Single-Stage Hierarchical Rectification for Weakly Supervised Histopathology Segmentation

Trong Duc Nguyen[1,2,3], Hoang Long Nguyen[1,2,3], Thanh Ha Do[4], and Huy Hieu Pham[1,2,3⋆]

[1] College of Engineering & Computer Science, VinUniversity, Hanoi, Vietnam
[2] VinUni-Illinois Smart Health Center, VinUniversity, Hanoi, Vietnam
[3] The Computer Vision and Medical AI Lab, VinUniversity, Hanoi, Vietnam
[4] Posts and Telecommunications Institute of Technology, Hanoi, Vietnam

Abstract. Existing weakly supervised semantic segmentation (WSSS) methods in computational pathology commonly use a multi-stage workflow involving class activation map (CAM) generation, offline pseudo-mask refinement, and fully supervised retraining. While effective, this decoupled workflow can incur high computational training costs and is vulnerable to error propagation, where local texture biases in shallow CNN layers may generate false-positive artifacts that subsequent refinement steps fail to correct. To mitigate these issues, we propose the Single-Stage Hierarchical Rectification (SSHR) framework. Rather than relying on post-hoc CAM refinement, SSHR rectifies intermediate feature representations during the forward pass. We introduce a Hierarchical Feature Rectification Module (HFRM) that uses deep global semantic context to guide CAM-producing shallow and intermediate features, suppressing local anomalies before activation maps are formed. This mechanism produces refined activation maps directly within a single training loop. Experiments on the LUAD-HistoSeg and BCSS datasets demonstrate that SSHR outperforms state-of-the-art multi-stage methods. Furthermore, SSHR reduces training duration by roughly 2 to 5 times. This efficiency lowers computational overhead and supports faster model development for large-scale histopathology workflows. The code is available at: https://github.com/trongduc-nguyen/SSHR.

Keywords: Weakly supervised learning · Semantic segmentation · Computational pathology · Feature rectification · Single-stage learning.

1 Introduction

Deep learning has revolutionized computational pathology, offering automated solutions for tumor diagnosis, prognosis, and treatment planning [11]. However, the success of these fully supervised models heavily relies on massive datasets with precise pixel-level annotations [13]. In digital pathology, acquiring such dense annotations is exceptionally capital intensive and time consuming, as

⋆ Corresponding author: Huy Hieu Pham (hieu.ph@vinuni.edu.vn).

it requires highly specialized domain knowledge from experienced pathologists [10,14]. Consequently, the scarcity of pixel-wise ground truth has become a major bottleneck, hindering the scalability of deep learning applications in clinical workflows [2]. To address this annotation burden, Weakly Supervised Semantic Segmentation (WSSS) has emerged as a promising paradigm, aiming to achieve segmentation performance comparable to fully supervised methods using only image-level labels, which are readily available in clinical reports [9,1].

Mainstream approaches in histopathological WSSS typically use a multi-stage workflow based on Class Activation Maps (CAMs) [19]. Early approaches focused on mitigating the noisy and sparse nature of CAMs caused by the morphological heterogeneity of tissue. For instance, HAMIL [18] used a two-stage framework with high-resolution maps to refine boundaries. More recently, techniques such as PBIP [12], WAWEHIS [6], and ESFAN [17] have improved CAM quality through edge-semantic synergy and prototype learning. Despite these advances, many recent techniques still follow a decoupled multi-stage workflow: (1) training a classification model, (2) applying post-processing algorithms to refine CAMs into pseudo-masks offline, and (3) retraining a separate segmentation network. This separation can propagate errors from the initial CAMs to the final segmentation model [4]. Shallow and intermediate CNN features preserve local tissue details but have limited receptive fields, making them susceptible to texture-driven false positives and isolated spatial outliers. Once these artifacts are encoded into pseudo-masks, the subsequent retraining stage may preserve or amplify them, making the final performance sensitive to the retraining architecture. Furthermore, repeated training and refinement stages increase computational cost and slow model development for large-scale pathology datasets [8].

To address these challenges, we propose the Single-Stage Hierarchical Rectification (SSHR) framework. Unlike conventional CAM-based methods that correct errors after CAM generation, SSHR intervenes directly at the feature level before activation maps are formed. We introduce an internal rectification mechanism that uses deep classification-supervised semantic context to guide CAM-producing shallow and intermediate features while enforcing local spatial consistency. By mitigating texture-driven artifacts during the forward pass, SSHR generates refined activation maps in a unified training loop. The main contributions of our work are as follows:

- To the best of our knowledge, SSHR is the first single-stage feature-rectification framework for CAM-based histopathological WSSS that outperforms multi-stage retraining pipelines. Our findings highlight the potential of internal rectification as an alternative to decoupled pseudo-mask generation and retraining.
- We propose the Hierarchical Feature Rectification Module (HFRM), a dual-branch mechanism combining Global Semantic Rectification (GSR) with Contextual Homogenization (CH). GSR uses deep semantic context to reweight CAM-producing channels, while CH suppresses isolated spatial outliers through lightweight local context aggregation.

- SSHR outperforms existing state-of-the-art multi-stage weakly supervised segmentation methods on both the LUAD-HistoSeg and BCSS datasets. By eliminating offline mask generation and repeated retraining rounds, SSHR reduces total training duration by roughly 2 to 5 times compared to conventional pipelines, supporting faster model development for large-scale workflows.

2 Methodology

2.1 Overall Framework

Conventional WSSS pipelines separate CAM generation and pseudo-mask refinement, relying on post-processing and retraining stages that may propagate false positives arising from local texture ambiguities in shallow and intermediate CNN features. To address this limitation, we propose the SSHR framework, which refines intermediate representations during the forward pass instead of correcting errors after CAM generation.

As shown in Fig. 1, the core component of SSHR is the HFRM, which integrates semantic guidance with spatial homogenization. By rectifying CAM-producing features before activation maps are formed, SSHR suppresses local outliers and produces refined CAMs without offline post-processing. The HFRM consists of two complementary branches described below.

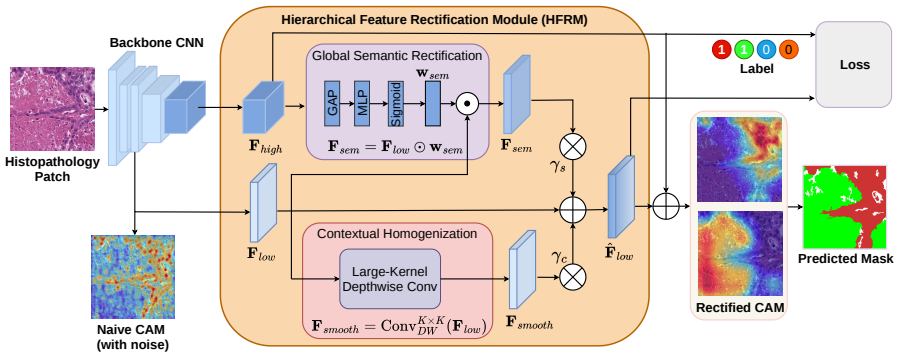

Fig. 1. Overview of the proposed Single-Stage Hierarchical Rectification framework.

2.2 Global Semantic Rectification

The Global Semantic Rectification (GSR) branch is designed to balance local tissue detail with image-level semantic context. In histopathology, tissue structures exhibit substantial morphological heterogeneity. Shallow and intermediate features preserve fine boundaries and local textures, but their limited receptive

fields make them prone to texture-driven false positives. In contrast, deep features provide stronger image-level semantic context, but are spatially coarse. GSR bridges this gap by using deep semantic context to guide CAM-producing intermediate representations.

Let $\mathbf{F}_{low} \in \mathbb{R}^{C\times H\times W}$ denote a CAM-producing intermediate feature map, and $\mathbf{F}_{high} \in \mathbb{R}^{D\times h\times w}$ denote the feature map from the deepest layer. In our ResNet38 implementation, HFRM is applied to three CAM-producing intermediate feature maps: one 56×56 feature map with 256 channels and two successive 28×28 feature maps with 512 and 1024 channels, respectively. Since $\mathbf{F}_{high}$ is directly supervised by image-level classification, it provides a compact semantic context about the tissue classes likely to be present in the image. We use this high-level context to reweight channels in $\mathbf{F}_{low}$ before CAM generation.

We first apply Global Average Pooling (GAP) to $\mathbf{F}_{high}$ to obtain a global context vector $\mathbf{v}_g \in \mathbb{R}^D$. This vector is mapped through a lightweight multilayer perceptron (MLP) and a Sigmoid activation to generate a channel-wise attention vector $\mathbf{w}_{sem} \in [0,1]^C$:

$$\mathbf{w}_{sem} = \sigma\left(\mathbf{W}_2\delta(\mathbf{W}_1\mathbf{v}_g)\right) \tag{1}$$

where $\mathbf{W}_1 \in \mathbb{R}^{\frac{D}{8}\times D}$ and $\mathbf{W}_2 \in \mathbb{R}^{C\times\frac{D}{8}}$ are learnable projection matrices, and δ denotes the ReLU activation. The semantically rectified feature is computed via a Hadamard product along the channel dimension, with $\mathbf{w}_{sem}$ broadcast over the spatial dimensions:

$$\mathbf{F}_{sem} = \mathbf{F}_{low} \odot \mathbf{w}_{sem}. \tag{2}$$

This mechanism down-weights feature channels that are likely to trigger texture-driven false positives inconsistent with the image-level semantic context.

2.3 Contextual Homogenization

While the semantic branch suppresses globally inconsistent activations, spatial outliers from semantically valid classes can still persist locally. To encourage spatial contiguity, the second branch performs contextual homogenization using a lightweight large-kernel depthwise convolution, providing an efficient local-context aggregation mechanism.

Given a spatial kernel of size $K\times K$, the spatially smoothed feature $\mathbf{F}_{smooth} \in \mathbb{R}^{C\times H\times W}$ is obtained by:

$$\mathbf{F}_{smooth} = \mathrm{Conv}_{DW}^{K\times K}(\mathbf{F}_{low}). \tag{3}$$

By aggregating features from a broader neighborhood, this branch reduces isolated spatial responses while preserving the computational efficiency of depthwise convolution.

To integrate these rectifications into the backbone without causing abrupt optimization changes, we formulate the final rectified feature $\hat{\mathbf{F}}_{low}$ using a zero-initialized residual connection:

$$\hat{\mathbf{F}}_{low} = \mathbf{F}_{low} + \gamma_s\mathbf{F}_{sem} + \gamma_c\mathbf{F}_{smooth}, \tag{4}$$

where γ_s and γ_c are zero-initialized learnable scalar parameters. This residual design makes the module initially behave like the original backbone and then gradually learn how much semantic and contextual rectification to incorporate.

We apply 1×1 convolutions to the rectified intermediate features and the final high-level feature to generate scale-specific CAMs, $\mathbf{M}_s \in \mathbb{R}^{N_c \times H_s \times W_s}$, for S different scales, where N_c denotes the number of tissue classes. Global Average Pooling extracts logits $\mathbf{p}_s$ to optimize the framework via an averaged multi-label soft margin loss:

$$\mathcal{L}_{cls} = \frac{1}{S} \sum_{s=1}^{S} \left(-\frac{1}{N_c} \sum_{k=1}^{N_c} \Big[y_k \log(\sigma(p_{s,k})) + (1 - y_k) \log(1 - \sigma(p_{s,k})) \Big] \right) \tag{5}$$

where $y_k \in \{0, 1\}$ is the image-level label for class k. During inference, the final high-resolution activation map $\mathbf{M}_{final}$ is obtained via a weighted fusion:

$$\mathbf{M}_{final} = \sum_{s=1}^{S} \alpha_s \cdot \mathrm{Norm}(\mathrm{Up}(\mathbf{M}_s)), \tag{6}$$

where $\sum_{s=1}^{S} \alpha_s = 1$. The final segmentation mask is derived by taking the pixel-wise argmax of $\mathbf{M}_{final}$.

3 Experiments

3.1 Datasets and Experimental Settings

LUAD-HistoSeg: This dataset contains whole slide images from 54 lung adenocarcinoma patients. It includes four tissue types: tumor epithelium (TE), tumor-associated stroma (TAS), necrosis (NEC), and lymphocyte (LYM). The dataset has 16,678 patches for training, 300 for validation, and 307 for testing [7].
BCSS: Derived from breast cancer whole slide images of 151 patients, this dataset also features four categories: tumor (TUM), stroma (STR), lymphocyte infiltration (LYM), and necrosis (NEC). It provides 23,422 patches for training, 3,418 for validation, and 4,986 for testing [3].

3.2 Experimental Details

All experiments were conducted on a NVIDIA RTX 3090 GPU. We use a ResNet38 backbone [16] for training. SGD optimizer with an initial learning rate of 1×10^{-2} and a polynomial decay policy. We evaluate performance using two standard metrics: mean Intersection-over-Union (mIoU) and mean Dice coefficient (mDice).

3.3 Comparison with state-of-the-art

Table 1 compares SSHR with advanced WSSS methods. SSHR achieves the best performance on both datasets within a single training round. On LUAD-HistoSeg, it reaches 77.93% mIoU, surpassing the final results of ESFAN (76.83%)

Table 1. Comparison with state-of-the-art methods on LUAD-HistoSeg and BCSS datasets. Phase 1 refers to initial CAMs/masks, while Phase 2 refers to the final segmentation after retraining. The results are reported as mean ± std. * indicates statistical significance compared with other methods ($p < 0.05$, two-tailed t-test). Best results are highlighted in bold.

Method	Venue	Backbone		mIoU (%)		mDice (%)	
		Phase 1	Phase 2	Phase 1	Phase 2	Phase 1	Phase 2
LUAD-HistoSeg							
Multi-Stage Methods							
MLPS [7]	MIA'22	ResNet38	ResNet101	74.18 ± 0.89	75.27 ± 0.13	85.42 ± 0.73	85.82 ± 0.08
ARML [5]	MICCAI'24	ResNet38	ResNet101	73.44 ± 0.27	76.55 ± 0.07	84.65 ± 0.18	86.67 ± 0.04
ESFAN [17]	MICCAI'25	ResNet38	ResNet200	75.76 ± 0.33	76.83 ± 0.38	86.17 ± 0.22	86.86 ± 0.24
PBIP [12]	CVPR'25	MiT-B1	ResNet200	74.46 ± 0.51	75.89 ± 0.26	85.62 ± 0.22	86.02 ± 0.13
WAWEHIS [6]	TMI'25	ResNet38	ResNet200	72.62 ± 0.31	74.69 ± 0.52	84.10 ± 0.21	85.46 ± 0.32
Single-Stage Methods							
DuPL [15]	CVPR'24	ViT-B	–	63.71 ± 0.11	–	73.06 ± 0.21	–
SSHR (Proposed)	–	ResNet38	–	$\mathbf{77.93 \pm 0.24}^*$	–	$\mathbf{87.54 \pm 0.19}^*$	–
BCSS							
Multi-Stage Methods							
MLPS [7]	MIA'22	ResNet38	ResNet101	64.11 ± 0.34	69.07 ± 0.35	77.88 ± 0.26	81.47 ± 0.28
ARML [5]	MICCAI'24	ResNet38	ResNet101	65.54 ± 0.33	69.58 ± 0.51	78.97 ± 0.25	81.83 ± 0.37
ESFAN [17]	MICCAI'25	ResNet38	ResNet200	70.34 ± 0.40	70.59 ± 0.41	82.41 ± 0.23	82.59 ± 0.29
PBIP [12]	CVPR'25	MiT-B1	ResNet200	66.74 ± 0.34	69.16 ± 0.41	79.06 ± 0.31	81.11 ± 0.23
WAWEHIS [6]	TMI'25	ResNet38	ResNet200	64.63 ± 0.54	70.08 ± 0.74	78.27 ± 0.42	82.19 ± 0.54
Single-Stage Methods							
DuPL [15]	CVPR'24	ViT-B	–	60.02 ± 0.51	–	72.11 ± 0.49	–
SSHR (Proposed)	–	ResNet38	–	$\mathbf{71.82 \pm 0.38}^*$	–	$\mathbf{83.52 \pm 0.27}^*$	–

and PBIP (75.89%), despite relying solely on a single-stage ResNet38 backbone, whereas these methods require heavier ResNet200 encoders during retraining. A similar advantage is observed on BCSS, where SSHR attains 71.82% mIoU and consistently outperforms all multi-stage pipelines.

Evaluation of DuPL [15], a representative natural vision method, underscores pathology's severe domain shift. Despite its robust ViT-B architecture, DuPL suffers massive degradation as it lacks specialized mechanisms for the morphological heterogeneity and ambiguous transitions inherent in tissue. Unlike contour reliant natural image models, SSHR resolves structural biases via internal hierarchical rectification, effectively adapting the efficient single stage paradigm to pathological constraints.

As shown in Fig. 2, SSHR achieves superior boundary adherence over multi-stage methods. Internal hierarchical rectification effectively overcomes CAM coarseness, yielding masks precisely aligned with intricate tissue morphologies.

3.4 Ablation Study

We conduct an ablation study to evaluate the individual contributions of the proposed components (Table 2). Combining Global Semantic Rectification (**GSR**) to suppress false positives and Contextual Homogenization (**CH**) to eliminate spatial outliers yields a strong synergistic performance boost over the baseline. Analyzing the spatial kernel size K shows that $K = 15$ optimally balances outlier

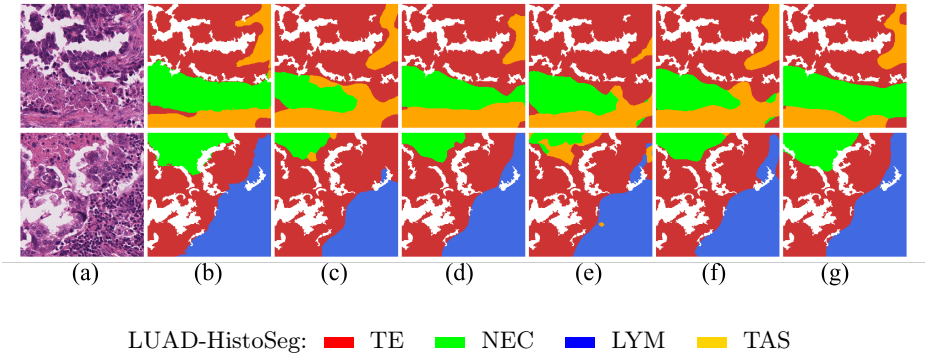

Fig. 2. Qualitative comparison of segmentation results. (a) Test image. (b) Ground truth. (c) ARML. (d) ESFAN. (e) MLPS. (f) WAWEHIS. (g) SSHR (Proposed method)

Table 2. Ablation study evaluating the individual contributions of the proposed components. **BL**: Baseline. The results are reported as mean ± std.

BL	GSR	CH	K	mIoU (%)		mDice (%)	
				LUAD-HistoSeg	BCSS	LUAD-HistoSeg	BCSS
✓			–	72.12 ± 0.31	63.54 ± 0.41	82.83 ± 0.28	72.02 ± 0.37
✓	✓		–	75.84 ± 0.28	68.31 ± 0.35	84.41 ± 0.26	78.65 ± 0.32
✓		✓	15	75.21 ± 0.25	66.95 ± 0.39	84.88 ± 0.23	77.42 ± 0.36
✓	✓	✓	7	76.55 ± 0.20	70.42 ± 0.37	86.26 ± 0.18	79.08 ± 0.34
✓	✓	✓	15	**77.93 ± 0.24**	**71.82 ± 0.38**	**87.54 ± 0.19**	**83.52 ± 0.27**
✓	✓	✓	21	75.41 ± 0.29	69.15 ± 0.49	84.02 ± 0.27	78.88 ± 0.44

suppression and boundary preservation. Smaller kernels ($K = 7$) fail to eliminate large outliers due to limited receptive fields, while larger kernels ($K = 21$) over-smooth valid tissue boundaries, causing performance drops.

3.5 Backbone Sensitivity and Instability in Stage 2

Evaluating diverse Stage 2 backbones (Table 3) reveals inherent instability in sequential pipelines. Despite increased capacity, performance fluctuates unpredictably in multi stage frameworks, with methods like MLPS and PBIP often suffering mIoU degradation on heavier encoders or falling below Phase 1. These observations suggest that gains from retraining on noisy pseudo labels may depend on backbone characteristics and do not always yield systematic improvements, while Phase 1 errors can propagate during retraining. Conversely, SSHR eliminates this dependency by resolving local texture bias internally during the forward pass. By bypassing decoupled retraining, SSHR derives its accuracy

Table 3. Backbone scaling instability on LUAD-HistoSeg (mIoU %). Red color and ↓ denote cases where the performance in Phase 2 drops below Phase 1. The results are reported as mean ± std.

Method	Phase 1	Phase 2 Backbone (mIoU %)			
		ResNet101	MiT-B4	ResNet152	EfficientNet-B8
Params (M)		42M	60M	64M	84M
MLPS [7]	74.18 ± 0.89	75.27 ± 0.18 ↑	72.61 ± 0.82 ↓	73.56 ± 0.37 ↓	70.55 ± 0.70 ↓
ARML [5]	73.43 ± 0.27	76.58 ± 0.06 ↑	73.53 ± 0.79 ↑	75.38 ± 0.25 ↑	74.71 ± 1.96 ↑
ESFAN [17]	75.75 ± 0.33	75.76 ± 0.88 ↑	75.10 ± 0.62 ↓	75.61 ± 0.78 ↓	72.52 ± 0.28 ↓
WAWEHIS [6]	72.62 ± 0.31	75.41 ± 0.11 ↑	68.39 ± 1.26 ↓	66.91 ± 0.34 ↓	73.04 ± 0.70 ↑
PBIP [12]	74.46 ± 0.51	73.02 ± 0.17↓	72.94 ± 0.58↓	74.91 ± 0.21 ↑	73.18 ± 0.77 ↓

from robust, rectified features, ensuring more stable performance compared to conventional multi-stage retraining pipelines.

3.6 Computational Efficiency

Table 4. Efficiency comparison on LUAD-HistoSeg. Ratio (×) denotes total training time relative to Ours. Latency (ms) is the average inference time per patch.

Method	ESFAN [17]	ARML [5]	WAWEHIS [6]	MLPS [7]	PBIP [12]	**Ours**
Ratio (×)	2.23×	4.12×	3.85×	2.73×	5.42×	**1.00**×
Latency (ms)	20.25	38.55	24.45	40.08	22.75	**9.10**

Table 4 highlights the efficiency of SSHR, which reduces training time by 2.23× to 5.42× compared to multi-stage methods. Since WSIs often reach gigapixel resolution, repeated training and offline refinement stages can become a practical bottleneck when scaling to large histopathology datasets. By reducing these repeated stages, SSHR supports faster model development for histopathology segmentation. Furthermore, SSHR achieves a superior inference latency of 9.10 ms per patch, outperforming all competitors and ensuring efficient inference.

4 Conclusion

SSHR highlights the potential of forward-pass feature rectification as a streamlined alternative to decoupled multi-stage pipelines in CAM-based pathology WSSS. By internalizing semantic and contextual rectification before CAM generation, SSHR mitigates error propagation from noisy pseudo-masks and produces refined activation maps within a single training loop. Experiments on LUAD-HistoSeg and BCSS show that SSHR achieves strong segmentation performance while reducing training duration, making it an efficient framework for histopathology segmentation model development.

Limitations and Future Work: SSHR is primarily evaluated within the CAM-based histopathological WSSS setting, and comparisons with broader non-CAM weakly supervised methods remain an important direction for future work. In addition, the current design uses lightweight semantic and contextual rectification modules, whose effectiveness may depend on the semantic quality of the backbone. SSHR also currently lacks modeling of long-range slide-level interactions. Future efforts will focus on integrating global slide context and exploring spatially adaptive weighting strategies to further enhance robustness across staining variations and tissue morphologies in real-world clinical environments.

Acknowledgments. This research was funded by the National Foundation for Science and Technology Development (NAFOSTED) through Project No. IZVSZ2_229539 (2025–2027).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Abdelsamea, M.M., Zidan, U., Senousy, Z., Gaber, M.M., Rakha, E., Ilyas, M.: A survey on artificial intelligence in histopathology image analysis. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery **12**(6), e1474 (2022)
2. Amgad, M., Atteya, L.A., Hussein, H., Mohammed, K.H., Hafiz, E., Elsebaie, M.A., Alhusseiny, A.M., AlMoslemany, M.A., Elmatboly, A.M., Pappalardo, P.A., et al.: Nucls: A scalable crowdsourcing approach and dataset for nucleus classification and segmentation in breast cancer. GigaScience **11**, giac037 (2022)
3. Amgad, M., Elfandy, H., Hussein, H., Atteya, L.A., Elsebaie, M.A., Abo Elnasr, L.S., Sakr, R.A., Salem, H.S., Ismail, A.F., Saad, A.M., et al.: Structured crowdsourcing enables convolutional segmentation of histology images. Bioinformatics **35**(18), 3461–3467 (2019)
4. Chong, Y., Chen, X., Tao, Y., Pan, S.: Erase then grow: Generating correct class activation maps for weakly-supervised semantic segmentation. Neurocomputing **453**, 97–108 (2021)
5. Feng, S., Chen, J., Liu, Z., Liu, W., Wang, Z., Lan, R., Pan, X.: Mining gold from the sand: Weakly supervised histological tissue segmentation with activation relocalization and mutual learning. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 414–423. Springer (2024)
6. Feng, S., Zhang, H., Zhao, X., Shi, L., Liu, Z., Lan, R., Shi, L., Pan, X.: Wave-aware weakly supervised histopathological tissue segmentation with cross-scale logits distillation. IEEE Transactions on Medical Imaging (2025)
7. Han, C., Lin, J., Mai, J., Wang, Y., Zhang, Q., Zhao, B., Chen, X., Pan, X., Shi, Z., Xu, Z., et al.: Multi-layer pseudo-supervision for histopathology tissue semantic segmentation using patch-level classification labels. Medical Image Analysis **80**, 102487 (2022)
8. Hanna, M.G., Pantanowitz, L., Dash, R., Harrison, J.H., Deebajah, M., Pantanowitz, J., Rashidi, H.H.: Future of artificial intelligence (ai)-machine learning (ml) trends in pathology and medicine. Modern Pathology p. 100705 (2025)

9. Montezuma, D., Oliveira, S.P., Neto, P.C., Oliveira, D., Monteiro, A., Cardoso, J.S., Macedo-Pinto, I.: Annotating for artificial intelligence applications in digital pathology: A practical guide for pathologists and researchers. Modern Pathology **36**(4), 100086 (2023)
10. Qu, L., Liu, S., Liu, X., Wang, M., Song, Z.: Towards label-efficient automatic diagnosis and analysis: a comprehensive survey of advanced deep learning-based weakly-supervised, semi-supervised and self-supervised techniques in histopathological image analysis. Physics in Medicine & Biology **67**(20), 20TR01 (2022)
11. Srinidhi, C.L., Ciga, O., Martel, A.L.: Deep neural network models for computational histopathology: A survey. Medical image analysis **67**, 101813 (2021)
12. Tang, Q., Fan, L., Pagnucco, M., Song, Y.: Prototype-based image prompting for weakly supervised histopathological image segmentation. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 30271–30280 (2025)
13. Van Eycke, Y.R., Foucart, A., Decaestecker, C.: Strategies to reduce the expert supervision required for deep learning-based segmentation of histopathological images. Frontiers in medicine **6**, 222 (2019)
14. Wang, X., Chen, H., Gan, C., Lin, H., Dou, Q., Tsougenis, E., Huang, Q., Cai, M., Heng, P.A.: Weakly supervised deep learning for whole slide lung cancer image analysis. IEEE transactions on cybernetics **50**(9), 3950–3962 (2019)
15. Wu, Y., Ye, X., Yang, K., Li, J., Li, X.: Dupl: Dual student with trustworthy progressive learning for robust weakly supervised semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3534–3543 (2024)
16. Wu, Z., Shen, C., Van Den Hengel, A.: Wider or deeper: Revisiting the resnet model for visual recognition. Pattern recognition **90**, 119–133 (2019)
17. Zhang, H., Feng, S., Huan, Z., Wang, H., Liu, Z., Lan, R., Pan, X.: Edge-semantic synergy fusion and adaptive noise-aware for weakly supervised pathological tissue segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 160–169. Springer (2025)
18. Zhong, L., Wang, G., Liao, X., Zhang, S.: Hamil: High-resolution activation maps and interleaved learning for weakly supervised segmentation of histopathological images. IEEE Transactions on Medical Imaging **42**(10), 2912–2923 (2023)
19. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A.: Learning deep features for discriminative localization. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2921–2929 (2016)