



## A dual branch and fine-grained enhancement network for pancreatic tumor segmentation in contrast enhanced CT images

Zhibang Zhou<sup>a</sup>, Yun Bian<sup>b</sup>, Shengxue Pan<sup>a</sup>, Qingquan Meng<sup>a</sup>, Weifang Zhu<sup>a</sup>, Fei Shi<sup>a</sup>, Xinjian Chen<sup>a</sup>, Chengwei Shao<sup>b,\*</sup>, Dehui Xiang<sup>a,\*</sup>

<sup>a</sup> School of Electronic and Information Engineering, Soochow University, Jiangsu 215006, China

<sup>b</sup> Department of Radiology, Changhai Hospital, The Navy Military Medical University, Shanghai, China

### ARTICLE INFO

#### Keywords:

Pancreatic tumor segmentation  
Computed tomography  
Dual-branch coding network  
Fine-grained enhancement network

### ABSTRACT

Segmentation of pancreatic tumors in CT images is important for clinical diagnosis and treatment, but it faces challenges of small size, low contrast, and large position difference. To address these issues, the abnormal pancreas is first segmented based on a dual branch coding network (DB-Net) using a coarse-to-fine segmentation strategy. In the encoder part, one branch extracts the semantic features of the pancreas and its surroundings, and the other branch captures the complex pancreas through wide-channel convolution and few down-sampling operations. An aggregation layer is used to fuse the different feature maps obtained by the two branches, and a U-Net decoder is used to segment the abnormal pancreas in CT images with pancreatic tumors. DB-Net is further trained to obtain the accurate pancreatic segmentation. Then, pancreatic tumors are segmented in the pancreas based on the fine-grained enhancement network (FE-Net). The FE-Net integrates a contrast enhancement block with a reverse attention block to extract detailed features and excavate effective information from the feature maps of the encoder and decoder to segment pancreatic tumors. In order to segment the tumor more accurately, the pancreatic tumor is segmented in the cropped pancreas. Experiments on 116 contrast-enhanced abdominal CT volumes of pancreatic cancer and 42 contrast-enhanced abdominal CT volumes of normal pancreas verify the effectiveness of the proposed framework in pancreatic tumor segmentation by using the two-fold cross-validation strategy. Compared to state-of-the-art deep learning segmentation network, the proposed method can achieve better segmentation of pancreas and pancreatic tumors.

### 1. Introduction

Pancreatic cancer is one of the common malignant tumors of the digestive system. The characteristics of pancreatic cancer include unobvious early symptoms, high degree of malignancy, rapid progression of the disease, and poor prognosis [1,2]. According to statistics from the World Health Organization (WHO), the incidence of pancreatic cancer in the world was 277,000, the incidence ranked 13th among malignant tumors. The deaths were about 266,000, and the mortality rate ranked 7th among malignant tumors [3]. CT is currently one of the most important imaging for pancreatic cancer due to the short scanning time, large scanning range, and high resolution [4,5]. CT can be used to detect lesions and make a reliable evaluation and estimation of the surgeries of pancreatic tumors [6,7]. The use of advanced algorithms to automatically segment pancreatic tumors in CT images is of great significance for the diagnosis and treatment of pancreatic cancer. Accurate and automatic pancreatic tumors and pancreas segmentation can save

physicians' efforts to annotate pancreas anatomy since it is tedious and time-consuming to label each voxel in huge number of slices in routine clinics.

Although CT have been widely used, pancreatic tumors and pancreas segmentation are still a challenging task in the medical image processing field. As shown in Fig. 1, the main research difficulties of pancreatic tumor segmentation in CT images include the following three points. First, the pancreas is very small, less than 0.5% of the entire CT volume [8], while pancreatic tumors are much smaller, and most pancreatic tumors account for less than 0.1% of the entire CT volume [9]. Second, the boundary of the pancreas and tumors is visually blurred, and the contrast between surrounding tissues and the pancreas is low, especially at the head of the pancreas. The boundary between the pancreas and the duodenum is difficult to distinguish, and the parenchyma and tumors of the pancreas are also difficult to be distinguished. The contrast between surrounding tissues and the

\* Corresponding authors.

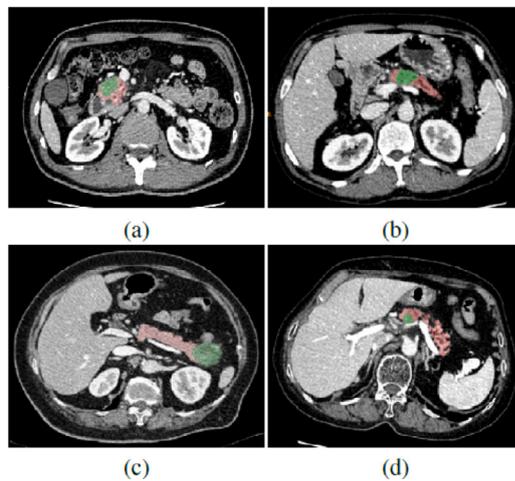
E-mail address: [xiangdehui@suda.edu.cn](mailto:xiangdehui@suda.edu.cn) (D. Xiang).

<https://doi.org/10.1016/j.bspc.2022.104516>

Received 3 August 2022; Received in revised form 29 October 2022; Accepted 18 December 2022

Available online 24 December 2022

1746-8094/© 2022 Elsevier Ltd. All rights reserved.



**Fig. 1.** Challenges in pancreas and tumor segmentation. The green region is the pancreatic tumor, and the red region is the rest of the pancreas. These images show the size, boundary, shape and location of the pancreas and tumors have great anatomical differences between different patients.

tumor is also relatively low [10]. Third, the shape and location of the pancreas and tumors have great anatomical differences between different patients, especially the position of pancreatic tumors, which can exist anywhere in the pancreas [11]. Therefore, to address the above issues, an effective segmentation framework is designed for robust pancreatic tumor segmentation with context adaptation and detailed feature extraction.

In routine clinics, pancreatic tumor is often with small number of images and only few pixels in CT slices. If the tumor is directly segmented in the entire CT slice, the segmented result is usually inaccurate. Therefore, generally, the pancreas in CT images is first segmented, and then the pancreatic area is cropped according to the results of the pancreas segmentation. Subsequently, the pancreatic tumor is segmented in the pancreas, and finally the segmented result of the pancreatic tumor is obtained [9,12]. Therefore, the final pancreatic tumor segmentation performance largely depends on the pancreas segmentation. However, the shape of the pancreas varies greatly from different patients, and the surrounding environment is complex. It is difficult for previous Convolutional Neural Networks (CNN) to extract sufficient semantic information and detailed information of the pancreas to determine the exact location and contour of the pancreas. In addition, in the stage of pancreatic tumor segmentation, previous methods cannot extract well the edges of the tumor to adapt the contrast between the pancreatic tumor and other surrounding tissues, resulting in inaccurate segmented results.

Based on the above observations, a novel CNN called DBFE-Net is designed to segment pancreatic tumors in CT images. A dual-branch encoder–decoder network (DB-Net) as the first part of DBFE-Net is designed to segment the pancreas. The semantic branch extracts rich semantic features of the pancreas and its surroundings through deep convolution and multiple down-sampling operations to overcome the influence of numerous blood vessels, tissues, and organs around the pancreas, and the detailed branch captures complex details of the pancreas through wide-channel convolution and few down-sampling operations to overcome the problem of unclear pancreatic edges. The two kinds of information are weighted and fused with the aggregation layer. Then, the segmented results are obtained by upsampling like U-Net [13]. Based on our previous work [14], we further follow the previous idea of pancreas segmentation and use a coarse-to-fine segmentation strategy to achieve accurate pancreas segmentation [15, 16]. A fine-grained enhancement encoder–decoder network (FE-Net) as the second part of DBFE-Net is then designed to segment pancreatic

tumors in the pancreas. First, a contrast enhancement block (CEB) is designed to extract the detailed features in the feature map and use the attention mechanism to integrate the features to solve the problem of low contrast between the pancreatic tumor and the surrounding structures. Then, a reverse attention block (RAB) is proposed to invert the decoder feature map to guide the mining of effective information in the encoder feature map. In summary, the main contributions of this work are summarized as follows.

1. DB-Net is proposed to segment the abnormal pancreas. In the encoder part, one branch extracts the rich semantic features of the pancreas and its surrounding tissues or organs through deep convolution and down-sampling operations, and the other branch captures the complex pancreas through wide-channel convolution and few down-sampling operations. DB-Net combines semantic information extraction branch and detailed information extraction branch, which can better segment the pancreatic head and tail regions. An aggregation layer is used to fuse the different feature maps obtained by the two branches, and an U-Net decoder is used to segment the abnormal pancreas in CT images with pancreatic tumors. DB-Net is further trained in the cropped initial segmentation to obtain the accurate pancreatic segmentation.

2. In order to solve the problem of small size and low contrast, FE-Net is designed to improve segmentation accuracy of pancreatic tumors. CEB is set after the encoder layer to extract the edge detail information of the pancreas in the feature map of the encoder layer. RAB is set between the encoder and the decoder. The output of each layer of the decoder features are used to fuse the complementary regions and details of the previous layer to make up for the tumor information lost in the down-sampling. In order to segment the tumor more accurately, the pancreatic tumor is segmented in the cropped pancreas.

The rest of this paper is organized as follows. We first review related work in Section 2. The technical details of the proposed pancreatic tumor segmentation scheme are described in Section 3. Section 4 presents the experimental results. The paper is finally concluded in Section 5.

## 2. Related work

With the development of deep learning [17–19], it has been widely used in many computer vision tasks, such as semantic segmentation [20–22], edge detection [23–25] and image denoising [26–28]. Medical image segmentation also used deep learning methods to achieve considerable performance [13,29–32]. Among them, the segmentation of pancreas and pancreatic tumors has also made some progress by using deep learning methods.

### 2.1. Pancreas segmentation

Regarding to the normal pancreas segmentation of abdominal CT images, many researchers have conducted a lot of researches. Roth et al. [33] used random forest algorithm and deep learning to obtain semantic clues of organ interior and boundary maps and integrated them to generate pixel-level label pancreas segmentation. Ma et al. [34] proposed a new Bayesian model that combined the segmented results of deep neural networks and statistical shape models to refine the segmented results of the pancreas. Both these methods combined traditional methods and deep learning methods to jointly optimize the segmentation of the pancreas. Due to the use of traditional algorithms, compared with the current emerging pure deep learning methods, the speed is slower to obtain the final segmented results. Roth et al. [35] subsequently used a nested convolutional neural network to implement the automatic positioning and segmentation of the pancreas using a two-stage strategy. Zhou et al. [36] used a two-stage coarse-to-fine method to segment the pancreas. They separately trained two fully convolutional neural network (FCN) to process the entire images and the regions cropped according to the bounding box. In the test phase, the coarse segmented results of the first network were sent to the second

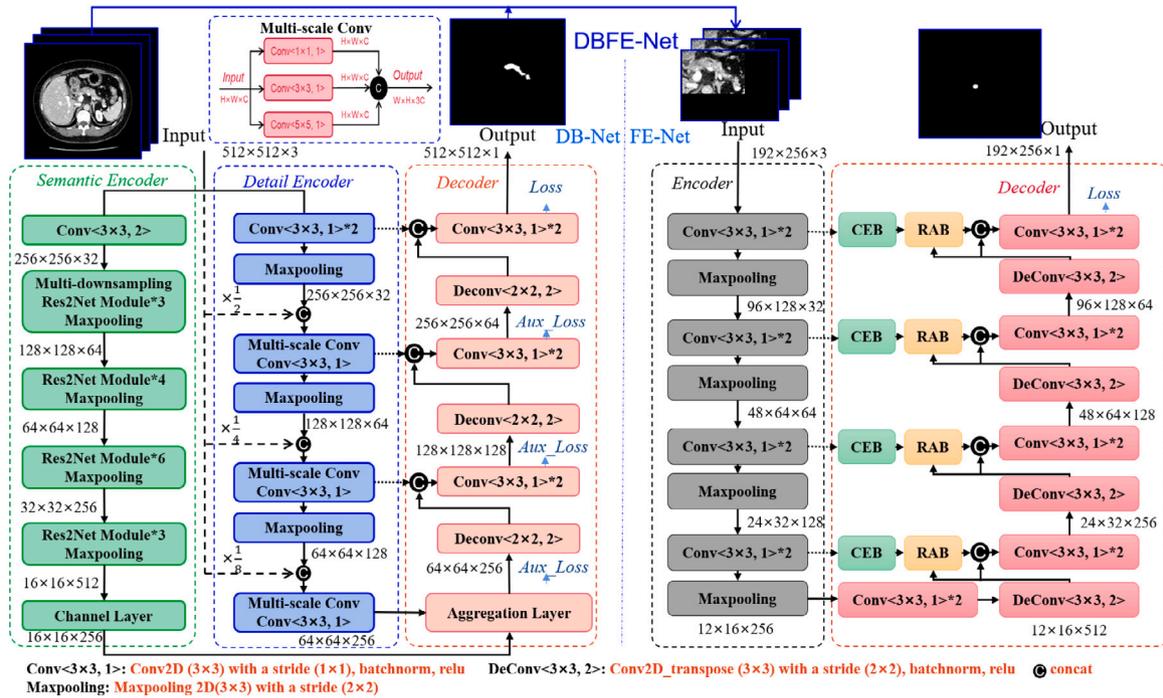


Fig. 2. Overview of the proposed DBFE-Net. DB-Net is used to extract semantic and fine-grained feature to achieve accurate segmentation of the pancreas with a coarse-to-fine strategy, and then FE-Net is used to extract fine-grained feature with higher contrast to segment the pancreatic tumor in the pancreas region.

network, and the fixed-point model was optimized through an iterative process to achieve more accurate segmentation. Man et al. [37] proposed a deep reinforcement learning strategy. First, a pancreas location bounding box was generated, and then a deformable deep U-shaped network method was proposed to further segment the pancreas with obvious non-rigid shape deformation. To solve the spatial non-smooth problem of segmentation between adjacent pancreatic image slices and improve the consistency of the shape of the pancreas after segmentation, Cai et al. [38] introduced a recurrent neural network to process the continuous initial segmented results of a two-dimensional convolutional neural network and effectively learn long-term dependence information between adjacent slices through LSTM. To use the three-dimensional information of the pancreas, Oktay et al. [39] used the spatial attention mechanism in three dimensions to guide the network to focus on the pancreas position in the skip-connection part. Zhu et al. [40] proposed a coarse-to-fine three-dimensional convolutional neural network framework, which divided three-dimensional CT volume image into many small cubes with the same size as input and used the three-dimensional spatial information for fine segmentation of the pancreas. However, all these methods did not consider the complex background interference around the pancreas.

### 2.2. Pancreatic tumor segmentation

Although the segmentation of normal pancreas in CT images still needs improvement, there are few studies on the segmentation of pancreatic tumors in CT images. The main reason is due to the small size of pancreatic tumors, low contrast with pancreatic parenchyma, and huge differences in the location of pancreatic tumors, it is difficult to accurately segment pancreatic tumors. Dmitriev et al. [9] proposed a semi-automatic segmentation algorithm for pancreatic cysts. Because the size and shape of the cysts are different, the authors used a new combination of random walk and region growing methods to delineate the boundary between the pancreas and the cyst. Some researchers have begun to use deep learning methods for the segmentation of pancreatic tumors. Zhou et al. [40] used a deep learning method to segment the pancreas first through FCN, and then cropped the original image to segment the pancreatic cyst through another FCN. Quo

et al. [12] combined contextual U-Net and a graph-based framework to segment 3D tumors. First, the tumor was manually cropped from the entire image into a cuboid. Then, each two-dimensional slice and its two adjacent slices were sent to U-Net to obtain the segmentation probability map. The final segmentation was obtained through the graph search algorithm. Zhou et al. [41] proposed a dual-path network to fuse dual-phase CT images of the arterial phase and the venous phase to segment pancreatic tumors and added a matching loss function to encourage the commonality between high-level feature representations of different phases. Zhu et al. [42] cropped CT images into three different sizes of three-dimensional small blocks and sent them to three three-dimensional U-Net with a deep supervision mechanism. Finally, the three prediction results of the network were averaged to obtain the final pancreatic tumor segmentation.

## 3. Method

The proposed DBFE-Net for pancreatic tumor segmentation is shown in Fig. 2. First, the designed DB-Net and a coarse-to-fine strategy is used to achieve accurate segmentation of the pancreas. Second, CT images are cropped according to the final segmentation of the pancreas to obtain the pancreas region. Third, the designed FE-Net is used to segment the pancreatic tumor in the pancreas region. Finally, the segmented pancreatic tumor is restored to the original size to obtain the final pancreatic tumor segmentation.

### 3.1. Pancreas segmentation

The overall structure of the dual branch coding network (DB-Net) is shown in Fig. 2. The idea of this structure comes from Bisenet V2 [43]. The encoder path is a dual-branch encoding structure. The left branch is the semantic feature extraction branch, and the semantic information is extracted in the deep layer of the network through sequent convolution and down-sampling operations. The right branch is the detail feature extraction branch. In the network shallow layer, multiple wide-channel convolution operations are used to extract detailed information. The deepest level of DB-Net is the aggregation layer to fuse semantic and detailed features.

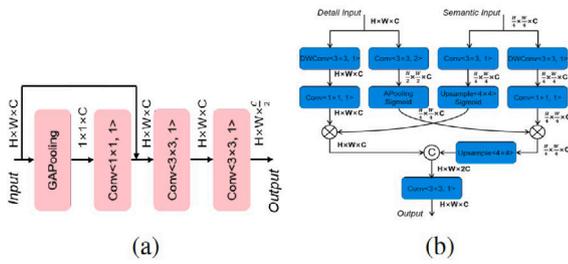


Fig. 3. (a) The structure of the channel layer, (b) The structure of the aggregation layer.

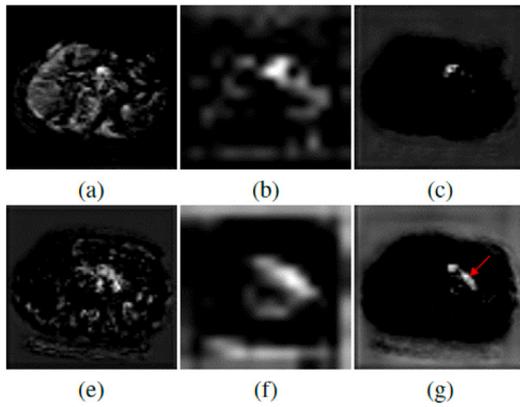


Fig. 4. Feature extraction by Bisenet V2 and DB-Net for the image in Fig. 1(d). (a) A typical feature map after the detail information extraction branch in Bisenet V2, (b) A typical feature map after the semantic information extraction branch in Bisenet V2, (c) A typical feature map after the aggregation layer in Bisenet V2, (d) A typical feature map after the detail information extraction branch in DB-Net, (e) A typical feature map after the semantic information extraction branch in DB-Net, (f) A typical feature map after the aggregation layer in DB-Net. As the red arrow points, the typical feature maps are enhanced by our semantic information extraction branch and detail information extraction branch in DB-Net stronger than those in Bisenet V2.

### 3.1.1. Semantic information extraction branch

As shown in Fig. 2, semantic information extraction branch mainly contains seven layers. The first layer is a convolution with a stride of 2 to extract image features and rapidly reduce the image size. The second layer is the multi-downsampling module like Bisenet V2. Different from Bisenet V2 using convolution and deep convolution, Res2Net module [44] is used in the third, fourth, fifth, and sixth layers to enhance the multi-scale representation ability with a finer granularity. It can combine features of different scales and numbers without adding additional calculations. In the seventh layer, a channel layer is introduced like Bisenet V2 to compress the number of channels of the semantic information extraction branch. The structure of the channel layer is shown in Fig. 3(a).

### 3.1.2. Detail information extraction branch

As shown in Fig. 2, the detail information extraction branch mainly contains four layers. Different from Bisenet V2, this branch extracts fine-grained features through a combination of ordinary convolution and multi-scale convolution. The multi-scale convolution layer includes three convolutions with convolution kernels of  $1 \times 1$ ,  $3 \times 3$  and  $5 \times 5$  to extract features of different scales. These features are concatenated together to aggregate all the features. The aggregation layer is designed like Bisenet V2 to fuse the output feature map of the detailed information extraction branch and the output feature map of the semantic information extraction branch. In order to reduce parameters, a combination of deep convolution and ordinary convolution is used to obtain semantic features, which are multiplied with the attention weight of another branch to achieve mutual guidance and transmission

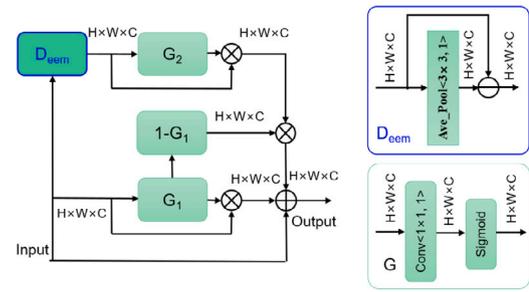


Fig. 5. Contrast enhancement block (CEB).

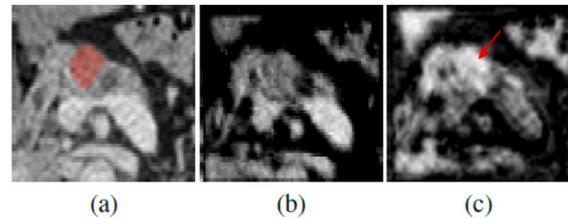


Fig. 6. The typical feature maps before and after CEB for pancreatic tumor segmentation. (a) Original image with manual annotation of pancreatic tumor (the red mask), (b) Input of CEB, (c) Output of CEB. As the red arrow points, the feature map is significantly enhanced by the CEB.

of information. Finally, the information is merged through convolution. The structure of the aggregation layer is shown in Fig. 3(b). In order to utilize the original image information, the original image is down-sampled to one-half, one-quarter, and one-eighth of its own size. Then, they are concatenated with the feature maps after the three max-pooling layers to enrich detailed information in the feature map.

Compared to Bisenet V2, Res2Net modules and multi-scale convolution are used the semantic information extraction branch and detail information extraction branch in DB-Net to extract the information of different sizes of receptive fields to make the information in the feature map richer and more accurate; and deconvolution and skip-connection operation are used in the decoder to gradually enlarge the feature map to the original image size like U-Net decoder. In addition, the deep supervision mechanism is also used to resample each layer of the decoder to its original size and the loss function is computed according to different levels of supervision. Fig. 4 shows the typical feature maps before and after the aggregation layer of DB-Net and Bisenet V2. As shown in Fig. 4, compared with Bisenet V2, detail information extraction branch, DB-Net enhances the detailed information of the pancreas region and suppresses the detailed information of non-pancreas, for semantic information extraction branch, DB-Net retains the semantic information of the pancreas very well. Therefore, after fusion through the aggregation layer, DB-Net can better segment the pancreas.

### 3.2. Pancreatic tumor segmentation

The overall structure of the fine-grained enhancement encoder-decoder network (FE-Net) is shown in Fig. 2. Since U-Net [13] achieves good results in medical image segmentation, U-Net framework is also used. In order to further improve pancreatic tumor segmentation performance, a CEB is proposed to guide the network to pay more attention to the edge details of the pancreatic tumor. Second, RAB is designed using decoder information to guide the mining of complementary discriminative regions. Pancreatic tumor is segmented within the segmented pancreas to eliminate most of the irrelevant background regions and improve the accuracy of pancreatic tumor segmentation.

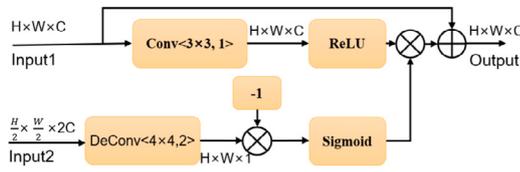


Fig. 7. Reverse attention block (RAB).

### 3.2.1. Contrast enhancement block

In the traditional U-Net, in order to use possible information extracted by the encoder branch in the decoder branch, a skip connection operation is designed to directly merge the feature maps of the encoder layer and the decoder layer by concatenation. But it is difficult to filter irrelevant information in the feature map of the encoder. To highlight the edge details of the pancreatic tumor in the feature map of the encoder layer, a CEB is proposed as shown in Fig. 5. The feature map of the encoder layer can be denoted as *Input*, and the output of the contrast enhancement block *Output* can be formulated as

$$output = (1 + G_1) \times input + (1 - G_1) \times (G_2 \times D_{em}), \quad (1)$$

where  $D_{em}$  is the edge extraction module as

$$D_{em} = input - AveragePooling(input), \quad (2)$$

The CEB use the self-attention mechanism to enhance the feature representation, and ultimately better extract the fine-grained features of pancreatic tumors. Average pooling can be used to extract local average information. The edge details can be obtained by subtracting the local average information from *Input*.  $G_1$  and  $G_2$  modules compute the deterministic contribution weight of each position in the feature map to the final expected result through the attention mechanism.  $G_1$  is multiplied by the *Input* to obtain deterministic information of each position in the *Input*. Similarly, the key information of the edge feature map is obtained by multiplying *Input* and  $D_{em}$ , and  $(1 - G_1)$  corresponds to the weight of the uncertain position in the *Input* feature map, and multiplying with the key information of the edge details can enhance and supplement edge detail information. The final output result introduces a residual structure, which not only enhances the key feature information, but also supplements the edge detail information, and can also avoid the difficulty of gradient backpropagation that may occur after deep convolution and the problem of difficult to improve accuracy. Fig. 6 shows the feature maps before and after CEB. It can be seen that the contrast between the tumor area and surrounding tissues is very low before passing CEB, and the boundary between the tumor area and surrounding tissues becomes clear after passing CEB.

### 3.2.2. Reverse attention block

The network produces the segmented results of abnormal pancreas through sequentially upsampling. It can only capture a relatively rough location without structural details. To solve this problem, a reverse attention block is designed as shown in Fig. 7. The block utilizes the output feature of each layer of decoder to mine the complementary regions and details of the previous layer. Specifically, given the high-level skip-connection inputs *Input1* and the reverse attention weight  $W_r$ , the attention feature map *Output* can be denoted as

$$output = \sigma_1(f_1(input1)) \times W_r + input1, \quad (3)$$

The reverse attention weight  $W_r$  can be formulated as

$$W_r = \sigma_2(-f_2(input2)), \quad (4)$$

where  $f_1$  denotes the convolution operation,  $\sigma_1$  denotes the ReLU activation function,  $f_2$  denotes the deconvolution operation,  $\sigma_2$  denotes the Sigmoid activation function, and *Input2* denotes the feature map of the current decoder layer. Inverting the rough feature map obtained by

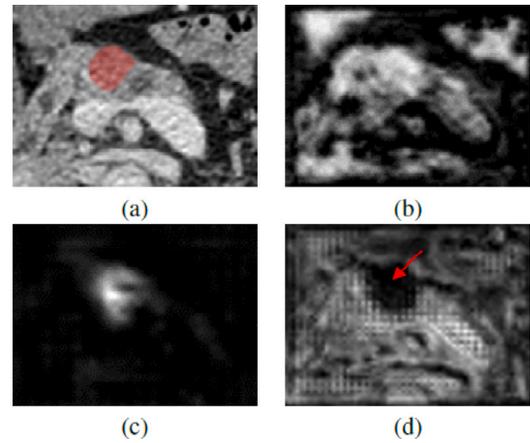


Fig. 8. The typical feature maps before and after RAB. (a) Original image, (b) Input1 of RAB, (c) Input2 of RAB, (d) Output of RAB. As the red arrow points, the feature map of pancreatic tumor area is significantly cleaned by the RAB.

deconvolution and multiplying it with the high-level encoder feature map can refine the inaccurate and rough estimation into an accurate and complete prediction map. Fig. 8 shows the feature maps before and after RAB. We can see that after RAB, the rough feature map of the decoder layer guides the feature map of the upper encoder layer to obtain a fine inverted pancreatic tumor region, and highlight the tumor region to obtain a more accurate boundary.

### 3.3. Loss function

As pancreas and pancreatic tumor segmentation are two-class segmentation tasks with class imbalance, Dice similarity coefficient loss (Dice loss) rather than cross entropy loss is used to supervise the training of the DBFE-Net. Given prediction map  $X$  and ground truth  $Y$ , Dice loss is defined as

$$loss = 1 - \frac{2|X \cap Y|}{|X| + |Y|}, \quad (5)$$

In DB-Net, deep supervision mechanism is used to assist supervision with loss functions behind each layer of the decoder. The overall loss function  $L_{total}$  can be defined as

$$Loss = loss_l + \sum_{i=1}^n aux\_loss_i, \quad (6)$$

where  $loss_l$  represents the loss function of the last layer of the network,  $aux\_loss_i$  represents the loss function of the  $i$ th decoder layer. In this network,  $n = 3$ , and the loss function is the Dice loss in Eq. (5).

### 3.4. Implementation details

#### 3.4.1. Preprocessing

The original intensity value of the CT image is truncated to  $[-100, 240]$  to enhance pancreatic details, and then each original CT image is normalized to a mean value of 0 and a variance of 1 to reduce data discrepancies caused by the medical image acquisition process. In order to prevent the model from being severely affected by the background, the experiment of pancreas segmentation is trained on the slices containing at least 50 pancreatic pixels. In order to obtain and utilize part of the spatial information in the three-dimensional image without increasing the amount of calculation, three consecutive slices are used as the input of our network

### 3.4.2. Pancreas segmentation

Pancreas segmentation is completed by two stages of training and predicting. In the first stage, the size of the network input is  $512 \times 512 \times 3$ . In the second stage, the original CT slices are cropped according to the segmented results predicted in the first stage, and the cropped region is to extend 20 pixels up, down, left, and right along the largest bounding box that can contain the pancreatic segmented results. Then, the cropped CT slice is used as the input of the second stage pancreas segmentation network. Since the size of the cropped result of each slice is not fixed, in order to meet the requirements of the fixed input size of the neural network, the right and bottom edges of the cropped CT slices are filled with zeros to unify their size to  $192 \times 256$ . Since the pancreas is a consecutive and complete organ, after the segmentation of our first-stage network and second-stage network, three-dimensional maximum connected regions are reserved to further improve the segmentation accuracy.

### 3.4.3. Pancreas tumor segmentation

The CT volumes of the pancreas are further cropped according to the results of the pancreas segmentation. The cropped regions are also to extend 20 pixels up, down, left, and right along the largest bounding box that can contain the pancreatic segmented results. Then, zero padding is also used on the right and bottom edges of the cropped CT slices to unify the size to  $192 \times 256$ . Since only a small part of the pancreatic slices contain pancreatic tumors, this experiment uses the slices containing pancreatic tumors and the slices without pancreatic tumors at a ratio of about 1:1 as the training set to train the network to prevent the model from being seriously affected by the background. In order to obtain and use part of the spatial information in the three-dimensional images without increasing the amount of calculation, three consecutive slices are also used as the input of the pancreatic tumor segmentation network.

### 3.4.4. Training parameters

All networks are trained on a workstation equipped with one NVIDIA Tesla K40 m GPU with 12G memory. Adam algorithm is used with an initial learning rate of  $10^{-4}$  to optimize the weights of the network in the training process. In our experiment, the batch size is 3. In the training of the pancreas segmentation network, the epoch is 20. In the training of the pancreatic tumor segmentation network, the epoch is 40. The total training time is 35.94 h, and the overall average test time for each CT volumes is 18.39 s. Among them, the training time of the first stage, second stage of DB-Net and FE-Net are 13.73 h, 4.63 h and 17.58 h respectively, and the average test time for each CT volumes are 9.87 s, 4.61 s and 3.91 s, respectively.

## 4. Experiments

In this section, we test the performance of pancreas segmentation and pancreatic tumor segmentation by performing comprehensive evaluations on our benchmark dataset. Additional details about experiments and results are reported as follows.

### 4.1. Dataset

The dataset is provided by Shanghai Changhai Hospital. CT images in the portal venous phase are used. The abnormal pancreas dataset contains 116 contrast-enhanced abdominal CT volumes, with 109 PDAC (pancreatic ductal adenocarcinoma) and 7 PASC (pancreatic adenosquamous carcinoma). The maximum diameters of tumors in 57 CT volumes are smaller than 2 cm, the maximum diameters of tumors in 56 CT volumes are between 2 cm and 4 cm, and the maximum diameters of tumors in 3 CT volumes are larger than 4 cm. Another normal pancreas dataset contains 42 contrast-enhanced abdominal CT volumes. The size of each CT volume is  $512 \times 512 \times L$ , where  $L \in [174, 376]$  is the number of sampling slices along the long axis of the body,

**Table 1**

Ablation experiment of pancreas segmentation results of dual-branch structure (Mean  $\pm$  Standard Deviation).

Methods	Recall (%)	Precision (%)	Jaccard (%)	F1 (%)
Semantic-only	70.15 $\pm$ 20.16	84.63 $\pm$ 8.31	61.44 $\pm$ 16.93	74.56 $\pm$ 15.15
Detail-only	78.65 $\pm$ 13.94	<b>90.75 <math>\pm</math> 5.35</b>	72.64 $\pm$ 12.65	83.44 $\pm$ 9.85
DB-Net	<b>85.55 <math>\pm</math> 12.05</b>	89.20 $\pm$ 5.59	<b>77.24 <math>\pm</math> 11.05</b>	<b>86.63 <math>\pm</math> 8.63</b>

**Table 2**

Comparisons of abnormal pancreas segmentation results between DB-Net and state-of-the-art networks (Mean  $\pm$  Standard Deviation).

Methods	Recall (%)	Precision (%)	Jaccard (%)	F1 (%)	p
FCN [15]	72.46 $\pm$ 16.23	83.07 $\pm$ 7.86	62.27 $\pm$ 12.97	75.82 $\pm$ 11.75	$8.22 \times 10^{-29}$
BisenetV2 [43]	75.21 $\pm$ 15.82	81.68 $\pm$ 9.44	64.03 $\pm$ 13.54	76.85 $\pm$ 12.20	$1.37 \times 10^{-20}$
ResDSN [16]	<b>91.39 <math>\pm</math> 10.20</b>	65.50 $\pm$ 14.24	61.84 $\pm$ 13.79	75.40 $\pm$ 12.05	$1.06 \times 10^{-16}$
SegNet [22]	73.73 $\pm$ 17.44	87.86 $\pm$ 6.14	66.14 $\pm$ 14.67	78.50 $\pm$ 13.06	$3.55 \times 10^{-18}$
nnU-Net [45]	91.07 $\pm$ 5.72	70.96 $\pm$ 15.01	66.10 $\pm$ 13.65	78.73 $\pm$ 10.51	$6.34 \times 10^{-13}$
CE-Net [31]	78.58 $\pm$ 15.19	87.03 $\pm$ 6.53	69.85 $\pm$ 12.92	81.38 $\pm$ 10.79	$1.45 \times 10^{-11}$
Att U-Net [39]	80.33 $\pm$ 15.73	89.39 $\pm$ 5.98	72.85 $\pm$ 13.81	83.38 $\pm$ 11.58	$2.45 \times 10^{-7}$
mU-Net [32]	80.67 $\pm$ 15.30	89.23 $\pm$ 5.30	73.14 $\pm$ 13.39	83.65 $\pm$ 10.95	$9.74 \times 10^{-8}$
U-Net [13]	80.63 $\pm$ 15.57	<b>90.26 <math>\pm</math> 5.33</b>	73.93 $\pm$ 13.98	84.11 $\pm$ 11.61	$3.64 \times 10^{-5}$
Swin-UNet [46]	84.42 $\pm$ 5.41	85.33 $\pm$ 4.87	73.36 $\pm$ 5.22	84.69 $\pm$ 3.54	$3.18 \times 10^{-6}$
HNN [35]	83.44 $\pm$ 12.56	87.86 $\pm$ 5.66	74.37 $\pm$ 11.14	84.73 $\pm$ 9.10	$3.93 \times 10^{-7}$
w-Net [47]	83.47 $\pm$ 10.25	88.09 $\pm$ 5.25	74.74 $\pm$ 8.89	85.23 $\pm$ 6.20	$3.72 \times 10^{-3}$
DB-Net	85.55 $\pm$ 12.05	89.20 $\pm$ 5.59	<b>77.24 <math>\pm</math> 11.05</b>	<b>86.63 <math>\pm</math> 8.63</b>	-

**Table 3**

Comparisons of normal pancreas segmentation results between DB-Net and state-of-the-art networks (Mean  $\pm$  Standard Deviation).

Methods	Recall (%)	Precision (%)	Jaccard (%)	F1 (%)	p
ResDSN [16]	86.96 $\pm$ 21.98	59.08 $\pm$ 13.73	53.50 $\pm$ 17.10	67.86 $\pm$ 16.82	$2.73 \times 10^{-8}$
Swin-UNet [46]	63.03 $\pm$ 17.62	77.53 $\pm$ 9.74	53.36 $\pm$ 15.05	68.14 $\pm$ 15.37	$9.22 \times 10^{-21}$
BisenetV2 [43]	66.99 $\pm$ 18.94	79.08 $\pm$ 12.67	57.05 $\pm$ 16.29	71.00 $\pm$ 16.16	$2.2 \times 10^{-8}$
FCN [15]	71.53 $\pm$ 14.69	80.67 $\pm$ 10.03	61.05 $\pm$ 13.23	74.90 $\pm$ 11.24	$1.77 \times 10^{-8}$
nnU-Net [45]	86.56 $\pm$ 11.61	73.19 $\pm$ 18.40	66.36 $\pm$ 17.59	78.23 $\pm$ 14.84	$7.13 \times 10^{-7}$
SegNet [22]	76.96 $\pm$ 15.91	83.76 $\pm$ 9.38	67.01 $\pm$ 14.67	79.22 $\pm$ 11.74	$1.67 \times 10^{-7}$
CE-Net [31]	80.53 $\pm$ 14.81	79.97 $\pm$ 13.01	67.81 $\pm$ 15.81	79.63 $\pm$ 12.73	$2.2 \times 10^{-8}$
HNN [35]	75.69 $\pm$ 17.19	89.95 $\pm$ 8.01	69.58 $\pm$ 16.23	80.85 $\pm$ 12.75	$1.36 \times 10^{-7}$
mU-Net [32]	79.77 $\pm$ 16.58	88.27 $\pm$ 6.38	72.26 $\pm$ 15.03	82.87 $\pm$ 11.87	$1.19 \times 10^{-6}$
w-Net [47]	83.51 $\pm$ 13.48	84.64 $\pm$ 9.39	72.69 $\pm$ 13.91	83.25 $\pm$ 12.12	$3.28 \times 10^{-19}$
U-Net [13]	82.68 $\pm$ 12.42	85.05 $\pm$ 9.69	72.40 $\pm$ 13.32	83.27 $\pm$ 9.43	$4.24 \times 10^{-7}$
Att U-Net [39]	82.74 $\pm$ 13.09	85.69 $\pm$ 10.23	72.69 $\pm$ 13.55	83.43 $\pm$ 9.63	$5.16 \times 10^{-7}$
DB-Net	<b>87.08 <math>\pm</math> 10.20</b>	<b>89.98 <math>\pm</math> 6.97</b>	<b>79.43 <math>\pm</math> 10.91</b>	<b>88.07 <math>\pm</math> 7.62</b>	-

the spatial resolutions are between  $0.579 \text{ mm} \times 0.579 \text{ mm} \times 0.8 \text{ mm}$  and  $0.746 \text{ mm} \times 0.746 \text{ mm} \times 1.5 \text{ mm}$ . All the CT volumes are accurately and manually labeled by an experienced radiologist. The abnormal pancreas dataset is randomly divided into two groups of different patients to perform the experiments according to the two-fold cross-validation strategy.

### 4.2. Evaluation metrics

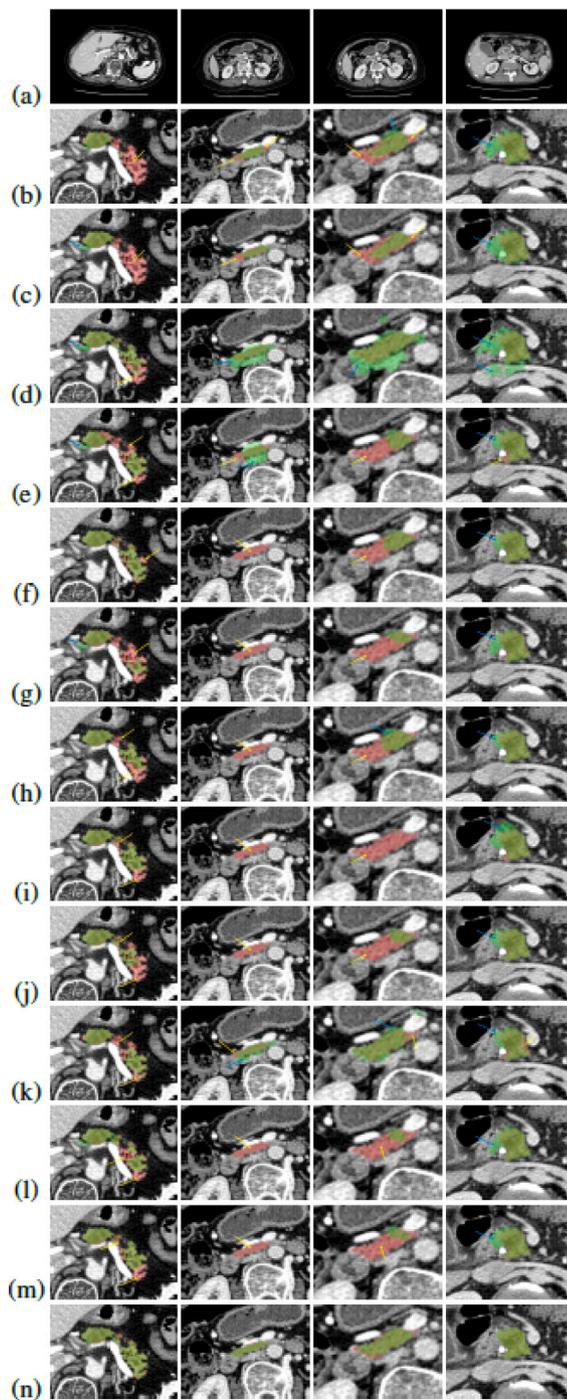
To quantitatively evaluate the performance of our proposed method, the segmented results are evaluated with ground truth based on the following four measures<sup>1</sup>: Precision, Recall, Jaccard, F1. F1 is mainly used to measure the similarity between the segmented results and ground truth. Wilcoxon test of F1<sup>1</sup> is performed to compare the difference between our method and related methods, and  $p < 0.05$  is considered to be statistically significant.

### 4.3. Pancreas segmentation results

#### 4.3.1. Ablation experiments of dual-branch structure

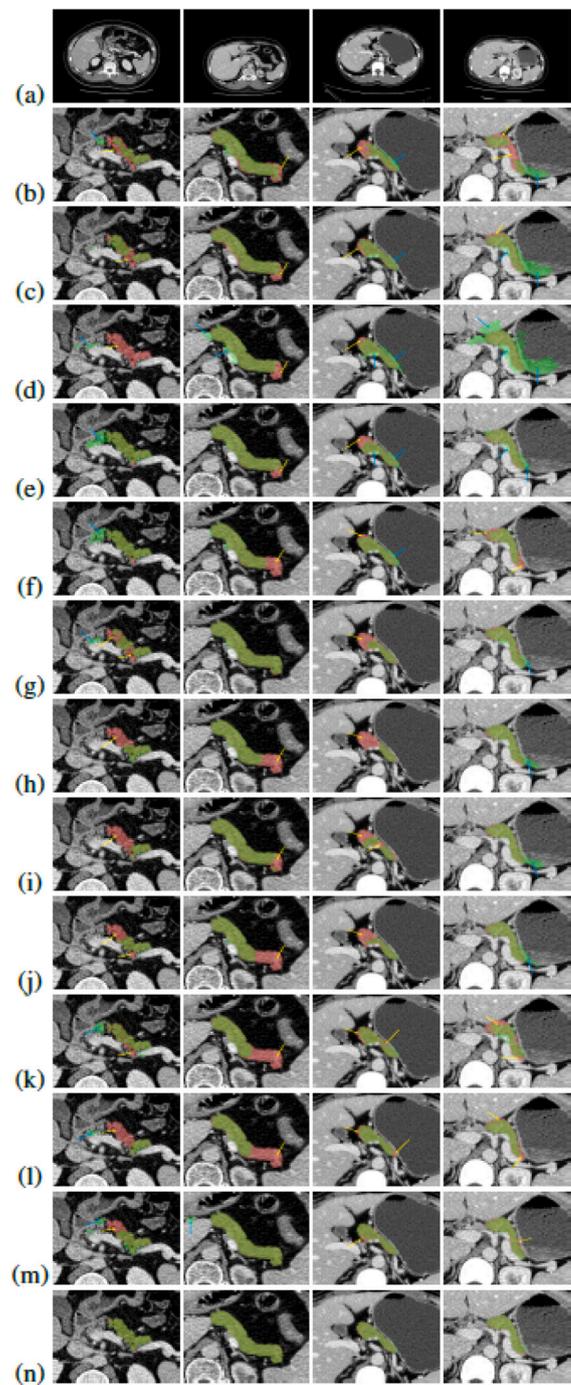
In order to prove the superiority of the proposed dual-branch coding structure, ablation experiment is done. First, semantic information extraction branch is only retained as the encoder, and the other parts of the network remain unchanged. Then, the detail information extraction branch is only retained as the encoder, and the other parts of the network remain unchanged. The experimental results are shown in Table 1. Recall, Precision, Jaccard and F1 scores of DB-Net are increased

<sup>1</sup> <https://scipy.org/>



**Fig. 9.** Abnormal pancreas segmentation of different methods. The red mask is ground truth, the green mask is the segmented result. The yellow arrow points out the under-segmentation and the cyan arrow points out the over-segmentation. (a) Original image, (b)–(l) Locally enlarged pancreas segmented results of different methods. (b) FCN [15], (c) BisenetV2 [43], (d) ResDSN [16], (e) SegNet [22], (f) nnU-Net [45], (g) CE-Net [31], (h) Att U-Net [39], (i) mU-Net [32], (j) U-Net [44], (k) Swin-UNet [46], (l) HNN [35], (m) w-Net [47], (n) DB-Net.

by 15.4, 4.57, 15.8, 12.07 compared to the network with only semantic coding branch, respectively. Recall, Jaccard and F1 scores of DB-Net are increased by 6.9, 4.6, 3.19 compared to the network with only detailed coding branch, respectively. This shows that DB-Net is better than the single-branch network for the pancreas segmentation task in CT images.



**Fig. 10.** Normal pancreas segmentation of different methods. The red mask is ground truth, the green mask is the segmented result. The yellow arrow points out the under-segmentation and the cyan arrow points out the over-segmentation. (a) Original image, (b)–(l) Locally enlarged pancreas segmented results of different methods. (b) FCN [15], (c) BisenetV2 [43], (d) ResDSN [16], (e) SegNet [22], (f) nnU-Net [45], (g) CE-Net [31], (h) Att U-Net [39], (i) mU-Net [32], (j) U-Net [44], (k) Swin-UNet [46], (l) HNN [35], (m) w-Net [47], (n) DB-Net.

#### 4.3.2. Comparison experiments of DB-Net

In order to prove that the performance of DB-Net is better, DB-Net is compared to state-of-the-art segmentation networks: FCN [15], BisenetV2 [43], ResDSN [16], SegNet [22], nnU-Net [45], CE-Net [31], Att U-Net [39], mU-Net [32], U-Net [13], Swin-UNet [46], w-Net [47] and HNN [35]. To make the comparison experiment results fair, segmented results only in the first stage of DB-Net are compared to those

**Table 4**

Comparison of two-stage segmentation of abnormal pancreas (Mean  $\pm$  Standard Deviation).

Methods	Recall (%)	Precision (%)	Jaccard (%)	F1 (%)
Stage_1	85.55 $\pm$ 12.05	<b>89.20 <math>\pm</math> 5.59</b>	77.24 $\pm$ 11.05	86.63 $\pm$ 8.63
Stage_2	<b>87.42 <math>\pm</math> 11.43</b>	89.07 $\pm$ 5.74	<b>78.72 <math>\pm</math> 10.60</b>	<b>87.62 <math>\pm</math> 8.12</b>

**Table 5**

Comparison of two-stage segmentation of normal pancreas (Mean  $\pm$  Standard Deviation).

Methods	Recall (%)	Precision (%)	Jaccard (%)	F1 (%)
Stage_1	87.08 $\pm$ 10.20	89.98 $\pm$ 6.97	79.43 $\pm$ 10.91	88.07 $\pm$ 7.62
Stage_2	<b>95.50 <math>\pm</math> 8.63</b>	<b>91.47 <math>\pm</math> 8.04</b>	<b>87.74 <math>\pm</math> 10.90</b>	<b>93.05 <math>\pm</math> 7.23</b>

by using state-of-the-art segmentation networks. Quantitative comparison of abnormal pancreas are shown in Table 2. It can be seen from the table that Jaccard and F1 scores of DB-Net reach the best to 77.24, 86.63, respectively. DB-Net statistically outperforms the state-of-the-art methods with  $p < 0.05$  of wilcoxon-test for F1. In order to make an intuitive comparison between the segmented results of DB-Net and the segmented results of other networks, Fig. 9 show some abnormal pancreas segmented results. As shown in the first column in Fig. 9, the pancreas is soft and lobulated, and the pancreas is not continuous in a single CT slice. Pancreatic lobules are prone to be severely under-segmented by traditional methods, and it is difficult to segment them completely and accurately. As shown in the second column in Fig. 9, the similarity between the pancreas and the surrounding tissues is too high and pancreas region is small, causing some methods to fail to identify this region, or it is difficult to distinguish the boundary. As shown in the third column in Fig. 9, due to the similarity of the pancreas to the surrounding tissues is too high, the traditional methods of segmentation is easy to cause severe under-segmentation, and it is difficult to segment accurately. As shown in the fourth column of Fig. 9, because the head region and the tail region of the pancreas are too similar to the surrounding tissues, and the boundary is not clear, the traditional methods of segmentation will lead to serious over-segmentation. However, due to the extraction and use of semantic information and detailed information, DB-Net can greatly reduce over-segmentation and under-segmentation, so as to achieve more accurate segmentation.

In order to better verify the performance superiority of DB-Net, the model that has been trained for the abnormal pancreas is used to segment the normal pancreas. Quantitative comparison of normal pancreas are shown in Table 3. It can be seen that for the segmentation of the normal pancreas, DB-Net can also achieve good results, Recall, Precision, Jaccard and F1 scores of DB-Net reach the best to 87.08, 89.98, 79.43, 88.07. DB-Net statistically outperforms the state-of-the-art methods with  $p < 0.05$  of wilcoxon-test for F1. Similarly, the comparison of the segmented results of normal pancreas is shown in Fig. 10. As shown in the first column in Fig. 10, the intensities of the pancreas are lower than those of the general pancreas, which makes it difficult to detect or accurately segment the pancreatic region with traditional methods. As shown in the second, third and fourth column in Fig. 10, because the head region and the tail region of the pancreas are too similar to the surrounding tissues, there is no obvious boundary, the traditional method is easy to be unable to accurately discriminate the boundary. It can be seen that DB-Net still achieves good segmentation performance, and can capture the boundary very well to achieve accurate segmentation.

#### 4.4.3. Comparison experiments of the two-stage segmentation strategy

In order to prove the effectiveness of the two-stage segmentation strategy, the segmented results of the first stage are compared to the final segmented results of the two stages. The quantitative comparison

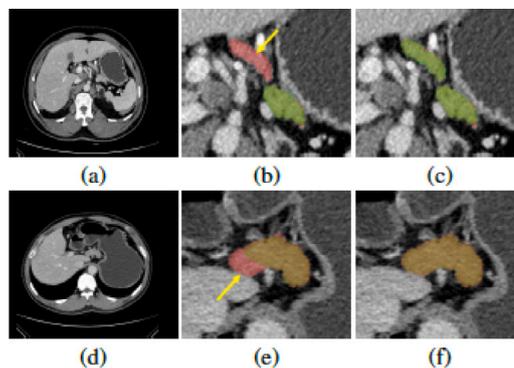


Fig. 11. Pancreas segmentation of two stages. The red mask is ground truth, and the green mask is the segmented result. The yellow arrows point out the under-segmentation. (a) Original image of abnormal pancreas, (d) Original image of normal pancreas, (b)(e) Locally enlarged pancreas segmented results of the first stage, (c)(f) Locally enlarged pancreas segmented results of the second stage.

**Table 6**

Ablation experiment results of pancreatic tumors by using CEB and RAB (Mean  $\pm$  Standard Deviation).

Methods	Recall (%)	Precision (%)	Jaccard (%)	F1 (%)
Baseline	64.38 $\pm$ 27.15	66.70 $\pm$ 25.81	44.47 $\pm$ 20.88	58.14 $\pm$ 23.40
Baseline+CEB	64.91 $\pm$ 26.26	69.60 $\pm$ 24.47	45.74 $\pm$ 20.39	59.65 $\pm$ 22.29
Baseline+RAB	<b>69.17 <math>\pm</math> 23.93</b>	64.16 $\pm$ 24.52	45.79 $\pm$ 19.18	60.15 $\pm$ 20.29
DBFE-Net	65.73 $\pm$ 25.29	<b>69.60 <math>\pm</math> 23.84</b>	<b>47.26 <math>\pm</math> 19.46</b>	<b>61.45 <math>\pm</math> 20.73</b>

**Table 7**

Comparisons of pancreatic tumor segmentation results between DBFE-Net and state-of-the-art networks (Mean  $\pm$  Standard Deviation).

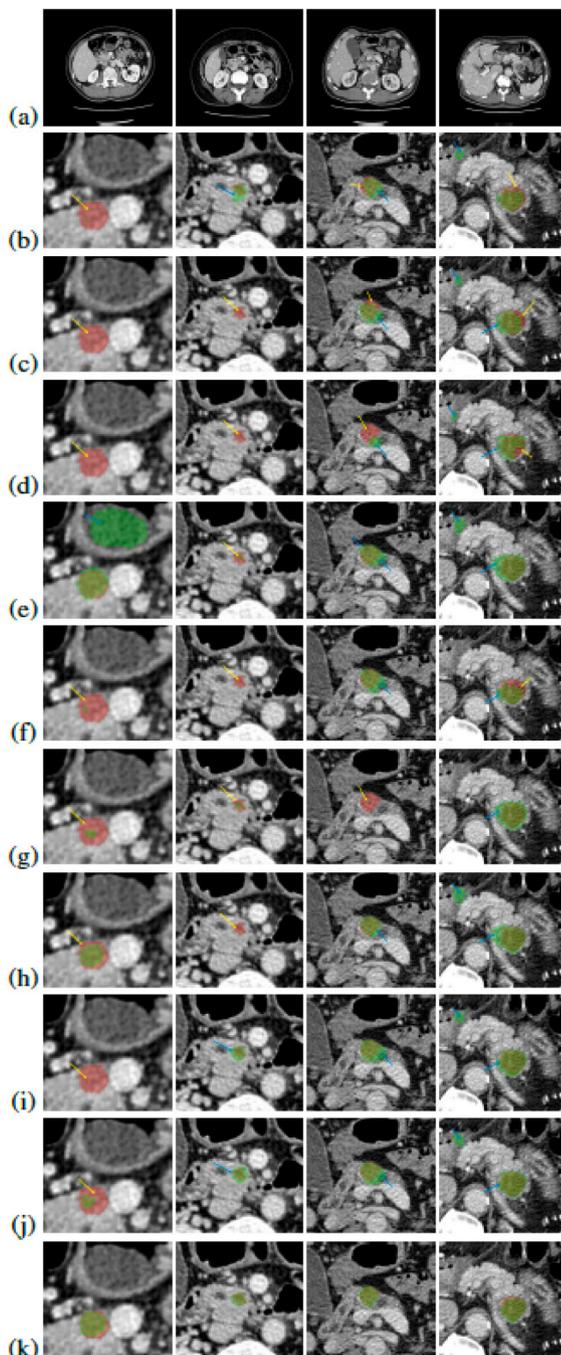
Methods	Recall (%)	Precision (%)	Jaccard (%)	F1 (%)	p
HNN [35]	64.65 $\pm$ 25.75	47.86 $\pm$ 29.33	31.87 $\pm$ 20.48	44.54 $\pm$ 24.88	$4.86 \times 10^{-15}$
FCN [40]	58.77 $\pm$ 31.81	58.23 $\pm$ 29.37	34.19 $\pm$ 21.95	46.64 $\pm$ 27.07	$2.22 \times 10^{-12}$
Swin-UNet [46]	54.58 $\pm$ 28.00	48.73 $\pm$ 23.44	34.05 $\pm$ 18.76	47.61 $\pm$ 23.24	$1.08 \times 10^{-10}$
w-Net [47]	60.76 $\pm$ 30.88	57.99 $\pm$ 27.28	37.36 $\pm$ 21.13	50.74 $\pm$ 24.23	$2.17 \times 10^{-8}$
SegNet [22]	57.51 $\pm$ 31.98	66.71 $\pm$ 27.71	38.53 $\pm$ 23.29	51.10 $\pm$ 27.56	$8.46 \times 10^{-9}$
nnU-Net [45]	65.57 $\pm$ 30.41	56.31 $\pm$ 29.93	38.81 $\pm$ 21.64	52.18 $\pm$ 24.50	$8.66 \times 10^{-6}$
Att U-Net [39]	65.06 $\pm$ 27.45	61.73 $\pm$ 26.18	42.47 $\pm$ 21.18	56.09 $\pm$ 23.53	$6.32 \times 10^{-5}$
U-Net [13]	64.38 $\pm$ 27.15	66.70 $\pm$ 25.81	44.47 $\pm$ 20.88	58.14 $\pm$ 23.40	$5.02 \times 10^{-3}$
mU-Net [32]	64.61 $\pm$ 25.86	69.60 $\pm$ 23.84	47.26 $\pm$ 19.46	59.14 $\pm$ 21.53	$4.75 \times 10^{-2}$
DBFE-Net	<b>65.73 <math>\pm</math> 25.29</b>	<b>69.60 <math>\pm</math> 23.84</b>	<b>47.26 <math>\pm</math> 19.46</b>	<b>61.45 <math>\pm</math> 20.73</b>	-

results are shown in Table 4. It can be seen from the table that the two-stage segmentation network has a better segmentation than the one-stage segmentation network. Recall, Jaccard, and F1 scores are increased by 1.87, 1.48, and 0.99, respectively. In order to better verify the superiority of the performance of the two-stage strategy, a comparison experiment with normal pancreas is also done. All the scores of Recall, Precision, Jaccard, and F1 are increased by 8.42, 1.49, 8.31 and 4.98. It shows that for normal pancreas without tumor interference, the two-stage segmentation strategy is very effective (see Table 5). Fig. 11 shows a comparison of abnormal pancreas and normal pancreas segmented results. The pancreas segmented results obtained by the two-stage segmentation network are more refined than the segmented results obtained only using the first stage. The over-segmentation and under-segmentation in the first stage can be corrected in the second stage to make the segmented result more accurate.

#### 4.4. Pancreatic tumor segmentation results

##### 4.4.1. Ablation experiments of CEB and RAB

In order to prove that CEB and RAB have a positive effect on improving the segmented results of pancreatic tumors, ablation experiments are performed on these two blocks, and performed quantitative and qualitative analysis. The baseline network is U-Net without CEB and RAB. The experimental results are shown in Table 6. As can be seen in



**Fig. 12.** Pancreatic tumor segmentation of different methods. The red mask is ground truth, and the green mask is the segmented result. The yellow arrow points out the under-segmentation and the cyan arrow points out the over-segmentation. (a) Original image, (b)–(j) Locally enlarged pancreas tumor segmented results of different methods. (b) HNN [35], (c) FCN [40], (d) Swin-UNet [46], (e) w-Net [47], (f) SegNet [22], (g) nnU-Net [45], (h) Att U-Net [39], (i) U-Net [13], (j) mU-Net [32], (k) DBFE-Net.

Table 6, compared to the baseline network, Recall, Precision, Jaccard, and F1 scores of the network with CEB are increased by 0.53, 2.9, 1.27, and 1.51, respectively. Compared to the baseline network, Recall, Jaccard, and F1 scores of the network with RAB are increased by 4.79, 1.32, and 2.01, respectively. When these two blocks are added to the network, the segmented result of the network is almost the best. CEB and RAB can help the network segment pancreatic tumors more accurate.

#### 4.4.2. Comparison experiments of DBFE-Net

DBFE-Net is also compared to state-of-the-art segmentation networks: HNN [35], FCN [40], Swin-UNet [46], w-Net [47], SegNet [22], nnU-Net [45], Att U-Net [39], U-Net [13] and mU-Net [32]. Quantitative comparisons are shown in Table 7. Recall, Precision, Jaccard, and F1 scores have reached 65.73, 69.60, 47.26, and 61.45, respectively and the results of pancreatic tumors segmented by DBFE-Net are more accurate than state-of-the-art networks. As shown in Table 7, DBFE-Net has better performance than these methods for the segmentation of pancreatic tumors with  $p < 0.05$  of wilcoxon-test for F1. Particularly, although HNN [35] outperforms other traditional methods in Jaccard and F1 for pancreas segmentation as show in Table 2, it performs poor for pancreatic tumor segmentation.

The results of different methods are visually compared, as shown in Fig. 12. As shown in Fig. 12, due to the supplement and enhancement of the details of pancreatic tumors by CEB and RAB, the results of pancreatic tumors segmented by DBFE-Net are more accurate than state-of-the-art networks. As shown in the first and the second column of Fig. 12, the tumor area is small, and the contrast with the surrounding pancreatic area is not obvious, the region is difficult to locate, and the boundary cannot be accurately segmented. For the case in the third and the fourth column of Fig. 12, the dark region is dilated pancreatic duct and its intensities are similar to those of tumors, traditional methods cannot determine the boundary well and cannot accurately segment. It can be seen that DBFE-Net can accurately locate and segment the tumor in some extreme situations such as small tumors and low contrast caused by pancreatic duct dilation.

## 5. Conclusion

In this paper, DBFE-Net is designed to segment pancreatic tumors in contrast-enhanced CT images. A dual-branch encoder–decoder network (DB-Net) as the first part of DBFE-Net is designed to segment the pancreas. The semantic branch can extract the rich semantic features of the pancreas and its surroundings through deep convolution and down-sampling operations to overcome the influence of numerous blood vessels, tissues, and organs around the pancreas, and the other branch can capture the complex pancreas through wide-channel convolution and few down-sampling operations to overcome the problem of unclear pancreatic edges. Finally, a U-Net decoder is used to obtain the accurate pancreatic segmented results. FE-Net is also designed to segment pancreatic tumors. CEB is set after the encoder layer to extract the edge detail information of the pancreas in the feature map of the encoder layer. RAB is set between the encoder and the decoder. The output of each layer of the decoder features are used to fuse the complementary areas and details of the previous layer to make up for the tumor information lost in the down-sampling. Experimental results show the proposed method can achieve better segmentation of pancreatic tumors than state-of-the-art segmentation networks.

### CRedit authorship contribution statement

**Zhibang Zhou:** Conceptualization, Methodology, Software, Validation. **Yun Bian:** Resources, Data curation, Investigation, Funding acquisition. **Shengxue Pan:** Writing – original draft, Methodology, Visualization. **Qingquan Meng:** Formal analysis, Visualization, Writing – review & editing. **Weifang Zhu:** Formal analysis, Visualization, Writing – review & editing. **Fei Shi:** Formal analysis, Visualization, Writing – review & editing. **Xinjian Chen:** Formal analysis, Visualization, Writing – review & editing. **Chengwei Shao:** Writing, Resources, Project administration, Funding acquisition. **Dehui Xiang:** Conceptualization, Writing – review & editing, Supervision, Project administration, Funding acquisition.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

The data that has been used is confidential.

## Acknowledgments

This study was supported in part by the National Nature Science Foundation of China (61971298, 81871352, 81871352, 82171915 and 82171930) and part by the National Key R&D Program of China (2018YFA0701700), supported in part by Clinical Research Plan of SHDC (SHDC2020CR4073), 234 Platform Discipline Consolidation Foundation Project, China (2019YPT001, 2020YPT001), and supported in part by The Natural Science Foundation of Shanghai Science and Technology Innovation Action Plan, China (21ZR1478500, 21Y11910300). Zhibang Zhou and Yun Bian equally contribute to this work. Dehui Xiang and Chengwei Shao are the corresponding authors.

## References

- [1] S. Kaur, M.J. Baine, M. Jain, A.R. Sasson, S.K. Batra, Early diagnosis of pancreatic cancer: Challenges and new developments, *Biomark. Med.* 6 (5) (2012) 597–612.
- [2] J. Kleff, M. Korc, M. Apte, C.L. Vecchia, C.D. Johnson, A.V. Biankin, R.E. Neale, D.A.T. Margaret Tempero, R.H. Hruban, J.P. Neoptolemos, Pancreatic cancer, *Nat. Rev. Dis. Primers* 2 (4, Sp. Iss. SI) (2016) 16023.
- [3] J. Ferlay, H.-R. Shin, F. Bray, D. Forman, C. Mathers, D.M. Parkin, Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008, *Int. J. Cancer* 127 (2010).
- [4] K. Stark, I. Schubert, U. Joshi, B. Kilani, P. Hoseinpour, M. Thakur, P. Grünauer, S. Pfeiler, T. Schmidgall, S. Stockhausen, et al., Distinct pathogenesis of pancreatic cancer microvesicle-associated venous thrombosis identifies new antithrombotic targets in Vivo, *Arterioscler. Thromb. Vasc. Biol.* 38 (4) (2018) 772–786.
- [5] J. Qiu, G. Yang, M. Feng, S. Zheng, Z. Cao, L. You, L. Zheng, T. Zhang, Y. Zhao, Extracellular vesicles as mediators of the progression and chemoresistance of pancreatic cancer and their potential clinical applications, *Mol. Cancer* 17 (1) (2018) 1–11.
- [6] J.R. Haaga, R.J. Alfidi, M.G. Zelch, T.F. Meany, M. Boller, L. Gonzalez, G.L. Jelden, Computed tomography of the pancreas, *Radiology* 120 (3) (1976) 589–595.
- [7] M. Xu, X. Jung, O.J. Hines, G. Eibl, Y. Chen, Obesity and pancreatic cancer: Overview of epidemiology and potential prevention by weight loss, *Pancreas* 47 (2) (2018) 158.
- [8] Y. Man, Y. Huang, J. Feng, X. Li, F. Wu, Deep q learning driven ct pancreas segmentation with geometry-aware u-net, *IEEE Trans. Med. Imaging* 38 (8) (2019) 1971–1980.
- [9] K. Dmitriev, I. Gutenko, S. Nadeem, A. Kaufman, Pancreas and cyst segmentation, in: *Medical Imaging 2016: Image Processing*, Vol. 9784, International Society for Optics and Photonics, 2016, p. 97842C.
- [10] K.M. Horton, E.K. Fishman, Adenocarcinoma of the pancreas: CT imaging, *Radiol. Clin.* 40 (6) (2002) 1263–1272.
- [11] M. Müller, C. Meyenberger, P. Bertschinger, R. Schaer, B. Marincek, Pancreatic tumors: Evaluation with endoscopic US, CT, and MR imaging, *Radiology* 190 (3) (1994) 745–751.
- [12] Z. Guo, L. Zhang, L. Lu, M. Bagheri, R.M. Summers, M. Sonka, J. Yao, Deep LOGISMOS: Deep learning graph-based 3D segmentation of pancreatic tumors on CT scans, in: *2018 IEEE 15th International Symposium on Biomedical Imaging, ISBI 2018, IEEE*, 2018, pp. 1230–1233.
- [13] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2015, pp. 234–241.
- [14] S. Pan, D. Xiang, Y. Bian, J. Lu, H. Jiang, J. Zheng, Automatic pancreas segmentation in abdominal CT image with contrast enhancement block, in: *Medical Imaging 2021: Image Processing*, Vol. 11596, International Society for Optics and Photonics, 2021, p. 115961B.
- [15] Y. Zhou, L. Xie, W. Shen, Y. Wang, E.K. Fishman, A.L. Yuille, A fixed-point model for pancreas segmentation in abdominal CT scans, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2017, pp. 693–701.
- [16] Z. Zhu, Y. Xia, W. Shen, E. Fishman, A. Yuille, A 3D coarse-to-fine framework for volumetric medical image segmentation, in: *2018 International Conference on 3D Vision, 3DV, IEEE*, 2018, pp. 682–690.
- [17] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, *Adv. Neural Inf. Process. Syst.* 25 (2012) 1097–1105.
- [18] W. Rawat, Z. Wang, Deep convolutional neural networks for image classification: A comprehensive review, *Neural Comput.* 29 (9) (2017) 2352–2449.
- [19] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, 2014, arXiv preprint arXiv:1409.1556.
- [20] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille, Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs, *IEEE Trans. Pattern Anal. Mach. Intell.* 40 (4) (2017) 834–848.
- [21] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3431–3440.
- [22] V. Badrinarayanan, A. Kendall, R. Cipolla, Segnet: A deep convolutional encoder-decoder architecture for image segmentation, *IEEE Trans. Pattern Anal. Mach. Intell.* 39 (12) (2017) 2481–2495.
- [23] W. Shen, X. Wang, Y. Wang, X. Bai, Z. Zhang, Deepcontour: A deep convolutional feature learned by positive-sharing loss for contour detection, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3982–3991.
- [24] W. Shen, B. Wang, Y. Jiang, Y. Wang, A. Yuille, Multi-stage multi-recursive-input fully convolutional networks for neuronal boundary detection, in: *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2391–2400.
- [25] Y. Liu, M.-M. Cheng, X. Hu, K. Wang, X. Bai, Richer convolutional features for edge detection, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 3000–3009.
- [26] J. Xie, L. Xu, E. Chen, Image denoising and inpainting with deep neural networks, *Adv. Neural Inf. Process. Syst.* 25 (2012) 341–349.
- [27] K. Zhang, W. Zuo, Y. Chen, D. Meng, L. Zhang, Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising, *IEEE Trans. Image Process.* 26 (7) (2017) 3142–3155.
- [28] J. Xu, L. Zhang, D. Zhang, X. Feng, Multi-channel weighted nuclear norm minimization for real color image denoising, in: *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 1096–1104.
- [29] F. Milletari, N. Navab, S.-A. Ahmadi, V-net: Fully convolutional neural networks for volumetric medical image segmentation, in: *2016 Fourth International Conference on 3D Vision, 3DV, IEEE*, 2016, pp. 565–571.
- [30] M.H. Hesamian, W. Jia, X. He, P. Kennedy, Deep learning techniques for medical image segmentation: Achievements and challenges, *J. Digit. Imaging* 32 (4) (2019) 582–596.
- [31] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, J. Liu, Ce-net: Context encoder network for 2d medical image segmentation, *IEEE Trans. Med. Imaging* 38 (10) (2019) 2281–2292.
- [32] H. Seo, C. Huang, M. Bassenne, R. Xiao, L. Xing, Modified U-net (mU-Net) with incorporation of object-dependent high level features for improved liver and liver-tumor segmentation in CT images, *IEEE Trans. Med. Imaging* 39 (5) (2019) 1316–1325.
- [33] H.R. Roth, L. Lu, A. Farag, A. Sohn, R.M. Summers, Spatial aggregation of holistically-nested networks for automated pancreas segmentation, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2016, pp. 451–459.
- [34] J. Ma, F. Lin, S. Wesarg, M. Erdt, A novel bayesian model incorporating deep neural network and statistical shape model for pancreas segmentation, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2018, pp. 480–487.
- [35] H.R. Roth, L. Lu, N. Lay, A.P. Harrison, A. Farag, A. Sohn, R.M. Summers, Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation, *Med. Image Anal.* 45 (2018) 94–107.
- [36] Y. Zhou, L. Xie, W. Shen, Y. Wang, E.K. Fishman, A.L. Yuille, A fixed-point model for pancreas segmentation in abdominal CT scans, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2017, pp. 693–701.
- [37] Y. Man, Y. Huang, J. Feng, X. Li, F. Wu, Deep q learning driven ct pancreas segmentation with geometry-aware u-net, *IEEE Trans. Med. Imaging* 38 (8) (2019) 1971–1980.
- [38] J. Cai, L. Lu, F. Xing, L. Yang, Pancreas segmentation in CT and MRI via task-specific network design and recurrent neural contextual learning, in: *Deep Learning and Convolutional Neural Networks for Medical Imaging and Clinical Informatics*, Springer, 2019, pp. 3–21.
- [39] O. Oktay, J. Schlemper, L.L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N.Y. Hammerla, B. Kainz, et al., Attention u-net: Learning where to look for the pancreas, 2018, arXiv preprint arXiv:1804.03999.
- [40] Y. Zhou, L. Xie, E.K. Fishman, A.L. Yuille, Deep supervision for pancreatic cyst segmentation in abdominal Ct scans, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2017, pp. 222–230.

- [41] Y. Zhou, Y. Li, Z. Zhang, Y. Wang, A. Wang, E.K. Fishman, A.L. Yuille, S. Park, Hyper-pairing network for multi-phase pancreatic ductal adenocarcinoma segmentation, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2019, pp. 155–163.
- [42] Z. Zhu, Y. Xia, L. Xie, E.K. Fishman, A.L. Yuille, Multi-scale coarse-to-fine segmentation for screening pancreatic ductal adenocarcinoma, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2019, pp. 3–12.
- [43] C. Yu, C. Gao, J. Wang, G. Yu, C. Shen, N. Sang, Bisenet v2: Bilateral network with guided aggregation for real-time semantic segmentation, 2020, arXiv preprint [arXiv:2004.02147](https://arxiv.org/abs/2004.02147).
- [44] S. Gao, M.-M. Cheng, K. Zhao, X.-Y. Zhang, M.-H. Yang, P.H. Torr, Res2net: A new multi-scale backbone architecture, *IEEE Trans. Pattern Anal. Mach. Intell.* (2019).
- [45] F. Isensee, P.F. Jaeger, S.A. Kohl, J. Petersen, K.H. Maier-Hein, nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation, *Nature Methods* 18 (2) (2021) 203–211.
- [46] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, M. Wang, Swin-Unet: Unet-like pure transformer for medical image segmentation, in: *Proceedings of the European Conference on Computer Vision Workshops, ECCVW, 2022*.
- [47] B. Wang, L. Wang, J. Chen, Z. Xu, T. Lukasiewicz, Z. Fu, w-Net: Dual supervised medical image segmentation model with multi-dimensional attention and cascade multi-scale convolution, *Neurocomputing* (2022).