Original Article

Semisupervised 3D segmentation of pancreatic tumors in positron emission tomography/computed tomography images using a mutual information minimization and cross-fusion strategy

Min Shao[1,2]^, Chao Cheng[3], Chengyuan Hu[4], Jian Zheng[1,2], Bo Zhang[5], Tao Wang[3], Gang Jin[6]^, Zhaobang Liu[1,2]^, Changjing Zuo[3]

[1]School of Biomedical Engineering (Suzhou), Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, China; [2]Department of Medical Imaging, Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences, Suzhou, China; [3]Department of Nuclear Medicine, the First Affiliated Hospital (Changhai Hospital) of Naval Medical University, Shanghai, China; [4]Department of AI Algorithm, Shenzhen Poros Technology Co., Ltd., Shenzhen, China; [5]Department of Radiology, the Second Affiliated Hospital of Soochow University, Suzhou, China; [6]Department of Hepatobiliary Pancreatic Surgery, the First Affiliated Hospital (Changhai Hospital) of Naval Medical University, Shanghai, China

Contributions: (I) Conception and design: Z Liu, G Jin, M Shao; (II) Administrative support: Z Liu, G Jin, J Zheng; (III) Provision of study materials or patients: C Zuo, G Jin, C Cheng; (IV) Collection and assembly of data: C Zuo, G Jin, C Cheng, T Wang; (V) Data analysis and interpretation: M Shao, J Zheng, C Hu, B Zhang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Gang Jin, PhD. Department of Hepatobiliary Pancreatic Surgery, the First Affiliated Hospital (Changhai Hospital) of Naval Medical University, 168 Changhai Road, Shanghai 200433, China. Email: jingang@smmu.edu.cn; Zhaobang Liu, PhD. School of Biomedical Engineering (Suzhou), Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, China; Department of Medical Imaging, Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences, 88 Keling Road, Suzhou 215163, China. Email: liuzb@sibet.ac.cn.

Background: Accurate segmentation of pancreatic cancer tumors using positron emission tomography/computed tomography (PET/CT) multimodal images is crucial for clinical diagnosis and prognosis evaluation. However, deep learning methods for automated medical image segmentation require a substantial amount of manually labeled data, making it time-consuming and labor-intensive. Moreover, addition or simple stitching of multimodal images leads to redundant information, failing to fully exploit the complementary information of multimodal images. Therefore, we developed a semisupervised multimodal network that leverages limited labeled samples and introduces a cross-fusion and mutual information minimization (MIM) strategy for PET/CT 3D segmentation of pancreatic tumors.

Methods: Our approach combined a cross multimodal fusion (CMF) module with a cross-attention mechanism. The complementary multimodal features were fused to form a multifeature set to enhance the effectiveness of feature extraction while preserving specific features of each modal image. In addition, we designed an MIM module to mitigate redundant high-level modal information and compute the latent loss of PET and CT. Finally, our method employed the uncertainty-aware mean teacher semi-supervised framework to segment regions of interest from PET/CT images using a small amount of labeled data and a large amount of unlabeled data.

Results: We evaluated our combined MIM and CMF semisupervised segmentation network (MIM-CMFNet) on a private dataset of pancreatic cancer, yielding an average Dice coefficient of 73.14%, an average Jaccard index score of 60.56%, and an average 95% Hausdorff distance (95HD) of 6.30 mm. In addition, to verify the broad applicability of our method, we used a public dataset of head and neck cancer, yielding an average Dice coefficient of 68.71%, an average Jaccard index score of 57.72%, and an average

^ ORCID: Min Shao, 0009-0002-0540-9334; Gang Jin, 0000-0001-6713-1185; Zhaobang Liu, 0000-0002-8412-9930.

95HD of 7.88 mm.

Conclusions: The experimental results demonstrate the superiority of our MIM-CMFNet over existing semisupervised techniques. Our approach can achieve a performance similar to that of fully supervised segmentation methods while significantly reducing the data annotation cost by 80%, suggesting it is highly practicable for clinical application.

Keywords: Pancreatic cancer; tumor segmentation; positron emission tomography/computed tomography (PET/CT); semisupervised learning; multimodal segmentation

Submitted Aug 16, 2023. Accepted for publication Dec 08, 2023. Published online Jan 23, 2024.
doi: 10.21037/qims-23-1153
View this article at: https://dx.doi.org/10.21037/qims-23-1153

Introduction

Pancreatic cancer is a highly malignant tumor with a 5-year survival rate of about 6–10% after diagnosis (1-3). The complex structure of tissues around the pancreas and the similar density of adjacent tissues render the segmentation of pancreatic tumors considerably challenging. [^{18}F]-fluorodeoxyglucose (^{18}F-FDG) positron emission tomography/computed tomography (PET/CT) combine two different imaging technologies: PET reflects the metabolic information of the lesion through the uptake of the tracer, while CT accurately describes the anatomical characteristics of the lesion. Obtaining PET/CT multimodal pancreatic cancer data is valuable but challenging. Wang *et al.* have proposed the only approach for multimodal pancreatic cancer tumor segmentation, which includes a multimodal fusion and calibration network (MFCNet) (4). The fused image of PET/CT provides simultaneous visualization of organ location, shape, and functional abnormalities (5), which cannot be provided by single-modality images [such as CT or magnetic resonance imaging (MRI)]. PET/CT is widely used in clinical practice for tumor diagnosis (6), staging (7), radiomics analysis (8), treatment evaluation, and prognosis assessment (9). Tumor segmentation is essential for quantitatively analyzing PET/CT images, and combining the image information of PET and CT can improve tumor segmentation accuracy (4,10). In *Table 1*, we provide the baseline characteristics of subjects in the pancreatic cancer dataset. However, exploiting the complementary information between PET and CT images to enhance segmentation performance remains a significant challenge, but one which may be overcome through multimodal feature fusion. Multimodal image segmentation networks employ various fusion strategies, such as input-level fusion (11,12), layer-level (13-16), and late-fusion networks (17,18). An input-level fusion network regards each modality as a channel of the input image, and these multimodal image channels are concatenated as the network input. A layer-level fusion network independently extracts features from each modality and fuses them at an intermediate level in the data flow. In a late-fusion network, each modality is passed through independent encoder-decoder networks, and the learned features are fused at the end of each stream. However, these traditional multimodal methods struggle to cross-correlate information from different modalities, limiting the mutual guidance of these extracted features. Therefore, we propose a cross-modal fusion (CMF) attention-encoding module, which merges multi-modal information while preserving single-modal features, simultaneously interacting with multimodal features to enforce constraint learning within the network. Most previous studies report achieving multimodal feature fusion by directly combining features of each modality (19), but this leads to redundant features and irrelevant information and fails to fully leverage the effective features from each modality. A multimodal segmentation scheme should be capable of managing the complementarity, redundancy, and cooperation between different modalities. Therefore, we designed a mutual information minimization (MIM) module to reduce modal redundancy. This module calculates the potential loss of both PET and CT modalities, incorporating it into the loss function to separate the salient feature distributions of PET images from those of CT images. In addition, the approaches in the aforementioned studies relied on fully supervised training, which requires a considerable amount of labeled data and constitutes a significant a limitation. Therefore, we sought to identify the semisupervised segmentation networks that could reduce reliance on labeled data while maintaining

Table 1 Subject baseline characteristics

Variables	Pancreatic cancer (n=93)
Age (years)	68.6±9.8
Male	58
Female	35
Weight (kg)	61.2±10.5
Tumor diameter (cm)	3.2±1.5
CT (HU)	17.3±35.9
PET (SUV)	9.0±3.9
Histological type	PDAC
Lesion location	
Head	22
Neck	28
Body	26
Tail	17
Histopathological diagnosis/cytological examination	
Exfoliative cytologic examination	8
Needle biopsy	23
Surgical biopsy	62
TNM stage	
I A	2
I B	25
II A	12
II B	19
III	27
IV	8

Continuous variables are expressed as the mean ± standard deviation. Categorical variables are expressed as numbers. CT, computed tomography; HU, Hounsfield unit; PET, positron emission tomography; SUV, standard uptake value; PDAC, pancreatic ductal adenocarcinoma; TNM, tumor-node-metastasis.

performance.

Compared with fully supervised learning, semisupervised learning utilizes a small amount of labeled data and a large amount of unlabeled data for model training. Popular semisupervised medical image segmentation techniques employ rule-based encoder-decoder segmentation networks as their backbone (20-23). In terms of learning strategies, these methods can be categorized into self-training (24,25), adversarial learning (26,27), cotraining (28,29), contrastive learning (30,31), and consistency regularization methods (32-39). Consistency regularization methods are widely applied in semisupervised medical image segmentation and facilitate consistent model predictions for the same input under different perturbations. The most classical method is mean teacher (MT) (40,41), which learns from labeled data in a supervised manner, uses a teacher model to provide pseudolabels for unlabeled data, and maintains prediction consistency of the teacher-student model for the unlabeled data through regularization. Finally, the supervised loss and consistency loss are combined and fed back to the network to update the student model. The teacher model's parameters are acquired via the moving average of the student model's parameters rather than via updating the gradient through loss backpropagation. This operation allows the teacher model to continuously accumulate historical prediction information for unlabeled data. However, the pseudolabels generated by the teacher model may be unreliable, leading to unstable training. Therefore, Yu *et al.* (35) proposed an uncertainty-aware mean teacher (UA-MT) framework, in which the student model undergoes multiple forward propagations, and the teacher model gradually learns more reliable targets based on uncertainty estimation. In addition, Luo *et al.* (37) introduced a dual-task consistency (DTC) regularization technique, employing a dual-task deep network to jointly predict the target's pixel-wise segmentation maps and geometry-aware level set. Wu *et al.* (39) proposed a mutual consistency network (MC-Net) containing two decoders and used the prediction difference between the two as model uncertainty information to regularize model training and enhance pseudolabel quality. Subsequent studies (42-45) have employed various consistency regularization strategies to improve semisupervised model performance. In our study, we aimed to leverage the differences and complementarity between CT and PET modalities to improve model prediction consistency, and thus direct use of the consistency regularization method was deemed appropriate (46). The proposed method uses the information interaction between modalities for consistency prediction on multimodal data, overcoming the limitation of previous methods that cannot directly use multimodal information. The proposed method was designed with the UA-MT framework used as the baseline.

To the best of our knowledge, only a few studies (46-48) have examined multimodal medical image semisupervised segmentation, and even fewer studies have done so

within the context of PET/CT images. Mondal *et al.* (47) employed generative adversarial learning for semisupervised segmentation of multimodal brain MRI images, which prevents overfitting by learning to discriminate between real and fake patches from the generator network. Chartsias *et al.* (48) proposed a dense attention fluid network (DAFNet) to segment multimodal cardiac and abdominal MRI images, using disentanglement, alignment, and fusion to construct a complex network for multimodal data fusion. In the context of PET/CT research, Zhang *et al.* (46) proposed using area-similarity contrastive loss to leverage cross-modal information and prediction consistency between different modalities for contrastive mutual learning (CML). They also included a soft pseudolabel relearning scheme to address potential performance gaps between various modalities, achieving good segmentation performance on PET/CT head and neck images and multimodal brain tumor MRI images. In contrast, our study focused on 3D volume data from PET/CT images of pancreatic cancer. We employed cross-modal feature fusion and minimized mutual information feature selection to more precisely segment pancreatic tumor edges, enhancing semisupervised segmentation performance.

This study aimed to achieve more accurate segmentation of multimodal PET/CT images of pancreatic cancer by leveraging a large amount of unlabeled multimodal data and a small amount of labeled data, following the classic UA-MT semisupervised segmentation framework. In addition, experiments on a public dataset of head and neck cancer demonstrated the universal applicability of the proposed module. We hope that this study will enable peers to focus on semisupervised studies on other multimodal disease images. Our main contributions are as follows: (I) we developed a CMF module to improve feature fusion effectiveness by fusing complementary multimodal features and preserving specific features of single-modal images; (II) we used MIM to reduce feature redundancy in each modality, screening out effective multimodal features; and (III) we combined semisupervised learning with multimodality to make full use of the limited labeled data and harness complementary information from various modalities. This combination can provide more effective and reliable solutions in the field of medical imaging analysis and has promising practical applications.

Methods

Overview

In this study's overall architecture (*Figure 1*), the student and teacher models have the same network structure, receiving PET and CT images as input. The teacher model's network parameters are acquired through the exponential moving average (EMA) of the student model. The teacher model generates targets and performes T times (T =8) forward propagation with Monte Carlo dropout to estimate the uncertainty of the targets. Subsequently, the teacher model filters out the relatively unreliable (high uncertainty) predictions, selecting specific predictions as learning targets for the student model. Finally, the student model is optimized by minimizing the supervised segmentation loss (L_{seg}) on labeled data, the prediction consistency loss (L_{con}) generated by the student-teacher model, and the latent loss (L_{latent}) generated by the student model.

Optimization objectives of the semisupervised tasks

The objective function of our semisupervised segmentation framework comprised L_{seg}, L_{con}, and L_{latent} as provided by the MIM module. The training set comprised N-labeled and M-unlabeled data, denoted as $D_L = \{(x_i, y_i)\}_{i=1}^{N}$ and $D_U = \{(x_i)\}_{i=N+1}^{M}$, respectively, where $x_{\text{i}} \in R^{\text{H}\times\text{W}\times\text{D}}$ represents the 3D input volume, and $y_{\text{i}} \in \{0,1\}^{\text{H}\times\text{W}\times\text{D}}$ corresponds to the ground-truth label. The objective function of our semisupervised segmentation framework can be formulated as follows:

$$l_{total}(\theta) = \min_{\theta} L_{seg}\left(f(x_i;\theta), y_i\right) + \lambda L_{con}\left(f\left(x_i;\theta',\eta'\right), f(x_i;\theta,\eta)\right) + \upsilon L_{latent} \quad [1]$$

where L_{seg} represents the supervised segmentation loss [binary cross-entropy (BCE) loss and Dice loss (49) are used here] to evaluate the network's segmentation quality for labeled data; L_{con} denotes the unsupervised consistency loss (computed using mean squared error) to measure the consistency between predictions of the student and teacher models under the same input x_{i} and different perturbations. In Eq. [1], $f(\cdot)$ denotes the segmentation neural network; θ and θ' represent the weights of the student and teacher models, respectively; η and η' denote different perturbation

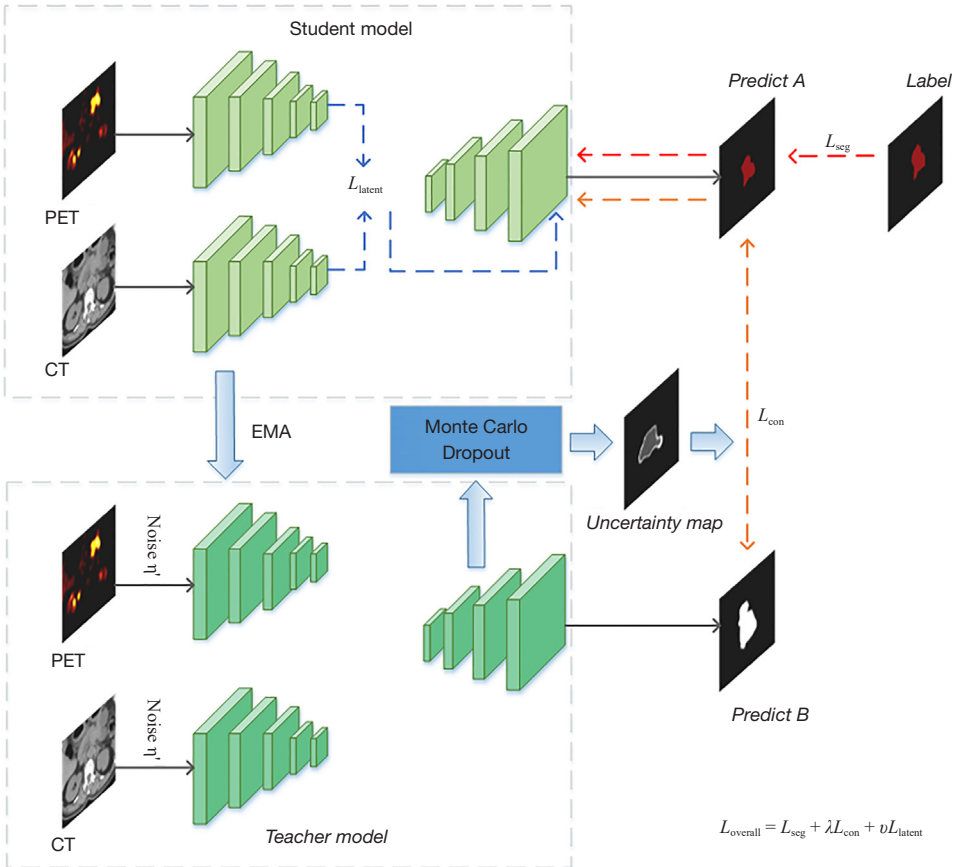

Figure 1 The proposed semisupervised segmentation framework for multimodal PET/CT medical images. PET, positron emission tomography; CT, computed tomography; EMA, exponential moving average; L_{seg}, segmentation loss; L_{con}, consistency loss; L_{latent}, latent loss.

operations (e.g., incorporating noise to the input or network dropout); λ is an ramp-up weighting coefficient used to balance L_{seg} and L_{con}, ensuring that most of the supervisory signal originates from labeled data during the early training stages; L_{seg} is computed from the labeled data only; L_{con} is unsupervised and used to supervise all the training data; and υ is a hyperparameter set to 1×10^{-5} in the pancreas dataset and 1×10^{-4} in the head and neck dataset to control the degree of minimizing mutual information and to adjust the correlation between modalities. Larger values of υ enhance the correlation between modalities, making the modalities more consistent, while smaller values of υ reduce the correlation between modalities, allowing for the difference between modalities.

Structure of the MIM-CMFNet

The improved structure of our student model, named MIM-CMFNet (*Figure 2*), consisted of four main components: two parallel encoders, one decoder, a CMF module (*Figure 3*), and an MIM module (*Figure 4*).

The encoder-decoder architecture was based on that of V-Net (23), comprising five encoding blocks for extracting PET and CT features independently. The PET and CT encoders shared the same network structure but had separate weights. We used 3×3×3 convolution kernels with a stride of 1 and BatchNorm (50) to maintain the consistency of input data distribution. Subsequently, rectified linear unit (ReLU) (51) with a negative slope of 0.01 was applied as the activation function to prevent overfitting and vanishing gradients during backpropagation. We used V-Net to act as a Bayesian network for uncertainty estimation by adding two dropout layers with a dropout rate of 0.5 in addition to the five layers of the L-stage and one layer of R-stage of V-Net. Dropout was activated during network training and uncertainty estimation but deactivated during testing since

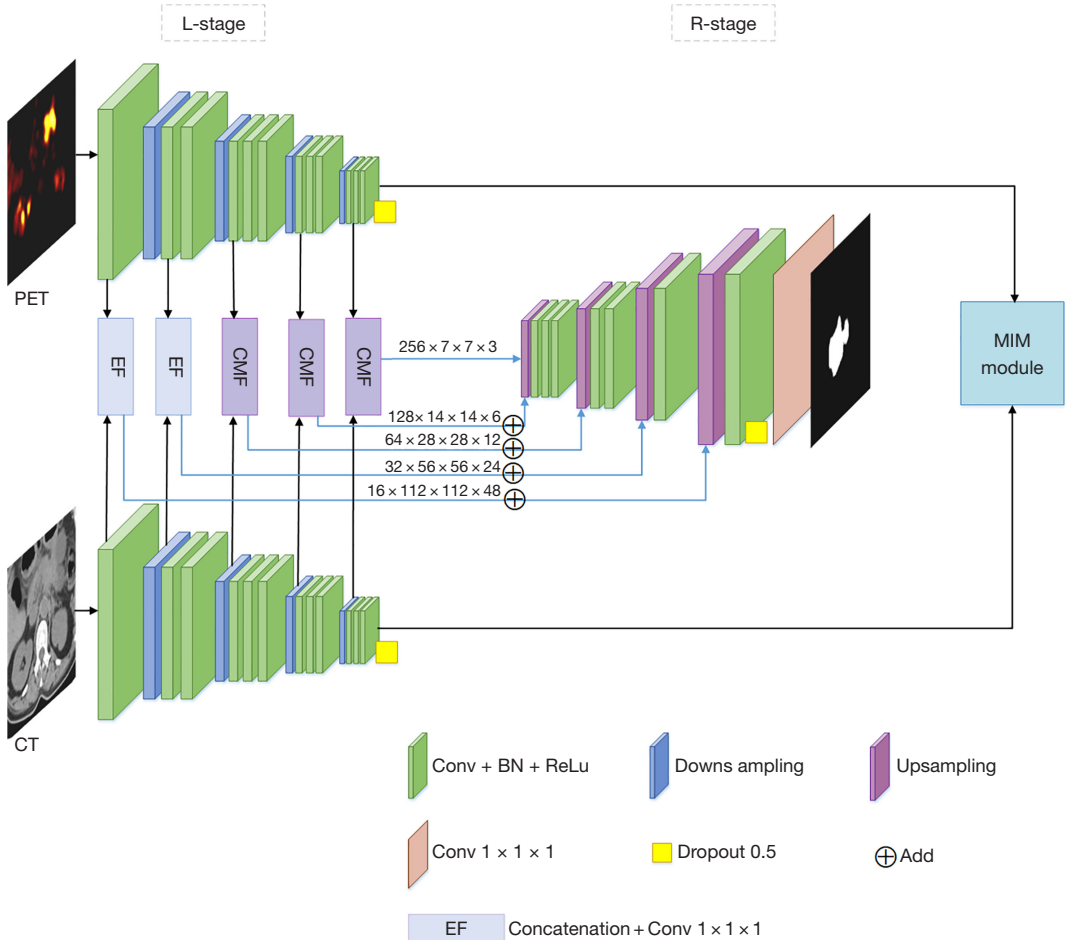

Figure 2 MIM-CMFNet structure. PET, positron emission tomography; CT, computed tomography; EF, EasyFusion; CMF, cross-modal fusion; MIM, mutual information minimization; Conv, convolution; BN, Batch Normalization; ReLu, rectified linear unit.

uncertainty estimation was unnecessary during testing.

The decoder consisted of four decoding blocks, each equipped with a transpose convolution layer and several convolutional layers. The decoder progressively upsampled the fused feature maps in four levels and used a 1×1×1 convolution kernel to obtain the final segmentation probability map.

EasyFusion (EF) was used for PET and CT fusion in layers 1 and 2 of the L-stage, while the CMF module was used for PET and CT fusion between layers 3 and 5 of the L-stage.

In layer 5 of the L-stage, high-level features from CT and PET were preserved and fed into the MIM module to compute the latent loss L_{latent}.

In summary, we developed a dual-parallel encoder feature fusion network for the student model that incorporated the CMF and MIM modules to facilitate the interaction of multimodal information. Finally, the decoder was used to recover the fused results, producing the final segmentation image.

CMF module

Inspired by the intrinsic correlation between PET and CT data and by the nonlocal block that encodes space-time relation (52), we adopted a similar module to encode pixel-level PET/CT information. We addressed the issue of the two modalities not mutually benefiting from each other by introducing the CMF module to facilitate information interaction between the PET and CT modalities, aiming to better incorporate CT information for guiding PET segmentation. PET was considered the primary modality

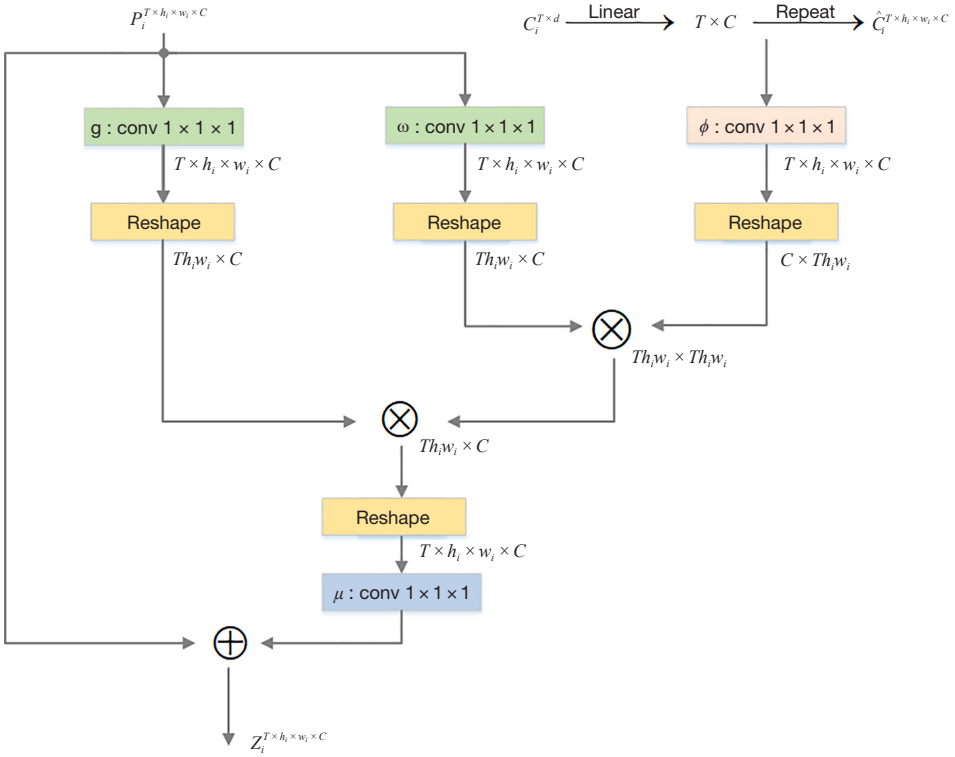

Figure 3 This is the CMF module in *Figure 2* for feature map fusion of PET and CT, which focuses on pixel-level modal interaction, takes the PET and CT feature output from layers 3–5 of the main network as input, denoted as P_i and C_i, respectively. CMF, cross-modal fusion; conv, convolution; CT, computed tomography; PET, positron emission tomography.

(*Figure 3*) due to its sensitivity to tissue metabolism. Variations in metabolic activity can be associated with abnormal tissues or lesions in various diseases, making PET images highly valuable for localizing pancreatic lesions and segmenting tumors. Combining high-resolution CT images can achieve more accurate image segmentation and localization. Building upon this foundation, the proposed CMF module allowed for the interactive fusion of PET and CT features, leveraging CT features to guide PET segmentation instead of simply adding or concatenating operations between the two modalities.

During stages 3–5 (*Figure 2*), the current PET feature map P_i and CT feature map C_i were fed into the CMF module to identify pixels in P_i that strongly responded to C_i across the entire PET. This modal interaction was measured as the dot product, and the updated feature map Z_i at the i-th stage was calculated as follows:

$$Z_i = P_i + \mu\left(\alpha_i g\left(P_i\right)\right), \text{ where } \alpha_i = \frac{\omega\left(P_i\right)\phi\left(\hat{C}_i\right)^T}{N} \quad [2]$$

where ω, ϕ, g and μ are 1×1×1 convolutions; $N = T\times h_i\times w_i$ is a normalization factor; and α_i is the modal similarity and $Z_i \in R^{T\times h_i\times w_i\times C}$. Each PET pixel interacted pixel-wise with all CT pixels through the CMF module.

MIM module

After obtaining the CT feature embeddings (CT_{feat}) and PET feature embeddings (PET_{feat}), we introduced an MIM module (*Figure 4*) to explicitly mitigate the redundancy between these two modalities. MIM (53,54) is widely used in representation learning to produce representations similar to the input; thus, MIM was used as a regularizer in our study to reduce feature redundancy for effective multimodal feature screening. We assumed that good PET and CT saliency features should contain common parts (semantic relevance) and distinct attributes (domain relevance). For the multimodal learning task, a well-trained mode was required that maximized the joint entropy of different modalities within the network's capacity range, equivalent

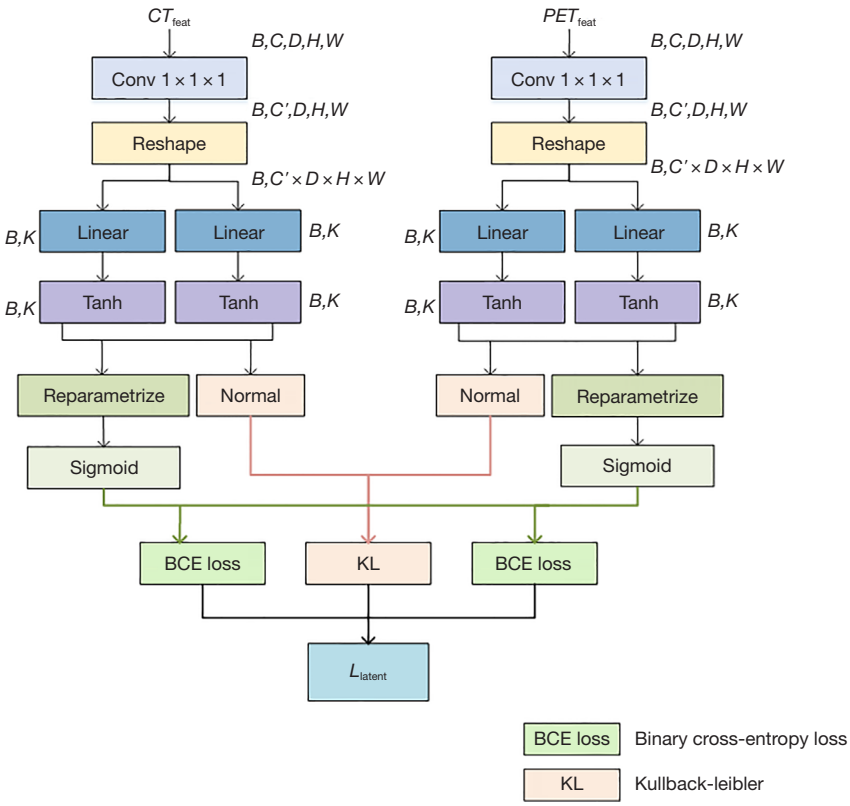

Figure 4 The MIM module accepts the CT_{feat} and PET_{feat} from the last layer of the V-Net network as inputs and then obtains latent loss after a series of calculations. CT_{feat}, CT feature embeddings; CT, computed tomography; PET_{feat}, PET feature embeddings; PET, positron emission tomography; L_{latent}, latent loss; MIM, mutual information minimization.

to minimizing mutual information and thus preventing redundant information in the network. This approach used a pixel-level interaction module to inject PET semantic information as guidance for CT segmentation to perform bimodal mapping interaction. In addition, the mutual information regularization term extracted salient features of each modality, explicitly modeling the redundancy between PET and CT features to make them distinct and effectively fusing PET and CT features under the constraint of MIM.

The MIM module received CT_{feat} and PET_{feat} output from the last layer of V-Net as input for complementary learning. Each modality's feature map was projected onto a lower-dimensional feature vector, and MIM was used as a regularizer to mitigate the redundancy between PET and CT features.

The specific steps for MIM were as follows:

(I) The input channels were reduced to hidden channels C' = 64 through a 1×1×1 convolution operation. After reshaping, the feature maps were separately mapped to two low-dimensional feature vectors with a size of K =24 through two distinct fully connected layers. A tanh activation function was then applied to obtain each sample's mean and variance in the latent space.

(II) An independent normal distribution was created based on the variance and mean, representing the feature distributions for CT and PET. The Kullback-Leibler (KL) divergence was computed between the CT and PET feature distributions, and the mean of the KL divergence was obtained. The KL divergence measured the discrepancy between two probability distributions. In some instances (55), the KL loss term was used as a measure of distribution similarity, whereas in our case, it was used to measure the similarity of modalities in multimodal learning.

(III) The reparameterization technique was employed to obtain the sampling results of the latent variables (latent features) based on the mean and variance.

The values were restricted to the [0,1] range using a sigmoid operation, and the BCE loss was computed for both values.

(IV) The final L_{latent} comprised two BCE losses and the KL divergence. Mutual information was used to measure the difference between entropy terms (correlation) as follows:

$$M_I(c,p) = H(c) + H(p) - H(c,p) \quad [3]$$

where $H(:)$ denotes the entropy function; $H(c)$ and $H(p)$ are the marginal entropies of c and p, respectively; and $H(c,p)$ is the joint entropy of c and p. The KL divergence of two latent variables (conditional entropy) was defined as follows:

$$KL(c \| p) = H_p(c) - H(c) \quad [4]$$

$$KL(p \| c) = H_c(p) - H(p) \quad [5]$$

where $H_p(c) = -\sum_x c(x) \log p(x)$ is the cross-entropy. Subsequently, summing Eqs. [3-5] results in the following:

$$M_I(c,p) = H_p(c) + H_c(p) - H(c,p) - \left(KL(c \| p) + KL(p \| c)\right) \quad [6]$$

Given a PET image and a CT image, $H(c,p)$ is nonnegative (the nonnegativity of entropy). The mutual information can be minimized by minimizing $M_I(c,p) = H_p(c) + H_c(p) - \left(KL(c \| p) + KL(p \| c)\right)$. Subsequently, when observing p, M_I measures an uncertainty reduction of c, and vice versa. As a multimodal task, each modality required learning new attributes from other modality's learning tasks. By minimizing M_I, we could explore the complementary attributes between the two modalities.

Datasets and preprocessing

We evaluated the proposed method using an internal dataset. It was obtained from the Department of Radiology of Shanghai Changhai Hospital. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013) and was reviewed and approved by the Ethics Committee of The First Affiliated Hospital (Changhai Hospital) of the Naval Medical University. Informed consent was waived due to the retrospective nature of the study. The private dataset comprised PET/CT data of 93 patients with pancreatic cancer (75 in the training set and 18 in the testing set). In order to minimize individual variations, the region of interest (ROI) for pancreatic lesions was manually delineated by an abdominal radiologist with over 5 years of experience in diagnosing pancreatic diseases using 3D Slicer software. This delineation was then validated and confirmed by another radiologist with more than 10 years of experience in pancreatic disease diagnosis. Importantly, both radiologists were kept blind to the patients' clinical outcomes (56,57). The pancreatic cancer datasets in this study were all confirmed by histopathological or cytological examination. Preprocessing steps involved strict registration between 3D CT and FDG-PET images to account for respiratory motion (58,59). We then obtained Hounsfield units (HUs) for CT images and standardized uptake values (SUVs) for FDG-PET images through numerical conversion (60). Subsequently, all images were resampled to an isotropic resolution of 1×1×1 mm using trilinear interpolation. Finally, we applied global normalization to the CT images, performing CT intensity value scaling based on the 0.5^{th} and 99.5^{th} percentiles of the foreground voxels from the training data. We normalized all CT images using the global foreground mean and standard deviation, while each PET image was independently normalized using the z-score method. All images were cropped to a fixed size of 144×144×48. Additionally, in order to demonstrate the universality of the study and further prove the validity of the module, we also used a public dataset as support. The public dataset was obtained from the HECKTOR (Head and Neck Tumor Segmentation and outcome prediction) challenge dataset at the 2021 Medical Image Computing and Computer Assisted Intervention Conference (61,62), comprising PET/CT data of 224 (180 in the training set and 44 in the testing set) patients with head and neck cancer from 5 medical centers. The data preprocessing steps were the same as mentioned above, with the images being cropped to a fixed size of 144×144×144.

Implementation

The framework was implemented in PyTorch, and training executed on a 11 GB NVIDIA GTX 2080Ti GPU. We used the stochastic gradient descent (SGD) optimizer to update the network parameters with a weight decay of 0.0001 and a momentum of 0.9. We ensured repeatability by fixing random seeds for all experiments. The initial learning rate was set to 0.01, and the network was trained for 30,000 iterations with progressive aggregation. The batch size was set to four, comprising two labeled and two unlabeled images. For the pancreatic cancer and

head and neck cancer datasets, patches with the sizes of 112×112×48 and 112×112×80 were randomly cropped as the network input, respectively. A sliding window strategy was used to acquire the final segmentation results. We used standard real-time data augmentation techniques to avoid overfitting (63), including random flipping and rotations of 90°, 180°, and 270° along the axial plane.

Evaluation criteria

We quantitatively evaluated our proposed method using the Dice coefficient, Jaccard coefficient, and 95% Hausdorff distance (95HD). The Dice coefficient indicates the overlap between model predictions and ground truth labels. The Jaccard coefficient describes the similarity between sets. 95HD refers to the maximum value of all the distances from the closest point in automatically segmented results to manually segmented results. A higher Dice score, a higher Jaccard index, and a lower 95HD indicate better segmentation results. The equations are described in Eqs. [7-9], where X represents the predicted result output by the network, and Y represents the true label.

$$\mathrm{Dice}(\mathrm{X},\mathrm{Y}) = \frac{2|\mathrm{X}\cap\mathrm{Y}|}{|\mathrm{X}|+|\mathrm{Y}|} \quad [7]$$

$$\mathrm{Jaccard}(\mathrm{X},\mathrm{Y}) = \frac{|\mathrm{X}\cap\mathrm{Y}|}{|\mathrm{X}\cup\mathrm{Y}|} \quad [8]$$

$$\mathrm{HD}(\mathrm{X},\mathrm{Y}) = \max\left(\max_{x\in\mathrm{X}}\left(\min_{y\in\mathrm{Y}}\|x-y\|_2\right), \max_{y\in Y}\left(\min_{x\in X}\|x-y\|_2\right)\right) \quad [9]$$

Results

Segmentation performance comparison

During training, we used only 20% of the labeled data, following the common practice in most semisupervised tasks (35,37,39). The baseline model was V-Net with two encoder and decoder branches. We conducted a three-fold cross-validation to evaluate the performance of our method on the head and neck cancer and pancreatic cancer datasets, which was compared with other semisupervised and fully supervised methods.

The segmentation results (*Table 2*) showed that the single-modal CT images performed the worst on both datasets when using 20% of the labeled data, with the Dice and Jaccard metrics at 34.03% and 22.33% for the pancreatic cancer dataset and 40.64% and 29.99% for the head and neck cancer dataset, respectively. The low contrast of the CT images challenged the segmentation algorithm to distinguish different tissues or structures accurately. In addition, substantial noise and artifacts further contributed to blurred boundaries or erroneous segmentation results. In contrast, segmentation accuracy significantly improved when PET images were used, with the Dice and Jaccard metrics increased by 34.42% and 32.14% on the pancreatic cancer dataset and 22.49% and 21.58% on the head and neck cancer dataset, respectively. However, tumor heterogeneity and partial volume effects in PET images still yielded several inaccuracies. Simultaneous segmentation using PET and CT yielded better results in our baseline, with Dice scores of 70.52% and 66.61% for the pancreatic cancer and head and neck cancer datasets, demonstrating that multimodal PET/CT segmentation outperformed single-modal segmentation.

Our proposed MIM-CMFNet achieved promising overall segmentation performance on both datasets. With only 20% of the labeled training data, the average Dice score on the pancreas dataset reached 73.14%, representing a 2.62% improvement over the baseline. The 95HD was reduced from 10.19 to 6.30 mm, indicating better segmentation boundaries. On the head and neck cancer dataset, the average Dice score reached 68.71%, showing a 2.1% improvement over the baseline. The 95HD was reduced from 13.40 to 7.88 mm, further demonstrating improved segmentation accuracy. Our semisupervised framework achieved results close to the state-of-the-art fully supervised frameworks while reducing the annotation cost by 80%, demonstrating its practicability.

Furthermore, we compared our method with several advanced semisupervised segmentation methods, including DTC (35), MC-Net (37), and CML (44), which are based on consistency-driven semisupervised segmentation networks. DTC and MC-Net were originally designed for single-modal cardiac image segmentation, and we adapted these two models for multimodal images. CML is specifically designed for PET/CT head and neck cancer multimodal image segmentation. We ensured fair comparisons using the same network backbone (Bayesian V-Net) in all methods. The results showed (*Table 2*) that our method outperformed these approaches in all evaluated metrics. *Figures 5,6* show the final segmentation effects of different methods on the pancreatic cancer dataset and the head and neck cancer dataset, respectively.

In addition, we observed that tumor segmentation on the head and neck cancer dataset was significantly more

Table 2 Segmentation performance comparison of state-of-the-art methods on the pancreatic cancer and H&N cancer datasets with PET/CT (UA-MT) as the baseline of the experiment

Dataset	Methods	L	U	Dice (%)	Jaccard (%)	95HD (mm)
Pancreas	CT (UA-MT)	15 (20%)	60	34.03±3.64	22.33±2.38	15.98±2.74
	PET (UA-MT)	15 (20%)	60	68.45±1.88	54.47±2.05	9.07±2.60
	PET/CT (UA-MT)	15 (20%)	60	70.52±1.45	56.72±2.11	10.19±1.29
	DTC (37)	15 (20%)	60	68.43±1.96	55.50±1.98	7.98±2.51
	MC-Net (39)	15 (20%)	60	71.42±1.15	57.86±2.13	6.61±1.63
	CML (46)	15 (20%)	60	71.86±2.76	58.15±2.31	7.37±2.31
	Xue *et al.* (15)	75 (100%)	0	69.31±2.32	55.05±2.63	23.56±7.13
	Zhong *et al.* (13)	75 (100%)	0	72.46±0.25	59.58±0.80	7.63±2.24
	Wang *et al.* (4)	75 (100%)	0	76.20±0.53	63.08±0.70	6.84±3.27
	Proposed	15 (20%)	60	73.14±2.71	60.56±2.35	6.30±2.49
H&N	CT (UA-MT)	36 (20%)	144	40.64±2.62	29.99±2.38	17.22±3.32
	PET (UA-MT)	36 (20%)	144	63.13±2.00	51.57±2.05	15.47±6.31
	PET/CT (UA-MT)	36 (20%)	144	66.61±1.97	56.37±2.47	13.40±2.85
	DTC (37)	36 (20%)	144	66.92±2.90	56.74±3.09	9.92±1.56
	MC-Net (39)	36 (20%)	144	67.62±0.37	57.13±0.61	9.08±1.05
	CML (46)	36 (20%)	144	66.15±1.91	56.47±0.89	10.74±1.33
	Xue *et al.* (15)	180 (100%)	0	63.88±5.38	51.54±5.27	24.55±7.45
	Zhong *et al.* (13)	180 (100%)	0	71.10±5.71	59.71±4.98	7.86±1.80
	Wang *et al.* (4)	180 (100%)	0	74.14±2.77	62.96±2.24	6.41±1.01
	Proposed	36 (20%)	144	68.71±1.16	57.72±1.38	7.88±1.77

The Dice score, Jaccard index, and 95HD are described as the mean ± standard deviation. 95HD, 95% Hausdorff distance; L, number of labeled samples (numbers in parentheses represent the proportion in the training set); U, number of unlabeled samples; CT, computed tomography; UA-MT, uncertainty-aware mean teacher; PET, positron emission tomography; DTC, dual-task consistency; MC-Net, mutual consistency network; CML, contrastive mutual learning; H&N, head and neck.

challenging compared to the pancreatic cancer dataset due to the significant variability in the shape, size, and location of head and neck cancer tumors, as they can occur in various locations within the head and neck region (4). Moreover, the head and neck cancer dataset came from five different medical centers, leading to differences in collection and quality, along with the presence of lymph nodes with high metabolic responses in PET images (56). These factors increased the difficulty of accurately segmenting the head and neck tumors.

Ablation study

We conducted ablation experiments to evaluate the effectiveness of the CMF and MIM modules (*Table 3*).

Compared with the baseline, adding the MIM module improved the Dice and Jaccard metrics by 0.65% and 0.50%, respectively, and reduced the 95HD by 4.64 mm on the pancreatic dataset. On the head and neck cancer dataset, the Dice and Jaccard metrics increased by 1.35% and 1.17%, respectively, while the 95HD decreased by 4.25 mm. The model with the CMF module increased the Dice and Jaccard metrics by 0.80% and 0.72%, respectively, and decreased the 95HD by 0.88 mm on the pancreatic cancer dataset. On the head and neck cancer dataset, the Dice and Jaccard metrics increased by 0.89% and 0.58%, respectively, while the 95HD decreased by 1.63 mm. Finally, the combination of both modules enhanced the

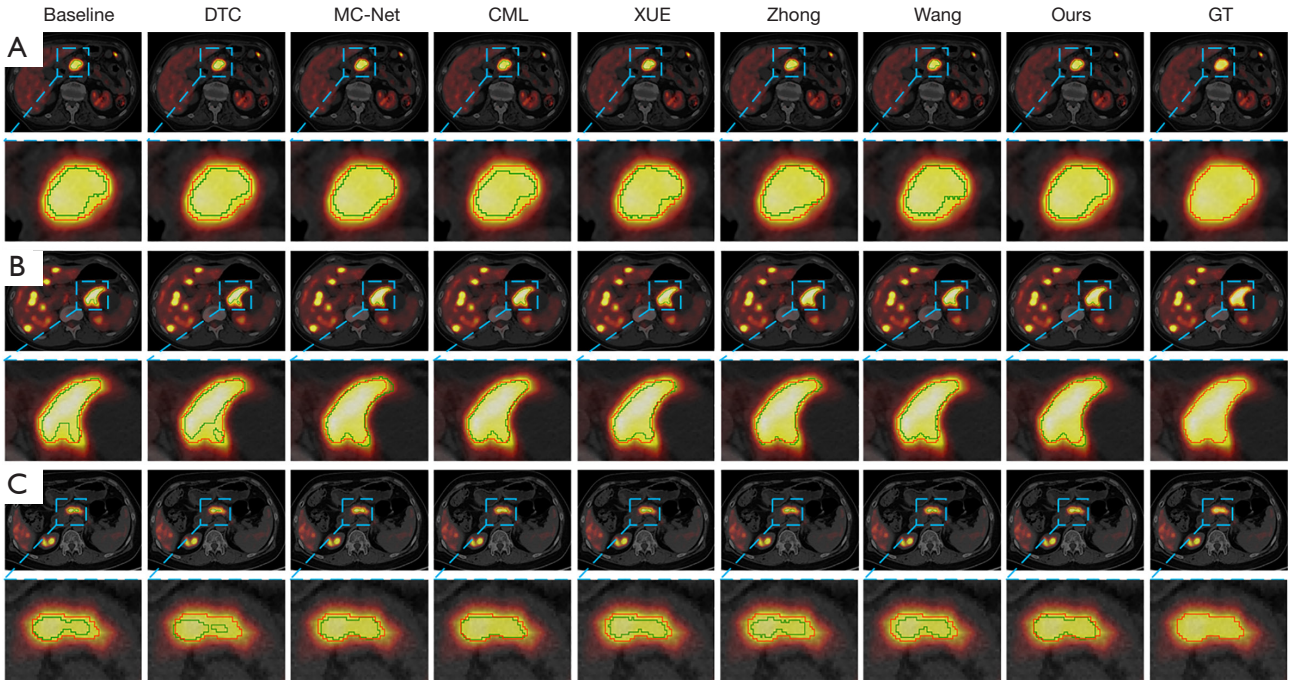

Figure 5 Visualization results of different segmentation methods for the pancreatic cancer dataset. (A-C) Each example shows the location of the lesion and the fine segmentation effect, and the blue box is used to indicate the fine segmentation effect. Green and red indicate the prediction and gold standard, respectively. From left to right: baseline, DTC (37), MC-Net (39), CML (46), Xue *et al.* (15), Zhong *et al.* (13), Wang *et al.* (4), our proposed method, and the ground truth. DTC, dual-task consistency; MC, mutual consistency; CML, contrastive mutual learning; GT, ground truth.

segmentation accuracy. On the pancreatic cancer dataset, the Dice and Jaccard metrics of both modules increased by 2.62% and 3.84%, respectively, while the 95HD decreased by 3.89 mm. On the head and neck cancer dataset, the Dice and Jaccard metrics of both modules increased by 2.1% and 1.35%, respectively, while the 95HD decreased by 5.52 mm. Moreover, the proposed model demonstrated excellent segmentation performance and generalization ability across both datasets.

Analysis of the unlabeled data

We analyzed the importance of labeled and unlabeled data by experimenting with the baseline model to evaluate the performance improvement of our semisupervised approach with the addition of unlabeled data. The results (*Table 4*) showed that using only 20% of the data (15 labeled samples) for fully supervised pancreatic cancer segmentation resulted in an accuracy of 68.82%. Incorporating 60 unlabeled cases enhanced the segmentation accuracy to 70.52%, verifying that including unlabeled data improved segmentation performance. Similarly, the segmentation accuracy for the head and neck cancer dataset also improved with the addition of unlabeled data.

Discussion

Pancreatic tumor segmentation is crucial in the diagnosis and radiotherapy of pancreatic cancer. Semisupervised learning addresses the challenge of limited labeled data in medical image segmentation tasks. However, most existing semisupervised studies focus on single-modal data and cannot leverage complementary information in multimodal medical images (46). In this study, we thus designed a multimodal deep learning approach that fused PET/CT images with limited labeled data, leading to more accurate segmentation results.

Module analysis

Pancreatic tumor segmentation involves inherent challenges. The pancreas is a complex organ composed of

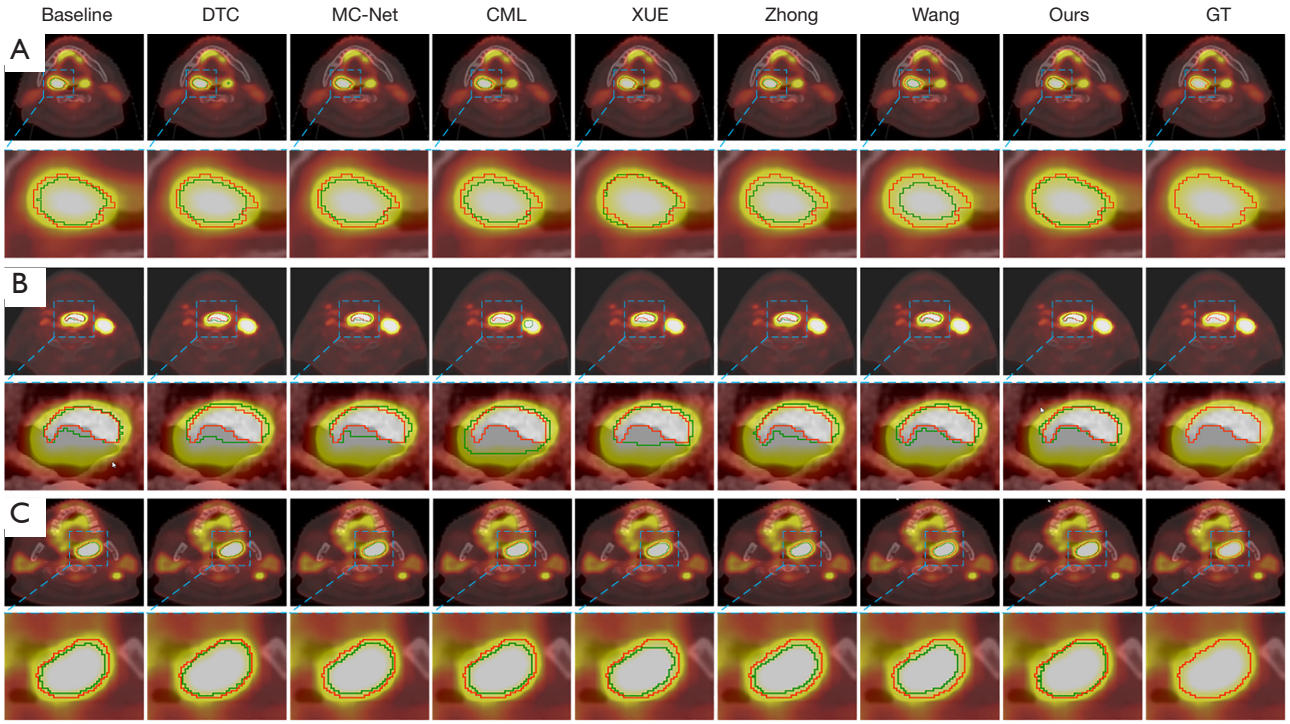

Figure 6 Visualization results of the different segmentation methods for the head and neck cancer dataset. Each example shows the location of the lesion and the fine segmentation effect, and the blue box is used to indicate the fine segmentation effect. Green and red indicate the prediction and gold standard, respectively. DTC, dual-task consistency; MC, mutual consistency; CML, contrastive mutual learning; GT, ground truth.

Table 3 Ablation experiments on pancreatic cancer and head and neck cancer datasets

Dataset	Method	Dice (%)	Jaccard (%)	95HD (mm)
Pancreatic cancer	Baseline	70.52±1.45	56.72±2.11	10.19±1.29
	Baseline +MIM	71.17±1.48	57.22±1.81	5.55±0.48
	Baseline + CMF	71.32±1.57	57.44±1.67	9.31±1.50
	Ours	73.14±2.71	60.56±2.35	6.30±2.49
Head and neck cancer	Baseline	66.61±1.97	56.37±2.47	13.40±2.85
	Baseline + MIM	67.96±0.77	57.54±0.72	9.15±1.93
	Baseline + CMF	67.50±2.23	56.95±1.66	11.77±1.80
	Ours	68.71±1.16	57.72±1.38	7.88±1.77

All results are described as the mean ± standard deviation. 95HD, 95% Hausdorff distance; MIM, mutual information minimization; CMF, cross-modal fusion.

three parts—the head, body, and tail—with a deep location and complex adjacent relationships, being surrounded by many important organs, blood vessels, lymphatic tissues, and other structures. In addition, pancreatic tumors often have similar densities to these surrounding tissues, leading to low contrast problems. Our proposed model outperformed the model of Xue *et al.* (15) (*Table 2*), demonstrating significant improvement in segmentation accuracy and tumor boundary delineation on both pancreatic cancer and head and neck cancer datasets. The

Table 4 Comparison of results of the models with and without unlabeled data

Dataset	L	U	Dice (%)	Jaccard (%)	95HD (mm)
Pancreatic cancer	15	0	68.82±0.82	55.54±1.07	13.95±3.65
	15	60	70.52±1.45	56.72±2.11	10.19±1.29
Head and neck cancer	36	0	64.46±1.55	54.36±1.75	14.21±3.18
	36	144	66.61±1.97	56.37±2.47	13.40±2.85

All results are described as the mean ± standard deviation. 95HD, 95% Hausdorff distance; L, number of labeled samples; U, number of unlabeled samples.

model proposed by Xue *et al.* (15) achieved reasonable results in liver tumor segmentation based on liver masks; however, its segmentation accuracy dropped significantly with more complex and variable tumors, particularly in pancreatic cancer and head and neck cancer. In contrast, our proposed MIM and CMF strategy demonstrated good segmentation performance on both datasets and effectively delineated tumor boundaries. The Dice and Jaccard segmentation metrics (*Table 3*) were improved for pancreatic cancer and head and neck cancers, while the 95HD significantly decreased from 15.19 to 6.30 mm for pancreatic cancer, indicating a substantial enhancement in tumor boundary delineation. This improvement can be attributed to two key factors. First, mutual information measures the degree of correlation and dependence between two variables. Our MIM module enhanced feature selection and refinement processes, eliminating irrelevant and redundant features and learning latent high-level features from PET and CT images, thus providing complementary biological metabolism and anatomical structural information, respectively. The extraction of latent features helped obtain essential information and patterns from the original data. By fusing their high-level features, our approach gained a better understanding of pancreatic tumors, particularly in pancreatic tumor edge segmentation, for which it distinguished regions with similar density to surrounding tissues in CT images, improving segmentation accuracy. Second, the CMF module guided the model to focus on cross-dimensional and cross-channel information, adaptively adjusting the weights of various features. The high uptake characteristics of PET allowed for rapid identification of pancreatic tumor regions, compensating for the low contrast issue in CT. Incorporating CT information provided more accurate anatomical details, mitigating the impact of PET partial volume effects. In addition, the CMF interaction at different levels enhanced the model's local and global perception, improving its ability to perceive features at various scales, thus demonstrating high discrimination for pancreatic tumor targets. The quality of feature maps (*Figure 7*) indicated that our approach produced feature maps with clearer boundary effects compared to the baseline model.

Importance of unlabeled data for semisupervised learning

The manual annotation of multimodal data by radiologists for segmentation tasks is laborious and resource-intensive. Unlike fully supervised segmentation approaches that rely heavily on considerable labeled data for training, our semisupervised multimodal segmentation method achieves near-fully supervised performance with only 20% labeled data. First, incorporating a large amount of unlabeled data improved the model's generalization and performance by expanding the training dataset and reducing overfitting. Second, the model could learn more diverse and rich feature representations. Through semisupervised learning on unlabeled data, the model explored the distribution and hidden structural features in a broader sample space. These learned feature representations were transferred to the labeled data to improve the model performance on the labeled data. Third, introducing unlabeled data acted as a regularization technique. By imposing the consistency constraint of unlabeled data, the model generated more consistent predictions, enhancing its robustness.

Limitations and future work

Although our method achieved good segmentation results on pancreatic cancer and head and neck datasets with limited labeled data, it still faced challenges in accurately segmenting extreme cases due to the lack of pixel-level annotated data. Therefore, future work may explore interactive techniques to assist in segmentation and to enhance segmentation performance. Moreover, our

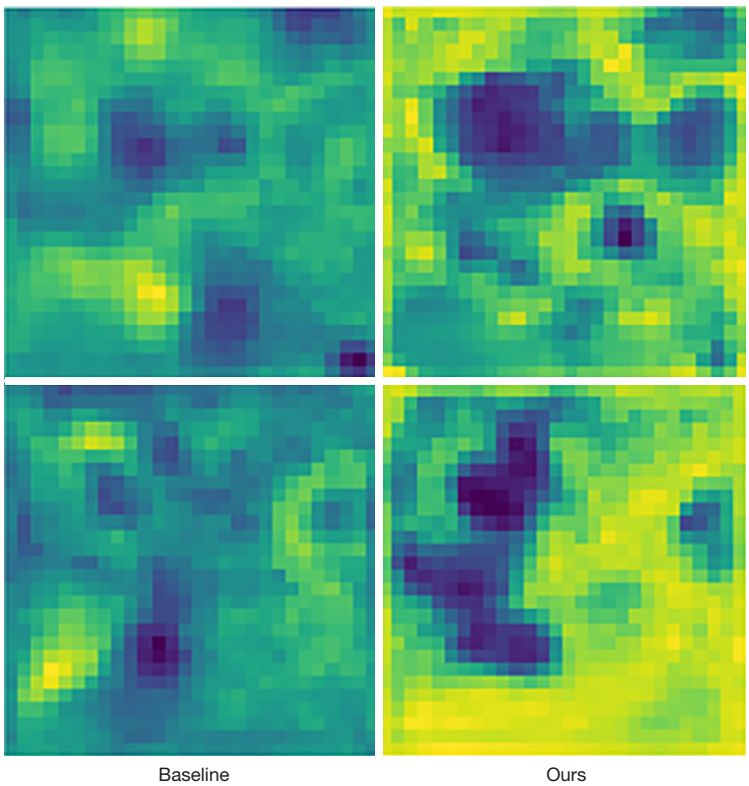

Figure 7 Comparison of the feature maps of baseline and proposed methods. Our approach produces feature maps with clearer edges, while the baseline exhibits stronger edge adhesion, indicating that our method better distinguishes the tumor from the surrounding tissue structures, resulting in improved segmentation accuracy.

network was based on strictly paired multimodal data, but in clinical practice, it is often challenging to acquire a large amount of paired PET/CT data. Therefore, in the future, it would be worth exploring the use of the recent and popular generative adversarial networks to generate synthetic PET images. This approach can help reduce the reliance on multimodal PET/CT image data and enhance the versatility of our algorithm in clinical applications.

Conclusions

This study proposed MIM-CMFNet for semisupervised medical image segmentation. Our method minimized mutual information, using mutual information regularization to extract salient features from each modality, reducing redundancy between the PET and CT modalities, and utilizing knowledge from the unlabeled data. In addition, our proposed CMF module facilitated information interaction between the two modalities, enabling better integration of CT information to guide PET segmentation. Experimental results on the pancreatic cancer and head and neck cancer datasets demonstrated that MIM-CMFNet outperformed existing semisupervised segmentation methods and achieved comparable performance to most fully supervised multimodal methods, thus exhibiting favorable generalization ability and potential for clinical application.

Acknowledgments

We would like to thank LetPub (www.letpub.com) for its linguistic assistance during the preparation of this manuscript.
Funding: This study was supported by the National Natural Science Foundation of China (Nos. 62101551 and 82172712), the "234 Discipline Climbing Plan" of the First Affiliated Hospital of Naval Medical University (Nos. 2019YPT002 and 2019YXK033), and the Special Foundation for Emerging Interdisciplinary Field Research of Shanghai Municipal Health Commission (No.

2022JC004).

Footnote

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://qims.amegroups.com/article/view/10.21037/qims-23-1153/coif). C.H. is the co-founder and Chief Technology Officer of Shenzhen Poros Technology Co., Ltd. Some of the work in the paper was completed during his tenure at the company. The other authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013) and was approved by the Ethics Committee of The First Affiliated Hospital (Changhai Hospital) of the Naval Medical University. Informed consent was waived due to the retrospective nature of the study.

References

1. Kamisawa T, Wood LD, Itoi T, Takaori K. Pancreatic cancer. Lancet 2016;388:73-85.
2. Mizrahi JD, Surana R, Valle JW, Shroff RT. Pancreatic cancer. Lancet 2020;395:2008-20.
3. Blackford AL, Canto MI, Klein AP, Hruban RH, Goggins M. Recent Trends in the Incidence and Survival of Stage 1A Pancreatic Cancer: A Surveillance, Epidemiology, and End Results Analysis. J Natl Cancer Inst 2020;112:1162-9.
4. Wang F, Cheng C, Cao W, Wu Z, Wang H, Wei W, Yan Z, Liu Z. MFCNet: A multi-modal fusion and calibration networks for 3D pancreas tumor segmentation on PET-CT images. Comput Biol Med 2023;155:106657.
5. Ahn PH, Garg MK. Positron emission tomography/computed tomography for target delineation in head and neck cancers. Semin Nucl Med 2008;38:141-8.
6. Zhang Y, Cheng C, Liu Z, Wang L, Pan G, Sun G, Chang Y, Zuo C, Yang X. Radiomics analysis for the differentiation of autoimmune pancreatitis and pancreatic ductal adenocarcinoma in (18) F-FDG PET/CT. Med Phys 2019;46:4520-30.
7. Sahani DV, Bonaffini PA, Catalano OA, Guimaraes AR, Blake MA. State-of-the-art PET/CT of the pancreas: current role and emerging indications. Radiographics 2012;32:1133-58; discussion 1158-60.
8. Liu Z, Li M, Zuo C, Yang Z, Yang X, Ren S, Peng Y, Sun G, Shen J, Cheng C, Yang X. Radiomics model of dual-time 2-[18F]FDG PET/CT imaging to distinguish between pancreatic ductal adenocarcinoma and autoimmune pancreatitis. Eur Radiol 2021;31:6983-91.
9. Cui Y, Song J, Pollom E, Alagappan M, Shirato H, Chang DT, Koong AC, Li R. Quantitative Analysis of (18) F-Fluorodeoxyglucose Positron Emission Tomography Identifies Novel Prognostic Imaging Biomarkers in Locally Advanced Pancreatic Cancer Patients Treated With Stereotactic Body Radiation Therapy. Int J Radiat Oncol Biol Phys 2016;96:102-9.
10. Yu K, Chen X, Shi F, Zhu W, Zhang B, Xiang D. A novel 3D graph cut based co-segmentation of lung tumor on PET-CT images with Gaussian mixture models. SPIE Medical Imaging, 2016. San Diego, California, USA: SPIE, 2016:787-93.
11. Andrearczyk V, Oreiller V, Vallières M, Castelli J, Elhalawani H, Jreige M, Boughdad S, Prior JO, Depeursinge A. Automatic Segmentation of Head and Neck Tumors and Nodal Metastases in PET-CT scans. Proc Mach Learn Res 2020;121:33-43.
12. Hu M, Maillard M, Zhang Y, Ciceri T, La Barbera G, Bloch I, Gori P. Knowledge Distillation from Multi-modal to Mono-modal Segmentation Networks. Medical Image Computing and Computer Assisted Intervention – MICCAI 2020. Cham: Springer, 2020:772-81.
13. Zhong Z, Kim Y, Plichta K, Allen BG, Zhou L, Buatti J, Wu X. Simultaneous cosegmentation of tumors in PET-CT images using deep fully convolutional networks. Med Phys 2019;46:619-33.
14. Kumar A, Fulham M, Feng D, Kim J. Co-Learning Feature Fusion Maps from PET-CT Images of Lung Cancer. IEEE Trans Med Imaging 2019. [Epub ahead of print]. doi: 10.1109/TMI.2019.2923601.
15. Xue Z, Li P, Zhang L, Lu X, Zhu G, Shen P, Ali Shah SA, Bennamoun M. Multimodal Co-Learning for Liver Lesion Segmentation on PET-CT Images. IEEE Trans Med

Imaging 2021;40:3531-42.
16. Fu X, Bi L, Kumar A, Fulham M, Kim J. Multimodal Spatial Attention Module for Targeting Multimodal PET-CT Lung Tumor Segmentation. IEEE J Biomed Health Inform 2021;25:3507-16.
17. Mo S, Cai M, Lin L, Tong R, Chen Q, Wang F, Hu H, Lwamoto Y, Han XH, Chen YW. Multimodal Priors Guided Segmentation of Liver Lesions in MRI Using Mutual Information Based Graph Co-Attention Networks. Medical Image Computing and Computer Assisted Intervention – MICCAI 2020. Cham: Springer, 2020:429-38.
18. Zhao J, Li D, Xiao X, Accorsi F, Marshall H, Cossetto T, Kim D, McCarthy D, Dawson C, Knezevic S, Chen B, Li S. United adversarial learning for liver tumor segmentation and detection of multimodality non-contrast MRI. Med Image Anal 2021;73:102154.
19. Cao W, Zheng J, Xiang D, Ding S, Sun H, Yang X, Liu Z, Dai Y. Edge and neighborhood guidance network for 2D medical image segmentation. Biomed Signal Process Control 2021;69:102856.
20. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Cham: Springer, 2015:234-41.
21. Lei T, Wang R, Zhang Y, Wan Y, Liu C, Nandi AK. DefED-Net: Deformable Encoder-Decoder Network for Liver and Liver Tumor Segmentation. IEEE Trans Radiat Plasma Med Sci 2021;6:68-78.
22. Wang J, Wei L, Wang L, Zhou Q, Zhu L, Qin J. Boundary-aware Transformers for Skin Lesion Segmentation. Medical Image Computing and Computer Assisted Intervention – MICCAI 2021. Cham: Springer, 2021:206-16.
23. Milletari F, Navab N, Ahmadi SA. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. 2016 Fourth International Conference on 3D vision (3DV). 2016:565-71.
24. Zhu Y, Zhang Z, Wu C, Zhang Z, He T, Zhang H, Manmatha R, Li M, Smola AJ. Improving Semantic Segmentation via Efficient Self-Training. IEEE Trans Pattern Anal Mach Intell 2021. [Epub ahead of print]. doi: 10.1109/TPAMI.2021.3138337.
25. Zou Y, Yu Z, Kumar BVK, Wang J. Domain Adaptation for Semantic Segmentation via Class-Balanced Self-Training. Proceedings of the European Conference on Computer Vision (ECCV). 2018:289-305.
26. Hung WC, Tsai YH, Liou YT, Lin YY, Yang MH. Adversarial Learning for Semi-Supervised Semantic Segmentation. ArXiv 2018. ArXiv:1802.07934.
27. Li S, Zhang C, He X. Shape-Aware Semi-supervised 3D Semantic Segmentation for Medical Images. Medical Image Computing and Computer Assisted Intervention – MICCAI 2020. Cham: Springer, 2020:552-61.
28. Xia Y, Yang D, Yu Z, Liu F, Cai J, Yu L, Zhu Z, Xu D, Yuille A, Roth H. Uncertainty-aware multi-view co-training for semi-supervised medical image segmentation and domain adaptation. Med Image Anal 2020;65:101766.
29. Wang P, Peng J, Pedersoli M, Zhou Y, Zhang C, Desrosiers C. Self-paced and self-consistent co-training for semi-supervised image segmentation. Med Image Anal 2021;73:102146.
30. Xiang J, Li Z, Wang W, Xia Q, Zhang S. Self-Ensembling Contrastive Learning for Semi-Supervised Medical Image Segmentation. ArXiv 2021. ArXiv:2105.12924.
31. Hu X, Zeng D, Xu X, Shi Y. Semi-supervised Contrastive Learning for Label-efficient Medical Image Segmentation. Medical Image Computing and Computer Assisted Intervention – MICCAI 2021. Cham: Springer, 2021:481-90.
32. Bortsova G, Dubost F, Hogeweg L, Katramados I, de Bruijne M. Semi-Supervised Medical Image Segmentation via Learning Consistency under Transformations. Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. Cham: Springer, 2019:810-8.
33. Hang W, Feng W, Liang S, Yu L, Wang Q, Choi KS, Qin J. Local and Global Structure-Aware Entropy Regularized Mean Teacher Model for 3D Left Atrium Segmentation. Medical Image Computing and Computer Assisted Intervention – MICCAI 2020. Cham: Springer, 2020:562-71.
34. Li X, Yu L, Chen H, Fu CW, Xing L, Heng PA. Transformation-Consistent Self-Ensembling Model for Semisupervised Medical Image Segmentation. IEEE Trans Neural Netw Learn Syst 2021;32:523-34.
35. Yu L, Wang S, Li X, Fu CW, Heng PA. Uncertainty-aware Self-ensembling Model for Semi-supervised 3D Left Atrium Segmentation. Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. Cham: Springer, 2019:605-13.
36. Wang Y, Zhang Y, Tian J, Zhong C, Shi Z, Zhang Y, He Z. Double-Uncertainty Weighted Method for Semi-supervised Learning. Medical Image Computing and Computer Assisted Intervention – MICCAI 2020. Cham: Springer, 2020:542-51.
37. Luo X, Chen J, Song T, Wang G. Semi-supervised

Medical Image Segmentation through Dual-task Consistency. Proceedings of the AAAI Conference on Artificial Intelligence. 2021:8801-9.
38. Laine S, Aila T. Temporal Ensembling for Semi-Supervised Learning. ArXiv 2016. ArXiv:1610.02242.
39. Wu Y, Xu M, Ge Z, Cai J, Zhang L. Semi-supervised Left Atrium Segmentation with Mutual Consistency Training. Medical Image Computing and Computer Assisted Intervention – MICCAI 2021. Cham: Springer, 2021:297-306.
40. Tarvainen A, Valpola H. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. Adv Neural Inf Process Syst 2017;30.
41. Dolz J, Desrosiers C, Ayed IB. Teach Me to Segment with Mixed Supervision: Confident Students Become Masters. Information Processing in Medical Imaging. Cham: Springer, 2021:517-29.
42. Sohn K, Berthelot D, Carlini N, Zhang Z, Zhang H, Raffel CA, Cubuk ED, Kurakin A, Li CL. FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence. NIPS'20: Proceedings of the 34th International Conference on Neural Information Processin. 2020:596-608.
43. Chen X, Yuan Y, Zeng G, Wang J. Semi-Supervised Semantic Segmentation with Cross Pseudo Supervision. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, TN, USA: IEEE, 2021:2613-22.
44. Ouali Y, Hudelot C, Tami M. Semi-Supervised Semantic Segmentation with Cross-Consistency Training. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2020:12674-84.
45. Li S, Zhao Z, Xu K, Zeng Z, Guan C. Hierarchical Consistency Regularized Mean Teacher for Semi-supervised 3D Left Atrium Segmentation. Annu Int Conf IEEE Eng Med Biol Soc 2021;2021:3395-8.
46. Zhang S, Zhang J, Tian B, Lukasiewicz T, Xu Z. Multimodal contrastive mutual learning and pseudo-label re-learning for semi-supervised medical image segmentation. Med Image Anal 2023;83:102656.
47. Mondal AK, Dolz J, Desrosiers C. Few-shot 3D Multimodal Medical Image Segmentation using Generative Adversarial Learning. ArXiv 2018. ArXiv:1810.12241.
48. Chartsias A, Papanastasiou G, Wang C, Semple S, Newby DE, Dharmakumar R, Tsaftaris SA. Disentangle, Align and Fuse for Multimodal and Semi-Supervised Image Segmentation. IEEE Trans Med Imaging 2021;40:781-92.
49. Yang X, Bian C, Yu L, Ni D, Heng PA. Hybrid Loss Guided Convolutional Networks for Whole Heart Parsing. Statistical Atlases and Computational Models of the Heart. Cham: Springer, 2018:215-23.
50. Santurkar S, Tsipras D, Ilyas A, Madry A. How Does Batch Normalization Help Optimization? Adv Neural Inf Process Syst 2018;31.
51. Glorot X, Bordes A, Bengio Y. Deep Sparse Rectifier Neural Networks. Proc Mach Learn Res 2011;15:315-23.
52. Wang X, Girshick R, Gupta A, He K. Non-local neural networks. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018:7794-803.
53. Linsker R. Self-organization in a perceptual network. Computer 1988;21:105-17.
54. Tschannen M, Djolonga J, Rubenstein PK, Gelly S, Lucic M. On Mutual Information Maximization for Representation Learning. ArXiv 2019. ArXiv:1907.13625.
55. Zhang J, Fan DP, Dai Y, Anwar S, Saleh F, Aliakbarian S, Barnes N. Uncertainty Inspired RGB-D Saliency Detection. IEEE Trans Pattern Anal Mach Intell 2022;44:5761-79.
56. Hu D, Jian J, Li Y, Gao X. Deep learning-based segmentation of epithelial ovarian cancer on T2-weighted magnetic resonance images. Quant Imaging Med Surg 2023;13:1464-77.
57. Zhang A, Khan A, Majeti S, Pham J, Nguyen C, Tran P, Iyer V, Shelat A, Chen J, Manjunath BS. Automated Segmentation and Connectivity Analysis for Normal Pressure Hydrocephalus. BME Front 2022;2022:9783128.
58. Mattes D, Haynor DR, Vesselle H, Lewellen TK, Eubank W. PET-CT image registration in the chest using free-form deformations. IEEE Trans Med Imaging 2003;22:120-8.
59. Pieper S, Halle M, Kikinis R. 3D slicer. IEEE International Symposium on Biomedical Imaging: Nano to Macro. IEEE, 2004:632-5.
60. Masa-Ah P, Soongsathitanon S. A novel standardized uptake value (SUV) calculation of PET DICOM files using MATLAB. Proceedings of the 10th WSEAS international conference on applied informatics and communications, and 3rd WSEAS international conference on Biomedical electronics and biomedical informatics. 2010: 413-6.
61. Andrearczyk V, Oreiller V, Boughdad S, Le Rest CC, Elhalawani H, Jreige M, Prior JO, Vallières M, Visvikis D, Hatt M, Depeursinge A. Overview of the HECKTOR Challenge at MICCAI 2021: Automatic Head and Neck

Tumor Segmentation and Outcome Prediction in PET/CT Images. Cham: Springer International Publishing, 2021:1-37.
62. Vallières M, Kay-Rivest E, Perrin LJ, Liem X, Furstoss C, Aerts HJWL, Khaouam N, Nguyen-Tan PF, Wang CS, Sultanem K, Seuntjens J, El Naqa I. Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer. Sci Rep 2017;7:10117.
63. Yu L, Cheng JZ, Dou Q, Yang X, Chen H, Qin J, Heng PA. Automatic 3D Cardiovascular MR Segmentation with Densely-Connected Volumetric ConvNets. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2017. Cham: Springer, 2017:287-95.

Cite this article as: Shao M, Cheng C, Hu C, Zheng J, Zhang B, Wang T, Jin G, Liu Z, Zuo C. Semisupervised 3D segmentation of pancreatic tumors in positron emission tomography/computed tomography images using a mutual information minimization and cross-fusion strategy. Quant Imaging Med Surg 2024;14(2):1747-1765. doi: 10.21037/qims-23-1153

