Boundary-Aware Dual Attention Guided Liver Segment Segmentation Model

Xibin Jia[1*], Chen Qian[1], Zhenghan Yang[2*], Hui Xu[2], Xianjun Han[2], Hao Ren[2], Xinru Wu[2], Boyang Ma[2], Dawei Yang[2], and Hong Min[3]

[1] Faculty of Information Technology, Beijing University of Technology, Beijing, 100124, China
[e-mail: jiaxibin@bjut.edu.cn]
[2] Department of Radiology, Beijing Friendship Hospital, Capital Medical University, Beijing, 100050, China
[e-mail: yangzhenghan@vip.163.com]
[3] Department of Computer Software Engineering, Soonchunhyang University, Asan, 31538, South Korea
[e-mail: mhong@sch.ac.kr]
[*]Corresponding authors: Xibin Jia, Zhenghan Yang

Received Novemver 8, 2021; revised December 28, 2021; accepted January 10, 2022; published January 31, 2022

Abstract

Accurate liver segment segmentation based on radiological images is indispensable for the preoperative analysis of liver tumor resection surgery. However, most of the existing segmentation methods are not feasible to be used directly for this task due to the challenge of exact edge prediction with some tiny and slender vessels as its clinical segmentation criterion. To address this problem, we propose a novel deep learning based segmentation model, called Boundary-Aware Dual Attention Liver Segment Segmentation Model (BADA). This model can improve the segmentation accuracy of liver segments with enhancing the edges including the vessels serving as segment boundaries. In our model, the dual gated attention is proposed, which composes of a spatial attention module and a semantic attention module. The spatial attention module enhances the weights of key edge regions by concerning about the salient intensity changes, while the semantic attention amplifies the contribution of filters that can extract more discriminative feature information by weighting the significant convolution channels. Simultaneously, we build a dataset of liver segments including 59 clinic cases with dynamically contrast enhanced MRI(Magnetic Resonance Imaging) of portal vein stage, which annotated by several professional radiologists. Comparing with several state-of-the-art methods and baseline segmentation methods, we achieve the best results on this clinic liver segment segmentation dataset, where Mean Dice, Mean Sensitivity and Mean Positive Predicted Value reach 89.01%, 87.71% and 90.67%, respectively.

Keywords: Segmentation model, liver segment, attention mechanism, boundary-aware

1. Introduction

Liver segmentation is a prerequisite for automatic image analysis, one important step is subdividing the liver into anatomical regions, that is, liver segments. According to the widely used "Couinaud Classification" [1, 2], lobes of the liver are divided into eight segments based on a transverse plane through the bifurcation of the main portal vein. The segmentation of the liver into independent units is significantly in surgical treatment, as segment with tumor involved can be resected individually without damaging the remaining segments, which could preserve the liver function as much as possible. Radiological images like computer tomography (CT) or magnetic resonance imaging (MRI) with contrast-agent administration could clearly show the anatomic structure such as liver vein, portal vein as well as their vascular branches, which are fundamental for the delineation of liver segments. Hence, accurate liver segment segmentation based on radiological images is essential and indispensable preoperationally for the possible resection management of liver tumor.

Most existing liver segment segmentation work is based on traditional image processing methods and employ similar processing procedures. First, segmentation of the blood vessels in the liver is done with kinds of traditional image processing methods. Then, the nearest neighbor approximation algorithm is used for liver segment segmentation while taking the relative vessel segmentation as a reference [3, 4, 5]. Although using the traditional segmentation methods provide the solution of the liver segment segmentation, there still exist some shortcomings. Specifically, traditional methods are not good at dealing with blurred boundaries, nor can they distinguish which blood vessels can be regarded as the reference for segmenting liver segments. Furthermore, these methods are not flexible in fitting to the varied characteristics of the data.

Nowadays, semantic segmentation based on deep learning has been applied in many medical scenarios, including organ segmentation [6, 7], vessel segmentation [8, 9], liver tumor segmentation [10, 11], 3D Reconstruction [12], and visual enhancement [13]. It is not difficult to think that applying deep learning to the liver segmentation task would be a feasible approach. However, most of the existing deep learning segmentation methods are not suitable for the liver segment segmentation task. Because one of the important basis of liver segment segmentation is specific vessels which are normally tiny and slender. These methods do not highlight the features of vessels and ignore the boundary details, which can lead to a decrease in segmentation accuracy. To address this problem, we propose a novel network framework for liver segment segmentation with combining two attention mechanism, named Boundary-Aware Dual Attention Guided Liver Segment Segmentation Model (BADA). In our BADA, a spatial attention module and a semantic attention module are built to perform attention weighting in parallel from both pixel and channel dimensions. It reuses the low-level feature map with richer boundary position information, and fuses it with the high-level feature map with richer semantic feature information as a gated signal to weight the boundary detail information. The proposed dual attention processing is employed at every layer at the decoding path of U-Net.

The main contributions of our work are as follows:

1) We propose a novel Boundary-Aware Dual Attention model (BADA) for liver segment segmentation. In our proposed network, the boundary between the liver and its surroundings and the boundaries between segments are highlighted. Accordingly, it enhances the accuracy of liver segment positioning and boundary recovery.

2) A dual attention module is proposed to make full use of both spatial and semantic information at every level of U-Net, which facilitate revealing the characteristics of boundaries at different scales.

3) A well-labeled dataset of liver segments is built from clinical cases. Comprehensive comparative and ablation experiments have been done on the built dataset. The experiments demonstrate that we achieve state-of-the-art results.

2. Related Work

2.1 Attention Mechanism

Attention mechanism was first used in natural language processing and has made great progress in computer vision tasks in recent years [14, 15]. The attention mechanism can capture the dependencies between pixels and highlight key areas in the image. SE-Net [16] is designed for image classification tasks, which employs a channel self-attention module to adjust the weight of channels. DA-Net [17] uses the dual attention of spatial and channel to enhance the discriminant ability of feature representations for scene segmentation. These methods employ the self-attention mechanism and consume a lot of computing resources. Moreover, the position information provided by the low-level feature map is not fully utilized.

2.2 Semantic Segmentation in Biomedical Image

Olaf Ronneberger et al. proposed a U-shape network for neuronal structure and cells segmentation, called U-Net [18], which has become the most common backbone network used in biomedical image segmentation. U-Net has symmetric encoding path and decoding path. Using skip connection between the same resolution stages, U-Net improves the accuracy of feature positioning and boundary recovery. At present, most of the segmentation models of biomedical images are variants of U-Net [19, 20, 21]. First, ResUNet [22] and DenseUNet [23] are inspired by residual connection and dense connection between convolutional layers, respectively. They replace each submodule of U-Net with the module with residual connections and dense connections. Although ResUNet is originally suitable for remotely sensed data segmentation, it is widely used as a baseline network in the biomedical image segmentation tasks, such as ResUNet++ [24].

Second, because of the multi-modality characteristics of biomedical images, some researchers work on the biomedical image segmentation by using multi-modal fusion method. Tseng et al. [25] design a cross-modality method for brain tumor segmentation, which fuses four MRI sequences Flair, T1, T1c, and T2. Jia et al. [26] employ a multi-path encoder to fuse the multi-modality data to complete the brain tumor segmentation task.

Currently, attention mechanism, which is widely used in natural language processing and natural image analysis, is also applied to biomedical image segmentation. Researchers hope to use attention mechanisms to highlight key areas in each image [27, 28]. Attention U-Net [29] designs a spatial attention gate to segment gastric cancer and the pancreas by using CT images. ET-Net [30] introduces a channel attention mechanism for retina vessel and the lung segmentation.

Meanwhile, semi-supervised learning, self-supervised learning and the use of transformer to process computer vision tasks are also relatively popular. Chen et al. [31] address the limitations of U-Net long-range dependency using the global self-attention mechanism innate to transformer. Li et al. and Agisilaos Chartsias et al. [32, 33] use a semi-supervised learning approach to solve the problem of lack of labeled samples and large number of unlabeled samples in medical image segmentation tasks.

Referring to the previous work and taking the live segment segmentation as problem, we propose a modified structure of two attention mechanism with a spatial attention module and a semantic attention module. We reuse the low-level and the high-level fused feature map as gated signal in the high-level feature learning, which facilitates to highlight the boundary impact and accordingly improve the accuracy of segmentation. The experiments in the later section demonstrates the effectiveness of our method.

3. Method

3.1 Overall architecture of proposed network

Our boundary-aware dual attention module is inspired by DANet [17]. DANet adopts both spatial and channel attention mechanism, which achieves good results for the natural image segmentation task. Similarly, we develop a boundary-aware dual gated attention mechanism, and apply it on the U-Net backbone network. Due to the symmetric network, U-Net is useful for recovering the image to the original resolution step by step at the decoding path, so that the small details like the vessel boundaries in the image can be maintained. Accordingly, deployed the dual attention module at each resolution level of U-Net, the boundary recovery performance is supposed to be fostered. Furthermore, the key point of our proposed dual gated attention is to use the low-level and the high-level fused feature maps as gated signal. So that we enhance awareness to the edges by reusing the information of the low-level feature map, which retains more prominent boundary information comparing to the high-level one. By feeding this information together with the current high-level feature into the dual gated attention module, the feature differences are increased and boundary information is more prominent. Therefore, instead of using self-attention, our model performs gated attention weighting with fusing high-level and low-level features.

The architecture of our proposed network is shown in **Fig. 1**. In our model, we select U-Net with residual blocks [34] as the backbone network. The dual attention blocks with rectangle boxes denoted in **Fig. 1** are added at the skip connection between the left decoding paths and the corresponding right encoding paths at each layer. As in the **Fig. 1**, the overall processing procedure of our proposed network is illustrated as follow. Firstly, for each dual attention module, the fused adjacent feature maps are loaded, i.e. the low-level feature map from the encoding path and upsampled high-level feature map from the decoding path. In our liver segment segmentation task, the low-level feature map contains more intensity detail information like edges reflecting the boundaries of the liver and the liver segment, while the high-level feature map contains more abstract semantic and category information for distinguishing the liver area from the background area or between liver segments. The fusion of the two feature maps can highlights the boundary and position information of the liver and liver segments and enhance the recognition of the boundary points.

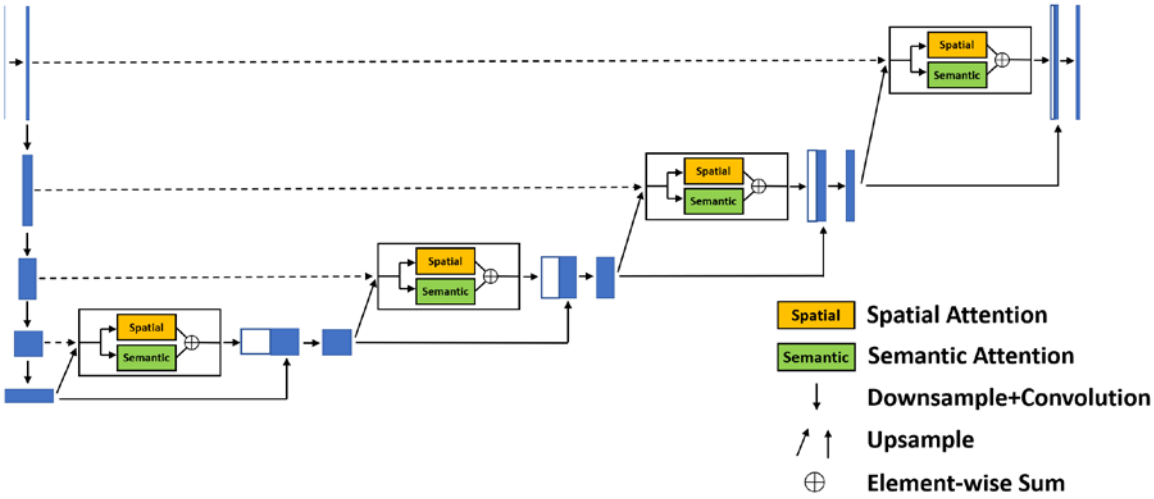


Fig. 1. Overview of network structure

Then, the fused feature map loaded into dual attention block is processed with counting the channel and pixel attention weight maps, respectively. In this way, the pixels around the boundary can obtain larger weights at the spatial level, which increases the feature difference. At the channel level, channels that contain more discriminative semantic category information can obtain larger weights while invalid channels are suppressed with less weights.

Finally, the original feature map is multiplied with two derived attention maps respectively. The element-wise sum is then employed to aggregate two attention-weighted feature maps and obtain the final output from each dual attention module.

3.2 Spatial Attention Module

Boundaries are the edges between the objects, while edges are a set of points with sharp intensity changes. Therefore, the possible boundaries between liver segments i.e. blood vessels have obvious spatial characteristics of edges with salient intensity changes. To obtain the discriminate expression of the image for revealing the boundary attribution, we develop the spatial attention module to weight the important spatial information of the feature map. In view that edge point positions in different channels of the feature maps ought to be consistence, we weight each channel in the feature map using the attention map generated by the same gated signal calculation. Thereby, important information such as the boundary information and position information of the liver segments will always be highlighted.

In this module, the fusion feature map is compressed into a single-channel map, and each pixel represents a weight, which is multiplied by the original feature map. The detail of the proposed spatial attention calculation is introduced as follows.

As illustrated in **Fig. 2(a)**. First, given the low-level feature $F_l \in \mathbb{R}^{C\times H\times W}$ from the decoding path and the high-level feature $F_h \in \mathbb{R}^{C\times H\times W}$ from the previous layer, we feed them into two different convolution kernels $W_l \in \mathbb{R}^{C\times C}$ and $W_h \in \mathbb{R}^{C\times C}$, and obtain two new feature maps $F_l^{'} \in \mathbb{R}^{C\times H\times W}$, $F_h^{'} \in \mathbb{R}^{C\times H\times W}$, respectively. Next, we concatenate the two new feature maps along the channel dimension to generate the low-level feature and high-lever feature fused map $F_{concat} \in \mathbb{R}^{2C\times H\times W}$. The fused map after concatenation is activated once by ReLu,

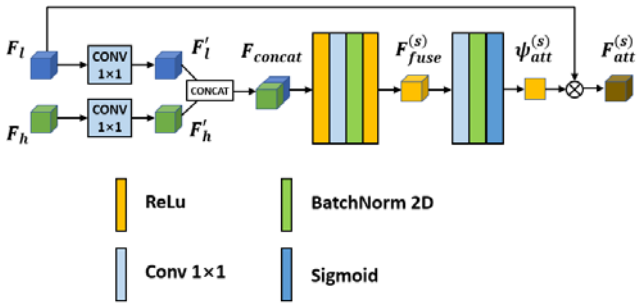
and then feed into a network consisting of a convolutional layer $W_c \in \mathbb{R}^{2C\times C}$, a BatchNorm layer and a ReLu activation function. This operation normalizes the channel dimension of the fused map to C, i.e. $F_{fuse}^{(s)} \in \mathbb{R}^{C\times H\times W}$. At this point, the low-level feature map and the high-level feature map are integrated while keeping the spatial information and the reduced dimension. After that, we feed the integrated feature map into a convolutional layer $W_f \in \mathbb{R}^{C\times 1}$ to compress it to a single-channel feature map. Then, we perform a Sigmoid function to normalize it to [0,1] to generate the spatial attention map $\psi_{att}^{(s)} \in \mathbb{R}^{1\times H\times W}$. Thus, the attention map is obtained, where each pixel represents the weight indicating its salient extent in the origin feature map from the perspective of spatial view. Finally, we perform an element-wise product between the spatial attention map $\psi_{att}^{(s)}$ and the low-level feature map F_l from the same resolution stage of the decoding path, to obtain the final weighted feature maps, i.e. $F_{att}^{(s)} \in \mathbb{R}^{C\times H\times W}$ from the spatial attention module. The entire process is formulated as Eq. (1)-Eq. (4):

$$F_{concat} = Concat(W_l^T F_l + b_l;\ W_h^T F_h + b_h) \tag{1}$$

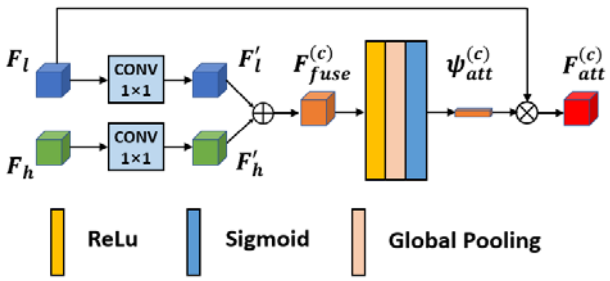
$$F_{fuse}^{(s)} = \sigma_1(W_c^T \sigma_1(F_{CONCAT}) + b_c) \tag{2}$$

$$\psi_{att}^{(s)} = \sigma_2(W_f^T F_{fuse}^{(s)} + b_f) \tag{3}$$

$$F_{att}^{(s)} = \psi_{att}^{(s)} \otimes F_l \tag{4}$$



(a) Spatial attention module



(b) Semantic attention module

Fig. 2. Boundary-aware dual attention module

where b_l, b_h, b_c, b_f are bias terms corresponding to different convolutional layers, $\sigma_1$and σ_2 denote the ReLu activation function and the Sigmoid activation function respectively.

With respect to merge the high-level and low-level feature maps, we don't use the element-wise sum method directly, which is employed in Attention U-Net [29]. Instead, we adopt the channel dimension concatenation method. In this way, we can use convolution operations for fusion to have one more feature learning opportunity, while maintaining the consistency of spatial information.

3.3 Semantic Attention Module

Each feature map generated from the different kernel filter reflect a certain attribution associated with a kind of semantic clues. So, highlighting the relative channels in revealing the boundary semantics and suppressing the irrelevant channels can improve the accuracy of segmentation. Inspired by the above idea and referring to the channel attention idea, we propose our semantic attention module which takes account of the kernel difference in reflecting the boundary semantic clues.

In the semantic attention module, the fused feature map is compressed into a one-dimensional column vector, which represents the weight of each channel. The semantic attention module detailed structure is illustrated in **Fig. 2(b)**, which employs the same gated method we proposed in our spatial attention module. Meanwhile, the same feature map F_l, F_h as input and shared parameters W_l, W_h of the convolutional layers for the low-level feature and the high-level feature are used as in the spatial attention module, respectively. Different from the spatial attention module, fused scheme of low-level feature and the high-level feature map in the semantic attention module performs an element-wise sum to generate the fused map $F_{fuse}^{(c)} \in \mathbb{R}^{C\times H\times W}$, for this operation is better for maintaining the consistency of semantic class (channel) information. Similarly, convolutional layer and fully connected layer are not used in this module, which reduces the impact on semantic information and also reduces the amount of calculation. After obtaining the fused feature map, we employ the global average pooling to downsample each channel map of $F_{fuse}^{(c)}$ to obtain a one-dimensional vector. After using the Sigmoid function to activate the one-dimensional vector, the semantic attention map $\psi_{att}^{(c)} \in \mathbb{R}^{C\times 1\times 1}$ is obtained. The last step of this module is performing an element-wise product between the semantic attention map $\psi_{att}^{(c)}$ and F_l. Let $b = b_l + b_h$, the entire process can be formulated as Eq. (5)-Eq. (7):

$$F_{fuse}^{(c)} = W_l^T F_l + W_h^T F + b \tag{5}$$

$$\psi_{att}^{(c)} = \sigma_2(GAP(\sigma_1(F_{fuse}^{(c)}))) \tag{6}$$

$$F_{att}^{(c)} = \psi_{att}^{(c)} \otimes F_l \tag{7}$$

3.4 Aggregation for Attention Module

The fused attention is obtained by aggregating the two outputs from the parallel attention modules. Specifically, we perform an element-wise sum between the outputs of spatial attention module $F_{att}^{(s)}$ and semantic attention module $F_{att}^{(c)}$ to generate a fused attention map. Then, we send the fused attention map into a convolution layer with a kernel size of 1×1 and obtain the final attention map. After that, the final attention map concatenates with the previous layer feature map to accomplish the skip connect of U-Net. The whole procedure is illustrated

in **Fig. 3**. As we can see, different from the input of traditional U-Net skip connection, our proposed model uses low-level feature maps weighted by boundary-aware dual attention as input to enhance the boundary, semantic, and location information.

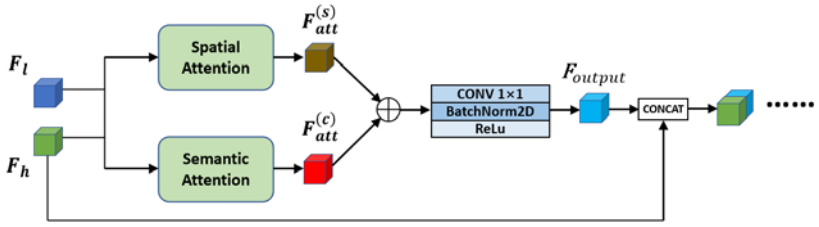


Fig. 3. Aggregation Procedure for Attention Module

3.5 Consideration in determination of Backbone Networks

FCNs (Fully Convolutional Networks) and U-Net are the two most commonly used backbone networks in semantic segmentation. For medical image segmentation tasks, it is consensus that U-Net has better performance. Therefore, in the paper, we prefer to use the U-net as backbone. However, to explore the effectiveness of the symmetric U-net and the asymmetric FCN as our backbone, we modify our proposed symmetric based network into several asymmetric forms (base on FCNs) in this section. Moreover, we set up an ablation experiment in section 4.4 to do the further comparison between the FCN-based and U-Net-based models.

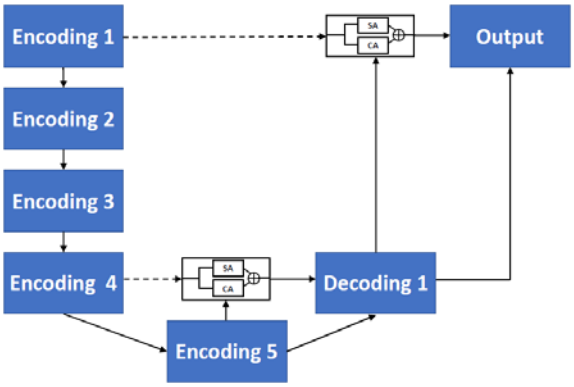


Fig. 4. Asymmetric network with 2 upsamples

The decoding path in our proposed U-Net-based model contains four upsample operating layers. The feature map loading on each upsample operating layer is weighted with the attention map from the dual gated attention module. Besides, three variant FCN-based models are built for the comparative analysis, where they perform the upsample operation and the attention weighting operation once, twice, and three times, respectively. These upsample operations are used to reshape the feature map to the same size of input. Take the FCN-based model with two upsample operations as an example, the network structure is shown in **Fig. 4**.

3.6 Mixed Loss Function

Dice loss is one of the most commonly used loss functions for image segmentation as shown in Eq. (8). It calculates the coincidence rate between the predicted image and the ground

truth. The value range of the dice loss function is [0,1], 0 means completely overlaping, and 1 indicates totally non-overlaping.

$$\mathrm{L}_{Dice} = 1 - \frac{2\left|A \cap B\right|}{\left|A\right| + \left|B\right|} \tag{8}$$

where A and B denote the predict pixels and the ground truth, respectively.

Cross-entropy loss is another commonly used function by measuring the difference between two probability distributions. In the image segmentation task, cross-entropy function is defined as Eq. (9) with calculating the average probability difference between pixels of the predicted image and the ground truth.

$$\mathrm{L}_{CE} = -\frac{1}{N}\sum_{i=1}^{N} y^{(i)} \log \hat{y}^{(i)} + (1 - y^{(i)}) \log(1 - \hat{y}^{(i)}) \tag{9}$$

where N denotes the number of pixels, *i* indicates the *i*-th pixel in the image, y and $\hat{y}$ denote the one-hot vector of segment mask of the ground truth and the predicted results, respectively.

Considering the cross-entropy loss is calculated at the pixel level, the training is prone to be dominated by the categories with more pixels, and it is not conducive for feature learning of small objects. Therefore, the cross-entropy loss is more suitable for the situation where sample categories are balanced. In contrast, the Dice loss can deal with unbalanced sample categories, but training might be unstable when sample categories are balanced. In view of the existence of both balanced categories and unbalanced categories in our data set, therefore, we combine these two loss functions. The adopted loss function is defined as Eq. (10):

$$L_{Mixed} = \lambda_1 L_{CE} + \lambda_2 L_{Dice} \tag{10}$$

where λ_1 and λ_2 are hyperparameters. In determining the value of these two hyperparameters, we do the following exploration. With the epoch of training increases, the gap between the dice loss and the cross-entropy loss is large, which may make the loss function retreat to the dice loss. To balance their contribution, we consider to set λ_1 and λ_2=0.1 in the experiment. However, experiments show that this hyperparameter setting affects the decline of the dice loss, which cause a quick convergence of the dice loss. To eliminate this problem and achieve a better result, the final hyperparameter is set with λ_1=λ_2=1

4. Experiments

In this section, we first introduce our liver segment segmentation dataset and evaluation metrics. Then, some ablation experiments are performed to evaluate the effectiveness of key modules in our method. Finally, we compare our model with several state-of-the-art methods from both qualitative and quantitative aspects.

(a) Couinaud liver segments method **(b)** 3D reconstruction label

Fig. 5. Couinaud Segmentation Method

4.1 Dataset

In this study, we collect a dataset with 59 liver MRI from Beijing Friendship Hospital. The cases in the dataset are deliberately selected from different liver diseases, such as cyst, focal nodular hyperplasia, hemangioma etc. Portal vein phase MRI after contrast-agent administration were used to label the main blood vessels such as hepatic veins and portal veins by four experienced radiologists.

We use the Couinaud liver segments method as the liver segment segmentation medical standard in the dataset, which is currently one mainstream liver segments definition methods [1, 2]. According to this medical criteria, the liver is separated into eight segments (numbered I to VIII) based on the vascular supply. Segment I is the caudate lobe. Segments II and III lie lateral to the falciform ligament with II superior to the portal venous supply and III inferior. Segment IV lies medial to the falciform ligament and is subdivided into IVa (superior) and IVb (inferior). Segments V to VIII make up the right part of the liver. Segment V is the most medial and inferior. Segment VI is located more posteriorly. SegmentVII is located above segment VI. Segment VIII sits above segment V in the superior-medial position. **Fig. 5(a)** illustrates the segment method with an example and **Fig. 5(b)** is its 3D reconstruction display. In the figure, we use nine different colors to mark and label different liver segments.

In the experiments, the dataset is randomly divided into three subsets. 75% of data is used for training, 10% of data is used for validation, and 15% of data is used for testing. In other words, the training set contains 45 cases, the validation set contains 5 cases, and the test set contains 9 cases. Considering that deep neural networks usually require large amount of training data, we slice 3D original data into 2D planes as samples. In the end, we have 2984 training images, 398 validation images and 597 test images. This dataset is annotated with 9 liver segment classes and one background class.

4.2 Evaluation Metrics

To evaluate the proposed method, we employ Dice, Sensitivity (Sens), and Positive Predicted Value (PPV) as the evaluation metrics, which are defined as formula 11-13 respectively.

$$Dice = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{11}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{12}$$

$$PPV = \frac{TP}{TP + FP} \tag{13}$$

where TP, TN, FP, FN denote true positives, true negatives, false positives, and false negatives.

4.3 Implementation Details

We use ResUNet as the backbone network. Specifically, considering the limitation of the dataset scale, we adopt the improved ResUNet introduced by He et al. [35], which is simple but effective. A 2×2 average pooling layer is added to the shortcut connection prior to the 1×1 convolutional layer for the transitioning blocks with stride of two. This method can make up for the loss of feature information caused by the downsampling using stride convolution.

We employ a poly learning rate policy that the initial learning rate will be multiplied by 0.1, if the dice metric on validation set does not reduce in 10 epochs. So that we can reduce the value of loss function as much as possible and make the training more thorough. The initial learning rate is set to 0.001 for all the methods used in experiments. The minimum learning rate is set to 0.00001. We employ Adam optimizer [36] and weight decay is set to 0.00004. Batchsize is set to 64. For data processing, the derived 2D data sliced from the 3D volumes are resized to 512×512 using bilinear interpolation. The pixel intensity is normalized to (0,1). We implement our method based on Pytorch. Two Nvidia GeForce GTX 1080ti GPUs are used to train our model several times to ensure the reliability of the results. For each time, the model will be trained 150 epochs to make the model fully converge.

4.4 Ablation Study

To verify the effectiveness of each module of BADA, we conduct ablation experiments addressing the performance of the boundary-aware dual attention mechanism and the symmetric U-shape network, respectively.

4.4.1 Ablation Study for Attention Modules

Boundary-aware dual attention can highlight significant parts of the feature maps that contribute most to segmentation from both the pixel and channel perspectives. To explore the impact of the two attention modules on the segmentation results, we adopt one attention scheme viz. the spatial attention module and the semantic attention module respectively to do the performance evaluation. The relative experiment results under the specific settings are shown in **Table 1**.

Here, we employ two models U-Net and ResUNet as baselines. As shown in **Table 1**, attention module brings a significant improvement in the experimental results. Our proposed model with spatial attention module achieves results of 87.39%, 86.30%, and 89.19% on Mean Dice, Mean Sens, and Mean PPV, respectively. Simultaneously, the model that employs semantic attention module reaches results of 86.25%, 85.82%, 88.7% on those three metrics. Compared with the baselines, the model with spatial attention module improves about 7-9%, while the model with semantic attention module improves about 6-8%. This shows that the spatial attention module can play a more important role.

With the gated attention modules aggregate the attention from the above two angles in the same model, the results improve further, reaching 89.01%, 87.71% and 90.67% on Mean Dice, Mean Sens and Mean PPV respectively. This demonstrates that the aggregated symmetrical gated attention module can make use of the information of the spatial dimension and the channel dimension and accordingly achieve a better segmentation effect. In other words, the spatial information reveals the local salient regions i.e. liver blood vessels and the semantic information reflects important feature expressions. Additionally, the attention modules employed at every resolution level are helpful to enhance the important features at different scales. Therefore, when the two modules are aggregated in the same network model, they can complement each other to foster the increasing of the segmentation accuracy.

Table 1. Ablation Study for Attention Module

Method	Mean Dice(%)	Mean Sens(%)	Mean PPV(%)
Baseline(U-Net) [13]	79.70	77.56	83.25
Baseline(ResUNet) [17]	85.14	85.71	87.14
Our method (with only Spatial)	87.39	86.30	89.19
Our method (with only Semantic)	86.25	85.82	88.79
Our method (with dual attention)	**89.01**	**87.71**	**90.67**

4.4.2 Ablation Study for Backbone Structure

The ablation experiment on structure of the backbone network is to study performance of network in the symmetric or asymmetric structure. Here, we mainly explore the difference between the U-Net based symmetric model and the FCNs based asymmetric model in the liver segments segmentation task. The optional network structure has been introduced in section 3.5, so this section mainly provides the experimental results and do the comparative analysis with some quantitative metrics and subjective visualization results. The comparative experiments conducted under different settings are shown in **Table 2**.

The models in the first three rows are in asymmetric structure, and the model in the last row is the symmetric model we proposed. Since our proposed attention module is applied after every upsample operation, the similar attention modules used by the four models in this table is the same as the number of upsample operations they perform. As the results are shown in **Table 2**, the model with more upsample operations and more attention mechanisms get a better performance in segmentation. Compared with the FCN-based models (the first three rows) which commonly used for scene segmentation, our proposed symmetric model (the last row) improves segmentation performance by 7.67%, 4.05%, 1.99%, respectively. Furthermore, we visualize the corresponding segmentation results as shown in **Fig. 6**. It can be seen that although the asymmetric model can roughly segment the liver sections, but there exist obvious deficiencies in boundary recovery, and even mosaic edges appear. However, with increasing upsample times from 1 to 4 times as in **Fig. 6**, the grade of the mosaic edge phenomenon degrades. Obviously, the effect of image boundary recovery and positioning gets better as well. It can be discovered that using the symmetric U-net model with the feature map concatenating between the encoding path and the decoding path at the same resolution stage is beneficial to improve the accuracy and precision of liver segment segmentation with feature details added at the decoding path. Therefore, adopting the U-Net-based symmetric model as basis, which doing the concatenation of feature maps at every scale level, will strive to obtain better effect of positioning and boundary recovery.

Table 2. Ablation Study for Backbone Network

Method	Mean Dice(%)	Mean Sens(%)	Mean PPV(%)
Proposed(with 1 Upsample)	81.34	80.76	82.32
Proposed(with 2 Upsample)	84.96	84.17	86.42
Proposed(with 3 Upsample)	87.02	85.99	88.09
Proposed(with 4 Upsample)	**89.01**	**87.71**	**90.67**

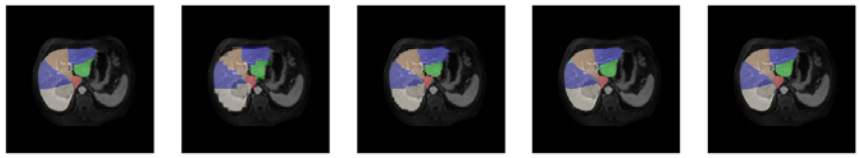


Fig. 6. Visualization results of symmetric and asymmetric model ablation experiments.

4.5 Comparative Experiments

To verify the performance of our proposed model, we conduct a comparative experiment with several methods on the liver segment dataset. For a fair comparison, we modify SE-ResNet and DANet from FCN-based to U-Net-based. Since Dice is the main metric, we record the Dice of each liver segment in this comparison experiment. For the other two metrics, we only show their average values across all liver segments. Results are shown in **Table 3** and **Table 4**. We can see that our method achieves the best performance among the comparative methods in all three metrics. Especially, our model outperforms SE-ResNet (Base on U-Net) [11] and DANet (Base on U-Net) [12] by a large margin. Analyzing the reason, we discover both these two methods use the self-attention mechanism. The experiment results show that self-attention does not exert an effective effect in the segmentation task of the liver segment, because it does not emphasize the boundary location information. In contrast, the gated attention mechanism in our method employs the fusion map of the high-level and the low-level feature as the gated signal. The experimental result demonstrates the effectiveness of our method with the proposed boundary-aware attention mechanism to highlight important features of liver segment positions and boundaries. In view of the main metric Dice, our method achieves the best segmentation performance on 8 of the 9 liver segments and on Mean Dice. Besides, our model outperforms the second-best models 2.62%, 2.37%, 2.29% in Mean Dice, Mean Sens, Mean PPV, respectively, and reaches 89.01%, 87.71%, 90.67%. Also, our proposed model (with Spatial) in the ablation experiment and Attention ResUNet only differs in the feature fusion method, however it achieves a 1.68% improvement. This proves that our adopted channel dimension concatenation method can effectively maintain the spatial feature consistency.

Table 3. Dice metric of per-class results of comparison study

Method	I (%)	II (%)	III (%)	IVA (%)	IVB (%)	V (%)	VI (%)	VII (%)	VIII (%)	Mean (%)
U-Net	74.35	66.15	79.12	76.01	79.34	87.04	85.45	84.75	85.07	79.70
U-Net++	76.97	73.72	83.55	84.17	90.57	90.16	90.04	91.41	87.23	85.32
ResUNet	79.30	77.64	83.44	86.19	79.94	91.13	89.69	90.72	**91.84**	85.54
Attention U-Net	77.30	76.28	84.61	82.42	88.00	91.50	91.41	88.98	89.82	85.59
Attention ResUNet	80.05	79.09	85.06	83.91	86.60	89.85	88.03	88.96	89.81	85.71
SE+ResUNet	78.12	77.97	84.64	87.64	86.43	91.09	90.38	89.84	91.36	86.39

DA+ResUNet	79.47	80.49	85.73	86.24	88.01	89.75	88.06	88.35	89.98	86.23
DANet	69.12	77.96	79.43	80.94	83.19	83.22	82.21	83.58	86.21	80.65
Our method	**80.49**	**83.80**	**86.84**	**89.48**	**91.89**	**92.39**	**91.49**	**91.58**	87.23	**89.01**

To have intuitive understanding of performance achieved in liver segment segmentation task, the comparative analysis with visualization results of the above experiments are done. Some segmentation results are shown in **Fig. 7**. It can be found that the segmentation results of our method are highly consistent with Ground Truth in terms of the region position and boundary. In contrast, there are some significant problems in the other five methods. As shown in second column in **Fig. 7**, the traditional U-Net has a poor performance on this task, where a number of obvious wrong segmentation errors in liver segment regions, and the inaccurate segmentation boundary comparing to that in Ground Truth. Attention U-Net and Attention ResUNet employ spatial gated attention method. As shown in third and fourth columns in **Fig. 7**, their results are significantly better than that of U-Net in the positioning of liver segment boundaries. However, there are still wrong and missing segmentation regions. Analyzing the reasons, it is considered that the semantic information of the key regions is not prominent enough, which leads to the bias of judgment on the category. Comparing with that, our methods with additional region semantic attention achieves better performance at this point as shown in the last column. SE-ResUNet performs a channel self-attention mechanism. As shown in sixth column in **Fig. 7**, there exist some visible inconsistencies in the boundary region pixel class judgments. From this, we can tell using single self-attention mechanism is not enough for revealing region position and boundary information. Otherwise, DANet uses a dual self-attention mechanism and the parallel attention module, its results are shown in the fifth column in **Fig. 7**. However, it can still be found from the results in the second row that this dual self-attention mechanism does not reuse the boundary location information in the low-level feature maps even if the category judgments are correct, which leads to poor boundary localization as well. This indicates that the self-attention mechanism by computing inter-pixel and inter-channel dependencies does not achieve as good results in the field of medical image segmentation as it does in the field of natural image segmentation. This may be because there are strong correlations between objects and objects in natural images, which is not the case in medical images. Comparing with others, BADA performs a boundary-aware dual attention processing with fused attention from paralleled spatial attention module and semantic attention module. Consequentially, our method can solve the problem of inaccurate positioning of liver segment boundaries and incorrect semantic category judgment in this task. The best results from visualization view have been obtained in the experiment.

Table 4. Dice metric of per-class results of comparison study

Method	Backbone	Mean Sens(%)	Mean PPV(%)
U-Net	U-Net	77.56	83.25
U-Net++	U-Net	84.17	88.65
ResUNet	ResNet	85.71	87.14
Attention U-Net	U-Net	84.28	88.10

Attention ResUNet	ResNet	84.78	87.72
SE+ResUNet	ResNet	85.34	88.38
DA+ResUNet	ResNet	85.25	88.05
DANet	ResNet	79.80	82.14
Our method	**ResNet**	**87.71**	**90.67**

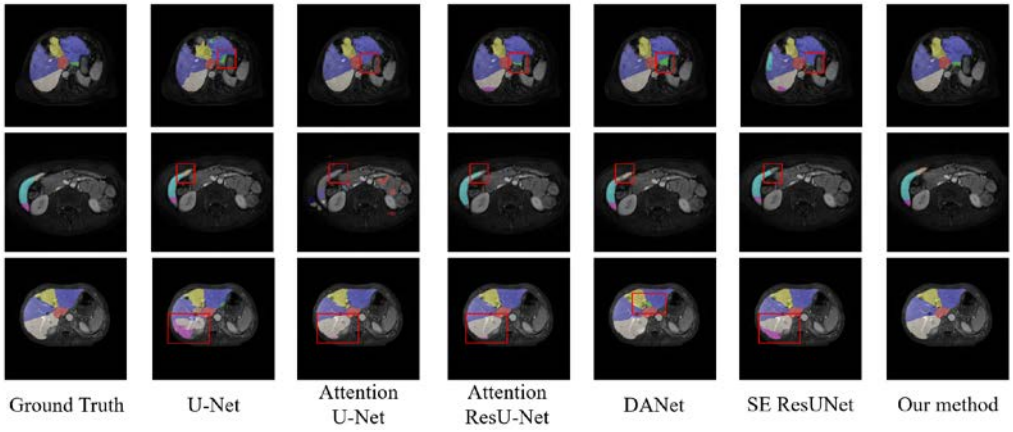


Fig. 7. Examples of visualization of comparison experiment results on test set.

Furthermore, to explore the prominence of our method for the boundary information and the location information of the region to be segmented, we visualize the heat map of the attention mechanism, the weighted feature map and the output of the intermediate layer of the network. We visualize the regions of interest of our method and some comparison methods for the same image by using Grad-Cam [37], and show them as heatmaps in **Fig. 8**. Here we only compare the two most classic and most widely used medical image segmentation methods U-Net and Attention U-Net. As shown in the figure, the heatmap for U-Net without using the attention mechanism reflects no reasonable salient areas. In contrast, Attention U-Net and our method use the attention mechanism. The corresponding heatmaps of both methods reveal that prominent areas are weighted with high attention especially that around the liver contour. In addition, our method achieves the expected result with highlighting the possible boundaries like the two blood vessels depicted with the white boxes in heatmap of our method as in **Fig. 8**.

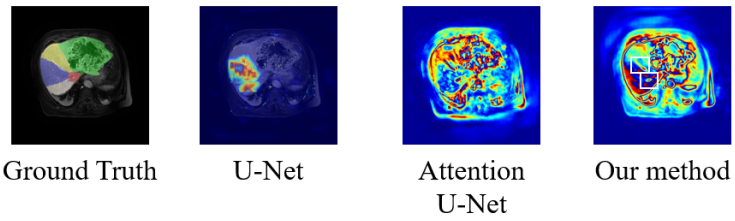


Fig. 8. Examples of visualization of heatmaps on test set.

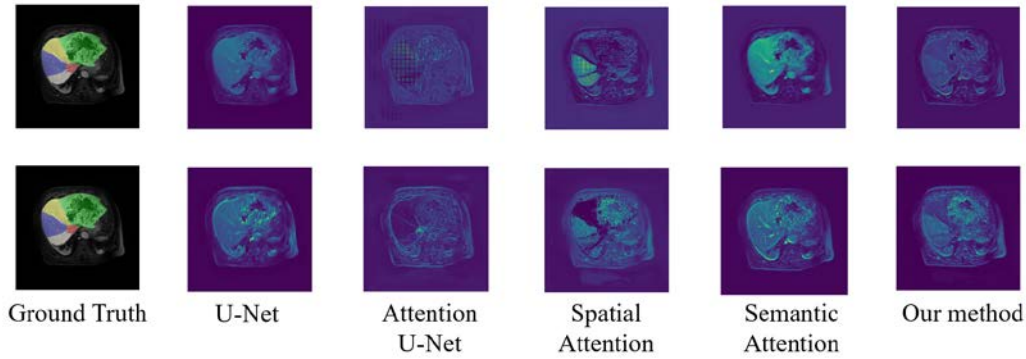


Fig. 9. Examples of visualization of the feature maps after attention weighting.

To have idea of gradually improving results of segmentation, the feature maps after dual attention weighting and the intermediate layer prediction results at the decoding path are shown in **Fig. 9** and **Fig. 10**, respectively. The first row and the second row display the corresponding results at the fourth layer (after 4 times upsample) and the third layer (after 3 times upsample) respectively in **Fig. 9** and **Fig. 10**. As shown in **Fig. 9**, we can tell that feature obtained from U-Net without any attention mechanism can hardly distinguish the liver segments with similar representation of all segments. Comparing with it, features from Attention U-Net capturing the relative boundary clues between the liver segments, but it is still not very distinguishable with obvious different expression. The fourth and fifth columns of the figure show the weighting effect with spatial attention and semantic attention alone in our method, respectively. We can clearly see that the spatial attention branch distinguishes the liver segments well, making the boundaries between liver segments clear and distinct. The semantic attention branch has a clear and prominent representation of the liver region and the blood vessels used to segment the liver segments. The weighted feature maps after the dual attention fusion at the last column combines the advantages of the above, which show clearly distinguishable feature varieties among the liver segments and obvious boundaries between segments. In addition, the boundary between liver and its surroundings presents better distinguishable attribution with a kind of highlighted pixels along the liver contour. Therefore, from the perspective of feature map visualization, it can be seen that our method has learned more discriminative representation and facilitates the liver segment segmentation with accurate boundary positions. In **Fig. 10**, it is not difficult to see that even in the output of the intermediate layers of the network, the feature map using boundary-aware dual attention is significantly clearer in boundary recovery and liver segments position than the feature map using other methods. Similarly, as the feature maps in **Fig. 9**, the output of the intermediate layer of U-Net can hardly tell the liver segment with only the same feature of the entire liver area reflecting at the intermedia layer. Attention U-Net with using the gated attention mechanism achieves a significant improvement in the recovery of liver segment boundaries,

but there still exists some obvious errors. Specifically, the upper boundary of the liver is not recovered clearly. Comparatively, our method achieves intuitively highest positioning accuracy and vivid liver segment boundaries.

In summary, both qualitative and quantitative experiments demonstrate that our proposed BADA play effective roles in liver segment segmentation and achieves the better performance than several baselines and state-of-art segmentation methods

4.6 Discussion of Limitation

To make the proposed BADA method more practical, there are still some works worth to be addressed in the future to overcome some limitations in the current method, which mainly from two aspects.

The first is the limitation of the method. We find in the experiment results as shown in **Fig. 8**, that some boundaries in the non-interest region are enhanced together with the boundaries of the liver segment. Analyzing the reason, it is because that our boundary-aware attention mechanism enhances candidate boundaries where the pixel intensity changes significantly, rather than focusing solely on specific liver segment boundaries. However, this limitation does not bring a significant negative affection on the current experimental results. Our method can still accurately learn the discriminative feature as shown in **Fig. 9** and **Fig. 10**. This is because the constraint of segmentation masks on the location of liver segments, which can guide the network model to distinguish the boundaries of liver segments and the boundaries of the non-interest region to some extent. Under the joint effect of our proposed dual attention and the empirical risk minimization, the overall segmentation performance is improved and achieves the best result. In spite of this, we realize that exploring more advanced and reasonable methods to distinguish the boundary type is worthy to research in the future. Therefore, in the follow-up research, we will focus on this issue to find the possible solution taking account of the context attention mechanism between focal and neighboring areas and cross-modality attention mechanism to making use of association relationship among multi-modal MRIs.

The second is the data limitation. Our current experiment is finished on the clinic dataset with only 59 cases. Current experiment results demonstrate the effectiveness of our proposed method. However, to prove the generalization and performance of proposed methods, we will continue to collect and label the additional cases to expand the dataset capacity in the future.

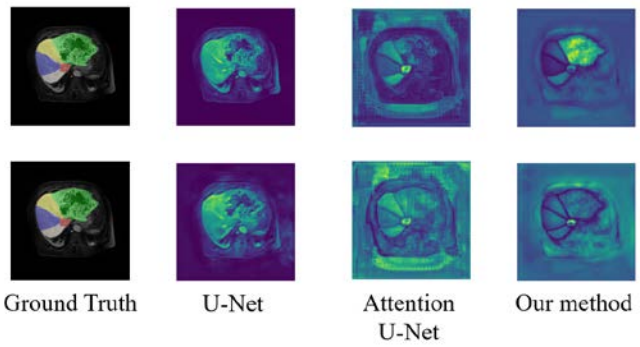


Fig. 10. Examples of visualization of intermediate layer results

5. Conclusion

In this paper, we have presented a Boundary-Aware Dual Attention Guided U-Net for liver segment segmentation, which fuses the low-level and high-level feature maps at the encoding and decoding path of U-Net. The fused feature map is regarded as the gated signal to calculate the attention-weighted the feature map to highlight key areas and suppress irrelevant areas. Specifically, our proposed boundary-aware dual attention module is composed of the spatial attention module and the semantic attention module in a parallel way, so as to calculate the attention weights from both spatial and semantic aspects to enhance boundary-related information. We conduct comprehensive comparative and ablate experiments to evaluate the performance of our approach. Experimental results show that BADA outperforms several baselines and state-of-the-art segmentation approaches. The visualization analysis demonstrate that discriminative features are learned with reflecting segment region and boundary information using two aspects attention weighting, where the spatial attention highlights the boundary and the semantic attention enhances the feature representation of the liver segment region and the liver blood vessel position. In all, the overall segmentation performance is improved with our proposed BADA guided mechanism. In the future, we will investigate how to incorporate context correlation information between liver segments to distinguish between liver segment boundaries and non-interest region boundaries. Moreover, we will do the further research of segmentation methods with making full use of multi-modal MRIs instead of using vein sequence MRI only. More complementary information among clinical MRI modalities can be considered in boundary attention calculating to improve the accuracy of segmentation.

Acknowledgement

This work is partly supported by National Natural Science Foundation of China (No.61871276, 82071876, 62171298), Beijing Natural Science Foundation (No.4202004, 7184199), Capital's Funds for Health Improvement and Research (No.2018-2-2023), National Key Research and Development Program of China (2016YFC0106900, 2019YFE0107800) and the National Research Foundation of Korea (NRF-2019K1A3A1A20093097).

References

[1] F. Sutherland, J. Harris, “Claude couinaud: a passion for the liver,” *Archives of surgery*, vol. 137, no. 11, pp. 1305-1310, 2002. Article (CrossRef Link)

[2] M. Lafortune, F. Madore, H. Patriquin, G. Breton, “Segmental anatomy of the liver: a sonographic approach to the couinaud nomenclature,” *Radiology*, vol. 181, no. 2, pp. 443-448, 1991. Article (CrossRef Link).

[3] R. Beichel, T. Pock, C. Janko, R. B. Zotter, B. Reitinger, A. Bornik, K. Palagyi, E. Sorantin, G. Werkgartner, H. Bischof, et al., “Liver segment approximation in ct data for surgical resection planning,” in *Medical Imaging 2004: Image Processing*, San Diego, California, United States, 2004, pp. 1435-1446. Article (CrossRef Link)

[4] X. Yang, J. Do Yang, H. P. Hwang, H. C. Yu, S. Ahn, B.-W. Kim, H. You, “Segmentation of liver and vessels from CT images and classification of liver segments for preoperative liver surgical planning in living donor liver transplantation,” *Computer methods and programs in biomedicine*, vol. 158, pp. 41-52, 2018. Article (CrossRef Link)

[5] D. Selle, B. Preim, A. Schenk, H.-O. Peitgen, "Analysis of vasculature for liver surgical planning," *IEEE transactions on medical imaging*, vol. 21, no. 11, pp. 1344-1357, 2002. Article (CrossRef Link)
[6] A. Sinha, J. Dolz, "Multi-scale self-guided attention for medical image segmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 1, pp. 121-130, 2021. Article (CrossRef Link)
[7] Cai J, Xia Y, Yang D, et al., "End-to-end adversarial shape learning for abdomen organ deep segmentation," in *Proc. of International Workshop on Machine Learning in Medical Imaging*, Shenzhen, China, pp. 124-132, 2019. Article (CrossRef Link)
[8] Liu Q, Chen C, Qin J, et al., "FedDG: Federated domain generalization on medical image segmentation via episodic learning in continuous frequency space," in *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1013-1023, 2021.
[9] G. Noyel, C. Vartin, P. Boyle, L. Kodjikian, "Retinal vessel segmentation by probing adaptive to lighting variations," in *Proc. of 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, Iowa City, United States, pp. 1246-1249, 2020. Article (CrossRef Link)
[10] Y. Liu, X. Jia, Z. Yang, D. Yang, "Style consistency constrained fusion feature learning for liver tumor segmentation," in *Proc. of Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*, Xi'an, China, pp. 390-396, 2019. Article (CrossRef Link)
[11] Zhang Y, Peng C, Peng L, et al., "Multi-phase Liver Tumor Segmentation with Spatial Aggregation and Uncertain Region Inpainting," in *Proc. of International Conference on Medical Image Computing and Computer-Assisted Intervention*, Strasbourg, France, pp. 68-77, 2021. Article (CrossRef Link)
[12] Y. Chen, P. Sun, "The research and practice of medical image enhancement and 3d reconstruction system," in *Proc. of 2017 International Conference on Robots & Intelligent System (ICRIS)*, Huai'an, China, pp. 350-353, 2017. Article (CrossRef Link)
[13] M. Moradi, P. Abolmaesumi, D. R. Siemens, E. E. Sauerbrei, A. H. Boag, P. Mousavi, "Augmenting detection of prostate cancer in transrectal ultrasound images using svm and rf time series," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 9, pp. 2214-2224, 2009. Article (CrossRef Link)
[14] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, X. Tang, "Residual attention network for image classication," in *Proc. of the IEEE conference on computer vision and pattern recognition*, Honolulu, Hawaii, United States, pp. 3156-3164, 2017. Article (CrossRef Link)
[15] X. Wang, R. Girshick, A. Gupta, K. He, "Non-local neural networks," in *Proc. of the IEEE conference on computer vision and pattern recognition*, Salt Lake City, Utah, United States, pp. 7794-7803, 2018. Article (CrossRef Link)
[16] J. Hu, L. Shen, G. Sun, "Squeeze-and-excitation networks," in *Proc. of the IEEE conference on computer vision and pattern recognition*, Salt Lake City, Utah, United States, pp. 7132-7141, 2018. Article (CrossRef Link)
[17] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, H. Lu, "Dual attention network for scene segmentation," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach*, California, United States, pp. 3146-3154, 2019. Article (CrossRef Link)
[18] O. Ronneberger, P. Fischer, T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. of International Conference on Medical image computing and computer-assisted intervention*, Munich, Germany, pp. 234-241, 2015. Article (CrossRef Link)
[19] O. Cicek, A. Abdulkadir, S. S. Lienkamp, T. Brox, O. Ronneberger, "3d u-net: learning dense volumetric segmentation from sparse annotation," in *Proc. of International conference on medical image computing and computerassisted intervention*, Athens, Greece, pp. 424-432, 2016. Article (CrossRef Link)
[20] V. Iglovikov, A. Shvets, "Ternausnet: U-net with vgg11 encoder pre-trained on imagenet for image segmentation," *arXiv preprint arXiv:1801.05746*, 2018.
[21] N. Ibtehaz, M. S. Rahman, "Multiresunet: Rethinking the u-net architecture for multimodal biomedical image segmentation," *Neural Networks*, vol. 121, pp. 74-87, 2020. Article (CrossRef Link)

[22] Z. Zhang, Q. Liu, Y. Wang, "Road extraction by deep residual u-net," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 5, pp. 749-753, 2018. Article (CrossRef Link)
[23] 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 transactions on medical imaging*, vol. 37, no. 12, pp. 2663-2674, 2018. Article (CrossRef Link)
[24] D. Jha, P. H. Smedsrud, M. A. Riegler, D. Johansen, T. De Lange, P. Halvorsen, H. D. Johansen, "Resunet++: An advanced architecture for medical image segmentation," in *Proc. of 2019 IEEE International Symposium on Multimedia (ISM)*, Taichung, China, pp. 225-2255, 2019. Article (CrossRef Link)
[25] K.-L. Tseng, Y.-L. Lin, W. Hsu, C.-Y. Huang, "Joint sequence learning and cross-modality convolution for 3d biomedical segmentation," in *Proc. of the IEEE conference on Computer Vision and Pattern Recognition*, Honolulu, Hawaii, United States, pp. 6393-6400, 2017. Article (CrossRef Link)
[26] X. Jia, Y. Liu, Z. Yang, D. Yang, "Multi-modality self-attention aware deep network for 3d biomedical segmentation," *BMC Medical Informatics and Decision Making*, vol. 20, no. 3, pp. 1-7, 2020. Article (CrossRef Link)
[27] X. Wang, S. Han, Y. Chen, D. Gao, N. Vasconcelos, "Volumetric attention for 3d medical image segmentation and detection," in *Proc. of International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, Shenzhen, China, pp. 175-184, 2019. Article (CrossRef Link)
[28] J. Zhang, Z. Jiang, J. Dong, Y. Hou, B. Liu, "Attention gate resu-net for automatic mri brain tumor segmentation," *IEEE Access*, vol. 8, pp. 58533-58545, 2020. Article (CrossRef Link)
[29] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, et al., "Attention u-net: Learning where to look for the pancreas," *arXiv preprint arXiv:1804.03999*, 2018.
[30] Z. Zhang, H. Fu, H. Dai, J. Shen, Y. Pang, L. Shao, "Et-net: A generic edge-attention guidance network for medical image segmentation," in *Proc. of International Conference on Medical Image Computing and Computer-Assisted Intervention*, Shenzhen, China, pp. 442-450, 2019. Article (CrossRef Link)
[31] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, Y. Zhou, "Transunet: Transformers make strong encoders for medical image segmentation," *arXiv preprint arXiv:2102.04306*, 2021.
[32] Li S, Zhang C, He X, "Shape-aware semi-supervised 3d semantic segmentation for medical images," in *Proc. of International Conference on Medical Image Computing and Computer Assisted Intervention*, Lima, Peru, pp. 552-561, 2020. Article (CrossRef Link)
[33] A. Chartsias, G. Papanastasiou, C. Wang, S. Semple, D. Newby, R. Dharmakumar, S. Tsaftaris, "Disentangle, align and fuse for multimodal and semi-supervised image segmentation," *IEEE Transactions on Medical Imaging*, vol. 40, no. 3, pp. 781-792, 2021. Article (CrossRef Link)
[34] K. He, X. Zhang, S. Ren, J. Sun, "Deep residual learning for image recognition," in *Proc. of the IEEE conference on computer vision and pattern recognition*, Las Vegas, United States, pp. 770-778, 2016. Article (CrossRef Link)
[35] T. He, Z. Zhang, H. Zhang, Z. Zhang, J. Xie, M. Li, "Bag of tricks for image classification with convolutional neural networks," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, Long Beach, California, United States, pp. 558-567, 2019. Article (CrossRef Link)
[36] D. P. Kingma, J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
[37] Selvaraju R R, Cogswell M, Das A, et al., "Grad-cam: Visual explanations from deep networks via gradient-based localization," *International Journal of Computer Vision*, vol. 128, pp. 336-359, 2020. Article (CrossRef Link)

Xibin Jia received the B.S. degree in wireless technology from Chongqing University, Chongqing, China in 1991, received the M.S. degree in intelligent instrument from North China Institute of Technology in 1996 and the Ph.D. degree in computer science and technology from Beijing University of Technology, Beijing, China, in 2007. Now, she is a Professor in the Faculty of Information at the Beijing University of Technology (BJUT) in Beijing, China.

Chen Qian received the B.S. degree in Beijing University of Technology, Beijing, China in 2019. He is currently pursuing a master degree at Beijing University of Technology (BJUT), Beijing, China. His current research interests include computer vision and biomedical image segmentation.

Zhenghan Yang is a chief physician and professor of radiology at Beijing Friendship Hospital Affiliated to Capital Medical University. He received Ph.D. degree in Beijing Medical University, Beijing, China in 1999. His current main research interests include imaging diagnosis of abdominal diseases, early imaging diagnosis of hepatocellular carcinoma.

Hui Xu received the Ph.D degree from Peking University Third Hospital. He is currently the attending physician in the Radiology Department of Beijing Friendship Hospital Affiliated to Capital Medical University. His current research interest include the deep learning in liver disease diagnosis.

Xianjun Han is a currently pursuing her Ph.D and graduated with a master's degree from Beijing Friendship Hospital Affiliated to Capital Medical University. Research interests: Artificial Intelligence and cardiothoracic imaging diagnostics.

Hao Ren is a Ph.D. candidate at Capital Medical University in Beijing, China, and has obtained a master's degree from Tai'an Medical College, Tai'an, China. His current research interests include deep learning in diffuse liver disease diagnosis.

Xinru Wu received the B.S.degree from Capital Medical University, Beijing, China and is a currently pursuing her M.S.degree from Beijing Friendship Hospital Affiliated to Capital Medical University, Beijing, China.

Boyang Ma graduated from North China Coal Medicine School, she is currently pursuing for a master's degree in medical imaging from Capital Medical University and has 9 years of work experience in radiology.

Dawei Yang received the B.S.degree from China Medical university, Shenyang, China and M.S.degree from Beijing Hospital, Ministry of Health Beijing, China. His current research interest include the deep learning in liver disease diagnosis.

Min Hong is a professor of Department of Computer Software Engineering at Soonchunhyang University. He received his B.S. from Soonchunhyang University, M.S. from University of Colorado at Boulder, and Ph.D. received from University of Colorado at Denver and Health Sciences Center in 1995, 2001, and 2005 respectively. His research interests are in computer graphics, physically-based modeling and simulation, bioinformatics, and image processing related applications. Currently he is the Director of Computer Graphics Laboratory.