

C²RM-Seg: Causal Counterfactual Reasoning with Structural-Semantic Priors for Weakly Supervised Histopathological Tissue Segmentation

Hualong Zhang¹, Siyang Feng¹, Zihan Huan¹, Yi Qian¹, Zhenbing Liu¹, Rushi Lan^{1,2(✉)}, and Xipeng Pan^{1,3}

¹ Guangxi Key Laboratory of Image and Graphic Intelligent Processing, Guilin University of Electronic Technology, Guilin 541004, China

² International Joint Research Laboratory of Spatio-temporal Information and Intelligent Location Services, Guilin University of Electronic Technology, Guilin 541004, China

³ School of Computer Science and Information Security, Guilin University of Electronic Technology, Guilin 541004, China
rslan@guet.edu.cn

Abstract. Histopathological tissue segmentation is essential for computer-aided diagnosis, yet weakly supervised methods often suffer from noisy pseudo-labels generated by Class Activation Mapping (CAM). Existing CAM approaches tend to focus on staining-driven appearance cues rather than true causal tissue morphology, resulting in spurious localization and poor structural consistency. To address this issue, we propose C²RM-Seg, a two-stage framework that integrates causal pseudo-label refinement with structure-aware semantic enhancement. For classification, we introduce a Causal Counterfactual Reasoning Module (C²RM) that decomposes features into latent factors and performs counterfactual intervention via a learned causal structure matrix, suppressing confounding context and producing morphology-aligned CAMs. For segmentation, we design a Dual-Path Structural-Semantic Architecture that combines fine-grained structural features from ResNeSt with global semantic priors from frozen DINOv3 foundation model. A cross-path gating mechanism adaptively regulates semantic injection using local structural cues to preserve boundary fidelity. To further mitigate residual pseudo-label noise, we propose an Uncertainty-Gated Margin (UGM) loss, which dynamically balances margin enforcement and confidence learning based on prediction uncertainty. Extensive experiments on two public histopathological tissue datasets show that C²RM-Seg achieves state-of-the-art performance. Code is available at <https://github.com/OceanPetal/C2AM-Seg>.

Keywords: Weakly Supervised Learning · Tissue Segmentation · Counterfactual Reasoning · Foundation Models

1 Introduction

Histopathological tissue segmentation plays an important role in computer-aided diagnosis, enabling quantitative analysis for cancer staging and prognosis assessment [1,2,3,4]. Accurate delineation of pathological regions provides structured representations that facilitate downstream clinical interpretation [5,6,7]. However, fully supervised segmentation requires dense pixel-level annotations, which are costly and highly dependent on expert knowledge, especially for large and heterogeneous whole-slide images [8,9]. Weakly Supervised Semantic Segmentation (WSSS) alleviates this burden by learning dense predictions from only image-level labels, offering a scalable alternative for histopathology analysis [10,11,12].

Most WSSS methods adopt Class Activation Mapping (CAM) to generate coarse seeds and then train a segmentation model with the resulting pseudo-masks [13,14,15,16]. Recent efforts refine CAMs through activation enhancement and uncertainty modeling. Representative approaches include Wave-aware CAM [17], UAM [13], and UM-CAM [14], as well as histopathology-oriented variants such as ARML [18] and HAMIL [19]. Despite improved localization, existing CAM-based refinements often conflate causal morphology with confounded cues (e.g., stain variations and background tissue patterns). Without explicitly modeling latent confounders or performing causal intervention, CAM responses can be systematically skewed toward context, leading to biased pseudo-masks. Beyond CAM refinement, structural modeling strategies such as MLPS [20], ES-FAN [11], and expert-guided multi-granularity fusion [21] aim to better preserve fine-grained morphology and boundary consistency. Noise-aware optimization techniques like OEEM [22] further attempt to mitigate pseudo-label noise. However, when pseudo-labels are systematically biased by uncontrolled confounding factors, architectural refinement or loss reweighting alone may be insufficient to resolve the underlying causal misalignment.

To address these challenges, we propose **C²RM-Seg**, a framework that combines causal deconfounding with foundation model priors for robust weakly supervised histopathology segmentation. We introduce a Causal Counterfactual Reasoning Module (C²RM) that formulates CAM generation under a Structural Causal Model (SCM) and performs counterfactual intervention to subtract confounding effects, producing cleaner seeds. We further design a Dual-Path Structural-Semantic Architecture that integrates a ResNeSt [23] branch for local morphology and a frozen DINOv3 [24] branch for global semantic priors, coupled with a cross-path gating mechanism for adaptive fusion. Finally, an Uncertainty-Gated Margin (UGM) loss enables noise-tolerant optimization from imperfect pseudo-labels. Our main contributions are summarized as follows:

- We identify that CAM-based methods are biased toward staining-driven appearance rather than causal tissue morphology, and propose C²RM, a counterfactual causal module that removes confounding context to produce high-fidelity pseudo-labels.
- We propose a Dual-Path Structural-Semantic segmentation framework where morphological structure guides foundation semantic injection via cross-path

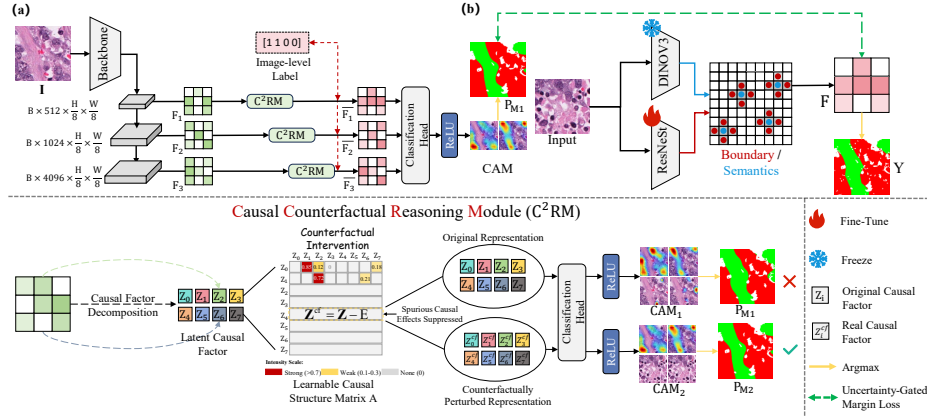


Fig. 1: Overview of C²RM-Seg. (a) Pseudo-Mask Generation: The C²RM rectifies CAMs by performing counterfactual intervention $Z^{cf} = Z - E$ on latent factors Z using a learnable causal matrix A . (b) Segmentation Training: A Dual-Path Structural-Semantic Architecture synergizes ResNeSt [23] features with frozen DINOv3 [24] priors via cross-path gating. The model is optimized using a UGM loss that dynamically re-weights supervision based on prediction uncertainty to mitigate label noise.

gating, integrating ResNeSt with a frozen DINOv3 for boundary-preserving global semantics, together with an Uncertainty-Gated Margin (UGM) loss for noise-tolerant optimization.

- Extensive experiments on public histopathology benchmarks demonstrate that C²RM-Seg outperforms state-of-the-art WSSS methods in both segmentation accuracy and boundary quality.

2 Method

2.1 Overview

As shown in Fig. 1, given a histopathology image I with an image-level label y , C²RM-Seg adopts a two-stage weakly supervised pipeline. First, we generate pseudo-masks by producing CAMs with the proposed C²RM, which reduces spurious contextual bias via counterfactual de-confounding. Second, we train a segmentation network with a Dual-Path Structural-Semantic Architecture that fuses local morphological cues and global foundation priors through cross-path gating. Finally, a UGM loss is used to alleviate residual noise in pseudo-masks during optimization.

2.2 Causal Counterfactual Reasoning Module

Spurious correlations are prevalent in histopathological images, where label-relevant tissue morphology, such as tumor, necrotic, lymphocyte-rich, and stro-

mal regions, is often entangled with label-irrelevant acquisition and style factors, including staining variability, scanner-specific responses, tissue-processing artifacts, and co-occurring background textures. We regard these style-related variations as latent confounders and introduce C²RM, an SCM-inspired counterfactual reasoning module that mitigates their influence by performing de-confounding through counterfactual feature correction.

Let $X \in \mathbb{R}^{C \times H \times W}$ be encoder features. We assume latent morphology causes M and latent style/acquisition variables S jointly affect the learned factors $\mathbf{Z}$, and $\mathbf{Z}$ determines the prediction $\hat{Y}$ (i.e., $S \rightarrow \mathbf{Z} \rightarrow \hat{Y}$ and $M \rightarrow \mathbf{Z} \rightarrow \hat{Y}$). Our goal is to estimate counterfactual factors where incoming confounding influences are removed.

We project channels into K factor subspaces ($D = C/K$) with a learnable 1×1 projection $\Phi(\cdot)$ and compute global factor statistics by average pooling:

$$\mathbf{z}_k = \frac{1}{HW} \sum_{h=1}^H \sum_{w=1}^W \Phi_k(X_{h,w}), \quad k \in \{1, \dots, K\}, \quad \mathbf{z}_k \in \mathbb{R}^D. \quad (1)$$

We learn a directed factor graph with adjacency $A_c \in \mathbb{R}^{K \times K}$, where $A_{c,i,j}$ represents the influence of factor j on factor i . To avoid self-loops, we mask the diagonal and apply row-wise softmax:

$$\mathbf{A} = \text{softmax}(A_c \odot (\mathbf{1} - I)), \quad A_{i,i} = 0. \quad (2)$$

The confounding contribution to factor i is estimated by aggregating its parents:

$$\mathbf{E}_i = \sum_{j \neq i} A_{i,j} \mathbf{z}_j. \quad (3)$$

Under a linear additive SCM approximation for factor interactions, removing incoming confounding edges yields a counterfactual factor that subtracts the estimated parental contribution:

$$\mathbf{z}_i^{cf} \approx \mathbf{z}_i - \mathbf{E}_i, \quad \mathbf{Z}^{cf} = \mathbf{Z} - \mathbf{A}\mathbf{Z}, \quad (4)$$

where $\mathbf{Z} \in \mathbb{R}^{K \times D}$ stacks $\{\mathbf{z}_k\}_{k=1}^K$. This operation suppresses components explainable by co-varying style/background factors (e.g., stain-driven textures) and preserves morphology-dominant signals.

Finally, we inject $\mathbf{Z}^{cf}$ back to recalibrate spatial features via a gated residual correction:

$$\mathbf{G} = \sigma(\mathcal{R}(\text{Broadcast}(\mathbf{Z}^{cf}))), \quad \hat{X} = X + \gamma \cdot (\mathbf{G} \odot X), \quad (5)$$

where $\mathcal{R}(\cdot)$ is a 1×1 bottleneck block, σ is Sigmoid, $\text{Broadcast}(\mathbf{Z}^{cf}) \in \mathbb{R}^{C \times H \times W}$ replicates the channel-wise counterfactual factors along spatial dimensions, and γ is a learnable scalar.

2.3 Dual-Path Structural-Semantic Architecture

To bridge the gap between local morphological fidelity and global pathological semantics, we propose a Dual-Path Structural-Semantic Architecture. This framework is designed to synergize high-resolution structural textures with robust features from large-scale foundation models, while simultaneously addressing the label noise inherent in pseudo-mask supervision.

The architecture comprises two specialized branches. The Structural Branch utilizes a ResNeSt-200e backbone followed by a pyramid pooling module to extract multi-scale spatial features, denoted as $F_S \in \mathbb{R}^{C \times H \times W}$. In parallel, the semantic branch leverages a vision transformer (ViT-B/16) pre-trained via the DINOv3 objective. To exploit the generalized representations of the foundation model while preventing overfitting to noisy pseudo-labels, the ViT backbone parameters Θ_{ViT} remain frozen. A learnable semantic adapter $\mathcal{A}_c$ projects the ViT embeddings into the structural manifold:

$$F_V = \text{Up}(\sigma(\mathcal{B}(\mathbf{W}_{ad} \cdot \mathcal{E}_{ViT}(X; \Theta_{ViT})))) , \quad (6)$$

where $\mathcal{E}_{ViT}$ represents the frozen encoder, $\mathbf{W}_{ad}$ denotes the adapter weights, and $\text{Up}(\cdot)$ denotes bilinear upsampling for spatial alignment.

To integrate these heterogeneous representations, we employ a cross-path gating mechanism. We posit that global semantic priors should be modulated by local structural evidence to maintain boundary integrity. A spatial gating map $G = \sigma(\Psi(F_S))$ is derived from F_S to selectively weight the semantic contribution:

$$\hat{F} = F_S + \sigma(\Psi(F_S)) \odot F_V, \quad (7)$$

where Ψ denotes a 1×1 convolutional bottleneck and $\odot$ represents the element-wise Hadamard product. This gating ensures that semantic information is concentrated on regions with high structural saliency.

Supervision via pseudo-masks introduces aleatoric uncertainty, particularly at tissue interfaces. To robustify the learning process, we formulate the Uncertainty-Gated Margin (UGM) loss. For each pixel i , let $p_{i,y}$ be the predicted probability for the target class y and $p_{i,s} = \max_{j \neq y} p_{i,j}$ be the probability of the most competitive alternative class. We define the pixel-wise margin violation $\mathcal{M}_i$ and the confidence penalty $\mathcal{C}_i$ as:

$$\mathcal{M}_i = [m - (p_{i,y} - p_{i,s})]_+, \quad \mathcal{C}_i = -\log(p_{i,y} + \epsilon), \quad (8)$$

where m is a predefined safety margin, $[\cdot]_+$ denotes the ReLU operation, and ϵ is a small constant for numerical stability.

The total optimization objective is governed by a dynamic uncertainty weight α_i , which adaptively shifts the focus between margin enforcement and confidence maximization. Given the local uncertainty $u_i = 1 - p_{i,y}$, the integrated loss function $\mathcal{L}$ for a set of valid pixels Ω is defined as:

$$\mathcal{L} = \frac{1}{|\Omega|} \sum_{i \in \Omega} \left[\frac{1}{1 + e^{-\kappa(u_i - 0.5)}} \mathcal{C}_i + \left(1 - \frac{1}{1 + e^{-\kappa(u_i - 0.5)}} \right) \mathcal{M}_i \right], \quad (9)$$

Table 1: Quantitative comparison on the BCSS-WSSS and LUAD-HistoSeg datasets. The best and second-best results are highlighted in **bold** and underlined, respectively. The reported standard deviations are computed over five independent runs.

Sup. Method	Venue	Efficiency		BCSS-WSSS			LUAD-HistoSeg			
		GFLOPs↓	FPS↑	mIoU↑	bIoU↑	HD95↓	mIoU↑	bIoU↑	HD95↓	
$\mathcal{F}$	SAM [26]	ICCV'23	5471.62	2.11	66.35 _{0.78}	44.26 _{0.46}	33.45 _{0.82}	69.37 _{0.37}	48.38 _{0.26}	24.81 _{0.63}
	SAM2 [27]	arXiv'24	121.68	135.45	69.44 _{0.93}	<u>48.91_{0.78}</u>	29.87 _{0.75}	77.02 _{0.29}	45.77 _{0.13}	20.35 _{0.48}
	MedSAM [28]	NC'24	738.01	9.59	65.23 _{0.61}	48.07 _{0.52}	34.12 _{1.04}	75.23 _{0.51}	43.22 _{0.38}	21.64 _{0.72}
$\mathcal{W}$	TPRO [29]	MICCAI'23	67.79	203.34	68.17 _{0.83}	45.84 _{0.47}	31.56 _{0.94}	76.53 _{0.92}	46.67 _{0.41}	20.92 _{0.55}
	HAMIL [19]	TMI'23	120.63	136.63	66.33 _{0.72}	43.07 _{1.18}	33.78 _{1.12}	75.02 _{0.96}	46.10 _{0.32}	22.14 _{0.84}
	MLPS [20]	MedIA'22	416.64	39.56	68.14 _{1.03}	46.84 _{0.83}	32.10 _{0.88}	74.11 _{1.38}	45.96 _{0.97}	23.47 _{0.91}
	SIPE [30]	CVPR'22	611.04	26.97	60.51 _{0.96}	36.71 _{0.65}	39.24 _{1.45}	70.55 _{0.38}	41.01 _{0.57}	26.89 _{1.10}
	ARML [18]	MICCAI'24	403.66	40.83	68.57 _{1.21}	43.33 _{0.87}	30.94 _{0.96}	76.24 _{1.05}	45.26 _{0.77}	21.08 _{0.68}
	OEEM [22]	MICCAI'22	<u>105.02</u>	<u>156.94</u>	67.39 _{1.58}	45.08 _{0.73}	32.85 _{1.20}	74.59 _{1.69}	44.17 _{1.12}	22.73 _{0.85}
	ESFAN [11]	MICCAI'25	154.43	147.08	<u>71.41_{0.26}</u>	47.07 _{0.63}	<u>28.08_{0.48}</u>	<u>79.29_{0.35}</u>	<u>49.02_{0.67}</u>	<u>18.55_{0.52}</u>
	Ours	–	191.47	86.08	72.17_{0.83}	49.52_{0.61}	27.31_{0.42}	79.62_{0.77}	49.47_{0.58}	18.07_{0.39}

where κ is a scaling factor controlling the transition steepness. This formulation ensures that for certain predictions (low u_i), the model prioritizes expanding the decision boundary margin, whereas for uncertain or potentially mislabeled regions (high u_i), it reverts to a conservative confidence-based penalty to avoid over-penalizing ambiguous samples.

3 Experiments

3.1 Experimental Setup

Datasets. We evaluate C²RM-Seg on two public datasets. **LUAD-HistoSeg** [20] (lung adenocarcinoma) comprises 54 WSIs with four phenotypes (TE, TAS, NEC, LYM). It contains 16,678 training patches (image-level labels) and 607 validation/testing patches with dense pixel-level annotations. **BCSS** [25] (breast cancer) includes 151 WSIs targeting TUM, STR, LYM, and NEC, split into 23,422 training patches and 8,404 validation/testing patches with ground-truth masks.

Implementation Details. Experiments are conducted in PyTorch on a NVIDIA RTX 3090. The classification stage uses ResNet38d with $K=8$ latent factors, trained for 20 epochs (batch size 20) via PolyOptimizer (initial LR 1×10^{-2}). For segmentation, the structural branch (ResNeSt-200e) is optimized by SGD (LR 1×10^{-2}), while the semantic branch (frozen DINOv3-Base + adapter) is trained via AdamW (LR 5×10^{-3}). We set the loss hyperparameters to $m=0.3$ and $\kappa=5$. Performance is reported using mIoU, bIoU, and HD95.

3.2 Comparison Study

As reported in Table 1, C²RM-Seg establishes the state-of-the-art performance on both benchmarks. On BCSS-WSSS, it achieves 72.17% mIoU and a HD95

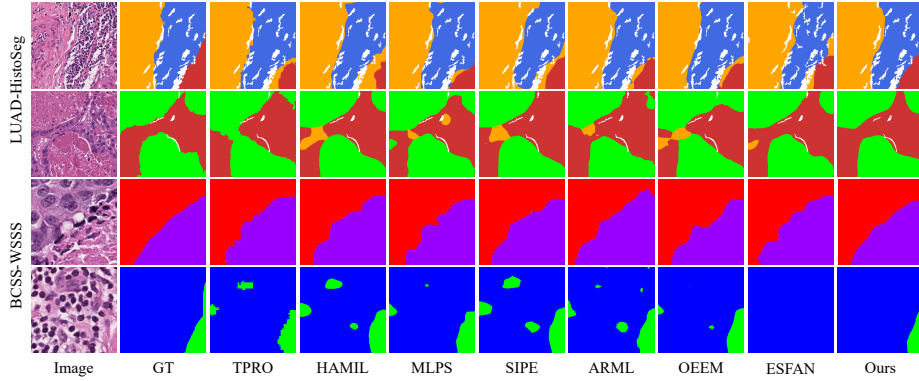


Fig. 2: Qualitative segmentation comparisons on the LUAD-HistoSeg and BCSS-WSSS datasets. **Color legend:** LUAD-HistoSeg (**TE**, **NEC**, **LYM**, **TAS**); BCSS-WSSS (**TUM**, **STR**, **LYM**, **NEC**).

Table 2: Ablation study of main modules on LUAD-HistoSeg.

C ² RM	Dual-Path	UGM	GFLOPs↓	FPS↑	mIoU↑	bIoU↑	HD95↓
×	×	×	112.35	208.42	71.45 _{1.12}	39.23 _{0.94}	27.84 _{1.56}
✓	×	×	124.68	182.15	74.12 _{0.98}	42.56 _{0.82}	23.45 _{1.21}
×	✓	×	148.92	134.56	73.28 _{1.05}	41.87 _{0.88}	24.62 _{1.34}
×	×	✓	135.24	168.37	75.89 _{0.92}	44.75 _{0.79}	21.36 _{1.12}
✓	✓	×	161.25	118.49	76.42 _{0.87}	45.64 _{0.75}	20.55 _{0.95}
✓	×	✓	147.57	152.04	78.56 _{0.81}	48.34 _{0.64}	19.12 _{0.78}
×	✓	✓	171.81	102.73	77.94 _{0.84}	47.12 _{0.71}	19.88 _{0.82}
✓	✓	✓	191.47	86.08	79.62 _{0.77}	49.47 _{0.58}	18.07 _{0.39}

of 27.31, surpassing the leading WSSS competitor, ESFAN. For rigorous evaluation, foundation models (e.g., SAM2, MedSAM) are prompted using our first-stage pseudo-masks; even under this strict condition, our end-to-end framework exhibits superior accuracy. This robust generalization is similarly observed on LUAD-HistoSeg (79.62% mIoU, 18.07 HD95). Visual comparisons (Fig. 2) corroborate these quantitative gains. Compared to baselines, C²RM-Seg delineates complex tissue interfaces with higher boundary fidelity and significantly suppresses false-positive predictions in morphologically ambiguous regions. Despite integrating causal reasoning and foundation priors, our model maintains a competitive computational footprint (191.47 GFLOPs, 86.08 FPS), confirming its practical scalability for large-scale histopathology analysis.

3.3 Ablation Study

As shown in Table 2, ablations on LUAD-HistoSeg show that each module independently improves the baseline, indicating complementary corrections of ap-

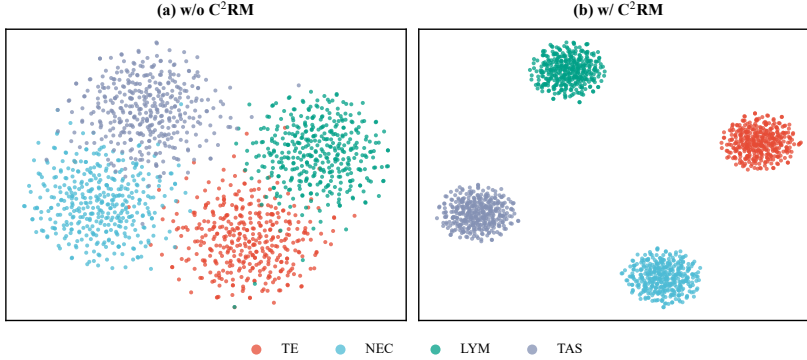


Fig. 3: t-SNE visualization. **(a)** Baseline features are severely entangled by spurious staining confounders. **(b)** C²RM eliminates these biases via counterfactual intervention, yielding compact clusters and distinct inter-class separability.

pearance bias, morphology–semantic mismatch, and label noise. In particular, C²RM notably improves both region overlap and boundary quality, suggesting that removing staining-driven confounding restores morphology-aligned pseudo-masks. This deconfounding capability is visually validated by the t-SNE feature space analysis in Fig. 3, where C²RM transforms severely entangled, stain-biased representations into highly compact and semantically separable clusters. Introducing the Dual-Path architecture further improves boundary fidelity by injecting foundation semantics under structural guidance, reducing morphology–semantic mismatch. The UGM loss significantly decreases HD95, indicating effective suppression of residual pseudo-label noise through uncertainty-aware margin constraints. When combining two modules, performance continues to increase, confirming that causal pseudo-label correction, structure-guided semantic fusion, and noise-aware optimization act synergistically. The full model achieves the best results across all metrics (79.62% mIoU, 49.47% bIoU, 18.07 HD95), demonstrating that correcting appearance–morphology bias and enforcing structure-consistent semantics jointly drive the overall performance gain. Despite increased cost, the full model remains computationally efficient.

4 Conclusion

In this paper, we present C²RM-Seg, a novel framework for weakly supervised histopathology image segmentation that fundamentally addresses the pervasive issue of histological confounding. Recognizing that traditional CAM-based pseudo-labels are inherently biased by spurious staining-driven correlations rather than causal tissue morphology, we formulate CAM generation within a rigorous Structural Causal Model. By explicitly executing counterfactual interventions via C²RM, our approach disentangles these confounding factors, synthesizing high-fidelity, morphologically faithful pseudo-masks. To fully leverage these causal

priors, we further propose a synergistic Dual-Path Structural–Semantic Architecture that harmonizes task-specific structural features with frozen foundation-model representations via adaptive gating. Optimized with the UGM loss to ensure robustness against residual noise, C²RM-Seg establishes new state-of-the-art performance on the BCSS-WSSS and LUAD-HistoSeg benchmarks.

Acknowledgments. This work was supported in part by Guangxi Natural Science Foundation (Nos. 2024GXNSFFA010014, AB25069496), National Natural Science Foundation of China (Nos. 82360356, T2541065), and Innovation Project of Guangxi Graduate Education (No. YCSW2025362, YCSW2026398).

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

References

1. Andani, S., Chen, B., Ficek-Pascual, J., Heinke, S., Casanova, R., Hild, B.F., Sobottka, B., Bodenmiller, B., Koelzer, V.H., et al.: Histopathology-based protein multiplex generation using deep learning. *Nature Machine Intelligence* **7**(8), 1292–1307 (2025)
2. Wang, X., Zhao, J., Marostica, E., Yuan, W., Jin, J., Zhang, J., Li, R., Tang, H., Wang, K., Li, Y., et al.: A pathology foundation model for cancer diagnosis and prognosis prediction. *Nature* **634**(8035), 970–978 (2024)
3. Feng, S., Wang, H., Han, C., Liu, Z., Zhang, H., Lan, R., Pan, X.: Weakly supervised gland segmentation with class semantic consistency and purified labels filtration. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. vol. 39, pp. 2987–2995 (2025)
4. An, Y., Chen, J., Lin, H., Liu, Z., Feng, S., Zhang, H., Lan, R., Liu, Z., Pan, X.: Ca-mlif: Cross-attention and multimodal low-rank interaction fusion framework for tumor prognostic prediction. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. vol. 39, pp. 1764–1772 (2025)
5. 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)
6. Cheng, X., Ding, W., Huang, J., Ju, H., Zhou, T., Guo, J., Pedrycz, W.: Fcaformer: Fuzzy-enhanced class-aware attention based transformer for weakly supervised histopathology image segmentation. *IEEE Transactions on Fuzzy Systems* (2025)
7. Pan, X., Zhang, H., Deng, H., Wang, H., Li, L., Liu, Z., Wang, L., An, Y., Lu, C., Liu, Z., et al.: Weakly supervised histopathology tissue semantic segmentation with multi-scale voting and online noise suppression. *Engineering Applications of Artificial Intelligence* **156**, 111100 (2025)
8. Chikontwe, P., Sung, H.J., Jeong, J., Kim, M., Go, H., Nam, S.J., Park, S.H.: Weakly supervised segmentation on neural compressed histopathology with self-equivariant regularization. *Medical Image Analysis* **80**, 102482 (2022)
9. Li, K., Qian, Z., Han, Y., Chang, E.I.C., Wei, B., Lai, M., Liao, J., Fan, Y., Xu, Y.: Weakly supervised histopathology image segmentation with self-attention. *Medical Image Analysis* **86**, 102791 (2023)
10. Fan, J., Sun, Q., Di, Y., Bao, J., Lv, T., Liu, Y., Hu, X., Li, L., Cui, X., Pan, X.: Dipathmamba: A domain-incremental weakly supervised state space model for pathology image segmentation. *Medical Image Analysis* **103**, 103563 (2025)

11. 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)
12. Huan, Z., Pan, X., Zhang, H., Feng, S., Lan, R., Wang, H., Lu, H., Liu, Z.: Bridging rgb and hematoxylin components: An interleaved guidance and fusion framework for point supervised nuclei segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 37518–37527 (2026)
13. Kang, Y., Li, H., Shi, X., Zhang, X., Xing, Y., Wen, Y., Wang, Y., Cui, L., Feng, J., Yang, L.: Exploring unbiased activation maps for weakly supervised tissue segmentation of histopathological images. *IEEE Transactions on Medical Imaging* (2025)
14. Fu, J., Wang, G., Lu, T., Yue, Q., Vercauteren, T., Ourselin, S., Zhang, S.: Um-cam: Uncertainty-weighted multi-resolution class activation maps for weakly-supervised segmentation. *Pattern Recognition* **160**, 111204 (2025)
15. Feng, S., Pan, X., Wang, H., Liu, Z., Zhang, W., Lan, R.: Long-tailed and inter-class homogeneity matters in multi-class weakly supervised tissue segmentation of histopathology images. *IEEE Transactions on Image Processing* **35**, 2513–2528 (2026)
16. Feng, S., Pan, X., Zhang, W., Pan, M., Han, C., Lan, R.: Qupas: Sam-based semi-supervised histopathological image segmentation with quantum force field finetuning and adversarial estimation. *IEEE transactions on medical imaging* (2026)
17. 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)
18. 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)
19. 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)
20. 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)
21. Hu, X., Cao, Y., Hu, W., Wang, S., Mukhopadhyay, S.C., Liu, Y., Li, H., Li, Y., Cai, Q., Liu, Z.: Expert knowledge-guided multi-granularity multi-scale fusion for weakly-supervised histological segmentation. *Information Fusion* **125**, 103432 (2026)
22. Li, Y., Yu, Y., Zou, Y., Xiang, T., Li, X.: Online easy example mining for weakly-supervised gland segmentation from histology images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 578–587. Springer (2022)
23. Zhang, H., Wu, C., Zhang, Z., Zhu, Y., Lin, H., Zhang, Z., Sun, Y., He, T., Mueller, J., Manmatha, R., et al.: Resnest: Split-attention networks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 2736–2746 (2022)
24. Siméoni, O., Vo, H.V., Seitzer, M., Baldassarre, F., Oquab, M., Jose, C., Khali-dov, V., Szafraniec, M., Yi, S., Ramamonjisoa, M., et al.: Dinov3. *arXiv preprint arXiv:2508.10104* (2025)

25. 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)
26. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., et al.: Segment anything. In: *Proceedings of the IEEE/CVF international conference on computer vision*. pp. 4015–4026 (2023)
27. Ravi, N., Gabeur, V., Hu, Y.T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L., et al.: Sam 2: Segment anything in images and videos. *arXiv preprint arXiv:2408.00714* (2024)
28. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. *Nature Communications* **15**(1), 654 (2024)
29. Zhang, S., Zhang, J., Xie, Y., Xia, Y.: Tpro: Text-prompting-based weakly supervised histopathology tissue segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 109–118. Springer (2023)
30. Chen, Q., Yang, L., Lai, J.H., Xie, X.: Self-supervised image-specific prototype exploration for weakly supervised semantic segmentation. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 4288–4298 (2022)