



Research article

mfeeU-Net: A multi-scale feature extraction and enhancement U-Net for automatic liver segmentation from CT Images

Jun Liu¹, Zhenhua Yan¹, Chaochao Zhou², Liren Shao¹, Yuanyuan Han¹ and Yusheng Song^{3,*}

¹ Department of Information Engineering, Nanchang Hangkong University, Nanchang 330063, Jiangxi, China

² Department of Radiology, Feinberg School of Medicine, Northwestern University, Chicago 60611, Illinois, U.S.

³ Interventional Radiology, The People's Hospital of Ganzhou, Ganzhou 34100, Jiangxi, China

* **Correspondence:** Email: superookie@126.com; Tel: +8618855038096.

Abstract: Medical image segmentation of the liver is an important prerequisite for clinical diagnosis and evaluation of liver cancer. For automatic liver segmentation from Computed Tomography (CT) images, we proposed a Multi-scale Feature Extraction and Enhancement U-Net (mfeeU-Net), incorporating Res2Net blocks, Squeeze-and-Excitation (SE) blocks, and Edge Attention (EA) blocks. The Res2Net blocks which are conducive to extracting multi-scale features of the liver were used as the backbone of the encoder, while the SE blocks were also added to the encoder to enhance channel information. The EA blocks were introduced to skip connections between the encoder and the decoder, to facilitate the detection of blurred liver edges where the intensities of nearby organs are close to the liver. The proposed mfeeU-Net was trained and evaluated using a publicly available CT dataset of LiTS2017. The average dice similarity coefficient, intersection-over-union ratio, and sensitivity of the mfeeU-Net for liver segmentation were 95.32%, 91.67%, and 95.53%, respectively, and all these metrics were better than those of U-Net, Res-U-Net, and Attention U-Net. The experimental results demonstrate that the mfeeU-Net can compete with and even outperform recently proposed convolutional neural networks and effectively overcome challenges, such as discontinuous liver regions and fuzzy liver boundaries.

Keywords: liver segmentation; U-Net; multi-scale; feature extraction and enhancement; Res2Net; squeeze-and-excitation; edge attention

1. Introduction

According to a cancer statistical analysis conducted in the United States, it is estimated that the annual number of new cases of liver cancer reaches 42,230, ranking the sixth incidence among malignant tumors, and liver cancer causes 30,230 deaths annually, ranking the second mortality rate among malignant tumors [1]. Computed Tomography (CT) is a widely used medical imaging technique with the advantages of non-invasiveness and low cost. Liver CT images have been routinely used in the clinical diagnosis and staging of liver cancer, and liver segmentation based on CT images is essential for the planning of liver cancer treatments.

Many segmentation methods, such as manual segmentation, semi-automatic segmentation, and automatic segmentation, have been developed to segment the liver from CT images. Manual segmentation is time-consuming and segmentation accuracy is highly dependent on the experience and professional knowledge of operators [2]. Most semi-automatic segmentation methods, such as the level set method [3], threshold insertion method [4], and region growing method [5], are based on the intensities and gradient information in CT images. As a result, under-segmentation or over-segmentation may occur, when the intensities of a CT image are relatively uniform [6]. In addition, there are parameters to manually set in these methods, and different parameter settings can significantly influence segmentation performance.

Automatic segmentation methods are mainly referred to as those methods that rely on neural networks. In recent years, CNNs have been widely implemented in semantic classification and segmentation tasks, where the outputs of CNNs can represent the classification probabilities of the input image at either image or pixel levels. As a milestone work, the fully convolutional network (FCN) was proposed by Long et al. [7]. The FCN can perform the segmentation of images with varying sizes by introducing deconvolution layers, which can upsample the feature map of the last convolution layer to the same size as the input image. The upsampled feature map represents the probability of whether a pixel of the input image belongs to a target object to segment. However, the segmentation using the FCN is not fine enough and less sensitive to image details.

Based on the FCN, Ronneberger et al. [8] proposed the U-Net composed of an encoder and a decoder, both of which were connected via skip connections, effectively mitigating gradient vanishing in the training of deep CNNs. The U-Net has achieved great success in the field of medical image segmentation and inspired many researchers to continuously improve prediction accuracy. Zhang et al. [9] further introduced ResNet blocks into the U-Net to replace traditional convolution layers in the encoder, resulting in the Residual U-Net (Res-U-Net). The Res-U-Net can significantly reduce training time, as the number of parameters is approximately a quarter as many as the U-Net [10]. Because skip connections in the U-Net are a major cause of losing important feature information of images, Oktay et al. [11] incorporated Attention Gates into the U-Net to maintain important features and suppress irrelevant information in skip connections. Through the proposed Attention U-Net, they showed that Attention Gates can improve the performance of the U-Net without compromising computational efficiency.

Many scholars have conducted extensive research using U-Net on liver segmentation. For example, Li et al. [12] applied the attention mechanism to the U-Net++ [13] and proposed the Attention U-Net++. The network introduced an attention gate between the nested convolutional blocks, so that the features extracted at different levels can be merged with a task-related selection. Kushnure et al. [14] proposed a multi-scale approach (MS-U-Net) to improve the receptive field of convolution operations in U-Net encoder-decoder stages that extract multi-scale (global and local) features at a more granular level. They also recalibrated channel-wise responses of the aggregated multi-scale features to enhance

the high-level feature description ability of the network. Based on the U-Net framework, Wang et al. [15] proposed the EAR-U-Net, in which they applied the EfficientNetB4 as the encoder, introduced the attention gate in the skip connection, replaced the traditional convolution of the decoder with a residual block. This network can extract feature information effectively, while it eliminates irrelevant feature responses and prevents gradient vanishing effectively. Lv et al. [16] proposed a new 2.5D lightweight network named RIU-Net with the U-Net framework, which leverages the techniques from the residual and Inception networks and employs a hybrid loss function combining the binary cross-entropy (BCE) and Dice losses to speed up the convergence and improve accuracy. Wang et al. [17] proposed a 3D UNet called MAD-UNet for automatic liver segmentation from CT. They introduced multi-scale attention and deep supervision mechanisms, and replaced the ordinary skip connections with the long-short skip connections (LSSC) to preserve more edge detail. Fan et al. [18] noticed that the U-Net network tends to fuse semantically dissimilar feature maps via simple skip connections between the encoder and decoder path, resulting in semantic gaps between feature maps. Therefore, they proposed a Multi-Scale Nested U-Net (MSN-Net) to obtain semantically similar feature maps and alleviate the semantic gaps caused by simple skip connections, and showed that the MSN-Net can effectively improve liver segmentation accuracy. Lv et al. [19] proposed an improved ResU-Net framework (iResU-Net) for automatic liver CT segmentation. By employing a new loss function and data augmentation strategy, the accuracy of liver segmentation was improved in terms of two public datasets LiTS17 and SLiver07. Araújo et al. [20] proposed a cascade network model for liver segmentation by combining deep convolutional neural network models and image processing techniques. They demonstrated that liver segmentation, even when lesions are present in CT images, can be efficiently performed using their approach. Although automatic segmentation can significantly reduce human labor, contemporary automatic segmentation is not ideal in the face of challenging scenarios. For example, in liver CT slices, segmentation regions of the liver may be discontinuous, and the boundaries of the liver may be blurred due to low contrast to its surrounding tissues. Therefore, it is of great significance to improve the accuracy and robustness of neural network algorithms for liver disease diagnosis and treatments.

A deep CNN can extract feature maps with different levels, which contain different information about the object to segment [21]. The low-level feature maps extracted by the shallow layers of the network contain texture information and minor details of the image, such as the boundaries of organs. The high-level feature maps built on top of low-level features by the deep layers of the network can represent the shape and location information of organs [22]. To promote multi-scale feature extraction in liver segmentation, this paper proposes a new CNN model, mfeeU-Net, based on the architecture of U-Net [8], including multiple new modules. Firstly, the Res2Net block [23] was used as the backbone of the encoder to extract multi-scale features of the liver. Secondly, the Squeeze-and-Excitation block [24] was introduced to strengthen important channel information. Finally, the Edge Attention block was proposed to facilitate the detection of liver boundaries which are often obscure due to low contrast.

2. Methods

2.1. Model architecture

The architecture of the mfeeU-Net is shown in Figure 1. With a U-shaped architecture, the input feature map is gradually encoded by a series of downsamplings and decoded by a series of upsamplings. A prediction map is an output as the probabilities corresponding to liver pixels. Different from the traditional U-Net, three new modules were introduced to the mfeeU-Net for liver segmentation from

CT images. Considering the difficulty in extracting features of livers with diverse shapes and sizes, we used Res2Net blocks as the backbone of the downsampling path, while we also introduced ResNet blocks as the backbone of the upsampling path (Figure 1). To better extract channel information, we introduce the Squeeze-and-Excitation block into each Res2Net. Additionally, we proposed a new module called the Edge Attention block to fuse low-level features about boundary information and high-level features about position information from the encoder to the decoder through skip connections. Details of the Res2Net, Squeeze-and-Excitation, and Edge Attention blocks are described as follows.

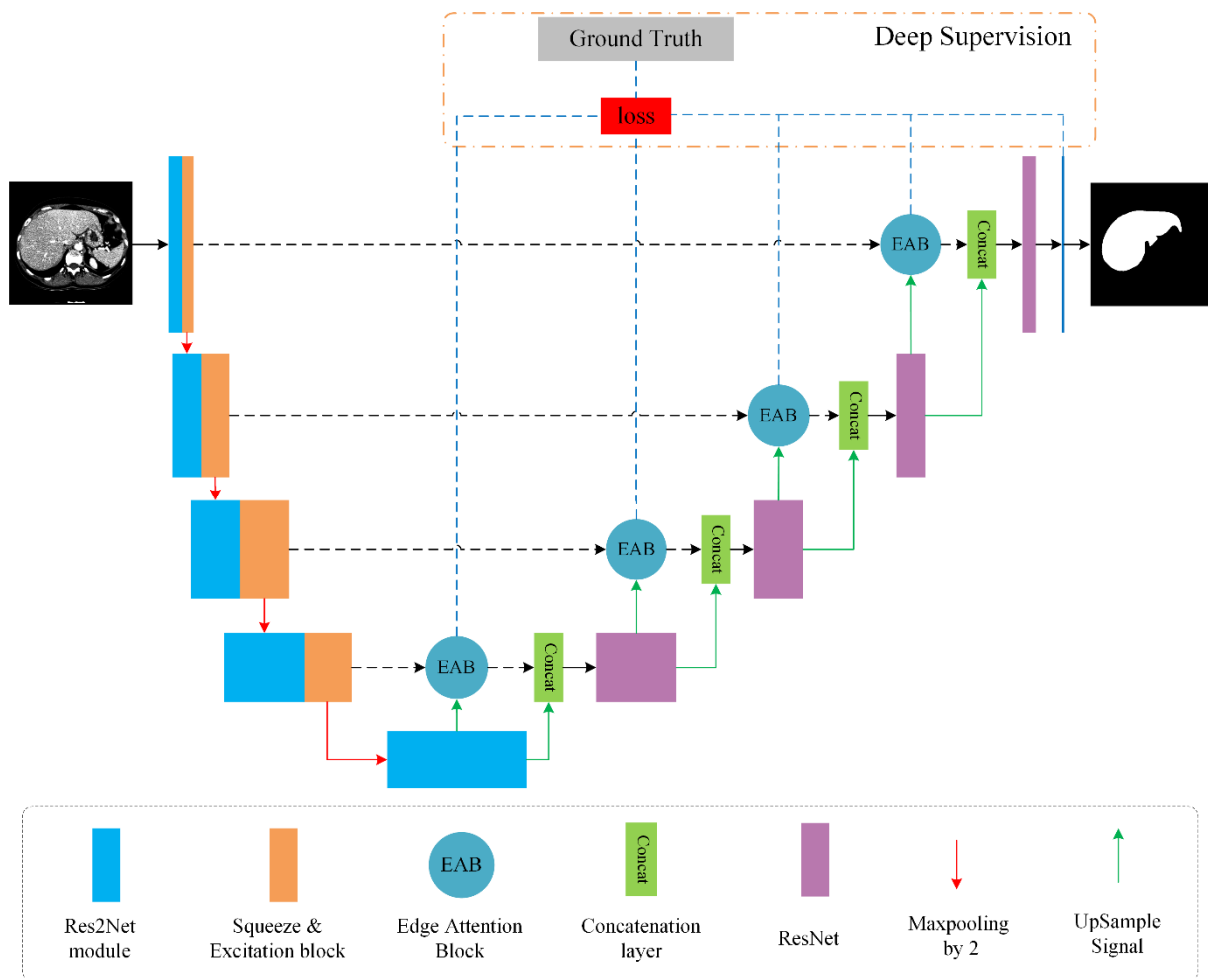


Figure 1. The diagram of the mfeeU-Net architecture.

2.2. Res2Net block

Although the learning ability of a CNN is improved as the increasing number of layers, training of a deep CNN often suffers gradient vanishing or gradient explosion [25]. This problem was well resolved by the ResNet blocks proposed by He et al. [10], in which input features were concatenated with output features by skip connections. To more effectively extract multi-scale features, Szegedy et al. [26] proposed the Inception network, which includes multiple convolutions with different kernel sizes. In this study, we introduced Res2Net which combined the ResNet and the Inception structure [23], to replace the convolutional layers in the downsampling path of the traditional U-Net. The structure of Res2Net

is shown in Figure 2. The Res2Net enables multi-scale feature extraction by learning the features of an input feature map through multiple branches, as formulated below:

$$y_i = \begin{cases} X_i & i = 1 \\ K_i(X_i) & i = 2 \\ K_i(X_i + y_i) & 2 < i \leq s \end{cases} \quad (1)$$

where X_i and Y_i are the input and output maps, respectively, on each branch denoted by i . K_i represent convolutions with different kernel sizes on each branch. In this work, we adopted four branches in the Res2Net block, i.e., $s = 4$.

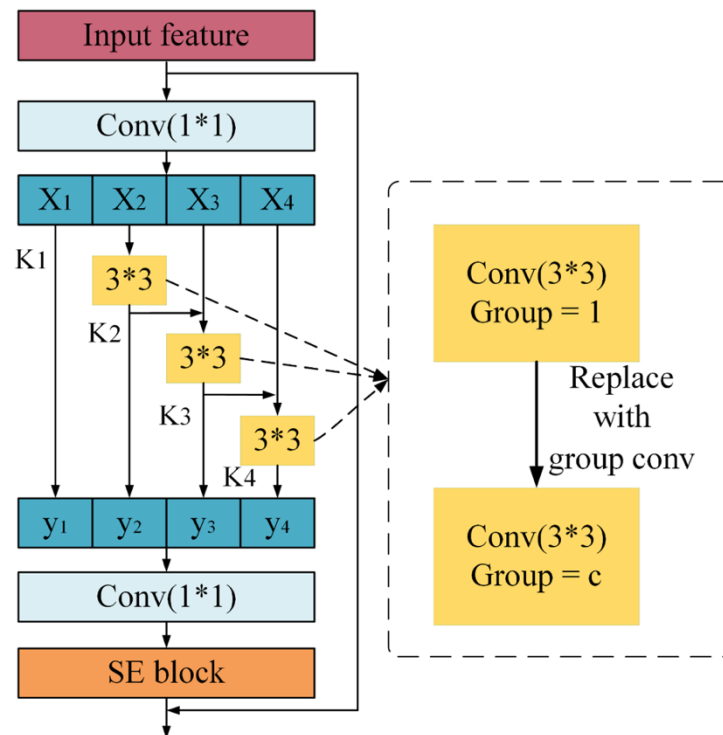


Figure 2. The diagram of a Res2Net block.

2.3. Squeeze and excitation block

After passing each Res2Net block (Figure 2), we set a Squeeze-and-Excitation (SE) block to further extract channel information of feature maps [24]. The structure of the SE block is presented in Figure 3. The SE block includes squeeze and excitation operations. First, an input feature map with a size of $W \times H \times C$ was squeezed to a tensor with a size of $1 \times 1 \times C$ by global average pooling. Then, the $1 \times 1 \times C$ tensor was excited by a two-layer fully connected network. Finally, the excited $1 \times 1 \times C$ tensor as weights was multiplied with each channel of the input feature map, resulting in an output feature map with the same size as the input feature map.

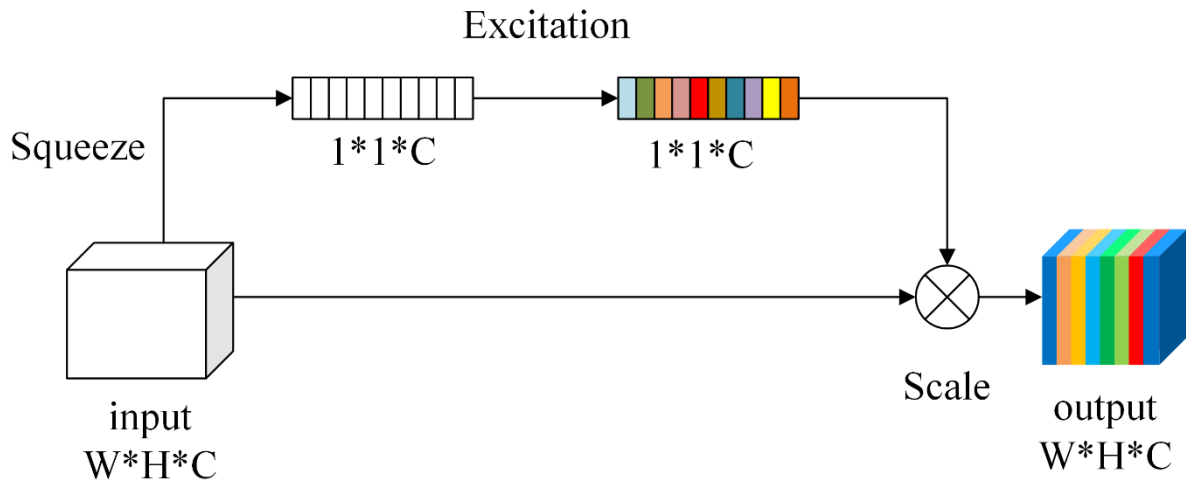


Figure 3. The diagram of an SE Block.

2.4. Edge attention block

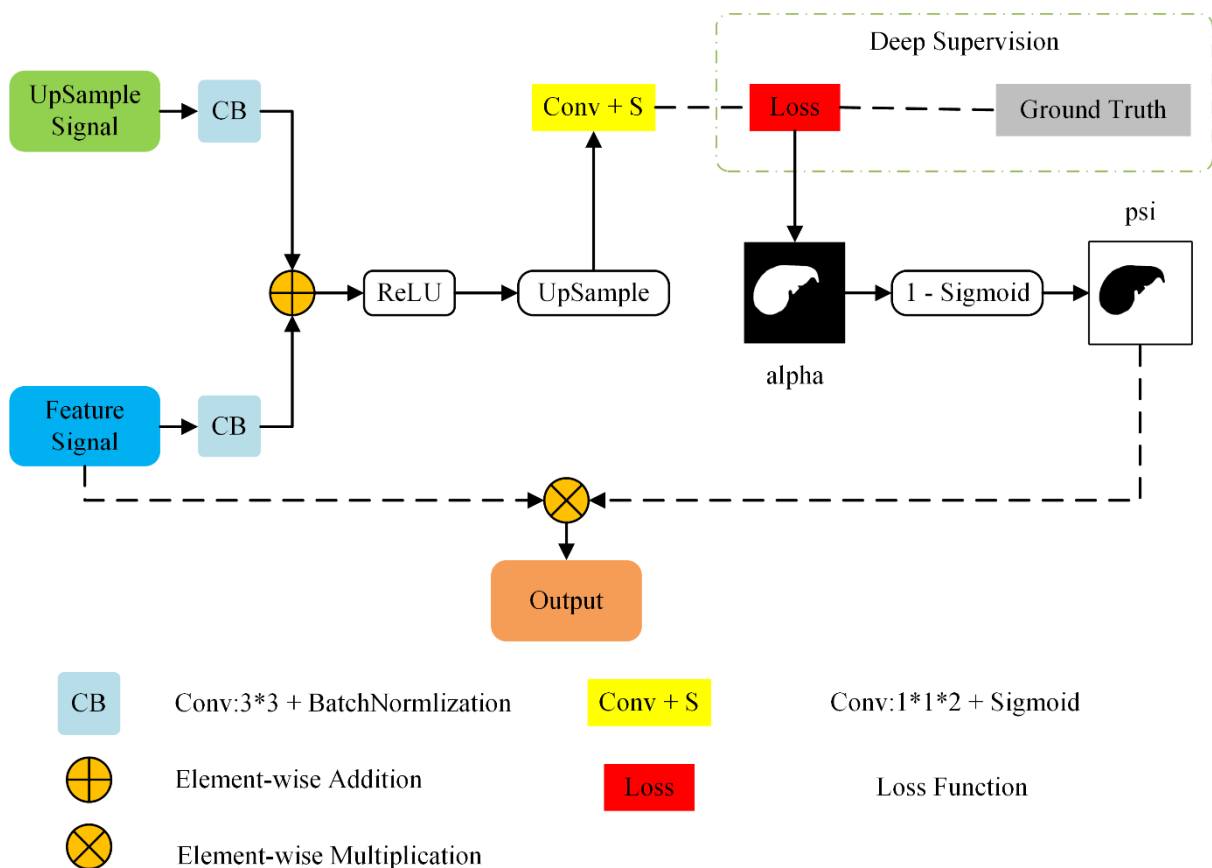


Figure 4. The diagram of an EA Block.

In most previous works [27–29], attention mechanisms for learning edge features were not specifically developed for liver segmentation using deep CNNs. During deep learning, low-level features contain spatial information (e.g., edges), while high-level features contain semantic

information (e.g., target locations) [30]. It is anticipated that both the low-level and high-level features are useful to identify the edges of the target object to segment. Therefore, we proposed an Edge Attention (EA) block to improve the detection of liver features at different levels. As shown in Figure 4, both the downsampling feature map from the encoder and the upsampling feature map from the decoder were fed to an EA block. Then, both the feature maps were linearly combined and activated by a ReLU function. The activated feature map was further processed by an upsampling operation and a convolution with sigmoid activation, to obtain a feature map of alpha. In each EA block, the loss between the alpha map and the ground-truth label map was minimized by gradient descent (see Section 2.5 Deep supervision). Furthermore, the alpha feature map was activated by “1-sigmoid” to obtain another feature map of psi. The psi feature map as pixel-by-pixel weights was multiplied with the input downsampling feature map to obtain the output feature map containing enhanced edge information of the liver.

2.5. Deep supervision

Previous work has demonstrated that the convergence speed and recognition ability in image classification can be improved by supervising the training of hidden layers in a deep CNN [31]. Therefore, we introduced deep supervision into each EA block of the mfeeU-Net (Figure 4). Specifically, we used the nearest neighbor interpolation algorithm to upscale the low-level features and exert a sigmoid function on the upscaled features to obtain a feature map of alpha with the same dimension as the ground-truth label map. In the meantime, the alpha feature map was trained to match to the ground-truth label map by minimizing their difference. In this way, weights of hidden layers in the mfeeU-Net were effectively trained with a lower risk of gradient vanishing.

The loss function for deep supervision consists of a binary cross-entropy (BCE) loss and a soft dice loss, as shown below,

$$Loss = Loss_{BCE}(A, B) + Loss_{Dice}(A, B) \quad (2)$$

where A and B represent the predicted probability map and ground-truth label map, respectively. The BCE and Dice losses are defined below:

$$Loss_{BCE}(A, B) = -B \log(A) - (1 - B) \log(1 - A) \quad (3)$$

$$Loss_{Dice}(A, B) = 1 - \frac{2 \sum_{i,j} B_{ij} A_{ij} + \epsilon}{\sum_{i,j} B_{ij}^2 + \sum_{i,j} A_{ij}^2 + \epsilon} \quad (4)$$

In Eq (4), ϵ is a small number that is introduced to avoid an undefined fraction.

3. Experimental setup

3.1. CT dataset

In this experiment, a publicly available CT dataset of LiTS2017 [32] was adopted for model training and testing. This dataset contains abdominal CT scans of 131 patients with doctor-manually annotated ground-truth liver labels. The number of slices in each CT volume ranges from 75 to 987 with a spacing ranging from 0.45 to 6.0 mm, and the resolution of each CT slice is 512×512 pixels, with a pixel size from 0.55 to 1.0 mm. In data preprocessing, the Hounsfield unit values of CT images

were set between -200 and 200 to exclude irrelevant details, and histogram equalization was used to enhance the contrast of the CT image. Figure 5 presents representative CT slices and annotated liver labels. In particular, there are two main challenges to segment the liver, including blurred liver edges (Figure 5(a)) and discontinuous liver regions (Figure 5(b)).

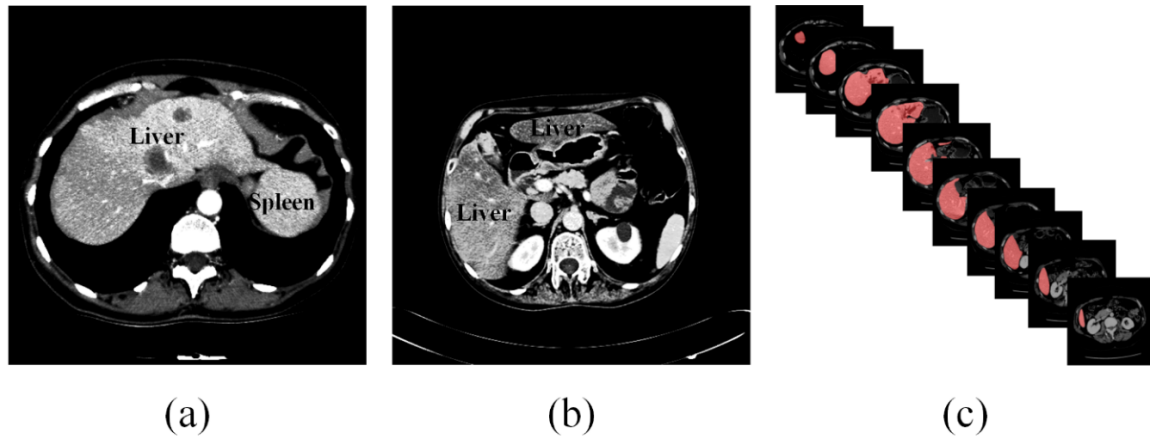


Figure 5. Representative CT slices with blurred liver edges (a) and discontinuous liver regions (b), as well as a sequence of CT slices with ground-truth liver labels (red).

3.2. Model training

In total, 131 CT volumes from the LiTS2017 dataset after pre-processing were chosen for CNN model training and testing. 22,880 paired slices and labels from 111 CT volumes were randomly selected as the training set, and 2599 paired slices and labels from the remaining 20 CT volumes were selected as the test set. During training, a 10-fold cross-validation scheme was adopted. 22,880 paired slices and labels in the training set were randomly divided into ten groups. In each iteration, one group was used for cross-validation, and the remaining nine groups were used for model training until the 10-fold cross-validation using each group was completed.

CNN models were developed and implemented using the TensorFlow deep learning framework. The Adam optimizer was used to optimize model weights using a learning rate of 1×10^{-4} . Model training was performed on a desktop with the hardware configuration of Intel(R) Xeon(R) E5-2678 V3, Nvidia GeForce RTX 3090, and 64 G Memory.

3.3. Evaluation metrics

To evaluate the segmentation performance of the mfeeU-Net proposed in this work, four commonly used evaluation metrics, including dice similarity coefficient (Dice) [33], intersection-over-union ratio (IOU) [18], sensitivity (SEN) [34] and relative value difference (RVD) [35] were adopted. These metrics are formulated below:

$$Dice = \frac{2|A \cap B|}{|A| + |B|} \quad (5)$$

$$IOU = \frac{|A \cap B|}{|A \cup B|} \quad (6)$$

$$SEN = \frac{|A \cap B|}{|A|} \quad (7)$$

$$RVD = \frac{|B| - |A|}{|A|} \quad (8)$$

where A is the image with the predicted liver mask and B is the image with the ground-truth liver label. The Dice, IOU, and SEN have a range between 0 and 1, and a larger metric indicates better performance. The RVD is unbounded, and the best performance corresponds to $RVD = 0$.

4. Experimental results

4.1. Training performance

The training performance of the mfeeU-Net was compared with three well-established CNNs, including the U-Net, Res-U-Net, and Attention U-Net. Figure 6 depicts the learning curves of different models in terms of the losses (Eq (2)) on the training set and cross-validation set, respectively. It can be seen that the loss curves of the mfeeU-Net were decreased more rapidly on both the training and cross-validation sets during 100 epochs of training, suggesting better convergence and generalizability of the mfeeU-Net.

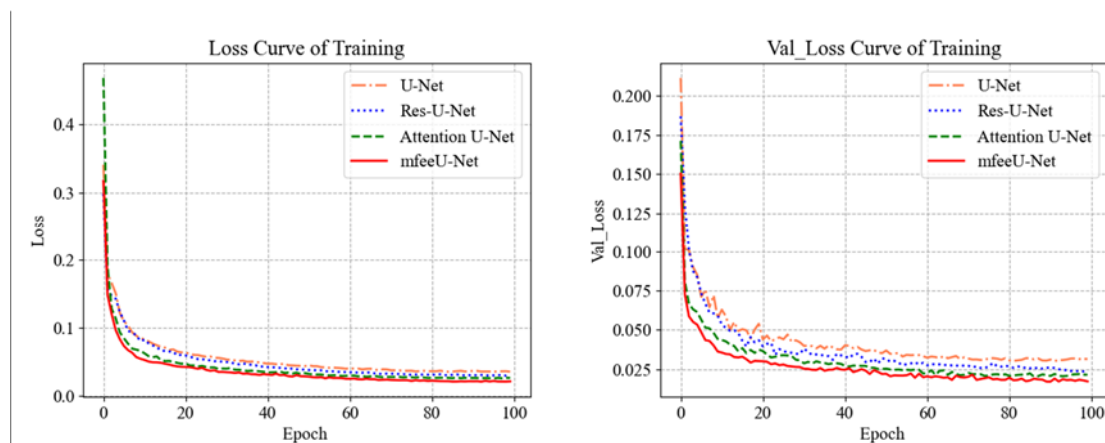


Figure 6. Learning curves of different models in terms of the losses on the training set (left) and cross-validation set (right) during 100 epochs of training.

4.2. Segmentation performance

The segmentation performance of the proposed mfeeU-Net on the test set (that was not used in model training) was compared with the U-Net, Res-U-Net, and Attention U-Net. As shown in Figure 7, the average Dice, IOU, and SEN of the mfeeU-Net for liver segmentation were 95.32%, 91.67%,

and 95.53%, respectively, and all these metrics were better than those of U-Net, Res-U-Net, and Attention U-Net. Furthermore, the RVD of the mfeeU-Net was 0.0026, that is closer to 0, also indicating better performance than other models. Compared with U-Net, the mfeeU-Net has remarkably improved the average Dice, IOU, and Sensitivity by 4.96%, 5.36%, and 7.57%, respectively.

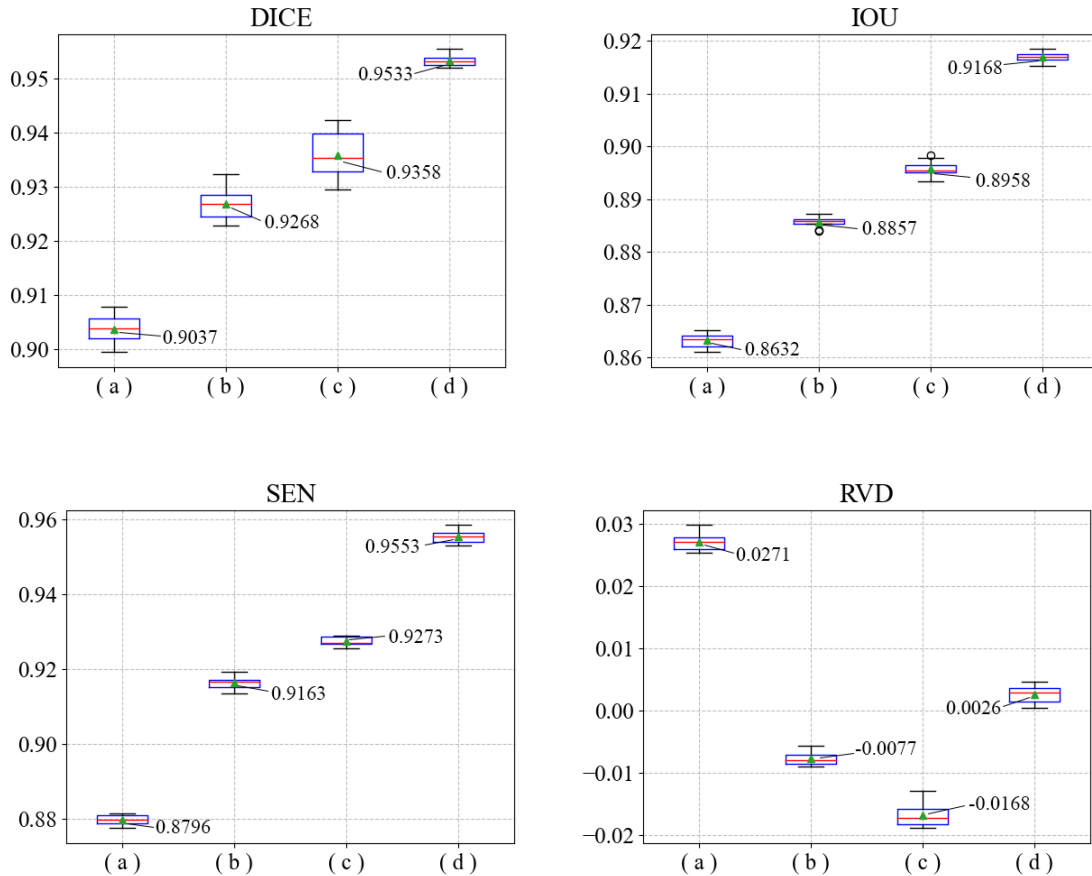


Figure 7. Comparisons of DICE, IOU, SEN, and RVD of four models. In each boxplot, (a) U-Net, (b) Res-U-Net, (c) Attention U-Net, and (d) mfeeU-Net. The green triangle in each box represents the average of the evaluation metric, and the red line in each box represents the median of the evaluation metric. In particular, the averages of each metric are presented.

Figure 8 visually compares the ground-truth 2D segmentation with prediction results using the proposed mfeeU-Net and other three classic models, in terms of the two challenging segmentation scenarios of blurred liver edges and discontinuous liver regions. It can be seen that the mfeeU-Net can provide more appropriate segmentation in detail. In contrast, the Attention U-Net, Res-U-Net, and U-Net led to over-segmentation in the scenario of blurred liver edges, but under-segmentation in the scenario of discontinuous liver regions.

Figure 9 further visually presents 3D liver models, which were reconstructed by assembling slices of 2D segmentation predicted by different CNN models. In this case, it can be observed that the reconstructed 3D liver model using the mfeeU-Net is more consistent with the ground-truth 3D liver model. Although the main portions of these reconstructed 3D liver models using the Attention U-Net, Res-U-Net, and U-Net can match the ground-truth 3D liver model, it is obvious that there are redundant parts predicted by these classic CNN models.

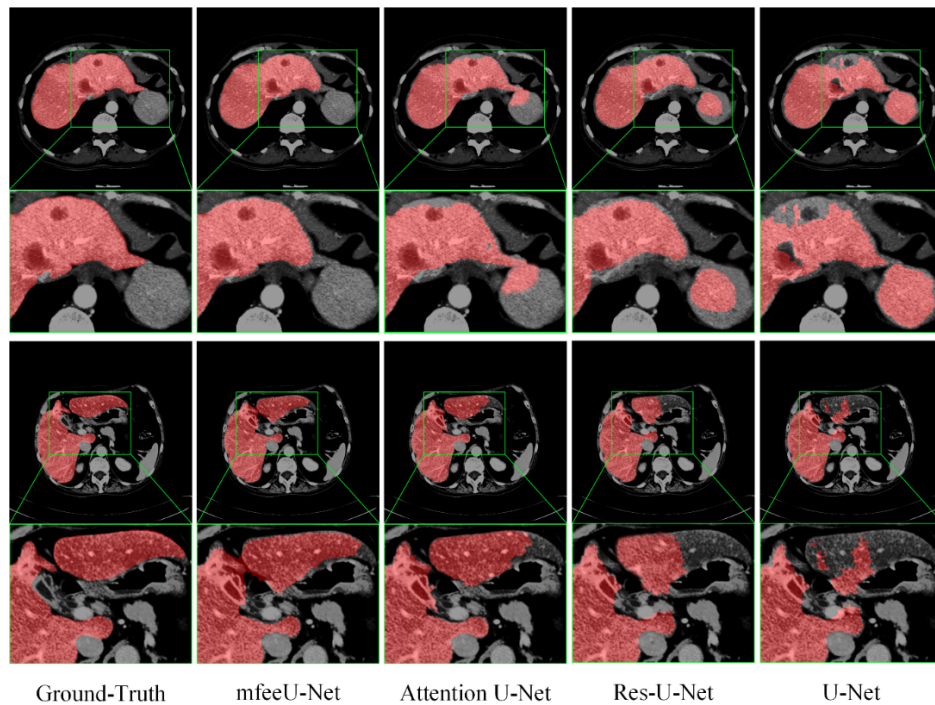


Figure 8. Visualization of 2D liver segmentation results in terms of challenging cases using the mfeeU-Net, Attention U-Net, Res-U-Net and U-Net, respectively. Rows 1: the scenario of blurred liver edges. Rows 2: the scenario of discontinuous liver regions. In particular, zoom-in views below each plot present detailed differences in the predictions using different models.

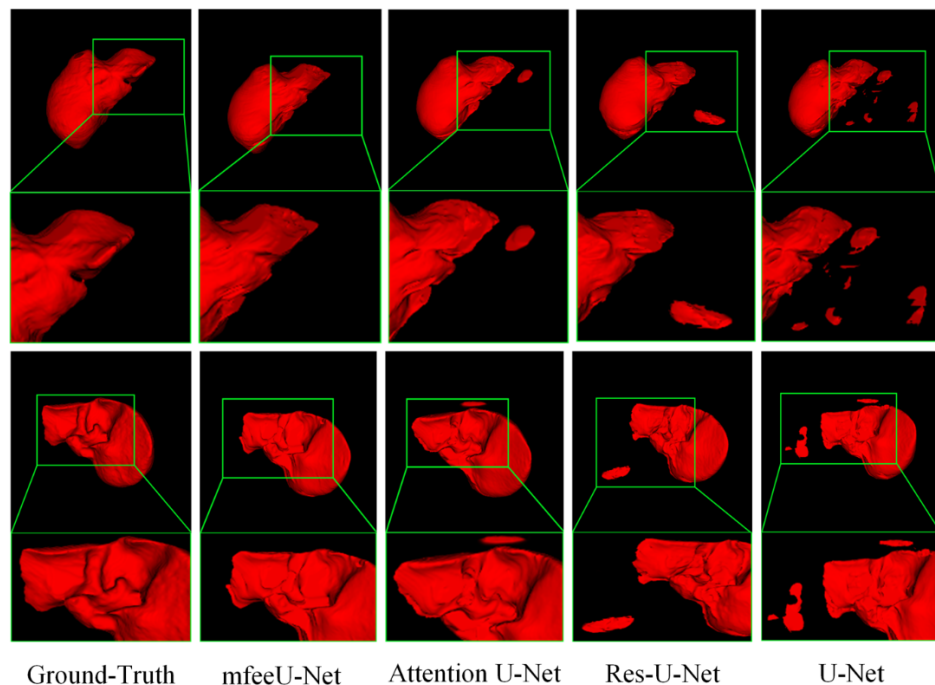


Figure 9. Visualization of 3D liver segmentation results using the mfeeU-Net, Attention U-Net, Res-U-Net and U-Net, respectively. Rows 1: the anterior view of the liver. Rows 2: the posterior view of the liver. In particular, zoom-in views below each plot present detailed differences in the predictions using different models.

4.3. Comparison with recently proposed methods

In order to further verify the robustness and effectiveness of the mfeeU-Net, we further compared its segmentation performance with several recently proposed semantic segmentation models. The results of these models including MSN-Net [18], EAR-U-Net [15], iResU-Net [19], and Araújo et al [20] were collected directly from the publications. As shown in Table 1, in terms of the DICE, IOU, SEN, and RVD metrics, the mfeeU-Net can compete with and even outperform these recently proposed models.

Table 1. Comparison of the segmentation performance of the mfeeU-Net with state-of-the-art methods. All experiments were performed on the LiTS17 dataset.

Method	DICE	IOU	SEN	RVD
MSN-Net (2021)	0.9424	0.9075	-	-
EAR-U-Net (2021)	0.9595	-	-	0.0050
iResU-Net (2022)	0.9428	-	-	-0.0025
Araújo et al (2022)	0.9564	-	0.9545	-0.0041
mfeeU-Net	0.9533	0.9168	0.9553	0.0026

4.4. Ablation study

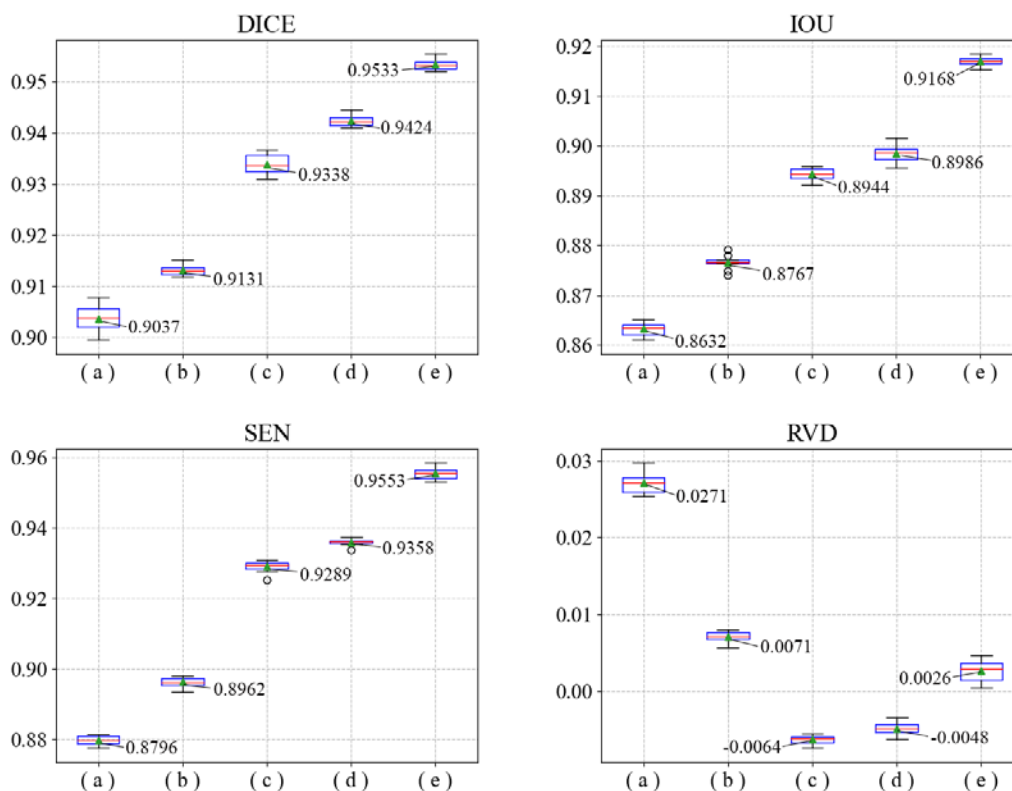


Figure 10. Comparisons of DICE, IOU, SEN, and RVD of different models in the ablation study. In each boxplot, (a) U-Net, (b) U-Net + Res2Net, (c) U-Net + Res2Net + SE, (d) U-Net + Res2Net + EA, and (e) mfeeU-Net. The green triangle in each box represents the average of the evaluation metric, and the red line in each box represents the median of the evaluation metric. In particular, the averages of each metric are presented.

To demonstrate the efficacy of each introduced module in the proposed mfeeU-Net, we performed an ablation study on the test set (that was not used in model training). Here, the Res2Net block, SE block, and EA block were individually added to the U-Net as the baseline, resulting in five different models: 1) U-Net, 2) U-Net + Res2Net, 3) U-Net + Res2Net + SE, 4) U-Net + Res2Net + EA, and 5) mfeeU-Net (i.e., U-Net + Res2Net + SE + EA). As shown in Figure 10, after introducing the Res2Net and SE / EA blocks to the U-Net, all metrics improved distinctly. Particularly, it is noted that the performance of U-Net + Res2Net + SE was similar to that of U-Net + Res2Net + EA. However, after introducing all the modules of Res2Net, SE, and EA blocks simultaneously, the resulting mfeeU-Net achieved the best segmentation performance. This ablation study clearly suggests that each module in the mfeeU-Net is indispensable in the improvement of liver segmentation performance.

The 2D masks of the liver segmented using different CNN models defined in the ablation study were visualized, as shown in Figure 11. In general, the 2D segmentation results using the proposed mfeeU-Net outperform other models missing Res2Net, SE, and/or EA blocks. Most importantly, it appears that the proposed mfeeU-Net can well tackle the challenges such as blurred liver edges (Figure 11, rows 1 and 2) and discontinuous liver regions (Figure 11, rows 3 and 4). In contrast, it is obvious that the U-Net (the last column in Figure 11) resulted in over-segmentation for the scenarios of blurred liver edges (Figure 11, rows 1 and 2), as well as under-segmentation for the scenarios of discontinuous liver regions (Figure 11, rows 3 and 4).

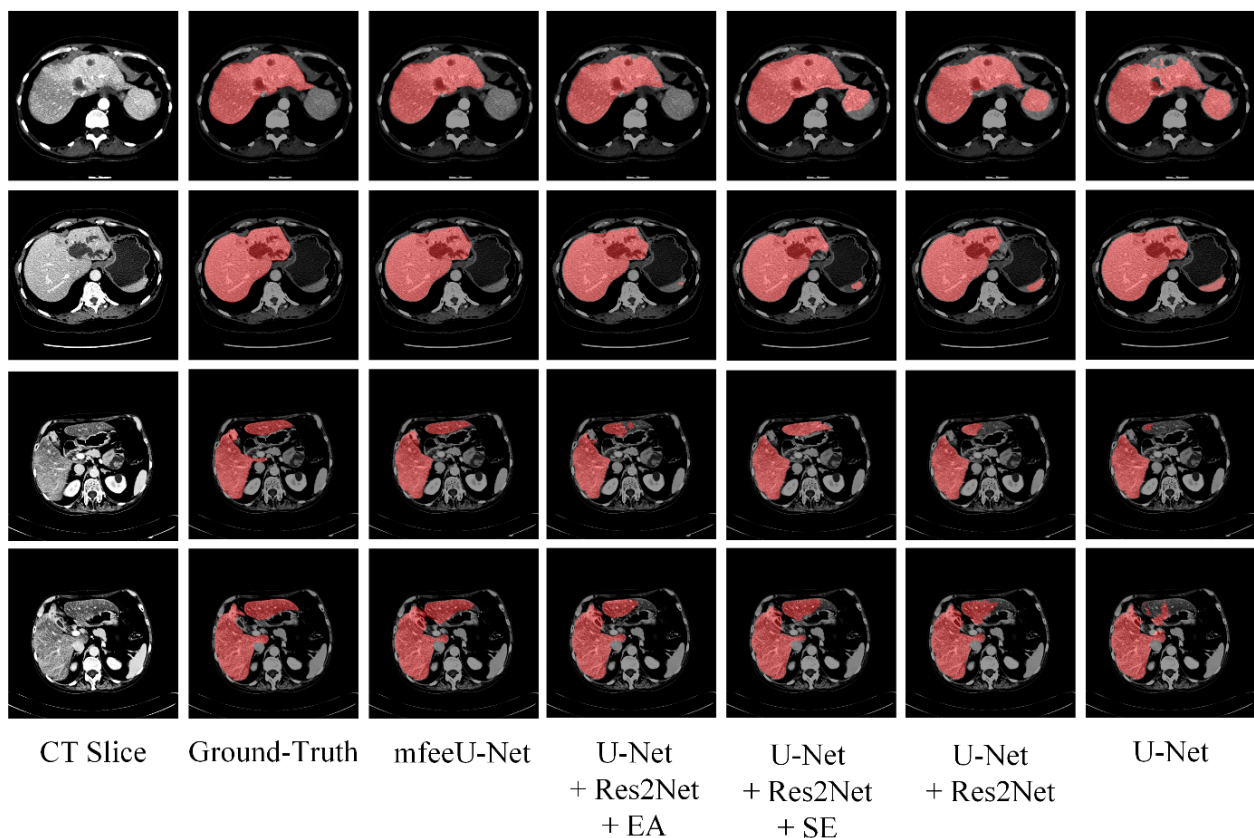


Figure 11. 2D liver segmentation on CT slices of a CT volume using different models defined in the ablation study. Rows 1 and 2 represent the scenarios of blurred liver edges. Rows 3 and 4 represent the scenarios of discontinuous liver regions.

Furthermore, we demonstrate that 3D liver models can be more accurately reconstructed by integrating 2D masks (Figure 11) predicted by the mfeeU-Net than other reduced models, as shown in Figure 12. It can be clearly seen that the reconstructed 3D liver model using the mfeeU-Net better matches to the ground-truth 3D liver model. However, larger discrepancy with respect to the ground-truth 3D liver model occurs, as more modules such as Res2Net, EA, and SE blocks are removed from the mfeeU-Net.

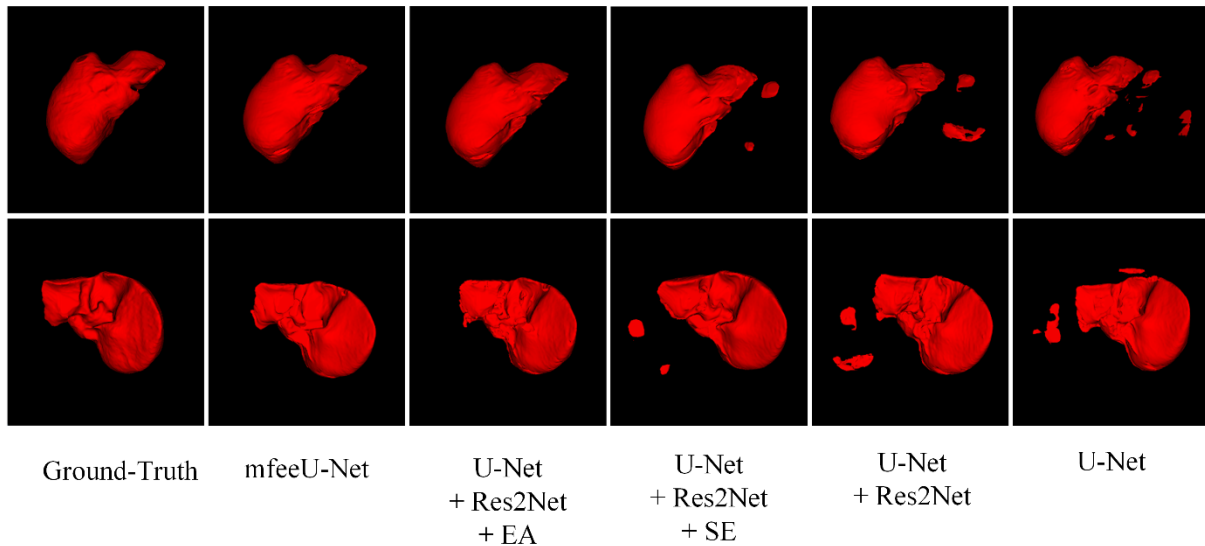


Figure 12. 3D reconstruction of a liver using 2D masks predicted using different models defined in the ablation study. (Rows 1 and 2 present the 3D liver model in the anterior and posterior views, respectively).

5. Discussion

Liver segmentation is an important prerequisite for the diagnosis and surgical planning of liver cancer [36]. However, there are two major challenges causing high uncertainties of liver segmentation from CT images using deep learning. The shapes and sizes of the liver in a CT slice sequence constantly change with discontinuous regions, and the CT intensities of the liver may be very close to the surrounding tissues or nearby organs [20]. In this work, we proposed and implemented the mfeeU-Net to tackle both challenges. Specifically, we introduced Res2Net and Squeeze-and-Excitation blocks for multi-scale feature extraction to improve the segmentation of discontinuous liver regions in a CT slice. Meanwhile, we proposed the Edge Attention block to enhance the detection of blurred liver boundaries.

Our ablation study demonstrated that enriched feature and channel information through multi-scale feature extraction combined with channel attention are beneficial to the improvement of segmentation performance. For example, as shown in our ablation study, after the Res2Net and Squeeze-and-Excitation blocks were introduced into the U-Net, the *Dice* and *IOU* increased from 90.37% and 86.32% to 93.38% and 89.44%, respectively (Figure 10). Furthermore, the Edge Attention block located on each skip connection in hidden layers with different network depths can promote the extraction of edge information from low-level and high-level feature maps. For example, after Res2Net and Edge Attention blocks were introduced into U-Net, the *Dice* and *IOU* increased from 91.31% and 87.67% to 94.24% and 89.86%, respectively (Figure 10). In general, each module including the Res2Net, Squeeze-and-Excitation, and Edge Attention blocks is essential, as the best segmentation performance

was achieved using the complete model with all modules, compared to other reduced CNN models (Figure 10). Correspondingly, the mfeeU-Net enabled more accurate 2D segmentation and 3D reconstruction than other reduced CNN models (Figures 11 and 12).

In this experiment, we also compared the training and segmentation performance of the proposed mfeeU-Net with three well-established CNN models, including U-Net, Res-U-Net, and Attention U-Net. In terms of training performance, the loss curve of the mfeeU-Net was decreased more rapidly on both the training and cross-validation sets during 100 epochs of training, highlighting better convergence and generalizability of the mfeeU-Net (Figure 6). For segmentation performance on the test set, all metrics of the mfeeU-Net are superior to other CNN models (Figure 7). Meanwhile, as seen from the visual results of 2D segmentation and reconstructed 3D models in Figures 8 and 9, the mfeeU-Net performs better than other classic models in terms of liver segmentation details and challenging segmentation scenarios, such as blurred liver edges and discontinuous liver regions. Because other CNN models lack mechanisms of multi-scale feature extraction or edge attention introduced in the mfeeU-Net, it further emphasizes that each module including the Res2Net, Squeeze-and-Excitation, and Edge Attention blocks are necessary, and they need to be simultaneously included in a CNN model for improving segmentation performance.

In this work, we adopted a publicly available CT dataset of LiTS2017 for model training and testing. As there were a small number of abdominal CT scans available in LiTS2017, the dataset may not be sufficient to train a 3D convolutional neural network for direct 3D reconstruction. Therefore, we adopted 2D segmentation predictions and developed the mfeeU-Net based on the 2D U-Net architecture. In this way, the size of training data can be significantly expanded, to avoid overfitting and better generalize the trained model. For example, a CT volume with an average dimension of $512 \times 512 \times 120$ in LiTS2017 includes 120 CT slices with an image size of 512×512 . Moreover, by predicting the 2D segmentation of livers on each CT slice, the resulting liver masks can be assembled to create 3D liver models, as demonstrated in Figures 9 and 12. However, we acknowledge that the 3D context in a CT volume cannot be directly leveraged in 2D segmentation [37]. Future work should further investigate whether 3D CNN models can achieve excellent performance for the 3D reconstruction of livers. In addition, more comprehensive tests of 2D liver segmentation using the mfeeU-Net are required for clinical applications.

6. Conclusions

In this work, a novel CNN architecture, mfeeU-Net, is proposed for liver CT image segmentation. To extract the multi-scale features of the liver in CT images more effectively, we used Res2Net blocks as the backbone of the encoder, while we introduced the Squeeze-and-Excitation blocks to strengthen the channel information and eliminate the influence of irrelevant information. Furthermore, we incorporated a new Edge Attention block to improve the detection of blurred liver boundaries due to the low contrast of the liver with surrounding tissues. Compared with well-established CNN models, the mfeeU-Net can provide more accurate 2D segmentation with better performance metrics. Our ablation study also shows that all introduced modules including the Res2Net, Squeeze-and-Excitation, and Edge Attention blocks are effective to achieve superior segmentation performance. Therefore, the mfeeU-Net is promising to assist in clinical diagnosis and decision-making for liver cancer treatments, whereas more rigorous tests using a larger CT dataset are necessary.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 61961028.

Conflict of interest

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

Reference

1. R. L. Siegel, K. D. Miller, H. E. Fuchs, A. Jemal, Cancer statistics, 2021, *CA Cancer J. Clin.*, **71** (2021), 7–33. <https://doi.org/10.3322/caac.21654>
2. S. Gul, M. S. Khan, A. Bibi, A. Khandakar, M. A. Ayari, M. E. H. Chowdhury, Deep learning techniques for liver and liver tumor segmentation: A review, *Comput. Biol. Med.*, **147** (2022), 105620. <https://doi.org/10.1016/j.compbiomed.2022.105620>
3. X. Shu, Y. Yang, B. Wu, Adaptive segmentation model for liver CT images based on neural network and level set method, *Neurocomputing*, **453** (2021), 438–452. <https://doi.org/10.1016/j.neucom.2021.01.081>
4. L. Soler, H. Delingette, G. Malandain, J. Montagnat, N. Ayache, C. Koehl, et al., Fully automatic anatomical, pathological, and functional segmentation from CT scans for hepatic surgery, *Comput. Aided Surg.*, **6** (2010) 131–142. <https://doi.org/10.3109/10929080109145999>
5. X. Lu, J. Wu, X. Ren, B. Zhang, Y. Li, The study and application of the improved region growing algorithm for liver segmentation, *Optik*, **125** (2014), 2142–2147. <https://doi.org/10.1016/j.ijleo.2013.10.049>
6. J. Wang, Y. Cheng, C. Guo, Y. Wang, S. Tamura, Shape-intensity prior level set combining probabilistic atlas and probability map constrains for automatic liver segmentation from abdominal CT images, *Int. J. Comput. Assist. Radiol. Surg.*, **11** (2016), 817–826. <https://doi.org/10.1007/s11548-015-1332-9>
7. J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2015), 3431–3440. <https://doi.org/10.1109/TPAMI.2016.2572683>
8. O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networks for biomedical image segmentation, in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*, (2015), 234–241. https://doi.org/10.1007/978-3-319-24574-4_28
9. Z. Zhang, Q. Liu, Y. Wang, Road extraction by deep residual u-net, *IEEE Geosci. Remote Sens. Lett.*, **15** (2018), 749–753. <https://doi.org/10.1109/LGRS.2018.2802944>
10. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2016), 770–778. <https://doi.org/10.1109/CVPR.2016.90>
11. 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. <https://doi.org/10.48550/arXiv.1804.03999>

12. C. Li, Y. Tan, W. Chen, X. Luo, Y. Gao, X. Jia, et al., Attention Unet++: A nested attention-aware U-net for liver CT image segmentation, in *2020 IEEE International Conference on Image Processing (ICIP)*, (2020), 345–349. <https://doi.org/10.1109/ICIP40778.2020.9190761>
13. 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*, (2018), 3–11.
14. D. T. Kushnure, S. N. Talbar, MS-UNet: A multi-scale UNet with feature recalibration approach for automatic liver and tumor segmentation in CT images, *Comput. Med. Imaging Graph.*, **89** (2021), 101885. <https://doi.org/10.1016/j.compmedimag.2021.101885>
15. J. Wang, X. Zhang, P. Lv, L. Zhou, H. Wang, EAR-U-Net: EfficientNet and attention-based residual U-Net for automatic liver segmentation in CT, preprint, arXiv:2110.01014. <https://doi.org/10.48550/arXiv.2110.01014>
16. P. Lv, J. Wang, H. Wang, 2.5D lightweight RIU-Net for automatic liver and tumor segmentation from CT, *Biomed. Signal Process. Control*, **75** (2022), 103567. <https://doi.org/10.1016/j.bspc.2022.103567>
17. J. Wang, X. Zhang, L. Guo, C. Shi, S. Tamura, Multi-scale attention and deep supervision-based 3D UNet for automatic liver segmentation from CT, *Math. Biosci. Eng.*, **20** (2023), 1297–1316. <https://doi.org/10.3934/mbe.2023059>
18. T. Fan, G. Wang, X. Wang, Y. Li, H. Wang, MSN-Net: A multi-scale context nested U-Net for liver segmentation, *Signal Image Video P*, **15** (2021), 1089–1097. <https://doi.org/10.1007/s11760-020-01835-9>
19. P. Lv, J. Wang, X. Zhang, C. Ji, L. Zhou, H. Wang, An improved residual U-Net with morphological-based loss function for automatic liver segmentation in computed tomography, *Math. Biosci. Eng.*, **19** (2022), 1426–1447. <https://doi.org/10.3934/mbe.2022066>
20. J. D. L. Araújo, L. B. da Cruz, J. O. B. Diniz, J. L. Ferreira, A. C. Silva, A. C. de Paiva, et al., Liver segmentation from computed tomography images using cascade deep learning, *Comput. Biol. Med.*, **140** (2022), 105095. <https://doi.org/10.1016/j.compbiomed.2021.105095>
21. H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, et al., Unet 3+: A full-scale connected Unet for medical image segmentation, in *ICASSP 2020—2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (2020), 1055–1059. <https://doi.org/10.1109/ICASSP40776.2020.9053405>
22. S. Sun, Z. Cao, D. Liao, R. Lv, A magnified adaptive feature pyramid network for automatic microaneurysms detection, *Comput. Biol. Med.*, **139** (2021), 105000. <https://doi.org/10.1016/j.compbiomed.2021.105000>
23. S. H. Gao, M. M. Cheng, K. Zhao, X. Y. Zhang, M. H. Yang, P. Torr, Res2Net: A new multi-scale backbone architecture, *IEEE Trans. Pattern Anal. Mach. Intell.*, **43** (2019), 652–662. <https://doi.org/10.1109/TPAMI.2019.2938758>
24. J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2018), 7132–7141. <https://doi.org/10.1109/CVPR.2018.00745>
25. A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, *Commun. ACM*, **60** (2017), 84–90. <https://doi.org/10.1145/3065386>

26. C. Szegedy, S. Ioffe, V. Vanhoucke, A. Alemi, Inception-v4, inception-resnet and the impact of residual connections on learning, in *Proceedings of the AAAI Conference on Artificial Intelligence*, **31** (2017). <https://doi.org/10.1609/aaai.v31i1.11231>
27. 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>
28. D. T. Kushnure, S. Tyagi, S. N. Talbar, LiM-Net: Lightweight multi-level multiscale network with deep residual learning for automatic liver segmentation in CT images, *Biomed. Signal Process. Control*, **80** (2023), 104305. <https://doi.org/10.1016/j.bspc.2022.104305>
29. T. Fan, G. Wang, Y. Li, H. Wang, MA-Net: A multi-scale attention network for liver and tumor segmentation, *IEEE Access*, **8** (2020), 179656–179665. <https://doi.org/10.1109/ACCESS.2020.3025372>
30. D. P. Fan, G. P. Ji, T. Zhou, G. Chen, H. Fu, J. Shen, et al., PraNet: Parallel reverse attention network for polyp segmentation, in *Medical Image Computing and Computer Assisted Intervention—MICCAI 2020*, (2020), 263–273. https://doi.org/10.1007/978-3-030-59725-2_26
31. C. Y. Lee, S. Xie, P. Gallagher, Z. Zhang, Z. Tu, Deeply-supervised nets, in *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics*, **38** (2015), 562–570. <https://doi.org/10.48550/arXiv.1409.5185>
32. P. Bilic, P. F. Christ, E. Vorontsov, G. Chlebus, H. Chen, Q. Dou, et al., The liver tumor segmentation benchmark (lits), *Med. Image Anal.*, **84** (2023), 102680. <https://doi.org/10.1016/j.media.2022.102680>
33. C. Zhang, Q. Hua, Y. Chu, P. Wang, Liver tumor segmentation using 2.5D UV-Net with multi-scale convolution, *Comput. Biol. Med.*, **133** (2021), 104–424. <https://doi.org/10.1016/j.combiomed.2021.104424>
34. Q. Jin, Z. Meng, C. Sun, H. Cui, R. Su, RA-UNet: A hybrid deep attention-aware network to extract liver and tumor in CT scans, *Front. Bioeng. Biotechnol.*, **8** (2020), 1471. <https://doi.org/10.3389/fbioe.2020.605132>
35. J. Li, X. Ou, N. Shen, J. Sun, J. Ding, J. Zhang, et al., Study on strategy of CT image sequence segmentation for liver and tumor based on U-Net and Bi-ConvLSTM, *Expert Syst. Appl.*, **180** (2021), 115008. <https://doi.org/10.1016/j.eswa.2021.115008>
36. R. Bi, C. Ji, Z. Yang, M. Qiao, P. Lv, H. Wang, Residual based attention-Unet combing DAC and RMP modules for automatic liver tumor segmentation in CT, *Math. Biosci. Eng.*, **19** (2022), 4703–4718. <https://doi.org/10.3934/mbe.2022219>
37. S. Shao, X. Zhang, R. Cheng, C. Deng, Semantic segmentation method of 3D liver image based on contextual attention model, in *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, (2021), 3042–3049. <https://doi.org/10.1109/SMC52423.2021.9659018>

