IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 23, 2020, accepted December 21, 2020, date of publication December 28, 2020,
date of current version January 7, 2021.

Digital Object Identifier 10.1109/ACCESS.2020.3047861

A Multiple Layer U-Net, U^n-Net, for Liver and Liver Tumor Segmentation in CT

SONG-TOAN TRAN[1,2], CHING-HWA CHENG[3], (Member, IEEE), AND DON-GEY LIU[1,3]

[1]Ph.D. Program of Electrical and Communications Engineering, Feng Chia University, Taichung 40724, Taiwan, R.O.C.
[2]Department of Electrical and Electronics, Tra Vinh University, Tra Vinh 87000, Vietnam
[3]Department of Electronic Engineering, Feng Chia University, Taichung 40724, Taiwan, R.O.C.

Corresponding author: Song-Toan Tran (tstoan1512@tvu.edu.vn)

This work was supported in part by the Ministry of Science and Technology (MOST), Republic of China, under the contact No. of 109-2218-E-035-005. The technical supports of the computer center of Feng Chia University on GPU resources is also acknowledged.

ABSTRACT Medical image segmentation is one of the crucial tasks in diagnosis as well as pre-surgery. Recently, deep learning has significantly contributed to improving the efficiency of medical image extraction. The U-Net network has been a favored network model, which has been used as a platform architecture, for medical image segmentation. For the success of these studies, most of these models were primarily focused on the changing of the interconnection between the nodes in the network, and changing the structure of the convolution units. This would result in the ignorance of the output features of convolution units in the nodes. In this study, a U^n-Net, an n-fold network architecture, was proposed based on the traditional U-Net. In the U^n-Net model, the output features of the convolution units are taken as the skip connection. Therefore, the U^n-Net network exploits the output features of the convolution units in the nodes. In this study, we investigated a U^2-Net and a U^3-Net for segmentation of the liver and liver tumors. Besides, dilated convolution (DC) and dense structure were also used in the nodes of our networks. The efficiency of our models was evaluated on two public datasets: LiTS and 3DIRCADb. The Dice's Similarity Coefficient (DSC) of our proposed models achieved 96.38% for liver segmentation and 73.69% for tumor segmentation on the LiTS dataset. For the 3DIRCADb dataset, the results achieved 96.45% in DSC for the liver segmentation and 73.34% for the tumor segmentation. The experimental results show that our proposed networks achieved better results than the recent networks. And it is convinced that our network would be useful for practical deployments.

INDEX TERMS Dilated convolution, liver segmentation, liver tumor segmentation, medical image segmentation, U-net architecture.

I. INTRODUCTION

The liver cancer has the sixth-highest incidence of all cancers and ranks fourth in mortality in the world, with about 841,000 new cases at a rate of 93 cases per million peoples and about 782,000 deaths as 85 cases per million peoples in 2018 [1]. The challenge of liver and lesion segmentation task is to identify the voxels that depicted liver and lesion regions in medical images. In recent years, the issue of liver and tumor extraction has received considerable attention. The studies on liver and tumor segmentation have been quadrupled in amount [2].

For liver segmentation, the main challenge is the low contrast between the liver and the adjacent organs. This issue can be alleviated by windowed Hounsfield unit values [3]. The liver tumor segmentation would be more difficult than liver segmentation because the size, shape, and location of lesions are often different for each patient [4]. Moreover, the boundaries of some lesions are ambiguous like fuzzy which make it difficult to detect them by edge-based segmentation methods. Deep learning may give a chance for researchers to develop new tools to help medical doctors in segmenting organs and tumors.

In literature, the approaches for liver and tumor segmentation can be categorized into three types: manual, semi-automatic, and automatic extraction [5]. The manual segmentation, not only depends on human ability but also consumes a lot of time. This has been rarely used in practical applications after computers are introduced in the investigation tasks. Semi-automatic segmentation also requires a

The associate editor coordinating the review of this manuscript and approving it for publication was Vishal Srivastava.

person to implement with the help of automated algorithms on computers. This method saves time but depends on the computer's performance. Fully automatic extraction would be eagerly demanded to reduce down the burden of medical staff. Therefore, the automatic segmentation is the end target for researchers [6]. According to the situations in modern hospitals, liver and tumor extraction achieved nowadays in Computed Tomography (CT) technologies have shown significant improvement [7]. Automatic segmentation is promising with available image processing techniques. However, traditional image techniques still have limitations in organ or tumor segmentation. Deep learning gives the inspiration to make the automatic segmentation more accurate. Especially for smaller tumors that cannot easily be recognized by human eyes, deep learning may make smart diagnosis feasible.

Several methods of automatic segmentation of liver and lesion have been proposed, consisting of level set parameter [8], [9], fast fuzzy c-means and adaptive watershed [10], [11], fully convolutional networks (FCNs) [12]–[15], segnet [16], encoder-decoder structure [17]–[23]. The most popular encoder-decoder architecture is the U-Net model [24] that has been modified to implement a lot of applications on medical image segmentation such as ischemic stroke lesion [25], pancreas [26], [27], retina vessel [28], [29], prostate [30], colorectal tumor [31], and brain tumor [32], etc.

There were several variations of U-Net models by changing the skip connection path or the connection between the nodes. Several researchers have invoked U-Net models for medical image processing. Li *et al.* [17] combined a 2D DenseUnet network that extracted the intra-slice features and the 3D counterpart for hierarchically aggregating volumetric contexts, for liver and lesion segmentation. Jiang *et al.* [18] used a cascade structure to segment the liver tumor. They combined the soft and hard attention mechanisms, long and short skip connections. A joint dice loss function was implemented to reduce the cases of false positives. Chen *et al.* [19] introduced an end-to-end network by adding a spatial channel-wise convolution in the module of the conventional U-Net network. Ibtehaz and Rahman [33] modified the U-net model by developing the MultiRes block. This block replaced the 3×3, 5×5, and 7×7 filters in parallel by multiple 3×3 filters. They also changed the skip connection path by ''Res Path'' which consisted of 3×3 filters in the convolutional layers and 1×1 filters accompany the residual connections. Zang *et al.* [34] used multi-scale dense connections to replace the node in the traditional U-Net. The skip connection was also changed by adding more connections from the encoder node to the counterpart decoder node. Zhou *et al.* [35] changed the skip connection path by a nested path to exploit the multi-scale features in the U-Net structure. The new skip connections consisted of the convolution units which connect the others as nested. They evaluated their method with six different datasets of medical images. Huang *et al.* [36] also proposed the same skip connections, but combined the low-features with the high-features from feature maps in different scales. They used deep supervision to learn the full-scale aggregated feature maps. Liu *et al.* [37] have enhanced the Unet++ [35] by integrating the multi-scale input, multi-scale side output, and an attention mechanism segmentation on optical coherence tomography images.

Despite the U-Net's satisfactory results, there still were some limitations which are mentioned in [35]. For example, the depth of the encoder-decoder structure cannot cover all of the applications. It was case by case with application difficulties. And the skip connection is unnecessary and restrictive. There were many studies that proposed the solutions to these drawbacks. They focused on the connection between the node and skip connection renovation. However, most of the new approaches and the conventional U-Net only concentrated on the last output of the convolution node. Therefore, some features were not aware when going through the convolution node. Because of the employment of skip connections in the decoder part, the efficiency of the network is also affected. To address this problem, we proposed a new structure for the convolution nodes. All the output features are used for the next layers and the skip connections. The renovations enhance the effectiveness of deep and comprehensive learning. FIGURE 1 illustrates the difference between the traditional U-Net's node structure and our proposed structure. The renovations would be expected to enhance the effectiveness of deep and comprehensive learning.

In this study, we also considered using the dilated convolution [38] in the convolution node. The advantage of the dilated convolution is to extend the convolution region without the pooling function. Therefore, the output features can cover a wider information area while preserving spatial information. The effectiveness of dilated convolution was also verified by Pang *et al.* [12]. Finally, we use the dense structure [39] for the nodes. The dense network not only overcomes the vanishing-gradient problem but also pushes the feature propagation more efficiently. Another advantage of dense-net is the feature reuse which reduces the parameter numbers in calculation.

In summary, the main contributions of this study are listed as below:

(1) We introduced a new U-Net model architecture named U^n-Net. The convolution node structure of the new model is redesigned. This structure supplies more feature information for the decoder nodes and next convolution nodes in the U-Net structure.

(2) We investigated two versions of this structure that were U^2-Net and U^3-Net for liver and liver tumor segmentation in CT images. The power numbers, '2' and '3', indicate the number of convolutions units in the nodes of the network.

(3) We found that the introduction of dilated convolution in the convolution nodes would improve the liver and lesion segmentation results.

II. PROPOSED METHOD

This section describes the structure and the technique of the U^n-Net network in details. The loss function is an

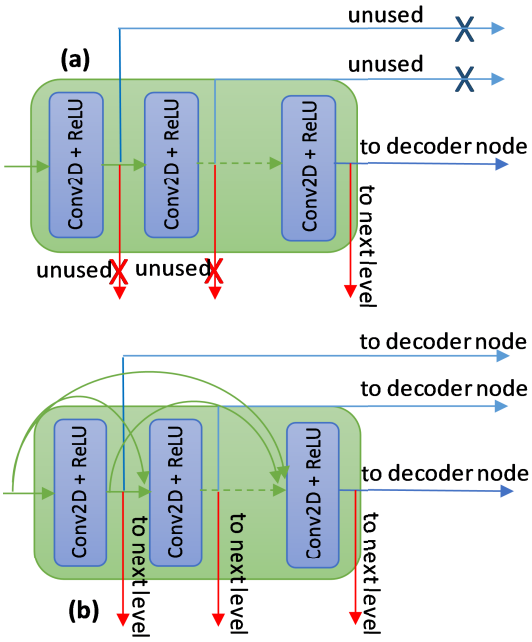

FIGURE 1. The difference between the encoder node structure of (a) traditional U-Net network and (b) our proposed network. Only the last output feature is used for the next layer and transfers to the decoder node in conventional U-Net. In the proposed network, however, all of the output features are exploited.

important factor that significantly affects the performance of the network, which has been implemented and evaluated by Xi *et al.* [22]. Hence, the detail of the loss function is also presented in this section.

A. U^n-Net NETWORK

The architecture of the U^n-Net network is presented in FIGURE 2, which was designed based on traditional U-Net. The encoder-decoder structure is the dominant factor in our model. The skip connection path, pooling path, and the up-convolution path are changed in the node structure. For the conventional U-Net and most of the U-Net based models, only the output features of the last convolution unit of the nodes are used as the input for the next layers and the decoder node. Therefore, the output features of the previous convolution units are skipped. To improve this weakness, we redesigned the structure of the nodes in the conventional U-Net network. With the new structure, all of the output features in the node are connected to the next nodes and the same level encoder nodes.

The detail of the structure of the nodes and the connectivity is illustrated in FIGURE 3. Considering the node in FIGURE 3(a), which consists of n convolution units. The convolution unit consists of a simple convolution function followed by the ReLU activation (FIGURE 3(f)). In the first encoder node N_E^1 (FIGURE 3(a)), the first convolution unit $C_E^{1,1}$ gets only the input, and the output feature is used for the next node in encoder part through the pooling path and the decoder part as skip connection path. For the next units $C_E^{1,2} \rightarrow C_E^{1,n}$,

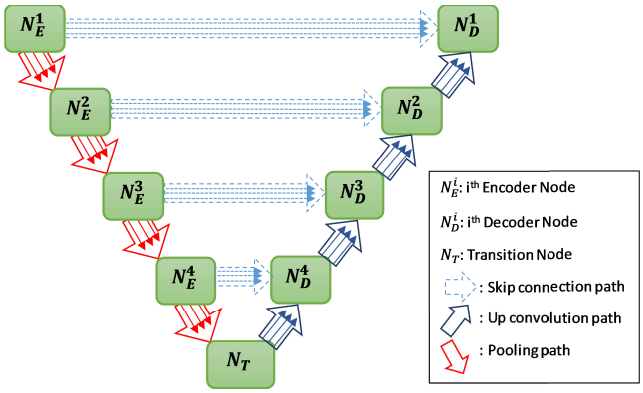

FIGURE 2. The overview structure of the U^n-Net.

the dense connections are applied. From the second to the fourth encoder node $N_E^2 \rightarrow N_E^4$ (FIGURE 3(b)), the convolution units get the pooled features from the upper nodes and concatenate with the features of previous convolution units. The transition node N_T (FIGURE 3(c)) receives the pooled features from the fourth encoder node, then transfers the output features to the fourth decoder node through up-convolution path. The decoder nodes from the second to the fourth $N_D^2 \rightarrow N_D^4$ (FIGURE 3(d)), are similar structure and connectivity, get the concatenated features from the same level skip connection path and the lower up-convolution path. The node on the top of the decoder part N_D^1 (FIGURE 3(e)) is connected to the output. In this scheme, one can find the connectivity is more complex than the other decoder nodes but provides important information for the output. The inputs of N_D^1 node are similar to the other decoder nodes but differ in the output connectivity. We implemented the deep supervision by multiple side-outputs fusion (MSOF) [40]. In this study, the output features were concatenated after the sigmoid function was applied to every output feature. The final output was also obtained by using sigmoid operation.

The renovation of the node structure would make our model be complex in interconnection. To make it easier to understand, FIGURE 4 illustrates the network architecture in detail. According to this diagram, the U^n-Net model includes n copies of U-Net networks that are arranged in parallel connection. The sub-U-Net networks are connected as $\{C_E^{1,i} \rightarrow C_E^{2,i} \rightarrow C_E^{3,i} \rightarrow C_E^{4,i} \rightarrow C_T^i \rightarrow C_D^{4,i} \rightarrow C_D^{3,i} \rightarrow C_D^{2,i} \rightarrow C_D^{1,i}\}$, where $i \in [1, n]$. The $\{f_{U_1}, f_{U_2}, \ldots, f_{U_n}\}$ denotes the output features of the sub-U-Net networks.

B. THE CONNECTIVITY DETAILS

Let $x_{E/D}^{i,j}$ denotes the output feature of the j^{th} convolution unit in the i^{th} node. The E/D indicates the node that is in the encoder/decoder sub-network. In U^n-Net network, the node consists of n outputs. Thus, the output features of the node can be defined as:

$$X_{E/D}^i = \left\{x_{E/D}^{i,1}, x_{E/D}^{i,2}, \ldots, x_{E/D}^{i,n}\right\}, \quad i \in [1; 4] \qquad (1)$$

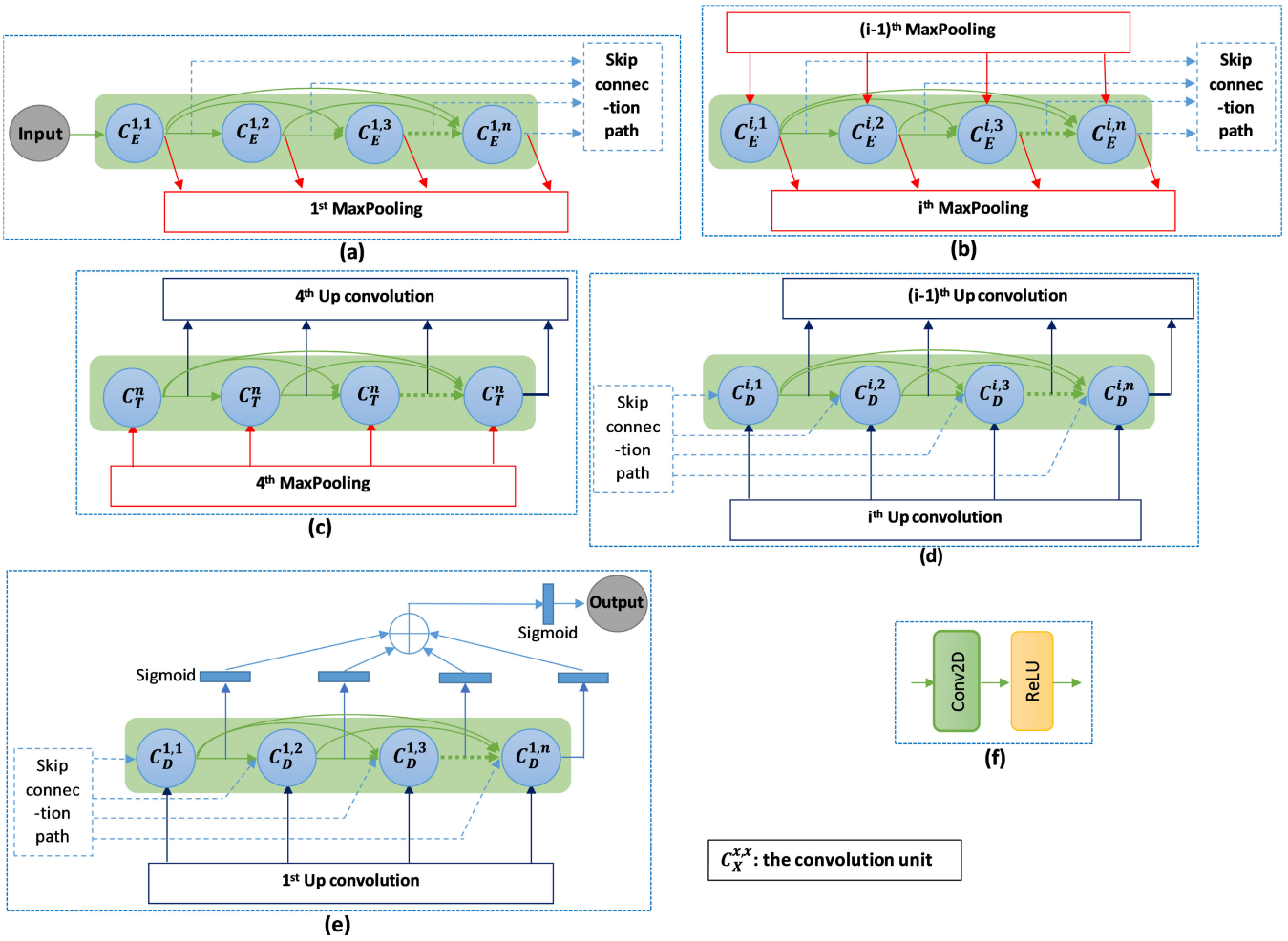

FIGURE 3. The structure and the connectivity of the node that consists of n convolution units. The illustration of (a) the first encoder node, (b) the second to the fourth encoder nodes, (c) the transition node, (d) the second to the fourth decoder nodes, (e) the first decoder node, and (f) the convolution unit. The ⊕ denotes the concatenation operation.

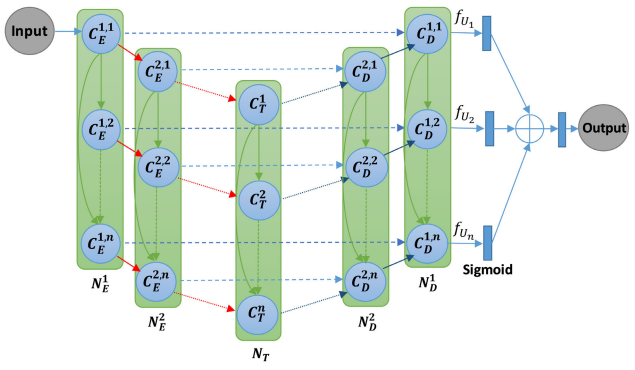

FIGURE 4. The U^n-Net architecture, which includes n of convolution units in one node, consists of *n* of sub-U-Net networks that are arranged in parallel connection. The ⊕ denotes the concatenation operation.

$$X_T = \left\{x_T^1, x_T^2, \ldots, x_T^n\right\} \tag{2}$$

where the $X_{E/D}^i$ indicates the output feature of the i^{th} node, X_T and x_T are the output features of the transition node and convolution unit in transition node, respectively. Basically, as shown in FIGURE 3(a), the convolution units of the first encoder node only receive the input from the input tensor and previous convolution units of this node. The expression for calculating the features of the first node is described as:

$$x_E^{1,j} = \begin{cases} C^j\,(input) & j = 1 \\ C^j\left(\left[x_E^{1,k}\right]_{k=1}^{j-1}\right) & j > 1 \end{cases} \tag{3}$$

where the function $C^j(\cdot)$ is a dilated convolution function with a dilation rate equal to j, and $[\cdot]$ denotes the concatenation operation. The second to the fourth encoder nodes (FIGURE 3(b)) get the inputs from the pooling path and previous convolution units in the same nodes. The output features of these nodes are computed as follows,

$$x_E^{i,j} = \begin{cases} C^j\left(P\left(x_E^{i-1,j}\right)\right) & j = 1 \\ C^j\left(\left[\left[x_E^{i,k}\right]_{k=1}^{j-1}, P\left(x_E^{i-1,j}\right)\right]\right) & j > 1 \end{cases} \quad with\ i \in [2, 4] \tag{4}$$

where the function $P(\cdot)$ denotes a max pooling function. The transition node gets the input from the fourth encoder node. The output is described as:

$$x_T^j = \begin{cases} C^j\left(P\left(x_E^{4,j}\right)\right) & j = 1 \\ C^j\left(\left[\left[x_T^k\right]_{k=1}^{j-1}, P\left(x_E^{4,j}\right)\right]\right) & j > 1 \end{cases} \quad (5)$$

In the decoder sub-network, the decoder nodes receive the inputs from the encoder sub-network and lower decoder nodes. Because of the dense structure, the convolution units also get the inputs from the previous units in the same node. The output features of the decoder nodes are computed as follows,

$$x_D^{4,j} = \begin{cases} C^j\left(\left[U\left(x_T^j\right), x_E^{4,j}\right]\right) & j = 1 \\ C^j\left(\left[\left[x_D^{4,k}\right]_{k=1}^{j-1}, U\left(x_T^j\right), x_E^{4,j}\right]\right) & j > 1 \end{cases} \quad (6)$$

$$x_D^{i,j} = \begin{cases} C^j\left(\left[U\left(x_D^{i+1,j}\right), x_E^{i,j}\right]\right) & j = 1 \\ C^j\left(\left[\left[x_D^{i,k}\right]_{k=1}^{j-1}, U\left(x_D^{i+1,j}\right), x_E^{i,j}\right]\right) & j > 1 \end{cases}$$
$$with\ i \in [1, 3] \quad (7)$$

where $U(\cdot)$ indicates a transposed convolution. The MOSF method is implemented in our study. The detail of the connectivity is illustrated in FIGURE 4. For the output features of the sub-U-Nets, $\{f_{U_1}, f_{U_2}, \ldots, f_{U_n}\}$, a 1×1 convolution and sigmoid function are applied to achieve the output results $\{Y_{U_1}, Y_{U_2}, \ldots, Y_{U_n}\}$. The final output Y_{U^n} is also obtained by using 1×1 convolution and sigmoid function. The formula is described as follow,

$$Y_{U^n} = sigmoid\left(\left[Y_{U_1}, Y_{U_2}, \ldots, Y_{U_n}\right]\right) \quad (8)$$

In this study, we investigated two versions of U^n-Net that are the U^2-Net and U^3-Net. The details of the U^2-Net and U^3-Net network architecture are shown in TABLE 1. The 3×3 convolution kernel is used for all of the convolution units. The number of filters in the encoder and decoder sub-network are similar for the nodes that are at the same level. The filter numbers are 32, 64, 128, and 256 for the first, second, third, and fourth nodes, respectively. The transition node is applied 512 of filters.

C. LOSS FUNCTION

The liver and lesion segmentation challenges have a problem that is the extreme imbalance between the background class and the foreground class. To address this problem, we use a combination of the weighted binary cross-entropy loss function and the logarithm of the dice coefficient. By using the hybrid loss, it not only handles the class imbalance problem but also smooths the gradient [41]. The hybrid loss is defined as:

$$L_{total} = L_{WBC} + L_{DC}, \quad (9)$$

where L_{WBC} and L_{DC} represent the weighted binary cross-entropy and the logarithm of dice coefficient respectively. Mathematically, the L_{WBC} and L_{DC} are expressed by

$$L_{WBC} = -\frac{1}{N}\sum_{i=1}^{N}[(1-w)\, g_i \log p_i + w\,(1-g_i)\,\log(1-p_i)] \quad (10)$$

$$L_{DC} = -\log\left(\frac{2\sum_{i=1}^{N}(g_i p_i) + \varepsilon}{\sum_{i=1}^{N}(g_i + p_i) + \varepsilon}\right) \quad (11)$$

where p_i is the probability that voxel i is predicted belongs to the foreground (liver or tumor), and g_i indicates the probability of voxel i that is the ground truth. The N depicts the total of the voxels that are predicted, the w denotes the weight attributed to the foreground class, and the ε is the smooth value.

III. EXPERIMENTS

A. DATASETS

The datasets were used in this paper are two public datasets: LiTS dataset and 3DIRCADb dataset. The LiTS dataset consists of 131 CT volumes for training and 70 testing volumes. The ground truth is not included for the testing part. In this study, therefore, we only used the training volumes for both training and testing. The slice number in the CT scans is greatly different. There are 58638 slices in total in the training part, which consists of 19163 slices containing the liver and 7183 slices containing the lesion. The CT scans were collected from seven hospitals and research institutions. The ground truth was manually created by three radiologists [2]. The parameters of the dataset, which are different, are 0.55mm to 1.0mm for in-plane resolution and 0.45mm to 6.0mm for section spacing. The 3DIRCADb dataset consists of 20 CT volumes and it is contained in the LiTS dataset (from volume 28 to volume 47) [2], [18]. Hence, we consider the LiTS dataset includes 111 volumes (after removing the 3DIRCADb volumes).

In order to train and evaluate the network performance, the LiTS dataset and 3DIRCADb were divided into three parts: ninety CT scans for training (include eighty-five LiTS dataset volumes and five 3DIRCADb volumes), eleven LiTS volumes for validation, and thirty CT scans for testing (fifteen LiTS volumes and fifteen 3DIRCADb volumes).

In the LiTS dataset, the slices, which have the liver and tumor on the image, comprise a small fraction. In order to tackle the data imbalance, we excluded 2/3 of total slices without the liver on the image. The training data include all slices that contain the liver and 1/3 of slices without the liver. There is a total of 22109 slices for training and 4494 slices for validation in our experiments.

TABLE 1. Details of the architecture of U^2-Net and U^3-Net with the dilated convolution.

Node	U^2-Net		U^3-Net	
	Encoder	Decoder	Encoder	Decoder
1	32 x (DiConv|1 + ReLU) 32 x (DiConv|2 + ReLU) [(2x2) maxpooling]^2	[(2x2) Tconv]^2 32 x (DiConv|1 + ReLU) 32 x (DiConv|2 + ReLU)	32 x (DiConv|1 + ReLU) 32 x (DiConv|2 + ReLU) 32 x (DiConv|3 + ReLU) [(2x2) maxpooling]^3	[(2x2) Tconv]^3 32 x (DiConv|1 + ReLU) 32 x (DiConv|2 + ReLU) 32 x (DiConv|3 + ReLU)
2	64 x (DiConv|1 + ReLU) 64 x (DiConv|2 + ReLU) [(2x2) maxpooling]^2	[(2x2) Tconv]^2 64 x (DiConv|1 + ReLU) 64 x (DiConv|2 + ReLU)	64 x (DiConv|1 + ReLU) 64 x (DiConv|2 + ReLU) 64 x (DiConv|3 + ReLU) [(2x2) maxpooling]^3	[(2x2) Tconv]^3 64 x (DiConv|1 + ReLU) 64 x (DiConv|2 + ReLU) 64 x (DiConv|3 + ReLU)
3	128 x (DiConv|1 + ReLU) 128 x (DiConv|2 + ReLU) [(2x2) maxpooling]^2	[(2x2) Tconv]^2 128 x (DiConv|1 + ReLU) 128 x (DiConv|2 + ReLU)	128 x (DiConv|1 + ReLU) 128 x (DiConv|2 + ReLU) 128 x (DiConv|3 + ReLU) [(2x2) maxpooling]^3	[(2x2) Tconv]^3 128 x (DiConv|1 + ReLU) 128 x (DiConv|2 + ReLU) 128 x (DiConv|3 + ReLU)
4	256 x (DiConv|1 + ReLU) 256 x (DiConv|2 + ReLU) [(2x2) maxpooling]^2	[(2x2) Tconv]^2 256 x (DiConv|1 + ReLU) 256 x (DiConv|2 + ReLU)	256 x (DiConv|1 + ReLU) 256 x (DiConv|2 + ReLU) 256 x (DiConv|3 + ReLU) [(2x2) maxpooling]^3	[(2x2) Tconv]^3 256 x (DiConv|1 + ReLU) 256 x (DiConv|2 + ReLU) 256 x (DiConv|3 + ReLU)
Transition	512 x (DiConv|1 + ReLU) 512 x (DiConv|2 + ReLU)		512 x (DiConv|1 + ReLU) 512 x (DiConv|2 + ReLU) 512 x (DiConv|3 + ReLU)	

Note that *'n x (Diconv|r + ReLU)'* denotes the 3x3 dilated convolution, *'r'* is the dilation rate, and *'n'* indicates the filter number.
The *'[(2x2) maxpooling]^X'* depicts the max pooling function is repeated *X* times.
The *'[(2x2) Tconv]^Y'* depicts the transposed convolution is repeated *Y* times. All convolutional layers include the dropout.

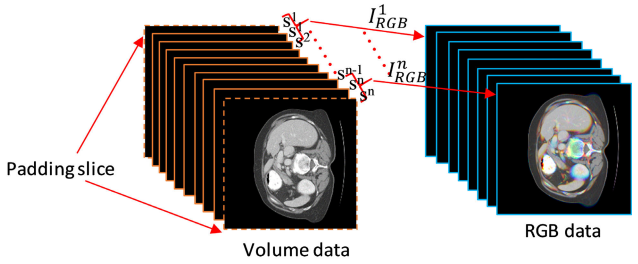

FIGURE 5. The conversion from volume data to RGB data. Three adjacent slices $\left[s^{i-1}, s^i, s^{i+1}\right]$ are combined to become an RGB image I^i_{RGB}.

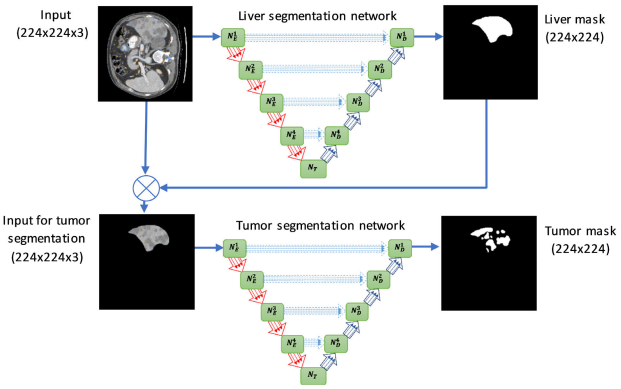

FIGURE 6. The cascade structure in our experiments. The first U^n-Net model is used to extract the liver and the second one is used for tumor segmentation. The ⊗ denotes the bitwise-and operator.

B. DATA PRE-PROCESSING

In the medical image segmentation challenge, data pre-processing is a crucial step. To enhance the contrast and remove the irrelevant organs and tissues, the Hounsfield unit values in the range of $[-200, 250]$ was applied to present the CT images. We converted three adjacent slices into the three channels image to exploit the z-dimensional information of the slices. FIGURE 5 illustrates the conversion from slices to the RGB images in detail. Let $S = \left[s^1, s^2, \ldots, s^n\right] \in R^{n\times512\times512\times1}$ is the volume data, where s^i denotes the i^{th} slice, n denotes the slice number of the volume, and $I_{RGB} = \left[I^1_{RGB}, I^2_{RGB}, \ldots, I^n_{RGB}\right] \in R^{n\times512\times512\times3}$ is the RGB data. The i^{th} RGB image I^i_{RGB} is defined as follow,

$$I^i_{RGB} = \begin{cases} \left[s^i, s^i, s^{i+1}\right], & i = 1 \\ \left[s^{i-1}, s^i, s^{i+1}\right], & i \in (1, n) \\ \left[s^{i-1}, s^i, s^i\right], & i = n \end{cases} \quad (12)$$

To improve the fraction of the foreground region and reduce the computation time, we cropped the images to a size of $448 \times 448 \times 3$ then re-scaled to the size of $224 \times 224 \times 3$. Finally, we applied the min-max normalization on every batch of images to input the network.

C. TRAINING PROCESS

Our training process consists of two stages. Firstly, the liver segmentation was implemented. In the second stage, we used the best weight from liver segmentation model to train the tumor segmentation model. The cascade structure was implemented in our system, this architecture was used in many studies such as [16], [18], [20], [40], and [41]. The advantage of the cascade structure is to reduce the cases of false positives. FIGURE 6 illustrates the cascade structure was implemented in our experiments in detail. The first U^n-Net network is used to segment the liver from the original image. The liver mask from the first model is multiplied with the original image to get the input for the second U^n-Net that is tumor segmentation model.

In the first stage, we used the training strategy with two steps. Firstly, the model under training was fed with a selected dataset only with slices containing the liver. Those slices without liver inside were ignored in this stage. After a set of converged weighting parameters was obtained, then in the second stage, these best weight in the first stage were employed to train the model by feeding all the slices as the general dataset no matter the slices containing the liver or not. This strategy is named ''easy-to-hard'' (E2H). The effectiveness of this strategy will be presented in section IV-B. For lesion segmentation stage, the best weight from the first step is also applied to train the network.

D. EXPERIMENTAL SETTING

Our experiments are conducted by the Keras package with Tensorflow as the backend. The he-normal distribution initializer, which was proposed by He *et al.* [42], is used to initialize the weights. The optimizer is used in the model is Adam optimizer, and the networks are trained with a learning rate of 3e-4 for 20 first epochs, 1e-5 for the next 20 epochs, and 3e-5 for the last 20 epochs. For the second step in the first stage (liver segmentation stage), the learning rate is 3e-5 in 30 epochs. The models are trained with a batch size of 8, and a dropout rate of 0.2 is applied for preventing over-fitting. We also implemented an early-stopping mechanism for training stages. All experiments are powered by a workstation with Intel Xeon Silver 4114 CPU, GRID Virtual GPU V100D-8Q, and 32GB of RAM memory.

IV. RESULTS AND DISCUSSION

To authenticate the effectiveness and robustness of our proposed networks, we implemented the conventional U-Net and another U-Net network that includes three convolution units in the node, named by U3C-Net. The U-Net++ [35] is also implemented in our experiments for comparison with our proposed networks. In addition, the effectiveness of the training strategy, the dilated convolution, and node structure are detailed in this section. In summary, we implemented a total of eight models that include U-Net, U-Net++, U3C-Net, U3C-Net+, U^2-Net, U^2-Net+, U^3-Net, and U^3-Net+. The U^2-Net+, U3C-Net+, and U^3-Net+ indicate that the dilated convolution is applied on the U^2-Net, U3C-Net, and U^3-Net models, respectively.

A. EVALUATION METRICS

In order to evaluate the segmentation results, the evaluation metrics were selected, which consist of three evaluation values from all the metrics: Dice's Similarity Coefficient (DSC), Volumetric Overlap Error (VOE), and Relative Volume Difference (RVD) as employed in [2]. The equations of the three metrics are expressed by

$$DSC(G, P) = \frac{2\,|G \cap P|}{|G| + |P|} \tag{13}$$

$$VOE(G, P) = 1 - \frac{|G \cap P|}{|G \cup P|} \tag{14}$$

$$RVD(G, P) = \frac{|P| - |G|}{|G|} \tag{15}$$

where G denotes the case number of the ground truth and P is that of positive prediction. The smaller value of *VOE* and *RVD* indicates a better segmentation result. For Dice's Similarity Coefficient, the greater value close to 1 indicates a better result.

B. EFFECTIVENESS OF THE TRAINING STRATEGY

The training strategy affects the performance of the model. In this study, we applied training by the easy-to-hard (E2H) strategy. The results showed that the performance improves significantly. FIGURE 7 shows the effectiveness of training strategy on liver segmentation. Observing the results, we found that, by comparing with the training from scratch, the DSC values of E2H method improved 0.46%, 0.35%, 0.93%, 0.61%, 0.83%, 0.58%, 0.28%, and 0.77%, respectively, in corresponding to U-Net, U^2-Net, U^2-Net+, U3C-Net, U3C-Net+, U-Net++, U^3-Net, and U^3-Net+ networks on the LiTS dataset. And the DSC became 0.44%, 0.66%, 0.98%, 0.86%, 1.5%, 0.32%, 0.44%, and 0.98% on the 3DIRCADb dataset.

It was noted in our experience in the training stage, the convergence of the loss function varied a lot and sometimes early-stop was employed to obtain the optimal results. This fact might depend on the fine tune of the learning rate and the initial guess of the weighting parameters in the network models. A better way for training may invoke a pre-trained backbone such as VGG [43] or ResNet [44]. Therefore, training the model with the popular datasets may help to create better initialization weights for the full learning process. In our experience, the training strategy has a significant effect on the performance of the models. Our E2H strategy can be regarded as a self-transfer learning technique. However, the techniques of transfer learning might be out of the scope of this paper. The effects of transfer learning will be an interesting topic in related studies.

C. EFFECTIVENESS OF DILATED CONVOLUTION

We also analyze the effect of DC on the segmentation results. All models used for comparison are trained with the same training settings and strategies. We compare DSC results on both the LiTS and the 3DIRCADb dataset. As shown in FIGURE 8, for the liver segmentation, it is obvious that the models that applied the DC get better DSC results than the models without the DC. As seen in FIGURE 8, the DSC of the models that use the DC increased by 0.8%, 0.26%, and 1.28% on the LiTS dataset and 0.39%, 0.94%, and 0.63% on the 3DIRCADb dataset on U^2-Net, U3C-Net, and U^3-Net, respectively. For lesion segmentation, the DSC metrics of the models that use the DC, increased by 0.23%, 2.6%, and 1.18% on the LiTS dataset and 4.54%, 2.06%, and 0.56% on the 3DIRCADb dataset on U^2-Net, U3C-Net, and U^3-Net, respectively.

From FIGURE 8, one can find the exploit of DC can improve the accuracy. The dominant characteristic of DC

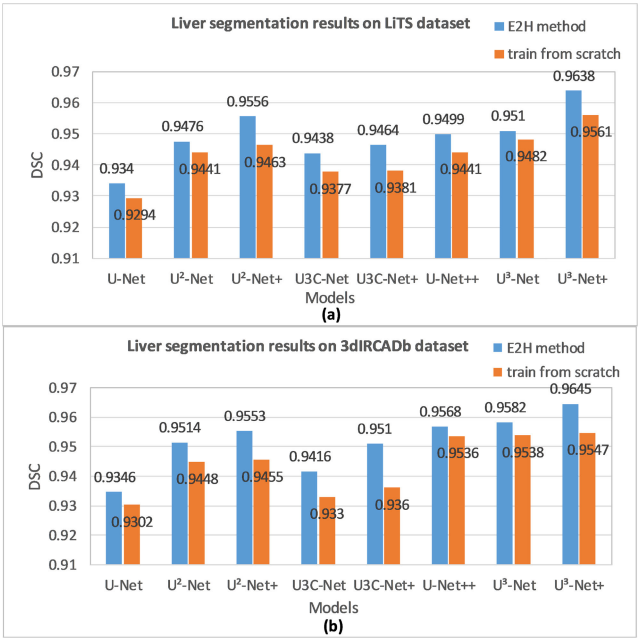

FIGURE 7. Illustrate the effectiveness of the E2H method on liver segmentation of (a) LiTS dataset and (b) 3DIRCADb dataset.

is to expand the convolutional region. Therefore, the output features are covered in a larger area with the same number of parameters. We believed DC can improve the accuracy as the results we have obtained. However, the dilation rate does not have much effect on small tumors. Larger objects such as liver and large tumors can be efficiently segmented by dilated convolution. In our experiments, the dilation rates were 1, 2, and 3 for the first, the second, and the third convolution units, respectively. The improvement of dilated convolution on the segmentation performance can be up to 4.54% in FIGURE 8(d) for the lesion segmentation by the U^2-Net and 2.6% in FIGURE 8(c) by U3C-Net.

FIGURE 9 illustrates the feature maps of Unet and the proposed method (U^2-Net+ and U^3-Net+). In FIGURE 9(a), we can observe that some information on features in the Unet model is lost after some nodes. Whereas the U^2-Net+ and U^3-Net+ still keep the semantic information. This finding would prove the effectiveness of the dilated convolution. The final segmentation results are presented in FIGURE 9(b). We can see that the segmentation results from our proposed networks are better than Unet model.

D. LIVER AND LIVER TUMOR SEGMENTATION RESULTS

The details of the liver and lesion segmentation results are shown in TABLE 2 and TABLE 3. TABLE 2 shows the segmentation results on the LiTS dataset while TABLE 3 details the results on the 3DIRCADb dataset. As seen in this table, the U-Net model has the worst performance. The segmentation results are better when we apply three convolution units for the node (U3C-Net). The U-Net++ achieves better results than U-Net, U3C-Net, U3C-Net+, and U^2-Net. The obvious results show that our proposed network, U^3-Net+, achieves

TABLE 2. The details of segmentation results on the LiTS dataset. All metrics (mean ± std.) are in %.

	Liver			Tumor		
	DSC	VOE	RVD	DSC	VOE	RVD
U-Net	93.40± 0.13	11.53± 0.99	3.39± 0.08	64.93± 2.04	49.12± 19.49	-30.31± 8.33
U-Net++	94.99± 0.07	9.08± 0.56	2.07± 0.02	72.49± 0.96	39.12± 10.68	-16.35± 6.49
U3C-Net	94.38± 0.10	10.61± 0.76	2.47± 0.06	67.64± 1.84	45.45± 16.06	-22.27± 8.12
U3C-Net+DC	94.64± 0.09	9.37± 0.62	2.26± 0.04	70.24± 1.48	42.31± 13.54	-19.29± 8.01
U^2-Net	94.76± 0.08	9.34± 0.61	2.21± 0.03	70.38± 1.33	42.38± 13.05	-19.05± 8.09
U^2-Net+DC	95.56± 0.06	8.83± 0.58	2.05± 0.01	70.61± 1.16	41.09± 11.80	-18.34± 7.66
U^3-Net	95.10± 0.07	8.69± 0.52	2.05± 0.01	72.51± 0.95	38.66± 10.34	-16.46± 7.42
U^3-Net+DC	**96.38± 0.06**	**6.36± 0.41**	**1.99± 0.01**	**73.69± 0.86**	**37.80± 9.82**	**-15.78± 6.82**

TABLE 3. The details of extraction results on the 3DIRCADb dataset. All metrics (mean ± std.) are in %.

	Liver			Tumor		
	DSC	VOE	RVD	DSC	VOE	RVD
U-Net	93.46± 0.11	11.79± 1.02	3.36± 0.09	61.93± 2.36	54.27± 17.08	-32.98± 8.83
U-Net++	95.68± 0.05	8.07± 0.81	2.04± 0.01	70.22± 0.97	43.79± 11.19	-17.54± 6.56
U3C-Net	94.16± 0.09	11.21± 0.72	2.41± 0.06	64.83± 1.98	50.08± 15.31	-26.71± 8.24
U3C-Net+DC	95.10± 0.07	9.15± 0.64	2.23± 0.05	66.89± 1.80	48.07± 13.48	-23.03± 8.21
U^2-Net	95.14± 0.06	7.51± 0.61	2.07± 0.03	67.05± 1.57	46.13± 12.88	-22.31± 8.02
U^2-Net+DC	95.53± 0.06	7.50± 0.52	2.05± 0.02	71.59± 1.13	41.16± 10.92	-18.82± 7.45
U^3-Net	95.82± 0.05	7.27± 0.54	2.03± 0.02	72.78± 0.94	39.12± 10.25	-17.17± 7.11
U^3-Net+DC	**96.45± 0.04**	**6.14± 0.44**	**1.98± 0.01**	**73.34± 0.84**	**37.34± 9.12**	**-15.82± 6.30**

the best result on both segmentation tasks and on both the datasets. The U^2-Net+ network also gave good results indicating that it may be better than U-Net++ for liver segmentation in the LiTS dataset and for lesion segmentation on the 3DIRCADb dataset. The possible reason can be attributed that the livers in the LiTS dataset were larger in size. Therefore, DC significantly leads to better segmentation efficiency. This is similar to the case of tumors in 3DIRCADb.

FIGURE 10 presents the examples of extraction results obtained by U-Net, U3C-Net, U^2-Net+, U-Net++, and U^3-Net+ on the LiTS dataset. The results prove that our proposed networks, U^2-Net+ and U^3-Net+, achieve better results than other methods. FIGURE 11 illustrates the results achieved by implementation the networks on the 3DIRCADb dataset. We observed that the liver and lesion can be segmented well by our models.

E. COMPARISON OF CONVENTIONAL U-Net AND U^2-Net, U3C-Net AND U^3-Net

Our proposed method exploits the output features from all of the convolution units. We observe in TABLE 2 and TABLE 3,

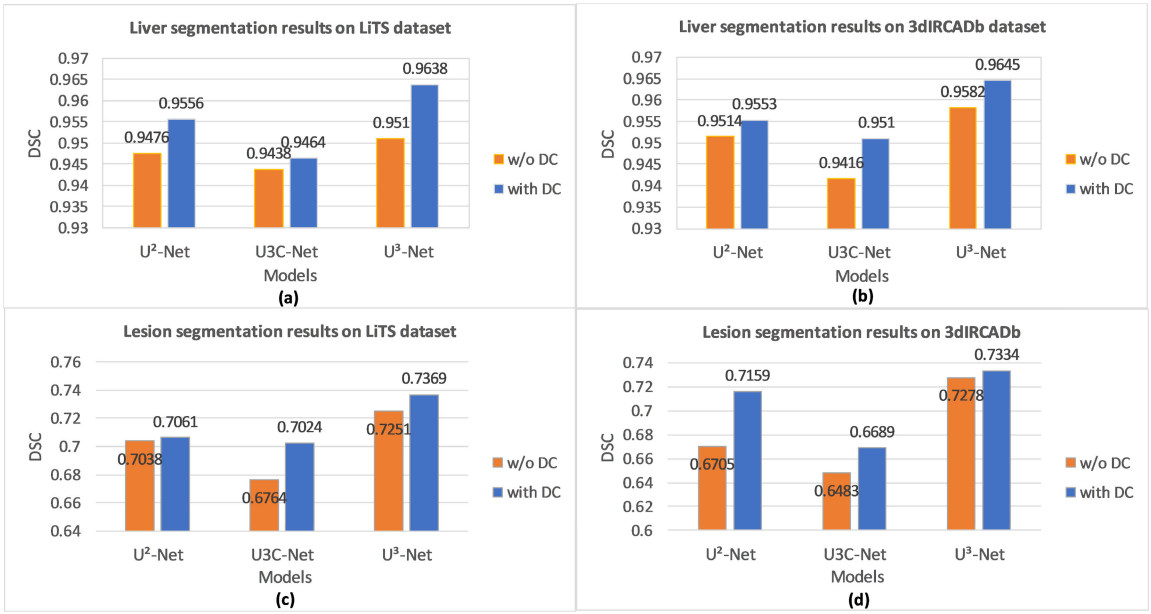

FIGURE 8. Illustrate the effectiveness of dilated convolution (DC) on liver and lesion segmentation. (a) and (c) show the results on the LiTS dataset. (b) and (d) present the results on the 3DIRCADb dataset.

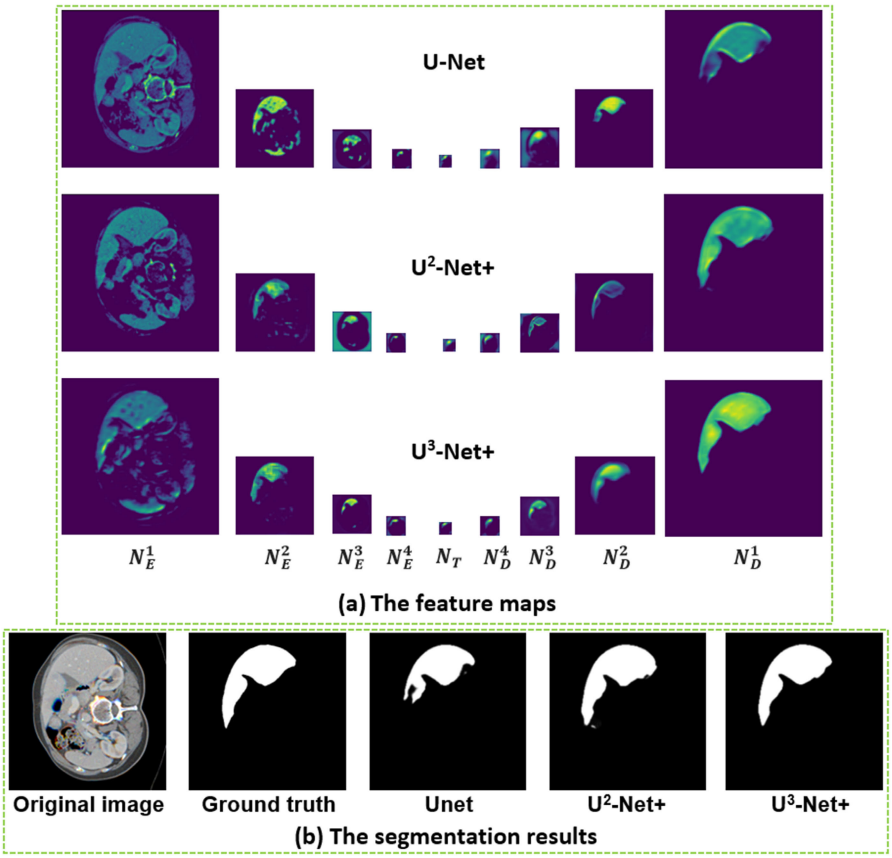

FIGURE 9. Visualization of (a) feature maps from each layer in different models; and (b) the related segmentation results by U-Net, U^2-Net+, and U^3-Net+.

the segmentation results of U^2-Net and U^3-Net are better than the results of U-Net and U3C-Net on both challenges and both datasets. For U-Net and U^2-Net, there are two convolution units in the node. However, the DSC of liver and lesion segmentation of the U^2-Net is greater by 1.36% and 5.45% respectively on the LiTS dataset than U-Net. Similarly, on the 3DIRCADb dataset, the improvements on the liver and tumor segmentation are 1.68% and 5.12% of DSC, respectively.

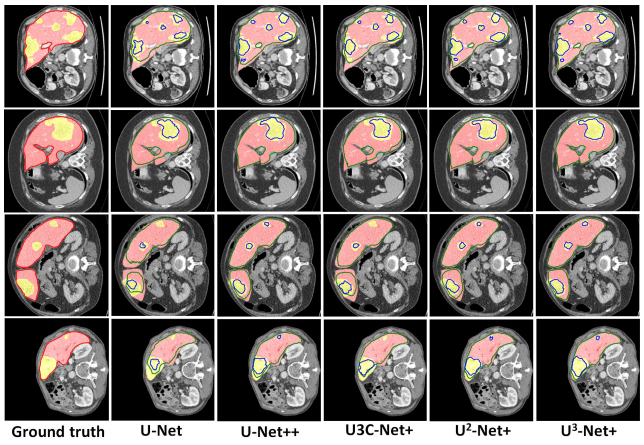

FIGURE 10. The slices in the LiTS dataset are segmented by the networks that are implemented in this study. The ground truth is presented in the first column. The second to the sixth columns show the segmentation results of the networks. The red areas depict the true liver regions and the yellow ones show the true lesion areas. The red curves indicate the true boundaries of the livers. The green curves denote the boundaries of segmented livers while the blue ones show the boundaries of the segmented tumors.

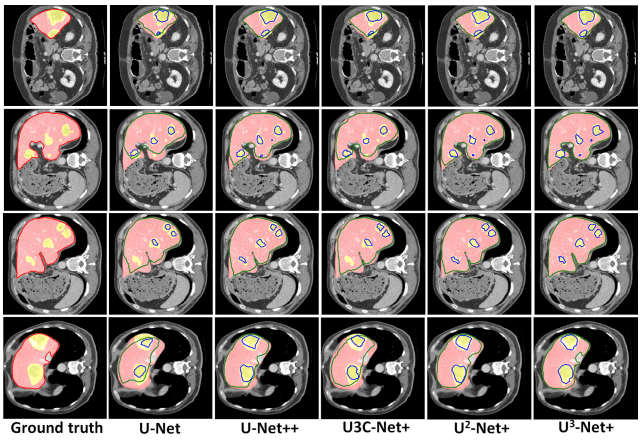

FIGURE 11. The slices in the 3DIRCADb dataset are segmented by the networks that are implemented in this study.

The U3C-Net and U^3-Net also have the same number of convolution units in the node. However, the metric values of U^3-Net are much better than U3C-Net. Specifically, the dice score of liver and tumor segmentation improve 0.72% and 4.87% respectively on LiTS dataset. The increments of DSC of liver and tumor on the 3DIRCADb dataset are 1.66% and 7.95%, respectively. For the VOE and RVD value, the proposed models have also achieved better results than U-Net and U3C-Net. The experimental results demonstrate that the exploitation of output features had a significant effect on the performance of the networks.

FIGURE 12 presents the training loss and validation loss during the training process. As seen in FIGURE 12, the loss functions decrease quickly at the beginning and keep below 0.1 stably which indicates good convergence of the parameters during calculations. Furthermore, one can find our

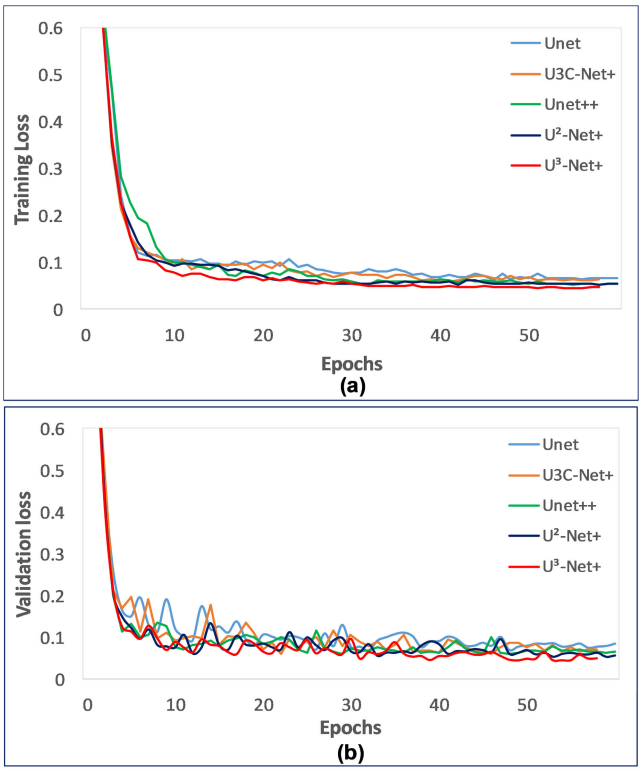

FIGURE 12. The learning curves of five models implemented in our experiments on liver segmentation for (a) training loss and (b) validation loss.

proposed models, U^2-Net+ and U^3-Net+, yield the lowest training and validation loss. The above findings would prove the effectiveness of our proposed models for practical applications.

TABLE 4 compares the parameters number and the calculation time of the models that are implemented in our experiments. As shown in TABLE 4, our proposed networks are more complex than the traditional U-Net. For detail, the parameter numbers increase 3.8 million, from 7.7 million (U-Net) to 11.5 million (U^2-Net) and 10 million, from 11.7 million (U3C-Net) to 21.7 million (U^3-Net). Therefore, the training time per epoch and the testing time per slice have also increased. One would concern that U^2-Net, U^2-Net+, and U^3-Net would have much more parameters in the convolution nodes. In this study, however, we got less training and testing time than that in U-Net++ (325s, 350s, and 480s against 510s). This fact can be attributed to the complexity of the U-Net++ connection. The U^3-Net+ has a large training time and test time (580s and 14.5s) due to the large calculation time for DC. Even for U^3-Net+ with the dilation convolution, the calculation times were still acceptable, only less than 20% and 30% of time increase in training and testing. In the meanwhile, the capacity of memory requirements maybe doubled in our network as compared to U-Net++. As the cost of memory may be not a critical concern, we think our networks can be feasible for practical applications. Furthermore, for the medical imaging segmentation task, these increments are not

TABLE 4. The parameters number, training time, and testing time of each method.

	Number of parameters	Training time per epoch (second)	Testing time per slice (millisecond)
U-Net	7.7 M	300	6.5
U-Net++	9.5 M	510	11
U3C-Net	11.7 M	345	8
U3C-Net+DC	11.7 M	390	10
U^2-Net	11.5 M	325	7.5
U^2-Net+DC	11.5 M	350	9.5
U^3-Net	21.7 M	480	11.5
U^3-Net+DC	21.7 M	570	14.5

TABLE 5. Comparing with popular networks on the LiTS dataset.

	Liver			Tumor		
	DSC (%)	VOE (%)	RVD (%)	DSC (%)	VOE (%)	RVD (%)
U-Net	93.40	11.53	3.39	64.93	49.12	-30.31
U-Net++	94.99	9.08	2.07	72.49	39.12	-16.35
CU-Net	89.40	-	-	59.5	-	-
RA-UNet	96.10	7.40	2.00	59.5	38.90	-15.2
Xi *et al.* [22]	94.90	9.50	2.10	75.2	37.90	-15.9
U^2-Net+DC	95.56	8.83	2.05	70.61	41.09	-18.34
U^3-Net+DC	**96.38**	**6.36**	**1.99**	**73.69**	**37.80**	**-15.78**

TABLE 6. Comparing with popular networks on the 3DIRCADb dataset.

	Liver			Tumor		
	DSC (%)	VOE (%)	RVD (%)	DSC (%)	VOE (%)	RVD (%)
U-Net	93.46	11.79	3.36	61.93	54.27	-32.98
Christ *et al.*	94.30	10.70	-1.4	-	-	-
U-Net++	95.68	8.07	2.04	70.22	43.79	-17.54
AHC-Net	95.30	-	-	66.80	1.354	12.9
ResNet	93.80	11.65	-3.00	67.00	45.00	40.00
U^2-Net+DC	95.53	7.50	2.03	71.59	41.16	-18.82
U^3-Net+DC	**96.45**	**6.14**	**1.98**	**73.34**	**37.34**	**-15.82**

significant. The crucial mission of medical image segmentation task is to increase segmentation performance.

In this study, the cascade structure was used to reduce the case number of false positives. At the first glance, this approach might have a drawback of heavy computation. This concern was not a serious problem. In fact, the worst case in our proposed models can process an image with a maximum of only 14.5ms as seen in TABLE 4. The average slice number of the CT volume was about 500 slices. The average total calculation time for a patient in our cascade structure would be 14.5 seconds only. This fact indicates that our model may be applicable to real clinical scenarios. In future studies, we try using simultaneously training combine with the post-processing methods.

F. COMPARISON OF PERFORMANCE WITH OTHER NETWORKS

To prove the effectiveness and feasibility of our proposed networks, we compare the segmentation results with popular networks. TABLE 5 compares the segmentation results of our models with U-Net [24], U-Net ++ [35], CU-Net [20], RA-Unet [23], and Cascade U-ResNets [22] models on the LiTS dataset. For the 3DIRCADb dataset, the U-Net, U-Net++, AHC-Net [18], ResNet [45], and the model that was proposed by Christ *et al.* [46] are compared with our models. The details of the comparisons of the networks on the 3DIRCADb are presented in TABLE 6.

Based on the results, one can find that our models outperformed the other models on the evaluation parameters except for the DSC for liver tumors on the LiTS. In this study, we did not use the post-processing methods, even in the preprocessing, only simple image processing algorithms were employed. Therefore, it is more convenient and simpler to train and test by our models. However, in the meanwhile, there is also a drawback of our networks, i.e., if the number of convolution unit increases, the connection in the model will become cumbersome.

V. CONCLUSION

In conclusion, this paper introduces a new network architecture to exploit the intra-features of the node in U-Net architecture. We also propose two models for liver and tumor segmentation in CT scan images. By innovating the architecture of the convolution node, the intra-features are used more effectively. The flexibility of the new network has been demonstrated in this study. Specifically, the variation of the convolution unit number makes the model easier to fit into problems with different numbers of data samples. In addition, we discover the training strategy also improves the segmentation results. We further demonstrate the effectiveness of dilated convolution on the model performance. We believe that this model can be applied to other types of medical images such as PET or ultrasound. We would like to recommend our network for other medical image segmentation tasks.

REFERENCES

[1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA, Cancer J. Clinicians*, vol. 68, no. 6, pp. 394–424, Nov. 2018.

[2] P. Bilic *et al.*, "The liver tumor segmentation benchmark (LiTS)," 2019, *arXiv:1901.04056*. [Online]. Available: http://arxiv.org/abs/1901.04056

[3] H. Lee, M. Kim, and S. Do, "Practical window setting optimization for medical image deep learning," 2018, *arXiv:1812.00572*. [Online]. Available: http://arxiv.org/abs/1812.00572

[4] M. Moghbel, S. Mashohor, R. Mahmud, and M. I. B. Saripan, "Review of liver segmentation and computer assisted detection/diagnosis methods in computed tomography," *Artif. Intell. Rev.*, vol. 50, no. 4, pp. 497–537, Dec. 2018.

[5] A. Gotra, L. Sivakumaran, G. Chartrand, K.-N. Vu, C. Kauffmann, S. Kadoury, F. Vandenbroucke-Menu, B. Gallix, J. A. D. Guise, and A. Tang, "Liver segmentation: Indications, techniques and future directions," *Insights Imag.*, vol. 8, no. 4, pp. 377–392, Aug. 2017.

[6] P. Campadelli, E. Casiraghi, and A. Esposito, "Liver segmentation from computed tomography scans: A survey and a new algorithm," *Artif. Intell. Med.*, vol. 45, nos. 2–3, pp. 185–196, Feb. 2009.

[7] S. A. Azer, "Deep learning with convolutional neural networks for identification of liver masses and hepatocellular carcinoma: A systematic review," *World J. Gastrointestinal Oncol.*, vol. 11, no. 12, pp. 1218–1230, Dec. 2019.

[8] Z. Deng, Q. Guo, and Z. Zhu, "Dynamic regulation of level set parameters using 3D convolutional neural network for liver tumor segmentation," *J. Healthcare Eng.*, vol. 2019, pp. 1–17, Feb. 2019.

[9] 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," *Med. Image Anal.*, vol. 37, pp. 46–55, Apr. 2017.

[10] 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, pp. 1–11, Sep. 2017.

[11] A. M. Anter and A. E. Hassenian, "CT liver tumor segmentation hybrid approach using neutrosophic sets, fast fuzzy C-means and adaptive watershed algorithm," *Artif. Intell. Med.*, vol. 97, pp. 105–117, Jun. 2019.

[12] Y. Pang, D. Hu, and M. Sun, "A modified scheme for liver tumor segmentation based on cascaded FCNs," in *Proc. Int. Conf. Artif. Intell., Inf. Process. Cloud Comput. (AIIPCC)*, Sanya, China, Dec. 2019, pp. 1–6.

[13] S. Zheng, B. Fang, L. Li, M. Gao, Y. Wang, and K. Peng, "Automatic liver tumour segmentation in CT combining FCN and NMF-based deformable model," in *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*. Milton Park, U.K.: Taylor & Francis, Jun. 2019, pp. 1–10.

[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," *Artif. Intell. Med.*, vol. 83, pp. 58–66, Nov. 2017.

[15] G. Chlebus, A. Schenk, J. H. Moltz, B. van Ginneken, H. K. Hahn, and H. Meine, "Automatic liver tumor segmentation in CT with fully convolutional neural networks and object-based postprocessing," *Sci. Rep.*, vol. 8, no. 1, p. 15497, Dec. 2018.

[16] S. Almotairi, G. Kareem, M. Aouf, B. Almutairi, and M. A.-M. Salem, "Liver tumor segmentation in CT scans using modified SegNet," *Sensors*, vol. 20, no. 5, p. 1516, Mar. 2020.

[17] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P.-A. Heng, "H-DenseUNet: Hybrid densely connected UNet for liver and tumor segmentation from CT volumes," *IEEE Trans. Med. Imag.*, vol. 37, no. 12, pp. 2663–2674, Dec. 2018.

[18] H. Jiang, T. Shi, Z. Bai, and L. Huang, "AHCNet: An application of attention mechanism and hybrid connection for liver tumor segmentation in CT volumes," *IEEE Access*, vol. 7, pp. 24898–24909, Feb. 2019.

[19] Y. Chen, K. Wang, X. Liao, Y. Qian, Q. Wang, Z. Yuan, and P.-A. Heng, "Channel-unet: A spatial channel-wise convolutional neural network for liver and tumors segmentation," *Frontiers Genet.*, vol. 10, p. 1110, Nov. 2019.

[20] A. A. Albishri, S. J. H. Shah, and Y. Lee, "CU-Net: Cascaded U-Net model for automated liver and lesion segmentation and summarization," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, San Diego, CA, USA, Nov. 2019, pp. 1416–1423.

[21] H. Seo, C. Huang, M. Bassenne, R. Xiao, and L. Xing, "Modified U-Net (mU-Net) with incorporation of object-dependent high level features for improved liver and liver-tumor segmentation in CT images," *IEEE Trans. Med. Imag.*, vol. 39, no. 5, pp. 1316–1325, May 2020.

[22] X.-F. Xi, L. Wang, V. S. Sheng, Z. Cui, B. Fu, and F. Hu, "Cascade U-ResNets for simultaneous liver and lesion segmentation," *IEEE Access*, vol. 8, pp. 68944–68952, Apr. 2020.

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

[24] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," 2015, *arXiv:1505.04597*. [Online]. Available: http://arxiv.org/abs/1505.04597

[25] J. Dolz, I. B. Ayed, and C. Desrosiers, "Dense multi-path U-Net for ischemic stroke lesion segmentation in multiple image modalities," 2018, *arXiv:1810.07003*. [Online]. Available: http://arxiv.org/abs/1810.07003

[26] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: Learning where to look for the pancreas," 2018, *arXiv:1804.03999*. [Online]. Available: http://arxiv.org/abs/1804.03999

[27] J. Li, X. Lin, H. Che, H. Li, and X. Qian, "Probability map guided bi-directional recurrent UNet for pancreas segmentation," 2019, *arXiv:1903.00923*. [Online]. Available: http://arxiv.org/abs/1903.00923

[28] T. Laibacher, T. Weyde, and S. Jalali, "M2U-Net: Effective and efficient retinal vessel segmentation for resource-constrained environments," 2018, *arXiv:1811.07738*. [Online]. Available: http://arxiv.org/abs/1811.07738

[29] R. Li, M. Li, J. Li, and Y. Zhou, "Connection sensitive attention U-NET for accurate retinal vessel segmentation," 2019, *arXiv:1903.05558*. [Online]. Available: http://arxiv.org/abs/1903.05558

[30] W. Chen, Y. Zhang, J. He, Y. Qiao, Y. Chen, H. Shi, and X. Tang, "Prostate segmentation using 2D bridged U-Net," 2018, *arXiv:1807.04459*. [Online]. Available: http://arxiv.org/abs/1807.04459

[31] Y.-J. Huang, Q. Dou, Z.-X. Wang, L.-Z. Liu, Y. Jin, C.-F. Li, L. Wang, H. Chen, and R.-H. Xu, "3D RoI-aware U-Net for accurate and efficient colorectal tumor segmentation," 2018, *arXiv:1806.10342*. [Online]. Available: http://arxiv.org/abs/1806.10342

[32] J. Zhang, X. Lv, H. Zhang, and B. Liu, "AResU-Net: Attention residual U-net for brain tumor segmentation," *Symmetry*, vol. 12, no. 5, p. 721, May 2020.

[33] N. Ibtehaz and M. S. Rahman, "MultiResUNet : Rethinking the U-Net architecture for multimodal biomedical image segmentation," *Neural Netw.*, vol. 121, pp. 74–87, Jan. 2020.

[34] J. Zhang, Y. Jin, J. Xu, X. Xu, and Y. Zhang, "MDU-Net: Multi-scale densely connected U-Net for biomedical image segmentation," 2018, *arXiv:1812.00352*. [Online]. Available: http://arxiv.org/abs/1812.00352

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

[36] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Y.-W. Chen, and J. Wu, "UNet 3+: A full-scale connected UNet for medical image segmentation," in *Proc. ICASSP-IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Barcelona, Spain, May 2020, pp. 1055–1059.

[37] W. Liu, Y. Sun, and Q. Ji, "MDAN-UNet: Multi-scale and dual attention enhanced nested U-Net architecture for segmentation of optical coherence tomography images," *Algorithms*, vol. 13, no. 3, p. 60, Mar. 2020.

[38] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," 2015, *arXiv:1511.07122*. [Online]. Available: http://arxiv.org/abs/1511.07122

[39] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," 2016, *arXiv:1608.06993*. [Online]. Available: http://arxiv.org/abs/1608.06993

[40] D. Peng, Y. Zhang, and H. Guan, "End-to-end change detection for high resolution satellite images using improved UNet++," *Remote Sens.*, vol. 11, no. 11, p. 1382, Jun. 2019.

[41] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso, "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," 2017, *arXiv:1707.03237*. [Online]. Available: http://arxiv.org/abs/1707.03237

[42] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Santiago, Chile, Dec. 2015, pp. 1026–1034.

[43] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556

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

[45] X. Han, "Automatic liver lesion segmentation using a deep convolutional neural network method," *Med. Phys.*, vol. 44, no. 4, pp. 1408–1419, Apr. 2017.

[46] P. F. Christ, F. Ettlinger, F. Grün, J. Lipkova, M. E. A. Elshaera, S. Schlecht, F. Ahmaddy, S. Tatavarty, M. Bickel, P. Bilic, M. Rempfler, F. Hofmann, M. D Anastasi, S.-A. Ahmadi, G. Kaissis, J. Holch, W. Sommer, R. Braren, V. Heinemann, and B. Menze, "Automatic liver and tumor segmentation of CT and MRI volumes using cascaded fully convolutional neural networks," 2017, *arXiv:1702.05970*. [Online]. Available: http://arxiv.org/abs/1702.05970

SONG-TOAN TRAN was born in Tra Vinh, Vietnam, in 1984. He received the B.S. degree from Can Tho University (CTU), Can Tho City, Vietnam, in 2007, and the M.S. degree from the Ho Chi Minh University of Technology (HCMUT), Ho Chi Minh City, Vietnam, in 2013. He is currently pursuing the Ph.D. degree in electrical and communication engineering with Feng Chia University (FCU), Taichung, Taiwan, R.O.C. His research interests include medical image processing, deep learning, computer vision, and virtual reality-augmented reality and applications.

CHING-HWA CHENG (Member, IEEE) is currently a Full Professor of electronic engineering with Feng Chia University. His research interests include endoscope system design, image processing, and low-power VLSI design/EDA/testing related issues, such as image guiding surgery and low-power chip designed by multi-voltage. His research interests include theoretical model, design flow development, cell library generation, physical chip implementation, and system validation. His work differs from that of other researchers in this field in that most of his research designs are silicon proven by a practical system.

DON-GEY LIU was born in Tainan, Taiwan, in 1963. He received the B.S. degree from the Department of Electrical Engineering, National Taiwan University (NTU), in 1986, the M.S. degree from the Institute of Electrical Engineering, National Tsing Hua University (NTHU), in 1988, and the Ph.D. degree from the Institute of Electronics, National Chiao Tung University (NCTU), in 1992, respectively.

He served for the Chinese Air Forces as a Lieutenant from 1992 to 1994. He was an Officer in charge of maintaining wireless equipment. In 1994, he joined the Electronics Research and Service Organization, Industrial Technology Research Institute (ERSO/ITRI) in developing flat-panel displays by field-emission devices (FEDs). In 1998, he was invited as a Consultant of process of semiconductor wastes in research laboratories and hi-tech companies in Taiwan. Then, he went to the Department of Electronic Engineering, Feng Chia University (FCU), Taichung, Taiwan, as an Associate Professor. He has been a Professor since 2001 and the Head of the Department from 2004 to 2008. Since 2014, he has also been serving as the Director of the Ph.D. Program of Electrical and Communications Engineering. Since 2017, he has also been the Director of the Master Program of Biomedical Informatics and Biomedical Engineering. He has filed six patents, and published more than 40 journal articles, and 70 conference papers. His current research interests include high-speed integrated circuit design for RF and analog applications. He has been awarded more than 30 research contracts and grants from government agencies, such as National Science Council (NSC) and Ministry of Education (MOE), and industries in Taiwan.

Dr. Liu served as a member of Technical Program Committee for the International Meeting on Information Display (IMID) held in Daegu, South Korea, from 2007 to 2010. He also was a member of the Board of Reviewers for proposals in NSC from 2004 to 2007. He received several awards in teaching and research from Feng Chia University and MOE. Since 2010, he has been invited as a Keynote Speaker of several International conferences held in Mainland China. In addition, he serves as a Technical Reviewer for several articles in the Institute of Electrical and Electronic Engineering (IEEE) and the Chinese Institute of Electrical Engineering (CIEE) journals. From 2004 to 2006, he was the Co-Chairperson in holding National Competition of Communication Researches. In his term of chairperson, he conducted the department to obtain the accreditation of Institute of Engineering Education, Taiwan (IEET), which has joined in Washington Covenants. In 2017, he also helped the Advisory Review of the IEET accreditation for Zhongshan Polytechnic University, Guan Dong, China.