Cascaded MultiTask 3-D Fully Convolutional Networks for Pancreas Segmentation

Jie Xue, Kelei He, Dong Nie, Ehsan Adeli, Zhenshan Shi, Seong-Whan Lee, Yuanjie Zheng, Xiyu Liu, Dengwang Li, and Dinggang Shen, *Fellow, IEEE*

***Abstract*—Automatic pancreas segmentation is crucial to the diagnostic assessment of diabetes or pancreatic cancer. However, the relatively small size of the pancreas in the upper body, as well as large variations of its location and shape in retroperitoneum, make the segmentation task challenging. To alleviate these challenges, in this article, we propose a cascaded multitask 3-D fully convolution network (FCN) to automatically segment the pancreas. Our cascaded network is composed of two parts. The first part focuses on fast locating the region of the pancreas, and the second part uses a multitask FCN with dense connections to refine the segmentation map for fine voxel-wise segmentation. In particular, our multitask FCN with dense connections is implemented to simultaneously complete tasks of the voxel-wise segmentation and skeleton extraction from the pancreas. These two tasks are complementary, that is, the extracted skeleton provides rich information about the shape and size of the pancreas in retroperitoneum, which can boost the segmentation of pancreas. The multitask FCN is also designed to share the low- and mid-level features across the tasks. A feature consistency module is further introduced to enhance the connection and fusion of different levels of feature maps. Evaluations on two pancreas datasets demonstrate the robustness of our proposed method in correctly segmenting the pancreas in various settings. Our experimental results outperform both baseline and state-of-the-art methods. Moreover, the ablation study shows that our proposed parts/modules are critical for effective multitask learning.**

***Index Terms*—Multitask FCN, pancreas segmentation, skeleton extraction.**

Manuscript received October 25, 2018; revised March 22, 2019 and October 7, 2019; accepted November 18, 2019. Date of publication December 18, 2019; date of current version March 17, 2021. The work of J. Xue was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 61802234 and Grant 61640201, in part by the China Post-Doctoral Project under Grant 2017M612339, and in part by the China Scholarship Council under Grant 201708370073. The work of Y. Zheng was supported by NSFC under Grant 81871508. The work of X. Liu was supported by NSFC under Grant 61876101. The work of D. Li was supported by NSFC under Grant 61773246. This article was recommended by Associate Editor D. Goldgof. *(Corresponding author: Dinggang Shen.)*

J. Xue is with the Business School, Shandong Key Laboratory of Medical Physics and Image Processing, Shandong Normal University, Jinan 250014, China, and also with the Department of Radiology and Biomedical Research Imaging Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27510 USA.

K. He is with the Medical School, National Institute of Healthcare Data Science, Nanjing University, Nanjing 210023, China.

D. Nie and E. Adeli are with the Department of Radiology and Biomedical Research Imaging Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27510 USA.

Z. Shi is with the Department of Radiology, First Affiliated Hospital of Fujian Medical University, Fuzhou 350005, China.

S.-W. Lee is with the Department of Brain and Cognitive Engineering, Korea University, Seoul 02841, South Korea.

Y. Zheng is with the School of Information Science and Engineering, Shandong Normal University, Jinan 250014, China.

X. Liu is with the Business School, Shandong Normal University, Jinan 250014, China.

D. Li is with the School of Physics and Electronics, Shandong Normal University, Jinan 250014, China.

D. Shen is with the Department of Radiology and Biomedical Research Imaging Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27510 USA, and also with the Department of Brain and Cognitive Engineering, Korea University, Seoul 02841, South Korea (e-mail: dgshen@med.unc.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCYB.2019.2955178.

Digital Object Identifier 10.1109/TCYB.2019.2955178

I. Introduction

PANCREATIC cancer, like ductal adenocarcinoma, has a high mortality rate with a low five-year survival rate, and is one of the most challenging cancers to treat [3]. Patients are frequently examined by the early parenchyma phase abdominal CT [7]. In upper abdominal surgery, such as laparoscopic gastrectomy or pancreatectomy, the location of the pancreas is required for enabling safer surgical procedure [10] whereas, manual delineation for the pancreas is time consuming and often irreproducible. Therefore, there is a calling need for developing an efficient computer-aided segmentation method to help physicians diagnose and assess the progression of diabetes or pancreatic cancer, as done in other applications [11]–[14]. However, accurate segmentation of pancreas is challenging due to the following two reasons: 1) the pancreas has especially large intersubject variability in its location, size, and shape (see Fig. 1) and 2) the pancreas has a thinner shape in the abdomen, compared with other abdominal organs, as can also be observed in Fig. 1. As shown in Fig. 1, the intensities of voxels in the pancreatic region are very similar to those of the neighboring structures (i.e., the stomach wall, duodenum, and intestines).

With the recent advances, deep-learning methods obtained superior performance in the segmentation of medical images [15]–[20]. For accurate pancreas segmentation, several convolutional-neural-network-based (CNN) [21] methods have been developed. Generally, we can classify the existing deep-learning frameworks for this segmentation task into two kinds, that is, 1) the one-stage methods and 2) the two-stage methods [2], [4]–[6], [8], [9], [22]. The one-stage methods directly segment organ(s) in a whole image, while the two-stage methods first locate the organ(s) and then perform segmentation on the localized region(s).

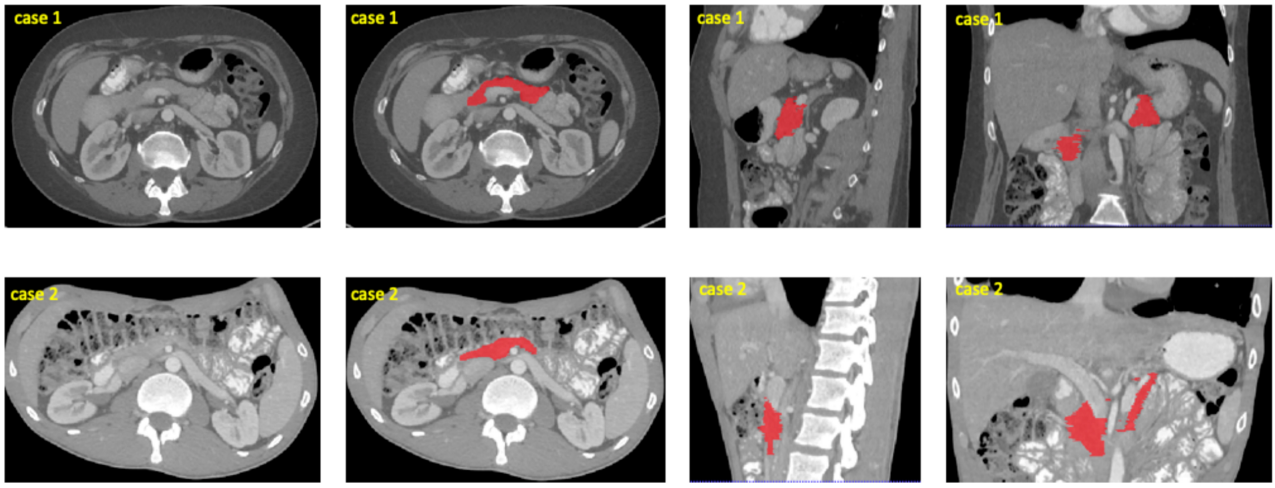

Fig. 1. Two examples of the pancreas from the NIH dataset [1], shown in three views (raw CT image, and the axial, sagittal and coronal views from left to right, respectively). Pancreas regions are shown as red, from which we can observe. 1) large shape, appearance and location variations of the pancreas across cases; 2) thinner shape in the abdomen; and 3) similar intensities in pancreas and neighboring structures.

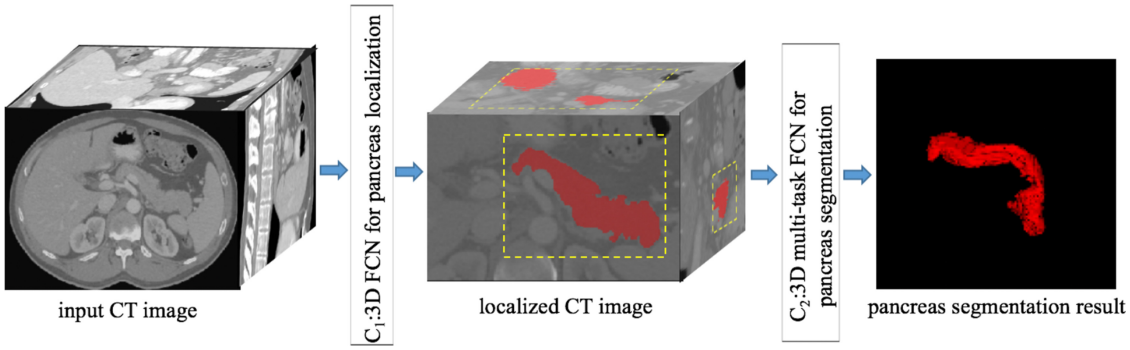

Fig. 2. Flowchart of our proposed cascaded multitask FCN for automatic pancreas localization and segmentation.

As the one-stage method, Farag *et al.* [2] used a CNN model with dropout [24] to conduct a classification of the pancreas and nonpancreas regions [25]. In another work, Cai *et al.* [9] added a convolutional long short-term memory (LSTM) [26] network to the output layer of CNN to complete the segmentation on 2-D slices of the pancreas. These methods directly apply CNNs on the entire CT images. However, pancreas is relatively small in size (i.e., less than 0.5% of the entire CT volume [4]). For such organs, deep-learning methods can be disrupted by the nontarget region, which often occupies a large fraction in the abdomen CT images. Therefore, the segmentation results are often not satisfied. To overcome the above-mentioned challenge, the two-stage methods have been developed. These methods only focused on the target region (i.e., the pancreas), which allows achieving more accurate segmentation [5] in the second stage. For example, Roth *et al.* [6] presented a two-stage method to localize and segment pancreas, respectively. Specifically, they employed holistically nested convolutional networks (HNNs) [27] on three views to do the task. A similar strategy was adopted in Zhou *et al.* [4] by applying the fixed-point models to shrink the input region. In their method, a 2-D fully convolution network (FCN) model was used. Besides, Yu *et al.* [8] added a recurrent saliency transformation module into the coarse-to-fine model, which achieved the best performance among all existing methods in terms of dice ratio. But, all of these methods merge contexts of different views of 2-D slices of CT images for segmentation, which unavoidably miss some spatial information across slices. Recently, Zhu *et al.* [5] proposed a 3-D coarse-to-fine segmentation method by using 3-D U-net [28] with residual connections. Similarly, Roth *et al.* [22], [23] employed a 3-D U-net with concatenation and summation skip connections to segment pancreas. Our method is also a two-stage framework which can also focus on small organ regions, and also works on 3-D by using 3-D patches of CT volumes as the input to leverage spatial information along all three axes.

Existing methods for pancreas segmentation [2], [4]–[6], [8], [9], [22] mainly use standard segmentation approaches in the literature for medical (or natural) image segmentation, ignoring problem-specific challenges (i.e., varying locations and shapes of the pancreas across different subjects). We argue that to segment the detailed and fine structures, like pancreas, shape-specific cues can significantly improve the segmentation performance.

To this end, in this article, we propose a cascaded 3-D FCN, composed of two major cascaded stages (see Fig. 2 for an overview of our approach). In the first part of the cascade (i.e., C_1 in Fig. 2), we concentrate on fast locating the

region of the pancreas, since pancreas is relatively small in size in the CT volume. Based on the obtained region, in the second part (C_2), we construct a novel multitask FCN with dense connections to adopt guidance from the organ skeleton to improve the accuracy and stability of the final segmentation, which consists of two branches with shared learned features. One branch of the deep multitask network is a regression network, aiming at describing the shapes of the pancreas, and another branch segments the pancreas. We use skeleton as a shape representation to more accurately and efficiently represent the pancreas shapes in the abdomen CT images. Extraction of object skeletons from images has been well studied and successfully applied to shape-based object matching and recognition [29]–[31]. The skeleton is a useful structure-based object descriptor, which can deliver significant information about the presence, shape, and size of the object [32]–[34], since the shapes of the pancreas are variable across different subjects and also thinner than other abdominal organs. Moreover, pancreas can be found as two separate parts in some axial views. In our method, we describe the pancreas with skeletons to capture its shape and preserve its geometric properties. Since extracting skeletons and segmenting pancreas are interrelated, the multitask [35]–[41] framework can be used to optimize both tasks and boost performance for pancreas segmentation. Besides, we also propose a feature consistency module to further enhance the connection and fusion of different levels of feature maps for improving the performance. To remove small false segments, we finally employ 3-D fully connected conditional random field (CRF) [42]–[45] as a post-processing step for pancreas segmentation.

For comparison, we evaluate our approach on the NIH pancreas segmentation dataset [1], which has been adopted by many previous methods. The average dice similarity coefficient (DSC) of our method reaches 86.4%, outperforming those previous methods. To check the contribution of each proposed module in our method, we also evaluate it on the NIH dataset. Besides the NIH dataset, we further evaluate our method on another in-house dataset, to show the robustness of our method across different datasets.

The contributions of this article can be summarized as follows.

1) We propose a cascaded shape-specific cues-guided FCN, composed of two major cascaded stages. The first stage focuses on fast locating the region of pancreas. The second stage employs a 3-D multitask dense-U-Net architecture to perform accurate segmentation on the located pancreas region. With this two-stage method, small pancreas can also be accurately segmented.
2) We adopt guidance from the pancreas skeleton to help the segmentation network to better learn the segmentation task. In particular, the estimated pancreas skeleton can provide coarse shape information, which can alleviate both issues of the low contrast in the boundary and the high geometric variability of pancreas in the CT images.
3) We propose a novel 3-D multitask framework for volumetric pancreas segmentation. Our proposed 3-D segmentation framework can leverage rich spatial information along all three axes for accurate segmentation.

II. Method

The overall framework of our proposed method is shown in Fig. 2, where a cascaded multitask 3-D FCN is proposed to first automatically localize the pancreas region(s) and then segment the located pancreas in detail. Since the raw CT image of the upper body contains a large region while the target pancreas is relatively small, we design a cascaded framework to first conduct an initial segmentation in a coarse level, and then segment pancreas in a fine level. The first stage of the cascaded framework, denoted as C_1, is implemented with an FCN, which is designed to localize the pancreas from the raw CT image. The details of C_1 are introduced in Section II-A. Then, a multitask FCN with dense connections is utilized in the second stage of the cascaded framework, denoted as C_2, to accurately segment pancreas based on the detected region(s) from C_1. This multitask FCN consists of two interrelated steps, that is, pancreas skeleton extraction and segmentation. Here, the pancreas skeleton serves as supplementary guidance to help the segmentation network to better learn the segmentation task. The details of C_2 are introduced in Section II-B. Afterward, a 3-D fully connected CRF is employed as a post-processing step to achieve smoother predictions, as described in Section II-C.

A. Pancreas Localization using 3-D FCN

We design C_1 to localize pancreas in raw CT images, which will propose regions for C_2. This can be considered as a coarse segmentation of pancreas and can be regarded as a binary-classification problem. In particular, FCN is devised to fast segment pancreas from the down-sampled CT image. Specifically, the original CT image is down-sampled to 1/4 of its original resolution in our case. After inference with the proposed FCN, we upsample the coarse segmentation result to the original resolution. Then, the pancreas region can be obtained from this upsampled segmentation result.

We use a variant of FCN, U-Net [28], as the base architecture of our network, which is illustrated in Fig. 3 (cascade C_1). The entire network contains a contracting path and an expanding path. The contracting path is consist of three blocks, each containing one or two convolutional layer(s), followed by a max-pooling layer with a kernel size of $2 \times 2 \times 2$. Each $3 \times 3 \times 3$ convolutional layer [46] with strides and padding of one is followed by a rectified linear unit (ReLU) [47]. The expanding path includes the same number of blocks as the contracting path. Each block has one transposed convolutional layer with several convolutional layers. The transposed convolutional layer has the kernel size of $2 \times 2 \times 2$ with strides of two. The output feature maps of the deconvolutional layer are concatenated with the feature maps of the convolutional layer in the corresponding scale of the contracting path. Then, output features are fed into the subsequent convolutional layers.

We employ patch-wise training, rather than entire-image training, due to a small number of training samples. Particularly, in this article, the images are cropped to 3-D patches with the size of $16 \times 64 \times 64$, according to the effect

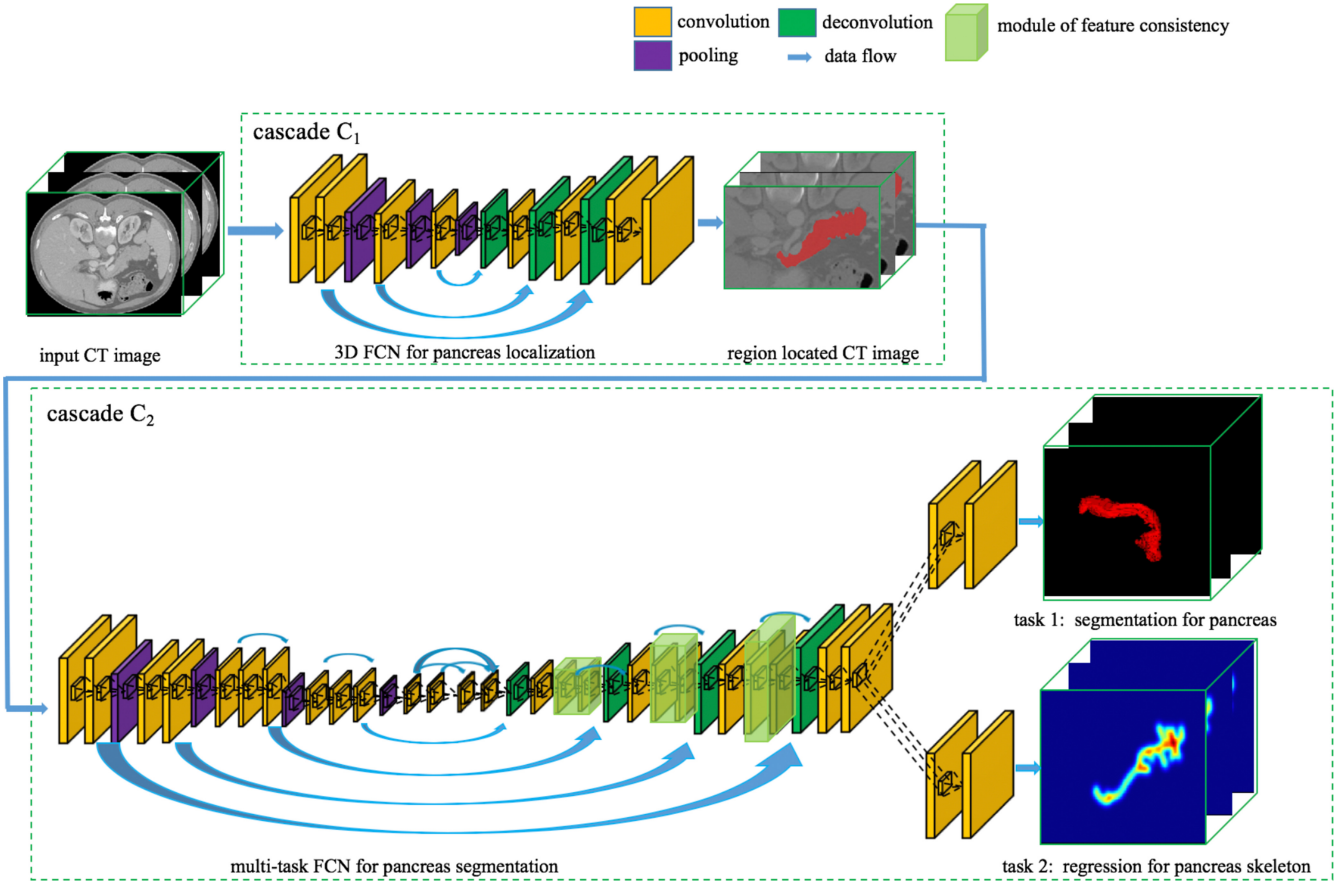

Fig. 3. Network architecture of multitask 3-D FCN for pancreas segmentation.

of patch sizes analyzed in Section III. Then, the bounding box of pancreas is obtained by morphological operation on this coarse segmentation. The intermediate point of the bounding box is selected as the centroid to crop the region of size $128\times224\times224$ from the raw CT images. The region size is ensured to cover the entire pancreas. This located region is finally fed into the next stage of the cascaded framework, C_2.

B. Pancreas Segmentation Using 3-D MultiTask FCN

In the second stage of the cascaded framework, C_2, we use a multitask FCN with dense connections for accurate pancreas segmentation, based on the localized region(s) from C_1. As discussed earlier, due to variable shapes of the pancreas across different patients, the guidance from the estimated pancreas skeleton can help better segment pancreas; on the other hand, the segmentation result can also help estimate pancreas skeleton. The network architecture is outlined in Fig. 3 (cascade C_2).

1) Pancreas Skeleton: One of the main difficulties of pancreas segmentation is their variable shapes across different patients. To deal with this issue, we estimate the skeleton of pancreas to provide a reliable reference for the pancreas shape, thus helping the segmentation network to better capture morphological information of pancreas in the CT image. Note that the ground-truth skeleton is only needed in the training phase, which can be actually obtained by morphological operation, followed by smoothing with a Gaussian filter.

2) Network Architecture: We adopt a 3-D multitask dense-U-Net to conduct the pancreas segmentation. To make full use of features and also strengthen feature propagation, we employ dense connections in the network [48], [49]. Especially, we first crop patches from the proposed region (generated by C_1). Then, we feed the cropped patches into the network. Since the two tasks of estimating pancreas skeleton and segmenting pancreas are interrelated, they are designed to share the entire encoder and a part of the decoder. In particular, the encoder part of the network includes five blocks of convolutional layers and pooling layers (See the left part of cascade C_2 in Fig. 3). Each block includes several convolutional layers with a kernel size of $3\times3\times3$, followed by a pooling layer with the size of $2\times2\times2$. The decoder is consist of a corresponding number of blocks, each with one transposed convolutional layer, followed by several convolutional layer(s). The transposed convolutional layer has the kernel size of $2\times2\times2$ with strides of two. After the decoder (see the right part of cascade C_2 in Fig. 3), each task has two extra convolutional layers to continue learning the task-specific features. Note that ReLU is adopted as the activation function for each convolutional layer.

Through the structure of U-net, we can make full use of context information from the coarse feature maps learned in low-level layers (in the encoder part) and fuse it with fine information from the dense feature maps learned in the high-level layers (in the decoder part). To enhance the fusion of

these low-level and high-level features, we propose to use several modules of feature consistency (denoted by three green cubes in Fig. 3), which consist of additional convolutional layers, to generate more precise output after the concatenation.

In the learning process, the segmentation loss is defined by the cross entropy loss (L_{CE}), and the regression loss by the Euclidean loss (L_{REG}), respectively. Hence, the final loss of the network is

$$L = \lambda_1 L_{\mathrm{CE}} + \lambda_2 L_{\mathrm{REG}} \tag{1}$$

with

$$L_{\mathrm{CE}} = -\frac{1}{m}\left[\sum_{i=1}^{m}\sum_{j=1}^{k} I\left\{y^{(i)}, j\right\} \log \frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^{k} e^{\theta_l^T x(i)}}\right] \tag{2}$$

$$I\{y^{(i)}, j\} = \begin{cases} 1, & y^{(i)} = j \\ 0, & y^{(i)} \neq j \end{cases} \tag{3}$$

$$L_{\mathrm{REG}} = \frac{1}{2m}\sum_{i=1}^{m}(p_i - g_i)^2 \tag{4}$$

where λ_1 and λ_2 are the weights of L_{CE} and L_{REG} in our experiments. m is the number of subjects. In (2), k is the total number of classes. x and y represent the input data and labels, respectively. l is the layer index, and θ is the parameter. In (4), p_i is the predict value, and g_i is the corresponding ground-truth value.

C. Post-Processing of Pancreas Segmentation by 3-D Fully Connected Conditional Random Field

Although the segmentation results of 3-D FCN are smooth, there are still some small isolated regions, caused by the independent and identically distributed inference of the network. We employ a 3-D fully connected CRF [42]–[45] as a post-processing step, which is able to connect all pairs of individual voxels in the image to refine the boundaries between pancreas and background and also remove isolated false positive segmentation.

For an input image I and the ground-truth segmentation S, the Gibbs energy in a CRF model is given by (5)

$$G(S) = \sum_i \Psi_\mu(S_i) + \sum_{i,j,i\neq j} \psi_p(S_i, S_j) \tag{5}$$

$$\psi_\mu(S_i) = -\log P(S_i|I). \tag{6}$$

$P(S_i|I)$ is the FCN's output for voxel i

$$\psi_p(S_i, S_j) = \mu(S_i, S_j) k(f_i, f_j) \tag{7}$$

$$\mu(S_i, S_j) = [S_i = S_j] \tag{8}$$

$$k(f_i, f_j) = \omega^{(1)} k^{(1)}(f_i, f_j) + \omega^{(2)} k^{(2)}(f_i, f_j) \tag{9}$$

where $\omega^{(1)}$ and $\omega^{(2)}$ define the weights of $k^{(1)}$ and $k^{(2)}$. k is a linear combination of Gaussian kernels (9), which is defined over an arbitrary feature space, with f_i, f_j being the feature vectors of the pair of voxels. There are two types of k, that is, one is the smoothness function $k^{(1)}$, and another is the appearance function $k^{(2)}$. $k^{(1)}$ and $k^{(2)}$ are defined in (10) and (11), respectively

$$k^{(1)}(f_i, f_j) = \exp\left(-\sum_{d=\{x,y,z\}} \frac{|p_{i,d} - p_{j,d}|^2}{2\sigma_{\alpha,d}^2}\right) \tag{10}$$

where $p_{i,d}$ represents the voxel coordinates and $\sigma_{\alpha,d}$ denotes the size and shape of neighbors that same labels are inspired

$$k^{(2)}(f_i, f_j) = \exp\left(-\sum_{d=\{x,y,z\}} \frac{|p_{i,d} - p_{j,d}|^2}{2\sigma_{\beta,d}^2} - \frac{|I_i - I_j|^2}{2\sigma_\gamma^2}\right) \tag{11}$$

where σ_γ can be viewed as how strong to implement the same appearance.

III. Experiment and Discussion

A. Data Acquisition

We compare our proposed method with previous methods on the NIH pancreas dataset [1]. The NIH pancreas segmentation dataset includes 82 contrast-enhanced abdominal CT scans. The image size is $512 \times 512 \times (181 \sim 466)$ with slice thickness as $1.5 - 2.5$ mm, acquired on Philips and Siemens MDCT scanners (120 kVp tube voltage). We also applied our proposed method to our own dataset (denoted as the Fujian Medical University (FMU) dataset), which was collected from the First Affiliated Hospital of FMU, China and approved by the Institutional Review Board of FMU. All experiments were performed in compliance with the Declaration of Helsinki. Written informed consent was acquired from each patient or next of kin. This dataset has a total number of 59 contrast-enhanced abdominal CT volumes, along with manual delineations of the pancreas by experienced physicians. The size of CT volumes in our own dataset is $512 \times 512 \times (37–224)$ with slice thickness as $2.5 - 5.0$ mm, acquired on Toshiba Aquilion one 320 CT kv 120 MA 193. Different from the NIH healthy pancreas dataset, some subjects in our own dataset include benign/malignant pathological cysts, which impact the morphology of the pancreas [50], thus making this dataset extremely challenging for segmentation due to large variation. Our experiments are conducted on splitting of 82 patients (from NIH) and 59 patients (from FMU) into their own four folds of 20, 20, 21, and 21 patients, and 15, 15, 15, and 14 cases, respectively. In each round of 4-fold standard cross-validation (CV-4), we employ three folds of data as training cases and the remaining fold for testing. 10% of the training cases are randomly selected for validation.

B. Evaluation Metrics

To evaluate the segmentation performance, four metrics (DSC, Jaccard similarity coefficient, precision, and recall) are used, with their definitions given below. Let V_s denote the voxel set of automatic segmentation volume, and V_g denote the voxel set of ground-truth volume. DSC and Jaccard metrics are means of measuring the correct overlap between the automatic segmentation and the ground-truth segmentation. Precision (or positive predictive value) and recall (or sensitivity) measure the fractions of relevant segmented voxels

$$\mathrm{DSC} = \frac{2||V_s \cap V_g||}{||V_s + V_g||} \tag{12}$$

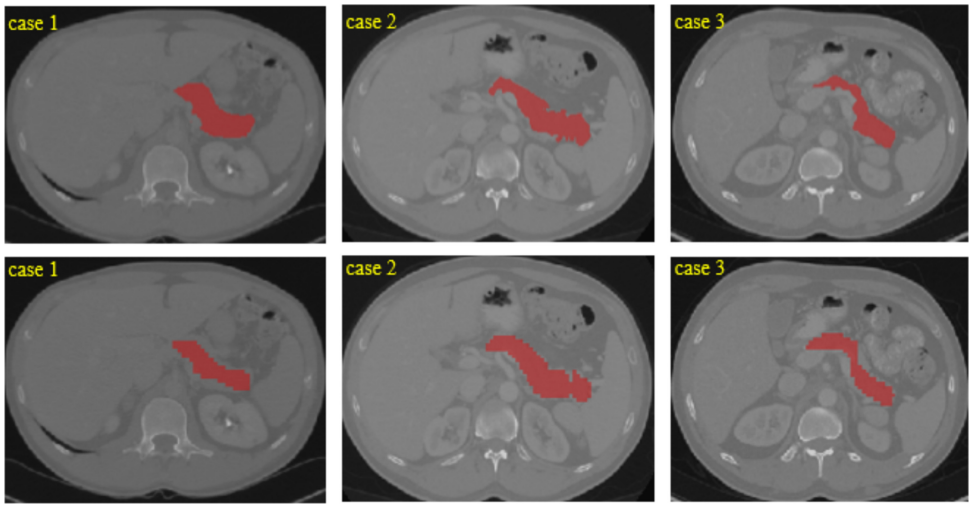

Fig. 4. Coarse pancreas segmentation results by the cascade C_1 for three different examples (case 1, case 2, and case 3). The first row shows the ground truth, and the second row shows the coarse segmentations.

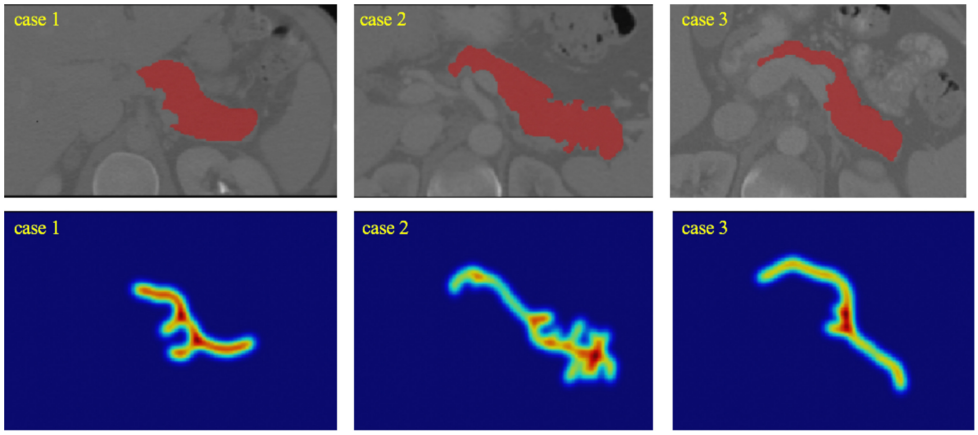

Fig. 5. Proposed regions (first row) and pancreas skeletons (second row) for images shown in Fig. 4.

$$\text{Jaccard} = \frac{||V_s \cap V_g||}{||V_s \cup V_g||} \tag{13}$$

$$\text{precision} = \frac{||V_s \cap V_g||}{||V_s||} \tag{14}$$

$$\text{recall} = \frac{||V_s \cap V_g||}{||V_g||}. \tag{15}$$

C. Parameters Setting

Our method was implemented based on the widely used open-source framework Caffe [51] customized to support 3-D operations for all necessary layers [17]. We train our network via the standard stochastic gradient descent (SGD) algorithm using a step learning rate. The learning rate is initialized at 10^{-2} and decayed over the training iterations with a rate of 10^{-1} until it reaches 10^{-6}. We use the momentum of 0.9 to make a tradeoff between the last observed image and the newly observed image. The batch size is 20. All the network parameters are initialized by Xavier's [52] method.

We resample all images to a unified resolution (i.e., $1 \times 1 \times 1$ mm^3) and then crop the image to delete nonabdomen regions by selecting the maximum connected area automatically. The input data from each image is decremented by the means of the whole image first. Then, we normalize intensities into the range of $(-1, 1)$ by dividing the maximum intensity value.

In the training process, we randomly crop patches through the whole image in the cascade C_1 and the images of pancreas regions in the cascade C_2. We employ the patch size of $16\times64\times64$ as the input image size with the consideration of computational expenses. In the testing phase, we crop patches with a fixed step size of $4\times16\times16$, as explained in detail in the following section.

D. Evaluation of Pancreas Localization on the NIH Dataset

Recall that the cascade C_1 conducts coarse segmentation of the pancreas (Fig. 4). The DSC value of pancreas in C_1 is 71.9%. Although the DSC value of pancreas in C_1 is not high, the proposed region can cover the whole pancreas (Fig. 5). Pancreas skeletons are also shown in Fig. 5. Moreover, the efficiency is significantly improved by using the down-sampled image, as the computational time of C_1 on an NVIDIA TITAN XP GPU is only 4 seconds.

E. Evaluation of Pancreas Segmentation on the NIH Dataset

1) Comparison With the State-of-the-Art Methods: In this section, we compare our proposed method with the

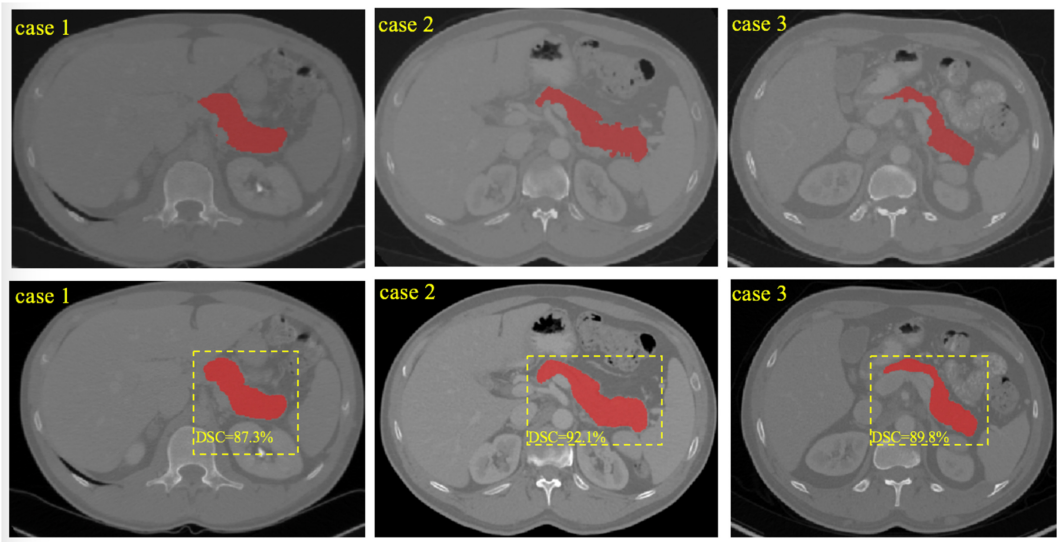

Fig. 6. Pancreas segmentation results for three examples (case 1, case 2, and case 3). The first row shows the ground truth, and the second row shows the final segmentation results.

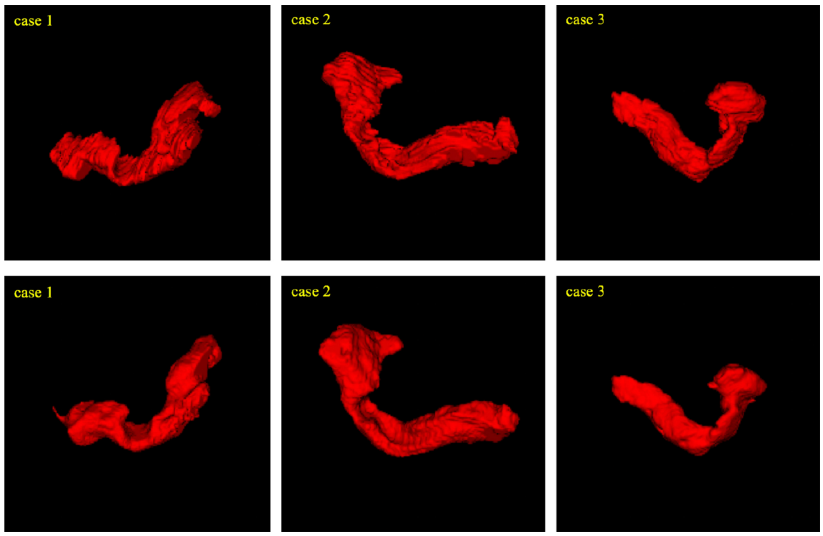

Fig. 7. 3-D visualization of the ground-truth segmentations (first row) and our segmentations (second row) for the results shown in Fig. 6.

state-of-the-art methods for the pancreas segmentation, as briefly introduced below.

1) Roth *et al.* [6] introduced HNNs on the three orthogonal axial, sagittal, and coronal views to do the localization and the segmentation of pancreas.
2) Farag *et al.* [2] adopted a CNN model with drop out to conduct a classification of pancreas and nonpancreas regions based on super-pixel patches.
3) Zhou *et al.* [4] used a fixed-point model to shrink the input region. The 2-D FCN model was employed in two stages for coarse and fine segmentation.
4) Cai *et al.* [9] utilized RNN with Jaccard loss function to conduct the segmentation on 2-D slices of pancreas.
5) Zhu *et al.* [5] expanded the 2-D coarse-fine FCN model of [4] into a 3-D framework for volumetric pancreas segmentation.
6) Yu *et al.* [8] added the recurrent saliency transformation module into their previous models (Zhou *et al.* [4]), which achieved the state-of-the-art performance in terms of DSC.

Table I compares the segmentation performance of our proposed method with six state-of-the-art methods, using mean DSC, Jaccard, precision, and recall (with standard deviation).

The four indices over 82 samples increase from 84.6% to 85.9%, 71.8% to 75.7%, 84.5% to 87.6%, and 82.8% to 85.2%, compared to the state-of-the-art methods. Three examples of the ground-truth segmentations and our segmentations are shown in Figs. 6 and 7. As can be observed from Figs. 6 and 7, the similarity with manual delineations is higher by our proposed method, in spite of diverse shapes and locations of pancreas in CT images. We compared our results with previous methods [2], [4], [6], [8], [9] through t-test

TABLE I
DSC, JACCARD, PRECISION, AND RECALL ARE COMPUTED UNDER 4-FOLD CROSS-VALIDATION. BEST PERFORMING METHODS ARE SHOWN IN **BOLD**. (THE PERFORMANCE IS DESCRIBED BY MEAN ± STD [MIN, MAX])

Method	No.	DSC (%)	Jaccard (%)	Precision (%)	Recall (%)	Protocol
Farag et al.(2017) [2]	80 (CT)	70.7±13.0 [24.4,85.3]	57.9±13.6 [13.9,74.4]	71.6±10.5 [34.8,85.8]	74.4±15.1 [15.0,90.9]	CV-6
Zhou et al.(2017) [4]	82 (CT)	82.4±5.7 [62.4,90.9]	–	–	–	CV-4
Zhu et al. (2017) [5]	82 (CT)	84.6±4.9 [69.6,91.5]	–	–	–	CV-4
Roth et al. (2018) [6]	82 (CT)	81.3±6.3 [50.7, 80.9]	68.9±8.1 –	–	–	CV-4
Yu et al. (2018) [8]	82 (CT)	84.5±5.0 [62.8,91.0]	–	–	–	CV-4
Cai et al. (2018) [9]	82 (CT)	83.3±5.6 [59.0,91.0]	71.8±7.7 [41.8,83.5]	84.5±6.2 [60.7,96.7]	82.8± 8.4 [56.4,94.6]	CV-4
Our proposed method (before post-processing)	82 (CT)	**85.9±5.1** [69.9,92.1]	**75.7±7.6** [54.4,85.4]	**87.6±4.7** [69.8,95.5]	**85.2±8.9** [57.5,95.3]	CV-4

TABLE II
t-VALUE AND p-VALUE FOR OUR METHOD AND THE METHODS LISTED IN TABLE I

Method	Degree of freedom	t-value	p-value
Farag et al. (2017) [2]	79	9.86	$p=4.4\times10^{-18}$ (<0.001)
Zhou et al. (2017) [4]	81	4.14	$p=6.1\times10^{-5}$ (<0.001)
Roth et al. (2018) [6]	81	5.14	$p=9.1\times10^{-7}$ (<0.001)
Yu et al. (2018) [8]	81	1.77	p=0.077 (<0.1)
Cai et al. (2018) [9]	81	3.11	p=0.003 (<0.05)

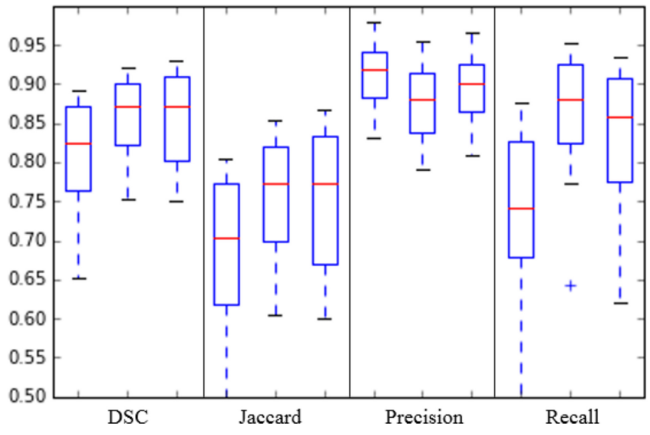

Fig. 8. Changes of values of four evaluation metrics *with respect to* three different patch sizes. For each metric, the first bar, second bar, and last bar correspond to the patch sizes of $16\times32\times32$, $16\times64\times64$, and $16\times128\times128$, respectively. Leave-one-subject-out cross-validation is used for obtaining all these results.

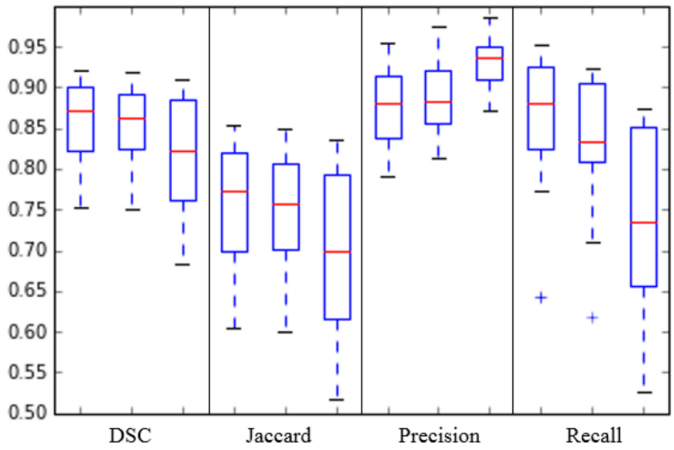

Fig. 9. Changes of values of four evaluation metrics *with respect to* three different step sizes at the testing phase. For each metric, the first bar, second bar, and last bar correspond to the step sizes of $2\times8\times8$, $4\times16\times16$, and $8\times32\times32$, respectively. Leave-one-subject-out cross-validation is used for obtaining all these results.

(Table II). The t-values with 81 degrees of freedom are 9.86, 4.14, 5.14, 1.77, and 3.11, respectively. The corresponding p-values are $p < 0.001$, $p < 0.001$, $p < 0.001$, $p < 0.1$, and $p < 0.05$, respectively. Therefore, our proposed method has statistically significant improvements ($p < 0.001$) compared with other methods [2], [4], [6]. We improve the results obviously ($p < 0.05$) compared with method [9]. However, the improvements do not seem significant compared with the recent state-of-the-art recurrent saliency transformation network (RSTN) method proposed by Yu *et al.* [8]. To further verify the performance of our method, we compared our results with RSTN [8] through a paired-sample t-test. In RSTN, all the intensity values were first saturated into $[-100, 240]$. The FCN model pretrained on PascalVOC is then adapted to conduct the segmentation on Caffe. The coarse-stage segmentation ran 60 000 iterations with the learning rate of 10^{-5}. The saliency transformation module is implemented by two 3×3 convolutional layers. The fine-scaled segmentation model ran 60,000 iterations with the learning rate of 10^{-5} with images cropped from the coarse-scaled segmentation mask. The mean and standard deviation of the differences in DSC in the paired-sample t-test are +2.1% and 5.0%, respectively. The t-value is 2.48 with 81 degrees of freedom ($p < 0.05$). Therefore, our proposed method achieves statistically significant improvements.

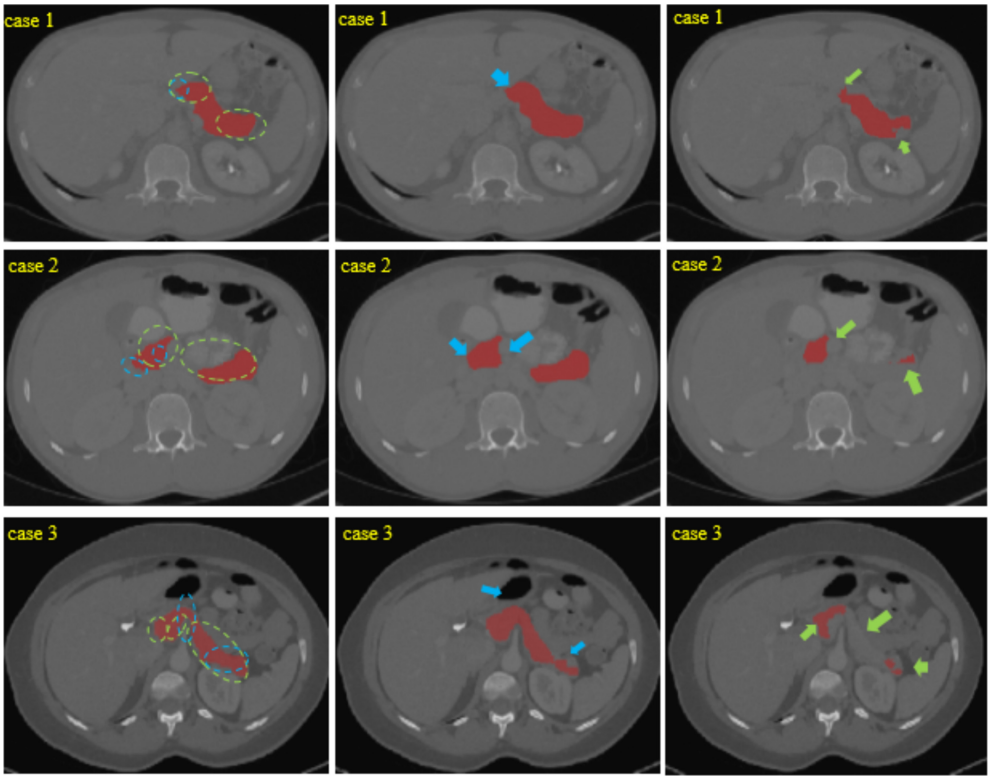

Fig. 10. Comparison of segmentations by the skeleton-guided model and the nonskeleton-guided model, along with the ground truth, on three examples (case 1, case 2, and case 3). In each row, from left to right, the ground-truth and the segmentations by skeleton-guided model and the nonskeleton-guided model are shown. Blue and green arrows indicate different segmentation results (compared to the ground truth) by the skeleton-guided model and the nonorgan-guided model. These disagreement regions are indicated in the ground-truth segmentation images (first column) by blue and green dashed ellipses, respectively.

2) Ablation Study:

a) Evaluation on the impact of input patch size: Since different patch sizes change the receptive field of the network which contributes to different region accuracies, we conduct experiments using three different input patch sizes, that is, 16×32×32, 16×64×64, and 16×128×128, for training the same network architecture shown in Fig. 3. As shown in Fig. 8, our method obtains the best results with the patch size of 16×64×64. Due to the small observation of contexture, the segmentation performance is the lowest with the patch size of 16×32×32. However, we found that the performance did not become obviously better with the patch size of 16×128×128. This is because a smaller number of input samples could be used to train the network, when using large patch size.

b) Evaluation on the impact of step size: In the testing phase, we crop patches with a fixed step size. After these patches are fed into the trained model, all the predicted label patches from the same subject are combined into a single label image by averaging the label values of the overlapping image regions. To find a step size that balances between the segmentation accuracy and computational complexity, we extract patches from CT images with a step size of 2×8×8, 4×16×16, and 8×32×32 for testing. The average time consumed for these three different step sizes are 706.69, 53.62, and 7.89 s, respectively. Small step size increases the workload and costs more time.

The performance results are given in Fig. 9. The step size of 2×8×8 did not obviously perform better than the step size of 4×16×16. Therefore, we select the step size of 4×16×16 for the testing stage.

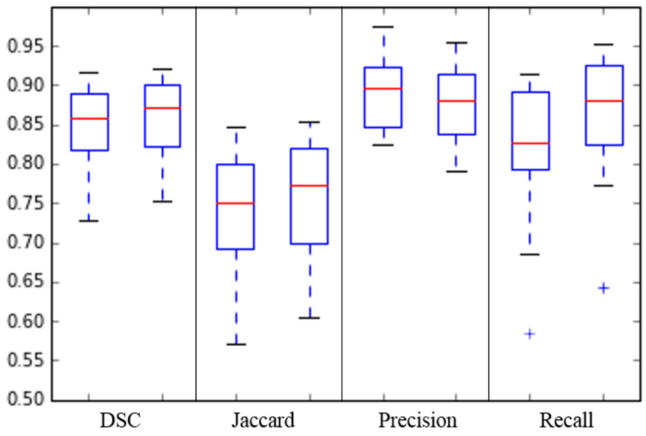

Fig. 11. Comparison between models with and without dense connections. For each metric, the first bar is the result of the model without dense connections, and the second bar is the result of the model with dense connections. Leave-one-subject-out cross validation is used to generate these results.

c) Comparison with nonskeleton-guided model: To evaluate the contribution of skeleton in our network, we compare our proposed network with our downgraded network without using guidance from the pancreas skeleton (i.e., the single task model). The DSC (mean±std [max, min]) value of pancreas after segmentation is 83.0%±5.7% [91.15%, 59.32%], which is significantly lower than our proposed method. The p-value for the single task model and the proposed method is less

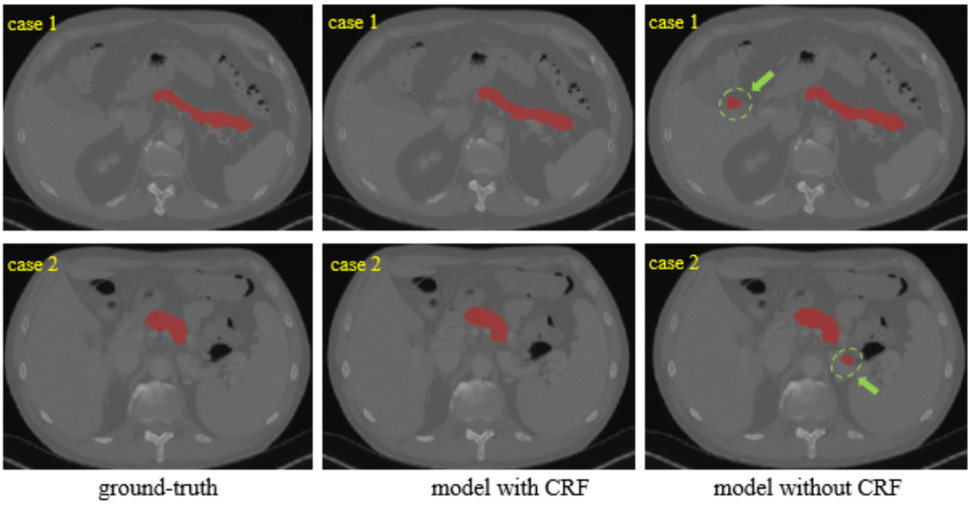

Fig. 12. Performance comparison on segmentation results by the multitask models with and without 3-D fully connected CRF, along with the ground truth, on two examples (cases 1 and 2). In each row, from left to right, the ground-truth segmentation and the segmentations by the multitask models with and without 3-D fully connected CRF are shown. Green arrows and green dashed ellipses indicate the isolated false positive segmentations by the multitask model without 3-D fully connected CRF, which are removed after using 3-D fully connected CRF.

than 0.05. Therefore, our proposed method with the guidance of skeletons improves the segmentation accuracy significantly.

In the conventional U-Net or FCN, the pixel labels equally contribute to the training of the network. However, in this situation, organs like pancreas will not be completely segmented because of the noise in CT images. Therefore, to enhance the discriminative ability of the network for pancreas, we provide the network with an additional guidance to improve the significance of the region where pancreas is located. This methodology teaches the network to focus more on the pancreas area. In the multitask learning strategy, the learning process is guided by the segmentation loss and the regression loss simultaneously. Complementary information from segmentation and regression tasks are better used. The regression task of pancreas skeleton provides a strong reference for the organ shape, which alleviates the under-estimation caused by large shape variation for pancreas in CT images. From the visualization results shown in Fig. 10, we can observe that, for cases 2 and 3, with large variable pancreas shapes, the single task model under-estimated pancreas severely. However, in the proposed model, the extracted skeleton provides rich information about the shape and size of the pancreas. The final results are improved obviously. On the other hand, for case 1 with relatively regular shape, the single task model obtains a similar result as our proposed method, although still worse than our method due to unclear boundaries of pancreas. The results verify the effectiveness of using the skeleton guidance for organs with large variability in shape.

d) Comparison with the multiTask model without dense connections: To investigate the impact of dense connections in the model, we conduct another experiment using our proposed model and the multitask model without dense connections. Fig. 11 shows the results of the four metrics for the two models. As confirmed in Fig. 11, dense connections are useful for training pancreas segmentation models.

e) Effectiveness of feature consistency: To demonstrate the effectiveness of using modules of feature consistency, we also run the model without modules of feature consistency. From Fig. 13, we can see that the results with the use of modules are more accurate.

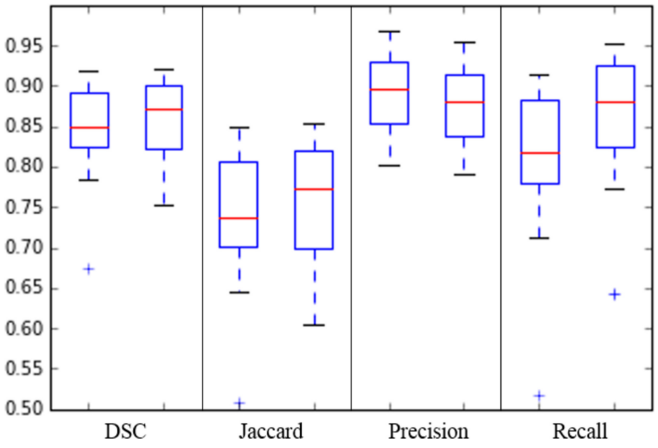

Fig. 13. Comparison between models with and without modules of feature consistency. For each metric, the first bar is the result of the model without modules of feature consistency, and the second bar is the result of the model with modules of feature consistency. Leave-one-subject-out cross validation is used to generate these results.

TABLE III
QUANTITATIVE COMPARISONS OF DICE, JACCARD, PRECISION, AND RECALL FOR PANCREAS SEGMENTATION ON THE CT IMAGES OF 59 PATIENTS. (THE BEST RESULTS ARE INDICATED IN **BOLD**, MEAN ± STD)

Method	DSC (%)	Jaccard (%)	Precision (%)	Recall (%)
RSTN (Yu et al.(2018) [8])	82.2±6.3	71.5±7.9	85.3±5.1	79.1±8.6
Our proposed method	**86.9±4.6**	**77.3±6.8**	**91.0±3.1**	**83.5±6.6**

f) Evaluation on the effectiveness of the 3-D fully connected conditional random field: As described in Section II-C, we utilize 3-D fully connected CRF to refine the segmentation results of our proposed model. The four indices over

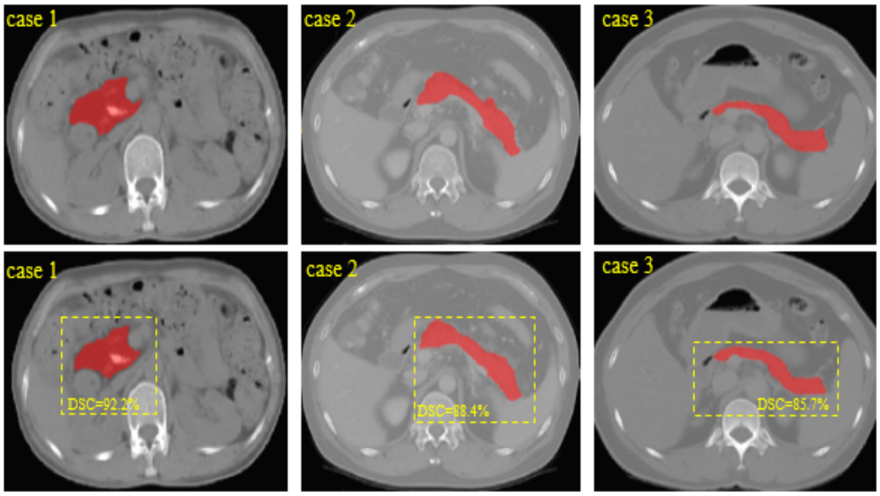

Fig. 14. Pancreas segmentation results on three examples of the FMU dataset (case 1, case 2, and case 3). The first row shows the ground-truth segmentations, and the second row shows our automatic segmentations.

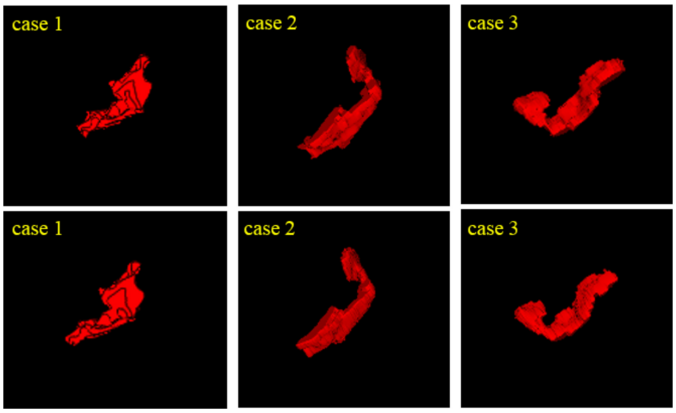

Fig. 15. 3-D visualization of the ground-truth segmentations (first row) and our segmentations (second row), shown in Fig. 14.

82 samples increases from 85.9% to 86.4%, 75.7% to 76.2%, 87.6% to 88.3%, and 85.2% to 85.3%. Two examples of the ground-truth segmentations and our segmentations are shown in Fig. 12. As can be seen, the 3-D fully connected CRF can effectively remove isolated false positive segmentations.

F. Evaluation of Pancreas Segmentation on the FMU Dataset

We trained the networks for the FMU dataset with the same training settings as the networks for the NIH dataset. Three examples of the ground-truth segmentations and our segmentations are shown in Figs. 14 and 15. As can be observed, the segmentations generated by our proposed method are highly consistent with the ground-truth segmentations, in spite of the shape and size variation of pancreas in CT images. To further verify the effectiveness of our proposed 3-D multitask FCN, the RSTN is also used for segmenting the pancreas on the FMU dataset. We fine-tuned the RSTN for the NIH dataset to make the model converge on the FMU dataset. As can be seen in Table III, our proposed method can achieve better segmentation precision than the RSTN method, indicating potential feasibility in real clinical applications.

IV. Conclusion

In this article, we have proposed a cascaded multitask 3-D FCN to address the challenging pancreas segmentation problem in abdominal CT images. Since pancreas is relatively small in shape and may appear in different locations of the abdomen, we first fast locate the pancreas in the raw CT image using a standard FCN architecture. Then, we apply a second stage of the cascade to only the localized region for final fine-grained segmentation in a multitask scheme. More importantly, we have proposed a skeleton-guided network to grasp the organ's morphological information, which is shown critical for accurate segmentation, especially in the challenging cases. The experimental results on the two challenging pancreas datasets, that is, the NIH dataset and the FMU dataset, indicate that our proposed method is more accurate than the state-of-the-art methods, and is also more robust across two different datasets. Finally, the ablation study also demonstrates that our proposed module does contribute to the performance gain.

References

[1] (2013). *The Cancer Imaging Archive (TCIA) Public Access*. [Online]. Available: https://wiki.cancerimagingarchive.net/display/Public/Pancreas-CT

[2] A. Farag, L. Lu, H. R. Roth, J. Liu, E. Turkbey, and R. M. Summers, "A bottom-up approach for pancreas segmentation using cascaded superpixels and (deep) image patch labeling," *IEEE Trans. Image Process.*, vol. 26, no. 1, pp. 386–399, Jan. 2017.

[3] J. K. Bjerregaard, M. B. Mortensen, H. A. Jensen, M. Nielsen, and P. Pfeiffer, "Prognostic factors for survival and resection in patients with initial nonresectable locally advanced pancreatic cancer treated with chemoradiotherapy," *Int. J. Rad. Oncol. Biol. Phys.*, vol. 83, no. 3, pp. 909–915, Jul. 2012.

[4] Y. Zhou, L. Xie, W. Shen, Y. Wang, E. K. Fishman, and A. L. Yuille, "A fixed-point model for pancreas segmentation in abdominal CT scans," in *Proc. Med. Image Comput. Comput. Assist. Intervent.*, vol. 10433, 2017, pp. 693–701.

[5] Z. Zhu *et al.*, "A 3D coarse-to-fine framework for automatic pancreas segmentation," *arXiv preprint:1712.00201*, vol. 2, pp. 1–10, 2017.

[6] H. R. Roth *et al.*, "Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation," *Med. Image Anal.*, vol. 45, pp. 94–107, Apr. 2018.

[7] H. Takizawa, T. Suzuki, H. Kudo, and T. Okada, "Interactive segmentation of pancreases in abdominal computed tomography images and its evaluation based on segmentation accuracy and interaction costs," *Biomed. Res. Int.*, vol. 2017, Aug. 2017, Art. no. 5094592.

[8] Q. Yu, L. Xie, Y. Wang, Y. Zhou, E. K. Fishman, and A. L. Yuille, "Recurrent saliency transformation network: Incorporating multi-stage visual cues for small organ segmentation," in *Proc. CVPR*, 2018, pp. 8280–8289. [Online]. Available: https://github.com/198808xc/OrganSeg
[9] J. Cai, L. Lu, Y. Xie, F. Xing, and L. Yang, "Improving deep pancreas segmentation in CT and MRI images via recurrent neural contextual learning and direct loss function," *arXiv preprint:1707.04912*, pp. 1–8, 2017.
[10] K. Dmitriev, I. Gutenko, S. Nadeem, and A. E. Kaufman, "Pancreas and cyst segmentation," in *Proc. Med. Imag. Image Process.*, vol. 9784, 2016, Art. no. 97842C.
[11] Z. Xue, D. Shen, and C. Davatzikos, "CLASSIC: Consistent longitudinal alignment and segmentation for serial image computing," *NeuroImage*, vol. 30, no. 2, pp. 388–399, 2006.
[12] Y. Guo, Y. Gao, and D. Shen, "Deformable MR prostate segmentation via deep feature learning and sparse patch matching," *IEEE Trans. Med. Imag.*, vol. 35, no. 4, pp. 1077–1089, Apr. 2016.
[13] Y. Zhan and D. Shen, "Automated segmentation of 3D U.S. prostate images using statistical texture-based matching method," in *Proc. MICCAI*, 2003, pp. 688–696.
[14] Q. Feng, M. Foskey, W. Chen, and D. Shen, "Segmenting CT prostate images using population and patient-specific statistics for radiotherapy," *Med. Phys.*, vol. 37, no. 8, pp. 4121–4132, 2010.
[15] S. K. Zhou, H. Greenspan, and D. Shen, *Deep Learning for Medical Image Analysis*. London, U.K.: Academic, 2017.
[16] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annu. Rev. Biomed. Eng.*, vol. 19, pp. 221–248, Jun. 2017.
[17] D. Nie, L. Wang, E. Adeli, C. Lao, W. Lin, and D. Shen, "3-D fully convolutional networks for multimodal isointense infant brain image segmentation," *IEEE Trans. Cybern.*, vol. 49, no. 3, pp. 1123–1136, Mar. 2019.
[18] K. Men, J. R. Dai, and Y. X. Li, "Automatic segmentation of the clinical target volume and organs at risk in the planning CT for rectal cancer using deep dilated convolutional neural networks," *Med. Phys.*, vol. 44, no. 12, pp. 6377–6389, 2017.
[19] M. Liu, J. Zhang, C. Lian, and D. Shen, "Weakly-supervised deep learning for brain disease prognosis using MRI and incomplete clinical scores," *IEEE Trans. Cybern.*, to be published.
[20] H. Fu *et al.*, "Angle-closure detection in anterior segment OCT based on multilevel deep network," *IEEE Trans. Cybern.*, to be published.
[21] Y. Lecun *et al.*, "Handwritten digit recognition with a back-propagation network," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 1990, pp. 396–404.
[22] H. Roth *et al.*, "Towards dense volumetric pancreas segmentation in CT using 3D fully convolutional networks," in *Proc. Med. Imag. Image Process.*, vol. 10574, 2018, Art. no. 105740B.
[23] H. Roth *et al.*, "An application of cascaded 3D fully convolutional networks for medical image segmentation," *Comput. Med. Imag. Graph.*, vol. 66, pp. 90–99, Jun. 2018.
[24] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
[25] A. Farag, L. Lu, E. Turkbey, J. Liu, and R. M. Summers, "A bottom-up approach for automatic pancreas segmentation in abdominal CT scans," in *Proc. MICCAI*, vol. 8676, 2014, pp. 103–113.
[26] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[27] S. Xie and Z. Tu, "Holistically-nested edge detection," in *Proc. ICCV*, 2015, pp. 1395–1403.
[28] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in *Proc. MICCAI*, 2016, pp. 424–432.
[29] X. Bai and L. J. Latecki, "Path similarity skeleton graph matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 7, pp. 1282–1292, Jul. 2008.
[30] N. H. Trinh and B. B. Kimia, "Category-specific object recognition and segmentation using a skeletal shape model," in *Proc. BMVC*, vol. 9, 2009, pp. 7–10.
[31] N. H. Trinh and B. B. Kimia, "*Skeletonsearch*: Category-specific object recognition and segmentation using a skeletal shape model," *Int. J. Comput. Vis.*, vol. 94, no. 2, pp. 215–240, 2011.
[32] W. Shen, K. Zhao, Y. Jiang, Y. Wang, X. Bai, and A. Yuille, "DeepSkeleton: Learning multi-task scale-associated deep side outputs for object skeleton extraction in natural images," *IEEE Trans. Image Process.*, vol. 26, no. 11, pp. 5298–5311, Nov. 2017.
[33] P. K. Saha, G. Borgefors, and G. S. di Baja, "A survey on skeletonization algorithms and their applications," *Pattern Recognit. Lett.*, vol. 76, pp. 3–12, Jun. 2016.
[34] W. Shen, K. Zhao, Y. Jiang, Y. Wang, Z. Zhang, and X. Bai, "Object skeleton extraction in natural images by fusing scale-associated deep side outputs," in *Proc. CVPR*, 2016, pp. 222–230.
[35] R. Ranjan, V. M. Patel, and R. Chellappa, "HyperFace: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 1, pp. 121–135, Jan. 2019.
[36] X. Li *et al.*, "DeepSaliency: Multi-task deep neural network model for salient object detection," *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3919–3930, Aug. 2016.
[37] S. Li, Z.-Q. Liu, and A. B. Chan, "Heterogeneous multi-task learning for human pose estimation with deep convolutional neural network," in *Proc. CVPR Workshops*, 2014, pp. 482–489.
[38] Z. Cai, Q. Fan, R. S. Feris, and N. Vasconcelos, "A unified multi-scale deep convolutional neural network for fast object detection," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 354–370.
[39] S. Pan, J. Wu, X. Zhu, G. Long, and C. Zhang, "Task sensitive feature exploration and learning for multitask graph classification," *IEEE Trans. Cybern.*, vol. 47, no. 3, pp. 744–758, Mar. 2017.
[40] Y. Yang, C. Deng, D. Tao, S. Zhang, W. Liu, and X. Gao, "Latent max-margin multitask learning with skelets for 3-D action recognition," *IEEE Trans. Cybern.*, vol. 47, no. 2, pp. 439–448, Feb. 2017.
[41] S. Yang, Q. Liu, and J. Wang, "A multi-agent system with a proportional-integral protocol for distributed constrained optimization," *IEEE Trans. Autom. Control*, vol. 62, no. 7, pp. 3461–3467, Jul. 2017.
[42] P. Krähenbühl and V. Koltun, "Efficient inference in fully connected CRFs with Gaussian edge potentials," in *Proc. Adv. Neural Inf. Process. Syst. 24 (NIPS)*, 2011, pp. 109–117.
[43] K. Kamnitsas *et al.*, "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Med. Image Anal.*, vol. 36, pp. 61–78, Feb. 2017.
[44] J. I. Orlando, E. Prokofyeva, and M. B. Blaschko, "A discriminatively trained fully connected conditional random field model for blood vessel segmentation in fundus images," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 1, pp. 16–27, Jan. 2017.
[45] H. Fu, Y. Xu, D. W. K. Wong, and J. Liu, "Retinal vessel segmentation via deep learning network and fully-connected conditional random fields," in *Proc. IEEE 13th Int. Symp. Biomed. Imag. (ISBI)*, 2016, pp. 698–701.
[46] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 818–833.
[47] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. ICML*, 2010, pp. 807–814.
[48] M. Arsalan, R. A. Naqvi, D. S. Kim, P. H. Nguyen, M. Owais, and K. R. Park, "IrisDenseNet: Robust Iris segmentation using densely connected fully convolutional networks in the images by visible light and near-infrared light camera sensors," *Sensors*, vol. 18, no. 5, pp. 1–30, 2018.
[49] 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," *arXiv preprint:1709.07330*, pp. 1–13, 2017.
[50] A. A. Lasboo, P. Rezai, and V. Yaghmai, "Morphological analysis of pancreatic cystic masses," *Acad. Radiol.*, vol. 17, no. 3, pp. 348–351, 2010.
[51] Y. Jia *et al.*, "Caffe: Convolutional architecture for fast feature embedding," in *Proc. 22nd ACM Int. Conf. Multimedia*, 2014, pp. 675–678.
[52] B. Y. G. Xavier, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. Artif. Intell. Stat.*, 2010, pp. 249–256.

Jie Xue received the B.S. and Ph.D. degrees in management science and engineering from Shandong Normal University, Jinan, China, in 2010 and 2015, respectively.

She is currently an Associate Professor with Business School, Shandong Normal University. Her current research interests include image processing, medical image analysis, and membrane computing.

Dr. Xue won a National Visiting Scholar Program with the University of North Carolina from 2017 to 2018.

Kelei He received the Ph.D. degree in computer science and technology from Nanjing University, Nanjing, China.

He is currently an Assistant Professor with the Medical School of Nanjing University. His research interests includes medical image analysis, computer vision, and deep learning.

Dong Nie received the B.Eng. degree in computer science from Northeastern University, Shenyang, China, the M.Sc. degree in computer science from the University of Chinese Academy of Sciences, Beijing, China, and the Ph.D. degree in computer science from the University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.

His current research interests include image processing, medical image analysis, and natural language processing.

Ehsan Adeli received the Ph.D. degree from the Iran University of Science and Technology, Tehran, Iran.

He is a Postdoctoral Research Fellow with Stanford University, Stanford, CA, USA. He was a Postdoctoral Researcher with the University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. He was a Visiting Research Scholar with the Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, USA. His current research interests include machine learning, computer vision, medical image analysis, and computational neuroscience.

Zhenshan Shi received the Ph.D. degree from Fujian Medical University, Fujian, China.

He was a Visiting Scholar with the University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. His current research interests include medical image analysis and machine learning.

Seong-Whan Lee (F'10) received the B.S. degree in computer science and statistics from Seoul National University, Seoul, South Korea, in 1984, and the M.S. and Ph.D. degrees in computer science from the Korea Advanced Institute of Science and Technology, Seoul, in 1986 and 1989, respectively.

He is currently the Head of the Department of Artificial Intelligence, Korea University, Seoul. His current research interests include pattern recognition, machine learning, and brain engineering.

Dr. Lee is a fellow of IAPR and the Korea Academy of Science and Technology.

Yuanjie Zheng received the Ph.D. degree from Shanghai Jiaotong University, Shanghai, China.

He is currently a Professor with the School of Information Science and Engineering, Shandong Normal University, Jinan, China, and a Taishan Scholar of People's Government of Shandong Province of China. He used to be a Senior Research Investigator with the Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. His research is in the fields of medical image analysis, translational medicine, computer vision, and computational photography. His ultimate research goal is to enhance patient care by creating algorithms for automatically quantifying and generalizing the information latent in various medical images for tasks, such as disease analysis and surgical planning, through the applications of computer vision and machine learning approaches to medical image analysis tasks and development of strategies for image-guided intervention/surgery.

Xiyu Liu received the Ph.D. degree in mathematical sciences from Shandong University, Jinan, China, in 1990.

He is currently working as the Dean of the Academy of Management Science and Engineering, Shandong Normal University, Jinan, China. He has authored two books and more than 140 articles. His research interests include membrane computing, data mining, computational intelligence, and nonlinear analysis.

Dr. Liu was awarded the "Taishan Scholar" of Management Science and Engineering, and the Vice President of the Computer Education Research Association of China Higher Normal Universities, and the Shandong Computer Society.

Dengwang Li received the B.S. and Ph.D. degrees in electronic engineering from Shandong University, Jinan, China, in 2006 and 2011, respectively.

He is currently a Professor with the Shandong Key Laboratory of Medical Physics and Image Processing, School of Physics and Electronics, Shandong Normal University, Jinan. His research focuses on signal processing and biomedical engineering.

Prof. Li won a joint Ph.D. program with the University of Sydney from 2009 to 2010.

Dinggang Shen (F'18) received the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China.

He is a Jeffrey Houpt Distinguished Investigator, and a Professor of Radiology, Biomedical Research Imaging Center (BRIC), Computer Science, and Biomedical Engineering with the University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. He is currently directing the Center for Image Analysis and Informatics, the Image Display, Enhancement, and Analysis Lab with the Department of Radiology, and also the medical image analysis core with the BRIC. He was a Tenure-Track Assistant Professor with the University of Pennsylvania, Philadelphia, PA, USA, and a Faculty Member with the Johns Hopkins University, Baltimore, MD, USA. He has published more than 1000 papers in the international journals and conference proceedings, with H-index 93. His research interests include medical image analysis, computer vision, and pattern recognition.

Mr. Shen serves as an Editorial Board Member for eight international journals. He has also served on the Board of Directors, the Medical Image Computing and Computer Assisted Intervention (MICCAI) Society, from 2012 to 2015, and the General Chair for MICCAI 2019. He is a fellow of the American Institute for Medical and Biological Engineering and the International Association for Pattern Recognition.