of prediction, and g_{ij} to denote the probability that pixel j of the real image belongs to class i. If pixel j in the input image belongs to class 0, then $g_{0j} = 1$, and g_{1j}, g_{2j}, ..., $g_{lj} = 0$. In addition, two parameters α and β are used to adjust the ratio between false positives and false negatives, and the sum of α and β is always equal to 1.

However, the training process and the segmentation effect are not stable enough. Therefore, we developed a hybrid loss function that combines the two parts with its equation shown in Eq (2).

$$L_H(\alpha,\beta)=\left(1-\frac{n}{N}\right)L_{CE}+\rho\frac{n}{N}L_{TL}(\alpha,\beta) \tag{2}$$

where L_{CE} represents a loss function based on cross-entropy, L_{TL} denotes the Tversky loss function, n denotes the number of iterations of the current training, and N denotes the total number of training iterations. ρ represents the equalization factor used to control two-loss values on the same order of magnitude. α and β represent the weighted proportions of Tversky false-negative and false-positive, respectively.

For semantic segmentation of images, the loss function based on cross-entropy treats each pixel independently without paying attention to the global information. As a result, the optimization process is simple, and the training is stable. Therefore, the cross-entropy based on the pixel points can be used to segment the target quickly and stably in the early stage of training. In the later stage of training, the rough results obtained from the previous segmentation are refined based on the result-oriented Tversky loss function. This coarse-to-fine strategy can ensure that the hybrid loss function can take advantage of different strengths at different stages.

2.3. Evaluation metrics

We used five metrics to evaluate the proposed method in the experiment, including *Sensitivity*, *Specific,* Intersection over Union (*IoU*), *Dice* coefficient, and *Hausdorff.* Their definitions are provided in Eqs (3)–(7) below:

$$Sensitivity=\frac{TP}{TP+FN} \tag{3}$$

$$Specifity=\frac{TN}{FP+TN} \tag{4}$$

$$IoU=\frac{TP}{TP+FP+FN} \tag{5}$$

$$Dice=\frac{2*TP}{2*TP+FP+FN} \tag{6}$$

$$d_H\{X,Y\} = \max\{d_{XY}, d_{YX}\} = \max\{max_{x\in X}mind_{y\in Y}(x,y), max_{y\in Y}mind_{x\in X}(x,y) \qquad (7)$$

where *TP*, *TN*, *FP*, and *FN* represent true positive, true negative, false positive, and false negative. d_{XY}/d_{YX} represents the maximum value of the shortest distance from any point on edge *X/Y* to edge *Y/X*. The *Hausdorff* coefficient of edge *X* and *Y* is the larger value of d_{XY} and d_{YX}, which is used to measure the correlation measure of the surface distance between the labeled tumor and the predicted tumor.

3. Experiments and results

3.1. Data and implementation

The experiment is evaluated in the LiTS17 [22] and 3Dircadb (https://www.ircad.fr/research/3dircadb/) databases. The LiTS17 dataset contains 201 CT datasets, and each set is annotated by three different doctors for the location of the liver and the tumor. The dataset is divided into the training set and test set, of which 131 for training and 70 for the test. The slice spacing was between [0.45 mm, 6.0 mm], the cross-sectional resolution distribution was between [0.55 mm, 1.0 mm], and the number of Z-axis slices varied from 42 to 1026.

The 3Dircadb dataset contains venous phase abdominal-enhanced CT data of ten men and ten women, 15 of which include liver tumors with ground truth.

In our experiments, 15 sets of LiTS dataset were randomly selected for testing, 5 sets of 3Dircadb were randomly selected for testing, and the remaining data were used in a 3:1 ratio for training and validation. We did not perform data augmentation in this experiment.

All the experiments are run on a workstation with Ubuntu 18.04 operating system, graphics card RTX2080Ti, memory 64G, single CPU Intel Xeon Silver 4110, and using the Pytorch1.4 deep learning framework for implementation. The initial learning rate of Adam of the network is set to 0.01, and the initial learning rate of Adam of other networks is 1e-4.

3.2. Comparison of the different loss function

To illustrate the impact of the hybrid loss function hyperparameter setting more directly on the segmentation results. We compared the segmentation results of the CE-Att-UNet network using different hyperparameters in Figure 4.

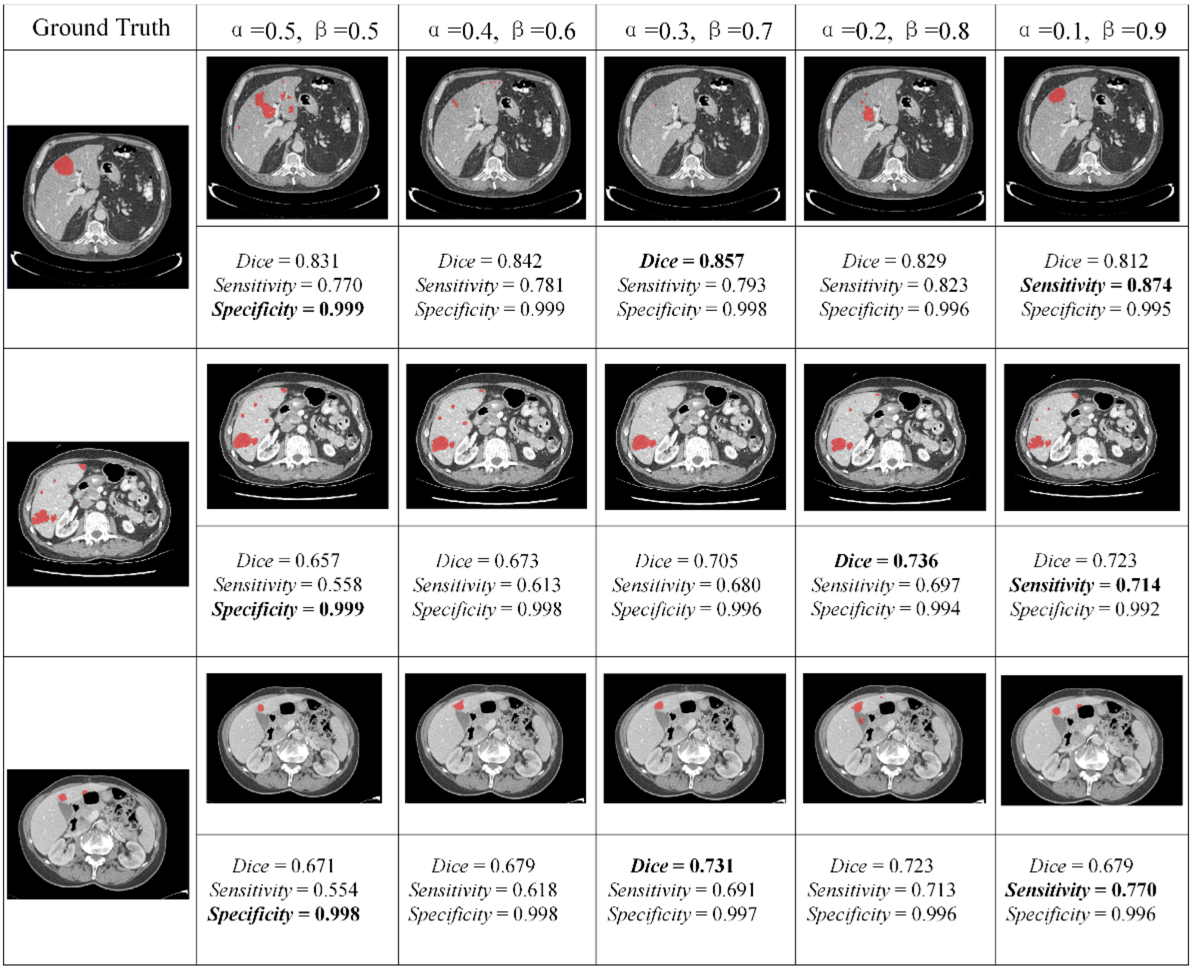
It can be seen from Figure 4 that, when the parameter α of the hybrid loss function is relatively large, the proportion of non-tumor pixels in CT slices that are incorrectly classified as tumor pixels is relatively low. It indicates that a larger α can effectively suppress false positives. Meanwhile, When the loss function β parameter is relatively large, the segmentation effect of the tumor edge and small target tumor in the CT slice is better, which indicates that a larger β can effectively suppress false negatives.

Figure 5 shows the change curve of the loss function value of the CE-Att-UNet network during training. The red/blue/green dashed line represents the Dice change of the cross-entropy/Tversky/mixed loss training model on the training set.

It can be seen from Figure 5 that, the cross-entropy loss function based on single-pixel point optimization can converge quickly at the beginning of model training. In the first training round, the Dice of the model reaches 0.60. Because the cross-entropy function has a higher weight at the beginning of training, the convergence speed of the hybrid loss function is also faster.

In addition, Figure 5 also shows that the training networks using cross-entropy loss and hybrid loss are close to full convergence when the model training reaches the seventh round. At the beginning of training, the segmentation effect of the Tversky is poor, so the function converges slowly. When the training reaches the 12th round, the training effect of the Tversky function exceeds the cross-entropy loss function, and it will not approach full convergence until the 16th round.

Figure 5 demonstrates that the hybrid loss function combines the advantages of the cross-entropy and the Tversky loss function. At the beginning of training, the hybrid loss function enables a quick converge, meanwhile refines the segmentation result at the end of the training, thus proving in a superior performance than the other two-loss functions.



Ground Truth	α =0.5, β =0.5	α =0.4, β =0.6	α =0.3, β =0.7	α =0.2, β =0.8	α =0.1, β =0.9
	Dice = 0.831 *Sensitivity* = 0.770 ***Specificity* = 0.999**	*Dice* = 0.842 *Sensitivity* = 0.781 *Specificity* = 0.999	***Dice* = 0.857** *Sensitivity* = 0.793 *Specificity* = 0.998	*Dice* = 0.829 *Sensitivity* = 0.823 *Specificity* = 0.996	*Dice* = 0.812 ***Sensitivity* = 0.874** *Specificity* = 0.995
	Dice = 0.657 *Sensitivity* = 0.558 ***Specificity* = 0.999**	*Dice* = 0.673 *Sensitivity* = 0.613 *Specificity* = 0.998	*Dice* = 0.705 *Sensitivity* = 0.680 *Specificity* = 0.996	***Dice* = 0.736** *Sensitivity* = 0.697 *Specificity* = 0.994	*Dice* = 0.723 ***Sensitivity* = 0.714** *Specificity* = 0.992
	Dice = 0.671 *Sensitivity* = 0.554 ***Specificity* = 0.998**	*Dice* = 0.679 *Sensitivity* = 0.618 *Specificity* = 0.998	***Dice* = 0.731** *Sensitivity* = 0.691 *Specificity* = 0.997	*Dice* = 0.723 *Sensitivity* = 0.713 *Specificity* = 0.996	*Dice* = 0.679 ***Sensitivity* = 0.770** *Specificity* = 0.996

Figure 4. Comparative results of the hybrid loss function using different hyperparameters.

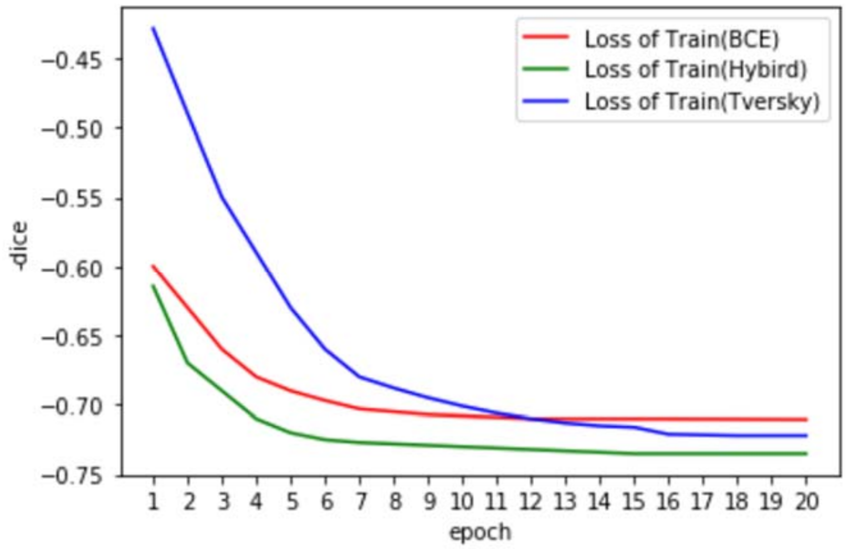


Figure 5. Comparation on Dice value using different loss functions during the training.

3.3. Ablation

This section evaluates the effectiveness of our proposed network framework by employing different structures and loss functions.

Table 1 shows that CE-Att-Unet(H) and ResCEAttUnet (H) obtained the best performance with the hybrid loss function of α = 0.3, β = 0.7. Furthermore, for the soft-dice loss functions, the Dice score of ResCEAttUnet(s), which added the ResNet-BN module, is higher than that of CE-Att-Unet(s). In contrast, the ResCEAttUnet(H) improves the Dice score compared with CE-Att-Unet(H) for the hybrid loss functions. Thus, we can conclude that the enhanced ResCEAttUnet based on residual and BN convolution modules can improve the effect of liver tumor segmentation.

Table 1. Comparison of experimental results of different network structures.

Network	Loss	Dice	IoU	Hausdorff	Sensitive	Specificity
U-Net(s)	soft-dice	0.6662	0.5163	3.0939	0.6657	0.9979
Attention-Unet(s)	soft-dice	0.6896	0.5478	3.0346	0.6390	**0.9987**
CE-Net(H)	Hybrid	0.6972	0.5507	3.0429	0.6691	0.9915
CE-Att-UNet(s)	soft-dice	0.7068	0.5641	3.0586	0.6657	0.9978
CE-Att-UNet(H)	Hybrid	0.7256	0.5839	2.9766	0.7148	0.9973
ResCEAttUnet(s)	soft-dice	0.7210	0.5862	2.9490	0.6892	0.9984
ResCEAttUnet(H)	Hybrid	0.7280	0.5872	**2.8170**	0.6932	0.9985
Pre-ResCEAttUnet(H)	Hybrid	**0.7294**	**0.5892**	2.8733	**0.7157**	0.9980

Note: Bold font indicates the best value for each metric.

The ResCEAttUnet network trained by loading the pre-training model weight of Resnet-34 improves the Dice coefficient more than the network trained from zero. Besides, the attention-Unet showed the best performance on the specificity, while the ResCEAttUnet network is slightly inferior to the benchmark network. Therefore, the RMP module in the CE structure weakens attention's ability on suppressing false positives. Nevertheless, the ResCEAttUnet network is optimal on Dice and other evaluation metrics. Therefore, the most significant effect of migration learning by loading the pre-training model is speed up the convergence.

In Figure 6, we visualize the role of adding residual structure and the effect of loading the pre-training model for migration learning through the change process of the loss function. Finally, we intercept the changes of the hybrid loss function in CE-Att-Unet, ResCEAttUnet, and the migration learning network loaded with pre-training network Resnet-34 in the training and verification set in the first 14 epochs.

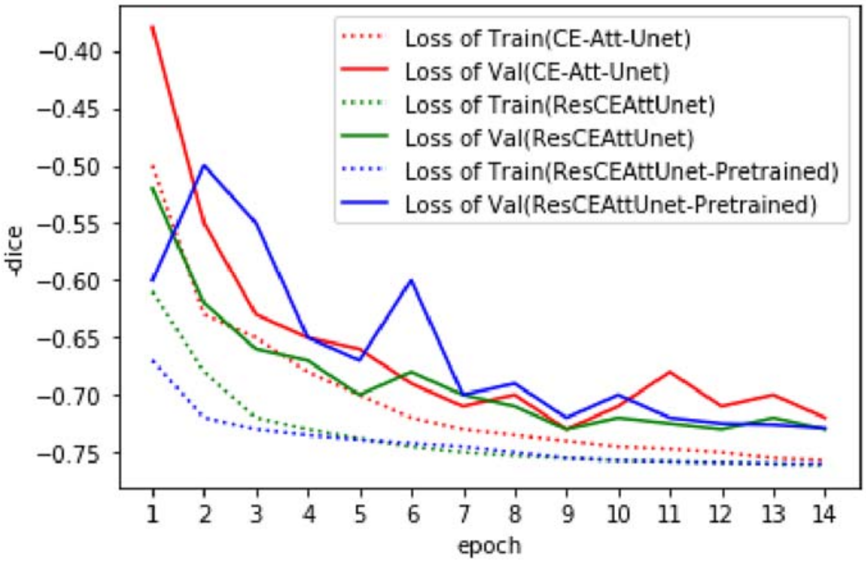


Figure 6. Comparison of loss functions using different networks.

The loss functions of the above networks are dynamic weighted hybrid loss functions, with α = 0.3, β = 0.7. In the training process, we can see that the convergence speed of ResCEAttUnet is faster. The reasons are as follows: (i) The convolution module of ResCEAttUnet has a BN layer, effectively reducing the internal skew variable shift problem. Therefore, the initial learning rate of the network is set relatively large. (ii) The convolution module of ResCEAttUnet employs the residual structure, which converts the original mapping into identity mapping, speeding up the model's convergence.

By comparing the loss function curves of CE-Att-Unet and ResCEAttUnet, we can see that ResCEAttUnet shows a superior segmentation effect on the verification set. Besides, the convergence speed of the model has been significantly improved by directly loading the weight of the pre-training model. In the primary stage of model training, although the effect of the verification set fluctuates, it eventually tends to be stable.

We can conclude from the ablation analysis that with the continuous improvement of the network structure, the final ResCEAttUnet based on the residuals and BN convolution module effectively improves the accuracy of liver tumor segmentation, and finally, with the ResCEAttUnet network loaded with the pre-trained model weights of Resnet-34. Furthermore, experimental results show that migration learning by loading pre-trained models can further improve segmentation accuracy and accelerate model convergence.

3.4. Comparison with state-of-the-art methods

To verify the effectiveness and robustness of the ResCEAttUnet network proposed in this paper for liver tumor segmentation, we selected some classical semantic segmentation models for comparison, including UNet, R2UNet, D-LinkNet, and UNet++.

Table 2 compares the segmentation effects of the four commonly used benchmark networks and our proposed ResCEAttUnet. On Dice, IOU, Hausdorff, and sensitivity, the proposed ResCEAttUnet in this paper are optimal, but only the specificity is slightly inferior to D-LinkNet.

Table 2. Comparative results on five main metrics using different networks.

Network	DICE	IoU	Hausdorff	Sensitive	Specificity
U-Net [8]	0.666190	0.516327	3.093943	0.639046	0.997883
R2UNet [13]	0.685249	0.547796	3.052701	0.687426	0.998020
D-LinkNet [23]	0.689817	0.552248	3.034576	0.639046	**0.998747**
UNet++ [15]	0.698363	0.559524	2.935850	0.649452	0.997992
ResCEAttUnet	**0.729378**	**0.589197**	**2.873382**	**0.716825**	0.998072

Note: Bold font indicates the best value for each metric.

Since the CE structure of the D-LinkNet network deploys only a DAC module without the multi-scale pooling module, the reason could be that the RMP module weakens the ability to suppress false positives. Nevertheless, our proposed in this paper are significantly superior to other networks on the other metrics.

Figure 7 shows some typical segmentation results of ResCEAttUnet and its baseline network Attention-Unet on some of the test sets. The red, blue, and green lines indicate the tumor contours declined by the ground truth, the attention-UNet, and our proposed ResCEAttUnet, respectively.

These test datasets are from the validation and test sets that do not participate in the model training. The main challenges include multiple liver tumors in some slices, uneven edges, and low contrast between tumors and normal tissues.

From the experimental results, some normal tissues are wrongly segmented as tumors by the attention-Unet (Figure 7(d),(f),(h),(i)). Still, our proposed ResCEAttUnet successfully avoids this kind of wrong segmentation.

Besides, for a single tumor with distinct contour and high contrast, the segmentation results of the two networks are both satisfying (Figure 7(e),(j)). However, when two tumors are close to each other, Attention-Unet tends to segment them into one (Figure 7(c)).

Furthermore, in Figure 7(a),(b), the attention-UNet mistakenly segmented many tumor edge pixels into the background. Therefore, the Dice is much lower than that of the ResCEAttUnet network. Also, for the large tumors in Figure 7(f),(g),(h), the segmentation results of ResCEAttUnet are very close to the ground truth.

Finally, for the complex tumor edges in Figure 7(e),(f), the segmentation effect of the ResCEAttUnet network is also significantly improved.

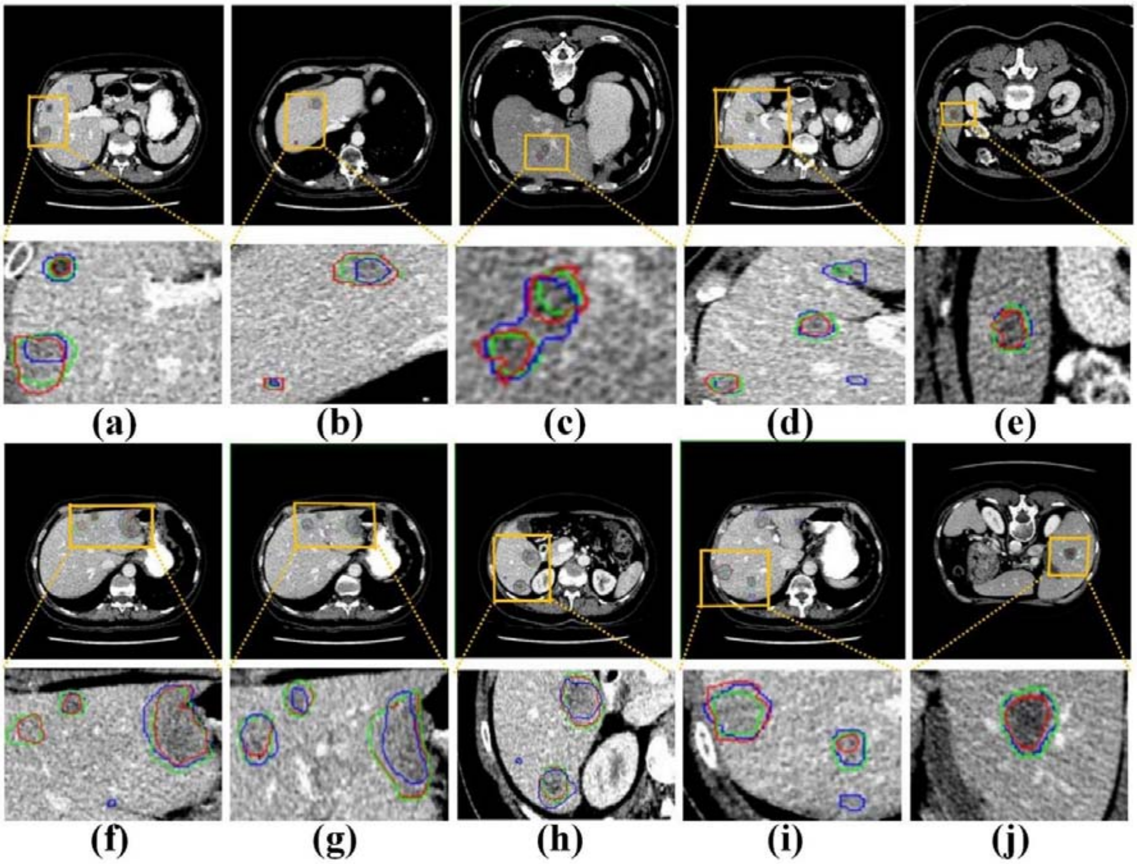


Figure 7. Comparative results on tumor segmentation using different methods. Red/blue/green line represents the ground truth/Attention-Unet/ResCEAttUnet.

3.5. Comparison of running time of different models

Tables 3 and 4 list the training and testing time (per case) using different models on datasets of LiTS17 and 3DIRCADb, respectively. It can be seen from the tables that, the training and testing time of the proposed Pre-ResCEAttUnet are the highest. Specifically, due to the introductions of the residual module, CE, and Attention-Unet in the ablation experiment, the training and testing time of ResCEAttUnet is gradually increased compared with the other methods. However, more image features are obtained through a deeper network, and the accuracy of the proposed ResCEAttUnet is greatly improved, which has been verified in Section 3.3. On the whole, the training time of ResCEAttUnet is higher than other models, while the test time is only slightly higher than other models. Nevertheless, this strategy of improving accuracy at the time cost is still relatively meaningful in computer-aided diagnosis.

Table 3. Training and testing times for ablation analysis.

Networks	Training time	Testing time
U-Net(s)	14 h 32 min	1 min 17 sec
Attention-Unet(s)	15 h 54 min	2 min 25 sec
CE-Att-UNet(s)	17 h 36 min	3 min 48 sec
CE-Att-UNet(H)	17 h 46 min	3 min 56 sec
ResCEAttUnet(s)	19 h 07 min	4 min 57 sec
ResCEAttUnet(H)	19 h 23 min	5 min 10 sec
Pre-ResCEAttUnet(H)	20 h 31 min	6 min 01 sec

Table 4. Training and testing times compared to other state-of-the-art models.

Networks	Training time	Testing time
U-Net	14 h 32 min	1 min 17 sec
R2UNet	16 h 24 min	2 min 55 sec
D-LinkNet	18 h 14 min	4 min 17 sec
UNet++	19 h 12 min	5 min 36 sec
ResCEAttUnet	20 h 31 min	6 min 01 sec

4. Conclusions

This paper proposed a hybrid network for liver tumor segmentation, which leverages residual, BN, CE technologies to improve the attention-Unet. Compared with the attention-Unet, our proposed network structure achieved the following advantages:

(i) All convolution modules on the encoder and decoder paths adopt the ResNet-BN structure, which can accelerate the convergence speed of the model, and the shrinking path is consistent with ResNet-34. Furthermore, we can load pre-trained ResNet-34 weights for migration learning to make the model converge faster and segmentation better. (ii) The DAC and RMP modules in the context information extraction structure proposed to enhance the network's ability to extract multi-scale features and improve the network's preservation of spatial details. In addition, the CE structure can effectively improve segmentation performance for easily misclassified targets, such as liver tumor edges and small target tumors. (iii) The proposed hybrid loss function based on cross-entropy and Tversky can dynamically allocate the weights of the two primary functions through the current number of iterations to fully play the advantages of different loss functions in various stages. Although our proposed method obtains a higher dice score than other methods, it still has limitations. For one, the large number of network parameters requires significant computing resources. For another, the proposed method tends to result in poor segmentation accuracy when dealing with tumors affected by adjacent organs, as shown in Figure 8.

In conclusion, compared with the Attention-Unet, our proposed network accelerates the convergence speed. Meanwhile, it improved the segmentation performance on small target tumors, the tumor edge pixels, and liver tumors with similar contrast to normal tissues. Thus, it could be a promising tool for liver tumor segmentation in clinical assistance. In addition, the growth and quantification of tumors over time need to be premised on accurate contour segmentation and then calculate its max diameter or volume. However, the segmentation accuracy of liver tumors by state-of-the-art methods is still not high. Therefore, we will focus on the growth of the tumor in future work based on further accurate segmentation.

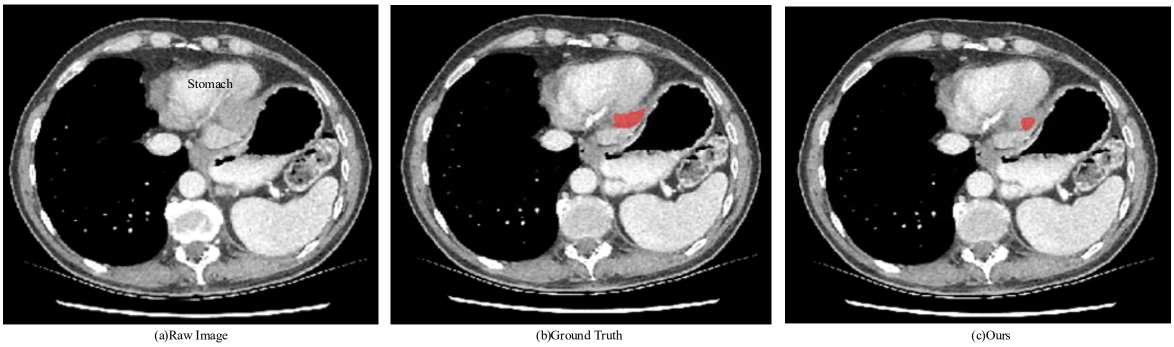


Figure 8. Tumor affected by adjacent organs. (a) original image. (b) gold standard. (c) segmentation results of our method.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (No. 61741106), and the Heilongjiang Provincial Natural Science Foundation of China (No. LH2019F023).

Conflict of interest

The authors declare there is no conflict of interest in this study.

References

1. J. Ferlay, H. R. Shin, F. Bray, D. Forman, C Mathers, D. M. Parkin, Estimates of worldwide burden of cancer in 2008: Globocan 2008, *Int. J. Cancer*, **27** (2010), 2893–2917. https://doi.org/10.1002/ijc.25516
2. K. M. Ratheesh, L. K. Seah, V. M. Murukeshan, Spectral phase-based automatic calibration scheme for swept source-based optical coherence tomography systems, *Phy. Med. Biol.*, **21** (2016), 7652. https://doi.org/10.1088/0031-9155/61/21/7652
3. R. K. Meleppat, M. V. Matham, L. K. Seah, An efficient phase analysis-based wavenumber linearization scheme for swept source optical coherence tomography systems, *Laser Phys. Lett.*, **5** (2015), 055601. https://doi.org/10.1088/1612-2011/12/5/055601
4. R. K. Meleppat, M. V. Matham, L. K. Seah, Optical frequency domain imaging with a rapidly swept laser in the 1300nm bio-imaging window, in *International Conference on Optical and Photonic Engineering (ICOPEN 2015)*, *International Society for Optics and Photonics*, (2015), 9524: 95242R. https://doi.org/10.1117/12.2190530
5. N. Mu, H. Wang, Y. Zhang, J. Jiang, J. Tang, Progressive global perception and local polishing network for lung infection segmentation of COVID-19 CT images, *Pattern Recognit.*, **120** (2021), 108168. https://doi.org/10.1016/j.patcog.2021.108168

6. F. Zhu, Z. Gao, C. Zhao, Z. Zhu, J. Tang, Y. Liu, et al., Semantic segmentation using deep learning to extract total extraocular muscles and optic nerve from orbital computed tomography images, *Optik*, **244** (2021), 167551. https://doi.org/10.1016/j.ijleo.2021.167551
7. C. Zhao, Y. Xu, Z. He, J. Tang, Y. Zhang, J Han, et al., Lung segmentation and automatic detection of COVID-19 using radiomic features from chest CT images, *Pattern Recognit.*, **119** (2021), 108071. https://doi.org/10.1016/j.patcog.2021.108071
8. 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*, (2015), 234–241. https://doi.org/10.1007/978-3-319-24574-4_28
9. 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, preprint, arXiv:1804.03999.
10. L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A. L Yuille, Deep lab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs, *IEEE Trans. Pattern Anal. Mach. Intell.*, **40** (2017), 834–848. https://doi.org/10.1109/TPAMI.2017.2699184
11. E. Shelhamer, J. Long, T. Darrell, Fully convolutional networks for semantic segmentation, *IEEE Trans. Pattern Anal. Mach. Intell*, **39** (2016), 640–651. https://doi.org/10.1109/TPAMI.2016.2572683
12. L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A. L. Yuille, Semantic image segmentation with deep convolutional nets and fully connected crfs, preprint, arXiv:1412.7062.
13. M. Z. Alom, M. Hasan, C. Yakopcic1, T. M. Taha, V. K. Asari1, Recurrent residual convolutional neural network based on U-Net (R2U-Net) for nedical image segmentation, preprint, arXiv:1802.06955.
14. 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
15. Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, J. Liang, Unet++: a nested u-net architecture for medical image segmentation, in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Springer, Cham, (2018), 3–11.
16. F. Milletari, N. Navab, S. A. Ahmadi, V-net: fully convolutional neural networks for volumetric medical image segmentation, in *2016 IEEE Fourth International Conference on 3D Vision (3DV)*, (2016), 565–571. https://doi.org/10.1109/3DV.2016.79
17. 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
18. R. Mehta, J. Sivaswamy, M-net: A convolutional neural network for deep brain structure segmentation, in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, (2017), 437–440. https://doi.org/10.1109/ISBI.2017.7950555
19. C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, N. Sang, Bisenet: bilateral segmentation network for real-time semantic segmentation, in *Proceedings of the European Conference on Computer Vision (ECCV)*, (2018), 325–341. https://doi.org/10.1007/978-3-030-01261-8_20

20. Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, et al., Ce-net: Context encoder network for 2d medical image segmentation, *IEEE Trans. Med. Imaging*, **38** (2019), 2281–2292. https://doi.org/10.1109/TMI.2019.2903562
21. S. Wiesler, H. Ney, A convergence analysis of log-linear training, *Adv. Neural Inf. Process. Syst.*, **24** (2011), 657–665.
22. E. Vorontsov, A. Tang, C. Pal, S Kadoury, Liver lesion segmentation informed by joint liver segmentation, in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, (2018), 1332–1335. https://doi.org/10.1109/ISBI.2018.8363817
23. L. Zhou, C. Zhang, M. Wu, D-LinkNet: LinkNet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, (2018), 182–186. https://doi.org/10.1109/CVPRW.2018.00034