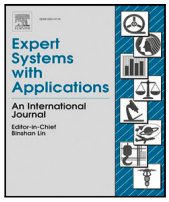
Expert Systems With Applications 238 (2024) 122093

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa



Inter- and intra-uncertainty based feature aggregation model for semi-supervised histopathology image segmentation

Qiangguo Jin [a], Hui Cui [b], Changming Sun [c], Yang Song [d], Jiangbin Zheng [a], Leilei Cao [a], Leyi Wei [e], Ran Su [f,*]

[a] School of Software, Northwestern Polytechnical University, Shaanxi, China
[b] Department of Computer Science and Information Technology, La Trobe University, Melbourne, Australia
[c] CSIRO Data61, Sydney, Australia
[d] School of Computer Science and Engineering, University of New South Wales, Sydney, Australia
[e] School of Software, Shandong University, Shandong, China
[f] School of Computer Software, College of Intelligence and Computing, Tianjin University, Tianjin, China

ARTICLE INFO

Keywords:
Semi-supervised learning
Feature aggregation
Uncertainty regularization
Histopathology image segmentation

ABSTRACT

Acquiring pixel-level annotations is often limited in applications such as histology studies that require domain expertise. Various semi-supervised learning approaches have been developed to work with limited ground truth annotations, such as the popular teacher-student models. However, hierarchical prediction uncertainty within the student model (intra-uncertainty) and image prediction uncertainty (inter-uncertainty) have not been fully utilized by existing methods. To address these issues, we first propose a novel inter- and intra-uncertainty regularization method to measure and constrain both inter- and intra-inconsistencies in the teacher-student architecture. We also propose a new two-stage network with pseudo-mask guided feature aggregation (PG-FANet) as the segmentation model. The two-stage structure complements with the uncertainty regularization strategy to avoid introducing extra modules in solving uncertainties and the aggregation mechanisms enable multi-scale and multi-stage feature integration. Comprehensive experimental results over the MoNuSeg and CRAG datasets show that our PG-FANet outperforms other state-of-the-art methods and our semi-supervised learning framework yields competitive performance with a limited amount of labeled data.

1. Introduction

Accurate instance segmentation in histology images is important to analyze the morphology of various structures, such as nuclei and glands, which is essential for disease diagnosis (Graham et al., 2019), prognostic prediction (Lu et al., 2021), and tissue phenotyping (Javed et al., 2020). However, it is impractical to segment numerous nuclei/glands manually due to the subtle contrast between the objects of interest and background tissues, and high complexity of morphological features. Therefore, automated segmentation methods are highly demanded for histopathology images.

Deep learning based methods have demonstrated superiority in histopathology image segmentation (Graham et al., 2019; Su et al., 2015) and achieved outstanding performance under full supervision. Although these solutions have shown superior performance under full supervision, these segmentation methods rely heavily on huge numbers of pixel-level annotations, which are difficult to obtain. The difficulties in delineating histopathology images can cause extremely high workload on pathologists. Semi-supervised learning (SSL) is one of the approaches to train with limited supervision. However, semi-supervised instance segmentation remains a challenging task in histopathology image processing due to image characteristics and domain problems. First, histopathology objects, such as nuclei and glands, are often closely adjacent or overlapping with each other, making it difficult to delineate separate instances. Second, areas of uncertainty that exist within the objects and around the boundaries may not be well captured under limited supervision.

Current state-of-the-art SSL methods for biomedical image segmentation can be roughly divided in four categories. The first category is based on consistency regularization (Luo et al., 2022; Tarvainen & Valpola, 2017; Wang et al., 2020; Wu et al., 2022; Xu et al., 2023; Yu, Wang, Li, Fu, & Heng, 2019), which minimizes the prediction variance on a given unlabeled example and its perturbed version. The second type is to generate pseudo labels by learning from labeled data, and then use the pseudo labels to enhance the learning from unlabeled data (Bai, Chen, Li, Shen, & Wang, 2023; Chen, Yuan, Zeng, & Wang,

* Corresponding author.
E-mail address: ran.su@tju.edu.cn (R. Su).

https://doi.org/10.1016/j.eswa.2023.122093
Received 25 June 2023; Received in revised form 8 October 2023; Accepted 8 October 2023
Available online 14 October 2023

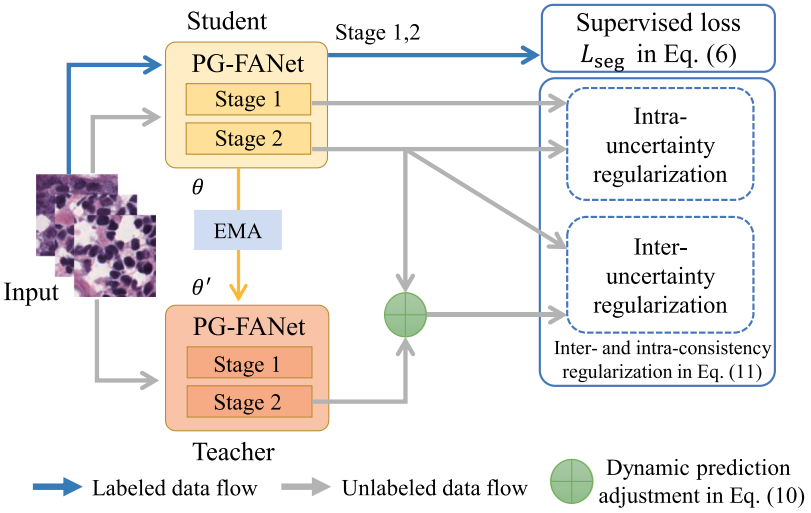


Fig. 1. The overall architecture of the proposed semi-supervised histopathology image segmentation model using two-stage PG-FANet and inter- and intra-uncertainty and consistency regularization. EMA denotes exponential moving average.

2021; Li, Chen, Xie, Ma and Zheng, 2020; Li et al., 2020; Zhao et al., 2023; Zheng et al., 2020). The third one is adversarial learning, which is introduced (Lei et al., 2022; Xu et al., 2022; Zhang et al., 2017) to enforce higher-level consistency between labeled data and unlabeled data. The last category utilizes contrastive learning (CL) (Basak & Yin, 2023; Chaitanya, Erdil, Karani, & Konukoglu, 2023; Gu et al., 2022; Shi, Gong, Wang, & Li, 2022; Zhang, Zhang, Tian, Lukasiewicz, & Xu, 2023), which forces the network to learn representative features in similar and dissimilar regions for segmentation. As for histopathology image segmentation, Li, Chen et al. (2020) generated pseudo labels as guidance for nuclei segmentation. Xie et al. (2020) proposed a pairwise relation-based semi-supervised (PRS^2) model for gland segmentation on histology images. Shi et al. (2022) and Wu, Wang, Song, Yang and Qin (2022) proposed contrast learning based semi-supervised model for tissue and nuclei segmentation.

However, the above models have several limitations. First, most of the previous SSL methods (Li, Yu et al., 2020; Yu et al., 2019; Zhou, Chen, Lin, & Heng, 2020) are based on the mean teacher (Tarvainen & Valpola, 2017) architecture. As unlabeled data do not have ground truth (GT), a common strategy is to use the predictions by the teacher model as guidance. Unfortunately, it is not guaranteed that the teacher model always generates better results than the student model from unlabeled data. Such prediction discrepancies relies crucially on the uncertainty of each target prediction by the teacher and student model (i.e., inter-uncertainty). Second, in convolutional neural networks (CNNs), extracting features at one particular layer could affect the following layers (Dou et al., 2019; Jin et al., 2022; Zheng & Yang, 2021), which may cause inconsistencies during information propagation. Such inconsistencies caused by the hierarchical network are neglected. Hence, we suggest leveraging uncertainty among the student network (i.e., intra-uncertainty) to reduce such prediction inconsistencies. Third, the pseudo labels may contain noise that results in ambiguous guidances, especially along object boundaries. Finally, methods based on adversarial learning are difficult to train, and CL based methods may have issues in distinguishing the foreground pixels and background pixels in high-level feature maps due to low spatial resolution.

To address these challenges existing in fully- or semi-supervised histopathology image segmentation, we propose an inter- and intra-uncertainty regularization based semi-supervised segmentation framework with a multi-scale multi-stage feature aggregation network. In our SSL setting, we gradually add fully annotated images to the labeled dataset to simulate annotating processes by pathologists. The overall architecture is shown in Fig. 1. To learn from both labeled and unlabeled data for annotation-efficient segmentation, we exploit the learning strategy in the mean teacher framework, which is simple yet effective. In the mean teacher architecture, the student model is trained by using both labeled and unlabeled data, and the teacher model is an average of succeeding student models. The major innovations and contributions of the proposed method include:

- We propose a novel pseudo-mask guided feature aggregation network (PG-FANet)[1], which consists of two-stage sub-networks, a mask-guided feature enhancement (MGFE) module, and a multi-scale multi-stage feature aggregation (MMFA) module. The MGFE drives the attention of the network to the region of interest (ROI) under coarse semantic segmentation. The MMFA enables simultaneous aggregation of multi-scale and multi-stage features, so as to avoid the impact of feature incompatibilities in conventional U-shape skip connections.
- We propose an inter- and intra-uncertainty modeling and measurement mechanism to penalize both inter- and intra-uncertainties in the teacher-student architecture. The inter-uncertainty regularization formula is enhanced by a newly introduced shape attention mechanism, aiming to improve the consistency of contour-relevant predictions from the student and the teacher models.
- The performances of the major components in the proposed model are validated by extensive experiments over two public histopathology image datasets for nuclei and gland segmentation. Experimental results show that our PG-FANet outperforms other fully-supervised state-of-the-art models and our proposed semi-supervised learning architecture achieves desirable performance with limited labeled data when compared with recent semi-supervised learning schemes.

2. Related work

2.1. Histopathology image segmentation

Automatically segmenting histopathology images is a challenging task. Many methods have been proposed in segmenting histopathology images under full supervision. For nuclei segmentation, Chen, Huang, Chen, Qian, and Yu (2023), Su et al. (2015), and Xiang et al. (2020, 2022) handled the large variations of shapes and inhomogeneous intensities of nuclei, while Chen, Qi, Yu, and Heng (2016), Graham

[1] Source code will be released at https://github.com/qgking/PG-FANet.

et al. (2019), and Xie, Lu, Zhang, Shen, and Xia (2019) proposed to reduce the ambiguity for glands of various sizes. For both tasks, a full-resolution convolutional neural network (FullNet) was proposed by Qu, Yan, Riedlinger, De, and Metaxas (2019), where a variance constrained cross-entropy loss was introduced to explicitly learn the instance-level relation between pixels in nuclei/gland images. Yang, Dasmahapatra, and Mahmoodi (2023) proposed a nested U-Net that combines cascade training and AdaBoost algorithm for histopathology image segmentation.

The aforementioned methods are all fully-supervised approaches, which require pixel-level annotations by pathologists. Considering the large number of nuclei (e.g., 28,846 in Kumar et al. (2019)) in a histopathology image dataset, reducing the workload of pathologists in clinical practice by leveraging reduced annotations via semi-supervised approach is worth investigation.

To deal with limited annotations, researchers attempted to exploit useful information from unlabeled data with semi-supervised techniques. Self-loop (Li, Chen et al., 2020) was proposed to utilize the generated pseudo label as guidance to optimize the neural network. Zhou et al. (2020) explored a mask-guided feature distillation mechanism for nuclei instance segmentation. Xie et al. (2020) proposed a segmentation network (S-Net) and a pairwise relation network (PR-Net) for gland segmentation. However, the vanilla pseudo labels may contain noise resulting in ambiguous guidances, and the U-Net (Ronneberger, Fischer, & Brox, 2015) based methods may introduce feature incompatibilities (Ibtehaz & Rahman, 2020).

In our work, we explore a pseudo label based two-stage model to help the feature representation ability for both nuclei and gland segmentation tasks.

2.2. Feature aggregation

Recently, deep learning networks have been proposed to enhance feature representation and aggregation in biomedical analysis tasks (Cao et al., 2022; Chen et al., 2018; Cui et al., 2022; Jiang, Zhang, Zhou, Wang, & Chen, 2023; Li et al., 2022; Sundaresan, Zamboni, Rothwell, Jenkinson, & Griffanti, 2021; Yan, Lv, Guo, Peng and Liu, 2023; Yang & Farsiu, 2023; Yu & Koltun, 2015; Yu et al., 2020, 2023; Zhong et al., 2020). The networks for biomedical image segmentation can be roughly divided into U-shape based architectures and none-U-shape based architectures.

Since U-Net (Ronneberger et al., 2015) showed its power in dealing with biomedical image segmentation, many U-Net variations with skip-connection have been proposed for a better feature aggregation. Those methods explored the feature aggregation of the encoder and decoder to integrate global and local features (Cao et al., 2022; Ji et al., 2020; Liu et al., 2019; Qin et al., 2020; Qu et al., 2019; Sundaresan et al., 2021; Xu et al., 2021; Yu et al., 2023; Zhao et al., 2020). The skip connections, however, may introduce feature incompatibilities (Ibtehaz & Rahman, 2020) and bring in discrepancies throughout the propagation.

Regarding none-U-shape based architectures, Zheng et al. (2019) proposed an ensemble learning (Yan, Guo and Liu, 2023) framework for 3D biomedical image segmentation that combined the merits of 2D and 3D models. Zhang, Xie, Xia, and Shen (2019) proposed an attention residual learning CNN model (ARL-CNN) to leverage features at difficult stages. Li et al. (2022) introduced an automated segmentation network known as NPCNet, which comprises a position enhancement module (PEM), a scale enhancement module (SEM), and a boundary enhancement module (BEM) for the segmentation of primary nasopharyngeal carcinoma tumors and metastatic lymph nodes. Jiang et al. (2023) presented a de-overlapping network (DoNet) within a decompose-and-recombined strategy. They aggregated rich semantic features for both overlapping and non-overlapping regions using fusion units for cytology instance segmentation. Yang and Farsiu (2023) proposed a directional connectivity-based segmentation network (DconnNet) designed to separate the directional subspace from the shared latent space. They then employed the extracted directional features to enhance the overall data representation. Considering that harnessing pseudo masks and multi-scale features to differentiate nuclei/glands from the background could benefit segmentation, we aim to propose a pseudo-mask guided feature aggregation approach. Different from previous feature aggregation work (Ji et al., 2020; Liu et al., 2019; Qin et al., 2020; Qu et al., 2019; Xiang et al., 2020, 2022; Zhao et al., 2020), the proposed feature aggregation incorporates pseudo-mask as guidance to attentively aggregate multi-scale features in different stages.

2.3. Semi-supervised biomedical image segmentation

In biomedical image segmentation, different types of semi-supervised techniques have been proposed, which can be roughly categorized in four categories: (1) Consistency regularization (Luo et al., 2022; Tarvainen & Valpola, 2017; Wang et al., 2020; Wu, Ge et al., 2022; Xu et al., 2023; Yu et al., 2019). Learning from consistency can be regarded as learning from stability under perturbations. Numerous consistency-based methodologies have been proposed, deriving supplementary supervisory cues from unannotated data. (2) Use of pseudo label (Bai et al., 2023; Chen et al., 2021; Li, Chen et al., 2020; Li, Yu et al., 2020; Zhao et al., 2023; Zheng et al., 2020). Several methods have achieved robust representations through the process of acquiring supplementary information in the form of pseudo-labels. Nevertheless, owing to the inadequate class separability within the feature space, these pseudo-labels may contain potential noise and provide ambiguous guidance. (3) Adversarial learning (Lei et al., 2022; Xu et al., 2022; Zhang et al., 2017). The adversarial learning model aligns the distributions of segmented objects across different patients, ensuring robust predictions. However, the training process can be laborious and prone to instability. (4) Contrastive learning (CL) (Basak & Yin, 2023; Chaitanya et al., 2023; Gu et al., 2022; Shi et al., 2022; Zhang et al., 2023). Recently proposed CL-based methods have demonstrated their effectiveness in distinguishing between similar and dissimilar regions within the feature space. However, it is worth noting that pixel-wise features can be challenging to differentiate, particularly in high-level feature maps with lower resolution. Among these methods, mean teacher architecture shows its learning ability and simple training procedure in consistency regularization. Hence, we introduce mean teacher based methods for a broad review.

Based on the mean teacher (Tarvainen & Valpola, 2017) architecture, Li, Yu et al. (2020) introduced a transformation consistent self-ensembling model for medical image segmentation. Yu et al. (2019) estimated the uncertainty with the Monte Carlo dropout for semi-supervised 3D left atrium segmentation. Zhou et al. (2020) constrained the teacher and the student networks under mask-guided feature distillation with a perturbation-sensitive sample mining mechanism for nuclei instance segmentation. Xu et al. (2023) selected the consistency targets to integrate informative complementary clues during training. Nevertheless, these approaches suffer from limitations. Without the ground truth, the teacher network does not always generate better performance than the student network, which leads to misguiding segmentation. These limitations motivate our approach. Our proposed solution leverages the uncertainties in the teacher-student architecture and constrains the inter- and intra-inconsistencies during training, so that the proposed consistency regularization strategy can be more robust.

2.4. Uncertainty estimation

For deep learning, Bayesian deep networks are widely used to measure prediction uncertainty (Gal & Ghahramani, 2016; Gustafsson, Danelljan, & Schon, 2020; Kendall & Gal, 2017; Nielsen & Jensen, 2009; Wang et al., 2023, 2021, 2022; Zheng, Xu, & Wei, 2022; Zhu, Bolsterlee, Chow, Song, & Meijering, 2023) because of its robustness

and effectiveness. For example, Kwon, Won, Kim, and Paik (2020) proposed to estimate the aleatoric and epistemic uncertainty in medical image classification using a Bayesian neural network. As transformation operations in data augmentation may have an impact on segmentation results, Wang et al. (2019) analyzed the effect of such transformations by introducing a test-time augmentation-based aleatoric uncertainty. Yu et al. (2019) generated the uncertainty map by utilizing 8 stochastic forward passes on the teacher model under random dropout. Li, Yu et al. (2020) estimated uncertainties by introducing a transformation-consistent regularization strategy when ensembling models. Li, Wang, Yu and Heng (2020) encouraged the student model to produce similar outputs as the exponential moving average (EMA) teacher model under small perturbation operations in order to eliminate the uncertainty predictions among the architecture. Wang et al. (2020) claimed that without ground truth for unlabeled data, it cannot be guaranteed that the teacher model can provide accurate predictions. Hence, they proposed a feature-uncertainty and segmentation-uncertainty estimation method (DUW) for left atrium (LA) segmentation. Wang et al. (2021, 2022) introduced a foreground and background reconstruction task, along with a signed distance field (SDF) prediction task. They investigated the mutual enhancement between these two auxiliary tasks using a mean teacher architecture. Furthermore, they developed a triple-uncertainty guided framework to extract more reliable knowledge from the teacher model for medical image segmentation. Zheng et al. (2022) introduced a double noise mean teacher self-ensembling model for semi-supervised 2D tumor segmentation. Wang et al. (2023) explored multi-scale information using a dual multi-scale mean teacher network for COVID-19 segmentation. Zhu et al. (2023) introduced a hybrid dual mean teacher (HD-Teacher) model that incorporates hybrid, semi-supervised, and multi-task learning techniques to achieve semi-supervised segmentation of MRI scans. In addition to the prediction variances from the teacher and student models, the intra-uncertainties caused by the hierarchical CNN architecture are neglected by previous work, which motivated our approach. Inspired by the works mentioned above, we leverage the intra-uncertainties in the teacher-student network by enforcing the small perturbations at extra feature-level in our novel segmentation model.

3. Methodology

3.1. Problem formulation

Given a labeled dataset $(\mathcal{X}_l, \mathcal{Y}_l) = \{(x_i, y_i)\}_{i=1}^{M}$ and an unlabeled dataset $\mathcal{X}_u = \{(x_i)\}_{i=1}^{N}$, where M is the number of images with known segmentation results, each image x_i has a corresponding segmented mask y_i, N denotes the number of unlabeled images, and $\mathcal{X}_u$ represents images without labeled masks during the training process. The segmentation task aims to learn a mapping function F from input images $\mathcal{X}$ to segmentation $\mathcal{Y}$. In semi-supervised learning, the parameters θ of F are optimized on the labeled and the unlabeled datasets as follows:

$$\min_{\theta} \sum_{i=1}^{M} L_{\text{seg}}\left(F\left(x_i|\theta\right), y_i\right) + \lambda L_{\text{c}}(\theta, (\mathcal{X}_l, \mathcal{Y}_l), \mathcal{X}_u), \tag{1}$$

where L_{seg} is the supervised loss function, L_{c} is the unsupervised consistency loss, and λ is a weighting factor to enforce the consistency between the two datasets. As discussed before, the intra-uncertainty of the student model is neglected in the recent teacher-student network during the learning process. Furthermore, without ground truth labels for unlabeled data, it is difficult to determine whether the teacher model provides more accurate results than the student model or not.

To address these issues, we propose an uncertainty modeling mechanism to measure the intra-model uncertainties within the student network and the inter-model uncertainties between the student and teacher networks. Accordingly, we propose an inter-uncertainty consistency loss (L_{inter}) and a new intra-uncertainty penalization term (L_{intra}), and the learning process and overall loss function in Eq. (1) are revised as:

$$\begin{aligned}&\min_{\theta} \sum_{i=1}^{M} L_{\text{seg}}\left(F\left(x_i|\theta\right), y_i\right) \\ &+ \lambda(t)(L_{\text{inter}}(\theta, (\mathcal{X}_l, \mathcal{Y}_l), \mathcal{X}_u) + \lambda_{\text{intra}} L_{\text{intra}}),\end{aligned} \tag{2}$$

where L_{inter} denotes the unsupervised consistency loss for minimizing the inter-uncertainty, L_{intra} represents the additional regularization, which incorporates intra-uncertainty into the optimization objective to model uncertainties, $\lambda(t)$ denotes the step-related ramp-up weight factor for the consistency loss, t represents the tth training step, and λ_{intra} is a weight factor to control the regularization.

3.2. Pseudo-mask guided feature aggregation network (PG-FANet)

For effectively improving learning ability, we propose a feature aggregation network for both supervised and semi-supervised learning processes. The architecture of PG-FANet is illustrated in Fig. 2. The PG-FANet consists of three major components, i.e., two-stage sub-networks, mask-guided feature enhancement (MGFE) modules, and multi-scale multi-stage feature aggregation (MMFA) modules. MGFE is to force the attention of the network to the ROI under the guidance of the semantic pseudo-mask at feature level. MMFA is designed to extract and aggregate multi-scale and multi-stage features simultaneously to avoid the problems of feature incompatibilities in the U-shape skip connections (Ibtehaz & Rahman, 2020). The final output of our PG-FANet is obtained by fusing the output of the second stage and the aggregated features.

3.2.1. Two-stage sub-networks and MGFE

As shown in Fig. 2, the first stage is for coarse pseudo-mask generation and the second stage is for refinement. The two-stage sub-networks follow the same architecture where each stage consists of four residual blocks (RB), an atrous spatial pyramid pooling (ASPP) (Chen, Papandreou, Kokkinos, Murphy, & Yuille, 2017) module, and a final convolutional layer. The output feature maps of the convolution block (CB) progressively flow to the second stage. Those features and the pseudo mask of a sub-network in an early stage are then concatenated as an input to a later stage sub-network. Afterwards, the mask-guided features are adapted by a 1×1 convolutional layer for further propagation as shown in Fig. 2(b). In this way, the pseudo masks can serve as feature selectors to extract features from object instances with various sizes and shapes.

3.2.2. MMFA

Given the extracted features at different scales, shapes, and densities from the two-stage sub-networks, we use the MMFA module for the aggregation of multi-scale and multi-stage features.

Multi-scale feature aggregation. Multi-scale feature aggregation (Fig. 2(c)) is to combine low-level features of each stage. The aggregation is formulated as Eq. (3). For the sth stage, the ith RB can be defined as $\phi_s^i(\cdot)$, where s is in $\{1, \ldots, S\}$ and i is in $\{1, \ldots, I\}$. The multi-scale feature aggregation process can be formulated as:

$$\mathbf{X_m} = \sum_{s=1}^{S} \sum_{i=1}^{I} \text{Up}\left(\delta\left(\mathcal{B}(\text{Conv}(\phi_s^i(\mathbf{X}_s^{i-1})))\right)\right), \tag{3}$$

where $\mathbf{X_m}$ represents the aggregated feature map, Up is the upsampling operation, δ denotes the parametric rectified linear unit (PReLU) (He, Zhang, Ren, & Sun, 2015), $\mathcal{B}$ denotes batch normalization (BN) (Ioffe & Szegedy, 2015), Conv represents the convolutional layer, and $\mathbf{X}_s^{i-1}$ denotes the output feature maps from the $(i-1)$th RB in the sth stage. In this work, S is set to 2, and I is set to 3. The multi-scale feature aggregation process reuses the mask-guided information and gains a better representation for further propagation.

Multi-stage feature aggregation. As the network goes deeper, low-level features and spatial information such as region boundaries may

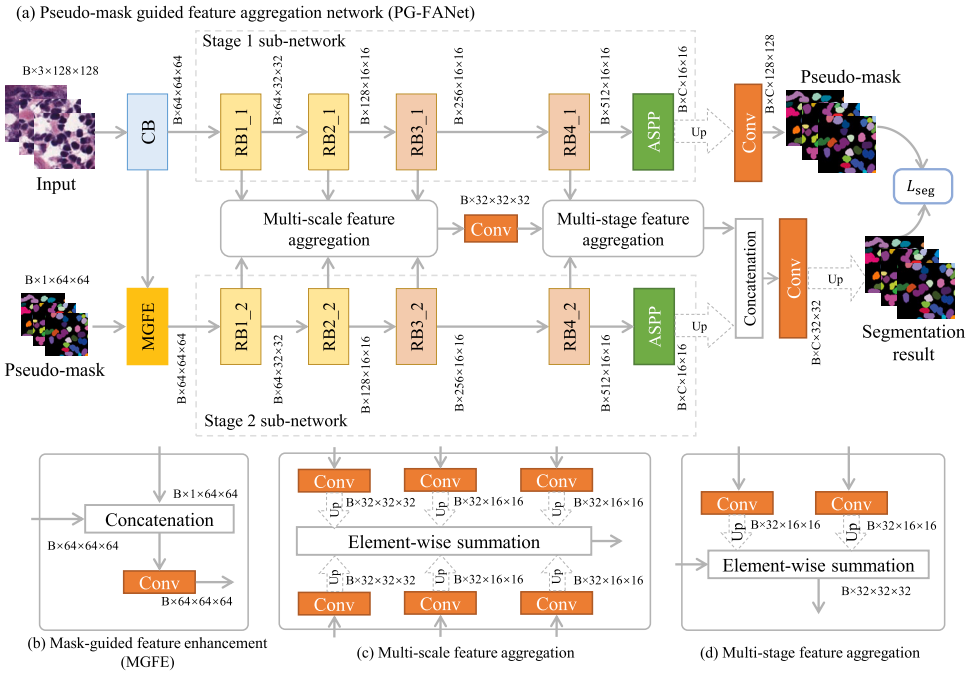


Fig. 2. Overview of our (a) PG-FANet with two-stage sub-networks, (b) mask-guided feature enhancement (MGFE) module, (c) multi-scale feature aggregation, and (d) multi-stage feature aggregation. Both stages share the same convolution block (CB). RBi_s denotes the ith residual blocks in stage s. RB3_s and RB4_s are dilated residual blocks with dilation rates of 2 and 4 respectively to generate feature maps with various receptive fields. Conv denotes the convolutional layer. ASPP denotes the atrous spatial pyramid pooling module. Up denotes the upsampling operation. The size of the output feature maps is given by batch size × channel size × height × width (B × C × H × W).

be lost (Li, Xiong, Fan, & Sun, 2019). Instead of introducing U-shape skip connections, which suffer from feature incompatibilities (Ibtehaz & Rahman, 2020), we fuse the early-stage features and the later-stage semantic features using the multi-stage feature aggregation module (Fig. 2(d)). The early-stage features before the ASPP and convolutional layers are more label-like ones, and the aggregation of those features would contribute to a more accurate and robust result. Formally, the multi-stage feature aggregation can be computed as:

$$\mathbf{X_h} = \mathbf{X'_m} + \sum_{s=1}^{S} \mathrm{Up}\left(\delta\left(\mathcal{B}(\mathrm{Conv}(\phi_s^4(\mathbf{X}_s^3)))\right)\right), \tag{4}$$

where $\mathbf{X'_m}$ denotes the fused features after the multi-scale aggregation operation, and $\mathbf{X_h}$ represents the output feature map after multi-stage aggregation.

3.2.3. Loss function for PG-FANet

As our PG-FANet has two-stage sub-networks, we optimize each stage by the standard cross-entropy loss L_{ce} and the Dice loss L_{Dice}. Hence, the main loss of each stage for supervised training is $L_{ce} + L_{Dice}$. The cross-entropy loss, Dice loss, or the combination of them are commonly used in many medical image segmentation methods and achieve remarkable success (Wong, Moradi, Tang, & Syeda-Mahmood, 2018). However, histopathology image segmentation requires not only to segment the nuclei/glands from the background, but also to separate each individual object from others. To consider the spatial relationship between objects, we also exploit the variance constrained cross (vcc) (Qu et al., 2019) loss as an auxiliary loss in the final stage for refined segmentation. Formally, given a minibatch $B \subseteq \mathcal{X}_l$, the L_{vcc} loss is defined as (Qu et al., 2019):

$$L_{vcc} = \frac{1}{D}\sum_{d=1}^{D}\frac{1}{|B_d|}\sum_{i=1}^{|B_d|}\left(\mu_d - p^i\right)^2 \tag{5}$$

where D is the number of instances in B, B_d denotes all the pixels that belong to instance d in the minibatch, $|B_d|$ represents the number of pixels in B_d, μ_d is the mean value of the probabilities of all the pixels in B_d, and p^i is the probability of the correct class for a pixel i in the final stage. In summary, L_{seg} in Eq. (2) for supervised segmentation is defined as:

$$L_{seg} = L_{ce} + \lambda_{Dice} L_{Dice} + \lambda_{vcc} L_{vcc}, \tag{6}$$

where λ_{Dice} and λ_{vcc} are parameters for adjusting weights.

3.3. Inter- and intra-uncertainty and consistency regularization

In a teacher-student framework, the teacher model is to evaluate images under perturbations for better performances (Li, Wang et al., 2020; Li, Yu et al., 2020). The weights θ'_t of the teacher model at training step t are updated as $\theta'_t = \alpha\theta'_{t-1} + (1-\alpha)\theta_t$ by the exponential moving average (EMA) weights of the student model F with weights θ_t. α is the EMA decay rate to update the weights of the student model with gradient descent in a total of T training steps. Although these methods model the uncertainty between the student and the teacher models to some extent, the uncertainty within the student network is neglected. Due to the hierarchical architecture of the convolutional networks, the features heavily rely on those from previous layers. Thus, discrepancies may exist within the student model. Fig. 8 shows prediction inconsistencies that exist near region boundaries, and the network contains ambiguous predictions between two stages. Additionally, without ground truth for $\mathcal{X}_u$, accurate prediction by the teacher model cannot always be guaranteed (Wang et al., 2020). To address this issue, we propose a novel uncertainty and consistency regularization strategy to quantify inter- and intra-uncertainties and inconsistencies in the teacher-student architecture. We hypothesize that the inter- and intra-consistencies can provide stronger regularization, thereby enabling a better performed student network.

3.3.1. Inter- and intra-uncertainty estimation

To address the inconsistency issue, previous works (Li, Wang et al., 2020; Li, Yu et al., 2020) model the uncertainty via inter-prediction variance of the student and the teacher networks as:

$$U_{\text{inter}} = \sum_{i=1}^{N} \mathbb{E} \left\| F_{s2}\left(x_i|\theta\right) - F_{s2}\left(x_i|\theta'\right) \right\|^2, \tag{7}$$

where $F_{s2}\left(x_i|\theta\right)$ represents the output of the second stage of the student model, and $F_{s2}\left(x_i|\theta'\right)$ denotes the output of the second stage of the teacher model in our work. However, intra-discrepancy exists in the different stages within the student network because of the hierarchical architecture, which will result in inconsistent predictions from the output of different stages. To remedy discrepancies, the resulting predictions from the output of each stage must be extremely consistent. Thus, we additionally estimate the intra-uncertainty (U_{intra}) as:

$$U_{\text{intra}} = \sum_{i=1}^{N} \mathbb{E} \left\| F_{s1}\left(x_i|\theta\right) - F_{s2}\left(x_i|\theta\right) \right\|^2, \tag{8}$$

where $F_{s1}\left(x_i|\theta\right)$ denotes the output of the first stage of the student model as shown in Fig. 2(a). To enforce consistency and improve the robustness, we further introduce small perturbations to hidden features in the two-stage of the student model.

On the one hand, the supervised learning process continuously improves the ability of the two-stage student model according to Eq. (6). On the other hand, the semi-supervised learning process forces the final prediction of the student model to be consistent with that of the teacher model (Eq. (7)). At the same time, the student model pushes the prediction of the first stage to be consistent with its final prediction (Eq. (8)). With such operations, the prediction of the first stage could be more accurate, which benefits the final prediction.

3.3.2. Inter- and intra-consistency regularization

As aforementioned, the teacher model cannot be guaranteed to produce more accurate predictions than the student model. To dynamically prevent the teacher model from obtaining high uncertainty estimation, inspired by Wang et al. (2020), we introduce a learnable loss function for penalizing the uncertainty generated by the teacher model. Firstly, we calculate the uncertainty of the ith image in the unlabeled dataset as:

$$u_i' = -\hat{q}_{\text{tea}}^i \log \hat{q}_{\text{tea}}^i, \tag{9}$$

where $\hat{q}_{\text{tea}}^i$ is the probability prediction provided by the teacher model, and u_i' represents the rectified uncertainty of the prediction on the ith sample. Secondly, we dynamically adjust the prediction of the teacher model as follows:

$$\hat{q}_{\text{tea}}^{i\prime} = \left(1 - u_i'\right) \hat{q}_{\text{tea}}^i + u_i' \hat{q}_{\text{stu}}^i, \tag{10}$$

where $\hat{q}_{\text{tea}}^i$ and $\hat{q}_{\text{stu}}^i$ denote the final predictions of the teacher model and the student model respectively, $\hat{q}_{\text{tea}}^{i\prime}$ represents the learnable prediction logits of the teacher model. When the teacher model provides unreliable results (high uncertainty), $\hat{q}_{\text{tea}}^{i\prime}$ is approximating $\hat{q}_{\text{stu}}^i$. On the contrary, when the teacher model is confident (with low uncertainty), $\hat{q}_{\text{tea}}^{i\prime}$ remains the same with $\hat{q}_{\text{tea}}^i$, and certain predictions are provided as targets for the student model to learn from. Finally, we incorporate the inter- and intra-uncertainty (i.e., U_{inter} and U_{intra}) into the training objective. The loss function of inter- and intra-consistency regularization can be formulated as:

$$\begin{aligned} L_{\text{inter}} &= L_{\text{mse}}\left(F_{s2}(x_i|\theta), F_{s2}'(x_i|\theta')\right), \\ L_{\text{intra}} &= L_{\text{mse}}(F_{s1}(x_i|\theta), F_{s2}(x_i|\theta)), \end{aligned} \tag{11}$$

where $F_{s2}'(x_i|\theta')$ denotes the learnable prediction of the teacher model in Eq. (10), L_{mse} denotes the mean square error loss function, and L_{inter} and L_{intra} are the terms in Eq. (2).

3.3.3. Shape attention weighted consistency regularization

In addition to promoting the complete segmentation for histopathology images, we leverage shape information to enhance the attention to boundary regions for better segmentation prediction. The shape consistency enhancement is crucial, given the observation that the increase of uncertainty in medical images mainly comes from ambiguous boundary regions. In this regard, we propose an attentive shape weight (U_{shape}) to explicitly promote the contour-relevant predictions from the student and the teacher models. The U_{shape} can be formulated as:

$$\begin{aligned} u_{\text{shape}} &= \left\| \text{Softmax}(F_{s2}(x_i|\theta)) - \text{Softmax}(F_{s2}(x_i|\theta')) \right\|, \\ U_{\text{shape}} &= -u_{\text{shape}} \log u_{\text{shape}}. \end{aligned} \tag{12}$$

As the final stage produces more accurate predictions than the first stage, additional shape information could benefit final predictions. Thus, we fuse the shape attention to L_{inter}. Finally, the inter-consistency regularization with shape attention weights can be defined as:

$$L_{\text{inter}} = (1 + \sigma(U_{\text{shape}}))L_{\text{mse}}\left(F_{s2}(x_i|\theta), F_{s2}'(x_i|\theta')\right), \tag{13}$$

where σ denotes the min–max normalization to scale the uncertainty to [0,1]. In this way, differences in boundary predictions between student and teacher models can be weighted by a shape attention mechanism, allowing more targeted adjustments during training to preserve complete shapes of segmented objects.

4. Experiments

4.1. Datasets and pre-processing

MoNuSeg: The multi-organ nuclei segmentation dataset (Kumar et al., 2019) consists of 44 H&E stained histopathology images, which are collected from multiple hospitals. These images are of 1000 × 1000 pixel resolution. The training set contains 30 histopathological images with hand-annotated nuclei, while the test set consists of 14 images. We use the randomly sampled 27 images as training data. The remaining 3 images are used as validation data.

CRAG: There are 213 H&E CRA images with different cancer grades taken from 38 whole slide images (WSIs) in the colorectal adenocarcinoma gland (CRAG) dataset (Awan et al., 2017). The CRAG dataset is split into 173 training images, where 153 images are used for training and 20 images are used for validation, and 40 test images. Most of the images are of 1512 × 1516 pixel resolution with instance-level ground truth.

We crop patches from each training image using a sliding window. For MoNuSeg, the patch size is 128 × 128, resulting in 1728 patches. For CRAG, we extract 5508 patches with 480 × 480 pixels from the 153 images. We further perform online data augmentations including random scale, flip, rotation, and affine operations. All these training images are normalized by using the mean and standard deviation for the images in ImageNet (Deng et al., 2009).

4.2. Experimental settings and parameters

We implement PG-FANet in PyTorch. Our experiments are conducted using an NVIDIA GeForce RTX 3090 graphics card. The batch size is set as 16 for MoNuSeg and 4 for CRAG. The Adam optimizer is applied with a learning rate at 2.5×10^{-4}. The optimizer is used with a polynomial learning rate policy, where the initial learning rate is multiplied by $\left(1 - \frac{iter}{total\_iter}\right)^{power}$ with $power$ fixed at 0.9. The total number of training iterations is set to 300 $*$ ($iter\_per\_epoch$), which is equivalent to 300 epochs as introduced in Jin et al. (2022). λ_{Dice} and λ_{vcc} in Eq. (6) are both set to 1. λ_{intra} is empirically set to 1 in Eq. (2). The weight factor $\lambda(t)$ in Eq. (2) is calculated by a Gaussian ramp-up function $\lambda(t) = k * e^{(-5(1-t/T)^2)}$, as introduced in Li, Yu et al. (2020). T is set to be equivalent to the total training epoch 300, and k is set to 0.1/5.0 for the MoNuSeg/CRAG dataset empirically. The EMA decay rate α is set to 0.99 empirically. We use the ImageNet pre-trained ResNet34 (He, Zhang, Ren, & Sun, 2016) as backbone. All the performances are averages over 3 runs. For semi-supervised learning, we use the same settings as above, except for the growing percentage of training samples. The percentages of training images are 5% (1/8), 10% (3/15), 20% (5/31), and 50% (14/76) for nuclei/gland.

Table 1
Performance comparison of the proposed PG-FANet and state-of-the-art methods on the MoNuSeg dataset.

Method	F1	Dice	IoU	AJI	95HD	Params(M)
Micro-Net (Raza et al., 2019)	0.810	0.723	0.602	0.457	9.753	186.74
U2-Net (Qin et al., 2020)	0.886	0.813	0.704	0.598	6.747	**1.14**
R2U-Net (Alom, Yakopcic, Taha, & Asari, 2018)	0.866	0.824	0.718	0.593	6.127	39.09
LinkNet (Chaurasia & Culurciello, 2017)	0.892	0.825	0.718	0.614	6.093	11.53
FullNet (Qu et al., 2019)	0.897	0.827	0.722	0.625	5.876	1.78
BiO-Net (Xiang et al., 2020)	0.894	0.824	0.720	0.619	6.008	14.97
MedFormer (Gao et al., 2022)	0.891	0.829	0.725	0.629	5.951	28.07
HARU-Net (Chen et al., 2023)	0.895	0.829	0.723	0.624	5.964	44.08
ADS_UNet (Yang et al., 2023)	0.894	0.831	0.728	0.619	5.885	26.72
Micro-Net[a] (Raza et al., 2019)	–	0.819	0.696	–	–	186.74
M-Net[a] (Mehta & Sivaswamy, 2017)	–	0.813	0.686	–	–	1.56
R2U-Net[a] (Alom et al., 2018)	–	0.815	0.683	–	–	9.09
LinkNet[a] (Chaurasia & Culurciello, 2017)	–	0.767	0.625	–	–	11.53
FullNet[a] (Qu et al., 2019)	0.857	0.802	–	0.600	–	1.78
BiO-Net[a] (Xiang et al., 2020)	–	0.825	0.704	–	–	14.97
BiX-NAS[a] (Xiang et al., 2022)	–	0.822	0.699	–	–	–
PG-FANet	**0.900**	**0.839**	**0.736**	**0.645**	**5.420**	42.78

[a] Copied directly from original papers.

4.3. Comparison methods and evaluation metrics

4.3.1. Comparison methods

To demonstrate the effectiveness of our proposed method, we compare our approach with several fully- or semi-supervised segmentation models. We choose the following representative methods in natural scene segmentation, nuclei/gland segmentation, and other medical image segmentation applications.

Fully supervised methods: The fully-supervised methods can be categorized as U-Net based models and none-U-Net based models: (1) For U-Net based models, we choose R2U-Net (Alom et al., 2018), BiO-Net (Xiang et al., 2020), BiX-NAS (Xiang et al., 2022), M-Net (Mehta & Sivaswamy, 2017), HARU-Net (Chen et al., 2023), and ADS_UNet (Yang et al., 2023) for comparison. (2) For none-U-Net based models, LinkNet (Chaurasia & Culurciello, 2017), Micro-Net (Raza et al., 2019), FullNet (Qu et al., 2019), DCAN (Chen et al., 2016), MILD-Net (Graham et al., 2019), DSE (Xie et al., 2019), PRS^2 (Xie et al., 2020), and MedFormer (Gao et al., 2022) are selected.

Semi-supervised methods: The number of published peer-reviewed SSL methods for nuclei/gland segmentation is relatively small, and we choose Self-loop (Wang et al., 2020) and PRS^2 (Xie et al., 2020) for comparison. Apart from that, we find that the recently proposed consistency regularization based semi-supervised learning models in other medical image segmentation approaches can be adapted (e.g., UA-MT Yu et al., 2019, ICT Verma, Lamb, Kannala, Bengio, & Lopez-Paz, 2019, TCSM Li, Yu et al., 2020, DUW Wang et al., 2020, CPS Chen et al., 2021, URPC Luo et al., 2022, and MC-Net Wu, Ge et al., 2022).

4.3.2. Evaluation metrics

Evaluation measures for nuclei segmentation include F1-score (F1), intersection over union (IoU), average Dice coefficient (Dice), aggregated Jaccard index (AJI), and 95% Hausdorff distance (95HD) as introduced in Kumar et al. (2019), Liu et al. (2019), and Xiang et al. (2020). For gland segmentation, F1, object-level Dice coefficient ($Dice_{obj}$), object-level Hausdorff distance ($Haus_{obj}$), and 95% object-level Hausdorff distance ($95HD_{obj}$) are used for detailed evaluation as described in Chen et al. (2016), Graham et al. (2019), and Xie et al. (2019).

4.4. Comparison with fully supervised state-of-the-arts

Following existing literature, we firstly conduct experiments on PG-FANet with full supervision, termed as PG-FANet, on the MoNuSeg and CRAG datasets. On the one hand, we re-implement some of existing models on the MoNuSeg and CRAG datasets under the same experimental settings for fair comparisons. On the other hand, for those complex models without source codes, we directly copy the values from their original papers. All the re-implemented models follow the same data augmentation and training strategies.

4.4.1. Nuclei segmentation

On the MoNuSeg dataset, we evaluate the performance of our proposed model on the test data in comparison with state-of-the-art methods. As shown in Table 1, PG-FANet outperforms all the other models in terms of all the evaluation metrics under both the same and different experimental settings. With the same experimental setting, the improvement by 1.6% of our model is substantial compared with the second best model MedFormer (Gao et al., 2022) on AJI score, which is a key metric for nuclei segmentation. When comparing with models under different experimental settings, the improvement of our PG-FANet is also desirable on Dice and IoU scores. Due to the missing of several metrics in their original work, F1, AJI, and 95HD cannot be explicitly compared. Even with limited evaluation metrics, the experimental results demonstrate that our PG-FANet outperforms all the other models.

4.4.2. Gland segmentation

We evaluate the gland segmentation performance of PG-FANet with other methods including U-Net (Ronneberger et al., 2015), DCAN (Chen et al., 2016), MILD-Net (Graham et al., 2019), DSE (Xie et al., 2019), PRS^2 (Xie et al., 2020), MedFormer (Gao et al., 2022), HARU-Net (Chen et al., 2023), and ADS_UNet (Yang et al., 2023). As shown in Table 2, PG-FANet consistently achieves the best performance in terms of F1, $Dice_{obj}$, and $Haus_{obj}$ among all the models.

In summary, the experimental results demonstrate that our PG-FANet with MGFE and MMFA outperforms the state-of-the-art methods and improves the fully supervised segmentation results.

4.5. Segmentation results using limited amount of labeled data

The major advantage of our semi-supervised framework, denoted by PG-FANet-SSL, is to use easily available unlabeled images to facilitate model training, leading to (1) less requirement on training data with annotations and (2) substantially improved segmentation performance, particularly when the number of images in the densely annotated training dataset is small. To evaluate the performance of PG-FANet-SSL, we conduct experiments by gradually increasing the proportion of labeled data. We also compare our framework with recent semi-supervised models including the mean teacher model (MT) (Tarvainen & Valpola,

Table 2
Performance comparison of the proposed PG-FANet and state-of-the-art methods on the CRAG dataset.

Method	F1	$Dice_{obj}$	$Haus_{obj}$	$95HD_{obj}$	Params (M)
U-Net (Ronneberger et al., 2015)	0.733	0.832	188.031	160.567	3.35
MedFormer (Gao et al., 2022)	0.813	0.885	118.204	96.861	28.07
HARU-Net (Chen et al., 2023)	0.841	0.875	137.774	107.728	44.08
ADS_UNet (Yang et al., 2023)	0.749	0.835	164.886	143.721	26.72
DCAN[a] (Chen et al., 2016)	0.736	0.794	218.760	–	**1.75**
MILD-Net[a] (Graham et al., 2019)	0.825	0.875	160.140	–	55.69
DSE[a] (Xie et al., 2019)	0.835	0.889	120.127	–	–
PRS^2[a] (Xie et al., 2020)	0.843	0.892	113.100	–	–
PG-FANet	**0.860**	**0.901**	**102.683**	**80.181**	42.78

[a] Copied directly from original papers.

Table 3
Experimental results on MoNuSeg and CRAG using our PG-FANet and state-of-the-art semi-supervised learning methods with different percentages of labeled data. Note that 5% (1/8) denotes 5% labeled data and the corresponding number of labeled data is 1/8 for MoNuSeg/CRAG dataset.

Labeled data	Methods	MoNuSeg					CRAG			
		F1	Dice	IoU	AJI	95HD	F1	$Dice_{obj}$	$Haus_{obj}$	$95HD_{obj}$
5% (1/8)	PG-FANet	0.822	0.767	0.646	0.505	8.998	0.718	0.773	246.130	208.665
	MT (Tarvainen & Valpola, 2017)	0.785	0.790	0.677	0.486	8.125	0.735	0.827	168.896	145.527
	UA-MT (Yu et al., 2019)	0.804	0.791	0.678	0.498	8.090	0.719	0.804	194.210	162.306
	ICT (Verma et al., 2019)	0.789	0.793	0.680	0.495	7.846	0.646	0.785	226.636	197.786
	TCSM (Li, Yu et al., 2020)	0.800	0.794	0.681	0.502	7.874	0.759	0.825	176.449	149.699
	DUW (Wang et al., 2020)	0.815	0.750	0.630	0.465	10.913	0.682	0.803	195.196	153.082
	CPS (Chen et al., 2021)	**0.844**	0.767	0.644	0.535	7.018	0.723	0.828	174.738	150.858
	URPC (Luo et al., 2022)	0.811	0.787	0.677	0.503	8.351	0.492	0.643	376.769	324.264
	MC-Net (Wu, Ge et al., 2022)	0.836	0.800	0.689	0.548	7.405	0.527	0.652	364.860	309.034
	PG-FANet-SSL	0.837	**0.809**	**0.698**	**0.564**	**6.641**	**0.807**	**0.869**	**136.498**	**112.232**
10% (3/15)	PG-FANet	0.874	0.798	0.686	0.580	7.071	0.770	0.821	192.773	169.371
	MT (Tarvainen & Valpola, 2017)	0.881	0.817	0.708	0.598	6.299	0.802	0.866	139.056	118.682
	UA-MT (Yu et al., 2019)	0.880	0.818	0.710	0.598	6.285	0.802	0.857	143.842	125.867
	ICT (Verma et al., 2019)	0.879	0.817	0.709	0.597	6.272	0.799	0.864	138.699	120.673
	TCSM (Li, Yu et al., 2020)	0.883	**0.819**	0.711	0.602	6.181	0.806	0.857	149.408	127.688
	DUW (Wang et al., 2020)	0.885	0.807	0.697	0.591	6.941	0.790	0.853	163.726	131.580
	CPS (Chen et al., 2021)	0.883	0.807	0.696	0.599	**6.105**	0.771	0.862	150.635	123.897
	URPC (Luo et al., 2022)	0.876	**0.819**	**0.713**	0.596	6.179	0.691	0.794	206.021	174.280
	MC-Net (Wu, Ge et al., 2022)	0.866	0.817	0.710	0.588	6.473	0.543	0.690	335.600	281.984
	PG-FANet-SSL	**0.886**	0.813	0.703	**0.609**	6.416	**0.822**	**0.878**	**116.901**	**97.906**
20% (5/31)	PG-FANet	0.884	0.811	0.702	0.602	6.726	0.828	0.870	128.166	110.100
	MT (Tarvainen & Valpola, 2017)	0.887	0.822	0.716	0.609	6.182	0.836	0.886	116.070	**96.658**
	UA-MT (Yu et al., 2019)	0.887	0.826	**0.721**	0.614	6.000	0.834	0.884	118.373	99.370
	ICT (Verma et al., 2019)	0.885	0.823	0.717	0.611	6.089	0.821	0.877	122.433	105.343
	TCSM (Li, Yu et al., 2020)	0.888	0.824	0.717	0.614	6.078	0.817	0.873	127.817	107.789
	DUW (Wang et al., 2020)	0.887	0.822	0.715	0.616	6.146	**0.837**	0.876	125.471	101.802
	CPS (Chen et al., 2021)	**0.891**	0.820	0.712	0.625	**5.694**	0.793	0.869	139.266	115.265
	URPC (Luo et al., 2022)	0.888	**0.832**	0.729	0.621	5.830	0.736	0.814	193.536	161.891
	MC-Net (Wu, Ge et al., 2022)	0.872	0.822	0.716	0.593	6.484	0.629	0.741	283.179	236.890
	PG-FANet-SSL	**0.891**	0.825	0.718	**0.629**	5.717	0.819	**0.888**	**112.694**	96.968
50% (14/76)	PG-FANet	0.884	0.809	0.700	0.598	6.731	0.836	0.887	123.998	95.639
	MT (Tarvainen & Valpola, 2017)	**0.891**	0.827	0.722	0.622	5.900	0.843	0.883	121.485	95.565
	UA-MT (Yu et al., 2019)	0.885	0.829	0.723	0.620	5.902	0.816	0.880	120.328	101.421
	ICT (Verma et al., 2019)	0.889	0.828	0.723	0.622	5.865	**0.846**	0.871	143.577	112.132
	TCSM (Li, Yu et al., 2020)	0.888	0.827	0.722	0.621	5.913	0.844	0.886	120.630	96.050
	DUW (Wang et al., 2020)	0.887	0.804	0.691	0.594	6.784	0.843	0.886	114.495	99.170
	CPS (Chen et al., 2021)	0.886	0.826	0.720	**0.630**	**5.555**	0.834	0.888	122.204	96.662
	URPC (Luo et al., 2022)	0.886	**0.834**	**0.732**	0.625	5.573	0.725	0.830	166.954	145.954
	MC-Net (Wu, Ge et al., 2022)	0.880	0.807	0.696	0.595	6.245	0.634	0.774	222.888	194.171
	PG-FANet-SSL	0.890	0.826	0.720	0.626	5.777	0.822	**0.889**	**112.926**	**93.775**
100% (27/153)	PG-FANet	0.900	0.839	0.736	0.645	5.420	0.860	0.901	102.683	80.181

Table 4
Experimental results on MoNuSeg and CRAG using our PG-FANet and state-of-the-art semi-supervised learning methods for nuclei/gland segmentation with different percentages of labeled data.

Labeled data	Methods	MoNuSeg	CRAG	
		F1	F1	$Dice_{obj}$
20%	Self-loop (Wang et al., 2020)	0.771	–	–
	PRS^2 (Xie et al., 2020)	–	0.807	0.850
	PG-FANet-SSL	**0.891**	**0.819**	**0.888**
50%	Self-loop (Wang et al., 2020)	0.791	–	–
	PRS^2 (Xie et al., 2020)	–	**0.823**	0.870
	PG-FANet-SSL	**0.890**	0.822	**0.889**

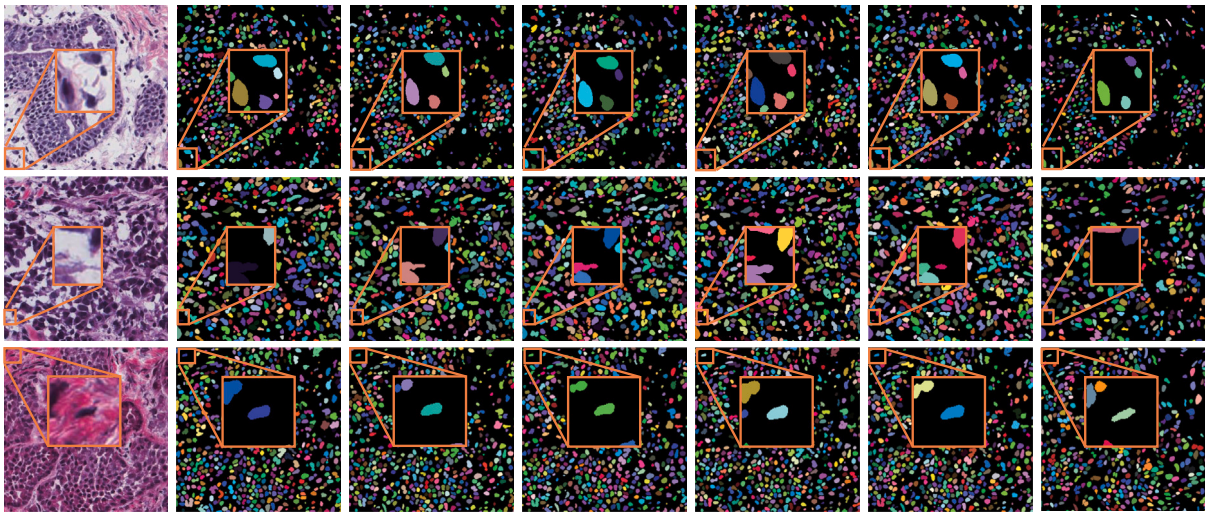


Fig. 3. Segmentation results on the MoNuSeg dataset with each add-on component in PG-FANet-SSL using 5% labeled data.

2017), uncertainty-aware self-ensembling model (UA-MT) (Yu et al., 2019), interpolation consistency training model (ICT) (Verma et al., 2019), transformation-consistent self-ensembling model (TCSM) (Li, Yu et al., 2020), double-uncertainty weighted model (DUW) (Wang et al., 2020), cross pseudo supervision (CPS) (Chen et al., 2021), uncertainty rectified pyramid consistency (URPC) (Luo et al., 2022), and mutual consistency learning (MC-Net) (Wu, Ge et al., 2022). For URPC and MC-Net, we utilize the original backbone network instead of PG-FANet, as it is not compatible with these methods. All other SSL methods mentioned above are re-implemented using PG-FANet as backbone and conducted under the same settings. Apart from the above methods, we compare our framework with the recent semi-supervised models, i.e., Self-loop (Wang et al., 2020) and PRS^2 (Xie et al., 2020), for nuclei/gland segmentation as well. We directly copy the values from the Self-loop and PRS^2 papers which we cannot obtain the source codes.

4.5.1. Nuclei segmentation

As shown in Tables 3 and 4, our PG-FANet-SSL yields significant improvements over PG-FANet using the same proportion of labeled data. Specifically, with the increasing number of labeled images, our model shows 5.9%, 2.9%, 2.7%, and 2.8% improvements on AJI compared with fully supervised training. It is noted that the AJI score has an obvious improvement when the labeled data proportion increases from 5% to 50%, which reveals that the increasing number of the labeled data has a significant impact when there is only a small amount of data with annotations. Furthermore, our PG-FANet-SSL method achieved overall better performances on AJI over all the semi-supervised learning models in comparison. It is interesting that with 5% labeled data, the other SSL methods gain marginal performances compared to the baseline PG-FANet. While with the increasing of labeled data, the other SSL methods show their learning ability on unlabeled data and achieve obvious improvements. We explain the finding as that our PG-FANet-SSL obtains more robust results with the inter- and intra-uncertainty regularization than the other methods.

4.5.2. Gland segmentation

The performance of our semi-supervised method is further demonstrated using the CRAG dataset as shown in Tables 3 and 4. Compared with the fully supervised baseline, PG-FANet, our PG-FANet-SSL method significantly improves F1, Dice_{obj}, and Haus_{obj} by 8.9%/5.2%, 9.6%/5.7%, and 109.632/75.872 respectively when using 5%/10% of labeled data. When compared with state-of-the-art methods, our PG-FANet-SSL shows significant improvement with only 5%/10% labeled data for training.

In summary, our first finding is that with the increasing number of labeled data, the performance of segmentation improves steadily. Second, our proposed PG-FANet-SSL achieves better performances compared to the fully supervised PG-FANet when it is trained with the same proportion of labeled data. Third, our PG-FANet-SSL maintains competitive performance when using 50% labeled data compared with other fully supervised methods with 100% labeled data. Fourth, PG-FANet-SSL outperforms the recent state-of-the-art semi-supervised learning methods especially with 5%, 10%, and 20% labeled data for training, and this shows the effectiveness of our uncertainty modeling strategy. Last but not the least, it is interesting that the performance discrepancy between semi-supervised and supervised learning methods becomes marginal with the increase in the number of labeled data. We explain this finding as the reason that the diversity of these two datasets is limited, and a whole histopathology image may contain a certain number of labeled nuclei/gland instances. Thus, only a limited number of labeled data is needed for training to obtain state-of-the-art performance on nuclei/gland segmentations.

4.6. Ablation studies

4.6.1. Effectiveness of MGFE and MMFA

We perform ablation studies to evaluate the contributions of different components in our framework. We first remove all the components, degrade the two-stage network to a single-stage network (i.e., DeepLabV2 Chen et al., 2017 with an extra convolutional layer), and gradually add proposed components (i.e., mask-guided, multi-scale, and multi-stage) to the model. As shown in Table 5, the overall performance evaluation metric, AJI score, increases by 0.8% on MoNuSeg when we add one more stage to DeepLabV2, which indicates that the extra backbone cannot improve the ability of the model. The AJI metric, however, increases by 1.4% when the mask-guided feature enhancement module is introduced to the two-stage network. The MGFE module utilizes the coarse segmentation results as an enhancement and finally improves the learning ability of the model. Improvement can be observed with the application of multi-scale and multi-stage feature aggregation modules.

4.6.2. Effectiveness of inter- and intra-uncertainty and consistency regularization

We use 5% labeled NoNuSeg and CRAG to demonstrate the effectiveness of the uncertainty and consistency regularization. Table 6

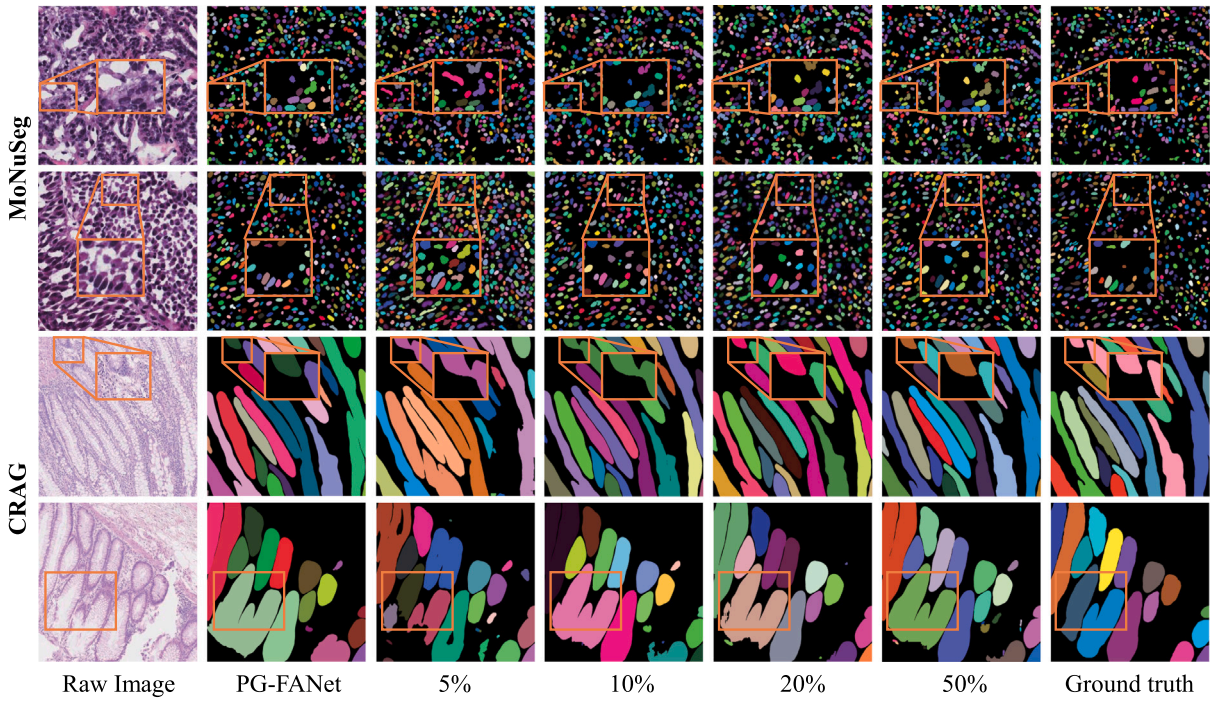


Fig. 4. Segmentation results on the MoNuSeg and CRAG datasets using our fully supervised PG-FANet with 100% labeled data and semi-supervised learning with 5%, 10%, 20%, and 50% of the labeled data.

Table 5
Effectiveness analysis of different modules in PG-FANet on MoNuSeg using 100% labeled data.

Two- stage	Mask-guided	Multi- scale	Multi- stage	F1	Dice	IoU	AJI	95HD
✗	✗	✗	✗	0.876	0.807	0.698	0.589	7.020
✓	✗	✗	✗	0.882	0.809	0.701	0.597	6.826
✓	✓	✗	✗	0.886	0.821	0.715	0.615	6.173
✓	✓	✓	✗	0.896	0.837	0.734	0.640	5.531
✓	✓	✓	✓	0.900	0.839	0.736	0.645	5.420

Two-stage: Two-stage sub-networks, **Mask-guided**: Mask-guided feature enhancement, **Multi-scale**: Multi-scale feature aggregation, **Multi-stage**: Multi-stage feature aggregation.

Table 6
Effectiveness analysis of regularizations in PG-FANet-SSL on MoNuSeg and CRAG using 5% labeled data.

Inter-	Intra-	Shape-	MoNuSeg					CRAG			
			F1	Dice	IoU	AJI	95HD	F1	$Dice_{obj}$	$Haus_{obj}$	$95HD_{obj}$
✗	✗	✗	0.822	0.767	0.646	0.505	8.998	0.718	0.773	246.130	208.665
✓	✗	✗	0.801	0.797	0.684	0.520	7.437	0.796	0.858	148.080	123.161
✗	✓	✗	0.818	0.796	0.683	0.524	7.620	0.764	0.860	140.692	119.236
✗	✗	✓	0.782	0.794	0.680	0.495	7.726	0.775	0.850	168.771	131.135
✓	✓	✗	0.826	0.803	0.691	0.547	7.072	0.797	0.861	145.948	116.866
✓	✓	✓	0.837	0.809	0.698	0.564	6.641	0.807	0.869	136.498	112.232

Inter-: Inter-consistency regularization, **Intra-**: Intra-consistency regularization, **Shape-**: Shape attention weighted consistency regularization.

presents the ablation studies of our key components. On the one hand, adding the inter-consistency regularization strategy improves AJI/$Dice_{obj}$ metric by 1.5%/8.5% on the MoNuSeg/CRAG dataset. On the other hand, reducing intra-uncertainty increases the AJI/$Dice_{obj}$ metric by 1.9%/8.7% as well. Furthermore, shape attention weighted consistency regularization also preserves the complete shape of segmentation in medical images. To visually illustrate the effectiveness of each component on the MoNuSeg dataset, we present the segmentation results in Fig. 3. As depicted, the occurrences of false positive predictions significantly diminish with the aid of inter-uncertainty reduction (i.e., the third column in Fig. 3) compared to the baseline method (i.e., the second column in Fig. 3). This outcome underscores the profound impact of strategically reducing inter-uncertainty in achieving substantial performance advancements. Regarding intra-uncertainty reduction, the extent of false predictions is also smaller compared to those produced by the baseline, affirming the efficacy of intra-uncertainty and consistency regularization. As for the shape enhancement component, the false predictions become slightly pronounced, potentially due to the smaller size and ambiguous boundaries of nuclei compared to glands. Nevertheless, the incorporation of all the consistency regularization strategies mitigates the challenge of enforcing appearance consistency, ultimately resulting in improved performance.

In summary, Table 6 and Fig. 3 indicate that (1) inconsistency within the student model exists whilst our PG-FANet-SSL approach can model the inconsistency in a better manner, (2) without the inter-uncertainty strategy, AJI/$Dice_{obj}$ continues to decrease since the inter-uncertainty strategy can dynamically leverage the uncertainty obtained by the teacher model, and (3) additional boundary information and shape enhancement benefit complete object segmentation for histopathology images.

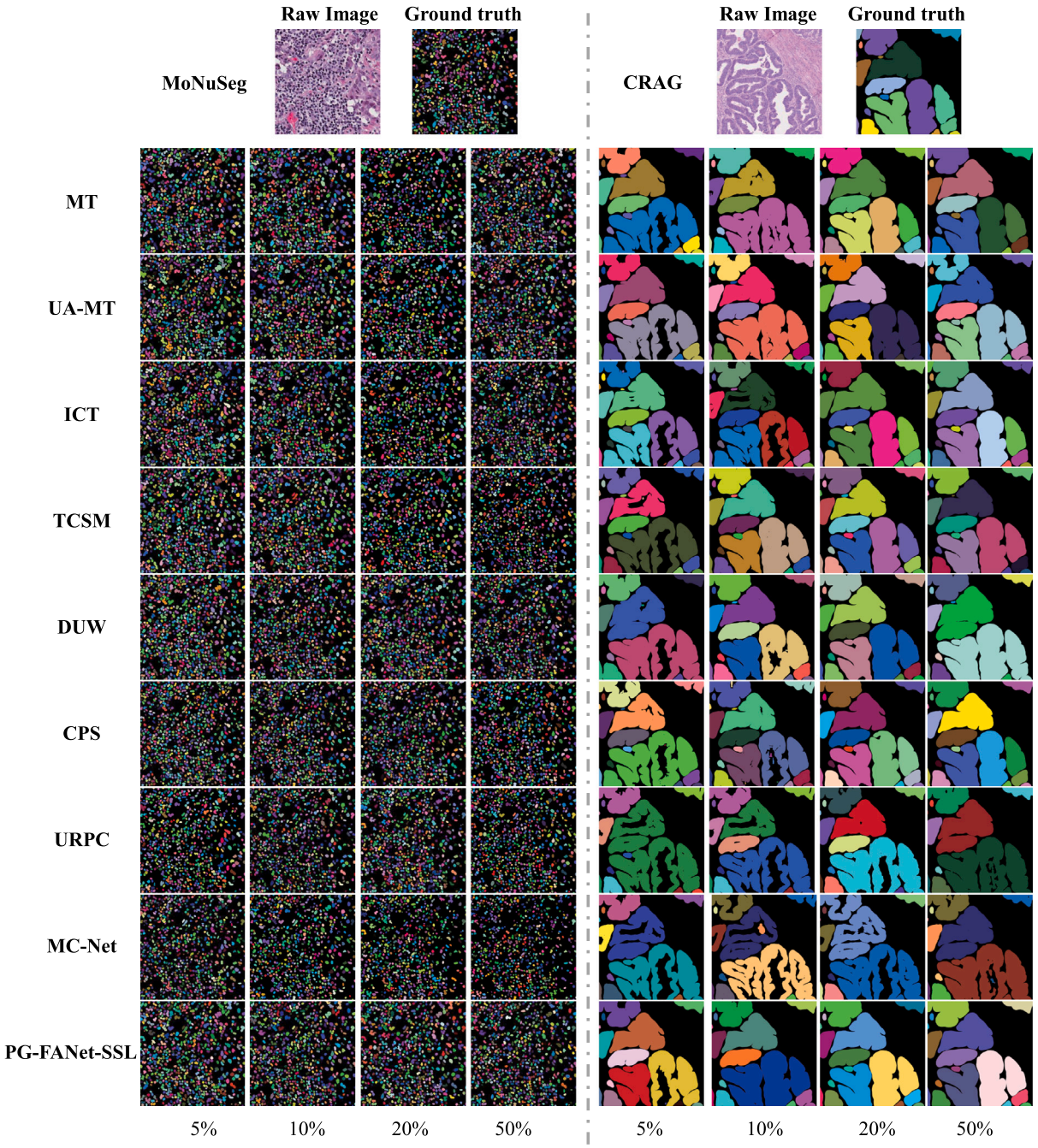


Fig. 5. A representative segmentation outcome achieved with our method on the MoNuSeg and CRAG datasets, compared with results from other state-of-the-art approaches.

4.7. Qualitative results

4.7.1. Segmentation visualization

Qualitative results on the nuclei and gland segmentations via full supervision and semi-supervision are shown in Fig. 4. Compared with the fully supervised PG-FANet, the PG-FANet-SSL has a competitive confident prediction near object boundaries. With the increase in the proportion of labeled data, the regions of false predictions become smaller, and the boundaries of nuclei/glands become clearer. Furthermore, we typically visualize samples generated by our PG-FANet-SSL and other state-of-the-art methods in Fig. 5. As the proportion of labeled data increases, there is a noticeable enhancement in the performance of all the compared methods. This improvement can be attributed to the integration of a greater amount of valuable information into the learning process. Notably, PG-FANet-SSL stands out in terms of competitive performance when compared to all the other methods. It demonstrates clearer boundaries in the MoNuSeg and CRAG datasets, underscoring its efficacy in the context of these specific datasets.

We explain the finding as that uncertainty usually exists near object boundaries because of the subtle contrast between the foreground and background regions in histopathology images. The proposed PG-FANet-SSL framework enables the learning process to focus on such uncertainties, yielding more reliable segmentation results.

4.7.2. Uncertainty visualization

We also provide the visualization results to show the inter- and intra-uncertainty differences between PG-FANet and PG-FANet-SSL trained with 5% labeled data. As shown in Fig. 6, we observe that the SSL with inter- or shape-attention weighted consistency regularization provides more confident boundary predictions when compared with

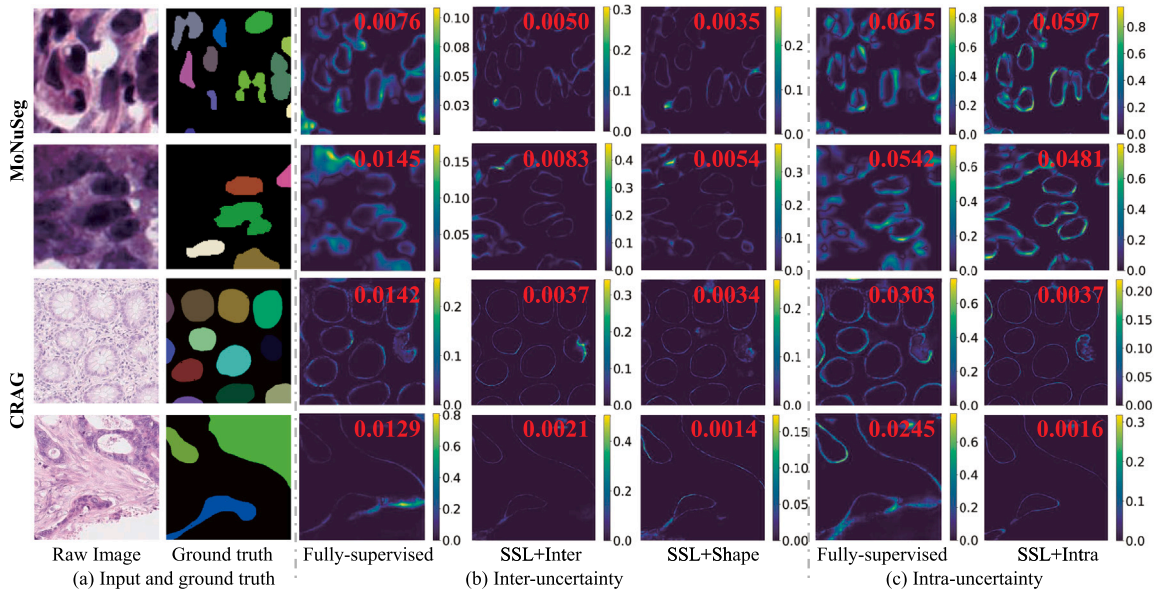


Fig. 6. Histopathology image, ground truth, inter- and intra-uncertainties of our fully-supervised method, and our semi-supervised method (denoted by SSL). The average prediction variance scores in red show the inconsistencies near object boundaries. It is noted that the uncertainties between the teacher and the student models are in (b), while the uncertainties between two-stage networks are in (c).

supervised PG-FANet. Moreover, SSL with the intra-consistency regularization mechanism also reduces the prediction discrepancies within the student model.

4.8. Discussion

4.8.1. Existence of inter-uncertainties

To explore the learning ability of the teacher model and the student model, we adopt the mean-teacher architecture and train the baseline model on 5% labeled data and 95% unlabeled data. The performances are depicted in Fig. 7 on the validation dataset. Our primary finding is that the teacher model gains desirable stability when compared with the student model. The teacher model, however, may not always generate better results than the student model, especially at the beginning of the training process. We explain this experimental finding by the reason that the weights of the teacher mode are updated by the EMA weights of the student model which demonstrates the existence of the inter-uncertainties. With the reduction of the inter-uncertainties and prediction discrepancy, the representation ability of the model could be improved.

4.8.2. Existence of the intra-uncertainties

To further show the intra-uncertainties, we depict the two-stage prediction variances in Fig. 8. The reason for such a prediction discrepancy is that different receptive fields introduce inconsistencies. As shown in Fig. 2, stage 1 is located at the relatively shallow layer, while stage 2 learns from deeper layers. The hierarchical architecture and different receptive fields increase the intra-uncertainties, which leads to prediction differences. Based on this finding, we enforce an invariance for predictions of the two stages over small perturbations applied to the hidden features. As a result, the learned model will be robust to such intra-uncertainties.

4.8.3. Effectiveness of loss functions

To demonstrate the performance gain obtained from the proposed segmentation loss (i.e., L_{seg}), we also attempt to train PG-FANet with different loss functions (i.e., L_{ce}, L_{Dice}, and L_{vcc}) and their combinations. The results in Table 7 reveal that, although using L_{ce} alone results in desirable performances on the MoNuSeg dataset, adding L_{Dice} and L_{vcc} individually to the segmentation loss improves the performance

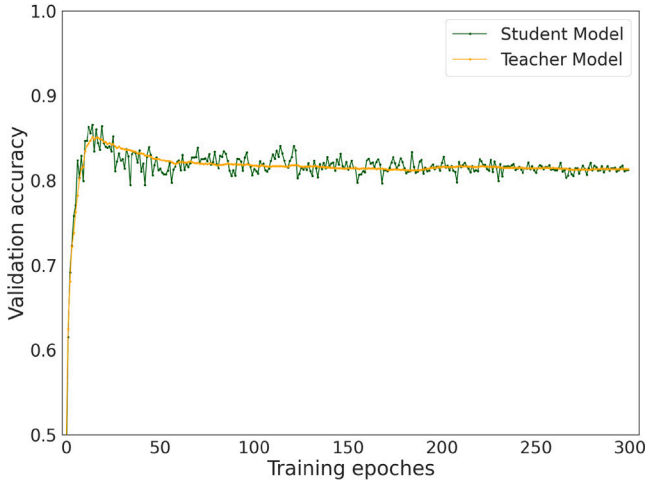


Fig. 7. Segmentation accuracy on the MoNuSeg validation dataset with 5% of the labeled data.

Table 7
Effectiveness analysis of loss functions on MoNuSeg using 100% labeled data.

Methods	F1	Dice	IoU	AJI	95HD
L_{ce}	0.887	0.824	0.715	0.625	5.456
$L_{ce}+L_{Dice}$	0.887	0.827	0.720	0.628	5.336
$L_{ce}+L_{vcc}$	0.891	0.834	0.731	0.635	5.665
$L_{ce}+L_{Dice}+L_{vcc}$	0.900	0.839	0.736	0.645	5.420

continuously. Meanwhile, the superior performance of our L_{seg} confirms the effectiveness of the combination of L_{ce}, L_{Dice}, and L_{vcc} loss functions, which have constraints on each individual instance.

4.8.4. The flexibility of the proposed algorithm

To demonstrate the flexibility of the proposed consistency regularization algorithm, we also degrade the two-stage network. In Table 8, PG-FANet and DeepLabV2 (Chen et al., 2017) denote fully supervised two-stage PG-FANet and degraded PG-FANet trained with only 5% labeled data, respectively. Furthermore, we adopt DeepLabV2 (Chen

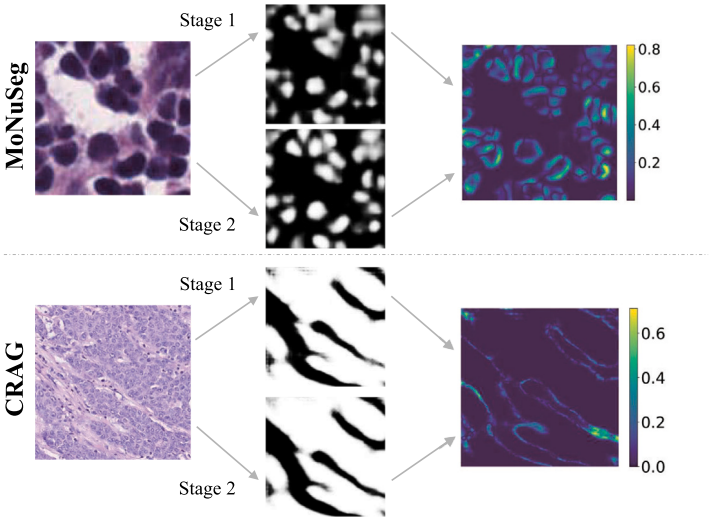


Fig. 8. Illustration of the intra-uncertainty within the student model. The segmentation probability maps of input (a) between two-stage networks are in (b). In (c), brighter colors indicate higher variance. Object boundaries are with higher uncertainties.

Table 8
Flexibility analysis of single-stage PG-FANet on MoNuSeg using 5% labeled data.

Methods	F1	Dice	IoU	AJI	95HD
PG-FANet	0.822	0.767	0.646	0.505	8.998
DeepLabV2 (Chen et al., 2017)	0.797	0.717	0.590	0.440	12.140
MT (Tarvainen & Valpola, 2017)	0.782	0.771	0.654	0.474	9.087
MT (Tarvainen & Valpola, 2017) + Inter	0.812	0.778	0.662	0.509	8.493
MT (Tarvainen & Valpola, 2017) + Shape	0.784	0.773	0.656	0.477	8.978
MT (Tarvainen & Valpola, 2017) + Inter+Shape	0.819	0.781	0.665	0.517	8.289

et al., 2017) as our supervised training backbone in the MT (Tarvainen & Valpola, 2017) architecture, and gradually add proposed components. It is noted that without the two-stage architecture, intra-consistency regularization is not applicable. As shown in Table 8, when the two-stage architecture is not adopted, the performance dramatically decreases from 50.5% to 44.0% on AJI. With the proposed consistency constraint and regularization, our model is substantially more accurate on nuclei segmentation. This experimental results reveal that (1) our two-stage network with MGFE and MMFA substantially improves the segmentation performance, (2) the proposed inter-consistency and shape attention weighted consistency regularization strategy also functions well on degraded models, and (3) the proposed module and learning strategy are simple yet flexible for extending to other models.

5. Conclusion

In this paper, we first propose a novel pseudo-mask guided feature aggregation network to leverage multi-scale and multi-stage features for histopathology image segmentation. Our semi-supervised learning framework models uncertainties in the mean teacher architecture by a novel inter- and intra-uncertainty and consistency regularization strategy. Comprehensive experimental results show that our semi-supervised segmentation model achieves competitive results for nuclei and gland segmentations using partially labeled data when compared with fully- or semi-supervised models. One limitation of our work pertains to the large number of parameters, primarily attributable to the selected backbone. Our future work includes the parameter reduction and extension to other 3D image segmentation tasks.

CRediT authorship contribution statement

Qiangguo Jin: Conducted the experiments. **Hui Cui:** Manuscript writing. **Changming Sun:** Designed the experiments, Edited the manuscript. **Yang Song:** Designed the experiments, Edited the manuscript. **Jiangbin Zheng:** Designed the experiments, Edited the manuscript. **Leilei Cao:** Designed the experiments, Edited the manuscript. **Leyi Wei:** Manuscript writing. **Ran Su:** Manuscript writing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work was supported by the Fundamental Research Funds for the Central Universities, China, the National Natural Science Foundation of China [Grant No. 62201460, No. 62222311 and No. 62072329], and the National Key Technology R&D Program of China [Grant No. 2018YFB1701700]. All authors approved the version of the manuscript to be published.

References

Alom, M. Z., Yakopcic, C., Taha, T. M., & Asari, V. K. (2018). Nuclei segmentation with recurrent residual convolutional neural networks based U-net (R2U-net). In *NAECON 2018-IEEE national aerospace and electronics conference* (pp. 228–233). IEEE.

Awan, R., Sirinukunwattana, K., Epstein, D., Jefferyes, S., Qidwai, U., Aftab, Z., et al. (2017). Glandular morphometrics for objective grading of colorectal adenocarcinoma histology images. *Scientific Reports*, *7*(1), 1–12.

Bai, Y., Chen, D., Li, Q., Shen, W., & Wang, Y. (2023). Bidirectional copy-paste for semi-supervised medical image segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 11514–11524).

Basak, H., & Yin, Z. (2023). Pseudo-label guided contrastive learning for semi-supervised medical image segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 19786–19797).

Cao, X., Chen, H., Li, Y., Peng, Y., Zhou, Y., Cheng, L., et al. (2022). Auto-DenseUNet: Searchable neural network architecture for mass segmentation in 3D automated breast ultrasound. *Medical Image Analysis*, *82*, Article 102589.

Chaitanya, K., Erdil, E., Karani, N., & Konukoglu, E. (2023). Local contrastive loss with pseudo-label based self-training for semi-supervised medical image segmentation. *Medical Image Analysis*, *87*, Article 102792.

Chaurasia, A., & Culurciello, E. (2017). LinkNet: Exploiting encoder representations for efficient semantic segmentation. In *2017 IEEE visual communications and image processing (VCIP)* (pp. 1–4). IEEE.

Chen, L.-C., Collins, M., Zhu, Y., Papandreou, G., Zoph, B., Schroff, F., et al. (2018). Searching for efficient multi-scale architectures for dense image prediction. In *Advances in neural information processing systems* (pp. 8699–8710).

Chen, J., Huang, Q., Chen, Y., Qian, L., & Yu, C. (2023). Enhancing nucleus segmentation with HARU-net: A hybrid attention based residual U-blocks network. arXiv preprint arXiv:2308.03382.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2017). DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *40*(4), 834–848.

Chen, H., Qi, X., Yu, L., & Heng, P.-A. (2016). DCAN: deep contour-aware networks for accurate gland segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2487–2496).

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. 2613–2622).

Cui, F., Li, S., Zhang, Z., Sui, M., Cao, C., Hesham, A. E.-L., et al. (2022). Deepmc-iNABP: Deep learning for multiclass identification and classification of nucleic acid-binding proteins. *Computational and Structural Biotechnology Journal*, *20*, 2020–2028.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248–255). IEEE.

Dou, Q., Ouyang, C., Chen, C., Chen, H., Glocker, B., Zhuang, X., et al. (2019). Pnp-AdaNet: Plug-and-play adversarial domain adaptation network at unpaired cross-modality cardiac segmentation. *IEEE Access*, *7*, 99065–99076. http://dx.doi.org/10.1109/ACCESS.2019.2929258.

Gal, Y., & Ghahramani, Z. (2016). Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In *International conference on machine learning* (pp. 1050–1059).

Gao, Y., Zhou, M., Liu, D., Yan, Z., Zhang, S., & Metaxas, D. N. (2022). A data-scalable transformer for medical image segmentation: architecture, model efficiency, and benchmark. arXiv preprint arXiv:2203.00131.

Graham, S., Chen, H., Gamper, J., Dou, Q., Heng, P.-A., Snead, D., et al. (2019). MILD-net: minimal information loss dilated network for gland instance segmentation in colon histology images. *Medical Image Analysis*, *52*, 199–211.

Gu, R., Zhang, J., Wang, G., Lei, W., Song, T., Zhang, X., et al. (2022). Contrastive semi-supervised learning for domain adaptive segmentation across similar anatomical structures. *IEEE Transactions on Medical Imaging*, *42*(1), 245–256.

Gustafsson, F. K., Danelljan, M., & Schon, T. B. (2020). Evaluating scalable Bayesian deep learning methods for robust computer vision. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops* (pp. 318–319).

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision* (pp. 1026–1034).

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778).

Ibtehaz, N., & Rahman, M. S. (2020). MultiResUNet: Rethinking the U-net architecture for multimodal biomedical image segmentation. *Neural Networks*, *121*, 74–87.

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167.

Javed, S., Mahmood, A., Fraz, M. M., Koohbanani, N. A., Benes, K., Tsang, Y.-W., et al. (2020). Cellular community detection for tissue phenotyping in colorectal cancer histology images. *Medical Image Analysis*, *63*, Article 101696.

Ji, Y., Zhang, R., Li, Z., Ren, J., Zhang, S., & Luo, P. (2020). Uxnet: Searching multi-level feature aggregation for 3D medical image segmentation. In *International conference on medical image computing and computer assisted intervention* (pp. 346–356). Springer.

Jiang, H., Zhang, R., Zhou, Y., Wang, Y., & Chen, H. (2023). Donet: Deep de-overlapping network for cytology instance segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 15641–15650).

Jin, Q., Cui, H., Sun, C., Zheng, J., Wei, L., Fang, Z., et al. (2022). Semi-supervised histological image segmentation via hierarchical consistency enforcement. In *International conference on medical image computing and computer-assisted intervention* (pp. 3–13). Springer.

Kendall, A., & Gal, Y. (2017). What uncertainties do we need in Bayesian deep learning for computer vision? In *Advances in neural information processing systems* (pp. 5574–5584).

Kumar, N., Verma, R., Anand, D., Zhou, Y., Onder, O. F., Tsougenis, E., et al. (2019). A multi-organ nucleus segmentation challenge. *IEEE Transactions on Medical Imaging*, *39*(5), 1380–1391.

Kwon, Y., Won, J.-H., Kim, B. J., & Paik, M. C. (2020). Uncertainty quantification using Bayesian neural networks in classification: Application to biomedical image segmentation. *Computational Statistics & Data Analysis*, *142*, Article 106816.

Lei, T., Zhang, D., Du, X., Wang, X., Wan, Y., & Nandi, A. K. (2022). Semi-supervised medical image segmentation using adversarial consistency learning and dynamic convolution network. *IEEE Transactions on Medical Imaging*.

Li, Y., Chen, J., Xie, X., Ma, K., & Zheng, Y. (2020). Self-loop uncertainty: A novel pseudo-label for semi-supervised medical image segmentation. In *International conference on medical image computing and computer assisted intervention* (pp. 614–623). Springer.

Li, Y., Dan, T., Li, H., Chen, J., Peng, H., Liu, L., et al. (2022). NPCNet: jointly segment primary nasopharyngeal carcinoma tumors and metastatic lymph nodes in MR images. *IEEE Transactions on Medical Imaging*, *41*(7), 1639–1650.

Li, K., Wang, S., Yu, L., & Heng, P.-A. (2020). Dual-teacher: Integrating intra-domain and inter-domain teachers for annotation-efficient cardiac segmentation. In *International conference on medical image computing and computer assisted intervention* (pp. 418–427). Springer.

Li, H., Xiong, P., Fan, H., & Sun, J. (2019). DFANet: Deep feature aggregation for real-time semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 9522–9531).

Li, X., Yu, L., Chen, H., Fu, C.-W., Xing, L., & Heng, P.-A. (2020). Transformation-consistent self-ensembling model for semi-supervised medical image segmentation. *IEEE Transactions on Neural Networks and Learning Systems*, 1–12.

Liu, D., Zhang, D., Song, Y., Zhang, C., Zhang, F., O'Donnell, L., et al. (2019). Nuclei segmentation via a deep panoptic model with semantic feature fusion. In *Proceedings of the Twenty-Eighth international joint conference on artificial intelligence, IJCAI-19* (pp. 861–868). http://dx.doi.org/10.24963/ijcai.2019/121.

Lu, C., Koyuncu, C., Corredor, G., Prasanna, P., Leo, P., Wang, X., et al. (2021). Feature-driven local cell graph (flock): new computational pathology-based descriptors for prognosis of lung cancer and HPV status of oropharyngeal cancers. *Medical Image Analysis*, *68*, Article 101903.

Luo, X., Wang, G., Liao, W., Chen, J., Song, T., Chen, Y., et al. (2022). Semi-supervised medical image segmentation via uncertainty rectified pyramid consistency. *Medical Image Analysis*, *80*, Article 102517.

Mehta, R., & Sivaswamy, J. (2017). M-net: A convolutional neural network for deep brain structure segmentation. In *2017 IEEE 14th international symposium on biomedical imaging (ISBI 2017)* (pp. 437–440). IEEE.

Nielsen, T. D., & Jensen, F. V. (2009). *Bayesian networks and decision graphs*. Springer Science & Business Media.

Qin, X., Zhang, Z., Huang, C., Dehghan, M., Zaiane, O. R., & Jagersand, M. (2020). U2-net: Going deeper with nested U-structure for salient object detection. *Pattern Recognition*, *106*, Article 107404.

Qu, H., Yan, Z., Riedlinger, G. M., De, S., & Metaxas, D. N. (2019). Improving nuclei/gland instance segmentation in histopathology images by full resolution neural network and spatial constrained loss. In *International conference on medical image computing and computer assisted intervention* (pp. 378–386). Springer.

Raza, S. E. A., Cheung, L., Shaban, M., Graham, S., Epstein, D., Pelengaris, S., et al. (2019). Micro-net: A unified model for segmentation of various objects in microscopy images. *Medical Image Analysis*, *52*, 160–173.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International conference on medical image computing and computer assisted intervention* (pp. 234–241). Springer.

Shi, J., Gong, T., Wang, C., & Li, C. (2022). Semi-supervised pixel contrastive learning framework for tissue segmentation in histopathological image. *IEEE Journal of Biomedical and Health Informatics*, *27*(1), 97–108.

Su, H., Xing, F., Kong, X., Xie, Y., Zhang, S., & Yang, L. (2015). Robust cell detection and segmentation in histopathological images using sparse reconstruction and stacked denoising autoencoders. In *International conference on medical image computing and computer assisted intervention* (pp. 383–390). Springer.

Sundaresan, V., Zamboni, G., Rothwell, P. M., Jenkinson, M., & Griffanti, L. (2021). Triplanar ensemble U-net model for white matter hyperintensities segmentation on MR images. *Medical Image Analysis*, *73*, Article 102184.

Tarvainen, A., & Valpola, H. (2017). Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In *Advances in neural information processing systems* (pp. 1195–1204).

Verma, V., Lamb, A., Kannala, J., Bengio, Y., & Lopez-Paz, D. (2019). Interpolation consistency training for semi-supervised learning. In *Proceedings of the 28th international joint conference on artificial intelligence* (pp. 3635–3641). AAAI Press.

Wang, G., Li, W., Aertsen, M., Deprest, J., Ourselin, S., & Vercauteren, T. (2019). Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks. *Neurocomputing*, *338*, 34–45.

Wang, L., Wang, J., Zhu, L., Fu, H., Li, P., Cheng, G., et al. (2023). Dual multiscale mean teacher network for semi-supervised infection segmentation in chest CT volume for COVID-19. *IEEE Transactions on Cybernetics*, *53*(10), 6363–6375.

Wang, K., Zhan, B., Zu, C., Wu, X., Zhou, J., Zhou, L., et al. (2021). Tripled-uncertainty guided mean teacher model for semi-supervised medical image segmentation. In *Medical image computing and computer assisted intervention* (pp. 450–460). Springer.

Wang, K., Zhan, B., Zu, C., Wu, X., Zhou, J., Zhou, L., et al. (2022). Semi-supervised medical image segmentation via a tripled-uncertainty guided mean teacher model with contrastive learning. *Medical Image Analysis*, *79*, Article 102447.

Wang, Y., Zhang, Y., Tian, J., Zhong, C., Shi, Z., Zhang, Y., et al. (2020). Double-uncertainty weighted method for semi-supervised learning. In *International conference on medical image computing and computer assisted intervention* (pp. 542–551). Springer.

Wong, K. C., Moradi, M., Tang, H., & Syeda-Mahmood, T. (2018). 3D segmentation with exponential logarithmic loss for highly unbalanced object sizes. In *International conference on medical image computing and computer assisted intervention* (pp. 612–619). Springer.

Wu, Y., Ge, Z., Zhang, D., Xu, M., Zhang, L., Xia, Y., et al. (2022). Mutual consistency learning for semi-supervised medical image segmentation. *Medical Image Analysis*, *81*, Article 102530.

Wu, H., Wang, Z., Song, Y., Yang, L., & Qin, J. (2022). Cross-patch dense contrastive learning for semi-supervised segmentation of cellular nuclei in histopathologic images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 11666–11675).

Xiang, T., Zhang, C., Liu, D., Song, Y., Huang, H., & Cai, W. (2020). Bio-net: Learning recurrent bi-directional connections for encoder-decoder architecture. In *International conference on medical image computing and computer assisted intervention* (pp. 74–84). Springer.

Xiang, T., Zhang, C., Wang, X., Song, Y., Liu, D., Huang, H., et al. (2022). Towards bi-directional skip connections in encoder-decoder architectures and beyond. *Medical Image Analysis*, *78*, Article 102420.

Xie, Y., Lu, H., Zhang, J., Shen, C., & Xia, Y. (2019). Deep segmentation-emendation model for gland instance segmentation. In *International conference on medical image computing and computer assisted intervention* (pp. 469–477). Springer.

Xie, Y., Zhang, J., Liao, Z., Verjans, J., Shen, C., & Xia, Y. (2020). Pairwise relation learning for semi-supervised gland segmentation. In *International conference on medical image computing and computer assisted intervention* (pp. 417–427). Springer.

Xu, X., Lian, C., Wang, S., Zhu, T., Chen, R. C., Wang, A. Z., et al. (2021). Asymmetric multi-task attention network for prostate bed segmentation in computed tomography images. *Medical Image Analysis*, *72*, Article 102116.

Xu, Z., Wang, Y., Lu, D., Luo, X., Yan, J., Zheng, Y., et al. (2023). Ambiguity-selective consistency regularization for mean-teacher semi-supervised medical image segmentation. *Medical Image Analysis*, *88*, Article 102880.

Xu, C., Wang, Y., Zhang, D., Han, L., Zhang, Y., Chen, J., et al. (2022). BMAnet: Boundary mining with adversarial learning for semi-supervised 2D myocardial infarction segmentation. *IEEE Journal of Biomedical and Health Informatics*, *27*(1), 87–96.

Yan, K., Guo, Y., & Liu, B. (2023). PreTP-2L: Identification of therapeutic peptides and their types using two-layer ensemble learning framework. *Bioinformatics*, *39*(4), Article btad125.

Yan, K., Lv, H., Guo, Y., Peng, W., & Liu, B. (2023). Samppred-GAT: prediction of antimicrobial peptide by graph attention network and predicted peptide structure. *Bioinformatics*, *39*(1), Article btac715.

Yang, Y., Dasmahapatra, S., & Mahmoodi, S. (2023). ADS_UNet: A nested unet for histopathology image segmentation. *Expert Systems with Applications*, *226*, Article 120128.

Yang, Z., & Farsiu, S. (2023). Directional connectivity-based segmentation of medical images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 11525–11535).

Yu, F., & Koltun, V. (2015). Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:1511.07122.

Yu, C., Wang, J., Gao, C., Yu, G., Shen, C., & Sang, N. (2020). Context prior for scene segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 12416–12425).

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 *International conference on medical image computing and computer assisted intervention* (pp. 605–613). Springer.

Yu, X., Yang, Q., Zhou, Y., Cai, L. Y., Gao, R., Lee, H. H., et al. (2023). UNesT: local spatial representation learning with hierarchical transformer for efficient medical segmentation. *Medical Image Analysis*, *90*, Article 102939.

Zhang, J., Xie, Y., Xia, Y., & Shen, C. (2019). Attention residual learning for skin lesion classification. *IEEE Transactions on Medical Imaging*, *38*(9), 2092–2103.

Zhang, Y., Yang, L., Chen, J., Fredericksen, M., Hughes, D. P., & Chen, D. Z. (2017). Deep adversarial networks for biomedical image segmentation utilizing unannotated images. In *International conference on medical image computing and computer assisted intervention* (pp. 408–416). Springer.

Zhang, S., Zhang, J., Tian, B., Lukasiewicz, T., & Xu, Z. (2023). Multi-modal contrastive mutual learning and pseudo-label re-learning for semi-supervised medical image segmentation. *Medical Image Analysis*, *83*, Article 102656.

Zhao, B., Chen, X., Li, Z., Yu, Z., Yao, S., Yan, L., et al. (2020). Triple U-net: Hematoxylin-aware nuclei segmentation with progressive dense feature aggregation. *Medical Image Analysis*, *65*, Article 101786.

Zhao, X., Qi, Z., Wang, S., Wang, Q., Wu, X., Mao, Y., et al. (2023). RCPS: Rectified contrastive pseudo supervision for semi-supervised medical image segmentation. arXiv preprint arXiv:2301.05500.

Zheng, H., Motch Perrine, S. M., Pitirri, M. K., Kawasaki, K., Wang, C., Richtsmeier, J. T., et al. (2020). Cartilage segmentation in high-resolution 3D micro-CT images via uncertainty-guided self-training with very sparse annotation. In *International conference on medical image computing and computer assisted intervention* (pp. 802–812). Springer.

Zheng, K., Xu, J., & Wei, J. (2022). Double noise mean teacher self-ensembling model for semi-supervised tumor segmentation. In *ICASSP 2022-2022 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (pp. 1446–1450). IEEE.

Zheng, Z., & Yang, Y. (2021). Rectifying pseudo label learning via uncertainty estimation for domain adaptive semantic segmentation. *International Journal of Computer Vision*, *129*(4), 1106–1120.

Zheng, H., Zhang, Y., Yang, L., Liang, P., Zhao, Z., Wang, C., et al. (2019). A new ensemble learning framework for 3D biomedical image segmentation. In *Proceedings of the AAAI conference on artificial intelligence, Vol. 33* (pp. 5909–5916).

Zhong, Z., Lin, Z. Q., Bidart, R., Hu, X., Daya, I. B., Li, Z., et al. (2020). Squeeze-and-attention networks for semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 13065–13074).

Zhou, Y., Chen, H., Lin, H., & Heng, P.-A. (2020). Deep semi-supervised knowledge distillation for overlapping cervical cell instance segmentation. In *International conference on medical image computing and computer assisted intervention* (pp. 521–531). Springer.

Zhu, J., Bolsterlee, B., Chow, B. V., Song, Y., & Meijering, E. (2023). Hybrid dual mean-teacher network with double-uncertainty guidance for semi-supervised segmentation of MRI scans. arXiv preprint arXiv:2303.05126.