ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/319977080

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

Article *in* IEEE Transactions on Medical Imaging · September 2017
DOI: 10.1109/TMI.2018.2845918

CITATIONS
8

READS
249

6 authors, including:

Hao Chen
The Chinese University of Hong Kong
52 PUBLICATIONS **754** CITATIONS
SEE PROFILE

Xiaojuan Qi
The Chinese University of Hong Kong
14 PUBLICATIONS **500** CITATIONS
SEE PROFILE

Qi Dou
The Chinese University of Hong Kong
37 PUBLICATIONS **504** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Deep learning for 3D medical image processing View project

Radiology Project View project

All content following this page was uploaded by Qi Dou on 21 January 2018.

The user has requested enhancement of the downloaded file.

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

Xiaomeng Li[1], Hao Chen[1,2], Xiaojuan Qi[1], Qi Dou[1], Chi-Wing Fu[1], and Pheng-Ann Heng[1]

[1]Department of Computer Science and Engineering, The Chinese University of Hong Kong
[2]Imsight Medical Technology, Inc

arXiv:1709.07330v2 [cs.CV] 22 Nov 2017

***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 2D and 3D FCNs, serve as the back-bone in many volumetric image segmentation. However, 2D convolutions can not fully leverage the spatial information along the third dimension while 3D 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 2D DenseUNet for efficiently extracting intra-slice features and a 3D counterpart for hierarchically aggregating volumetric contexts under the spirit of the auto-context algorithm for liver and tumor segmentation. We formulate the learning process of 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 (HFF) layer. We extensively evaluated our method on the dataset of MICCAI 2017 Liver Tumor Segmentation (LiTS) Challenge. 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.

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 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.

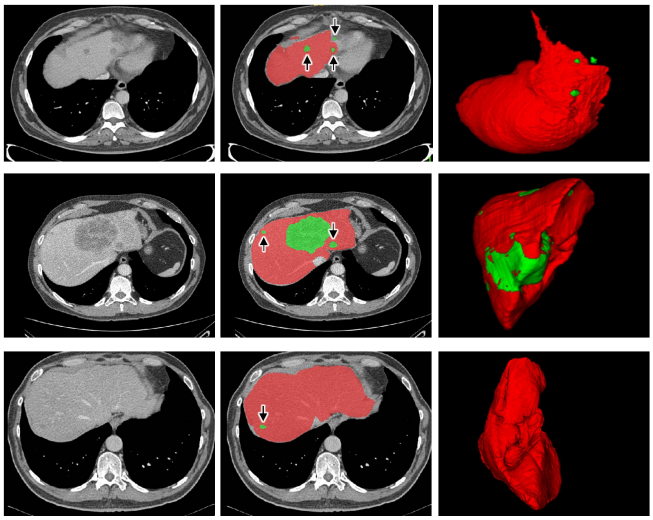

Figure 1: Examples of contrast-enhanced CT scans showing the large variations of shape, size, location of liver lesion. Each row shows a CT scan acquired from individual patient. The *red* regions denote the liver while the *green* ones denote the lesions (see the black arrows above).

To tackle these difficulties, the remarkable ability for extracting visual features is required. Recently, fully convolutional neural networks (FCNs) have achieved great success on a broad array of recognition problems [4–12]. 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 [13], the multi-channel FCN [14], and the FCN based on VGG-16 [15]. (2) 3D FCNs, where 2D convolutions are replaced by 3D convolutions with volumetric data input [16, 17].

In clinical diagnosis, an 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 [14, 15] 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 [4], that is, three 2D FCNs are applied on orthogonal planes (e.g., the xy, yz, and xz planes) and voxel prediction results are generated by the average of these probabilities. However, this approach still cannot probe sufficient volumetric contexts, leading to limited representation ability.

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 [18]. The heavy computation of 3D convolutions also impedes its 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 [19, 20]. 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 pushes the limit further than other works with technical contributions on the following two key factors:

Increased network depth. First, to fully extract high-level intra-slice features, we design a very deep and efficient training network based on 2D convolutions, called 2D DenseUNet, which inherits the advantages of both densely connected path [21] and UNet connections [5]. Densely connected path is derived from densely connected network (DenseNet), where the improved information flow and parameters efficiency make it easy for training a deep network. Different from DenseNet [21], we add the UNet connections, i.e., long-range skip connections, between the encoding part and the decoding part in our architecture; hence, we can enable low-level spatial feature preservation for better intra-slice context exploration.

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 mechanism. 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 to improve the performance. To our knowledge, this is the first work that pushes the depth of 2D networks for liver and tumor segmentation to 167 layers.
- We propose a H-DenseUNet framework to explore hybrid (intra-slice and inter-slice) features for liver and tumor segmentation, which elegantly tackles the problems that 2D convolutions neglect the volumetric contexts and 3D convolutions suffer from heavy computational cost.
- Our framework is an end-to-end system that jointly fuses and optimizes the hybrid features through the HFF layer, which can be served as a new paradigm for effectively exploiting 3D contexts. Compared with other state-of-the-art methods, our method ranked the 1st on lesion segmentation and achieved very competitive performance on liver segmentation in the 2017 LiTS Leaderboard.

II. Related Work

A. Hand-crafted feature based methods

In the past decades, a lot of algorithms, including thresholding [22, 23], region growing, deformable model based methods [24, 25] and machine learning based methods [26–30] 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. [24] 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 [31]. For example, Jimenez-Carretero et al. [25] 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. [26] proposed to employ the random feature subspace ensemble-based extreme learning machine (ELM) for liver lesion segmentation. Vorontsov et al. [27] 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. [29] proposed to learn SVM classifier with texture feature vector for liver tumor segmentation. Le et al. [28] 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. [32] 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. [15] proposed to use a FCN for liver segmentation and liver-metastasis detection in CT examinations. Christ et al. [13, 33] 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. [14] 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 [34], 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. [35] and Chlebus et al. [36] achieved the second place in the ISBI challenge. Vorontsov et al. [35] 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 [34]. Chlebus et al. [36] 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. To reduce the overall computation time, a simple ResNet architecture [34] 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 [21], 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 [21] 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 DesneUNet. 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 liver and tumor segmentation,

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

where $\mathbf{X_{2d}}$ is the feature map from layer "*upsample5_conv*" (see Table I) and $\hat{y_{2d}}$ is the predicted pixel-wise probabilities corresponding to the input three adjacent slices.

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 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.

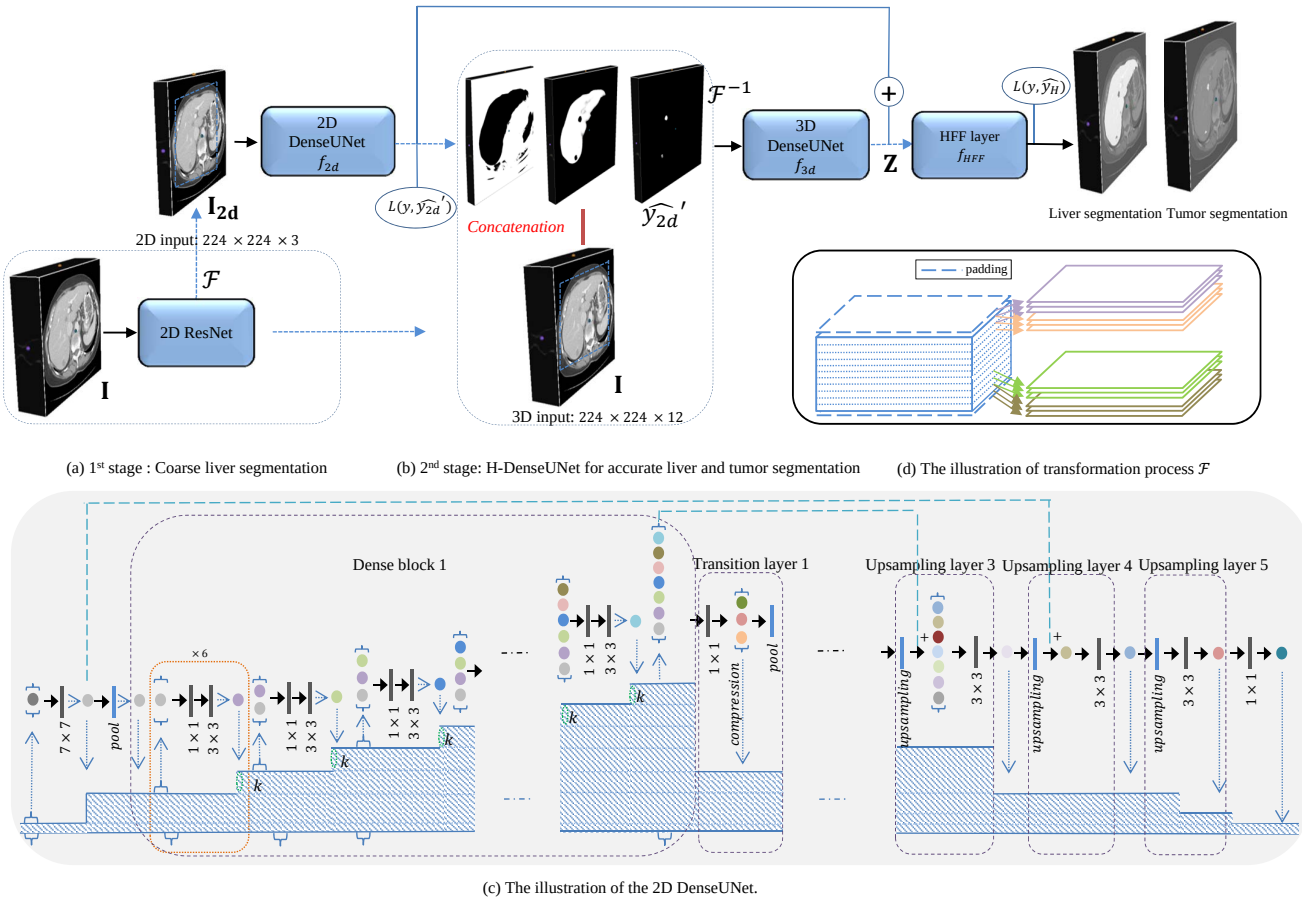

Figure 2: The illustration of the pipeline for liver and lesion segmentation. (a) A ResNet is employed for coarse liver segmentation to reduce computation time. (b) The H-DenseUNet consists of 2D DenseUNet and 3D counterpart, which are responsible for hierarchical features extraction from intra-slice and inter-slice, respectively. Hybrid feature fusion (HFF) layer is proposed to fuse intra-slice and inter-slice features and optimize together for better liver and tumor segmentation. $L(y, \hat{y_{2d}}')$ and $L(y, \hat{y_H})$ are jointly employed to supervise this end-to-end learning process. (c) The illustration of the 2D DenseUNet. The structure in the orange block is a micro-block and k denotes the growth-rate. (d) The transformation of volumetric data to three adjacent slices (Best viewed in color).

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:

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

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 "*upsample5_conv*" 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,

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

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. The second and forth column indicate the output size of the current stage in two architectures, respectively. Note that each "conv" corresponds the sequence BN-ReLU-Conv.

	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
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

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 [34] 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) [21] 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:

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

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.

IV. Experiments and Results

A. Dataset and Pre-processing

We tested our method on a competitive dataset of MICCAI 2017 LiTS Challenge, which 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.

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, all the

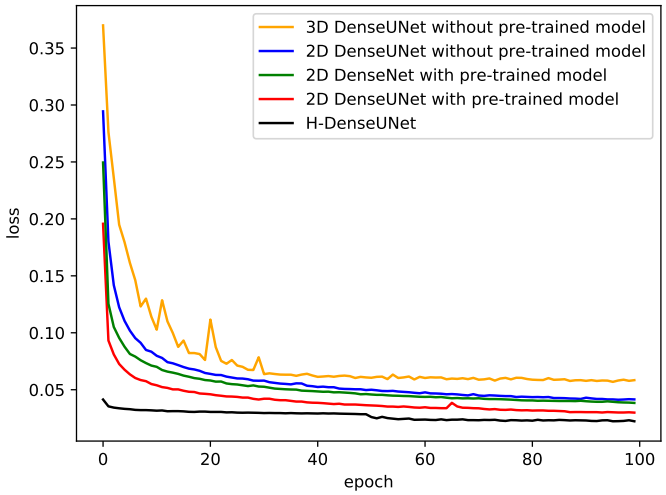

Figure 3: Training losses of 2D DenseUNet with and without pre-trained model, 2D DenseNet with pre-trained model, 3D DenseUNet without pre-trained model as well as H-DenseUNet (Best viewed in color).

training images were resampled to a fixed resolution of $0.69 \times 0.69 \times 1.0$ mm^3. Since the liver lesions are extremely small in some training cases, we trained with the original images in the networks to avoid possible artifact from image resampling.

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.

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 [37]. 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 3D counterpart model cost about 9 hours under the same settings. 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

In this section, we conduct comprehensive experiments to analyze the effectiveness of our proposed H-DenseUNet.

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 DesneUNet-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 (63 hours) to converge compared to 2D DenseUNet (21 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.

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

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 [36]	65.0	-	-	-
ResNet [34]	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**

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 [34]	67.0	-	-	-	-
medical	66.1	78.3	95.1	95.1	0.023
deepX [38]	65.7	82.0	**96.3**	**96.7**	0.017
Njust768	65.5	76.8	4.10	13.5	0.920
Medical [35]	65.0	-	-	-	-
Gchlebus [36]	65.0	-	-	-	-
predible	64.0	77.0	95.0	95.0	0.020
Lei [39]	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.

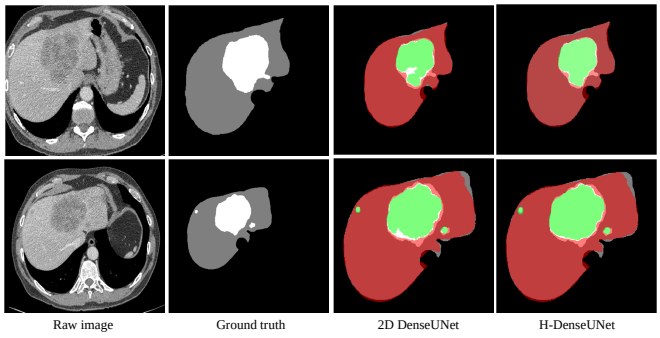

Figure 4: Examples of segmentation results by 2D DenseUNet and H-DenseUNet on the validation dataset. The *red* regions denote the segmented liver while the *green* ones denote the segmented lesions. The *gray* regions denote the true liver while the *white* ones denote the true lesions.

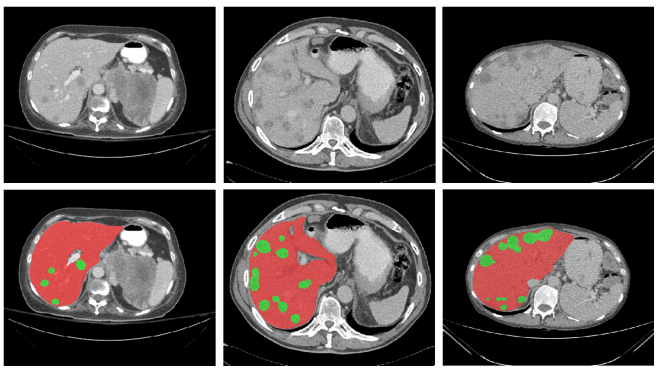

Figure 5: Examples of liver and tumor segmentation results of H-DenseUNet from the test dataset. The *red* regions denote the liver and the *green* ones denote the tumors.

the 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 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

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, including both ISBI and MICCAI challenges, are listed in Table III. Our method 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 our result was achieved by a single model.

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 [34], Vorontsov et al. [35] and Bi et al. [39] all adopted 2D deep FCNs, where ResNet-like residual blocks were employed as the building blocks. In addition, Chlebus et al. [36] 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 [38] 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 [38] employed multi-model combination strategy to improve the results, showing the efficiency of our method in the clinical practice.

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 paper, we present an end-to-end training system to explore hybrid features for automatic liver lesion segmentation, where we probe 3D contexts effectively 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 [9, 16], 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 have a better understanding about the performance gains, we analyze the effectiveness of our method regarding the tumor size in each patient. Figure 6 shows the tumor size

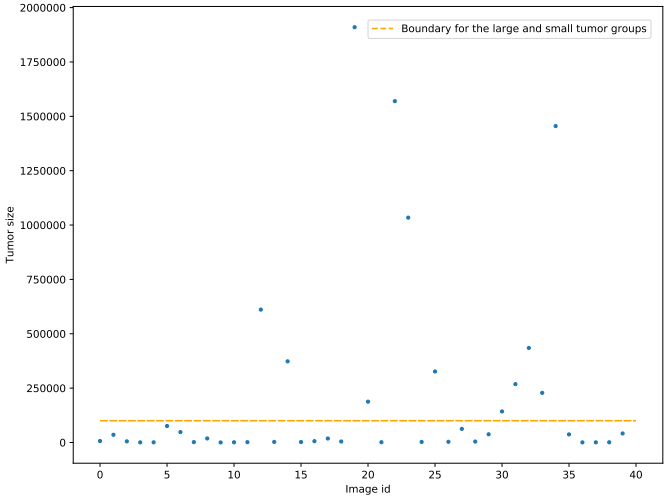

Figure 6: Tumor size in each patient of the validation dataset.

Table IV: 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.

value of 40 CT volume data in the validation dataset, where 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 6. From Table IV, 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 tumors. Several potential directions will be taken into considerations, i.e., multi-scale representation structure [40] and deep supervision [16]. Recently, perceptual generative adversarial network (GAN) [41] has been proposed for small object detection. They generate 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, which will be another possible direction to further improve the 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 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 Leaderboard. In addition, our network architecture is inherently general and can be easily extended to other applications.

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," *International journal of cancer*, vol. 127, no. 12, pp. 2893–2917, 2010.

[2] R. Lu, P. Marziliano, and C. H. Thng, "Liver tumor volume estimation by semi-automatic segmentation method," in *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*. IEEE, 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," *Artificial Intelligence Review*, pp. 1–41, 2017.

[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 *International conference on medical image computing and computer-assisted intervention*. Springer, 2013, pp. 246–253.

[5] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, 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 *Advances in Neural Information Processing Systems*, 2015, pp. 2998–3006.

[7] H. R. Roth, L. Lu, A. Farag, H.-C. Shin, J. Liu, E. B. Turkbey, and R. M. Summers, "Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 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 *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 20–27.

[9] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3d u-net: learning dense volumetric segmentation from sparse annotation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2016, pp. 424–432.

[10] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, and H. Larochelle, "Brain tumor segmentation with deep neural networks," *Medical image analysis*, vol. 35, pp. 18–31, 2017.

[11] 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*, 2017.

[12] X. Wang, Y. Zheng, L. Gan, X. Wang, X. Sang, X. Kong, and J. Zhao, "Liver segmentation from ct images using a sparse priori statistical shape model (sp-ssm)," *PloS one*, vol. 12, no. 10, p. e0185249, 2017.

[13] P. F. Christ, M. E. A. Elshaer, F. Ettlinger, S. Tatavarty, M. Bickel, P. Bilic, M. Rempfler, M. Armbruster, F. Hofmann, M. D'Anastasi *et al.*, "Automatic liver and lesion segmentation in ct using cascaded fully convolutional neural networks and 3d conditional random fields," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2016, pp. 415–423.

[14] C. Sun, S. Guo, H. Zhang, J. Li, M. Chen, S. Ma, L. Jin, X. Liu, X. Li, and X. Qian, "Automatic segmentation of liver tumors from multiphase contrast-enhanced ct images based on fcns," *Artificial Intelligence in Medicine*, 2017.

[15] A. Ben-Cohen, I. Diamant, E. Klang, M. Amitai, and H. Greenspan, "Fully convolutional network for liver segmentation and lesions detection," in *International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis*. Springer, 2016, pp. 77–85.

[16] 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 *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2016, pp. 149–157.

[17] F. Lu, F. Wu, P. Hu, Z. Peng, and D. Kong, "Automatic 3d liver location and segmentation via convolutional neural network and graph cut," *International journal of computer assisted radiology and surgery*, vol. 12, no. 2, pp. 171–182, 2017.

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

[19] H. Chen, D. Ni, J. Qin, S. Li, X. Yang, T. Wang, and P. A. Heng, "Standard plane localization in fetal ultrasound via domain transferred deep neural networks," *IEEE journal of biomedical and health informatics*, vol. 19, no. 5, pp. 1627–1636, 2015.

[20] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1299–1312, 2016.

[21] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[22] L. Soler, H. Delingette, G. Malandain, J. Montagnat, N. Ayache, C. Koehl, O. Dourthe, B. Malassagne, M. Smith, D. Mutter *et al.*, "Fully automatic anatomical, pathological, and functional segmentation from ct scans for hepatic surgery," *Computer Aided Surgery*, vol. 6, no. 3, pp. 131–142, 2001.

[23] J. H. Moltz, L. Bornemann, V. Dicken, and H. Peitgen, "Segmentation of liver metastases in ct scans by adaptive thresholding and morphological processing," in *MICCAI workshop*, vol. 41, no. 43, 2008, p. 195.

[24] D. Wong, J. Liu, Y. Fengshou, Q. Tian, W. Xiong, J. Zhou, Y. Qi, T. Han, S. Venkatesh, and S.-c. Wang, "A semi-automated method for liver tumor segmentation based on 2d region growing with knowledge-based constraints," in *MICCAI workshop*, vol. 41, no. 43, 2008, p. 159.

[25] D. Jimenez-Carretero, L. Fernandez-de Manuel, J. Pascau, J. M. Tellado, E. Ramon, M. Desco, A. Santos, and M. J. Ledesma-Carbayo, "Optimal multiresolution 3d level-set method for liver segmentation incorporating local curvature constraints," in *Engineering in medicine and biology society, EMBC, 2011 annual international conference of the IEEE*. IEEE, 2011, pp. 3419–3422.

[26] W. Huang, Y. Yang, Z. Lin, G.-B. Huang, J. Zhou, Y. Duan, and W. Xiong, "Random feature subspace ensemble based extreme learning machine for liver tumor detection and segmentation," in *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*. IEEE, 2014, pp. 4675–4678.

[27] E. Vorontsov, N. Abi-Jaoudeh, and S. Kadoury, "Metastatic liver tumor segmentation using texture-based omni-directional deformable surface models," in *International MICCAI Workshop on Computational and Clinical Challenges in Abdominal Imaging*. Springer, 2014, pp. 74–83.

[28] T.-N. Le, H. T. Huynh *et al.*, "Liver tumor segmentation from mr images using 3d fast marching algorithm and single hidden layer feedforward neural network," *BioMed research international*, vol. 2016, 2016.

[29] 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 *Computer, Information and Telecommunication Systems (CITS), 2017 International Conference on*. IEEE, 2017, pp. 128–132.

[30] P.-H. Conze, V. Noblet, F. Rousseau, F. Heitz, V. de Blasi, R. Memeo, and P. Pessaux, "Scale-adaptive supervoxel-based random forests for liver tumor segmentation in dynamic contrast-enhanced ct scans," *International journal of computer assisted radiology and surgery*, vol. 12, no. 2, pp. 223–233, 2017.

[31] A. Hoogi, C. F. Beaulieu, G. M. Cunha, E. Heba, C. B. Sirlin, S. Napel, and D. L. Rubin, "Adaptive local window for level set segmentation of ct and mri liver lesions," *Medical image analysis*, vol. 37, pp. 46–55, 2017.

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

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

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

[35] E. Vorontsov, G. Chartrand, A. Tang, C. Pal, and S. Kadoury, "Liver lesion segmentation informed by joint liver segmentation," *arXiv preprint arXiv:1707.07734*, 2017.

[36] G. Chlebus, H. Meine, J. H. Moltz, and A. Schenk, "Neural network-based automatic liver tumor segmentation with random forest-based candidate filtering," *arXiv preprint arXiv:1706.00842*, 2017.

[37] F. Chollet *et al.*, "Keras," https://github.com/fchollet/keras, 2015.

[38] Y. Yuan, "Hierarchical convolutional-deconvolutional neural networks for automatic liver and tumor segmentation," *arXiv preprint arXiv:1710.04540*, 2017.

[39] L. Bi, J. Kim, A. Kumar, and D. Feng, "Automatic liver lesion detection using cascaded deep residual networks," *arXiv preprint arXiv:1704.02703*, 2017.

[40] K. Kamnitsas, C. Ledig, V. F. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, "Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation," *Medical image analysis*, vol. 36, pp. 61–78, 2017.

[41] J. Li, X. Liang, Y. Wei, T. Xu, J. Feng, and S. Yan, "Perceptual generative adversarial networks for small object detection," in *IEEE CVPR*, 2017.

View publication stats