

Deep causal learning for pancreatic cancer segmentation in CT sequences

Chengkang Li ^{a,b}, Yishen Mao ^c, Shuyu Liang ^{a,b}, Ji Li ^{c,*}, Yuanyuan Wang ^{a,b,*}, Yi Guo ^{a,b,*}

^a School of Information Science and Technology of Fudan University, Shanghai 200433, China

^b Key Laboratory of Medical Imaging Computing and Computer Assisted Intervention (MICCAI) of Shanghai, Shanghai 200032, China

^c Department of Pancreatic Surgery, Pancreatic Disease Institute, Huashan Hospital, Shanghai Medical College, Fudan University, Shanghai 200040, China

ARTICLE INFO

Keywords:

Deep causal learning
Pancreatic cancer co-segmentation
Counterfactual inference
Pancreatic cancer risk assessment

ABSTRACT

Segmenting the irregular pancreas and inconspicuous tumor simultaneously is an essential but challenging step in diagnosing pancreatic cancer. Current deep-learning (DL) methods usually segment the pancreas or tumor independently using mixed image features, which are disrupted by surrounding complex and low-contrast background tissues. Here, we proposed a deep causal learning framework named CausegNet for pancreas and tumor co-segmentation in 3D CT sequences. Specifically, a causality-aware module and a counterfactual loss are employed to enhance the DL network's comprehension of the anatomical causal relationship between the foreground elements (pancreas and tumor) and the background. By integrating causality into CausegNet, the network focuses solely on extracting intrinsic foreground causal features while effectively learning the potential causality between the pancreas and the tumor. Then based on the extracted causal features, CausegNet applies a counterfactual inference to significantly reduce the background interference and sequentially search for pancreas and tumor from the foreground. Consequently, our approach can handle deformable pancreas and obscure tumors, resulting in superior co-segmentation performance in both public and real clinical datasets, achieving the highest pancreas/tumor Dice coefficients of 86.67%/84.28%. The visualized features and anti-noise experiments further demonstrate the causal interpretability and stability of our method. Furthermore, our approach improves the accuracy and sensitivity of downstream pancreatic cancer risk assessment task by 12.50% and 50.00%, respectively, compared to experienced clinicians, indicating promising clinical applications.

1. Introduction

Pancreatic cancer is one of the malignant diseases with the highest cancer mortality rate globally. According to previous reports, only 7.2% of newly diagnosed pancreatic cancer patients survive for more than 5 years (Gunderson et al., 2016). One of the significant challenges is the difficulty in detecting early-stage pancreatic tumors, which are typically cancerous or accompanied by metastatic disease upon diagnosis (Kleeff et al., 2016). 3D computed tomography (CT) plays a crucial role in identifying and diagnosing pancreatic cancer by capturing essential imaging characteristics of both the pancreas and tumors, such as shape, intensity, cyst fluid, and pancreatic duct dilation (Wolfgang et al., 2013). Simultaneously segmenting the pancreas and tumor in 3D CT sequences can help identify early pancreatic tumors, which is vital for clinical preoperative diagnosis and surgical navigation. However, manual segmentation is a burdensome task due to the large volume of 3D CT scans and the complex anatomical structure of the pancreas. It typically takes around an hour to segment each case manually. Thus, there is an urgent

need for a fast and accurate automated segmentation method.

Deep learning (DL) has shown significant advancements in automatic organ segmentation in medical images (Khan et al., 2023; Kuang et al., 2023; Litjens et al., 2017; Li et al., 2022; Pang et al., 2021; Shen et al., 2017; Valanarasu et al., 2021; Trinh, 2023; Dumitru et al., 2023). Specifically, in pancreas segmentation tasks, there are three main types of current DL methods: (1) Manual cropping-based methods (Chen et al., 2021; Guo et al., 2018; Zhou et al., 2019) cropped foreground region (pancreas and tumor) from 3D CT images before applying a DL segmentation model. (2) Single-stage methods (Cai et al., 2016; Chen et al., 2020; Isensee et al., 2021; Mo et al., 2020; Jiang et al., 2023; Tang et al., 2022, and Li et al., 2023) utilized a single DL model to directly segment entire 3D CT images without manual interventions. (3) Multi-stage coarse-to-fine approaches (Dai et al., 2023; Dogan et al., 2021; Farag et al., 2016; Man et al., 2019; Roth et al., 2015; Roth et al., 2018; Wang et al., 2020; Zhao et al., 2019; Zhu et al., 2019) used a DL network to roughly locate organs, followed by another fine segmentation network. Generally, these methods have made various attempts, but there are still

* Corresponding authors.

E-mail addresses: liji@huashan.org.cn (J. Li), yywang@fudan.edu.cn (Y. Wang), guoyi@fudan.edu.cn (Y. Guo).

shortcomings from three perspectives. First, most of the current works only perform single-target segmentation of the pancreas (Dai et al., 2023; Dogan et al., 2021; Farag et al., 2016; Man et al., 2019; Wang et al., 2020; Roth et al., 2015; Roth et al., 2018; Zhao et al., 2019; Zhu et al., 2019) or tumors (Chen et al., 2021; Guo et al., 2018) rather than their co-segmentation which plays a significant role in pancreatic cancer diagnosis. Second, due to the insufficient attention and limited encoding ability for the complex pancreas and tumors, these models struggle to handle difficult cases that are commonly encountered in real-world clinical data (e.g., abnormal morphology of the pancreas, unobvious solid tumors, and oversized tumors). Third, as shown in Fig. 1(a), the model that introduces confounding features during the inference process is referred to as a biased model. In the context of segmentation tasks, the background tissues surrounding the pancreas can be considered as a form of confounding bias, impeding network's learning and subsequently reducing the segmentation accuracy. The aforementioned methods utilize all image features for inference, thereby being categorized as biased segmentation models. Overall, a fully automated precise pancreas and tumor co-segmentation method is highly required to address the above challenges and demonstrate its clinical diagnostic ability.

Recently, causal learning has garnered great attention (Luo et al., 2020; Prosperi et al., 2020; Richens et al., 2020). Researchers employ counterfactual methods to effectively search for the causal relationship between mixed input factors and output results. However, to the best of our knowledge, the use of causal learning in image segmentation tasks is unexplored. Here, a strong prior anatomical causality for the pancreas and tumor co-segmentation task is assumed: the CT image can be divided into two main parts, foreground and background. The foreground is an absolute causal factor that determines the segmentation result, while the background is a confounding factor and irrelevant bias. As depicted in Fig. 1(b), we consider integrating causality enables networks to effectively eliminate redundant background bias, while focusing on significant foreground causal features, rather than involving manual cropping or multi-stage approaches.

Therefore, in this work, a single-stage unbiased deep causal learning model, named CausegNet, is proposed for the fully-automated pancreas and tumor co-segmentation in 3D CT images. Specifically, based on the prior causality, a counterfactual unbiased feature is generated by a causality-aware module, which separates the pancreas and tumor features from the background organs, and only retains the foreground causal features. During the training process, this counterfactual feature is used as a template, then a counterfactual loss is applied to ensure that extracted features are as similar as possible to the template. Through loss optimization, CausegNet gradually eliminates interference from complex backgrounds, concentrating on the foreground to encode causal relationships between the pancreas and tumor. Sequentially, the network performs counterfactual inference only relying on the acquired instinct foreground causal features, so as to precisely predict the irregular edges of the pancreas and search for hidden tumors. To validate the segmentation performance and robustness of CausegNet, we evaluate it on multiple scenarios, including several public datasets and real clinic data. The feature extraction of the network is visualized to enhance the causal interpretability of the DL black box model. In addition, our method's clinical practicality is verified by comparing its accuracy of pancreatic cancer risk assessment with expert clinicians. The data and code of this study are available. Overall, our contributions are as follows:

1. As a pioneering work, we introduce causal learning into DL image segmentation, proposing a deep causal learning model named CausegNet for pancreas and tumor co-segmentation in 3D CT images.
2. CausegNet integrates prior anatomical causality through a causality-aware module and a counterfactual loss. This enables the network to extract instinct causal features of the irregular pancreas and inconspicuous tumor while eliminating background organ intervention through counterfactual inference.
3. Understanding causality allows our method to achieve superior co-segmentation performances in both public and real clinical datasets, resulting in the pancreas/tumor Dice of 86.67%/84.28%.

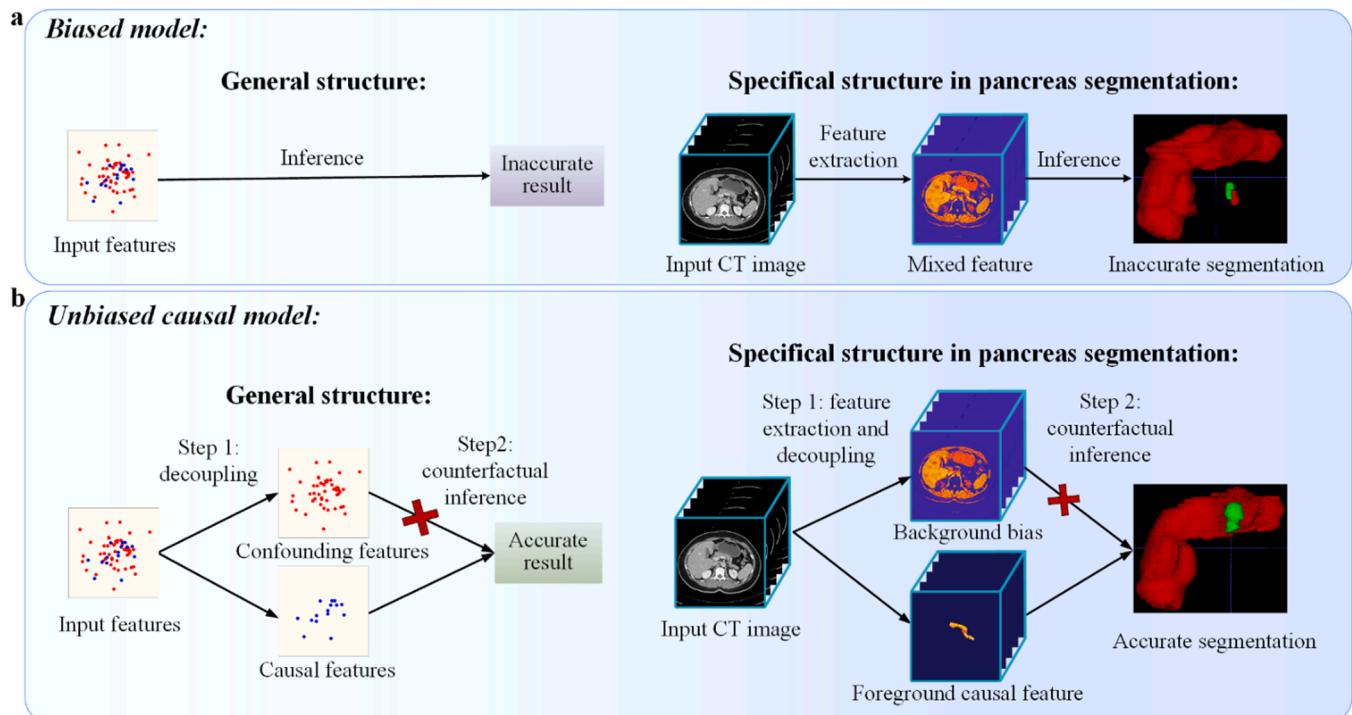


Fig. 1. The structure of the biased model and the unbiased causal model. **a** Biased model structures. The biased model performs inference based on the mixed features, resulting in an inaccurate segmentation. **b** Unbiased causal model structures. The unbiased causal model first decouples the background bias from the foreground causal feature, and then conducts counterfactual inference only based on the foreground causal feature, leading to an accurate segmentation.

4. A pancreatic cancer risk assessment system is proposed based on CausegNet, which improves diagnostic accuracy and sensitivity by 12.50% and 50.00%, compared to experienced clinicians. This presents promising clinical applications.

The remaining sections of this paper are structured as follows: In Chapter 2, the related work is presented. Chapter 3 offers comprehensive details of the proposed CausegNet. The experimental data, planning, and implementation details are outlined in Chapter 4. Detailed experimental results are provided in Chapter 5. Finally, Chapters 6 and 7 engage in comprehensive discussions and present a summary of the research.

2. Related works

2.1. DL-based pancreas segmentation

DL-based pancreatic-related segmentation research can be broadly classified into three types according to the image input and network structure: manual cropping-based segmentation, entire image-based single-stage segmentation, and coarse-to-fine multi-stage segmentation.

The manual cropping-based segmentation network aims to address the challenge of complex background interference in pancreas segmentation. For instance, Chen et al. (2021) have employed manual cropping techniques to select a spherical region of interest (ROI) containing the pancreatic tumor from 3D CT scans. By removing unnecessary background organ information, these methods enhanced the accuracy of pancreatic cancer segmentation. Another approach, used by Guo et al. (2018), involves selecting a square ROI and incorporating boundary information to improve the segmentation results of pancreatic tumors. In addition, based on the cropped ROI, Zhou et al. (2019) combined arterial and portal phases CT images to enhance the segmentation performance of pancreatic ductal adenocarcinoma. These manual cropping methods reduce the computational effort and background interference to a certain extent. However, it is important to note that their segmentation results may be subject to the initial cropping and require clinical anatomical knowledge, limiting their broader application.

The entire image-based single-stage network enables fully automatic segmentation without the need for human intervention. For instance, Isensee et al. (2021) employed a single-stage nnUNet to directly segment the pancreas and tumors from whole CT images. Similarly, Mo et al. (2020) and Chen et al. (2020) incorporated a multi-scale feature fusion strategy within a single-stage network to enhance the representation of irregular pancreatic features. Cai et al. (2016) proposed a convolutional/recurrent architecture network to ensure consistent contextual information for pancreatic segmentation. Tang et al. (2022) combined the advantages of both the Transformer and CNN architectures in a single network for pancreatic cancer segmentation. Jiang et al. (2023) used a network that utilizes contextual information of three 2D views to enhance the accuracy of segmenting small-volume 3D pancreas. Li et al. (2023) introduced a temperature-guided segmentation network to address the volume differences between the pancreas and tumors. Although these single-stage approaches eliminate the subjectivity associated with manual cropping, they may encounter challenges when dealing with morphologically altered pancreas and obscured tumors using a single DL network.

The coarse-to-fine approach enhances the ability to capture low-contrast pancreatic features by employing cascading multi-stage networks. Several studies, including Roth et al. (2018), Dogan et al. (2021), Zhao et al. (2019), and Man et al. (2019), have utilized the coarse-to-fine concept to progressively focus on the pancreas. Initially, a network is employed to roughly localize the pancreatic region, followed by a separate network for precise pancreas segmentation. In addition, Farag et al. (2016) and Roth et al. (2015) adopted a bottom-up strategy for progressively segmenting the pancreas using a cascade of super pixels.

Wang et al. (2020) proposed a coarse-to-fine federated segmentation architecture to address data privacy constraints in pancreas and tumor co-segmentation. These multi-stage approaches improve the network's learning capability but also introduce algorithmic complexity, potentially leading to a propagation of segmentation errors.

2.2. Deep causal learning

In recent years, an increasing number of researchers have been striving to integrate the concept of causal learning into DL models. (Huynh et al., 2016; Imbens et al., 2015; Cui et al., 2022; Luo et al., 2020). Generally, current data-driven DL approaches establish correlations between inputs and outputs, essentially learning a mapping function. As shown in Fig. 1, the result is jointly determined by various input features. Data-driven models use all factors for inference, making them biased models. They can be influenced by confounding features that are unrelated to the result, leading to false correlations and conclusions. Conversely, causal learning aims to identify the causal relationship between inputs and outputs, thus eliminating confounding factors and constructing an unbiased causal model based on causal factors. Unbiased causal models typically possess advantages such as stability, generalizability, and interpretability compared to biased models. One crucial approach to discovering causal relationships involves using counterfactuals. This method removes or replaces certain input factors to observe their influence on the result. For example, in Fig. 1, the removal of confounding features, as well as background bias, produces an accurate result, suggesting that they do not play a decisive causal role.

Specifically, in the field of image analysis, some researchers have been introducing causality into DL models. For instance, in image classification tasks, Liu et al. (2019) employed counterfactual methods to change the predicted image category, thereby enhancing the interpretability of DL classification models. Suzuki et al. (2021) generated counterfactual images to gain insights into the reasons for misclassifications made by DL models. Dash et al. (2022) leveraged causal relationships among image attributes to generate counterfactuals, reducing bias in pre-trained classification models. Similarly, in image recognition tasks, Yue et al. (2021) addressed Zero-Shot and Open-Set recognition problems by learning category differences through counterfactual changes in sample categories. Mao et al. (2021) improved recognition performance by learning robust visual representations using a deep causal model. In scene graph generation tasks, Chen et al. (2019) and Tang et al. (2020) both used counterfactual training methods to enhance the reliability of the generated results. In visual question-answering tasks, Abbasnejad et al. (2020), Niu et al. (2021), and Kolling et al. (2022) constructed causal models using counterfactual methods to eliminate prior language biases. Moreover, causal learning and counterfactual methods have demonstrated improvements in the accuracy of medical diagnosis (Prosperi et al., 2020; Wu et al., 2022; Richens et al., 2020). As an example, Wang et al. (2021) used a deep causal model to generate counterfactual mammogram features. The counterfactual features proposed a hypothetical scenario where a breast tumor patient has never experienced any breast disease, thus aiding in the diagnosis of benign and malignant breast tumors.

However, to the best of our knowledge, previous research has not explored the integration of causal learning into DL image segmentation. This can be attributed to two reasons, 1. previous works have not provided a clear definition of causal and confounding factors in the context of image segmentation tasks, 2. there has been a lack of effective supervision to disentangle these causal and confounding factors. As pioneers in this field, we have successfully developed a deep causal model named CausegNet that is specifically tailored for image segmentation tasks. In this work, we raise a counterfactual question: if the image contains only foreground regions, how does this affect the training and prediction of the network? To explore this question, we first introduced a prior causality that defines the causal and confounding factors for

image segmentation. Then, a counterfactual unbiased feature is generated to simulate the feature extraction scenario where the image solely comprises the foreground. The counterfactual unbiased feature is regarded as additional supervised information, and an advanced training method is designed to decouple the causal and confounding factors. Finally, the network conducts counterfactual inference based on foreground causal features to observe changes in the segmentation results compared to mixed feature inference. In summary, the proposed CausegNet inherently eliminates interference from complex backgrounds and focuses solely on learning the intricate anatomical properties of the pancreas and tumor. This enables our network to achieve accurate and fully automated co-segmentation of the pancreas and tumor using a single-stage network without requiring manual intervention or multi-stage cascades. Moreover, the experiments indicate that our method has the potential to be applied to various image segmentation tasks.

3. Method

CausegNet is a single-stage unbiased DL network for pancreas and tumor co-segmentation. An overview of CausegNet is illustrated in Fig. 2. Generally, CausegNet extracts features from a whole 3D CT image input. It then uses a causality-aware module to learn the anatomical prior causality between the pancreas/tumor and the background. By

relying on the highly correlated instinct features of the pancreas and tumor, our method employs counterfactual inference to complete their co-segmentation.

3.1. Training phase of CausegNet

In the training phase (Fig. 2(a)), CausegNet consists of three parts, including a biased feature extraction, a causality-aware module, and an inference module.

3.1.1. Biased feature extraction

As shown in Fig. 2(a), the biased feature extraction module extracts high-dimensional mixed features that contain both background biases and pancreatic region features from a whole 3D CT image. It is a multi-scale convolutional architecture containing four layers. For the brief presentation, only one scale of features is shown as an example in Fig. 2. Its detailed structure is presented in Fig. 3, where the shallow layer and deep layer extract detailed information and high-level semantic knowledge, respectively. The distribution of mixed features can be expressed as $P(F_M | I, \theta)$. Where I is the input image ($I \in \mathbb{R}^{B \times X \times Y \times Z}$, \mathbb{R} is a high-dimensional space, B is the batch size, and $X \times Y \times Z$ is the input size). θ represents the feature extractor parameterized by θ . F_M is the extracted mixed features ($F_M \in \mathbb{R}^{N \times B \times L \times W \times H}$, N is the number of

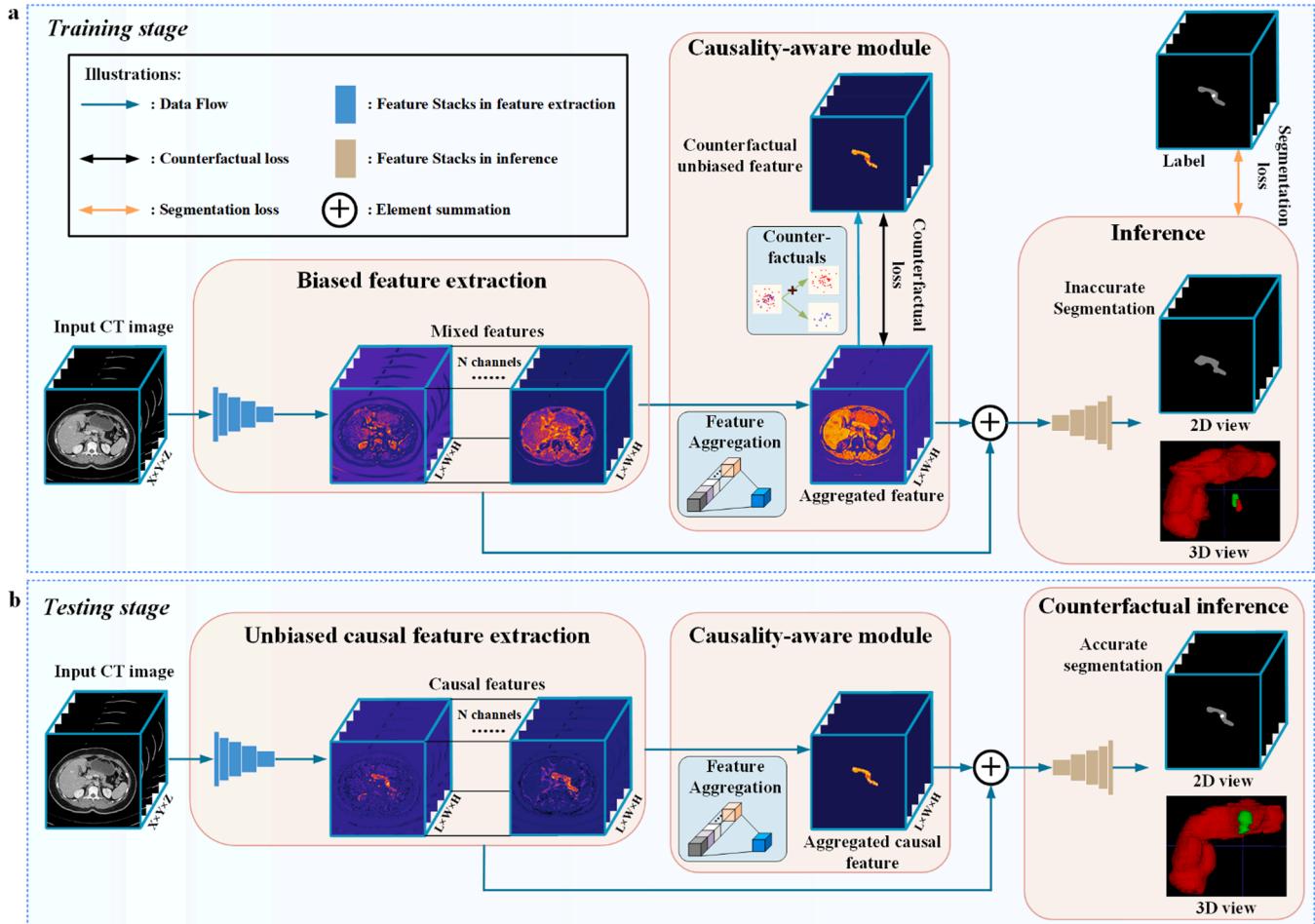


Fig. 2. Overview of CausegNet. a An overall training framework of CausegNet, which comprises three major steps: (1) biased feature extraction, which extracts high-dimensional mixed features, (2) a causality-aware module that embeds prior causality into the network, which includes a feature aggregation and a counterfactual method, and (3) inference, which outputs inaccurate segmentation. Two loss functions are applied to optimize the network, including a counterfactual loss and a segmentation loss. b In the testing framework, the three steps differ from the training process. Firstly, an unbiased causal feature extraction extracts causal features instead of mixed features; then the causality-aware module aggregates the high-dimensional features to a low-dimensional representation; finally, a counterfactual inference outputs accurate segmentation.

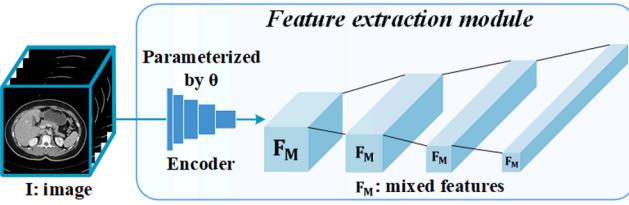


Fig. 3. Specific structure of feature extraction.

channels, and $L \times W \times H$ represents the size of the features. Usually, from the first layer to the last layer, $L \times W \times H$ gradually decreases from $X \times Y \times Z$ to $X/32 \times Y/32 \times Z/16$, $X \times Y \times Z \times Y$ to $X/32 \times Y/32 \times Z/16$, meanwhile, N gradually increases from 32 to 330).

3.1.2. Causality-aware module

The causality-aware module embeds the prior anatomical causality into the CausegNet. It consists of a feature aggregation module and a counterfactual method. The primary objective of the causality-aware module is to teach the network how to decouple foreground features from background biases, enabling the removal of a substantial amount of irrelevant background information while facilitating a thorough exploration of the causal relationship between the pancreas and tumors within the foreground. However, directly constraining the high-dimensional features extracted by the neural network presents challenges, involving considerable computational complexity and potential class imbalance. Therefore, we initially employ a feature aggregation module to learn a low-dimensional representation of the high-dimensional mixed features extracted by the neural network. As shown in Fig. 4, the feature aggregation module contains three $1 \times 1 \times 1$ convolution and relu functions, which gradually reduces the dimension of a mixed feature to obtain a low-dimensional aggregated feature. Each layer of mixed feature is followed with a feature aggregation module, resulting in four multi-scale single-dimensional aggregated features in Fig. 5. The distribution of the aggregated feature can be expressed as $P_{FA}(F_A | F_M)$, where F_M is the abovementioned mixed feature, FA and F_A are the feature aggregation module and the aggregated feature respectively ($F_A \in \mathbb{R}^{(1 \times B \times L \times W \times H)}$ has the same size with F_M but has only one single channel).

Then, a counterfactual method guided by prior causality is applied to the above multi-scale aggregated features to generate multi-scale ideal counterfactual unbiased features, as shown in Fig. 5. This counterfactual method is exclusively used during the training process and acts as a parameter-free function that does not require optimization. Its primary purpose is to completely remove the background organs of the aggregation feature while employing a linear mapping to accentuate the pancreas-related causal feature, enhancing its distinctiveness from the background. Specifically, the distribution of the counterfactual unbiased features can be expressed as $P_{CF}(F_C | F_A)$. Where F_A is the former aggregated features, $F_C \in \mathbb{R}^{N \times 1 \times L \times W \times H}$ is the counterfactual

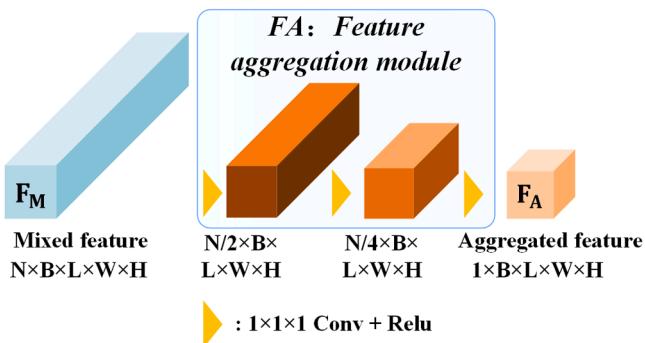


Fig. 4. Specific structure of feature aggregation module.

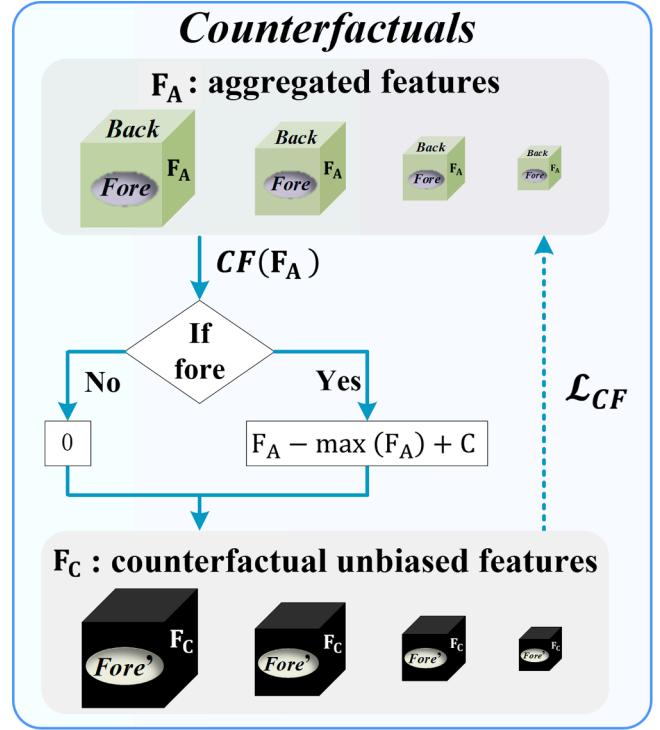


Fig. 5. Specific structure of counterfactual method. The Back and Fore are the image's background and foreground respectively. The counterfactual method is only performed during the training stage.

unbiased features, and CF is the counterfactual method that can be formulated as follows.

$$CF(F_A) = F_C = \begin{cases} 0, & F_A \in \text{Background} \\ F_A - \max(F_A) + C, & F_A \in \text{Foreground} \end{cases} \quad (1)$$

where C is a constant that effectively ensures a distinct value difference between causal features and background bias. $\max(F_A)$ means the maximum value of F_A in the foreground area. By subtracting $\max(F_A)$, we prevent the value of F_C from approaching infinity during the optimization process.

To enhance the network's understanding of anatomical causality, our objective is to minimize the distribution distance between the aggregated features and the counterfactual unbiased features. This approach not only allows the network to spontaneously separate foreground causal features from background biases but also enables it to learn the causal relationship between the pancreas and tumors within the foreground region. We incorporated multi-scale supervision during the training process, a critical step in achieving causal segmentation networks. This approach facilitates the efficient transfer of causal information across different network layers, empowering the network with feature purification and error correction capabilities. Specifically, for multi-scale training, we calculate the counterfactual loss between the distribution of multi-scale F_A and the distribution of multi-scale F_C , which can be expressed as:

$$\mathcal{L}_{CF} = \mathcal{D}(P_{FA}(F_A | F_M), P_{CF}(F_C | F_A)) \quad (2)$$

With the optimization of \mathcal{L}_{CF} , F_A will be closer to F_C , which means the aggregated features will become more and more unbiased. To establish gradient bi-connection and backpropagation path, the low-dimensional aggregated features and the high-dimensional mixed features are summed together at the element level and sent to the subsequent inference module. This bi-connection effectively prevents isolated optimization of the feature aggregation module while enhancing the foreground region within the mixed features. Through backpropagation,

the biased feature extractor will converge towards an unbiased causal feature extractor. As a result, the mixed features will also become more unbiased and focus on encoding deformable pancreas and inconspicuous tumors.

3.1.3. Inference

$$\frac{\partial E}{\partial w_i} = (A - A_{cuf}) \cdot \left(\frac{\partial A}{\partial w_i} - \frac{\partial A_{cuf}}{\partial A} \cdot \frac{\partial A}{\partial w_i} \right) = (A - A_{cuf}) \cdot \frac{\partial A}{\partial w_i} = (A - A_{cuf}) \cdot \sigma' \left(\sum_{i=1}^n w_i f_i + b \right) \cdot f_i \quad (9)$$

The inference module is responsible for generating the co-segmentation results of the pancreas and tumor, which is designed in a symmetrical structure similar to the encoder. This module only uses the causal features within the foreground while progressively restoring low-resolution features into high-resolution ones. During this process, the inference module initially focuses on refining the edges of the pancreas within the foreground region. Then according to acquired position causality, it infers the precise location and boundary of the tumor inside the pancreas. The distribution of the output can be expressed as $P(O | F)$, where $F = F_A + F_M \in \mathbb{R}^{N \times B \times L \times W \times H}$. $O \in \mathbb{R}^{B \times X \times Y \times Z}$. The segmentation loss is calculated between the segmentation output and the label.

$$\mathcal{L}_{seg} = \mathcal{D}(P(O | F), P(Label)) \quad (3)$$

With the iteration of \mathcal{L}_{seg} , inaccurate segmentation will become more accurate and closer to the ground truth.

3.1.4. CausegNet optimization

This section shows the optimization and loss function in CausegNet training. The total loss function of CausegNet is a linear weighted sum of a counterfactual loss \mathcal{L}_{CF} and a segmentation loss \mathcal{L}_{seg} .

$$\mathcal{L}_{Total} = \lambda_1 \mathcal{L}_{CF} + \lambda_2 \mathcal{L}_{seg} \quad (4)$$

Where the λ_1 and λ_2 are the weights. \mathcal{L}_{CF} aims to embed causality into CausegNet. It instructs the network to focus on the foreground region and extract causal features of the pancreas and tumor. The subsequent \mathcal{L}_{seg} ensures that the network accurately predicts the location and morphology of the pancreas and tumor.

Specifically, in Eq. (2), \mathcal{L}_{CF} minimizes the distribution distance between the aggregated feature F_A and the counterfactual unbiased feature F_C . Here, the distance between the two distributions is measured by mean squared error. Suppose A is a voxel point in F_A with the position of $(1, b, l, w, h)$, then

$$A = F_A(1, b, l, w, h) = \sigma \left(\sum_{i=1}^n w_i f_i + b \right) \quad (5)$$

Where f_i is the i -th dimensional feature of the mixed feature F_M , which can be expressed as

$$f_i = F_M(i, b, l, w, h), \quad i \in [1, N] \quad (6)$$

w_i and b are the weight and bias of convolution. $\sigma(\cdot)$ is the active function. According to the counterfactual method in Eq. (1), the counterfactual unbiased feature at voxel point A is A_{cuf} :

$$A_{cuf} = CF(A) = \begin{cases} 0, & A \in Background \\ A + C - \max(A), & A \in foreground \end{cases} \quad (7)$$

We minimize the distance between A and A_{cuf} to make them as similar as possible. Therefore, the objective function at this one voxel point can be written as

$$\min E = \min \mathcal{D}(A, A_{cuf}) = \min \frac{1}{2} (A - A_{cuf})^2 \quad (8)$$

The optimization of the network uses gradient descent, so we need to calculate $\frac{\partial E}{\partial w_i}$ and $\frac{\partial E}{\partial b}$.

Bringing Eq. (7) into Eq. (9) yields

$$\frac{\partial E}{\partial w_i} = \begin{cases} A \cdot \sigma' \left(\sum_{i=1}^n w_i f_i + b \right) \cdot f_i, & A \in background \\ (\max(A) - C) \cdot \sigma' \left(\sum_{i=1}^n w_i f_i + b \right) \cdot f_i, & A \in foreground \end{cases} \quad (10)$$

Then

$$\frac{\partial E}{\partial b} = \begin{cases} A \cdot \sigma' \left(\sum_{i=1}^n w_i f_i + b \right), & A \in background \\ (\max(A) - C) \cdot \sigma' \left(\sum_{i=1}^n w_i f_i + b \right), & A \in foreground \end{cases} \quad (11)$$

From the above analyses, the $\frac{\partial E}{\partial w_i}$ and $\frac{\partial E}{\partial b}$ can be applied for the back-propagation algorithm. The parameters are updated as follows:

$$w_i^+ = w_i - \eta * \frac{\partial E}{\partial w_i}; b^+ = b - \eta * \frac{\partial E}{\partial b} \quad (12)$$

Where the η represents the learning rate. For the overall objective function of \mathcal{L}_{CF} can be written as

$$\min E = \frac{1}{2 * B * L * W * H} \sum_{b=1}^B \sum_{l=1}^L \sum_{w=1}^W \sum_{h=1}^H (A - A_{cuf})^2 \quad (13)$$

As the model's parameters continue to update, the distribution of A gradually approaches A_{cuf} , leading to a more unbiased aggregated feature F_A . Because there is a gradient connection between F_A and F_M , meaning that as the backpropagation progresses, F_M also becomes more unbiased.

In Eq. (3), the segmentation loss \mathcal{L}_{seg} minimizes the distribution distance between the segmentation output O and the label. The distribution distance is measured by Dice and cross-entropy. The optimization of \mathcal{L}_{seg} can be found in Milletari et al. (2016), which is omitted here due to space constraints.

3.2. Testing phase of CausegNet

In Fig. 2(b), the testing phase of CausegNet is an end-to-end structure that also contains three parts. The whole 3D CT image is inputted to CausegNet to extract unbiased causal features of the pancreas and tumor. Then the feature aggregation module compresses the high-dimensional causal features into an aggregated causal feature. Finally, without the counterfactual loss, CausegNet directly performs counterfactual inference based on the causal features to eliminate the background bias while getting the accurate pancreas and tumor co-segmentation result.

Table 1

Characteristics of internal and external CT datasets from multi-center.

Dataset Name	Internal dataset (MSD pancreas)	External datasets		
		Huashan pancreas	MSD liver	MSD spleen
Patient number	281	269	131	41
Training/Testing	85% / 15%	85% / 15%	80% / 20%	75% / 25%
Label	Pancreas and tumor	Pancreas and tumor	Liver and tumor	Spleen
Modality	Portal venous phase 3D CT sequences	Portal venous phase 3D CT sequences	Portal venous phase 3D CT sequences	Portal venous phase 3D CT sequences
Tumor type		High-risk (162): IPMN, MCN, SPN	low-risk (107): SCN	
In-plane resolution (mm)	0.61–0.98	0.55–0.98	0.50–1.00	0.50–1.00
Slice thickness (mm)	1.50–7.50	0.45–0.75	0.45–6.00	2.50–5.00
Data source	Memorial Sloan Kettering Cancer Center, USA	Huashan hospital, Shanghai, China	Liver Tumor Segmentation (LiTS) challenge	Memorial Sloan Kettering Cancer Center, USA

Abbreviation: intraductal papillary mucinous neoplasm (IPMN), mucinous cystic neoplasms (MCN), solid pseudopapillary neoplasms (SPN), serous cystic neoplasm (SCN).

4. Experiments

4.1. Datasets and study design

In this work, we collected four datasets containing 722 3D CT sequences from multiple centers. The patient and image characteristics are displayed in Table 1. Among them, 281 pancreatic tumor cases published by the Medical Segmentation Decathlon challenge (MSD) (Antonelli et al., 2022) were used as an internal dataset (MSD pancreas) to develop our unbiased causal segmentation model. The remaining three datasets were used as external datasets to validate the method performance, including 269 real clinical pancreatic tumor cases from the Department of Pancreatic Surgery of Huashan Hospital of Fudan University (Huashan pancreas), 131 liver tumor cases and 41 healthy spleen cases from MSD (MSD liver dataset and MSD spleen dataset), respectively. The segmentation labels of 3D CT sequences were annotated by experienced clinicians. For detailed information on the dataset, please refer to the supplementary materials.

In our experiments, firstly, we randomly selected 85% of cases from the internal MSD pancreas dataset to train our CausegNet for pancreatic cancer co-segmentation. To determine the appropriate hyperparameters, cross-validation was applied during training. The remaining 15% of cases were used to independently test the performance of the model. Secondly, the external real clinical Huashan dataset was applied to assess the CausegNet performance from two perspectives: segmentation stability and clinical practicality. For segmentation stability, we extended CausegNet from the internal MSD pancreas dataset to the external Huashan dataset to examine its segmentation effectiveness. For clinical practicality, it is worth noting that the cases in the Huashan dataset can be classified into high-risk and low-risk types according to postoperative pathological examinations. Therefore, based on the co-segmentation results, we developed a preoperative pancreatic cancer risk assessment DL system to research the significance of CausegNet for clinical diagnoses. Thirdly, the external MSD liver and MSD spleen datasets were collected to further evaluate the availability of the proposed prior causality and CausegNet across various segmentation tasks (i.e., liver/liver tumor co-segmentation, and spleen segmentation). Overall, our data and experiments

4.2. Implementation details

For data preprocessing, the window width and level of CT images were selected to 350 and 40. The image values were normalized to [0, 1]. Data augmentation was performed using random rotation, B-sample deformation, and random scaling. Data preprocessing was performed using MATLAB 2021a and ITK-SNAP 3.8 (<http://itksnap.com>).

Training is performed on a 32 GB Tesla V100-SXM2 GPU. The images are cropped into several patches and fed into the network, with a patch size of $224 \times 224 \times 40$ along the X, Y, and Z axes. The predicted patches are aggregated back to the original resolution using Gaussian weighting. The training batch size was set to 2. The initial learning rate was set to 0.01. The Adam optimizer was used with a weight decay of 0.00003. In this work, to ensure a balance between \mathcal{L}_{CF} and \mathcal{L}_{Seg} , the weights λ_1 and λ_2 were both set to 1. The constant C in Formula 1 was set to 3. Cross-validation is used to determine the remaining hyperparameters during the training process. The network was implemented in PyTorch.

4.3. Evaluation metrics

For the segmentation tasks, two common matrices are used to evaluate segmentation performance: the Dice coefficient (Dice) (Milletari et al., 2016) and Normalized Surface Distance (NSD) (Antonelli et al., 2022; Nikolov et al., 2021). Higher Dice and NSD values indicate better area accuracy and boundary overlap, respectively. In addition, an auxiliary metric called foreground-normalized mean square error (FNMSE) is proposed to assess the unbiasedness of the segmentation model. Specifically, in the extracted features by a segmentation model, if the foreground and background values are significantly different meanwhile the background features are sufficiently suppressed, the FNMSE value is lower which means the model is more unbiased. The detailed calculations and explanations of evaluation metrics are shown in the supplementary materials.

For the pancreatic cancer risk assessment task, three common metrics are used to evaluate classification accuracy: area under the subject curve (AUC), accuracy (ACC), and sensitivity (SEN), respectively.

4.4. Comparison methods

Several representative pancreatic-related segmentation algorithms were replicated and compared, including LOGISMOS (Guo et al., 2018), MSCFS (Zhu et al., 2019), nnUNet (Isensee et al., 2021), SwinUNETR (Tang et al., 2022), and APAUNet (Jiang et al., 2023). LOGISMOS is a representative manual cropping-based pancreatic cancer segmentation method. It artificially clips the original large-volume CT images to the target bounding box and then uses a DL network to perform segmentation. MSCFS is an excellent coarse-to-fine multi-stage pancreas and tumor co-segmentation method, which first uses a network to roughly localize the pancreas region and then employs a separate network to finely segment the pancreas and tumor. The remaining three are the latest and most representative single-stage pancreatic cancer segmentation networks. Among them, nnUNet reported state-of-the-art (SOTA) performance in the pancreas and tumor co-segmentation task published

in the MSD Challenge, making it a benchmark for segmentation tasks. nnUNet is a whole-image-based single-stage approach, which adaptively adjusts network hyperparameters according to the dataset and task. Swin UNETR is a transformer-based self-supervised pre-trained single-stage segmentation network that utilizes a hierarchical encoder to extract multi-resolution features. The transformer structure therein can be regarded as a self-attention mechanism. APAUNet employs axis projection attention mechanisms to compute three 2D-view attentions for 3D features, enhancing segmentation accuracy for small targets. The training process of the above algorithms is consistent with CausegNet, in which some training parameters are adapted to achieve optimal performances.

5. Results

5.1. Pancreas and tumor co-segmentation performance

The segmentation results are shown in Table 2 and Fig. 6. From Table 2, it is evident that our pancreas segmentation achieved satisfactory results in both region and edge segmentation on the internal MSD pancreas dataset, with Dice and NSD scores of 81.15% and 96.88%, respectively. We also observed promising accuracy in tumor segmentation, with Dice and NSD scores of 71.09% and 90.97%, respectively. Our method's effectiveness in the real clinical data was validated on the internal Huashan pancreas dataset. Specifically, we obtained the Dice scores of 86.67% (pancreas) and 84.28% (tumor) and NSD scores of 96.97% (pancreas) and 95.12% (tumor). The result indicates that CausegNet can be extended to clinical datasets, which is of significant importance for the application of AI algorithms in real clinical settings.

In Table 2, our method outperformed all compared methods both on MSD and Huashan datasets. Unlike LOGISMOS, our CausegNet eliminates the subjectivity and inaccuracy caused by simple manual cropping. As a result, compared to LOGISMOS, we achieved a 2.29% improvement in pancreas Dice and a 22.55% improvement in tumor Dice on the MSD pancreas dataset. Compared to MSCFS, our method does not rely on multi-stage algorithms, thus avoiding the accumulation and propagation of segmentation errors from one stage to the next. This led to 4.73% and 19.14% improvements in pancreas and tumor Dice. As for single-stage methods such as nnUNet, Swin UNETR, and APAUNet, which use mixed image features for inference, our method performs counterfactual inference, improving the pancreas/tumor Dice by 1.56%/6.05%, 1.11%/8.17% and 1.83%/5.28%, respectively. Additionally, as indicated by statistical analysis in Fig. 6(d), our method demonstrated significant statistical differences in segmentation performance compared to other methods.

From Fig. 6(a), (b), and (c), based on the acquired causality, our method outperforms other comparative methods in handling challenging segmentation scenarios that are frequently encountered in clinical practice. First, Fig. 6(a) shows a scenario of an exceptionally large tumor in the pancreatic tail that highly distorts the pancreatic shape. The biased model such as MSCFS failed to recognize the oversized tumor, yielding broken tumor segmentation with only 1.58% tumor

Dice, as indicated by the white arrow in the fourth row of Fig. 6(a). Additionally, nnUNet and APAUNet also yielded discontinuous pancreas segmentation, with pancreatic Dice scores of only 66.48% and 39.60%, respectively. However, CausegNet is capable of capturing causality, specifically, that “the pancreas is contiguous, and the tumor must appear inside the pancreas”. To achieve this, it first determines the relative positions of the pancreas and tumor, and then completely segments them based on interlayer continuity. This effectively addresses the discontinuous and inaccurate segmentation, resulting in a Dice score of 94.73%. Second, Fig. 6(b) presents a common scenario in analyzing an inconspicuous pancreatic head solid tumor, which lacks contrast with surrounding tissue. Pancreatic duct dilation usually has a strong positional causal relationship with solid tumors and can serve as an indirect indicator in tumor detection. Unfortunately, due to the absence of the above causality, the biased models such as LOGISMOS, MSCFS, nnUNet, Swin UNETR, and APAUNet mistakenly identified pancreatic duct dilation as a tumor. In addition, disturbed by the complex background organs, they may segment some irrelevant regions, as shown in the white arrow in the fourth row of Fig. 6(b). In contrast, CausegNet accurately inferred the tumor location by relying on pancreatic duct dilation as a reliable indicator, eliminating background interference and outliers. This resulted in significantly improved accuracy, with 87.47% and 69.63% Dice scores for the pancreas and tumor, respectively, surpassing the SOTA performance in this case by 3.59% and 15.18%. Fig. 6(c) shows pancreatic duct dilation appearing as a low-density shadow, challenging its differentiation from low-density cystic tumors. Biased models misclassified the dilation as a tumor, yielding a maximum tumor Dice score of 77.02%. However, CausegNet correctly differentiated between the two, achieving pancreas and tumor Dice scores of 88.69% and 85.17%, respectively.

In summary, our CausegNet method offers three key advantages over current DL pancreatic segmentation approaches. Firstly, it achieves superior accuracy in pancreatic and tumor co-segmentation, which is crucial for the development of effective computer-aided diagnosis systems. Secondly, it is capable of handling common and challenging clinical cases and reconstructing a complete 3D pancreatic morphology that is highly consistent with anatomical structure. This facilitates the detection of early, inconspicuous pancreatic tumors, as well as malignant tumors with dilated pancreatic ducts. Third, to the best of our knowledge, CausegNet is the first to incorporate causality into the segmentation network. This allows us to exclude background organ interference and learn the causal relationship between the pancreas and tumor without relying on manual selection or multi-stage methods. This may be the essential reason for the above competitive results.

5.2. Unbiasedness and interpretability of CausegNet

To evaluate the unbiasedness of different networks, we visualized the multi-scale features from each layer, providing insights into the black box processing of DL models. Fig. 7 shows the features of two typical pancreatic types, from which we can make three observations.

First, when comparing feature extraction methods on the same scale,

Table 2

Pancreas cancer segmentation evaluation matrix of the compared algorithms on two datasets.

Algorithms	MSD pancreas				Huashan pancreas			
	Dice (%): mean (std)		NSD (%): mean (std)		Dice (%): mean (std)		NSD (%): mean (std)	
	Pancreas	Tumor	Pancreas	Tumor	Pancreas	Tumor	Pancreas	Tumor
LOGISMOS	78.86 (6.62)	48.54 (28.66)	96.73 (4.14)	72.12 (32.08)	82.03 (14.76)	75.93 (24.71)	96.00 (6.98)	87.45 (25.88)
MSCFS	76.42 (8.72)	51.95 (27.98)	94.87 (7.14)	73.97 (32.81)	79.93 (16.17)	70.65 (28.17)	94.58 (10.78)	83.11 (26.78)
nnUNet	79.59 (7.79)	65.04 (26.22)	95.90 (5.29)	83.89 (27.39)	84.91 (13.33)	79.69 (18.93)	94.48 (13.42)	91.41 (14.90)
Swin UNETR	80.04 (8.61)	62.92 (27.28)	96.15 (4.98)	85.28 (25.24)	85.06 (12.57)	77.25 (23.50)	96.02 (8.90)	90.73 (14.76)
APAUNet	79.32 (9.41)	65.81 (22.79)	95.68 (6.98)	86.88 (23.24)	84.80 (13.34)	80.60 (19.35)	95.68 (9.18)	90.90 (15.78)
CausegNet	81.15 (7.80)	71.09 (21.41)	96.88 (3.81)	90.97 (21.63)	86.67 (11.71)	84.28 (16.43)	96.97 (7.86)	95.12 (9.23)

generate multiple tasks to simulate real-world situations that may arise during the practical application of our model.

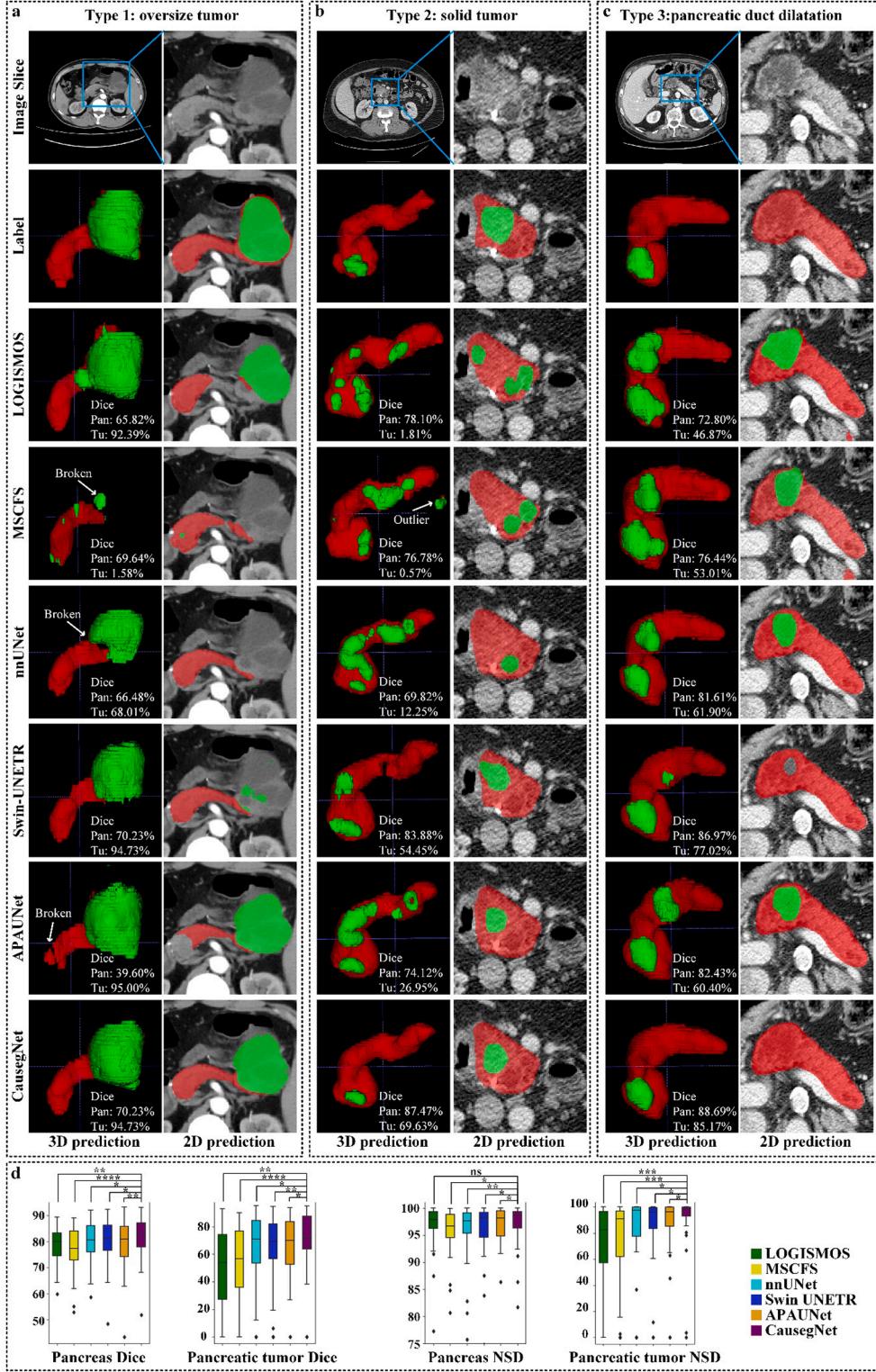


Fig. 6. Pancreas and tumor co-segmentation results of typical types with highly segmented difficulty. **a** Patient accompanied by an extremely oversized tumor. The first row is the image in 2D view, where the blue box enlarges the foreground area to show it more clearly. The second to sixth rows are label and segmentation results of various methods from 2D and 3D views. The red and green regions represent the pancreas and tumor respectively. The white numbers in the 3D views indicate the pancreas and tumor Dice scores. **b** Low-contrast solid tumor in the pancreatic head. **c** Severe pancreatic duct dilation. **d** Dice and NSD of the compared methods. ****: $p < 0.0001$; ***: $0.0001 \leq p < 0.001$; **: $0.001 \leq p < 0.01$; *: $0.01 \leq p < 0.05$; ns, not significant, $p \geq 0.05$.

our CausegNet demonstrates high unbiasedness. Specifically, Fig. 7(a) shows a narrow pancreas and a small pancreatic head tumor, along with heterogeneous background organs. In such a scenario, biased models extracted mixed features that encompassed irrelevant background

organs. Despite attempts at manual cropping and coarse-to-fine methods, these models encountered challenges in accurately concentrating on the pancreas and tumor. For example, the first layer of LOGISMOS and Swin UNETR features erroneously focused on the kidney

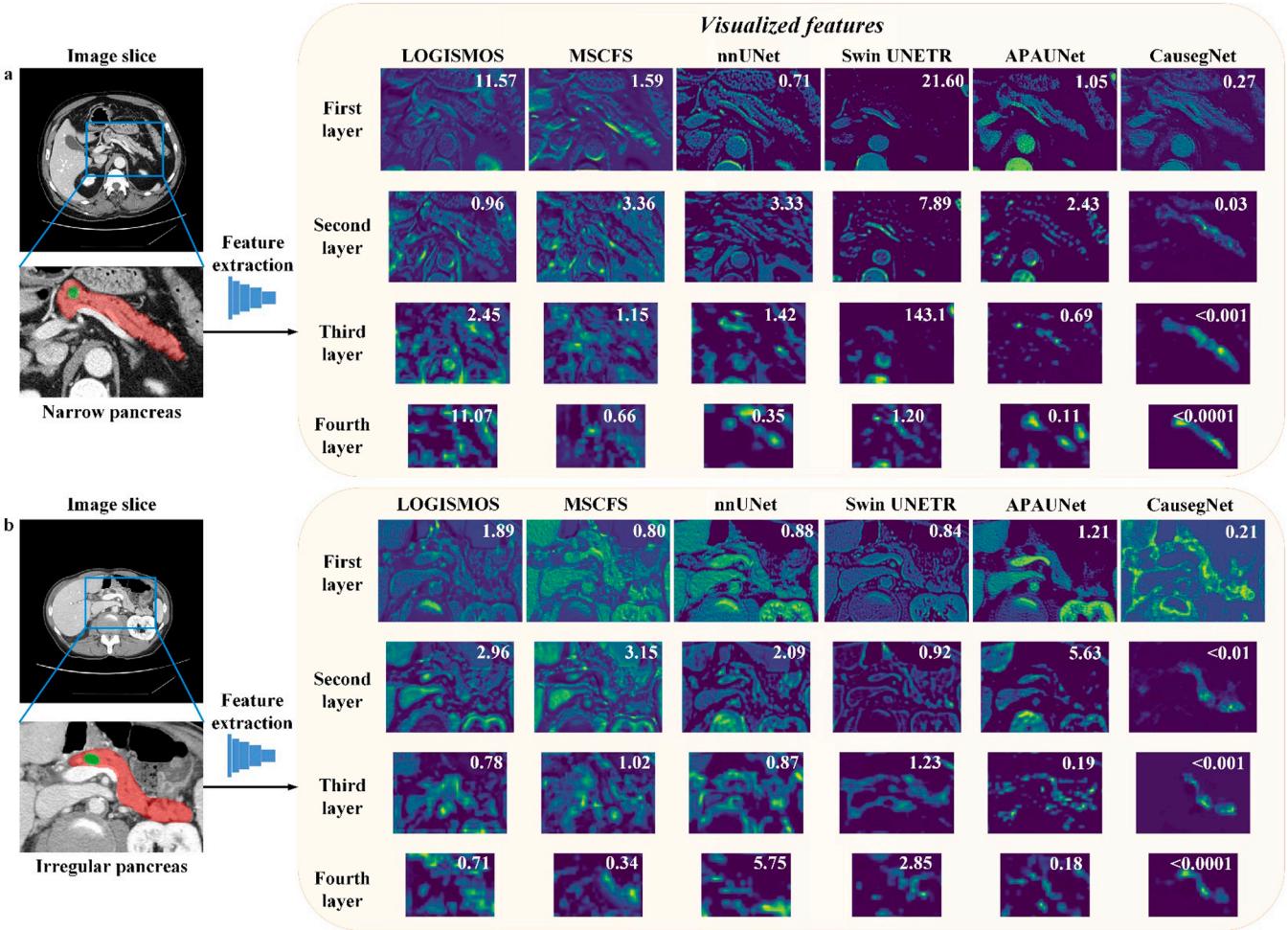


Fig. 7. Extracted multi-scale features of typical cases. a and b are the features of a narrow pancreas with a small tumor, and a bent pancreas with low-contrast and irregular boundaries. The first column is a 2D image slice, where the blue box enlarges the foreground area to show it more clearly. The red and green regions represent the pancreas and tumor respectively. The visualized features show the first to fourth layer features of various methods. The white numbers in the features indicate its FNMSE.

or spinal region, resulting in an abnormal FNMSE of 11.57 and 21.60. In addition, some attention-based methods, like APAUNet, while extracting pancreatic information, may still exhibit limitations in effectively filtering the background. In contrast, CausegNet successfully decouples the foreground causal features and sufficiently suppresses the background bias. In the first layer, CausegNet's FNMSE was 0.27, significantly lower compared to other methods.

Second, when comparing features at different scales, our causal model effectively corrects erroneous features. Biased networks, as shown in Fig. 7(a), tend to amplify irrelevant background bias during multi-scale feature extraction. In contrast, CausegNet gradually filtered out useless information and accurately focused on the pancreatic region as the network went deeper. This is evident from the drop in CausegNet's FNMSE from 0.27 to less than 0.0001 from the first to the fourth layer. At the final layer, CausegNet fully concentrated on the pancreatic region, highlighting important features like the tumor and dilated pancreatic duct.

Third, our model exhibits robustness in handling deformable pancreases with irregular boundaries. In Fig. 7(b), we observe a low-contrast pancreas surrounded by intricate background tissues. A biased model fails to distinguish the irregular boundaries of the pancreas, resulting in a cluttered and non-targeted feature extraction. In contrast, CausegNet effectively separates the features and progressively reduces the FNMSE from 0.21 to below 0.0001.

In summary, CausegNet offers several advantages. Our model

extracts more meaningful features by embedding causality, and the visualized features enhance its causal interpretability. In addition, it has a correction effect that prevents the propagation of erroneous semantic information in the multi-scale network. Hence, our method is applicable to various pancreatic morphologies, ensuring its versatility.

5.3. Prediction invariance of CausegNet

Prediction invariance asserts that a causal model should remain stable despite changes in usage scenarios and slight perturbations since causality is an intrinsic anatomical property of images. In this section, we aim to demonstrate the prediction invariance of CausegNet through two aspects: model generalizability and noise resistance.

To test model generalizability, we used both the internal MSD pancreas dataset and the external Huashan dataset to simulate the scenario changes when a trained AI system is delivered to actual clinical deployment. First, a model was trained on the MSD pancreas dataset and the external Huashan dataset to simulate the scenario changes when a trained AI system is delivered to actual clinical deployment. First, a model was trained on the MSD pancreas dataset, and then its performance was tested on the unseen Huashan dataset. The results are shown in Fig. 8(a). In the right subfigure, our method shows a higher NSD and enhanced stability, resulting in a lesser decline in NSD when compared to other methods. In the left panel of Fig. 8(a), the variation in Dice coefficients followed a similar pattern to that of NSD, proving a good

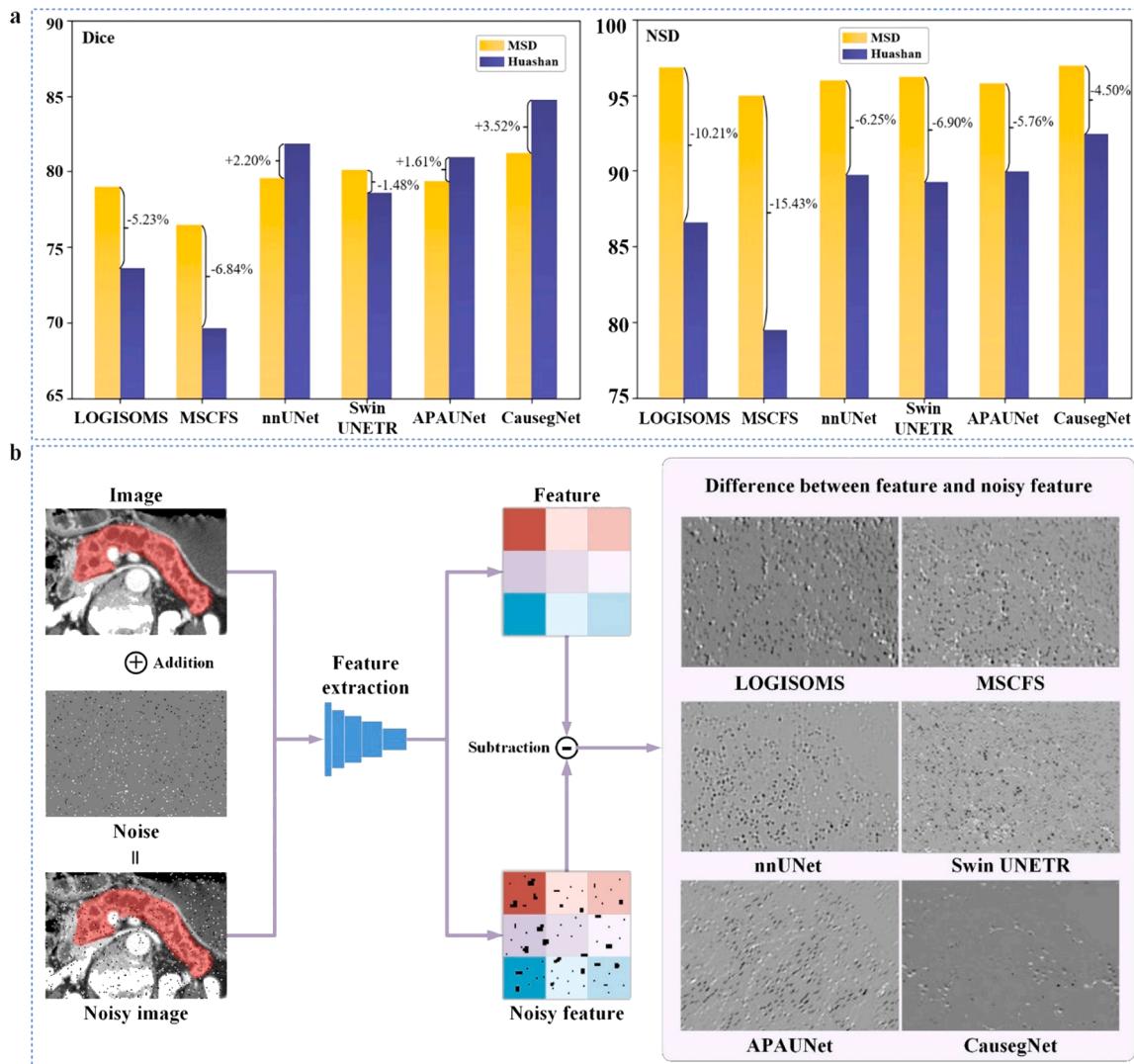


Fig. 8. Comparison of the prediction invariance. a Comparison of segmentation performance changes from trained dataset to unseen dataset. This bar graph illustrates the changes in segmentation performance between two datasets: MSD (shown in orange) and Huashan (shown in blue). b Comparison of the noise resistance of different models from the feature extraction perspective.

ability to cope with unseen datasets.

To assess the noise resistance of different models, we added random salt and pepper noise to the original 3D CT images, as shown in Fig. 8(b). Then the differences between the two features (extracted from the original image and noisy image) were calculated to observe the noise resistance of various networks (i.e., the smaller difference means the stronger noise resistance). Our findings indicate that as the five biased models extract global features from the whole image, they were significantly influenced by the noise existing in the large volume background region, leading to a considerable number of redundant points. In contrast, CausegNet is insensitive to the irrelevant region, thus the feature difference was relatively small, indicating it can maintain stable feature extraction even in the presence of slight perturbations.

Overall, CausegNet achieves better prediction invariance by focusing exclusively on foreground regions while disregarding background interference. Consequently, during the deployment of AI systems, our approach is more stable and generalizable in dealing with noisy and unseen datasets.

5.4. Transportability of the prior causality

Transportability means the license to transfer causal effects from one

Table 3

Liver cancer and spleen segmentation evaluation matrixes of the compared algorithms.

Algorithms	MSD liver				MSD spleen	
	Dice		NSD		Dice	NSD
	Liver	Tumor	Liver	Tumor		
LOGISMOS	95.17 (1.47)	74.51 (18.74)	97.46 (1.51)	89.65 (19.09)	92.19 (6.08)	98.93 (1.97)
MSCFS	94.73 (1.89)	68.85 (21.89)	96.58 (2.66)	84.47 (20.97)	91.36 (7.60)	98.68 (1.92)
nnUNet	95.95 (1.17)	76.22 (19.48)	97.26 (1.61)	87.84 (20.64)	93.16 (7.62)	94.09 (10.04)
Swin UNETR	96.00 (1.37)	74.56 (23.00)	96.92 (2.01)	85.16 (25.78)	95.26 (3.89)	96.29 (7.16)
APAUNet	95.90 (1.39)	76.07 (23.36)	96.78 (2.16)	87.40 (26.12)	96.19 (1.41)	98.47 (3.28)
CausegNet	96.48 (0.98)	79.74 (18.33)	98.10 (1.18)	90.67 (19.69)	97.17 (1.36)	99.85 (0.20)

task to another. To be specific, the abovementioned prior causality should be widely applicable to various segmentation tasks. Here, we tested it on two additional segmentation tasks involving adjacent organs

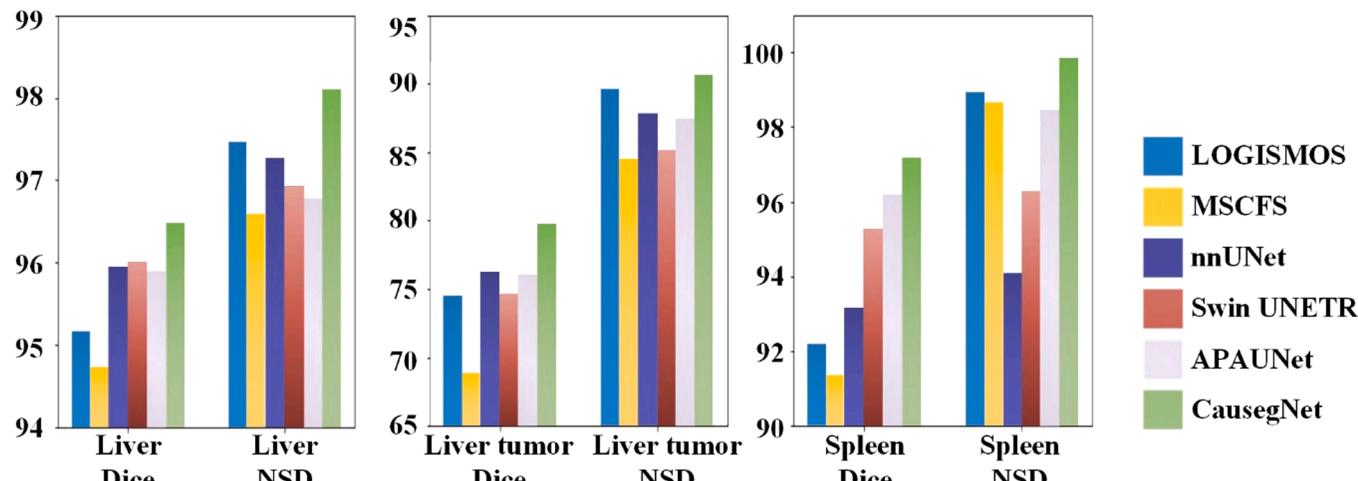


Fig. 9. Liver cancer and spleen segmentation evaluation matrixes. The left, middle, and right subplots display the segmentation matrixes of the liver, liver tumor, and spleen, respectively.

of the pancreas: liver and liver tumor co-segmentation, and spleen segmentation. The compared methods and their configurations were the same as those used in the pancreas cancer segmentation task, while the hyperparameters were modified. The experimental results are presented in Table 3, Fig. 9. And the visualized features are in Supplementary Fig. 3.

Based on the above results, we can draw two important observations. On the one hand, our method achieves the best segmentation results in both Table 3 and Fig. 9, proving our method is widely available in various segmentation tasks. For example, in liver tumor segmentation, we achieved improvements of 3.52% and 1.02% in Dice and NSD, respectively, compared to the best performance obtained by other methods. With regards to the spleen segmentation, CausegNet obtained a Dice of 97.17% and an NSD of 99.85%, which reflects an increase of 0.98% and 0.92%, respectively. On the other hand, as shown in Supplementary Fig. 3, our network demonstrates a consistently high degree of unbiasedness across these two segmentation tasks, thereby indicating the robustness and validity of the prior causality principle.

The experiments demonstrate our method is highly transportable beyond the pancreatic cancer co-segmentation task. Incorporating causality into DL networks may be a technique that could be beneficial in various segmentation tasks.

5.5. Parameter configurations of CausegNet

We evaluated the effects of parameter settings on CausegNet from two perspectives: the ratios of training samples and the weights of loss functions.

First, we assessed how the number of training samples affected our algorithm's performance. Specifically, on the MSD pancreas dataset, while keeping the test set consistent, we trained segmentation models sequentially using 85% and 70% of the training samples, and then evaluated their segmentation performance on the test set. Fig. 11 illustrates the changes in the segmentation performance of different methods. As the training sample size decreases, all segmentation methods experience a decrease in pancreas/tumor Dice scores by 9.58%/15.26%, 7.97%/13.80%, 6.63%/12.66%, 6.38%/11.29%, and 7.49%/13.06%, 5.31%/8.25%, respectively. This happens because, with a smaller training set, there are fewer samples for network learning, making it more likely to encounter unseen samples during testing. In addition, the reduction in pancreas/tumor Dice scores for our CausegNet is comparatively minor compared to other segmentation approaches, by 4.27%/7.01%, 2.66%/5.55%, 1.32%/4.41%, 1.07%/3.04%, and 2.18%/4.81%, respectively. The reason is that other biased methods

learn a correlation between training set images and segmentation results, making them more sensitive to variations in training samples. In contrast, our approach learns causal relationships within images, a more fundamental and universal relationship than correlation, which is less affected by changes in the training set.

Second, we conduct a parameter analysis of λ_1 and λ_2 in Eq. (4) across two datasets and three organs. λ_1 and λ_2 are the weights of counterfactual loss \mathcal{L}_{CF} and segmentation loss \mathcal{L}_{Seg} , respectively. They adjust the degree of constraint on network unbiasedness and segmentation accuracy during the network optimization process. To simulate both unbalanced and balanced scenarios between \mathcal{L}_{CF} and \mathcal{L}_{Seg} , we define three cases: $\lambda_1=0.5$, $\lambda_2=1.5$ ($\lambda_1 < \lambda_2$); $\lambda_1=\lambda_2=1$; $\lambda_1=1.5$, $\lambda_2=0.5$ ($\lambda_1 > \lambda_2$). The impacts of these different settings on segmentation performance are presented in Table 4 and Supplementary Table 3.

5.6. Pancreatic cancer risk assessment

In this section, based on the former co-segmentation results, we design a DL pancreatic cancer risk assessment system to demonstrate the importance of co-segmentation in disease diagnosis tasks. We compare the accuracy of our DL system with clinical diagnosis performed by experienced doctors, showcasing the potential of our co-segmentation method for clinical applications.

In Table 1, among four subtypes of tumors in the Huashan dataset, SCN is considered a low-risk tumor that typically does not require surgical treatment and can be closely monitored for changes. In contrast, other subtypes are deemed to have high malignant risk and surgical resection is usually recommended in clinical practice. However, distinguishing SCN from other subtypes manually is challenging due to similarities in imaging and clinical presentation, even for experienced doctors. Therefore, preoperative pancreatic cancer risk assessment is crucial for developing treatment plans and tumor prognosis.

To address this clinical need, a DL computer-aided diagnostic system is developed for preoperative pancreatic cancer risk assessment. A detailed network framework is available in the supplementary

Table 4

Dice scores under different loss weight settings for the MSD pancreas dataset.

Weight settings	MSD Pancreas dataset	
	Pancreas Dice	Tumor Dice
$\lambda_1=0.5$, $\lambda_2=1.5$	79.82 (8.32)	69.73 (21.96)
$\lambda_1=1.5$, $\lambda_2=0.5$	80.03 (8.13)	69.01 (22.38)
$\lambda_1=\lambda_2=1$	81.15 (7.80)	71.09 (21.41)

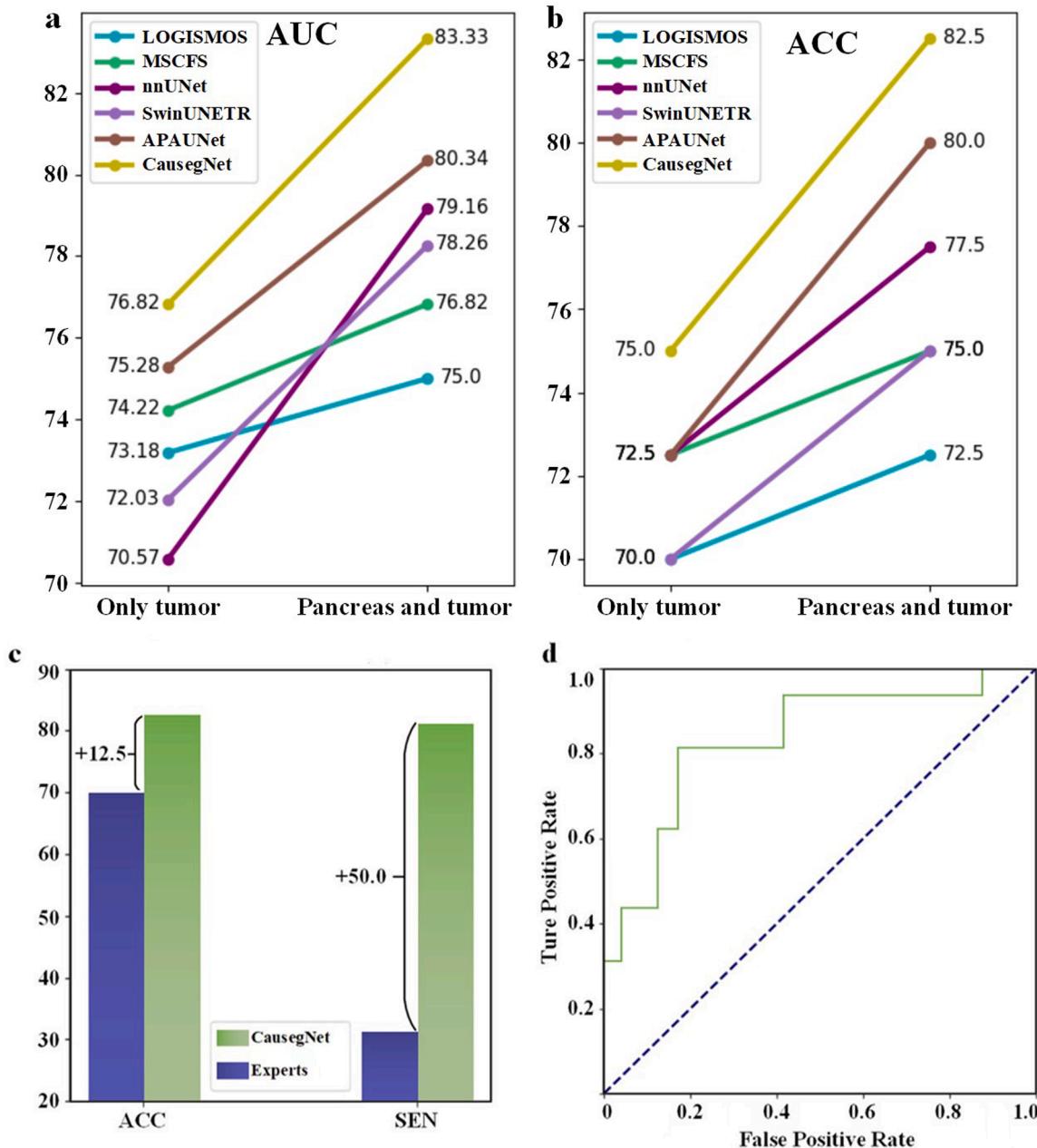


Fig. 10. Performance of pancreatic cancer risk assessment. a and b AUC and ACC change across various methods when using single target segmentation or co-segmentation as inputs. c A comparison of diagnostic performance between our DL system and expert diagnosticians. d Diagnostic ROC curve of our DL system.

materials. Here, briefly speaking, according to the co-segmentation results, the pancreas and tumor images are combined as co-input and sent to a ResNet to predict a binary diagnostic result. As a comparison, the tumor image alone is fed independently into the network to simulate the single-target segmentation input. In the evaluation metrics, a higher SEN indicates that more SCN were identified correctly.

The experimental results are shown in Fig. 10. Compared to only a single tumor as input, using co-segmentation can improve diagnostic accuracy, which is widely acknowledged by all segmentation algorithms. For instance, in Fig. 10(a) and (b), the AUC and ACC of CausegNet increased from 76.82% and 75.00% to 83.33% and 82.50%. Similar conclusions can be drawn for other segmentation algorithms, demonstrating that co-segmentation provides inherent and widespread advantages. Another important discovery is that when using co-segmentation as input, our method outperforms others due to its higher segmentation accuracy. In Fig. 10(a), our method achieved an

AUC improvement of 8.33%, 6.15%, 5.07%, 4.17%, and 2.99%, respectively, compared to other competing methods. Likewise, in Fig. 10(b), our method achieved an increase in ACC of 10.00%, 7.50%, 7.50%, 5.00%, and 2.5%, respectively.

It is worth emphasizing that our method exceeds the diagnostic accuracy of experienced experts, which makes our system promising to provide a reference for clinical diagnosis. For expert manual classification, we directly utilized the authentic clinical diagnostic records of these cases. These records encompass comprehensive patient data and diagnostic outcomes documented by a team of pancreatic surgeons at Huashan Hospital of Fudan University. The surgeons hold an average of over 10 years of clinical experience. They diagnosed the patients pre-operatively as either low-risk SCN or other high-risk subtypes of PCN. Subsequently, all these patients underwent surgical resection at Huashan Hospital, with intraoperative pathology results serving as the gold standard. Fig. 10(c) shows that only 31.25% of low-risk SCNs were

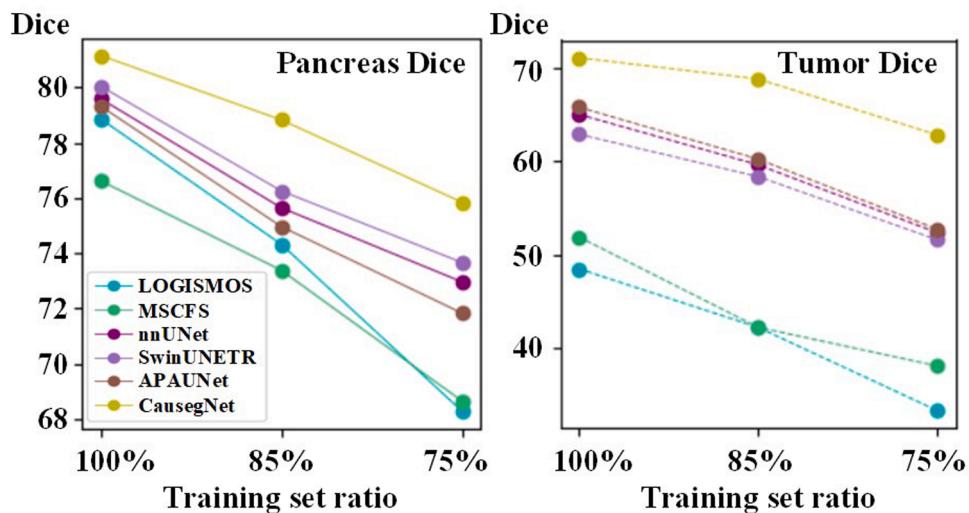


Fig. 11. Dice changes as the training sample size decreases. a represents the changes in pancreatic Dice. b represents the changes in tumor Dice.

correctly diagnosed by experts before surgery, which leads to unnecessary surgery and a waste of medical resources. In contrast, our DL diagnostic system reached 82.50% ACC and 81.25% SEN, improving accuracy by 12.50% and sensitivity by 50.00% compared to experienced experts. The p-value between expert diagnoses and network predictions was $0.002 < 0.05$, showing a significant statistical difference. This is an exciting result that may demonstrate the potential of our unbiased causal co-segmentation approach and DL diagnostic system in avoiding overtreatment.

Overall, the DL diagnostic system based on our co-segmentation method enhances the accuracy of preoperative pancreatic cancer risk prediction. In fact, it surpasses the accuracy achieved by experienced clinicians. This outcome is significant in terms of developing and implementing diagnostic protocols, as well as improving patient prognosis.

6. Discussion

In this study, we propose a deep causal learning model called CausegNet for pancreas and tumor co-segmentation in 3D CT images, which is the first to incorporate causal relationships into a DL segmentation network.

One of the key properties of our approach is its unbiasedness and embedded causality, which allows us to obtain the best segmentation performance and causal interpretability. Firstly, compared to existing biased networks, our method fundamentally removes interference from the background. Currently, one of the biggest difficulties in the pancreatic segmentation task is the long-tailed distribution, where the volume of background organs may account for 99.9% of a whole 3D CT image. The extreme category imbalances may cause network weights to be biased towards the background even ignoring foreground segmentation. Our method essentially eliminates complex background biases, preventing it from focusing on irrelevant areas and learning incorrect semantic information, which is a common issue in biased networks. For example, as shown in Fig. 7(a), the biased LOGISMOS erroneously focused on the kidney region in the first layer, while our method thoroughly filters out the background area. It is worth emphasizing that, despite incorporating attention mechanisms such as MSCFS, SwinUNETR, and APAUNet, due to the absence of additional supervision for decoupling foreground and background, as well as the insufficient learning and inference capabilities for causality, they still cannot fully concentrate on the foreground region or adequately filter the background. Additionally, we also remove segmentation outliers that occur in low-contrast backgrounds. For instance, in Fig. 6(b), the biased

MSCFS produced segmentation outliers in irrelevant regions, which were eliminated by our CausegNet. Secondly, our network pays sufficient attention to learning the causal relationship between the pancreas and tumors. Another challenge in the pancreas and tumor segmentation task is the complex pancreas morphology and the hidden tumor characteristics. Current biased DL networks struggle to encode and learn diverse features, and thus cannot handle some difficult-to-segment cases that are abundant in clinical data. Instead, CausegNet learns a large number of causalities between the pancreas and tumors to assist in segmentation. Consequently, our approach solves challenges such as abnormal pancreas morphology caused by oversized tumors, solid tumors with no obvious grayscale differences, and severe pancreatic duct dilation, as demonstrated in Fig. 6. Meanwhile, the visualization in Fig. 7 demonstrates features in CausegNet are fully decoupled and targeted, increasing the causal interpretability of the DL model. Thirdly, compared to manual cropping-based methods, the CausegNet directly segments the entire 3D CT image without requiring human involvement, increasing the algorithm's objectivity and reproducibility. In addition, this single-stage approach effectively eliminates the propagation of segmentation errors between different stages in coarse-to-fine methods.

Another significant contribution of this study is the comprehensive experiments in large-scale multi-center multi-organ datasets, which demonstrated the stability, noise resistance, and generalizability of our approach. On the one hand, our causal model shows good prediction invariance, which facilitates the clinical deployment of AI systems. Inconsistencies and noise in the data can arise due to the use of different instruments and sampling environments across various hospitals. Biased segmentation models are unable to grasp the unique essential part of an image. In such cases, performing inference based on the confused image features is problematic since data discrepancies can significantly impact segmentation performance. As shown in Fig. 8(a), biased models suffer from substantial performance degradation when applied to unseen datasets. In contrast, causality is an inherent anatomical property between the pancreas, tumor, and background, which remains constant despite data variations. By leveraging the causality, our unbiased model remains focused on the foreground region in unseen data, leading to better stability. Additionally, when the data is noisy, biased models tend to extract global features that are influenced by the noise in the background, leading to a lot of noise points in the features. While our unbiased causal model is inherently insensitive to background information, regardless of whether it contains noise or not. As a result, as shown in Fig. 8(b), our method exhibits a better ability to suppress noise. On the other hand, our causal model exhibits superior transportability in liver and tumor co-segmentation as well as spleen segmentation, proving that

causality can be considered in various segmentation tasks. Among the extended target organs, liver tumors are difficult to identify due to their multiplicity, irregularity, and significant volume variation. By understanding the transported causality in the images, CausegNet focuses more on the target area, extracting valuable features such as tumor location, edges, and organ morphology. This enables our model to perform better in the segmentation of liver tumors and spleen.

Last but not least, our work contributes by showcasing the positive impact on the downstream DL pancreatic cancer risk assessment, highlighting its clinical practical perspectives. The experimental results suggest that combining pancreas and tumor information improves the accuracy of pancreatic cancer risk assessment. Most pancreas-related segmentation tasks only perform single-objective segmentation of the pancreas or tumor, rather than co-segmentation. When only tumor information is used, the classification network focuses on learning the differences between low-risk and high-risk tumors, such as their morphology, texture, intensity, and internal cyst. In contrast, combining pancreas and tumor information involves comprehensive features, such as tumor location, pancreatic duct dilation, and pancreatic shape abnormalities, which are pivotal for pancreatic cancer diagnosis. With the accurate segmentation results, the DL diagnostic system we developed has surpassed clinical doctors in terms of diagnostic accuracy, particularly in sensitivity. It is attributed to the fact that our DL system can uncover high-throughput image features that are difficult for the naked eye to detect, thus mitigating subjectivity.

7. Conclusion

In conclusion, this study presents an accurate, generalized, and clinically applicable unbiased causal segmentation model for pancreatic and tumor co-segmentation in 3D CT images. Our method was validated through comprehensive experiments on large-scale multicenter data. It highlights the potential of deep causal learning in clinical pancreatic cancer diagnosis.

CRediT authorship contribution statement

Chengkang Li: Writing – review & editing, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization, Writing – original draft. **Yishen Mao:** Writing – review & editing, Investigation, Data curation. **Shuyu Liang:** Writing – original draft, Visualization. **Ji Li:** Supervision, Funding acquisition, Data curation. **Yuanyuan Wang:** Supervision, Funding acquisition, Data curation. **Yi Guo:** Funding acquisition, Formal analysis, Data curation, Conceptualization, Investigation, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare no competing interests in this work.

Data and code availability

The data used in this study can be retrieved from <https://pan.baidu.com/s/1zfe3ZeFNSlTChnw53jILLg> with a password of v15d. Code and trained models are available at <https://pan.baidu.com/s/15UWVbyXwarBmphai5wc9GA> with a password of b6pj.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grants 81830058 and 82227803), and the Science and Technology Commission of Shanghai Municipality (Grants 22ZR1404800 and 22DZ1100101).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neunet.2024.106294.

References

- Abbasnejad, E., Teney, D., Parvaneh, A., Shi, J., & Hengel, A. van den (2020). Counterfactual vision and language learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 10044–10054).
- Antonelli, M., Reinke, A., Bakas, S., Farahani, K., et al. (2022). The medical segmentation decathlon. *Nature Communications*, 13(1), 4128.
- Cai, J., Lu, L., Zhang, Z., Xing, F., Yang, L., & Yin, Q. (2016). Pancreas segmentation in MRI using graph-based decision fusion on convolutional neural networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 442–450).
- Chen, L., Zhang, H., Xiao, J., He, X., Pu, S., & Chang, S. (2019). Counterfactual critic multi-agent training for scene graph generation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 4613–4623).
- 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 Transactions on Medical Imaging*, 41(1), 75–87.
- Chen, Z., Wang, X., Yan, K., & Zheng, J. (2020). Deep multi-scale feature fusion for pancreas segmentation from CT images. *International Journal of Computer Assisted Radiology and Surgery*, 15, 415–423.
- Cui, P., & Athey, S. (2022). Stable learning establishes some common ground between causal inference and machine learning. *Nature Machine Intelligence*, 4(2), 110–115.
- Dai, S., Zhu, Y., Jiang, X., Yu, F., Lin, J., & Yang, D. (2023). TD-Net: Trans-Deformer network for automatic pancreas segmentation. *Neurocomputing*, 517, 279–293.
- Dash, S., Balasubramanian, V. N., & Sharma, A. (2022). Evaluating and mitigating bias in image classifiers: A causal perspective using counterfactuals. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (pp. 915–924).
- Dogan, R. O., Dogan, H., Bayrak, C., & Kayikcioglu, T. (2021). A two-phase approach using mask R-CNN and 3D U-Net for high-accuracy automatic segmentation of pancreas in CT imaging. *Computer Methods and Programs in Biomedicine*, 207, Article 106141.
- Dumitru, R., Peteleaza, D., & Craciun, C. (2023). Using DUCK-Net for polyp image segmentation. *Scientific Reports*, 13(1), 9803.
- Farag, A., Lu, L., Roth, H. R., Liu, J., Turkbey, E., & Summers, R. M. (2016). A bottom-up approach for pancreas segmentation using cascaded superpixels and (deep) image patch labeling. *IEEE Transactions on Image Processing*, 26(1), 386–399.
- Gunderson, A. J., Kaneda, M. M., Tsujikawa, T., Nguyen, A. V., et al. (2016). Bruton tyrosine kinase-dependent immune cell cross-talk drives pancreas cancer. *Cancer Discovery*, 6(3), 270–285.
- Guo, Z., Zhang, L., Lu, L., Bagheri, M., Summers, R. M., Sonka, M., et al. (2018). Deep LOGISMOS: Deep learning graph-based 3D segmentation of pancreatic tumors on CT scans. *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, 1230–1233.
- Huynh, V., Kreinovich, V., & Sriboonchitta, S. (2016). *Causal inference in econometrics*. Springer.
- Imbens, G. W., & Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.
- Isensee, 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.
- Jiang, Y., Zhang, Z., Qin, S., Guo, Y., Li, Z., & Cui, S. (2023). APAUNet: Axis Projection Attention UNet for Small Target in 3D Medical Segmentation. In *13846. 2022 Asian Conference on Computer Vision (ACCV 2022)* (pp. 21–36).
- Khan, T. M., Naqvi, S. S., Robles-Kelly, A., & Razzaq, I. (2023). Retinal vessel segmentation via a Multi-resolution Contextual Network and adversarial learning. *Neural Networks*, 165, 310–320. <https://doi.org/10.1016/j.neunet.2023.05.029>
- Kleeff, J., Korc, M., Apte, M., La Vecchia, C., Johnson, C. D., Biankin, A. V., et al. (2016). Pancreatic cancer. *Nature Reviews Disease Primers*, 2(1), 1–22.
- Kolling, C., More, M., Gavenski, N., Pooch, E., Parraga, O., & Barros, R. C. (2022). Efficient counterfactual debiasing for visual question answering. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (pp. 3001–3010).
- Kuang, S., Woodruff, H. C., Granzier, R., Nijhatten, T. J. A. van, Lobbes, M. B. I., Smidt, M. L., et al. (2023). MSCDA: Multi-level semantic-guided contrast improves unsupervised domain adaptation for breast MRI segmentation in small datasets. *Neural Networks*, 165, 119–134.
- Li, C., Mao, Y., Guo, Y., Li, J., & Wang, Y. (2022). Multi-dimensional cascaded net with uncertain probability reduction for abdominal multi-organ segmentation in CT sequences. *Computer Methods and Programs in Biomedicine*, 221, Article 106887.
- Li, Q., Liu, X., He, Y., Li, D., & Xue, J. (2023). Temperature guided network for 3D joint segmentation of the pancreas and tumors. *Neural Networks*, 157, 387–403.
- Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., et al. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42, 60–88.
- Liu, S., Kailkhura, B., Loveland, D., & Han, Y. (2019). Generative counterfactual introspection for explainable deep learning. *2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 1–5.
- Luo, Y., Peng, J., & Ma, J. (2020). When causal inference meets deep learning. *Nature Machine Intelligence*, 2(8), 426–427.

- Man, Y., Huang, Y., Feng, J., Li, X., & Wu, F. (2019). Deep Q learning driven CT pancreas segmentation with geometry-aware U-Net. *IEEE Transactions on Medical Imaging*, 38 (8), 1971–1980.
- Mao, C., Cha, A., Gupta, A., Wang, H., Yang, J., & Vondrick, C. (2021). Generative interventions for causal learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 3947–3956).
- Milletari, F., Navab, N., & Ahmadi, S. (2016). V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 Fourth International Conference on 3D Vision (3DV)* (pp. 565–571).
- Mo, J., Zhang, L., Wang, Y., & Huang, H. (2020). Iterative 3D feature enhancement network for pancreas segmentation from CT images. *Neural Computing and Applications*, 32, 12535–12546.
- Nikolov, S., Blackwell, S., Zverovitch, A., Mendes, R., Livne, M., et al. (2021). Clinically applicable segmentation of head and neck anatomy for radiotherapy: Deep learning algorithm development and validation study. *Journal of Medical Internet Research*, 23 (7), e26151.
- Niu, Y., Tang, K., Zhang, H., Lu, Z., Hua, X., & Wen, J. (2021). Counterfactual vqa: A cause-effect look at language bias. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 12700–12710).
- Pang, S., Du, A., Orgun, M. A., Wang, Y., & Yu, Z. (2021). Tumor attention networks: Better feature selection, better tumor segmentation. *Neural Networks*, 140, 203–222. <https://doi.org/10.1016/j.neunet.2021.03.006>
- Prosperi, M., Guo, Y., Sperrin, M., Koopman, J. S., Min, J. S., He, X., et al. (2020). Causal inference and counterfactual prediction in machine learning for actionable healthcare. *Nature Machine Intelligence*, 2(7), 369–375.
- Richens, J. G., Lee, C. M., & Johri, S. (2020). Improving the accuracy of medical diagnosis with causal machine learning. *Nature Communications*, 11(1), 3923.
- Roth, H. R., Lu, L., Farag, A., Shin, H., Liu, J., Turkbey, E. B., et al. (2015). Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation. In , 18. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference* (pp. 556–564). October 5–9, 2015, Proceedings, Part I.
- Roth, H. R., Lu, L., Lay, N., Harrison, A. P., Farag, A., Sohn, A., et al. (2018). Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation. *Medical Image Analysis*, 45, 94–107.
- Shen, D., Wu, G., & Suk, H. (2017). Deep learning in medical image analysis. *Annual Review of Biomedical Engineering*, 19, 221–248.
- Suzuki, M., Kamcya, Y., Kutsuna, T., & Mitsumoto, N. (2021). Understanding the reason for misclassification by generating counterfactual images. In *2021 17th International Conference on Machine Vision and Applications (MVA)* (pp. 1–5).
- Tang, K., Niu, Y., Huang, J., Shi, J., & Zhang, H. (2020). Unbiased scene graph generation from biased training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 3716–3725).
- Tang, Y., Yang, D., Li, W., Roth, H., Landman, B., Xu, D., et al. (2022). Self-Supervised Pre-Training of Swin Transformers for 3D Medical Image Analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 27030–27040).
- Trinh, Q. (2023). Meta-Polyp: A baseline for efficient Polyp segmentation. In *2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS)* (pp. 742–747).
- Valanarasu, J. M. J., Oza, P., Hacihaliloglu, I., & Patel, V. M. (2021). Medical Transformer: Gated Axial-Attention for Medical Image Segmentation. *Medical Image Computing and Computer Assisted Intervention– MICCAI*, 2021, 36–46.
- Wang, C., Li, J., Zhang, F., Sun, X., Dong, H., Yu, Y., et al. (2021). Bilateral asymmetry guided counterfactual generating network for mammogram classification. *IEEE Transactions on Image Processing*, 30, 7980–7994.
- Wang, P., Shen, C., Roth, H.R., Yang, D., et al. (2020). Automated pancreas segmentation using multi-institutional collaborative deep learning. Domain Adaptation and Representation Transfer, and Distributed and Collaborative Learning: Second MICCAI Workshop, DART 2020, and First MICCAI Workshop, DCL 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4–8, 2020, Proceedings 2, 192–200.
- Wolfgang, C. L., Herman, J. M., Laheru, D. A., Klein, A. P., Erdek, M. A., Fishman, E. K., et al. (2013). Recent progress in pancreatic cancer. *CA: A Cancer Journal for Clinicians*, 63(5), 318–348.
- Wu, J. Q., Horeweg, N., de Bruyn, M., Nout, R. A., et al. (2022). Automated causal inference in application to randomized controlled clinical trials. *Nature Machine Intelligence*, 4(5), 436–444.
- Yue, Z., Wang, T., Sun, Q., Hua, X., & Zhang, H. (2021). Counterfactual zero-shot and open-set visual recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 15404–15414).
- Zhao, N., Tong, N., Ruan, D., & Sheng, K. (2019). Fully automated pancreas segmentation with two-stage 3D convolutional neural networks. In , 22. *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference* (pp. 201–209). October 13–17, 2019, Proceedings, Part II.
- Zhou, Y., Li, Y., Zhang, Z., Wang, Y., Wang, A., Fishman, E. K., et al. (2019). Hyper-pairing network for multi-phase pancreatic ductal adenocarcinoma segmentation. In , 22. *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference* (pp. 155–163). October 13–17, 2019, Proceedings, Part II.
- Zhu, Z., Xia, Y., Xie, L., Fishman, E. K., & Yuille, A. L. (2019). Multi-scale coarse-to-fine segmentation for screening pancreatic ductal adenocarcinoma. In , 22. *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference* (pp. 3–12). October 13–17, 2019, Proceedings, Part VI.