Study on the Segmentation Method of Multi-Phase CT Liver Tumor Based on Dual Channel U-Nets

Changchang Xu 1, Dong Hu 1,2,3*, YaoDong Zhang 4, and Yuwei Pang 3

1 Jiangsu Province’s Key Lab of Image Procession and Image Communications, Nanjing University of Posts and Telecommunications, Nanjing, 210003, China
2 Education Ministry’s Key Lab of Broadband Wireless Communication and Sensor Network Technology, Nanjing University of Posts and Telecommunications, Nanjing, 210003, China
3 Education Ministry’s Engineering Research Center of Ubiquitous Network and Heath Service, Nanjing University of Posts and Telecommunications, Nanjing, 210003, China
4 The first clinical medical college of Nanjing Medical University, Nanjing University of Posts and Telecommunications, Nanjing, 210003, China

*Email: hud@njupt.edu.cn

Abstract. As a common malignant tumor disease, hepatocellular carcinoma is the most common cancers in the world. The incidence of hepatocellular carcinoma in China is higher than that in the world. Therefore, it is very important for doctors to separate liver and tumor from CT images by means of computer-aided diagnosis and treatment. In this paper, a multiscale DC-CUNets network liver tumor segmentation method is proposed to enhance the fusion of multi-phase image features in CT, the scale of liver tumors, and the optimization of network training process. (1) A multistage CT liver tumor segmentation method based on two-channel cascaded U-Nets (DC-CUNets) is proposed. The liver was segmented using the first-order U-Net, and then the segmented area of interest of the liver was input into the second-order U-Net network to segment liver tumors. We designed two-channel U-Nets to learn the image characteristics of CT images in arterial and venous phases respectively, and to achieve two-channel feature fusion through feature cascade to improve the overall accuracy of liver tumor segmentation. (2) A multistage CT liver tumor segmentation method based on multiscale DC-CUNets was proposed. For the scale problem of liver tumors, we designed a two-layer multiscale void convolution module to obtain image features at different scales for large, medium and small tumors, and fuse the multiscale features at the output of the module. We have replaced the convolution layer of the fourth module in the second-order two-channel liver tumor segmentation U-Nets by the two-layer multiscale cavity convolution module to implement multiscale DC-CUNets.

1. Introduction
The liver is responsible for the metabolism and detoxification of the human body. It also plays a role in protein synthesis and other functions. It is an important organ to maintain human life, and is also the largest substantial tissue in the body. Radiologists often use enhanced CT to better observe liver details. Enhanced CT [1] improve the contrast between different tissues, especially between liver lesions and normal tissues, compared with unenhanced CT, and helps radiologists to observe tumors...
more clearly and make more accurate disease diagnosis accordingly. In recent years, with the expansion of data size and the enhancement of computing power, Convolutional Neural NetWork (CNN) has achieved remarkable results in many target recognition problems in the field of computer vision.

Fully Convolutional Neural NetWork (Fully Convolutional Neural NetWork) FCN [2]) has evolved from CNN removing the full connection layer to deconvolution layer, and has achieved many excellent segmentation results in medical image segmentation tasks. For example, Ben-Cohen et al. proposed FCN-8s based on VGG-16 for liver segmentation and lesion detection [3]. This model can detect the location information of liver lesions better and lay a foundation for subsequent lesion segmentation. Christ et al. proposed the use of cascaded FCNs (CFCNs [4]) to automate the segmentation of liver and liver tumors, and use a 3D Conditional Random Field (CRF [5]) to refine the segmentation results. CFCNs work by using a first-order FCN to train the liver segmentation network, and then input the results into a second-order FCN as an area of interest to train the network to segment liver tumors. This cascade segmentation method effectively reduces the probability of false-positive segmentation and improves the accuracy of liver tumor segmentation.

Li et al. also proposed H-DenseUNet [6], which combines 2D dense U-Net to extract information in slices and uses a 3D corresponding to aggregate contextual information in different volumes at the same time. This greatly improves the segmentation accuracy of large tumors, but has limited improvement for small tumors.

2. Related Works

2.1. FCN

Full convolution neural network FCN is a more suitable model for medical image segmentation. It is based on the network structure of CNN. On the one hand, it replaces the last full connection layer with deconvolution layer to achieve any size of image input and maintain the same input and output. On the other hand, it proposes a skip structure that combines shallow and deep image features. Compared with the one-dimensional vector of CNN output, the output of FCN is a two-dimensional image, which can better preserve the two-dimensional spatial information of the image, and greatly improve the segmentation accuracy and speed. In FCN, it is proposed to link the shallow image features with the deep image features by Skip Connection to supplement the location and spatial relationship between the pixels in the shallow image features with the high-level image features so as to improve the overall performance of the network [7].

Figure 1 is a skip connection diagram in FCN. When the input image passes through a structure consisting of conv 1 and pool1, the image size is reduced to 1/2 of the original size, and so on, after a structure consisting of conv5 and pool5, the image resolution is reduced to 1/32 of the input image. Then, when 1/32 of the small feature map is sampled twice at a time, the image resolution can be restored to 1/16 of the original image. It is not difficult to find that the resolution is the same as that of the feature map output after conv4 and pool4, Skip linking refers to iterating forward from conv5, supplementing the output of conv5 with the feature map from conv4 and pool4, and then supplementing the feature map from conv3 and pool3 to restore the image resolution. By skipping connections, the deep semantic information of the network can be fused with the shallow fine-grained surface information, which provides effective performance optimization.
2.2. U-net

U-Net [8] is a kind of FCN, its overall network structure is shown in Figure 1. The left half is the shrinking path used to extract the features, and the right half is the expanding path used to recover the size of the image by up-sampling, which is generally "U". This structure is also known as the Encoder-Decoder structure. The shrink path of U-Net repeats four times a module consisting of two 3*3 convolution layers and a maximum pooled layer with a convolution kernel of 2*2. After each module, the image size is reduced to half of the original size, the number of characteristic channels is doubled, and the ReLu [9] function is used after each convolution layer to improve the non-linear expression of the network. U-Net inherits the idea of FCN and uses deconvolution to achieve up-sampling. As a result, U-Net's extended path continuously uses two 3*3 convolution layers and a 2*2 deconvolution layer module to extract deep features and restore image size. Finally, a 1*1 convolution layer is used to map the feature layer to the output layer to output the final probability image.

It should be noted that although U-Net inherits the idea of FCN skip joining, it combines shallow features with deep features by summing corresponding pixels, while U-Net splits feature channels in this process. At the same time, compared with the structure of FCN, U-Net achieves a fully symmetric network structure by increasing the convolution layer of the expansion path and increasing the network depth.

![Figure 1. FCN Network Structure Diagram [3].](image)

![Figure 2. U-Net Network Architecture [8].](image)
3. Proposed Works

It should be noted in section 3.1, a liver tumor segmentation method based on two-channel cascade U-Nets is proposed to make full use of the different image features contained in the enhanced CT images during each period to improve the whole segmentation accuracy of liver tumors. A multi-stage CT liver tumor segmentation method based on multi-scale DC-CUNets is proposed in section 3.2, which combines the feature that void convolution can enhance the pixel field without affecting the image resolution.

3.1. Dual Channel Cascaded U-Nets

(1) The first liver was divided into U-Net. In the first U-Net for liver segmentation, only venous phase (PV) CT was used as the input to segment the liver region for one reason: For the image data of different stages of enhanced CT, the number of CT sections at different stages was the same as that at the same slice thickness, and the hepatic parenchyma was strong in vena cava phase CT. Significantly, the liver contour is clearer than other stages, and using only intravenous CT image data can further speed up network training and improve network convergence rate. However, there is also a problem with liver segmentation using only venous phase CT images, that is, arterial and venous phases are taken at different times after injection of contrast media [10].

From the perspective of the movement of the living body itself, even if the slice thickness is the same, slight distortion and displacement may exist in the same slice corresponding to arterial and venous phases. Without image registration, slight distortion and displacement may occur in the same slice. Output masks that result in a first-order liver segmentation network do not accurately cover the liver area on arterial phase CT. To solve this problem, this paper uses non-rigid registration technology [11] to register arterial and venous phase images to improve the rationality and reliability of network design.

(2) Segmentation of U-Nets by second-grade dual-channel tumors. As shown in Figure 3, the second-stage dual-channel U-Nets are used to segment liver tumors that exist in the liver region. As concluded in this paper, radiologists usually diagnose liver lesions using the imaging features of arterial and venous phases in multi-phase contrast-enhanced CT. Therefore, when designing a second-level tumor segmentation network, a dual-channel U-Nets structure is proposed, and the deep features learnt from the two channels are fused in a cascade of features to improve the accuracy of tumor segmentation. The final output is a probability map of liver tumors on CT images. Compared with the single-channel model, dual-channel U-Net can train CT images of arterial and venous phases separately. Each channel can obtain training parameters that are not shared.

That is, for arterial and venous phases, CT can learn different deep image characteristics according to their own image characteristics. It is worth noting that enhanced CT usually consists of three phases, but considering the small density difference between the liver parenchyma and the liver tumor in the balanced phase, the segmentation method proposed in this paper does not include all three phases into the segmentation network, but selects the arterial and venous phase CT images for liver tumor segmentation. The input of the network is a predictive probability map of the output of a trained U-Net liver segment on an abdominal CT image, where the value in each pixel represents the probability that the pixel belongs to the liver. By setting a threshold, the output image can be transformed into a binary liver mask and fed into a second-level tumor segmentation network as input. Before segmentation, arterial and venous CT masks the liver area on the original image, shielding other unrelated organs, and obtaining the liver area of interest in arterial and venous phases.
3.2. Multiscale DC-CUNets Network Design

The implementation of the multiscale void convolution module is to design the void convolution with different expansion rates based on the analysis of the scales in liver tumors. Therefore, the analysis of the scale of liver tumors is the key to the implementation of this module. We performed a scale-based statistical analysis of abdominal enhanced CT images obtained from hospitals and their corresponding radiodiagnostic reports and liver tumors contained in existing open datasets. All abdominal CT images used in this study were 512*512 pixels in size.

The results show that the liver tumors included in CT images can be divided into three categories according to the size of large, medium and small tumors. The size of small liver tumors is 40 pixels or less; the size of medium liver tumors is between 60 pixels and 90 pixels; and the size of large liver tumors is generally between 150 pixels and 200 pixels. Therefore, for three different scales of liver tumors, this chapter designs three convolutions with different expansion rates to preserve deeper image information in varying degrees. The multiscale void convolution module is implemented by designing void convolutions with different expansion rates based on scale analysis in liver tumors. The multiscale void convolution module is shown in Figure 4.

Figure 3. Dual Channel U-Net Network Structure.

Figure 4. Multiscale Void Convolution Module.
Considering that the convolution of voids with different resolutions can produce different size local sensory fields, we keep the size of the convolution core unchanged at 3×3, and set the expansion ratio of the first layer void convolution to 1-dilated, 2-dilated and 5-dilated, respectively, while the resolution of the second layer void convolution to 2-dilated, 3-dilated, and 6-dilated to accommodate different size types. For liver tumors, the multiscale void convolution expansion rate distribution table is shown in Table 1.

**Table 1.** Multiscale void convolution expansion rate distribution table.

| Tumour      | Original tumour size (pixel) | 1/8 tumour size (pixel) | Convolution rate of the first layer of holes (dilation rate) | Convolution rate of the second layer of holes (dilation rate) | Feel the wild (pixel) |
|-------------|-----------------------------|-------------------------|-------------------------------------------------------------|-------------------------------------------------------------|----------------------|
| Large volume| 150~200                     | 18~25                   | 1-dilated                                                  | 2-dilated                                                  | 23×23                |
| Medium volume| 60~90                     | 8~12                    | 2-dilated                                                  | 3-dilated                                                  | 11×11                |
| Small volume| Below 40                    | Below 5                 | 5-dilated                                                  | 6-dilated                                                  | 7×7                  |

3.3. **DC-CUNets Split Network with Multiscale Void Convolution Module**

When the multiscale void convolution is fused into the DC-CUNets segmented network proposed in Chapter 3, we replace the fourth module in the U-Net structure used in each channel of the two-channel liver tumor segmented network with the multiscale void convolution module. The U-Net network structure after the multiscale void convolution module is fused is shown in Figure 5.

![Figure 5. U-Net after fusing the multiscale void convolution module.](image)

(1) Replace the double-layer convolution layer and the pooled layer in the fourth module of the network with the empty convolution module, and remove the first deconvolution layer in the expansion path. Because the feature of void convolution is to increase the local perception field
without reducing the image resolution, the image size will not be reduced by 1/2 as the original U-Net structure after multiscale void convolution module. Therefore, in the improved multiscale DC-CUNets structure, the first deconvolution layer in the U-Net expansion path is also removed, directly expanding the output connections of multiscale void convolution module. The first two-layer convolution layer in the path cascades the left and right branches of the U-Net.

(2) Increase skip connections to maintain U-Net network depth and number of image features. The U-Net network itself has the structure feature of jump connection, which combines shallow image features with deep image features through stitching to obtain global and local information of the image. When the double-layer convolution layer and pooling layer of the fourth module in the original U-Net structure and the corresponding deconvolution layer are removed, a jump connection is also reduced, which reduces the depth and expansion path of the U-Net network. The number of feature image channels will be reduced. For this point, we connect the output of the first layer of void convolution with the output of the first 3 *3 convolution layer in the diffusion path in a splice way to achieve a skip connection, which keeps the network depth constant and at the same time maintains the degree to which the shallow and deep features within the network are fused.

4. Experiments
Experiments are carried out under GPU of Nvidia GeForce GTX 2080 Ti. We use deep learning framework Tensorflow and public dataset is 3D-IRCADb-01 and the MICCAI 2017 LiTS Challenge dataset. The LiTS dataset contains 131 and 70 contrast-enhanced 3D abdominal CT scans for training and testing. The 3D-IRCADb-01 consists of 20 volumes 3D-enhanced CT venous phase images. We select 16 volumes for training while 4 volumes for testing. The input image size is 512*512 in this paper. The format of this dataset is DICOM medical image format, so we need to first convert the pictures to HU values, then add windows. Based on the range of CT values of different organs and the scanning properties of the scanner, we set the window position (WL) to 35 and the window width (WW) to 200 to increase the contrast between the liver and other parenchyma of the abdomen, where the range of CT values within the window is [-75, 135].

4.1. Experimental Results of Multistage CT Liver Tumor Segmentation with Two-Channel Cascade U-Nets
In the process of DC-CUNets training, we train the network by random out of the plan parameters. On our hardware platform, the training phase of the first-level liver segmentation U-Net requires 80 hours, while the second-level dual-channel liver tumor segmentation U-Nets requires 150 hours because the network parameters between channels are not shared. Figure 6 The results of each segmentation method. To be consistent with other segmentation methods, we outline the liver probability map and liver tumor probability map obtained from DC-CUNets network on intravenous CT images.

![Image](image_url)

(a) Arterial phase  (b) Venous phase  (c) standard  (d) FCN[3]
(e) CFCNs[4]  (f) H-DenseUNet[6]  (g) improved CFCNs[12]  (h) DC-CUNets

Figure 6. DC-CUNets liver and tumor comparisons (liver outline in blue and tumor outline in green).

The following points can be seen from the Table 2: (1) Comparing single-layer FCN[3], CFCNs[4] and improved CFCNs [12], it can be proved that using cascaded automatic liver and tumor segmentation network can effectively improve the overall segmentation accuracy; (2) Comparing DC-CUNets with other segmentation methods[6], in the case of liver tumor segmentation network, using dual-channel structure to fuse multi-phase CT image features, the accuracy of tumor segmentation was improved by 4% to 5%. (3) Because DC-CUNets fuses the image characteristics of multi-phase CT images at different stages after contrast agent injection, it can effectively reduce the false positivity of liver tumor segmentation compared with other segmentation methods. The accuracy and recall rate of DC-CUNets in Table 2 are higher than those of other segmentation methods.

| Method              | VOE (%) | RVD (%) | ASD (mm) | MSD (mm) | Precision (%) | Recall (%) | Liver DICE (%) | Tumor DICE (%) |
|---------------------|---------|---------|----------|----------|---------------|------------|----------------|----------------|
| FCN[3]              | 80.19   | -2.57   | 12.41    | 115      | 72.13         | 83.49      | 83.52          | 71.96          |
| CFCNs[4]            | 44.02   | 1.79    | 5.23     | 43.7     | -             | -          | 86.49          | 75.38          |
| H-DenseUNet[6]      | 15.35   | 0.46    | 0.74     | 19.6     | -             | -          | 88.35          | 78.71          |
| Improved CFCNs[12]  | 16.74   | -0.37   | 1.08     | 25.33    | 80.72         | 82.05      | 87.93          | 79.94          |
| DC-CUNets           | 11.68   | -0.35   | 0.92     | 17.3     | 84.24         | 87.87      | 90.37          | 83.41          |

Table 2. Comparison of liver tumor segmentation indices.

4.2. Multiscale DC-CUNets Multiphase CT Liver Tumor Segmentation Method Experimental Results

The proposed multi-scale DC-DUnets can accurately segment tumors of different scales, greatly reducing the false-negative and false-positive rates of tumor segmentation.
Figure 7. Experimental liver and tumor segmentation of multiscale DC-CUNets (liver outline in blue and tumor outline in green).

As we can see, it is shown in Table 3. CFCNs use conditional random fields to optimize the split boundary. Each slice takes 0.8s to split. H-DenseUNets combines three-dimensional spatial data. It takes 0.61 seconds to split each slice, 0.48 seconds for the original DC-CUNets, and 0.63 seconds for the multiscale DC-CUNets to split a slice. To sum up, compared with other state-of-the-art segmentation methods, multiscale DC-CUNets can improve the accuracy of tumor segmentation while still maintaining a competitive segmentation speed.

### Table 3. Comparison of liver tumor segmentation results.

| Method               | VOE (%)  | RVD (%) | ASD (mm) | MSD (mm) | Precision (%) | Recall (%) | Liver DICE (%) | Tumor DICE (%) | Time (s) |
|----------------------|----------|---------|----------|----------|---------------|------------|----------------|----------------|----------|
| FCN[3]               | 80.19    | -2.57   | 12.41    | 115      | 72.13         | 83.49      | 83.52          | 71.96          | 0.4      |
| CFCNs[4]             | 44.02    | 1.79    | 5.23     | 43.7     | -             | -          | 86.49          | 75.38          | 0.8      |
| H-DenseUNets[6]      | 15.35    | 0.46    | 0.74     | 19.6     | -             | -          | 88.35          | 78.71          | 0.61     |
| Improved CFCNs[12]   | 16.74    | -0.37   | 1.08     | 25.33    | 80.72         | 82.05      | 87.93          | 79.94          | 0.57     |
| DC-CUNets            | 11.68    | -0.35   | 0.92     | 17.3     | 84.24         | 87.87      | 90.37          | 83.41          | 0.48     |
| Multi-scale DC-CUNets| 9.24     | 0.27    | 1.03     | 15.45    | 86.31         | 90.04      | 90.37          | 86.29          | 0.63     |

5. Conclusions
This paper presents a multistage liver tumor segmentation method based on multiscale DC-CUNets. By analyzing the scale characteristics of liver tumors and the ability of void convolution to increase the local receptive field without reducing the image size, we designed a two-layer multiscale void convolution module for three different size tumors to obtain the local receptive field at different scales so as to learn different image characteristics. At the same time, we replaced the fourth module of the two-channel U-Nets for splitting liver tumors with the multiscale void convolution module to achieve liver tumor splitting based on multiscale DC-CUNets.

References
[1] Masuda Y, Tateyama T, Xiong W 2011 Liver tumor detection in CT images by adaptive contrast enhancement and the EM/MPM algorithm *IEEE International Conference on Image Processing* IEEE pp 1421-4

[2] Wen S P, Chen M Z Q, Yu X H, Zeng Z G and Huang T W 2017 Fuzzy control for uncertain vehicle active suspension systems via dynamic sliding-mode approach *IEEE Transactions on Systems, Man and Cybernetics: Systems* 47(1) 24-32

[3] Ben-Cohen A, Diamant I, Klang E, et al. 2016 Fully convolutional network for liver segmentation and lesions detection. *Deep Learning and Data Labeling for Medical Applications*. (Springer, Cham) pp 77-85

[4] Christ P F, Elshaer M E A, Ettinger F, et al. 2016 Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields *International Conference on Medical Image Computing and Computer-Assisted Intervention*. (Springer, Cham) pp 415-423

[5] Wen S P, Chen M Z Q, Zeng Z G, Huang T W and Li C J 2017 Adaptive neural-fuzzy sliding-mode fault-tolerant control for uncertain nonlinear systems *IEEE Transactions on Systems, Man and Cybernetics: Systems* 47(8) 2268-78

[6] Li X, Chen H, Qi X, et al. 2018 H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes *IEEE Transactions on Medical Imaging* 2018 2663-74

[7] Wen S P, Zeng Z G, Chen M Z Q and Huang T W 2017 Synchronization of switched neural networks with communication delays via the event-triggered method *IEEE Transactions on Neural Networks and Learning Systems* 28(10) 2334-43

[8] Ronneberger O, Fischer P, Brox T 2015 U-net: Convolutional networks for biomedical image segmentation *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer) pp 234-241

[9] Wen S P, Huang T W, Yu X H, Chen M Z Q and Zeng Z G 2016 Aperiodic sampled-data sliding-mode control of fuzzy systems with communication delays via the event-triggered method *IEEE Transactions on Fuzzy Systems* 24(5) 1048-57

[10] Wen S P, Yu X H, Zeng Z G and Wang J J 2016 Event-triggering load frequency control for multi-area power systems with communication delay *IEEE Transactions on Industrial Electronics* 63(2) 1308-17

[11] Dong C, Chen Y, Seki T, et al. 2015 Non-rigid image registration with anatomical structure constraint for assessing locoregional therapy of hepatocellular carcinoma *Computerized Medical Imaging and Graphics* 45 75-8

[12] Pang Y, Hu D, Sun M A 2019 modified scheme for liver tumor segmentation based on cascaded FCNs *Proceedings of the International Conference on Artificial Intelligence, Information Processing and Cloud Computing* pp 1-6