DefED-Net: Deformable Encoder-Decoder Network for Liver and Liver Tumor Segmentation

Tao Lei, *Senior Member, IEEE*, Risheng Wang, Yuxiao Zhang, Yong Wan, Chang Liu, and Asoke K. Nandi, *Fellow, IEEE*

***Abstract*—Deep convolutional neural networks have been widely used for medical image segmentation due to their superiority in feature learning. Although these networks are successful for simple object segmentation tasks, they suffer from two problems for liver and liver tumor segmentation in CT images. One is that convolutional kernels of fixed geometrical structure are unmatched with livers and liver tumors of irregular shapes. The other is that pooling and strided convolutional operations easily lead to the loss of spatial contextual information of images. To address these issues, we propose a deformable encoder-decoder network (DefED-Net) for liver and liver tumor segmentation. The proposed network makes two contributions: 1) the deformable convolution is used to enhance the feature representation capability of DefED-Net, which can help the network to learn convolution kernels with adaptive spatial structuring information and 2) we design a ladder-atrous-spatial-pyramid-pooling (Ladder-ASPP) module using multiscale dilation rate (Ladder-ASPP) and apply the module to learn better context information than the atrous spatial pyramid pooling for CT image segmentation. The proposed DefED-Net is evaluated on two public benchmark datasets, the LiTS, and the 3DIRCADb. Experiments demonstrate that the DefED-Net has better capability of feature representation as well as provides higher accuracy on liver and liver tumor segmentation than state-of-the-art networks. The available code of DefED-Net we propose can be found from https://github.com/SUST-reynole/DefED-Net.**

***Index Terms*—Deep learning, deformable convolution (DC), image segmentation, ladder-atrous-spatial-pyramid-pooling (Ladder-ASPP), U-Net.**

Manuscript received July 31, 2020; revised October 23, 2020, December 4, 2020, and February 3, 2021; accepted February 10, 2021. Date of publication February 16, 2021; date of current version December 30, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 61871259, Grant 61811530325 (IEC\NSFC\170396, Royal Society, U.K.), Grant 61871260, Grant 61672333, and Grant 61873155; and in part by the Science and Technology Program of Shaanxi Province of China under Grant 2020NY-172. *(Corresponding authors: Tao Lei; Yong Wan.)*

Tao Lei and Risheng Wang are with the School of Electronic Information and Artificial Intelligence and the Shaanxi Joint Laboratory of Artificial Intelligence, Shaanxi University of Science and Technology, Xi'an 710021, China (e-mail: leitao@sust.edu.cn).

Yuxiao Zhang is with the School of Electrical and Control Engineering, Shaanxi University of Science and Technology, Xi'an 710021, China (e-mail: xyu@sust.edu.cn).

Yong Wan is with the Department of Geriatric Surgery, First Affiliated Hospital, Xi'an Jiaotong University, Xi'an 710061, China (e-mail: rareyong@qq.com).

Chang Liu is with the Department of Hepatobiliary Surgery, First Affiliated Hospital, Xi'an Jiaotong University, Xi'an 710061, China (e-mail: liuchangdoctor@163.com).

Asoke K. Nandi is with the Department of Electronic and Electrical Engineering, Brunel University London, Uxbridge UB8 3PH, U.K., and also with the School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: asoke.nandi@brunel.ac.uk).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TRPMS.2021.3059780.

Digital Object Identifier 10.1109/TRPMS.2021.3059780

I. Introduction

LIVER cancer is one of the most common and most lethal cancers in the world, which threatens life and health of humans seriously [1], [2]. In the clinical context a liver is a common site for both primary (e.g., hepatocellular carcinoma) or secondary (e.g., hepatic metastases due to colorectal cancer) tumor development. Accurate liver and liver tumor segmentation from enhanced abdominal CT images can help doctors to assess the function of livers and make a decision for disease diagnosis and treatment. However, as livers have a similar density with other neighboring organs and the liver tumors show very low contrast and serious intensity inhomogeneities in abdominal CT images, it is difficult to find accurate liver and liver tumor boundaries depending on human vision [3]. Manually labeling liver and liver tumor areas not only suffers from subjective judgment and limited accuracy, but also is tedious and inefficient. Therefore, semi-automatic or fully automatic approaches for liver and liver tumor segmentation have been a research goal in the field of medical image analysis to help in clinical applications [4].

Before the advent of deep learning techniques [5], liver and liver tumor segmentation were often semi-automatic and they mainly relied on image segmentation algorithms based on model-driven, such as region growing [6], active contour models [7], graph cut [8], shape statistical models [9], etc. These approaches can be roughly categorized into three groups: 1) pixel-based approaches; 2) graph-based approaches; and 3) contour-based approaches. The first type of approach mainly includes thresholding and region merging. The type of approach only achieves low segmentation accuracy for liver and liver tumor segmentation due to the employment of low-level features and limited capability of model representation. Graph-based approaches show clear superiority than pixel-based approaches, since they employ the max-flow/min-cut algorithm to find a minimum-cost closed set [10]. This kind of semi-automatic approach can achieve accurate liver segmentation by simply labeling the foreground and background, and it does not even require the iterative operation [11]. However, image segmentation results are easily influenced by labeling results, and graph cuts require high computational cost for

high-resolution images since each pixel of images is viewed as a note [12]. Consequently, researchers often employ the combination of graph cuts and other algorithms, such as watershed [13], shape constrain [14], multiscale registration [15], etc., to improve the segmentation accuracy and computational efficiency for liver and liver tumor segmentation.

Compared to the first two kinds of approaches, contour-based liver and liver tumor segmentation attracts more researchers' attention since they can provide better segmentation results using curve or shape evolution. Level-set [16] is one of the most popular algorithms in medical image segmentation, since the level-set utilizes energy optimization to evolve a given curve into the real boundaries of objects. A large number of improved level-set algorithms have been proposed by introducing partial differential equations into the evolution process to improve the convergence speed and segmentation accuracy [17]–[19]. The statistical shape model [20] is another popular algorithm for contour-based liver and liver tumor segmentation. Different from the level-set, this algorithm often constructs first a training set of liver and liver tumor shapes, and then employs machine learning algorithms, such as the random forest [21], [57], [58], support vector machine (SVM) [22], and adaboosted histogram [59] to learn an effective classifier. As a result, each liver and liver tumor shape can be represented by some corresponding patches from liver and liver tumor surface in the training set [23], [24]. The advantage of this kind of approach is that they can provide better segmentation results than unsupervised approaches, but the disadvantage is that segmentation results depend on the selection of training set and classifiers. Although numerous algorithms have been proposed for liver and liver tumor segmentation, they only provide good segmentation results for some slices with clear liver or tumor boundaries, and they are often unavailable for slices with blurred liver contour or intensity inhomogeneities tumors in practical applications. Ren *et al.* [54] proposed an automatic framework for atlas-based multiorgan segmentation in abdominal dynamic PET images with three different methods (4D-pair, 4D-PCA, and 3-D), incorporating probabilistic atlas information into the segmentation as a spatial prior using maximum *a posteriori* (MAP) estimation. This provides a powerful and reliable region of interest (ROI) for dynamic abdominal PET multi-organ segmentation for better segmentation results. Although atlas-based segmentation can easily capture anatomical variation and thus offers higher segmentation accuracy, it suffers from a clear shortcoming of ravenous appetite for computational resources because analyzing, manipulating, and processing all atlases typically demands a substantial amount of memory and time. It is believed that this is one of the main reasons why atlas-based segmentation has not been widely used in clinical applications.

In recent years, with the rapid development of deep learning [25] in the field of computer vision, researchers prefer to use fully convolutional neural networks (FCN) [26] to achieve image semantic segmentation in an end-to-end way [27]. These networks usually adopt multilevel encoder-decoder structures, and the encoder and decoder are often composed of a large number of standard convolutional or deconvolutional layers. In addition, there is a residual or long-range connection between encoders and decoders. This kind of design can automatically remove insignificant features and maintain interesting features through the contraction and expansion paths; it can also achieve the fusion of low-level and high-level features. Compared with FCN, U-Net proposed by Ronneberger *et al.* [28] obtained great success for medical image segmentation, since the encoder and decoder of U-Net are perfectly symmetrical and upsampling gradually makes it possible to obtain finer segmentation results. Since then, researchers focused on the improvements of U-Net [55], [56]. The most common way is to use the backbone of classic convolutional neural networks with pretrained parameters, such as VGG [29], ResNet [30], DenseNet [31], GhostNet [50], etc., to replace the encoder achieving transfer learning [60]. The other popular way of improving U-Nets is to add attention mechanisms [32] between encoders and decoders to focus on interesting regions, such as attention U-Net [33] and RA-UNet [34]. To exploit further potentially useful information in feature maps, R2-UNet [35] introduces recurrent convolution that is able to extract features using the same layer many times. UNet++ [36] employs U-Nets with different depths instead of long-range connections to avoid the rough fusion of low-level and high-level features. Recently, mU-Net [37] believes that small targets may disappear after pooling since the skip connection in U-Net repeatedly processes low-resolution feature information. Therefore, mU-Net achieves better liver and liver tumor segmentation by adding a residual path with deconvolution and activation operations to the skip connection of U-Net. These improved 2-D networks not only show better performance in medical image segmentation, but also achieve simpler design of data augmentation schemes while keeping the lower memory requirement than 3-D networks. However, they cannot capture the spatial information along the z-axis due to the employment of 2-D convolution kernels, which may degrade the performance in volumetric segmentations.

To extract the spatial information along the 3-D, Ji *et al.* [38] employed 3-D convolution kernels to achieve 3-D CNN, which makes it possible to process 3-D volume data directly. Based on 3-D CNN and U-Net, Çiçek *et al.* and Milletari *et al.* proposed 3-D U-Net [39] and V-Net [40], respectively. The V-Net applies 3-D convolutions together with residual connection to the feature encoder stage, and deepens the network depth to obtain better segmentation results than 3-D-UNet. Furthermore, by introducing the strategy of depth supervision, both Med3D [41] and 3-D DSN [42] achieve faster and more accurate segmentation of volumetric medical images. More application of 3-D CNNs can be seen in [43]. Although these 3-D networks can simultaneously explore the spatial information of interslice and inner-slice, these networks suffer from some new problems, such as more parameters, much memory usage, and much narrow reception fields than 2-D networks. To combine the advantages of 2-D and 3-D networks, researchers proposed H-DenseUNet [44]. This network first uses a 2-D network to extract image features and perform segmentation tasks on a slice-by-slice basis. The pixel-wise probabilities produced by the 2-D network are

then concatenated with the original 3-D volume and fed into a 3-D network for a refinement. The H-DenseUNet finally achieves excellent liver and liver tumor segmentation. In addition, Vu *et al.* [53] applied the overlay of adjacent slices as input to the central slice prediction, and then fed the obtained 2-D feature maps into a standard 2-D network for model training. Although these pseudo-3-D approaches can segment objects from 3-D volume data, they only obtain limited accuracy improvement due to the utilization of local temporal information. Compared to pseudo-3-D networks, hybrid cascading 2-D and 3-D networks are more popular for medical image segmentation.

Although the networks mentioned above can perform end-to-end liver and liver tumor segmentation well, the use of vanilla convolution limits the further improvement of segmentation accuracy. Since standard convolution kernels have a regular sampling grid, they are unable to capture accurately liver and liver tumor features with variable shapes in different slices. Besides, some improved networks, such as CE-Net [45] and MSB-Net [51] employ multiscale feature fusion to enhance feature representation of networks, but many network branches lead to the requirement of more parameters. To address these issues, we propose a deformable encoder-decoder network (DefED-Net) to improve liver and liver tumor segmentation. The proposed DefED-Net includes following advantages.

1) Feature extraction layers of the DefED-Net are constructed by using the deformable convolution (DC) with residual design. The design can more effectively extract the spatial context information of images while maintaining high-level features.
2) The feature fusion module of the DefED-Net depends on a ladder-atrous-spatial-pyramid-pooling (Ladder-ASPP) that employs multiscale dilated convolution kernels using variable dilation rate to obtain better spatial context information.
3) The DefED-Net provides higher segmentation accuracy for liver and liver tumor than state-of-the-art approaches, and it requires smaller memory usage due to the employment of depth separable convolution.

The remainder of this article is organized as follows. In the next section, we detailedly introduced the design of network architecture and advantages of our proposed DefED-Net. To demonstrate the superiority of DefED-Net, we introduced our experimental environment and pretreatment, performed ablation studies and comparative experiments, and analyzed the experimental results in Section III, followed by the conclusion in Section IV.

II. Method

In this work, we propose a DefED-Net and apply it to liver and liver tumor segmentation. Fig. 1 shows the architecture of the DefED-Net. As can be seen from Fig. 1 that the DefED-Net is an enhanced U-net and it is composed of three parts, including an encoder, a middle processing module, and a decoder. In contrast with the U-net, the DefED-Net employs the DC with residual structure to generate feature maps. Moreover, the

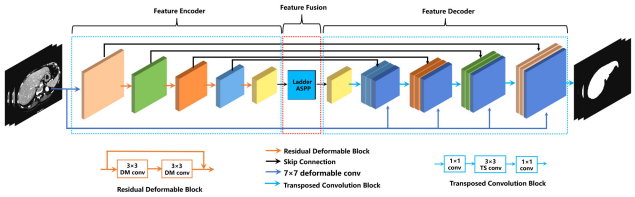

Fig. 1. Architecture of the proposed DefED-Net. First, the feature encoder employs DC using the residual connection. Second, the Ladder-ASPP block is used to extract richer context information. Finally, both the skip connection and the dense connection of original images are used for the fusion of feature maps in decoder.

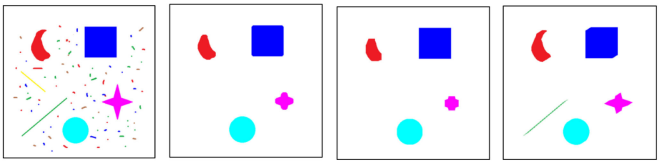

Fig. 2. Image filtering using a morphological opening filter with different structuring elements. (a) Original image. (b) SE is a disk of size 20×20. (c) SE is a square of size 20×20. (d) SE is a line that the length is 20 and the orientation is $1/6\pi$.

original image is concatenated with outputs at different layers of the decoder to obtain better feature representation. Different from general pyramid pooling (PP) modules [45], [46], we design a better feature fusion module, namely, Ladder-ASPP and apply it to our DefED-Net. Although the Ladder-ASPP adopts the way of dense connection, it only requires smaller memory usage due to the utilization of the depth separable convolution. It is worth mentioning that the DefED-Net is designed in 2-D domain.

A. Deformable Encoding

Although a large number of improved U-Nets have been proposed for medical image segmentation, they provide limited segmentation accuracy for livers and liver tumors in CT images. Here, are two reasons that limit the performance of U-Nets. First, convolutional kernels with fixed geometric structures are employed by the U-Nets, which ignores the shape information of objects in an image. Second, the operation of polling and strided convolution leads to the loss of spatial context detail information.

To illustrate the first reason, we presented an example of image filtering as shown in Fig. 2. Fig. 2 shows that the morphological opening filter is able to smooth noise effectively by employing different structuring elements (SEs). However, these filtering results depend on the choice of SEs. Fig. 2(b) shows that a circular SE is useful for preserving the details of circular objects and Fig. 2(c) shows that a square SE is effective for square objects. Similarly, Fig. 2(d) shows that a linear SE can maintain the details of linear objects. Therefore, it is better to adopt multiple different SEs for an image including many different objects. In other words, we should consider adaptive filters that can obtain better filtering effect due to the consideration of geometrical shape information of objects. In addition, the design of convolution

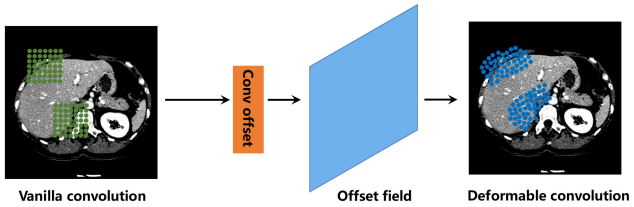

Fig. 3. Comparison of vanilla convolution and DC for liver segmentation. In contrast with the standard convolution, the DC requires offset locations for each sampling location.

kernels also plays the same important role for convolutional neural networks. In practical applications, researchers often employ fixed-shape square convolutional kernels to perform feature learning, such as U-Net, PSP-Net, CE-Net, etc. Since convolution kernels with fixed shape show weak ability for the extraction of image contextual information, these aforementioned networks only provide tolerable accuracy for liver and liver tumor segmentations. Instead, we use DC kernels to extract richer geometry information of the liver and liver tumor, which can better accommodate the irregular shape of liver and liver tumor and lead to better segmentation results. In Fig. 3, the DC shows better adaption for liver in a CT image than the vanilla convolution. In fact, Sun *et al.* [62] have started to explore the utilization of DC on automatic segmentation networks for gastric cancer, and their proposed network achieves better segmentation results than vanilla U-Net [28] and ResU-Net [61].

The DC is able to provide convolutional kernels with arbitrary shapes by learning offset locations, and thus adaptively decide scales of receptive field with fine localization. Therefore, the DefED-Net possesses better capability of modeling geometric transformation than common U-Nets due to the employment of deformation convolution. However, the implementation of DC is more complex than vanilla convolution since additional spatial offset locations are limited. Based on learned offset locations, the convolution kernels can achieve the deformation of different scales, shapes, and orientations. Fig. 3 illustrates the principle of deformation convolution on liver segmentation.

In practical applications, a DC is composed of four layers: 1) a convolutional layer; 2) a convolutional offset layer; 3) a batch normalization layer; and 4) an activation layer. The principle of DC is given as follows. Let x and y be the input and the output feature map, respectively. The L denotes a regular grid in 2-D domain.

When performing the convolution operation on x using the L, the output is denoted by

$$y(e_0) = \sum_{e_n \in L} w(e_n) \times x(e_0 + e_n) \tag{1}$$

where w denotes the weight, e_0 denotes the location of a pixel, and e_n denotes the location of neighboring pixels falling into L. If we perform the DC on x, the output can be represented by

$$\widetilde{y}(e_0) = \sum_{e_n \in \widetilde{L}} w(e_n) \times x(e_0 + e_n + \Delta e_n) \tag{2}$$

where $\widetilde{L}$ is the deformation result of L. Compared to L, $\widetilde{L}$ is an irregular grid including offset locations Δe_n.

The offset $\triangle e_n$ is usually a float number and the sampling position of the DC becomes irregular, so the bilinear interpolation is used to perform the process of determining the pixel value of the final sampling position. The pixel value $x(e)$ at the final sampling position is defined as

$$x(e) = B\big(w_i, q_j\big) \tag{3}$$

where w_i denotes the corresponding weight, q_j denotes the four surrounding pixels involved in the computation at the irregular sampling position, and $B(,)$ is the bilinear interpolation kernel. Note that B is 2-D and is defined as

$$B\big(w_i, q_j\big) = w_1 q_1 + w_2 q_2 + w_3 q_3 + w_4 q_4. \tag{4}$$

For instance, if the coordinates we got from the sampling position is (2.2, 4.6), then its nearest pixel is (2, 4),(2, 5), and (3, 4), and (3, 5). Therefore, in the actual program calculation, we will use the bilinear interpolation of (2, 4), (2, 5), (3, 4), and (3, 5) pixels for the sampled location (2.2, 4.6) pixels.

As shown in Fig. 3, the offset is obtained by applying a convolutional layer. Note that the convolutional layer to obtain the offset needs to have the same spatial resolution and dilation rate as the convolutional layer to extract the features in the offset feature map. For each layer of the DC, when the input of a convolutional layer is a feature map with N channels, the corresponding offset map includes $2N$ channels in this convolutional layer because each channel includes two offset maps in the x and y directions, separately. Note that the offset map of the output has the same spatial resolution as the input map in a convolutional layer. During training, the offset can be learned through the back propagation of (3) and (4). After the pixel values of all sampled positions are obtained, a new feature map will be generated. Although the DC is superior to vanilla convolution due to the employment of convolutional kernels with flexible shape, it can be further improved by using multiscale convolutional kernels instead of single-scale kernels. For liver segmentation task, a large convolutional kernel is better than small ones for capturing coarse liver areas. However, a small convolutional kernel is more useful for obtaining accurate contour details. Therefore, here we use a large convolutional kernel 7×7 for the first DC layer while using small convolutional kernel 3×3 for subsequent layers. The proposed multiscale DC is able to achieve better feature representation than single-scale DC, and thus leads to better liver and liver tumor segmentation results due to more accurate liver and liver tumor contours. In addition, the residual design is integrated in the proposed deformable encoder to avoid vanishing gradients and speeds up the convergence of networks.

B. Ladder-ASPP

Both PP and atrous spatial pyramid pooling (ASPP) are two popular ways for encoding context information due to wider receptive fields than standard pooling. Since the PP directly performs pooling operation using multiscale pooling kernels, it often causes irreversible information loss leading to poor segmentation results for small objects such as liver

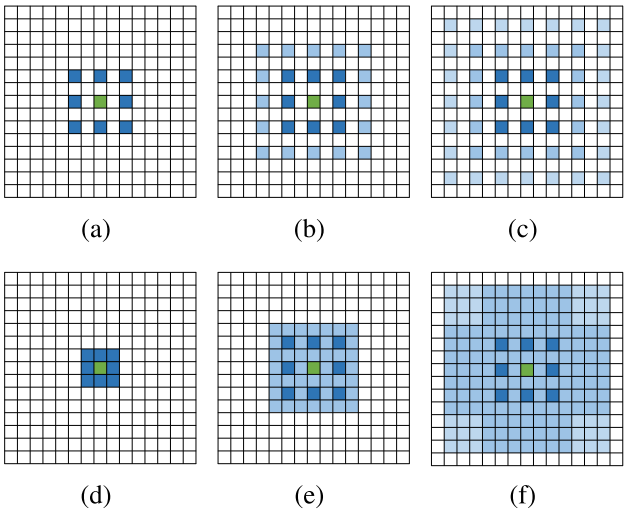

Fig. 4. Comparison of standard atrous convolution and the atrous convolution with variable dilation rate. (a) Convolutional kernel: 3×3, rate = 2. (b) Cascade of two convolutional kernels: 3×3, rate = 2. (c) Cascade of three convolutional kernels: 3×3, rate = 2. (d) Convolutional kernel: 3×3, rate = 1. (e) Cascade of two convolutional kernels: 3×3, rate = 1, 2. (f) Cascade of three convolutional kernels: 3×3, rate = 1, 2, and 3. Note that although (c) and (f) have similar receptive fields 13×13, (c) has 70% pixel loss compared to (f).

tumor. However, the ASPP performs atrous convolution using multiple dilation convolutional kernels instead of multiscale pooling kernels. Compared with PP, ASPP provides better context information since atrous convolution is superior to pooling operation for the preservation of detail information. However, ASPP still faces two challenges in practical applications: 1) the fixed dilation rate is used for ASPP, which causes gridding effect as shown in Fig. 4(a)–(c); some pixels falling into receptive fields cannot take part in the convolutional operation and 2) ASPP ignores the global context information. To address these issues, we proposed a novel Ladder-ASPP as shown in Fig. 5.

The standard atrous convolution easily leads to the loss of spatial detail information. To overcome the drawback, we use variable dilation-rate instead of fixed dilation rate leads to better receptive fields. It is clear that each pixel in the receptive fields is covered as shown in Fig. 4(d)–(f). Therefore, atrous convolution with variable dilation rate can overcome gridding effect caused by the standard atrous convolution.

Based on atrous convolution with variable dilation rate, we design a Ladder-ASPP to improve context encoding. Fig. 5 shows the architecture of the Ladder-ASPP.

First, the Ladder-ASPP employs variable dilation rate to achieve atrous convolution that was mentioned previously.

Second, the Ladder-ASPP uses densely ladder connection that is helpful for ASPP to achieve better feature fusion. However, the dense connection easily leads to the increase of the number of parameters and high memory requirement. To reduce the number of parameters to obtain a lightweight network, we introduce depthwise separable convolution (DSC) [52] to Ladder-ASPP. Compared to the standard convolution in which spatial features and channel features are often coupled together, the DSC can achieve the decoupling

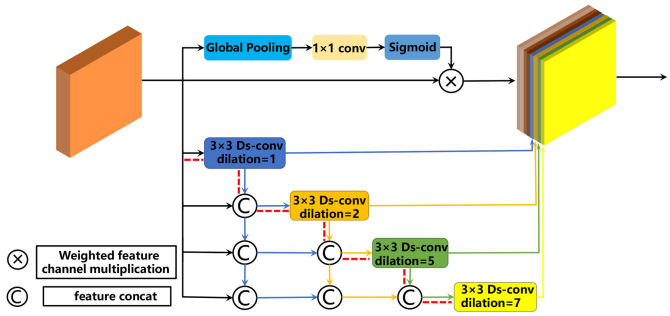

Fig. 5. Architecture of the Ladder-ASPP. The output feature maps are concatenated by two parts. The first one is the output from global pooling and the second one is the densely connected feature fusion liking ladder.

computation between spatial features and channel features leading to the requirement of fewer parameters.

It is well-known that the standard convolution requires parameters $D_K \times D_K \times M \times N$, where M is the dimension of input feature maps, N is the dimension of output feature maps, and D_K is the space-resolution of convolution kernels. In the DSC, the depthwise convolution only requires parameters $D_K \times D_K \times 1 \times M$ and the pointwise convolution only requires parameters $1 \times 1 \times M \times N$. Therefore, the number of parameters of DSC is $(1/N + 1/D_K^2)$ of the standard convolution. Here, the proposed Ladder-ASPP employs four kernels of size 3×3. Consequently, the Ladder-ASPP only requires 36% parameters compared to the one without using DSC.

Finally, to improve feature representation of ASPP, the global pooling is integrated into Ladder-ASPP since it can achieve the priority of channels including more important information. We can see from Fig. 5 that the information hidden in both space dimension and channel dimension is exploited simultaneously. The final feature maps fuse both the global and local information.

To illustrate the proposed Ladder-ASPP, let Y be the output feature map, y_1 be the global pooling result, and y_2 be the output from the module of ladder autrous convolution. It is clear that $Y = y_1 + y_2$. The y_1 is defined as

$$y_1 = B[C_1[GP_S(x)]] \times x \tag{5}$$

where x is the feature map obtained from the feature encoder, followed by global pooling denoted by $GP_S(x)$, and C_1 represents the weight of each feature channel through 1×1 convolution, and B is the normalization of feature weight.

In our Ladder-ASPP, we adopt variable dilation rate, i.e., 1, 2, 5, and 7. Let $G_{K,D}$ be the output of densely connected PP, where K is the level of pyramid and D is dilation rate, we get

$$y_2 = G_{1,1}(x^{(1)}) \oplus G_{2,2}(x^{(2)}) \oplus G_{3,5}(x^{(3)}) \oplus G_{4,7}(x^{(4)}) \tag{6}$$

where

$$\begin{cases} x^{(1)} = x \\ x^{(2)} = x^{(1)} \oplus G_{1,1}(x^{(1)}) \\ x^{(3)} = x^{(2)} \oplus G_{2,2}(x^{(2)}) \\ x^{(4)} = x^{(3)} \oplus G_{3,5}(x^{(3)}) \end{cases} \tag{7}$$

where $\oplus$ denotes concatenation operation.

According to (5)–(7), we can see that the output from Ladder-ASPP includes richer information than the original input. The Ladder-ASPP can help our DefED-Net to achieve better segmentation results due to the exploitation of significant spatial information.

C. Loss Function

Our framework is an end-to-end deep learning system. As illustrated in Fig. 1, we need to train the proposed method to predict each pixel as foreground or background, which is a pixel-wise classification problem. The cross entropy is one of the most popular loss functions and it is defined as

$$L_{\text{cross}} = -(p\log(\hat{p}) + (1-p)\log(1-\hat{p})) \tag{8}$$

where p and $\hat{p}$ are the ground truth and predicted segmentation, respectively.

However, the tumor often occupies a small region in an image. The cross entropy loss is not optimal for such tasks. It is worth noting that the Dice loss [40] is suitable for uneven samples. This metric is essentially a measure of overlap between a segmentation result and corresponding ground truth. The Dice loss is defined as

$$L_{\text{dice}} = 1 - \frac{2 < p, \hat{p} >}{\|p\|_1 + \left\|\hat{p}\right\|_1} \tag{9}$$

where $p \in (0, 1)$ and $0 \leq \hat{p} \leq 1$. The p and $\hat{p}$ are the ground truth and predicted segmentation, respectively, and $<p, \hat{p}>$ denotes dot product.

However, the use of the Dice loss easily influences the back propagation and leads to a training difficulty. Therefore, the final loss function is defined as a combination of both losses

$$L_{\text{loss}} = L_{\text{cross}} + L_{\text{dice}}. \tag{10}$$

D. Post Processing

Generally, the task of liver and liver tumor segmentation aims to obtain a binary image where the foreground denotes liver and liver tumor and the background denotes other areas. Based on Sections II-A and II-B, we can obtain a coarse segmentation result for livers and liver tumors. However, the segmented image often includes a lot of small and isolated areas or some holes. In practical applications, binary image filtering is often used to remove false liver areas or fill holes within livers. For binary image filtering, morphological filters are very popular for the removal of small segmentation areas.

Although both classic morphological opening and closing operations can effectively improve binary segmentation results, they often smooth the boundaries of main objects as well. It is difficult to remove false objects while maintaining the boundary accuracy of real objects. For this problem, morphological reconstruction is an excellent tool and it has been widely used for object extraction [48]. Morphological reconstruction is able to achieve binary image filtering while maintaining the large objects unchanged. The operation requires to set the parameter of SEs. If the parameter is large, more small areas would be removed. On the contrary, fewer areas are removed in the case of small value of parameters. To address the issue, we propose an adaptive morphological reconstruction to optimize liver and liver segmentation results from the DefED-Net.

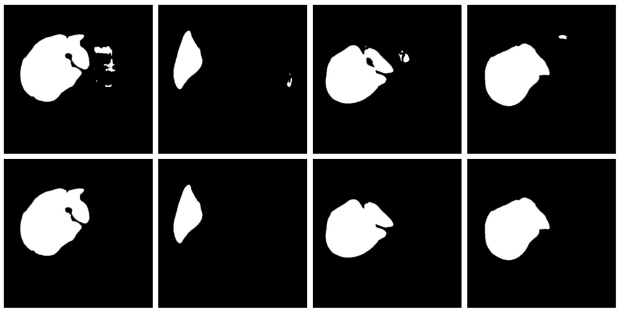

Fig. 6. Post-processing results using adaptive morphological reconstruction. Top: segmentation results from the DefED-Net. Bottom: post-processing results.

We first compute the proportion of the maximal connected component in an image to the total area of the image. If the value is large, then a large SE will be adopted. Conversely, a small SE will be adopted when the value is small. Here, the SE is a disk and its radius is denoted by r

$$r = 30 \times \text{round}(R/(H \times W)) + 1 \tag{11}$$

where R denotes the area of the maximal connected component in the segmentation result, and H and W denote the height and width of the input image, respectively. Fig. 6 shows post-processing results using the proposed adaptive morphological reconstruction. Note that it is unnecessary to make post-processing for liver tumor segmentation since the area of liver tumors is generally small.

III. Results and Discussion

A. Dataset and Preprocessing

Two public contrast-enhanced CT scans datasets: liver tumor segmentation challenge (LiTS-ISBI2017) and the 3-D image reconstruction for comparison of algorithm and database (3Dircadb) datasets are considered as experimental data. The LiTS dataset is a large dataset that contains 130 3-D abdominal CT scans, where the image size is 512×512, slice thickness varied from 0.55 to 6 mm, pixel spacing varied from 0.55 to 1 mm. The 3DIRCADb is a small dataset that contains 22 3-D data, where the image size is 512×512, slice thickness varied from 1 to 4 mm, pixel spacing varied from 0.56 to 0.86 mm, and slice number varied from 184 to 260. We constructed the training set and validation set using 90 patients (total 43 219 axial slices) and ten patients (total 1,500 axial slices), respectively. Then the other 30 patients (total 15 419 axial slices) are considered as the test set. For the 3DIRCADb, it was split into 17 patients for training and ten patients for test.

Medical CT axial slices are different from normal axial slices, the former is able to obtain wider range of values from −1000 to 3000 than the latter from 0 to 255. To remove interferences and enhance liver areas, we truncated the image intensity values of all scans of [−200, 250] HU and performed the normalization on these scans. In our experiments, the given models are independently and separately performed for liver and liver tumor segmentation.

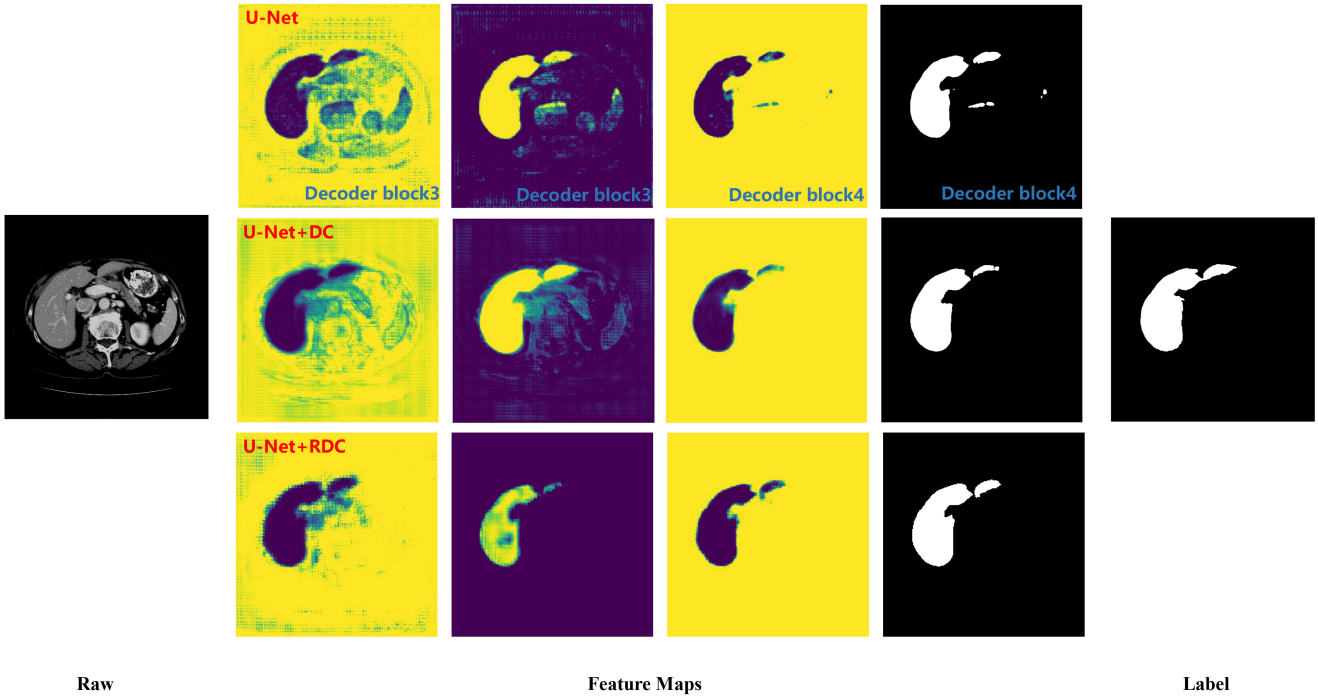

Fig. 7. Comparison of feature maps generated by U-Net, U-Net+DC, and U-Net+RDC, respectively.

B. Experimental Setup and Evaluation Metrics

All algorithms were implemented on a desktop PC with double NVIDIA GeForce RTX 2080 Ti with 11-GBVRAM. The convolutional neural networks were performed and trained using the framework of Pytorch 1.3.0.

On the model training, we set the initial learning rate (lr) to 0.001, and define the decay strategy for learning rate during training as

$$lr = lr \times (1 - i/t_i)^{0.9} \tag{12}$$

where i denotes the number of iterations of this training and t_i denotes the total number of iterations. Note that the DC requires two learning rates compared to one for vanilla convolution. We set $lr_2 = lr \times 0.01$ for offset convolutional layers used for DC networks, and used the Adam gradient descend with momentum to optimize the model.

Five popular evaluation metrics are used to measure the accuracy of segmentation results, such as dice score (DICE) [49], volumetric overlap error (VOE), relative volume difference (RVD), average symmetric surface distance (ASD), and root mean square symmetric surface distance (RMSD). The tumor burden of the liver is a measure of the fraction of the liver afflicted by cancer. In particular, as a metric, we measure the root mean square error (RMSE) in tumor burden estimates from lesion predictions. The value of DICE ranges from 0 to 1, and a perfect segmentation yields a DICE value of 1. In fact, the DICE is one of the most important metrics in image segmentation evaluation. The VOE is the complement of the Jaccard coefficient, and thus a perfect segmentation yields a VOE value of 0. The RVD is an asymmetric metric, and a smaller value of RVD means a better segmentation result. Both ASD and RMSD are used to measure the surface distance between segmentation results and ground truths, the former is used to compute the average distance but the latter is used to compute the maximal distance. Consequently, a better segmentation result corresponds to high values of DICE but low values of VOE, RVD, ASD, and RMSD. Note that we evaluate segmentation results based on 3-D volumes.

C. Ablation Study

This article focuses on liver and liver tumor segmentation. Two contributions are highlighted, one is that the DC is used to instead of the vanilla convolution; the other is that Ladder-ASPP is integrated into the proposed DefED-Net to improve the context information. To demonstrate the two contributions and the effectiveness of the DefED-Net, we conducted comprehensive experiments on both LiTS liver and liver tumor datasets.

Effectiveness of the DC: We analyzed the performance of DC and residual DC (RDC), respectively. Fig. 7 shows the comparison of U-Net, U-Net+DC, and U-Net+RDC on liver segmentation. It is clear that both DC and RDC can help U-Net to focus on the interesting regions and remove irrelevant background information, but RDC can help the network converge faster and obtain more accurate edge predictions. In the third column of Fig. 7, the feature maps provided by U-Net+RDC include less information that is unrelated with the liver. Consequently, U-Net obtains more fake liver regions than U-Net+RDC and U-Net+DC in the fifth column of Fig. 7. Table I demonstrates the effectiveness of the first contribution. We can see that the utilization of DC effectively raises the segmentation accuracy of U-Net. The residual design not only speeds up the convergence of U-Net, but also further improves segmentation accuracy.

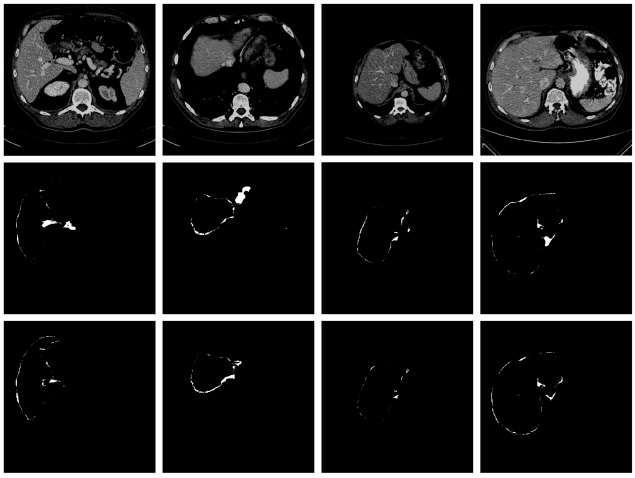

Fig. 8. Difference of prediction results and ground truths. Top: Input images, Middle: U-Net results, and Bottom: U-Net+Ladder-ASPP results.

TABLE I
COMPARISON OF ABLATION STUDY ON LITS TEST DATASETS. THE BEST VALUES ARE IN BOLD

Method	Liver	Tumor
	DICE (%)	DICE (%)
U-Net [28]	93.99±1.18	82.16±6.26
U-Net+post-processing	94.25±1.06	82.16±6.26
U-Net+DC	95.23±1.13	84.57±6.21
U-Net+RDC	95.80±1.12	85.68±6.18
U-Net+ASPP	94.30±1.16	85.32±6.08
U-Net+DenseASPP	95.32±1.14	85.43±6.02
U-Net+Ladder-ASPP	95.50±1.09	86.72±5.87
DefED-Net(without post-processing)	96.02±1.04	**87.52**±5.32
DefED-Net	**96.30**±1.01	**87.52**±5.32

Effectiveness of Ladder-ASPP: Both U-Net+ASPP and U-Net+Ladder-ASPP use the idea of context encoding to improve feature representation of networks. The difference is that Ladder-ASPP uses atrous convolution with variable dilation rate and dense connection to obtain better context information than ASPP. Experimental results in Table I consistently demonstrate that both ASPP and Ladder-ASPP can help U-Net to improve segmentation accuracy of livers, and the latter is superior to the former. Fig. 8 shows the difference of segmentation results between prediction results and ground truths, where the foreground is the difference and the background is the same. It is clear that the prediction result obtained by U-Net+Ladder-ASPP is closer to the ground truth than U-Net. Furthermore, U-Net+ASPP only improves the representation capability of models on the capture of spatial context information, which is unavailable for the optimization of channel dimension. Therefore, U-Net+Ladder-ASPP provides higher DICE than U-Net+ASPP as shown in Table I.

In addition, post-processing is also useful for improving segmentation accuracy. Table I shows that RDC plays a more important role than ASPP and post-processing for improving segmentation accuracy. The results further demonstrate that location information is more important than feature fusion for image segmentation. Fig. 9 shows the comparison of segmentation boundaries, which further illustrates the ablation study. All these comparison results demonstrate the effectiveness of the DC, Ladder-ASPP, and post-processing on liver and liver tumor segmentation.

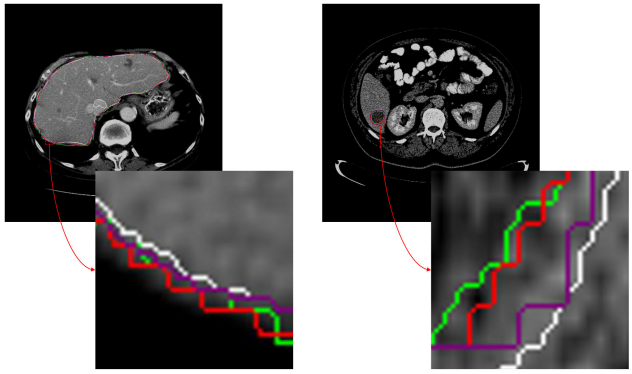

Fig. 9. Comparison of segmentation boundaries using different approaches. The green denotes ground truth, the white denotes the result provided by U-Net, the purple denotes the result provided by CE-Net, and the red denotes the result provided by DefED-Net.

D. Experimental Comparison on Test Datasets

To validate the superiority of the proposed DefED-Net, six state-of-the-art networks used for liver and liver tumor segmentation are considered as comparative approaches. These networks can be grouped into three categories: 1) 2-D networks; 2) 3-D networks; and 3) hybrid networks with 2-D and 3-D, where 2-D networks include U-Net, U-Net++, and CE-Net, 3-D networks include 3-D U-Net and V-Net, hybrid networks include H-DenseUNet. Note that we do not give experimental results obtained by 3-D U-Net and V-Net in Tables V and VI due to high risk of over-fitting on the 3DIRCADb dataset.

It is known that 3-D networks can provide better segmentation results than 2-D networks due to their exploitation of information between slices. Tables II and III demonstrate that both 3-D U-Net and V-Net provide higher segmentation accuracy than U-Net. However, CE-Net is superior to U-Net since it employs SPP to achieve feature fusion. In contrast with those networks mentioned above, H-DenseUNet provides better segmentation accuracy since it balances the advantages of both 2-D networks and 3-D networks. The proposed DefED-Net provides the best quantitative scores (DICE, VOE, and ASD) than comparative approaches. As the DefED-Net belongs to 2-D networks, it obtains lower values of RMSD than 3-D networks, such as 3-D U-Net and V-Net. Since the 3DIRCADb is a small dataset, 3-D networks including a mountain of parameters easily lead to over-fitting for the dataset. Therefore, we only show the comparison results of U-Net, U-Net++, CE-Net, H-DenseUNet, and DefED-Net in Tables IV and V, which demonstrates the DefED-Net outperforms those comparative networks on the 3DIRCADb dataset. DICE, VOE, and RVD are all overlap measures while ASD and RMSD are surface distance measures. The former focuses more on the interior of the segmentation target, while the latter focuses

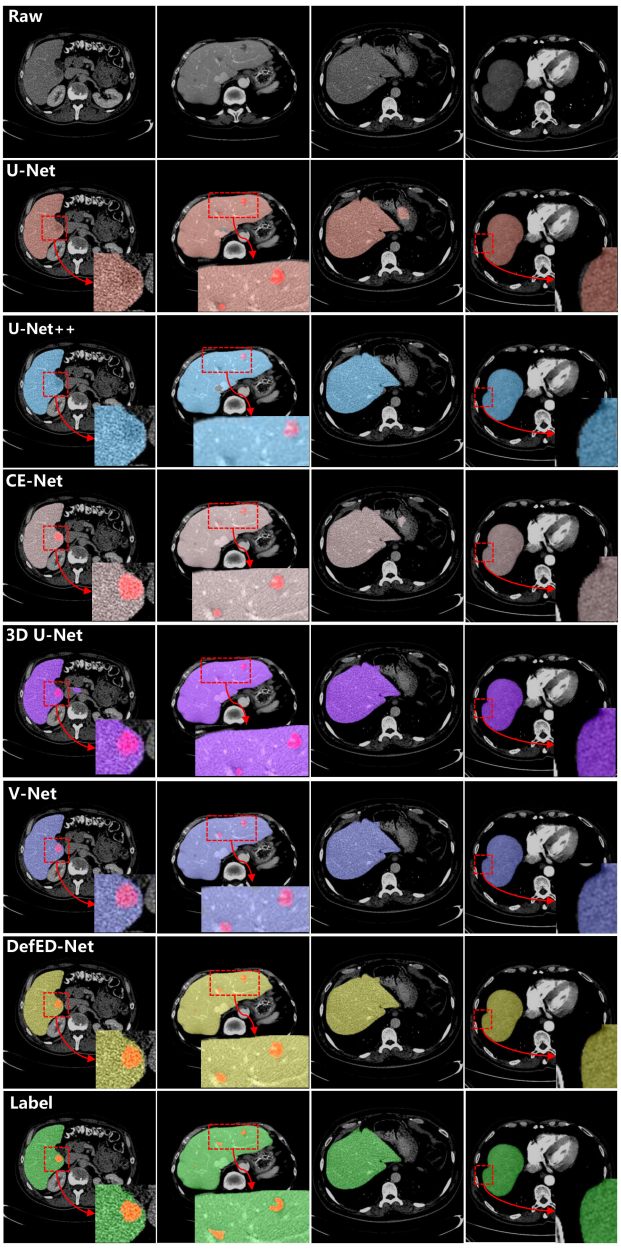

Fig. 10. Liver and liver tumor segmentation results using different approaches.

TABLE II
QUANTITATIVE SCORES OF THE LIVER SEGMENTATION RESULTS USING DIFFERENT APPROACHES ON THE LITS DATASET. THE BEST VALUES ARE IN BOLD

Method	LITS-Liver				
	DICE (%)	VOE (%)	RVD (%)	ASD (mm)	RMSD (mm)
U-Net [28]	93.99±1.23	11.13±2.47	3.22±0.20	5.79±0.53	123.57±6.28
U-Net++ [36]	94.01±1.18	11.12±2.37	2.36±0.15	5.23±0.45	120.36±5.03
CE-Net [45]	94.04±1.15	11.03±2.31	6.19±0.16	4.11±0.51	115.40±5.82
3D U-Net [39]	94.10±1.06	11.13±2.23	**1.42**±0.13	2.61±0.45	**36.43**±5.38
V-Net [40]	94.25±1.03	10.65±2.17	1.92±**0.11**	2.48±0.38	38.28±5.05
H-DenseUNet [44]	96.10±1.02	7.02±**2.00**	1.53±0.12	1.56±0.28	37.26±**3.64**
DefED-Net	**96.30±1.01**	**6.88**±2.10	1.46±0.12	**1.37±0.23**	77.60±4.26

TABLE III
QUANTITATIVE SCORES OF THE LIVER TUMOR SEGMENTATION RESULTS USING DIFFERENT APPROACHES ON THE LITS DATASET. THE BEST VALUES ARE IN BOLD

Method	LITS-Tumor					Tumor Burden
	DICE (%)	VOE (%)	RVD (%)	ASD (mm)	RMSD (mm)	RMSE
U-Net [28]	82.16±6.26	26.85±16.21	3.54±0.18	22.26±0.30	155.15±5.62	0.020
U-Net++ [36]	83.23±6.36	26.03±15.39	2.16±0.17	21.36±0.25	112.36±4.89	0.017
CE-Net [45]	84.02±6.15	25.62±15.21	1.59±0.17	20.79±0.28	100.29±5.23	0.018
3D U-Net [39]	85.13±5.87	25.13±15.02	1.23±0.14	20.32±0.27	62.36±5.16	0.018
V-Net [40]	85.87±5.42	24.52±14.86	1.09±0.15	19.23±0.25	68.32±4.52	0.017
H-DenseUNet [44]	86.23±**5.13**	24.46±**13.25**	0.53±0.13	18.83±**0.22**	**54.32±4.32**	**0.015**
DefED-Net	**87.52**±5.32	**23.85**±14.62	**0.52±0.10**	**17.41**±0.28	64.25±4.87	0.016

TABLE IV
QUANTITATIVE SCORES OF THE LIVER SEGMENTATION RESULTS USING DIFFERENT APPROACHES ON THE 3DIRCADB DATASET. THE BEST VALUES ARE IN BOLD

Method	3DIRCADb-Liver				
	DICE (%)	VOE (%)	RVD (%)	ASD (mm)	RMSD (mm)
U-Net [28]	92.30±1.27	11.78±3.62	-2.83±0.38	4.25±1.56	75.67±5.68
U-Net++ [36]	93.60±1.29	10.36±3.90	1.21±0.23	3.87±1.36	56.39±4.76
CE-Net [45]	94.28±1.22	10.02±3.53	-1.80±0.24	3.52±1.25	30.29±4.82
H-DenseUNet [44]	95.72±1.14	9.88±2.91	0.39±0.12	2.85±0.89	**9.63**±3.95
DefED-Net	**96.60±1.08**	**5.65±2.81**	**0.23±0.11**	**2.61±0.84**	12.76±**3.43**

TABLE V
QUANTITATIVE SCORES OF THE LIVER TUMOR SEGMENTATION RESULTS USING DIFFERENT APPROACHES ON THE 3DIRCADB DATASET. THE BEST VALUES ARE IN BOLD

Method	3DIRCADb-Tumor					Tumor Burden
	DICE (%)	VOE (%)	RVD (%)	ASD (mm)	RMSD (mm)	RMSE
U-Net [28]	51.25±8.28	50.75±18.26	-1.11±0.48	16.72±0.92	130.54±5.64	0.023
U-Net++ [36]	60.36±7.36	46.72±17.64	1.36±0.78	14.76±0.63	118.23±5.64	0.018
CE-Net [45]	60.25±7.18	40.36±17.26	0.93±0.32	12.42±0.79	110.67±4.92	0.020
H-DenseUNet [44]	65.47±**6.54**	36.74±**12.86**	**-0.74±0.18**	12.21±**0.51**	**32.52**±3.28	**0.015**
DefED-Net	**66.25**±6.62	**34.28**±13.43	0.81±0.20	**11.21**±0.63	70.05±**3.10**	0.018

more on the shape similarity of the segmentation target. It is important to note that the shape and size of liver tumors vary greatly among patients as well as in the same patient at different times compared to the liver, which make it more difficult to achieve fully automatic segmentation of liver tumors. Therefore, as recorded in Tables III–V, the ASD and RMSD values for liver tumors are obviously larger than the values for livers.

Fig. 10 shows the segmentation results from different approaches. First, from the segmentation results obtained by 2-D networks, both U-Net and U-Net++ fail to identify large liver tumors but CE-Net is successful in the first column of results. U-Net++ obtains poorer segmentation results than U-Net and CE-Net for small liver tumor as shown in the second column of results. It is clear that CE-Net is able to recognize a larger range of liver tumors due to the employment of SPP module with multiscale receptive fields. In the third column, both U-net and CE-Net obtain more false liver areas, but U-Net++ shows better performance for large liver target recognition because it has stronger generalization capability and more dense feature representation. In the fourth column of Fig. 10, U-Net and U-Net++ are inaccurate in identifying liver boundaries, while CE-Net, which uses multiple atrous convolutional parallel modules, provides higher accuracy for

TABLE VI
COMPARISON OF THE EFFICIENCIES OF DIFFERENT NETWORKS. THE FIRST TWO BEST VALUES ARE IN BOLD

Network	Operations(GFLOPS)	Training Parameters	ModelSize (MB)
U-Net [28]	**123.96**	**13,394,242**	**51.15**
U-Net++ [36]	**25.90**	**9,041,700**	**35.34**
CE-Net [45]	35.78	29,003,668	110.77
3D U-Net [39]	1032.80	16,320,322	62.27
V-Net [40]	516.12	65,173,903	248.69
DefED-Net (without DSC)	547.81	35,580,976	139.19
DefED-Net	222.44	14,529,959	56.96

liver boundary detection. Second, it is well known that 3-D networks can provide better segmentation results of liver and liver tumors than 2-D networks as they can capture the temporal information of volumetic data. In both the first and second columns, it can be seen that the tumor boundaries obtained by the 3-D network are clearer than results provided by 2-D networks, and the tumor boundaries obtained by V-Net are clearer and more accurate than results obtained by 3-D U-Net due to the utilization of feature extraction block with residual connection. In the third column, 3-D networks are clearly superior to 2-D networks since the former do not suffer form the problem of over-detection. Finally, it is evident from the results obtained by DefED-Net in the first and second columns that they provide more accurate segmentations of both liver and liver tumors than the above mentioned 2-D and 3-D networks. In addition, in the third and fourth columns, the DefED-Net focuses on the relevant liver region while suppressing the influence of surrounding organs, it thus provides smoother segmentation boundaries than comparative approaches. In general, Fig. 10 shows that the DefED-Net achieves better feature encoding and context information extraction, which is helpful for improving the segmentation accuracy of liver and liver tumor.

E. Model-Size Comparison

We also counted the number of training parameters and computational costs of networks as shown in Table VI. Compared with 2-D networks, 3-D networks require much more memory and higher computational cost due to the employment of 3-D convolutional kernels. The number of parameters of V-Net is greatly larger than one of 3-D U-Net since V-Net uses a deeper network structure than 3DU-Net, uses more convolutions and uses residual connections. On the efficiency of models, the DefED-Net is similar to U-Net.

In fact, the DefED-Net adds Ladder-ASPP block compared to U-Net. The Ladder-ASPP is a densely connected block, and thus it shows high computational complexity and requires a large of parameters as shown in Table VI. In this article, we utilize depth separable convolution to decouple the operation of spatial-dimension and channel-dimension, which efficiently reduces the number of parameters. Thus, the added Ladder-ASPP is a very small block compared to the size of U-Net. Finally, the DefED-Net achieves excellent liver and liver tumor segmentation with low computational cost.

IV. CONCLUSION

Liver and liver tumor segmentations attract attentions of many researchers due to their importance in medical image analysis. Deep convolutional neural networks, especially U-Nets, are very useful and popular for liver and liver tumor segmentations. For improved CNNs, DC is very important for the capture of context information, but received little consideration in liver segmentation. In this article, we have introduced DC into U-Nets to achieve better feature encoding. Furthermore, although ASPP is effective for improving the context information, atrous convolution and pooling lead to the loss of detail information. We have suggested the Ladder-ASPP for feature encoding and fusion. The Ladder-ASPP is superior to ASPP due to the dense connection and atrous convolution with variable dilation rate. Finally, the proposed DefED-Net provides the best liver segmentation results without increasing the size of models. Our studies also show that utilization of spatial information is more important than feature fusion via modifying the network architecture for liver and liver tumor segmentations. Experiments demonstrate the advantages of the proposed DefED-Net on improving segmentation accuracies and reducing model-size for liver and liver tumor segmentations.

REFERENCES

[1] M. Seehawer *et al.*, "Necroptosis microenvironment directs lineage commitment in liver cancer," *Nature*, vol. 562, no. 7725, pp. 69–75, Oct. 2018.
[2] A. B. Ryerson *et al.*, "Annual report to the nation on the status of cancer, 1975–R2012, featuring the increasing incidence of liver cancer," *Cancer*, vol. 122, no. 9, pp. 1312–1337, May 2016.
[3] G. Li, X. Chen, F. Shi, W. Zhu, J. Tian, and D. Xiang, "Automatic liver segmentation based on shape constraints and deformable graph cut in CT images," *IEEE Trans. Image Process.*, vol. 24, pp. 5315–5329, Dec. 2015.
[4] D. Furukawa, A. Shimizu, and H. Kobatake, "Automatic liver segmentation method based on maximum a posterior probability estimation and level set method," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, 2017, pp. 117–124.
[5] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.
[6] D. Wong *et al.*, "A semi-knowledge method for liver tumor segmentation based on 2D region growing with knowledge-based constraints," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, 2008, pp. 1–10.
[7] C. Li *et al.*, "A likelihood and local constraint level set model for liver tumor segmentation from CT volumes," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 10, pp. 2967–2977, Oct. 2013.
[8] G. Chartrand, T. Cresson, R. Chav, A. Gotra, A. Tang, and J. A. De Guise, "Liver segmentation on CT and MR using Laplacian mesh optimization," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 9, pp. 2110–2121, Sep. 2017.
[9] A. Saito, S. Nawano, and A. Shimizu, "Joint optimization of segmentation and shape prior from level-set-based statistical shape model, and its application to the automated segmentation of abdominal organs," *Med. Image Anal.*, vol. 28, no. 33, pp. 46–65, Feb. 2016.
[10] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222–1239, Nov. 2001.
[11] S. Vicente, V. Kolmogorov, and C. Rother, "Graph cut based image segmentation with connectivity priors," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Anchorage, AK, USA, 2018, pp. 1–8.
[12] S. Paris, F. X. Sillion, and L. Quan, "A surface reconstruction method using global graph cut optimization," *Int. J. Comput. Vis.*, vol. 66, no. 2, pp. 141–161, Jan. 2006.
[13] J. Stawiaski, E. Decencière, and F. Bidault, "Interactive liver tumor segmentation using graph-cuts and watershed," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, 2008, pp. 1–12.

[14] E. Wisse, "An electron microscopic study of the fenestrated endothelial lining of rat liver sinusoids," *J. Ultrastruct. Res.*, vol. 31, nos. 1–2, pp. 125–150, May 1970.

[15] R. Kéchichian, S. Valette, and M. Desvignes, "Automatic multiorgan segmentation via multiscale registration and graph cut," *IEEE Trans. Med. Imag.*, vol. 37, no. 12, pp. 2739–2749, Dec. 2018.

[16] A. Khadidos, V. Sanchez, and C.-T. Li, "Weighted level set evolution based on local edge features for medical image segmentation," *IEEE Trans. Image Process.*, vol. 26, pp. 1979–1991, 2017.

[17] G. Chen, L. Gu, L. Qian, and J. Xu, "An improved level set for liver segmentation and perfusion analysis in MRIs," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 1, pp. 94–103, Jan. 2009.

[18] C. Li, R. Huang, Z. Ding, J. C. Gatenby, D. N. Metaxas, and J. C. Gore "A level set method for image segmentation in the presence of intensity inhomogeneities with application to MRI," *IEEE Trans. Image Process.*, vol. 20, pp. 2007–2016, 2011.

[19] B. Wang, X. Yuan, X. Gao, X. Li, and D. Tao, "A hybrid level set with semantic shape constraint for object segmentation," *IEEE Trans. Cybern.*, vol. 49, no. 5, pp. 1558–1569, May 2019.

[20] D. Shen, Y. Zhan, and C. Davatzikos, "Segmentation of prostate boundaries from ultrasound images using statistical shape model," *IEEE Trans. Med. Imag.*, vol. 22, no. 4, pp. 539–551, Apr. 2003.

[21] D. F. Polan, S. L. Brady, and R. A. Kaufman, "Tissue segmentation of computed tomography images using a Random Forest algorithm: A feasibility study," *Phys. Med. Biol.*, vol. 61, no. 17, pp. 6553–6569, Aug. 2016.

[22] E. Vorontsov, N. Abi-Jaoudeh, and S. Kadoury, "Metastatic liver tumor segmentation using texture-based omni-directional deformable surface models," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, 2014, pp. 74–83.

[23] S. Tomoshige, E. Oost, A. Shimizu, H. Watanabe, H. Kobatake, and S. Nawano, "Relaxed conditional statistical shape models and their application to non-contrast liver segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, 2012, pp. 126–136.

[24] X. Zhang, J. Tian, K. Deng, Y. Wu, and X. Li, "Automatic liver segmentation using a statistical shape model with optimal surface detection," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 10, pp. 2622–2626, Oct. 2010.

[25] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006.

[26] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.

[27] D. Nie, L. Wang, Y. Gao, and D. Shen, "Fully convolutional networks for multi-modality isointense infant brain image segmentation," in *Proc. IEEE Conf. Int. Symp. Biomed. Imag.*, 2016, pp. 1342–1345.

[28] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, 2015, pp. 234–241.

[29] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014. [Online]. Available: arXiv:1409.1556

[30] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Las Vegas, NV, USA, 2016, pp. 770–778.

[31] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recogn.*, Honolulu, HI, USA, 2017, pp. 2261–2269.

[32] N. Abraham and N. M. Khan, "A novel focal tversky loss function with improved attention U-Net for lesion segmentation," in *Proc. IEEE 16th Conf. Int. Symp. Biomed. Imag.*, Venice, Italy, 2019, pp. 683–687.

[33] O. Oktay *et al.*, "Attention U-Net: Learning where to look for the pancreas," 2018. [Online]. Available: arXiv:1804.03999

[34] Q. Jin, Z. Meng, C. Sun, L. Wei, and R. Su, "RA-UNet: A hybrid deep attention-aware network to extract liver and tumor in CT scans," 2018. [Online]. Available: arXiv:1811.01328

[35] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, "Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation," 2018. [Online]. Available: arXiv:1802.06955

[36] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: Redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Trans. Med. Imag.*, vol. 39, no. 6, pp. 1856–1867, Jun. 2020, doi: 10.1109/TMI.2019.2959609.

[37] 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 Trans. Med. Imag.*, vol. 39, no. 5, pp. 1316–1325, May 2020, doi: 10.1109/TMI.2019.2948320.

[38] S. Ji, W. Xu, M. Yang, and K. Yu, "3D convolutional neural networks for human action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 221–231, Jan. 2013.

[39] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, 2016, pp. 424–432.

[40] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: fully convolutional neural networks for volumetric medical image segmentation," in *Proc. Int. Conf. 3D Vis.*, Stanford, CA, USA, 2016, pp. 565–671.

[41] S. Chen, K. Ma, and Y. Zheng, "Med3D: Transfer learning for 3D medical image analysis," 2019. [Online]. Available: arXiv:1904.00625

[42] Q. Dou *et al.*, "3D deeply supervised network for automated segmentation of volumetric medical images," *Med. Image Anal.*, vol. 41, pp. 40–54, Oct. 2017.

[43] T. Lei, W. Zhou, Y. Zhang, R. Wang, H. Meng, and A. K. Nandi, "Lightweight V-Net for liver segmentation," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Barcelona, Spain, 2020, pp. 1379–1383.

[44] 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 Trans. Med. Imag.*, vol. 37, no. 12, pp. 2663–2674, Dec. 2018.

[45] Z. Gu *et al.*, "CE-Net: Context encoder network for 2D medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 38, no. 10, pp. 2281–2292, Oct. 2019.

[46] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Honolulu, HI, USA, 2017, pp. 2881–2890.

[47] J. Dai *et al.*, "Deformable convolutional networks," in *Proc. IEEE Conf. Comput. Vis.*, 2017, pp. 764–773.

[48] T. Lei, X. Jia, Y. Zhang, S. Liu, H. Meng and A. K. Nandi, "Superpixel-based fast fuzzy C-means clustering for color image segmentation," *IEEE Trans. Fuzzy Syst.*, vol. 27, no. 9, pp. 1753–1766, Sep. 2019.

[49] P. Bilic *et al.*, "The liver tumor segmentation benchmark (LiTS)," 2019. [Online]. Available: arXiv:1901.04056

[50] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, and C. Xu, "GhostNet: More features from cheap operations," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 1580–1589.

[51] Q. Shao, L. Gong, K. Ma, H. Liu, and Y. Zheng, "Attentive CT lesion detection using deep pyramid inference with multi-scale booster," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, 2019, pp. 301–309.

[52] A. G. Howard *et al.*, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017. [Online]. Available: arXiv:1704.04861

[53] M. H. Vu, G. Grimbergen, T. Nyholm, and T. Löfstedt, "Evaluation of multi-slice inputs to convolutional neural networks for medical image segmentation," *Med. Phys.*, vol. 47, no. 12, pp. 6216–6231, 2020.

[54] S. Ren, P. Laub, Y. Lu, M. Naganawa, and R. E. Carson, "Atlas-based multiorgan segmentation for dynamic abdominal PET," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 4, no. 1, pp. 50–62, Jan. 2020.

[55] Z. Guo, X. Li, H. Huang, N. Guo, and Q. Li, "Deep Learning-Based image segmentation on multimodal medical imaging," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 3, no. 2, pp. 162–169, Mar. 2019.

[56] C. De Sio, J. J. Velthuis, L. Beck, J. L. Pritchard and R. P. Hugtenburg, "r-UNet: Leaf position reconstruction in upstream radiotherapy verification," *IEEE Trans. Radiat. Plasma Med. Sci.*, early access, May 15, 2020, doi: 10.1109/TRPMS.2020.2994648.

[57] P.-H. Conze, F. Rousseau, V. Noblet, F. Heitz, R. Memeo, and P. Pessaux, "Semi-automatic liver tumor segmentation in dynamic contrast-enhanced CT scans using random forests and supervoxels," in *Proc. Int. Workshop Mach. Learn. Med. Imag.*, 2015, pp. 212–219.

[58] P.-H. Conze *et al.*, "Scale-adaptive supervoxel-based random forests for liver tumor segmentation in dynamic contrast-enhanced CT scans," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 12, no. 2, pp. 223–233, 2017.

[59] Y. Li, S. Hara, and K. Shimura, "A machine learning approach for locating boundaries of liver tumors in CT images," in *Proc. 18th Int. Conf. Pattern Recognit.*, Hong Kong, China, 2006, pp. 400–403.

[60] P.-H. Conze *et al.*, "Abdominal multi-organ segmentation with cascaded convolutional and adversarial deep networks," 2020. [Online]. Available: arXiv:2001.09521

[61] X. Xiao, S. Lian, Z. Luo, and S. Li, "Weighted res-UNet for high-quality retina vessel segmentation," in *Proc. Int. Conf. Inf. Technol. Med. Educ. (ITME)*, Hangzhou, China, 2018, pp. 327–331.

[62] M. Sun, G. Zhang, H. Dang, X. Qi, X. Zhou, and Q. Chang, "Accurate gastric cancer segmentation in digital pathology images using deformable convolution and multi-scale embedding networks," *IEEE Access*, vol. 7, pp. 75530–75541, 2019.