

# Pancreas segmentation with probabilistic map guided bi-directional recurrent UNet

Jun Li, Xiaozhu Lin, Hui Che, Hao Li, and Xiaohua Qian

**Abstract**—Pancreas segmentation in medical imaging data is of great significance for clinical pancreas diagnostics and treatment. However, the large population variations in the pancreas shape and volume cause enormous segmentation difficulties, even for state-of-the-art algorithms utilizing fully convolutional neural networks (FCNs). Specifically, pancreas segmentation suffers from the loss of spatial information in 2D methods, and the high computational cost of 3D methods. To alleviate these problems, we propose a probabilistic-map-guided bi-directional recurrent U-Net (PBR-UNet) architecture, which fuses intra-slice information and inter-slice probabilistic maps into a local 3D hybrid regularization scheme, which is followed by bi-directional recurrent network optimization. The PBR-UNet method consists of an initial estimation module for efficiently extracting pixel-level probabilistic maps and a primary segmentation module for propagating hybrid information through a 2.5D U-Net architecture. Specifically, local 3D information is inferred by combining an input image with the probabilistic maps of the adjacent slices into multi-channel hybrid data, and then hierarchically aggregating the hybrid information of the entire segmentation network. Besides, a bi-directional recurrent optimization mechanism is developed to update the hybrid information in both the forward and the backward directions. This allows the proposed network to make full and optimal use of the local context information. Quantitative and qualitative evaluation was performed on the NIH Pancreas-CT dataset, and our proposed PBR-UNet method achieved better segmentation results with less computational cost compared to other state-of-the-art methods.

**Index Terms**—Pancreas segmentation, Deep learning, Medical image segmentation.

## I. INTRODUCTION

ACCURATE segmentation of the human pancreas in medical imaging data is an essential prerequisite for relevant medical image analysis and surgical navigation systems. However, pancreas segmentation is quite challenging due to the considerable variations in the shape, and the pancreas vulnerability to elastic deformations resulting from breathing and heartbeat. Therefore, developing a robust and accurate pancreas segmentation model is of profound significance for performance improvement in computer-assist surgery techniques.

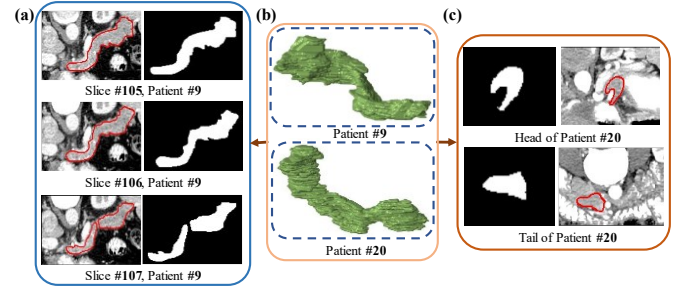


Fig. 1. Examples of pancreas scans: (a) Three adjacent slices with high correlation; (b) 3D pancreas data from two subjects showing different spatial shapes; (c) A small correlation exists between the shapes of the pancreas head and tail.

Nowadays, developing satisfactory methods for pancreas segmentation is still challenging. Compared with some other human organs such as the heart and liver, the pancreas exhibits a higher anatomical variability [1], as shown in Fig. 1. Variations in the background tissues and dramatic volume changes can undermine the performance of any start-of-the-art method for pancreas segmentation [2]. Consequently, the pancreas has been typically considered among the most complex organs for segmentation [3].

Pancreas segmentation methods can be roughly divided into two categories, namely methods based on top-down multi-atlas registration and label fusion (MALF) [4]–[7], and methods based on deep learning [1]–[3]. For a MALF method, volumetric multiple-atlas registration is combined with a robust label-fusion scheme to optimize the per-pixel pancreas segmentation [8]. Due to the high shape variability and blurred boundaries of the pancreas, the accuracies of MALF-based methods on benchmark datasets range merely from 69.6% to 78.5% [4]–[7]. Nevertheless, this performance can be considerably improved by deep learning methods. State-of-the-art deep-learning segmentation models have been proposed, such as FCN [9], U-Net [10], and DeepLab [11]. The encoder-decoder and skip-connection techniques are widely used in these models to increase the final output resolution, and accurately locate and distinguish the pancreas from surrounding tissues [12], [13].

Many existing deep-learning segmentation methods are based on 2D data. For example, 2D segmentation methods process each slice of a volume as a separate input. Then, the segmentation results of all slices are combined to construct a 3D pancreas object [1], [8], [14], [15]. Because three non-identical views give quite different images of the pancreas, Zhou *et al.*

J. Li, H. Li and X. Qian is with School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai 200030, China.

X. Qian is the corresponding author. (e-mail: xiaohua.qian@sjtu.edu.cn).

X. Lin is with Department of Radiology, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China.

H. Che is with the Biomedical Engineering Department, Rutgers University, New Jersey 08901, USA.

[1] proposed training three 2D models for segmentation in each of the coronal, sagittal, and axial views, and then merged these three results via majority voting to produce coarse 3D segmentation. While this method introduces 3D features, it does not make full use of the 3D information. Fu *et al.* [8] proposed a richer convolutional feature network, which performs pancreas segmentation by extracting multiscale information via a multi-layer upsampling structure. Although this method enhances the extraction of intra-slice information, it completely ignores the 3D information. Thus, 2D networks neglect the relationship between adjacent slices, and the output of such networks cannot be interpreted in a 3D context [16]. This impedes the extraction of high-level features and restricts the pancreas segmentation performance.

Since 2D networks cannot capture 3D volume information, 3D networks were proposed by using the CT volume as the network input [17]–[23]. However, these 3D networks require significant computational and memory resources and hence rely excessively on high-performance servers [16], [24]. Oktay *et al.* [17] proposed an attention gate model, based on a 3D U-Net architecture. While this model can suppress irrelevant regions in the network input and highlight useful salient features, the model suffers from irreversible positioning errors and large computational costs. To reduce GPU memory requirements and acquire 3D information, Roth *et al.* [25] proposed using a two-stage segmentation scheme, in which the second 3D FCN has reduced computations and is focused on the segmentation of the target organ. However, the sub-volume approach they adopted may cause segmentation discontinuities or inconsistencies at overlapping window boundaries [26]. The performance and costs of a complete 3D convolutional architecture was further investigated. It was found that such 3D architecture provided slightly better performance in comparison to 2D methods, but caused a significant and disproportionate increase in computing costs [27]. The requirement of a high-memory footprint limits two performance improvement factors, namely the network depth and the filter field of view [28]. So, the basic 3D networks can hardly achieve satisfactory performance in pancreas segmentation.

Two-stage learning frameworks for coarse-to-fine pancreas segmentation were also proposed to overcome the susceptibility of one-pass learning strategies to background interference [1], [20], [24]–[27]. In such frameworks, the region of interest (ROI) is roughly localized by the initial segmentation, and then a finer segmentation is carried out by focusing on the localized ROI. However, a part of the pancreas could be irreversibly discarded at the initial segmentation. This results in unreliable and unstable pancreas localization, due to the lack of an effective error correction mechanism. Yu *et al.* [29] presented two failure cases using the coarse-to-fine scheme in [1] and indicated that the fine-scale segmentation deteriorated in some cases due to the lack of contextual information. Overall, the performance of a two-stage learning framework for pancreas segmentation is limited by overlooking the contextual information and possibly missing the target area.

The success of the long short-term memory (LSTM) networks is essentially due to their effective consideration of long-span dependencies [30], and contextual information. In [14], [31], [32], LSTM networks were applied to directly embed

contextual information as time series into models. But in the LSTM-based segmentation framework, the LSTM module usually serves as a single refinement module following the main segmentation networks [31]. Indeed, the LSTM makes limited segmentation improvement, enlarges the entire network, and requires more computing resources.

The pancreas globally shows a significant anatomical variability, while it locally exhibits strong morphology and pattern correlation among adjacent CT slices (Fig. 1(a)). These global and local pancreas characteristics demonstrate that the local 3D information or inter-slice information is critical for developing pancreas segmentation models of high precision and robustness. However, while earlier segmentation models exploited 2D features or global 3D information, few models accounted for the local 3D context without incurring high computational costs. In general, multi-channel networks have achieved better results than single-channel networks without a significant increase in the computational burden [27].

Motivated by the above observations, we developed a novel pancreas segmentation model based on local 3D hybrid information and a bi-directional recurrent 2.5D U-Net architecture, namely the probabilistic-map-guided bi-directional recurrent U-Net (PBR-UNet). In this model, the original map of a CT slice is combined with probabilistic maps of the adjacent slices to infer local 3D hybrid information that can be used for guiding the segmentation of the center slice. This information is propagated in the 2.5D U-Net and then optimized through a bi-directional recurrent structure in order to improve and refine the segmentation results. Specifically, we firstly apply an initial estimation model to extract, for each slice, a pixel-level probabilistic map, which represents the per-pixel probability of belonging to the pancreas. Then, the initial probabilistic maps of the adjacent slices are combined with the map of the center slice into multi-channel hybrid data, which contains local 3D hybrid information of the center slice. Under the constraints of the local 3D context, the segmentation of the center slice could be constrained, resulting in stable results. Finally, a bi-directional recurrent structure is applied to the primary segmentation to optimize the local 3D hybrid information. We use each primary segmentation output to update the probabilistic maps in the multi-channel data, make the local 3D hybrid information more precise, and boost the final segmentation performance.

In summary, our PBR-UNet framework has the following two technical contributions:

- **Introducing local 3D hybrid information.** A probabilistic-map-guided segmentation model is developed to combine intra-slice information and probabilistic maps of adjacent slices to form the local 3D hybrid information. The proposed model balances the requirements for high efficiency in spatial information utilization and low computational costs, and thus avoids the problems of lack of context in 2D models, and the impact of anatomical variability or computational costs on the 3D ones.
- **Constructing a bi-directional recurrent U-Net.** A bi-directional recurrent update scheme was proposed to optimize local 3D hybrid information in 2.5D U-Net. This information is propagated and updated in both the forward and backward directions to make full use of the local

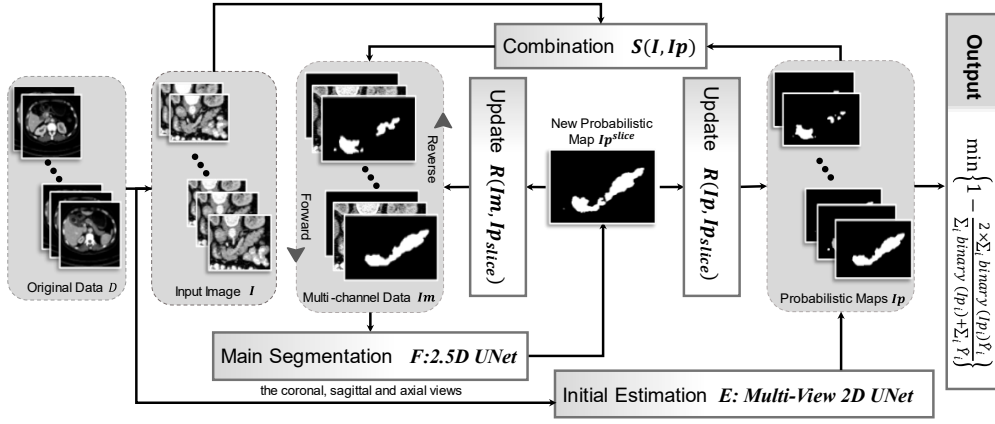


Fig. 2. Illustration of the PBR-UNet pipeline for pancreas segmentation. The pipeline includes modules for the initial estimation ( $E$ ), the primary segmentation ( $F$ ) using the local 3D hybrid information from the combination ( $S$ ) of the original map and probabilistic maps, followed by the bi-directional recurrent update ( $R$ ) scheme.

context. Under the guidance of the optimized local contextual information, the burden of searching for the optimal pancreas segmentation is well-relieved, and high-precision results can be achieved. In addition, this bi-directional recurrent update scheme can be embedded in most segmentation models.

The remainder of this paper is organized as follows. Section II gives the details of the methods used in our proposed model. Section III presents the experimental results, Section IV discusses our findings, and Section VI highlights key conclusions.

## II. METHODOLOGY

Figure 2 shows the flowchart of the PBR-UNet framework for pancreas segmentation. We begin by a problem definition (Sec. A), followed by a detailed description of the initial estimation (Sec. B) and primary segmentation (Sec. C) stages. Finally, we summarize our inference schemes (Sec. D).

### A. Problem Definition

In this section, we formulate the problem of pancreas segmentation from 3D CT scans in terms of basic mathematical notations. Let  $I \in \mathbb{R}^{m \times h \times w \times c}$  be the 3D scan data of a patient, where  $m$  denotes the total number of slices,  $h$  and  $w$  refer to the slice height and width, respectively, and  $c$  denotes the number of channels. The annotation of  $I$  is a binary segmentation mask  $\hat{Y}_{m,h,w,c} \in \mathbb{R}^{m \times h \times w \times c}$ , and it is defined as

$$\hat{Y}_{m,h,w,c} = \begin{cases} 0, & I(m,h,w,c) \notin \text{pancreas} \\ 1, & I(m,h,w,c) \in \text{pancreas} \end{cases} \quad (1)$$

where a value of 1 means that the voxel belongs to the pancreas, 0 means that the voxel belongs to the background. The mapping function  $\Omega$  is constructed based on the ground truth  $\hat{Y}$ , and this function outputs the pixel-wise segmentation maps  $\Omega(I)$  for given 3D scan  $I$ . The mapping function should be constructed such that the similarity between  $\Omega(I)$  and the ground truth map  $\hat{Y}$  is as high as possible.  $\Omega$  can be written as the composition of two functions:  $E$  and  $F$ . The function  $E$  returns the initial segmentation estimate, i.e., the pixel-wise probabilistic map of each slice in the 3D scan  $I$ . The function  $F$  returns the primary segmentation resulting from bi-directionally propagating and optimizing local 3D hybrid information. To make full use of the

local context information and spread the local 3D hybrid information, we design the primary segmentation function  $F$  in both forward and backward directions. This design can be expressed as  $\text{backward}(\text{forward}(I))$ , and the design details will be given in the next section. Thus, the pancreas segmentation problem can be formulated as the problem of minimizing the loss function

$$\mathcal{L}(\Omega, I, \hat{Y}) = \min \left( 1/m \sum_{i=1}^m (F(E(I_i)) - \hat{Y}_i)^2 \right) \quad (2)$$

### B. Initial Estimation of Probabilistic Maps

In the first stage of the proposed PBR-UNet framework, intra-slice features are extracted, and a probabilistic map is obtained for each slice using the initial estimation module, which is represented by the function  $E$  (Fig. 2). The probabilistic maps of a CT volume  $I$  can be denoted mathematically as  $Ip_\theta = E(I)$ ,  $Ip_\theta \& I \in \mathbb{R}^{m \times h \times w \times c}$ ,  $\|Ip_\theta\| \in [0, 1]$ . Specifically, to utilize the volume information, we use a multi-view 2D U-Net (Multi-UNet) architecture [11] to quickly obtain pixel-wise probabilistic maps. The Multi-UNet model contains three 2D U-Net modules for conducting segmentation along the coronal, sagittal, and axial views. These modules have the same structure and trained separately with data associated with their respective views. Then, arithmetic averaging is applied to merge these three results.

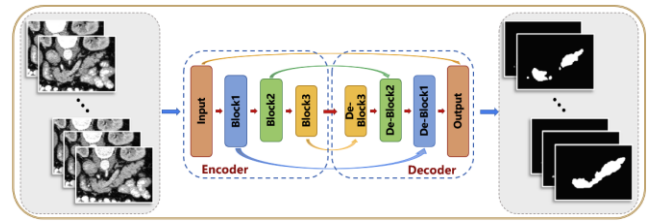


Fig. 3. The 2D U-Net architecture for extracting probabilistic maps from the axial view. The outputs are the pixel-wise probabilistic maps, which indicates the likelihood that the pixel belongs to the pancreas.

Fig. 3 shows the overall 2D segmentation structure, which is composed of a pair of encoder and decoder modules, where each module consists of 4 blocks. Each block contains two convolutional layers and two rectified linear unit activation functions. A pooling layer is added at the end of each block in the encoder module, while a deconvolutional layer is added at the

start of each block in the decoder module. Moreover, skip connections are used to restore the full spatial resolution of the network output [9]. With these skip connections, deep semantic information obtained by the encoder module is transferred directly to the decoder module. Thus, the output probabilistic maps will contain more discriminative features, including deep and shallow semantic information, which represents useful contextual information for the next segmentation stage.

### C. Bi-directional Recurrent Segmentation

In this section, we present details of combining multi-channel data and primary segmentation with the bi-directional recurrent structure. Due to the volumetric continuity of the pancreas, experienced radiologists typically localize the pancreas in a CT slice according to the adjacent slices along the Z-axis. However, the 2D U-Net architecture can only capture intra-slice features but not the context information. Some 3D segmentation networks [17][19]–[22], [33] were proposed for extracting context information. However, these networks suffer from dramatic anatomical variability. As shown in Fig. 1, the pancreas context information is not generally relevant but there is a greater similarity between adjacent slices. Thus, we propose to leverage the probabilistic maps of adjacent slices to guide the segmentation process with local 3D hybrid information. This approach can improve the segmentation performance without the need for high computational resources.

To fuse the intra-slice and context information into local 3D hybrid information, each center slice is combined with the probabilistic maps of its adjacent slices into multi-channel hybrid data  $Im_\theta$ , which can guide the bi-directional recurrent segmentation process without introducing irrelevant information. Specifically,  $Im_\theta$  is formed by combining the estimated probabilistic map  $Ip_\theta$  and the original data  $I$  using the transformation  $\mathcal{S}$ , i.e., the 3-channel hybrid data can be written as  $\mathcal{S}(I, Ip_\theta) = [(Ip_{\theta,1}, I_1, Ip_{\theta,2}), \dots, (Ip_{\theta,n-1}, I_n, Ip_{\theta,n})]$ .

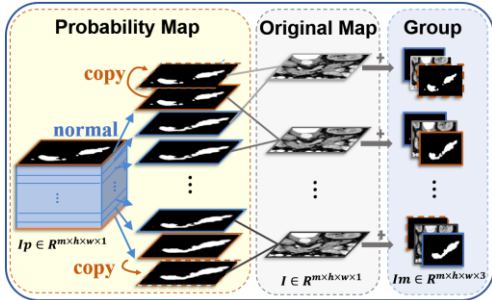


Fig. 4. Construction of the local 3D hybrid information by combining the original map and the probabilistic map to form multi-channel hybrid data (in this case, 3 channels are used). To maintain the data integrity, the first and last slices are duplicated.

Fig. 4 describes the transformation process for 3-channel data. Since the first slice does not have a preceding slice and the last slice does not have a succeeding slice, the first and the last slices are duplicated. After fusion, the local 3D hybrid information of the multi-channel data can be propagated through the 2.5D U-Net to guide the primary segmentation and update the multi-channel data through the bi-directional recurrent update scheme. The 2.5D U-Net has the same architecture as that of the initial estimation module, but  $c$  input and output channels are assumed.

The output of the 2.5D U-Net is a probabilistic map, which is used to update the corresponding part of the multi-channel hybrid data. The final output is obtained after the bi-directional recurrent update and binarization steps.

Fig. 5 illustrates the bi-directional recurrent update scheme with multi-channel data, where three channels are used for illustration. Let  $I_t$  be the  $t$ -th slice of scan  $I$ . As described above,  $I_t$ ,  $Ip_{t-1}$  and  $Ip_{t+1}$  form a three-channel hybrid data sample  $Im_t$  (shown as a blue solid circle in Fig. 5), which is fed into the 2.5D U-Net (indicated by  $F$  in Fig. 5) to output a new probabilistic map  $Ip_t$  of the  $t$ -th slice (shown as an orange solid circle in Fig. 5). Moreover,  $Ip_t$  is contained in the data samples  $Im_{t-1}$  and  $Im_{t+1}$  (shown as blue solid circles in Fig. 5). So, the new probabilistic map  $Ip_t$  updates the corresponding part in  $Im_{t-1}$  and  $Im_{t+1}$  based on the function  $R$ . The probabilistic map is updated by averaging to be  $Ip_{new} = (Ip_{segment} + Ip_\theta)/2$ . The segmentation output is binarized to get the final result  $Z$  (indicated by  $\phi$  in Fig. 5),

$$Z = \varphi(Ip_o) = \begin{cases} 1, & \text{probability value} > 0.5 \\ 0, & \text{probability value} \leq 0.5 \end{cases} \quad (3)$$

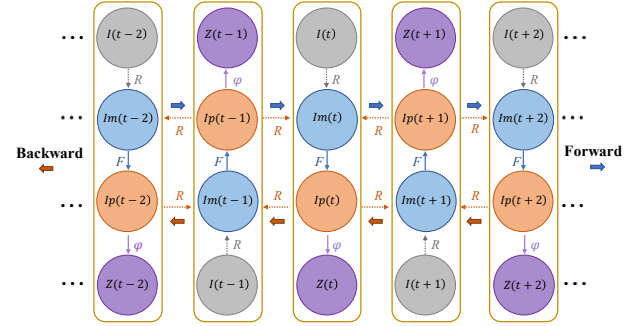


Fig. 5. The pipeline for the primary segmentation stage with a bi-directional recurrent update scheme (where 3 channels are used in this case).

A two-way propagation method is employed to propagate the local 3D hybrid information to each slice. As shown in Fig. 5, the blue and red arrows represent information flow in the forward and backward directions, respectively. This bi-directional flow avoids loss of local 3D information between adjacent slices and ensures that this information is fully integrated. When the segmentation of each slice is completed and made closer to the manual annotation, the corresponding probabilistic map of multi-channel data is updated. This ensures that the final output is guided by the local optimal 3D hybrid information.

Since any function involving a loop can be modeled by a recurrent neural network (RNN) [34], we compare here the bi-directional recurrent update structure with an RNN structure. Yu *et al.* [29] directly optimized an RNN-based segmentation model, at a high computational cost in the training phase. Yang *et al.* [31] used LSTM in the segmentation refinement module with serialized volume data. This approach required mapping the input sequence  $(\dots, x_{t-2}, x_{t-1}, x_{t+1}, x_{t+2}, \dots)$  into the hidden unit  $h_t$ , and this led to a large increase in model parameters and complexity. In this work, we select the adjacent slices to guide the segmentation process. Thus, unlike using a RNN structure, our proposed model has the following two advantages [34]: a) the model does not need additional parameters because it only transfers the guidance information from one probability

distribution to another; b) the recurrent update process is avoided in the training phase when the primary segmentation model  $F$  is used at each step. In brief, our PBR-UNet framework does not require many training parameters and has a greatly reduced computational cost in the training phase. More importantly, under the guidance of the effective contextual information within a certain range, the burden of obtaining an optimal refined pancreas segmentation result is highly reduced.

#### D. Loss Function and Inference Procedure

We follow the strategy of using the DSC for training and similarity characterization [19]. Given the ground truth map  $\hat{Y}$  and the final output  $\Omega(X)$ , then the loss function can be defined as

$$Loss(\Omega(X), \hat{Y}) = 1 - \frac{2 \times \sum_i \Omega(X_i) \hat{Y}_i}{\sum_i \Omega(X_i) + \sum_i \hat{Y}_i} \quad (4)$$

---

#### Algorithm 1: Probabilistic Maps Guided Bi-directional Recurrent U-Net

---

**Input:** input volume  $I$ ;

**Output:** segmentation volume  $Z$ ;

```

1:  $Ip_\theta \leftarrow E(I)$ ,  $I, Ip_\theta \in R^{m \times h \times w \times c}$ ;
2:  $Im_\theta \leftarrow S(I, Ip_\theta)$ ,  $Im_\theta \in R^{m \times h \times w \times 3c}$ ;
3: bi-directional process:
4:   for  $i \leftarrow 1$  to  $n$  (forward):
5:      $Im_{\theta,i} = (Ip_{\theta,i-1}, I_i, Ip_{\theta,i+1})$ ,  $Im_\theta \in R^{m \times h \times w \times 3c}$ ;
6:      $Ip_{\theta,i} \leftarrow F(Im_{\theta,i})$ ;
7:      $Im_{\theta,i-1}^{(2)}, Im_{\theta,i+1}^{(0)} \leftarrow R(Ip_{\theta,i}, Im_{\theta,i-1}^{(2)}, Im_{\theta,i+1}^{(0)})$ ;
8:      $Ip_{\theta,i} \leftarrow R(Ip_{\theta,i}, Ip_{\theta,i})$ ;
9:   for  $j \leftarrow n$  to 1 (backward):
10:     $Im_{\theta,j} = (Ip_{\theta,j-1}, I_j, Ip_{\theta,j+1})$ ,  $Im_\theta \in R^{m \times h \times w \times 3c}$ ;
11:     $Ip_{\theta,j} \leftarrow F(Im_{\theta,j})$ ;
12:     $Im_{\theta,j-1}^{(2)}, Im_{\theta,j+1}^{(0)} \leftarrow R(Ip_{\theta,j}, Im_{\theta,j-1}^{(2)}, Im_{\theta,j+1}^{(0)})$ ;
13:     $Ip_{\theta,j} \leftarrow R(Ip_{\theta,j}, Ip_{\theta,j})$ ;
14:    $Z^{[t]} = \varphi(Ip_\theta \geq 0.5)$ ;

```

**Return:**  $Z \leftarrow Z^{[t]}$

---

The overall flow of our PBR-UNet is shown in Algorithm 1. The variables used are defined as follows. The volumetric CT data  $I$  is the input, and  $Z$  is the output.  $Im_{\theta,i-1}^{(j)}$  denotes the  $j$ th channel of the  $(i-1)$ st data sample in  $Im_\theta$ . Before the bi-directional recurrent update scheme, the Multi-UNet architecture is used to get the probabilistic map  $Ip_\theta$ , which is denoted as the output of the function  $E$  in Algorithm 1. Next, each input  $I$  and the corresponding initial estimated probabilistic maps  $Ip_\theta$  are combined by the transformation function  $S$  to obtain the multi-channel data sample  $Im_\theta$  which is then fed into the 2.5D U-Net for the primary segmentation stage. Forward and backward information flow procedures in the segmentation process are described as follows:

##### 1) Forward Flow

The output of the 2.5D U-Net is a new probabilistic map  $Ip_{\theta,i}$ , which is used to update the corresponding part of the multi-channel data samples  $Im_{\theta,i-1}$  and  $Im_{\theta,i+1}$  to get the new multi-channel data samples  $Im_{\theta,i-1}^{(2)}$  and  $Im_{\theta,i+1}^{(0)}$ , respectively. This

update process is performed after each primary segmentation stage until all slices are segmented.

##### 2) Backward Flow

After the forward flow procedure, most of the  $Ip_\theta$  and  $Im_\theta$  values have been updated. But one-way flow only gets the information from the previous slice while the information of the next slice is lost. Therefore, we repeat the same process in the backward direction to complete a bi-directional recurrent update scheme. Finally, the output probabilistic map is binarized (as defined in Equation 4) to obtain the final segmentation result  $Z$ .

### III. EXPERIMENTS AND RESULTS

We used three evaluation metrics (Sec. III.B) to evaluate the performance of our proposed model by quantitative and qualitative analyses (Sec. III.D) on an authoritative public dataset (Sec. III.A). Moreover, the experiments were aimed to evaluate the efficiency of our model (Sec. III.E). The hardware setup and training settings were reported in Sec. III.C.

#### A. Dataset and Pre-processing

Following most of the earlier approaches, we used the NIH Pancreas-CT dataset [35] to extensively and quantitatively evaluate our proposed model. This dataset contains the abdominal CT scans of 82 patients where each scan has a size of  $512 \times 512 \times L$ , and  $L \in [181, 466]$  is the number of slices in each CT scan. We empirically truncated the CT radio-density values to the range of  $[-100, 200]$  HU and normalized them to have a zero mean and unit variance. All CT scans were cropped to a size of  $[192, 240]$ , which still can fully cover the pancreas in the scans. A 4-fold cross-validation scheme was used, we split the dataset into 4 fixed folds. Every time, we train the models on 3 out of 4 subsets and test them independently on the remaining one and all parameters remain the same. Finally, we report the average of the four folds.

#### B. Evaluation Metrics

To evaluate the pancreas segmentation performance, we mainly used the DSC. We also used the root-mean-square error (RMSE) and the Hausdorff distance (HD) [36] to evaluate the results of modeling the inter-slice shape continuity. The RMSE metric is the square root of the sum of the squared deviations between the observed and true values where this sum is normalized by the number of observations,

$$RMSE(x, y; \Omega) = \left( (1/m) \sum_{i=1}^m ((\Omega(x^i) - y^i)^2) \right)^{1/2} \quad (5)$$

The HD index represents the maximum deviation between two-point sets or surfaces,

$$HD(X, \hat{Y}; \Omega) = \max_{x \in X} \left\{ \min_{\hat{y} \in \hat{Y}} \{ \|\Omega(x), \hat{y}\| \} \right\} \quad (6)$$

We also used the recall, precision, and intersection over union (IOU) to evaluate the pancreas localization performance. The IOU measure is the ratio of the intersection to the union of the predicted and real borders,

$$IOU(X, \hat{Y}; \Omega) = (\Omega(X) \cap \hat{Y}) / (\Omega(X) \cup \hat{Y}) \quad (7)$$

The recall indicates the proportion of the pancreas pixels (or voxels) that are correctly segmented to the total number of true pancreas pixels (or voxels),  $Recall(X, \hat{Y}; \Omega) = |\Omega(X) \cap \hat{Y}| / \hat{Y}$ . The precision reflects the proportion of the pancreas pixels that



are correctly segmented to the total number of pixels that are labeled as a part of the pancreas,  $Precision(X, \hat{Y}; \Omega) = |\Omega(X) \cap \hat{Y}| / \Omega(X)$ . Moreover, we measured the standard deviation, maximum and minimum values, and then calculated the average of these metrics over all test cases.

### C. Implementation details

Our model was implemented in the Keras framework based on TensorFlow [37]. We built the initial estimate model and the primary segmentation model on an NVIDIA GeForce GTX 1080Ti GPU. For the initial estimate model, the initial learning rate was 0.0001 for 300 epochs of the stochastic gradient descent (SGD) algorithm, and 0.00001 for 400 epochs of the adaptive moment estimation (Adam) algorithm. The batch size was set to 1. For the primary segmentation model, the initial learning rate was 0.00001 for 300 epochs of the Adam optimizer. To prevent the model from falling into a local minimum, we used a method of learning rate annealing.

### D. Quantitative and Qualitative Analyses

In this section, we conduct comprehensive experiments to analyze the effectiveness of our proposed PBR-UNet. We first introduce the performance of our proposed PBR-UNet and compare the results of different stages (the three stages: U-Net, Multi-UNet and PBR-UNet) for ablation testing (part 1 and part 2). Then we introduce parameter selection and time consumption (part 3 and part 4), and finally we compare our results with other state-of-the-art methods (part 5). The models involved in the following experiments are defined as follows: we use U-Net as our baseline, we use the three-view method (called Multi-UNet) on U-Net to extract probability maps, and finally we add Multi-UNet for primary. The segmentation bidirectional recurrent network constitutes our proposed PBR-UNet.

#### 1) Segmentation performance

We quantitatively evaluated the segmentation results using the DSC, RMSE, and HD indicators. The values of these indicators are reported in Table I, which shows that the mean DSC, RMSE, and HD values of our model reached  $84.19 \pm 5.73\%$ ,  $3.60 \pm 2.57$  mm, and  $3.19 \pm 0.40$  mm, respectively. Moreover, The DSC of the four folds are 81.91%, 85.11%, 84.79%, and 84.93%. Fig. 6 shows our segmentation results, most of the errors occurred at the pancreas edges, while the main part of the pancreas was correctly segmented. This demonstrates a satisfactory performance of our proposed method.

TABLE I  
SEGMENTATION RESULT FOR THE NIH PANCREAS-CT DATASET.

Evaluation	Min	Max	Mean	Std
DSC (%)	53.61	91.08	84.19	5.73
HD (mm)	2.24	4.32	3.19	0.40
RMSE (mm)	1.40	18.05	3.60	2.57

Fig. 7(a) shows that the DSC values of most of our results were distributed above 85%. This indicates that our method could improve the model generalization performance for different CT volumes. For a slice-wise assessment of the performance of our model, we also investigated the DSC distribution of each slice of the 23rd patient at a DSC value of 79.61%, which is below the mean DSC value of the test cases (See Fig.

7(b)). Except for the last two slices at the pancreas tail, and due to the small target area, the DSC values of most slices were all stable above 60%, which was still an encouraging performance. Thus, our model had excellent generalization performance for different patient data both on the volume and slice levels.

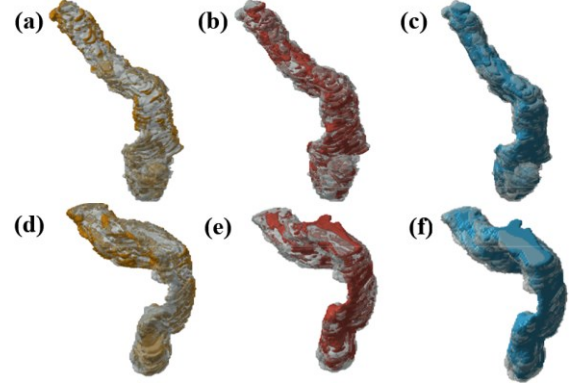


Fig. 6. Three-dimensional representations of the PBR-UNet segmentation results compared to the manually labeled reference standards (best viewed in color). The first and second row show the segmentation results of the 10th patient (DSC 90.74%) and the 7th patient (DSC 90.76%), respectively. The columns from left to right show the under-segmentation, over-segmentation, and overall segmentation results, respectively.

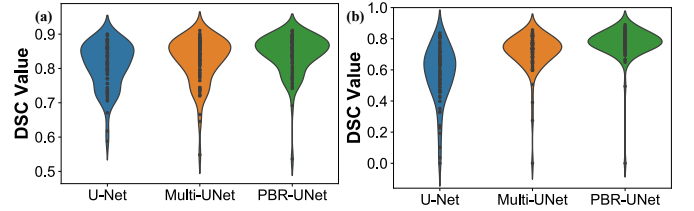


Fig. 7. The DSC distribution for the segmentation results of three different deep architectures. (a) The DSC distribution for all test cases; (b) The DSC distribution for the 23rd patient.

TABLE II  
COMPARISON OF THE PANCREAS LOCALIZATION PERFORMANCE BASED ON THREE METRICS AND THREE DEEP ARCHITECTURES.

Methods	Recall [%]	Precision [%]	IOU [%]
U-Net	76.86 $\pm$ 11.40	79.86 $\pm$ 7.33	67.04 $\pm$ 10.78
Multi-UNet	74.58 $\pm$ 11.26	<b>85.65 <math>\pm</math> 6.38</b>	72.03 $\pm$ 9.01
PBR-UNet	<b>82.23 <math>\pm</math> 9.14</b>	81.44 $\pm$ 7.53	<b>72.62 <math>\pm</math> 9.09</b>

We applied the recall, precision, and IOU metrics to quantitatively evaluate the pancreas localization performance of our model. These three metrics are mainly focused on the shape and position of the segmented object. As shown in Table II, the recall, precision, and IOU reached 82.23%, 81.44%, and 72.62%, respectively. This shows that our model could accurately locate the pancreas in a global search. We also noticed that the Multi-UNet architecture achieved the best precision results because the averaging of the results of the three views would eliminate many uncertain pixels, as we discuss later.

To justify the complementary roles of the initial estimation and the primary segmentation in our proposed framework, we compared the results of different stages (the three stages: U-Net, Multi-UNet, and PBR-UNet) in ablation experiments. First, we focused on the role of the initial estimation stage. Specifically, the basic U-Net model was used for segmentation with respect

to three views, and then the results from these views were averaged. To verify the validity of this averaging process, we compared the results of Multi-UNet and U-Net. As shown in Fig. 7(a), the Multi-UNet segmentation results were mostly concentrated at a DSC value of 0.85 compared to the dispersed results of U-Net. This indicates that most of the segmentation results were in this interval. For the results of the 23rd patient, Fig. 7(b) shows a more significant improvement, with the DSC values of most of the slices being increased from 0.62 to around 0.78. The performance based on the IOU and precision metrics was also improved, while the IOU value of the Multi-UNet segmentation result was 5.58% higher than that obtained by U-Net as Table II shows. The above statistics illustrate the effectiveness of the adopted averaging mechanism in our proposed model.

TABLE III  
THE NUMBER OF SLICES IN THE DSC INTERVALS FOR SEGMENTATION METHODS.

Methods	U-Net	Multi-UNet	PBR-UNet
DSC $\in [0,0.5)$	647 (9.42%)	1540 (22.42%)	<b>357 (5.19%)</b>
DSC $\in [0.5,0.6)$	338 (4.92%)	<b>483 (7.03%)</b>	202 (2.94%)
DSC $\in [0.6,0.7)$	614 (8.94%)	<b>886 (12.90%)</b>	413 (6.01%)
DSC $\in [0.7,0.8)$	1165 (16.96%)	<b>1583 (23.04%)</b>	1159 (16.87%)
DSC $\in [0.8,0.9)$	2764 (40.24%)	1983 (28.87%)	<b>3133 (45.61%)</b>
DSC $\in [0.9,1)$	1340 (19.51%)	393 (5.72%)	<b>1604 (23.35%)</b>

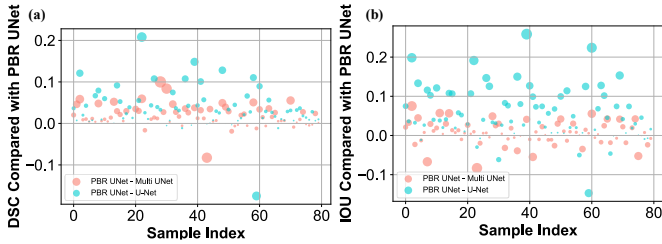


Fig. 8. Comparison of the segmentation and localization performance for three deep architectures. (a) Comparison of the segmentation results; (b) Comparison of the localization results.

Then, we assessed the performance of the bi-directional recurrent update scheme with local 3D hybrid information. Fig. 7 and Table II indicate that, in comparison with the results obtained by Multi-UNet, our PBR-UNet achieved further improvements in the pancreas segmentation and localization on the volume and slice levels as demonstrated by the recall and IOU values. Furthermore, according to the improved DSC and IOU values for each patient (as shown in Fig. 8), the best performance for most patients was achieved by the proposed PBR-UNet. This highlights the effectiveness of the bi-directional recurrent update with local 3D hybrid information. Fig. 9(a) demonstrates that U-Net and Multi-UNet happened to obtain a DSC value of 0. This means that their segmentation ultimately failed, while our model still achieved a DSC of 77%. Irrespective of whether U-Net and Multi-UNet had poor or good results, our proposed model still achieved better performance with a DSC exceeding 80% (Fig. 9(b) and (c)). Thus, our proposed PBR-UNet model did not only improve the performance of localization but also enhanced the pixel-level segmentation performance.

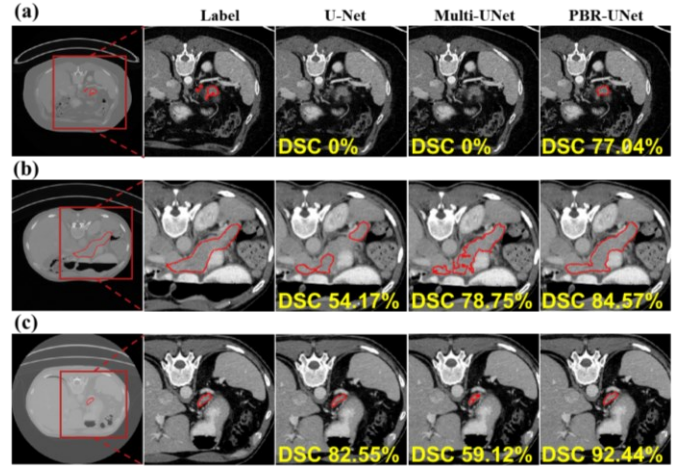


Fig. 9. Visual segmentation results for U-Net, Multi-UNet, and PBR-UNet. Parts a, b and c denote the 85th slice of the 6th patient, the 123rd slice of the 23rd patient, and the 83rd slice of the 20th patient, respectively.

## 2) Reliability analysis

Fig. 10 demonstrates the volume-wise reliability of our model. The ordinate (reliability) of each point (DSC, reliability) shows the ratio of the test results whose DSC was greater than the horizontal coordinate (DSC). Our proposed PBR-UNet model achieved DSC values greater than 0.8 in 81% of the cases. The PBR-UNet curve showed a significant downward trend at DSC=0.75. This means that most segmentation results had DSC values higher than 75%. By contrast, the Multi-UNet model exceeded DSC=0.8 for 72% of the cases, while the U-Net model reached a DSC value of 0.8 in only 61% of the cases. Therefore, our proposed PBR-UNet method could significantly improve segmentation stability and ensure that most test cases achieve excellent segmentation performance.

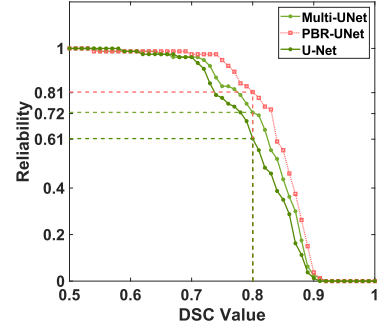


Fig. 10. The segmentation reliability for the U-Net, Multi-UNet, and PBR-UNet.

The slice-based segmentation performance was also important for the integrity of pancreas segmentation. Thus, we analyzed the distribution of the number of slices in the DSC intervals for different models, as shown in Table III. The slices of 69.96% of the cases for the proposed PBR-UNet method were distributed in  $[0.8,1.0)$ , while only the slices of 5.19% of the cases had a DSC value below 0.5. Compared to the distributions of the U-Net and Multi-UNet architectures, our method successfully improved the segmentation performance for a large number of slices. Therefore, our model not only had excellent volume-wise performance but also had an acceptable slice-wise distribution of the results. This indicates that our model had a reliable performance at the volume and slice levels.

TABLE IV  
THE DSC OF THE SEGMENTATION RESULTS FOR SLICES WITH SMALL PANCREAS SIZES.

Methods	Head and tail, 480 cases		size $\leq 300$ , 496 cases	
	Failed slices	Mean	Failed slices	Mean
U-Net	102	$50.07 \pm 31.89$	112	$46.52 \pm 30.94$
Multi-UNet	175	$27.80 \pm 29.22$	193	$29.48 \pm 31.37$
PBR-UNet	<b>84</b>	<b><math>54.40 \pm 30.77</math></b>	<b>90</b>	<b><math>51.43 \pm 29.73</math></b>

In Table III, we can still find that poorly segmented slices accounted for a small percentage of the results of the method we proposed. In fact, most of the poorly segmented slices had a small target area. Thus, we focused on slices with small-sized targets. Without loss of generality, we calculated the segmentation results of 3 slices of the pancreas head and tail and slices at which the pancreas size is less than 300 pixels (See Table IV). The segmentation results suffered a dramatic drop due to the small size in the selected slices. However, despite the lack of sufficient context information, our model could still use local 3D hybrid information to optimize the segmentation of the poorly segmented slices and achieve the highest average DSC value of approximately 54%. Compared with the DSC results below 30% in Multi-UNet, our proposed method almost doubled the DSC values. Moreover, the number of slices with a DSC segmentation result of 0 was reduced by more than one half compared to Multi-UNet. Overall, our approach could significantly improve the segmentation performance of U-Net and Multi-UNet in terms of segmentation small target areas.

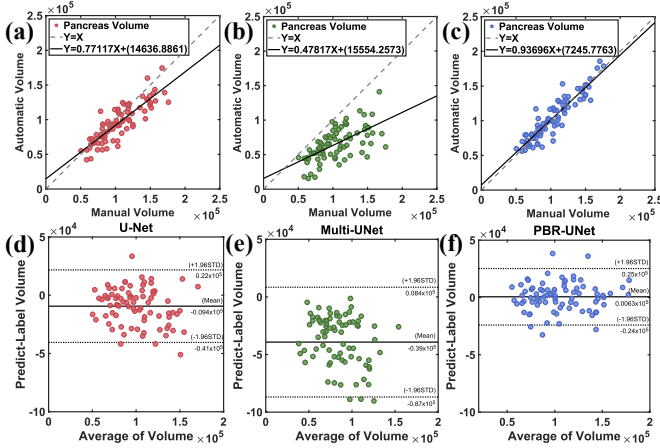


Fig. 11. Segmentation volumes of the U-Net, Multi-UNet, and PBR-UNet architectures versus the reference standard volumes. The first row shows the correlation of the segmentation volumes with the manually labeled volumes. The second row shows the results of the Bland-Altman agreement test which again compares between the manually segmented volumes and the segmentation volumes.

To further assess the agreement between the automatic and human-guided segmentation results, we compared the volumetric correlation between the expert annotation and the results of our model. Specifically, we adopted the correlation coefficient and the Bland-Altman agreement tests for a comprehensive assessment. The slope of the linear regression curve represents the consistency of the model-based pancreas volume and the manually labeled reference volume. The closer the slope of the curve to 1 is, the better the segmentation result is. As Fig. 11(c)

shows, the linear regression curves indicate a strong correlation between the automated and the manual segmentation results. While the linear fitting results of Multi-UNet produced significant degradation, our model could still correct deviations to a high correlation coefficient. Furthermore, our model not only achieved a high correlation coefficient (0.9369), but more than 93% of the measurements were still within the  $\pm 1.96$  standard deviation in Bland-Altman test (Fig. 11(f)). This further demonstrates the reliability of our model.

### 3) Parameter selection

The range of contextual information used for guiding the segmentation was limited to a fixed depth. We conducted comparative experiments to explore the impact of the guidance depth on the segmentation results. We set the navigation depth to 1, 2, and 3, respectively, while the channels of the multi-channel data would be 3, 5, and 7, respectively. All experiments were conducted under the same experimental settings. From Table V, PBR-UNet with a guidance depth of 1 achieved the best overall performance. Obviously, the results at a guidance depth of 1 and 3 were much better than those at a depth of 2. Increasing the guidance depth would enrich the local 3D hybrid information but would also introduce irrelevant information and increase computational resource consumption. Thus, we use a guidance depth of 1 for subsequent experiments.

TABLE V  
COMPARISON OF THE PBR-UNET SEGMENTATION RESULTS WITH DIFFERENT GUIDANCE DEPTHS.

Depth	DSC [%]	RMSE [mm]	HD [mm]
1	<b><math>84.19 \pm 5.73</math></b>	$3.60 \pm 2.57$	$3.19 \pm 0.40$
2	$83.06 \pm 6.03$	<b><math>3.25 \pm 2.26</math></b>	$3.22 \pm 0.39$
3	$84.03 \pm 5.72$	$3.77 \pm 2.64$	<b><math>3.19 \pm 0.39</math></b>

TABLE VI  
TIME CONSUMPTION COMPARISON WITH DIFFERENT DEEP LEARNING MODELS.

Methods	Time [s]		
	Min	Max	Mean
U-Net	1.05	4.04	$1.49 \pm 0.45$
Multi-UNet	5.66	12.74	$7.02 \pm 1.14$
PBR-UNet	7.95	18.24	$10.18 \pm 1.61$

### 4) Time consumption

Our proposed bi-directional recurrent network based on probabilistic map guidance represents a lightweight solution with low computational and time resource consumption. The data for different patients had different slices. So, we calculated the test time for each patient data sample to quantify the model efficiency. As shown in Table VI, after loading the model, the time required for each stage is 1.49s, 7.02s, and 10.18s, respectively. In conclusion, the time consumption of our model was much better than that of cascaded FCNs with 3D dense conditional random fields (CRFs) [38], which took just below 100 s per volume with NVIDIA Titan X GPU with 12 GB VRAM.

### 5) Comparison with other methods

As shown in Table VII, our model exhibits a competitive performance against recent state-of-the-art pancreas segmentation



methods. For our model, a DSC value of 84.19% was achieved. This figure is exceedingly close to the highest result of 84.50% which was obtained by Yu et al. [29]. Although our results were 0.31% worse than the best result, the relatively small number of parameters in our model made the test time much shorter. Yu et al. [29] needed 1.3 minutes with Titan-X Pascal GPU, and Zhou et al. needed 3 minutes with a modern GPU [1], which were much longer than the average time (10 seconds) of the proposed model. Moreover, our method performs better than other 3D methods and most 2.5D methods, which demonstrates the effectiveness of our strategy of updating the local 3D hybrid information in a bi-directional loop.

TABLE VII  
COMPARISON OF PANCREAS SEGMENTATION RESULTS WITH THE STATE-OF-THE-ART METHODS (MEASURED BY DSC [%]).

Models	Year	Method	Folds	DSC
Roth et al. [3]	2015	2.5D	4-CV	71.42 $\pm$ 10.11
Farag et al. [39]	2016	2.5D	6-CV	70.70 $\pm$ 13.00
Zhou et al. [1]	2017	3D	4-CV	82.37 $\pm$ 5.68
Karasawa et al. [7]	2017	MALF	LOOCV	78.50 $\pm$ 14.00
Roth et al. [26]	2018	2.5D	4-CV	81.27 $\pm$ 6.27
Oktay et al. [17]	2018	3D	5-CV	81.48 $\pm$ 6.23
Fu et al. [8]	2018	2D	5-CV	76.36 $\pm$ 14.34
Yu et al. [29]	2018	2.5D	4-CV	<b>84.50 <math>\pm</math> 4.97</b>
Asaturyan et al. [40]	2019	3D	Hold-out	79.3 $\pm$ 4.4
Ours	2019	2.5D	4-CV	84.19 $\pm$ 5.73

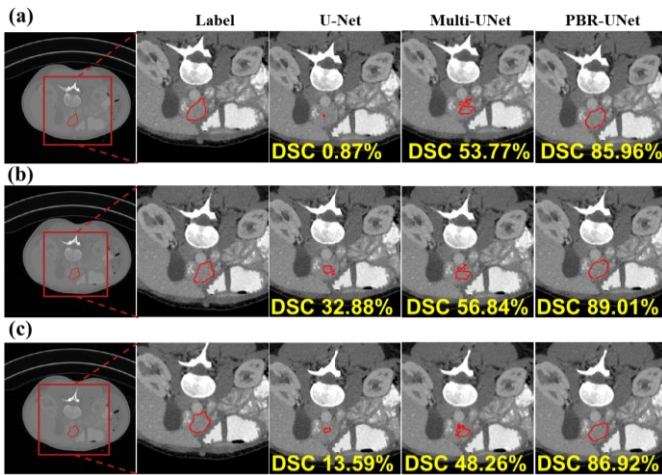


Fig. 12. The segmentation results of three adjacent slices with small target areas. The parts a, b and c in the figure denote the 145th, 146th, and 147th slices, respectively for the 3rd patient.

On the other hand, most pancreas segmentation approaches paid considerable attention to the overall segmentation performance, with little focus on the segmentation of small targets. However, due to the large shape and volume variability of the human pancreas, the ability to segment small-sized targets is an important indicator of segmentation stability. Therefore, we discussed the segmentation results of slices with small sizes in the above section on reliability analysis. Indeed, our model could optimize the segmentation results by propagating local

3D hybrid information. Fig. 12 shows the segmentation results of three adjacent slices (the 145th, 146th, and 147th slices) of the 3rd patient. Due to the small size of the pancreas in these three slices, the DSC value for the initial segmentation results was below 60%. Based on the excellent segmentation results of the adjacent slices of the specified three slices (e.g. DSC = 91.56% for the 143rd slice), the segmentation results for the three slices all reached a DSC exceeding 85%. This emphasized that our model effectively used local 3D hybrid information, making large improvements in slices with poor segmentation.

#### IV. DISCUSSION

##### A. Comparison on Computing Resources

Pancreas segmentation is of great significance for clinical computer-aided diagnosis. The segmentation results can provide accurate location and contour information of the pancreas. In this paper, we present a segmentation network guided by probabilistic maps, which aims to extract local 3D hybrid information without requiring excessive computational resources. We used an NVIDIA GeForce GTX 1080Ti GPU with 11GB VRAM to explore the need for computing resources in 3D networks. We adopted a basic 3D U-Net [21] model with an input size of  $120 \times 120 \times 120$ . The training and testing of this network couldn't continue because of memory insufficiency. Therefore, the network we proposed can be used as a lightweight solution when the 3D model cannot be performed smoothly with poor hardware resources.

##### B. On Merging the Segmentation Results of Three Views

In Fig. 11(b), we find that the linear regression curve of Multi-UNet showed a large shift, with a correlation coefficient of only 0.47. This indicated that the volume of Multi-UNet segmentation results was significantly different from the expert annotation. The reason for this result is the merging mechanism for the segmentation results of the three views. The merging method in this paper is simply an averaging one. Hence, a pixel could be regarded within the pancreas area only when two views assert its presence within the pancreas area. So, many uncertain pixel areas could be discarded after averaging. Nevertheless, this mechanism gives higher confidence in the surviving pixels. As shown in Table II, the precision value of Multi-UNet is the highest. Overall, the result of fusing the segmentation outputs of the three views is a valuable and practicable scheme [1], [29], as verified by our experiments.

#### V. CONCLUSIONS

The current key challenge in pancreas segmentation is that 3D networks require high computing resources, while 2D networks cannot capture contextual information. Therefore, we propose a PBR-UNet for pancreatic segmentation. This network combines the intra-slice information and probabilistic maps of the adjacent slices into local 3D hybrid information for guiding segmentation. This information could be optimized by a bi-directional recurrent updating scheme. Extensive experiments have demonstrated that our proposed PBR-UNet method could achieve competitive results compared with other state-of-the-art methods. Besides, this new paradigm with the local 3D

hybrid information and bi-directional recurrent updating scheme can be integrated with other segmentation models for compromising the trade-offs between 2D and 3D segmentation networks. Thus, our proposed PBR-UNet method not only provides a useful tool for pancreatic segmentation but also a potential paradigm for research in medical image segmentation.

## REFERENCES

- [1] 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," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10433 LNCS, pp. 693–701, 2017.
- [2] H. R. Roth, A. Farag, L. Lu, E. B. Turkbey, and R. M. Summers, "Deep convolutional networks for pancreas segmentation in CT imaging," no. March 2015, 2015.
- [3] H. R. Roth *et al.*, "Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9349, pp. 556–564, 2015.
- [4] M. Oda *et al.*, "Regression Forest-Based Atlas Localization and Direction Specific Atlas Generation for Pancreas Segmentation," 2016, pp. 556–563.
- [5] R. Wolz, C. Chu, K. Misawa, M. Fujiwara, K. Mori, and D. Rueckert, "Automated abdominal multi-organ segmentation with subject-specific atlas generation," *IEEE Trans. Med. Imaging*, vol. 32, no. 9, pp. 1723–1730, 2013.
- [6] T. Tong *et al.*, "Discriminative dictionary learning for abdominal multi-organ segmentation," *Med. Image Anal.*, vol. 23, no. 1, pp. 92–104, 2015.
- [7] K. Karasawa *et al.*, "Multi-atlas pancreas segmentation: Atlas selection based on vessel structure," *Med. Image Anal.*, vol. 39, pp. 18–28, 2017.
- [8] M. Fu *et al.*, "Hierarchical combinatorial deep learning architecture for pancreas segmentation of medical computed tomography cancer images," *BMC Syst. Biol.*, vol. 12, no. Suppl 4, 2018.
- [9] E. Shelhamer, J. Long, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," pp. 1–12, 2016.
- [10] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," pp. 1–8, 2015.
- [11] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, 2018.
- [12] M. Drozdal, E. Vorontsov, G. Chartrand, S. Kadoury, and C. Pal, "The Importance of Skip Connections in Biomedical Image Segmentation," 2016.
- [13] G. González, G. R. Washko, and R. San José Estépar, "Multi-structure Segmentation from Partially Labeled Datasets. Application to Body Composition Measurements on CT Scans," in *Image Analysis for Moving Organ, Breast, and Thoracic Images*, 2018, pp. 215–224.
- [14] 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," pp. 1–8, 2017.
- [15] J. Cai, L. Lu, F. Xing, and L. Yang, "Pancreas Segmentation in CT and MRI Images via Domain Specific Network Designing and Recurrent Neural Contextual Learning," no. Cv, pp. 1–11, 2018.
- [16] 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. Imaging*, no. 1, pp. 1–13, 2018.
- [17] O. Oktay *et al.*, "Attention U-Net: Learning Where to Look for the Pancreas," no. Midl, 2018.
- [18] Q. Dou, H. Chen, Y. Jin, L. Yu, J. Qin, and P.-A. Heng, "3D Deeply Supervised Network for Automatic Liver Segmentation from CT Volumes," in *Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2016*, 2016, pp. 149–157.
- [19] S. Liu *et al.*, "3D anisotropic hybrid network: Transferring convolutional features from 2D images to 3D anisotropic volumes," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11071 LNCS, pp. 851–858, 2018.
- [20] Z. Zhu, Y. Xia, W. Shen, E. Fishman, and A. Yuille, "A 3D coarse-to-fine framework for volumetric medical image segmentation," *Proc. - 2018 Int. Conf. 3D Vision, 3DV 2018*, pp. 682–690, 2018.
- [21] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-net: Learning dense volumetric segmentation from sparse annotation," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9901 LNCS, pp. 424–432, 2016.
- [22] F. Milletari, N. Navab, and S. A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," *Proc. - 2016 4th Int. Conf. 3D Vision, 3DV 2016*, pp. 565–571, 2016.
- [23] Z. Quo *et al.*, "Deep LOGISMOS: Deep learning graph-based 3D segmentation of pancreatic tumors on CT scans," *Proc. - Int. Symp. Biomed. Imaging*, vol. 2018-April, no. Isbi, pp. 1230–1233, 2018.
- [24] 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," no. October, pp. 1–13, 2018.
- [25] H. R. Roth *et al.*, "An application of cascaded 3D fully convolutional networks for medical image segmentation," *Comput. Med. Imaging Graph.*, vol. 66, no. October 2017, pp. 90–99, 2018.
- [26] 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, 2018.
- [27] M. Lai, "Deep Learning for Medical Image Segmentation," 2017. [Online]. Available: <https://arxiv.org/abs/1505.02000>.
- [28] C. Chung *et al.*, "Very deep convolutional networks for large-scale image recognition," *Comput. Sci.*, 2014.
- [29] 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 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2018*, 2018, vol. 2, pp. 8280–8289.
- [30] S. Hochreiter and J. Schmidhuber, "Long Short-term Memory," *Neural Comput.*, vol. 9, pp. 1735–1780, 1997.
- [31] X. Yang *et al.*, "Towards Automated Semantic Segmentation in Prenatal Volumetric Ultrasound," *IEEE Trans. Med. Imaging*, vol. PP, no. c, pp. 1–1, 2018.
- [32] Y. Hua, L. Mou, and X. X. Zhu, "Recurrently Exploring Class-wise Attention in A Hybrid Convolutional and Bidirectional LSTM Network for Multi-label Aerial Image Classification," 2018.
- [33] Y. Xia *et al.*, "3D Semi-Supervised Learning with Uncertainty-Aware Multi-View Co-Training," 2018.
- [34] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. The MIT Press, 2016.
- [35] K. Clark *et al.*, "The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository," *J. Digit. Imaging*, vol. 26, no. 6, pp. 1045–1057, 2013.
- [36] A. A. Taha and A. Hanbury, "An Efficient Algorithm for Calculating the Exact Hausdorff Distance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 11, pp. 2153–2163, 2015.
- [37] F. Chollet *et al.*, "Keras," <https://github.com/keras-team/keras>, 2015.
- [38] P. F. Christ *et al.*, "Automatic Liver and Tumor Segmentation of CT and MRI Volumes using Cascaded Fully Convolutional Neural Networks," pp. 1–20, 2017.
- [39] A. Farag *et al.*, "A Bottom-up Approach for Pancreas Segmentation using Cascaded Superpixels and ( Deep ) Image Patch Labeling," *IEEE Trans. Image Process.*, vol. 26, no. 1, pp. 1–14, 2016.
- [40] H. Asaturyan, A. Gligorievski, and B. Villarini, "Morphological and multi-level geometrical descriptor analysis in CT and MRI volumes for automatic pancreas segmentation," *Comput. Med. Imaging Graph.*, vol. 75, pp. 1–13, 2019.