H-DenseUNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation From CT Volumes

Xiaomeng Li, Hao Chen, *Member, IEEE*, Xiaojuan Qi, Qi Dou, *Student Member, IEEE*, Chi-Wing Fu, *Member, IEEE*, and Pheng-Ann Heng, *Senior Member, IEEE*

***Abstract*— Liver cancer is one of the leading causes of cancer death. To assist doctors in hepatocellular carcinoma diagnosis and treatment planning, an accurate and automatic liver and tumor segmentation method is highly demanded in the clinical practice. Recently, fully convolutional neural networks (FCNs), including 2-D and 3-D FCNs, serve as the backbone in many volumetric image segmentation. However, 2-D convolutions cannot fully leverage the spatial information along the third dimension while 3-D convolutions suffer from high computational cost and GPU memory consumption. To address these issues, we propose a novel hybrid densely connected UNet (H-DenseUNet), which consists of a 2-D DenseUNet for efficiently extracting intra-slice features and a 3-D counterpart for hierarchically aggregating volumetric contexts under the spirit of the auto-context algorithm for liver and tumor segmentation. We formulate the learning process of the H-DenseUNet in an end-to-end manner, where the intra-slice representations and inter-slice features can be jointly optimized through a hybrid feature fusion layer. We extensively evaluated our method on the data set of the MICCAI 2017 Liver Tumor Segmentation Challenge and 3DIRCADb data set. Our method outperformed other state-of-the-arts on the segmentation results of tumors and achieved very competitive performance for liver segmentation even with a single model.**

***Index Terms*— CT, liver tumor segmentation, deep learning, hybrid features.**

I. Introduction

LIVER cancer is one of the most common cancer diseases in the world and causes massive deaths every year [1], [2]. The accurate measurements from CT, including tumor volume, shape, location and further functional liver volume, can assist doctors in making accurate hepatocellular carcinoma evaluation and treatment planning. Traditionally, the liver and liver lesion are delineated by radiologists on a slice-by-slice basis, which is time-consuming and prone to inter- and intra-rater variations. Therefore, automatic liver and liver tumor segmentation methods are highly demanded in the clinical practice.

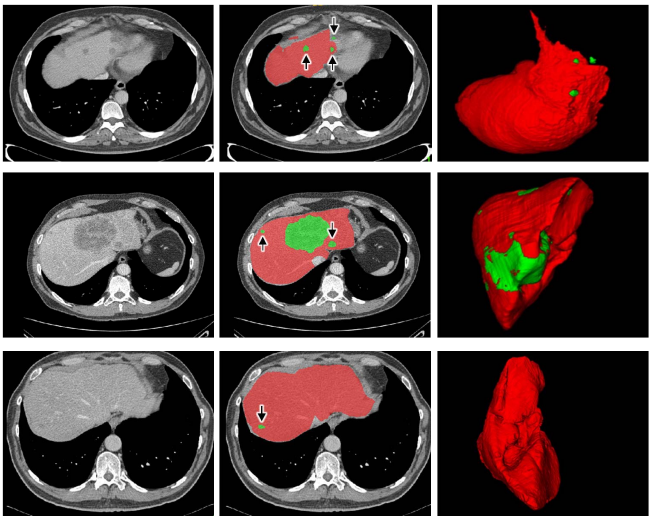

Fig. 1. Examples of contrast-enhanced CT scans showing the large variations of shape, size, location of liver lesion. Each row shows a CT scan acquired from individual patient. The *red* regions denote the liver while the *green* ones denote the lesions (see the black arrows above).

Automatic liver segmentation from the contrast-enhanced CT volumes is a very challenging task due to the low intensity contrast between the liver and other neighboring organs (see the first row in Figure 1). Moreover, radiologists usually enhance CT scans by an injection protocol for clearly observing tumors, which may increase the noise inside the images on the liver region [3]. Compared with liver segmentation, liver tumor segmentation is considered to be a more challenging task. First, the liver tumor has various size, shape, location and numbers within one patient, which hinders the automatic segmentation, as shown in Figure 1. Second, some lesions do not have clear boundaries, limiting the performance of solely

Manuscript received May 7, 2018; accepted June 2, 2018. Date of publication June 11, 2018; date of current version November 29, 2018. This work was supported by the Research Grants Council of the Hong Kong Special Administrative Region under Project GRF 14202514 and Project GRF 14203115. *(Corresponding author: Hao Chen.)*

X. Li, X. Qi, Q. Dou, C.-W. Fu, and P.-A. Heng are with the Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong (e-mail: xmli@cse.cuhk.edu.hk).

H. Chen is with the Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, and also with Imsight Medical Technology, Inc., Shenzhen 518000, China (e-mail: hchen@cse.cuhk.edu.hk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMI.2018.2845918

edge based segmentation methods (see the lesions in the third row of Figure 1). Third, many CT scans consist of anisotropic dimensions with high variations along the z-axis direction (the voxel spacing ranges from 0.45mm to 6.0mm), which further poses challenges for automatic segmentation methods.

To tackle these difficulties, many segmentation methods have been proposed, including intensity thresholding, region growing, and deformable models. These methods, however, rely on hand-crafted features, and have limited feature representation capability. Recently, fully convolutional neural networks (FCNs) have achieved great success on a broad array of recognition problems [4]–[14]. Many researchers advance this stream using deep learning methods in the liver and tumor segmentation problem and the literature can be classified into two categories broadly. (1) 2D FCNs, such as UNet architecture [15], the multi-channel FCN [16], and the FCN based on VGG-16 [17]. (2) 3D FCNs, where 2D convolutions are replaced by 3D convolutions with volumetric data input [18], [19].

In the clinical diagnosis, the experienced radiologist usually observes and segments tumors according to many adjacent slices along the z-axis. However, 2D FCN based methods ignore the contexts on the z-axis, which would lead to limited segmentation accuracy. To be specific, single or three adjacent slices cropped from volumetric images are fed into 2D FCNs [16], [17] and the 3D segmentation volume is generated by simply stacking the 2D segmentation maps. Although adjacent slices are employed, it is still not enough to probe the spatial information along the third dimension, which may degrade the segmentation performance. To solve this problem, some researchers proposed to use tri-planar schemes or RNN to probe the 3D contexts [4], [20], [21]. For example, Prasoon *et al.* [4] applied three 2D FCNs on orthogonal planes (e.g., the xy, yz, and xz planes) and voxel prediction results are generated by the average of these probabilities. Compared to 2D FCNs, 3D FCNs suffer from high computational cost and GPU memory consumption. The high memory consumption limits the depth of the network as well as the filter's field-of-view, which are the two key factors for performance gains [22]. The heavy computation of 3D convolutions also impedes the application in training a large-scale dataset. Moreover, many researchers have demonstrated the effectiveness of knowledge transfer (the knowledge learnt from one source domain efficiently transferred to another domain) for boosting the performance [23], [24]. Unfortunately, only a dearth of 3D pre-trained model exists, which restricts the performance and also the adoption of 3D FCNs.

To address the above problems, we proposed a novel end-to-end system, called hybrid densely connected UNet (H-DenseUNet), where intra-slice features and 3D contexts are effectively probed and jointly optimized for accurate liver and lesion segmentation. Our H-DenseUNet has the following two technical achievements:

A. Deep and Efficient Network

First, to fully extract high-level intra-slice features, we design a very deep and efficient network based on the pre-defined design principles by 2D convolutions, called 2D DenseUNet, where the advantages of both densely connected path [25] and UNet connections [5] are fused together. Densely connected path is derived from densely connected network (DenseNet), where the improved information flow and parameters efficiency alleviate the difficulty for training the deep network. Different from DenseNet [25], we add the UNet connections, i.e., long-range skip connections, between the encoding part and the decoding part in our architecture; hence, the network can enable low-level spatial feature preservation for better intra-slice context exploration.

B. Hybrid Feature Exploration

Second, to explore the volumetric feature representation, we design an end-to-end training system, called H-DenseUNet, where intra-slice and inter-slice features are effectively extracted and then jointly optimized through the hybrid feature fusion (HFF) layer. Specifically, 3D DenseUNet is integrated with the 2D DenseUNet by the way of auto-context [26] mechanism, which is a general form of stacked generality [27]. With the guidance of semantic probabilities from 2D DenseUNet, the optimization burden in the 3D DenseUNet can be well alleviated, which contributes to the training efficiency for 3D contexts extraction. Moreover, with the end-to-end system, the hybrid feature, consisting of volumetric features and the high-level representative intra-slice features, can be automatically fused and jointly optimized together for better liver and tumor recognition. In summary, this work has the following achievements:

- We design a DenseUNet to effectively probe hierarchical intra-slice features for liver and tumor segmentation, where the densely connected path and UNet connections are carefully integrated based on pre-defined design principles to improve the liver tumor segmentation performance.
- We propose a H-DenseUNet framework to explore hybrid (intra-slice and inter-slice) features for liver and tumor segmentation. The hybrid feature learning architecture well sidesteps the problems that 2D networks neglect the volumetric contexts and 3D networks suffer from heavy computational cost, and can be served as a new paradigm for effectively exploiting 3D contexts.
- Our method ranked the 1st on lesion segmentation, achieved very competitive performance on liver segmentation in the 2017 LiTS Leaderboard, and also achieved the state-of-the-art results on the 3DIRCADb Dataset.

II. Related Work

A. Hand-Crafted Feature Based Methods

In the past decades, a lot of algorithms, including thresholding [28], [29], region growing, deformable model based methods [30], [31] and machine learning based methods [32]–[36] have been proposed to segment liver and liver tumor. Threshold-based methods classified foreground and background according to whether the intensity value is above a threshold. Variations of region growing algorithms were also popular in the liver and lesion segmentation task. For example, Wong *et al.* [30] segmented tumors by a 2D region growing

method with knowledge-based constraints. Level set methods also attracted attentions from researchers with the advantages of numerical computations involving curves and surfaces [37]. For example, Jimenez-Carretero *et al.* [31] proposed to classify tumors by a multi-resolution 3D level set method coupled with adaptive curvature technique. A large variety of machine learning based methods have also been proposed for liver tumor segmentation. For example, Huang *et al.* [32] proposed to employ the random feature subspace ensemble-based extreme learning machine (ELM) for liver lesion segmentation. Vorontsov *et al.* [33] proposed to segment tumors by support vector machine (SVM) classifier and then refined the results by the omnidirectional deformable surface model. Similarly, Kuo *et al.* [35] proposed to learn SVM classifier with texture feature vector for liver tumor segmentation. Le *et al.* [34] employed the fast marching algorithm to generate initial regions and then classified tumors by training a noniterative single hidden layer feedforward network (SLFN). To speed up the segmentation algorithm, Chaieb *et al.* [38] adopted a bootstrap sampling approach for efficient liver tumor segmentation.

B. Deep Learning Based Methods

Convolutional neural networks (CNNs) have achieved great success in many object recognition problems in computer vision community. Many researchers followed this trend and proposed to utilize various CNNs for learning feature representations in the application of liver and lesion segmentation. For example, Ben-Cohen *et al.* [17] proposed to use a FCN for liver segmentation and liver-metastasis detection in CT examinations. Christ *et al.* [15], [39] proposed a cascaded FCN architecture and dense 3D conditional random fields (CRFs) to automatically segment liver and liver lesions. In the meanwhile, Sun *et al.* [16] designed a multi-channel FCN to segment liver tumors from CT images, where the probability maps were generated by the feature fusion from different channels.

Recently, during the 2017 ISBI LiTS challenge, Han [40], proposed a 2.5D 24-layer FCN model to segment liver tumors, where the residual block was employed as the repetitive building blocks and the UNet connection was designed across the encoding part and decoding part. 2.5D refers to using 2D convolutional neural network with the input of adjacent slices from the volumetric images. Both Vorontsov *et al.* [41] and Chlebus *et al.* [42] achieved the second place in the ISBI challenge. Vorontsov *et al.* [41] also employed ResNet-like residual blocks and UNet connections with 21 convolutional layers, which is a bit shallower and has fewer parameters compared to that proposed by Han [40]. Chlebus *et al.* [42] designed a 28-layer UNet architecture in two individual models and subsequently filtered the false positives of tumor segmentation results by a random forest classifier. Instead of using 3D FCNs, all of the top results employed 2D FCNs with different network depths, showing the efficacy of 2D FCNs regarding the underlying volumetric segmentation problem. However, all these networks are shallow and ignore the 3D contexts, which limit the high-level feature extraction capability and restrict the recognition performance.

III. Method

Figure 2 shows the pipeline of our proposed method for liver and tumor segmentation. We employed the cascaded learning strategy to reduce the overall computation time, which has also been adopted in many recognition tasks [43]–[46]. First, a simple ResNet architecture [40] is trained to get a quick but coarse segmentation of liver. With the region of interest (ROI), our proposed H-DenseUNet efficiently probes intra-slice and inter-slice features through a 2D DenseUNet f_{2d} and a 3D counterpart f_{3d}, followed by jointly optimizing the hybrid features in the hybrid feature fusion (HFF) layer for accurate liver and lesion segmentation.

A. Deep 2D DenseUNet for Intra-Slice Feature Extraction

The intra-slice feature extraction part follows the structure of DenseNet-161 [25], which is composed of repetitive densely connected building blocks with different output dimensions. In each densely connected building block, there are direct connections from any layer to all subsequent layers, as shown in Figure 2(c). Each layer produces k feature maps and k is called *growth rate*. One advantage of the dense connectivity between layers is that it has fewer output dimensions than traditional networks, avoiding learning redundant features. Moreover, the densely connected path ensures the maximum information flow between layers, which improves the gradient flow, and thus alleviates the burden in searching for the optimal solution in a very deep neural network.

However, the original DenseNet-161 [25] is designed for the object classification task while our problem belongs to the segmentation topics. Moreover, a deep FCN network for segmentation tasks actually contains several max-pooling and upsampling operations, which may lead to the information loss of low-level (i.e., high resolution) features. Given above two considerations, we develop a 2D DenseUNet, which inherits both advantages of densely connected path and UNet-like connections [5]. Specifically, the dense connection between layers is employed within each micro-block to ensure the maximum information flow while the UNet long range connection links the encoding part and the decoding part to preserve low-level information.

Let $\mathbf{I} \in R^{n\times 224\times 224\times 12\times 1}$ denote the input training samples (for $224 \times 224 \times 12$ input volumes) with ground-truth labels $\mathbf{Y} \in R^{n\times 224\times 224\times 12\times 1}$, where n denotes the batch size of the input training samples and the last dimension denotes the channel. $\mathbf{Y}_{i,j,k} = c$ since each pixel (i, j, k) is tagged with class c (background, liver and tumor). Let function $\mathscr{F}$ denote the transformation from the volumetric data to three adjacent slices. Specifically, every three adjacent slices along z-axis are stacked together and the number of groups can be transformed to the batch dimension. For example, $\mathbf{I_{2d}} = \mathscr{F}(\mathbf{I})$, where $\mathbf{I_{2d}} \in R^{12n\times 224\times 224\times 3}$ denotes the input samples of 2D DenseUNet. The detailed transformation process is illustrated in Figure 2(d). Because of the transformation, the 2D and 3D DenseUNet can be jointly trained, which will be described in detail in section B. For convenience, $\mathscr{F}^{-1}$ denotes the inverse transformation from three adjacent slices to the volumetric data. The 2D DenseUNet conducts

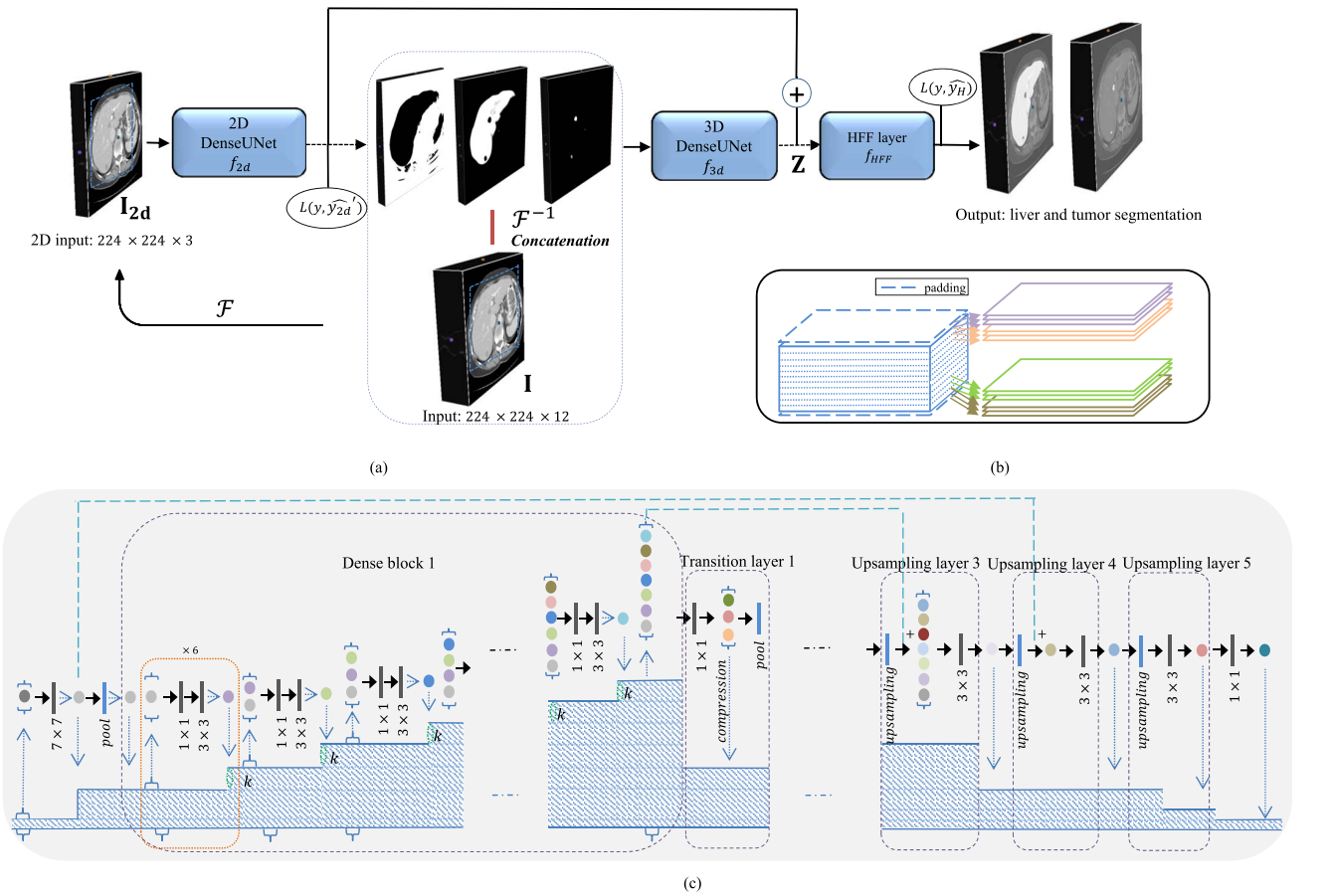

Fig. 2. The illustration of the pipeline for liver and lesion segmentation. Each 3D input volume is sliced into adjacent slices through transformation process $\mathcal{F}$ and then fed into 2D DenseUNet; Concatenated with the prediction volumes from 2D network, the 3D input volumes are fed into the 3D network for learning inter-slice features; Then, the HFF layer fused and optimized the intra-slice and inter-slice features for accurate liver and tumor segmentation. (a) The structure of H-DenseUNet, including the 2D DenseUNet and the 3D counterpart. (b) The transformation of the volumetric data to three adjacent slices. (c) The network structure of the 2D DenseUNet. The structure in the orange block is a micro-block and k denotes the growth-rate. (Best viewed in color).

liver and tumor segmentation,

$$\begin{aligned} \mathbf{X_{2d}} &= f_{2d}(\mathbf{I_{2d}}; \theta_{2d}), \quad \mathbf{X_{2d}} \in R^{12n \times 224 \times 224 \times 64}, \\ \hat{y_{2d}} &= f_{2dcls}(\mathbf{X_{2d}}; \theta_{2dcls}), \quad \hat{y_{2d}} \in R^{12n \times 224 \times 224 \times 3} \end{aligned} \tag{1}$$

where $\mathbf{X_{2d}}$ is the feature map from layer "upsampling layer 5" (see Table I) and $\hat{y_{2d}}$ is the corresponding pixel-wise probabilities for input $\mathbf{I_{2d}}$.

The illustration and detailed structure of 2D DenseUNet are shown in Figure 2(c) and Table I, respectively. The depth of 2D DenseUNet is extended to 167 layers, referred as 2D DenseUNet-167, which consists of 167 convolution layers, pooling layers, dense blocks, transition layers and upsampling layers. The dense block denotes the cascade of several micro-blocks, in which all layers are directly connected, see Figure 2(c). To change the size of feature-maps, the transition layer is employed, which consists of a batch normalization layer and a 1×1 convolution layer followed by an average pooling layer. A compression factor is included in the transition layer to compress the number of feature-maps, preventing the expanding of feature-maps (set as 0.5 in our experiments). The upsampling layer is implemented by the bilinear interpolation, followed by the summation with low-level features (i.e., UNet connections) and a 3×3 convolutional layer. Before each convolution layer, the batch normalization and the Rectified Linear Unit (ReLU) are employed in the architecture.

B. H-DenseUNet for Hybrid Feature Exploration

2D DenseUNet with deep convolutions can produce high-level representative in-plane features but neglect the spatial information along the z dimension while 3D DenseUNet has large GPU computational cost and limited kernel's field-of-view as well as the network depth. To address these issues, we propose H-DenseUNet to jointly fuse and optimize the learned intra-slice and inter-slice features for better liver tumor segmentation.

To fuse hybrid features from the 2D and 3D network, the feature volume size should be aligned. Therefore, the feature maps and score maps from 2D DenseUNet are transformed to the volumetric shape as follows:

$$\begin{aligned} \mathbf{X_{2d}}' &= \mathscr{F}^{-1}(\mathbf{X_{2d}}), \mathbf{X_{2d}}' \in R^{n \times 224 \times 224 \times 12 \times 64}, \\ \hat{y_{2d}}' &= \mathscr{F}^{-1}(\hat{y_{2d}}), \hat{y_{2d}}' \in R^{n \times 224 \times 224 \times 12 \times 3}, \end{aligned} \tag{2}$$

TABLE I

ARCHITECTURES OF THE PROPOSED H-DENSEUNET, CONSISTING OF THE 2D DENSEUNET AND THE 3D COUNTERPART. THE SYMBOL –[ ] DENOTES THE LONG RANGE UNET SUMMATION CONNECTIONS WITH THE LAST LAYER OF THE DENSE BLOCK. THE SECOND AND FORTH COLUMN INDICATE THE OUTPUT SIZE OF THE CURRENT STAGE IN TWO ARCHITECTURES, RESPECTIVELY. NOTE THAT "1 × 1, 192 CONV" CORRESPONDS TO THE SEQUENCE BN-RELU-CONV LAYER WITH CONVOLUTIONAL KERNEL SIZE OF 1 × 1 AND 192 FEATURES. "[ ]×*d*" REPRESENTS THE DENSE BLOCK IS REPEATED FOR *d* TIMES

	Feature size	2D DenseUNet-167 (k=48)	Feature size	3D DenseUNet-65 (k=32)
input	224 × 224	-	224 × 224 × 12	-
convolution 1	112 × 112	7 × 7, 96, stride 2	112 × 112 × 6	7 × 7 × 7, 96, stride 2
pooling	56 × 56	3 × 3 max pool, stride 2	56 × 56 × 3	3 × 3 × 3 max pool, stride 2
dense block 1	56 × 56	$\begin{bmatrix} 1\times1, 192 & \text{conv} \\ 3\times3, 48 & \text{conv} \end{bmatrix} \times 6$	56 × 56 × 3	$\begin{bmatrix} 1\times1\times1, 128 & \text{conv} \\ 3\times3\times3, 32 & \text{conv} \end{bmatrix} \times 3$
transition layer 1	56 × 56	1 × 1 conv	56 × 56 × 3	1 × 1 × 1 conv
	28 × 28	2 × 2 average pool	28 × 28 × 3	2 × 2 × 1 average pool
dense block 2	28 × 28	$\begin{bmatrix} 1\times1, 192 & \text{conv} \\ 3\times3, 48 & \text{conv} \end{bmatrix} \times 12$	28 × 28 × 3	$\begin{bmatrix} 1\times1\times1, 128 & \text{conv} \\ 3\times3\times3, 32 & \text{conv} \end{bmatrix} \times 4$
transition layer 2	28 × 28	1 × 1 conv	28 × 28 × 3	1 × 1 × 1 conv
	14 × 14	2 × 2 average pool,	14 × 14 × 3	2 × 2 × 1 average pool
dense block 3	14 × 14	$\begin{bmatrix} 1\times1, 192 & \text{conv} \\ 3\times3, 48 & \text{conv} \end{bmatrix} \times 36$	14 × 14 × 3	$\begin{bmatrix} 1\times1\times1, 128 & \text{conv} \\ 3\times3\times3, 32 & \text{conv} \end{bmatrix} \times 12$
transition layer 3	14 × 14	1 × 1 conv	14 × 14 × 3	1 × 1 × 1 conv
	7 × 7	2 × 2 average pool	7 × 7 × 3	2 × 2 × 1 average pool
dense block 4	7 × 7	$\begin{bmatrix} 1\times1, 192 & \text{conv} \\ 3\times3, 48 & \text{conv} \end{bmatrix} \times 24$	7 × 7 × 3	$\begin{bmatrix} 1\times1\times1, 128 & \text{conv} \\ 3\times3\times3, 32 & \text{conv} \end{bmatrix} \times 8$
upsampling layer 1	14 × 14	2 × 2 upsampling – [dense block 3], 768, conv	14 × 14 × 3	2 × 2 × 1 upsampling – [dense block 3], 504, conv
upsampling layer 2	28 × 28	2 × 2 upsampling – [dense block 2], 384, conv	28 × 28 × 3	2 × 2 × 1 upsampling – [dense block 2], 224, conv
upsampling layer 3	56 × 56	2 × 2 upsampling – [dense block 1], 96, conv	56 × 56 × 3	2 × 2 × 1 upsampling – [dense block 1], 192, conv
upsampling layer 4	112 × 112	2 × 2 upsampling – [convolution 1], 96, conv	112 × 112 × 6	2 × 2 × 2 upsampling – [convolution 1], 96, conv
upsampling layer 5	224 × 224	2 × 2 upsampling, 64, conv	224 × 224 × 12	2 × 2 × 2 upsampling, 64, conv
convolution 2	224 × 224	1 × 1, 3	224 × 224 × 12	1 × 1 × 1, 3

Then the 3D DenseUNet distill the visual features with 3D contexts by concatenating the original volumes $\mathbf{I}$ with the contextual information $\hat{y_{2d}}'$ from the 2D network. Specifically, the detectors in the 3D counterpart trained based not only on the features probed from the original images, but also on the probabilities of a large number of context pixels from 2D DenseUNet. With the guidance from the supporting contexts pixels, the burden in searching for the optimal solution in the 3D counterpart has also been well alleviated, which significantly improves the learning efficiency of the 3D network. The learning process of 3D DenseUNet can be described as:

$$\begin{aligned} \mathbf{X_{3d}} &= f_{3d}(\mathbf{I}, \hat{y_{2d}}'; \theta_{3d}), \\ \mathbf{Z} &= \mathbf{X_{3d}} + \mathbf{X_{2d}}', \end{aligned} \tag{3}$$

where $\mathbf{X_{3d}}$ denotes the feature volume from layer "*upsampling layer 5*" in 3D DenseUNet-65. $\mathbf{Z}$ denotes the hybrid feature, which refers to the sum of intra-slice and inter-slice features from 2D and 3D network, respectively. Then the hybrid feature is jointly learned and optimized in the HFF layer,

$$\begin{aligned} \mathbf{H} &= f_{HFF}(\mathbf{Z}; \theta_{HFF}), \\ \hat{y_H} &= f_{HFFcls}(\mathbf{H}; \theta_{HFFcls}) \end{aligned} \tag{4}$$

where H denotes the optimized hybrid features and $\hat{y_H}$ refers to the pixel-wise predicted probabilities generated from the HFF layer $f_{HFFcls}(\cdot)$. In our experiments, the 3D counterpart of H-DenseUNet cost only 9 hours to converge, which is significantly faster than training the 3D counterpart with original data solely (63 hours).

The detailed structure of the 3D counterpart is shown in the Table I, called 3D DenseUNet-65, which consists of 65 convolutional layers and the growth rate is 32. Compared with 2D DenseUNet counterpart, the number of micro-blocks in each dense block is decreased due to the high memory consumption of 3D convolutions and the limited GPU memory. The rest of the network setting is the same with the 2D counterpart.

C. Loss Function, Training and Inference Schemes

In this section, we present more details regarding the loss function, training and the inference schemes.

1) Loss Function: To train the networks, we employed weighted cross-entropy function as the loss function, which is described as:

$$L(y, \hat{y}) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{3} w_i^c y_i^c \log \hat{y}_i^c \tag{5}$$

where $\hat{y}_i^c$ denotes the probability of voxel i belongs to class c (background, liver or lesion), w_i^c denotes the weight and y_i^c indicates the ground truth label for voxel i.

2) Training Scheme: We first train the ResNet in the same way with Han [40] to get the coarse liver segmentation results. The parameters of the encoder part in 2D DenseUNet f_{2d} are initialized with DenseNet's weights (object classification-trained) [25] while the decoder part are trained with the random initialization. Since the weights are initialized with a random distribution in the decoder part, we first warm up the network without UNet connections. After several iterations, the UNet connections are added to jointly fine tune the model.

To effectively train the H-DenseUNet, we first optimize $f_{2d}(\cdot)$ and $f_{2dcls}(\cdot)$ with cross entropy loss $L(y, \hat{y_{2d}}')$ on

our dataset. Secondly, we fix parameters in $f_{2d}(\cdot)$ and $f_{2dcls}(\cdot)$, and focus on training $f_{3d}(\cdot)$, $f_{HFF}(\cdot)$ and $f_{HFFcls}(\cdot)$ with cross entropy loss $L(y, \hat{y_H})$, where parameters are all randomly initialized. Finally, The whole network is jointly fine-tuned with following combined loss:

$$L_{total} = \lambda L(y, \hat{y_{2d}}') + L(y, \hat{y_H}) \quad (6)$$

where λ is the balanced weight and set as 0.5 in our experiments empirically.

3) Inference Scheme: In the test stage, we first get the coarse liver segmentation result. Then H-DenseUNet can generate accurate liver and tumor predicted probabilities within the ROI. The thresholding is applied to get the liver tumor segmentation result. To avoid the holes in the liver, a largest connected component labeling is performed to refine the liver result. After that, the final lesion segmentation result is obtained by removing lesions outside the final liver region.

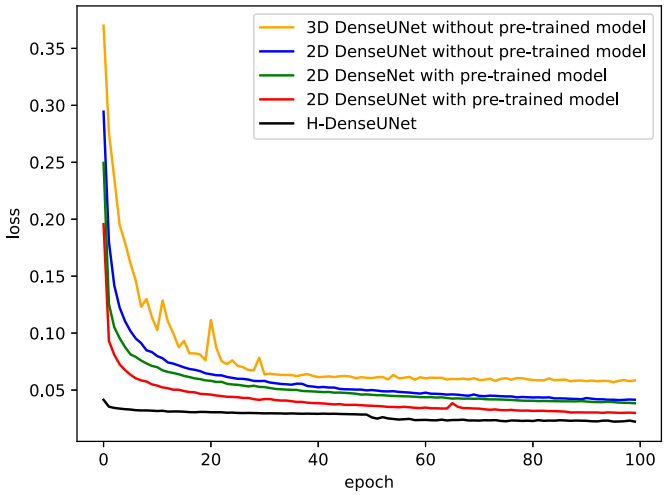

Fig. 3. Training losses of 2D DenseUNet with and without pre-trained model, 2D DenseNet with pre-trained model, 3D DenseUNet without pre-trained model as well as H-DenseUNet (Best viewed in color).

IV. Experiments and Results

A. Dataset and Pre-Processing

We tested our method on the competitive dataset of MICCAI 2017 LiTS Challenge and 3DIRCADb Dataset. The LiTS dataset contains 131 and 70 contrast-enhanced 3D abdominal CT scans for training and testing, respectively. The dataset was acquired by different scanners and protocols from six different clinical sites, with a largely varying in-plane resolution from 0.55 mm to 1.0 mm and slice spacing from 0.45 mm to 6.0 mm. The 3DIRCADb dataset contains 20 venous phase enhanced CT scans, where 15 volumes have hepatic tumors in the liver.

For image preprocessing, we truncated the image intensity values of all scans to the range of [-200, 250] HU to remove the irrelevant details. For coarse liver segmentation in the first stage, we trained a simple network from resampled images with the same resolution $0.69 \times 0.69 \times 1.0$ mm^3. In the test stage, we also employ the resampled images for coarse liver segmentation. For lesion segmentation in the second stage, the network is trained on the images with the original resolution. This is because in some training cases liver lesions are notably small, thus we use images with the original resolution to avoid possible artifacts from image resampling. In this test stage, we also employ the images with original resolution for accurate liver and lesion segmentation.

B. Evaluation Metrics

According to the evaluation of 2017 LiTS challenge, we employed Dice per case score and Dice global score to evaluate the liver and tumor segmentation performance respectively. Dice per case score refers to an average Dice score per volume while Dice global score is the Dice score evaluated by combining all datasets into one. Root mean square error (RMSE) is also adopted to measure the tumor burden.

In the 3DIRCADb dataset, five metrics are used to measure the accuracy of segmentation results, including the volumetric overlap error (VOE), relative volume difference (RVD), average symmetric surface distance (ASD), root mean square symmetric surface distance (RMSD) and DICE. For the first four evaluation metrics, the smaller the value is, the better the segmentation result. The value of DICE refers to the same measurement as Dice per case in the LiTS dataset.

C. Implementation Details

In this section, we present more details regarding the implementation environment and data augmentation strategies. The model was implemented using *Keras* package [47]. The initial learning rate was 0.01 and decayed according to the equation $lr = lr * (1 - iterations/total\_iterations)^{0.9}$. We used stochastic gradient descent with momentum.

For data augmentation, we adopted random mirror and scaling between 0.8 and 1.2 for all training data to alleviate the overfitting problem. The training of 2D DenseUNet model took about 21 hours using two NVIDIA Titan Xp GPUs with 12 GB memory while the end-to-end system fine-tuning cost approximately 9 hours. In other words, the total training time for H-DenseUNet took about 30 hours. In the test phase, the total processing time of one subject depends on the number of slices, ranging from 30 seconds to 200 seconds.

D. Ablation Analysis of H-DenseUNet on LiTS Dataset

In this section, we conduct comprehensive experiments to analyze the effectiveness of our proposed H-DenseUNet. Figure 3 shows the training losses of 2D DenseUNet with and without pre-trained model, 2D DenseNet with pre-trained model, 3D DenseUNet without pre-trained model as well as H-DenseUNet. Note that 3D DenseUNet costs around 60 hours, nearly 3 times than 2D networks. H-DenseUNet costs nearly 30 hours, where 21 hours are spent for 2D DenseUNet training and 9 hours are used to fine-tune the whole architecture in the end-to-end manner. It is worth mentioning that all of the models are run with NVIDIA Titan Xp GPUs with full memory.

TABLE II
SEGMENTATION RESULTS BY ABLATION STUDY OF OUR METHODS ON THE TEST DATASET (DICE: %)

Model	Lesion		Liver	
	Dice per case	Dice global	Dice per case	Dice global
3D DenseUNet without pre-trained model	59.4	78.8	93.6	92.9
UNet [42]	65.0	-	-	-
ResNet [40]	67.0	-	-	-
2D DenseUNet without pre-trained model	67.7	80.1	94.7	94.7
2D DenseNet with pre-trained model	68.3	81.8	95.3	95.9
2D DenseUNet with pre-trained model	70.2	82.1	95.8	96.3
H-DenseUNet	**72.2**	**82.4**	**96.1**	**96.5**

1) Effectiveness of the Pre-Trained Model: One advantage in the proposed method is that we can train the network by transfer learning with the pre-trained model, which is crucial in finding an optimal solution for the network. Here, we analyze the learning behaviors of 2D DenseUNet with and without the pre-trained model. Both two experiments were conducted under the same experimental settings. From Figure 3, it is clearly observed that with the pre-trained model, 2D DenseUNet can converge faster and achieve lower loss value, which shows the importance of utilizing the pre-trained model with transfer learning. The test results in Table II demonstrated that the pre-trained model can help the network achieve better performance consistently. Our proposed H-DenseUNet inherits this advantage, which plays an important role in achieving the promising results.

2) Comparison of 2D and 3D DenseUNet: We compare the inherent performance of 2D DenseUNet and 3D DenseUNet to validate that using 3D network solely maybe defective. The number of parameters is one of key elements in measuring the model representation capability, thus both 2D DenseUNet-167 and 3D DenseUNet-65 are designed with the same level of model complexity (around 40M parameters).

We compare the learning behaviors of two experiments without using the pre-trained model. From Figure 3, it shows that the 2D DenseUNet achieves better performance than the 3D DenseUNet, which highlights the effectiveness and efficiency of 2D convolutions with the deep architecture. This is because the 3D kernel consumes large GPU memory so that the network depth and width are limited, leading to weak representation capability. In addition, 3D DenseUNet took much more training time (approximately 60 hours) to converge compared to 2D DenseUNet (around 20 hours).

Except for the heavy computational cost of the 3D network, another defective is that only a dearth of pre-trained model exists for the 3D network. From Table II, compared with the results generated by 3D DenseUNet, 2D DenseUNet with pre-trained model achieved 8.9 and 3.0 (Dice: %) improvements on the lesion segmentation results by the measurement of Dice per case and Dice global score, respectively.

3) Effectiveness of UNet Connections: We analyze the effectiveness of UNet connections in our proposed framework. Both 2D DenseNet and DenseUNet are trained with the same pre-trained model and training strategies. The difference is that DenseUNet contains long range connections between the encoding part and the decoding part to preserve high-resolution features. As the results shown in Figure 3, it is obvious that DenseUNet achieves lower loss value than DenseNet, demonstrating the UNet connections actually help the network converge to a better solution. The experimental results in Table II consistently demonstrated that the lesion segmentation performance can be boosted by a large margin with UNet connections embedded in the network.

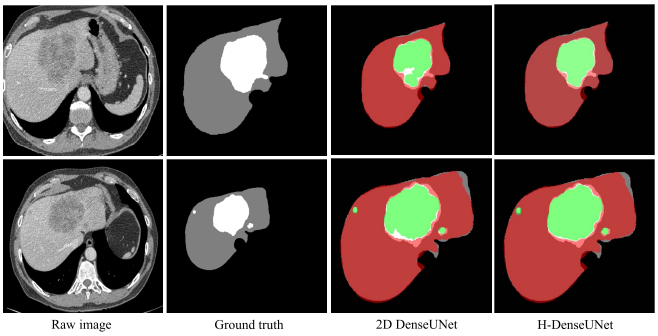

Fig. 4. Examples of segmentation results by 2D DenseUNet and H-DenseUNet on the validation dataset. The *red* regions denote the segmented liver while the *green* ones denote the segmented lesions. The *gray* regions denote the true liver while the *white* ones denote the true lesions.

4) Effectiveness of Hybrid Feature Fusion: To validate the effectiveness of the hybrid architecture, we compare the learning behaviors of H-DenseUNet and 2D DenseUNet. It is observed that the loss curve for H-DenseUNet begins around 0.04. This is because we fine tune the H-DenseUNet on the 2D DenseUNet basis, which serves as a good initialization. Then the loss value decreases to nearly 0.02, which is attributed to the hybrid feature fusion learning. Figure 3 shows that H-DenseUNet can converge to the smaller loss value than the 2D DenseUNet, which indicates that the hybrid architecture can contribute to the performance gains. Compared with 2D DenseUNet, our proposed H-DenseUNet advances the segmentation results on both two measurements for liver and tumor segmentation consistently, as shown in Table II. The performance gains indicate that contextual information along the z dimension, indeed, contributes to the recognition of lesion and liver, especially for lesions that have much more blurred boundary and considered to be difficult to recognize. Figure 4 shows some segmentation results achieved by 2D DenseUNet and H-DenseUNet on the validation dataset. It is observed that H-DenseUNet can achieve much better results than 2D DenseUNet. Moreover, we trained H-DenseUNet in an end-to-end manner, where

TABLE III
LEADERBOARD OF 2017 LIVER TUMOR SEGMENTATION (LITS) CHALLENGE (DICE: %, UNTIL 1ST NOV. 2017)

Team	Lesion		Liver		Tumor Burden
	Dice per case	Dice global	Dice per case	Dice global	RMSE
our	**72.2**	**82.4**	**96.1**	**96.5**	**0.015**
IeHealth	70.2	79.4	**96.1**	96.4	0.017
hans.meine	67.6	79.6	96.0	**96.5**	0.020
superAI	67.4	81.4	0.0	0.0	1251.447
Elehanx [40]	67.0	-	-	-	-
medical	66.1	78.3	95.1	95.1	0.023
deepX [48]	65.7	82.0	**96.3**	**96.7**	0.017
Njust768	65.5	76.8	4.10	13.5	0.920
Medical [41]	65.0	-	-	-	-
Gchlebus [42]	65.0	-	-	-	-
predible	64.0	77.0	95.0	95.0	0.020
Lei [49]	64.0	-	-	-	-
ed10b047	63.0	77.0	94.0	94.0	0.020
chunliang	62.5	78.8	95.8	96.2	0.016
yaya	62.4	79.2	95.9	96.3	0.016

Note: - denotes that the team participated in ISBI competition and the measurement was not evaluated.

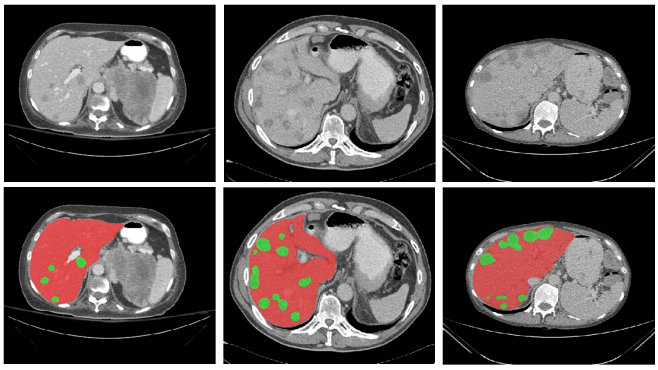

Fig. 5. Examples of liver and tumor segmentation results of H-DenseUNet from the test dataset. The *red* regions denote the liver and the *green* ones denote the tumors.

the 3D contexts can also help extract more representative in-plane features. The end-to-end system jointly optimizes the 2D and 3D networks, where the hybrid feature can be fully explored. Figure 5 presents some examples of liver and tumor segmentation results of our H-DenseUNet on the test dataset. We can observe that most small targets as well as large objects can be well segmented.

E. Comparison With Other Methods on LiTS Dataset

There were more than 50 submissions in 2017 ISBI and MICCAI LiTS challenges. Both challenges employed the same training and test datasets for fair performance comparison. Different from the ISBI challenge, more evaluation metrics have been added in the MICCAI challenge for comprehensive comparison. The detailed results of top 15 teams on the leaderboard,[1] including both ISBI and MICCAI challenges, are listed in Table III. Our method (team name: xjqi, entry date: Nov. 17, 2017) outperformed other state-of-the-arts on the segmentation results of tumors and achieved very competitive performance for liver segmentation. For tumor burden evaluation, our method achieved the lowest estimation error and ranked the 1st place among all the teams. It is worth mentioning that we used ten entries on the test dataset for ablation analysis of our method. Since there is no validation set provided by challenge organizers, the ablation experiments were performed on test dataset for fair comparison. Please note that the final result is just one of these entries, instead of multiple entries averages.

Most of the top teams in the challenges employed deep learning based methods, demonstrating the effectiveness of CNN based methods in medical image analysis. For example, Han [40] Vorontsov *et al.* [41] and Bi *et al.* [49] all adopted 2D deep FCNs, where ResNet-like residual blocks were employed as the building blocks. In addition, Chlebus *et al.* [42] trained the UNet architecture in two individual models, followed by a random forest classifier. In comparison, our method with a 167-layer network consistently outperformed these methods, which highlighted the efficacy of 2D DenseUNet with pre-trained model. Our proposed H-DenseUNet further advanced the segmentation accuracy for both liver and tumor, showing the effectiveness of the hybrid feature learning process.

Our method achieved the 1st place among all state-of-the-arts in the lesion segmentation and very competitive result to DeepX [48] for liver segmentation. Note that our method surpassed DeepX by a significant margin in the Dice per case evaluation for lesion, which is considered to be notoriously challenging and difficult. Moreover, our result was produced by the single model while DeepX [48] employed multi-model combination strategy to improve the results, showing the efficiency of our method in the clinical practice.

F. Comparison With Other Methods on 3DIRCADb Dataset

To validate the effectiveness and robustness of our method, we also conduct experiments on 3DIRCADb dataset [56], which is publicly available and offers a higher variety and complexity of livers and lesions. Table IV and Table V show

[1] https://competitions.codalab.org/competitions/17094#results

TABLE IV
COMPARSION OF TUMOR SEGMENTATION RESULTS ON 3DIRCADB DATASET

Model	Year	VOE(%)	RVD(%)	ASD(mm)	RMSD(mm)	DICE
Unet [42]	2017	62.55 ± 22.36	0.380 ± 1.95	11.11 ± 12.02	16.71 ± 13.81	0.51 ± 0.25
Christ et al. [39]	2017	-	-	-	-	0.56 ± 0.26
ResNet [40]	2017	56.47 ± 13.62	-0.41 ± 0.21	6.36 ± 3.77	11.69 ± 7.60	0.60 ± 0.12
ours		49.72 ± 5.2	-0.33 ± 0.10	5.293 ± 6.15	11.11 ± 29.14	0.65 ± 0.02
Foruzan and Chen [50]*	2016	30.61 ± 10.44	15.97 ± 12.04	4.18 ± 9.60	5.09 ± 10.71	0.82 ± 0.07
Wu et al. [51]*	2017	29.04 ± 8.16	-2.20 ± 15.88	0.72 ± 0.33	1.10 ± 0.49	0.83 ± 0.06
Li et al. [52] †	2013	14.4 ± 5.3	-8.1 ± 2.1	2.4 ± 0.8	2.9 ± 0.7	-
Moghbel et al. [53] †	2016	22.78 ± 12.15	8.59 ± 18.78	-	-	0.75 ± 0.15
Sun et al. [16] †	2017	15.6 ± 4.3	5.8 ± 3.5	2.0 ± 0.9	2.9 ± 1.5	-
ours †		**11.68 ± 4.33**	**-0.01 ± 0.05**	**0.58 ± 0.46**	**1.87 ± 2.33**	**0.937 ± 0.02**

Note: ∗ denotes the semi-automatic methods; † denotes the method use additional datasets; - denotes the result is not reported.

TABLE V
COMPARSION OF LIVER SEGMENTATION RESULTS ON 3DIRCADB DATASET

Model	Year	VOE(%)	RVD(%)	ASD(mm)	RMSD(mm)	DICE
Unet [42]	2017	14.21 ± 5.71	-0.05 ± 0.10	4.33 ± 3.39	8.35 ± 7.54	0.923 ± 0.03
ResNet [40]	2017	11.65 ± 4.06	-0.03 ± 0.06	3.91 ± 3.95	8.11 ± 9.68	0.938 ± 0.02
Christ et al. [39]	2017	10.7	-1.4	1.5	24.0	0.943
ours		10.02 ± 3.44	-0.01 ± 0.05	4.06 ± 3.85	9.63 ± 10.41	0.947 ± 0.01
Li et al. [54] †	2015	9.15 ± 1.44	-0.07 ± 3.64	1.55 ± 0.39	3.15 ± 0.98	-
Moghbel et al. [55]†	2016	5.95	7.49	-	-	0.911
Lu et al. [19] †	2017	9.36 ± 3.34	0.97 ± 3.26	1.89 ± 1.08	4.15 ± 3.16	-
ours †		**3.57 ± 1.66**	**0.01 ± 0.02**	**1.28 ± 2.02**	**3.58 ± 6.58**	**0.982 ± 0.01**

Note: † denotes the method use additional datasets. - denotes the result is not reported.

the comparison of the tumor and liver segmentation performance on the 3DIRCADb dataset. We compared our method with the state-of-the-art method [39] on the 3DIRCADb dataset by running experiments through cross-validation, as the way used in [39]. We can see that our method achieved the better performance than [39] on both lesion and liver segmentation accuracy, with 9.0% and 0.4% improvement on DICE, respectively. To further validate the effectiveness of our method, we ran experiments with methods of Unet [42] and ResNet architecture [40] respectively, where the training setting keeps the same with Christ *et al.* [39]. From Table IV and Table V, we can see that our method still outperforms Unet [42] and ResNet [40] on the 3DIRCADb dataset, with 14.0% and 5.0% improvement on DICE for tumor segmentation respectively. The experimental comparison validated the superiority of our proposed method in comparison with other methods.

To have a comprehensive comparison with liver tumor segmentation methods, we listed the reported tumor and liver segmentation results on the 3DIRCADb dataset below the bold line in Table IV and Table V, respectively. Note that except experiments [40] and [42], all other experiment results are the reported values in the original papers. It is worth noting that most liver tumor segmentation methods [16], [19], [52]–[55] utilized additional datasets for training and tested on the 3DIRCADb dataset. For example, Li *et al.* [52], Sun *et al.* [16] and Lu *et al.* [19] collected additional clinical data from hospitals as the training set. Moghbel *et al.* [53] utilized additional the MIDAS dataset while Li *et al.* [54] used the SLIVER07 dataset in the training, respectively. In addition, Foruzan and Chen [50] and Wu *et al.* [51] achieved good results on tumor segmentation by semi-automatic methods. Actually, these methods cannot be compared directly with each other due to the differences in the training dataset and whether is fully-automatic or not. However, to some extent, the reported results on the 3DIRCADb dataset can reflect the state-of-the-art performance for the lesion and liver segmentation task. Here, we employed the LiTS dataset as the additional dataset. Specifically, we directly tested the well-trained model from 2017 LiTS dataset on the 3DIRCADb dataset. As shown in Table IV and Table V, our method achieves the best tumor and liver segmentation results on the 3DIRCADb dataset, surpassing the state-of-the-art result largely, with 10.7% and 7.1% improvement on DICE for tumor and liver segmentation respectively. The promising result indicates the effectiveness and good generalization capability of our method. On the other hand, such a good result is also attributed to the LiTS dataset, which contains a huge amount of training data with large variations, and the ability of our method to extract discriminative features from this dataset. Figure 6 shows some examples of the results on the 3DIRCADb dataset. It is obvious that our method can well segment the liver and liver lesions from challenging raw CT scans.

V. DISCUSSION

Automatic liver and tumor segmentation plays an important role in clinical diagnosis. It provides the precise contour of the liver and any tumors inside the anatomical segments of the liver, which assists doctors in the diagnosis process. In this

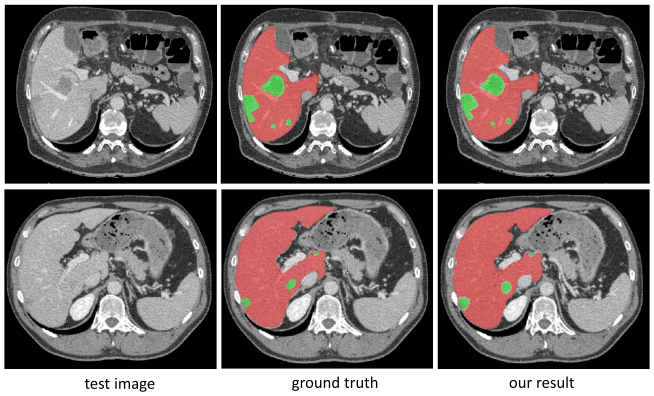

Fig. 6. Examples of our segmentation results on the 3DIRCADb dataset.

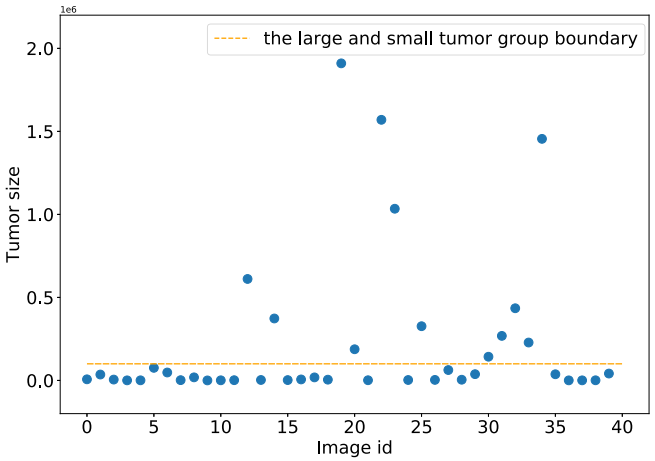

Fig. 7. Tumor size (tumor voxels number) in each patient of our validation dataset. We define the orange line to seperate the large-tumor and the small-tumor group.

paper, we present an end-to-end training system to explore hybrid features for automatic liver lesion segmentation, where the 3D contexts are effectively probed under the auto-context mechanism. Through the hybrid fusion learning of intra-slice and inter-slice features, the segmentation performance for liver lesion has been improved, which demonstrates the effectiveness of our H-DenseUNet. Moreover, compared with other 3D networks [10], [18], our method probes 3D contexts efficiently. This is crucial in the clinical practice, especially when huge amount of 3D images, containing large image size and a number of slices, are increasingly accumulated in the clinical sites.

To show the generalization capability of our method in the clinical practice, we tested our trained model from the LiTS dataset on the 3DIRCADb dataset, and it achieved the state-of-the-art results on both liver and tumor segmentation, with 98.2% and 93.7% on DICE. The promising results achieved on the 3DIRACDb dataset also validated that our method is not simple overtraining, but actually is effective to generalize to different dataset under different data collection conditions.

To have a better understanding about the performance gains, we analyze the effectiveness of our method regarding the liver tumor size in each patient. Figure 7 shows the tumor size value of 40 CT volume data in our validation dataset, where

TABLE VI
EFFECTIVENESS OF OUR METHOD REGARDING TO THE TUMOR SIZE (DICE: %)

	Total	Large-tumor group	Small-tumor group
Baseline	43.56	58.24	41.08
H-DenseUNet	45.04 (+1.48)	60.59 (+2.35)	42.18 (+1.1)

Note: Baseline is the 2D DenseUNet with pre-trained model.

the tumor size is obtained by summing up tumor voxels in each ground-truth image. It is observed that the dataset has large variations of the tumor size. For comparison, we divide the dataset into the large-tumor group and the small-tumor group by the orange line in Figure 7. From Table VI, we can observe that our method improves the segmentation accuracy by 1.48 (Dice:%) in the whole validation dataset. We can also observe that the large-tumor group achieves 2.35 (Dice:%) accuracy improvements while the score for the small-tumor group is slightly advanced, with 1.1 (Dice:%). From the comparison, we claim that the performance gain is mainly attributed to the improvement of the large-tumor data segmentation results. This is mainly because that the H-DenseUNet mimics the diagnosis process of radiologists, where tumors are delineated by observing several adjacent slices, especially for tumors have blurred boundaries. Once the blurred boundaries are well segmented, the segmentation accuracy for the large-tumor data can be improved by a large margin. Although the hybrid feature still contributes to the segmentation of small tumors, the improvement is limited since small tumors usually occur in fewer slices. In the future, we will focus on the segmentation for small liver tumors. Several potential directions will be taken into considerations for tackling small liver tumor problem, i.e., multi-scale representation structure [57] and deep supervision [18]. Recently, perceptual generative adversarial networks (GANs) have been proposed for small object detection and classification [58], [59]. For example, Li *et al.* [58] generated superresolved representations for small objects by discovering the intrinsic structural correlations between small-scale and large-scale objects, which may also be a potential direction for handling this challenging problem.

Another key that should be explored in the future study is the potential depth for the H-DenseUNet. In our experiments, we trained the network using data parallel training, which is an effective technique to speed up the gradient descent by paralleling the computation of the gradient for a mini-batch across mini-batch elements. However, the model complexity is restricted by the GPU memory. In the future, to exploit the potential depth of the H-DenseUNet, we can train the network using model parallel training, where different portions of the model computation are done on distributed computing infrastructures for the same batch of examples. This strategy maybe another possible direction to further improve the liver tumor segmentation performance.

VI. CONCLUSION

We present an end-to-end training system H-DenseUNet for liver and tumor segmentation from CT volumes, which is a new paradigm to effectively probe high-level representative

intra-slice and inter-slice features, followed by optimizing the features through the hybrid feature fusion layer. The architecture gracefully addressed the problems that 2D convolutions ignore the volumetric contexts and 3D convolutions suffer from heavy computational cost. Extensive experiments on the dataset of 2017 LiTS and 3DIRCADb dataset demonstrated the superiority of our proposed H-DenseUNet. With a single-model basis, our method excelled others by a large margin on lesion segmentation and achieved very competitive result on liver segmentation on the LiTS Leaderboard.

REFERENCES

[1] J. Ferlay, H. R. Shin, F. Bray, D. Forman, C. Mathers, and D. M. Parkin, "Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008," *Int. J. Cancer*, vol. 127, no. 12, pp. 2893–2917, Dec. 2010.

[2] R. Lu, P. Marziliano, and C. H. Thng, "Liver tumor volume estimation by semi-automatic segmentation method," in *Proc. 27th Annu. Int. Conf. Eng. Med. Biol. Soc. (IEEE-EMBS)*, Jan. 2006, pp. 3296–3299.

[3] M. Moghbel, S. Mashohor, R. Mahmud, and M. I. B. Saripan, "Review of liver segmentation and computer assisted detection/diagnosis methods in computed tomography," in *Artificial Intelligence Review*. Springer, 2017, pp. 1–41.

[4] A. Prasoon, K. Petersen, C. Igel, F. Lauze, E. Dam, and M. Nielsen, "Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*. Berlin, Germany: Springer, 2013, pp. 246–253.

[5] 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)*. Cham, Switzerland: Springer, 2015, pp. 234–241.

[6] M. F. Stollenga, W. Byeon, M. Liwicki, and J. Schmidhuber, "Parallel multi-dimensional LSTM, with application to fast biomedical volumetric image segmentation," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 2998–3006.

[7] H. R. Roth *et al.*, "DeepOrgan: Multi-level deep convolutional networks for automated pancreas segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*. Cham, Switzerland: Springer, 2015, pp. 556–564.

[8] J. Wang, J. D. MacKenzie, R. Ramachandran, and D. Z. Chen, "Detection of glands and villi by collaboration of domain knowledge and deep learning," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*. Cham, Switzerland: Springer, 2015, pp. 20–27.

[9] X. Li, Q. Dou, H. Chen, C.-W. Fu, and P.-A. Heng, "Multi-scale and modality dropout learning for intervertebral disc localization and segmentation," in *Proc. Int. Workshop Comput. Methods Clin. Appl. Spine Imag. (MICCAI)*. Cham, Switzerland: Springer, 2016, pp. 85–91.

[10] 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)*. Cham, Switzerland: Springer, 2016, pp. 424–432.

[11] M. Havaei *et al.*, "Brain tumor segmentation with deep neural networks," *Med. Image Anal.*, vol. 35, pp. 18–31, Jan. 2017.

[12] H. Chen, Q. Dou, L. Yu, J. Qin, and P.-A. Heng, "VoxResNet: Deep voxelwise residual networks for brain segmentation from 3D MR images," *NeuroImage*, vol. 170, pp. 446–455, Apr. 2017.

[13] X. Wang *et al.*, "Liver segmentation from CT images using a sparse priori statistical shape model (SP-SSM)," *PLoS ONE*, vol. 12, no. 10, p. e0185249, 2017.

[14] X. Li *et al.*, "3D multi-scale FCN with random modality voxel dropout learning for intervertebral disc localization and segmentation from multi-modality MR images," *Med. Image Anal.*, vol. 45, pp. 41–54, Apr. 2018.

[15] P. F. Christ *et al.*, "Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*. Cham, Switzerland: Springer, 2016, pp. 415–423.

[16] C. Sun *et al.*, "Automatic segmentation of liver tumors from multiphase contrast-enhanced CT images based on FCNs," *Artif. Intell. Med.*, vol. 83, pp. 58–66, Nov. 2017.

[17] A. Ben-Cohen, I. Diamant, E. Klang, M. Amitai, and H. Greenspan, "Fully convolutional network for liver segmentation and lesions detection," in *Deep Learning and Data Labeling for Medical Applications*. Cham, Switzerland: Springer, 2016, pp. 77–85.

[18] Q. Dou, H. Chen, Y. Jin, L. Yu, J. Qin, and P.-A. Heng, "3D deeply supervised network for automatic liver segmentation from CT volumes," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*. Cham, Switzerland: Springer, 2016, pp. 149–157.

[19] F. Lu, F. Wu, P. Hu, Z. Peng, and D. Kong, "Automatic 3D liver location and segmentation via convolutional neural network and graph cut," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 12, no. 2, pp. 171–182, 2017.

[20] K.-L. Tseng, Y.-L. Lin, W. Hsu, and C.-Y. Huang. (2017). "Joint sequence learning and cross-modality convolution for 3D biomedical segmentation." [Online]. Available: https://arxiv.org/abs/1704.07754

[21] J. Cai, L. Lu, Y. Xie, F. Xing, and L. Yang. (2017). "Improving deep pancreas segmentation in CT and MRI images via recurrent neural contextual learning and direct loss function." [Online]. Available: https://arxiv.org/abs/1707.04912

[22] K. Simonyan and A. Zisserman. (2014). "Very deep convolutional networks for large-scale image recognition." [Online]. Available: https://arxiv.org/abs/1409.1556

[23] H. Chen *et al.*, "Standard plane localization in fetal ultrasound via domain transferred deep neural networks," *IEEE J. Biomed. Health Inform.*, vol. 19, no. 5, pp. 1627–1636, Sep. 2015.

[24] N. Tajbakhsh *et al.*, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1299–1312, May 2016.

[25] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2017, pp. 4700–4708.

[26] Z. Tu, "Auto-context and its application to high-level vision tasks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2008, pp. 1–8.

[27] D. H. Wolpert, "Stacked generalization," *Neural Netw.*, vol. 5, no. 2, pp. 241–259, 1992.

[28] L. Soler *et al.*, "Fully automatic anatomical, pathological, and functional segmentation from CT scans for hepatic surgery," *Comput. Aided Surgery*, vol. 6, no. 3, pp. 131–142, Jan. 2001.

[29] J. H. Moltz, L. Bornemann, V. Dicken, and H. Peitgen, "Segmentation of liver metastases in ct scans by adaptive thresholding and morphological processing," in *Proc. MICCAI Workshop*, 2008, pp. 1–8.

[30] D. Wong *et al.*, "A semi-automated method for liver tumor segmentation based on 2D region growing with knowledge-based constraints," in *Proc. MICCAI Workshop*, 2008, pp. 1–10.

[31] D. Jimenez-Carretero *et al.*, "Optimal multiresolution 3D level-set method for liver segmentation incorporating local curvature constraints," in *Proc. Annu. Int. Conf. Eng. Med. Biol. Soc. (EMBC)*, Aug./Sep. 2011, pp. 3419–3422.

[32] W. Huang *et al.*, "Random feature subspace ensemble based Extreme Learning Machine for liver tumor detection and segmentation," in *Proc. 36th Annu. Int. Conf. Eng. Med. Biol. Soc. (EMBC)*, Aug. 2014, pp. 4675–4678.

[33] E. Vorontsov, N. Abi-Jaoudeh, and S. Kadoury, "Metastatic liver tumor segmentation using texture-based omni-directional deformable surface models," in *Proc. Int. MICCAI Workshop Comput. Clin. Challenges Abdominal Imag.* Cham, Switzerland: Springer, 2014, pp. 74–83.

[34] T.-N. Le, P. T. Bao, and H. T. Huynh, "Liver tumor segmentation from MR images using 3D fast marching algorithm and single hidden layer feedforward neural network," *BioMed Res. Int.*, vol. 2016, Jul. 2016, Art. no. 3219068.

[35] C.-L. Kuo, S.-C. Cheng, C.-L. Lin, K.-F. Hsiao, and S.-H. Lee, "Texture-based treatment prediction by automatic liver tumor segmentation on computed tomography," in *Proc. Int. Conf. Comput., Inf. Telecommun. Syst. (CITS)*, Jul. 2017, pp. 128–132.

[36] 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. surgery*, vol. 12, no. 2, pp. 223–233, 2017.

[37] A. Hoogi *et al.*, "Adaptive local window for level set segmentation of CT and MRI liver lesions," *Med. Image Anal.*, vol. 37, pp. 46–55, Apr. 2017.

[38] F. Chaieb, T. B. Said, S. Mabrouk, and F. Ghorbel, "Accelerated liver tumor segmentation in four-phase computed tomography images," *J. Real-Time Image Process.*, vol. 13, no. 1, pp. 121–133, 2017.

[39] P. F. Christ *et al.* (2017). "Automatic liver and tumor segmentation of CT and MRI volumes using cascaded fully convolutional neural networks." [Online]. Available: https://arxiv.org/abs/1702.05970

[40] X. Han. (2017). "Automatic liver lesion segmentation using a deep convolutional neural network method." [Online]. Available: https://arxiv.org/abs/1704.07239

[41] E. Vorontsov, A. Tang, C. Pal, and S. Kadoury. (2017). "Liver lesion segmentation informed by joint liver segmentation." [Online]. Available: https://arxiv.org/abs/1707.07734
[42] G. Chlebus, H. Meine, J. H. Moltz, and A. Schenk. (2017). "Neural network-based automatic liver tumor segmentation with random forest-based candidate filtering." [Online]. Available: https://arxiv.org/abs/1706.00842
[43] Y. Zhou, L. Xie, W. Shen, Y. Wang, E. K. Fishman, and A. L. Yuille, "A fixed-point model for pancreas segmentation in abdominal CT scans," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*. Cham, Switzerland: Springer, 2017, pp. 693–701.
[44] A. Farag, L. Lu, H. R. Roth, J. Liu, E. Turkbey, and R. M. Summers, "A bottom-up approach for pancreas segmentation using cascaded superpixels and (deep) image patch labeling," *IEEE Trans. Image Process.*, vol. 26, no. 1, pp. 386–399, Jan. 2017.
[45] Y. Zhou, L. Xie, E. K. Fishman, and A. L. Yuille, "Deep supervision for pancreatic cyst segmentation in abdominal CT scans," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*. Cham, Switzerland: Springer, 2017, pp. 222–230.
[46] H. R. Roth *et al.* (2017). "Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation." [Online]. Available: https://arxiv.org/abs/1702.00045
[47] F. Chollet *et al.* (2015). *Keras*. [Online]. Available: https://github.com/fchollet/keras
[48] Y. Yuan. (2017). "Hierarchical convolutional-deconvolutional neural networks for automatic liver and tumor segmentation." [Online]. Available: https://arxiv.org/abs/1710.04540
[49] L. Bi, J. Kim, A. Kumar, and D. Feng. (2017). "Automatic liver lesion detection using cascaded deep residual networks." [Online]. Available: https://arxiv.org/abs/1704.02703
[50] A. H. Foruzan and Y.-W. Chen, "Improved segmentation of low-contrast lesions using sigmoid edge model," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 11, no. 7, pp. 1267–1283, 2016.
[51] W. Wu, S. Wu, Z. Zhou, R. Zhang, and Y. Zhang, "3D liver tumor segmentation in CT images using improved fuzzy C-means and graph cuts," *BioMed Res. Int.*, vol. 2017, Sep. 2017, Art. no. 5207685.
[52] 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.
[53] M. Moghbel, S. Mashohor, R. Mahmud, and M. I. B. Saripan, "Automatic liver tumor segmentation on computed tomography for patient treatment planning and monitoring," *EXCLI J.*, vol. 15, pp. 406–423, Jun. 2016.
[54] 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, no. 12, pp. 5315–5329, Dec. 2015.
[55] M. Moghbel, S. Mashohor, R. Mahmud, and M. I. B. Saripan, "Automatic liver segmentation on Computed Tomography using random walkers for treatment planning," *EXCLI J.*, vol. 15, pp. 500–517, Aug. 2016.
[56] L. Soler *et al.*, "3d image reconstruction for comparison of algorithm database: A patient-specific anatomical and medical image database," IRCAD, Strasbourg, France, Tech. Rep., 2010.
[57] K. Kamnitsas *et al.*, "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Med. Image Anal.*, vol. 36, pp. 61–78, Feb. 2017.
[58] J. Li, X. Liang, Y. Wei, T. Xu, J. Feng, and S. Yan, "Perceptual generative adversarial networks for small object detection," in *Proc. IEEE CVPR*, Jul. 2017, pp. 1951–1959.
[59] M. Frid-Adar, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "Gan-based data augmentation for improved liver lesion classification," in *Proc. 1st Conf. Med. Imag. with Deep Learn.*, 2018, pp. 1–3.