

WeaveSeg: Iterative Contrast-weaving and Spectral Feature-refining for Nuclei Instance Segmentation

Jiajia Li¹ Huisi Wu^{1*} Jing Qin²

¹ College of Computer Science and Software Engineering, Shenzhen University

² Centre for Smart Health, School of Nursing, The Hong Kong Polytechnic University

2400101028@mails.szu.edu.cn, hswu@szu.edu.cn, harry.qin@polyu.edu.hk

Abstract

Automatic and accurate nuclei instance segmentation in histopathology images is a fundamental task in computational pathology. It is also a very challenging task due to complex nuclei morphologies, ambiguous boundaries, and staining variations. Existing methods often struggle to precisely delineate overlapping nuclei and handle class imbalance. We introduce WeaveSeg, a novel deep learning framework that synergistically integrates two key innovations: an adaptive spectral refinement (ASR) module to enhance high-frequency boundary details often blurred by standard convolutions, and an iterative contrast-weaving (ICW) module. Guided by a specialized uncertainty-aware loss, this module leverages contrastive attention and a novel local self-adaptive mechanism to progressively resolve ambiguous instances. Extensive experiments on MoNuSeg, CoNSeP, and CPM17 demonstrate WeaveSeg’s SOTA performance over existing models. The code is available at <https://github.com/jj-sterne/WeaveSeg>.

1. Introduction

Accurate segmentation of cell nuclei within histopathological images is a fundamental task in computational pathology, crucial for advancing cancer diagnosis, treatment planning, and prognosis [8, 15, 25, 33]. The morphological characteristics of nuclei, including size, shape, texture, and spatial arrangement, provide critical information for histopathologists to evaluate tumor malignancy and guide clinical decisions [29, 35, 42]. Automating this process through deep learning models has significant potential to enhance diagnostic precision and efficiency, and thus improve treatment outcomes [18, 34, 37, 40]. Despite advances in deep learning-based medical image segmentation, the precise segmentation of nuclei in histopathological images remains a formidable challenge [21].

*Corresponding Author

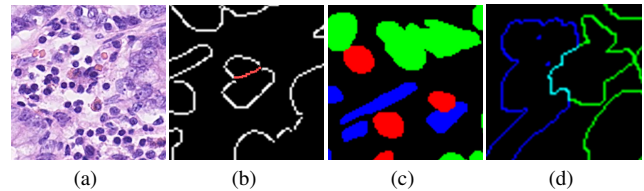


Figure 1. **Challenges in Nuclei Segmentation.** (a) a representative colorectal adenocarcinoma image; (b) ambiguous boundaries between nuclei and surrounding stroma; (c) morphological heterogeneity of nuclei; and (d) densely packed and overlapping nuclei

As illustrated in Figure 1, several inherent complexities within histopathology images impede accurate nuclei boundary delineation. First, because of weak staining, inflammation, or obscuring cellular structures, there exist many ambiguous boundaries between nuclei and the surrounding stroma (Figure 1 (b)), making it quite difficult to accurately identify these boundaries. Second, histopathological nuclei exhibit substantial morphological heterogeneity (Figure 1 (c)); variations in shape, size, and texture across cell types and disease states complicate the definition of universally applicable features. Third, nuclei frequently appear in densely packed clusters (Figure 1(d)), leading to significant overlap and occlusion, hindering the separation of individual nuclei, especially with indistinct boundaries. Furthermore, staining intensity and color balance variations across samples, along with potential class imbalance (rarity of certain nuclei types), present additional challenges.

While existing deep learning methods for nuclei segmentation have shown promise, they often fall short of comprehensively addressing these challenges. Encoder-decoder architectures, exemplified by U-Net [31] and its variants [23, 45], while effective for multi-scale feature fusion, tend to blur fine boundary details due to repeated downsampling and upsampling. Instance segmentation approaches [11], are usually confounded by severe nuclei overlap and class imbalance, resulting in missed detections or false positives. Furthermore, boundary-based methods [3, 44], while directly leveraging boundary informa-

tion, are vulnerable to noise and indistinct boundaries. Distance-based methods [8, 12, 25], designed to separate touching nuclei, still exhibit limitations when confronted with highly irregular nuclei shapes. Topology-aware approaches [24, 38], despite enhancing shape representation, do not directly address boundary ambiguity. Attention and recurrent mechanisms [7, 13, 41], aimed at improving feature learning, lack specific focus on challenges like nuclei overlap or staining variations.

A critical limitation of conventional models is their reliance on spatial-only convolutions. While effective for general feature extraction, repeated downsampling and upsampling operations inherently blur the subtle yet vital high-frequency details. Unlike in natural images, histopathological nuclei depend on these high-frequency cues for accurate delineation. Existing frequency-based methods [4, 9, 30], often address global characteristics and fail to capture the localized, boundary-specific spectral information crucial for separating instances.

To comprehensively address above limitations, in this paper, we present *WeaveSeg*, a novel deep learning model for nuclei segmentation that integrates an iterative contrast enhancement strategy with an adaptive activation mechanism, and a frequency-aware detail adaptation across multiple scales. The core of the *WeaveSeg* is the iterative contrast-weaving (ICW) module, which goes far beyond traditional simple feature fusion. ICW sufficiently exploits the intrinsic contrast between nuclei, surrounding tissue, and ambiguous boundary regions, iteratively disentangling and refining feature representations using a unique "weaving" process. This process, combined with an adaptive activation mechanism (AGLU) [1], allows the network to progressively resolve uncertainties and enhance the discriminability of nuclei features. Complementing ICW, the adaptive spectral detail refinement (ASR) module explicitly addresses frequency-domain inconsistencies that often arise from feature fusion in encoder-decoder architectures. The ASR leverages adaptive filtering to enhance both semantic consistency within nuclei and boundary sharpness. Our main contributions can be summarized as follows.

- We propose *WeaveSeg*, a framework that synergistically integrates spectral and spatial refinement. It explicitly targets blurred boundaries and overlapping instances by coupling ASR with ICW.
- We introduce two innovative components: the ASR module for adaptive, boundary-aware frequency adaptation, and the ICW module, which leverages contrastive attention and is guided by a specialized uncertainty-aware loss for feature disentanglement.
- Extensive experiments on publicly available histopathological nuclei segmentation datasets demonstrate that *WeaveSeg* achieves state-of-the-art performance.

2. Related Works

2.1. Nuclei Instance Segmentation

Early deep learning methods primarily relied on encoder-decoder architectures like U-Net [31] and its variants [27, 45]. While effective for multi-scale feature fusion, they often suffer from boundary blurring due to repeated downsampling and upsampling, particularly affecting the delineation of small and irregularly shaped nuclei. To address boundary blurring, boundary-based methods [3, 44] utilize contour information, but their performance is highly sensitive to image noise and inherent indistinctness of nuclei boundaries in histopathology images. Distance-based methods [8, 12, 25] improve separation of touching nuclei, but struggle with highly irregular shapes and often requires precise shape priors. Topology-aware approaches [24, 38] offer improvements in shape representation by incorporating topological constraints, but their ability to resolve fundamental issues like ambiguous boundaries and significant morphological variations remains limited. Attention mechanisms [13] and recurrent processing [7] are helpful in enhancing feature discriminability, but often lack explicit mechanisms to directly tackle boundary ambiguity and severe nuclei overlap, crucial factors in accurate instance segmentation.

In contrast, *WeaveSeg*'s iterative refinement with contrastive attention and uncertainty loss is proposed to comprehensively address these multifaceted challenges.

2.2. Frequency-Domain Feature Enhancement

Frequency-domain analysis offers a complementary perspective in the realm of medical image analysis. Qin et al. introduced FcaNet [30], which pioneers the application of channel attention mechanisms within the frequency domain. Subsequently, Guo et al. [9] explored the utility of adaptive frequency filters for capturing long-range dependencies in feature representations. More recently, Li et al. [20] leveraged a high-frequency prior to enhance super-resolution performance. While these methodologies effectively demonstrate the inherent benefits of incorporating frequency information, they predominantly concentrate on global frequency characteristics.

In contrast to these approaches, the ASR module of *WeaveSeg* is specifically designed to focus on localized, boundary-aware frequency information. By adaptively fusing frequency components at a granular, local level, *WeaveSeg* aims to address the critical challenge of discerning subtle textural and structural disparities between nuclei and the surrounding stroma.

2.3. Contrastive and Decoupling Learning

Contrastive learning methodologies [5, 10, 18] have garnered considerable attention and demonstrated substantial

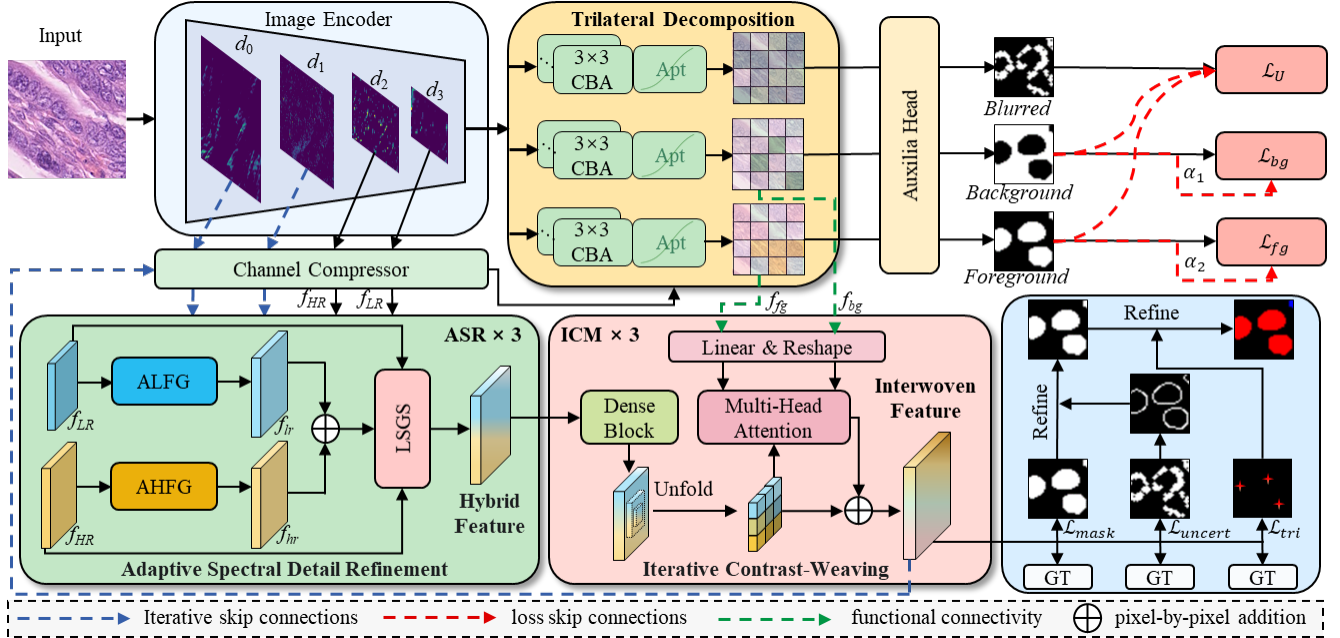


Figure 2. **Overview of the WeaveSeg framework.** The architecture features three main components: foreground, background, and uncertainty regions, and incorporates two key modules: iterative contrast-weaving (ICW) and adaptive spectral detail refinement (ASR).

promise in facilitating the learning of highly discriminative feature representations. Nevertheless, the majority of existing applications of contrastive learning within medical image segmentation [2, 19, 39] predominantly employ global or pixel-level contrastive strategies.

Distinctly, the ICW module of *WeaveSeg* uniquely integrates localized, boundary-aware contrastive learning principles. While decoupling methodologies have proven effective in enhancing feature representation capabilities, they frequently lack dedicated mechanisms to effectively manage uncertainty inherent in ambiguous regions. *WeaveSeg* directly addresses this critical limitation through the incorporation of a specialized uncertainty-aware loss function, explicitly designed for the sophisticated modeling of ambiguity associated with nuclei boundaries and overlaps.

3. Methodology

3.1. Overall architecture

WeaveSeg (Figure 2) is an encoder-decoder architecture for precise nuclei instance segmentation in histopathology images. The encoder extracts multi-scale feature hierarchies from input imagery. Subsequently, the decoder processes these features via two pathways: frequency-aware refinement and contrast enhancement. Notably, both pathways converge within the iterative contrast-weaving (ICW) module for core feature integration.

The adaptive spectral detail refinement (ASR) module (Section 3.2) lies within the first pathway. ASR selectively enhances high-frequency components associated with nu-

clei boundaries and fine textures, while preserving low-frequency information crucial for overall nuclei structure, providing a refined feature representation. The second pathway decouples the encoder features into foreground (nuclei), background (surrounding tissue), and uncertain region components, facilitating targeted contrast enhancement in ambiguous boundary areas.

The iterative contrast-weaving module (Section 3.3) is central to *WeaveSeg* as the core processing unit. ICW iteratively refines features via foreground-background contrast, guided by ASR’s frequency-adapted features. Critically, a closed-loop iteration feeds intermediate ICW outputs back to ASR. This ASR-ICW interplay (ASR refines spectra and ICW enhances contrast) ultimately resolves ambiguities, generating a precise nuclei instance segmentation map as the final ICW output.

3.2. Adaptive Spectral Detail Refinement (ASR)

Standard upsampling and feature fusion methods often struggle to preserve the high-frequency details crucial for accurate nuclei delineation, consequently leading to blurred boundaries. To mitigate this issue, we introduce the ASR module, illustrated in Figure 3(a), which directly integrates frequency-aware processing into the feature fusion and up-sampling pipeline. ASR is designed to promote both intra-nuclei semantic consistency and sharp inter-nuclei boundaries.

ASR takes as input a low-resolution, high-level feature map, $\mathbf{F}_{\text{high}} \in \mathbb{R}^{C \times H \times W}$, from a deeper decoder layer, and a high-resolution, low-level feature map, $\mathbf{F}_{\text{low}} \in$

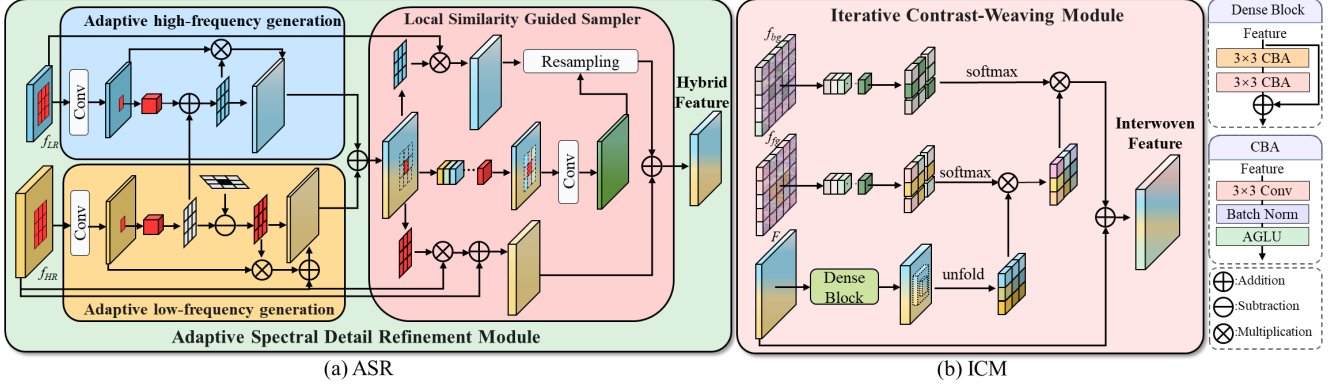


Figure 3. **Detailed architecture of the proposed modules.** (a) ASR enhances features by adaptively fusing frequency components and employing Local Similarity Guided Sampling for high-fidelity upsampling. (b) ICW iteratively refines features using contrastive attention, dense connections, and an adaptive gated linear unit (AGLU) for modulation.

$\mathbb{R}^{C \times 2H \times 2W}$, from a shallower decoder layer. It produces a refined output, $\mathbf{F}_{\text{out}} \in \mathbb{R}^{C \times 2H \times 2W}$, enriched with high-frequency details.

The ASR module adaptively generates both low- and high-frequency feature components through distinct branches. For *low-frequency enhancement*, aiming to promote intra-nuclei consistency, ASR learns a spatially-varying low-pass filter, $\mathbf{W}_{\text{LP}} \in \mathbb{R}^{K^2 \times 2H \times 2W}$. This learnable filter, implicitly modulated by a Hamming window due to localized convolution, allows for adaptive smoothing based on feature content. A low-pass filtered version of \mathbf{F}_{high} , denoted $\mathbf{F}_{\text{high}}^{\text{LP}}$, is then obtained via depthwise convolution (\otimes) and pixel shuffling:

$$\mathbf{F}_{\text{high}}^{\text{LP}} = \text{PixelShuffle}(\mathbf{W}_{\text{LP}} \otimes \mathbf{F}_{\text{high}}) \quad (1)$$

Complementing this, to effectively recover fine details, ASR employs Local Similarity-Guided Resampling (LSGR). LSGR leverages a local cosine similarity map, $\mathbf{S} \in \mathbb{R}^{8 \times 2H \times 2W}$, computed from the combined features \mathbf{F}_{comb} . This similarity map guides the prediction of spatial offsets, $\mathbf{O} \in \mathbb{R}^{2G \times 2H \times 2W}$, enabling content-aware resampling that preserves high-frequency details:

$$\mathbf{O} = \text{Conv}_{3 \times 3}([\mathbf{F}_{\text{comb}}, \mathbf{S}]) \quad (2)$$

These offsets resample $\mathbf{F}_{\text{high}}^{\text{LP}}$ to produce a refined high-level map, $\mathbf{F}_{\text{high}}^{\text{R}}$:

$$\mathbf{F}_{\text{high}}^{\text{R}}(i, j) = \mathbf{F}_{\text{high}}^{\text{LP}}(i + u, j + v) \quad (3)$$

In parallel, a high-frequency enhancement branch, mirroring the low-frequency branch’s structure, operates on \mathbf{F}_{low} to generate a high-frequency-enhanced map, $\mathbf{F}_{\text{low}}^{\text{HF}}$, adaptively generating high-pass filters analogous to the low-pass filter generation.

Finally, element-wise addition fuses the resampled high-level map, $\mathbf{F}_{\text{high}}^{\text{R}}$, and the high-frequency-enhanced low-level map, $\mathbf{F}_{\text{low}}^{\text{HF}}$:

$$\mathbf{F}_{\text{out}} = \mathbf{F}_{\text{high}}^{\text{R}} + \mathbf{F}_{\text{low}}^{\text{HF}} \quad (4)$$

Through this frequency-modulated and adaptively-refined fusion process, ASR effectively produces feature maps that exhibit both semantic coherence within nuclei and sharp, precise boundaries, which are essential for accurate instance segmentation performance.

3.3. Iterative Contrast-Weaving Module (ICW)

Complementing the per-scale feature enhancement of the ASR module, the ICW is designed to focus on contextual relationships and progressive refinement. ICW transcends simple feature fusion by employing a novel “weaving” process. This process dynamically integrates local contrastive attention and adaptive gating within a feedback loop to iteratively enhance feature discriminability, a critical factor for precise nuclei segmentation, especially in the presence of ambiguous boundaries and class imbalances. Specifically, it mitigates the foreground-background imbalance by employing a contrastive attention mechanism that amplifies nuclei-specific features while suppressing background noise, thus enhancing the nuclei-background distinction. The architecture of the ICW module is detailed in Figure 3(b).

ICW operates through an iterative refinement process. Each iteration t refines the feature representation by taking as input frequency-adapted features, $\mathbf{F}_{\text{ASR}}^{(t-1)} \in \mathbb{R}^{C \times H \times W}$, from the ASR module, and semantic context maps, $\mathbf{S}^f, \mathbf{S}^b, \mathbf{S}^u \in \mathbb{R}^{H \times W}$. These context maps, representing foreground, background, and uncertain regions respectively. The initial input, $\mathbf{F}_{\text{ASR}}^{(0)}$, is the output \mathbf{F}_{out} from the ASR module.

Each ICW iteration performs three core operations, conceptualized as a refinement step towards improved instance separation. First, a dense block, $\text{DenseBlock}(\cdot)$, enhances inter-layer feature propagation, yielding an intermediate feature map $\mathbf{F}_{\text{dense}}^{(t)}$:

$$\mathbf{F}_{\text{dense}}^{(t)} = \text{DenseBlock}(\mathbf{F}_{\text{ASR}}^{(t-1)}) \quad (5)$$

This dense connectivity facilitates feature reuse and mitigates the vanishing gradient problem, promoting robust feature learning.

Next, a contrastive attention (CA) mechanism selectively attends to features based on local context within a $K \times K$ neighborhood $\mathcal{N}(x, y)$ around each spatial location (x, y) . Attention weights, $\alpha_{(x,y)}^{(t)}$, are computed using downsampled foreground ($\mathbf{S}^{\text{f}\downarrow}$) and background ($\mathbf{S}^{\text{b}\downarrow}$) context maps, thus incorporating a broader contextual window to guide attention. The context-aware representation, $\mathbf{V}_{\text{CA}}^{(t)}(x, y)$, is then formulated as:

$$\mathbf{V}_{\text{CA}}^{(t)}(x, y) = \sum_{(i,j) \in \mathcal{N}(x,y)} \alpha_{(x,y)}^{(t)}(x, y, i, j) \cdot \mathbf{V}^{(t)}(i, j) \quad (6)$$

where $\mathbf{V}^{(t)}$ represents value features obtained through a linear transformation:

$$\mathbf{V}^{(t)} = \mathbf{W}_v \mathbf{F}_{\text{dense}}^{(t)} \quad (7)$$

and the attention weight $\alpha_{(x,y)}^{(t)}$ is defined as:

$$\alpha_{(x,y)}^{(t)} = \frac{\exp(\mathbf{w}_f^T \mathbf{S}_{(x,y)}^{\text{f}\downarrow} - \mathbf{w}_b^T \mathbf{S}_{(x,y)}^{\text{b}\downarrow})}{\sum_{(m,n) \in \mathcal{N}(x,y)} \exp(\mathbf{w}_f^T \mathbf{S}_{(m,n)}^{\text{f}\downarrow} - \mathbf{w}_b^T \mathbf{S}_{(m,n)}^{\text{b}\downarrow})} \quad (8)$$

with learnable weight vectors \mathbf{w}_f and \mathbf{w}_b . This local contrastive learning mechanism, distinct from global self-attention, drives the network to learn features that are discriminative for the foreground (nuclei) relative to the background. It effectively amplifies features that contribute to instance separation while suppressing irrelevant ones, resulting in a contrast-enhanced map, $\mathbf{F}_{\text{CA}}^{(t)}$.

Subsequently, an adaptive gated linear unit (AGLU) modulates the influence of $\mathbf{F}_{\text{CA}}^{(t)}$, functioning as a dynamic gate to control information flow:

$$\mathbf{F}_{\text{gated}}^{(t)} = \mathbf{F}_{\text{CA}}^{(t)} \odot \sigma(\kappa \cdot \mathbf{F}_{\text{CA}}^{(t)} + \lambda) \quad (9)$$

where \odot denotes element-wise multiplication, σ is the sigmoid function, and κ and λ are learnable parameters. This adaptivity ensures that feature modulation is neither too aggressive (suppressing useful information) nor too passive (allowing noise to propagate), dynamically balancing contrast enhancement with feature fidelity across diverse image characteristics and learning stages.

Finally, a residual connection integrates the gated features $\mathbf{F}_{\text{gated}}^{(t)}$ with the input features $\mathbf{F}_{\text{ASR}}^{(t-1)}$, generating the refined features $\mathbf{F}_{\text{ASR}}^{(t)}$ for the subsequent iteration:

$$\mathbf{F}_{\text{ASR}}^{(t)} = \mathbf{F}_{\text{ASR}}^{(t-1)} + \mathbf{F}_{\text{gated}}^{(t)} \quad (10)$$

The iteratively refined features from the final ICW iteration are then utilized to derive the ultimate segmentation output.

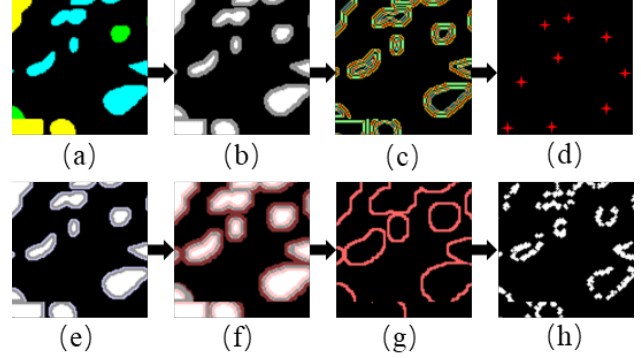


Figure 4. **Illustration of the uncertainty map generation process.** (a) Ground truth mask; (b) Grayscale ground truth mask, delineating nuclei interior (white), boundary (gray), and background (black) regions; (c) Gradient magnitude map; (d) Centroid plot; (e) Mask erosion, marking nuclei interior (blue); (f) Mask dilation, expanding nuclei region (red); (g) Uncertainty behavior map; (h) Final uncertainty map.

Through this iterative "weaving" process—integrating dense propagation, local contrastive attention, and adaptive gating within a closed-loop system with ASR. *WeaveSeg* progressively resolves ambiguities, enhances boundaries, and achieves improved segmentation accuracy. Each iteration facilitates a refinement of the feature representation, iteratively optimizing towards improved instance separation in the feature space. This iterative refinement strategy, leveraging contrastive learning and adaptive gating, holds potential for broader application in other image segmentation tasks demanding robust handling of complex contextual dependencies and feature ambiguities.

3.4. Loss Function

WeaveSeg's training process is guided by a comprehensive multi-task loss function, $\mathcal{L}_{\text{Total}}$, meticulously designed to promote accurate instance segmentation, semantic consistency, and robust handling of ambiguous regions inherent in histopathology images. This loss function is composed of several distinct terms: binary cross-entropy (BCE) losses for the primary segmentation mask ($\mathcal{L}_{\text{mask}}$) and auxiliary foreground, background predictions (\mathcal{L}_{fg} , \mathcal{L}_{bg}), a trilateral decomposition-inspired loss (\mathcal{L}_{tri}), and a novel uncertainty loss ($\mathcal{L}_{\text{uncert}}$). The $\mathcal{L}_{\text{uncert}}$ term is specifically introduced to explicitly address ambiguous regions, particularly prevalent at nuclei boundaries and overlaps.

The uncertainty loss, $\mathcal{L}_{\text{uncert}}$, serves as a key component for robustly handling ambiguous regions. It operates on the predicted uncertainty map, $\mathbf{S}^u \in \mathbb{R}^{H \times W}$, with the objective of encouraging the model to express high uncertainty in areas where the ground truth segmentation is inherently ambiguous (typically boundaries and overlaps) and conversely, low uncertainty in regions with clear foreground (nuclei interiors) or background. To facilitate this, a target uncertainty map, $\mathbf{U}^* \in \mathbb{R}^{H \times W}$, is generated, with its generation

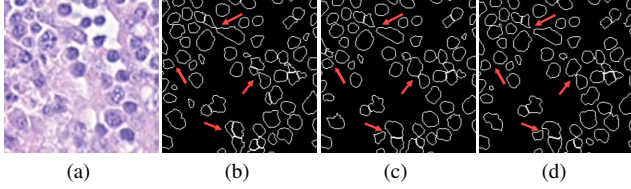


Figure 5. **Qualitative comparison of edge prediction performance.** (a) Input image; (b) Baseline model; (c) Baseline + ICW, showing improved boundary delineation; (d) Baseline + ICW + ASR (WeaveSeg), demonstrating the accurate edge prediction.

process illustrated in Figure 4. The uncertainty loss is then defined as the mean squared error (MSE) between the predicted uncertainty map, \mathbf{S}^u , and the target uncertainty map, \mathbf{U}^* :

$$\mathcal{L}_{\text{uncert}} = \frac{1}{N} \sum_{i=1}^N \left(\mathbf{S}^{u(i)} - \mathbf{U}^{*(i)} \right)^2 \quad (11)$$

where N denotes the total number of pixels, and (i) indexes a specific pixel location. This loss term effectively encourages the model to refrain from making over-confident, yet potentially erroneous, predictions in ambiguous regions. Consequently, it guides the ICW module to prioritize the learning of discriminative features while explicitly acknowledging and accommodating boundary uncertainty. Further details regarding the target uncertainty map generation are provided in the supplementary material.

The final multi-task loss, $\mathcal{L}_{\text{Total}}$, is formulated as a weighted sum of these individual loss terms:

$$\mathcal{L}_{\text{Total}} = \lambda_{\text{mask}} \mathcal{L}_{\text{mask}} + \lambda_{\text{fg}} \mathcal{L}_{\text{fg}} + \lambda_{\text{bg}} \mathcal{L}_{\text{bg}} + \lambda_{\text{tri}} \mathcal{L}_{\text{tri}} + \lambda_{\text{uncert}} \mathcal{L}_{\text{uncert}} \quad (12)$$

The loss weights ($\lambda_{\text{mask}} = 1.0$, $\lambda_{\text{fg}} = 0.5$, $\lambda_{\text{bg}} = 0.5$, $\lambda_{\text{tri}} = 0.3$, and $\lambda_{\text{uncert}} = 1.0$), were empirically determined via validation tuning.

4. Experiments

4.1. Datasets and Evaluation Metrics

We rigorously evaluated WeaveSeg on three publicly available histopathology benchmarks: CoNSEP [8], MoNuSeg [17] and the CPM17 [36] dataset. These datasets feature diverse cell types and challenging conditions, such as significant nuclei overlap and staining variations, making them suitable for comprehensive evaluation. Following standard protocols, we used the official data splits for training and testing. Segmentation performance was quantitatively assessed using four widely-accepted metrics: F1-score (F1), Dice coefficient (Dice), Hausdorff distance (HausD), Panoptic Quality (PQ) score and Aggregated Jaccard Index (AJI) [17, 26].

Models	F1	Dice	HausD	AJI
Baseline	0.725	0.828	27.381	0.515
w/o ICW	0.738	0.835	26.933	0.529
w/o ASR	0.746	0.844	26.354	0.547
Full WeaveSeg	0.759	0.852	25.835	0.563

Table 1. **Ablation study on the CoNSEP dataset, evaluating the impact of different components.** "w/o" denotes removing the corresponding component.

Iter.	F1	Dice	HausD	AJI
0	0.859	0.823	7.141	0.582
1	0.865	0.830	6.754	0.622
2	0.872	0.838	6.312	0.637
3	0.874	0.841	6.042	0.648
4	0.868	0.836	6.453	0.635

Table 2. **Performance comparison across different iterations of the ICW module on the MoNuSeg dataset.** The baseline refers to the initial framework without ICW or ASR (Iter. = 0).

4.2. Implementation Details

WeaveSeg was implemented using PyTorch 1.8 [28] and CUDA 10.1. Training was conducted on a workstation with dual NVIDIA RTX 2080Ti GPUs. Network weights were initialized using ImageNet pre-trained weights [6], leveraging transfer learning methodologies [43]. The Adam optimizer [16] was employed with an initial learning rate of 10^{-4} , which was reduced by a factor of 0.1 every 25 epochs, reaching a final learning rate of 10^{-7} . A batch size of 32 was utilized, and training was performed for 100 epochs. The network processed 256×256 RGB image patches, normalized to the range $[0, 1]$. Standard data augmentation techniques, including random color transformations, horizontal flips, rotations, elastic transformations, and cropping, were applied throughout training. As depicted in Figure 8, the κ and λ parameters within the AGLU module were adaptively adjusted throughout the training process.

4.3. Ablation Studies

To rigorously validate the individual contributions of the ASR and ICW modules, ablation studies were performed on the CoNSEP validation set, utilizing a U-Net variant as the baseline model. Table 1 presents the ablation study results. The complete *WeaveSeg* model consistently outperforms both the baseline and the ablation variants (ablated of ASR or ICW modules) across all evaluation metrics, unequivocally demonstrating the synergistic effect of these two proposed modules.

As detailed in Table 1, the removal of the ICW module leads to a more pronounced performance degradation compared to the removal of the ASR module. Specifically, the Dice coefficient decreases by 0.017 without ICW, whereas it decreases by 0.008 without ASR. A similar trend

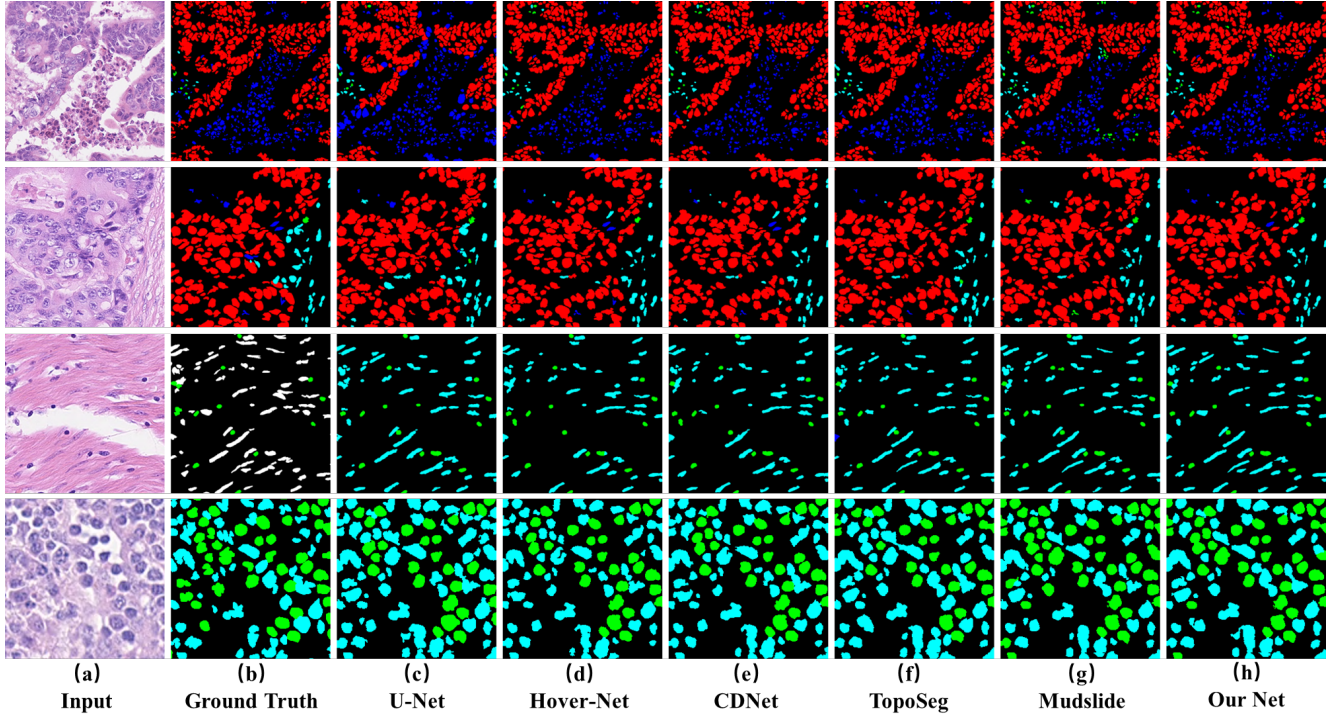


Figure 6. Visualization of semantic segmentation results with different methods. Different colors represent different types.

\mathcal{L}_{mask}	$\mathcal{L}_{fg} \ \& \ \mathcal{L}_{bg}$	\mathcal{L}_{tri}	\mathcal{L}_{uncert}	MoNuSeg				CoNSEP			
				F1	Dice	HausD	AJI	F1	Dice	HausD	AJI
✓				0.861	0.821	6.962	0.588	0.733	0.816	30.527	0.505
✓	✓			0.868	0.828	6.794	0.603	0.748	0.834	28.759	0.536
✓	✓	✓		0.877	0.832	6.751	0.624	0.764	0.839	27.683	0.542
✓	✓	✓	✓	0.874	0.836	6.435	0.644	0.756	0.844	26.354	0.547

Table 3. Ablation study on the impact of different loss function components within the ICW module.

is observed for the instance-level AJI metric. This underscores the critical role of the iterative contrast enhancement provided by the ICW module in achieving accurate segmentation and instance discrimination, particularly in challenging scenarios involving overlapping nuclei and ambiguous boundaries. The ICW module’s capacity for refining features based on contextual relationships and addressing class imbalance appears crucial for attaining optimal performance. Figure 5 visually corroborates the enhanced edge prediction performance resulting from the progressive integration of ICW and ASR.

Furthermore, we investigated the individual impact of different loss components within the ICW module itself (Table 3), isolating their contributions by excluding the ASR module. Progressive incorporation of the foreground/background losses (\mathcal{L}_{fg} & \mathcal{L}_{bg}), the trilateral decomposition loss (\mathcal{L}_{tri}), and the novel uncertainty loss (\mathcal{L}_{uncert}), in addition to the baseline mask loss (\mathcal{L}_{mask})—reveals a consistent performance improvement trend. On MoNuSeg, the incorporation of \mathcal{L}_{fg} and \mathcal{L}_{bg} yields improvements in Dice and AJI. Incorporating \mathcal{L}_{tri} further enhances F1-score. Cru-

cially, adding \mathcal{L}_{uncert} results in the highest Dice, HausD, and AJI values, albeit with a slight decrease in F1-score. This indicates that \mathcal{L}_{uncert} effectively prioritizes precise boundary delineation and instance separation, which is paramount for accurate nuclei segmentation. Consistent trends are observed on CoNSEP, with progressive enhancements in Dice, HausD, and AJI corresponding to the inclusion of each successive loss component. Table 2 performance peaks at the 3rd iteration, demonstrating the effectiveness of the iterative refinement process. This empirically validates our choice of 3 iterations as the optimal configuration for balancing performance and computational efficiency.

4.4. Comparative Experiments

We conducted a comprehensive comparison of WeaveSeg against several recent state-of-the-art methods, including Hover-Net [8], CD-Net [12], Mudslide [38] and CellViT-B [14]. As presented in Table 4, our proposed WeaveSeg demonstrates superior performance across all three benchmark datasets. On MoNuSeg, WeaveSeg achieves an AJI of 0.648 and a PQ of 0.644, outperforming all competitors.

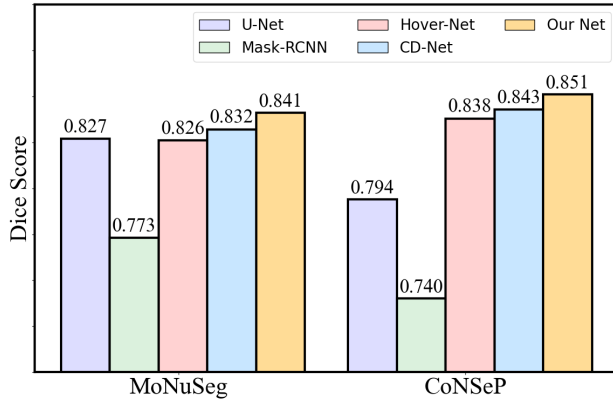


Figure 7. **Dice score comparison on MoNuSeg and CoNSEP datasets.** WeaveSeg (Our Net) achieves superior performance compared to other methods.

Method	MoNuSeg		CoNSEP		CMP17	
	AJI	PQ	AJI	PQ	AJI	PQ
U-Net [31]	0.597	0.581	0.501	0.492	0.655	0.578
Mask-RCNN [11]	0.546	0.509	0.494	0.460	0.684	-
DCAN [3]	0.525	0.492	0.289	-	0.561	0.545
DIST [25]	0.559	0.443	0.502	0.398	0.616	0.504
Hover-Net [8]	0.618	0.597	0.536	0.532	0.705	0.697
CD-Net [12]	0.637	0.619	0.545	0.524	0.729	-
TopoSeg [24]	0.641	0.625	-	-	0.743	0.705
MedSAM [22]	0.622	0.618	0.517	0.505	0.707	0.694
CellViT-B [14]	0.639	0.654	0.547	0.539	0.737	0.713
NucSeg-B [32]	0.634	0.626	0.542	0.533	0.731	0.726
Mudslide [38]	0.642	0.637	0.547	0.542	0.740	0.714
Our Net	0.648	0.644	0.563	0.548	0.751	0.733

Table 4. Performance comparisons on MoNuSeg, CoNSEP and CPM17. AJI and PQ (%) are reported. ‘-’ indicates value not available in original SOTA paper.

Similarly, on CoNSEP, it sets a new SOTA with an AJI of 0.563 and a PQ of 0.548. Notably, our method also excels on the CPM17 dataset, achieving the highest AJI (0.751) and PQ (0.733) scores. These results consistently underscore WeaveSeg’s effectiveness and its robust capability to handle diverse and challenging nuclei segmentation scenarios. The qualitative results in Figure 6 further visualize our model’s proficiency in accurately segmenting challenging instances with overlapping and densely packed nuclei compared to other methods.

Further details on ablation studies and comparative experiments can be found in the Supplementary Material.

4.5. Limitation

Although *WeaveSeg* effectively addresses numerous scenarios featuring ambiguous boundaries and overlapping nuclei, it still encounters limitations in particularly challenging instances involving severely clustered nuclei with nearly

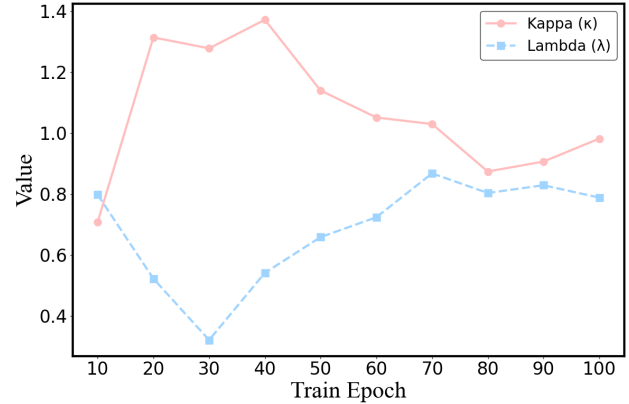


Figure 8. **Visualization of hyperparameter adaptation during training.** The parameters κ and λ in the AGLU module are dynamically adjusted to optimize feature modulation.

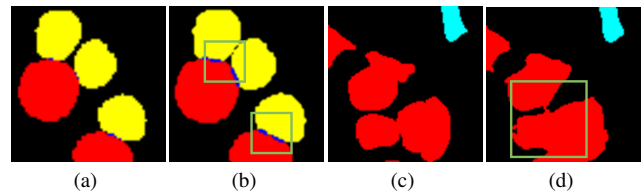


Figure 9. **Illustration of challenging cases where WeaveSeg shows limitations.** (a) Ground truth and (b) failure case on MoNuSeg, where it has limitation; (c) ground truth, and (d) failure case on CoNSEP.

indistinguishable boundaries, often further complicated by staining variations. Figure 9 illustrates representative failure cases. Furthermore, while our method has been validated on three diverse datasets, its generalization capability could be further strengthened by evaluation on a broader spectrum of tissue types and staining protocols. Future research will focus on integrating more sophisticated context modeling mechanisms to handle severe clustering.

5. Conclusion

In summary, *WeaveSeg* introduces a novel framework for nuclei instance segmentation in histopathology images, effectively addressing persistent challenges of ambiguous boundaries, complex morphologies, and nuclei overlap. Achieving state-of-the-art performance, *WeaveSeg* synergistically integrates frequency-aware feature adaptation via the ASR module, iterative contrast enhancement through the ICW module, and a dedicated uncertainty loss function. Empirical validation on MoNuSeg and CoNSEP datasets, substantiated by ablation studies, demonstrates significant performance gains over existing methodologies, underscoring the contribution of iterative refinement. While *WeaveSeg* exhibits robust performance, future efforts will enhance resilience to staining variations in clustered nuclei.

Acknowledgments

This work was supported partly by the National Natural Science Foundation of China (No.62273241), and the Natural Science Foundation of Guangdong Province, China (No.2024A1515011946).

References

- [1] Konstantinos P. Alexandridis, Jianfei Deng, Anh Nguyen, Ser-Nam Lim, and Timothy M. Hospedales. Adaptive parametric activation. In *European Conference on Computer Vision*, pages 455–476, Cham, 2024. Springer Nature Switzerland. 2
- [2] Krishna Chaitanya, Neeraj Karani, Sriram Konjeti, Ardeshir Katouzian, and Debdoot Sheet. Contrastive learning of global and local features for medical image segmentation with limited annotations. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, pages 586–596. Springer, 2020. 3
- [3] Hao Chen, Xiaojuan Qi, Lequan Yu, and Pheng-Ann Heng. Dcan: Deep contour-aware networks for object instance segmentation from histology images. *Medical Image Analysis*, 36:135–146, 2016. 1, 2, 8
- [4] Lei Chen, Yuxi Fu, Li Gu, et al. Frequency-aware feature fusion for dense image prediction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024. 2
- [5] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey E. Hinton. A simple framework for contrastive learning of visual representations. In *International Conference on Machine Learning, ICML 2020*, pages 1597–1607. PMLR, 2020. 2
- [6] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. IEEE, 2009. 6
- [7] Prerna Gehlot, Dinesh Kumar Jha, Nidhi Agrawal, Kshitij Gupta, Sumit Kumar, Poonam Johri, and Poonam Tiwari. Ednfc-net: an encoder decoder-based nested feature cascade network for biomedical image segmentation. *Multimedia Tools and Applications*, pages 1–25, 2024. Online First. 2
- [8] Simon Graham, Quoc Dang Vu, Syed EA Raza, Atif Azam, Yee Wah Tsang, Jin Tae Kwak, and Nasir Rajpoot. Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Medical image analysis*, 58:101563, 2019. 1, 2, 6, 7, 8
- [9] Ming-Hui Guo, Cheng-Ze Lu, Qing Hou, Zhen-Ning Liu, Ming-Ming Cheng, and Shi-Min Hu. Adaptive frequency filters as efficient global token mixers. In *CVPR*, pages 11645–11654, 2023. 2
- [10] Raia Hadsell, Sumit Chopra, and Yann LeCun. Dimensionality reduction by learning an invariant mapping. In *CVPR (2)*, pages 1735–1742. IEEE Computer Society, 2006. 2
- [11] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017. 1, 8
- [12] Xin He, Sheng-Kun Zhou, Yi Zhou, Zheng Chen, Yefeng Wang, Pheng-Ann Heng, Danny Z. Chen, and Jian Wu. Cd-net: Centripetal directional network for nuclear instance segmentation. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 1255–1259. IEEE, 2021. 2, 7, 8
- [13] Xiao He, Sheng Zhao, Hang Zhang, Yutong Chen, Bo Zhou, Pheng-Ann Heng, and Sheng-Kun Zhou. HAN-Net: a hybrid attention nested U-shaped network for pathological image segmentation. *IEEE Transactions on Medical Imaging*, 2024. Early Access. 2
- [14] F. Hörst, M. Rempe, L. Heine, and et al. Cellvit: Vision transformers for precise cell segmentation and classification. *Medical Image Analysis*, 94:103143, 2024. 7, 8
- [15] Lijing Hou, Ayush Agarwal, Dimitris Samaras, Tahsin M Kurc, Rajarsi R Gupta, and Joel H Saltz. Unsupervised histopathology image synthesis. *IEEE transactions on medical imaging*, 39(2):382–394, 2020. 1
- [16] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 6
- [17] Nishant Kumar, Rakesh Verma, Sachin Sharma, Saurabh Bhargava, Ali Vahadane, and Anant Sethi. A dataset and a technique for generalized nuclear segmentation for computational pathology. *IEEE Transactions on Medical Imaging*, 36(7):1550–1560, 2017. 6
- [18] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015. 1, 2
- [19] Meng Lei, Huisi Wu, Xiaohua Lv, et al. Condseg: A general medical image segmentation framework via contrast-driven feature enhancement. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 4571–4579, 2025. 3
- [20] Shangchen Li, Xinyi Shi, Qiang Ye, Xin Yang, Xun Lin, Xiaopeng Xie, Wenming Cao, Mingming Yang, Reynold Cheng, and Wangmeng Zuo. Feature modulation transformer: Cross-refinement of global representation via high-frequency prior for image super-resolution. In *The Twelfth International Conference on Learning Representations, ICLR 2024*, 2024. 2
- [21] Dongnan Liu, Donghao Zhang, Yang Song, Heng Huang, and Weidong Cai. Panoptic feature fusion net: A novel instance segmentation paradigm for biomedical and biological images. *IEEE Transactions on Image Processing*, 30:2045–2059, 2021. 1
- [22] Jican Ma, Yang He, Fan Li, Li Wang, Hao Wang, Kan Li, Weixin Liu, Jie Zhang, Li Xu, Xuan Han, et al. Segment anything in medical images. *Nature Communications*, 15(1): 654, 2024. 8
- [23] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)*, pages 565–571. IEEE, 2016. 1
- [24] Anna Mośńska, Florian Dubost, Jonathan Saltz, Peter J. Wild, Katja A. Frei, Florian Wempe, Maximilian Baust, Nasir Rajpoot, and Dinesh Jha. Toposeg: Topology-aware nuclear instance segmentation. In *Medical Image Computing*

- and *Computer Assisted Intervention – MICCAI 2023*, pages 303–313. Springer, 2023. 2, 8
- [25] Peter Naylor, Mathieu Laé, Frédéric Rey, and Thomas Walter. Segmentation of nuclei in histopathology images by deep regression of the distance map. *IEEE transactions on medical imaging*, 38(2):448–459, 2018. 1, 2, 8
- [26] Davy Neven, Bert De Brabandere, Marc Proesmans, and Luc Van Gool. Instance segmentation by jointly optimizing spatial embeddings and clustering bandwidth. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8837–8845, 2019. 6
- [27] Ozan Oktay, Joachim Schlemper, Loic Le Folgoc, Matthew Lee, Martin Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nathalie Y. Hammerla, Bernhard Kainz, and et al. Attention u-net: Learning where to look for the pancreas. *Medical Image Analysis*, 61:101640, 2020. arXiv preprint arXiv:1804.03999. 2
- [28] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killea, Zeming Lin, Yuqing Yang, Gregory DeVito, Zachary Raison, Alykhan Tejani, Sasank Chilimuntala, Benoit Prabhath, et al. Pytorch: An imperative style, high-performance deep learning library. In *Advances in neural information processing systems*, pages 8026–8037, 2019. 6
- [29] H. Pinckaers, W. Bulten, J. van der Laak, and G. Litjens. Detection of prostate cancer in whole-slide images through end-to-end training with image-level labels. *IEEE Transactions on Medical Imaging*, 40(7):1817–1826, 2021. 1
- [30] Zheng Qin, Pengyi Zhang, Feifei Wu, and Xiao Li. Fcanet: Frequency channel attention networks. In *ICCV*, pages 783–792, 2021. 2
- [31] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 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*, pages 234–241. Springer International Publishing, 2015. 1, 2, 8
- [32] Z. Shui, Y. Zhang, K. Yao, and et al. Unleashing the power of prompt-driven nucleus instance segmentation. In *European Conference on Computer Vision*, pages 288–304. Springer Nature Switzerland, Cham, 2024. 8
- [33] Kriangsak Sirinukunwattana, Syed EA Raza, Weilin Ya, Quanfeng Ba, Chi-Liang Chen, Fan Xing, Josien PW Pluim, Pheng-Ann Heng, Ursula Sanchez, Elodie Bruni, et al. Gland segmentation in colon histology images: the glas challenge contest. *Medical image analysis*, 35:489–502, 2017. 1
- [34] Eric Upschulte, Stefan Harmeling, Katrin Amunts, and Timo Dickscheid. Uncertainty-aware contour proposal networks for cell segmentation in multi-modality high-resolution microscopy images. In *Competitions in Neural Information Processing Systems*, pages 1–12. PMLR, 2023. 1
- [35] Mitko Veta, Yi-Ju Heng, Nikolaos Stathonikos, Babak Ehteshami Bejnordi, Florentin Beca, Thomas Wollmann, Kay Rohr, Muhammad Arsalan Shah, Dayong Wang, Michel Rousson, et al. Predicting breast tumor proliferation from whole-slide images: the tupac16 challenge. *Medical image analysis*, 54:111–121, 2019. 1
- [36] Q. D. Vu, S. Graham, T. Kurc, and et al. Methods for segmentation and classification of digital microscopy tissue images. *Frontiers in bioengineering and biotechnology*, 7:53, 2019. 6
- [37] Hong Wang, Huisi Wu, and Jing Qin. Incremental nuclei segmentation from histopathological images via future-class awareness and compatibility-inspired distillation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11408–11417, 2024. 1
- [38] J. Wang. Mudslide: A universal nuclear instance segmentation method. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11673–11682, 2024. 2, 7, 8
- [39] Zikun Wang, Liang Zheng, Yichen Li, Shenglin Lu, Bowen Zhang, Zongyuan Zhou, Ruimao Zhang, Xiaopeng Li, Jian Yang, Xiaoguang Li, and et al. Self-supervised nuclei segmentation via local and global contrastive learning. *arXiv preprint arXiv:2403.08474*, 2024. 3
- [40] Huisi Wu, Zhaoze Wang, Youyi Song, Lin Yang, and Jing Qin. Cross-patch dense contrastive learning for semisupervised segmentation of cellular nuclei in histopathologic images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11666–11675, 2022. 1
- [41] Ziyuan Xiang, Wen Luo, Zhiqiang Li, and Jun Su. Bionet: Bi-directional o-shape network for nuclei instance segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 612–621. Springer, 2024. 2
- [42] Fan Xing and Lin Yang. Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: A comprehensive review. *IEEE reviews in biomedical engineering*, 9:234–263, 2016. 1
- [43] Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. How transferable are features in deep neural networks? In *Advances in neural information processing systems*, pages 3320–3328, 2014. 6
- [44] Bo Zhou, Hao Chen, Jian Yang, Wei Wang, Xiaojuan Qi, and Pheng-Ann Heng. Cia-net: Robust nuclei instance segmentation with contour-aware information aggregation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 326–336. Springer, 2020. 1, 2
- [45] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. 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, Granada, Spain, September 20, 2018, proceedings 4*, pages 3–11. Springer International Publishing, 2018. 1, 2