

PAPER

CTG-Net: an efficient cascaded framework driven by terminal guidance mechanism for dilated pancreatic duct segmentation

To cite this article: Liwen Zou *et al* 2023 *Phys. Med. Biol.* **68** 215006

View the [article online](#) for updates and enhancements.

You may also like

- [Increase in breath hydrogen concentration was correlated with the main pancreatic duct stenosis](#)

Daisuke Sakai, Yoshiki Hirooka, Hiroki Kawashima et al.

- [Development of a high-throughput micropatterned agarose scaffold for consistent and reproducible hPSC-derived liver organoids](#)

Shanqing Jiang, Fang Xu, Menglong Jin et al.

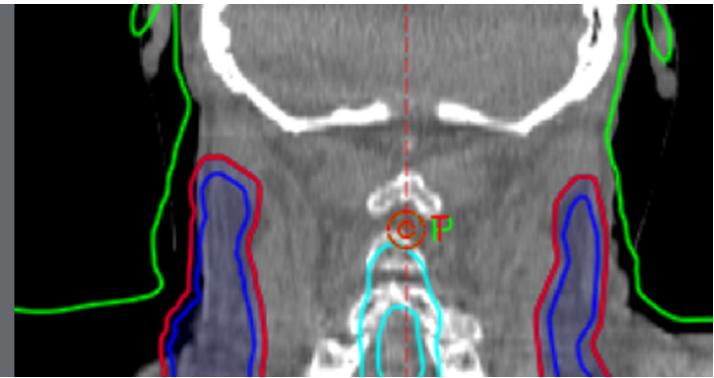
- [High-speed atomic force microscopy coming of age](#)

Toshio Ando

Rethink re-plans.

See how SunCHECK® automates
in-vivo monitoring.

ASTRO Booth #1835



**PAPER****RECEIVED**
24 May 2023**REVISED**
29 July 2023**ACCEPTED FOR PUBLICATION**
16 August 2023**PUBLISHED**
23 October 2023

CTG-Net: an efficient cascaded framework driven by terminal guidance mechanism for dilated pancreatic duct segmentation

Liwen Zou¹ , Zhenghua Cai², Yudong Qiu³, Luying Gui⁴, Liang Mao^{3,*} and Xiaoping Yang^{1,*}¹ Department of Mathematics, Nanjing University, Nanjing, 210093, People's Republic of China² Medical School, Nanjing University, Nanjing, 210007, People's Republic of China³ Department of General Surgery, Nanjing Drum Tower Hospital, Nanjing, 210008, People's Republic of China⁴ School of Mathematics and Statistics, Nanjing University of Science and Technology, Nanjing, 210094, People's Republic of China

* Authors to whom any correspondence should be addressed.

E-mail: maoliang@njglyy.com and xpyang@nju.edu.cn**Keywords:** pancreatic duct dilation, terminal guidance, cascaded strategy, distance transform, distraction attention, medical image segmentation**Abstract**

Pancreatic duct dilation indicates a high risk of various pancreatic diseases. Segmentation for dilated pancreatic duct (DPD) on computed tomography (CT) image shows the potential to assist the early diagnosis, surgical planning and prognosis. Because of the DPD's tiny size, slender tubular structure and the surrounding distractions, most current researches on DPD segmentation achieve low accuracy and always have segmentation errors on the terminal DPD regions. To address these problems, we propose a cascaded terminal guidance network to efficiently improve the DPD segmentation performance. Firstly, a basic cascaded segmentation architecture is established to get the pancreas and coarse DPD segmentation, a DPD graph structure is build on the coarse DPD segmentation to locate the terminal DPD regions. Then, a terminal anatomy attention module is introduced for jointly learning the local intensity from the CT images, feature cues from the coarse DPD segmentation and global anatomy information from the designed pancreas anatomy-aware maps. Finally, a terminal distraction attention module which explicitly learns the distribution of the terminal distraction regions is proposed to reduce the false positive and false negative predictions. We also propose a new metric called tDice to measure the terminal segmentation accuracy for targets with tubular structures and two other metrics for segmentation error evaluation. We collect our dilated pancreatic duct segmentation dataset with 150 CT scans from patients with five types of pancreatic tumors. Experimental results on our dataset show that our proposed approach boosts DPD segmentation accuracy by nearly 20% compared with the existing results, and achieves more than 9% improvement for the terminal segmentation accuracy compared with the state-of-the-art methods.

1. Introduction

Pancreatic cancer is one of the most intractable cancers with only 10% of five-year relative survival rate in the USA. The most common pancreatic cancer occurs in the main pancreatic duct, known as pancreatic ductal adenocarcinoma (PDAC) (Mizrahi *et al* 2020). Pancreatic duct dilation is identified as a high risk of PDAC in several clinical studies (Tanaka *et al* 2002, Edge *et al* 2007). However, the pancreatic disease taxonomy related to dilated pancreatic duct (DPD) includes other types of tumors such as intraductal papillary mucinous neoplasm (IPMN), serous cystic intraductal papillary mucinous neoplasm (SCN), intraductal papillary mucinous neoplasm (MCN) and solid pseudopapillary tumor (SPT) (Springer *et al* 2019). A comparison of pancreatic duct dilations with five different pancreatic tumors on computed tomography (CT) images are shown in figure 1. Because of the slender shape and tiny size of the pancreatic duct, the normal pancreatic duct region is almost invisible on contrast-enhanced abdominal CT. Visibility of pancreatic duct from CT image could be a warning

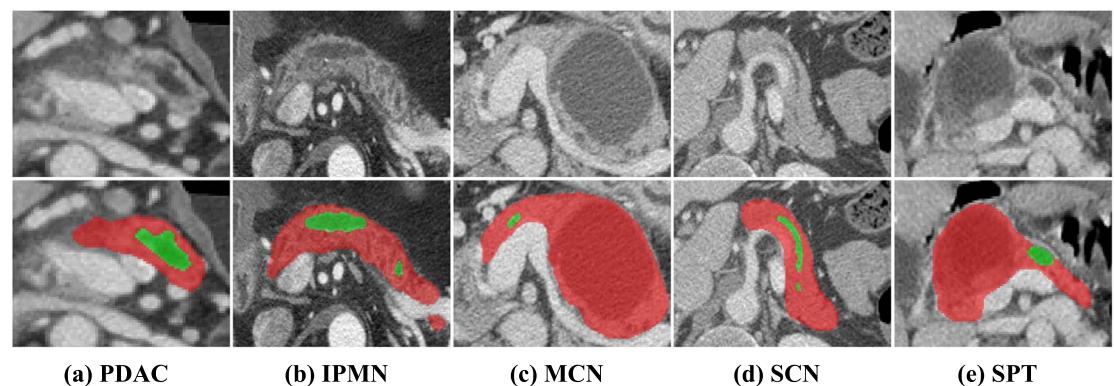


Figure 1. Visual comparison of dilated pancreatic ducts (DPDs) on CT images of patients with five types of pancreatic tumors. Top: the CT images. Bottom: the corresponding masks where red denotes the pancreas and green denotes the DPD.

sign for PDAC and other pancreatic tumors. Hence, automated segmentation for DPD on CT images is promising for the early diagnosis, surgical planning and prognosis of pancreatic tumors.

The pancreas has a small size in the abdomen CT scan, therefore, a pancreatic duct inside the pancreas has a much smaller size, which makes the segmentation quite difficult. Additionally, the intensity similarity, slender shape, and distractions from surrounding tissues and lesions usually lead to the false positive (FP) or false negative (FN) predictions outside the pancreas or on the terminal regions of the DPD predictions. The general experience motivating the solution is that when radiologists look for the DPDs, they usually concentrate on the pancreas regions and ignore other regions' contexts. This suggests that a deep neural network should also mainly concentrate on the pancreas when it is designed for DPD segmentation. Therefore, the cascaded strategy of sequentially segmenting pancreas and DPD shows better performance than directly segmenting the DPD on the whole CT scans (Shen *et al* 2021a). However, previous methods for DPD segmentation still cannot achieve satisfactory performance mostly because of the DPD's slender shape and confusing tissues such as bile ducts and tumors around the DPDs. In fact, since the DPDs have tubular structures, almost all of the errors are located in the terminal regions of the predictions.

To address this problem, we propose a novel cascaded terminal guidance network (CTG-Net) to explicitly take the terminal guidance into account for DPD segmentation. Our CTG-Net includes three stages: coarse, fine and refine stage. The coarse stage uses a basic cascaded segmentation framework composed of a basic pancreas segmentation network (BPSN) and a basic duct segmentation network (BDSN) to sequentially segment the pancreas and DPD. A DPD graph structure is build on the coarse DPD segmentation to locate the terminal DPD regions. In the fine stage, to focus on the terminal errors of the coarse predictions, we introduce a terminal anatomy attention module (TAAM) to jointly learn the local intensity from the terminal CT images, feature cues from the coarse predictions and global anatomy information from our designed pancreas anatomy-aware (PAA) maps. In the refine stage, we propose a terminal distraction attention module (TDAM) to learn the terminal distraction information and distinguish the false positive and false negative regions to improve the fine predictions. We collect our dilated pancreatic duct segmentation (IDPDSeg) dataset with 150 CT scans from patients with five types of pancreatic tumors. Experimental results on our dataset show the superiority to the previous techniques especially for the terminal segmentation accuracy. Figure 2 shows two experimental examples to illustrate the benefits of the proposed CTG-Net. Moreover, we propose a new metric, called terminal Dice (tDice), to measure the terminal segmentation accuracy for targets with tubular structures, and two other metrics for segmentation error evaluation.

The main contributions of our work are summarized as follows.

- To the best of our knowledge, this is the first work to consider the terminal errors caused by general deep learning-based segmentation methods for targets with tubular structures, and we propose a novel cascaded segmentation framework driven by terminal guidance mechanism for the dilated pancreatic duct (DPD) segmentation to solve the problem of terminal errors.
- We design a TAAM, which focuses on the terminal regions of the DPD and takes the local intensity, feature cues of the targets and global anatomy information into account to improve the terminal segmentation accuracy of DPDs.
- We design a TDAM which explicitly learns the distribution of the distraction tissues within the terminal regions of DPD to reduce the false positive and false negative predictions.

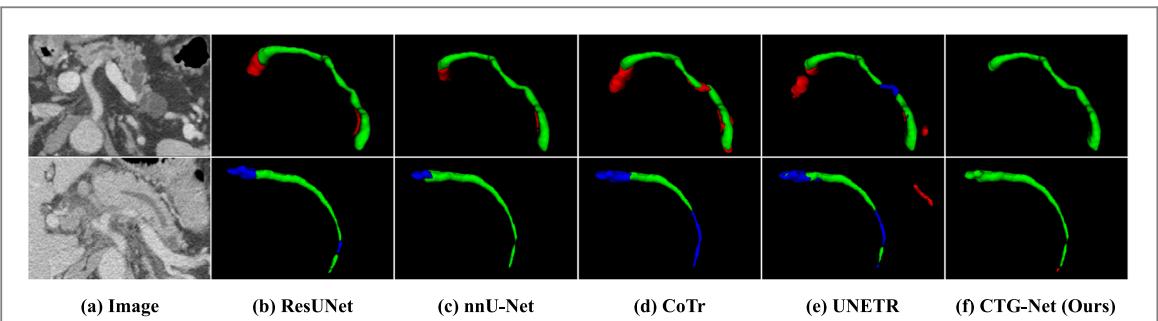


Figure 2. Two examples to illustrate the benefits of the proposed CTG-Net. (a) CT images and the corresponding predictions between (b) ResUNet, (c) nnU-Net, (d) CoTr, (e) UNETR and (f) proposed CTG-Net. The green, red and blue voxels denote the TP, FP and FN segmentation, respectively. It can be found that the result of CTG-Net has fewer terminal segmentation errors compared with the results of these previous methods.

- We propose a new metric called tDice to measure the terminal segmentation accuracy for targets with tubular structures and two other metrics for segmentation error evaluation.
- Our proposed DPDSeg dataset includes not only the PDAC patients but also the cases with other four types of pancreatic tumors. Experimental results on our DPDSeg dataset show that our proposed approach boosts DPD segmentation accuracy by nearly 20% compared with the existing results, and achieves more than 9% improvement for the terminal segmentation accuracy compared with the state-of-the-art (SOTA) methods.

2. Related work

2.1. Deep learning for medical image analysis

Deep learning-based methods have successfully solved many challenging tasks in image processing, such as classification (He *et al* 2016), semantic segmentation (Ma *et al* 2020, 2021), object detection (Zou *et al* 2022), and image quality improvement (Higaki *et al* 2019). With the rapid development of the fully convolutional network (FCN) (Long *et al* 2015), its extensions (Ronneberger *et al* 2015, Milletari *et al* 2016, Oktay *et al* 2018, Isensee *et al* 2021) are widely applied in the segmentation task during medical image analysis. Encoder and decoder structures greatly improve the performance on pixel-wise segmentation compared with many traditional methods which need designing manual features. Semantic segmentation for medical images is important for developing computer-aided diagnosis (CAD) and computer-assisted surgery (CAS) systems.

2.2. Previous techniques for DPD segmentation

Only few studies are related to DPD segmentation (Zhou *et al* 2019, Xia *et al* 2020, Shen *et al* 2021a, 2021b). Zhou *et al* (2019) investigated a dual-path network for pancreas, PDAC tumor, and DPD segmentation. Both arterial and venous phase CT volumes are required to conduct the segmentation processes. They achieved a Dice score of 56.77% on their collected PDAC dataset including 239 cases with DPDs. Xia *et al* (2020) proposed a segmentation method based on multi-phase CT alignment that achieved better performance of 64.38% Dice score on the same dataset. Shen *et al* (2021a) used a cascaded segmentation framework with a pancreas segmentation network to get the pancreas region of interest (ROI) and segment the DPD with the following U-like network with bottleneck attention block, and they got a Dice score of 49.87% on their collected PDAC dataset including 30 cases with DPDs. Shen *et al* (2021b) embedded pancreas-guided attention mechanism to the segmentation network and improved the segmentation accuracy to 54.16% Dice score. However, almost none of the previous techniques have paid attention to the terminal segmentation errors caused by the tiny size and slender shape of the DPDs. Actually, the low accuracy for DPD segmentation can significantly affect the clinical diagnosis and treatment of the patients with pancreatic tumors. Additionally, almost all of previous studies focused on DPD segmentation for PDAC patients, segmentation DPD for patients with other pancreatic tumors needs to be addressed.

2.3. Cascaded strategy for medical image segmentation

Multiple applications of cascaded methods are used for medical image segmentation. Chang and Teng (2007) presented a two-stage self-organizing map approach, which can precisely identify dominant color components to discover the region of interest for diagnosis purposes. Roth *et al* (2018) claimed a two-stage, coarse-to-fine approach for three anatomical organ segmentation (liver, spleen, pancreas). They firstly used a 3D FCN to roughly define a candidate region and then fed the region to the second 3D FCN, which focuses on more detailed

segmentation of the organs and vessels. Zhang *et al* (2021) introduced a cascaded framework combining with deep neural networks, multi-atlas registration and 3D level-set for pancreas segmentation. They firstly obtained a coarse segmentation through multi-atlas based 3D diffeomorphic registration and fusion. Then, a 3D patch-based convolutional neural network (CNN) and three 2D slice-based CNNs are jointly used to predict a fine segmentation based on a bounding box determined from the coarse segmentation. Finally, a 3D level-set method is used, with the fine segmentation being one of its constraints, to integrate information of the original image and the CNN-derived probability map to achieve a refine segmentation. For segmentation of lesions within organs, the cascaded strategy is an effective framework to create an attention mechanism to guide the model where it should focus.

2.4. Distance transform for tubular structure segmentation

Distance transform (Rosenfeld and Pfaltz 1968) is a classical image processing operator to produce a distance map with the same size of the input image, each value on the pixel is the distance from the foreground pixel to the foreground boundary. Distance transform is also known as the basis of one type of skeletonization algorithms (Ge and Fitzpatrick 1996). Thus, the distance map encodes the geometric characteristics of the tubular structure. Wang *et al* (2020) proposed a geometry-aware tubular structure segmentation method which combines intuitions from the classical distance transform for skeletonization and modern deep segmentation networks. In Ma *et al* (2020), the signed distance function was used to measure the small segmentation errors, and a corresponding loss function was designed based on the signed distance function to train the segmentation network. Inspired by these works above, we observe that the distance transform map of the binary pancreas mask can effectively reflect the anatomy structure of the pancreas. Each value on the pixel of the distance transform map shows the distance to the boundary of the pancreas, which can provide more anatomy details inside the pancreas for the DPD segmentation.

2.5. Distraction attention mechanism

Distraction concepts have been explored in many computer vision tasks, such as semantic segmentation (Huang *et al* 2017), saliency detection (Xiao *et al* 2018, Chen *et al* 2020) and visual tracking (Zhu *et al* 2018). Most of previous methods suppress negative high-level representations or filter out the distracting regions (Huang *et al* 2017, Xiao *et al* 2018, Chen *et al* 2020). Instead, Zheng *et al* (2019) used distraction cues to improve the performance of shadow distraction. They split shadow distraction into false negative estimates and false positive estimates, and designed specific architectures to efficiently integrate the two types of distraction semantics. Zhao *et al* (2022) proposed a cascaded two-stage U-Net model to explicitly take the ambiguous region information into account. In their model, the first stage generates a global segmentation for the whole input CT volume and predicts latent distraction regions, which contain both false negative areas and false positive areas, against the segmentation ground truth. The second stage embeds the distraction region information into local segmentation for volume patches to further discriminate the target regions. Inspired by this method, we develop a TDAM to explicitly learn semantic features of the terminal distraction regions, and embed these features in an encoder-decoder structure to reduce the false positive and false negative predictions in the DPD terminal regions.

3. Methods

3.1. Overview

The overview of our proposed CTG-Net is illustrated in figure 3. The whole network is mainly divided into three stages: (1) coarse stage with a basic cascaded segmentation architecture to get the pancreas segmentation and coarse DPD segmentation; (2) fine stage with our designed TAAM to improve the terminal segmentation performance of DPD by jointly learning the local intensity from the CT image, feature cues from coarse DPD segmentation and global anatomy information from the designed PAA map; (3) refine stage with the proposed TDAM to reduce the false positive and false negative predictions by explicitly learning the distribution of the terminal distraction regions. The goal of our proposed CTG-Net is to improve the segmentation performance of DPD by establishing the above two terminal attention mechanisms, TAAM and TDAM, which embeds anatomy guidance and distraction guidance, respectively.

3.2. Coarse stage with a basic cascaded segmentation architecture

The coarse stage is composed of a basic pancreas segmentation network (BPSN) for segmenting the pancreas from the CT image and obtaining the pancreas region of interest (ROI), and a basic duct segmentation network (BDSN) for obtaining the coarse DPD segmentation from the pancreas ROI.

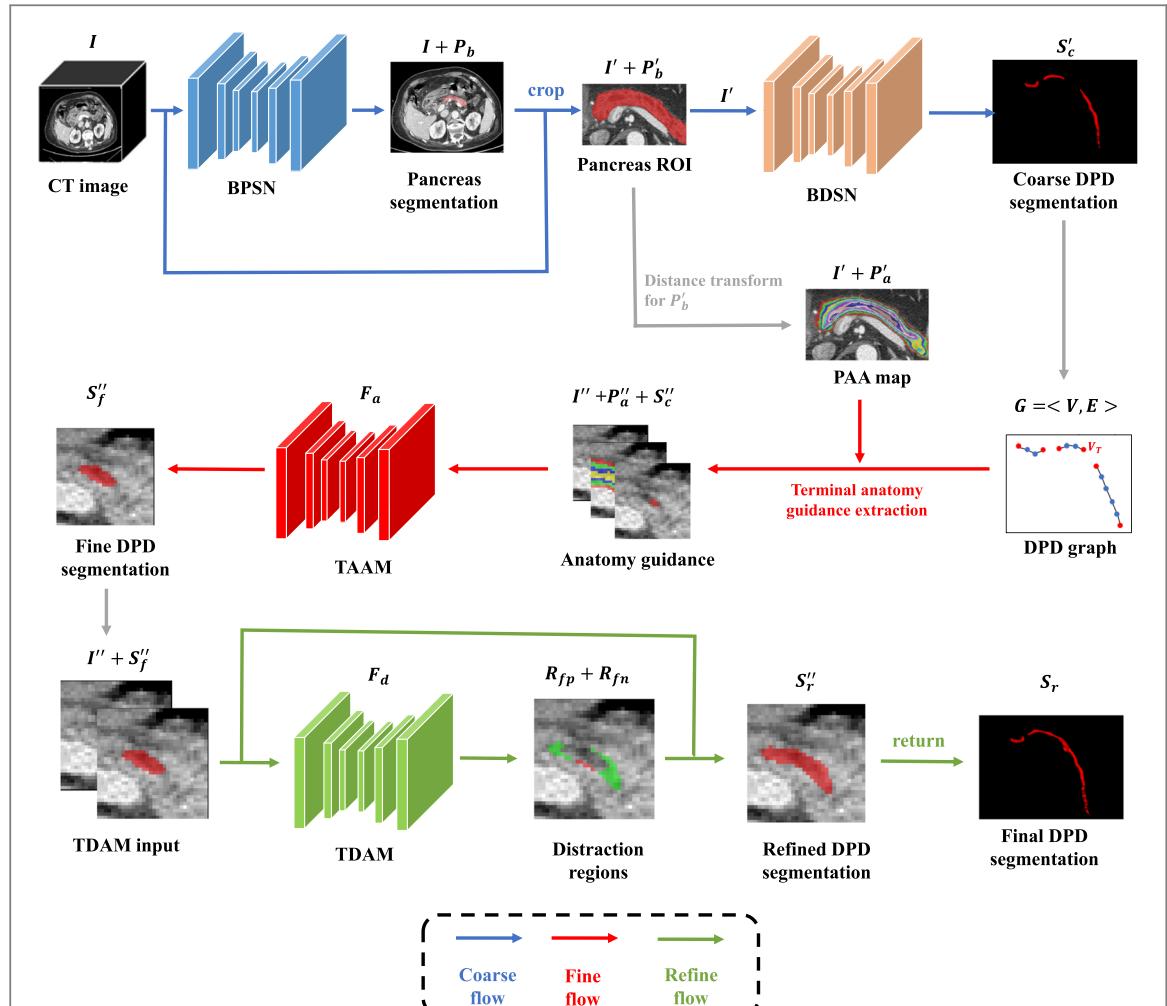


Figure 3. Pipeline of our proposed CTG-Net for dilated pancreatic duct (DPD) segmentation. The coarse stage composed of the BPSN and BDSN gets the pancreas segmentation and coarse DPD segmentation. The fine stage with our designed TAAM learns from the terminal anatomy guidance to obtain the fine DPD segmentation. The refine stage with the proposed TDAM learns the distribution of the terminal distraction regions to reduce the false positive and false negative predictions.

Mathematically, given a CT image $I \in \mathbb{R}^{W \times H \times L}$, the BPSN firstly takes I as the input and outputs the binary pancreas segmentation $P_b \in \{0, 1\}^{W \times H \times L}$. Then, the pancreas ROI is obtained by cropping the original CT image I according to the bounding box of the pancreas segmentation P_b . To differentiate from the original image I and pancreas segmentation P_b , we mark the pancreas ROI image and pancreas segmentation as $I' \in \mathbb{R}^{W_1 \times H_1 \times L_1}$ and $P'_b \in \{0, 1\}^{W_1 \times H_1 \times L_1}$, respectively. The following BDSN takes the pancreas ROI image I' as input and outputs the coarse DPD segmentation $S'_c \in \{0, 1\}^{W_1 \times H_1 \times L_1}$.

The coarse DPD segmentation S'_c of the pancreas ROI level can be returned to be the coarse DPD segmentation of the original level $S_c \in \{0, 1\}^{W \times H \times L}$ as follows:

$$S_c(x) = \begin{cases} S'_c(x), & P_b(x) = 1 \\ 0, & P_b(x) = 0 \end{cases}. \quad (1)$$

3.3. Fine stage with TAAM

In the fine stage, a DPD graph structure $G = \langle V, E \rangle$ is built on the coarse DPD segmentation S'_c to get the positions of the terminal regions, where V and E are the vertexes and edges of the graph G . We firstly extract the centerline of S'_c using the method proposed in Lee *et al* (1994) to get V which is the set of points on the centerline of S'_c . The neighbor relationships between these points in V are the edges E of the graph G . The terminal vertexes V_T can be automatically found based on the law that they only have one adjacent vertex. Once we obtain the set of the terminal vertexes V_T , we can crop the terminal ROI cubes centered on the terminal vertexes with the size of $W_2 \times H_2 \times L_2$. The cube can be the CT image I , coarse DPD segmentation S'_c and the following designed PAA map.

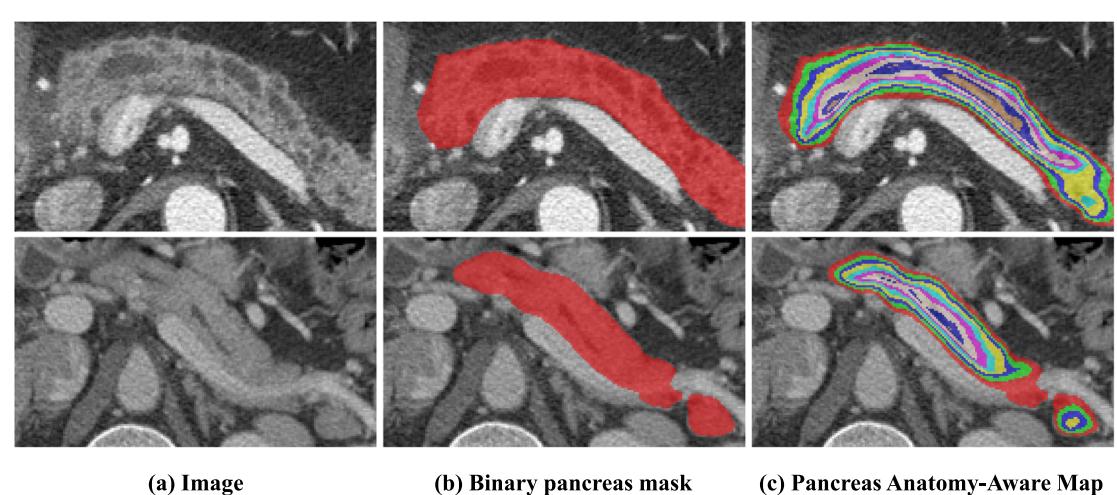


Figure 4. Visual comparison between the (b) binary pancreas masks and (c) our designed pancreas anatomy-aware (PAA) maps on (a) two CT images in our DPDSeg dataset. In the PAA maps, different colors denote different distances to the boundary of the binary pancreas masks.

For DPD segmentation with attention mechanism, Shen *et al* (2021b) adopted the binary pancreas segmentation as the attention source to guide the segmentation of DPD. However, the binary pancreas mask only provides the information of the pancreas location without the structure hints inside the pancreas. In order to get the global anatomy information to improve the terminal segmentation performance of DPD, we calculate the distance transform map of the pancreas segmentation P'_b as the PAA map $P'_a \in \mathbb{R}^{W_1 \times H_1 \times L_1}$ to guide the segmentation in DPD terminal regions, which can be calculated as follows:

$$P'_a(x) = \begin{cases} \min_{y \in \partial P'_b} d(x, y), & P'_b(x) = 1 \\ 0, & \text{otherwise} \end{cases}, \quad (2)$$

where $\partial P'_b$ denotes the boundary of P'_b , and $d(x, y)$ is the Euclidean distance between x and y . Figure 4 shows the difference between the binary pancreas mask and the PAA map. Then, the terminal anatomy guidance is extracted from the terminal CT image I' , coarse DPD segmentation S'_c and the calculated PAA map P'_a . All the three cubes are cropped centered on the terminal vertexes V_T of the DPD graph G to be the terminal ROI image $I'' \in \mathbb{R}^{W_2 \times H_2 \times L_2}$, coarse DPD segmentation $S''_c \in \{0, 1\}^{W_2 \times H_2 \times L_2}$ and PAA map $P''_a \in \mathbb{R}^{W_2 \times H_2 \times L_2}$, respectively.

The TAAM parameterized by F_a is designed with a two-channel encoder-decoder architecture, it concatenates the extracted terminal anatomy guidance as the network input and outputs the fine DPD segmentation $S''_f \in \{0, 1\}^{W_2 \times H_2 \times L_2}$, which can be formulated as follows:

$$S''_f = F_a(I'', P''_a, S''_c; \theta_a), \quad (3)$$

where θ_a is the parameter of F_a to be trained. The fine DPD segmentation S_f'' of the terminal ROI level can be returned to be the fine DPD segmentation of the original level $S_f \in \{0, 1\}^{W \times H \times L}$ as follows:

$$S_f(x) = \begin{cases} 0, & P_b(x) = 0 \\ S''_f(x), & P_b(x) = 1 \text{ and } x \in C, \\ S_c(x), & \text{otherwise} \end{cases} \quad (4)$$

where C denotes the set of locations in the terminal ROI cubes.

3.4. Refine stage with TDAM

In the refine stage, we introduce the TDAM parameterized by F_d which aims to explicitly learn the distribution of the terminal distraction regions. The TDAM F_d takes the terminal CT image I'' and the fine DPD segmentation S_f'' as the network input and learns to find the distraction regions (i.e. the false positive predictions $R_{fp} \in \{0, 1\}^{W_2 \times H_2 \times L_2}$ and the false negative predictions $R_{fn} \in \{0, 1\}^{W_2 \times H_2 \times L_2}$) and make corresponding refinement on the fine DPD segmentation to obtain the refined DPD segmentation $S_r'' \in \{0, 1\}^{W_2 \times H_2 \times L_2}$, which can be formulated as follows:

$$R_{fp}, R_{fn} = F_d(I'', S_f''; \theta_d), \quad (5)$$

$$S_r'' = S_f'' + R_{fn} - R_{fp}, \quad (6)$$

where θ_d is the parameter of F_d to be trained. The refined DPD segmentation S_r'' of the terminal ROI level can also be returned to be the refined DPD segmentation of the original level $S_r \in \{0, 1\}^{W \times H \times L}$ as follows:

$$S_r(x) = \begin{cases} 0, & P_b(x) = 0 \\ S_r''(x), & P_b(x) = 1 \text{ and } x \in C. \\ S_f(x). & \text{otherwise} \end{cases} \quad (7)$$

4. Experiments

4.1. Dataset

In this work, we evaluate our proposed methods on our DPDSeg dataset collected from Nanjing Drum Tower Hospital. It consists of 150 patients with surgical pathology-confirmed pancreatic tumors (103 PDACs, 32 IPMNs, 7 SCNs, 5 SPTs, 3 MCNs) and DPDs. Each patient has a venous phase abdominal CT scan and each CT volume contains 165–789 slices with size of 512×512 . The manual annotations of DPD were performed by two experienced pancreatic imaging radiologists. The DPD voxels labeled by the experts range from 238 to 49490. Our DPDSeg dataset is randomly divided into five folds, and all the experiments in this work are performed using five-fold cross validation. We fixed all parameters and reported average results of five folds.

To verify the generalization ability of our proposed method, we collect another dataset containing CT images of 40 PDAC patients from Jiangsu Province Hospital of Chinese Medicine (JPHCM) for testing. We term it as JPHCM dataset. Each patient has a venous phase abdominal CT scan and each CT volume contains 174–691 slices with size of 512×512 . The DPD voxels labeled by the experts range from 272 to 35788.

4.2. Evaluation metrics

For quantitative evaluation of the segmentation performance of DPD, we applied two commonly used evaluation metrics, namely the Dice similarity coefficient (Dice) and Hausdorff distance (HD), which can be calculated as follows

$$\text{Dice}(X, Y) = \frac{2|X \cap Y|}{|X| + |Y|} \quad (8)$$

$$h(X, Y) = \max_{x \in X} \min_{y \in Y} d(x, y) \quad (9)$$

$$HD(X, Y) = \max(h(X, Y), h(Y, X)), \quad (10)$$

where X and Y denote the prediction mask and ground truth mask, respectively. In addition to the above two traditional metrics, we propose two other metrics for segmentation error evaluation. We define the false positive segmentation rate (FPSR) and false negative segmentation rate (FNSR) as follows:

$$\text{FPSR}(X, Y) = \frac{|X - Y|}{|X| + |Y|}, \quad (11)$$

$$\text{FNSR}(X, Y) = \frac{|Y - X|}{|X| + |Y|}. \quad (12)$$

It can be found that there is $\text{Dice} + \text{FPSR} + \text{FNSR} = 1$, which can intuitively reflect the false positive and false negative segmentation errors.

We also employ the clDice (Shit *et al* 2021) to measure the topology preservation of the DPD predictions. Moreover, we propose a new metric developed from clDice, called terminal Dice (tDice), to measure the terminal segmentation performance for targets with tubular structures. As shown in figure 5, We define Y_T and X_T as the terminal regions of Y and X , respectively. Y_T and X_T are two binary 3D masks as the same sizes as Y and X . The values of Y_T and X_T can be calculated as follows.

$$Y_T(y) = \begin{cases} Y(y), & \left\| y - \underset{z \in Y_E}{\operatorname{argmin}} d(y, z) \right\|_\infty \leq \frac{d}{2} \\ 0, & \text{otherwise} \end{cases} \quad (13)$$

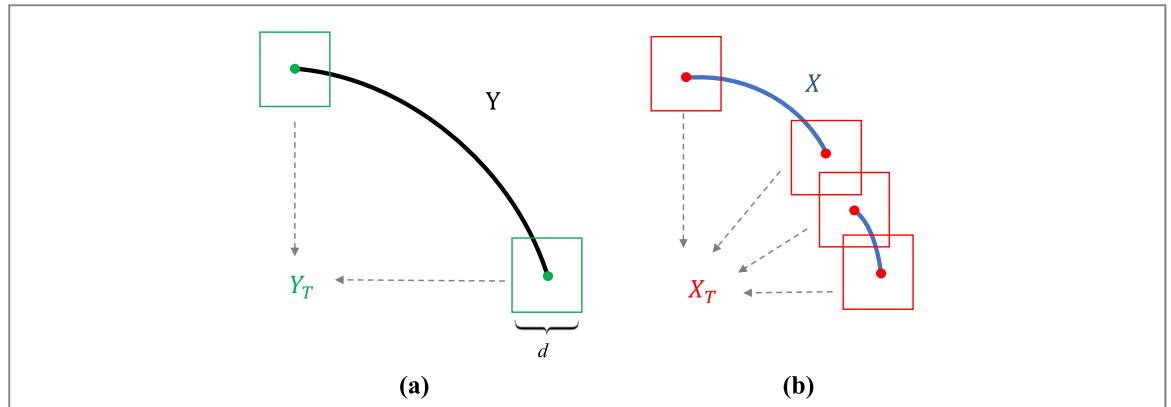


Figure 5. A schematic illustration for the proposed tDice metric. (a) The ground truth Y and its terminal regions Y_T of the dilated pancreatic duct. (b) The segmentation X and its terminal regions X_T of the dilated pancreatic duct.

$$X_T(x) = \begin{cases} X(x), & \left\| x - \underset{z \in X_E}{\operatorname{argmin}} d(x, z) \right\|_{\infty} \leq \frac{d}{2}, \\ 0, & \text{otherwise} \end{cases} \quad (14)$$

where $Y_T(y)$ and $X_T(x)$ denote the values of Y_T and X_T on the location y and x , respectively. Y_E and X_E denote the sets of endpoints on the centerlines extracted from the Y and X respectively, and d denotes the size of the terminal ROI. The infinite norm is defined as $\|x\|_{\infty} = \max_{1 \leq i \leq N} |x_i|$ where x_i is the i th component of x with N dimensions.

Then, we calculate the terminal precision ($Tprec$) and terminal sensitivity ($Tsens$) as defined below.

$$Tprec(X_T, Y) = \frac{|X_T \cap Y|}{|X_T|} \quad (15)$$

$$Tsens(X, Y_T) = \frac{|Y_T \cap X|}{|Y_T|}. \quad (16)$$

The measure $Tprec$ is related to the false positive predictions while the measure $Tsens$ reflects the false negative predictions. Since we want to maximize both precision and sensitivity, we define tDice to be the harmonic mean of both the measures:

$$tDice(X, Y) = 2 \times \frac{Tprec(X_T, Y) \times Tsens(X, Y_T)}{Tprec(X_T, Y) + Tsens(X, Y_T)}. \quad (17)$$

Therefore, The tDice metric can intuitively reflect the segmentation accuracy for the terminal regions of the targets with tubular structures.

4.3. Implementation details

We have no requirements for the backbones of BPSN, BDSN, TAAM and TDAM, and empirically keep them the same in our experiments. We use the 3D nnU-Net as the network backbones for all of them. Images are normalized to the range of 0-1. Specifically, all foreground voxels in the training set are collected, and an automated level-window-like clipping of intensity values is performed based on the 0.5 and 99.5th percentile of these values. The data is then normalized with the global foreground mean and standard deviation. All spacings within the training data are collected and for each axis the median is chosen as the target spacing. All training cases are then resampled with third order spline interpolation. We set the size of the terminal ROI as $W_2 = H_2 = L_2 = d = 32$ in our experiments which is an optimal hyperparameter after considering the sizes of the pancreas and pancreatic ducts.

4.4. Ablation study

To verify our improvement of the proposed TAAM (from coarse to fine stage) and TDAM (from fine to refine stage), we show the ablation study results in table 1. We present the segmentation results of the coarse DPD segmentation S_c , fine DPD segmentation S_f and refined DPD segmentation S_r in terms of tDice, cIDice, Dice, FPSR and FNSR.

4.4.1. Effectiveness of TAAM

From coarse to fine stage, the segmentation performance with TAAM is to be improved 3.47% in tDice, 1.98% in cIDice, 2.46% in Dice, leading to decrease 1.62% in FPSR, and 0.84% in FNSR.

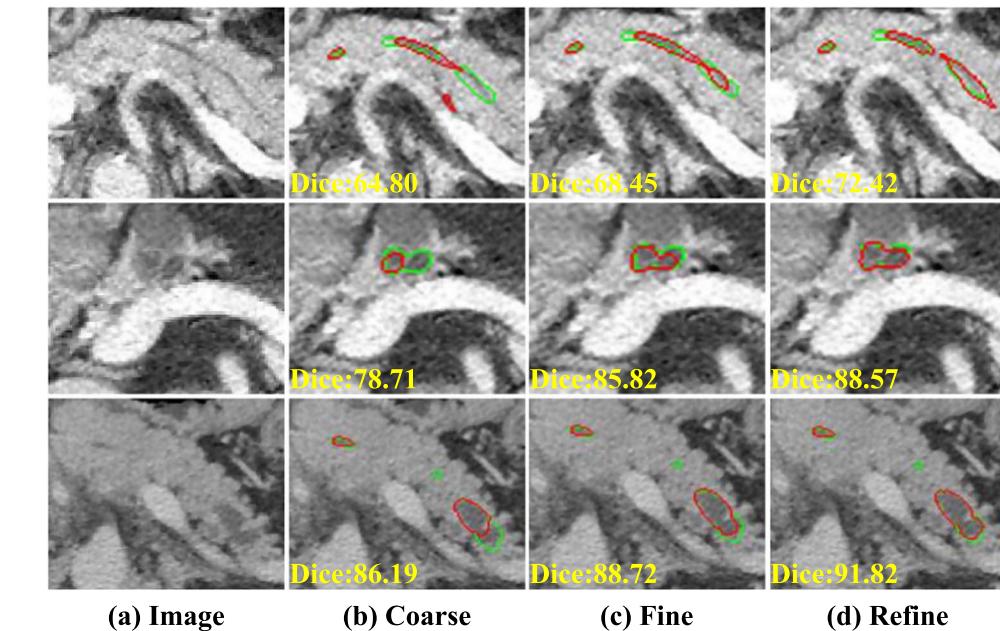


Figure 6. Visual comparison of the ablation study of CTG-Net in the axial plane. The green and red contours denote the ground truth and the predicted segmentation, respectively. Dice score of each case is written in the bottom left of each image.

Table 1. Ablation study for the proposed CTG-Net. All the results are based on five-fold cross validation. * denotes that the refine stage achieves statistically significant ($p < 0.01$) better performance.

Stage	tDice \uparrow	clDice \uparrow	Dice \uparrow	FPSR \downarrow	FNSR \downarrow
Coarse	$67.25 \pm 16.50^*$	$80.46 \pm 11.62^*$	$78.82 \pm 12.18^*$	$11.09 \pm 10.50^*$	$10.09 \pm 8.27^*$
Fine	$70.72 \pm 15.59^*$	$82.44 \pm 10.16^*$	$81.28 \pm 11.31^*$	$9.47 \pm 9.34^*$	$9.25 \pm 8.08^*$
Refine	76.29 ± 15.96	85.32 ± 9.45	84.17 ± 9.16	8.23 ± 7.37	7.60 ± 5.94

4.4.2. Effectiveness of TDAM

From fine to refine stage, the segmentation performance with TDAM is to be improved 5.57% in tDice, 2.88% in clDice, 2.89% in Dice, leading to decrease 1.24% in FPSR, and 1.65% in FNSR.

4.4.3. Effectiveness of TAAM and TDAM

We can see that both the TAAM (from coarse to fine stage) and TDAM (from fine to refine stage) can significantly improve the segmentation performance especially the terminal segmentation accuracy in terms of the tDice. Additionally, the TDAM used in the refine stage shows its effectiveness to capture terminal distraction regions especially for false negative regions. We show the visualization comparison of the ablation study results in figure 6 (axial plane) and figure 7 (3D render), respectively.

4.4.4. Effectiveness of PAA map

Moreover, we also show the effectiveness of our designed PAA map for guiding DPD segmentation comparing to general methods in table 2. One-channel encoder-decoder architecture with the terminal ROI image I'' is presented as the baseline. The two-channel encoder-decoder architecture with the terminal ROI image I'' and the coarse DPD segmentation S_c'' slightly improve the DPD segmentation performance in terms of the tDice, clDice, Dice and FNSR, leading to decrease 0.39% in FNSR comparing to the baseline method. Similarly, the three-channel encoder-decoder architecture with the terminal ROI image I'' , coarse DPD segmentation S_c'' and binary pancreas mask P_b'' slightly improve the DPD segmentation performance in terms of the tDice, clDice, Dice, FPSR and FNSR, which confirms that the binary pancreas mask has limited guidance information to segment DPD. The three-channel encoder-decoder architecture with the terminal ROI image I'' , coarse DPD segmentation S_c'' and binary pancreas mask P_a'' (i.e. the TAAM version) significantly improves the segmentation performance in terms of clDice and Dice comparing to the above three methods ($p < 0.01$), and improves the FPSR ($p < 0.05$) and FNSR ($p < 0.01$) comparing to the binary pancreas mask method.

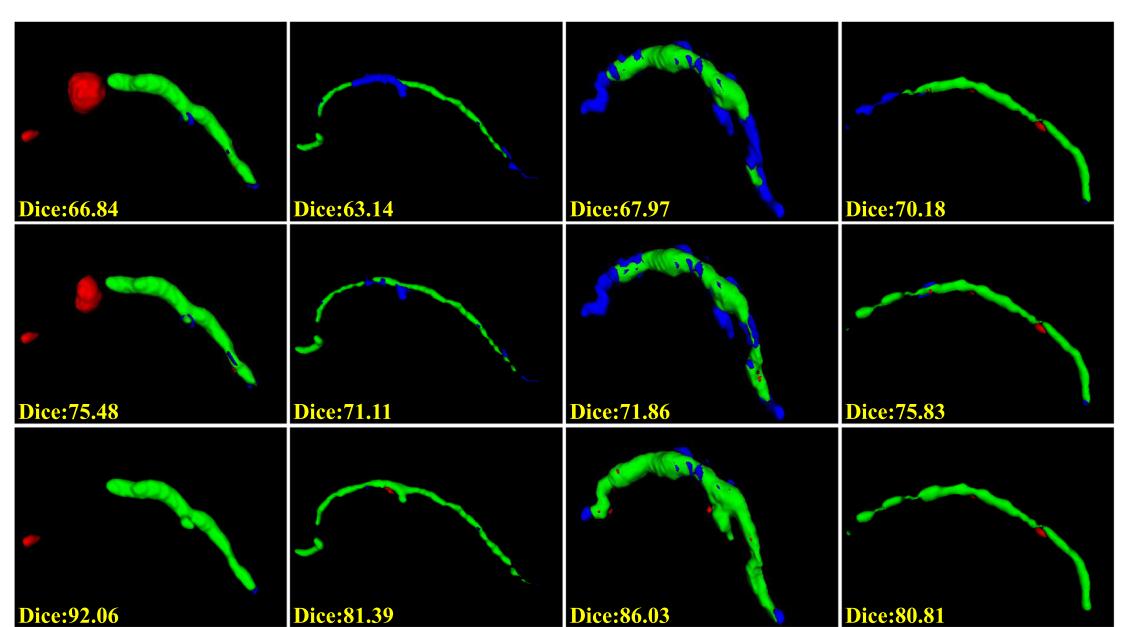


Figure 7. Visual comparison of the ablation study of CTG-Net in 3D render. The green, red and blue voxels denote the TP, FP and FN segmentation, respectively. Dice score of each case is written in the bottom left of each image. Top: coarse segmentation. Middle: fine segmentation. Bottom: refined segmentation.

Table 2. Ablation study for the proposed terminal guidance segmentation module (TAAM). All the results are based on five-fold cross validation. ⁺ and ^{*} denote the statistically significant level $p < 0.05$ and $p < 0.01$, respectively, when compared with TAAM.

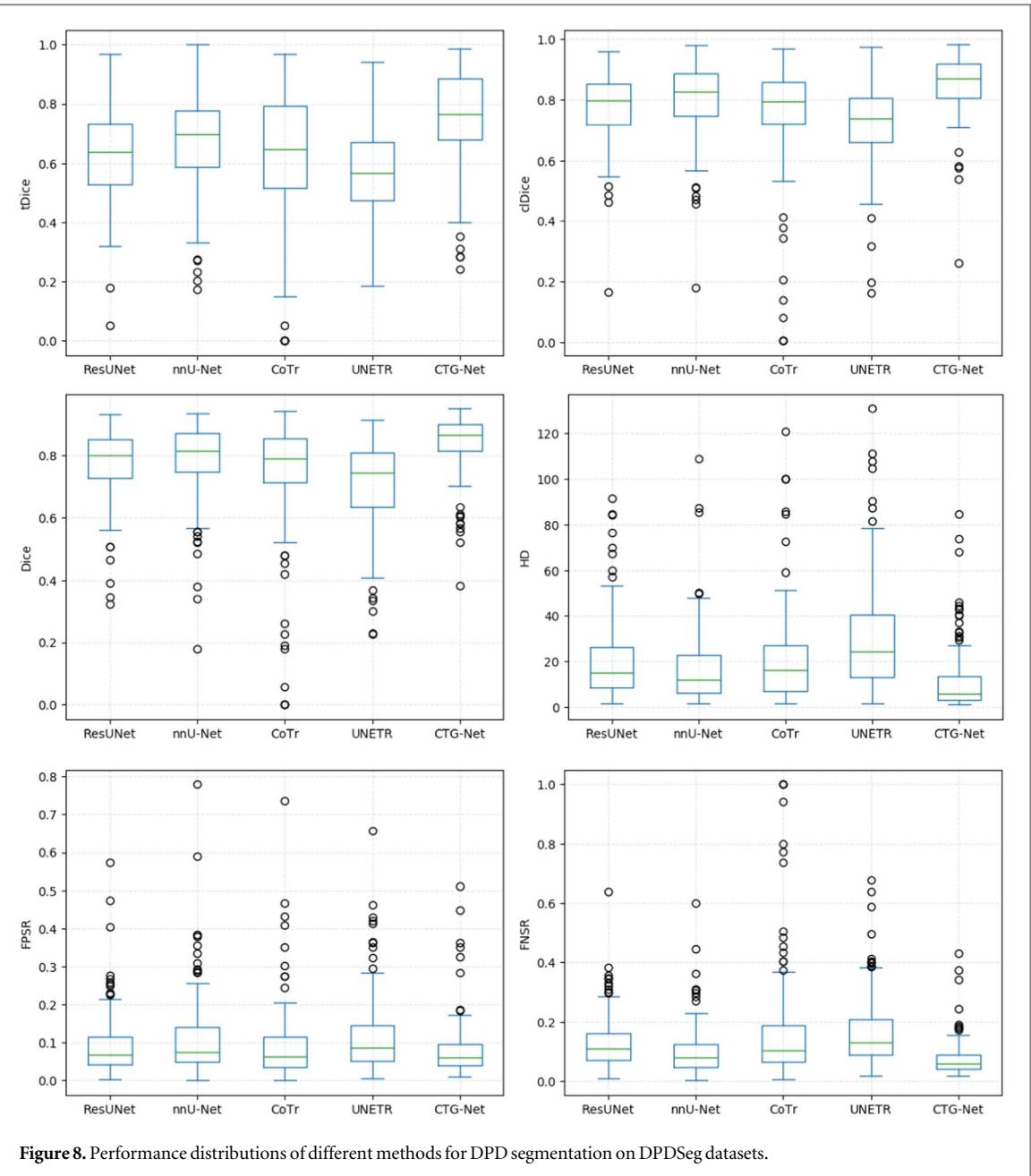
Methods	tDice \uparrow	clDice \uparrow	Dice \uparrow	FPSR \downarrow	FNSR \downarrow
One-channel with I'' (baseline)	69.39 ± 16.56	$80.61 \pm 10.89^*$	$79.44 \pm 11.30^*$	9.85 ± 9.85	$10.71 \pm 7.95^*$
Two-channel with I'' and S_c''	69.46 ± 16.82	$81.19 \pm 11.06^*$	$79.86 \pm 11.14^*$	10.24 ± 9.36	$9.89 \pm 8.44^+$
Three-channel with I'', S_c'' and P_b''	69.83 ± 17.42	$81.39 \pm 10.76^*$	$80.06 \pm 11.38^*$	$10.12 \pm 9.88^+$	$9.82 \pm 8.16^*$
Three-channel with I'', S_c'' and P_a'' (TAAM)	70.72 ± 15.59	82.44 ± 10.16	81.28 ± 11.31	9.47 ± 9.34	9.25 ± 8.08

Table 3. Quantitative five-fold cross validation results of the comparison with ResUNet (Diakogiannis *et al* 2020), nnU-Net (Isensee *et al* 2021), CoTr (Xie *et al* 2021) and UNETR (Hatamizadeh *et al* 2022) on DPDSeg dataset. The mean \pm standard deviation of tDice, clDice, Dice, HD, FPSR and FNSR are presented. ^{*} denotes that CTG-Net achieves statistically significant ($p < 0.01$) better performance.

Methods	tDice \uparrow	clDice \uparrow	Dice \uparrow	HD \downarrow	FPSR \downarrow	FNSR \downarrow
ResUNet	$63.52 \pm 15.43^*$	$78.17 \pm 11.33^*$	$77.68 \pm 11.02^*$	$20.16 \pm 17.26^*$	$9.34 \pm 8.26^*$	$12.99 \pm 8.84^*$
nnU-Net	$67.25 \pm 16.50^*$	$80.46 \pm 11.62^*$	$78.82 \pm 12.18^*$	$16.97 \pm 16.20^*$	$11.09 \pm 10.50^*$	$10.09 \pm 8.27^*$
CoTr	$62.31 \pm 20.85^*$	$75.76 \pm 18.05^*$	$74.28 \pm 18.55^*$	$21.28 \pm 21.05^*$	$9.01 \pm 9.58^*$	$16.72 \pm 19.03^*$
UNETR	$56.90 \pm 16.04^*$	$72.13 \pm 13.50^*$	$71.57 \pm 13.81^*$	$30.83 \pm 24.37^*$	$11.54 \pm 9.96^*$	$16.90 \pm 11.97^*$
CTG-Net	76.29 ± 15.96	85.32 ± 9.45	84.17 ± 9.16	11.11 ± 13.71	8.23 ± 7.37	7.60 ± 5.94

4.5. Comparison with the state-of-the-art segmentation methods

We compare our CTG-Net with other state-of-the-art (SOTA) methods for medical segmentation on our DPDSeg dataset. Recently, Transformer (Vaswani *et al* 2017), a sequence-to-sequence prediction framework, has been considered as an alternative architecture, and has achieved competitive performance on many computer vision tasks, like image recognition (Dosovitskiy *et al* 2020), semantic segmentation (Liu *et al* 2021), object detection (Carion *et al* 2020) and low-level vision (Parmar *et al* 2018). CoTr (Xie *et al* 2021) and UNETR (Hatamizadeh *et al* 2022) are the combination of convolutional neural network (CNN) and transformer for computer vision and achieve SOTA performance in medical image segmentation challenges. We compare these two transformer-based methods as long as the CNN-based methods, nnU-Net (Isensee *et al* 2021) and ResUNet (Diakogiannis *et al* 2020), with our proposed CTG-Net on our DPDSeg dataset for DPD segmentation. All the SOTA methods use the same cascaded strategy (i.e. segmenting the DPD from the pancreas ROI). All the experiments follow the same five-fold cross validation setting.



The segmentation results are shown in table 3 where we can see that our CTG-Net achieves the best performance in tDice, clDice, Dice, FPSR and FNSR with statistical significance of $p < 0.01$. We achieve 9.04% improvement for the terminal segmentation accuracy compared with the best performance by nnU-Net. Additionally, the transformer-combined methods do not show their superiority to the CNN-based methods, which may be caused by the limited scale of our collected dataset. Figure 8 presents the performance distributions of different methods. It can be found that our CTG-Net has a more compact distribution with fewer outliers. Figures 9 and 10 show the visual comparison of the axial plane and 3D render results, respectively.

4.6. Comparison with the previous techniques for DPD segmentation

We also compare with the previous techniques for DPD segmentation in table 4. Because direct comparison is not possible, we list the reported Dice score in these studies. We boost the DPD segmentation accuracy by nearly 20% compared with the existing results. We outperforms the methods of Shen *et al* (2021a, 2021b) because they only used the binary pancreas mask as the attention guidance, which might ignore the anatomy structure hints inside the pancreas for DPD segmentation. We also beat the methods of Zhou *et al* (2019), Xia *et al* (2020) because they directly segmented the DPD from the abdominal CT scans and the DPD is a extremely tiny component which only occupies a small part of the CT image, which confirms the superiority of the coarse-to-fine framework. Additionally, all the above works did not concentrate on the terminal errors while we obtain

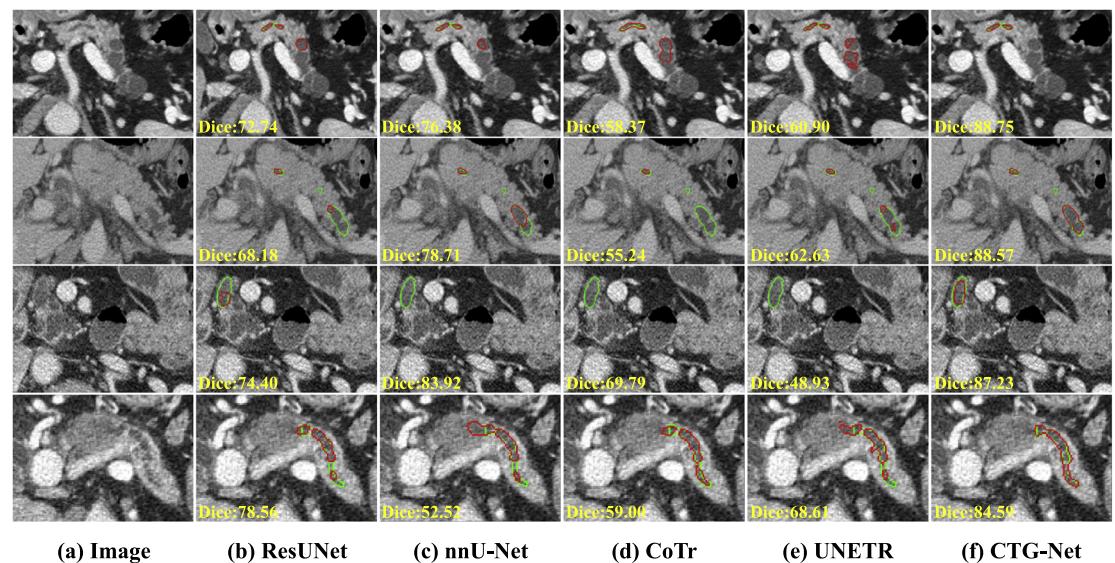


Figure 9. Qualitative results of the comparison with ResUNet (Diakogiannis *et al* 2020), nnU-Net (Isensee *et al* 2021), CoTr (Xie *et al* 2021) and UNETR (Hatamizadeh *et al* 2022) in the axial plane. The green and red contours denote the ground truth and the predicted segmentation, respectively. Dice score of each case is written in the bottom left of each image.

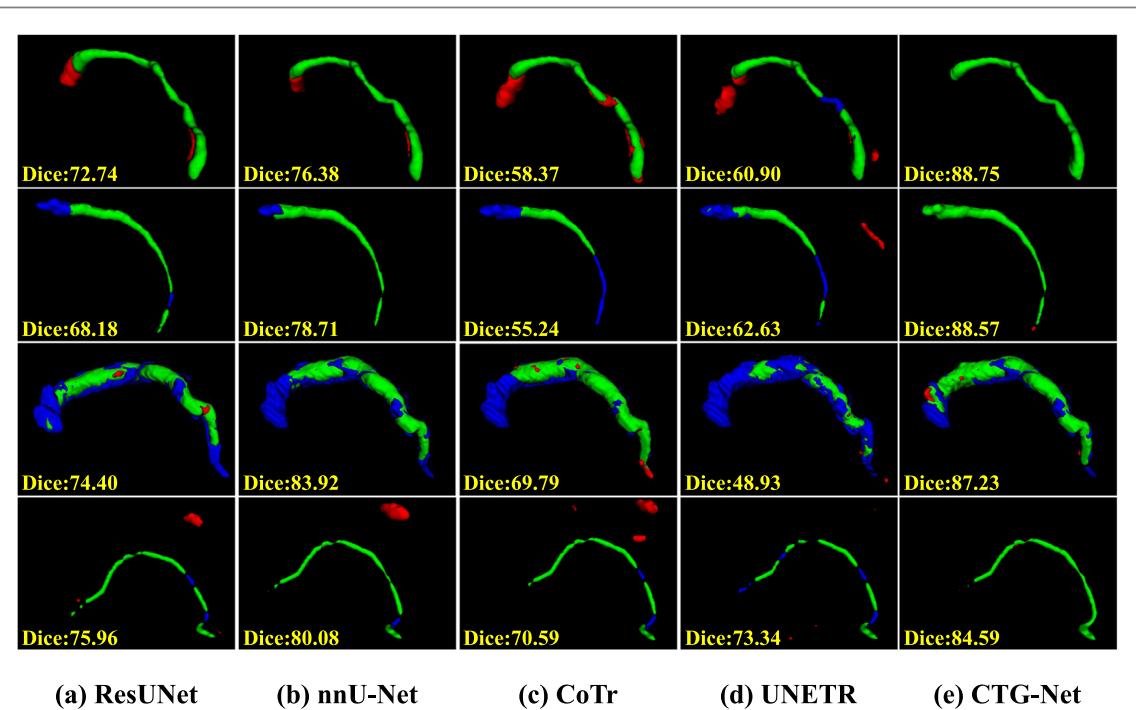


Figure 10. Qualitative results of the comparison with ResUNet (Diakogiannis *et al* 2020), nnU-Net (Isensee *et al* 2021), CoTr (Xie *et al* 2021) and UNETR (Hatamizadeh *et al* 2022) in 3D render. The green, red and blue voxels denote the TP, FP and FN segmentation, respectively. Dice score of each case is written in the bottom left of each image.

Table 4. Comparison to previous DPD segmentation works.

Methods	CT phase	Data	Tumor types	Dice
SE-Dense Unet (Shen <i>et al</i> 2021a)	Single	30	1	49.87
MPA Net (Shen <i>et al</i> 2021b)	Single	30	1	54.16
HPN (Zhou <i>et al</i> 2019)	Multi	239	1	56.77
Alignment Ensemble (Xia <i>et al</i> 2020)	Multi	239	1	64.38
CTG-Net (Ours)	Single	150	5	84.17

Table 5. Generalization validation and statistical analysis on the JPHCM dataset. * denotes that the CTG-Net achieves statistically significant ($p < 0.01$) better performance.

Methods	tDice \uparrow	clDice \uparrow	Dice \uparrow	FPSR \downarrow	FNSR \downarrow
ResUNet	$61.86 \pm 20.21^*$	$75.77 \pm 22.84^*$	$74.37 \pm 21.46^*$	10.52 ± 8.80	$15.10 \pm 16.96^*$
nnU-Net	$65.54 \pm 19.22^*$	$77.82 \pm 21.28^*$	$76.12 \pm 20.83^*$	12.79 ± 11.30	11.10 ± 19.11
CoTr	$53.25 \pm 28.53^*$	$64.31 \pm 32.92^*$	$61.36 \pm 32.84^*$	4.21 ± 6.83	$34.43 \pm 33.31^*$
UNETR	$48.47 \pm 14.74^*$	$66.64 \pm 21.88^*$	$66.09 \pm 20.94^*$	11.07 ± 8.82	$22.84 \pm 17.80^*$
CTG-Net	73.95 ± 16.17	83.05 ± 17.96	82.58 ± 15.91	10.32 ± 10.38	7.10 ± 6.86

Table 6. Segmentation performance of TDAM for distraction regions.

Target	Dice \uparrow	HD \downarrow	FPSR \downarrow	FNSR \downarrow
R_{fp}	54.09 ± 26.78	8.98 ± 5.89	13.08 ± 12.89	23.77 ± 18.26
R_{fn}	46.84 ± 27.61	7.50 ± 4.90	15.68 ± 16.90	28.65 ± 24.05

significant improvement for the terminal segmentation performance. Although we cannot make a direct comparison to these techniques, our method achieves the highest reported Dice for DPD segmentation with moderate quantity and the most tumor types. Moreover, our performance is based on single-phase CT volumes. Multi-phase CT volumes can help the model to learn more information about the DPDs.

4.7. Generalization ability validation

To demonstrate the generalization ability of our proposed method in different clinical situations, we conduct extra comparison experiment on the JPHCM dataset. We train our CTG-Net and the comparison methods on the DPDSeg dataset and test on the JPHCM dataset from another institution. The experimental results are shown in table 5. It can be observed that our proposed method achieves the best performance in tDice, clDice, Dice and FNSR metrics, which shows a satisfactory generalization ability. Besides, we notice that the CoTr method tends to have conservative predictions in the JPHCM dataset by minimizing the false positive predictions as much as possible, despite causing many false negative predictions. Therefore, the CoTr method has a better FPSR than ours, but leading to the worst performance in FNSR.

5. Discussions

5.1. Experimental results of TDAM

We have successfully evaluated the segmentation performance of proposed CTG-Net on our DPDSeg dataset. We give the prediction results of TDAM in our experiments. Table 6 shows the segmentation results for R_{fp} and R_{fn} predicted by TDAM. The Dice, HD, FPSR and FNSR are calculated on the terminal regions. Because of the tiny size of FP and FN regions, minor errors can make great drop to the segmentation metrics, which makes the Dice score of the TDAM's predictions within terminal ROIs not high. Nevertheless, since the predictions are only focus the terminal regions, over 45% terminal distraction accuracy can still make a big improvement for the final results of the whole CTG-Net framework, which shows the excellent robustness of our proposed method.

5.2. Influence of the previous stage on the subsequent results

As for cascaded model, the results of previous stage may affect the results of the next stage in general. Therefore, the adequate performance of the previous stage is required. To study the influence of the potential bad segmentation at the first model, we add extra studies on the JPHCM dataset as follows.

Firstly, we evaluate the pancreas segmentation performance on the coarse stage and the mean Dice score achieves 82.93%. We observe that the poor pancreas segmentation performance in our experiment (6 cases with pancreas Dice below 73%) generally cannot have positive influence for the subsequent results (average DPD segmentation performance decrease by 3.85%). Secondly, in order to study whether the BPSN on the coarse stage has achieved satisfactory pancreas segmentation performance, we conduct a further experiment. In the fine stage, we use the ground truth of the pancreas to replace the predicted results. The PAA maps are generated by these two different pancreas segmentation masks. The original version is termed as PAA v1 and the ground truth version is termed as PAA v2. Then, we compare the DPD performance of these two methods in the fine stage.

The experimental results are shown in table 7. It can be observed that the PAA maps generated by the ground truth can only slightly improve the DPD performance without statistical significance ($p > 0.1$). Therefore, we verify that the BPSN has achieved satisfactory performance in our experiment. We analyze the following possible

Table 7. The experimental results of fine stage with different versions of PAA map. The p-values for the t-test are shown in the right side of the corresponding metrics.

Methods	Dice	p-value	tDice	p-value	clDice	p-value
PAA v1	79.81 ± 18.09	0.892	69.95 ± 17.91	0.840	80.53 ± 19.33	0.472
PAA v2	80.01 ± 20.84		69.50 ± 18.52		81.24 ± 20.42	

reasons for the above results: (1) the proposed TAAM is with three-channel architecture. In addition to the PAA map generated by the pancreas segmentation results, the original CT image and the coarse DPD segmentation also have great learning weights for the coarse-to-fine improvement. (2) Because the pancreas segmentation errors generally occur at the boundaries and the pancreatic duct is located near the centerline of the pancreas, the predicted PAA maps generated by these imperfect pancreas masks can still provide rich structure information inside the pancreas to improve the DPD segmentation when the pancreas segmentation achieves satisfactory performance. (3) The TAAM is learned with the PAA maps generated by these imperfect pancreas segmentation results, so the network tends to have a certain ability for fault tolerance.

5.3. Limitations and future works

For the limitations, our proposed method follows the intuitive and effective coarse-fine-refine strategy for medical segmentation. However, the results of previous stage directly affect the results of the next stage, which makes it difficult to find the global optimal parameters for the whole network. Besides, the multi-stage framework also increases the computational cost.

In fact, all the three stages containing four networks in our CTG-Net can be trained in an end-to-end manner to find the global optimal parameters for the whole framework, which is studied in our further work. Moreover, the tDice metric proposed in this work can be used as a loss function to train a segmentation network for improving the terminal segmentation accuracy. Additionally, because of the advantages of the proposed CTG-Net for extracting tubular structures and enhancing terminal features in medical images, our method can be applied to other medical imaging modalities such as the retinal vascular segmentation on retinal images and bile duct segmentation on magnetic resonance cholangiopancreatography (MRCP) images. As for the clinical adoption, our proposed method can be used to pancreatic tumor screening, pancreatic duct dilation diagnosis and DPD diameter measurement for patients with pancreatic diseases.

6. Conclusions

In this paper, we propose a CTG-Net for DPD segmentation. Our approach is motivated by the effective coarse-to-fine strategy for lesion segmentation and the insufficient segmentation for the targets with tubular structures by previous methods. Our CTG-Net includes three stages: coarse, fine and refine stage. The coarse stage uses a basic cascaded segmentation framework composed of a basic pancreas segmentation network (BPSN) and a basic duct segmentation network (BDSN) to sequentially segment the pancreas and DPD. A DPD graph structure is build on the coarse DPD segmentation to locate the terminal DPD regions. In the fine stage, we introduce a TAAM to jointly learn the local intensity from the terminal CT images, feature cues from the coarse predictions and global anatomy information from our designed PAA maps. In the refine stage, we propose a terminal distraction attention module to learn the terminal distraction information and distinguish the false positive and false negative regions to improve the fine predictions. Experimental results on our dataset show that our proposed approach boosts DPD segmentation accuracy by nearly 20% compared with the existing results, and achieves more than 9% improvement for the terminal segmentation accuracy compared with the state-of-the-art methods. Our method can apply to other medical segmentation tasks such as organ-specific tumors or targets with tubular structures.

Acknowledgments

This work is supported by Ministry of Science and Technology of the People's Republic of China (No. 2020YFA0713800), National Natural Science Foundation of China (No. 11971229, 12001273) and Postgraduate Research & Practice Innovation Program of Jiangsu Province (No. KYCX22_0082). Thanks the Department of Radiology, Jiangsu Province Hospital of Chinese Medicine for providing the JPHCM dataset to verify the generalization ability of the proposed method.

Data availability statement

The data cannot be made publicly available upon publication because they contain sensitive personal information. The data that support the findings of this study are available upon reasonable request from the authors.

ORCID iDs

Liwen Zou  <https://orcid.org/0000-0003-4085-4003>

References

- Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A and Zagoruyko S 2020 End-to-end object detection with transformers *Computer Vision—ECCV 2020: 16th European Conf., Glasgow, UK, Proc., Part I 16* (Springer) pp 213–29
- Chang P-L and Teng W-G 2007 Exploiting the self-organizing map for medical image segmentation *Twentieth IEEE Int. Symp. on Computer-Based Medical Systems (CBMS'07)* (IEEE) pp 281–8
- Chen S, Tan X, Wang B, Lu H, Hu X and Fu Y 2020 Reverse attention-based residual network for salient object detection *IEEE Trans. Image Process.* **29** 3763–76
- Diakogiannis F I, Waldner F, Caccetta P and Wu C 2020 Resunet-a: a deep learning framework for semantic segmentation of remotely sensed data *ISPRS J. Photogramm. Remote Sens.* **162** 94–114
- Dosovitskiy A et al 2020 *arXiv arXiv:2010.11929* 22 Oct 2020 An image is worth 16×16 words: transformers for image recognition at scale
- Edge M D, Hoteit M, Patel A P, Wang X, Baumgarten D A and Cai Q 2007 Clinical significance of main pancreatic duct dilation on computed tomography: single and double duct dilation *World J. Gastroenterol.: WJG* **13** 1701
- Ge Y and Fitzpatrick J M 1996 On the generation of skeletons from discrete euclidean distance maps *IEEE Trans. Pattern Anal. Mach. Intell.* **18** 1055–66
- Hatamizadeh A, Tang Y, Nath V, Yang D, Myronenko A, Landman B, Roth H R and Xu D 2022 Unetr: transformers for 3D medical image segmentation *Proc. of the IEEE/CVF Winter Conf. on Applications of Computer Vision* pp 574–84
- He K, Zhang X, Ren S and Sun J 2016 Deep residual learning for image recognition *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition* pp 770–8
- Higaki T, Nakamura Y, Tatsugami F, Nakaura T and Awai K 2019 Improvement of image quality at CT and MRI using deep learning *Jpn. J. Radiol.* **37** 73–80
- Huang Q, Xia C, Wu C, Li S, Wang Y, Song Y and Kuo C-C J 2017 *arXiv arXiv:1707.06426* 20 Jul 2017 Semantic segmentation with reverse attention
- Isensee F, Jaeger P F, Kohl S A, Petersen J and Maier-Hein K H 2021 nnUNet: a self-configuring method for deep learning-based biomedical image segmentation *Nat. Methods* **18** 203–11
- Lee T-C, Kashyap R L and Chu C-N 1994 Building skeleton models via 3D medial surface axis thinning algorithms *CVGIP, Graph. Models Image Process.* **56** 462–78
- Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, Lin S and Guo B 2021 Swin transformer: hierarchical vision transformer using shifted windows *Proc. of the IEEE/CVF Int. Conf. on Computer Vision* 10012–22
- Long J, Shelhamer E and Darrell T 2015 Fully convolutional networks for semantic segmentation *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition* pp 3431–3440
- Ma J, He J and Yang X 2020 Learning geodesic active contours for embedding object global information in segmentation cnns *IEEE Trans. Med. Imaging* **40** 93–104
- Ma J et al 2021 Abdomenct-1k: is abdominal organ segmentation a solved problem? *IEEE Trans. Pattern Anal. Mach. Intell.* **44** 6695–714
- Milletari F, Navab N and Ahmadi S-A 2016 V-net: fully convolutional neural networks for volumetric medical image segmentation *2016 Fourth Int. Conf. on 3D Vision (3DV)* (IEEE) pp 565–571
- Mizrahi J D, Surana R, Valle J W and Shroff R T 2020 Pancreatic cancer *The Lancet* **395** 2008–20
- Oktay O et al 2018 *arXiv arXiv:1804.03999* 11 Apr 2018 Attention u-net: learning where to look for the pancreas
- Parmar N, Vaswani A, Uszkoreit J, Kaiser L, Shazeer N, Ku A and Tran D 2018 Image transformer *Int. Conf. on Machine Learning, PMLR* pp 4055–4064
- Ronneberger O, Fischer P and Brox T 2015 U-net: convolutional networks for biomedical image segmentation *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th Int. Conf. Proc. Part III 18* (Springer) pp 234–241
- Rosenfeld A and Pfaltz J L 1968 Distance functions on digital pictures *Pattern Recognit.* **1** 33–61
- Roth H R, Oda H, Zhou X, Shimizu N, Yang Y, Hayashi Y, Oda M, Fujiwara M, Misawa K and Mori K 2018 An application of cascaded 3D fully convolutional networks for medical image segmentation *Comput. Med. Imaging Graph.* **66** 90–9
- Shen C, Roth H, Hayashi Y, Oda M, Miyamoto T, Sato G and Mori K 2022 A cascaded fully convolutional network framework for dilated pancreatic duct segmentation *Int. J. Comput. Assist. Radiol. Surg.* **17** 343–54
- Shen C, Roth H R, Yuichiro H, Oda M, Miyamoto T, Sato G and Mori K 2021 Attention-guided pancreatic duct segmentation from abdominal ct volumes *Clinical Image-Based Procedures, Distributed and Collaborative Learning, Artificial Intelligence for Combating COVID-19 and Secure and Privacy-Preserving Machine Learning: 10th Workshop, CLIP 2021, Second Workshop, DCL 2021, First Workshop, LL-COVID19 2021, and First Workshop and Tutorial, PPML 2021, Held in Conjunction with MICCAI 2021, Proc. 2* (Springer) pp 46–55
- Shit S, Paetzold J C, Sekuboyina A, Ezhov I, Unger A, Zhylyka A, Pluim J P, Bauer U and Menze B H 2021 cldice-a novel topology-preserving loss function for tubular structure segmentation *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition* pp 16560–16569
- Springer S et al 2019 A multimodality test to guide the management of patients with a pancreatic cyst *Sci. Transl. Med.* **11** eaav4772
- Tanaka S, Nakazumi A, Ioka T, Oshikawa O, Uehara H, Nakao M, Yamamoto K, Ishikawa O, Ohigashi H and Kitamra T 2002 Main pancreatic duct dilatation: a sign of high risk for pancreatic cancer *Jpn. J. Clin. Oncol.* **32** 407–11
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser L and Polosukhin I 2017 Attention is all you need *Advances in Neural Information Processing Systems* (New York: Curran Associates, Inc.) 30

- Wang Y, Wei X, Liu F, Chen J, Zhou Y, Shen W, Fishman E K and Yuille A L 2020 Deep distance transform for tubular structure segmentation in ct scans *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition* pp 3833–3842
- Xia Y, Yu Q, Shen W, Zhou Y, Fishman E K and Yuille A L 2020 Detecting pancreatic ductal adenocarcinoma in multi-phase CT scans via alignment ensemble *Medical Image Computing and Computer Assisted Intervention—MICCAI 2020: 23rd Int. Conf. Proc. Part III* 23 (Springer) pp 285–295
- Xiao H, Feng J, Wei Y, Zhang M and Yan S 2018 Deep salient object detection with dense connections and distraction diagnosis *IEEE Trans. Multimedia* **20** 3239–51
- Xie Y, Zhang J, Shen C and Xia Y 2021 Cotr: efficiently bridging cnn and transformer for 3D medical image segmentation *Medical Image Computing and Computer Assisted Intervention—MICCAI 2021: 24th Int. Conf. Proc. Part III* 24 (Springer) pp 171–180
- Zhang Y, Wu J, Liu Y, Chen Y, Chen W, Wu E X, Li C and Tang X 2021 A deep learning framework for pancreas segmentation with multi-atlas registration and 3D level-set *Med. Image Anal.* **68** 101884
- Zhao J, Dang M, Chen Z and Wan L 2022 Dsu-net: distraction-sensitive u-net for 3D lung tumor segmentation *Eng. Appl. Artif. Intell.* **109** 104649
- Zheng Q, Qiao X, Cao Y and Lau R W 2019 Distraction-aware shadow detection *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition* pp 5167–5176
- Zhou Y, Li Y, Zhang Z, Wang Y, Wang A, Fishman E K, Yuille A L and Park S 2019 Hyper-pairing network for multi-phase pancreatic ductal adenocarcinoma segmentation *Medical Image Computing and Computer Assisted Intervention—MICCAI 2019: 22nd Int. Conf. Proc. Part II* 22 (October 13–17, 2019)(Springer) pp 155–163
- Zhu Z, Wang Q, Li B, Wu W, Yan J and Hu W 2018 Distractor-aware siamese networks for visual object tracking *Proc. of the European Conf. on Computer Vision (ECCV)* pp 101–117
- Zou L, Luo X, Zeng D, Ling L and Zhao L-C 2022 Measuring the rogue wave pattern triggered from gaussian perturbations by deep learning *Phys. Rev. E* **105** 054202