IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 7, 2019, accepted December 14, 2019, date of publication December 20, 2019, date of current version January 6, 2020.

Digital Object Identifier 10.1109/ACCESS.2019.2961125

Automatic Pancreas Segmentation via Coarse Location and Ensemble Learning

SHANGQING LIU[1,2], XINRUI YUAN[1,2], RUNYUE HU[1,2], SHUJUN LIANG[1,2], SHAOHUA FENG[3], YUHUA AI[1,4], AND YU ZHANG[1,2]

[1]School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China
[2]Guangdong Provincial Key Laboratory of Medical Image Processing, Southern Medical University, Guangzhou 510515, China
[3]Radiology Department, Manzhouli People's Hospital, Manzhouli 021400, China
[4]Department of Information Center, Nanfang Hospital, Southern Medical University, Guangzhou 510515, China

Corresponding authors: Yuhua Ai (aiyuhua@sina.com) and Yu Zhang (yuzhang@smu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61671230 and Grant 61971213, and in part by the Basic and Applied Basic Research Foundation of Guangdong Province under Grant 2019A151010417.

ABSTRACT Automatic and reliable segmentation of the pancreas is an important but difficult task for various clinical applications, such as pancreatic cancer radiotherapy and computer-aided diagnosis (CAD). The main challenges for accurate CT pancreas segmentation lie in two aspects: (1) large shape variation across different patients, and (2) low contrast and blurring around the pancreas boundary. In this paper, we propose a two-stage, ensemble-based fully convolutional neural network (FCN) to solve the challenging pancreas segmentation problem in CT images. First, candidate region generation is performed by classifying patches generated by superpixels. Second, five FCNs based on the U-Net architecture are trained with different objective functions. For each network, 2.5D slices are used as the input to provide 3D image information complementarily without the need for computationally expensive 3D convolutions. Then, an ensemble model is utilized to combine the five output segmentation maps and achieve the final segmentation. The proposed method is extensively evaluated on a publicly available dataset of 82 manually segmented CT volumes via 4-fold cross-validation. Experimental results show its superior performance compared with several state-of-the-art methods with a Dice coefficient of 84.10±4.91% and Jaccard coefficient of 72.86±6.89%.

INDEX TERMS Superpixel, ResNet, fully convolutional neural networks, ensemble learning, pancreas segmentation.

I. INTRODUCTION

The pancreas, as an important organ of the human body, has internal and external secretion functions and is susceptible to various diseases. Pancreatic cancer, which is one of the most prevalent cancers in the world, is a devastating malignant disease with a median survival of 3–6 months and a 5-year survival rate of less than 5% [1]. Contrast-enhanced CT is now the worldwide imaging modality of choice for pancreatic disease evaluation and may be the best modality of the resectability of pancreatic cancer [1]. The segmentation of pancreas in CT images can support clinical workflows, including pancreas cancer diagnosis, treatment planning, and surgical assistance, in multiple domains [2]. Therefore, a robust, accurate, and automatic segmentation method for the pancreas is worth exploring.

However, automatic pancreas segmentation remains a challenge due to the following reasons: 1) low soft tissue contrast in CT images. As shown in Fig. 1 (a) and (b), the contrast among the pancreas and surrounding organs, such as liver, stomach, and spleen, is remarkably low, and the intensities of voxels in the pancreas and surrounding organs are in similar ranges, making it difficult to distinguish the organ boundaries. Moreover, the similarity in appearance and texture patterns of the pancreas and neighboring tissues makes their identification difficult. 2) Large anatomical variations. The pancreas exhibits high anatomical variability in terms of size and its location in the abdominal cavity of patients [3]–[6]. Furthermore, the pancreas is a deformable soft tissue. Thus, the shape and appearance of the pancreas have large variations across different individuals, as shown in Fig. 1.

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

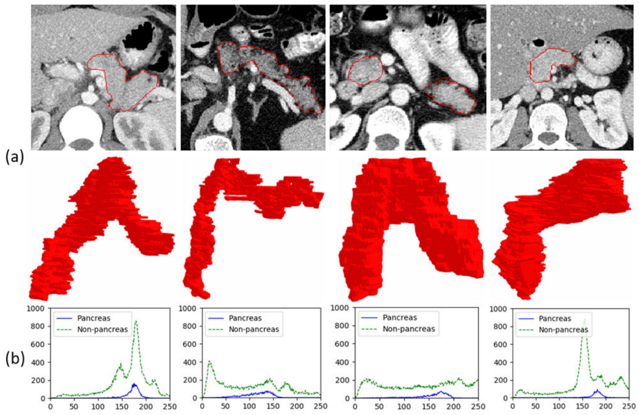

FIGURE 1. (a) Examples of variations in appearance, shape, size and location of the pancreas as seen in contrast-enhanced CT. Manual ground truth annotations are shown in red. (b) Intensity distributions of pancreas and the surrounding organs like liver, stomach and spleen.

Recently, learning-based methods have achieved satisfactory performance in pancreas segmentation [6]–[8]. Erdt *et al.* [6] built a pancreas tissue classifier incorporating spatial relationships among the pancreas, surrounding organs, vessels, and meaningful texture features. Classification was then used to guide a constrained statistical shape model to fit the data. Cross-validation on 40 datasets obtained a Dice coefficient of 61.2%. Farag *et al.* [7] proposed a fully automatic bottom-up method, which differentiates the pancreas from other elements by classifying superpixels with a two-level cascade of supervised random forests. The evaluation of the proposed approach was conducted on CT volumes of 80 patients from a publicly available dataset in 6-fold cross-validation with a Dice coefficient of 70.7% and Jaccard coefficient (JC) of 57.9%.

One main limitation of the aforementioned learning-based methods is the need to predefine the features for a specific learning model. In this case, the distinctiveness of features may considerably influence the learning accuracy. To address this limitation, deep learning methods have been widely used, in which features can be learned automatically and effectively through convolutional neural networks (CNNs [9]) [10]-[16]. To segment the pancreas, Roth *et al.* [13] proposed a coarse-to-fine classifier on image patches and regions via CNN and achieved a Dice coefficient of 71.8 ± 10.5% in testing. The authors further improved the segmentation (Dice coefficient of 81.27 ± 6.27%) by introducing pancreas localization and 3D information [15]. Zhou *et al.* [16] designed a fixed-point model with a predicted segmentation mask to shrink the input region and update the results iteratively. 4-fold cross-validation was performed on the datasets of 82 patients with CT scans and obtained a Dice coefficient of 82.37%. Various deep learning-based methods has been proposed to overcome the challenges, but the variable shape, size, and location in the abdomen of the pancreas still limit the segmentation accuracy of these deep learning methods.

In this paper, to segment the pancreas accurately in CT images, we propose an ensemble-based multiloss fully convolutional neural network (FCN). Fig. 2 illustrates the flowchart of the proposed method. Given that the CT image covers a large region of the human body but the target (pancreas) is relatively small, a two-stage framework is further designed to segment the pancreas robustly from the coarse level to the fine level. The first stage presents the dense labeling of local image patches generated by superpixels via residual neural network (ResNet). This labeling is designed for the pancreas region detection of the entire CT image. Section II-A provides the details of this stage. The second stage involves multiloss FCNs, which are designed for accurate pancreas segmentation on the basis of the detected regions from the first stage. In this stage, multiloss FCNs initially learn the probability maps by separately using five different loss functions. Section II-B elaborates the details of this stage. Finally, an averaging-based ensemble algorithm is proposed to integrate these probability maps from the second stage to produce a final segmentation result. Section II-C describes the details of this part.

The main contributions of this study are summarized as follows:

1) We present a strategy for generating candidate regions by classifying patches generated by superpixels. This approach not only obtains increased location accuracy but also remarkably improves the segmentation speed.

2) We ensemble multiple same-architecture networks with different loss functions to help enhance the accuracy and robustness of the conventional deep network for this challenging segmentation task. To the best of our knowledge, this is the first work to use different loss functions for pancreas segmentation.

3) We demonstrate that our proposed method outperforms state-of-the-art methods on the public dataset for pancreas segmentation.

The remainder of this paper is organized as follows. The technical motivation and details of our proposed approach are described in Section II. The experiments and results are presented in Section III and IV. The discussion and conclusions are drawn in Section V.

II. METHOD

A. PANCREAS LOCATION

The first stage aims to detect the target organ location in CT image effectively and then provide the region proposal of this organ (resulting in a bounding box) to the second stage for segmentation. This approach helps reduce the input candidate space, which not only efficiently helps our architecture to learn the differences between the pancreas and surrounding anatomy but also reduces the occupied graphics memory and improves the segmentation speed.

Regression is a widely used method for object detection and localization [17]-[19]. However, a local classification-based approach is more robust than an approach using global

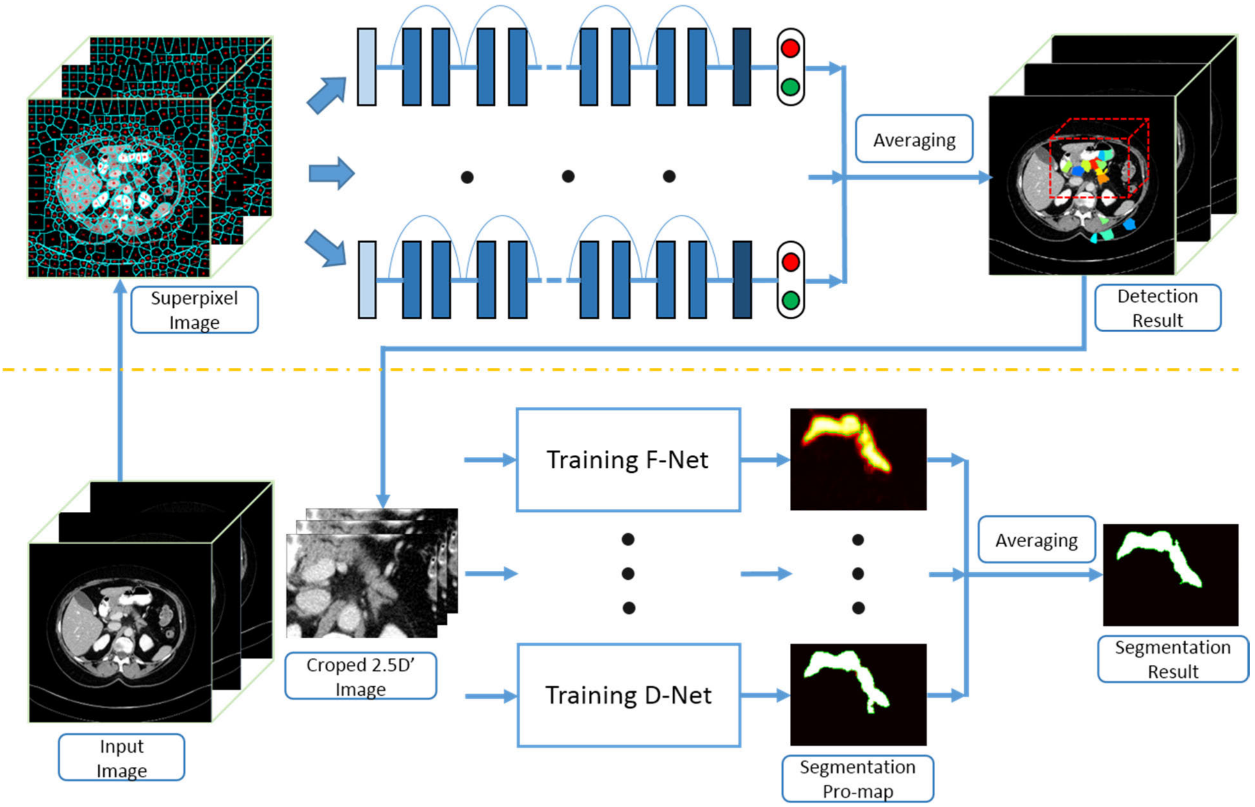

FIGURE 2. The flowchart of the proposed method. First, candidate region generation via dense labeling of local image patches generated by superpixels via residual neural network. Second, accurate pancreas segmentation with multiloss FCNs. Finally, averaging-based ensemble to get the final result.

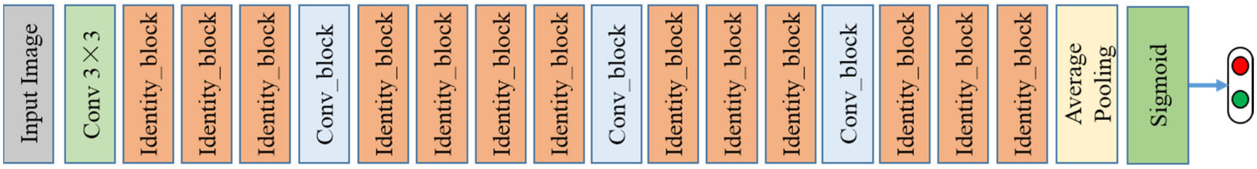

FIGURE 3. The detailed architecture of classification network.

context due to the high anatomical variability of the pancreas inside the abdomen [20]. In this work, we use superpixel [21]-based over-segmentation to divide the images into small perceptually meaningful regions. The superpixel-based approach has the following advantages: (1) the superpixel being an irregular patch, the pixels in which demonstrate similarities in color, texture, and intensity, (2) computationally fast, and (3) ease of use.

We only need to determin which superpixels belong to the pancreas and then specify the bounding box on the basis of the labeled superpixels. We use ResNet to classify the central pixels of superpixel patches. If the central pixel is labeled as the pancreas, then we classify the entire superpixel patch as the pancreas. Our strategy for classifying central pixels is presented as follows: We extract different scale image patches centered on the superpixel centers, classify them into pancreas and nonpancreas via ResNet, and ensemble the classification results of different scale image patches to obtain the final label of the central pixel. In our experiment, we select three scales (64, 48, and 32) of the patch size for training.

ResNet [22] permanently utilizes shortcut connections between shallow and deep layers to control and adjust the training error rate. We select a simple architecture to overcome the limitations in training time and the size of input patches. Fig. 3 shows that a total of 13 identity and 3 convolution blocks exist. Each identity block has two convolutional layers with a 3 × 3 convolution kernel size, which increases the width of residual networks. Every convolution block has three convolutional layers with one 3 × 3 convolution surrounded by dimensionality-reducing and dimensionality-expanding 1 × 1 convolution layers. Batch normalization (BN) [23] and rectified linear units (ReLUs) are adopted after every convolutional layer. The former ensures that the network activation follows a unit Gaussian distribution after each update to prevent internal covariate migration and overfitting.

The localization of superpixel-labeled pancreas will provide us with the initial bounding box region of the pancreas while removing a large amount of background in the image. To ensure 100% recall, the bounding box is then enlarged by adding additional pixels of the margin. Fig. 4 presents the examples of superpixel segmentation, patch labeling, and bounding box generation.

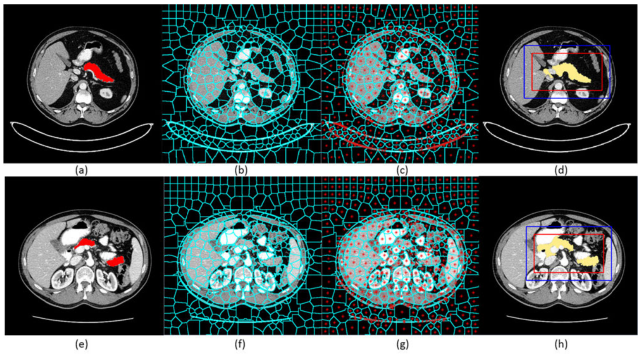

FIGURE 4. Examples of superpixel segmentation, patch labelling and bounding box generation. (a)(e) depicts two slices with pancreas segmentation map in red. (b)(f) depicts over-segmentation results with the contours of each superpixel region superimposed on the image. Red stars in (c)(g) show the centers of superpixels which are used to extract patches next.(d)(h) show the results of patch labelling and bounding box generation. Light yellow regions are the patches belong to pancreas. Red rectangular is the generated bounding box from labeled result and blue rectangular is the bounding box after enlarged.

B. FINE-SEGMENTATION BY MULTI-LOSS FCNs

On the basis of the bounding box generated in the first stage, we ensemble multiple same-architecture deep networks with different loss functions to achieve accurate and robust segmentation of pancreas. Moreover, to enhance the segmentation without remarkably increasing the computational burden, we use 2.5D context input.

1) NETWORK ARCHITECTURE

In Fig. 5, we provide the schematic of our variant U-Net [24] model, which consists of an encoder path that captures global features and a decoder path that enables precise segmentation. MobileNet [25] architecture based on depthwise separable convolution is applied in the encoder path. This architecture is an extreme case of the inception module, wherein a separate spatial convolution is applied for each channel and denoted as depthwise convolutions. Then, a 1 × 1 convolution with all the channels is used to merge the output, which is denoted as the pointwise convolutions. On the one hand, separation in depthwise and pointwise convolutions is used to improve computational efficiency. On the other hand, it improves accuracy as the cross channel and spatial correlation mapping are learned separately.

The following decoder path includes operations arranged in five increasing resolution level. The blue blocks function as the convolution units with the following layers: (1) a convolution layer with a learned kernel, (2) batch normalization to accelerate robust gradient propagation and reduce overfitting, and (3) a nonlinear ReLU represents the nonlinear functions. Mathematically, the convolutional units are denoted as

$$c(X, W, \beta) = r(b((X * W), \beta)), \tag{1}$$

where W is a convolutional kernel, batch normalization $b(X, \beta)$ transforms the mean of each channel to 0 and the variance to a learned per-channel scale parameter β, and the ReLU $r(X) = \max(0, X)$ induces non-linearity. The yellow blocks perform 2 ×2 upsampling. At each resolution level, a skip connection is included to fuse the upsampled feature maps with the same-level feature maps obtained from the previous encoder path and complementarily combine global contextual information with spatial details for the precise detection and location of pancreas. The final green block performs 1 × 1 convolution and sigmoid activation to calculate the pixel-wise pancreas probability map from highly dimensional feature maps. Similar to the U-Net, the concatenation operation is adopted as a combination strategy to fuse of feature maps with the same resolution in both paths.

2) 2.5D CONTEXT INPUT

Specifically, the pancreas region proposals extracted from consecutive slices are the inputs of the network instead of the 3D inputs, and the output is the corresponding label region proposal of the middle slice. In this way, the adequate neighborhood information of the 3D image can be efficiently leveraged to train the segmentation model while avoiding the heavy burden of 3D computing.

3) MULTI-LOSS

We use five different loss functions to train our network. Dice loss [26] is a common used semantic segmentation loss function on the basis of Dice coefficient, and the network trained with dice loss function as denoted as D-Net. n represents the number of pixels in the considered image or mini-batch. Let $Y = \{y_1, y_2, \ldots, y_n\}$ be the ground-truth segmentation probabilistic maps over n pixels and $\hat{Y} = \{\overset{\wedge}{y_1}, \overset{\wedge}{y_2}, \ldots, \overset{\wedge}{y_n}\}$ be the predicted probabilistic maps over n pixels. In this study, s was set to 1 and used to ensure the loss function stability by avoiding the division by 0. The dice loss function can be expressed as

$$loss_{(dice)} = -\frac{2\sum_{n=1}^{N} y_i \cdot \overset{\wedge}{y_i} + s}{\sum_{n=1}^{N} y_i + \sum_{n=1}^{N} \overset{\wedge}{y_i} + s}. \tag{2}$$

Focal loss [27] introduces a tunable focusing parameter γ to reshape the loss function into down-weight easy examples and focus training on hard negatives; a balancing parameter α is used to balance the importance of positive/negative examples. The network trained with focal loss function is denoted

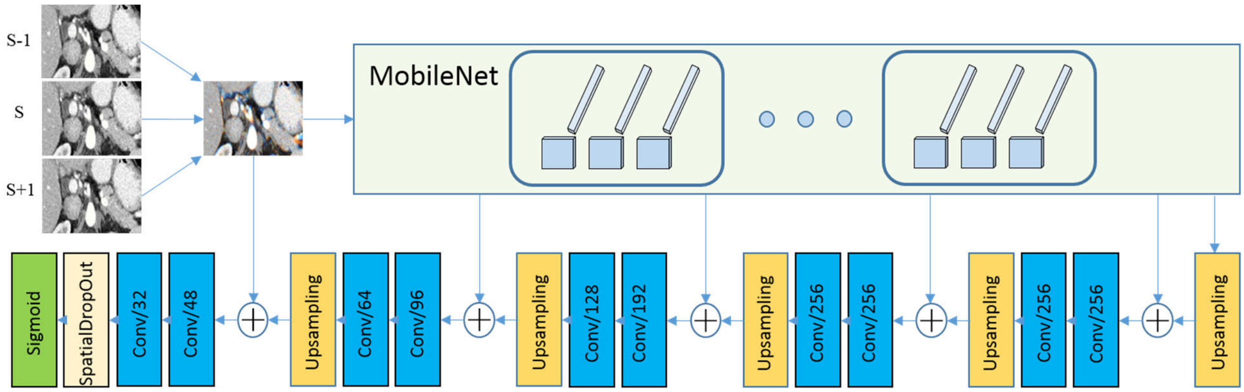

FIGURE 5. **Architecture of segmentation network based on U-Net.**

as F-Net. In our experiments, we set α and γ to 0.6 and 2, respectively. The focal loss function can be expressed as:

$$L_{(each-element)} = \begin{cases} -\alpha(1-\hat{y_i})^{\gamma}\log(\hat{y_i}) & y_i = 1 \\ -(1-\alpha)\hat{y_i}^{\gamma}\log(1-\hat{y_i}) & y_i = 0, \end{cases} \quad (3)$$

$$loss_{(focal)} = \sum_{i=1}^{n} L_{(each-element)}. \quad (4)$$

Jaccard distance loss is derived from common evaluation measures for the semantic segmentation of the Jaccard Coefficient (JC) [28], [29], which measures the intersection over the union of the labeled segments for each class and obtains the average. The network trained with Jaccard distance loss function is denoted as J-Net. The Jaccard distance loss function can be expressed as:

$$loss_{(Jaccard)} = -\frac{\sum_{n=1}^{N} y_i \cdot \hat{y_i} + s}{\sum_{n=1}^{N} y_i + \sum_{n=1}^{N} \hat{y_i} - \sum_{n=1}^{N} y_i \cdot \hat{y_i} + s}. \quad (5)$$

We use a simpler strategy as the complementary policy, wherein a class-balancing weight is introduced on a per-pixel term basis to offset the imbalance between positive/negative examples. The network trained with class-balanced cross-entropy loss function is denoted as C-Net and is expressed as follows.

$$loss = -\sum_{i=1}^{n} [\beta \cdot y_i \log(\hat{y_i}) + (1-\beta) \cdot (1-y) \log(1-\hat{y_i})]. \quad (6)$$

where $\beta = n_-/n$ and $1-\beta = n_+/n$. n_- and n_+ denote the number of negative and positive pixels, respectively.

A commonly used and fundamental loss function binary cross-entropy loss [11] is also adopted and expressed as follows:

$$loss_{(BCE)} = -\sum_{i=1}^{n} [y_i \log(\hat{y_i}) + (1-y_i) \log(1-\hat{y_i})]. \quad (7)$$

The network trained with this loss is denoted as B-Net.

C. AVERAGING-BASED ENSEMBLE FCNS

Ensemble techniques are helpful in reducing overfitting problems in the training data of complex models [30], [31]. This type of approach combines multiple learning models to obtain better predictive performance than any of the constituent learning algorithm when used alone. The use of ensemble models includes two aspects: 1) networks trained with different loss functions can learn different attributes of the training data during batch learning; thus, their ensemble can boost the segmentation results. 2) Bias–variance tradeoff. Bias and variance are critical for determining the behavior of prediction models and understanding the occurrence of overfitting and underfitting. This study aims to lower the model variance by averaging the model output. An FCN with millions of parameters and overtrained on different bootstrapped/subsampled training sets can qualify for unbiased and highly variant models.

As shown in Fig. 2, five FCN models with different loss functions are trained with random parameter initialization and shuffle data in the batch learning process. This training creates sufficient diversity in the trained models to allow the averaged predictions of the ensemble to outperform the individual models significantly. We attempt to add diversity to the models by varying the sets of data that each model sees, but our results do not change significantly. Each FCN model generates a probability segmentation map when a test image is given. Then, sample averaging will be applied to transform five score maps into a binary segmentation map.

III. EXPERIMENTS

A. DATASET

The National Institutes of Health Clinical Center performed 82 abdominal contrast-enhanced 3D CT scans with a resolution of 512 × 512 pixels, varying pixel sizes, and slice thickness between 1.5 mm and 2.5 mm. Such scans were acquired on Philips and Siemens MDCT scanners. Our experiments were conducted on 4-fold cross-validation, and the dataset was randomly grouped into four groups of 20, 20, 21, and 21. The division was performed at the patient level. Hence, all scans of a given patient were either in the training or test set. Image intensities were rescaled within [0, . . ., 255] using a soft-tissue window of [−110,190] HU to increase the contrast

TABLE 1. The average (±standard deviation) performance by different number of input channels.

Input channel	Dice (%)	JC (%)	Precision (%)	Recall (%)
2D	82.14±5.82	70.08±7.89	82.60±6.94	82.58±9.28
2.5D(3 slices)	**84.10±4.91**	**72.86±6.89**	**83.60±5.85**	**85.33±8.24**
2.5D(5 slices)	83.44±6.26	71.58±8.41	81.91±7.74	83.93±9.44
2.5D(7 slices)	83.48±6.27	71.64±8.51	83.14±6.80	82.76±9.93

TABLE 2. The average (±standard deviation) performance by different combination strategies.

Ensemble strategy	Dice (%)	JC (%)	Precision (%)	Recall (%)
Sample averaging	**84.10±4.91**	**72.86±6.89**	83.60±5.85	**85.33±8.24**
Majority voting	84.06±5.07	72.85±7.11	**84.35±5.57**	84.59±8.70
MLR	83.12±6.64	71.81±8.92	83.83±6.62	84.43±8.09

of soft tissues and show additional details of the abdominal organs.

B. EVALUATION METRICS

To measure segmentation performance, we used four different evaluation metrics, namely, the Dice coefficient, JC, recall, and precision. The Dice coefficient interprets the overlap between sample sets, and the JC computes the similarities between the segmentation result and the reference standard. These metrics are defined as follows:

(a) Dice coefficient: $2(|A \cap B|) / (|A| + |B|)$.

(b) JC: $(|A \cap B|) / (|A \cup B|)$, where A and B refer to the algorithm output and manual ground-truth pancreas segmentation, respectively.

(c) Recall: TP / (TP+FN).

(d) Precision: TP / (TP+FP), where TP and FP indicate the numbers of true and false positives, respectively, and FN denotes the number of false negatives.

C. IMPLEMENTATION DETAILS

Our method was implemented with Python based on the Keras package[32] with the TensorFlow library[33] as the backend. During training, the probability of each pixel belonging to the pancreas was computed with a sigmoid classifier. The weights of the network were optimized via RMSprop optimizer with a mini-batch size of 16. The learning rate was set to 0.001 with a momentum coefficient of 0.9 and reduced by a factor of 0.2 after five consecutive epochs without improving the validation loss. During segmentation, training takes approximately 5 hours on a single NVIDIA Tian X and the testing time per patient is around 1.99 seconds.

IV. SEGMENTATION RESULTS

A. EVALUATION ON 2.5D CONTEXT INFORMATION

In our method, we use the sequential slices as the input to predict the segmentation map of the middle slice. Table 1 lists the influence of the 2.5D context information in Stage 2. As shown in the table, the networks with the 2.5D context information obtain better results compared with 2D input in terms of Dice coefficient, JC and precision. These findings prove that the use of neighbor slices information improves the accuracy of distinguishing the difference between pancreas and nonpancreas tissues. This differentiation is important for segmenting the pancreas. Meanwhile, Table 1 shows that additional slices are used as the input, but the results are not sensitive to the number of input slices when it reaches three. Moreover, using additional input slices makes the network computationally expensive. Thus, we finally select three as the number of 2.5 D input slices. This solution may be important for deep learning methods applied to medical images without adequate training samples.

B. EVALUATION ON ENSEMBLE METHOD

We further evaluate our ensemble model to show its effectiveness. Three of the most common combination strategies in ensemble learning are compared. The first is sample averaging. The second and third strategies are majority voting and multiresponse linear regression (MLR) [30], respectively. MLR is a basic method of stacking which takes the output class probability of the basic classifier as the input attributes. From Table 2, we have following observations: First, sample averaging gets the best result compared with majority voting and MLR in terms of the three evaluation criteria (i.e., Dice coefficient, JC and recall). Second, MLR obtains higher standard deviations than sample averaging and majority voting.

As a main contribution of this study, the proposed ensemble method adopts five loss functions to obtain five different networks and combines the advantage of these networks. To demonstrate its effectiveness, Fig. 6 shows the four evaluation metrics of the different networks and ensemble model. As the sample averaging of the five networks, the ensemble model effectively improves the overall segmentation performance and further refines the average Dice coefficient (from

TABLE 3. Performance of different methods. k means the folds of the cross validation. MALF means the "Multi-Atlas and Label Fusion".

Method	Test data/k	Dice (%)	JC (%)	Precision (%)	Recall (%)
MALF[7]	80/6-fold	52.5±20.8	38.1±18.3	-	-
Wolz et al. [4]	100/-	65.5	49.6	70.7	62.9
Wolz et al.[5]	150/-	69.6±16.7	55.5±17.1	67.9±18.2	74.1±17.1
Farag et al.[7]	80/6-fold	70.7±13.0	57.9±13.6	71.6±10.5	74.4±15.1
Holger et al.[13]	82/4-fold	71.8±10.7	-	-	-
Holger et al.[14]	80/4-fold	78.01±8.2	-	-	-
Holger et al.[15]	82/4-fold	81.27±6.27	68.87±8.12	-	-
Zhou et al.[16]	82/4-fold	82.37±5.68	-	-	-
Proposed	80/4-fold	**84.10±4.91**	**72.86±6.89**	83.60±5.85	**85.33±8.24**

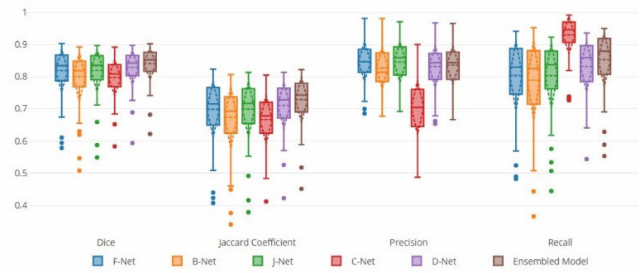

FIGURE 6. Performance of different networks and ensemble model in terms of Dice, Jaccard Coefficient, Precision and Recall. Dotted line in the box is the mean and standard deviation, real line is means the median values.

81.92±6.53% generated by F-Net, 80.05±7.46% generated by B-Net, 81.94±6.47% generated by J-Net, 79.73±5.53% generated by C-Net and 82.86±5.14% generated by D-Net to 84.10±4.91% generated by ensemble modal). Specifically, the proposed model makes an adjustment in JC (from 69.87±8.73%, 67.33±9.55%, 69.88±8.61%, 66.63±7.35%, 71.04±7.08% to 72.86±6.89%). Fig. 6 implies that the combination of several deep learning techniques into one predictive model does improve the segmentation performance and decrease variance. A possible reason could be that the models are independent of one another, whereas individual models have high variance. We have performed McNemar's test to compare the segmentation results achieved by five networks and ensemble model. Our test obtains small (<0.001) *p*-values, thereby suggesting that the performance difference among the six models is significant.

C. VISUAL RESULTS

As a qualitative illustration, six automatic segmentations are produced by F-Net, J-Net, B-Net, C-Net, D-Net, and the ensemble model and compared in Fig. 7. Each column in the figure demonstrates the results for a specific subject. We can observe that the ensemble model obtains more accurate segmentation than the five other networks. All these results demonstrate the effectiveness of our proposed method.

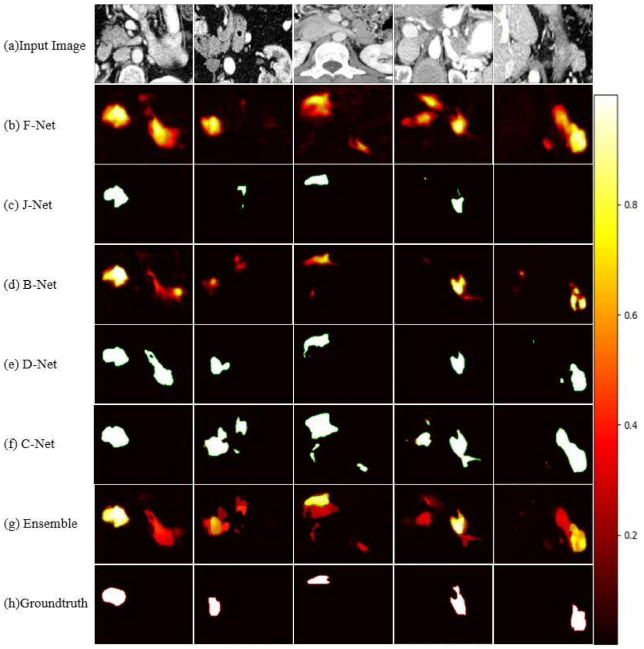

FIGURE 7. Typical pancreas segmentation results of the same image: (a) Original input image to the network, (b) the output of F-Net, (c) the output of B-Net, (d) the output of J-Net, (e) the output of C-Net, (f) the output of D-Net and (g) the output of Ensemble model. (h) The manual ground-truth segmentations. Green line is the boundary where the output probability value is processed through the 0.5 threshold.

D. COMPARISON WITH STATE-OF-THE-ART METHODS

Table 3 compares the performance of our method with certain state-of-the-art methods. Based on 80 and 82 CT datasets, our results are comparable and substantially better than those of recent studies [4]–[8], [12]–[16]. For example, the Dice coefficient of 84.10±4.91% is obtained (4-fold CV), versus 70.7±13.0% in [7] (6-fold CV), 71.8±10.7%

in [13], 78.01±8.2% in [14] and 81.27±6.27% in [15]. Table 3 demonstrates the superiority of deep learning-based methods over the methods based on atlas and the powerful capabilities of deep CNN in feature learning and classification.

V. DISCUSSION AND CONCLUSION

In this paper, we present a novel segmentation strategy for CT pancreas images. The key points of this novel approach includes three folds: (1) coarse pancreas location via superpixel over-segmentation and classification; (2) an ensemble model combined with five FCNs, which were trained with different objective functions; (3) 2.5D image input. The proposed method is extensively evaluated on the dataset containing 82 pancreas CT images. In comparison with several state-of-the-art CT pancreas segmentation methods, our method demonstrates superior performance in segmentation accuracy.

The main component of our method is the combination of five basic CNNs into one predictive model. This approach can address the challenges of increasing the robustness of the feature representation on the large appearance variations of the pancreas. During our experiments, we prove that the ensemble model notably improves the accuracy of segmentation and obtains low variance. Empirically, ensembles tend to yield improved results when a significant diversity exists among the submodels[34]. We calculate the measure of the pairwise diversity matrix called the double-fault measure for the five networks and obtain the averaged value of 0.965, which indicates some diversity. Moreover, we train and test the networks in 2.5D slice-by-slice. This strategy can not only effectively utilize context information, but also reduce the heavy burden of using 3D input.

Coarse location of abdominal organs plays a meaningful role in the automatic segmentation of abdominal organs. For example, Dice coefficient of 56% was obtained with the original images in this experiment directly segmented via 2D U-Net. It can not only improve the accuracy and robustness of segmentation but also reduce the computational cost and time. Currently, candidate region generation and coarse segmentation is a useful pre-segmentation step for abdomen organs segmentation [26], [35]. However, the common methods are volumetric image pre-segmentation based on deep convolutional neural networks, which are high computational and memory cost. Meanwhile, the downsampling and upsampling in the pre-segmentation process will lose lots of image details, especially in axial direction. In our method, we use superpixel-based over-segmentation and classification to generate the candidate regions. The experimental results demonstrate the effectiveness of this simple approach.

The proposed method is limited by its two separate stages for detection and segmentation, which may lead to slow training and testing. In the future, research will focus on transforming the method to an end-to-end learning framework. The investigation of additional diverse classifiers, such as different architecture networks and ensemble strategies that take the intensity images and dissimilarity between basic classifiers' outputs into considerations will be our work next too. At the same time, some failed segmentations still occur because of the inconstant position. Hence, the spatial relationships of splenic, portal and superior mesenteric veins with pancreas will be considered in the future work.

REFERENCES

[1] Q. Lin, Y. Feng, J. Chen, and D.-L. Fu, "Current status and progress of pancreatic cancer in China," *World J. Gastroenterol.*, vol. 26, pp. 7988–8003, 2015, doi: 10.3748/wjg.v21.i26.7988.

[2] T. Okada, M. G. Linguraru, Y. Yoshida, M. Hori, R. M. Summers, and Y.-W. Chen, "Abdominal multi-organ segmentation of CT images based on hierarchical spatial modeling of organ interrelations," in *Proc. Int. MICCAI Workshop Comput. Clin. Challenges Abdominal Imag.*, New York, NY, USA, Springer-Verlag, 2011, pp. 173–180.

[3] A. Shimizu, T. Kimoto, H. Kobatake, S. Nawano, and K. Shinozaki, "Automated pancreas segmentation from three-dimensional contrast-enhanced computed tomography," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 5, no. 1, pp. 85–98, 2010.

[4] R. Wolz, C. Chu, K. Misawa, K. Mori, and D. Rueckert, "Multi-organ abdominal CT segmentation using hierarchically weighted subject-specific atlases," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Berlin, Germany, Springer, 2012, pp. 10–17.

[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. Imag.*, vol. 32, no. 9, pp. 1723–1730, Sep. 2013.

[6] M. Erdt, M. Kirschner, K. Drechsler, S. Wesarg, M. Hammon, and A. Cavallaro, "Automatic pancreas segmentation in contrast enhanced CT data using learned spatial anatomy and texture descriptors," in *Proc. IEEE Int. Symp. Biomed. Imag. From Nano Macro*, Chicago, IL, USA, Mar./Apr. 2011, pp. 2076–2082.

[7] A. Farag *et al.*, "A bottom-up approach for automatic pancreas segmentation in abdominal CT scans," in *Proc. MICCAI Abdominal Imag. Workshop*, vol. 8676. Boston, MA, USA: Springer, 2014, pp. 103–113.

[8] A. Farag, L. Lu, H. R. Roth, J. Liu, E. Turkbey, and R. M. Summers, "A bottom-up approach for pancreas segmentation using cascaded Superpixels and (deep) image patch labeling," *IEEE Trans. Image Process.*, vol. 26, no. 1, pp. 386–399, Jan. 2016.

[9] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[10] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2014.

[11] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for scene segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[12] H. Roth, A. Farag, L. Lu, E. B. Turkbey, and R. M. Summers, "Deep convolutional networks for pancreas segmentation in CT imaging," *Proc. SPIE*, vol. 9413, Mar. 2015, Art. no. 94131G.

[13] H. R. Roth, L. Lu, A. Farag, H.-C. Shin, J. Liu, E. B. Turkbey, and R. M. Summers, "DeepOrgan: Multi-level deep convolutional networks for automated pancreas segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Munich, Germany, 2015, pp. 556–564.

[14] H. R. Roth, L. Lu, A. Farag, A. Sohn, and R. M. Summers, "Spatial aggregation of holistically-nested networks for automated pancreas segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Athens, Greece, 2016, pp. 451–459.

[15] H. R. Roth, L. Lu, N. Lay, A. P. Harrison, A. Farag, A. Sohn, and R. M. Summers, "Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation," *Med. Image Anal.*, vol. 45, pp. 94–107, Apr. 2018.

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

[17] A. Criminisi, J. Shotton, D. Robertson, and E. Konukoglu, "Regression forests for efficient anatomy detection and localization in CT studies," in *Proc. Int. MICCAI Workshop Med. Comput. Vis.*, Beijing, China, 2010, pp. 106–111.

[18] N. Lay, N. Birkbeck, J. Zhang, and S. K. Zhou, "Rapid multi-organ segmentation using context integration and discriminative models," in *Information Processing in Medical Imaging*. Asilomar, CA, USA: Springer, 2013, pp. 450–462.

[19] S. Liang, F. Tang, X. Huang, K. Yang, T. Zhong, R. Hu, S. Liu, X. Yuan, and Y. Zhang, "Deep-learning-based detection and segmentation of organs at risk in nasopharyngeal carcinoma computed tomographic images for radiotherapy planning," *Eur. Radiol.*, vol. 29, no. 4, pp. 1961–1967, 2019.

[20] Y. F. Zheng *et al.*, "Deep learning based automatic segmentation of pathological kidney in CT: Local versus global image context," in *Deep Learning and Convolutional Neural Networks for Medical Image Computing*. Cham, Switzerland: Springer, 2017, ch. 14, pp. 241–255. [Online]. Available: https://link.springer.com/book/10.1007/978-3-319-42999-1

[21] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.

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

[23] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.

[24] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Munich, Germany, 2015, pp. 234–241.

[25] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*. [Online]. Available: https://arxiv.org/abs/1704.04861

[26] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis. (3DV)*, Oct. 2016, pp. 565–571.

[27] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, Jun. 2017, pp. 2999–3007.

[28] G. Csurka, D. Larlus, F. Perronnin, and F. Meylan, "What is a good evaluation measure for semantic segmentation?" in *Proc. British Machine Vision Association*, 2013.

[29] M. Berman, A. R. Triki, and M. B. Blaschko, "The lovasz-softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4413–4421.

[30] Z. Zhou, "Ensemble learning," in *Encyclopedia of Biometrics*. New York, NY, USA: Springer, 2009, pp. 270–273.

[31] L. K. Hansen and P. Salamon, "Neural network ensembles," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 10, pp. 993–1001, Oct. 1990.

[32] F. Chollet. (2015). *Keras*. [Online]. Available: https://keras.io

[33] M. Adadi. 2015. *Tensorflow: Large-Scale Machine Learning o Heterogeneous Systems*. [Online]. Available: https://tensorflow.org.

[34] L. I. Kuncheva and C. J. Whitaker, "Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy," *Mach. Learn.*, vol. 51, no. 2, pp. 181–207, 2003.

[35] P. Hu, F. Wu, J. Peng, Y. Bao, F. Chen, and D. Kong, "Automatic abdominal multi-organ segmentation using deep convolutional neural network and time-implicit level sets," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 12, no. 3, pp. 399–411, 2016.

SHANGQING LIU was born in China, in 1994. She received the B.S. degree from the Department of Biomedical Engineering, Southern Medical University, Guangzhou, China, in 2017, where she is currently pursuing the Ph.D. degree with the Guangdong Provincial Key Laboratory of Medical Image Processing. Her research interests include image segmentation, medical image processing and analysis, machine learning, and computerized-aid diagnosis.

XINRUI YUAN was born in China, in 1995. She received the B.S. degree in biomedical engineering from Southern Medical University, Guangzhou, China, in 2017, where she is currently pursuing the M.S. degree in engineering with the Department of Biomedical Engineering. Her research interests include the medical image analysis and computerized-aid diagnosis.

RUNYUE HU was born in China, in 1994. He received the B.S. degree in biomedical engineering from Southern Medical University, Guangzhou, China, in 2017, where he is currently pursuing the M.S. degree in engineering with the Department of Biomedical Engineering. His research interests include the medical image analysis and image registration.

SHUJUN LIANG received the B.Eng. degree in biomedical engineering from Southern Medical University, Guangzhou, China, where she is currently pursuing the Ph.D. degree in biomedical engineering. Her current research interests include image segmentation and medical image analysis.

SHAOHUA FENG received the bachelor's degree in clinical medicine from Inner Mongolia Medical University. He is currently working as the Director of the Radiology Department, Manzhouli People's Hospital. The research "the diagnostic comparison of CT virtual endoscopy and colonoscope" won the Third Prize of Inner Mongolia Medical Association Science and Technology, in 2013.

YUHUA AI received the M.S. degree in biomedical engineering and the Ph.D. degree in communication and information system from the South China University of Technology, in 2000 and 2007, respectively. He is currently a Senior Engineer with the Computer Center of Nanfang Hospital Affiliated to Southern Medical University, China. His research interests include biomedical signal/image processing, the Internet of Things, data mining, machine learning, pervasive and ubiquitous computing, and service computing in healthcare domain.

YU ZHANG received the Ph.D. degree in biomedical engineering from First Military Medical University, China, in 2003. He is currently the Vice Dean with the School of Biomedical Engineering, Southern Medical University, China. His research focus has been mainly in the area of medical image processing and analysis, data mining, machine learning, and feature engineering. He has published a series of related articles in the IEEE TRANSACTIONS ON MEDICAL IMAGING, the IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS, *Medical Physics*, *Physics in Medicine and Biology*, and *Neuroimage*.