

Research paper

Automated peripancreatic vessel segmentation and labeling based on iterative trunk growth and weakly supervised mechanism

Liwen Zou^a, Zhenghua Cai^b, Liang Mao^c, Ziwei Nie^a, Yudong Qiu^c, Xiaoping Yang^{a,*}^a Department of Mathematics, Nanjing University, Nanjing, 210093, China^b Medical School, Nanjing University, Nanjing, 210007, China^c Department of General Surgery, Nanjing Drum Tower Hospital, Nanjing, 210008, China

ARTICLE INFO

Keywords:

Peripancreatic vessel
Segmentation
Anatomical labeling
Iterative trunk growth
Branch proposal
Weakly supervised learning

ABSTRACT

Peripancreatic vessel segmentation and anatomical labeling are pivotal aspects in aiding surgical planning and prognosis for patients with pancreatic tumors. Nevertheless, prevailing techniques often fall short in achieving satisfactory segmentation performance for the peripancreatic vein (PPV), leading to predictions characterized by poor integrity and connectivity. Besides, unsupervised labeling algorithms usually cannot deal with complex anatomical variation while fully supervised methods require a large number of voxel-wise annotations for training, which is very labor-intensive and time-consuming. To address these two problems, we propose an Automated Peripancreatic vEsSEL Segmentation and lAbeling (APESA) framework, to not only highly improve the segmentation performance for PPV, but also efficiently identify the peripancreatic artery (PPA) branches. There are two core modules in our proposed APESA framework: iterative trunk growth module (ITGM) for vein segmentation and weakly supervised labeling mechanism (WSLM) for artery labeling. The ITGM is composed of a series of iterative submodules, each of which chooses the largest connected component of the previous PPV segmentation as the trunk of a tree structure, seeks for the potential missing branches around the trunk by our designed branch proposal network, and facilitates trunk growth under the connectivity constraint. The WSLM incorporates the rule-based pseudo label generation with less expert participation, an anatomical labeling network to learn the branch distribution voxel by voxel, and adaptive radius-based postprocessing to refine the branch structures of the labeling predictions. Our achieved Dice of 94.01% for PPV segmentation on our collected dataset represents an approximately 10% accuracy improvement compared to state-of-the-art methods. Additionally, we attained a Dice of 97.01% for PPA segmentation and competitive labeling performance for PPA labeling compared to prior works. Our source codes will be publicly available at <https://github.com/ZouLiwen-1999/APESA>.

1. Introduction

Pancreatic cancer ranks as the third leading cause of cancer-related fatalities in the United States [1]. Among pancreatic tumors, pancreatic ductal adenocarcinoma (PDAC) stands out as the most prevalent, constituting over 90% of all pancreatic malignancies [2]. PDAC poses a formidable challenge in clinical management due to its intractability and grim prognosis, with tumor infiltration of surrounding vessels, particularly arteries, significantly impacting operability and survival prospects. Peripancreatic vascular anatomy is complex and tumors often tend to infiltrate both arteries and veins. Fig. 1 shows the relationship between the pancreas, pancreatic tumor, peripancreatic artery (PPA) and peripancreatic vein (PPV) on the contrast computed tomography (CT) image. Vascular infiltration occurs when a tumor invades

the vessel wall. Tumor resectability in case of vessel wall contact or invasion depends on the function of the specific artery and the possibility to reconstruct the vessel [3]. Moreover, tumor-vascular contact is very useful imaging features to reflect tumor's pathological heterogeneity, location, and vascular contact for building a more accurate prognosis model [4]. Hence, automated segmentation and anatomical labeling for peripancreatic vessels are promising for the surgery planning and prognosis to the patients with pancreatic tumors.

To the best of our knowledge, only one prior study has focused on peripancreatic vessel segmentation [3]. This study achieved nearly perfect segmentation performance for PPA on multispectral CT images. However, it only provided one argument for segmenting the arteries, but the veins. Moreover, it did not emphasize the anatomical labeling of artery branches which is a critical aspect for addressing clinical

* Corresponding author.

E-mail addresses: yudongqiu510@nju.edu.cn (Y. Qiu), xpyang@nju.edu.cn (X. Yang).<https://doi.org/10.1016/j.artmed.2024.102825>

Received 5 May 2023; Received in revised form 4 January 2024; Accepted 23 February 2024

Available online 24 February 2024

0933-3657/© 2024 Published by Elsevier B.V.

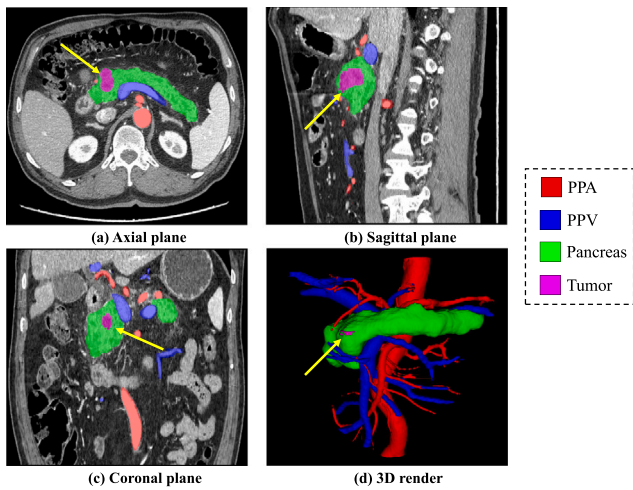


Fig. 1. Visualization of relationship between the pancreas, pancreatic tumor and peripancreatic vessels. The tumor areas are specially indicated by the yellow arrows.

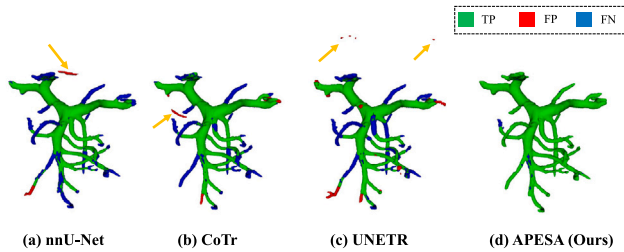


Fig. 2. An example to illustrate the benefits of the proposed APESA method for peripancreatic vein segmentation. (a) nnU-Net (b) CoTr (c) UNETR and (d) the proposed APESA. Green, red and blue voxels denote the true positive (TP), false positive (FP) and false negative (FN) predictions, respectively. It can be found that the result of proposed method has fewer FN predictions and almost no FP predictions. Besides, our prediction is fully connected while the previous methods can bring some noisy components indicated by yellow arrows. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

concerns. Recent investigations indicate a notably less clinical demand for labeling veins compared to arteries according to the impact of surgical resection of vessels on prognosis [5,6]. Therefore, our primary objective is to tackle two specific challenges: PPV segmentation and PPA labeling.

In CT images, PPV on the venous phase have lower contrast of intensity than PPA on the arterial phase, which brings extra difficulties for segmentation. Besides, previous techniques have demonstrated the shortcomings in vein segmentation, particularly concerning the terminal integrity and connectivity. Fig. 2 shows the experimental results that we compared the existing deep learning-based techniques with our proposed APESA method, where we can see that our proposed APESA method can effectively deal with the terminal errors and disconnection while the existing methods cannot.

PPA is composed of critical branches of the abdominal artery system surrounding the pancreas. There are various studies focusing on the abdominal artery anatomical labeling. Most of these studies employed unsupervised rule-based algorithms [7] or fully supervised learning-based methods [8,9]. The rule-based algorithms usually cannot deal with the anatomical variation and are always with low efficiency, while the fully supervised learning-based methods require a large number of voxel-wise annotations by the experts, which is labor-intensive and time-consuming. In this work, we propose a weakly supervised labeling mechanism which aims to efficiently identify the PPA branches with very less participation of the experts. Fig. 3 shows an illustrative example of our targeted PPA branches. According to the National

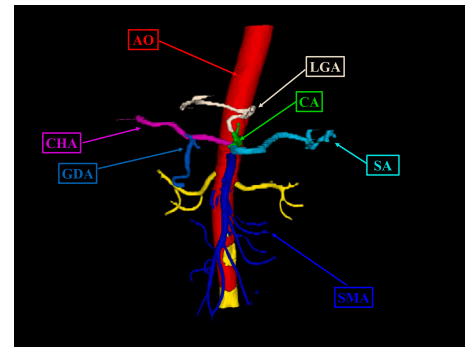


Fig. 3. Visualization of the anatomical labeling for the peripancreatic artery (PPA).

Comprehensive Cancer Network (NCCN) clinical practice guidelines for the resectability classification of pancreatic cancer [6], labeling the abdominal aorta (AO), celiac artery (CA), superior mesenteric artery (SMA), splenic artery (SA), common hepatic artery (CHA), left gastric artery (LGA) and gastroduodenal artery (GDA) of PPA has the highest priority. Because they are the closest artery branches to the pancreatic tumor, the resection and reconstruction surgery related to them will have a significant impact on the prognosis and postoperative complications [10–12]. Therefore, our goal of peripancreatic vessel labeling in this work is to identify the above seven PPA branches.

The contributions of our work can be summarized as follows.

- We introduce a novel Automated Peripancreatic vEsSel Segmentation and lAbeling (APESA) framework to not only highly improve the integrity and connectivity of PPV segmentation but also efficiently identify PPA branches with less expert involvement. Our proposed method can also be applied to other segmentation and labeling tasks for targets with tubular structures.
- We present an iterative trunk growth module (ITGM) composed of a series of iterative submodules, each of which chooses the largest connected component of the previous vein segmentation as the trunk of a tree structure, seeks for the potential missing branches around the trunk by our proposed branch proposal network, and facilitates trunk growth under the connectivity constraint. The designed iterative process can effectively guide a raw vessel trunk to be more complete and fully connected.
- We propose a weakly supervised labeling mechanism (WSLM) to efficiently identify the PPA branches with very less participation of experts. Our proposed WSLM consists of the rule-based pseudo label generation with patient-level judgement by the experts, an anatomical labeling network to learn the voxel-wise distribution of the PPA branches, and an adaptive radius-based postprocessing to refine the branch structures of labeling predictions.
- Experimental results on our collected dataset show that our proposed APESA approach boosts the PPV segmentation accuracy by nearly 10% (Dice: 94.01%), and achieves more than 20% (cIDice: 87.96%) improvement for the topological integrity compared with the state-of-the-art (SOTA) methods. Besides, we also achieve the best (Dice: 97.01%) and competitive performance on segmentation and labeling for PPA, respectively compared with the previous works.

2. Related work

2.1. Abdominal vessel segmentation in CT images

Deep learning techniques have been widely applied in medical image segmentation tasks [13–15]. In recent years, a myriad of dedicated blood vessel segmentation algorithms for different body regions and

modalities have been developed [16]. Ibragimov et al. [17] employed convolutional neural networks (CNNs) to discern the consistent appearance patterns of the portal vein (PV) and achieved an impressive Dice score of 83% in their experiments. Oda et al. [18] utilized a fully convolutional network (FCN) [19] to segment the abdominal arteries in CT images, achieving an F-measure of 87.1% on the BTCV public dataset [20]. Golla et al. [21] used a novel ratio-based sampling method to train 2D and 3D versions of the U-Net [22], V-Net [23] and DeepVesselNet [24] for abdominal vessel segmentation, resulting in Dice scores of 75.8% and 83.8% for veins and arteries, respectively. Dima et al. [3] employed a U-Net based model to perform binary segmentation of the peripancreatic arteries in multispectral CT images, achieving near perfect performance with a Dice score of 95.05% by their optimal model. Mahmoudi et al. [25] introduced a CNN-based framework for the segmentation of PDAC masses and surrounding vessels in CT images by incorporating powerful classic features, obtaining Dice scores of 73% and 81% on superior mesenteric vein (SMV) and SMA, respectively. Zhu et al. [26] enhanced the FCN with two auxiliary tasks to extract the skeleton context of abdominal arteries, yielding a Dice score of 93.2% on the BTCV public dataset. However, as of our knowledge, simultaneous segmentation of peripancreatic arteries and veins remains unexplored, representing an urgent clinical need for patients with pancreatic tumors.

2.2. Region proposal and distraction attention

The utilization of region proposal strategies is prevalent in computer vision, particularly in the object detection tasks. Ren et al. [27] introduced a region proposal network to efficiently generate high-quality detection region proposals. These proposals undergo subsequent evaluation by a training classifier to determine their veracity. The distraction concepts have been explored across various computer vision tasks, such as semantic segmentation [28], saliency detection [29,30] and visual tracking [31]. Zhao et al. [32] proposed a cascaded two-stage U-Net-based model to explicitly incorporate ambiguous region information. In their approach, the first stage generates a global segmentation for the entire input CT volumes and predicts latent distraction regions containing both false negative and false positive areas against the ground truth. The second stage integrates the distraction region information into segmentation for volume patches to enhance the discrimination of target regions. Inspired by these methods and considering the characteristics of vessel segmentation, we reformulate the vessel segmentation task to involve finding missing branches in predictions as possible based on a reliable trunk segmentation. Consequently, we develop a new branch proposal network by learning the distraction regions to provide missing branch proposals based on raw vessel trunk predictions. Instead of utilizing a training classifier in [27], we leverage the fundamental prior knowledge that vessels are fully connected to assess the rationality of generated proposals.

2.3. Anatomical labeling for abdominal vessels

Significant advancements have been made in anatomical labeling for abdominal vessels in recent years. Oda et al. [7] presented an automated anatomical labeling method for upper and lower abdominal arteries by designed rules and machine learning, achieving recall and precision rates of 79.01% and 80.41%, respectively. Matsuzaki et al. [8] adopted a similar strategy combining rule-based preprocessing, machine learning techniques and rule-based postprocessing to label abdominal arteries and hepatic portal system. They achieved labeling F-measures of 85.7% and 72.2% on manually and automatically extracted arteries, respectively. Kitasaka et al. [33] presented a method for vessel labeling by representing a vessel tree as a probabilistic graphical model using conditional random fields [34], achieving an F-measure of 94.4% for abdominal arteries. Liu et al. [9] introduced a hypergraph representation of the abdominal arterial system as a family tree model

with a probabilistic hypergraph matching framework for automated vessel labeling, achieving an average F-measure of 93.0% for abdominal arteries. However, the unsupervised labeling algorithms cannot deal with complex anatomical variation and fully supervised methods require a large number of voxel-wise annotations for training, which is very labor-intensive and time-consuming. Recently, partially supervised and knowledge-guided learning methods have been greatly developed in medical image analysis [35–37]. Inspired by these techniques, we propose a weakly supervised labeling mechanism to efficiently label the peripancreatic artery branches by embedding anatomical knowledge rules to generate high-quality pseudo labels, learning the branch distribution voxel by voxel and designing adaptive postprocessing.

3. Methods

3.1. Overview

As shown in Fig. 4, we introduce a novel framework, Automated Peripancreatic Vessel Segmentation and Labeling (APESA), for peripancreatic vein (PPV) segmentation and peripancreatic artery (PPA) labeling in CT images. Our APESA method comprises two processing flows: one for arteries and another for veins. Upon receiving a CT scan, a basic segmentation network (BSN) generates initial segmentation results for PPA and PPV, forming the basis for subsequent processing.

For the vein flow, we present an iterative trunk growth module (ITGM) composed of a series of trunk growth submodules (TGMs) to obtain more complete and fully connected PPV predictions based on the basic PPV segmentation results. Each TGM consists of the largest connected constraint (LCC) operation to get the PPV trunk, and a designed branch proposal network (BPN) to find potential missing branches around the trunk. The ITGM can guide the trunk to gradually grow to be more completed and fully connected throughout the iterative process.

For the artery flow, considering the satisfactory performance by the basic U-Net architecture for PPA segmentation [3], we only focused on the PPA labeling problem instead of further segmentation improvement. The basic PPA segmentation is directly fed to the proposed weakly supervised labeling mechanism (WSLM) for branch identification. Our proposed WSLM consists of the rule-based pseudo label generation (RPLG) with patient-level judgement by the experts, an anatomical labeling network (ALN) to learn the voxel-wise distribution of the PPA branches, and an adaptive radius-based postprocessing (ARP) to refine the branch structures of labeling predictions.

3.2. Iterative Trunk Growth Module (ITGM)

We introduce the iterative trunk growth module (ITGM) presented in Fig. 5 to obtain more complete and fully connected PPV segmentation based on the basic PPV predictions. Our proposed ITGM is composed of a series of trunk growth submodules (TGMs). Considering the basic PPV segmentation as a tree, the TGM firstly gets the trunk of the tree throughout a largest connected constraint (LCC) operation which selects the largest connected component of the PPV segmentation result. The trunk serves as a reliable foundation with little false positive (FP) predictions and much false negative (FN) predictions, which can provide valuable guidance for the subsequent growth processing. A branch proposal network (BPN) is designed to find potential missing branches or proposals by learning the similar features around the trunk. Although these proposals generated by BPN may not be absolutely true predictions, those proposal branches connected to the raw trunk have high probabilities to be the missing branches required. Therefore, another LCC operation is adopted to eliminate branches far from the trunk, ensuring trunk growth under connectivity constraints. This iterative mechanism can guide the trunk toward completeness and full connectivity. We formulate the ITGM as follows:

$$T_i = \begin{cases} LCC(S_0) & i = 0 \\ LCC(BPN(T_{i-1}; T) + T_{i-1}), & i > 0 \end{cases} \quad (1)$$

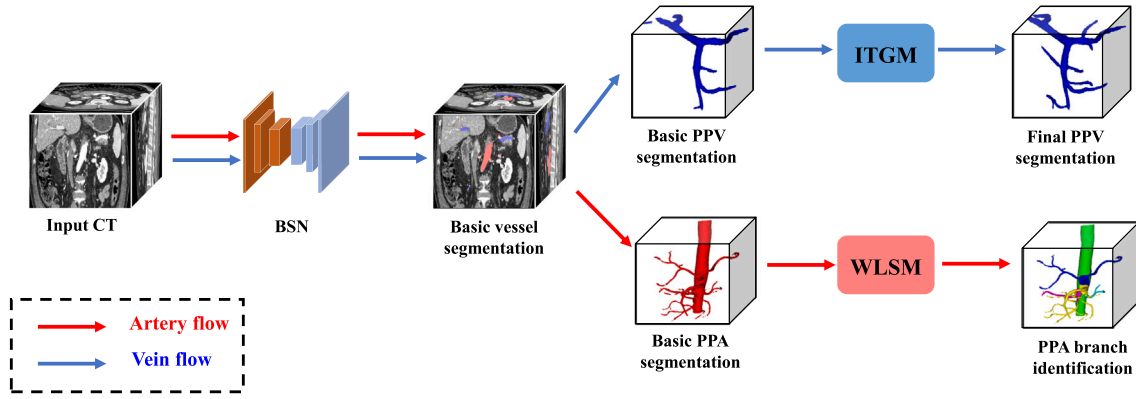


Fig. 4. Pipeline of our proposed Automated Peripancreatic Vessel Segmentation and Labeling (APESA) framework for PPV segmentation and PPA labeling. The basic PPV and PPA segmentation is fed to the proposed ITGM and WLSM, respectively for the following segmentation improvement and efficiently branch labeling.

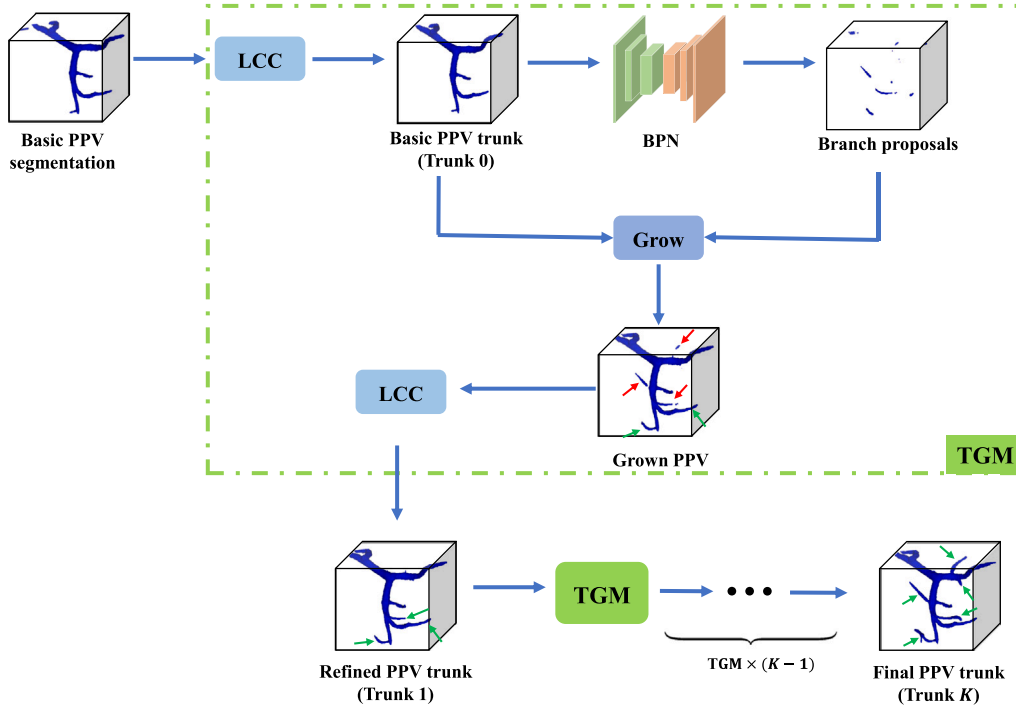


Fig. 5. Overview of our proposed Iterative Trunk Growth Module (ITGM). We can see that the proposed BPN can find the potential missing branches (shown by arrows) around the trunk. The connected ones (shown by green arrows) are preserved, while the disconnected ones (shown by red arrows) are eliminated by the LCC operation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

where T_i , I and S_0 denote the i th trunk prediction, CT image and the basic PPV segmentation by the previous BSN, respectively. $LCC(S)$ denotes the largest connected component of the segmentation S , and $BPN(S; I)$ denotes the missing branches predicted by the BPN with the input of segmentation S and CT image I .

The proposed BPN is designed to find the potential false negative regions or missing branches around the trunk. The details of the BPN is illustrated in Fig. 6. In the training stage, the difference between the trunk prediction and ground truth (GT) will be calculated as the GT missing branches to train the BPN. In the inference stage, we apply the trained BPN to obtain potential missing branch proposals around the trunk.

3.3. Weakly supervised labeling mechanism (WLSM)

We propose a weakly supervised labeling mechanism (WLSM) to efficiently identify the seven crucial PPA branches mentioned above. Our WLSM is composed of the rule-based pseudo label generation

(RPLG) with very less participation of experts, an anatomical labeling network (ALN) to learn the branch distribution voxel by voxel, and an adaptive radius-based postprocessing (ARP) to refine the branch structures of the labeling predictions. The details of WLSM is illustrated in Fig. 7.

3.3.1. Rule-based pseudo label generation (RPLG)

As depicted in the purple dash rectangle in Fig. 7, after obtaining the basic PPA segmentation results from the previous basic segmentation network (BSN), we skeletonize the 3D volumes to derive the centerline using the method proposed in [38]. A graph structure is then established based on the centerline, with points on the centerline serving as nodes and the neighbor relationships as edges. To prevent the closed-loop situations, a tree structure is created using the Prim algorithm [39] which is a classic minimum spanning tree algorithm.

Subsequently, we use the rule-based labeling algorithm to predict the branch category of each node on the PPA tree structure. The process of the rule-based labeling algorithm is illustrated in Fig. 8. We firstly

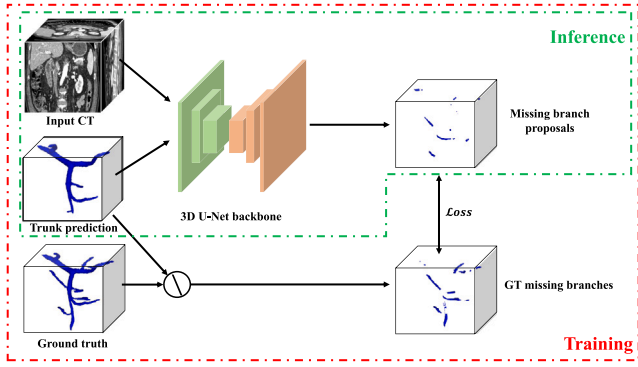


Fig. 6. Overview of our proposed Branch Proposal Network (BPN). In the training stage, the difference between the trunk prediction and ground truth (GT) will be calculated as the GT missing branches to train the BPN. In the inference stage, we apply the pretrained BPN to the previous trunk prediction to get potential missing branch proposals around the trunk.

find the key nodes (i.e., the terminal nodes and the junctions) of the PPA tree. Then, the shortest paths between these key nodes are labeled as corresponding branches based on the peripancreatic vessel anatomical rules in order. More details about the implementation of the rule-based labeling algorithm can be found in Section A of the Supplementary Materials. Once the node categories are predicted, the volume of the segmentation result is labeled as the same category as the nearest node.

After we obtain the PPA labeling results by the rule-based strategy, the experts only need to give the patient-level annotations of “good” or “bad” judgement. The “good” judgement is given when the true positive voxels account for the majority on all focused branches and there is no negative impact on clinical issues according to these predictions. The “good” predictions are used as pseudo branch labels for subsequent procedures while the “bad” predictions are not used.

It should be pointed that, although the rule-based labeling algorithm may have low yield rate (i.e., cannot generate many “good” cases)

because of the complex anatomical variation, more “good” cases can be produced when we apply this unsupervised algorithm to larger-scale dataset with patient-level judgement by the experts, which can save a lot of time and labor costs compared with voxel-level annotation strategy. Additionally, the “good” cases are judged to be the gold standard for all seven branches and they contains rich supervision information to train the following anatomical labeling network.

3.3.2. Anatomical labeling network (ALN)

With pseudo labels for the PPA branch identification, an anatomical labeling network (ALN) is trained to learn the branch distribution voxel by voxel. ALN takes CT images and the point clouds of the basic PPA segmentation results as inputs and is trained with pseudo labels generated by the RPLG. It is designed to learn the voxel-level branch distribution based on the “good” cases judged by the experts. On the one hand, the ALN is trained in a weakly supervised manner with less expert participation. On the other hand, voxel-wised feature learning can efficiently inherit the knowledge extracted from the rule-based algorithm to deal with the anatomical variation between different patients.

3.3.3. Adaptive radius-based postprocessing (ARP)

As shown in Fig. 9(a), point Q might be misclassified as CA instead of AO in the RPLG stage based on the distance between Q and the centerline. The large gap of radius between AO and other branches can bring “disk” branch predictions shown in 9(b), which makes negative impact on labeling evaluation for subsequent branches of AO especially CA. To address this problem, we introduce adaptive radius-based postprocessing (ARP) as follows.

Considering the wrong “disk” have larger radius than the true branch, we calculate the radius of each point on the branch centerline. The K-Means algorithm is then used to divide these points into two clusters based on the calculated radius. The cluster with larger radius is predicted as AO. Subsequently, the whole 3d volumes can be labeled as follows:

$$L(p) = \begin{cases} KM(r(p)) & p \in U_{cl} \\ L(\arg \min_{i \in U_{cl}} \|i - p\|_2) & otherwise \end{cases} \quad (2)$$

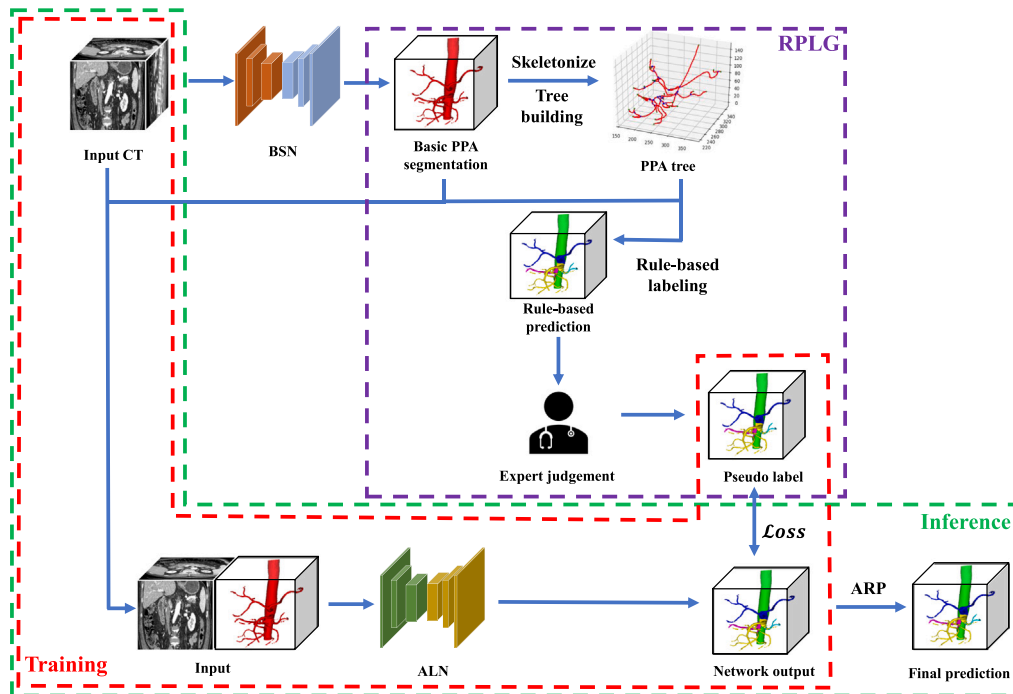


Fig. 7. Overview of the proposed Weakly Supervised Labeling Mechanism (WSLM). The RPLG is designed to generate the pseudo labels of branch identification with patient-level judgement of experts. The ALN is trained with the pseudo labels and the ARP is only applied in the inference stage.

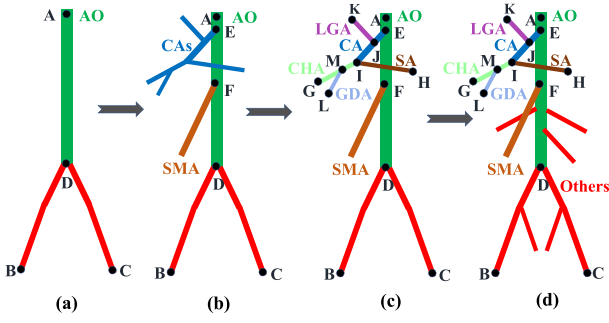


Fig. 8. The process of the rule-based labeling algorithm for PPA branch identification. (a) The first order PPA branch (AO) labeling. (b) The second order PPA branch (SMA, CA and its subordinate branches) labeling. (c) The third order PPA branch (CHA, SA, LGA and GDA) labeling. (d) Other branches labeling.

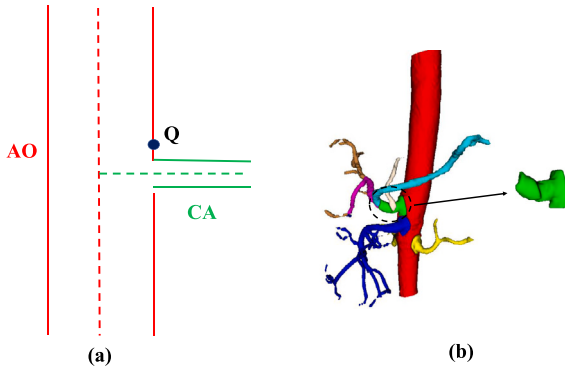


Fig. 9. (a) A schematic diagram to illustrate the cause of “disk” prediction, the dash lines denote the centerlines of the branches. Point Q on AO is closer to CA’s centerline instead of AO’s. (b) A visual example for the “disk” branch prediction in 3D render. We can see that CA and other subsequent branches of AO tend to have “disk” predictions on the AO’s surface in the RPLG stage.

where p is a volume of the segmentation result, $L(p)$ is the branch labeling prediction of volume p , U_{cl} is the set of points on the centerline, $KM(r(p))$ denotes the prediction of K-Means algorithm based on the radius of point p .

4. Experiments

4.1. Datasets

We evaluated our proposed method on two datasets collected from Nanjing Drum Tower Hospital. Besides the PDAC cases, patients with other pancreatic tumors such as the intraductal papillary mucinous neoplasm (IPMN), serous cystic intraductal papillary mucinous neoplasm (SCN), intraductal papillary mucinous neoplasm (MCN) and solid pseudopapillary tumor (SPT) are also included in our datasets. The first dataset, termed as Drum Tower Peripancreatic Vein (DTPV) dataset, consists of 272 venous phase CT scans from patients with surgical pathology-confirmed pancreatic tumors (136 PDACs, 43 IPMNs, 43 SCNs, 24 SPTs, 26 MCNs). The second dataset, termed as Drum Tower Peripancreatic Artery (DTPA) dataset, consists of 338 arterial phase CT scans from patients with surgical pathology-confirmed pancreatic tumors (191 PDACs, 46 IPMNs, 40 SCNs, 34 SPTs, 27 MCNs).

All the CT images are stored as NIFTI files with an axial plane resolution of 512×512 , a slice number ranging from 117 to 1049, and an original spacing of $0.75 \times 0.75 \times 1.25 \text{ mm}^3$. Data preprocessing involves normalization and resampling, following the strategy outlined in [40].

Each CT scan in the DTPV or DTPA datasets has voxel-wise manual annotation of the PPV or PPA which is performed by two pancreatic imaging radiologists using the open-source application ITK-SNAP.¹

4.2. Evaluation metrics

4.2.1. Segmentation metrics

For segmentation evaluation, we adopt the Dice similarity coefficient, calculated as follows:

$$Dice = \frac{2 \times TP}{2 \times TP + FP + FN}, \quad (3)$$

where TP , FP and FN denote the numbers of true positive, false positive and false negative predicted voxels, respectively. We also employ the cDice [41] metric which was widely used in recent segmentation studies [42,43] to measure the topology preservation of the segmentation for targets with tubular structures. Unlike the pixel-level region matching of the Dice metric, the cDice focuses on the morphological (skeleton) matching between the segmentation result and the ground truth of the targets with tubular structures. It can be calculated as follows:

$$T_{prec}(S_p, V_G) = \frac{|S_p \cap V_G|}{|S_p|}, \quad (4)$$

$$T_{sens}(S_G, V_p) = \frac{|S_G \cap V_p|}{|S_G|}, \quad (5)$$

$$cDice(V_p, V_G) = 2 \times \frac{T_{prec} \times T_{sens}}{T_{prec} + T_{sens}}, \quad (6)$$

where V_G and V_p are the ground truth and the predicted mask, respectively. S_G and S_p are the skeletons extracted from V_G and V_p , respectively. In addition to the above metrics for overall evaluation of segmentation results, we also use the following two segmentation metrics for measuring the distractions. The false positive segmentation rate (FPSR) and false negative segmentation rate (FNSR) are employed in our experiments, which can intuitively reflect the false positive and false negative segmentation errors [44]. They can be calculated as follows:

$$FPSR = \frac{FP}{2 \times TP + FP + FN}, \quad (7)$$

$$FNSR = \frac{FN}{2 \times TP + FP + FN}. \quad (8)$$

Besides, we also calculate the number of the connected components (NCC) to evaluate the connectivity-preserving performance of segmentation. All the segmentation experiments are performed using five-fold cross validation.

4.2.2. Labeling metrics

Branch-wise and voxel-wise metrics for anatomical labeling are both utilized in our experiments. The precision, recall and F-measure (or F_1 score) as follows:

$$Precision = \frac{TP}{TP + FP}, \quad (9)$$

$$Recall = \frac{TP}{TP + FN}, \quad (10)$$

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}, \quad (11)$$

where TP , FP and FN denote the numbers of TP, FP and FN branches or voxels, respectively. We define a branch labeling prediction as a TP branch when the TP voxels on this branch account for the majority compared with FP and FN voxels. FP and FN branches can be defined in the similar way.

¹ <http://www.itksnap.org/>.

4.3. Implementation details

The network backbones of BSN, BPN and ALN are all based on the generic 3D U-Net architecture, which includes skip connections between different resolution stages of the encoder and decoder. Both the encoder and decoder are composed of stacks of convolutional blocks. In the encoder, we use 9 convolutional blocks, each of which consists of two sets of consecutive $3 \times 3 \times 3$ convolution layers followed by instance normalization and leaky ReLU activation. The dropout layers are also included. In the decoder, we use 8 convolutional blocks, each of which takes the output of the previous block, upsamples feature maps with a $2 \times 2 \times 2$ transposed convolutional layer, and concatenates the output with the corresponding shortcut output from the encoder. The BSN, BPN and ALN are trained for 1000, 500 and 200 epochs, respectively. Stochastic gradient descent with Nesterov momentum ($\mu = 0.99$) and an initial learning rate of 0.01 is used for learning network weights. The learning rate is decayed throughout the training following the poly learning rate policy in [45]

$$lr = base_lr \times (1 - \frac{epoch}{max_epoch})^{0.9}, \quad (12)$$

where lr and $base_lr$ are the current and previous learning rate on this epoch, respectively. max_epoch denotes the maximum of the training epochs. The loss function is the sum of cross-entropy (CE) and Dice loss [46], which is calculated as follows.

$$L_{Dice} = -\frac{2 \sum_{i=1}^N s_i g_i}{\sum_{i=1}^N s_i^2 + \sum_{i=1}^N g_i^2}, \quad (13)$$

$$L_{CE} = -\frac{\sum_{i=1}^N g_i \ln p_i}{N}, \quad (14)$$

$$Loss = L_{Dice} + L_{CE}, \quad (15)$$

where N denotes the number of the input CT voxels, s_i , g_i and p_i are the binary segmentation prediction, the ground truth and the *softmax* output on voxel i , respectively. The training batch size and the iterations K in the ITGM are both set to be 2. All the experiments are performed on an NVIDIA A100 GPU machine with 40 GB memory.

4.4. PPV segmentation

In this section, we assess the performance of our proposed method for PPV segmentation on the collected DTPV dataset. We conduct a detailed analysis of segmentation performance across various patients with pancreatic tumors. Additionally, we present an ablation study of our proposed ITGM for PPV segmentation, comparing it with state-of-the-art (SOTA) methods and related works.

4.4.1. PPV segmentation results on DTPV dataset

The PPV segmentation results of the proposed APESA for different pancreatic tumor cases on the collected DTPV dataset are shown in Table 1. The highest metric is shown in bold and the lowest one is shown in red. It can be observed that the non-PDAC cases achieve better segmentation performance than the PDAC cases, which may be caused by the invasion of peripancreatic vessels by pancreatic malignant tumors.

4.4.2. Ablation study for ITGM

To verify the improvement of our proposed ITGM, we conduct the ablation study and the experimental results are shown in Table 2. We compare the segmentation performance of the basic segmentation network (BSN) marked as basic, and the two following TGM results marked as trunk 1 and trunk 2. Besides, we also list the performance of largest connected component of the basic prediction marked as trunk 0, and the iterative grown predictions before the second LCC operation which are marked as TGM $\times 1$ and TGM $\times 2$, respectively. From Table 2, we can see that the proposed ITGM can effectively boost the topological

Table 1

Comparison of PPV segmentation performance for patients with different pancreatic tumors on the DTPV dataset in terms of four metrics (arrows indicate which direction is better). Bold and red denote the best and worst performance, respectively.

Case	clDice (%) \uparrow	Dice (%) \uparrow	FPSR (%) \downarrow	FNSR (%) \downarrow
PDAC	84.17 \pm 14.68	91.44 \pm 11.16	5.66 \pm 9.70	2.90 \pm 2.92
IPMN	89.75 \pm 11.01	95.32 \pm 5.89	2.94 \pm 5.77	1.74 \pm 0.90
MCN	92.51 \pm 6.67	97.16 \pm 3.73	0.83 \pm 3.59	2.02 \pm 1.34
SCN	94.14 \pm 5.26	97.73 \pm 2.75	0.36 \pm 2.14	1.91 \pm 1.83
SPT	90.34 \pm 13.19	96.24 \pm 5.42	2.41 \pm 5.37	1.36 \pm 0.69
Total	87.96 \pm 12.93	94.01 \pm 8.96	3.65 \pm 7.85	2.34 \pm 2.34

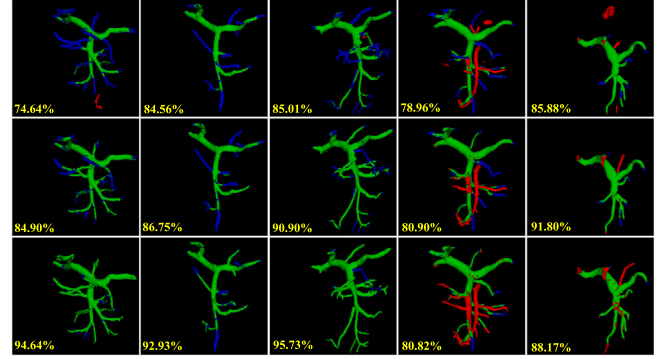


Fig. 10. Visual comparison in ablation study of ITGM in 3D render. The green, red and blue voxels denote the true positive, false positive and false negative segmentation, respectively. Dice score of each case is written in the bottom left of each image. **Top:** Basic segmentation. **Middle:** Trunk 1 segmentation. **Bottom:** Trunk 2 segmentation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

integrity (clDice) and the overall segmentation (Dice) performance. Additionally, the proposed method can significantly alleviate the FN predictions and keep comparable FP segmentation performance compared to the basic predictions. It can be also observed that the proposed ITGM can completely ensure the connectivity and has better performance in structure integrity than the largest connected component of the basic predictions (trunk 0). The visualization of the different segmentation results is presented in Fig. 10.

4.4.3. Comparison with SOTA segmentation methods

We also compare our method with state-of-the-art (SOTA) models for PPV segmentation on the DTPV dataset. Recently, Transformer [47], 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 [48], semantic segmentation [49], object detection [50] and low-level vision [51]. CoTr [52] and UNETR [53] are the combination of convolutional neural network and transformer for computer vision and achieve SOTA performance in medical image segmentation challenges. We compared our proposed APESA method to the two SOTA transformer-based methods and the SOTA CNN-based nnU-Net [40] for PPV segmentation on our DTPV dataset. All the experiments follow the same five-fold cross validation setting.

The segmentation results are presented in Table 3 where it can be observed that our proposed method achieves the best performance in clDice, Dice, FNSR and NCC metrics. Additionally, the transformer-based methods do not show their superiority to the CNN-based method, which may be caused by the limited scale of the dataset. Fig. 11 presents the performance distributions for the comparison experiment. It can be found that our proposed APESA has a more compact distribution with fewer outliers. Fig. 12 illustrates the visual comparisons. It should be pointed that the LCC operation is not used in the comparison methods. The details about the comparison experiments with LCC operation can be found in Section B of the Supplementary Materials.

Table 2

Ablation study results of ITGM on our DTPV dataset in terms of five metrics (arrows indicate which direction is better). All the results are based on five-fold cross validation.

Methods	clDice (% ↑)	Dice (% ↑)	FPSR (% ↓)	FNSR (% ↓)	NCC (↓)
Basic	68.52 ± 9.22*	84.57 ± 7.90*	3.43 ± 8.54	12.00 ± 4.31	5.08 ± 2.72*
Trunk 0	68.66 ± 8.76*	85.52 ± 6.86*	1.44 ± 5.28	13.04 ± 4.89	1.00 ± 0.00
TGM × 1	78.68 ± 9.75*	89.53 ± 7.33*	3.16 ± 6.73	7.31 ± 3.31	4.33 ± 2.29*
Trunk 1	79.57 ± 9.15*	89.98 ± 7.04*	2.46 ± 6.31	7.56 ± 3.53	1.00 ± 0.00
TGM × 2	86.12 ± 12.79*	93.41 ± 8.53*	4.45 ± 7.99	2.13 ± 1.76	6.26 ± 3.32*
Trunk 2	87.96 ± 12.93	94.01 ± 8.96	3.65 ± 7.85	2.34 ± 2.34	1.00 ± 0.00

* Denotes that trunk 2 achieves statistically significant ($p < 0.01$) better performance using paired t-test.

Table 3

Quantitative five-fold cross validation results of the comparison with nnU-Net [40], CoTr [52] and UNETR [53] on DTPV dataset. The mean ± standard deviation of clDice, Dice, FPSR, FNSR and NCC are presented.

Methods	clDice (% ↑)	Dice (% ↑)	FPSR (% ↓)	FNSR (% ↓)	NCC (↓)
nnU-Net [40]	68.52 ± 9.22*	84.57 ± 7.90*	3.43 ± 8.54	12.00 ± 4.31*	5.08 ± 2.72*
CoTr [52]	65.30 ± 10.10*	82.47 ± 8.56*	4.00 ± 7.17	13.52 ± 6.82*	6.64 ± 3.56*
UNETR [53]	62.25 ± 10.21*	80.90 ± 8.69*	5.42 ± 9.24*	13.68 ± 4.70*	15.13 ± 7.36*
APESA (Ours)	87.96 ± 12.93	94.01 ± 8.96	3.65 ± 7.85	2.34 ± 2.34	1.00 ± 0.00

* Denotes that APESA achieves statistically significant ($p < 0.01$) better performance using paired t-test.

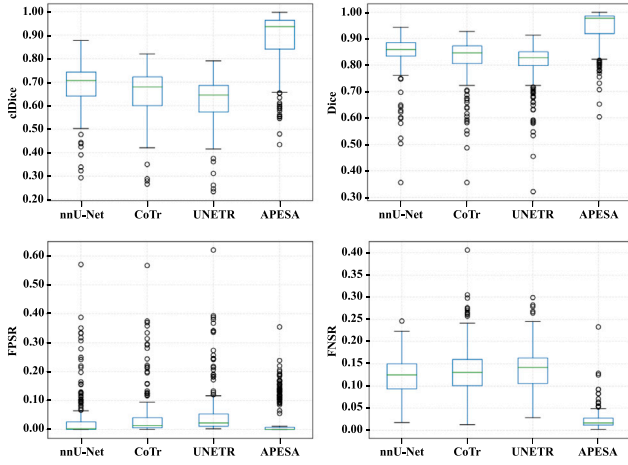


Fig. 11. Performance distributions of different methods for PPV segmentation on DTPV datasets.

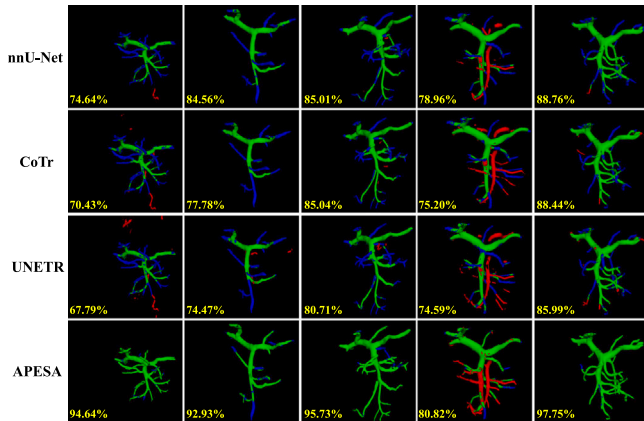


Fig. 12. Qualitative results of the comparison with nnU-Net [40], CoTr [52] and UNETR [53] in 3D render. The green, red and blue voxels denote the true positive, false positive and false negative segmentation, respectively. Dice score of each case is written in the bottom left of each image. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 4

Comparison to the works related to PPV segmentation. N.R. stands for not reported.

Methods	Data	Target	Tumor	Dice (%)
Ibragimov [17]	72	Portal vein	N.R.	83.00
Golla [21]	20	Portal vein	N.R.	75.80
Mahmoudi [25]	138	SMV	1	73.00
APESA (Ours)	273	PPV	5	94.01

4.4.4. Comparison with the works related to PPV segmentation

To the best of our knowledge, this is the first work concentrating on the PPV segmentation and there is no related public dataset to evaluate and compare. Therefore, we compare with the state-of-the-art methods for the portal vein or other related vein branch segmentation in Table 4. Because direct comparison is not possible, The reported Dice scores are listed in these studies only for reference. It can be observed that our methods achieve the highest reported Dice score for vein segmentation task. Besides, since our dataset includes the patients with 5 types of pancreatic tumors, the proposed method has more reliable performance for clinical diagnosis and treatment.

4.5. PPA segmentation

In this section, we present the PPA segmentation performance by the basic segmentation network (BSN) on our collected DTPA dataset. We conduct a comprehensive analysis of segmentation performance among different patients with pancreatic tumors and compare the results with relevant works.

4.5.1. PPA segmentation results on DTPA dataset

The PPA segmentation results of the BSN for different pancreatic tumor cases on the collected DTPA dataset are shown in Table 5. The highest metric is shown in bold and the lowest is shown in red. Similar to the PPV segmentation performance, the non-PDAC cases achieve better PPA segmentation performance than the PDAC cases. Fig. 13 presents the qualitative results of the PPA segmentation. It can be found that the PPA segmentation almost achieves perfect performance comparing to the PPV segmentation, which may be attributed to the higher intensity contrast.

4.5.2. Comparison with the works related to PPA segmentation

We conduct a comparative analysis of our proposed method with the state-of-the-art techniques for PPA segmentation as summarized in Table 6. Dima et al. [3] achieved a Dice score of 95.05% for PPA

Table 5

Comparison of PPA segmentation performance for patients with different pancreatic tumors on the DTPA dataset in terms of four metrics (arrows indicate which direction is better). Bold and red denote the best and worst performance, respectively.

Case	clDice (% ↑)	Dice (% ↑)	FPSR (% ↓)	FNSR (% ↓)
PDAC	78.98 ± 8.07	96.25 ± 2.90	1.59 ± 2.40	2.16 ± 1.14
IPMN	88.15 ± 3.63	98.28 ± 0.34	0.77 ± 0.29	0.95 ± 0.29
MCN	87.13 ± 5.59	98.08 ± 0.66	0.77 ± 0.45	1.15 ± 0.40
SCN	85.81 ± 4.39	97.85 ± 0.79	0.91 ± 0.63	1.25 ± 0.46
SPT	87.23 ± 3.13	97.77 ± 0.59	0.95 ± 0.30	1.28 ± 0.43
Total	82.52 ± 7.81	97.01 ± 2.39	1.27 ± 1.87	1.72 ± 1.04

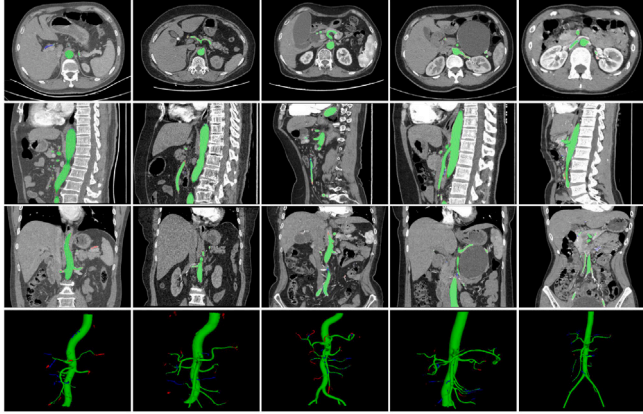


Fig. 13. Qualitative results: In the first row our segmentations are presented in the axial view, in the second row as sagittal view, in the third row as coronal view, in the last row as a 3D rendering. Green, blue, and red represent true positive, false negative and false positive predictions, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 6

Comparison to the works related to PPA segmentation. N.R. stands for not reported.

Methods	Data	Target	Tumor	Dice
Oda [18]	20	AO	N.R.	87.10
Golla [21]	20	AO	N.R.	83.80
Dima [3]	143	PPA	1	95.05
Mahmoudi [25]	138	SMA	1	81.00
Zhu [26]	20	AO	N.R.	93.20
APESA (Ours)	345	PPA	5	97.01

segmentation using 3D U-Net with multiple inputs of Iodine and arterial images. The BSN in our APESA framework attains the highest reported Dice score for PPA segmentation only by single-phase input with more tumor types.

4.6. PPA labeling

We apply the rule-based pseudo label generation (RPLG) to 108 cases from the DTPA dataset and 24 cases are judged as “good” by the experienced experts. We train the labeling network with these 24 pseudo annotations. There are 30 independent cases from the DTPA dataset with voxel-wise branch manual annotations provided by experts for testing.

4.6.1. PPA labeling results on DTPA dataset

Table 7 displays the anatomical labeling results for the seven PPA branches. Branch-wise and voxel-wise metrics are shown in blue and red, respectively. **Fig. 14** presents the qualitative results of the anatomical labeling predictions, and the difference between predictions and ground truth is marked by arrows.

Table 7

Experimental results of WSLM for PPA branch identification. The branch-wise metrics Precision, Recall and F-measure are shown in blue, the voxel-wise metric Dice is shown in red.

Artery	Precision (%)	Recall (%)	F-measure (%)	Dice (%)
AO	100	100	100	95.67
CA	100	86.67	92.86	67.79
SMA	100	100	100	86.09
SA	100	100	100	94.42
CHA	100	100	100	89.92
LGA	100	96.67	98.31	89.37
GDA	96.55	96.55	96.55	81.95

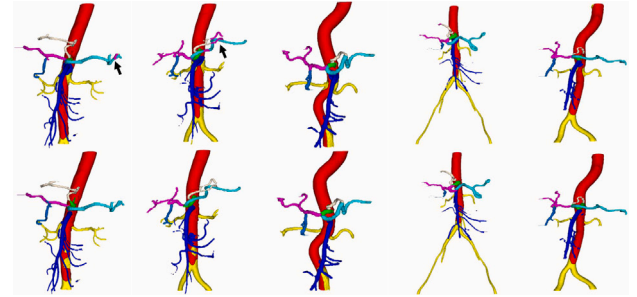


Fig. 14. The 3D render comparison results of the anatomical labeling predictions for the PPA. **Top:** Predictions. **Bottom:** Ground truth. The black arrows show the difference between prediction and ground truth.

Table 8

Ablation study for WSLM. w/o means removing the corresponding module. The ARP is only applied for CA labeling improvement in the experiment.

Methods	AO	CA	SMA	SA	CHA	LGA	GDA
WSLM w/o ALN, ARP	100	68.8	64.5	95.0	89.5	80.0	76.5
WSLM w/o ARP	100	84.6*	100*	100**	100*	98.3*	96.6*
WSLM	100	92.9*	100	100	100	98.3	96.6

* Denotes statistically significant $p < 0.0001$ performance improvement compared with the previous method using the bootstrap method.

** Denotes statistically significant $p < 0.01$ performance improvement compared with the previous method using the bootstrap method.

4.6.2. Ablation study for WSLM

In order to verify the effectiveness of our WSLM for PPA labeling, we conduct the corresponding ablation study. For statistical analysis, we follow the previous studies [4,54] and apply a bootstrap method. For each method, 10 cases in test set are randomly sampled for calculating the branch-level F-measure. In order to match the sample size of the test set, this is repeated 30 times for each method. The Wilcoxon signed rank test is then used to compare the F-measure distributions. Experimental results are presented in **Table 8**. It can be observed that the anatomical labeling network (ALN) indeed improves the labeling performance and has better generalization ability for almost all the targeted branches compared to the rule-based pseudo label generation (RPLG). Additionally, the proposed adaptive radius-based postprocessing (ARP) significantly improve the labeling performance of CA by alleviating the “disk” predictions. The visual comparison between the labeling results with and without ARP can be found in Section C of the Supplementary Materials.

4.6.3. Comparison with the SOTA vessel labeling methods

Given that most abdominal artery labeling studies encompass the PPA labeling task, we present the reported experimental results for the related branches in **Table 9**. It can be observed that our approach achieves the best performance on six PPA branches and achieves comparable performance for CA labeling compared to the existing labeling techniques.

Table 9

Comparison to the existing artery labeling works, the branch-wise F-measure is presented as the evaluation metric. N.R. stands for not reported.

Methods	Data	AO	CA	SMA	SA	CHA	LGA	GDA
Oda [7]	23	100	64.6	83.3	57.1	33.3	N.R.	N.R.
Kitasaka [33]	50	95.8	93.8	97.6	91.1	87.0	89.5	75.8
Liu [9]	37	98.7	95.5	95.9	84.8	88.4	N.R.	N.R.
APESA (Ours)	30	100	92.9	100	100	100	98.3	96.6

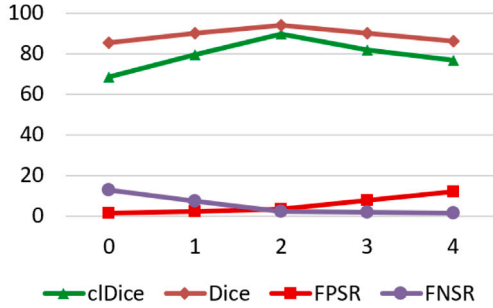


Fig. 15. PPV segmentation performance (vertical axis) comparison between different iteration number K (horizontal axis) in the ITGM.

5. Discussions

5.1. Optimal iteration of the ITGM

An extra experiment is conducted to find the optimal iteration K in the ITGM. We applied the ITGM with different iteration K for PPV segmentation on the DTPV dataset and compared the performance between these iterations. Fig. 15 presents the performance comparison between the different iteration K . It is observed that the overall segmentation performance achieves the optimal level when $K = 2$. While the ITGM effectively mitigates false negative predictions, an increase in false positive predictions with the number of iterations leads to an overall performance decline. Possible reasons for these results are analyzed: (1) Adhesion between PPV and other similar targets such as PPA leads to incorrect branch growth under the LCC constraint for certain special cases (e.g., the cases shown in the last two columns of Fig. 10). (2) Due to the low contrast at the terminal regions of the PPV on CT images, it is difficult to fully covering these regions in manual annotations, resulting in the grown branches being erroneously judged as false positive predictions. Quantitative experimental results are detailed in Section D of the Supplementary Materials.

5.2. Training ALN with different number of “good” cases

In order to verify the supervision ability of the 24 cases used in the ALN training process, we train the ALN with different number of “good” cases under the unified training settings. The trained models are then evaluated on the same test dataset for labeling performance comparison. Especially, we obtained 28 “good” cases when we applied the rule-based labeling algorithm to 125 samples from the collected dataset for training ALN with larger number of “good” cases. The labeling performance comparison is presented in Table 10. The statistical analysis is also provided following the setting in Section 4.6.2. It is observed that the ALN achieves satisfactory labeling performance for most PPA branches even trained with 16 “good” cases which demonstrates the high-quality of the pseudo labels generated by the RPLG. However, it is also noted that the smaller branched structures such as CA and GDA require a larger quantity of training data to improve the labeling performance.

Table 10

Experimental results for PPA labeling with different number of ALN training cases.

Training cases	AO	CA	SMA	SA	CHA	LGA	GDA
16	100	63.6	100	98.3	100	96.6	88.9
20	100	77.6*	100	96.6	98.3**	100*	92.7**
24	100	84.6*	100	100**	100**	98.3**	96.6*
28	100	86.8**	100	100	100	98.3	98.3**

* Denotes statistically significant $p < 0.0001$ performance improvement or drop compared with the previous result using the bootstrap method.

** Denotes statistically significant $p < 0.01$ performance improvement or drop compared with the previous result using the bootstrap method.

5.3. Limitations and future works

To our knowledge, this is the first work that simultaneously focuses on segmentation and labeling of peripancreatic vessels. Prior studies pertaining to vessel segmentation and labeling outlined in the Tables 4, 6, and 9 primarily focused on the abdominal vessel system or specific branches. Moreover, almost all the datasets used in these studies are not publicly available. Therefore, it is difficult to make the direct comparison. We can only list the reported performance and the used datasets for reference. The iteration number K is a hyper-parameter of the ITGM in our experiments. We conclude in Section 5.1 that the overall segmentation performance achieves the optimal level when $K = 2$. However, an adaptive iteration setting can be proposed for each specific case. Additionally, our ITGM can effectively alleviate the false negative predictions while ignoring the increase in false positive predictions caused by incorrect growth. Our method can be further improved if we design an error growth recognition mechanism and use it as an adaptive iteration termination condition, which is the problem we aim to address in our subsequent studies. In the WSLM, because CA has smaller size and is more affected by the “disk” predictions than the other branches, we only apply the ARP for CA labeling improvement in our experiments. The other slightly affected branches such as SMA can be processed in the same mechanism, which will be explored in our following researches.

Since previous techniques have achieved nearly perfect segmentation performance on PPA segmentation, the ITGM in our proposed APESA framework is designed only for PPV segmentation improvement. However, this mechanism can also be applied for PPA segmentation to improve the terminal error predictions. Moreover, the proposed WSLM can also be applied for PPV labeling as long as the corresponding PPV anatomy rules are designed, which will be further studied in our future works.

6. Conclusions

In this paper, we introduce a novel Automated Peripancreatic Vessel Segmentation and Labeling (APESA) framework, for peripancreatic vein (PPV) segmentation and peripancreatic artery (PPA) labeling. Our approach is driven by two core modules, iterative trunk growth module (ITGM) for PPV segmentation and weakly supervised labeling mechanism (WSLM) for PPA branch identification. The ITGM is composed of a series of iterative submodules, each of which chooses the largest connected component of the previous PPV segmentation as the trunk of the vein tree, seeks for the potential missing branches around the trunk by our designed branch proposal network, and facilitates trunk growth under the connectivity constraint. The WSLM incorporates the rule-based pseudo label generation with less expert participation, an anatomical labeling network to learn the branch distribution voxel by voxel, and adaptive radius-based postprocessing to refine the branch structures of the labeling predictions.

Our experiments demonstrate the significant advancements of the proposed method with an almost 10% enhancement in PPV segmentation accuracy compared to state-of-the-art algorithms and an approximate 20% improvement in topological integrity performance compared

to previous techniques. Additionally, we achieve superior and competitive performance for PPA segmentation and labeling, respectively. The proposed framework can also be applied to other segmentation and labeling task for targets with tubular structures.

CRedit authorship contribution statement

Liwen Zou: Funding acquisition, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. **Zhenghua Cai:** Investigation, Resources, Validation, Writing – original draft, Writing – review & editing. **Liang Mao:** Investigation, Resources, Supervision, Validation, Writing – review & editing. **Ziwei Nie:** Funding acquisition, Investigation, Supervision, Validation, Writing – review & editing. **Yudong Qiu:** Conceptualization, Investigation, Project administration, Resources, Supervision, Validation, Writing – review & editing. **Xiaoping Yang:** Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing.

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.

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. 12090023, 12301661), Natural Science Foundation of Jiangsu Province, China (No. BK20210173) and Postgraduate Research and Practice Innovation Program of Jiangsu Province, China (No. KYCX22_0082).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.artmed.2024.102825>.

References

- [1] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA: Cancer J Clin* 2019;69(1):7–34.
- [2] Wild C, Weiderpass E, Stewart BW. World cancer report: cancer research for cancer prevention. International Agency for Research on Cancer; 2020.
- [3] Dima A, Paetzold JC, Jungmann F, Lemke T, Raffler P, Kaissis G, Rueckert D, Braren R. Segmentation of peripancreatic arteries in multispectral computed tomography imaging. In: Machine learning in medical imaging: 12th international workshop, MLMI 2021, held in conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, proceedings 12. Springer; 2021, p. 596–605.
- [4] Yao J, Shi Y, Cao K, Lu L, Lu J, Song Q, Jin G, Xiao J, Hou Y, Zhang L. DeepPrognosis: Preoperative prediction of pancreatic cancer survival and surgical margin via comprehensive understanding of dynamic contrast-enhanced CT imaging and tumor-vascular contact parsing. *Med Image Anal* 2021;73:102150.
- [5] Groen JV, van Manen L, van Roessel S, van Dam JL, Bonsing BA, Doukas M, van Eijck CH, Sarasqueta AF, Putter H, Vahrmeijer AL, et al. Resection of the portal-superior mesenteric vein in pancreatic cancer: pathological assessment and recurrence patterns. *Pancreas* 2021;50(8):1218–29.
- [6] Tempero MA, Malafa MP, Al-Hawary M, Behrman SW, Benson AB, Cardin DB, Chiorean EG, Chung V, Czito B, Del Chiaro M, et al. Pancreatic adenocarcinoma, version 2.2021, NCCN clinical practice guidelines in oncology. *J Natl Compr Cancer Netw* 2021;19(4):439–57.
- [7] Oda M, Hoang BH, Kitasaka T, Misawa K, Fujiwara M, Mori K. Automated anatomical labeling method for abdominal arteries extracted from 3D abdominal CT images. In: Medical imaging 2012: image processing. vol. 8314, SPIE; 2012, p. 707–12.
- [8] Matsuzaki T, Oda M, Kitasaka T, Hayashi Y, Misawa K, Mori K. Automated anatomical labeling of abdominal arteries and hepatic portal system extracted from abdominal CT volumes. *Med Image Anal* 2015;20(1):152–61.
- [9] Liu Y, Wang X, Wu Z, Lopez-Linares K, Macia I, Ru X, Zhao H, Ballester MAG, Zhang C. Automated anatomical labeling of a topologically variant abdominal arterial system via probabilistic hypergraph matching. *Med Image Anal* 2022;75:102249.
- [10] Shibata Y, Uemura K, Kondo N, Sumiyoshi T, Okada K, Seo S, Otsuka H, Murakami Y, Arihiro K, Takahashi S. Long-term survival after distal pancreatectomy with celiac axis resection and hepatic artery reconstruction in the setting of locally advanced unresectable pancreatic cancer. *Clin J Gastroenterol* 2022;15(3):635–41.
- [11] Mollberg N, Rahbari NN, Koch M, Hartwig W, Hoeger Y, Buechler MW, Weitz J. Arterial resection during pancreatectomy for pancreatic cancer: a systematic review and meta-analysis. *Ann Surg* 2011;254(6):882–93.
- [12] Schneider M, Strobel O, Hackert T, Büchler MW. Pancreatic resection for cancer—the Heidelberg technique. *Langenbeck's Arch Surg* 2019;404:1017–22.
- [13] Li X, Chen H, Qi X, Dou Q, Fu C-W, Heng P-A. H-DenseUNet: hybrid densely connected unet for liver and tumor segmentation from CT volumes. *IEEE Trans Med Imaging* 2018;37(12):2663–74.
- [14] Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, Liang J. Unet++: A nested u-net architecture for medical image segmentation. In: Deep learning in medical image analysis and multimodal learning for clinical decision support: 4th international workshop, DLMIA 2018, and 8th international workshop, ML-CDS 2018, held in conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, proceedings 4. Springer; 2018, p. 3–11.
- [15] Wang Y, Wei X, Liu F, Chen J, Zhou Y, Shen W, Fishman EK, Yuille AL. Deep distance transform for tubular structure segmentation in ct scans. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020, p. 3833–42.
- [16] Moccia S, De Momi E, El Hadji S, Mattos LS. Blood vessel segmentation algorithms—review of methods, datasets and evaluation metrics. *Comput Methods Programs Biomed* 2018;158:71–91.
- [17] Ibragimov B, Toesca D, Chang D, Koong A, Xing L. Combining deep learning with anatomical analysis for segmentation of the portal vein for liver SBRT planning. *Phys Med Biol* 2017;62(23):8943.
- [18] Oda M, Roth HR, Kitasaka T, Misawa K, Fujiwara M, Mori K. Abdominal artery segmentation method from CT volumes using fully convolutional neural network. *Int J Comput Assist Radiol Surg* 2019;14:2069–81.
- [19] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2015, p. 3431–40.
- [20] Landman B, Xu Z, Igelsias J, Styner M, Langerak T, Klein A. MICCAI multi-atlas labeling beyond the cranial vault-workshop and challenge (2015). 2015, <http://dx.doi.org/10.7303/syn3193805>.
- [21] Golla A-K, Bauer DF, Schmidt R, Russ T, Nörenberg D, Chung K, Tönnies C, Schad LR, Zöllner FG. Convolutional neural network ensemble segmentation with ratio-based sampling for the arteries and veins in abdominal CT scans. *IEEE Trans Biomed Eng* 2020;68(5):1518–26.
- [22] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: Medical image computing and computer-assisted intervention—MICCAI 2015: 18th international conference, Munich, Germany, October 5–9, 2015, proceedings, part III 18. Springer; 2015, p. 234–41.
- [23] Milletari F, Navab N, Ahmadi S-A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 fourth international conference on 3D vision. 3DV, IEEE; 2016, p. 565–71.
- [24] Tetteh G, Efremov V, Forkert ND, Schneider M, Kirschke J, Weber B, Zimmer C, Piraud M, Menze BH. Deepvesselnet: Vessel segmentation, centerline prediction, and bifurcation detection in 3-d angiographic volumes. *Front Neurosci* 2020;14:1285.
- [25] Mahmoudi T, Kouzahan ZM, Radmard AR, Kafieh R, Salehnia A, Davarpanah AH, Arabalibeik H, Ahmadian A. Segmentation of pancreatic ductal adenocarcinoma (PDAC) and surrounding vessels in CT images using deep convolutional neural networks and texture descriptors. *Sci Rep* 2022;12(1):3092.
- [26] Zhu R, Oda M, Hayashi Y, Kitasaka T, Misawa K, Fujiwara M, Mori K. A skeleton context-aware 3D fully convolutional network for abdominal artery segmentation. *Int J Comput Assist Radiol Surg* 2023;18(3):461–72.
- [27] Ren S, He K, Girshick R, Sun J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv Neural Inf Process Syst* 2015;28.
- [28] Huang Q, Xia C, Wu C, Li S, Wang Y, Song Y, Kuo C-CJ. Semantic segmentation with reverse attention. 2017, arXiv preprint arXiv:1707.06426.
- [29] Chen S, Tan X, Wang B, Lu H, Hu X, Fu Y. Reverse attention-based residual network for salient object detection. *IEEE Trans Image Process* 2020;29:3763–76.
- [30] Xiao H, Feng J, Wei Y, Zhang M, Yan S. Deep salient object detection with dense connections and distraction diagnosis. *IEEE Trans Multimed* 2018;20(12):3239–51.
- [31] Zhu Z, Wang Q, Li B, Wu W, Yan J, Hu W. Distractor-aware siamese networks for visual object tracking. In: Proceedings of the European conference on computer vision. ECCV, 2018, p. 101–17.
- [32] Zhao J, Dang M, Chen Z, Wan L. DSU-Net: Distraction-Sensitive U-Net for 3D lung tumor segmentation. *Eng Appl Artif Intell* 2022;109:104649.
- [33] Kitasaka T, Kagajo M, Nimura Y, Hayashi Y, Oda M, Misawa K, Mori K. Automatic anatomical labeling of arteries and veins using conditional random fields. *Int J Comput Assist Radiol Surg* 2017;12(6):1041–8.

- [34] Sutton C, McCallum A, et al. An introduction to conditional random fields. *Found Trends[®] Mach Learn* 2012;4(4):267–373.
- [35] Dou Q, Liu Q, Heng PA, Glocker B. Unpaired multi-modal segmentation via knowledge distillation. *IEEE Trans Med Imaging* 2020;39(7):2415–25.
- [36] Zhao T, Cao K, Yao J, Nogues I, Lu L, Huang L, Xiao J, Yin Z, Zhang L. 3D graph anatomy geometry-integrated network for pancreatic mass segmentation, diagnosis, and quantitative patient management. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2021, p. 13743–52.
- [37] Bai Y, Chen D, Li Q, Shen W, Wang Y. Bidirectional copy-paste for semi-supervised medical image segmentation. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2023, p. 11514–24.
- [38] Lee T-C, Kashyap RL, Chu C-N. Building skeleton models via 3-D medial surface axis thinning algorithms. *CVGIP: Graph Models Image Process* 1994;56(6):462–78.
- [39] Prim RC. Shortest connection networks and some generalizations. *Bell Syst Tech J* 1957;36(6):1389–401.
- [40] Isensee F, Jaeger PF, Kohl SA, Petersen J, Maier-Hein KH. NnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods* 2021;18(2):203–11.
- [41] Shit S, Paetzold JC, Sekuboyina A, Ezhov I, Unger A, Zhylka A, Pluim JP, Bauer U, Menze BH. cDice-a novel topology-preserving loss function for tubular structure segmentation. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2021, p. 16560–9.
- [42] Yu H, Zhao J, Zhang L. Vessel segmentation via link prediction of graph neural networks. In: *International workshop on multiscale multimodal medical imaging*. Springer; 2022, p. 34–43.
- [43] Shamsi NI, Gjestey LA, Chavez D, Snyder M, Eastwood BS, Fay MG, O'Connor NJ, Glaser JR, Gerfen CR, Brattain LJ. Self-supervised learning to improve topology-optimized axon segmentation and centerline detection. In: *2023 IEEE 20th international symposium on biomedical imaging. ISBI, IEEE; 2023*, p. 1–4.
- [44] Zou L, Cai Z, Qiu Y, Gui L, Mao L, Yang X. CTG-Net: An efficient cascaded framework driven by terminal guidance mechanism for dilated pancreatic duct segmentation. *Phys Med Biol* 2023.
- [45] Chen L-C, Papandreou G, Kokkinos I, Murphy K, Yuille AL. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans Pattern Anal Mach Intell* 2017;40(4):834–48.
- [46] Drozdal M, Vorontsov E, Chartrand G, Kadoury S, Pal C. The importance of skip connections in biomedical image segmentation. In: *International workshop on deep learning in medical image analysis, international workshop on large-scale annotation of biomedical data and expert label synthesis*. Springer; 2016, p. 179–87.
- [47] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. *Adv Neural Inf Process Syst* 2017;30.
- [48] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, et al. An image is worth 16x16 words: Transformers for image recognition at scale. 2020, arXiv preprint arXiv:2010.11929.
- [49] Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, Lin S, Guo B. Swin transformer: Hierarchical vision transformer using shifted windows. In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2021, p. 10012–22.
- [50] Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A, Zagoruyko S. End-to-end object detection with transformers. In: *European conference on computer vision*. Springer; 2020, p. 213–29.
- [51] Parmar N, Vaswani A, Uszkoreit J, Kaiser L, Shazeer N, Ku A, Tran D. Image transformer. In: *International conference on machine learning*. PMLR; 2018, p. 4055–64.
- [52] Xie Y, Zhang J, Shen C, Xia Y. Cotr: Efficiently bridging cnn and transformer for 3d medical image segmentation. In: *Medical image computing and computer assisted intervention—mICCAI 2021: 24th international conference, strasbourg, France, September 27–October 1, 2021, proceedings, part III 24*. Springer; 2021, p. 171–80.
- [53] Hatamizadeh A, Tang Y, Nath V, Yang D, Myronenko A, Landman B, Roth HR, Xu D. Unetr: Transformers for 3d medical image segmentation. In: *Proceedings of the IEEE/CVF winter conference on applications of computer vision*. 2022, p. 574–84.
- [54] Lou B, Doken S, Zhuang T, Wingerter D, Gidwani M, Mistry N, Ladic L, Kamen A, Abazeed ME. An image-based deep learning framework for individualising radiotherapy dose: a retrospective analysis of outcome prediction. *Lancet Digit Health* 2019;1(3):e136–47.