



Segment Like A Doctor: Learning reliable clinical thinking and experience for pancreas and pancreatic cancer segmentation

Liwen Zou^a, Yingying Cao^b, Ziwei Nie^a, Liang Mao^c, Yudong Qiu^c, Zhongqiu Wang^b,
Zhenghua Cai^{c,d,*}, Xiaoping Yang^{a, **}

^a School of Mathematics, Nanjing University, Nanjing, 210093, China

^b Department of Radiology, Affiliated Hospital of Nanjing University of Chinese Medicine, Nanjing, 210000, China

^c Department of Pancreatic Surgery, Nanjing Drum Tower Hospital, Nanjing, 210008, China

^d Medical School, Nanjing University, Nanjing, 210007, China

ARTICLE INFO

Keywords:

Clinical thinking and experience
Pancreatic cancer segmentation
Self-supervised learning
Causality reasoning
Boundary calibration

ABSTRACT

Pancreatic cancer is a lethal invasive tumor with one of the worst prognosis. Accurate and reliable segmentation for pancreas and pancreatic cancer on computerized tomography (CT) images is vital in clinical diagnosis and treatment. Although certain deep learning-based techniques have been tentatively applied to this task, current performance of pancreatic cancer segmentation is far from meeting the clinical needs due to the tiny size, irregular shape and extremely uncertain boundary of the cancer. Besides, most of the existing studies are established on the black-box models which only learn the annotation distribution instead of the logical thinking and diagnostic experience of high-level medical experts, the latter is more credible and interpretable. To alleviate the above issues, we propose a novel Segment-Like-A-Doctor (SLAD) framework to learn the reliable clinical thinking and experience for pancreas and pancreatic cancer segmentation on CT images. Specifically, SLAD aims to simulate the essential logical thinking and experience of doctors in the progressive diagnostic stages of pancreatic cancer: organ, lesion and boundary stage. Firstly, in the organ stage, an Anatomy-aware Masked AutoEncoder (AMAE) is introduced to model the doctors' overall cognition for the anatomical distribution of abdominal organs on CT images by self-supervised pretraining. Secondly, in the lesion stage, a Causality-driven Graph Reasoning Module (CGRM) is designed to learn the global judgment of doctors for lesion detection by exploring topological feature difference between the causal lesion and the non-causal organ. Finally, in the boundary stage, a Diffusion-based Discrepancy Calibration Module (DDCM) is developed to fit the refined understanding of doctors for uncertain boundary of pancreatic cancer by inferring the ambiguous segmentation discrepancy based on the trustworthy lesion core. Experimental results on three independent datasets demonstrate that our approach boosts pancreatic cancer segmentation accuracy by 4%–9% compared with the state-of-the-art methods. Additionally, the tumor-vascular involvement analysis is also conducted to verify the superiority of our method in clinical applications. Our source codes will be publicly available at <https://github.com/ZouLiwen-1999/SLAD>.

1. Introduction

Pancreatic cancer is one of the most lethal malignant tumors and ranks as the seventh-highest leading cause of cancer-related fatalities worldwide (Rawla et al., 2019). It is characterized by delayed diagnosis, difficult treatment and high mortality (Kamisawa et al., 2016). The overall five-year survival rate among patients is less than 8% (Chhoda et al., 2019). Pancreatic cancer always shows growing invasion potential on the disease progression, which directly affects the clinical

decision-making such as the tumor resectability evaluation. The Dutch Pancreatic Cancer Group (DPCG) classifies tumor resectability of pancreatic cancer based on the degrees of tumor-vessel contact (Versteijne et al., 2016). Fig. 1 presents the visual comparison of pancreatic cancer with different resectability on contrast computed tomography (CT) images. Therefore, accurate and reliable segmentation of pancreatic cancer plays an extremely important role in the process of diagnosis and treatment.

* Corresponding author at: Department of Pancreatic Surgery, Nanjing Drum Tower Hospital, Nanjing, 210008, China.

** Corresponding author.

E-mail addresses: df21350003@mail.nju.edu.cn (Z. Cai), xpyang@nju.edu.cn (X. Yang).

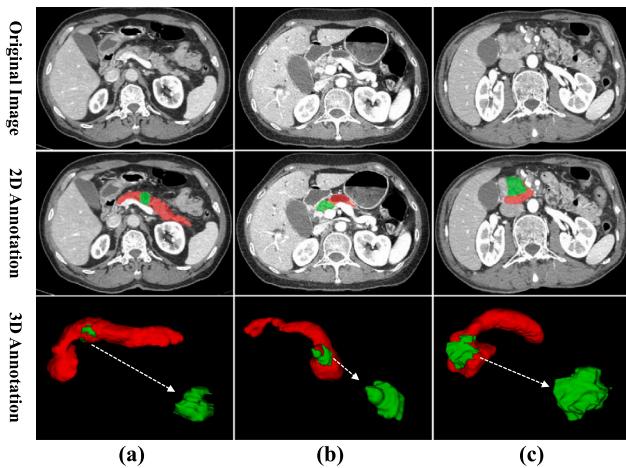


Fig. 1. Visual comparison of pancreatic cancers with different resectability on CT images. (a) Resectable pancreatic cancer. (b) Borderline resectable pancreatic cancer. (c) Locally advanced pancreatic cancer. Pancreas and tumor are marked in red and green, respectively.

Recently, certain deep learning-based methods have been tentatively applied to pancreatic cancer segmentation. Zhao et al. (2021) proposed a holistic segmentation-mesh-classification network for pancreatic mass segmentation by utilizing the geometry and location information. Chen et al. (2021) proposed a model-driven deep learning method for pancreatic cancer segmentation based on spiral transformation. Li et al. (2023b) proposed a 3D fully convolutional neural network with three temperature guided modules to realize joint segmentation of the pancreas and tumors. Qu et al. (2023) proposed a transformer guided progressive fusion network that utilizes the global representation for 3D pancreas and pancreatic mass segmentation. Generally, the above methods have made various attempts but suffered from two perspectives. First, most of existing works struggled to handle difficult cases in real-world clinical data due to the challenges brought by the tiny size, irregular shape and extremely uncertain boundary of the pancreatic cancer (Wang et al., 2021; Cao et al., 2023). Second, almost all the previous works were established on the black-box models which only learned the annotation distribution and lacked credibility and interpretability (Yao et al., 2023; Zhao et al., 2024).

Therefore, in this work, a novel Segment-Like-A-Doctor (SLAD) framework is proposed to learn the reliable clinical thinking and experience for pancreas and pancreatic cancer segmentation on CT images. Fig. 2 illustrates the main idea of our proposed method. In short, our SLAD aims to simulate the essential logical thinking of doctors in the progressive diagnostic stages of pancreatic cancer: organ, lesion and boundary stage, each of which contains rich and credible medical experience for tumor analysis. Specifically, when a high-level medical expert segments the pancreatic cancer, he or she always first makes an overall assessment of the abdominal organs on CT image to obtain a general cognition of the anatomical distribution, so as to preliminarily locates the pancreas (Zhou et al., 2017). Then, a global judgment within the pancreas volumes is made for detecting the core part of the pancreatic cancer by seeking the essential feature difference such as the intensity-based heterogeneity and shape-based abnormality (Li et al., 2023a). Finally, doctors usually calibrate the tumor boundary based on the trustworthy lesion core by exploring additional clinic-related imaging information such as the tumor-vascular involvement or peri-neural invasion (Mahmoudi et al., 2022). Thus, we establish our SLAD by learning the above reliable clinical thinking and experience with artificial neural networks to achieve more accurate and credible segmentation for pancreatic cancer.

The main contributions of our work can be summarized as follows:

- We propose a novel Segment-Like-A-Doctor (SLAD) framework to learn the reliable clinical thinking and experience in the progres-

sive diagnostic stages for more accurate and credible segmentation of pancreatic cancer.

- We introduce an Anatomy-aware Masked AutoEncoder (AMAE) to model the doctors' overall cognition for the anatomical distribution of abdominal organs on CT images by self-supervised pretraining.
- We design a Causality-driven Graph Reasoning Module (CGRM) to simulate the global judgment of doctors for lesion detection by exploring topological feature difference between the causal lesion and the non-causal organ.
- We develop a Diffusion-based Discrepancy Calibration Module (DDCM) to fit the refined understanding of doctors for uncertain boundary of pancreatic cancer by inferring the ambiguous segmentation discrepancy based on the trustworthy lesion core.
- Experimental results on three independent datasets demonstrate that our approach boosts pancreatic cancer segmentation accuracy by 4%–9% compared with the state-of-the-art methods. Additionally, the tumor-vascular involvement analysis is also conducted to verify the superiority of our method in clinical applications.

2. Related work

2.1. Self-supervised learning in medical images

Self-supervised learning (SSL) schemes have been developed to learn a generic representation of unlabeled data, and the obtained models can have faster convergence and higher accuracy when finetuned for downstream tasks (Xing et al., 2024). Recently, a growing body of work is exploring to apply the idea of masked image modeling (MIM) for self-supervised learning. Masked Autoencoder (MAE) (He et al., 2022), built on Vision Transformer (ViT) architecture, masks random image patches and reconstructs the missing pixels by an asymmetric transformer autoencoder. Xing et al. (2024) proposed a hybrid masked image modeling framework for 3D medical image segmentation pretraining by introducing a multi-level SSL strategy that learns semantic features of 3D medical images from the pixel-level, region-level, and sample-level, respectively. Although promising results have been demonstrated, most of these MIM methods only considered the intensity-based reconstruction pretext task while largely ignored the importance of integrating high-level semantics such as the essential anatomical information into model representations, which heavily hinders the performance of downstream tasks (Li et al., 2024; Bai et al., 2024). On the other hand, position-related SSL methods towards model pretraining in terms of structure distribution are also explored in a number of previous works. Haghghi et al. (2021) proposed an SSL model based on transferable visual words to learn common visual representation and leverage the semantics associated with the recurrent anatomical patterns across medical images. Wu et al. (2024c) proposed a volume contrast framework to leverage the contextual position priors for pretraining by predicting contextual positions. However, most of these position-related SSL methods preserving high-level semantics did not contain enough local information, which is vital in medical image analysis especially for image-based diagnosis and tumor segmentation (Zhou et al., 2023b; Taher et al., 2024). In order to address the importance of integrating pixel-level information into the high-level representations with rich semantics, we introduce our Anatomy-aware Masked AutoEncoder (AMAE) for simultaneously prelearning the intensity-based texture information and anatomy-aware position distribution of CT images to model the doctors' overall cognition for the anatomical distribution of abdominal organs.

2.2. Graph reasoning network

Graph-based methods have been very popular in recent years and shown to be an efficient way of relation reasoning. Wang and Gupta

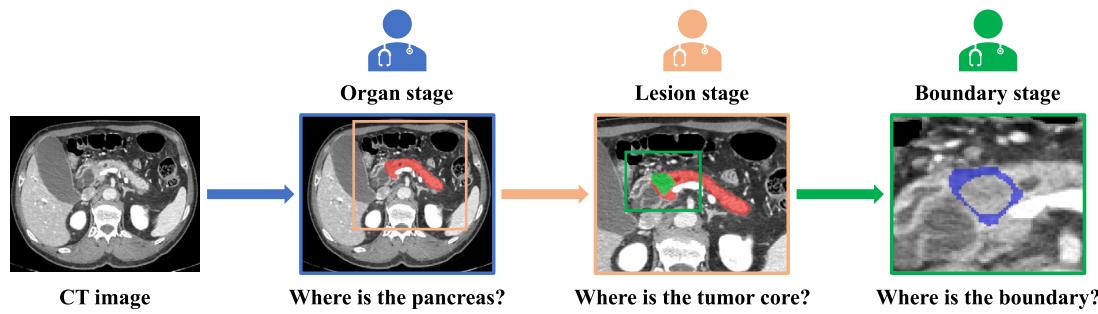


Fig. 2. Illustration of the main idea of the proposed method which aims to simulate the essential logical thinking and experience of doctors in the progressive diagnostic stages of pancreatic cancer: organ, lesion and boundary stage. The blue, orange and green rectangles denote the vision attention of the doctors on the organ, lesion and boundary stage, respectively. Pancreas, tumor and ambiguous region around the tumor boundary are marked in red, green and blue, respectively.

(2018) proposed to use Graph Convolution Network (GCN) (Kipf and Welling, 2016) to capture relations between objects in video recognition tasks. Chen et al. (2019) presented a global reasoning approach that can be implemented by projecting information from the coordinate space to nodes in an interaction space graph. However, most previous graph-based reasoning works focused on natural image analysis and novel graph reasoning strategies for medical images remain largely unexplored (Huang et al., 2024). Latterly, Meng et al. (2021) built a GCN-based deep learning framework with multiple graph reasoning modules to leverage both region and boundary features in an end-to-end manner for polyp image segmentation. Meanwhile, Wang et al. (2022) contributed a contour-guided graph reasoning network for capturing the semantic relations between object regions and contours for ambiguous biomedical image segmentation. Nevertheless, these graph-based reasoning works for biomedical image segmentation only considered the region-contour relation of the lesion, while ignored the essential logical thinking and experience of doctors for lesion detection which is the causality relationship between causal features (generic lesion) and non-causal features (normal organ). Recent works have demonstrated the importance and effectiveness of the causality concept in tumor diagnosis (Li et al., 2023a; Qu et al., 2024). Therefore, in this work, we design our Causality-driven Graph Reasoning Module (CGRM) for exploring the topological feature difference between the normal and abnormal regions by graph-based relation reasoning, aiming to learn the global judgment of doctors for lesion detection.

2.3. Calibration in medical image segmentation

Calibration concept refers to obtaining more accurate and reliable results based on one or more preliminary predictions with uncertainty. It has been explored in many computer vision tasks, such as semantic segmentation and saliency detection. In medical segmentation, calibration is usually implemented by estimating the uncertain regions (false positive and false negative predictions) and refining the previous predictions to the final results based on these estimated regions. Zhao et al. (2022) proposed a cascaded two-stage U-Net model to take the ambiguous region information (referred as distraction region) into account for 3D lung tumor segmentation. Ren et al. (2024) introduced a framework to model the discrepancy between the outputs of a U-Net like segmentation model and the ground truth by the denoising diffusion probabilistic model (DDPM) (Ho et al., 2020) for calibrating medical image segmentation. However, small object segmentation remains a significant challenge for deep learning techniques (Liu et al., 2021). Additionally, the ambiguous boundary regions of the pancreatic tumor are much smaller than the tumor itself, which makes both convolutional neural networks (CNNs) and diffusion models perform poorly in predicting tiny discrepancy regions without additional regularization (Zou et al., 2023; Rahman et al., 2023). Recently, signed distance function (SDF) loss has been adopted for medical image segmentation and demonstrated the advantage in reducing small noise predictions (Ma et al.,

2020; Wang et al., 2023). Therefore, we develop our Diffusion-based Discrepancy Calibration Module (DDCM) combining DDPM with SDF regularization to fit the refined understanding of doctors for uncertain boundary of pancreatic cancer.

3. Method

3.1. Overview

As shown in Fig. 3, we propose a novel Segment-Like-A-Doctor (SLAD) framework to learn the reliable clinical thinking and experience for pancreas and pancreatic cancer segmentation on CT images. Our SLAD aims to simulate the essential logical thinking and experience of doctors in the progressive diagnostic stages of pancreatic cancer: organ, lesion and boundary stage. Mathematically, Let the pretraining, training and testing set be denoted by $D^p = \{X_k^p\}_{k=1}^{N_p}$, $D^r = \{X_i^r, Y_i^r\}_{i=1}^{N_r}$ and $D^s = \{X_j^s, Y_j^s\}_{j=1}^{N_s}$, where $X_k^p, X_i^r, X_j^s \in \mathbb{R}^{H \times W \times D}$ are the input CT images, $Y_i^r, Y_j^s \in \{0, 1, 2\}^{H \times W \times D}$ are the segmentation labels, label 1 and 2 represent pancreas and pancreatic cancer, respectively, N_p, N_r, N_s are the numbers of the datasets.

Firstly, for learning the organ stage thinking of doctors, an Anatomy-aware Masked AutoEncoder (AMAE) parameterized by $F_o(\cdot; \theta_E, \theta_D)$ is introduced to model the doctors' overall cognition for the anatomical distribution of abdominal organs on CT images by self-supervised pretraining. AMAE is established on a ViT-based encoder-decoder architecture, θ_E and θ_D denote the parameters of the encoder and decoder, respectively. We pretrain the AMAE on the dataset D^p to calculate the optimal encoding parameters θ_E^* which contains rich anatomical prior on abdominal organs. Then, we freeze the encoder θ_E^* and fine-tune the decoder to the optimal parameters θ_D^* on the training dataset D^r . Thus, a coarse segmentation \hat{Y}^o for the pancreas and pancreatic cancer in the organ stage is obtained.

Secondly, for simulating the lesion stage logic of the doctors, a Causality-driven Graph Reasoning Module (CGRM) parameterized by $F_l(\cdot; \eta_E, \eta_D)$ is designed to learn the global judgment of doctors for lesion detection by exploring topological feature difference between the causal lesion and the non-causal organ. We extract the causal and non-causal features based on the coarse segmentation \hat{Y}^o and project them onto the graph domain followed by a graph-based relation reasoning. Then, a fined segmentation \hat{Y}^l in the lesion stage is predicted by decoding the reprojected vertex features.

Finally, for modeling the boundary stage experience, a Diffusion-based Discrepancy Calibration Module (DDCM) parameterized by $F_b(\cdot; \xi_E, \xi_D)$ is developed to fit the refined understanding of doctors for uncertain boundary of pancreatic cancer by inferring the ambiguous segmentation discrepancy based on the trustworthy lesion core. We add successive noise on the segmentation discrepancy between \hat{Y}^l and the ground truth which indicates the ambiguous region around the tumor boundary. A regularized DDPM is developed to denoise the discrepancy map and calibrate the final refined segmentation \hat{Y}^b in the boundary stage.

The details of each component are illustrated below.

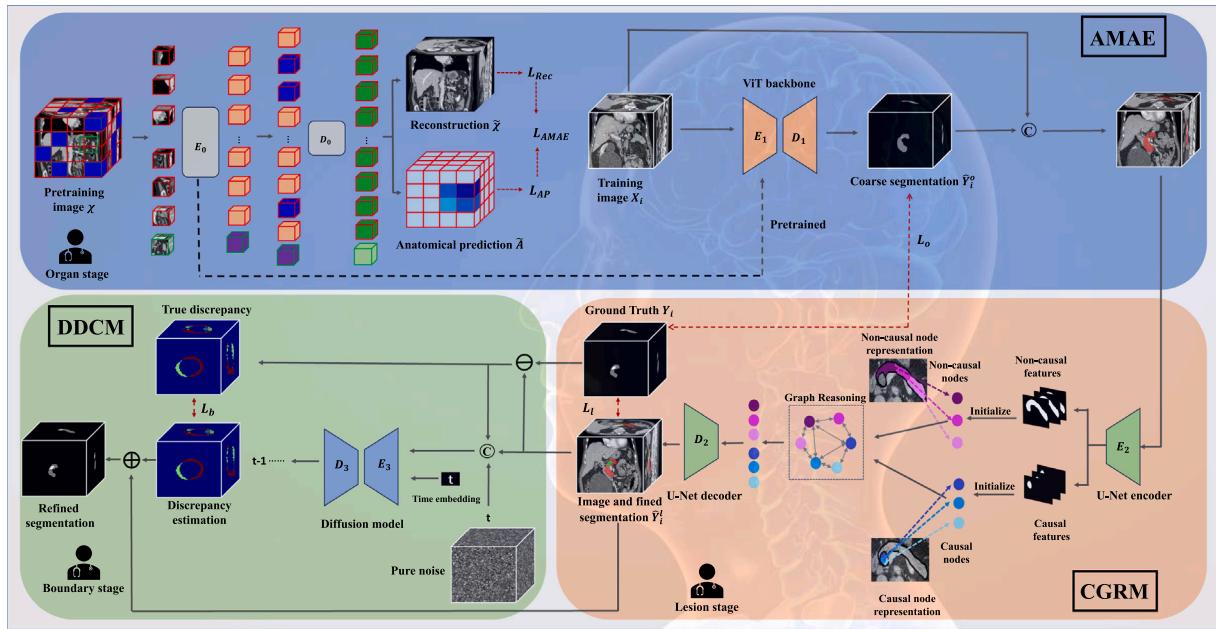


Fig. 3. Pipeline of our proposed Segment-Like-A-Doctor (SLAD) framework to learn the reliable clinical thinking and experience for pancreas and pancreatic cancer segmentation. SLAD aims to simulate the essential logical thinking of doctors in the progressive diagnostic stages of pancreatic cancer: organ, lesion and boundary stage. For the organ stage, an Anatomy-aware Masked AutoEncoder (AMAE) is introduced model the doctors' overall cognition for the anatomical distribution of abdominal organs on CT images by self-supervised pretraining. For the lesion stage, a Causality-driven Graph Reasoning Module (CGRM) is designed to learn the global judgment of doctors for lesion detection by exploring topological feature difference between the causal lesion and the non-causal organ. As for the final boundary stage, a Diffusion-based Discrepancy Calibration Module (DDCM) is developed to fit the refined understanding of doctors for uncertain boundary of pancreatic cancer by inferring the ambiguous segmentation discrepancy based on the trustworthy lesion core.

3.2. Anatomy-aware Masked AutoEncoder (AMAE)

In the organ stage of the doctors' clinical thinking and experience, it is essential to have a comprehensive cognition for the anatomical structure and distribution of abdominal organs, which is beneficial for discovering the diseased organs and lesions. Therefore, we introduce our Anatomy-aware Masked AutoEncoder (AMAE) to simulate this behavior by self-supervised pretraining. The objective of self-supervised pretraining is to effectively encode anatomical information of the human body in unlabeled image data. In this work, our pretraining is accomplished with two proxy tasks: image reconstruction and anatomical position prediction, which learn the comprehensive semantic representation and anatomical distribution pattern of the medical images, respectively. Fig. 4 illustrates the specific structure and pretraining process of our AMAE.

Given an input volume $\chi \in D^p$, we crop it into n non-overlap sub-volumes $\{r_1, \dots, r_n\}$ which cover the whole input volume. On the one hand, we randomly crop a query volume \tilde{r} with the same size of sub-volumes from χ for the anatomical prediction task. On the other hand, we also randomly select the divided sub-volumes with ratio γ to do the masking for the reconstruction task. The random masking divides the sub-volumes into two groups: the visible sub-volume set R^V and the masked sub-volume set R^M . Subsequently, we employ a ViT-based encoder E_0 to extract the features \tilde{z} , Z^V and Z^M for \tilde{r} , R^V and R^M , respectively. Z^V and Z^M present the prototype-level features from different anatomical patterns. Z^V is concatenated with the mask token and position embedding B^p to obtain the reconstruction image $\hat{\chi} = \{\tilde{r}_1, \dots, \tilde{r}_n\}$ by the ViT decoder D_0 . Then, we can calculate the reconstruction loss based on the L_2 distance between the reconstructed region r_i and the target region \tilde{r}_i :

$$L_{Rec} = \frac{1}{|R^M|} \sum_{r_i \in R^M} \|r_i - \tilde{r}_i\|_2. \quad (1)$$

Following previous SSL works (Wu et al., 2024c; Chen et al., 2020), a projector with linear layers and 3D adaptive average pooling is used to project \tilde{z} , Z^V and Z^M into latent features $q \in \mathbb{R}^{1 \times C}$ and $p_1, \dots, p_n \in \mathbb{R}^{1 \times C}$, where C is the number of channels and p_i is the projected

feature of each sub-volume. Because each of the sub-volumes represents a unique anatomical position of abdominal structure, the anatomical prediction task is equivalent to finding the optimal sub-volumes which are most similar to the randomly given query volume. Therefore, we calculate the anatomical prediction \tilde{A} composed of similarity logits a_i between q and p_i . Specifically, we use cosine similarity to compute a_i as follows:

$$a_i = CosSim(q, p_i) = \frac{q \cdot p_i}{\|q\| \|p_i\|}. \quad (2)$$

The proportions of the overlap area between \tilde{r} and r_i are assigned as anatomical labels $y = y_1, \dots, y_n$ where

$$y_i = \frac{|\tilde{r} \cap r_i|}{|\tilde{r}|}. \quad (3)$$

Therefore, the anatomical prediction loss can be defined as

$$L_{AP} = -\frac{1}{n} \sum_{i=1}^n \log(1 - |y_i - a_i|). \quad (4)$$

The overall loss function for pretraining the AMAE is calculated as follows:

$$L_{AMAE} = w_{Rec} L_{Rec} + w_{AP} L_{AP}, \quad (5)$$

where w_{Rec} and w_{AP} are balance factors for the reconstruction and anatomical prediction term, respectively.

When we complete the pretraining for AMAE on D^p , the optimal parameters θ_E^* of the encoder E_0 are inherited to the same encoder E_1 of the subsequent segmentation model $F_o(\cdot; \theta_E, \theta_D)$ in the organ stage to fine-tune the decoder head D_1 on the training dataset D^r . Then, a coarse segmentation \hat{Y}_o for the pancreas and pancreatic cancer in the organ stage can be obtained as follows:

$$\hat{Y}_o = F_o(X_i; \theta_E^*, \theta_D), \quad (6)$$

where $X_i \in D^r$ is the input training image. The segmentation loss L_o in the organ stage is based on the Dice and cross entropy (CE) loss.

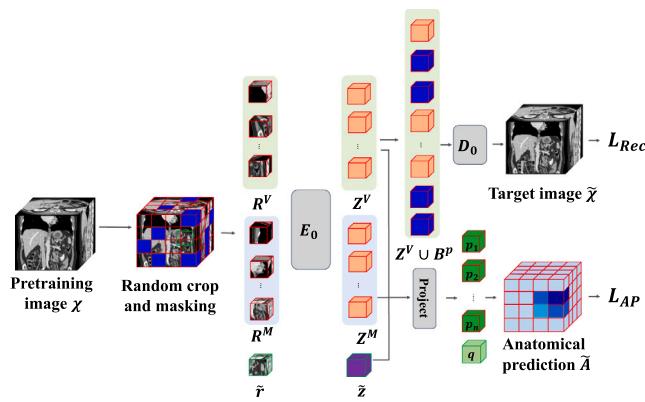


Fig. 4. Specific structure and pretraining process of the Anatomy-aware Masked AutoEncoder (AMAE).

Therefore, the fine-tuning for decoder D_1 can be formulated as:

$$\theta_D^* = \arg \min_{\theta_D} \sum_{i=1}^{N_r} L_o(\hat{Y}_i^o, Y_i), \quad (7)$$

where

$$L_o(\hat{Y}_i^o, Y_i) = L_{dice}(\hat{Y}_i^o, Y_i) + L_{ce}(\hat{Y}_i^o, Y_i), \quad (8)$$

and Y_i is the segmentation label for X_i .

3.3. Causality-driven Graph Reasoning Module (CGRM)

In the lesion stage of the doctors' clinical thinking and experience, it is natural to perform global reasoning on abnormal (causal) and normal (non-causal) features based on the clues from the previous diagnostic stage within the organ region of interest (ROI) for inferring the core part of the lesion. Therefore, in this section, we present our Causality-driven Graph Reasoning Module (CGRM) to explore the global feature difference between the normal and abnormal regions. The detailed architecture of CGRM is shown in Fig. 5.

The coarse segmentation \hat{Y}_i^o is divided into two binary segmentation maps of pancreas and pancreatic cancer, called non-causal map \mathcal{N} and causal map \mathcal{C} , respectively. Given the input feature map $\mathcal{F} \in \mathbb{R}^{L_1 \times C_1}$ obtained by the encoder $E_2(\cdot; \eta_E)$ of the segmentation model $F_l(\cdot; \eta_E, \eta_D)$ in the lesion stage experience, where $L_1 = H_1 \times W_1 \times D_1$ is the spatial resolution, C_1 is the number of channels. We build two projection matrices, P_c and P_n , by mapping \mathcal{F} to the vertices of a graph guided by \mathcal{C} and a graph guided by \mathcal{N} , respectively. Specifically, the projection matrices can be calculated as follows:

$$P_c = SoftMax(\mathcal{P}(\phi(\mathcal{F}) \odot \mathcal{C}) \cdot \phi(\mathcal{F})^T), \quad (9)$$

$$P_n = SoftMax(\mathcal{P}(\phi(\mathcal{F}) \odot \mathcal{N}) \cdot \phi(\mathcal{F})^T), \quad (10)$$

where \mathcal{P} denotes the average pooling operation, $\phi(\cdot)$ is the $1 \times 1 \times 1$ convolution operation, \odot represents the Hadamard product. After the projection matrices are obtained, we project the input feature map \mathcal{F} onto the causality-driven graph domain to obtain the topological representations of the causal region G_c and the non-causal region G_n , which can be formulated as follows:

$$G_c = P_c \phi(\mathcal{F}), \quad (11)$$

$$G_n = P_n \phi(\mathcal{F}). \quad (12)$$

It should be pointed that the vertices of the graph represent the independent subspaces in the high-dimensional latent space rather than separate regions in the image space. Subsequently, we pick up the connectivity between vertices from G_c and G_n . Meanwhile, we learn higher-level semantic information by propagating information between

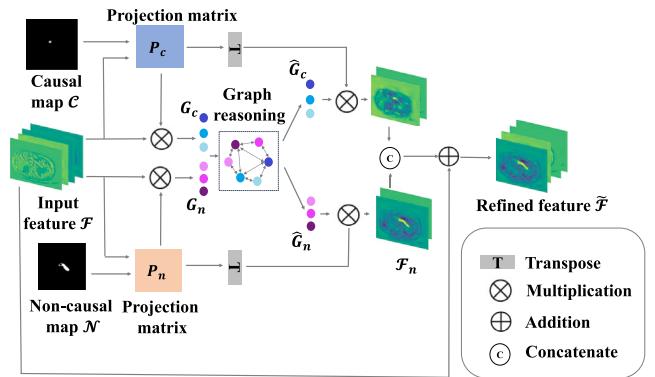


Fig. 5. Specific structure of the Causality-driven Graph Reasoning Module (CGRM).

vertices to reason about these potential relations. This strategy can be elegantly implemented with a one-layer GCN. To be specific, the vertex features G_c and G_n are fed into a first-order approximation of the spectral graph convolution:

$$\hat{G} = ReLu((I - A)Cat(G_c, G_n)W_G), \quad (13)$$

where A denotes the adjacency matrix encoding the connectivity of the graph to be learned and is randomly initialized, $Cat(\cdot)$ is the concatenate operation, W_G is the weight of the GCN, I represents the identity matrix. After the relation reasoning, we can model the global judgment of doctors for the lesion detection by the refined topological features $\hat{G} = [\hat{G}_c, \hat{G}_n]$. Then, we reproject the extracted vertex features to be the original pixel grids of the graph domain. Following the previous works (Chen et al., 2019; Wang et al., 2022), we consider the reproject matrix as a transpose matrix. The reproject can be defined as follows:

$$\mathcal{F}_c = \phi(P_c^T \hat{G}_c), \quad (14)$$

$$\mathcal{F}_n = \phi(P_n^T \hat{G}_n). \quad (15)$$

Therefore, a fined segmentation of pancreas and pancreatic cancer \hat{Y}^l in the lesion stage can be obtained by decoding the refined features with decoder D_2 :

$$\hat{Y}^l = D_2(\phi(Cat(\mathcal{F}_c, \mathcal{F}_n)); \eta_D). \quad (16)$$

The segmentation loss in the lesion stage is also established on the Dice and CE loss. We optimize the CGRM parameterized by $F_l(\cdot; \eta_E, \eta_D)$ as follows:

$$\eta_E^*, \eta_D^* = \arg \min_{\eta_E, \eta_D} \sum_{i=1}^{N_r} L_l(\hat{Y}_i^l, Y_i), \quad (17)$$

where

$$L_l(\hat{Y}_i^l, Y_i) = L_{dice}(\hat{Y}_i^l, Y_i) + L_{ce}(\hat{Y}_i^l, Y_i). \quad (18)$$

3.4. Diffusion-based Discrepancy Calibration Module (DDCM)

As for the final boundary stage of the doctors' clinical thinking and experience, it is crucial to have a refined calibration of lesion boundary by exploring latent clinic-related imaging information such as the tumor-vascular involvement or peri-neural invasion. Therefore, in this part, we propose our Diffusion-based Discrepancy Calibration Module (DDCM) to fit the refined understanding of doctors for uncertain boundary of pancreatic cancer by inferring the ambiguous segmentation discrepancy based on the trustworthy lesion core. The details of our DDCM are illustrated in Fig. 6. Specifically, the designed DDCM builds upon the DiffUNet (Xing et al., 2023), which uses conditional Denoising

Diffusion Probabilistic Model (DDPM) (Ho et al., 2020). DDPM can take a randomly drawn noise and successively denoise it over several steps to generate a sample from the distribution of images once it has been trained. Previous works have verified the effectiveness of the DDPM for calibration in medical image segmentation (Ren et al., 2024; Wu et al., 2024a,b).

Mathematically, we first obtain the true segmentation discrepancy d_0 between segmentation result \hat{Y}^l from the lesion stage and the ground truth Y by the subtraction operation. Then, we add successive t step noise ϵ on the discrepancy map d_0 to get the noised map:

$$d_t = \sqrt{\bar{\alpha}_t} d_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon, \quad (19)$$

where

$$\bar{\alpha}_t = \prod_{s=1}^t (1 - \beta_s), \quad (20)$$

and β_1, \dots, β_t are the forward process variances. Meanwhile, we extract the latent features \tilde{I}_f of the original image X and segmentation result Y^l by the encoder E_3 of the diffusion model $F_b(\cdot; \xi_E, \xi_D)$ in the boundary stage experience:

$$\tilde{I}_f = E_3(Cat(X, Y^l); \xi_E). \quad (21)$$

\tilde{I}_f is the essential basis containing rich information of the lesion core to infer the ambiguous boundary discrepancy. Then, we employ the decoder D_3 to conduct the denoising process and obtain the discrepancy estimation:

$$\hat{d}_0 = D_3(Cat(X, d_t, t, \tilde{I}_f), \quad (22)$$

where t is the time embedding. However, in practice, we observe that both CNNs and diffusion models perform poorly in predicting tiny discrepancy regions and usually lead to unreasonable outlier prediction. To alleviate this problem, we adopt the signed distance function (SDF) regularization to alleviate this problem. Specially, the SDF map of a binary mask M can be calculated as follows:

$$SDF(x, M) = \begin{cases} \min_{m \in \partial M} -d(x, m), & x \in M_{in}, \\ 0, & x \in \partial M, \\ \min_{m \in \partial M} d(x, m), & x \in M_{out}, \end{cases} \quad (23)$$

where $x \in \Gamma$ denotes the location on the map, $d(x, m)$ is the Euclidean distance between location x and m . M_{in} , M_{out} and ∂M denote the inside, the outside and the boundary of the object, respectively. Then we get the SDF regularization loss by calculating the mean square error (MSE) between the SDFs of discrepancy estimation \hat{d}_0 and the true discrepancy d_0 :

$$L_{sdf}(d_0, \hat{d}_0) = \frac{\sum_{x \in \Gamma} |SDF(x, d_0) - SDF(x, \hat{d}_0)|^2}{|\Gamma|}. \quad (24)$$

The Dice and CE loss are also adopted for training the DDCM. Therefore, we optimize the DDCM parameterized by $F_b(\cdot; \xi_E, \xi_D)$ as follows:

$$\xi_E^*, \xi_D^* = \arg \min_{\xi_E, \xi_D} \sum_{i=1}^{N_r} L_b(\hat{d}_i, d_i), \quad (25)$$

where

$$L_b(\hat{d}_i, d_i) = L_{dice}(\hat{d}_i, d_i) + L_{ce}(\hat{d}_i, d_i) + L_{sdf}(\hat{d}_i, d_i), \quad (26)$$

$$d_i = Y_i - \hat{Y}_i^l, \quad (27)$$

$$\hat{d}_i = F_b(X_i, \hat{Y}_i^l; \xi_E, \xi_D). \quad (28)$$

3.5. Inference

When we complete all the training processes on the three diagnostic stages, we can infer the segmentation result of the input image $X_j \in D^s$ as follows:

$$\hat{Y}_j^o = F_o(X_j; \theta_E^*, \theta_D^*), \quad (29)$$

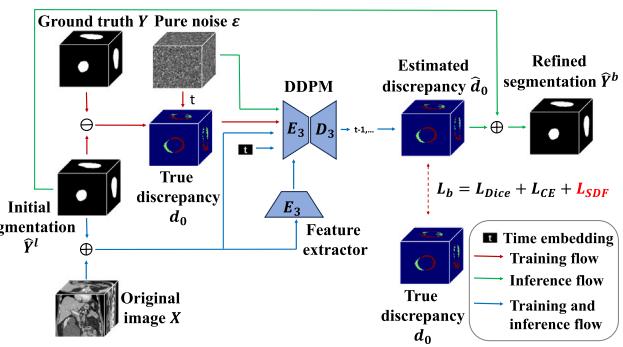


Fig. 6. Specific structure of the Diffusion-based Discrepancy Calibration Module (DDCM).

$$\hat{Y}_j^l = F_l(X_j, \hat{Y}_j^o; \eta_E^*, \eta_D^*), \quad (30)$$

$$\hat{Y}_j^b = \hat{Y}_j^l + \hat{d}_j. \quad (31)$$

We employ the calibrated segmentation \hat{Y}_j^b as the final prediction of our method. Thus, our SLAD has completed the simulation of doctors' logical thinking and medical experience in the progressive diagnostic stages.

3.6. Implementation details

The network backbones of AMAE, CGRM and DDCM are based on the SwinUNETR (Hatamizadeh et al., 2021), generic UNet (Ronneberger et al., 2015) and Denoising UNet (Xing et al., 2023), respectively. For pretraining, we set 100K training steps and employ AdamW optimizer with cosine learning rate scheduler. For fine-tuning, we run 1000 epochs with batch size of 2 on the training set. Stochastic gradient descent (SGD) optimizer 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 policy in Isensee et al. (2021). We set the sub-volume size to $32 \times 32 \times 32$, set the masking ratio γ to 75%, the number of feature channels C to 2048. All the pretraining and training procedures have been performed on five NVIDIA A100 GPUs, each of which has 40 GB memory.

4. Experiment

4.1. Datasets

Pretraining dataset: We collect a total of 4010 3D CT volumes to construct our pretraining dataset. It comes from four independent sources: the public FLARE21 dataset (Ma et al., 2022), public NIH dataset (Roth et al., 2016), self-collected CT dataset from General Hospital of Eastern Theater Command of China (GHETC), and self-collected CT dataset from Jiangsu Province Hospital of Chinese Medicine (JHCM).

Downstream segmentation datasets: To evaluate the effectiveness of our SLAD for 3D pancreas and pancreatic cancer segmentation on CT images, we conduct the downstream segmentation experiments on five independent sources: Nanjing Drum Tower Hospital (NDTH), JiangSu Province Hospital (JSPH), the First Affiliated Hospital of Wannan Medical College (FHW), Jiangsu Province Hospital of Chinese Medicine (JHCM) and the First Hospital of ChangZhou (FHCZ). All the CT scans are collected from patients with surgical pathology confirmed pancreatic cancer. All the CT images are stored as NIFTI files with an axial plane resolution of 512×512 , a slice number ranging from 182 to 1037, and an original spacing of $0.75 \times 0.75 \times 1.25 \text{ mm}^3$. Data preprocessing involves normalization and resampling, following the

Table 1

Detailed information of the utilized datasets in this work. The numbers in () represent the corresponding quantity of CT scans in each dataset. The numbers in [] stand for the volume sizes of the pancreatic tumor [mean \pm std, cm³] in the downstream segmentation datasets.

Split	Sources	CT Scans
Pretraining set	FLARE (1000), NIH (82), GHETC (2534), JHCM (394)	4010
Training set	NDTH (328) [20.3 \pm 25.7], JSPH (48) [10.1 \pm 8.2], FHWG (96) [10.4 \pm 10.1]	472
Internal testing set	NDTH (82) [24.0 \pm 32.5], JSPH (14) [7.0 \pm 4.1], FHWG (25) [8.8 \pm 9.1]	121
External testing set I	JHCM* (87) [13.3 \pm 17.8]	87
External testing set II	FHCZ (69) [18.3 \pm 12.8]	69
Overall	Five centers	749

*The CT scans from JHCM in the pretraining set are independent of those in the external testing set I.

Table 2

Quantitative comparison of segmentation and detection performance on three independent testing datasets. PS, TD and TS denote pancreas segmentation, tumor detection and tumor segmentation, respectively. Dice scores (mean \pm std, %) and 10% detection rates (%) are presented as segmentation and detection metrics, respectively. The best and second best performance are in red and blue, respectively.

Methods	Internal Testing Set (n = 121)			External Testing Set I (n = 87)			External Testing Set II (n = 69)		
	PS	TD	TS	PS	TD	TS	PS	TD	TS
SegResNet	71.3 \pm 12.8**	90.9	60.4 \pm 25.3**	67.5 \pm 11.8**	92.0	56.8 \pm 23.8**	63.6 \pm 11.3**	89.9	48.7 \pm 25.1**
ResUNet	77.6 \pm 13.1**	90.9	57.2 \pm 26.0**	72.7 \pm 11.1**	92.0	45.4 \pm 24.3**	71.8 \pm 11.5**	95.7	44.7 \pm 22.0**
nnUNet	89.8 \pm 10.0**	93.4	70.7 \pm 24.4**	88.3 \pm 9.9*	93.1	66.5 \pm 26.0**	87.7 \pm 7.1*	88.4	61.9 \pm 27.5**
nnFormer	77.0 \pm 13.1**	90.1	59.2 \pm 26.3**	73.9 \pm 11.9**	90.8	54.4 \pm 25.9**	71.2 \pm 14.3**	92.8	45.4 \pm 25.3**
UNETR	65.8 \pm 15.6**	88.4	49.6 \pm 26.0**	63.9 \pm 12.7**	89.7	49.6 \pm 23.4**	64.2 \pm 14.0**	82.6	41.6 \pm 25.8**
UNETR++	86.2 \pm 11.4**	83.5	63.7 \pm 31.3**	83.2 \pm 11.6**	81.6	58.1 \pm 31.1**	83.5 \pm 8.4**	79.7	53.5 \pm 31.0**
DiffUNet	88.7 \pm 8.8**	94.2	69.2 \pm 24.0**	86.8 \pm 10.4**	94.3	66.6 \pm 25.2**	84.7 \pm 7.4**	94.2	61.1 \pm 23.5**
SLAD (Ours)	91.1 \pm 9.5	94.2	75.0 \pm 22.7	89.5 \pm 9.9	94.3	71.3 \pm 24.0	89.2 \pm 8.2	95.7	70.9 \pm 21.4

** and * represent statistically significant $p < 0.01$ and $p < 0.05$ difference comparing with our method by student t-test, respectively.

strategy outlined in Isensee et al. (2021). Each CT scan has voxel-wise manual annotation of the pancreas and pancreatic cancer which was performed by two pancreatic imaging radiologists using the open-source application ITK-SNAP.¹ The NDTH, JSPH and FHWG datasets are divided as the training set and the internal testing set. The JHCM and FHCZ datasets are utilized as external testing sets I and II, respectively. More detailed information is provided in Table 1. Besides, in order to verify the effectiveness of our method in clinical applications, each case in the internal testing set, external testing set I and external testing set II has the annotated tumor resectability label and the degrees of tumor-vessel contact for tumor vascular involvement analysis. The annotation details are provided in Section A of the Supplementary Materials.

4.2. Evaluation metrics

For quantitative evaluation of the segmentation performance, we first employ the Dice similarity coefficient (Dice) to evaluate the overall segmentation accuracy, which can be calculated as follows:

$$Dice(X, Y) = \frac{2|X \cap Y|}{|X| + |Y|}, \quad (32)$$

where X and Y denote the prediction mask and ground truth mask, respectively. Then, the detection rate (DR) is also adopted to evaluate the accuracy of positioning, where a detection is considered successful if the intersection (between the ground truth and segmentation mask) over the ground truth is larger than 10% (counted as 1); otherwise, it is considered a misdetection (counted as 0). Moreover, the Hausdorff distance (HD) is used for evaluating the segmentation accuracy for boundaries, which can be calculated as follows:

$$h(X, Y) = \max_{x \in X, y \in Y} \min(x, y), \quad (33)$$

$$HD(X, Y) = \max(h(X, Y), h(Y, X)), \quad (34)$$

where $d(x, y)$ represents the Euclidean distance between x and y . For tumor-vascular analysis, we calculate the accuracy for tumor resectability and degree prediction.

4.3. Comparisons with state-of-the-art methods

In this section, we compare our SLAD with other state-of-the-art (SOTA) methods for medical image segmentation on our collected datasets. Specifically, we conduct a comparison between SLAD and seven known competitors: (1) SegResNet (Myronenko, 2019), (2) ResUNet (Diakogiannis et al., 2020) and (3) nnUNet (Isensee et al., 2021) are the mainstream segmentation baselines which are established on CNN. (4) nnFormer (Zhou et al., 2023a) is a strong self-configuration segmentation framework based on Transformer (Vaswani et al., 2017). (5) UNETR (Hatamizadeh et al., 2022) and (6) UNETR++ (Shaker et al., 2024) are general medical image segmentation frameworks which serially connect CNN and Transformer. (7) DiffUNet (Xing et al., 2023) integrates the diffusion model into a standard U-shaped architecture to extract semantic information from the input volume effectively, resulting in excellent pixel-level representations for medical volumetric segmentation. More implementation details can be found in Section B of the Supplementary Materials.

Segmentation results of the pancreas and pancreatic cancer in three testing datasets are presented in Table 2. Dice metric for pancreas segmentation (PS), DR metric for tumor detection (TD) and Dice metric for tumor segmentation (TS) are calculated. We also conduct the paired t-tests to calculate the statistical significance between our SLAD and the other competing methods. It can be observed that SLAD reaches the highest Dice scores for pancreas segmentation on all the three testing datasets, with statistically significant differences from other seven competing methods. For tumor detection, our SLAD achieves best performance on all three testing datasets, demonstrating the superiority of generalization ability and robustness compared to other methods. As for tumor segmentation, SLAD also outperforms other seven methods on all the three testing datasets with statistically significant differences. On these three datasets, the average Dice scores for tumor segmentation of SLAD are 4.3%, 4.7% and 9.0% higher than the sub-optimal results, respectively. Fig. 7 presents the performance distributions for the comparison experiments. It can be found that our SLAD has a more compact distribution with fewer outliers. Fig. 8 illustrates the visual comparisons. Fig. 8(a) and (b) present two cases with pancreatic tail cancer from internal testing dataset, (c) and (d) present two

¹ <http://www.itksnap.org/>

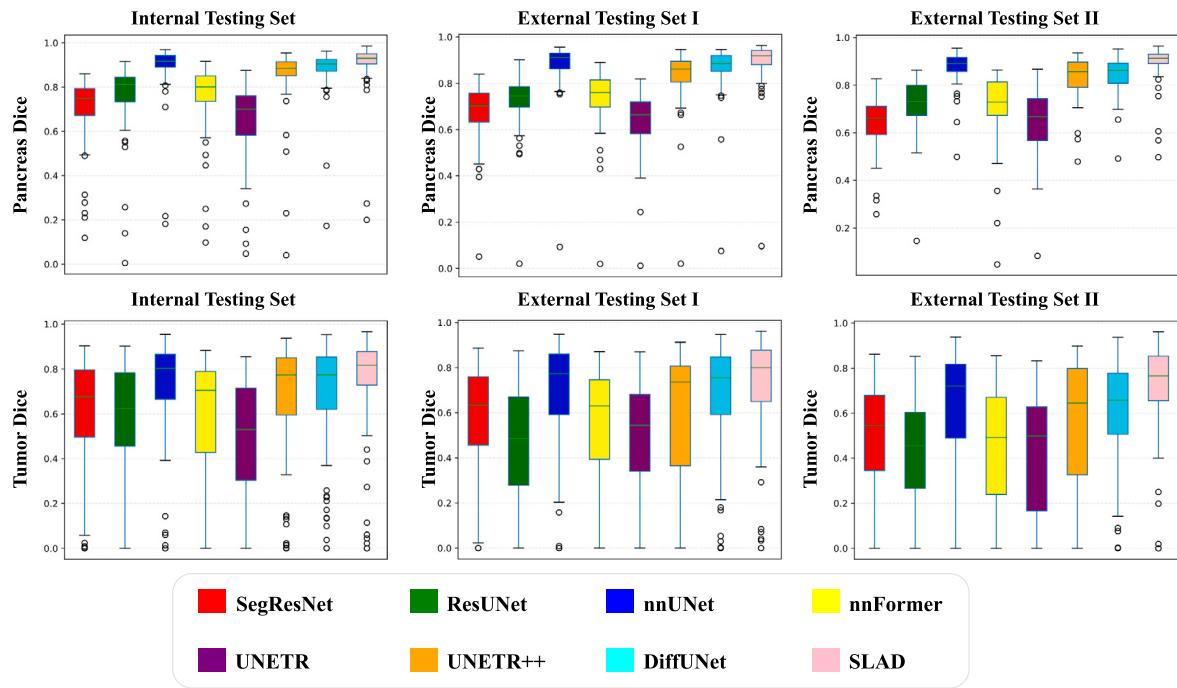


Fig. 7. Performance distributions of different methods on three testing datasets.

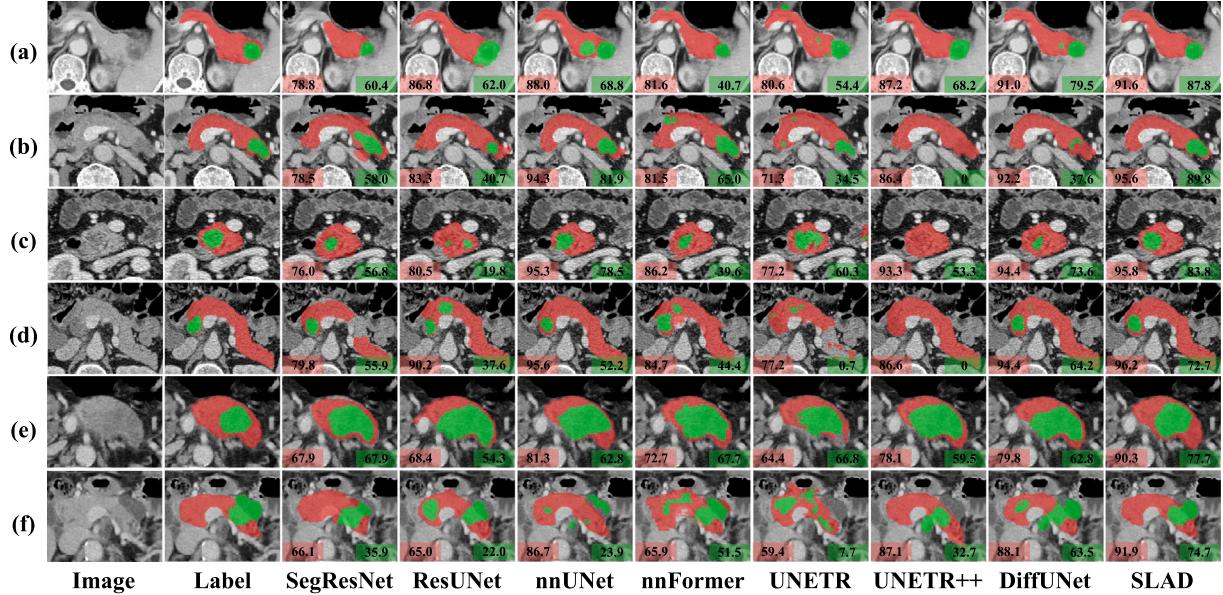


Fig. 8. Qualitative comparison of segmentation of pancreases (marked in red) and tumors (marked in green). (a) and (b) are taken from internal testing dataset, (c) and (d) are taken from external testing dataset I, (e) and (f) are taken from external testing dataset II. Six cases are shown in 2D slices in axial view. For each case, the Dice scores of the pancreas and tumor are presented in the red and green boxes below, respectively.

pancreatic head cancers from external testing dataset I, (e) and (f) present two pancreatic body cancers from external testing dataset II. It can be observed that our method well distinguishes pancreatic and extra-pancreatic tissues and gives the best tumor segmentation.

4.4. Ablation study

(1) The effects of individual components of SLAD

We conduct the ablation study for verifying the effectiveness of the three diagnostic stages of our proposed SLAD. As shown in Table 3,

row 1 presents the segmentation performance of the baseline model (SwinUNETR (Hatamizadeh et al., 2021)) in the organ stage without self-supervised pretraining. The proposed AMAE (row 2) contributes 4.0%, 7.4% and 6.7% Dice improvement with statistical significance for tumor segmentation on three testing datasets, respectively. And the DRs are improved by 3.3%, 3.4% and 4.3% on three datasets, respectively. The CGRM (row 3) contributes 5.2%, 7.2% and 5.8% Dice improvement with statistical significance for tumor segmentation on three testing datasets, respectively. And the DRs are improved by 3.3% and 3.4% on internal testing dataset and external testing dataset I, respectively. The HD metric also achieves significant improvement with both AMAE and

Table 3

Ablation study of the proposed SLAD method for pancreatic cancer segmentation on three independent testing datasets. Dice coefficient similarity (mean \pm std, %) and 10% detection rates (DR, %), 95% Hausdorff Distance (HD, mm) are presented. The best and second best performance are in red and blue, respectively.

AMAE	CGRM	DDCM	Internal Testing Set (n = 121)			External Testing Set I (n = 87)			External Testing Set II (n = 69)		
			Dice	DR	HD	Dice	DR	HD	Dice	DR	HD
✗	✗	✗	65.7 \pm 26.5**	90.1	60.2 \pm 54.5**	57.8 \pm 26.7**	89.7	121.4 \pm 58.6**	56.2 \pm 28.6**	89.9	80.0 \pm 51.1**
✓	✗	✗	69.7 \pm 24.3**	93.4	29.3 \pm 45.9**	65.2 \pm 24.0**	93.1	44.6 \pm 53.2**	62.9 \pm 25.4**	94.2	36.3 \pm 45.0**
✗	✓	✗	70.9 \pm 24.4**	93.4	26.6 \pm 39.3**	65.0 \pm 26.4**	93.1	38.3 \pm 47.5**	62.0 \pm 26.5**	89.9	39.6 \pm 43.2**
✓	✓	✗	73.6 \pm 23.1**	94.2	20.5 \pm 33.5**	68.9 \pm 26.0**	92.0	31.6 \pm 36.2**	67.6 \pm 24.6**	94.2	28.4 \pm 34.2**
✓	✓	✓	75.0 \pm 22.7	94.2	13.7 \pm 20.7	71.3 \pm 24.0	94.3	15.9 \pm 21.2	70.9 \pm 21.4	95.7	10.8 \pm 11.9

** represent statistically significant $p < 0.01$ difference comparing with method containing all three components by student t-test.

Table 4

Ablation study of the AMAE for pancreatic cancer segmentation on three independent testing datasets. Dice coefficient similarity (mean \pm std, %) and 10% detection rates (DR, %), 95% Hausdorff Distance (HD, mm) are presented. The best and second best performance are in red and blue, respectively. $w_{AP} = 0$ and $w_{Rec} = 0$ denote the supervised baseline without self-supervised pretraining.

w_{AP}	w_{Rec}	Internal Testing Set (n = 121)			External Testing Set I (n = 87)			External Testing Set II (n = 69)		
		Dice	DR	HD	Dice	DR	HD	Dice	DR	HD
0	0	65.7 \pm 26.5**	90.1	60.2 \pm 54.5**	57.8 \pm 26.7**	89.7	121.4 \pm 58.6**	56.2 \pm 28.6**	89.9	80.0 \pm 51.1**
0	1	68.5 \pm 25.5	91.7	30.1 \pm 43.0	61.4 \pm 25.2*	93.1	79.4 \pm 75.5**	61.2 \pm 24.2	94.2	52.8 \pm 49.5**
1	0	67.9 \pm 25.1	91.7	32.8 \pm 46.2	61.6 \pm 26.1*	92.0	86.3 \pm 81.0**	61.7 \pm 24.7	94.2	48.1 \pm 55.4**
1	1	69.7 \pm 24.3	93.4	29.3 \pm 45.9	65.2 \pm 24.0	93.1	44.6 \pm 53.2	62.9 \pm 25.4	94.2	36.3 \pm 45.0
1	0.5	68.6 \pm 26.5	91.7	26.5 \pm 39.3	63.8 \pm 25.8	92.0	67.8 \pm 70.0**	60.3 \pm 26.2	92.8	49.0 \pm 51.3**
1	1.5	69.6 \pm 25.8	92.6	32.5 \pm 50.7	64.4 \pm 25.2	93.1	54.6 \pm 66.1**	62.5 \pm 25.1	94.2	42.5 \pm 49.8*

** and * represent statistically significant $p < 0.01$ and $p < 0.05$ difference comparing with model of $w_{AP} = 1$ and $w_{Rec} = 1$ by student t-test, respectively.

CGRM on three datasets. With the fusion of AMAE and CGRM (row 4), the Dice scores and HDs are improved by 2.7%–3.9% and 22.9%–30.0% in the internal testing dataset, 2.7%–3.9% and 17.5%–29.1% in the external testing dataset I, 4.7%–5.6% and 21.8%–28.3% in the external testing dataset II, comparing to the individual one. Additionally, the DDCM (row 5) contributes 33.2%, 49.7% and 62.0% improvement in HD metrics on three testing datasets, respectively.

(2) The effects of loss weights of AMAE

We conduct ablation experiments to explore the effects of each component in the loss function (5) of the AMAE. Table 4 presents the quantitative results. By applying the reconstruction loss term only (row 2), Dice scores are increased by 2.8%, 3.6% and 5.0% on three testing datasets, respectively. Applying the anatomical prediction loss term only (row 3) improves the Dice scores on both datasets by 2.2%, 3.8% and 5.5%, respectively. The performance can be further improved by applying both the reconstruction loss term and the anatomical prediction loss term simultaneously (row 4), which demonstrates the effectiveness of the combination of these two proxy tasks for self-supervised learning. We also explore the two hyperparameters of w_{AP} and w_{Rec} in Eq. (5), and the experimental results are shown in rows 4–6 in Table 4. It can be observed that taking $w_{AP} = 1$, $w_{Rec} = 1$ and $w_{AP} = 1$, $w_{Rec} = 1.5$ reaches the optimal and suboptimal performances on three testing datasets, respectively. Besides, to evaluate the unbiasedness of the pretrained models with different loss weights, we visualize the multi-scale features from each layer of the encoder E_0 in Fig. 9, providing insights into the black box processing of deep learning models.

(3) The effects of different reasoning mechanisms

We also compare our CGRM with other reasoning mechanisms and the experimental results are illustrated in Table 5. Specifically, we conduct the comparison between our CGRM and (1) No graph reasoning (GR): training a generic UNet without graph reasoning module in the lesion stage with the inputs of original CT image and previous segmentation; (2) XOR-based: the non-causal feature map \mathcal{N} is expanded from normal pancreas region to non-tumor area in the abdominal CT image; (3) Contour-guided (Wang et al., 2022): the non-causal feature map

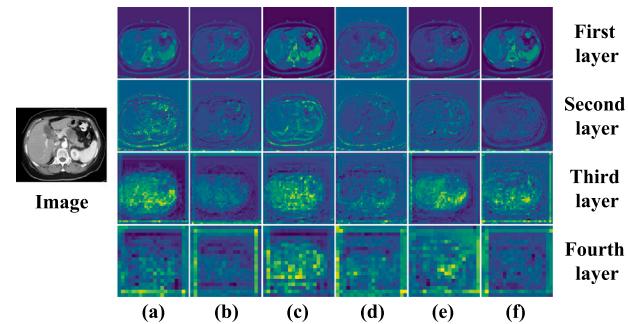


Fig. 9. Feature visualization of the pretrained AMAE with different training loss weights. (a) $w_{AP} = 0, w_{Rec} = 0$; (b) $w_{AP} = 0, w_{Rec} = 1$; (c) $w_{AP} = 1, w_{Rec} = 0$; (d) $w_{AP} = 1, w_{Rec} = 0.5$; (e) $w_{AP} = 1, w_{Rec} = 1$; (f) $w_{AP} = 1, w_{Rec} = 1.5$.

\mathcal{N} is changed from normal pancreas region to the edge of the tumor. It can be observed that simply concatenating a generic UNet without using any graph reasoning (row 2) only has slight improvement for tumor segmentation. The utilization of graph reasoning (rows 3–5) can significantly outperforms the baseline model (row 1) especially in Dice and HD metrics. Besides, our CGRM achieves best Dice scores on all three testing datasets comparing with XOR-based and contour-guided reasoning mechanisms, and has statistically significant differences from XOR-based method and contour-guided method on two testing datasets. Moreover, the t-distributed stochastic neighbor embedding (T-SNE) visualization is used to project the learned non-causal vertex features \hat{G}_n and causal vertex features \hat{G}_c onto two dimensions. Fig. 10 shows the feature distributions of different graph reasoning mechanisms on the internal testing dataset. We can observe that the graph reasoning mechanisms have satisfactory performance in distinguishing between the causal and non-causal features, and our CGRM performs more outstandingly among them.

(4) The effects of components in DDCM

Table 5

Performance comparison of different reasoning mechanisms for pancreatic cancer segmentation on three independent testing datasets. Dice coefficient similarity (mean \pm std, %) and 10% detection rates (DR, %), 95% Hausdorff Distance (HD, mm) are presented. The best and second best performance are in red and blue, respectively. GR stands for graph reasoning.

Methods	Internal Testing Set (n = 121)			External Testing Set I (n = 87)			External Testing Set II (n = 69)		
	Dice	DR	HD	Dice	DR	HD	Dice	DR	HD
Organ stage	69.7 \pm 24.3**	93.4	29.3 \pm 45.9**	65.2 \pm 24.0*	93.1	44.6 \pm 53.2**	62.9 \pm 25.4**	94.2	36.3 \pm 45.0**
No GR	70.4 \pm 24.1**	93.4	28.2 \pm 44.4**	64.3 \pm 25.2**	93.1	71.1 \pm 84.3**	60.3 \pm 26.2**	94.2	49.4 \pm 64.5**
XOR-based	72.4 \pm 24.6*	93.4	21.9 \pm 34.1	68.6 \pm 25.3	94.3	36.2 \pm 47.4	65.8 \pm 24.9*	92.8	27.6 \pm 33.8
Contour-guided	73.1 \pm 23.9	93.4	21.2 \pm 31.7	67.1 \pm 26.5*	93.1	26.9 \pm 32.8	65.2 \pm 26.5*	91.3	23.8 \pm 31.7
Causality-driven (Ours)	73.6 \pm 23.1	94.2	20.5 \pm 33.5	68.9 \pm 26.0	92.0	31.6 \pm 36.2	67.6 \pm 24.6	94.2	28.4 \pm 34.2

** and * represent statistically significant $p < 0.01$ and $p < 0.05$ difference comparing with our method by student t-test, respectively.

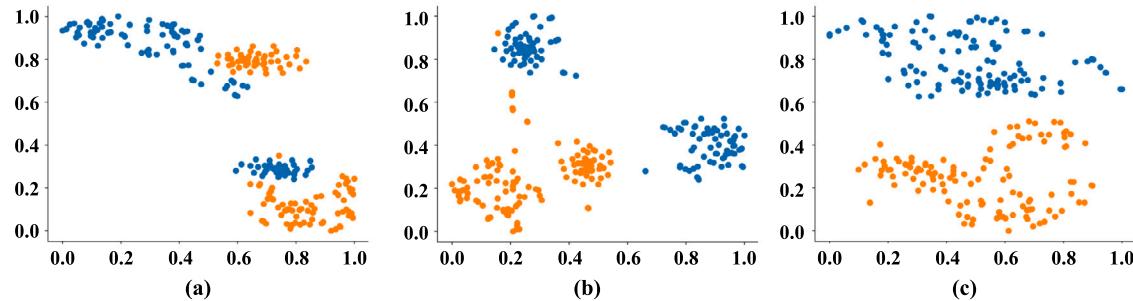


Fig. 10. T-SNE visualization of the vertex features obtained by different graph reasoning mechanisms on the internal testing datasets. (a) XOR-based; (b) Contour-guided; (c) Causality-driven (Ours). Blue and orange dots represent the causal vertex features and the non-causal vertex features, respectively.

Table 6

Ablation study of the components in DDCM for pancreatic cancer segmentation on three independent testing datasets. Dice coefficient similarity (mean \pm std, %) and 10% detection rates (DR, %), 95% Hausdorff Distance (HD, mm) are presented. The best and second best performance are in red and blue, respectively.

Methods	Internal Testing Set (n = 121)			External Testing Set I (n = 87)			External Testing Set II (n = 69)		
	Dice	DR	HD	Dice	DR	HD	Dice	DR	HD
Lesion stage	73.6 \pm 23.1**	94.2	20.5 \pm 33.5**	68.9 \pm 26.0**	92.0	31.6 \pm 36.2**	67.6 \pm 24.6**	94.2	28.4 \pm 34.2**
DDCM w/o DDPM, SDF	74.1 \pm 23.2**	94.2	15.1 \pm 22.1*	69.9 \pm 25.5**	92.0	17.8 \pm 22.6*	69.6 \pm 22.4**	94.2	11.8 \pm 12.8
DDCM w/o DDPM	74.1 \pm 23.1*	94.2	14.7 \pm 22.4	70.1 \pm 25.4**	92.0	17.2 \pm 22.2*	69.9 \pm 22.4*	94.2	12.1 \pm 13.0*
DDCM w/o SDF	74.4 \pm 22.9**	94.2	14.2 \pm 20.8	70.5 \pm 25.1*	93.1	16.9 \pm 22.3	70.0 \pm 22.2*	94.2	11.2 \pm 12.3
DDCM	75.0 \pm 22.7	94.2	13.7 \pm 20.7	71.3 \pm 24.0	94.3	15.9 \pm 21.2	70.9 \pm 21.4	95.7	10.8 \pm 11.9

** and * represent statistically significant $p < 0.01$ and $p < 0.05$ difference comparing with our DDCM by student t-test, respectively.

We also conduct the ablation experiment for the DDCM of the boundary stage, and the experimental results are shown in [Table 6](#). Without using DDPM (row 3), the performance slightly drops on all three testing datasets, leading to decrease 0.9%–1.2% in Dice with statistical significance, 7.3%–12.0% in HD. Without using SDF (row 4), the performance also slightly drops on all three testing datasets, leading to decrease 0.6%–1.2% in Dice with statistical significance, 3.6%–6.3% in HD. Additionally, it can be found that without using DDPM and SDF loss (row 2) leads to the worst performance in the boundary stage, indicating that both the DDPM and SDF are important to obtain performance improvements. [Fig. 11](#) illustrates the visual comparison of the ablation study of DDCM. Three presented cases are randomly taken from three testing datasets, respectively.

4.5. Application for tumor-vascular involvement analysis

In order to verify the benefits of our SLAD for clinical applications. We conduct additional comparison experiments for the assessment of tumor-vascular involvement and the surgical resectability prediction of pancreatic cancer. Specifically, we first calculate the degrees of tumor-vascular contact based on the tumor segmentation of our SLAD and other competing methods, along with the common peripancreatic vessel segmentation obtained by the pretrained model in work ([Zou et al., 2024](#)). The calculation is performed on a per 2D-axial slice basis. The length of contact borderline l between the tumor and vessel is calculated based on the adjacent pixels, which is followed by

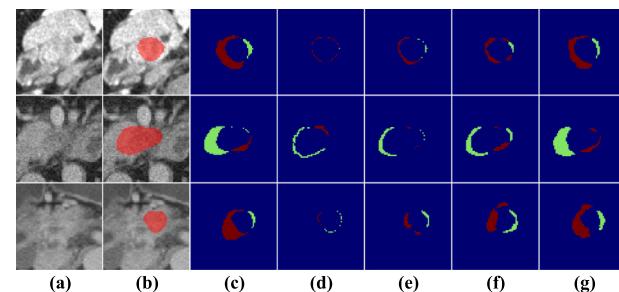


Fig. 11. Visual comparison of the ablation study of DDCM. (a) Original CT images; (b) Previous tumor segmentation (marked in red); (c) True discrepancy; (d) Predicted discrepancy without DDPM and SDF loss; (e) Predicted discrepancy without DDPM; (f) Predicted discrepancy without SDF loss; (g) Predicted discrepancy by Our DDCM. False positive and false negative discrepancies are shown in green and red in (c)–(g), respectively.

calculating the perimeter L of the vessel. The angle $\alpha = \frac{l}{L} \times 360^\circ$ is calculated on each contact 2D slice. The maximum angle is used to determine the degree of involvement. Based on the guideline of DPCG ([Versteijne et al., 2016](#)), the tumor is resectable if superior mesenteric vein (SMV) and portal vein (PV) have involvement degrees less than 90°, borderline resectable if the degrees are between 90° and 270° and locally advanced if the degrees are larger than 270°. For the arterial vasculature it is borderline resectable for degrees less than 90°

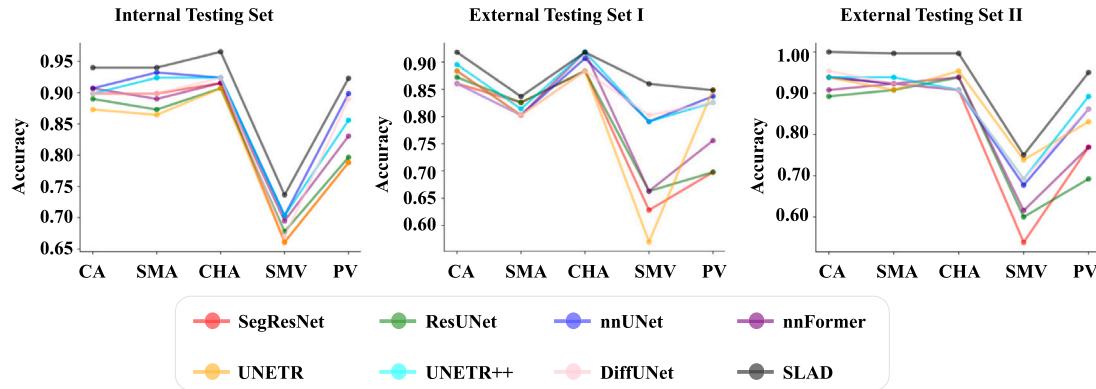


Fig. 12. Performance comparison in assessment of degrees of vascular involvement for five main peri-pancreatic vessels. CA: celiac artery; SMA: superior mesenteric artery, CHA: common hepatic artery; SMV: superior mesenteric vein; PV: portal vein.

Table 7

Performance comparison of evaluation the pancreatic tumor-vascular involvement on three independent testing datasets. The accuracy metrics for resectability prediction of resectable, borderline resectable or locally advanced pancreatic cancer are presented. The best performance are shown in bold.

Methods	Internal	External I	External II
SegResNet	83.9	69.8	84.6
ResUNet	80.5	73.3	86.2
nnUNet	84.7	70.9	84.6
nnFormer	85.6	75.6	84.6
UNETR	81.4	69.8	83.1
UNETR++	87.3	74.4	84.6
DiffUNet	87.3	72.1	83.1
SLAD (Ours)	90.7	79.1	92.3

and any amount of involvement more than 90° deems the tumor locally advanced. The accuracy scores are illustrated in Table 7 where we can observe that our SLAD achieves the best performance for resectability prediction on all three testing datasets, demonstrating the superiority of the potential clinical benefits. Besides, we also present the comparison results between our SLAD and other competing methods for calculating the degrees of the tumor-vascular involvement which are shown in Fig. 12.

5. Discussion

5.1. Limitations

(1) Model complexity and computational cost. Although our method significantly improves the performance for pancreatic cancer segmentation, the multi-stage framework inevitably increases the model complexity and computational cost. More detailed analysis of the network parameters and inference time for the proposed method and competing techniques is provided in Section C of the Supplementary Materials.

(2) Optimization problem for multi-stage training. Due to the multi-stage framework, 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.

(3) Lack of further interpretability. Our proposed method aims to simulate the essential logical thinking and experience of doctors in the progressive diagnostic stages to make the entire learning process more interpretable compared to the previous pure black box models. However, it should be pointed that each module of our method still lacks further interpretability because of their respective implementation by deep learning techniques.

5.2. Future works

(1) Training in an end-to-end manner. In the future, we will explore to design our SLAD into a completely end-to-end network structure to find the global optimal parameters and decrease the computational cost.

(2) Including healthy subjects. Our method aims to address the problem of accurate and reliable segmentation for pancreas and pancreatic cancer. Therefore, the healthy subjects are not included in this work. A multi-task framework including healthy subjects for both pancreatic cancer diagnosis and segmentation will be considered in our future studies.

(3) Embedding vascular involvement analysis into the framework. The tumor-vascular involvement analysis in this work is only conducted as an additional application to verify the superiority of the proposed method in clinical practice. However, a comprehensive framework embedding tumor-vascular involvement analysis can be explored in our future works.

6. Conclusion

In this paper, we propose a novel Segment-Like-A-Doctor (SLAD) framework to learn the reliable clinical thinking and experience for pancreas and pancreatic cancer segmentation on CT images. Specifically, our proposed method aims to simulate the essential logical thinking and experience of doctors in the progressive diagnostic stages of pancreatic cancer: organ, lesion and boundary stage. Firstly, in the organ stage, an Anatomy-aware Masked AutoEncoder (AMAE) is introduced to model the doctors' overall cognition for the anatomical distribution of abdominal organs on CT images by self-supervised pretraining. Secondly, in the lesion stage, a Causality-driven Graph Reasoning Module (CGRM) is designed to learn the global judgment of doctors for lesion detection by exploring topological feature difference between the causal lesion and the non-causal organ. Finally, in the boundary stage, a Diffusion-based Discrepancy Calibration Module (DDCM) is developed to fit the refined understanding of doctors for uncertain boundary of pancreatic cancer by inferring the ambiguous segmentation discrepancy based on the trustworthy lesion core. Experimental results on three independent datasets demonstrate that our approach boosts pancreatic cancer segmentation accuracy by 4%–9% compared with the state-of-the-art methods. Additionally, the tumor-vascular involvement analysis is also conducted to verify the superiority of the proposed method in clinical applications.

CRediT authorship contribution statement

Liwen Zou: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Data curation, Conceptualization. **Yingying Cao:** Validation, Resources, Data curation. **Ziwei Nie:** Writing – review & editing, Supervision, Funding acquisition. **Liang Mao:** Validation, Supervision, Resources, Data curation. **Yudong Qiu:** Validation, Supervision, Resources, Data curation. **Zhongqiu Wang:** Validation, Supervision, Resources, Data curation. **Zhenghua Cai:** Validation, Supervision, Resources, Investigation, Data curation. **Xiaoping Yang:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Formal analysis, Conceptualization.

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). We sincerely thank JiangSu Province Hospital, the First Affiliated Hospital of Wannan Medical College and the First Hospital of ChangZhou for providing the external validation data.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.media.2025.103539>.

Data availability

The authors do not have permission to share data.

References

- Bai, Y., Li, W., An, J., Xia, L., Chen, H., Zhao, G., Gao, Z., 2024. Masked autoencoders with handcrafted feature predictions: Transformer for weakly supervised esophageal cancer classification. *Comput. Methods Programs Biomed.* 244, 107936.
- Cao, L., Li, J., Chen, S., 2023. Multi-target segmentation of pancreas and pancreatic tumor based on fusion of attention mechanism. *Biomed. Signal Process. Control.* 79, 104170.
- Chen, X., Chen, Z., Li, J., Zhang, Y.-D., Lin, X., Qian, X., 2021. Model-driven deep learning method for pancreatic cancer segmentation based on spiral-transformation. *IEEE Trans. Med. Imaging* 41 (1), 75–87.
- Chen, T., Kornblith, S., Norouzi, M., Hinton, G., 2020. A simple framework for contrastive learning of visual representations. In: International Conference on Machine Learning, PMLR, pp. 1597–1607.
- Chen, Y., Rohrbach, M., Yan, Z., Shuicheng, Y., Feng, J., Kalantidis, Y., 2019. Graph-based global reasoning networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 433–442.
- Chhoda, A., Lu, L., Clerkin, B.M., Risch, H., Farrell, J.J., 2019. Current approaches to pancreatic cancer screening. *Am. J. Pathol.* 189 (1), 22–35.
- Diakogiannis, F.I., Waldner, F., Caccetta, P., Wu, C., 2020. ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. *ISPRS J. Photogramm. Remote Sens.* 162, 94–114.
- Haghghi, F., Taher, M.R.H., Zhou, Z., Gotway, M.B., Liang, J., 2021. Transferable visual words: Exploiting the semantics of anatomical patterns for self-supervised learning. *IEEE Trans. Med. Imaging* 40 (10), 2857–2868.
- Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H.R., Xu, D., 2021. Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. In: International MICCAI Brainlesion Workshop, Springer, pp. 272–284.
- Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H.R., Xu, D., 2022. Unetr: Transformers for 3d medical image segmentation. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 574–584.
- He, K., Chen, X., Xie, S., Li, Y., Dollár, P., Girshick, R., 2022. Masked autoencoders are scalable vision learners. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 16000–16009.
- Ho, J., Jain, A., Abbeel, P., 2020. Denoising diffusion probabilistic models. *Adv. Neural Inf. Process. Syst.* 33, 6840–6851.
- Huang, W., Hu, J., Xiao, J., Wei, Y., Bi, X., Xiao, B., 2024. Prototype-guided graph reasoning network for few-shot medical image segmentation. *IEEE Trans. Med. Imaging*.
- Iseñee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H., 2021. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods* 18 (2), 203–211.
- Kamisawa, T., Wood, L.D., Ito, T., Takaori, K., 2016. Pancreatic cancer. *Lancet* 388 (10039), 73–85.
- Kipf, T.N., Welling, M., 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.
- Li, X., Guo, R., Lu, J., Chen, T., Qian, X., 2023a. Causality-driven graph neural network for early diagnosis of pancreatic cancer in non-contrast computerized tomography. *IEEE Trans. Med. Imaging* 42 (6), 1656–1667.
- Li, Q., Liu, X., He, Y., Li, D., Xue, J., 2023b. Temperature guided network for 3D joint segmentation of the pancreas and tumors. *Neural Netw.* 157, 387–403.
- Li, Y., Luan, T., Wu, Y., Pan, S., Chen, Y., Yang, X., 2024. Anatomask: Enhancing medical image segmentation with reconstruction-guided self-masking. In: European Conference on Computer Vision, Springer, pp. 146–163.
- Li, Y., Sun, P., Wergeles, N., Shang, Y., 2021. A survey and performance evaluation of deep learning methods for small object detection. *Expert Syst. Appl.* 172, 114602.
- Ma, J., He, J., Yang, X., 2020. Learning geodesic active contours for embedding object global information in segmentation CNNs. *IEEE Trans. Med. Imaging* 40 (1), 93–104.
- Ma, J., Zhang, Y., Gu, S., An, X., Wang, Z., Ge, C., Wang, C., Zhang, F., Wang, Y., Xu, Y., et al., 2022. Fast and low-GPU-memory abdomen CT organ segmentation: the flare challenge. *Med. Image Anal.* 82, 102617.
- Mahmoudi, T., Kouzakhanian, Z.M., Radmard, A.R., Kafieh, R., Salehnia, A., Davarpanah, A.H., Aralabibeik, H., Ahmadian, A., 2022. Segmentation of pancreatic ductal adenocarcinoma (PDAC) and surrounding vessels in CT images using deep convolutional neural networks and texture descriptors. *Sci. Rep.* 12 (1), 3092.
- Meng, Y., Zhang, H., Zhao, Y., Yang, X., Qiao, Y., MacCormick, I.J., Huang, X., Zheng, Y., 2021. Graph-based region and boundary aggregation for biomedical image segmentation. *IEEE Trans. Med. Imaging* 41 (3), 690–701.
- Myronenko, A., 2019. 3D MRI brain tumor segmentation using autoencoder regularization. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part II 4, Springer, pp. 311–320.
- Qu, T., Li, X., Wang, X., Deng, W., Mao, L., He, M., Li, X., Wang, Y., Liu, Z., Zhang, L., et al., 2023. Transformer guided progressive fusion network for 3D pancreas and pancreatic mass segmentation. *Med. Image Anal.* 86, 102801.
- Qu, J., Xiao, X., Wei, X., Qian, X., 2024. A causality-inspired generalized model for automated pancreatic cancer diagnosis. *Med. Image Anal.* 94, 103154.
- Rahman, A., Valanarasu, J.M.J., Hacihamoglu, I., Patel, V.M., 2023. Ambiguous medical image segmentation using diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11536–11546.
- Rawla, P., Sunkara, T., Gaduputi, V., 2019. Epidemiology of pancreatic cancer: global trends, etiology and risk factors. *World J. Oncol.* 10 (1), 10.
- Ren, T., Sharma, A., Rivera, J.E.H., Rebala, L.H., Honey, E., Chopra, A., Kurt, M., 2024. Re-DiffiNet: Modeling discrepancy in tumor segmentation using diffusion models. In: Medical Imaging with Deep Learning.
- Ronneberger, O., Fischer, P., Brox, T., 2015. 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, pp. 234–241.
- Roth, H.R., Farag, A., Turkbey, E., Lu, L., Liu, J., Summers, R.M., 2016. Data from pancreas-ct: the cancer imaging archive. *IEEE Trans. Image Process.* 5.
- Shaker, A.M., Maaz, M., Rasheed, H., Khan, S., Yang, M.-H., Khan, F.S., 2024. UNETR++: delving into efficient and accurate 3D medical image segmentation. *IEEE Trans. Med. Imaging*.
- Taher, M.R.H., Gotway, M.B., Liang, J., 2024. Representing part-whole hierarchies in foundation models by learning localizability composable and decomposableability from anatomy via self supervision. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11269–11281.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2017. Attention is all you need. *Adv. Neural Inf. Process. Syst.* 30.
- Versteijne, E., van Eijck, C.H., Punt, C.J., Suker, M., Zwinderman, A.H., Dohmen, M.A., Grootenhuis, K.B., Busch, O.R., Besselink, M.G., de Hingh, I.H., et al., 2016. Preoperative radiochemotherapy versus immediate surgery for resectable and borderline resectable pancreatic cancer (PREOPANC trial): study protocol for a multicentre randomized controlled trial. *Trials* 17, 1–8.
- Wang, X., Gupta, A., 2018. Videos as space-time region graphs. In: Proceedings of the European Conference on Computer Vision, ECCV, pp. 399–417.
- Wang, Y., Tang, P., Zhou, Y., Shen, W., Fishman, E.K., Yuille, A.L., 2021. Learning inductive attention guidance for partially supervised pancreatic ductal adenocarcinoma prediction. *IEEE Trans. Med. Imaging* 40 (10), 2723–2735.

- Wang, K., Zhang, X., Lu, Y., Zhang, X., Zhang, W., 2022. CGRNet: Contour-guided graph reasoning network for ambiguous biomedical image segmentation. *Biomed. Signal Process. Control.* 75, 103621.
- Wang, B., Zou, L., Chen, J., Cao, Y., Cai, Z., Qiu, Y., Mao, L., Wang, Z., Chen, J., Gui, L., et al., 2023. A weakly supervised segmentation network embedding cross-scale attention guidance and noise-sensitive constraint for detecting tertiary lymphoid structures of pancreatic tumors. *IEEE J. Biomed. Heal. Informatics.*
- Wu, J., Fu, R., Fang, H., Zhang, Y., Yang, Y., Xiong, H., Liu, H., Xu, Y., 2024a. Medsegdiff: Medical image segmentation with diffusion probabilistic model. In: *Medical Imaging with Deep Learning*. PMLR, pp. 1623–1639.
- Wu, J., Ji, W., Fu, H., Xu, M., Jin, Y., Xu, Y., 2024b. Medsegdiff-v2: Diffusion-based medical image segmentation with transformer. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38, No. 6. pp. 6030–6038.
- Wu, L., Zhuang, J., Chen, H., 2024c. Voco: A simple-yet-effective volume contrastive learning framework for 3d medical image analysis. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 22873–22882.
- Xing, Z., Wan, L., Fu, H., Yang, G., Zhu, L., 2023. Diff-unet: A diffusion embedded network for volumetric segmentation. *arXiv preprint arXiv:2303.10326*.
- Xing, Z., Zhu, L., Yu, L., Xing, Z., Wan, L., 2024. Hybrid masked image modeling for 3D medical image segmentation. *IEEE J. Biomed. Heal. Informatics.*
- Yao, L., Zhang, Z., Keles, E., Yazici, C., Tirkes, T., Bagci, U., 2023. A review of deep learning and radiomics approaches for pancreatic cancer diagnosis from medical imaging. *Curr. Opin. Gastroenterol.* 39 (5), 436–447.
- Zhao, T., Cao, K., Yao, J., Nogues, I., Lu, L., Huang, L., Xiao, J., Yin, Z., Zhang, L., 2021. 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*. pp. 13743–13752.
- Zhao, G., Chen, X., Zhu, M., Liu, Y., Wang, Y., 2024. Exploring the application and future outlook of artificial intelligence in pancreatic cancer. *Front. Oncol.* 14, 1345810.
- Zhao, J., Dang, M., Chen, Z., Wan, L., 2022. DSU-net: Distraction-sensitive U-net for 3D lung tumor segmentation. *Eng. Appl. Artif. Intell.* 109, 104649.
- Zhou, H.-Y., Guo, J., Zhang, Y., Han, X., Yu, L., Wang, L., Yu, Y., 2023a. nnformer: Volumetric medical image segmentation via a 3d transformer. *IEEE Trans. Image Process.*
- Zhou, H.-Y., Lu, C., Chen, C., Yang, S., Yu, Y., 2023b. A unified visual information preservation framework for self-supervised pre-training in medical image analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* 45 (7), 8020–8035.
- Zhou, Y., Xie, L., Fishman, E.K., Yuille, A.L., 2017. Deep supervision for pancreatic cyst segmentation in abdominal CT scans. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 222–230.
- Zou, L., Cai, Z., Mao, L., Nie, Z., Qiu, Y., Yang, X., 2024. Automated peripancreatic vessel segmentation and labeling based on iterative trunk growth and weakly supervised mechanism. *Artif. Intell. Med.* 150, 102825.
- Zou, L., Cai, Z., Qiu, Y., Gui, L., Mao, L., Yang, X., 2023. CTG-Net: An efficient cascaded framework driven by terminal guidance mechanism for dilated pancreatic duct segmentation. *Phys. Med. Biol.*.