Semi-supervised Segmentation of Histopathology Images with Noise-Aware Topological Consistency

Meilong Xu*[1], Xiaoling Hu[2], Saumya Gupta[1], Shahira Abousamra[1], and Chao Chen[1]

[1] Stony Brook University, Stony Brook, NY, USA
[2] Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA

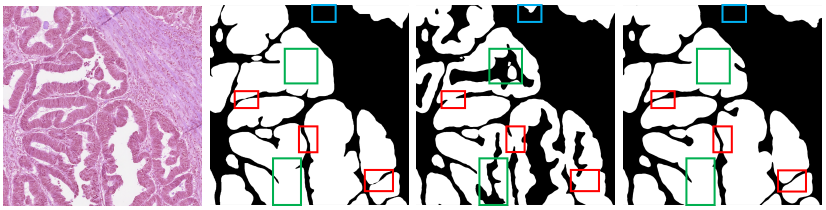
Abstract. In digital pathology, segmenting densely distributed objects like glands and nuclei is crucial for downstream analysis. Since detailed pixel-wise annotations are very time-consuming, we need semi-supervised segmentation methods that can learn from unlabeled images. Existing semi-supervised methods are often prone to topological errors, *e.g.*, missing or incorrectly merged/separated glands or nuclei. To address this issue, we propose *TopoSemiSeg*, the first semi-supervised method that learns the topological representation from unlabeled histopathology images. The major challenge is for unlabeled images; we only have predictions carrying noisy topology. To this end, we introduce a noise-aware topological consistency loss to align the representations of a teacher and a student model. By decomposing the topology of the prediction into signal topology and noisy topology, we ensure that the models learn the true topological signals and become robust to noise. Extensive experiments on public histopathology image datasets show the superiority of our method, especially on topology-aware evaluation metrics. Code is available at https://github.com/Melon-Xu/TopoSemiSeg.

Keywords: Histopathology Imaging · Semi-supervised Segmentation · Topological Consistency

1 Introduction

In digital pathology, histopathology images can provide crucial insights for clinical diagnoses and treatment planning. Pathologists can make diagnosis and prognosis decisions by studying the morphology of glands/nuclei and their spatial arrangements. For example, assessing gland morphology can help pathologists determine different stages of colon cancer [12] and prostate cancer [35]. Traditionally, this would rely on manual annotations by pathologists, which is costly, time-consuming, and error-prone. To alleviate this burden, deep learning methods have been proposed to automatically segment the objects of interest [3, 13, 22, 50, 65]. However, despite their satisfactory performances, these methods still rely on a large amount of high-quality annotations, which is expensive and requires a lot of domain expertise.

* Email: meixu@cs.stonybrook.edu.



(a) Input **(b)** GT **(c)** XNet [64] **(d)** Ours

Fig. 1: Illustration of the significance of topological correctness in gland segmentation. **(a)** an input image. **(b)** ground truth GT. **(c)** the result of SoTA semi-supervised segmentation method [64] devoid of any topological regularization. **(d)** our segmentation result. For the regions within boxes, the SoTA's result has errors that, while minor at the pixel level, significantly alter the semantic interpretation. The **red** boxes indicate prediction errors such as incorrectly merging adjacent glands, the **blue** box indicates false positive gland predictions, and the **green** boxes indicate the false negative holes in glands. These errors affect the pathologist's decision and analysis.

One of the dominant schemes to reduce the cost of annotation is semi-supervised learning (SemiSL) [11, 23, 26, 30, 43, 48, 54, 61, 62]. SemiSL leverages a small group of labeled data along with a large amount of unlabeled data to train a model. By harvesting the rich information in the unlabeled data, they can perform as well as fully-supervised methods. While existing works mostly focus on pixel-wise accuracy, limited progress has been made to address topological correctness. At regions where glands or nuclei are densely distributed, the model tends to make topological errors such as mistakenly merged/separated glands, or missing components. See Fig. 1 for an illustration. Even a strong semi-supervised method like [64] still fails to properly maintain glands' topological correctness, as highlighted by the boxed regions. Such topological errors significantly change their morphological measures. Similar issues also occur in the nuclei segmentation task due to dense distribution characteristics. If not addressed properly, these topological errors will significantly impact downstream analysis.

In this paper, we investigate how to help models learn the correct topological characteristics from unlabeled data. Traditional methods [6, 17, 18, 20, 44, 51] solve the segmentation problem by encoding topological properties as constraints during training. However, these methods require ground-truth annotations and a clean topology to train. Therefore, they are not able to exploit unlabeled data.

To address this issue, we propose *the first topology-aware solution for semi-supervised segmentation of histopathology images*. Unlike the clean topology of the ground truth (GT), the topology of the predictions is noisy and contains spurious structures (see Fig. 2). These **noisy structures** may be the holes inside large glands or small islands that are false positives. They will oscillate through training and significantly distract the learning from concentrating on the **true topological signals** (the structures we want to preserve).

To this end, based on the teacher-student framework [46], we propose to enforce noise-aware topological consistency between the predictions of different

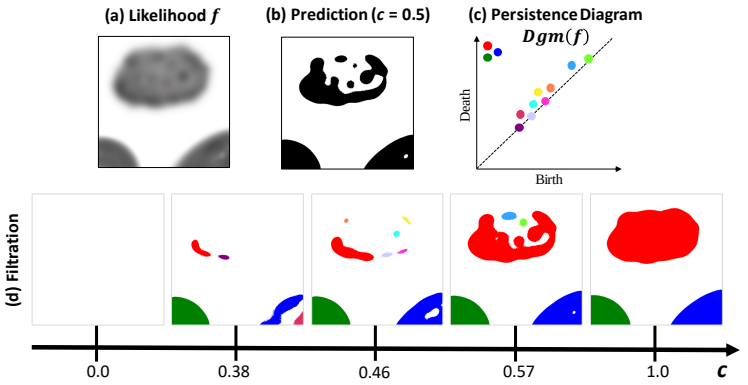


Fig. 2: **(a)** A predicted likelihood map f, **(b)** the binary prediction, and **(c)** the corresponding persistence diagram $Dgm(f)$, which tends to be noisy. In **(d)**, consider the filtration for different values of threshold c. Notice that there are three true, or *signal*, structures, denoted by colors **red**, **green**, and **blue**, which persist across the range of c. Hence the dots corresponding to these structures are located at the upper-left corner of $Dgm(f)$. The remaining colors denote several *noisy* structures which persist for a short range of c, and thus their dots appear closer to the diagonal. Note that we only show 0-dim persistent dots referring to connected components in $Dgm(f)$.

augmented inputs. More specifically, we propose to decompose the topological structures of a potentially noisy prediction into *signal topology* and *noisy topology*. This can be achieved by decomposing the topological features, formalized as the *persistence diagram* (PD) [10], into signal and noise. We only enforce the signal topology of the teacher and the student's prediction to be consistent. This is achieved by a *signal topology consistency loss* that matches the signal topological features using the Wasserstein distance [7, 8]. Meanwhile, for the noisy topological features, we introduce a *noisy topology removal loss*, based on a theoretical measure called *total persistence* [8]. It aggregates the saliency of all noisy topological structures. Minimizing it essentially removes all these noisy structures. Combining the proposed *signal topology consistency loss* and *noisy topology removal loss* with the classic pixel-wise consistency loss, our method achieves the desired goal and ensures the student model learns the robust topological representations from the unlabeled data.

We note that the method in [18] (which we refer to as TopoLoss) also designed a topology-aware loss based on the persistence diagram. However, this method is designed for a fully-supervised setting and relies heavily on the clean persistence diagrams of the GTs. The clean GTs facilitate the straightforward matching of the persistent dots. In contrast, SemiSL encounters a unique issue where we only have noisy predictions of the unlabeled data, containing a lot of spurious structures. A non-trivial solution is required to address the challenge of matching two noisy PDs. Hence, the previous TopoLoss cannot adapt to the

teacher-student network, where we are forced to compare the topology of noisy predictions.

We evaluate the proposed method by conducting experiments on three public histopathology image datasets. The results show that our method outperforms other SoTA semi-supervised methods on both pixel- and topology-wise evaluation metrics, across 10% and 20% labeled data settings. In summary, our contributions are as follows:

- We propose the first topology-aware semi-supervised framework that enforces topological consistency in segmenting densely distributed objects of interest in histopathology images.
- We propose a learning strategy that produces robust topological representations from the noisy topological feature space of the unlabeled images.
- Through extensive experiments on multiple histopathology imaging datasets, we show that our method effectively improves the segmentation quality both pixel- and topology-wise.

2 Related work

Segmentation with limited annotations. To address the scarcity of annotated data, semi-supervised learning (SemiSL) has emerged as a pivotal methodology in medical image segmentation [24]. The primary schemes in this domain encompass pseudo-labeling [39, 57, 62], consistency learning [21, 31, 37] and entropy minimization [2, 14, 53]. Pseudo-labeling-based methods aim to generate pseudo-labels for unlabeled data, which are then used to train the model further. To improve the quality of pseudo-labels, Wang *et al.* [52] propose a confidence-aware module to select pseudo labels with high confidence. Some works try to refine the pseudo-labels by morphological methods [47] or by adding additional refinement networks [40, 62]. By learning better representations that pull similar samples together and push dissimilar ones apart, contrastive learning is also applied in SemiSL [1, 58, 59].

Another main scheme in SemiSL is consistency learning, which emphasizes consistent predictions under various perturbations. Different perturbations at input or feature level are proposed to make the model robust [31, 32]. Also, most of these methods are the variants of Mean-Teacher framework [46], which encourages invariant predictions for perturbed inputs, like combining with uncertainty [60] and calculating different levels of consistency [4, 33].

Topology-aware image segmentation. There are existing methods enforcing segmentation to have correct topology [6, 15, 17, 18, 20, 41, 44, 49, 51, 56]. These methods compare the predictions and ground truth (GT) in terms of their topology, using differentiable loss functions based on tools such as persistent homology [6, 18, 44], discrete Morse theory [16, 19, 20], homotopy warping [17], topological interactions [15], and centerline comparison [41, 51]. Despite the success of these topology-aware segmentation methods, they rely heavily

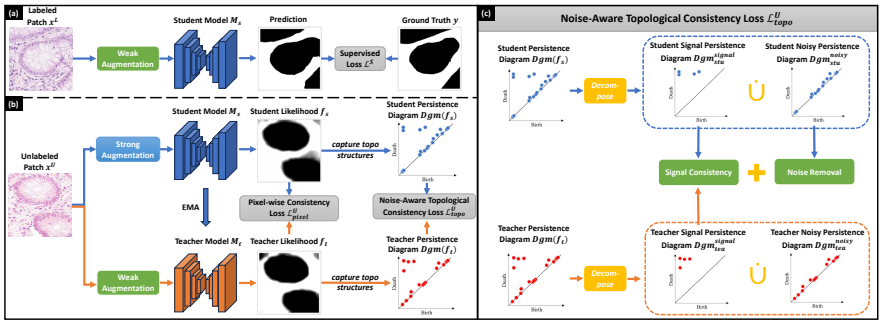


Fig. 3: An overview of our method. (**a**) denotes the labeled workflow. The student model learns from labeled images via the supervised loss $\mathcal{L}^S$. (**b**) denotes the unlabeled workflow. The student model learns from unlabeled images using $\mathcal{L}^U$, which consists of pixel-wise consistency loss $\mathcal{L}^U_{\text{pixel}}$ and noise-aware topological consistency loss $\mathcal{L}^U_{\text{topo}}$. (**c**) shows the details of our proposed noise-aware topological consistency loss $\mathcal{L}^U_{\text{topo}}$, which encompasses our decomposition and optimal matching strategy, resulting in signal topology consistency loss $\mathcal{L}^U_{\text{topo-cons}}$ and noisy topology removal loss $\mathcal{L}^U_{\text{topo-rem}}$.

on well-annotated, topologically correct labels, as well as the explicit topological information extracted from these labels. These methods are not suitable for a semi-supervised setting with limited annotations. Clough *et al.* [6] assume a fixed topology for input data and use a topology-preserving loss in a semi-supervised setting. However, their assumption is too strong and cannot adapt to histopathology images, where at different locations we have different numbers of glands/nuclei. Our work aims to break such limitations by unearthing essential topological information from the vast amount of unlabeled images.

3 Proposed Method

In this section, we first provide an overview of our proposed method in Sec. 3.1. Then, we give a brief introduction to the background of persistent homology in Sec. 3.2. Finally, we introduce our TopoSemiSeg in Sec. 3.3.

In SemiSL, we have a small set of labeled training samples and a much larger set of unlabeled samples. Let $\mathcal{D}_L = \{(x^L_1, y_1), (x^L_2, y_2), ..., (x^L_{N_L}, y_{N_L})\}$ be the dataset of N_L labeled samples, and $\mathcal{D}_U = \{x^U_1, x^U_2, ..., x^U_{N_U}\}$ be the unlabeled dataset of N_U images, where $N_L << N_U$. x^U_i denotes the i-th unlabeled image and x^L_i denotes the i-th labeled image with its corresponding pixel-wise label y_i.

The objective of SemiSL is to unearth the rich information within the unlabeled data, accompanied by limited guidance from labeled data. Most existing works only consider pixel-wise accuracy, ignoring the importance of topological correctness. Here, we take both of them into consideration.

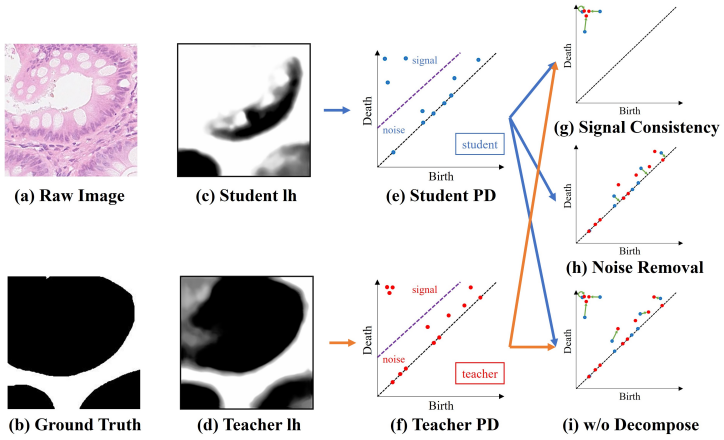


Fig. 4: Inituition of our decomposition and matching strategy. (**a**) the raw image. (**b**) the ground truth, included for reference. (**c**) the student likelihood (lh). (**d**) the teacher likelihood. (**e**) decomposition of the persistence diagram of the student likelihood. The **purple** line demonstrates the decomposition. (**f**) decomposition of the persistence diagram (tPD) of teacher likelihood. (**g**) the consistency between the signal topology. **Green** arrows show the matching process. (**h**) the noisy topology removal process. (**i**) the matching process without decomposition.

3.1 Overview of the Method

Fig. 3 provides an overview of our method. We adopt the popular teacher-student framework [46] in SemiSL. This framework contains two networks – a student and a teacher – with identical architecture. We denote the student network as M_s, parameterized by θ_s, and the teacher network as M_t, parameterized by θ_t. The student network learns from the teacher network. It is trained by minimizing the supervised loss $\mathcal{L}^S$ on the labeled data and the unsupervised loss $\mathcal{L}^U$ on the unlabeled data. More details can be found in Fig. 3(a) and (b). The overall training objective is formulated as

$$\mathcal{L} = \mathcal{L}^S + \mathcal{L}^U \tag{1}$$

To make full use of limited annotations, $\mathcal{L}^S$ is defined as the combination of cross-entropy loss (ℓ_{CE}) and Dice loss (ℓ_{Dice}) [45] between the predictions and the labels:

$$\mathcal{L}^S(D_L, M_s) = \sum_{i=1}^{N_L} \left[\lambda_1^L \ell_{CE}(M_s(x_i^L), y_i) + \lambda_2^L \ell_{Dice}(M_s(x_i^L), y_i)\right]$$

where $\lambda_*^L$are adjustable weights.

For unlabeled data, we apply strong (A_{strong}) and weak (A_{weak}) augmentations before passing them as input to the student and teacher networks, respectively. The unsupervised loss enforces the consistency between predictions of the

student and teacher models. It consists of two loss terms: pixel-wise consistency loss ($\mathcal{L}^U_{\text{pixel}}$) and the noise-aware topological consistency loss $\mathcal{L}^U_{\text{topo}}$.

$$\mathcal{L}^U = \lambda^U_1 \mathcal{L}^U_{\text{pixel}} + \lambda^U_2 \mathcal{L}^U_{\text{topo}} \tag{2}$$

where λ^U_* are adjustable weights.

We formulate the pixel-wise consistency loss as the cross-entropy (CE) loss between the outputs of the student and teacher models:

$$\mathcal{L}^U_{\text{pixel}}(D_U, M_s, M_t) = \sum_{i=1}^{N_U} \ell_{CE}(M_s(A_{strong}(x_i^U)), M_t(A_{weak}(x_i^U))) \tag{3}$$

The second part of $\mathcal{L}^U$ is the noise-aware topological consistency loss $\mathcal{L}^U_{\text{topo}}$, which is crucial for learning a robust topological representation from the unlabeled data. It will be explained in detail the next subsection.

During the training phase, the student network's parameters θ_s are updated by minimizing the overall loss (Eq. (1)). We update the teacher model's parameters θ_t based on the student model's parameters using exponential moving average (EMA) [46]. In particular, at the $(\tau+1)^{\text{th}}$ epoch, θ_t is updated as $\theta_t(\tau+1) = \alpha\theta_t(\tau) + (1-\alpha)\theta_s(\tau+1)$ where α is the EMA decay controlling the update rate.

3.2 Background: Persistent Homology

In algebraic topology [36], persistent homology [9] has emerged as a powerful tool for analyzing the topology of various kinds of real-world data, including images. It tracks the evolution of all topological structures, such as connected components and loops. All the topological structures and their birth/death times are captured in a so-called *persistence diagram*, providing a multi-scale topological representation (See Fig. 2).

A persistence diagram (PD) consists of multiple dots in a 2-dimensional plane, referred to as *persistent dots*. Each persistent dot $p \in Dgm(f)$ represents a topological structure. Its two coordinates denote the birth and death filtration values for the corresponding topological structure, i.e., $p = (b, d)$, where $b = birth(p)$ and $d = death(p)$. More details are in the Supplementary.

3.3 Noise-aware Topological Consistency Loss

We propose a noise-aware topological consistency loss to ensure that the teacher and the student models make consistent predictions in terms of topology. Given the likelihood maps of both the teacher and the student models, f_t and f_s, we first compute the persistence diagrams, $Dgm(f_t)$ and $Dgm(f_s)$. However, directly comparing the two diagrams is not desirable. As shown in Fig. 4, without supervision, both the student persistence diagram and the teacher persistence diagram are quite noisy. Direct comparison of the two diagrams can create a lot of unnecessary matching between the noisy structures. This will cause inefficiency in learning, and can potentially even derail the whole training.

To address this challenge, we propose to decompose $Dgm(f_s)$ and $Dgm(f_t)$ into signal and noise parts. The signal part is used to enforce teacher-student consistency via a signal topology consistency loss. The noise part will be removed through a novel noisy topology removal loss.

Signal-Noise Decomposition of a Persistence Diagram. We would like to decompose a diagram into signal and noise parts. However, in reality, without the ground truth, the decomposition cannot be guaranteed to be accurate. Hence, we use the classic measure of *persistence* to decide whether a dot in the persistence diagram is a signal or noise.

For a persistent dot $p \in Dgm(f)$, its persistence is simply its life span, i.e., the difference between its death and birth time: $per(p) = death(p) - birth(p)$. Persistence is a good heuristic approximating the significance of a topological structure; the greater the persistence, the longer the structure exists through filtration, and the more likely the structure is a true signal. This is theoretically justified. The celebrated stability theorem [7,8] implies that low-persistence dots are much easier to be "shed off" through perturbation of the input function f.

Formally, using a predetermined threshold ϕ, we decompose $Dgm(f)$ into disjoint signal and noise persistence diagrams based on the persistence:

$$\begin{aligned} Dgm(f) &= Dgm(f)^{signal} \dot{\bigcup} Dgm(f)^{noise} \\ Dgm(f)^{signal} &= \{p \in Dgm(f) \mid per(p) > \phi\} \\ Dgm(f)^{noise} &= \{p \in Dgm(f) \mid per(p) \leq \phi\} \end{aligned}$$

where $\dot{\bigcup}$ denotes the disjoint union. We apply the same decomposition to both teacher and student model outputs, acquiring their signal and noise parts respectively. The threshold ϕ is tuned empirically. These signal/noisy diagrams for the outputs will be used for $\mathcal{L}^U_{\text{topo}}$ in Eq. (2).

Signal Topology Consistency Loss. After the decomposition of both persistence diagrams, we obtain Dgm^{signal}_{stu} and Dgm^{signal}_{tea} representing the meaningful topological signals. Our first topology-aware loss is to ensure the two signal diagrams are the same. Similar to previous topological losses [18], we will use the classic Wasserstein distance between the two diagrams. Note: for any diagram $Dgm(g)$, we regard it as the generalized persistence diagram[1].

Definition 1 (Wasserstein distance between PDs [8]). *Given two diagrams $Dgm(g)$ and $Dgm(h)$, the p-th Wasserstein distance between them is defined as:*[2]

$$W_p(Dgm(g), Dgm(h)) = \left(\inf_{\gamma \in \Gamma} \sum_{x \in Dgm(g)} ||x - \gamma(x)||^p \right)^{\frac{1}{p}}$$

[1] A generalized persistence diagram is a countable multiset of points in $\mathbb{R}^2$ along with the diagonal $\Delta = \{(b, d) \mid b = d\}$, where each dot on the diagonal has infinite multiplicity.

[2] For ease of exposition, we change the original formulation and use the 2-norm instead of infinity norm for $||x - \gamma(x)||$. The difference is bounded by a $\sqrt{2}/2$ constant factor.

where Γ represents all bijections from $Dgm(g)$ to $Dgm(h)$.

See Fig. 4(g) and (h) for an illustration. The Wasserstein distance essentially finds an optimal matching between dots of the two diagrams. Unmatched dots are matched to their projection on the diagonal line. The distance is computed by aggregating over distance between all the matched pairs of dots. The optimal matching, as well as the distance, can be computed using either the classic Hungarian method, or more advanced algorithms [27, 29].

Next, we write the signal topology consistency loss in terms of the student's likelihood map, f_s. Denote by γ^* the optimal matching between Dgm_{stu}^{signal} and Dgm_{tea}^{signal}. Each student persistent dot $p_{stu}^{signal} \in Dgm_{stu}^{signal}$ is matched to either a teacher persistent dot, or its projection on the diagonal. We can now formulate our signal topology consistency loss $\mathcal{L}_{\text{topo-cons}}^{U}$ as the squared distance between every student signal dot and its match:

$$\mathcal{L}_{\text{topo-cons}}^{U} = \sum_{p \in Dgm_{stu}^{signal}} ||p - \gamma^*(p)||^2 \tag{4}$$

We still have to write the loss in terms of the student likelihood map. Note that in persistent homology, the birth and death times of every persistent dot are the function values of certain critical points. See Supplementary for more details and illustrations. For each 0-dimensional persistent dot p in a student diagram, the birth is at a local maxima x_p^b and the death is at a saddle point x_p^d, formally, $birth(p) = f_s(x_p^b)$ and $death(p) = f_s(x_p^d)$. Substituting into Eq. (4), we have

$$\mathcal{L}_{\text{topo-cons}}^{U}(f_s) = \sum_{p \in Dgm_{stu}^{signal}} \{[f_s(x_p^b) - birth(\gamma^*(p))]^2 + [f_s(x_p^d) - death(\gamma^*(p))]^2\} \tag{5}$$

which can be optimized with respect to the student network's parameters θ_s.

Noisy Topology Removal Loss. So far, we have introduced how to decompose the diagram and how the signal part of the diagrams can be used to enforce topological consistency. We also introduce a loss to remove the noisy topology from the student likelihood map. This turns out to be very powerful in practice: by removing the topological noise, we can stabilize the output of the student network, and eventually also stabilize the teacher network via EMA.

Our noisy topology removal loss is based on the concept of *Total Persistence*, which essentially measures the total amount of information a diagram carries. By minimizing the total persistence of the noise diagram, we are effectively removing all noise dots.

Definition 2 (Total Persistence [8]). *Given a persistence diagram, $Dgm(g)$, the p-th total persistence is :*

$$P_{total}(Dgm(g)) = \sum_{x \in Dgm(g)} [death(x) - birth(x)]^p \tag{6}$$

Similar to the consistency loss, we can define the loss in terms of the student likelihood map as follows:

$$\mathcal{L}^{U}_{\text{topo-rem}}(f_s) = P_{total}(Dgm^{noise}_{stu}) = \sum_{p \in Dgm^{noise}_{stu}} \left[f_s(x^b_p) - f_s(x^d_p)\right]^2 \quad (7)$$

The final noise-aware topological consistency loss $\mathcal{L}^{U}_{\text{topo}}$ becomes the sum of the two topology-aware loss terms: $\mathcal{L}^{U}_{\text{topo-cons}}$ and $\mathcal{L}^{U}_{\text{topo-rem}}$,

$$\mathcal{L}^{U}_{\text{topo}} = \mathcal{L}^{U}_{\text{topo-cons}} + \mathcal{L}^{U}_{\text{topo-rem}} \quad (8)$$

See Supplementary for the illustration of the differentiability of these two losses.

4 Experiments

We conduct extensive experiments on three public and widely used histopathology image datasets. We compare our method against SoTA semi-supervised segmentation methods on both pixel- and topology-wise evaluation metrics. **Implementation details** are in the Supplementary.

4.1 Datasets

We evaluate our proposed method on **Colorectal Adenocarcinoma Gland (CRAG)** [13], **Gland Segmentation in Colon Histology Images Challenge (GlaS)** [42], and **Multi-Organ Nuclei Segmentation (MoNuSeg)** [28]. More details are provided in the Supplementary.

4.2 Evaluation Metrics

We select three widely used pixel-wise evaluation metrics, **Object-level Dice coefficient (Dice_Obj)** [55], **Intersection over Union (IoU)** and **Pixel-wise accuracy**. Meanwhile, topology-relevant metrics measure the structural accuracy. Hence, we also select three topological evaluation metrics, **Betti Error** [18], **Betti Matching Error** [44], and **Variation of Information (VOI)** [34]. More details are provided in the Supplementary.

4.3 Results: Comparison with SoTA SemiSL methods

We conduct experiments on different fractions of labeled data, specifically, 10% and 20%. Training UNet++ [65] on 100% of the labeled data is treated as the performance upper bound. To indicate the effectiveness and superiority of our method, we select several SoTA semi-supervised methods for comparison both from pixel and topological perspectives. Quantitative results are shown in Tab. 1, and qualitative results are shown in Fig. 5. We discuss more below.

Table 1: Quantitative results on three histopathology image datasets. We compare our method with several state-of-the-art semi-supervised medical image segmentation methods on two settings of 10% and 20% labeled data. The best results are highlighted in **bold**, and * indicates that the method is re-implemented by ourselves.

Dataset	Labeled Ratio (%)	Method	Pixel-Wise			Topology-Wise		
			Accuracy ↑	Dice Obj ↑	IoU ↑	Betti Error ↓	Betti Matching Error ↓	VOI ↓
CRAG	10%	MT [46]	0.862	0.821	0.713	2.238	62.250	0.977
		EM [48]	0.834	0.789	0.688	2.178	80.100	1.027
		UA-MT [60]	0.874	0.837	0.728	1.703	66.450	0.947
		HCE* [25]	0.891	0.862	0.773	1.286	35.530	0.861
		URPC [33]	0.872	0.829	0.728	1.732	74.600	0.883
		XNet [64]	0.895	0.872	0.781	0.578	15.050	0.773
		TopoSemiSeg	**0.905**	**0.884**	**0.798**	**0.227**	**10.475**	**0.758**
	20%	MT [46]	0.887	0.858	0.759	2.603	99.025	0.867
		EM [48]	0.903	0.869	0.776	1.933	75.225	0.798
		UA-MT [60]	0.895	0.859	0.765	1.822	70.850	0.829
		HCE* [25]	0.910	0.881	0.809	0.875	17.400	0.769
		URPC [33]	0.881	0.849	0.744	2.489	99.500	0.912
		XNet [64]	0.907	0.883	0.792	0.422	10.900	0.735
		TopoSemiSeg	**0.912**	**0.898**	**0.820**	**0.226**	**8.575**	**0.709**
	100%	Fully-supervised	0.945	0.928	0.869	0.149	5.650	0.547
GlaS	10%	MT [46]	0.815	0.790	0.671	2.392	31.125	1.079
		EM [48]	0.833	0.819	0.708	1.431	19.188	1.051
		UA-MT [60]	0.728	0.845	0.829	2.086	26.650	1.018
		HCE* [25]	0.859	0.852	0.762	0.631	11.950	0.953
		URPC [33]	0.829	0.849	0.751	1.155	19.588	0.968
		XNet [64]	0.871	0.874	0.786	0.843	14.238	0.917
		TopoSemiSeg	**0.890**	**0.878**	**0.797**	**0.551**	**8.300**	**0.811**
	20%	MT [46]	0.870	0.863	0.771	2.126	29.963	0.925
		EM [48]	0.861	0.865	0.776	1.255	17.275	0.841
		UA-MT [60]	0.874	0.866	0.781	1.123	18.038	0.869
		HCE* [25]	0.864	0.871	0.779	0.871	16.213	0.824
		URPC [33]	0.876	0.878	0.794	0.759	14.350	0.837
		XNet [64]	0.886	0.884	0.804	0.735	10.188	0.816
		TopoSemiSeg	**0.896**	**0.895**	**0.818**	**0.510**	**9.825**	**0.808**
	100%	Fully-supervised	0.920	0.917	0.853	0.473	7.125	0.686
MoNuSeg	10%	MT [46]	0.889	0.748	0.607	10.210	292.857	0.874
		EM [48]	0.901	0.757	0.612	10.339	257.071	0.844
		UA-MT [60]	0.898	0.741	0.594	10.227	255.428	0.862
		HCE* [25]	0.882	0.761	0.617	14.210	377.928	0.890
		CCT [37]	0.892	0.766	0.624	8.063	225.500	0.839
		URPC [33]	0.896	0.774	0.633	6.829	214.428	0.863
		TopoSemiSeg	**0.909**	**0.783**	**0.646**	**6.661**	**196.357**	**0.789**
	20%	MT [46]	0.896	0.767	0.624	12.522	246.786	0.873
		EM [48]	0.905	0.777	0.637	7.160	198.571	0.805
		UA-MT [60]	0.904	0.772	0.632	9.406	246.857	0.826
		HCE* [25]	0.899	0.771	0.642	13.330	311.143	0.829
		CCT [37]	0.903	0.785	0.648	7.977	207.857	0.832
		URPC [33]	**0.909**	0.779	0.639	5.325	193.429	0.788
		TopoSemiSeg	0.908	**0.793**	**0.653**	**4.250**	**188.642**	**0.787**
	100%	Fully-supervised	0.929	0.817	0.702	2.491	142.429	0.657

Quantitative Results. For a comprehensive comparison, we select several classical and recent SoTA SemiSL methods like MT [46], EM [48], UA-MT [60], HCE [25], URPC [33], XNet [64] and CCT [37]. Note that HCE is re-implemented by ourselves due to code unavailability. As shown in Tab. 1, our method not only achieves comparable performance on pixel-wise evaluation metrics but also achieves the best results on all topology-wise metrics. This indicates that our proposed TopoSemiSeg is able to unearth and utilize topological information in unlabeled data well, without sacrificing pixel-level performance.

Qualitative Results. In Fig. 5, we provide the qualitative results of the methods on 20% labeled data for each dataset. Compared to other SoTA SemiSL methods, our method does better where topological errors are prone to occur, as shown in the **red** boxes. The proposed TopoSemiSeg ensures topological integrity: by enforcing signal consistency, we can maintain the thin separation between the densely distributed glands and cells. Additionally, the noise removal

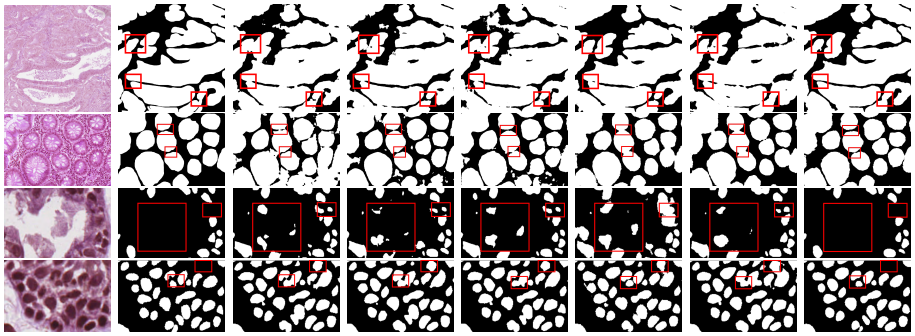


Fig. 5: Qualitative results on three histopathology image datasets using 20% labeled data for training. Locations prone to topological errors are shown within **red** boxes. Row 1: CRAG, Row 2: GlaS, Rows 3 & 4: MoNuSeg. Zoom in for better views.

component of our loss minimizes the occurrence of false positive cells, as can be seen in Row 3. This is in contrast to the results obtained from the other baseline methods, which contain a discernible presence of noise and unoccupied interspaces in and around the glandular and cellular structures. Our method can effectively address and rectify these issues. This is because we not only focus on the signal topology which should be preserved, but also remove all the noisy topology during training, thus making the model learn more robust and accurate topological representations from the unlabeled data. More qualitative results are provided in the Supplementary.

4.4 Ablation Studies

We conduct experiments to illustrate the effectiveness and robustness of our hyper-parameters selections and experimental settings. All experiments are performed on the CRAG dataset using 20% labeled data. **To save space, additional ablation studies are provided in ?? in the Supplementary.**

Weight of Noise-aware Topological Consistency loss λ_2^U. We study the effect of the weight of the noise-aware topological consistency loss λ_2^U introduced in Eq. (2). As shown in Tab. 2, at $\lambda_2^U = 0.002$, the model achieves the best Object-level Dice coefficient, Betti Matching Error, and VOI. Additionally, a reasonable range of λ_2^U always results in improvement. This demonstrates the efficacy and robustness of the proposed method.

Robustness of Persistence Threshold ϕ. To compute the noise-aware topological consistency loss, we define a persistence threshold ϕ to decompose both persistence diagrams into signal and noise parts. We conduct experiments on different values of ϕ. As we can see from Tab. 3, our method is not sensitive to the value of ϕ and a wide range of ϕ (from 0.5 to 0.9) results in improvements on topological metrics. This demonstrates the robustness of our method with respect to perturbations. Moreover, note that $\phi = 0$ means without decomposition

Table 2: Ablation study on loss weight λ_2^U.

λ_2^U	Pixel-Wise	Topology-Wise		
	Dice_Obj ↑	Betti Error ↓	Betti Matching Error ↓	VOI ↓
0	0.887	0.306	12.125	0.783
0.001	0.874	0.217	12.175	0.736
0.002	**0.898**	0.226	**8.575**	**0.709**
0.005	0.889	**0.213**	9.875	0.739
0.008	0.896	0.235	9.700	0.722
0.01	0.873	0.277	9.725	0.754

(direct matching). As in Tab. 3, this significantly hurts the performance. This is because both teacher and student predictions are noisy. Direct matching of diagrams will force the student to learn from many noisy structures, confusing the model and resulting in a significant performance drop.

Table 3: Ablation study on persistence threshold ϕ.

ϕ	Pixel-Wise	Topology-Wise		
	Dice_Obj ↑	Betti Error ↓	Betti Matching Error ↓	VOI ↓
0	0.880	0.317	14.025	0.762
0.50	0.881	0.241	8.950	0.753
0.60	0.895	0.219	9.600	0.725
0.70	**0.898**	0.226	**8.575**	**0.709**
0.80	0.896	**0.209**	9.000	0.722
0.90	0.889	0.231	10.150	0.717

Generalizability to Different Backbones. We verify the generalizability of our method by performing experiments on three different backbones, UNet [38], PSPNet [63], and DeepLabV3+ [5], keeping the same values of hyper-parameters for each. Tab. 4 shows that our method is robust to backbone selections and can obtain performance improvements with each of them. For those with poor topological performances, like PSPNet [63], our method significantly reduces the number of topological errors. This proves the effectiveness and generalizability of our method in that it can facilitate capturing topological information from the unlabeled data irrespective of the backbone.

Table 4: Comparison on different backbones. BE and BME respectively denote Betti Error and Betti Matching Error.

Method	Pixel-Wise	Topology-Wise		
	Dice_Obj↑	BE↓	BME↓	VOI↓
UNet [38]	0.892	0.266	10.775	0.790
UNet [38]+Ours	**0.893**	**0.236**	**8.700**	**0.722**
PSPNet [63]	0.773	1.809	70.625	1.337
PSPNet [63]+Ours	**0.775**	**1.021**	**44.150**	**1.040**
DeepLabV3+ [5]	0.883	0.293	14.000	0.725
DeepLabV3+ [5]+Ours	**0.891**	**0.265**	**11.725**	**0.713**
UNet++ [65]	0.887	0.306	12.125	0.783
UNet++ [65]+Ours	**0.898**	**0.226**	**8.575**	**0.709**

Ablation study on loss components. We conduct an ablation study on the signal consistency loss and noise removal loss to verify their effectiveness. The signal topology $\mathcal{L}^{\mathrm{U}}_{\text{topo-cons}}$ is used to guide the student to learn the meaningful structures from the teacher. Meanwhile, we use the noise removal loss $\mathcal{L}^{\mathrm{U}}_{\text{topo-rem}}$ to directly train the student model to avoid noise in its prediction. In Tab. 5, we present the ablation study of these two components and from the performance, we can see that both components are necessary for us to enhance the segmentation quality.

Table 5: Ablation study on loss components.

$\mathcal{L}^{\mathrm{U}}_{\text{topo-cons}}$	$\mathcal{L}^{\mathrm{U}}_{\text{topo-rem}}$	Dice_Obj ↑	BE ↓	BME ↓	VOI ↓
✗	✗	0.887	0.306	12.125	0.783
✓	✗	0.890	0.275	10.150	0.715
✗	✓	0.888	0.230	10.525	0.723
✓	✓	**0.898**	**0.226**	**8.575**	**0.709**

Limitations. In the teacher-student framework, the prediction of the teacher model is not always accurate. On the contrary, the teacher model's output may be unreliable sometimes. How to alleviate the uncertainty of the teacher model's topological output will be the one of topics we will be exploring in the future. Further, the current decomposition threshold is set manually and needs to be set adaptively according to the data distribution. Finally, our current implementation is only available for 2D datasets, extending the proposed method to 3D scenarios with acceptable computational cost will be of great value in practice. For the failure cases, we observe that when the objects are extremely densely distributed, our model sometimes cannot split the glands in close proximity.

5 Conclusion

This work introduces TopoSemiSeg, the first semi-supervised method that learns topological representation from unlabeled histopathology images. It consists of a novel and differentiable noise-aware topological consistency loss integrated into the teacher-student framework. We propose to decompose the calculated persistence diagrams into true signal and noisy components, and respectively formulate signal consistency and noise removal losses from them. These losses enforce the model to learn a robust representation of topology from unlabeled data and can be incorporated into any variant of the teacher-student framework. Extensive experiments on several histopathology image datasets show the effectiveness of the proposed method on both pixel- and topology-wise evaluation metrics.

Acknowledgements. We thank the anonymous reviewers for their constructive feedback. This research was partially supported by the National Science Foundation (NSF) grant CCF-2144901, the National Institute of General Medical Sciences (NIGMS) grant R01GM148970, and the Stony Brook Trustees Faculty Award.

References

1. Basak, H., Yin, Z.: Pseudo-label guided contrastive learning for semi-supervised medical image segmentation. In: CVPR (2023)
2. Berthelot, D., Carlini, N., Goodfellow, I., Papernot, N., Oliver, A., Raffel, C.A.: Mixmatch: A holistic approach to semi-supervised learning. In: NeurIPS (2019)
3. Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M.: Swin-unet: Unet-like pure transformer for medical image segmentation. In: ECCV (2022)
4. Chen, G., Ru, J., Zhou, Y., Rekik, I., Pan, Z., Liu, X., Lin, Y., Lu, B., Shi, J.: Mtans: Multi-scale mean teacher combined adversarial network with shape-aware embedding for semi-supervised brain lesion segmentation. NeuroImage (2021)
5. Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: ECCV (2018)
6. Clough, J.R., Byrne, N., Oksuz, I., Zimmer, V.A., Schnabel, J.A., King, A.P.: A topological loss function for deep-learning based image segmentation using persistent homology. TPAMI (2020)
7. Cohen-Steiner, D., Edelsbrunner, H., Harer, J.: Stability of persistence diagrams. In: Proceedings of the twenty-first annual symposium on Computational geometry (2005)
8. Cohen-Steiner, D., Edelsbrunner, H., Harer, J., Mileyko, Y.: Lipschitz functions have l p-stable persistence. Foundations of Computational Mathematics (2010)
9. Edelsbrunner, Letscher, Zomorodian: Topological persistence and simplification. Discrete & Computational Geometry (2002)
10. Edelsbrunner, H., Harer, J.L.: Computational topology: an introduction. American Mathematical Society (2022)
11. Fang, K., Li, W.J.: Dmnet: difference minimization network for semi-supervised segmentation in medical images. In: MICCAI (2020)
12. Fleming, M., Ravula, S., Tatishchev, S.F., Wang, H.L.: Colorectal carcinoma: Pathologic aspects. Journal of Gastrointestinal Oncology (2012)
13. Graham, S., Chen, H., Gamper, J., Dou, Q., Heng, P.A., Snead, D., Tsang, Y.W., Rajpoot, N.: Mild-net: Minimal information loss dilated network for gland instance segmentation in colon histology images. MedIA (2019)
14. Grandvalet, Y., Bengio, Y.: Semi-supervised learning by entropy minimization. In: NeurIPS (2004)
15. Gupta, S., Hu, X., Kaan, J., Jin, M., Mpoy, M., Chung, K., Singh, G., Saltz, M., Kurc, T., Saltz, J., et al.: Learning topological interactions for multi-class medical image segmentation. In: ECCV (2022)
16. Gupta, S., Zhang, Y., Hu, X., Prasanna, P., Chen, C.: Topology-aware uncertainty for image segmentation. In: NeurIPS (2023)
17. Hu, X.: Structure-aware image segmentation with homotopy warping. In: NeurIPS (2022)
18. Hu, X., Li, F., Samaras, D., Chen, C.: Topology-preserving deep image segmentation. In: NeurIPS (2019)
19. Hu, X., Samaras, D., Chen, C.: Learning probabilistic topological representations using discrete morse theory. In: ICLR (2023)
20. Hu, X., Wang, Y., Fuxin, L., Samaras, D., Chen, C.: Topology-aware segmentation using discrete morse theory. In: ICLR (2021)
21. Huang, W., Chen, C., Xiong, Z., Zhang, Y., Chen, X., Sun, X., Wu, F.: Semi-supervised neuron segmentation via reinforced consistency learning. TMI (2022)

22. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods (2021)
23. Jeong, J., Lee, S., Kim, J., Kwak, N.: Consistency-based semi-supervised learning for object detection. In: NeurIPS (2019)
24. Jiao, R., Zhang, Y., Ding, L., Cai, R., Zhang, J.: Learning with limited annotations: a survey on deep semi-supervised learning for medical image segmentation. arXiv preprint arXiv:2207.14191 (2022)
25. Jin, Q., Cui, H., Sun, C., Zheng, J., Wei, L., Fang, Z., Meng, Z., Su, R.: Semi-supervised histological image segmentation via hierarchical consistency enforcement. In: MICCAI (2022)
26. Jin, Y., Wang, J., Lin, D.: Semi-supervised semantic segmentation via gentle teaching assistant. In: NeurIPS (2022)
27. Kerber, M., Morozov, D., Nigmetov, A.: Geometry helps to compare persistence diagrams. In: 2016 Proceedings of the Eighteenth Workshop on Algorithm Engineering and Experiments (ALENEX). SIAM (2016)
28. Kumar, N., Verma, R., Anand, D., Zhou, Y., Onder, O.F., Tsougenis, E., Chen, H., Heng, P.A., Li, J., Hu, Z., et al.: A multi-organ nucleus segmentation challenge. TMI (2019)
29. Lacombe, T., Cuturi, M., Oudot, S.: Large scale computation of means and clusters for persistence diagrams using optimal transport. In: NeurIPS (2018)
30. Li, C., Hu, X., Abousamra, S., Chen, C.: Calibrating uncertainty for semi-supervised crowd counting. In: ICCV (2023)
31. Li, X., Yu, L., Chen, H., Fu, C.W., Xing, L., Heng, P.A.: Transformation-consistent self-ensembling model for semisupervised medical image segmentation. TNNLS (2020)
32. Li, Y., Luo, L., Lin, H., Chen, H., Heng, P.A.: Dual-consistency semi-supervised learning with uncertainty quantification for covid-19 lesion segmentation from ct images. In: MICCAI (2021)
33. Luo, X., Wang, G., Liao, W., Chen, J., Song, T., Chen, Y., Zhang, S., Metaxas, D.N., Zhang, S.: Semi-supervised medical image segmentation via uncertainty rectified pyramid consistency. MedIA (2022)
34. Meilă, M.: Comparing clusterings by the variation of information. In: Learning Theory and Kernel Machines: 16th Annual Conference on Learning Theory and 7th Kernel Workshop, COLT/Kernel 2003 (2003)
35. Montironi, R., Mazzuccheli, R., Scarpelli, M., Lopez-Beltran, A., Fellegara, G., Algaba, F.: Gleason grading of prostate cancer in needle biopsies or radical prostatectomy specimens: contemporary approach, current clinical significance and sources of pathology discrepancies. BJU International (2005)
36. Munkres, J.R.: Elements of algebraic topology (1984)
37. Ouali, Y., Hudelot, C., Tami, M.: Semi-supervised semantic segmentation with cross-consistency training. In: CVPR (2020)
38. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI (2015)
39. Seibold, C.M., Reiß, S., Kleesiek, J., Stiefelhagen, R.: Reference-guided pseudo-label generation for medical semantic segmentation. In: AAAI (2022)
40. Shi, Y., Zhang, J., Ling, T., Lu, J., Zheng, Y., Yu, Q., Qi, L., Gao, Y.: Inconsistency-aware uncertainty estimation for semi-supervised medical image segmentation. TMI (2021)

41. Shit, S., Paetzold, J.C., Sekuboyina, A., Ezhov, I., Unger, A., Zhylka, A., Pluim, J.P., Bauer, U., Menze, B.H.: cldice-a novel topology-preserving loss function for tubular structure segmentation. In: CVPR (2021)
42. 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.: Gland segmentation in colon histology images: The glas challenge contest. MedIA (2017)
43. Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C.A., Cubuk, E.D., Kurakin, A., Li, C.L.: Fixmatch: Simplifying semi-supervised learning with consistency and confidence. In: NeurIPS (2020)
44. Stucki, N., Paetzold, J.C., Shit, S., Menze, B., Bauer, U.: Topologically faithful image segmentation via induced matching of persistence barcodes. In: ICML (2023)
45. Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Jorge Cardoso, M.: Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, Held in Conjunction with MICCAI 2017 (2017)
46. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In: NeurIPS (2017)
47. Thompson, B.H., Di Caterina, G., Voisey, J.P.: Pseudo-label refinement using superpixels for semi-supervised brain tumour segmentation. In: ISBI (2022)
48. Vu, T.H., Jain, H., Bucher, M., Cord, M., Pérez, P.: Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In: CVPR (2019)
49. Wang, F., Liu, H., Samaras, D., Chen, C.: Topogan: A topology-aware generative adversarial network. In: ECCV (2020)
50. Wang, H., Cao, P., Wang, J., Zaiane, O.R.: Uctransnet: rethinking the skip connections in u-net from a channel-wise perspective with transformer. In: AAAI (2022)
51. Wang, H., Xian, M., Vakanski, A.: Ta-net: Topology-aware network for gland segmentation. In: WACV (2022)
52. Wang, X., Yuan, Y., Guo, D., Huang, X., Cui, Y., Xia, M., Wang, Z., Bai, C., Chen, S.: Ssa-net: Spatial self-attention network for covid-19 pneumonia infection segmentation with semi-supervised few-shot learning. MedIA (2022)
53. Wu, H., Wang, Z., Song, Y., Yang, L., Qin, J.: Cross-patch dense contrastive learning for semi-supervised segmentation of cellular nuclei in histopathologic images. In: CVPR (2022)
54. Wu, Y., Ge, Z., Zhang, D., Xu, M., Zhang, L., Xia, Y., Cai, J.: Mutual consistency learning for semi-supervised medical image segmentation. MedIA (2022)
55. Xie, Y., Lu, H., Zhang, J., Shen, C., Xia, Y.: Deep segmentation-emendation model for gland instance segmentation. In: MICCAI (2019)
56. Yang, J., Hu, X., Chen, C., Tsai, C.: A topological-attention convlstm network and its application to em images. In: MICCAI (2021)
57. Yao, H., Hu, X., Li, X.: Enhancing pseudo label quality for semi-supervised domain-generalized medical image segmentation. In: AAAI (2022)
58. You, C., Dai, W., Min, Y., Liu, F., Zhang, X., Feng, C., Clifton, D.A., Zhou, S.K., Staib, L.H., Duncan, J.S.: Rethinking semi-supervised medical image segmentation: A variance-reduction perspective. In: NeurIPS (2023)
59. You, C., Zhou, Y., Zhao, R., Staib, L., Duncan, J.S.: Simcvd: Simple contrastive voxel-wise representation distillation for semi-supervised medical image segmentation. TMI (2022)
60. Yu, L., Wang, S., Li, X., Fu, C.W., Heng, P.A.: Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation. In: MICCAI (2019)

61. Zhang, W., Zhu, L., Hallinan, J., Zhang, S., Makmur, A., Cai, Q., Ooi, B.C.: Boostmis: Boosting medical image semi-supervised learning with adaptive pseudo labeling and informative active annotation. In: CVPR (2022)
62. Zhang, Z., Tian, C., Bai, H.X., Jiao, Z., Tian, X.: Discriminative error prediction network for semi-supervised colon gland segmentation. MedIA (2022)
63. Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J.: Pyramid scene parsing network. In: CVPR (2017)
64. Zhou, Y., Huang, J., Wang, C., Song, L., Yang, G.: Xnet: Wavelet-based low and high frequency fusion networks for fully-and semi-supervised semantic segmentation of biomedical images. In: ICCV (2023)
65. Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J.: Unet++: A nested u-net architecture for medical image segmentation. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018 (2018)