Physics-Grounded Weakly Supervised Histopathological Tissue Segmentation via Frozen Tri-Domain Prototypes

Shuyu Zhang[1], Jiansong Fan[1], Wanglong Mao[1], Lihua Li[2], Yizhang Jiang[1], Pengjiang Qian[1], and Xiang Pan[1(✉)]

[1] School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214122, China

[2] Institute of Biomedical Engineering and Instrumentation, Hangzhou Dianzi University, Hangzhou 310018, China

xiangpan@jiangnan.edu.cn

Abstract. Tissue-level semantic segmentation is vital to computational pathology, yet pixel-wise annotation of whole-slide images remains prohibitively expensive, motivating weakly supervised approaches. Existing weakly supervised semantic segmentation methods face two coupled challenges: extreme intra-class morphological heterogeneity and pseudo-label noise accumulation, which jointly induce progressive semantic drift during training. In this paper, we propose PhysPro, a physics-semantics synergistic framework which stabilizes weak supervision by anchoring representation learning to immutable physical priors. Specifically, Physics-Derived Tri-Domain Static Prototypes (PhyTriSP) construct offline frozen anchors across the semantic, phase-morphological, and spectral-energy domains, which provide physically grounded references that mitigate intra-class variability. Building upon these anchors, Physics-Semantics Orchestrated Mixture of Experts (PhyOrchMoE) deterministically decouples spatial morphology from spectral energy via physics-steered tri-scale fusion and sub-band amplitude calibration, which yields complementary noise-insulated representations. A Physics-Guided Discrepancy Supervisor (PhyDS) then exploits inter-domain discrepancy across the frozen anchors, which disentangles label noise from tissue boundary ambiguity to enable tri-regime adaptive supervision with prototype-guided soft labels and category-aware spectral recalibration. Extensive experiments on BCSS-WSSS and LUAD-HistoSeg demonstrate that PhysPro consistently outperforms state-of-the-art methods.The code is available at https://github.com/zsyUp/PhysPro.

Keywords: Weakly Supervised Semantic Segmentation · Computational Pathology · Physical Prior · Frequency-Domain Decoupling

1 Introduction

Pathology image analysis plays a pivotal role in clinical diagnosis and disease prognosis [15]. Although deep learning has enabled automated histopathological analysis to approach expert-level performance [3,9], supervised segmentation

models rely heavily on exhaustive pixel-level annotations. Labeling billions of pixels in whole-slide images (WSIs) is prohibitively expensive and introduces considerable inter-observer variability. To reduce this burden, Weakly Supervised Semantic Segmentation (WSSS) has emerged as a promising paradigm that derives dense pixel-wise predictions from coarse image-level labels [2, 6, 26].

The prevailing WSSS paradigm typically synthesizes class activation maps (CAMs) from image-level supervision to serve as seed regions for pseudo-label generation, upon which a segmentation model is supervised [14, 18, 24]. While recent advances have incorporated patch-level constraints [10,29] or spatial structural priors [5] to enhance mask fidelity, histopathological WSSS remains confronted by twin challenges rarely encountered in generic computer vision tasks. Primarily, the extreme intra-class morphological heterogeneity causes the same tissue category to exhibit divergent structural manifestations across patients and regions [23], frequently yielding inconsistent or fragmented CAM activations. Exacerbating this is the recursive accumulation of pseudo-label noise; in the absence of dense pixel-level ground truth, erroneous labels inevitably corrupt the latent feature space and trigger a progressive performance decay as training evolves [22, 30]. Crucially, these factors form a detrimental feedback loop—morphological diversity amplifies initial label noise, which in turn destabilizes feature representation and precipitates irreversible semantic drift [16].

The core motivation of this work stems from the realization that frequency-domain representations provide physically-grounded descriptors that spatial features fundamentally lack. While spatial appearances in histopathology are highly sensitive to staining fluctuations and tissue pleomorphism, the Fourier phase spectrum encapsulates essential geometric and topological structures that remain largely invariant to intensity-level perturbations [28]. Simultaneously, the amplitude spectrum characterizes the intrinsic energy distribution of tissue textures, exhibiting remarkable stability across diverse appearance variations and scanning protocols [11]. We argue that anchoring representation learning to such physical invariants—spanning semantic, morphological, and spectral energy domains—can establish immutable reference anchors to effectively decouple tissue identity from appearance ambiguity. Such a physics-informed perspective offers a robust alternative to conventional spatial-only methods, representing a synergistic potential that remains largely unexplored in histopathological WSSS.

In this paper, we propose PhysPro, a physics-semantics synergistic framework designed as a cohesive system to dismantle the detrimental feedback loop between morphological noise and semantic drift. To break this cycle at its source, PhyTriSP first establishes a suite of offline frozen anchors across the semantic, phase-morphological, and spectral-energy domains. Unlike learnable prototypes adopted in recent prototype-based WSSS methods [12, 20] that risk collapsing or drifting under noisy weak supervision, these static references serve as a physical north star, providing a constant coordinate system to stabilize feature learning against extreme tissue pleomorphism. Building upon these stable anchors, PhyOrchMoE functions as the framework's structural core; it employs a deterministic physics-semantics gating mechanism to orthogonally decouple spatial morphol-

ogy from spectral energy into complementary, non-redundant expert streams. This orchestration ensures that the model preserves structural integrity while filtering out stochastic spatial distortions, effectively insulating the latent feature space. Finally, PhyDS closes the supervisory loop by treating the inter-domain discrepancy across the tri-domain anchors as a diagnostic signal. By exploiting this discrepancy, PhyDS disentangles genuine tissue boundary ambiguity from deceptive label noise, enabling a tri-regime adaptive supervision strategy that dynamically recalibrates the learning process with prototype-guided soft labels. We validate PhysPro on BCSS-WSSS [10] and LUAD-HistoSeg [1], demonstrating that our physics-anchored approach consistently and significantly outperforms state-of-the-art WSSS methods across all evaluation metrics. The main contributions of this work are as follows:

- We propose a physics-semantics synergistic framework named PhysPro, which stabilizes weak supervision by anchoring representation learning to immutable physical priors to mitigate the coupling of morphological heterogeneity and pseudo-label noise.
- We develop physics-derived tri-domain static prototypes, which construct offline frozen anchors across the semantic, phase-morphological, and spectral-energy domains to provide physically grounded references that suppress intra-class variability.
- We introduce a physics-semantics orchestrated mixture of experts, which deterministically decouples spatial morphology from spectral energy via physics-steered tri-scale fusion and sub-band amplitude calibration to yield complementary noise-insulated representations.
- We design a physics-guided discrepancy supervisor, which exploits inter-domain discrepancy across the frozen anchors to disentangle label noise from tissue boundary ambiguity, enabling tri-regime adaptive supervision with prototype-guided soft labels and category-aware spectral recalibration.

2 Methodology

2.1 Overview

Given an input image $\mathbf{X} \in \mathbb{R}^{H\times W\times 3}$ and image-level label $\mathbf{Y} \in \{0,1\}^C$, PhysPro produces pixel-wise segmentation via three sequentially coupled components, as illustrated in Fig. 1. Prior to training, PhyTriSP constructs three sets of frozen prototypes—semantic $\mathcal{P}_{\text{sem}}$, phase-morphological $\mathcal{P}_{\text{phase}}$, and spectral-energy $\mathcal{P}_{\text{energy}}$—via density-weighted aggregation, which remain fixed throughout training as shared physical references for downstream components. At each training step, the backbone extracts feature map $F \in \mathbb{R}^{H'\times W'\times d}$, which PhyOrchMoE routes through two deterministically specialized experts anchored to $\mathcal{P}_{\text{phase}}$ and $\mathcal{P}_{\text{energy}}$, yielding a noise-insulated fused representation X_{fuse}. PhyDS then computes per-pixel cross-domain discrepancy $\Delta = \text{Var}(s_{\text{sem}}, s_{\text{phase}}, s_{\text{energy}})$ to classify each pixel into reliable, boundary, or noise regimes, enabling adaptive supervision and online recalibration of $\mathcal{E}_{\text{freq}}$ via category-aware confidence.

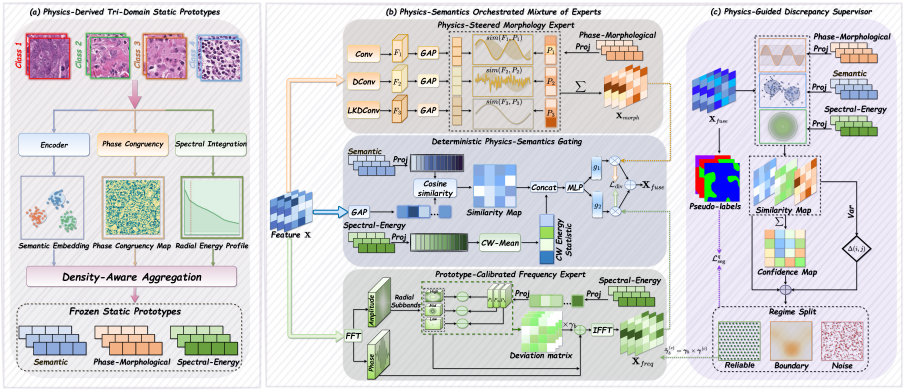

Fig. 1: The overview of the proposed PhysPro. (a) Physics-Derived Tri-Domain Static Prototypes for offline construction of semantic, phase-morphological, and spectral-energy anchors; (b) Physics-Semantics Orchestrated Mixture of Experts for deterministic spatial-spectral decoupling; (c) Physics-Guided Discrepancy Supervisor for inter-domain discrepancy-driven disentanglement of label noise from tissue boundary ambiguity, enabling tri-regime adaptive supervision and category-aware spectral recalibration.

2.2 Physics-Derived Tri-Domain Static Prototypes

PhyTriSP constructs category-wise reference prototypes in an offline calibration stage to stabilize feature learning against intra-class morphological heterogeneity and semantic drift under noisy weak supervision. For each tissue category c, descriptors aggregated over $\mathcal{S}_c$ via density-weighted averaging yield fixed d-dimensional anchors:

$$\mathcal{P}^c = \frac{\sum_{i\in\mathcal{S}_c} w_i\, f_i}{\sum_{i\in\mathcal{S}_c} w_i}, \qquad w_i = \exp\!\left(-\frac{\|f_i - \bar{f}_c\|_2}{\tau}\right), \tag{1}$$

where $\bar{f}_c$ is the unweighted category mean and τ down-weights outliers caused by residual staining artifacts. S_c contains single-category training patches, excluding mixed-category samples to reduce semantic contamination. All prototypes remain frozen throughout training to prevent reference drift.

Semantic Prototype ($\mathcal{P}_{\mathbf{sem}} \in \mathbb{R}^{C\times d}$). Training patches are encoded by the MedCLIP image encoder and aggregated via Eq. (1) per category. $\mathcal{P}_{\text{sem}}$ encodes high-level categorical identity with global discriminative structure robust to intra-class appearance variation.

Phase-Morphological Prototype ($\mathcal{P}_{\mathbf{phase}} \in \mathbb{R}^{C\times d}$). Phase congruency descriptors are extracted from the Fourier phase spectrum $\Phi(u,v) = \angle\mathcal{F}(x)$ via a multi-scale Log-Gabor filter bank. Invariance to monotonic intensity transformations ensures stability under staining variation while faithfully encoding geometric tissue structure—glandular circularity and nuclear spatial organization. The prototype is obtained via Eq. (1) over $\mathcal{S}_c$.

Spectral-Energy Prototype ($\mathcal{P}_{\text{energy}} \in \mathbb{R}^{C\times d}$). The log-amplitude spectrum $\hat{A}(u,v) = \log(|\mathcal{F}(x)|+1)$ is mapped to log-polar coordinates and integrated angularly to eliminate rotational dependence:

$$e(x) = \left[\frac{1}{2\pi}\int_0^{2\pi} \hat{A}(\log r,\, \theta)\, \mathrm{d}\theta\right]_{r=r_{\min}}^{r_{\max}}, \tag{2}$$

yielding a rotation-invariant energy profile across R logarithmic frequency bands that captures category-specific structural signatures such as stromal fiber density and cytoplasmic granularity. $\mathcal{P}^c_{\text{energy}}$ is obtained via Eq. (1) over $\mathcal{S}_c$. Together, the three frozen anchors provide physically grounded references that suppress semantic drift throughout weakly supervised training.

2.3 Physics-Semantics Orchestrated Mixture of Experts

Unlike conventional MoE architectures with stochastic routing, PhyOrchMoE enforces deterministic functional specialization by anchoring each expert to physics-derived prototypes, ensuring orthogonal decomposition of spatial morphology and spectral energy into complementary, noise-insulated streams.

Expert Specialization. The Physics-Steered Morphology Expert ($\mathcal{E}_{\text{morph}}$) captures hierarchical tissue structures via a tri-scale pathway (Conv, DilConv d=5, DilConv d=9) spanning nuclei-to-gland receptive fields. Each branch F_s is weighted by its cosine alignment with $\mathcal{P}_{\text{phase}}$, so that scale selection is governed by structural consistency with the frozen anchor rather than learned attention:

$$X_{\text{morph}} = \sum_{s=1}^{3} \text{sim}(F_s, \mathcal{P}_{\text{phase}}) \cdot F_s. \tag{3}$$

The Prototype-Calibrated Frequency Expert ($\mathcal{E}_{\text{freq}}$) decomposes the amplitude spectrum A into three radial sub-bands $\{A_b\}_{b=1}^3$, each calibrated against the corresponding log-polar sub-prototype $\mathcal{P}^b_{\text{energy}} \in \mathbb{R}^{d/3}$:

$$\hat{A} = \bigoplus_{b=1}^{3} \big[A_b + \gamma_b\, \delta_b(A_b,\, \mathcal{P}^b_{\text{energy}})\big], \tag{4}$$

where $\delta_b(\cdot)$ is a lightweight linear projection and γ_b is initialized to zero for progressive activation. The sub-band partition is physically consistent with the log-polar energy descriptor in Eq. (2), enabling differentiated calibration of low-frequency morphology and high-frequency boundary details. The phase spectrum Φ remains unmodified to preserve geometric integrity.

Deterministic Physics-Semantics Gating. The gating descriptor v concatenates the channel-wise energy statistic H with the semantic similarity map S (cosine similarity between features and $\mathcal{P}_{\text{sem}}$). A lightweight MLP yields normalized routing weights, producing the fused representation:

$$X_{\text{fuse}} = g_1(v)\, X_{\text{morph}} + g_2(v)\, X_{\text{freq}}, \quad g_1(v) + g_2(v) = 1. \tag{5}$$

Diversity Constraint. To preclude expert collapse, a diversity loss jointly promotes balanced routing and weight-space orthogonality:

$$\mathcal{L}_{\text{div}} = -\sum_{e} \bar{g}_e \log \bar{g}_e + \lambda \sum_{i \neq j} \frac{|\langle W_i, W_j \rangle|}{\|W_i\| \, \|W_j\|}. \tag{6}$$

2.4 Physics-Guided Discrepancy Supervisor

Conventional confidence estimators suppress uncertain pixels uniformly, conflating label noise with biologically meaningful tissue-boundary ambiguity. PhyDS resolves this by treating the cross-domain variance of frozen tri-domain anchor similarities, $\Delta(i,j) = \text{Var}(s_{\text{sem}}, s_{\text{phase}}, s_{\text{energy}})$, as a physics-grounded diagnostic signal. For each pixel (i,j), cosine similarities $s_d(i,j)$ are first computed between its feature vector in X_{fuse} and the frozen prototype of its pseudo-labeled category across $d \in \{\text{sem}, \text{phase}, \text{energy}\}$, then aggregated with local spatial coherence into a unified confidence score $q(i,j) \in [0,1]$. Adaptive thresholds are computed per mini-batch using the 30th/70th percentiles of Δ and the medians of q and s_{sem}.

The resulting Δ induces three physically interpretable regimes: (i) low Δ, high q: reliable; (ii) high Δ, high s_{sem}: genuine tissue boundary; (iii) high Δ, low s_{sem}: corrupted pseudo-label. This decomposition is only achievable via physically distinct tri-domain anchors; a single-domain estimator inevitably conflates the two failure modes. Regime membership then jointly governs per-pixel supervision weight and label target:

$$\big(w, \hat{y}\big) = \begin{cases} \big(q,\ y\big) & \text{(reliable)}, \\ \big(\beta q,\ \tilde{y}\big) & \text{(boundary)}, \\ \big(\tau,\ y\big) & \text{(noise)}, \end{cases} \tag{7}$$

where $\tilde{y}_k$=softmax($s_{\text{sem}}^{(k)}/T$) are prototype-guided soft labels, β<1 downweights boundary pixels, and τ prevents gradient starvation. The quality-aware segmentation objective is:

$$\mathcal{L}_{\text{seg}}^{q} = -\frac{1}{HW} \sum_{i,j} w(i,j) \sum_{k} \hat{y}_k(i,j) \log p_k(i,j). \tag{8}$$

Category-aware confidence $\bar{q}^{(c)}$ over reliable pixels further recalibrates $\mathcal{E}_{\text{freq}}$ via $\tilde{\gamma}_b^{(c)} = \gamma_b \cdot \bar{q}^{(c)}$, elevating the frozen anchors from passive references to active arbiters of supervision quality.

2.5 Overall Optimization Objective

PhysPro is trained end-to-end under weak slide-level supervision via a composite objective:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{seg}}^{q} + \eta \, \mathcal{L}_{\text{cls}} + \lambda \, \mathcal{L}_{\text{div}}, \tag{9}$$

where $\mathcal{L}_{\text{cls}}$ provides global semantic constraint from image-level labels, $\mathcal{L}_{\text{seg}}^{q}$ enforces tri-regime reliability-aware pixel supervision, and $\mathcal{L}_{\text{div}}$ regularizes expert specialization via balanced routing and weight-space orthogonality.

3 Experiments

Datasets. We evaluate PhysPro on two large-scale pathology benchmarks. LUAD-HistoSeg [1] contains 31,826 images (224×224) across four tissue categories: tumor epithelium, tumor-associated stroma, necrosis, and lymphocytes. BCSS-WSSS [10] comprises 17,286 images (224×224) across five categories: tumor, stroma, lymphocyte infiltration, necrosis, and others.

Table 1: Performance comparison with state-of-the-art methods on BCSS-WSSS and LUAD-HistoSeg. Best results in bold.

Method	BCSS-WSSS				LUAD-HistoSeg			
	mDice	mIoU	mP	mR	mDice	mIoU	mP	mR
HistoSegNet [3]	0.505	0.276	0.582	0.571	0.641	0.478	0.654	0.662
SC-CAM [4]	0.729	0.663	0.742	0.733	0.715	0.641	0.756	0.761
WSSS-Tissue [10]	0.767	0.697	0.786	0.758	0.818	0.756	0.826	0.822
SSC [5]	0.755	0.654	0.727	0.727	0.772	0.731	0.783	0.748
Mctformer+ [26]	0.772	0.707	0.790	0.766	0.825	0.763	0.830	0.825
DuPL [24]	0.781	0.715	0.799	0.782	0.823	0.769	0.838	0.831
Pbip [20]	0.820	0.695	0.826	0.814	0.866	0.764	0.874	0.859
Ours	**0.843**	**0.722**	**0.848**	**0.857**	**0.874**	**0.776**	**0.881**	**0.867**

Competing Methods and Evaluation Metrics. To comprehensively evaluate PhysPro, we compare it with multiple state-of-the-art weakly supervised segmentation methods, including HistoSegNet [3], SC-CAM [4], SSC [5], WSSS-Tissue [10], Mctformer+ [26], DuPL [24], and Pbip [20]. All methods are evaluated using four metrics: mean Dice coefficient (mDice), mean Intersection-over-Union (mIoU), mean Precision (mP), and mean Recall (mR).

Implementation Details. All experiments were conducted on an NVIDIA GeForce RTX A6000 GPU (48GB VRAM), with results averaged over five runs. The network was trained using Adam with a learning rate of 1×10^{-4} and weight decay of 5×10^{-4}, using a batch size of 16. The loss balancing parameters were set to $\eta = 0.5$ and $\lambda = 0.3$.

Comparison with SOTA Methods. Table 1 reports results on BCSS-WSSS and LUAD-HistoSeg. PhysPro consistently outperforms all competitors across both benchmarks. On BCSS-WSSS, PhysPro surpasses the strongest baseline Pbip by +2.3% mDice and +2.7% mIoU, and achieves gains of +6.2% and +0.7% over DuPL and Mctformer+, respectively. On LUAD-HistoSeg, PhysPro outperforms Pbip by +0.8% mDice and +1.2% mIoU, with improvements in both mP and mR confirming more precise boundary localization. As shown in

Fig. 2, PhysPro produces cleaner tissue boundaries and fewer misclassified regions across all competing methods.

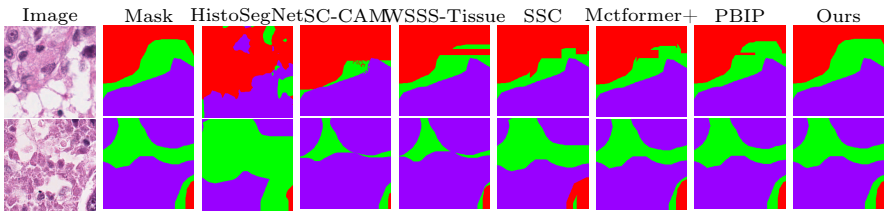

Fig. 2: Visualization of segmentation results on BCSS-WSSS.

Ablation Study. Table 2 evaluates each proposed component on BCSS-WSSS. The baseline yields 67.5% mDice and 54.6% mIoU. Incorporating PhyTriSP substantially boosts performance to 82.2% mDice and 70.1% mIoU, demonstrating that physically grounded frozen anchors stabilize feature learning under severe intra-class morphological heterogeneity. Further introducing PhyOrchMoE improves results to 82.8% mDice and 70.9% mIoU, confirming that spatial–spectral decoupling enhances representational discriminability. In contrast, applying PhyDS without PhyOrchMoE yields only 81.8% mDice and 70.7% mIoU, revealing that noise-boundary disentanglement requires well-decoupled representations as prerequisite. The complete model achieves 84.3% mDice and 72.2% mIoU, validating that all three components jointly address complementary sources of instability in weakly supervised semantic segmentation.

Table 2: Ablation analysis of different components in the proposed PhysPro on BCSS-WSSS.

Baseline	PhyTriSP	PhyOrchMoE	PhyDS	mDice	mIoU	mPrecision	mRecall
✓				0.675	0.546	0.682	0.693
✓	✓			0.822	0.701	0.796	0.851
✓	✓	✓		0.828	0.709	0.804	0.855
✓	✓		✓	0.818	0.707	0.822	0.820
✓	✓	✓	✓	**0.843**	**0.722**	**0.848**	**0.857**

4 Conclusion

We present PhysPro, a physics-semantics synergistic framework for weakly supervised histopathological segmentation that anchors representation learning to frozen tri-domain prototypes, decouples spatial-spectral representations, and exploits inter-domain discrepancy to disentangle label noise from tissue boundary ambiguity via tri-regime adaptive supervision. Experiments on BCSS-WSSS

and LUAD-HistoSeg demonstrate consistent improvements over state-of-the-art methods, validating the effectiveness of physics-prior-guided learning for robust weakly supervised semantic segmentation in computational pathology.

Acknowledgments. This work was supported in part by the National Natural Science Foundation of China under Grants W2411054, U21A20521, and 62271178; the National Foreign Expert Project of China under Grant G2023144009L; the Zhejiang Provincial Natural Science Foundation of China under Grant LR23F010002; the Wuxi Health Commission Precision Medicine Project under Grant J202106; the Jiangsu Provincial Six Talent Peaks Project under Grant YY-124; and the Construction Project of Shanghai Key Laboratory of Molecular Imaging under Grant 18DZ2260400.

Disclosure of Interests. The authors declare that they have no competing interests.

References

1. Amgad, M., Elfandy, H., Hussein, H., et al.: Structured crowdsourcing enables convolutional segmentation of histology images. Bioinformatics **35**(18), 3461–3467 (2019)
2. Araslanov, N., Roth, S.: Single-stage semantic segmentation from image labels. In: CVPR 2020, pp. 4253–4262 (2020)
3. Chan, L., Hosseini, M.S., Rowsell, C., et al.: Histosegnet: Semantic segmentation of histological tissue type in whole slide images. In: ICCV 2019, pp. 10662–10671 (2019)
4. Chang, Y.T., Wang, Q., Hung, W.C., et al.: Weakly-supervised semantic segmentation via sub-category exploration. In: CVPR 2020, pp. 8991–9000 (2020)
5. Chen, T., Yao, Y., Huang, X., et al.: Spatial structure constraints for weakly supervised semantic segmentation. IEEE Transactions on Image Processing **33**, 1136–1148 (2024)
6. Chen, Z., Wang, T., Wu, X., et al.: Class re-activation maps for weakly-supervised semantic segmentation. In: CVPR 2022, pp. 969–978 (2022)
7. Ding, M., Chen, W., Luo, X., Zhong, H., Shen, L.: FProtoSeg: Fine-grained prototype alignment for weakly supervised semantic segmentation of histopathology images. Pattern Recognition, 113126 (2026)
8. Du, Y., Fu, Z., Liu, Q., Wang, Y.: Weakly supervised semantic segmentation by pixel-to-prototype contrast. In: CVPR 2022, pp. 4320–4329 (2022)
9. Ehteshami Bejnordi, B., Veta, M., Johannes van Diest, P., et al.: Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. JAMA **318**(22), 2199–2210 (2017)
10. Han, C., Lin, J., Mai, J., et al.: Multi-layer pseudo-supervision for histopathology tissue semantic segmentation using patch-level classification labels. Medical Image Analysis **80**, 102487 (2022)
11. He, Y., Lu, Z., Huan, H.: SF3Net: Frequency-Domain Enhanced Segmentation Network for High-Resolution Remote Sensing Imagery. Remote Sensing **17**(22), 3734 (2025)
12. Le, K., Vu, A.M., Vo, T.K.T., et al.: LPD: Learnable prototypes with diversity regularization for weakly supervised histopathology segmentation. In: ISBI 2026, pp. 1–5 (2026)

13. Li, R., Mai, Z., Zhang, Z., et al.: TransCAM: Transformer attention-based CAM refinement. J. Vis. Commun. Image Represent. **92**, 103800 (2023)
14. Li, Y., Kuang, Z., Liu, L., et al.: Pseudo-mask matters in weakly-supervised semantic segmentation. In: ICCV 2021, pp. 6964–6973 (2021)
15. Litjens, G., Kooi, T., Bejnordi, B.E., et al.: A survey on deep learning in medical image analysis. Medical Image Analysis **42**, 60–88 (2017)
16. Pan, X., Zhang, H., Deng, H., et al.: Weakly supervised histopathology tissue semantic segmentation with multi-scale voting and online noise suppression. Engineering Applications of Artificial Intelligence **156**, 111100 (2025)
17. Ru, L., Zheng, H., Zhan, Y., Du, B.: Token contrast for weakly-supervised semantic segmentation. In: CVPR 2023, pp. 3093–3102 (2023)
18. Sun, W., Zhang, J., Barnes, N.: Inferring the class conditional response map for weakly supervised semantic segmentation. In: WACV 2022, pp. 2878–2887 (2022)
19. Tang, F., Xu, Z., Qu, Z., et al.: Hunting attributes: Context prototype-aware learning for weakly supervised semantic segmentation. In: CVPR 2024, pp. 3324–3334 (2024)
20. Tang, Q., Fan, L., Pagnucco, M., Song, Y.: Prototype-based image prompting for weakly supervised histopathological image segmentation. In: CVPR 2025, pp. 30271–30280 (2025)
21. Wang, J., Dai, T., Zhang, B., et al.: POT: Prototypical optimal transport for weakly supervised semantic segmentation. In: CVPR 2025, pp. 15055–15064 (2025)
22. Wang, Y., Zhang, J., Kan, M., Shan, S.: Learning pseudo labels for semi-and-weakly supervised semantic segmentation. Pattern Recognition **132**, 108925 (2022)
23. Wu, H., Zhang, H.: Superpixel boundary correction for weakly-supervised semantic segmentation on histopathology images. In: NNICE 2025, pp. 310–313 (2025)
24. Wu, Y., Ye, X., Yang, K., et al.: Dupl: Dual student with trustworthy progressive learning for robust weakly supervised semantic segmentation. In: CVPR 2024, pp. 3534–3543 (2024)
25. Xu, L., Bennamoun, M., et al.: Learning class-agnostic masks with cross-task refinement for weakly supervised semantic segmentation. Neural Computing and Applications **35**(27), 20189–20205 (2023)
26. Xu, L., Bennamoun, M., Boussaid, F., et al.: Mctformer+: Multi-class token transformer for weakly supervised semantic segmentation. IEEE TPAMI **46**(12), 8380–8395 (2024)
27. Yang, Z., Fu, K., Duan, M., et al.: Separate and conquer: Decoupling co-occurrence via decomposition and representation for weakly supervised semantic segmentation. In: CVPR 2024, pp. 3606–3615 (2024)
28. Yang, Y., Soatto, S.: FDA: Fourier domain adaptation for semantic segmentation. In: CVPR 2020, pp. 4085–4095 (2020)
29. Zhang, F., Gu, C., Zhang, C., Dai, Y.: Complementary patch for weakly supervised semantic segmentation. In: ICCV 2021, pp. 7242–7251 (2021)
30. Zhou, X., Tang, Z.: Enhanced pseudo-labels for end-to-end weakly supervised semantic segmentation with foundation models. Applied Sciences **15**(18), 10013 (2025)