Semi-Supervised Histopathology Image Segmentation with Feature Diversified Collaborative Learning

Thanh-Huy Nguyen[2, 5*], **Nguyen Lan Vi Vu**[1, 4*], **Hoang-Thien Nguyen**[3, 4], **Quang-Vinh Dinh**[4], **Xingjian Li**[5], **Min Xu**[5†]

[1]Ho Chi Minh University of Technology, Vietnam
[2]Université de Bourgogne, Dijon, France
[3]Posts and Telecommunications Institute of Technology, Vietnam
[4]AI Vietnam Research Lab, Vietnam
[5]Carnegie Mellon University, United States
mxu1@cs.cmu.edu

Abstract

Histopathology image segmentation plays a critical role in advancing disease diagnosis, prognosis, and treatment planning. However, it presents significant challenges due to the complexity of tissue structures, staining variability, and low contrast between tissue classes. Semi-supervised learning, employed to mitigate annotation scarcity, introduces additional difficulties, such as managing noisy pseudo-labels while ensuring robust performance with limited supervision. Traditional collaborative training methods, commonly used in medical image segmentation, often face issues like model coupling—where models become overly dependent on each other, propagating similar errors—or confirmation bias, where networks reinforce initial mistakes by relying on inaccurate pseudo-labels. Existing frameworks designed to tackle these challenges often suffer from complex pipelines and require extensive pre-training but fail to address the noise characteristics inherent in such datasets. To balance the efficiency of traditional co-training methods with dual networks while enhancing segmentation accuracy on noisy histopathological data, we propose Feature Diversified Collaborative Learning (FDCL). Our work aims to design an effective feature diversification loss that encourages the feature representations of sub-networks to be distinct, ensuring they capture different information to exchange with each other, thereby avoiding suboptimal solutions or, even worse, falling into the coupling problem. We benchmark our method on two well-known histopathology datasets and achieve state-of-the-art results on the GlaS dataset with only 10% of the labeled data. Code is available at https://github.com/vnlvi2k3/FDCL.

Introduction

In computer-aided diagnosis, accurately segmenting cells and glands from histological pictures is a crucial but difficult task (Graham et al. 2019; Xu et al. 2025). Deep learning has achieved state-of-the-art (SoTA) performance on histological image segmentation tasks with a substantial amount of labeled data (Sahasrabudhe et al. 2020). The fact that the data-hungry deep learning model needs a lot of high-quality, well-annotated data presents a difficult problem in histological image processing. However, for experts with subject knowledge, obtaining well-annotated data is a laborious and time-consuming effort. Semi-supervised learning (SSL), which simultaneously learns from a small quantity of labeled and unlabeled data, is suggested as a solution to this problem (Jin et al. 2022). In the realm of semi-supervised learning, ensuring consistency between labeled and unlabeled data remains a challenging problem. This issue has garnered significant attention in recent medical image analysis research (Jin et al. 2022; Xu et al. 2025). For instance, Yu et al. (Yu et al. 2019a) leveraged unlabeled data by incorporating uncertainty into the Mean-Teacher framework (Tarvainen and Valpola 2017), enforcing consistent predictions under various perturbations to improve the target model's performance. Similarly, (Ouali, Hudelot, and Tami 2020a) employed multiple decoders with diverse perturbations ap-

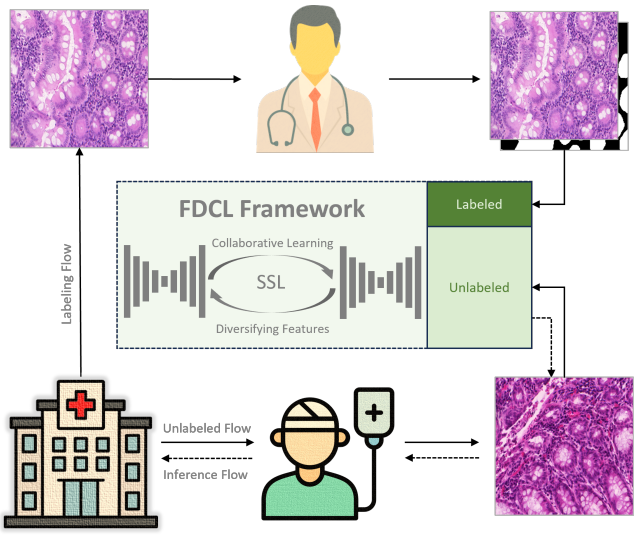


Figure 1: Our overview scenario of FDCL Framework implemented into the real-world clinical setting.

*These authors contributed equally.
†Corresponding Author

plied to the feature space, ensuring consistency between their outputs. In another direction, (Luo et al. 2022b) employed multi-scale consistency regularization at the model output level by leveraging pyramid predictions and quantifying the uncertainty between them. Additionally, (Zhou et al. 2023, 2024) employed fusion techniques integrated with wavelet transform generation to maximize the knowledge extracted from training samples. Another notable approach is the Co-Training method utilized by (Chen et al. 2021a), which initializes two models with different weights to generate pseudo-labels for each other. Expanding on this idea, (Luo et al. 2022a) shifted the focus from model weights to architecture, combining CNN and Transformer backbones in their co-training framework.

However, the idea of leveraging two networks that perform co-training will make these two models become similar after certain iterations. Similar to the coupling bias issue (Ke et al. 2019) from the teacher-student approach, the co-training student-student-like framework performing cross-view pseudo supervision may suffer the convergence of two similar structure networks. In order to reduce the negative effect of identical coupling problems from CPS (Chen et al. 2021b), we proposed the FDCL framework and its variants to perform collaborative training between two identical networks with different initialized weights. FDCL is built on the idea of using contrastive feature pairs to push features away from two network branches depending on a distance in feature space. The question is: where should we perform the features pushing between two identical networks to effectively diversify the network outputs? To answer this, we designed the ablation study with various positions setting of feature diversified module.

Our contributions can be summarized as follows:

- We proposed a simple yet novel semi-supervised segmentation framework on histopathology images called FDCL. Building on the original idea of CPS (Chen et al. 2021b), FDCL introduces a robust feature diversification loss that encourages distinct feature learning between sub-networks, thereby enhancing consistency learning of their predictions while mitigating confirmation bias and coupling problem.
- Through ablation studies, we identified the U-Net feature layers that encapsulate the most meaningful semantic information. These layers were found critical for learning distinct representations, which significantly improve the effectiveness of knowledge exchange between sub-networks.
- We achieved state-of-the-art performance on the GlaS dataset, outperforming CPS and all other frameworks, including both general medical image segmentation methods and those specifically fine-tuned for histopathological data.

Related Works

Histopathology Image Segmentation

In computer-aided analysis, nuclei segmentation is still a difficult task while being crucial to the study of H&E-stained histopathological images. For medical picture segmentation, DL-based techniques have been extensively investigated. In the International Symposium on Biomedical Imaging (ISBI) cell tracking challenge, for example, U-Net (Ronneberger, Fischer, and Brox 2015), an autoencoder-decoder-based model with a skip connection, performed the best. As a result, it has become the standard architecture for the segmentation task, particularly in medical applications such as nuclei segmentation (Nguyen et al. 2024).

Incorporating AI tools into the diagnostic workflow in pathology practice has advanced in a number of ways. By employing computer vision algorithms to extract many features from WSIs, diagnostic predictions can be made (Saltz et al. 2018; Balkenhol et al. 2019). A number of artificial intelligence (AI) techniques are being utilized more frequently to deliver information that is hard for pathologists to recognize (Ström et al. 2020). AI can also be utilized to quickly and efficiently increase the sensitivity of detection by identifying isolated tumor cells in lymph nodes that may be suspected of having metastatic carcinoma. Y-Net, as proposed by Farshad et al.(Farshad et al. 2022), combines a spectral encoder that uses FFC blocks to extract frequency domain characteristics with a spatial encoder that extracts local data. The findings demonstrate that Y-Net performs better than current models at identifying the fluid region in OCT images. By using two encoders—one for the extraction of spatial features from the input image and another for the extraction of features from the attention maps—DAINet, as described by Yeganeh et al. (Yeganeh et al. 2023), makes use of the semantic information stored in the attention maps of a pre-trained DINO. Depending on the tissue type, preparation, microscope settings, and imaging quality, H&E-stained histopathology photographs might vary in hue. It is exceedingly difficult to use histopathological images in DL-based analysis because of this color variance. To lessen color variability, stain normalization techniques like StainGAN (Shaban et al. 2019) have been developed. These techniques work by changing the source images' color distribution to resemble a target template image.

Semi-Supervised Semantic Segmentation

The development of Semi-Supervised Semantic Segmentation (SSSS) heavily relies on two common strategies: Consistency Regularization (Ouali, Hudelot, and Tami 2020a; Zou et al. 2021; Jin, Wang, and Lin 2022; Yang et al. 2023) and Pseudo-Labeling (Yang et al. 2022; Du et al. 2022; Ke et al. 2022). *Consistency Regularization*-based methods aim to ensure invariant model predictions under small perturbations or augmentations applied to the input. For instance, (Ouali, Hudelot, and Tami 2020a) introduced perturbations as strong augmentations in the feature space and employed multiple decoders to enforce consistency in their predictions. Similarly, (Jin, Wang, and Lin 2022) extended the idea of (Tarvainen and Valpola 2017) by introducing an additional teacher assistant model to enhance knowledge transfer between the teacher and student networks. *Pseudo-Labeling*-based methods, on the other hand, focus on generating pseudo-labels for an unlabeled dataset using a model pretrained on labeled data. For example, (Yang et al. 2022)

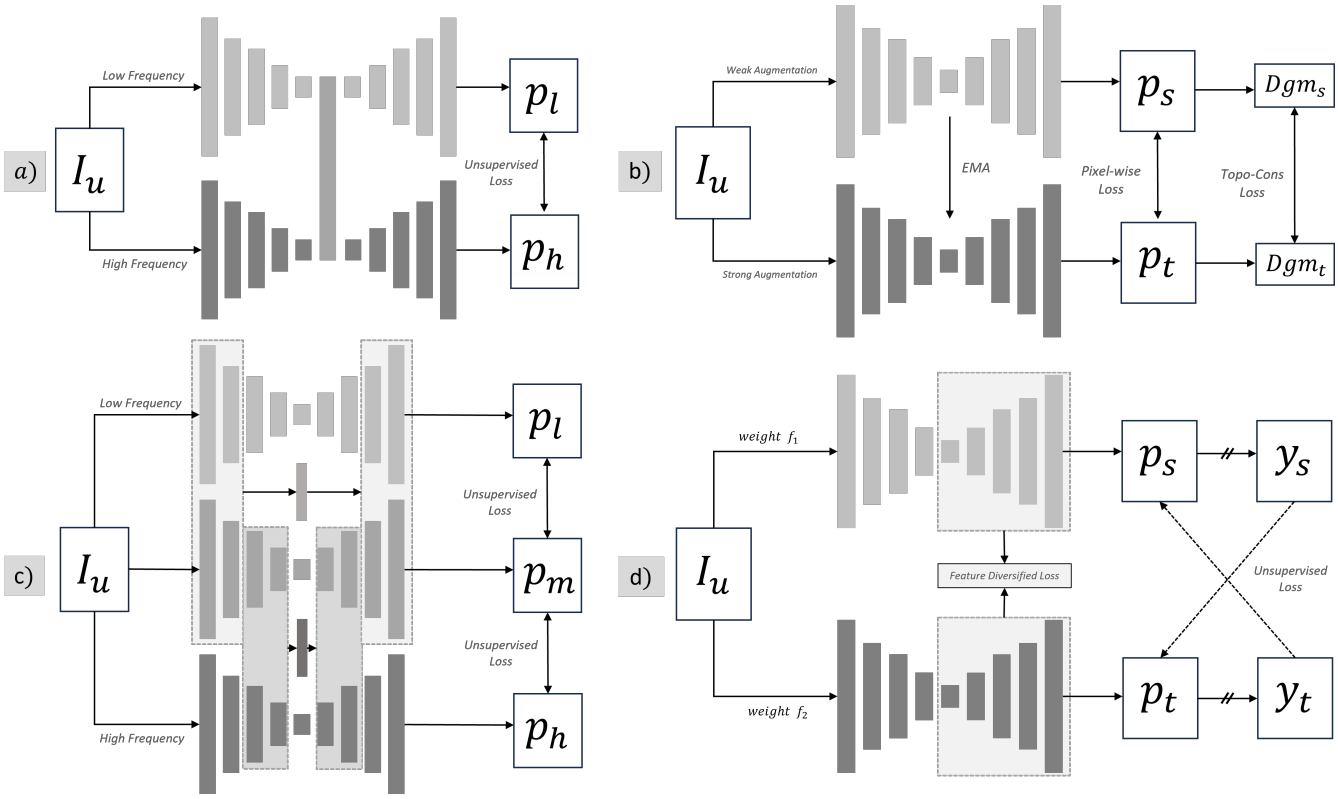


Figure 2: Demonstrating the network structures of (a) Wavelet-Based Low and High-Frequency Fusion Networks (XNet), (b) Topological Consistency with Mean-Teacher framework, (c) enhanced XNetv2, and (d) our proposed FDCL method. In this diagram, $\rightarrow$ denotes forward propagation, // indicates the stop-gradient operation, and both $\leftrightarrow$ and $\dashrightarrow$ represent loss supervision.

first trained the model on labeled data and then used the pretrained model to generate pseudo-labels for the unlabeled dataset. To improve the quality of pseudo-labels, (Ke et al. 2022) proposed a three-stage training process, generating pseudo-labels twice and assuming that the later pseudo-labels would carry more uncertainty than the earlier ones. While these methods achieve remarkable, and even state-of-the-art, results in SSSS, they face significant challenges, such as confirmation bias (Arazo et al. 2019), which arises from the use of inaccurate or incorrect pseudo-labels.

In recent years, another approach, known as Co-Training methods (Chen et al. 2021a; Fan et al. 2022; Wang et al. 2023), has also demonstrated promising results in SSSS. Co-training is based on the multi-view assumption, where two models are initialized independently and exchange knowledge with each other, thereby boosting diversity and improving performance during training. However, these Co-Training methods face the issue of model collapse, where the two models exhibit identical behavior, contradicting the multi-view assumption. To address this issue, (Wang et al. 2023) proposed a discrepancy loss between the extracted features of the two views to ensure distinct behavior in the feature representation space. However, their work lacks depth in analysis and explanation, resulting in a simple and naive formula.

Methodology

The objective of semi-supervised learning is to train a model by utilizing resources from both labeled and unlabeled datasets, denoted as D_L and D_U, respectively. To achieve this goal, our proposed method incorporates a Cross Pseudo Supervision module and a new Feature Diversified Loss, as illustrated in Figure 2

Supervised Phase

As in most semi-supervised methods, a small set of labeled data D_L is trained in a supervised manner to ensure that the proposed framework can generate meaningful predictions. First, two confidence maps P_1^L and P_2^L are computed as:

$$P_1^L = \text{Softmax}(M_1(x_l)) \tag{1}$$

$$P_2^L = \text{Softmax}(M_2(x_l)) \tag{2}$$

where $x_l \in D_L$ represents the labeled input samples.

Then, the supervised loss between these predictions and ground truth labels is calculated by applying the Dice Loss function, which can be defined as:

$$\mathcal{L}_{sup} = Dice(P_1^L, Y) + Dice(P_2^L, Y) \tag{3}$$

where Y denotes the ground truth labels in the labeled dataset D_L.

Labeled Ratio	Method	GlaS-2017		CRAG-2019	
		Dice ↑	Jaccard ↑	Dice ↑	Jaccard ↑
100%	**Fully- Supervised**	92.59 ± 0.01	86.2 ± 0.01	92.68 ± 0.01	86.36 ± 0.01
20%	MT (NeurIPS'2017)	89.46 ± 0.83	80.94 ± 1.35	89.77 ± 0.1	81.43 ± 0.15
	EM (CVPR'2019)	88.53 ± 0.34	79.42 ± 0.56	87.55 ± 0.5	77.93 ± 0.8
	UA-MT (MICCAI'2019)	89.27 ± 0.12	80.62 ± 0.19	88.84 ± 0.11	80.01 ± 0.06
	CCT (CVPR'2020)	89.49 ± 0.09	80.98 ± 0.15	89.5 ± 0.12	80.99 ± 0.19
	CPS (CVPR'2021)	$\underline{91.94 \pm 0.05}$	$\underline{85.08 \pm 0.08}$	89.42 ± 0.15	80.87 ± 0.25
	URPC (MedIA'2022)	88.92 ± 0.45	80.05 ± 0.74	88.75 ± 0.73	79.25 ± 0.25
	CT (PMLR'2022)	88.45 ± 0.3	79.29 ± 0.49	86.96 ± 1.19	75.9 ± 0.33
	XNet (ICCV'2023)	88.68 ± 0.42	79.67 ± 0.67	88.1 ± 0.15	79.1 ± 0.29
	TopoSemiSeg (ECCV'2024)	89.5 ± 0.0	81.8 ± 0.0	89.8 ± 0.0	82.0 ± 0.0
	XNetv2 (BIBM'2024)	90.8 ± 0.12	83.15 ± 0.2	$\mathbf{91.11 \pm 0.25}$	$\mathbf{83.75 \pm 0.45}$
	FDCL	$\mathbf{92.01 \pm 0.23}$	$\mathbf{85.16 \pm 0.41}$	$\underline{90.14 \pm 0.16}$	$\underline{82.02 \pm 0.12}$
10%	MT (NeurIPS'2017)	87.65 ± 0.78	78.02 ± 1.23	86.82 ± 0.09	76.71 ± 0.14
	EM (CVPR'2019)	83.64 ± 0.56	71.89 ± 0.79	84.13 ± 0.78	72.61 ± 1.17
	UA-MT (MICCAI'2019)	84.41 ± 0.56	73.03 ± 0.83	83.97 ± 0.46	72.75 ± 0.6
	CCT (CVPR'2020)	85.46 ± 0.42	74.51 ± 0.64	85.28 ± 0.12	74.34 ± 0.18
	CPS (CVPR'2021)	$\underline{89.86 \pm 0.7}$	$\underline{81.6 \pm 0.15}$	85.52 ± 0.31	74.7 ± 0.47
	URPC (MedIA'2022)	81.59 ± 0.62	68.9 ± 0.88	82.5 ± 0.31	72.1 ± 0.61
	CT (PMLR'2022)	81.65 ± 0.61	68.99 ± 0.87	80.33 ± 0.68	67.13 ± 0.95
	XNet (ICCV'2023)	84.44 ± 0.5	73.07 ± 0.75	86.67 ± 0.05	76.48 ± 0.09
	TopoSemiSeg (ECCV'2024)	87.8 ± 0.0	79.7 ± 0.0	$\mathbf{88.4 \pm 0.0}$	$\mathbf{79.8 \pm 0.0}$
	XNetv2 (BIBM'2024)	88.96 ± 0.46	80.11 ± 0.75	$\underline{88.3 \pm 0.12}$	$\underline{79.04 \pm 0.2}$
	FDCL	$\mathbf{90.95 \pm 0.14}$	$\mathbf{83.43 \pm 0.27}$	87.48 ± 0.17	77.74 ± 0.27

Table 1: Quantitative experiments compare our method with fully supervised results and other semi-supervised approaches on 20% and 10% labeled fractions of the GlaS and CRAG datasets. Bold and underlined results indicate the highest and second-best performance, respectively.

Unsupervised Co-Training Phase

Based on the idea of (Chen et al. 2021a), our framework is designed with two models, which are initialized with the same backbone but distinct parameters, called M_1 and M_2 respectively. For the unlabeled dataset D_U, two confidence maps are predicted by M_1 and M_2 after applying the $Softmax$ function, which can be written as:

$$P_1 = \text{Softmax}(M_1(x^u)) \tag{4}$$

$$P_2 = \text{Softmax}(M_2(x^u)) \tag{5}$$

where $x^u \in D_U$ is considered as unannotated samples.

Next, the confidence maps P_1 and P_2 are converted into hard pseudo-label maps by using the $argmax$ function:

$$\hat{Y}_1 = \arg\max(P_1) \tag{6}$$

$$\hat{Y}_2 = \arg\max(P_2) \tag{7}$$

Finally, the cross supervision loss $\mathcal{L}_{cps}$ is computed by cross-pairing confidence maps and the pseudo labels between two models M_1 and M_2, resulting in the pairs $\{P_1, \hat{Y}_2\}$ and $\{P_2, \hat{Y}_1\}$. The $\mathcal{L}_{cps}$ can be formulated as:

$$\mathcal{L}_{cps} = Dice(P_1, \hat{Y}_2) + Dice(P_2, \hat{Y}_1) \tag{8}$$

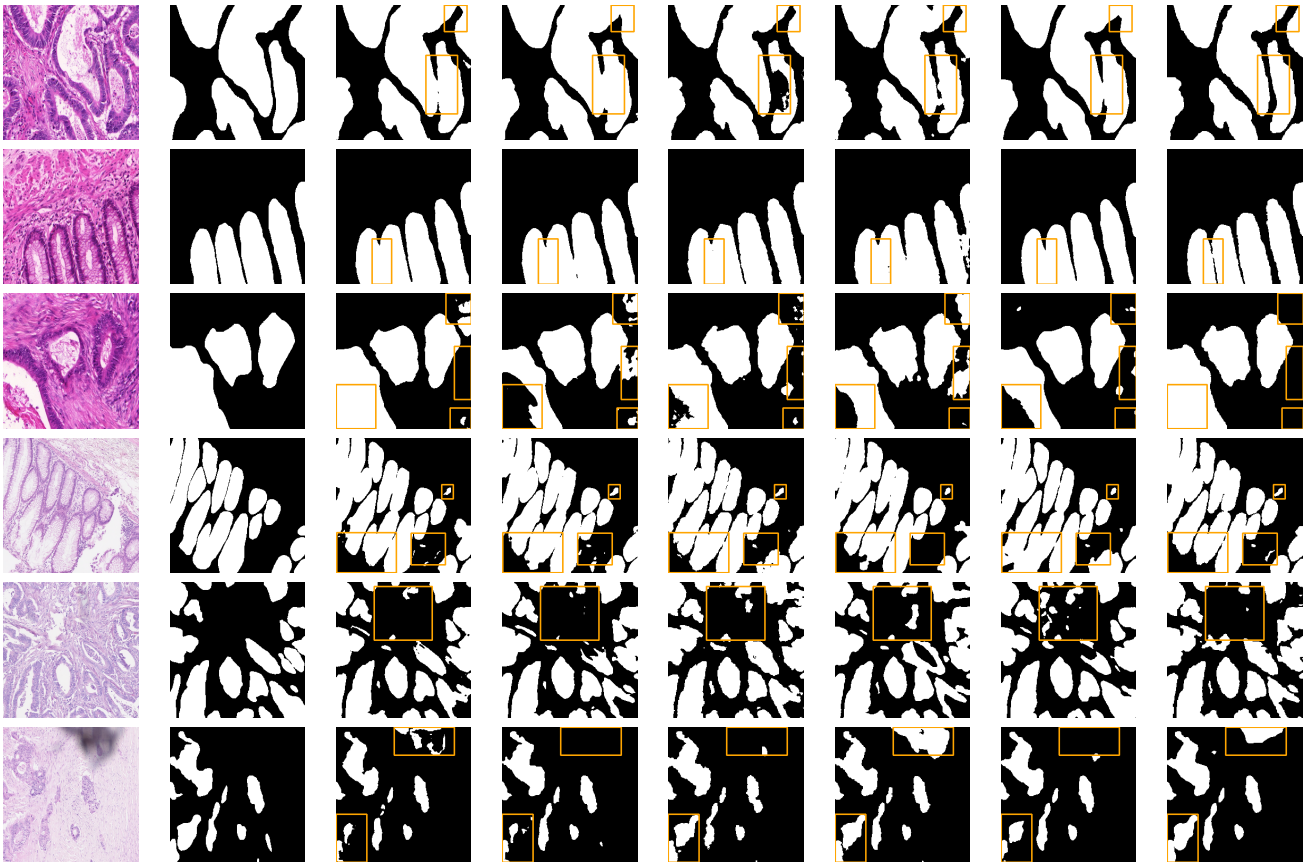
where $Dice$ represents the Dice Loss function.

Feature Diversified Collaborative Learning

We define a Feature Diversified Loss function with the objective of encouraging the two sub-nets to learn diverse and expanded feature spaces while maintaining consistent predictions. This loss function prevents the models from converging to identical representations as training progresses, thus avoiding the risk of both branches learning nothing distinct from each other.

$$\mathcal{L}_{fd} = \frac{1}{2}\left(\frac{1}{\|F_1 - \text{sg}(F_2)\|_2^2} + \frac{1}{\|F_2 - \text{sg}(F_1)\|_2^2}\right) \tag{9}$$

where F_1 and F_2 represent the features extracted from the layers just before the final classifier layer of M_1 and M_2, respectively (experiments on other feature layers can be found in the ablation study section), and $\text{sg}()$ denotes the stop-gradient operation. Through empirical experiments, we observe that this loss function plays a crucial role in mitigating coupling problem, based on the strong hypothesis that it compels the models to explore a broader complementary information search space, thereby avoiding convergence to suboptimal solutions. The conflicting features before the classifier enhance the pseudo-labels generated by each model, making them a more reliable version for supervising the other model prediction. The unsupervised loss is



(a) Image (b) GT (c) MT (d) CCT (e) CPS (f) URPC (g) XNetv2 **(h) FDCL**

Figure 3: Qualitative results for different semi-supervised methods on 10% labeled data (GlaS) and 20% labeled data (CRAG). Rows 1 to 3 correspond to GlaS, while Rows 4 to 6 correspond to CRAG. The differences in the results are emphasized by the orange boxes.

then computed as a combination of $\mathcal{L}_{cps}$ and $\mathcal{L}_{fd}$:

$$\mathcal{L}_{unsup} = \mathcal{L}_{cps} + \lambda_{fd}\mathcal{L}_{fd} \tag{10}$$

where λ_{fd} is a hyperparameter that controls the contribution of the Feature Diversified Loss to the overall unsupervised objective. The final training loss is formulated as:

$$\mathcal{L}_{final} = \mathcal{L}_{sup} + \lambda_{unsup}\mathcal{L}_{unsup} \tag{11}$$

Here, λ_{unsup} is another hyperparameter that control the balance between supervised loss and unsupervised loss.

Experiments

Experimental Settings

Dataset: We evaluate our method on two Histopathological Image datasets:

CRAG (Graham et al. 2019) focuses on gland segmentation in colon histopathology images, a critical task for grading colon cancer. The dataset includes 213 annotated images, with 173 for training and 40 for testing, most of which have a resolution of 1512×1516. Gland segmentation is particularly challenging due to the significant variability in glandular morphology.

GlaS (Sirinukunwattana et al. 2017) is tailored for gland segmentation in colorectal adenocarcinoma and captures images from various stages of cancer progression. It provides 85 annotated training images (37 benign and 48 malignant) and 80 test images, with resolutions commonly around 775×522 or 589×453.

We present the performance results of all models trained with 10% and 20% labeled images on both datasets.

Model details and Data Augmentation: We conduct all experiments on a single NVIDIA RTX 3090 GPU, using Python 3.11, PyTorch 2.3, and CUDA 12.2. For the GlaS dataset, we train XNet_{sb} (Zhou et al. 2023), a simplified variant of XNet with only its top branch, while for the CRAG dataset, we use U-Net (Ronneberger, Fischer, and Brox 2015) as the backbone. We train our model on GlaS for 200 epochs and CRAG for 250 epochs to achieve convergence. The input images are first resized to 256×256 and augmented using random flips, rotations, and transpositions. We use SGD with a momentum of 0.9, a weight decay of 5×10^{-5}, and an initial learning rate of 0.5. The batch size of 2 or 4 yields the best results, λ_{fd} is set to 1.0, and λ_{unsup} is modeled as an time-dependent warming up function, reach-

ing a maximum of 5.0.

Evaluation Metrics: Both Dice similarity and Jaccard coefficient are used to evaluate our performance model on validation dataset. These metrics can be defined as follows:

$$Dice = \frac{2TP}{2TP + FP + FN}$$

$$Jaccard = \frac{TP}{TP + FP + FN}$$

where TP, FP, TN, and FN represent true-positive, false-positive, true-negative, and false-negative predictions, respectively.

Comparison with state-of-the-art SSL methods

We evaluate the effectiveness of our proposed FDCL by benchmarking it against 10 SOTA semi-supervised segmentation methods. These methods encompass both general-purpose medical image segmentation techniques and specialized approaches tailored for histopathological image segmentation, including Mean Teacher (MT) (Tarvainen and Valpola 2017), Entropy Minimization (EM) (Vu et al. 2019), Uncertainty-Aware Mean Teacher (UA-MT) (Yu et al. 2019b), Cross Consistency Training (CCT) (Ouali, Hudelot, and Tami 2020b), Cross Pseudo-Supervision (CPS) (Chen et al. 2021b), Uncertainty-Rectified Pyramid Consistency (URPC) (Luo et al. 2022b), Cross Teaching Between CNN and Transformer (CT) (Luo et al. 2022a), Wavelet-Based Low and High Frequency Fusion Networks (XNet) (Zhou et al. 2023), Topological Consistency (TopoSemiSeg) (Xu et al. 2024), and the enhanced XNetv2 (Zhou et al. 2024). Due to challenges in reproducing the results of TopoSemiSeg, we directly report its original results, while for all other methods, we reran experiments three times to obtain the mean and standard deviation.

As shown in Table 1, our method demonstrates outstanding performance on the GlaS dataset, even with limited labeled data. At the 10% labeled ratio, our method achieves a Dice score of 90.95 ± 0.14 and a Jaccard index of 83.43 ± 0.27, surpassing all other approaches, including CPS (Dice = 89.86 ± 0.7, Jaccard = 81.6 ± 0.15), and closely approaching the fully-supervised results at the 20% labeled fraction. Mean Teacher-based architectures, such as MT and UA-MT, as well as single-network approaches like URPC, perform less effectively on the GlaS dataset.

The top three rows of Figure 3 further illustrate the qualitative superiority of our method. It successfully separates individual glands, whereas other methods tend to merge adjacent ones. Moreover, our method exhibits strong robustness to noise. For instance, in the third row, while other methods misclassify noisy edge regions as false positives, FDCL correctly classifies these areas.

On the CRAG dataset, at the 20% labeled fraction, our method achieves a Dice score of 90.14 ± 0.16 and a Jaccard index of 82.02 ± 0.12, ranking just behind XNetv2, which achieves 91.11 ± 0.25 and 83.75 ± 0.45, respectively. At the 10% labeled fraction, our method follows TopoSemiSeg and XNetv2 in performance. While XNetv2's 3-branch architecture enhances its accuracy, it requires 3 hours to train compared to just 1 hour for our method on the same hardware. Similarly, TopoSemiSeg demands pretraining with 12,000 iterations before applying the topological loss, whereas our method achieves competitive results without such extensive pretraining.

Figure 3 (bottom three rows) highlights the challenges posed by this dataset. In row 1, small noise artifacts are misclassified by most methods. In row 2, XNet and URPC are more affected by noise, while our method remains comparatively stable. In row 3, our method, along with MT, CPS, and URPC, misclassifies the black region at the top, whereas XNet and CCT handle it correctly. Despite these challenges, our method strikes an effective balance between efficiency and performance, making it practical for real-world settings.

Ablation Studies

Method	Feature layer			Metrics	
	Low	Med	High	Dice ↑	Jaccard ↑
FDCL	✓	✗	✗	**90.95**	**83.43**
	✗	✓	✗	90.46	82.57
	✗	✗	✓	90.69	82.96
	✓	✓	✗	90.13	82.03
	✓	✗	✓	90.88	83.28
	✗	✓	✓	90.00	81.81
	✓	✓	✓	90.88	83.28
	✗	✗	✗	89.23	80.56

Table 2: Ablation Study on Feature Layer Selection. Low, Mid, and High refer to features extracted from the layers before the classifier, intermediate layers, and the bottleneck layer of U-Net, respectively.

We perform a step-by-step ablation study to evaluate the effectiveness of using three different feature layers of U-Net in the Feature Diversified Loss $\mathcal{L}_{fd}$. These layers include: High-Level features (extracted from the encoder block), Medium-Level features (extracted from the immediate layers of the decoder block), and Low-Level features (from the last layers of the decoder block). The results in Table 2 show that using low-level features provides the highest performance, achieving a Dice score of 90.95% and a Jaccard score of 83.43%. In contrast, when $\mathcal{L}_{fd}$ is not applied, the performance drops significantly, with Dice and Jaccard scores of only 89.23% and 80.56%, respectively, highlighting the importance of $\mathcal{L}_{fd}$ in improving segmentation results. Using high-level features or medium-level features improves performance compared to the baseline but remains suboptimal, falling short of the performance achieved with low-level features. Interestingly, fusing high, medium, and low-level features results in varying outcomes depending on the combination. Notably, in cases when integrating FDCL for all mentioned layers or just both low-level and high-level, they reached the same metric scores with Dice at 90.88% and JC at 83.23%, which could indicate that the effect of medium-level features on model performance is trivial compared to other features. Furthermore, the combination of low-level and medium-level features outperformed the combination of medium-level and high-level features, indicating that low-level features have a more significant

impact on the model's performance than either medium- or high-level features.

Conclusion

We propose Feature Diversified Collaborative Learning (FDCL), a straightforward yet powerful method for semi-supervised histopathology image segmentation. FDCL enhances the power of collaborative training by eliminating a common challenge of this framework — confirmation bias and coupling problem. The core of FDCL is to encourage diverse feature learning across sub-networks, thereby guiding the model to produce robust confidence maps and enhancing the overall consistency of predictions.

Our extensive experiments reveal that FDCL achieves competitive performance across challenging histopathology image datasets and outperforms current state-of-the-art methods on the GlaS dataset. FDCL strikes a balance between training efficiency and segmentation performance. In future work, we aim to delve deeper into advanced data augmentation techniques tailored to the unique characteristics of histopathological data and extend the adaptability of FDCL to a broader range of medical image segmentation tasks.

Acknowledgments

This work was supported in part by U.S. NIH grants R01GM134020.

References

Arazo, E.; Ortego, D.; Albert, P.; O'Connor, N. E.; and McGuinness, K. 2019. Pseudo-Labeling and Confirmation Bias in Deep Semi-Supervised Learning. *2020 International Joint Conference on Neural Networks (IJCNN)*, 1–8.

Balkenhol, M. C.; Tellez, D.; Vreuls, W.; Clahsen, P. C.; Pinckaers, H.; Ciompi, F.; Bult, P.; and Van Der Laak, J. A. 2019. Deep learning assisted mitotic counting for breast cancer. *Laboratory investigation*, 99(11): 1596–1606.

Chen, X.; Yuan, Y.; Zeng, G.; and Wang, J. 2021a. Semi-Supervised Semantic Segmentation with Cross Pseudo Supervision. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2613–2622.

Chen, X.; Yuan, Y.; Zeng, G.; and Wang, J. 2021b. Semi-Supervised Semantic Segmentation with Cross Pseudo Supervision. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Du, Y.; Shen, Y.; Wang, H.; Fei, J.; Li, W.; Wu, L.; Zhao, R.; Fu, Z.; and LIU, Q. 2022. Learning from Future: A Novel Self-Training Framework for Semantic Segmentation. In *Advances in Neural Information Processing Systems*.

Fan, J.; Gao, B.; Jin, H.; and Jiang, L. 2022. UCC: Uncertainty guided Cross-head Cotraining for Semi-Supervised Semantic Segmentation. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 9937–9946.

Farshad, A.; Yeganeh, Y.; Gehlbach, P.; and Navab, N. 2022. Y-net: A spatiospectral dual-encoder network for medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 582–592. Springer.

Graham, S.; Chen, H.; Gamper, J.; Dou, Q.; Heng, P.-A.; Snead, D.; Tsang, Y. W.; and Rajpoot, N. 2019. MILD-Net: Minimal information loss dilated network for gland instance segmentation in colon histology images. *Medical image analysis*, 52: 199–211.

Jin, Q.; Cui, H.; Sun, C.; Zheng, J.; Wei, L.; Fang, Z.; Meng, Z.; and Su, R. 2022. Semi-supervised histological image segmentation via hierarchical consistency enforcement. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 3–13. Springer.

Jin, Y.; Wang, J.; and Lin, D. 2022. Semi-Supervised Semantic Segmentation via Gentle Teaching Assistant. In *Advances in Neural Information Processing Systems*.

Ke, R.; Aviles-Rivero, A. I.; Pandey, S.; Reddy, S.; and Schönlieb, C.-B. 2022. A Three-Stage Self-Training Framework for Semi-Supervised Semantic Segmentation. *IEEE Transactions on Image Processing*, 31: 1805–1815.

Ke, Z.; Wang, D.; Yan, Q.; Ren, J.; and Lau, R. W. 2019. Dual student: Breaking the limits of the teacher in semi-supervised learning. In *Proceedings of the IEEE/CVF international conference on computer vision*, 6728–6736.

Luo, X.; Hu, M.; Song, T.; Wang, G.; and Zhang, S. 2022a. Semi-supervised medical image segmentation via cross teaching between cnn and transformer. In *International conference on medical imaging with deep learning*, 820–833. PMLR.

Luo, X.; Wang, G.; Liao, W.; Chen, J.; Song, T.; Chen, Y.; Zhang, S.; Metaxas, D. N.; and Zhang, S. 2022b. Semi-supervised medical image segmentation via uncertainty rectified pyramid consistency. *Medical Image Analysis*, 80: 102517.

Nguyen, T.-H.; Ngo, T. K. N.; Vu, M. A.; and Tu, T.-Y. 2024. Blurry-Consistency Segmentation Framework with Selective Stacking on Differential Interference Contrast 3D Breast Cancer Spheroid. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5223–5230.

Ouali, Y.; Hudelot, C.; and Tami, M. 2020a. Semi-Supervised Semantic Segmentation With Cross-Consistency Training. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 12671–12681.

Ouali, Y.; Hudelot, C.; and Tami, M. 2020b. Semi-Supervised Semantic Segmentation With Cross-Consistency Training. In *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, 234–241. Springer.

Sahasrabudhe, M.; Christodoulidis, S.; Salgado, R.; Michiels, S.; Loi, S.; André, F.; Paragios, N.; and

Vakalopoulou, M. 2020. Self-supervised nuclei segmentation in histopathological images using attention. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part V 23*, 393–402. Springer.

Saltz, J.; Gupta, R.; Hou, L.; Kurc, T.; Singh, P.; Nguyen, V.; Samaras, D.; Shroyer, K. R.; Zhao, T.; Batiste, R.; et al. 2018. Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images. *Cell reports*, 23(1): 181–193.

Shaban, M. T.; Baur, C.; Navab, N.; and Albarqouni, S. 2019. Staingan: Stain style transfer for digital histological images. In *2019 Ieee 16th international symposium on biomedical imaging (Isbi 2019)*, 953–956. IEEE.

Sirinukunwattana, K.; Pluim, J. P.; Chen, H.; Qi, X.; Heng, P.-A.; Guo, Y. B.; Wang, L. Y.; Matuszewski, B. J.; Bruni, E.; Sanchez, U.; et al. 2017. Gland segmentation in colon histology images: The glas challenge contest. *Medical image analysis*, 35: 489–502.

Ström, P.; Kartasalo, K.; Olsson, H.; Solorzano, L.; Delahunt, B.; Berney, D. M.; Bostwick, D. G.; Evans, A. J.; Grignon, D. J.; Humphrey, P. A.; et al. 2020. Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: a population-based, diagnostic study. *The Lancet Oncology*, 21(2): 222–232.

Tarvainen, A.; and Valpola, H. 2017. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Advances in neural information processing systems*, 30.

Vu, T.-H.; Jain, H.; Bucher, M.; Cord, M.; and Pérez, P. 2019. Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2517–2526.

Wang, Z.; Zhao, Z.; Xing, X.; Xu, D.; Kong, X.; and Zhou, L. 2023. Conflict-Based Cross-View Consistency for Semi-Supervised Semantic Segmentation. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 19585–19595.

Xu, M.; Hu, X.; Gupta, S.; Abousamra, S.; and Chen, C. 2024. TopoSemiSeg: Enforcing Topological Consistency for Semi-Supervised Segmentation of Histopathology Images. In *ECCV*.

Xu, M.; Hu, X.; Gupta, S.; Abousamra, S.; and Chen, C. 2025. Semi-supervised Segmentation of Histopathology Images with Noise-Aware Topological Consistency. In *European Conference on Computer Vision*, 271–289. Springer.

Yang, L.; Qi, L.; Feng, L.; Zhang, W.; and Shi, Y. 2023. Revisiting Weak-to-Strong Consistency in Semi-Supervised Semantic Segmentation. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 7236–7246.

Yang, L.; Zhuo, W.; Qi, L.; Shi, Y.; and Gao, Y. 2022. ST++: Make Self-trainingWork Better for Semi-supervised Semantic Segmentation. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 4258–4267.

Yeganeh, Y.; Farshad, A.; Weinberger, P.; Ahmadi, S.-A.; Adeli, E.; and Navab, N. 2023. Transformers pay attention to convolutions leveraging emerging properties of ViTs by dual attention-image network. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2304–2315.

Yu, L.; Wang, S.; Li, X.; Fu, C.-W.; and Heng, P.-A. 2019a. 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, proceedings, part II 22*, 605–613. Springer.

Yu, L.; Wang, S.; Li, X.; Fu, C.-W.; and Heng, P.-A. 2019b. Uncertainty-aware Self-ensembling Model for Semi-supervised 3D Left Atrium Segmentation. In *MICCAI*.

Zhou, Y.; Huang, J.; Wang, C.; Song, L.; and Yang, G. 2023. XNet: Wavelet-Based Low and High Frequency Fusion Networks for Fully- and Semi-Supervised Semantic Segmentation of Biomedical Images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 21085–21096.

Zhou, Y.; Li, L.; Wang, Z.; Liu, G.; Liu, Z.; and Yang, G. 2024. XNet v2: Fewer Limitations, Better Results and Greater Universality. *arXiv preprint arXiv:2409.00947*.

Zou, Y.; Zhang, Z.; Zhang, H.; Li, C.-L.; Bian, X.; Huang, J.-B.; and Pfister, T. 2021. PseudoSeg: Designing Pseudo Labels for Semantic Segmentation. In *International Conference on Learning Representations*.