



OPEN ACCESS

EDITED BY
Qingli Li,
East China Normal University, China

REVIEWED BY
Meng Lv,
Beijing Institute of Technology, China
Haoyang Yu,
Dalian Maritime University, China

*CORRESPONDENCE
Jinghui Fang
✉ 2106020108@hhu.edu.cn

RECEIVED 06 March 2024
ACCEPTED 15 April 2024
PUBLISHED 19 June 2024

CITATION
Fang J (2024) A semi-supervised
segmentation method for microscopic
hyperspectral pathological images based on
multi-consistency learning.
Front. Oncol. 14:1396887.
doi: 10.3389/fonc.2024.1396887

COPYRIGHT
© 2024 Fang. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

A semi-supervised segmentation method for microscopic hyperspectral pathological images based on multi-consistency learning

Jinghui Fang*

College of Information Science and Engineering, Hohai University, Nanjing, China

Pathological images are considered the gold standard for clinical diagnosis and cancer grading. Automatic segmentation of pathological images is a fundamental and crucial step in constructing powerful computer-aided diagnostic systems. Medical microscopic hyperspectral pathological images can provide additional spectral information, further distinguishing different chemical components of biological tissues, offering new insights for accurate segmentation of pathological images. However, hyperspectral pathological images have higher resolution and larger area, and their annotation requires more time and clinical experience. The lack of precise annotations limits the progress of research in pathological image segmentation. In this paper, we propose a novel semi-supervised segmentation method for microscopic hyperspectral pathological images based on multi-consistency learning (MCL-Net), which combines consistency regularization methods with pseudo-labeling techniques. The MCL-Net architecture employs a shared encoder and multiple independent decoders. We introduce a Soft-Hard pseudo-label generation strategy in MCL-Net to generate pseudo-labels that are closer to real labels for pathological images. Furthermore, we propose a multi-consistency learning strategy, treating pseudo-labels generated by the Soft-Hard process as real labels, by promoting consistency between predictions of different decoders, enabling the model to learn more sample features. Extensive experiments in this paper demonstrate the effectiveness of the proposed method, providing new insights for the segmentation of microscopic hyperspectral tissue pathology images.

KEYWORDS

microscopic hyperspectral images, semi-supervised learning, medical image segmentation, mutual consistency, pseudo-labels

1 Introduction

Pathological images are considered the gold standard for clinical diagnosis and cancer grading (1). Automatic segmentation of pathological images is a fundamental and crucial step in constructing powerful computer-aided diagnostic systems. Quantitative analysis of the morphological properties of organs and tissues based on segmentation results provides valuable evidence for clinical diagnosis. Existing pathological image segmentation methods (2–4) typically utilize RGB datasets. However, common RGB images can only provide spatial information for cancer diagnosis. The similarity in biological tissue morphology affects the accuracy of diagnostic results.

With the advancement of imaging systems, medical microscopic hyperspectral images have been employed in various tumor recognition applications (5–7). The DMCA method proposed in (5) integrates the classifier for prediction into the extraction of deep features from MedHSIs. This integration ensures compatibility between the extracted features and the classifier, facilitating tumor diagnosis. In (6), Ravi et al. introduced a novel manifold embedding framework called FR-t-SNE. Using this framework, the outputs generated from hyperspectral imaging can be utilized as inputs for semantic segmentation classifiers of brain tissue *in vivo*. The proposed method aims to delineate tumor boundaries, preserve healthy brain tissue, and facilitate complete removal of malignant cells. Muniz et al. proposed, in (7), a method utilizing hyperspectral imaging and micro-FTIR spectroscopy to represent biological tissues based on their spectral characteristics. Subsequently, a deep learning-based classification approach was established to aid experts in distinguishing tissues affected by cancer or inflammation from healthy tissues. Microscopic hyperspectral imaging applied in medical image analysis relies on the following two fundamental principles: i) tissues with similar biochemical compositions are likely to exhibit similar spectra; and ii) variations in spectra can be quantified to delineate different tissues (8). Compared to conventional imaging modalities, medical microscopic hyperspectral pathological images offer additional spectral information, enabling further differentiation of various chemical constituents within biological tissues. Nevertheless, hyperspectral pathological images have higher resolution and larger area, and their annotation requires more time and clinical experience. Therefore, the lack of precise annotations limits the progress of research in pathological image segmentation.

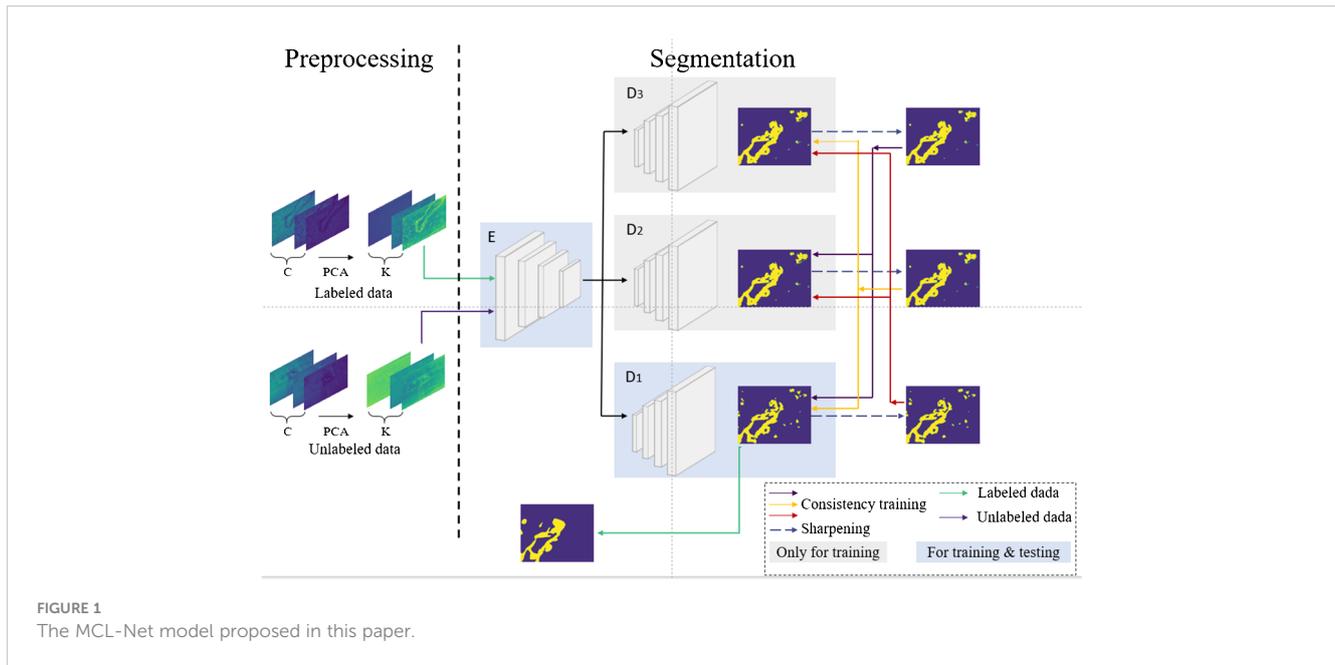
Semi-supervised learning is a method used to address the issue of limited labeled data. This approach typically involves joint training with a small amount of labeled data and a large amount of unlabeled data. The core of this method lies in effectively extracting useful information from both labeled and unlabeled data to achieve relatively stable segmentation results. To achieve this goal, many semi-supervised algorithms have been applied in this field. Common existing semi-supervised segmentation methods can be categorized into pseudo-labeling and consistency regularization methods (9). Firstly, pseudo-labeling is an intuitive approach where a model trained on labeled data is used to predict

pseudo-labels for unlabeled data. These new pseudo-labeled data are then combined with the original labeled set to further refine the model. However, the effectiveness of this method is constrained by the varying quality of the predicted pseudo-labels (9). Consistency regularization methods are based on the smoothness assumption (10). They explore an unsupervised way to leverage unlabeled data. This method (11–13) typically applies slight perturbations to the input data or the model, and learns from unlabeled data by ensuring consistency in model output under different perturbations. Many methods employ a single image to enforce consistency in their perturbations (14), which may lead to inaccurate segmentation results due to a lack of context information across volumes, thus limiting the effectiveness of consistency regularization.

Given the aforementioned issues with consistency regularization and pseudo-labeling methods, this paper introduces a semi-supervised segmentation method for microscopic hyperspectral pathological images based on *multi-consistency learning (MCL-Net)*, which combines both methods. The architecture of the model is illustrated in Figure 1. The model employs a shared encoder and multiple decoders, where the output of the encoder undergoes different perturbations before being fed into distinct decoders. Subsequently, we employ a novel pseudo-label generation method called Soft-Hard to transform the outputs of different decoders into pseudo-labels. Using these generated pseudo-labels as a basis, we devise a novel multi-consistency training approach, wherein soft pseudo-labels obtained from each decoder are treated as genuine labels for the other decoders and subjected to consistency constraints. Through this approach, we minimize discrepancies in output across multiple decoders during model training, thereby obtaining a more comprehensive feature representation. In summary, this paper makes the following four contributions:

- 1) A semi-supervised segmentation method for microscopic hyperspectral pathology images based on multi-consistency learning is proposed in this study. This method combines pseudo-labeling and consistency regularization techniques.
- 2) A multi-consistency learning approach that effectively integrates features extracted by different models is introduced in this research.
- 3) A novel pseudo-label generation method, Soft-Hard, which generates pseudo-labels that are closer to real labels, has been devised.
- 4) Extensive experiments demonstrate that our method outperforms five other state-of-the-art methods, providing new insights for the segmentation of microscopic hyperspectral pathology images.

Remaining sections of the paper are organized as follows: Section 2 provides a review of pathological image segmentation and semi-supervised segmentation methods for medical images. Section 3 introduces the MCL-Net method proposed in this paper. Section 4 outlines the experimental setup. Section 5 is dedicated to result analysis. Finally, Section 6 concludes and summarizes the paper.



2 Related work

2.1 Pathological image segmentation

Pathological images serve as the gold standard for cancer detection, with various segmentation methods being employed in different types of cancer detection. Musulin et al. (15) proposed a two-stage image segmentation method utilizing DeepLabv3+ as the backbone model for predicting oral squamous cell carcinoma in head and neck cancer. Zidan et al. (16) designed a Transformer-based approach constructing a Swin Transformer encoder block to mimic the global context of tumor-related regions in colorectal cancer. Additionally, a cascaded upsampler was devised to utilize supervised multiscale features from the encoder to assist in detecting tumor boundary regions. Jayachandran et al. (17) introduced a novel deep learning framework based on an encoder-decoder structure effectively incorporating attention mechanisms for segmenting osteosarcoma from histological images. Huang et al. (18) proposed an end-to-end ViT-AMCNet, possessing interpretable throat tumor grading capabilities and good interpretability. This model not only ensured good feature representation capabilities of ViT and AMC blocks but also enhanced the redundancy removal ability of the model fusion algorithm. In (19), Rashmi et al. proposed an unsupervised method for segmenting cell nuclei from breast tissue pathology images. A method for selecting template images for color normalization was introduced. An experiment determining a new color channel combination was conducted, which could distinguish cell nuclei from background regions. Furthermore, this work introduced an improved C-V model capable of effectively segmenting nuclei using multi-channel color information. To fully exploit the spectral characteristics of three-dimensional hyperspectral data, Wang et al. (8) applied deep convolutional networks for melanoma segmentation on hyperspectral

pathological images. They introduced a 3D fully convolutional network named Hyper-net for segmenting melanoma from hyperspectral pathological images. Zhang et al. (20) proposed a two-stage segmentation method for OSCC tumors with lymph node metastasis. In the learning stage, this method is employed for coarse segmentation of cancer cell nuclei. In the decision stage, the pathologist's prior knowledge is utilized to make lesion decisions based on the coarse segmentation mask of cancer nuclei, resulting in refined segmentation results. Gao et al. (21) proposed a semi-supervised segmentation method for microscopic hyperspectral pathological images based on shape priors and contrastive learning. They utilized shape priors and image-level contrastive learning to learn features from unlabeled data, enhancing semi-supervised segmentation performance and mitigating limitations posed by limited annotated data. Despite significant progress in tissue pathology image segmentation for cancer prediction, research on hyperspectral pathological images remains limited.

2.2 Consistency regularization

Consistency regularization refers to the similarity of predictions generated by a model under the same input data or model with added random noise. It is a crucial component of temporal ensemble techniques (22). Mean Teacher (23) is a classical temporal ensemble technique where both the student and teacher models adopt the same network structure. Through exponential moving average (EMA), the student network's output across different training iterations becomes similar to that of the teacher network. Various models based on temporal ensemble techniques have been developed based on Mean Teacher. In (24), Shu et al. proposed a novel cross-pollination learning and feature migration mechanism allowing the teacher model to provide higher confidence outputs for student model learning. This method

cross-pollinates unlabeled samples to enhance the segmentation network's generalization ability. It also introduces new cross-gradient monitors to reduce consistency failures caused by semantic gaps between teacher and student models. The average teacher model is enhanced into a novel Fuzzy Consistency Average Teacher (AC-MT) model, where Xu et al. (25) added a series of comprehensive plug-and-play strategies for fuzzy (informative) target selection based on Mean Teacher. This model stabilizes disturbances in regions, enabling more useful representations to be learned from unlabeled data. In (26), Zhang et al. proposed a novel uncertainty-guided mutual consistency learning framework for semi-supervised medical image segmentation. The model employs a dual-task backbone network with two output branches to simultaneously generate segmentation probability maps and signed distance maps. It performs intra-task consistency learning within self-ensemble tasks and utilizes task-level regularization for cross-task consistency learning to leverage geometric shape information. By estimating model segmentation uncertainty guidance, the framework effectively utilizes more reliable information from unlabeled data by selecting relatively determinis.

2.3 Pseudo-labeling

Pseudo-labeling involves generating targets for unlabeled data to obtain amplified, approximately fully labeled datasets (27). Common pseudo-labeling methods focus on effective pseudo-label generation strategies and how to generate high-quality segmentation results under the supervision of pseudo-labels. Wu et al. (28) proposed a novel Mutual Consistency Network (MC-Net) for semi-supervised left atrium segmentation in 3D MR images. MC-Net consists of an encoder and two slightly different decoders. It converts the prediction differences between the two decoders into unsupervised loss through a cyclic pseudo-labeling scheme to encourage mutual consistency. Building upon (28), Wu et al. (29) further introduced a model comprising a shared encoder and multiple slightly different decoders. This model represents the model's uncertainty by computing the statistical differences among the outputs of multiple decoders, indicating uncertain regions in unlabeled data. The model obtains soft pseudo-labels using a sharpening function and applies a novel mutual consistency constraint between the probability output of one decoder and the soft pseudo-labels of other decoders. Chaitanya et al. (30) proposed a joint training framework defining per-pixel contrastive loss on pseudo-labels of unlabeled and sparsely labeled images, while applying traditional segmentation loss only on the labeled set. This method performs pseudo-label-based self-training and trains the network by jointly optimizing the contrastive loss proposed on labeled and unlabeled sets and the segmentation loss on the sparsely labeled set. Chen et al. (31) proposed a semi-supervised tissue segmentation framework called FDCT. This framework introduces the SBOM boundary refinement strategy, utilizing the characteristics of distance maps to optimize the pseudo-labels generated by the model, making them closer to the ground truth labels.

3 Proposed method based on multi-consistency learning

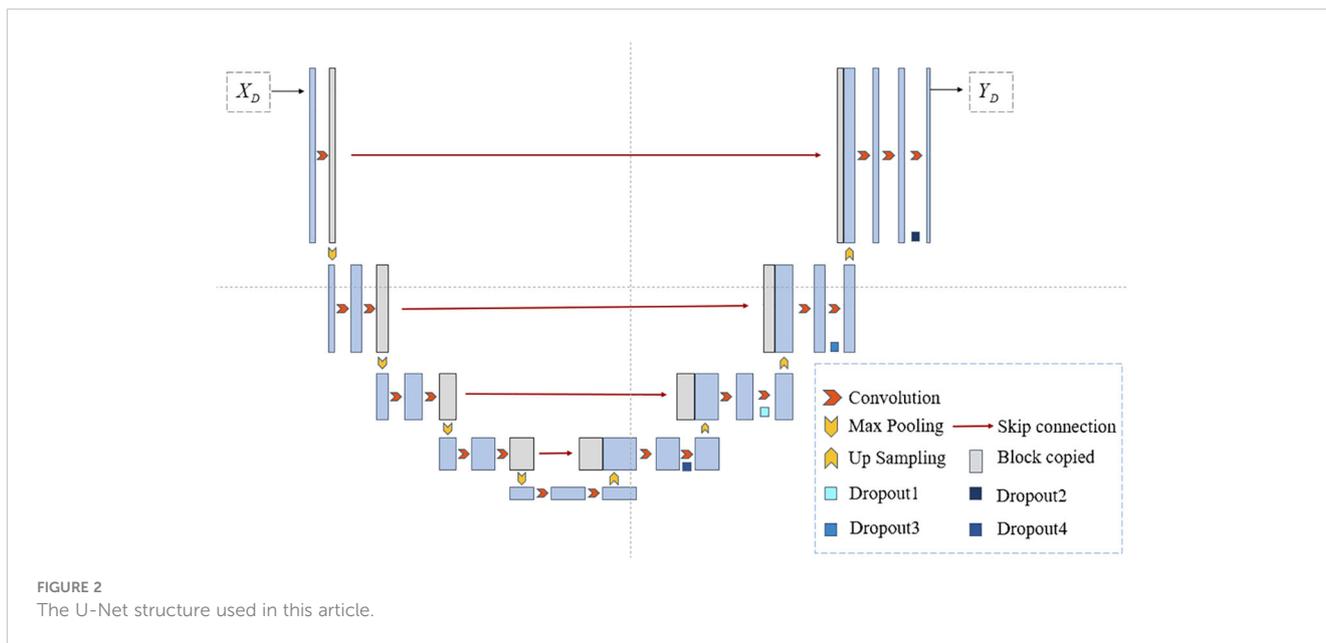
3.1 Overall structure

The task of semi-supervised segmentation of microscopic hyperspectral pathology images aims to learn more sample information by utilizing a small set of labeled samples and a large set of unlabeled samples. In this paper, we work with a dataset, denoted as D , which contains M labeled samples and N unlabeled samples, where $M \ll N$. We define the labeled dataset as $D^L = \{X^i, Y^i\}_{i=1}^M$ and the unlabeled dataset as $D^U = \{X^i\}_{i=M+1}^{M+N}$. Each sample $X^i \in \mathbb{R}^{H \times W \times C}$ in D is a microscopic hyperspectral image with a size of $H \times W$ and C channels. Correspondingly, $Y^i \in \mathbb{R}^{H \times W}$ is the segmentation label map associated with X^i . The objective of semi-supervised segmentation is to learn a segmentation model $f_s(\theta_s)$ parameterized by $D = D^L + D^U$ from θ_s , such that each pixel in the input image is mapped to its correct class.

The network model proposed in this paper is illustrated in Figure 1. During the training phase, in the preprocessing step on the left, Principal Component Analysis (PCA) is employed to perform dimensionality reduction on the input microscopic hyperspectral images. PCA serves a dual purpose: firstly, it helps distance the image from noise, thereby enhancing the data quality; secondly, it eliminates redundant spectral bands, reducing computational overhead and improving processing efficiency. We will use the data obtained by PCA dimensionality reduction as model input. In the segmentation step depicted in Figure 1, the feature vector FA is obtained by first passing through a shared encoder. Subsequently, various perturbations are applied to the feature vector FA, and the perturbed feature vectors are fed into different decoders. The outputs of multiple decoders are subjected to the proposed multi-consistency loss in this paper. All decoders update their model parameters during the training process. However, during testing, only one decoder is selected as the primary decoder, while the others are referred to as auxiliary decoders. Further details will be elucidated in Section 3.2. The U-Net model, known for its simple yet efficient structure, has found wide application in the field of medical image segmentation. Therefore, in our segmentation model, we employ an encoder-decoder structure based on U-Net, as depicted in Figure 2.

3.2 Training with multiple consistencies

After the input data D passes through the encoder E , it yields the feature vector FA. FA is directly fed into the primary decoder $D^L = \{X^i, Y^i\}_{i=1}^M$. Different noise is introduced to the inputs of the model's auxiliary decoders, represented as $FA_i = \text{featurenoise}(FA)$, where $n \geq i$ denotes the number of decoders and $\text{featurenoise}(FA)$ signifies the distinct noise added to FA. Specifically, we set n to 3 to strike a balance between effectiveness and training efficiency. Through experiments in Section 5.3, we ultimately determine $\text{featurenoise}()$ to be Gaussian noise. Further experimental details will be presented in Sections 5.3 and 5.4.



The feature vector is processed by different decoders to obtain distinct probability maps. For the labeled dataset $D^L = \{X^i, Y^i\}_{i=1}^M$ this paper only calculates the supervised loss between the probability map from the primary decoder and its corresponding ground truth label. In this paper, the supervised loss is computed using both cross-entropy loss and Dice loss, defined by Equation (1):

$$L_s = L_{CE}(y_{seg}^L, Y) + L_{dice}(y_{seg}^L, Y) \quad (1)$$

For the unlabeled data, we utilize the Soft-Hard process proposed in this paper to obtain pseudo-labels corresponding to different decoders for the same input. Firstly, we employ a sharpening function (32) to transform the probability maps into soft pseudo-labels. The computation process of the sharpening function is as Equation (2):

$$PS = \frac{p(y_{pred}|x; \theta)^{1/T}}{p(y_{pred}|x; \theta)^{1/T} + (1 - p(y_{pred}|x; \theta))^{1/T}} \quad (2)$$

Here, T is a hyperparameter controlling the sharpening temperature. By choosing an appropriate T , we can apply entropy minimization constraint to regularize our model without introducing additional noise that might interfere with model training. The specific value of the temperature coefficient T will be tested in Section 5.5. Afterwards, we obtain the corresponding hard labels generated by the above process, denoted as $PH = \arg \max(PS)$. The process of obtaining hard labels can be expressed as Equation (3):

$$PH = \arg \max \left(\frac{p(y_{pred}|x; \theta)^{1/T}}{p(y_{pred}|x; \theta)^{1/T} + (1 - p(y_{pred}|x; \theta))^{1/T}} \right) \quad (3)$$

By minimizing the consistency loss imposed on the update direction of the aforementioned constrained model, we enable the primary decoder to integrate sample features extracted by different decoders, thereby maximizing the learning of latent sample

information from unlabeled samples. Ultimately, the overall loss function of this study is formulated as Equation (4):

$$L_u = \sum_{i,j=1 \neq j}^n (L_{CE}(y_{seg}^i, PH_j) + L_{dice}(y_{seg}^i, PH_j)) \quad (4)$$

We treat the hard labels obtained from each decoder as the ground truth labels for the inputs of the other decoders. We then compute the multi-consistency loss between different decoders. Inspired by the findings in (32), we have adopted a strategy of weighting the supervised and unsupervised losses using hyperparameters to achieve improved experimental results. The formula for the multi-consistency loss is as Equation (5):

$$L = \lambda L_s + \beta L_u \quad (5)$$

where λ represents the weight of the supervised loss, and β represents the weight of the unsupervised loss. Due to the lack of true labels for unlabeled data, pseudo-labels generated by the Soft-Hard process might initially lead the model in the wrong direction during training. Therefore, this paper adopts the method from (33) by adding a time-varying Gaussian weighting function, denoted as $\beta = 0.001 \times \exp\left(-5 \times \left(1 - \frac{t}{t_{max}}\right)^2\right)$, to the unsupervised loss. This aims to balance the supervised loss and the consistency loss. Here, t denotes the current iteration count, and t_{max} represents the maximum number of iterations.

4 Experimental setup

4.1 Dataset

In this study, we utilized the cholangiocarcinoma micro-hyperspectral images from the multi-dimensional biliary tract database collected in (34). This dataset comprises 880 scenes from 174 individuals, with 689 scenes containing partially cancerous regions, 49 scenes representing complete cancerous regions, and

142 scenes devoid of any cancerous regions. The spatial resolution of these images is 1024×1280 pixels, each image encompassing 60 bands uniformly distributed from 550nm to 1000nm. While the database provides pixel-level labels for each image, it was observed through experimentation that these labels were somewhat coarse, failing to meet the accuracy requirements for semantic segmentation tasks. As illustrated in Figure 3A, shows a false-color image synthesized using the 5th, 15th, and 25th bands of the input data, while Figure 3B depicts the original labels provided by the database. It can be noted that, on one hand, the labels are disconnected at the opening of the circular structure, while in reality, they should be continuous. On the other hand, the boundaries between the tumor region and the normal region in the original labels are too sharp and abrupt, failing to accurately represent the true boundary information. Therefore, experienced researchers re-annotated the dataset, resulting in a total of 94 re-annotated images. The annotated results are shown in Figure 3C, demonstrating improved continuity and accuracy compared to the original labels.

4.2 Evaluation metrics

We employed four evaluation metrics to assess the performance of semi-supervised segmentation on histopathological images, including overall accuracy (OA), average accuracy (AA), Dice coefficient, and mean intersection over Union (MIoU). Better segmentation performance is indicated by higher values of OA, AA, Dice, and MIoU.

OA and AA represent the proportion of pixels correctly classified: We define the tumor region as positive samples and the non-cancer region (i.e., normal region) as negative samples. TP, TN, FP, and FN denote true positive pixels, true negative pixels, false positive pixels, and false negative pixels, respectively. Then, OA and AA can be defined as Equations (6) and (7):

$$OA = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

$$AA = \frac{1}{2} \left(\frac{TP}{TP + FN} + \frac{TN}{FP + TN} \right) \quad (7)$$

MIoU measures the ratio of the intersection to the union between the sets of true positive and predicted positive pixels as Equation (8). It provides a measure of how well the predicted

segmentation aligns with the ground truth segmentation.

$$MIoU = \frac{1}{2} \left(\frac{TP}{TP + FP + FN} + \frac{TN}{FP + FN + TN} \right) \quad (8)$$

The Dice coefficient is a metric that quantifies the similarity between the predicted segmentation (P) and the ground truth labels (Y) based on region overlap as Equation (9). It's widely used in image segmentation tasks to evaluate the accuracy of the segmentation results.

$$Dice = \frac{2 \cdot |P \cap Y|}{|P| + |Y|} \quad (9)$$

4.3 Implementation detail

In this study, we implemented the model in the environment of PyTorch 1.13.1 with CUDA 11.7 and Python 3.8. Training and testing were performed on an NVIDIA GeForce RTX 3090. The batch size was set to 2, with batch sizes of 1 for labeled and unlabeled data, respectively. In the data preprocessing stage, the input data was dimensionally reduced to 6 channels using PCA. We utilized the SGD optimizer for training the entire network for 100 epochs, with a learning rate of 0.01 and momentum set to 0.9. Gaussian noise with mean 1 and standard deviation 1.2 was added to one auxiliary decoder, and mean 1 and standard deviation 1.5 was added to the other auxiliary decoder.

5 Result

5.1 Comparative experiments

To demonstrate the effectiveness of our proposed semi-supervised method, we conducted comparative experiments on the Multidimensional Cholangiocarcinoma Dataset. We compared our method with six other approaches, including: (1) 2D U-Net (35); (2) Mean Teacher (MT) (23); (3) Uncertainty Aware Mean Teacher (UA-MT) (33); (4) Cross Consistency Training (CCT) (36); (5) Cross Pseudo Supervision (CPS) (14); (6) Uncertainty-aware Pseudo-label and Consistency (UPC) (37); Here, the 2D U-Net is trained in a fully supervised manner using a limited set of labeled samples. Our proposed method and the other five approaches utilize semi-

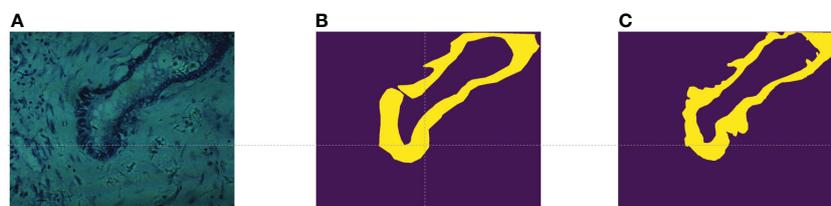


FIGURE 3

(A) False color images synthesized using bands 5,15, and 25; (B) The original label provided in (34); (C) Our re-labeling.

supervised learning algorithms with a certain proportion of labeled data and a large amount of unlabeled data.

Table 1 presents quantitative results obtained using various semi-supervised models with different labeling ratios. From Table 1, it can be observed that our proposed method outperforms the fully supervised approach in scenarios with different labeling ratios. Particularly, when using 20% labeled data, our method shows improvements of 1.23% in OA, 0.77% in AA, 2.04% in MIOU, and 1.97% in DICE compared to the fully supervised 2D U-Net. This indicates that our proposed method is able to better utilize the information embedded in the unlabeled data compared to the fully supervised approach. In comparison with other semi-supervised methods, our method achieves results close to the other methods, indicating that our proposed approach is suitable for histopathological microscopic hyperspectral images. Figure 4 displays the predicted results of the fully supervised and semi-supervised methods, including our proposed method, using different labeling ratios. It can be observed from the figure that our method's predicted results are closer to the ground truth labels.

5.2 Ablation study

In order to validate the effectiveness of the proposed method, ablation experiments were conducted on the multi-dimensional bile duct dataset. We removed the “multi-consistency” learning method and the Soft-Hard pseudo-label generation method, constructing a “basic” model that uses a shared encoder and three independent decoders. We then separately added the multi-consistency learning strategy and the Soft-Hard pseudo-label generation method, referred to as “basic+mcl” and “basic+s-h” respectively. Next, we incorporated both of these methods into the “basic” model to obtain

our final model, MCL-Net. The quantitative analysis results of the ablation study are presented in Table 2.

From Table 2, it can be observed that when using the “basic” model, the experimental results showed little improvement compared to the fully supervised approach. This indicates that merely employing a multi-decoder structure does not significantly enhance experimental performance. When we added our proposed multi-consistency learning strategy to the original model, the Dice coefficient and MIOU improved by 1.07% and 1.11% respectively. This demonstrates that the multi-consistency learning strategy effectively integrates features extracted by different decoders, extracting more sample information from unlabeled data.

When we separately added the Soft-Hard method to the model, the two evaluation metrics improved from 69.43% and 71.12% to 70.69% and 72.23%. This suggests that our proposed pseudo-label generation strategy can bring pseudo-labels closer to real labels to a greater extent. When we applied both of our proposed methods together on the model, experimental performance further improved. This indicates that our multi-consistency method and Soft-Hard method can collaborate synergistically to enhance experimental performance.

5.3 Influence of different data perturbation methods

To increase the diversity of input data among different decoders, we applied perturbations to the inputs of the two decoders, excluding the main decoder. We utilized three perturbation methods: adding Gaussian noise (Add-gn), adding salt-and-pepper noise (Add-sp), and adding Poisson noise (Add-pn). In our experiments, we employed strategies of using the same

TABLE 1 Quantitative experimental results of different methods on multi-dimensional common bile duct dataset.

Labeled data	Method	OA↑	AA↑	Dice↑	MIOU↑
6/66(10%)	MT (23)	88.69%	82.34%	66.00%	69.24%
	UA-MT (33)	88.62%	82.93%	66.59%	69.35%
	CCT (36)	87.83%	82.69%	65.24%	68.21%
	CPS (14)	88.72%	82.83%	66.38%	69.53%
	UPC (37)	88.50%	83.32%	66.81%	69.65%
	2D-Unet (35)	88.56%	82.20%	65.24%	68.77%
	Ours	88.81%	82.76%	65.50%	68.95%
	13/66(20%)	MT (23)	89.84%	86.50%	70.61%
UA-MT (33)		90.17%	86.05%	70.77%	72.67%
CCT (36)		89.74%	85.96%	71.05%	72.79%
CPS (14)		90.19%	86.03%	70.88%	72.59%
UPC (37)		90.32%	85.11%	70.40%	72.53%
2D-Unet (35)		89.25%	85.14%	69.15%	71.00%
Ours		90.48%	85.91%	71.19%	72.97%

Bold text represent the optimal result.

The symbol ‘↑’ signifies that a higher metric corresponds to better segmentation performance.

TABLE 2 Ablation results on multidimensional common bile duct dataset.

Labeled data	Method	Dice↑	MIoU↑
6/66(10%)	basic	69.43%	71.12%
	basic+mcl	70.50%	72.23%
	basic+s-h	70.69%	72.01%
	ours	71.19%	72.97%

The symbol '↑' signifies that a higher metric corresponds to better segmentation performance.

noise and adding different noises to different decoder inputs. Specifically, we denoted the use of Gaussian noise and salt-and-pepper noise as (Add-gspn), the use of Gaussian noise and Poisson noise as (Add-gpn), and the use of salt-and-pepper noise and Poisson noise as (Add-sppn). Table 3 records the quantitative results corresponding to the six methods mentioned above. For each noise adding method, we conducted numerous experiments, and the results in Table 3 represent the optimal settings for each method. From Table 3, it can be observed that the optimal experimental results are achieved when adding Gaussian noise to both auxiliary decoders. This may be attributed to the fact that Gaussian noise better simulates the existing noise patterns in images, aligning more closely with real-world requirements.

5.4 Impact of different decoder numbers

In order to obtain more comprehensive sample features, we designed a multi-decoder structure. To better understand the

influence of the number of decoders n on the experimental results, we set n to different values. When $n = 1$, the model corresponds to the classic 2D U-Net model. Since it is not possible to apply the proposed multi-consistency loss and Soft-Hard method in this case, the model operates in a fully supervised manner. When $n \geq 1$, as described in section 5.3, experiments were conducted by perturbing the input data to the decoders using Gaussian noise applied before the auxiliary decoders.

Figure 5 illustrates the influence of the number of decoders on the Dice similarity coefficient under different proportions of labeled data. It can be observed from the figure that as n increases from 1 to 2, the Dice coefficient significantly improves. This indicates that the multi-decoder structure and multi-consistency learning strategy can comprehensively learn the model features. As n further increases from 2 to 3, the Dice coefficient continues to improve, suggesting that appropriately increasing the number of decoders allows for more effective utilization of information from unlabeled data. However, when n is increased to 4, the Dice coefficient experiences only a slight increase. Therefore, to balance accuracy and efficiency in the experiments, we set $n = 3$.

5.5 Impact of temperature T

In Equation 2, we use a sharpening function to generate preliminary pseudo-labels. Figure 6 presents the Dice coefficients obtained by training our MCL-Net+ model on the multi-dimensional bile duct dataset with different temperature values T . Following the guidance from (29), we experimented with different values of T , specifically setting T to 0.01, 0.1, 0.5, and 1.

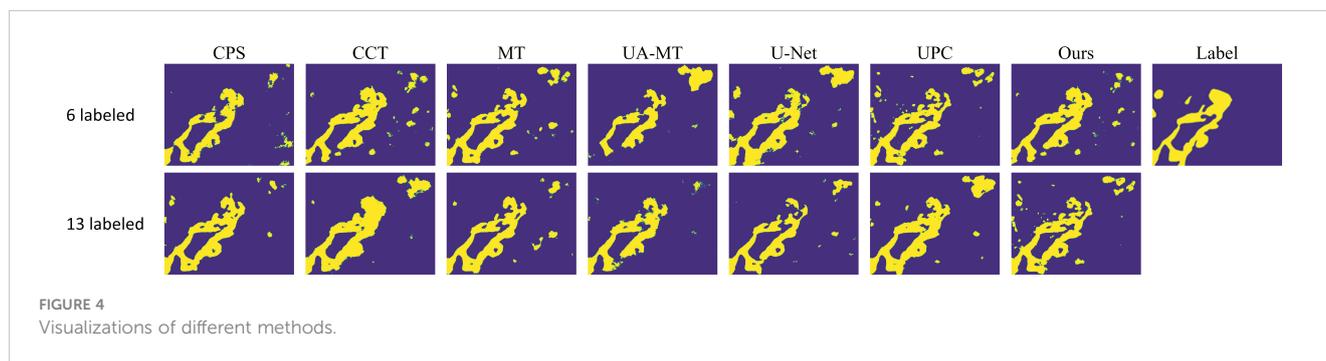


FIGURE 4 Visualizations of different methods.

TABLE 3 Influence of different data perturbation methods.

Labeled data	Method	OA↑	AA↑	Dice↑	MIoU↑
6/66(10%)	Add-gn	90.48%	85.91%	71.19%	72.97%
	Add-spn	89.60%	84.28%	69.24%	71.51%
	Add-pn	88.77%	84.07%	70.61%	71.94%
	Add-gpn	90.45%	84.50%	70.81%	72.85%
	Add-gspn	89.59%	84.36%	69.29%	71.54%
	Add-sppn	89.72%	84.85%	69.23%	71.40%

Bold text represent the optimal result.

The symbol '↑' signifies that a higher metric corresponds to better segmentation performance.

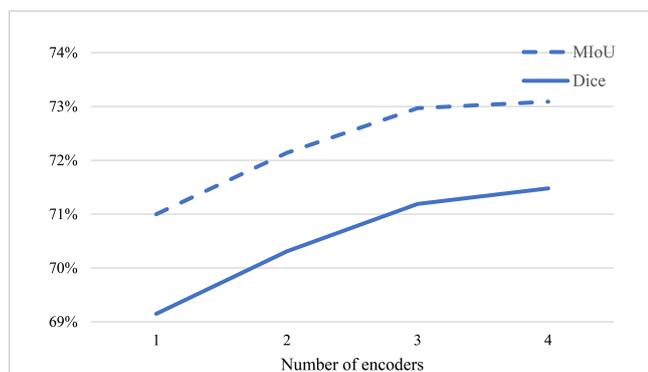


FIGURE 5
Impact of different decoder numbers on experimental results.

From Figure 6, it can be observed that when $T = 0.5$, the model achieves the optimal experimental results. When T is too large, the model may fail to generate reasonable soft pseudo-labels due to the inability to utilize entropy minimization. On the other hand, when T is too small, it may introduce noise into the pseudo-labels, leading to prediction errors. Therefore, this study ultimately selects a temperature coefficient of $T = 0.5$.

5.6 Impact of supervised loss weight λ

We further investigated the influence of the weight of the supervised loss term λ in the loss function. In Equation 5, the weight of the unsupervised loss is set according to a Gaussian warming function, while λ affects the balance between the two types of losses. Figure 7 illustrates how different weights impact the experimental performance.

When λ is too high, the model tends to focus more on extracting features from the labeled data, but this comes at the cost of neglecting the unlabeled data, as the proposed multi-consistency loss may not be effectively utilized. On the other hand, when λ is too low, the accurately labeled data is not effectively leveraged, resulting in poorer experimental results. Therefore, we finally chose $\lambda = 0.5$ as the final experimental setting.

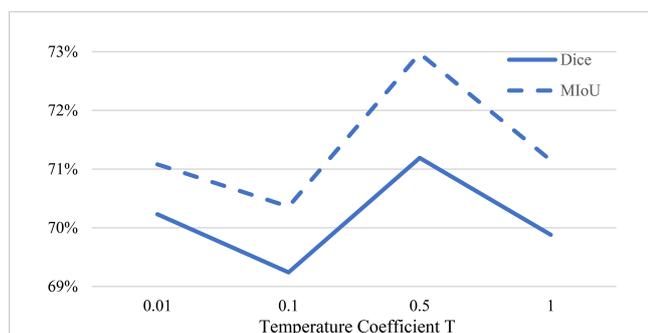


FIGURE 6
Impact of temperature coefficient T on experimental results.

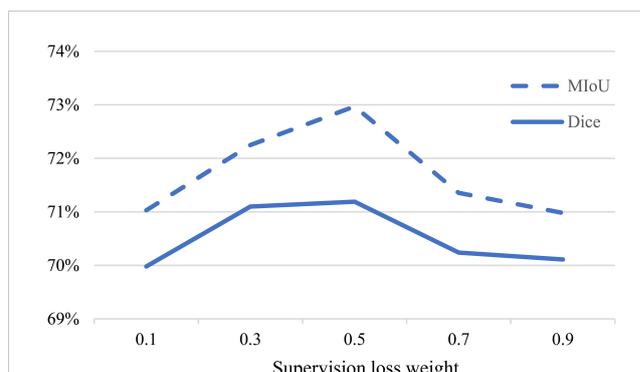


FIGURE 7
Impact of supervision loss weight on experimental results.

6 Conclusion and future work

In this paper, we proposed a novel semi-supervised segmentation model, MCL-Net, for microspectroscopic pathology images. The model combines consistency regularization and pseudo-labeling methods. MCL-Net employs a shared encoder and multiple independent decoders. Through the proposed Soft-Hard pseudo-labeling strategy, MCL-Net generates pseudo-labels that are closer to the real labels for pathological images. Additionally, we introduced a multi-consistency learning strategy, treating the pseudo-labels generated by the Soft-Hard process as real labels. This encourages consistency among predictions from different decoders, enabling the model to learn more sample features.

The effectiveness of this approach was demonstrated through extensive experiments, providing a new perspective for the segmentation of microspectroscopic pathological images. Despite the promising results, there are limitations. Specifically, when using only 10% labeled data for experiments, our method did not significantly improve performance. This might be attributed to the limited explicit application of spectral information, which is unique to microspectroscopy. In the future, we will further explore ways to utilize spectral information and consider both labeled and unlabeled samples from multiple angles.

Data availability statement

The datasets presented in this article are not readily available because this is for the use of this team only. Requests to access the datasets should be directed to JF, 2106020108@hhu.edu.cn.

Ethics statement

Ethical approval was not required for the study involving humans in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and the institutional requirements.

Author contributions

JF: Data curation, Formal Analysis, Investigation, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was supported by National Undergraduate Training Program for Innovation and Entrepreneurship under Grant 202310294055Z.

References

- Li Z, Li C, Luo X, Zhou Y, Zhu J, Xu C, et al. Toward source-free cross tissues histopathological cell segmentation via target-specific finetuning. *IEEE Trans Med Imaging*. (2023) 42:2666–77. doi: 10.1109/TMI.2023.3263465
- Hu W, Li X, Li C, Li R, Jiang T, Sun H, et al. A state-of-the-art survey of artificial neural networks for whole-slide image analysis: from popular convolutional neural networks to potential visual transformers. *Comput Biol Med*. (2023) 161:107034. doi: 10.1016/j.combiomed.2023.107034
- Mehta S, Lu X, Wu W, Weaver D, Hajishirzi H, Elmore JG, et al. End-to-End diagnosis of breast biopsy images with transformers. *Med Image Anal*. (2022) 79:102466. doi: 10.1016/j.media.2022.102466
- Gao Z, Jia C, Li Y, Zhang X, Hong B, Wu J, et al. Unsupervised representation learning for tissue segmentation in histopathological images: From global to local contrast. *IEEE Trans Med Imaging*. (2022) 41:3611–23. doi: 10.1109/TMI.2022.3191398
- Wang M, Xu Y, Wang Z, Xing C. Deep margin cosine autoencoder-based medical hyperspectral image classification for tumor diagnosis. *IEEE Trans Instrum Meas*. (2023) 72:1–12. doi: 10.1109/TIM.2023.3293548
- Ravi D, Fabelo H, Callic GM, Yang G-Z. Manifold embedding and semantic segmentation for intraoperative guidance with hyperspectral brain imaging. *IEEE Trans Med Imaging*. (2017) 36:1845–57. doi: 10.1109/TMI.42
- Muniz FB, Baffa MdeFO, Garcia SB, Bachmann L, Felipe JC. Histopathological diagnosis of colon cancer using micro-FTIR hyperspectral imaging and deep learning. *Comput Methods Programs BioMed*. (2023) 231:107388. doi: 10.1016/j.cmpb.2023.107388
- Wang Q, Sun L, Wang Y, Zhou M, Hu M, Chen J, et al. Identification of melanoma from hyperspectral pathology image using 3D convolutional networks. *IEEE Trans Med Imaging*. (2020) 40:218–27. doi: 10.1109/TMI.42
- Wang K, Zhan B, Zu C, Wu X, Zhou J, Zhou L, et al. Semi-supervised medical image segmentation via a tripled-uncertainty guided mean teacher model with contrastive learning. *Med Image Anal*. (2022) 79:102447. doi: 10.1016/j.media.2022.102447
- Van Engelen JE, Hoos HH. A survey on semi-supervised learning. *Mach Learn*. (2020) 109:373–440. doi: 10.1007/s10994-019-05855-6
- Zhao X, Qi Z, Wang S, Wang Q, Wu X, Mao Y, et al. RCPS: rectified contrastive pseudo supervision for semi-supervised medical image segmentation. *IEEE J BioMed Health Inform*. (2024) 28:251–61. doi: 10.1109/JBHI.2023.3322590
- Xu C, Yang Y, Xia Z, Wang B, Zhang D, Zhang Y, et al. Dual uncertainty-guided mixing consistency for semi-supervised 3D medical image segmentation. *IEEE Trans Big Data*. (2023) 9:1156–70. doi: 10.1109/TBDATA.2023.3258643
- Zhao Y, Lu K, Xue J, Wang S, Lu J. Semi-supervised medical image segmentation with voxel stability and reliability constraints. *IEEE J BioMed Health Inform*. (2023) 27:3912–23. doi: 10.1109/JBHI.2023.3273609
- Chen X, Yuan Y, Zeng G, Wang J. (2021). Semi-supervised semantic segmentation with cross pseudo supervision, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, . pp. 2613–22.
- Musuljin J, Štifićanić D, Zuljanić A, Čabov T, Dekanić A, Car Z. An enhanced histopathology analysis: An ai-based system for multiclass grading of oral squamous cell carcinoma and segmenting of epithelial and stromal tissue. *Cancers (Basel)*. (2021) 13:1784. doi: 10.3390/cancers13081784
- Zidan U, Gaber MM, Abdelsamea MM. SwinCup: Cascaded swin transformer for histopathological structures segmentation in colorectal cancer. *Expert Syst Appl*. (2023) 216:119452. doi: 10.1016/j.eswa.2022.119452

Conflict of interest

The author declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Jayachandran A, Ganesh S, Kumar SR. Multi-stage deep convolutional neural network for histopathological analysis of osteosarcoma. *Neural Comput Appl*. (2023) 35:1–14. doi: 10.1007/s00521-023-08837-x
- Huang P, He P, Tian S, Ma M, Feng P, Xiao H, et al. ViT-AMC network with adaptive model fusion and multiobjective optimization for interpretable laryngeal tumor grading from histopathological images. *IEEE Trans Med Imaging*. (2022) 42:15–28. doi: 10.1109/TMI.2022.3202248
- Rashmi R, Prasad K, Udupa CBK. Multi-channel Chan-Vese model for unsupervised segmentation of nuclei from breast histopathological images. *Comput Biol Med*. (2021) 136:104651. doi: 10.1016/j.combiomed.2021.104651
- Zhang X, Li Q, Li W, Guo Y, Zhang J, Guo C, et al. FD-net: feature distillation network for oral squamous cell carcinoma lymph node segmentation in hyperspectral imagery. *IEEE J BioMed Health Inform*. (2024) 28:1552–63. doi: 10.1109/JBHI.2024.3350245
- Gao H, Wang H, Chen L, Cao X, Zhu M, Xu P. Semi-supervised segmentation of hyperspectral pathological imagery based on shape priors and contrastive learning. *BioMed Signal Process Control*. (2024) 91:105881. doi: 10.1016/j.bspc.2023.105881
- Laine S, Aila T. Temporal ensembling for semi-supervised learning. *ArXiv Preprint ArXiv*. (2016) arXiv:1610.02242. doi: 10.48550/arXiv.1610.02242
- 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.
- Shu Y, Li H, Xiao B, Bi X, Li W. Cross-mix monitoring for medical image segmentation with limited supervision. *IEEE Trans Multimedia*. (2022) 25:1700–12. doi: 10.1109/TMM.2022.3154159
- Xu Z, Wang Y, Lu D, Luo X, Yan J, Zheng Y, et al. Ambiguity-selective consistency regularization for mean-teacher semi-supervised medical image segmentation. *Med Image Anal*. (2023) 88:102880. doi: 10.1016/j.media.2023.102880
- Zhang Y, Jiao R, Liao Q, Li D, Zhang J. Uncertainty-guided mutual consistency learning for semi-supervised medical image segmentation. *Artif Intell Med*. (2023) 138:102476. doi: 10.1016/j.artmed.2022.102476
- Wu H, Li X, Lin Y, Cheng K-T. Compete to win: enhancing pseudo labels for barely-supervised medical image segmentation. *IEEE Trans Med Imaging*. (2023) 42:3244–55. doi: 10.1109/TMI.2023.3279110
- Wu Y, Xu M, Ge Z, Cai J, Zhang L. (2021). Semi-supervised left atrium segmentation with mutual consistency training, in: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021*. pp. 297–306.
- Wu Y, Ge Z, Zhang D, Xu M, Zhang L, Xia Y, et al. Mutual consistency learning for semi-supervised medical image segmentation. *Med Image Anal*. (2022) 81:102530. doi: 10.1016/j.media.2022.102530
- Chaitanya K, Erdil E, Karani N, Konukoglu E. Local contrastive loss with pseudo-label based self-training for semi-supervised medical image segmentation. *Med Image Anal*. (2023) 87:102792. doi: 10.1016/j.media.2023.102792
- Chen Z, Hou Y, Liu H, Ye Z, Zhao R, Shen H. FDCT: Fusion-Guided Dual-View Consistency Training for semi-supervised tissue segmentation on MRI. *Comput Biol Med*. (2023) 160:106908. doi: 10.1016/j.combiomed.2023.106908
- Li X, He M, Li H, Shen H. A combined loss-based multiscale fully convolutional network for high-resolution remote sensing image change detection. *IEEE Geosci Remote Sens Lett*. (2022) 19:1–5. doi: 10.1109/LGRS.2021.3098774
- Yu L, Wang S, Li X, Fu C-W, Heng P-A. (2019). Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation, in: *Medical Image*

Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019. pp. 605–13.

34. Zhang Q, Li Q, Yu G, Sun L, Zhou M, Chu J. A multidimensional choledoch database and benchmarks for cholangiocarcinoma diagnosis. *IEEE Access*. (2019) 7:149414–21. doi: 10.1109/Access.6287639

35. Ronneberger O, Fischer P, Brox T. (2015). U-net: Convolutional networks for biomedical image segmentation, in: *Medical Image Computing and Computer-Assisted*

Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015. pp. 234–41.

36. Ouali Y, Hudelot C, Tami M. Semi-supervised semantic segmentation with cross-consistency training. *IEEE*. (2020). 12674–84. doi: 10.1109/CVPR42600.2020.01269

37. Lu L, Yin M, Fu L, Yang F. Uncertainty-aware pseudo-label and consistency for semi-supervised medical image segmentation. *BioMed Signal Process Control*. (2023) 79:104203. doi: 10.1016/j.bspc.2022.104203