Pancreas Segmentation via Spatial Context based U-net and Bidirectional LSTM

Hao Li, Jun Li, Xiaozhu Lin, Xiaohua Qian*

Abstract

Pancreas is characterized by small size and irregular shape, so achieving accurate pancreas segmentation is challenging. Traditional 2D pancreas segmentation network based on the independent 2D image slices, which often leads to spatial discontinuity problem. Therefore, how to utility spatial context information is the key point to improve the segmentation quality. In this paper, we proposed a divide-and-conquer strategy, divided the abdominal CT scans into several isometric blocks. And we designed a multiple channels convolutional neural network to learn the local spatial context characteristics from blocks called SCU-Net. SCU-Net is a partial 3D segmentation idea, which transforms overall pancreas segmentation into a combination of multiple local segmentation results. In order to improve the segmentation accuracy for each layer, we also proposed a new loss function for inter-slice constrain and regularization. Thereafter, we introduced the BiCLSTM network for stimulating the interaction between bidirectional segmentation sequence, thus making up the boundary defect and fault problem of the segmentation results. We trained SCU-Net+BiLSTM network respectively, and evaluated segmentation result on the NIH data set.

Keywords: Pancreas Segmentation, Convolutional Neural Networks, Recurrent Neural Networks, Deep Learning, Inter-slice Regularization

I. Introduction

Pancreas is an important organ in human body. Accurate pancreas segmentation technology can be applied in many computer-assisted diagnosis and treatment systems, which is of great help to the diagnosis of pancreatic cancer and other diseases. However, up to now, locating the boundary of the pancreas is still a difficult problem. Due to the small size and irregular shape of the pancreas, it is very challenging to segment pancreas from MRI or CT accurately. Other organs with larger targets, such as liver, heart or kidney, have achieved a high Dice accuracy rate (>90%) [1][2][3]. In contrast, the accuracy of pancreatic segmentation is basically below 75%, as shown in [4][5][6]. Traditional image segmentation algorithms have not achieved good results when faced with pancreas segmentation problem. Nowadays, deep learning technology has shown its amazing potential in the field of medical images. Deep CNN segmentation method can achieve higher accuracy and stability on organ segmentation, so deep learning has gradually become popular in pancreas segmentation.

In the deep learning algorithm of medical image segmentation, FCN [8] and its variant U-Net [9] are now commonly used tools for pancreas segmentation, similar algorithms are used or improved in [10][11][12][13]. FCN can classify images at pixel level and solve the problem of semantic segmentation. In order to reduce the
irreversible spatial information loss caused by continuous down sampling operation, U-Net [9] improves the segmentation effect of FCN by establishing skip join to merge feature maps from different semantic levels. On the basis of U-Net, [13] added several branches to the upsampling process, similar to U-Net++ [14], and therefore achieved a segmentation result with 83.3 ± 5.6. In addition, the Holistically-Nested Networks (HNN)[15] which based on Deep CNN can also achieve good results in the aspect of pancreas segmentation. The advantages of 2D segmentation network have been well proved in the segmentation of liver or heart [11]. However, different from liver or heart, the shape and position of the pancreas varies widely from patient to patient. Pancreas segmentation only at 2D level ignores the continuity of pancreas in 3d space, which has great limitations. As shown in figure 1, the shape, size and distribution of the pancreas in different CT slice are quite different.

![Figure 1](image_url)

Fig. 1. Illustration of (a) is a three-dimensional pancreatic model of a patient in HIN dataset and (b) from top to bottom: 106th, 132nd and 148th CT slice of this patient, the outline of pancreas is marked by red line.

In 2016, the proposed of V-Net [19] and 3D U-Net [20] expanded the dimension of 2D segmentation network to 3D. Paper [18] demonstrated the feasibility of 3DFCN [18] in medical image processing. 3D segmentation network takes 3D data as input, carries out image processing through corresponding 3D operations, including 3D convolutional layer, 3D maximum pooling layer and 3D deconvolutional layer, and finally obtains the result of 3D overall segmentation. 3DFCN, represented by V-Net and 3D U-Net, can directly extract features from 3D spatial information, thus avoiding the bottleneck problem of 2D segmentation network. [17] use the method of random forest plus 3DFCN for pancreas segmentation, achieving Dice rate of 89.7 ± 3.8 on 147 cases of enhanced abdominal CT scans. However, the computation burden of 3DFCN is much higher than 2D segmentation network, so 3DFCN will rely on computing resources to some extent. For example, the device in [16] is NVIDIA GeForce GTX TITAN X GPU (12GB), and the size of input data is 64*64*64. According to statistics, in the HIN dataset, the average three-dimensional space occupied by the pancreas is 86*192*256, which means that a larger computational resource is needed to meet the training requirements of 3DFCN.

In addition to 3DFCN, [13][16][21][22][23] have proved the recurrent neural network(RNN) is also an effective method to improve semantic segmentation result.
Variations of RNN include LSTM[24], GRU[25], BiRNN[26], etc. Among them, BiLSTM performs better in refining segmentation results. BILSTM can integrate the bidirectional information flow to make the overall segmentation result more continuous. [13] used BiRNN to optimize the results on the basis of 2D pancreatic segmentation, obtained higher edge fitting degree and Dice rate. However, since the basis of RNN optimization is the probability volumes from Preliminary segmented network, the quality of the probability volumes determines the upper limit of the optimization result of RNN. Many unnecessary errors may be introduced in the training process of RNN, if the segmentation result of the first part with poor quality.

To sum up, in the field of pancreas segmentation, 2D segmentation networks has its inherent limitations; 3D segmentation networks requires high computing resources; RNN refining result depends on the segmentation quality of the first part. This is the main motivation for the work in this paper.

In this paper, we propose a pancreas segmentation model based on spatial context information (SCU-Net) and bidirectional convolutional long-term memory networks (BiCLSTM). Our contributions are threefold: 1) Proposed SCU-Net to segment pancreas accurately from local spatial context information, which can make up the defect of 2D segmentation network. The abdominal CT scans was divided into equal-length block data as the input of the network, each block contains several CT slices. SCU-Net extracts features from three-dimensional block data in multi-channel manner, makes the segmentation result have more space continuity; 2) Proposed a new loss function to constrain and regular original loss function, so as to ensure the consistency between each layer and the overall segmentation results; 3) Utilize BiCLSTM network structure to optimize the results of SCU-Net and improve the edge fitting degree of the segmentation results, so as to obtain the final segmentation results. Finally, we obtained good verification on the public HIN data set, and our segmentation results have high accuracy and stability. In addition, our network has good portability and has great potential to be used in computer-assisted medical system in the future.

In a later section of this article, we will introduce the concrete implementation details of the algorithm in the section of Methodology, list our Experimental results and qualitative and quantitative analysis in section of Experimental Result, and give conclusions in the Discussion & Conclusion section.
II. Methodology

This paper focuses on the pixel-level pancreatic segmentation from abdominal CT scans. Define the CT scans input is \( X \in \mathbb{R}^{d \times l \times w} \), \( d, l, w \) are the depth, length and width of the CT scan data. The label \( Y \in \mathbb{R}^{d \times l \times w} \) of the segmentation result has the same size as \( X \). Prediction results of the network are represented by \( \hat{Y} \), the goal of our work is to get a prediction \( \hat{Y} \) as close as possible to \( Y \). In order to make the segmentation result have higher continuity and edge fitting degree, we used SCU-Net + BiCLSTM network framework to fuse spatial context information. On the basis of SCU-Net, we propose a new loss function to regularize the inter-slice relation of prediction results, and achieved better performance on pancreas segmentation. Figure 2 is the schematic diagram of our proposed framework. The input of network is a three-dimensional block data \( B \in \mathbb{R}^{k \times l \times w} \), contains \( k \)-slice CT scans, \( k < d \). According to the accurate labeling \( Y \) of pancreas, SCU-Net was trained by calculating new loss function and stochastic gradient descent algorithm. Thereafter, the BiCLSTM network is used to refine the initial segmentation results. Finally, we get the whole network segmentation result \( \hat{Y} \). The following detailed method introduction will be composed of three parts: 1) Preliminary pancreas segmentation with SCU-Net; 2) Inter-slice regularization with new loss function; 3) Prediction result refinement with BiCLSTM;

A. Preliminary pancreas segmentation with SCU-Net

2D segmentation network cannot consider the spatial context information, so the segmentation accuracy is often not high when segment small and irregular organs like pancreas. Therefore, on the basis of 2D segmentation, we designed the SCU-Net network to apply spatial characteristics to the segmentation process. SCU-Net network
is a kind of divide-and-conquer idea, that is, to conduct network training with three-dimensional block unit, extract local spatial features, and finally integrate them into the overall pancreatic segmentation results.

Fig. 3. Schematic view of SCU-Net network structure. CT scans data for each patient are divided into n blocks as the input of SCU-Net. In the network structure of SCU-Net, the blue block is the down sampling unit, the green block is the up sampling unit, and the yellow block is the transition unit.

In SCU-Net, we put three-dimensional block data \( B \in \mathbb{R}^{k \times l \times w} \) into the convolutional layer in a multi-channel way, calculate the loss function between the prediction result and precise label, and optimize the network parameter \( W \) by stochastic gradient descent method. Let function \( F \) represent the pancreas segmentation process of SCU-Net. We input block \( B_l \) into function \( F \), and then function \( F \) will get the prediction result according to the network parameter \( W \). In the end, we use the \( M \) function to map the image features to the same specifications as \( B \), and then obtain the prediction result probability volumes \( P_l \in \mathbb{R}^{k \times l \times w} \) through the sigmoid activation function. SCU-Net network model can be formulated,

\[
P_l = \sigma \cdot M[F(B_l; W_l); k], \quad B_l \in \mathbb{R}^{k \times l \times w} \quad (1)
\]

Where \( \sigma \) represents the sigmoid activation function. Since SCU-Net performs network training in units of three-dimensional block data, when calculating the loss function, we need to perform the overall Dice loss operation on the three-bit block data, as shown in Eq. 2,

\[
L_v(B_l, Y_l; W) = 1 - \frac{|Y_l \cap F(B_l; W)|}{|Y_l| + |F(B_l; W)|}, \quad B_l, Y_l \in \mathbb{R}^{k \times l \times w} \quad (2)
\]

In the specific network architecture design, we follow the U-Net classic network architecture, consisting of a down sampling path and an up sampling path. The down sampling path includes four down sampling units, each of which is composed of \( Conv + ReLu + BatchNorm + Maxpooling \); the up sampling path includes four upsampling units, each of which is composed of \( Upconcat + Conv + ReLu + BatchNorm \). Between the down sampling path and the up sampling path is the \( Bottom_{feat} \) transition layer. At the end of the network, a convolution layer is added to map the predicted results to multiple channels and restore them to the same size as
the input. The network structure is shown in Fig. 3.

In a patient's abdominal CT data, pancreatic segmentation results were only strongly correlated in CT slices of adjacent layers, while the correlation between the distant sections was minimal. Although SCU-Net cannot acquire the overall 3D-dimensional features of the pancreas like 3DFCN, it can effectively utilize the CT slice information of the upper and lower layers. Considering the slice information of the upper and lower layers helps to enhance the local spatial feature learning, thereby improving the accuracy of the pancreas segmentation as a whole. More importantly, SCU-Net can also save computing resources and improves network efficiency and portability.

**B. Inter-slice regularization with new loss function**

For pancreas segmentation, 2D segmentation networks generally use Dice loss=1−2(|Y∩Y^|)/(|Y|+|Y^|) or Jaccard loss=1−(|Y∩Y^|)/(|Y∪Y^|) as a loss function training network model[13][23]. This approach allows the segmentation results to achieve good DSC and JI indicators on the 2D level, but this does not take into account the spatial information of the pancreas, and the 2D segmentation results still have no spatial continuity. On the other hand, pancreas segmentation from the three-dimensional perspective is also neglecting the segmentation accuracy at the two-dimensional level. For such problem, we proposed a new loss function for inter-slice regularization. The new loss function is an improvement based on Eq. 2., which improves the accuracy of each layer segmentation result by means of a constraint similar to the L2 norm. Specifically, this method is to calculate the Dice loss function of each layer, and superimpose the loss results of each layer according to the L2 norm, denoted by ||L_i||_2^2. Finally, the overall loss function is superimposed by ||L_t||_2 and L_v in a certain proportion. New loss function can be formulated,

\[
L_t(B, Y ; w) = \lambda_v L_v(B, Y ; W) + \lambda_k ||L_t(x_i, y_i; w_l)||_2^2
\]  

(3)

Where W is the parameter of network; x_i ∈ B is the i-th slice in block data, y_i ∈ Y is the corresponding i-th label; \( \lambda_v \) and \( \lambda_k \) are the coefficients of the main term \( L_v \) and the regular term \( ||L_i||_2^2 \). The network generates a prediction result P according to B and W, and \( L_t \) is used as the overall loss function to measure the difference between P and Y. \( ||L_i||_2^2 \) is formulated as Eq. 4,

\[
||L_i(x_i, y_i; w_l)||_2^2 = \sum_{i=0}^{k} \left( 1 - \frac{|y_i \cap F(x_i; w_l)|}{|y_i| + |F(x_i; w_l)|} \right)^2
\]  

(4)

The L2 norm has great benefits for optimization training, which can effectively prevent over-fitting in the training process. We use this regularization strategy to add a regular term \( \Omega(w) = \lambda_i||L_i(x_i, y_i; w_l)||_2^2 \) to the target loss function \( L_t \), which is used to constrain the relationship between the whole and the individual. We can observe the regularization of weight decay by studying the gradient of the target loss function. Let \( s_t \) be the gradient direction of minibatch-SGD at time \( t \),

\[
s_t = -\varepsilon \sum_{l=0}^{k-1} (\nabla_w \lambda_v L_v + \nabla_w \lambda_k ||L_k||_2^2) + ps_{t-1}
\]  

(5)
Where $\varepsilon$ is the learning rate, $w$ is the network parameter, $k$ is the batch-size, $p<1$ is the momentum value, and $s_{t-1}$ is the gradient at the previous time. Adding weight attenuation causes a modification of the learning rules, that is, shrinking the weight before performing the usual gradient update at each step. Let $w^*$ be the weight vector before the regularization, $w^* = \arg \min_w L(w)$. And make a quadratic approximation to the objective function in the neighborhood of $w^*$. The approximate $\hat{L}(\theta)$ is as follows (Eq. 6):

$$\hat{L}(\theta) = L(w^*) + \frac{1}{2}(w - w^*)^T H(w - w^*)$$ (6)

Where $H$ is the Hessian matrix for $w$ and $w^*$. Since $w^*$ is defined as optimal of which gradient disappears to 0, there is no first-order term in the quadratic approximation. When $L^*$ is minimized, its gradient is 0,

$$\nabla_w \hat{L}(w) = H(w - w^*)$$ (7)

Add the weight attenuation gradient in the above equation, we use $\tilde{w}$ to represent the optimal point at this time,

$$\lambda \tilde{w} + H(\tilde{w} - w^*) = 0$$ (8)

$$\tilde{w} = (H + \lambda I)^{-1} H w^*$$ (9)

Since $H$ is real symmetric, we can decompose $H$ into a diagonal matrix $\Lambda$ and a standard orthogonal basis $Q$ of a set of eigenvectors, that is, $H = Q\Lambda Q^T$. Now $\tilde{w}$ can be expressed as,

$$\tilde{w} = (HQ\Lambda Q^T + \lambda I)^{-1} HQ\Lambda Q^T w^* = Q(\Lambda + \lambda I)^{-1}\Lambda Q^T w^*$$ (10)

It can be seen that the regularization has a smaller effect along the direction in which the $H$ eigenvalue is larger, and the component with a smaller eigenvalue shrinks to almost zero. In simple terms, only the parameters that are significantly reduced in the direction of the objective function will remain intact, the components corresponding to the unimportant direction are attenuated due to regularization during training. The L2 regularization allows the learning algorithm to perceive the input $x$ with a higher variance, so the weight of the feature with less covariance to the output target will shrink. When applied to the SCU-Net pancreas segmentation, new loss function will make the prediction result close to the label of the 3D block level as much as possible without losing the accuracy of the 2D layer.

\section*{C. Prediction result refinement with BiCLSTM}

Segmentation results of SCU-Net network depend on spatial concurrency in pancreatic features. Therefore, SCU-Net may not have good fitting effect at the boundary of segmentation results and has boundary defects to a certain extent. In addition, although SCU-Net can consider the spatial context information inside the block, the segmentation result may also have a large difference at the fault between the blocks. So, inspired by [13][16], we introduced RNN to make up for the shortcomings of SCU-Net.

By reconstructing the first step of semantic segmentation result data into a sequence...
format, the RNN can flexibly encode context information with dynamic hidden state and enhance local prediction. In this paper, we input the sequence of probability volumes output by SCU-Net into the RNN. The RNN can reveal the context dependencies, which proves to be the key to make up the boundary blur and missing of the segmentation results [16].

Fig. 4. The network framework of BiCLSTM, which consists of a forward CLSTM network and a backward CLSTM network. Where $\vec{h}_t$, $\vec{h}_t$ is the main form of CLSTM unit; $\hat{y}_1, \hat{y}_2, ..., \hat{y}_n$ is the output probability volumes sequence of SCU-Net, which is also the input of CLSTM network; The result sequence after refinement by BiCLSTM is $\tilde{y}_1, \tilde{y}_2, ..., \tilde{y}_t$.

LSTM is the most commonly used RNN variant. The standard LSTM network will directly vectorize the input data, which will sacrifice the spatial information encoded in the CNN output. In order to introduce CNN convolutional coding information, in CLSTM, the input sequence data is first preprocessed through a convolution layer and then placed in the LSTM unit for operation. Furthermore, since the segmentation results of each layer of pancreas are spatially correlated with the upper and lower layers, only the forward CLSTM will ignore the impact of the next segmentation result. For this reason, we extend the CLSTM network to BiCLSTM network for introducing contextual information from both directions.

Specifically, as shown in Fig. 4, BiCLSTM is composed of multiple CLSTM units in series, each of them consists of input gate $i_t$, forgetting gate $f_t$ and output gate $o_t$. For each CLSTM unit, its output state $\vec{h}_t$ is determined by the output $\vec{h}_{t-1}$ of the previous CLSTM unit, the cell state $c_{t-1}$, and the input probability map $\hat{y}_t$ at time $t$. The process of data flow in the BiCLSTM is shown in the following formula,

$$\vec{h}_t = W_h \vec{h}_{t-1} + W_c \vec{c}_{t-1} + W_{\hat{y}} \hat{y}_t + b_t \quad (11)$$

$$\vec{h}_t = W_h \vec{h}_{t-1} + W_c \vec{c}_{t-1} + W_{\hat{y}} \hat{y}_t + b_t \quad (12)$$

$$\tilde{y}_t = W_{\tilde{y}} \vec{h}_t + W_{\tilde{y}} \vec{h}_t + b_{\tilde{y}} \quad (13)$$
Pancreas Segmentation via Spatial Context based U-net and Bidirectional CLSTM

Where $W(c), b(c)$ represent parameters and bias values in the BiCLSTM model; $\hat{y}_t$ represent the prediction result at time $t$. Although the overall shape of the pancreas is irregular, there is a strong relationship between the upper and lower layers of the local space. Therefore, BiCLSTM can effectively make up the segmentation boundary defect and the fault between data blocks, and further improve the segmentation accuracy on the basis of SCU-Net.

**D. Algorithm pseudocode**

1) The algorithm procedure of training the SCU-Net network with the new loss function,

**Algorithm 1: The SUC-Net Training Phase**

**Input:** image serial $x_1, ..., x_n \in R^{d \times l \times w}$; mask serial $y_1, ..., y_n \in Y^{d \times l \times w}$; model weights $w$; max number of iterations $T$; the layers of blocks $k$; Threshold $thr$;

**Output:** prediction probability $\hat{y}_1, ..., \hat{y}_n$;

1) divide $X$ into several blocks $B_1, B_2, ..., B_s \in R^{k \times l \times w} (k < d)$;
2) load network weights $W, b$; iteration times $t \leftarrow 1$
3) repeat
4)   layer index of block: $l \leftarrow 1$;
5)   repeat
6)     $P_l = \sigma \cdot M(F(x_i; W_l; b_l); k), x_i \in B_l^{k \times l \times w}$;
7)     $L_t(B_l, Y_l; w) = \lambda_v L_v(B_l, Y_l; w) + \lambda_k \|L(x_i, y_i; W_l)\|_2^2, x_i \in B_l, y_i \in Y_l$
8)     $w \leftarrow w + \varepsilon \sum_{i=0}^{k-1} (V_w \lambda_v L_v + V_w \lambda_m \|L_m\|_2^2) + ps_{t-1}$
9)     $l \leftarrow l + 1$
10) until $l = s$;
11) until $t = T$;
12) $\hat{y}_t \leftarrow \hat{y}_t \geq thr, \hat{y}_t \in \hat{Y} = merge(P_1, ..., P_s)$;
13) **Return:** $\hat{y}_1, ..., \hat{y}_n$.

In the training preparation phase, we firstly divide CT scans into 3d block sequences $B_1, B_2, ..., B_s \in R^{k \times l \times w}$, and initialize the network parameters. In the specific training process, SCU-Net optimizes network parameters according to the new loss function and stochastic gradient descent algorithm (SGD). After the $T$ iteration process, the network parameters will converge to an ideal state, thus completing the network training.
2) The algorithm procedure of training the SCU-Net network

Algorithm 2: The BiLSTM Phase

**Input:** probability volumes serial $\hat{y}_1, ..., \hat{y}_n \in \mathbb{R}^{d \times t \times w}$; Threshold $thr$; model weights and bias $W, b$; CLSTM unit serial index $t \in [1, n]$; mask serial $y_1, ..., y_n \in \mathbb{R}^{d \times t \times w}$;

**Output:** prediction result $\tilde{y}_1, \tilde{y}_2, ..., \tilde{y}_n$

14 $t \leftarrow 1$, $\bar{y}_i \leftarrow \hat{y}_i \ast W_t$, $i \in [1, n]$;
15 **initialize BiLSTM weights and bias value $W, b$;**
16 **repeat**
17 $t \leftarrow t + 1$;
18 $(\bar{i}_t, \bar{o}_t) \leftarrow \sigma(W_y \ast \hat{y}_t + W_h \ast \tilde{h}_{t-1} + W_c \odot \tilde{c}_{t-1} + b_t)$;
19 $(\bar{f}_t, \bar{c}_t) \leftarrow \sigma(W_y \ast \hat{y}_t + W_h \ast \tilde{h}_{t+1} + W_c \odot \tilde{c}_{t-1} + b_t)$;
20 $\tilde{c}_t \leftarrow \bar{f}_t \odot \tilde{c}_{t-1} + \bar{i}_t \odot \tanh(\bar{y}_t, \tilde{h}_{t-1}; W_c, b_c)$;
21 $\bar{c}_t \leftarrow \bar{f}_t \odot \tilde{c}_{t-1} + \bar{i}_t \odot \tanh(\bar{y}_t, \tilde{h}_{t-1}; W_c, b_c)$;
22 $\bar{o}_t = \sigma(W_yo \ast \hat{y}_t + W_{ho} \ast \tilde{h}_{t-1} + W_{co} \odot \tilde{c}_t + b_o)$
23 $\bar{o}_t = \sigma(W_yo \ast \hat{y}_t + W_{ho} \ast \tilde{h}_{t-1} + W_{co} \odot \tilde{c}_t + b_o)$
24 $\tilde{h}_t \leftarrow \bar{o}_t \odot \tanh(\bar{c}_t)$; $\tilde{h}_t \leftarrow \bar{o}_t \odot \tanh(\bar{c}_t)$;
25 $\check{y}_t \leftarrow \sigma(\text{concat}(\tilde{h}_t, \tilde{h}_t) \ast g_t)$;
26 loss = DSC($\tilde{y}_t, y_t$) = $2(|\check{y}_t \cap y_t|)/(|\check{y}_t| + |y_t|)$;
27 $(W, b) = (W, b) - \nabla_{W,b} \text{DSC}(\check{y}_t, y_t);
28 **until** $t = n$;
29 $\tilde{y}_i \leftarrow \tilde{y}_i \geq thr$, $i \in [1, n]$;
30 **Return:** $\tilde{y}_1, ..., \tilde{y}_n$.

In the training preparation phase, we first use the convolution layer to extract features from the input of the probability volumes $\hat{y}_t$, and initializes the network parameters. After that, according to the LSTM formula, we calculate the value of input gate $\bar{i}_t$, forgetting gate $\bar{f}_t$, output gate $\bar{o}_t$ and current cell state $\bar{c}_t$, thereby obtain the output result $\tilde{h}_t$ of the unidirectional CLSTM network. Finally, the bidirectional LSTM output $\tilde{h}_t$, $\tilde{h}_t$ is merged into the prediction result $\tilde{y}_t$, the DICE loss function
is calculated, and the network weight is optimized according to the gradient descent method.

### III. Experimental Results

#### A. Datasets and Evaluation Criteria

We test our SCU-Net + BiCLSTM architecture on the NIH-CT dataset [6][15], which contains 82 abdominal enhanced 3D CT scans. In the direction of coronal plane, the CT slice size is 512*512 pixels, and the number of slices varies from 181 to 466 for different patient. We pre-estimate the approximate range of the pancreas according to the label, and reduce the 512*512 slice data to 192*256. Experimental result is validated by random 4-fold cross validation [11], that is, randomly shuffling patient index and splitting the dataset into 4 fixed fold. We used 3 out of 4 fold for training and the remaining one for testing. This random cross-validation process is repeated 20 times, each time generating a different patient sequence to verify the reliability of our network. Considering pancreas irregular shape, we augmented the training data set via rotation, horizontal flipping and vertical flipping. We use the dice similarity coefficient: \( DSC = \frac{2(|Y \cap \hat{Y}|)}{|Y| + |\hat{Y}|} \), Jaccard index: \( JI = \frac{|Y \cap \hat{Y}|}{|Y \cup \hat{Y}|} \), pixels-wise precision and recall to evaluate our segmentation results [13]. In addition, we use averaged Hausdorff distance (AVD) to evaluate the result of modeling inter-slice shape continuity and regularization [13].

#### B. Implementation Details

We implement the network framework on PyTorch[29] for the Python environment using NVIDIA GeForce GTX 1080Ti with 8GB memory. End-to-end training of the network as a whole is very challenging, so we train SCU-Net and BiCLSTM networks in two phases. At the beginning, we train the SCU-Net for 40 epochs to generate intermediate probability volumes. We set the layer number \( k \) of the input block to 7. After experimental verification, the block data with size of 7*192*256 can make SCU-Net get an optimal segmentation result, as shown in table 1. The learning rate of scunet is set to 1e-3 and the batch size is set to 1. Compared with ADAM, we found that SGD algorithm had higher training efficiency when applied to SCU-Net, so SGD algorithm was selected as the optimizer of SCU-Net, and the momentum value is set to 0.99. We apply the newly designed loss function to the training of SCU-Net, and set the values of \( \lambda_v \) and \( \lambda_m \) in Eq. 3 to 0.5. Afterwards, we train the BiCLSTM network with probability volumes which generated by SCU-Net for 120 epochs. Dice loss is taken as the loss function, and three consecutive slices of probability volumes are taken as the basic unit of BiCLSTM. Similar to SCUnet, we set the learning rate of BiCLSTM network as 1e-3, batch-size as 1, and the optimizer as SGD. The average time of model training is approximately ~6 hours for SCU-Net and ~4 hours for BiCLSTM on a single standard NVIDIA GeForce GTX 1080Ti(8GB).
Tabel I compares the performance of HNN, U-Net and SCU-Net with DSC, JI, Precision and Recall evaluation index. As shown in Table I, compared with HNN and U-net, the segmentation results of SCU-Net network have different degrees of improvement on these four indicators. In addition, we also evaluated the impact of different block numbers on the segmentation results. It was found that when slices number k=7, the segmentation results were the best, and the Dice average was 82.8%. Notably, the minimum value of DSC in the SCU-Net with 7 slices is 60.6%, which is significantly higher than the HNN value of 41.9% and the U-Net value of 43.3%. Moreover, the standard deviation of SCU-Net segmentation accuracy also decreases a lot, which proves that the SCU-Net network has high robustness and stability.

### IV. Conclusion

In this paper, we proposed a network structure of SCU-Net +BiCLSTM for pancreatic segmentation, and integrated spatial context information on the basis of 2D segmentation network, thus improving the segmentation accuracy and spatial continuity. The main contributions include three aspects: 1) Divide and conquer strategy is adopted to divide the 3D overall data into equal length block data, and a new network structure is designed to learn the local spatial characteristics of the pancreas, which is called SCU-Net; 2) A new loss function is proposed, which constrains the segmentation results of the whole local partial block and the single layer, and enables SCU-Net to improve the segmentation accuracy of the whole block on the basis of satisfying the precise segmentation of each layer; 3) Bidirectional long and short term memory (BiCLSTM) network is introduced to stimulate the interaction between information flows from bidirectional sequences, so as to make up the boundary defect and fault problem of segmentation results.

We validated the effectiveness of the framework in the pancreatic segmentation problem on the NIH public dataset. The final prediction accuracy was 82.8%, which achieved a good segmentation effect.

### Reference

[1] L. Sun, Z. Fan, Y. Huang, X. Ding, and J. Paisley, “Joint CS-MRI Reconstruction and Segmentation with a Unified Deep Network,” no. 1, pp. 1–8, 2018.
Pancreas Segmentation via Spatial Context based U-net and Bidirectional CLSTM

[2] C. Chu et al., “Multi-organ segmentation based on spatially-divided probabilistic atlas from 3D abdominal CT images,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 8150 LNCS, no. PART 2, pp. 165–172, 2013.

[3] R. Wolz, C. Chu, K. Misawa, M. Fujiiwara, 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.

[4] T. Okadaa, M. G. Linguraru, M. Horid, R. M. Summerse, N. Tomiyamad, and Y. Satof, “Abdominal multi-organ segmentation from CT images using conditional shape–location and unsupervised intensity priors Toshiyuki,” Med Image Anal. 2015 December, p. 26(1): 1–18. doi:10.1016/j.media.2015.06.009. Abdo, 2015.

[5] A. Farag, L. Lu, E. Turkbey, J. Liu, and R. M. Summers, “A bottom-up approach for automatic pancreas segmentation in abdominal CT scans,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 8676, pp. 103–113, 2014.

[6] 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.

[7] D. Shen, G. Wu, and H.-I. Suk, “Deep Learning in Medical Image Analysis,” Annu. Rev. Biomed. Eng., vol. 19, no. 1, pp. 221–248, 2017.

[8] J. Long*, E. Shelhamer*, and Trevor Darrell, “Fully Convolutional Networks for Semantic Segmentation,” CVPR, 2015.

[9] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 9351, pp. 234–241, 2015.

[10] Y. Zhou, L. Xie, W. Shen, Y. Wang, E. K. Fishman, and A. L. Yuille, “Pancreas Segmentation in Abdominal CT Scan: A Coarse-to-Fine Approach,” in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2016, vol. 10433 LNCS, pp. 693–701.

[11] P. F. Christ et al., “Automatic Liver and Lesion Segmentation in CT Using Cascaded Fully Convolutional Neural Networks and 3D Conditional Random Fields,” pp. 1–8.

[12] J. Cai, L. Lu, Z. Zhang, F. Xing, L. Yang, and Q. Yin, “Pancreas Segmentation in MRI using Graph-Based Decision Fusion on Convolutional Neural Networks,” pp. 1–13, 2017.

[13] 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.

[14] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, “UNet++: A Nested U-Net Architecture for Medical Image Segmentation,” 2018.

[15] H. R. Roth, L. Lu, A. Farag, A. Sohn, and R. M. Summers, “Spatial aggregation of holistically-nested networks for automated pancreas segmentation,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 9901 LNCS, pp. 451–459, 2016.

[16] X. Yang et al., “Towards Automated Semantic Segmentation in Prenatal Volumetric Ultrasound,” IEEE Trans. Med. Imaging, vol. PP, no. c, p. 1, 2018.

[17] H. Roth et al., “Towards dense volumetric pancreas segmentation in CT using 3D fully convolutional networks,” pp. 1–6.
Pancreas Segmentation via Spatial Context based U-net and Bidirectional CLSTM

[18] H. R. Roth et al., “An application of cascaded 3D fully convolutional networks for medical image segmentation,” Comput. Med. Imaging Graph., vol. 66, pp. 90–99, 2018.

[19] 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.

[20] 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.

[21] 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.

[22] J. Chen, L. Yang, Y. Zhang, M. Alber, and D. Z. Chen, “Combining Fully Convolutional and Recurrent Neural Networks for 3D Biomedical Image Segmentation,” no. Ii, 2016.

[23] 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 8th US National Combustion Meeting 2013, 2017, vol. 1, pp. 294–322.

[24] K. Greff, R. K. Srivastava, J. Koutnik, B. R. Steunebrink, and J. Schmidhuber, “LSTM: A Search Space Odyssey,” IEEE Trans. Neural Networks Learn. Syst., vol. 28, no. 10, pp. 2222–2232, 2017.

[25] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, “Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling,” IEEE Int. Conf. Rehabil. Robot., vol. 2015–Septe, pp. 119–124, 2015.

[26] M. Schuster and K. K. Paliwal, “Bidirectional recurrent neural networks,” IEEE Trans. Signal Process., vol. 45, no. 11, pp. 2673–2681, 1997.

[27] T. Chen, R. Xu, Y. He, and X. Wang, “Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN,” Expert Syst. Appl., vol. 72, pp. 221–230, 2017.

[28] H. Sak, A. Senior, and Franc¸oise Beaufays, “Long Short-Term Memory Recurrent Neural Network Architectures for Large Scale Acoustic Modeling,” INTERSPEECH, 2014.

[29] A. Paszke et al., “Automatic differentiation in PyTorch,” no. Nips, pp. 1–4, 2017.