

**Densely Connected U-Net with Criss-Cross Attention for  
Automatic Liver Tumor Segmentation in CT Images**

Journal:	<i>IEEE/ACM Transactions on Computational Biology and Bioinformatics</i>
Manuscript ID	TCBBSI-2022-05-0425
Manuscript Type:	SI - BIBM 2021
Keywords:	I.4 Image Processing and Computer Vision < I Computing Methodologies, I.4.6 Segmentation < I.4 Image Processing and Computer Vision < I Computing Methodologies

**SCHOLARONE™**  
Manuscripts

# Densely Connected U-Net with Criss-Cross Attention for Automatic Liver Tumor Segmentation in CT Images

Qiang Li, Hong Song, Fengbo Yang, Zenghui Wei, Jingfan Fan, Danni Ai, Yucong Lin, Jian Yang, and Xiaoling Yu

**Abstract**—Automatic liver tumor segmentation plays a key role in radiation therapy of hepatocellular carcinoma. In this paper, we propose a novel densely connected U-Net model with criss-cross attention (CC-DenseUNet) to segment liver tumors in computed tomography (CT) images. The dense interconnections in CC-DenseUNet ensure the maximum information flow between encoder layers when extracting intra-slice features of liver tumors. Moreover, the criss-cross attention is used in CC-DenseUNet to efficiently capture only the necessary and meaningful non-local contextual information of CT images containing liver tumors. We evaluated the proposed CC-DenseUNet on the LiTS dataset and the 3DIRCADb dataset. Experimental results show that the proposed method reaches the state-of-the-art performance for liver tumor segmentation. We further experimentally demonstrate the robustness of the proposed method on a clinical dataset comprising 20 CT volumes.

**Index Terms**—dense interconnection, criss-cross attention, U-Net, liver tumor segmentation, CT images

## 1 INTRODUCTION

LIVER cancer is one of the malignant diseases with high morbidity and mortality in the world [1]. Accurate liver tumor measurements, including size, shape and location, from CT images can help doctors make treatment plans, but which need radiologists to delineate the liver tumors slice-by-slice. That is laborious and time-consuming. Therefore, automatic liver tumor segmentation is very necessary in the clinical. However, automatic segmentation of liver tumors from CT images is a challenging task. Firstly, the contrast between liver tumors and other liver tissues is low. Secondly, liver tumors have significant variations in size, shape and location. Last but not least, the numbers of liver tumors across different patients are various.

Recently, convolutional neural network (CNN) based liver tumor segmentation methods have achieved promising performance, which can be divided into four categories: two-dimensional (2D), three-dimensional (3D), two and half-dimensional (2.5D) and hybrid 2D/3D. 2D CNN based methods [2–11] only focus on the 2D context by treating the CT volume as many independent image slices, which lose the inter-slice information of continuous slices. 3D CNN based methods [12–14] have high computational cost and

GPU memory consumption. As a result, 2.5D CNN based methods [15–18] and hybrid 2D/3D CNN based methods [19–22] were proposed to balance the computational overhead and 3D context.

To address the above problems, we proposed a novel 2.5D densely connected U-Net (DenseU-Net) with criss-cross attention (CCA) mechanism, called CC-DenseUNet, to segment liver tumors in abdominal contrast-enhanced CT volumes. CC-DenseUNet borrows the idea of DenseNet [23], in which dense interconnections introduce direct connections between two layers with the same feature map size to facilitate a better feature propagation across the network. Moreover, inspired by CCNet [24], CC-DenseUNet employs a novel CCA module to harvest the contextual information of all pixels on criss-cross paths. The contributions of this work are three-fold.

- We designed the 2.5D CC-DenseUNet to learn inter-slice temporal information in the z-direction. And we used dense interconnections to extract intra-slice features and enable feature reuse, which is suitable for liver tumor feature extraction with large scale and shape variation.
- We inserted the CCA module in the bottleneck layer of CC-DenseUNet to efficiently capture the necessary and meaningful non-local contextual information, which can enhance the expressive power of liver tumor features.
- The proposed method reaches the state-of-the-art segmentation performance for liver tumor segmentation on the datasets of MICCAI 2017 Liver Tumor Segmentation (LiTS) Challenge [25] and 3DIRCADb [26], and also obtains good performance on a clinical dataset, which confirmed the robustness of our

• Q. Li, H. Song, F. Yang and Z. Wei are with the School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China.  
E-mail: liqiang3235@foxmail.com, anniesun@bit.edu.cn, georgeyfb@163.com, bitkf@bit.edu.cn.

• J. Fan, D. Ai and J. Yang are with the School of Optics and Electronics, Beijing Institute of Technology, Beijing 100081, China.  
E-mail: fjj@bit.edu.cn, danni@bit.edu.cn, jyang@bit.edu.cn.

• Y. Lin is with the School of Medical Technology, Beijing Institute of Technology, Beijing 100081, China. E-mail: linyucong@bit.edu.cn.

• X. Yu is with the Department of Interventional Ultrasound, Chinese PLA General Hospital, Beijing 100853, China. E-mail: dyuxl301@aliyun.com.

Corresponding authors: Hong Song, Jingfan Fan, Xiaoling Yu.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
model.

Compared with our original conference paper [27], the following improvements are made: 1) we focus on the liver tumor segmentation as its good performance achieved by CC-DenseUNet; 2) we include more extensive experiments on the LiTS and 3DIRCADb datasets to verify the effectiveness of the proposed model; 3) we conduct experiments on a clinical dataset to demonstrate the robustness of our proposed method. The rest of this paper is organized as follows. We first review related works in Section 2 and describe our method in Section 3. In Section 4, ablation and comparison studies are given and experimental results are analyzed. Section 5 presents our conclusion.

## 2 RELATED WORKS

2D CNN based methods: Christ *et al.* [8] proposed a method to automatically segment liver and lesions in CT abdomen images using cascaded fully convolutional networks (CFCNs). An FCN is firstly trained to segment the liver as region of interest (ROI) input for a second FCN. Then the second FCN solely segments lesions within the predicted liver ROIs of step one. Tang *et al.* [9] proposed a two-stage framework for 2D liver and tumor segmentation. The second stage is an edge enhanced network (E<sup>2</sup>Net) for more accurate liver and tumor segmentation. The E<sup>2</sup>Net explicitly models complementary objects (liver and tumor) and their edge information within the network to preserve the organ and lesion boundaries. Song *et al.* [10] proposed a 2D bottleneck feature supervised (BS) U-Net and applied it to liver and tumor segmentation. The BS U-Net containing an encoding network and a segmentation network has better performance and smaller number of parameters than most state-of-the-art 2D U-Net [28] based networks. Seo *et al.* [11] proposed a modified U-Net (mU-Net) that incorporates object-dependent high-level features for improving liver and tumor segmentation in CT images. The mU-Net adds a residual path with deconvolution and activation operations to the skip connection of the U-Net to avoid duplication of low-resolution information of features, and has additional convolution layers in the skip connection in order to extract high-level global features of small object inputs as well as high-level features of high-resolution edge information of large object inputs.

3D CNN based methods: Lyu *et al.* [13] proposed a novel approach, namely CouinaudNet, to train CNNs for liver tumor segmentation using Couinaud segment annotations. The CouinaudNet can estimate pseudo tumor masks from the Couinaud segment annotations as pixel-wise supervision for training a fully supervised tumor segmentation model. Chi *et al.* [14] proposed a novel framework to segment liver and tumors in abdominal CT volumes, in which a multi-branch network is integrated into the back-bone DenseU-Net for better extracting intra-slice features of liver and tumors and a 3D U-Net [29] is simplified for inhibiting the influence from neighboring pixels and alleviating the computational burden greatly.

2.5D CNN based methods: Han [16] developed a particular 2.5D deep convolutional neural network (DCNN) model for fully automatic liver lesion segmentation. This DCNN model takes five adjacent slices as input and produces a

segmentation map that corresponds to the center slice, and makes use of both long-range concatenation connections of U-Net and short-range residual connections of ResNet [30]. Wang *et al.* [17] proposed a volumetric attention (VA) module and integrated it in the Mask R-CNN [31] for liver and tumor segmentation. The VA module operates on the Mask R-CNN feature pyramids extracted from a target 2.5D image (composed of 3 adjacent slices), and neighboring contextual 2.5D images. Lv *et al.* [18] proposed a new 2.5D lightweight network for fast and accurate liver and tumor segmentation from CT images. This method is based on the U-Net framework, which leverages the techniques from the residual and Inception theories.

Hybrid 2D/3D CNN based methods: Li *et al.* [19] proposed a novel hybrid DenseU-Net (H-DenseUNet), which consists of a 2D DenseU-Net for efficiently extracting intra-slice features and a 3D counterpart for hierarchically aggregating volumetric contexts under the spirit of the auto-context algorithm for liver and tumor segmentation. Zhang *et al.* [20] proposed a light-weight hybrid convolutional network (LW-HCN) to segment the liver and its tumors in CT volumes. The LW-HCN is an encoder-decoder network, which uses 2D convolutions at the bottom of the encoder to decrease the complexity and uses 3D convolutions in other layers to explore both spatial and temporal information. Dey and Hong [21] developed a cascaded system that combines both 2D and 3D CNNs to effectively segment hepatic lesions, in which a 2D CNN operates on a slice- by-slice basis to segment the liver and larger tumors while a 3D CNN detects small lesions that are often missed in the 2D segmentation network. Song *et al.* [22] proposed a novel full-context CNN for automatic liver and tumor segmentation. This network bridges the gap between 2D and 3D contexts, which can utilize the temporal information along the Z axis in CT volume while retaining the spatial detail in each slice. Specifically, they separately designed a 2D spatial network for intra-slice features extraction and a 3D temporal network for inter-slice features extraction. These two networks are then guided by the squeeze and excitation layer that allows the flow of 2D context and 3D temporal information.

## 3 METHODS

### 3.1 CC-DenseUNet

The architecture of the proposed CC-DenseUNet is shown in Fig. 1a. CC-DenseUNet consists of a downsampling path, a bottleneck module, and an upsampling path. In the downsampling path, fifty-eight repetitive densely connected convolution blocks with different output dimensions of DenseNet-121 [23] are used to extract intra-slice features. As shown in Fig. 1b, densely connected paths can ensure the maximum information flow between network layers. The bottleneck module includes a dense block and the CCA module [24], which can efficiently capture only the necessary and meaningful non-local contextual information. The local feature map outputted from the fourth dense block serves as the input of the CCA module. This attention module inserted in CC-DenseUNet keeps important contextual information for the feature maps.

In the upsampling path, five successive up-sampling layers are utilized to stepwisely recovery the size of fea-

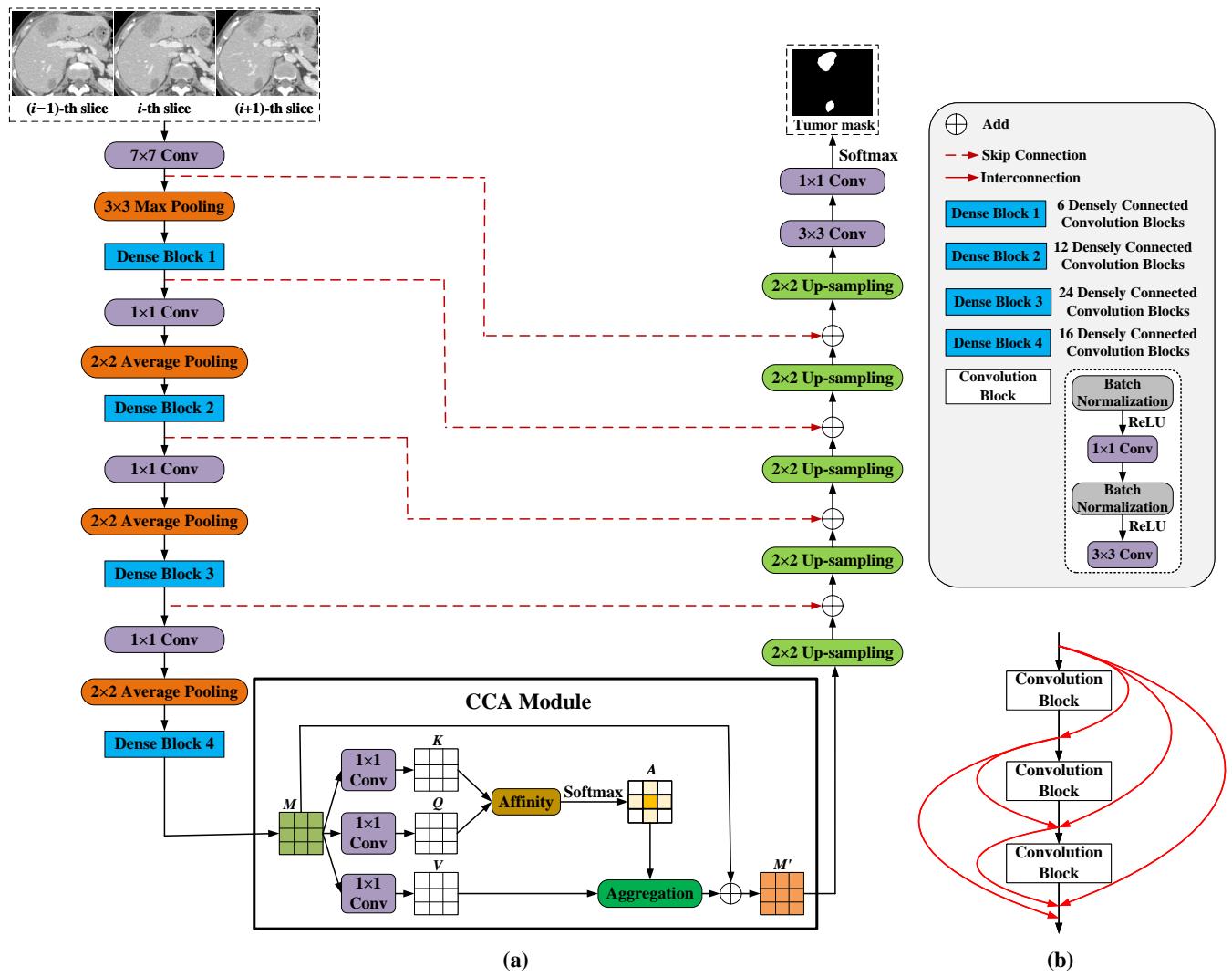


Fig. 1. (a) CC-DenseUNet architecture; (b) Schematic of dense block structure.

ture maps. Specifically, CC-DenseUNet is a 2.5D CNN model. Three adjacent slices ( $i - 1$  ( $i \geq 1$ ),  $i$ ,  $i + 1$  ( $i \leq \text{total number of slices} - 1$ )) cropped from volumetric images are fed into the model, and the 3D segmentation volume is generated by simply stacking the segmentation maps corresponding to the center slices.

As shown in Fig. 1a, the CCA module collects contextual information in vertical and horizontal directions for each pixel. Input images are passed through the network encoder to produce the feature map  $M$  with reduced dimension. Then,  $M$  is fed into the CCA module to generate a new feature map  $M'$ , which aggregates contextual information together for each pixel in its criss-cross path. The CCA module first applies two convolutional layers with  $1 \times 1$  filters on  $M \in \mathbb{R}^{C \times H \times W}$  to generate two feature maps  $K$  and  $Q$ , where  $\{K, Q\} \in \mathbb{R}^{C' \times H \times W}$ .  $C'$  is the number of channels, which is less than  $C$ . A vector  $Q_u \in \mathbb{R}^{C'}$  can be obtained at each position  $u$  of  $Q$ . A set  $\Omega_u \in \mathbb{R}^{(H+W-1) \times C'}$  can also be obtained by extracting feature vectors in the same column or row as position  $u$  from  $K$ . Attention map  $A \in \mathbb{R}^{(H+W-1) \times H \times W}$  is calculated by applying a softmax

layer on an affinity operation.

The affinity operation is formulated as:

$$d_{i,u} = Q_u \Omega_{i,u}^T \quad (1)$$

where  $d_{i,u} \in D$  is the degree of correlation between features  $Q_u$  and  $\Omega_{i,u}$ ,  $\Omega_{i,u} \in \mathbb{R}^{C'}$  is the  $i$ -th element of  $\Omega_u$ ,  $i = [1, \dots, |\Omega_u|]$ , and  $D \in \mathbb{R}^{(H+W-1) \times H \times W}$ .

The third convolutional layer applied on  $M$  generates  $V \in \mathbb{R}^{C \times H \times W}$  for feature adaption. A feature vector  $V_u \in \mathbb{R}^C$  and a set  $\Phi_u \in \mathbb{R}^{(H+W-1) \times H \times W}$  can be obtained at each position  $u$  in  $V$ . The set  $\Phi_u$  is a collection of feature vectors in  $V$ , which are in the same column or row as position  $u$ .

The contextual information is aggregated by

$$M'_u = \sum_{i=0}^{H+W-1} A_{i,u} \Phi_{i,u} + M_u \quad (2)$$

where  $M'_u$  is a feature vector in  $M' \in \mathbb{R}^{C \times H \times W}$  at position  $u$ ,  $\Phi_{i,u}$  is the  $i$ -th element of  $\Phi_u$ ,  $i = [0, \dots, H+W-1]$ , and  $A_{i,u}$  is a scalar value at channel  $i$  and position  $u$  in  $A$ .

### 3.2 Loss Function

We employed weighted cross-entropy function as the loss function of CC-DenseUNet, and it is defined as:

$$f(g, p) = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^3 w_i^c g_i^c \log p_i^c \quad (3)$$

where  $w_i^c$  denotes the weight,  $g_i^c$  is the ground truth label for voxel  $i$ , and  $p_i^c$  indicates the probability of voxel  $i$  belongs to class  $c$  (background, liver or tumor).

### 3.3 Implementation Details

We implemented CC-DenseUNet by using *Pytorch* package [32]. And we used stochastic gradient descent with a momentum of 0.9 and a weight decay of 0.0001 to optimize our model. A poly learning rate policy was employed in which the learning rate decayed according to the equation:

$$lr = lr * \left(1 - \frac{iter}{max\_iter}\right)^{0.9} \quad (4)$$

where  $iter$  denotes current iteration,  $max\_iter$  denotes maximum iterations, and the initial learning rate was 0.01. For data augmentation, we adopted random elastic deformation, rotation and scaling between 0.8 and 1.2 for the training data to teach the model the desired invariance properties.

In the training phase, the pre-trained DenseNet-121 on ImageNet [33] was used to initialize a part of weights of the encoder for our model. In the test phase, the original size of the image slice was adopted as liver tumors are extremely small in some CT volumes. We trained and tested all models using one NVIDIA GeForce 1080Ti GPU with 11 GB memory.

## 4 EXPERIMENTAL RESULTS AND DISCUSSION

### 4.1 Datasets

We carried experiments on LiTS, 3DIRCADb and a clinical dataset. The LiTS dataset contains 131 and 70 abdominal contrast-enhanced CT volumes for training and test, respectively. However, the ground truths of the LiTS test dataset are not publicly available. We randomly selected 31 CT volumes from the LiTS training dataset, called the custom LiTS testset, as test set. The remaining 100 CT volumes of the LiTS training dataset were used as training set. The 3DIRCADb dataset contains 20 venous phase-enhanced CT volumes. The dataset of 15 CT volumes with liver tumors was split into training and validation sets by 3-fold cross-validation. The clinical dataset is a real-life CT dataset acquired at Chinese PLA General Hospital. It compromises 20 portal venous phase CT scans from different patients. This clinical dataset was used to test the robustness of our model.

### 4.2 Performance Metrics

We used five metrics to evaluate the segmentation performance of liver tumors, and they are defined as follows.

i) Dice similarity coefficient (DSC)

$$DSC(Vol_{seg}, Vol_{ref}) = 2 \times \frac{|Vol_{seg} \cap Vol_{ref}|}{|Vol_{seg}| + |Vol_{ref}|} \quad (5)$$

where  $Vol_{seg}$  denotes segmented volume.  $Vol_{ref}$  denotes reference volume.

ii) volumetric overlap error (VOE)

$$VOE(Vol_{seg}, Vol_{ref}) = \left(1 - \frac{Vol_{seg} \cap Vol_{ref}}{Vol_{seg} \cup Vol_{ref}}\right) \times 100 \quad (6)$$

iii) relative absolute volume difference (RAVD)

$$RAVD(Vol_{seg}, Vol_{ref}) = \frac{|Vol_{seg} - Vol_{ref}|}{Vol_{ref}} \times 100 \quad (7)$$

iv) average symmetric surface distance (ASSD)

$$SD(Sur_{seg}, Sur_{ref}) = \sum_{a \in Sur_{seg}} \left[ \min_{b \in Sur_{ref}} \{dist(a, b)\} \right]$$

$$SD(Sur_{ref}, Sur_{seg}) = \sum_{b \in Sur_{ref}} \left[ \min_{a \in Sur_{seg}} \{dist(a, b)\} \right]$$

$$\begin{aligned} ASSD(Sur_{seg}, Sur_{ref}) &= ASSD(Sur_{ref}, Sur_{seg}) \\ &= \frac{SD(seg, ref) + SD(ref, seg)}{N_{Sur_{seg}} + N_{Sur_{ref}}} \end{aligned} \quad (8)$$

where  $Sur_{seg}$  and  $Sur_{ref}$  denote the surfaces of segmented and reference volumes.  $a$  and  $b$  are mesh points on  $Sur_{seg}$  and  $Sur_{ref}$ , respectively.  $dist(a, b)$  denotes the distance between  $a$  and  $b$ .  $N_{Sur_{seg}}$  and  $N_{Sur_{ref}}$  are the number of points on  $Sur_{seg}$  and  $Sur_{ref}$ , respectively.

v) root mean square of symmetric surface distance (RMSSD)

$$SD(Sur_{seg}, Sur_{ref})_{square} = \sum_{a \in Sur_{seg}} \left[ \min_{b \in Sur_{ref}} \{dist(a, b)\} \right]^2$$

$$SD(Sur_{ref}, Sur_{seg})_{square} = \sum_{b \in Sur_{ref}} \left[ \min_{a \in Sur_{seg}} \{dist(a, b)\} \right]^2$$

$$\begin{aligned} RMSSD(Sur_{seg}, Sur_{ref}) &= RMSSD(Sur_{ref}, Sur_{seg}) \\ &= \sqrt{\frac{SD(Sur_{seg}, Sur_{ref})_{square} + SD(Sur_{ref}, Sur_{seg})_{square}}{N_{Sur_{seg}} + N_{Sur_{ref}}}} \end{aligned} \quad (9)$$

### 4.3 Comparison with Other Methods

We compared our method with seven different methods (U-Net, 3D U-Net, CFCNs [8], DCNN [16], H-DenseUNet [19], LW-HCN [20], and nnU-Net [34]) on the custom LiTS testset and the 3DIRCADb dataset to validate its effectiveness. Table 1 shows that CC-DenseUNet outperforms 3D U-Net and CFCNs with 19.1% and 11.8% DSC improvements in liver tumor segmentation on the custom LiTS testset, respectively. Table 2 displays that CC-DenseUNet outperforms LW-HCN and nnU-Net with 2.1% and 2.5% DSC improvements in liver tumor segmentation on the 3DIRCADb dataset, respectively. Moreover, Table 1 and 2 present that CC-DenseUNet achieved a better performance than H-DenseUNet in terms of liver tumor segmentation with 1.9% and 1.3% DSC improvements on the custom LiTS testset and the 3DIRCADb dataset, respectively. Other quantitative metrics including VOE, RAVD, ASSD, and RMSSD, are also used to evaluate the performance of liver tumor segmentation. We made a comprehensive comparison between our method and other methods. The best values of these metrics are shown in bold.

TABLE 1

Quantitative liver tumor segmentation results compared with other methods on the custom LiTS testset.  $\uparrow$  means the value is the larger the better and  $\downarrow$  is the opposite. Best values are in bold.

Method	Performance Metrics				
	DSC $\uparrow$	VOE (%) $\downarrow$	RAVD (%) $\downarrow$	ASSD (mm) $\downarrow$	RMSSSD (mm) $\downarrow$
U-Net [28]	0.526 $\pm$ 0.225	26.80 $\pm$ 19.04	21.82 $\pm$ 20.81	12.30 $\pm$ 11.82	29.40 $\pm$ 10.41
3D U-Net [29]	0.550 $\pm$ 0.216	25.10 $\pm$ 15.39	12.30 $\pm$ 1.90	10.30 $\pm$ 1.49	22.40 $\pm$ 5.62
CFCNs [8]	0.623 $\pm$ 0.112	19.70 $\pm$ 15.21	10.90 $\pm$ 1.40	9.20 $\pm$ 0.43	18.30 $\pm$ 4.89
DCNN [16]	0.670 $\pm$ 0.212	45.00 $\pm$ 15.02	14.00 $\pm$ 1.20	8.30 $\pm$ 0.24	17.20 $\pm$ 5.23
H-DenseUNet [19]	0.722 $\pm$ 0.119	24.50 $\pm$ 14.86	<b>5.30<math>\pm</math>1.10</b>	6.66 $\pm$ 0.22	24.32 $\pm$ 4.32
LW-HCN [20]	0.730 $\pm$ 0.158	24.46 $\pm$ 14.43	9.10 $\pm$ 2.10	4.70 $\pm$ 0.26	24.25 $\pm$ 4.87
nnU-Net [34]	<b>0.763<math>\pm</math>0.062</b>	21.93 $\pm$ 13.00	7.20 $\pm$ 1.50	<b>1.54<math>\pm</math>0.20</b>	<b>6.36<math>\pm</math>5.16</b>
CC-DenseUNet (Ours)	0.741 $\pm$ 0.079	<b>16.90<math>\pm</math>3.00</b>	6.20 $\pm$ 1.80	6.09 $\pm$ 3.82	9.40 $\pm$ 3.16

TABLE 2

Quantitative liver tumor segmentation results compared with other methods on the 3DIRCADb dataset.  $\uparrow$  means the value is the larger the better and  $\downarrow$  is the opposite. Best values are in bold.

Method	Performance Metrics				
	DSC $\uparrow$	VOE (%) $\downarrow$	RAVD (%) $\downarrow$	ASSD (mm) $\downarrow$	RMSSSD (mm) $\downarrow$
U-Net [28]	0.510 $\pm$ 0.204	62.60 $\pm$ 24.21	11.20 $\pm$ 2.71	12.50 $\pm$ 4.61	18.30 $\pm$ 13.81
3D U-Net [29]	0.543 $\pm$ 0.122	50.20 $\pm$ 17.64	3.80 $\pm$ 2.40	11.10 $\pm$ 3.46	16.70 $\pm$ 5.64
CFCNs [8]	0.612 $\pm$ 0.129	40.40 $\pm$ 17.26	4.30 $\pm$ 1.50	8.40 $\pm$ 4.47	12.80 $\pm$ 4.52
DCNN [16]	0.600 $\pm$ 0.106	56.50 $\pm$ 18.36	4.10 $\pm$ 1.90	6.36 $\pm$ 2.26	11.70 $\pm$ 4.92
H-DenseUNet [19]	0.650 $\pm$ 0.101	<b>49.70<math>\pm</math>12.86</b>	<b>3.30<math>\pm</math>1.30</b>	<b>5.29<math>\pm</math>0.84</b>	11.10 $\pm$ 3.28
LW-HCN [20]	0.642 $\pm$ 0.102	39.83 $\pm$ 14.62	4.90 $\pm$ 2.70	7.49 $\pm$ 2.17	12.52 $\pm$ 3.57
nnU-Net [34]	0.638 $\pm$ 0.156	36.25 $\pm$ 13.25	4.10 $\pm$ 2.90	6.77 $\pm$ 1.65	<b>10.67<math>\pm</math>3.10</b>
CC-DenseUNet (Ours)	<b>0.663<math>\pm</math>0.093</b>	<b>34.30<math>\pm</math>13.82</b>	<b>3.20<math>\pm</math>1.50</b>	<b>2.23<math>\pm</math>1.49</b>	10.90 $\pm$ 1.50

Fig. 2 presents that CC-DenseUNet outperforms the remaining liver tumor segmentation methods, except nnU-Net. Fig. 3 shows that CC-DenseUNet outperforms the other seven liver tumor segmentation methods. The experimental results validated the superiority of our proposed method in comparison with other methods.

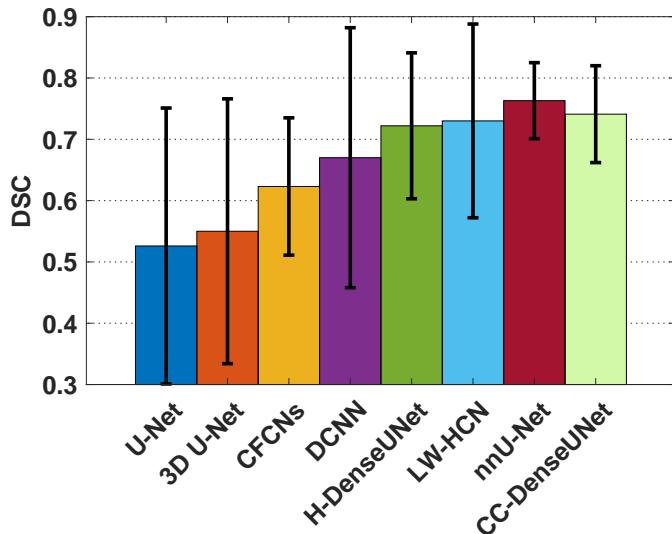


Fig. 2. Bar chart with error bars of DSC scores for different methods on the custom LiTS testset.

Fig. 4 and 5 show some of the qualitative segmentation results achieved by eight different methods on the custom LiTS testset and the 3DIRCADb dataset, respectively.

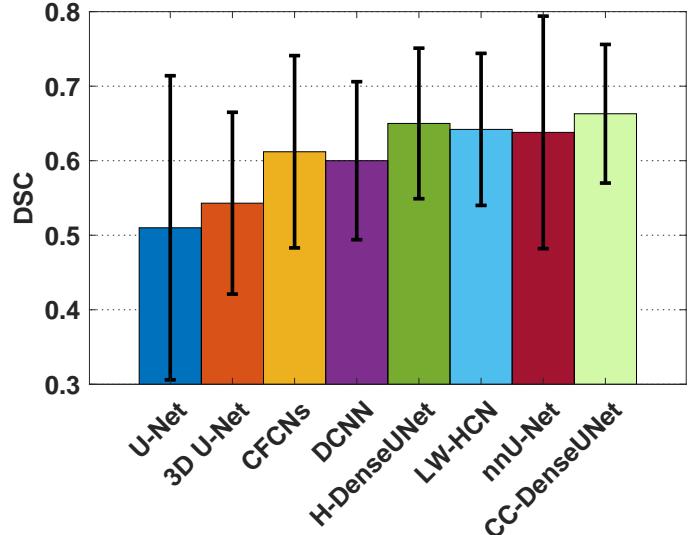


Fig. 3. Bar chart with error bars of DSC scores for different methods on the 3DIRCADb dataset.

#### 4.4 Ablation Study

We conducted ablation studies by whether to add dense interconnections on U-Net, whether to insert the CCA module into DenseU-Net, and whether to use the pre-trained model when training DenseU-Net and CC-DenseUNet.

One advantage of CC-DenseUNet is that we can train the network by transfer learning with the pre-trained model, which can help the network converge faster and achieve a lower loss value. The quantitative results of liver tumor segmentation in Table 3 and 4 demonstrate that the pre-trained model can help the network achieve a better performance consistently.

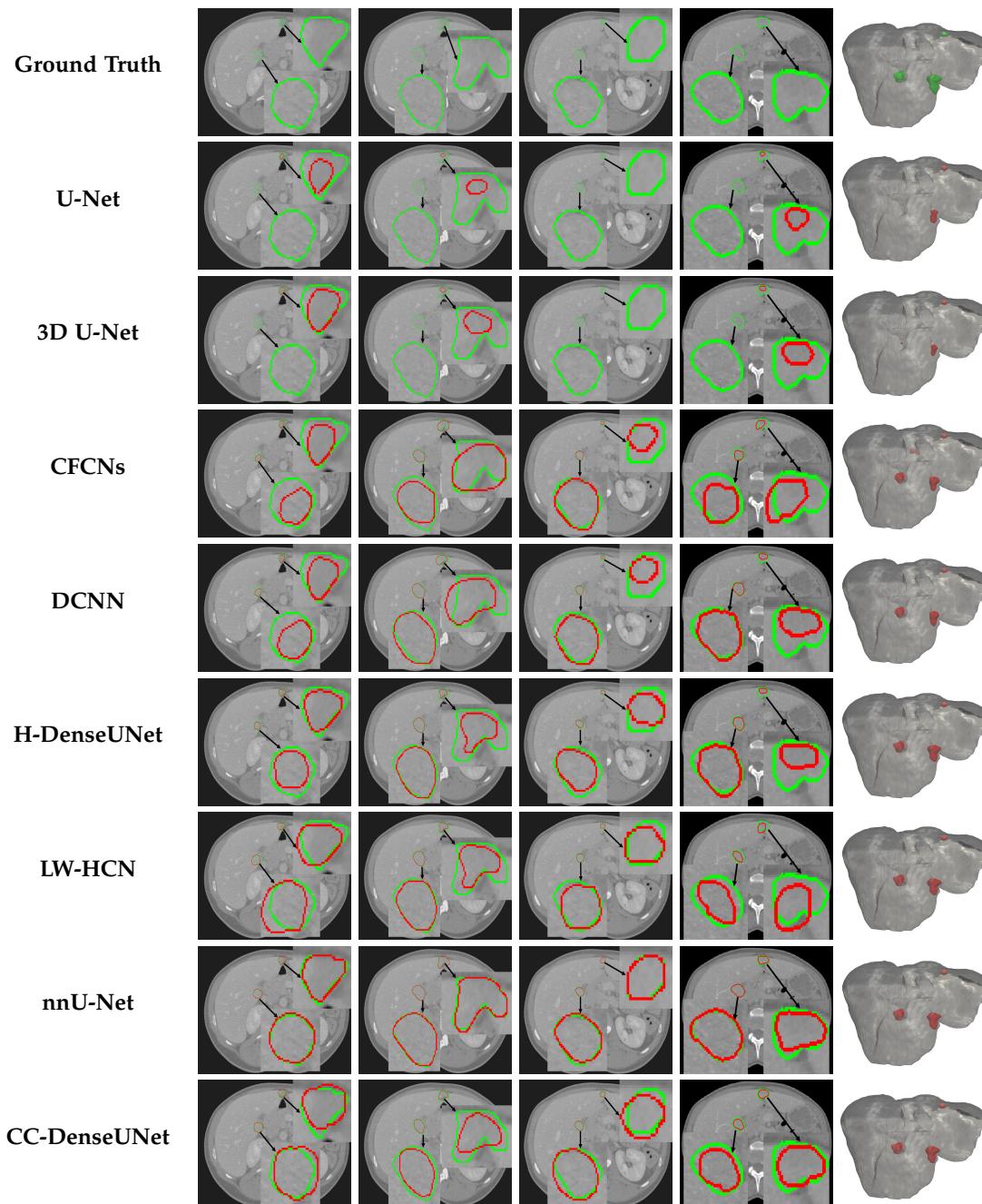


Fig. 4. Qualitative segmentation results by different methods on the custom LiTS testset. The green outlines and objects denote the true liver tumors while the red ones denote the segmented liver tumors.

TABLE 3  
Quantitative results for liver tumor segmentation by ablation models on the custom LiTS testset. ↑ means the value is the larger the better and ↓ is the opposite. Best values are in bold.

Model	Performance Metrics				
	DSC ↑	VOE (%) ↓	RAVD (%) ↓	ASSD (mm) ↓	RMSSSD (mm) ↓
U-Net	0.526±0.225	26.80±19.04	21.82±20.81	12.30±11.82	29.40±10.41
DenseU-Net without pre-trained model	0.607±0.167	17.70±13.06	10.00±2.70	10.50±9.29	17.50±9.68
DenseU-Net with pre-trained model	0.702±0.165	16.90± <b>3.00</b>	6.60±2.40	9.70±8.98	9.63±7.54
CC-DenseUNet without pre-trained model	0.738± <b>0.079</b>	15.26±6.31	<b>6.20</b> ±2.10	7.70± <b>3.82</b>	9.40±6.58
CC-DenseUNet with pre-trained model	<b>0.741</b> ±0.084	<b>13.60</b> ±4.43	6.40± <b>1.80</b>	<b>6.09</b> ±4.47	<b>9.10</b> ±3.16

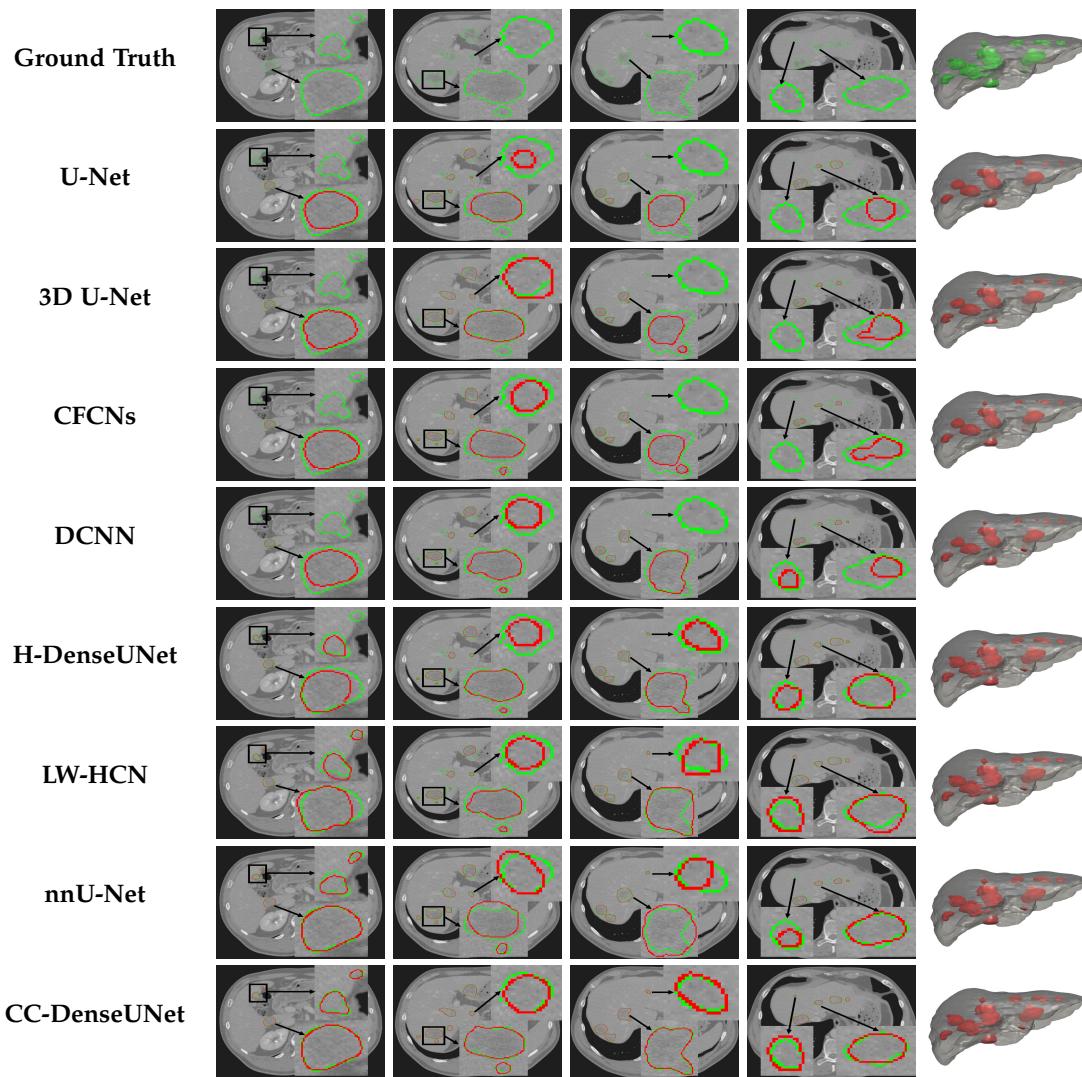


Fig. 5. Qualitative segmentation results by different methods on the 3DIRCADb dataset. The green outlines and objects denote the true liver tumors while the red ones denote the segmented liver tumors.

Another advantage of CC-DenseUNet is that we can utilize dense interconnections to improve the gradient flow and save time in searching for the optimal solution of a deep neural network. Table 3 and 4 show that the models with dense interconnections (DenseU-Net and CC-DenseUNet with/without the pre-trained model) achieved much performance improvement in liver tumor segmentation on the custom LiTS testset and the 3DIRCADb dataset, compared with the model without dense interconnections (U-Net). The last advantage of CC-DenseUNet is that we can employ the CCA module to effectively and efficiently obtain full-image contextual information. According to Table 3, compared with the experimental results generated by the DenseU-Net with the pre-trained model, the CC-DenseUNet with the pre-trained model achieved 3.9% DSC improvements in liver tumor segmentation results on the custom LiTS testset. Table 4 presents that the CC-DenseUNet with the pre-trained model achieved 4.3% DSC improvements in liver tumor segmentation results on the 3DIRCADb dataset, compared with the experimental results generated by the DenseU-Net with the pre-trained model.

Fig. 6 indicates that the CC-DenseUNet with the pre-

trained model outperforms the other four ablation models in liver tumor segmentation. Fig. 7 and 8 show some of the qualitative segmentation results achieved by five ablation models on the custom LiTS testset and the 3DIRCADb dataset, respectively.

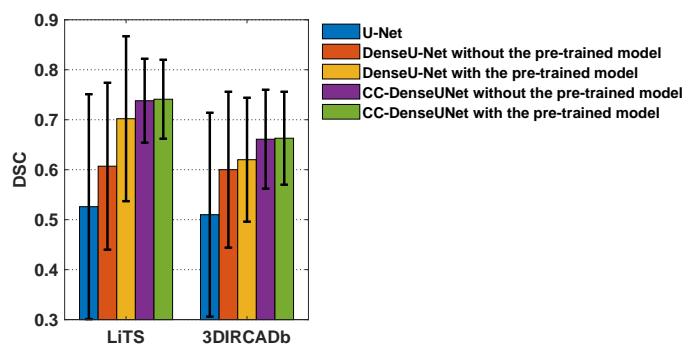


Fig. 6. Bar chart with error bars of DSC scores for different models on two datasets.

TABLE 4

Quantitative results for liver tumor segmentation by ablation models on the 3DIRCADb dataset. ↑ means the value is the larger the better and ↓ is the opposite. Best values are in bold.

Model	Performance Metrics				
	DSC ↑	VOE (%) ↓	RAVD (%) ↓	ASSD (mm) ↓	RMSSSD (mm) ↓
U-Net	0.510±0.204	62.60±24.21	11.20±2.71	12.50±4.61	18.30±13.81
DenseU-Net without pre-trained model	0.600±0.156	58.90±21.71	11.70±2.40	4.52±3.91	15.80±10.71
DenseU-Net with pre-trained model	0.620±0.124	53.20±17.58	9.20±2.10	4.20±3.46	12.60±7.60
CC-DenseUNet without pre-trained model	0.661±0.099	41.37± <b>13.43</b>	<b>3.20±1.50</b>	2.46±2.26	<b>10.90±2.33</b>
CC-DenseUNet with pre-trained model	<b>0.663±0.093</b>	<b>34.30±13.82</b>	4.20±1.90	<b>2.23±1.49</b>	11.69± <b>1.50</b>

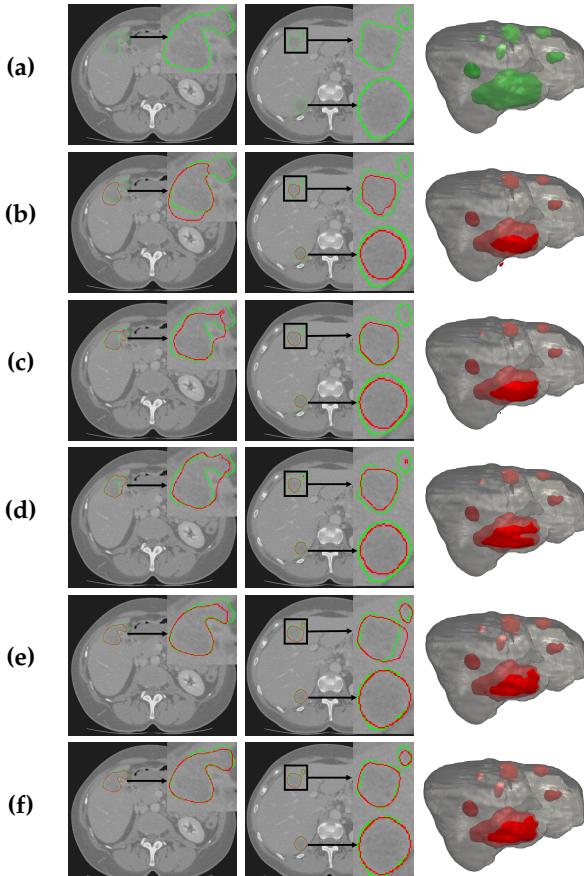


Fig. 7. Qualitative segmentation results by ablation experiments on the custom LiTS testset. The green outlines and objects denote the true liver tumors while the red ones denote the segmented liver tumors. (a) Ground truth; (b) U-Net; (c) DenseU-Net without the pre-trained model; (d) DenseU-Net with the pre-trained model; (e) CC-DenseUNet without the pre-trained model; (f) CC-DenseUNet with the pre-trained model.

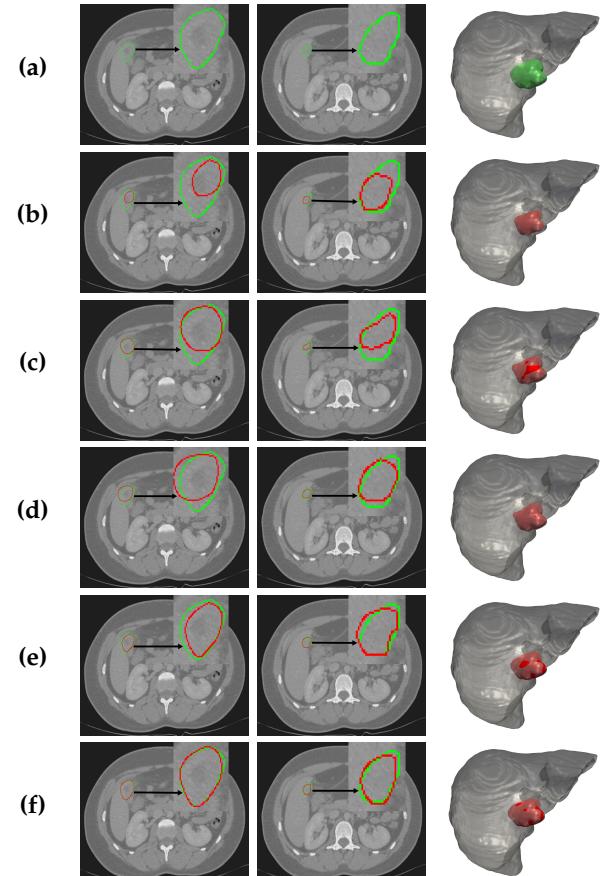


Fig. 8. Qualitative segmentation results by ablation experiments on the 3DIRCADb dataset. The green outlines and objects denote the true liver tumors while the red ones denote the segmented liver tumors. (a) Ground truth; (b) U-Net; (c) DenseU-Net without the pre-trained model; (d) DenseU-Net with the pre-trained model; (e) CC-DenseUNet without the pre-trained model; (f) CC-DenseUNet with the pre-trained model.

#### 4.5 Robustness Testing

To verify the robustness of the proposed method, we used the models trained on the LiTS dataset to test the performance of liver tumor segmentation on the clinical dataset, and the test results are presented in Table 5. Fig. 9 displays some of the qualitative segmentation results by the trained models. The results show that CC-DenseUNet outperforms the other seven methods, which demonstrated that the proposed method has good robustness for segmenting liver tumors in CT images.

## 5 CONCLUSION

In this paper, we proposed CC-DenseUNet for automatic liver tumor segmentation from CT images. This novel network, based on traditional U-Net architecture, integrates dense interconnections and the CCA module at the same time. Dense interconnections are used to improve gradient flow between encoder layers and save time in searching for the optimal solution of the network. The CCA module is employed to harvest the contextual information of all image pixels on criss-cross paths at the bottleneck stage. We

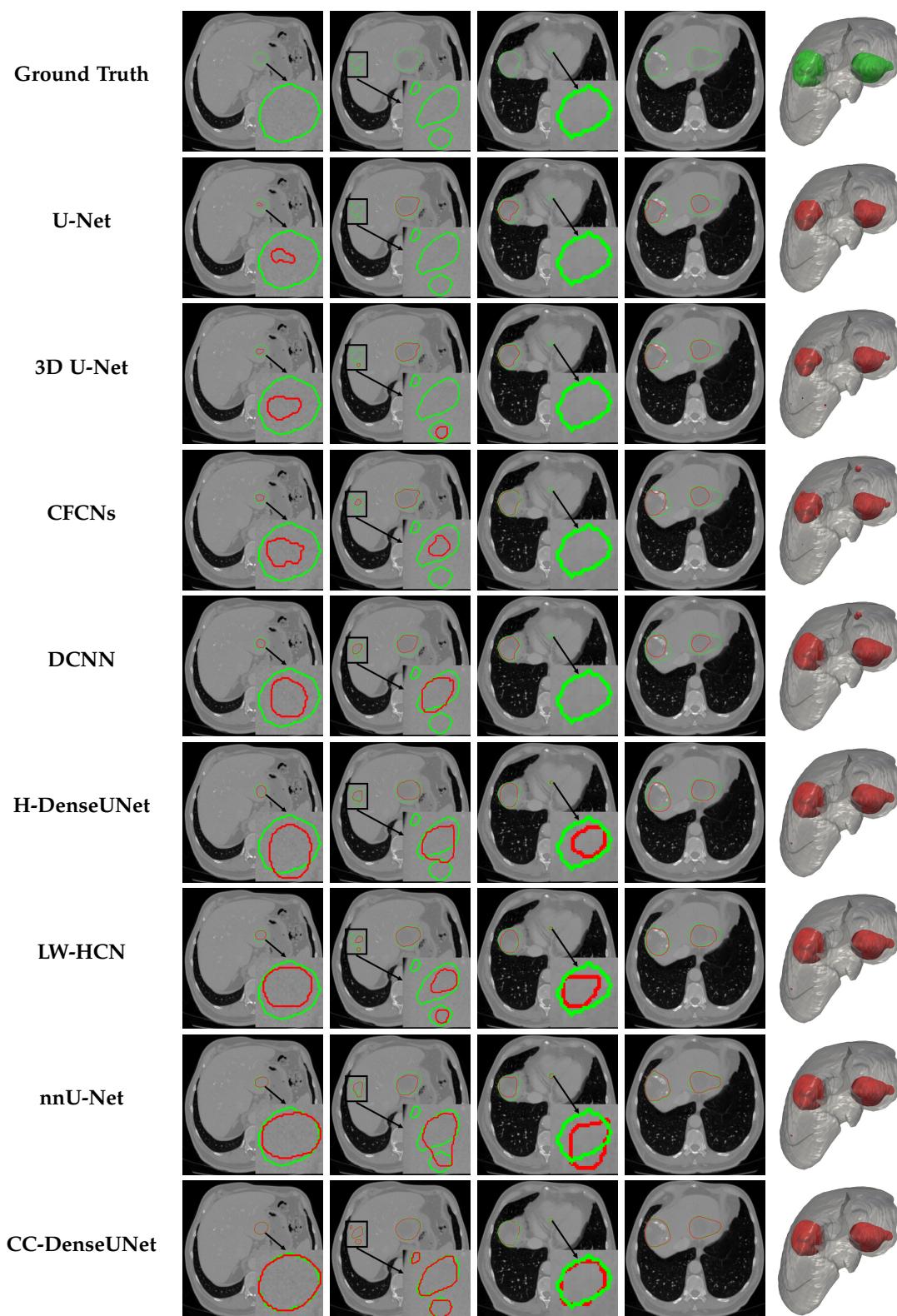


Fig. 9. Qualitative segmentation results on the clinical dataset. The green outlines and objects denote the true liver tumors while the red ones denote the segmented liver tumors.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Quantitative liver tumor segmentation results on the clinical dataset. ↑ means the value is the larger the better and ↓ is the opposite. Best values are in bold.

Method	Performance Metrics				
	DSC ↑	VOE (%) ↓	RAVD (%) ↓	ASSD (mm) ↓	RMSSD (mm) ↓
U-Net [28]	0.658±0.229	47.28±22.85	27.32±24.86	7.66±6.78	11.89±2.14
3D U-Net [29]	0.695±0.227	42.99±22.85	25.03±22.72	7.13±6.27	10.68±2.77
CFCNs [8]	0.698±0.228	41.51±23.15	24.51±22.19	6.73±6.23	10.16±2.79
DCNN [16]	0.708±0.228	41.25±23.18	22.86±20.22	6.58±6.25	9.65±3.74
H-DenseUNet [19]	0.727±0.226	39.03±22.88	21.73±5.35	5.27±3.05	8.48±1.56
LW-HCN [20]	0.735±0.225	37.97±22.85	21.88±5.42	<b>5.12±2.95</b>	<b>8.32±1.52</b>
nnU-Net [34]	0.736±0.226	37.86±22.89	20.11±3.58	5.33±3.19	8.86±1.15
CC-DenseUNet (Ours)	<b>0.745±0.225</b>	<b>36.76±22.80</b>	<b>18.65±4.70</b>	6.92±5.25	9.90±1.90

experimented with the proposed model by ablating dense interconnections or the CCA module, and compared the proposed method with the state-of-the-art methods. The experimental results on the LiTS and 3DIRCADb datasets demonstrated that CC-DenseUNet improved the liver tumor segmentation performance. In addition, we have confirmed the robustness of CC-DenseUNet on the clinical dataset. Therefore, CC-DenseUNet has potential in aiding clinicians with liver tumor segmentation models.

## ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (62171039, 62025104, 82172027, 61901031), Beijing Nova Program (Z201100006820004) from Beijing Municipal Science & Technology Commission.

## REFERENCES

- [1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a cancer journal for clinicians*, vol. 68, no. 6, p. 394–424, 2018.
- [2] D. Zhang, B. Chen, J. Chong, and S. Li, "Weakly-supervised teacher-student network for liver tumor segmentation from non-enhanced images," *Medical Image Analysis*, vol. 70, p. 102005, 2021.
- [3] J. Zhao, D. Li, X. Xiao, F. Accorsi, H. Marshall, T. Cossetto, D. Kim, D. McCarthy, C. Dawson, S. Knezevic *et al.*, "United adversarial learning for liver tumor segmentation and detection of multi-modality non-contrast mri," *Medical Image Analysis*, vol. 73, p. 102154, 2021.
- [4] J. Li, X. Ou, N. Shen, J. Sun, J. Ding, J. Zhang, J. Yao, and Z. Wang, "Study on strategy of ct image sequence segmentation for liver and tumor based on u-net and bi-convlstm," *Expert Systems with Applications*, vol. 180, p. 115008, 2021.
- [5] D. T. Kushnure and S. N. Talbar, "Hfru-net: High-level feature fusion and recalibration unet for automatic liver and tumor segmentation in ct images," *Computer Methods and Programs in Biomedicine*, vol. 213, p. 106501, 2022.
- [6] R. A. Khan, Y. Luo, and F.-X. Wu, "Rms-unet: Residual multi-scale unet for liver and lesion segmentation," *Artificial Intelligence in Medicine*, p. 102231, 2022.
- [7] S. Pang, A. Du, M. A. Orgun, Y. Wang, and Z. Yu, "Tumor attention networks: Better feature selection, better tumor segmentation," *Neural Networks*, vol. 140, p. 203–222, 2021.
- [8] P. F. Christ, M. E. A. Elshaer, F. Ettlinger, S. Tatavarty, M. Bickel, P. Bilic, M. Rempfler, M. Armbruster, F. Hofmann, and M. D'Anastasi, "Automatic liver and lesion segmentation in ct using cascaded fully convolutional neural networks and 3d conditional random fields," in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer, 2016, p. 415–423.
- [9] Y. Tang, Y. Tang, Y. Zhu, J. Xiao, and R. M. Summers, "E<sup>2</sup>net: An edge enhanced network for accurate liver and tumor segmentation on ct scans," in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer, 2020, p. 512–522.
- [10] L. Song, K. Geoffrey, and H. Kaijian, "Bottleneck feature supervised u-net for pixel-wise liver and tumor segmentation," *Expert Systems with Applications*, vol. 145, p. 113131, 2020.
- [11] H. Seo, C. Huang, M. Bassenne, R. Xiao, and L. Xing, "Modified u-net (mu-net) with incorporation of object-dependent high level features for improved liver and liver-tumor segmentation in ct images," *IEEE Transactions on Medical Imaging*, vol. 39, no. 5, p. 1316–1325, 2019.
- [12] Y. Zhang, J. Tian, C. Zhong, Y. Zhang, Z. Shi, and Z. He, "Darn: Deep attentive refinement network for liver tumor segmentation from 3d ct volume," in *the Twenty-Fifth International Conference on Pattern Recognition (ICPR)*. IEEE, 2021, p. 7796–7803.
- [13] F. Lyu, A. J. Ma, T. C.-F. Yip, G. L.-H. Wong, and P. C. Yuen, "Weakly supervised liver tumor segmentation using couinaud segment annotation," *IEEE Transactions on Medical Imaging*, 2021.
- [14] J. Chi, X. Han, C. Wu, H. Wang, and P. Ji, "X-net: Multi-branch unet-like network for liver and tumor segmentation from 3d abdominal ct scans," *Neurocomputing*, vol. 459, p. 81–96, 2021.
- [15] A. Aghamohammadi, R. Ranjbarzadeh, F. Naiemi, M. Mogharrebi, S. Dorosty, and M. Bendechache, "Tpcnn: Two-path convolutional neural network for tumor and liver segmentation in ct images using a novel encoding approach," *Expert Systems with Applications*, vol. 183, p. 115406, 2021.

- [16] X. Han, "Automatic liver lesion segmentation using a deep convolutional neural network method," *arXiv preprint arXiv:1704.07239*, 2017.
- [17] X. Wang, S. Han, Y. Chen, D. Gao, and N. Vasconcelos, "Volumetric attention for 3d medical image segmentation and detection," in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer, 2019, p. 175–184.
- [18] P. Lv, J. Wang, and H. Wang, "2.5d lightweight riu-net for automatic liver and tumor segmentation from ct," *Biomedical Signal Processing and Control*, vol. 75, p. 103567, 2022.
- [19] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and 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, p. 2663–2674, 2018.
- [20] J. Zhang, Y. Xie, P. Zhang, H. Chen, Y. Xia, and C. Shen, "Light-weight hybrid convolutional network for liver tumor segmentation," in *the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI)*, vol. 19, 2019, p. 4271–4277.
- [21] R. Dey and Y. Hong, "Hybrid cascaded neural network for liver lesion segmentation," in *the Seventeenth International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2020, p. 1173–1177.
- [22] L. Song, H. Wang, and Z. J. Wang, "Bridging the gap between 2d and 3d contexts in ct volume for liver and tumor segmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 9, p. 3450–3459, 2021.
- [23] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *International Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2017, p. 4700–4708.
- [24] Z. Huang, X. Wang, L. Huang, C. Huang, Y. Wei, and W. Liu, "Ccnet: Criss-cross attention for semantic segmentation," in *International Conference on Computer Vision (ICCV)*. IEEE, 2019, p. 603–612.
- [25] P. Bilic, P. F. Christ, E. Vorontsov, G. Chlebus, H. Chen, Q. Dou, C.-W. Fu, X. Han, P.-A. Heng, J. Hesser *et al.*, "The liver tumor segmentation benchmark (lits)," *arXiv preprint arXiv:1901.04056*, 2019.
- [26] L. Soler, A. Hostettler, V. Agnus, A. Charnoz, J. Fasquel, J. Moreau, A. Osswald, M. Bouhadjar, and J. Marescaux, "3d image reconstruction for comparison of algorithm database: A patient specific anatomical and medical image database," *IRCAD, Strasbourg, France, Tech. Rep.*, 2010.
- [27] Q. Li, H. Song, W. Zhang, J. Fan, D. Ai, Y. Lin, and J. Yang, "Cc-denseunet: Densely connected u-net with criss-cross attention for liver and tumor segmentation in ct volumes," in *International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2021, p. 966–971.
- [28] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer, 2015, p. 234–241.
- [29] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3d u-net: Learning dense volumetric segmentation from sparse annotation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer, 2016, p. 424–432.
- [30] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *International Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2016, p. 770–778.
- [31] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *International Conference on Computer Vision (ICCV)*. IEEE, 2017, p. 2961–2969.
- [32] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems*, vol. 32. Curran Associates, Inc., 2019, p. 8026–8037.
- [33] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *International Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2009, p. 248–255.
- [34] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnu-net: A self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, no. 2, p. 203–211, 2021.



**Qiang Li** received the M.S. degree in computer software and theory from Wenzhou University, China, in 2017. He is now pursuing the Ph.D. degree with the School of Computer Science and Technology, Beijing Institute of Technology, China. His research interests include medical image analysis, deep learning and computer vision.



**Hong Song** received the Ph.D. degree in computer science from Beijing Institute of Technology, China, in 2004. She is a professor with the School of Computer Science and Technology, Beijing Institute of Technology. Her research interests focus on medical image processing, augmented reality and computer vision.



**Fengbo Yang** received the B.S. degree in software engineering from Beijing Institute of Technology, China, in 2020. He is now pursuing the M.S. degree with the School of Computer Science and Technology, Beijing Institute of Technology. His research interests include computer vision and computational biology.



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
**Zenghui Wei** received the M.S. degree from Zhengzhou University, China, in 2009. He is now pursuing the Ph.D. degree with the School of Computer Science and Technology, Beijing Institute of Technology, China. His major interests include machine learning, medical image analysis and fine-grained image recognition.



**Xiaoling Yu** is a director with the Department of Interventional Ultrasound, Chinese PLA General Hospital, Beijing, China. She is good at color ultrasound and contrast-enhanced ultrasound diagnosis of abdominal and chest tumors, interventional ultrasound diagnosis and treatment, ultrasound-guided local therapy for multi-organ solid tumors, combined immunotherapy, combined embolization, and combined drug therapy.



**Jingfan Fan** received the Ph.D. degree in optical engineering from Beijing Institute of Technology, China, in 2016. He worked as post-doctoral researcher with University of North Carolina at Chapel Hill in US from 2017 to 2019. He is currently an associate professor with the School of Optics and Photonics, Beijing Institute of Technology. He focuses his research interests on medical image processing, computer vision and augmented reality.



**Danni Ai** received the Ph.D. degree from Ritsumeikan University, Japan, in 2011. She is currently an associate professor with the School of Optics and Photonics, Beijing Institute of Technology, China. Her research interests include medical image analysis, surgical navigation, virtual reality and augmented reality.



**Yucong Lin** received the Ph.D. degree in statistics from Tsinghua University, China, in 2021. He worked as visiting scholar with Harvard Medical School in US from 2020 to 2021. He is currently a post-doctoral researcher with the School of Medical Technology, Beijing Institute of Technology, China. His research interests include knowledge graph construction, natural language processing and artificial intelligence in medicine.



**Jian Yang** received the Ph.D. degree in optical engineering from Beijing Institute of Technology, China, in 2007. He is a professor with the School of Optics and Photonics, Beijing Institute of Technology. His research interests focus on medical image processing, augmented reality and computer vision.

# CC-DenseUNet: Densely Connected U-Net with Criss-Cross Attention for Liver and Tumor Segmentation in CT Volumes

Qiang Li<sup>1</sup>, Hong Song<sup>1,\*</sup>, Weiwei Zhang<sup>2</sup>, Jingfan Fan<sup>3</sup>, Danni Ai<sup>3</sup>, Yucong Lin<sup>3</sup>, Jian Yang<sup>3</sup>

<sup>1</sup>School of Computer Science & Technology, Beijing Institute of Technology, Beijing, China

<sup>2</sup>The Third Medical Center, Chinese PLA General Hospital, Beijing, China

<sup>3</sup>School of Optics and Electronics, Beijing Institute of Technology, Beijing, China

liqiang3235@foxmail.com, anniesun@bit.edu.cn, zhangweiweimedic@163.com,

fjf@bit.edu.cn, danni@bit.edu.cn, yuconglinhu@gmail.com, jyang@bit.edu.cn

**Abstract**—The automatic segmentation of liver and tumor is important for hepatic tumor surgery. In this paper, we propose a novel densely connected U-Net (CC-DenseUNet), which integrates criss-cross attention (CCA) module, to segment the liver and tumor in computed tomography (CT) volumes. The dense interconnections in CC-DenseUNet ensure the maximum information flow between encoder layers when extracting intra-slice features of liver and tumors. Moreover, the CCA module is used in CC-DenseUNet to efficiently capture only the necessary and meaningful non-local contextual information of CT images containing liver or tumors. We evaluated the proposed CC-DenseUNet on the Liver Tumor Segmentation Challenge and 3DIRCADb datasets. Experimental results show that our method outperformed the state-of-the-art methods in liver tumor segmentation and achieved a highly competitive performance in liver segmentation.

**Index Terms**—Criss-Cross Attention, CT, Dense Interconnections, Liver and Tumor Segmentation, U-Net

## I. INTRODUCTION

Liver cancer is one of the malignant diseases with high morbidity and mortality in the world [1]. Accurate liver and tumor measurements from computed tomography (CT) scans can help doctors in treatment planning. Delineating the liver and tumor tissues in each CT slice is laborious and time-consuming. Therefore, the demand for automatic liver and tumor segmentation in clinical practice is high.

However, automatic liver and tumor segmentation from CT volumes is a very challenging task. First, the contrast between the liver and other neighboring organs is low. Second, the liver tumor has various shape, size, location and numbers for each patient. Many segmentation methods [2–6] have been proposed to address the above problems.

In this paper, we proposed a novel densely connected U-Net with criss-cross attention (CCA) mechanism, called CC-DenseUNet, to segment liver and tumors in contrast-enhanced abdominal CT volumes. CC-DenseUNet borrows the idea of DenseNet [7], in which dense interconnections introduce direct connections between two layers with the same feature map size to facilitate a better feature propagation across the network. Moreover, inspired by CCNet [8], CC-DenseUNet employs a

novel CCA module to harvest the contextual information of all pixels on its criss-cross path. The main contributions of this work are as follows:

- We propose CC-DenseUNet, which jointly use dense interconnections and CCA for the effective and efficient segmentation of liver and tumors in CT volumes.
- We designed a 2.5D fully convolutional network (FCN) model to probe the spatial information along the third dimension and trained our model by transfer learning with pre-trained models, which is conducive to the fast convergence of the model.
- The proposed CC-DenseUNet model reaches the state-of-the-art segmentation performance on the LiTS and 3DIRCADb datasets.

## II. RELATED WORK

Convolutional neural networks (CNNs) [9–11] have achieved great success in medical image processing. Many researchers proposed various CNNs for liver and tumor segmentation. For example, Sun *et al.* [12] designed a multi-channel FCN to segment liver tumors from multiphase contrast-enhanced CT images. Seo *et al.* [13] proposed a modified U-Net that incorporates object-dependent high-level features for improved liver and tumor segmentation in CT images. Tang *et al.* [14] proposed an edge-enhanced network for accurate liver and tumor segmentation in CT scans.

Christ *et al.* [2, 3] proposed cascaded FCNs to automatically segment liver and tumors in CT abdomen images, and refined the segmentation results using dense 3D conditional random fields that account for both spatial coherence and appearance. Han [4] developed a 2.5D CNN model, which takes a stack of adjacent slices as input and produces a segmentation map that corresponds to the center slice. Li *et al.* [5] proposed a hybrid densely connected U-Net (H-DenseUNet), which is composed of a 2D DenseUNet and a 3D counterpart. The 2D DenseUNet is used to effectively extract features in the slice, while the 3D counterpart hierarchically aggregates volumetric contexts based on the auto-context algorithm. Zhang *et al.* [6] proposed a light-weight hybrid CNN to segment the liver and

1 tumors in CT volumes. They designed a depthwise and spatio  
 2 temporal separate factorization for 3D convolutions, which not  
 3 only reduces parameters dramatically but also improves the  
 4 performance of liver tumor segmentation.  
 5

### III. METHOD

8 Figure 1a shows the architecture of our proposed method  
 9 for liver and tumor segmentation. Our approach integrates the  
 10 CCA module [8] within the densely connected U-Net as an  
 11 extension of the bottleneck of U-Net [15] to efficiently capture  
 12 only the necessary and meaningful non-local contextual  
 13 information.

#### A. CCA Module

15 The CCA module [8] collects contextual information in  
 16 vertical and horizontal directions for each pixel. Input images  
 17 are passed through the network encoder to produce the feature  
 18 map  $M$  with reduced dimension. Then,  $M$  is fed into the CCA  
 19 module to generate a new feature map  $M'$ , which aggregates  
 20 contextual information together for each pixel in its criss-cross  
 21 path.

22 As shown in Figure 1c, the CCA module first applies two  
 23 convolutional layers with  $1 \times 1$  filters on  $M \in \mathbb{R}^{C \times H \times W}$   
 24 to generate two feature maps  $K$  and  $Q$ , where  $\{K, Q\} \in \mathbb{R}^{C' \times H \times W}$ .  $C'$  is the number of channels, which is less than  
 25  $C$ . A vector  $Q_u \in \mathbb{R}^{C'}$  can be obtained at each position  $u$  of  $Q$ .  
 26 A set  $\Omega_u \in \mathbb{R}^{(H+W-1) \times C'}$  can also be obtained by extracting  
 27 feature vectors in the same column or row as position  $u$  from  
 28  $K$ . Attention map  $A \in \mathbb{R}^{(H+W-1) \times H \times W}$  is calculated by  
 29 applying a softmax layer on an affinity operation.  
 30

31 The affinity operation is formulized as:

$$d_{i,u} = Q_u \Omega_{i,u}^T \quad (1)$$

32 where  $d_{i,u} \in D$  is the degree of correlation between features  
 33  $Q_u$  and  $\Omega_{i,u}$ ,  $\Omega_{i,u} \in \mathbb{R}^{C'}$  is the  $i$ -th element of  $\Omega_u$ ,  
 34  $i = [1, \dots, |\Omega_u|]$ , and  $D \in \mathbb{R}^{(H+W-1) \times H \times W}$ .

35 The third convolutional layer applied on  $M$  generates  $V \in \mathbb{R}^{C \times H \times W}$  for feature adaption. A feature vector  $V_u \in \mathbb{R}^C$  and  
 36 a set  $\Phi_u \in \mathbb{R}^{(H+W-1) \times H \times W}$  can be obtained at each position  
 37  $u$  in  $V$ . The set  $\Phi_u$  is a collection of feature vectors in  $V$ ,  
 38 which are in the same column or row as position  $u$ .

39 The contextual information is aggregated by

$$M'_u = \sum_{i=0}^{H+W-1} A_{i,u} \Phi_{i,u} + M_u \quad (2)$$

40 where  $M'_u$  is a feature vector in  $M' \in \mathbb{R}^{C \times H \times W}$  at position  
 41  $u$ ,  $\Phi_{i,u}$  is the  $i$ -th element of  $\Phi_u$ ,  $i = [0, \dots, H+W-1]$ ,  
 42 and  $A_{i,u}$  is a scalar value at channel  $i$  and position  $u$  in  $A$ .

#### B. CC-DenseUNet

43 The architecture of our model combines the concepts of  
 44 DenseNet [7], U-Net [15], and CCNet [8]. An overview  
 45 of the proposed CC-DenseUNet is shown in Figure 1. CC-  
 46 DenseUNet consists of a downsampling path, a bottleneck, and  
 47 an upsampling path. In the downsampling path, the structure  
 48 of DenseNet-121 [7], which is composed of repetitive densely  
 49

50 connected convolution blocks with different output dimensions,  
 51 is used to extract intra-slice features. One advantage  
 52 of the densely connected path is that it ensures the maximum  
 53 information flow between layers. The local feature map  $M$   
 54 outputted from the last block within the downsampling path  
 55 of the CC-DenseUNet serves as the input of the CCA module  
 56 [8]. This attention module is inserted in the bottleneck of CC-  
 57 DenseUNet to gather contextual information in the criss-cross  
 58 path of each pixel. Specifically, CC-DenseUNet is a 2.5D  
 59 FCN model. Three adjacent slices cropped from volumetric  
 60 images are fed into the model, and the 3D segmentation  
 61 volume is generated by simply stacking the segmentation maps  
 62 corresponding to the center slices.

#### C. Loss Function

63 Weighted cross-entropy function was employed as the loss  
 64 function of CC-DenseUNet, which is defined as:

$$f(g, p) = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^3 w_i^c g_i^c \log p_i^c \quad (3)$$

65 where  $w_i^c$  denotes the weight,  $g_i^c$  is the ground truth label for voxel  $i$ , and  $p_i^c$  indicates the probability of voxel  $i$  belongs to class  $c$  (background, liver or tumor).

## IV. EXPERIMENTS AND RESULTS

#### A. Datasets and Preprocessing

66 Our method was evaluated on the datasets of MICCAI  
 67 2017 LiTS Challenge [16] and 3DIRCADb [17]. The LiTS  
 68 dataset contains 131 and 70 contrast-enhanced abdominal CT  
 69 scans for training and test, respectively. However, the ground  
 70 truths of the LiTS test dataset are not publicly available. We  
 71 randomly selected 31 CT scans from the LiTS training dataset  
 72 called the custom LiTS test dataset to validate our model. The  
 73 remaining 100 CT scans of the LiTS training dataset were  
 74 used to train our model. The 3DIRCADb dataset contains 20  
 75 venous phase-enhanced CT volumes. We trained and tested  
 76 our model using 15 CT volumes with liver tumors through  
 77 twofold cross validation.

78 For the LiTS dataset, the Hounsfield unit (HU) values of all slices are windowed in the range  $[-200, 200]$  to exclude irrelevant organs and objects. For the 3DIRCADb dataset, the image intensity values of all scans were truncated to the range of  $[-100, 400]$  HU to ignore irrelevant details.

#### B. Evaluation Metrics

81 Five metrics (Dice similarity coefficient (DSC), volumetric  
 82 overlap error (VOE), relative absolute volume difference  
 83 (RAVD), average symmetric surface distance (ASD), root  
 84 mean square of symmetric surface distance (RMSD)) are used  
 85 to measure the accuracy of segmentation results, which are  
 86 described in [18].

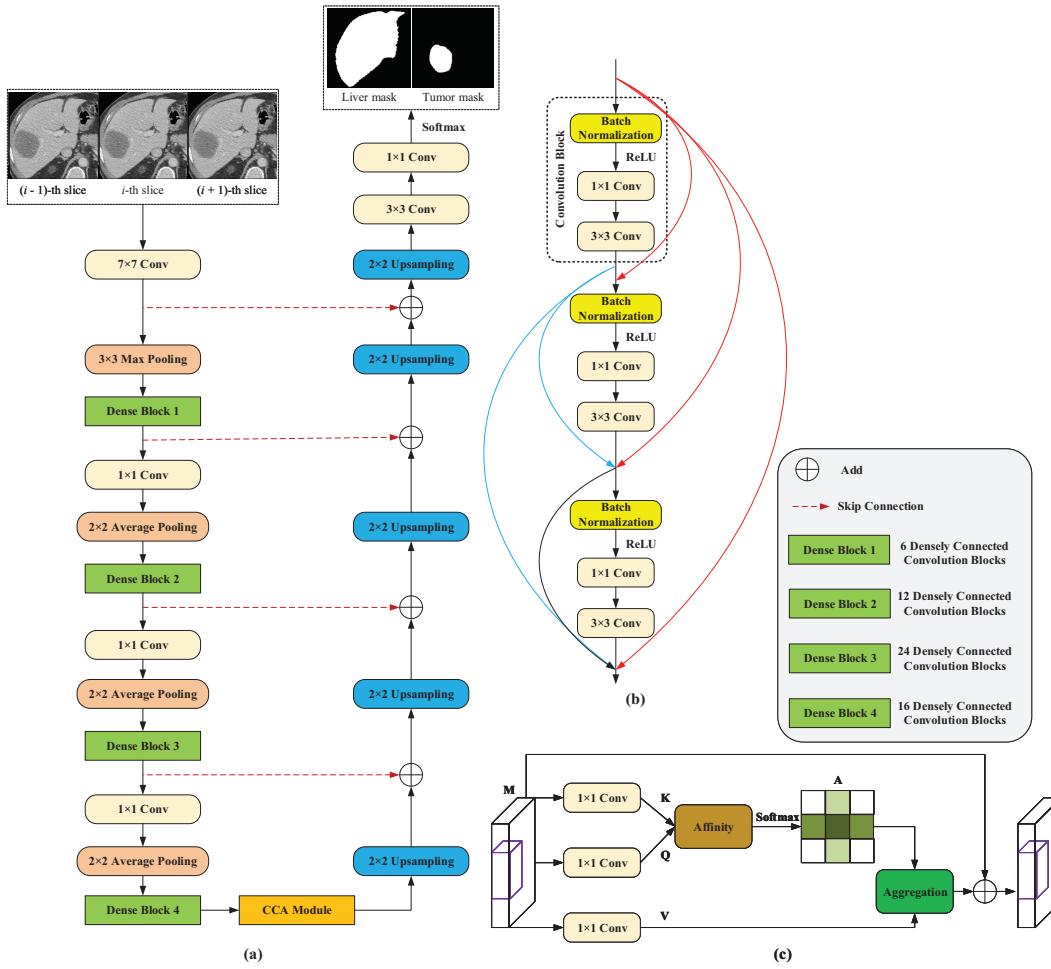


Fig. 1. CC-DenseUNet. (a) Network architecture of CC-DenseUNet; (b) Schematic of dense block structure; (c) Details of CCA module [8]

TABLE I  
COMPARISON OF SEGMENTATION RESULTS ON THE CUSTOM LiTS TEST DATASET

Method	Liver					Tumor				
	DSC	VOE (%)	RAVD (%)	ASD (mm)	RMSD (mm)	DSC	VOE (%)	RAVD (%)	ASD (mm)	RMSD (mm)
2D U-Net [15]	0.939	11.1	3.22	5.79	17.3	0.562	26.8	3.54	12.3	29.4
3D U-Net [19]	0.941	11.2	1.42	2.61	16.4	0.55	25.1	1.23	10.3	22.4
Cascaded U-Net [2]	0.943	10.7	1.92	2.48	8.3	0.623	19.7	1.09	9.2	18.3
DCNN [4]	0.95	10.0	2.36	5.23	8.6	0.67	45.0	4.0	8.3	17.2
H-DenseUNet [5]	0.961	<b>3.31</b>	<b>0.32</b>	<b>1.16</b>	9.17	0.722	24.5	<b>0.53</b>	<b>6.66</b>	<b>4.74</b>
CC-DenseUNet (Ours)	<b>0.962</b>	8.98	1.58	1.37	<b>7.3</b>	<b>0.741</b>	<b>16.9</b>	0.62	7.7	9.4

TABLE II  
COMPARISON OF SEGMENTATION RESULTS ON THE 3DIRCADB DATASET

Method	Liver					Tumor				
	DSC	VOE (%)	RAVD (%)	ASD (mm)	RMSD (mm)	DSC	VOE (%)	RAVD (%)	ASD (mm)	RMSD (mm)
2D U-Net [15]	0.729	39.0	87.0	19.4	15.7	0.51	62.6	1.11	12.5	18.3
3D U-Net [19]	0.923	14.2	0.05	4.33	8.35	0.543	50.2	0.38	11.1	16.7
Cascaded U-Net [2]	0.931	12.8	3.3	<b>2.3</b>	10.6	0.612	40.4	0.43	8.4	12.8
DCNN [4]	0.938	11.7	0.03	3.91	8.11	0.6	56.5	0.41	6.36	11.7
H-DenseUNet [5]	<b>0.947</b>	10.0	<b>0.01</b>	4.06	9.63	0.65	49.7	0.33	5.29	11.1
CC-DenseUNet (Ours)	0.943	<b>8.46</b>	0.02	3.45	<b>6.19</b>	<b>0.663</b>	<b>34.3</b>	<b>0.32</b>	<b>4.2</b>	<b>10.9</b>

### 1 C. Implementation Details

2 Our model was implemented using *Pytorch* package [20].  
 3 We used stochastic gradient descent with a momentum of 0.9  
 4 and a weight decay of 0.0001. A poly learning rate policy was  
 5 employed in which the learning rate decayed according to the  
 6 equation:  
 7

$$8 \quad lr = lr * \left(1 - \frac{iter}{max\_iter}\right)^{0.9} \quad (4)$$

9 where *iter* denotes current iteration, *max\_iter* denotes maximum iterations, and the initial learning rate was 0.01.

10 For data augmentation, we adopted random elastic deformation,  
 11 rotation and scaling between 0.8 and 1.2 for the training  
 12 data to teach the network the desired invariance properties.  
 13

14 In the training phase, the pre-trained DenseNet-121 [7] was  
 15 used to initialize our model. In the test phase, the original  
 16 size of the image slice was adopted because liver tumors are  
 17 extremely small in some CT volumes. We trained and tested  
 18 all models using one NVIDIA GeForce 1080Ti GPU with 11  
 19 GB memory.  
 20

### 21 D. Comparison with Other Methods

22 We compared our method with five different methods on  
 23 the custom LiTS test dataset and the 3DIRCADb dataset to  
 24 validate the effectiveness and robustness of our method. Table  
 25 I shows that our method outperformed 2D U-Net [15] and 3D  
 26 U-Net [19] on the custom LiTS test dataset with 2.3% and  
 27 2.1% DSC improvements in liver segmentation, respectively.  
 28 Table II displays that our method outperformed Cascaded U-  
 29 Net [2] and DCNN [4] on the 3DIRCADb dataset with 5.1%  
 30 and 6.3% DSC improvements in liver tumor segmentation,  
 31 respectively. Moreover, Tables I and II show that our method  
 32 achieved a better performance than H-DenseUNet [5] in terms  
 33 of liver tumor segmentation accuracy with 2.1% and 1.3%  
 34 DSC improvements on the custom LiTS test dataset and the  
 35 3DIRCADb dataset, respectively.  
 36

37 We also use other metrics, including VOE, RAVD, ASD,  
 38 and RMSD, to evaluate the segmentation performance of our  
 39 method and obtain a comprehensive comparison with other  
 40 liver and tumor segmentation methods. The best metric values  
 41 are shown in bold in Tables I and II.

42 Figures 2 and 3 show some of the segmentation results  
 43 achieved by six different methods on the custom LiTS test  
 44 dataset and the 3DIRCADb dataset, respectively. Figure 2  
 45 shows that CC-DenseUNet outperformed the five other meth-  
 46 ods in liver and tumor segmentation. Moreover, Figure 3  
 47 presents that CC-DenseUNet outperformed the remaining liver  
 48 segmentation methods, except H-DenseUNet [5], and also  
 49 outperformed the five other liver tumor segmentation methods.  
 50 The experimental results validated the superiority of our  
 51 proposed method in comparison with other methods.  
 52

### 53 E. Ablation Analysis of CC-DenseUNet

54 One advantage of CC-DenseUNet is that we can utilize  
 55 dense connectivity to improve the gradient flow and save time  
 56 in searching for the optimal solution of a deep neural network.  
 57

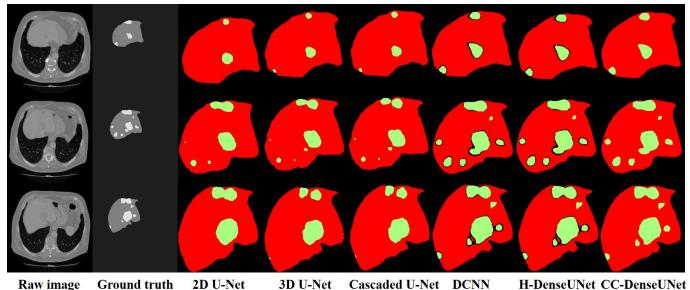


Fig. 2. Examples of segmentation results by different methods on the custom LiTS test dataset. The gray regions denote the true liver while the white ones denote the true tumors. The red regions denote the segmented liver while the green ones denote the segmented tumors.

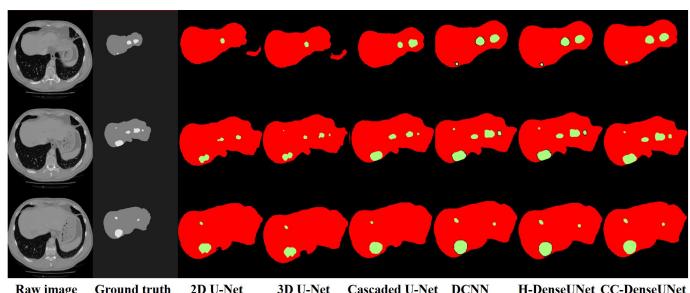


Fig. 3. Examples of segmentation results by different methods on the 3DIRCADb dataset. The gray regions denote the true liver while the white ones denote the true tumors. The red regions denote the segmented liver while the green ones denote the segmented tumors.

A comparison of Tables III and IV shows that compared with the results generated by 2D U-Net [15], DenseUNet without pre-trained model achieved 8.1% and 9% improvements in the DSC of the liver tumor segmentation results on the custom LiTS test dataset and the 3DIRCADb dataset, respectively.

Another advantage of CC-DenseUNet is that we can train the network by transfer learning with the pre-trained model, which can help the network converge faster and achieve a lower loss value. The liver and tumor segmentation results in Tables III and IV demonstrate that the pre-trained model can help the network achieve better a performance consistently.

Moreover, the last advantage of CC-DenseUNet is that we can employ CCA to effectively and efficiently obtain full-image contextual information. According to Table III, CC-DenseUNet achieved 0.4% and 3.9% DSC improvements in liver and tumor segmentation results, respectively, compared with the experimental results generated by the DenseUNet with pre-trained model on the custom LiTS test dataset. Table IV shows that CC-DenseUNet achieved 0.2% and 4.3% DSC improvements in liver and tumor segmentation results, respectively, compared with the experimental results generated by the DenseUNet with pre-trained model on the 3DIRCADb dataset.

Figures 4 and 5 show some of the segmentation results achieved by four ablation networks on the custom LiTS test dataset and the 3DIRCADb dataset, respectively. The results

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
TABLE III  
SEGMENTATION RESULTS BY ABLATION EXPERIMENTS ON THE CUSTOM LiTS TEST DATASET

Model	Liver					Tumor				
	DSC	VOE (%)	RAVD (%)	ASD (mm)	RMSD (mm)	DSC	VOE (%)	RAVD (%)	ASD (mm)	RMSD (mm)
2D U-Net [15]	0.939	11.1	3.22	5.79	17.3	0.526	26.8	3.54	12.3	29.4
DenseUNet without pre-trained model	0.947	8.29	1.1	2.35	15.3	0.607	17.7	1.0	9.7	17.5
DenseUNet with pre-trained model	0.958	<b>7.34</b>	<b>0.86</b>	2.37	9.6	0.702	<b>13.6</b>	0.66	10.5	<b>9.1</b>
CC-DenseUNet	<b>0.962</b>	8.98	1.58	<b>1.37</b>	<b>7.3</b>	<b>0.741</b>	16.9	<b>0.62</b>	<b>7.7</b>	9.4

10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
TABLE IV  
SEGMENTATION RESULTS BY ABLATION EXPERIMENTS ON THE 3DIRCADB DATASET

Model	Liver					Tumor				
	DSC	VOE (%)	RAVD (%)	ASD (mm)	RMSD (mm)	DSC	VOE (%)	RAVD (%)	ASD (mm)	RMSD (mm)
2D U-Net [15]	0.729	39.0	87.0	19.4	15.7	0.51	62.6	1.11	12.5	18.3
DenseUNet without pre-trained model	0.939	8.53	1.3	6.61	11.6	0.6	58.9	1.17	2.46	15.8
DenseUNet with pre-trained model	0.941	<b>7.44</b>	0.67	5.37	<b>5.5</b>	0.62	53.2	0.92	<b>2.23</b>	12.6
CC-DenseUNet	<b>0.943</b>	8.46	<b>0.02</b>	<b>3.45</b>	6.19	<b>0.663</b>	<b>34.3</b>	<b>0.32</b>	4.2	<b>10.9</b>

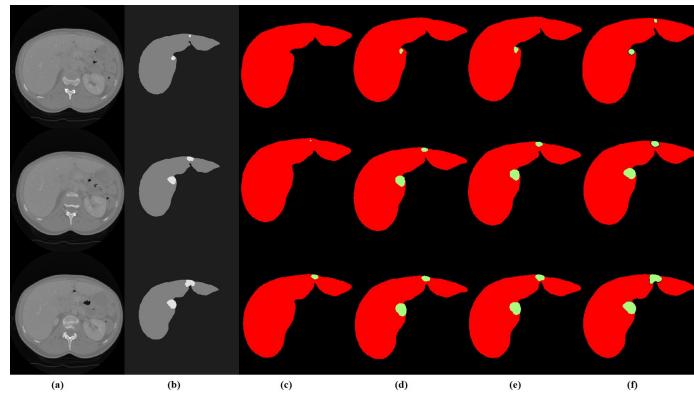


Fig. 4. Examples of segmentation results by ablation experiments on the custom LiTS test dataset. The gray regions denote the true liver while the white ones denote the true tumors. The red regions denote the segmented liver while the green ones denote the segmented tumors. (a) Raw image; (b) Ground truth; (c) 2D U-Net [15]; (d) DenseUNet without pre-trained model; (e) DenseUNet with pre-trained model; (f) CC-DenseUNet.

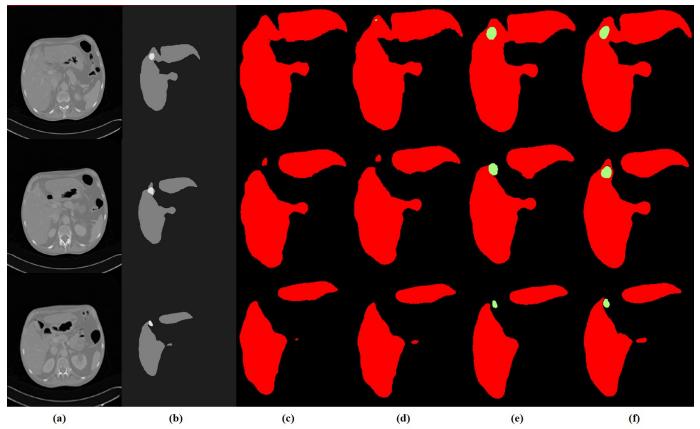


Fig. 5. Examples of segmentation results by ablation experiments on the 3DIRCADb dataset. The gray regions denote the true liver while the white ones denote the true tumors. The red regions denote the segmented liver while the green ones denote the segmented tumors. (a) Raw image; (b) Ground truth; (c) 2D U-Net [15]; (d) DenseUNet without pre-trained model; (e) DenseUNet with pre-trained model; (f) CC-DenseUNet.

indicate that CC-DenseUNet outperformed the three other ablation networks.

## V. CONCLUSION

In this paper, we proposed a novel FCN based on traditional U-Net architecture for automatic liver and tumor segmentation from CT volumes. The new network, called CC-DenseUNet, utilizes dense interconnections and CCA module. CC-DenseUNet was evaluated on the custom LiTS test dataset and the 3DIRCADb dataset. We can conclude from the evaluation results that CC-DenseUNet performs better than 2D U-Net [15], DenseUNet with or without pre-trained model, and four other networks as listed in Table I. Our novel model has potential in aiding clinicians with state-of-the-art liver and tumor segmentation models.

## ACKNOWLEDGMENT

This work was supported by National Natural Science Foundation of China (61771056).

## REFERENCES

- [1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, “Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries,” *CA: a cancer journal for clinicians*, vol. 68, no. 6, pp. 394–424, 2018.
- [2] P. F. Christ, M. E. A. Elshaer, F. Ettlinger, S. Tatavarty, M. Bickel, P. Bilic, M. Rempfler, M. Armbruster, F. Hofmann, and M. D’Anastasi, “Automatic liver and lesion segmentation in ct using cascaded fully convolutional neural networks and 3d conditional random fields,” in

- International Conference on Medical Image Computing and Computer-Assisted Intervention.* Springer, 2016, Conference Proceedings, pp. 415–423.

[3] P. F. Christ, F. Ettlinger, F. Grün, M. E. A. Elshaera, J. Lipkova, S. Schlecht, F. Ahmaddy, S. Tatavarty, M. Bickel, P. Bilic *et al.*, “Automatic liver and tumor segmentation of ct and mri volumes using cascaded fully convolutional neural networks,” *arXiv preprint arXiv:1702.05970*, 2017.

[4] X. Han, “Automatic liver lesion segmentation using a deep convolutional neural network method,” *arXiv preprint arXiv:1704.07239*, 2017.

[5] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and 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.

[6] J. Zhang, Y. Xie, P. Zhang, H. Chen, Y. Xia, and C. Shen, “Light-weight hybrid convolutional network for liver tumor segmentation,” in *Twenty-Eighth International Joint Conference on Artificial Intelligence IJCAI-19*, vol. 19, 2019, Conference Proceedings, pp. 4271–4277.

[7] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, Conference Proceedings, pp. 4700–4708.

[8] Z. Huang, X. Wang, L. Huang, C. Huang, Y. Wei, and W. Liu, “Cnnet: Criss-cross attention for semantic segmentation,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, Conference Proceedings, pp. 603–612.

[9] H. Song, L. Chen, Y. Cui, Q. Li, Q. Wang, J. Fan, J. Yang, and L. Zhang, “Denoising of mr and ct images using cascaded multi-supervision convolutional neural networks with progressive training,” *Neurocomputing*, 2021.

[10] L. Chen, H. Song, C. Wang, Y. Cui, J. Yang, X. Hu, and L. Zhang, “Liver tumor segmentation in ct volumes using an adversarial densely connected network,” *BMC bioinformatics*, vol. 20, no. 16, pp. 1–13, 2019.

[11] C. Zhang, Q. Hua, Y. Chu, and P. Wang, “Liver tumor segmentation using 2.5 d uv-net with multi-scale convolution,” *Computers in Biology and Medicine*, vol. 133, p. 104424, 2021.

[12] C. Sun, S. Guo, H. Zhang, J. Li, M. Chen, S. Ma, L. Jin, X. Liu, X. Li, and X. Qian, “Automatic segmentation of liver tumors from multiphase contrast-enhanced ct images based on fcns,” *Artificial Intelligence in Medicine*, vol. 83, pp. 58–66, 2017.

[13] H. Seo, C. Huang, M. Bassenne, R. Xiao, and L. Xing, “Modified u-net (mu-net) with incorporation of object-dependent high level features for improved liver and liver-tumor segmentation in ct images,” *IEEE transactions on medical imaging*, vol. 39, no. 5, pp. 1316–1325, 2019.

[14] Y. Tang, Y. Tang, Y. Zhu, J. Xiao, and R. M. Summers, “E<sup>2</sup>net: An edge enhanced network for accurate liver and tumor segmentation on ct scans,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2020, Conference Proceedings, pp. 512–522.

[15] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, Conference Proceedings, pp. 234–241.

[16] P. Bilic, P. F. Christ, E. Vorontsov, G. Chlebus, H. Chen, Q. Dou, C.-W. Fu, X. Han, P.-A. Heng, J. Hesser *et al.*, “The liver tumor segmentation benchmark (lits),” *arXiv preprint arXiv:1901.04056*, 2019.

[17] “3dircadb — ircad france,” <https://www.ircad.fr/research/3dircadb/>, accessed May, 2021.

[18] V. Yeghiazaryan and I. Voiculescu, “An overview of current evaluation methods used in medical image segmentation,” *Department of Computer Science, University of Oxford*, 2015.

[19] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, “3d u-net: learning dense volumetric segmentation from sparse annotation,” in *International conference on medical image computing and computer-assisted intervention*. Springer, 2016, Conference Proceedings, pp. 424–432.

[20] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, and L. Antiga, “Pytorch: An imperative style, high-performance deep learning library,” *Advances in neural information processing systems*, vol. 32, pp. 8026–8037, 2019.

## Summary of Changes

1. We changed the title “CC-DenseUNet: Densely Connected U-Net with Criss-Cross  
2 Attention for Liver and Tumor Segmentation in CT Volumes” to “Densely  
3 Connected U-Net with Criss-Cross Attention for Automatic Liver Tumor  
4 Segmentation in CT Images”.
5. We updated the sections of “Abstract” and “Introduction”.
6. We rewrote the section “Related Works”.
7. We merged the subsections “CCA Module” and “CC-DenseUNet” of the previous  
8 paper to one subsection “CC-DenseUNet”. And we redrew the figure of our  
9 proposed method.
10. We removed the contents of dataset preprocessing in the subsection “Datasets and  
11 Preprocessing” and renamed it as “Datasets”. And we also updated the renamed  
12 subsection.
13. We introduced the segmentation evaluation metrics in details.
14. We expanded the experiments by adding the segmentation results of LW-HCN and  
15 nnU-Net in the subsection “Comparison with Other Methods”.
16. We renamed the section “Experiments and Results” as “Experimental Results and  
17 Discussion” and added a new subsection “Robustness Testing” in the renamed  
18 section.
19. We replaced all the figures of qualitative segmentation results in the previous paper.
20. We updated the sections of “Conclusion”, “Acknowledgments” and “References”.