Edge-semantic Synergy Fusion and Adaptive Noise-aware for Weakly Supervised Pathological Tissue Segmentation

Hualong Zhang[1⋆], Siyang Feng[1⋆], Zihan Huan[1], Huadeng Wang[1], Zhenbing Liu[1], Rushi Lan[1,2(✉)], and Xipeng Pan[1,3(✉)]

[1] School of Computer Science and Information Security, Guilin University of Electronic Technology, Guilin 541004, China
pxp201@guet.edu.cn

[2] International Joint Research Laboratory of Spatio-temporal Information and Intelligent Location Services, Guilin University of Electronic Technology, Guilin 541004, China
rslan@guet.edu.cn

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

Abstract. Existing studies on weakly supervised pathological tissue segmentation predominantly rely on class activation maps (CAMs) to generate pixel-level pseudo-masks from image-level labels. However, CAMs tend to emphasize only the most discriminative regions, resulting in boundary noise that undermines the quality of pseudo-masks and degrades segmentation performance. To address these challenges, we propose a novel weakly supervised pathological tissue segmentation framework: Edge-semantic Synergy Fusion and Adaptive Noise-aware (ESFAN) mechanism. In the classification phase, the Edge-semantic Synergy Fusion (ESF) improves the quality of pseudo-masks by incorporating four synergistic components. The hybrid edge-aware transformer refines boundaries, while the pyramid context integrator captures multi-scale context. The context channel amplifier fine-tunes semantic features, and the adaptive fusion gating balances feature map contributions using learnable spatial weights. In the segmentation phase, we propose an Adaptive Noise-aware Mechanism (ANM) that incorporates adaptive weighted cross-entropy, uncertainty regularization, and spatial smoothing constraints to mitigate noise in pseudo-masks and enhance segmentation robustness. Extensive experiments on the LUAD-HistoSeg and BCSS datasets demonstrate that ESFAN significantly outperforms state-of-the-art methods. The code is available at: `https://github.com/Sameer-815/ESFAN`.

Keywords: Pathological tissue segmentation· Weakly supervised learning· Edge-semantic synergy fusion · Adaptive noise-aware

⋆ Hualong Zhang and Siyang Feng contributed equally to this work.

1 Introduction

Accurate pathological tissue segmentation is crucial for quantitative diagnosis and personalized treatment planning, particularly in characterizing tumor microenvironments. For example, spatial patterns of tumor-infiltrating lymphocytes (TILs), tumor-stroma ratio, and necrotic area proportion serve as important prognostic biomarkers, providing valuable insights into tumor progression and patient outcomes [9,11,3]. However, the scarcity of pixel-level annotated datasets poses significant challenges for current deep learning approaches, as the annotation of pathological images is prohibitively time-consuming and expensive, further complicated by inter-expert variability [7]. These limitations have propelled weakly supervised learning as a research hotspot, demonstrating remarkable potential to achieve competitive segmentation performance using cost-efficient image-level annotations while substantially reducing annotation burdens [2,5,4].

Weakly supervised semantic segmentation (WSSS) primarily relies on class activation maps (CAMs) [18] to generate coarse pseudo-masks. However, CAMs tend to highlight only the most discriminative regions while overlooking target boundaries. Existing enhancement strategies have attempted to mitigate these limitations. For instance, MLPS [7] suppresses salient regions through masking but neglects multi-scale morphological information. TPRO [15] provides text-based guidance yet remains insensitive to fine-grained edge details. Similarly, ARML [6] enhances local responses but struggles with boundary delineation. To address these challenges, we propose the ESF, which adaptively integrates high-frequency edge features with semantic representations through dynamic gating. Unlike conventional edge-aware methods that rely on static filters [13], ESF leverages multi-scale contextual cues and hybrid edge detection to achieve precise boundary localization. By incorporating channel-wise semantic attention, ESF effectively balances global context and fine-grained boundary details, improving CAMs completeness while enhancing segmentation accuracy.

The inherent noise in classification-generated pseudo-masks poses a critical challenge for model correction during segmentation. Existing methods, such as CPAL [10] or OEEM [8], struggle to suppress label noise and stabilize training. Our ANM addresses this by synergizing adaptive class cross-entropy, uncertainty constraints, and boundary-aware smoothing. Through confidence-guided dynamic optimization, ANM robustly mitigates pseudo-label inaccuracies.

The main contributions of our work are as follows:

- We identify that the primary challenges in weakly supervised pathological tissue segmentation arise from imprecise boundary localization and inherent noise in pseudo-masks generated by CAMs, which are often overlooked by existing WSSS methods.
- We propose the Edge-semantic Synergy Fusion (ESF) to enhance CAMs boundary precision and the Adaptive Noise-aware Mechanism (ANM) to suppress pseudo-mask noise, significantly improving segmentation accuracy.

- Our method achieves 79.29% mIoU on the LUAD-HistoSeg dataset and 71.41% on the BCSS dataset, setting state-of-the-art performance with only image-level labels, which substantially enhances the clinical applicability.

2 Method

2.1 Overall Framework

The framework of our proposed ESFAN is shown in Fig. 1. In the first stage, image-level labels are fed into the classification network, where the ESF integrates edge enhancement and semantic attention before the final layer. The output features from the last three layers are used to compute the classification loss $\mathcal{L}_{cls}$, optimizing the network's predictions. A fused class activation map is then generated by combining the maps from the last three layers and applying an argmax operation to create pseudo-masks. To mitigate noise in the pseudo-masks, we introduce an ANM in the segmentation stage, significantly improving segmentation accuracy.

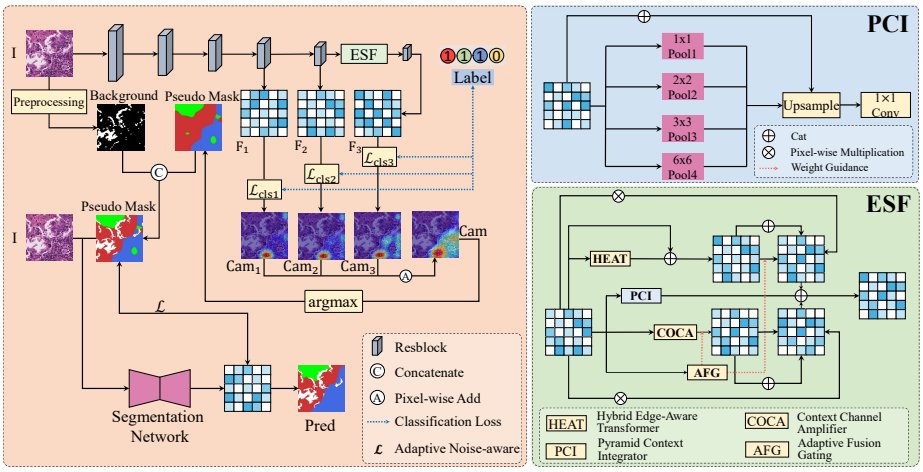

Fig. 1: Overall structure of our propoesd method. The classification network generates the pseudo-masks based on ESF, and the segmentation network optimizes the segmentation results based on ANM.

2.2 Edge-semantic Synergy Fusion

The CAMs generated during the classification stage often exhibit imprecise boundaries, resulting in significant loss of semantic information. To address this issue, we propose an ESF strategy during the classification stage. The framework

integrates four components: (1) a Hybrid Edge-Aware Transformer for boundary detection, (2) a Pyramid Context Integrator for multi-scale feature fusion, (3) a Context Channel Amplifier for semantic enhancement, and (4) an Adaptive Fusion Gating mechanism to balance contributions. Given an input feature map $X \in \mathbb{R}^{C\times H\times W}$, the network achieves pixel-level prediction through multiple feature interaction, with core operations formulated as:

$$\hat{X} = \mathcal{F}_{\text{fusion}}\left(\Psi_{\text{HEAT}} \otimes X + \Psi_{\text{COCA}} \otimes X + \Phi_{\text{PCI}}(X)\right), \tag{1}$$

where the feature fusion convolution $\mathcal{F}_{\text{fusion}}$ employs 3×3 kernels for cross-channel information integration, and $\otimes$ denotes element-wise multiplication. Ψ_{HEAT} represents the feature map generated by Hybrid Edge-Aware Transformer, Ψ_{COCA} represents the feature map generated by Context Channel Amplifier, and Ψ_{PCI} represents the feature map generated by Pyramid Context Integrator.

For Hybrid Edge-Aware Transformer, high-frequency features are extracted using a fixed Laplacian operator:

$$\Psi_{\text{edge}}(X) = \sum_{i=-1}^{1}\sum_{j=-1}^{1} W_{\text{lap}}^{(i+1,j+1)} * X_{p+i,q+j}, \quad W_{\text{lap}} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \tag{2}$$

where p and q represent the current pixel. The original features concatenated with high-frequency components are processed through a 3×3 convolutional layer to generate edge attention maps: $\Psi_{\text{Edge}} = \sigma\left(\mathcal{F}_{\text{Edge}}([X, \Psi_{\text{Edge}}(X)])\right)$, where σ represents the sigmoid activation. For channel-wise semantic enhancement, global average pooling establishes channel dependencies:

$$z_c = \frac{1}{H \times W}\sum_{i,j} X_c(i,j), \Psi_{\text{semantic}} = \sigma\left(W_2\delta\left(W_1 z_c\right)\right), \tag{3}$$

with a reduction ratio $r = 16$. The bottleneck structure consists of $W_1 \in \mathbb{R}^{C/r\times C}$ and $W_2 \in \mathbb{R}^{C\times C/r}$, where δ denotes ReLU activation. A Pyramid Context Integrator captures morphological diversity in pathological tissues:

$$\Phi_{\text{PCI}} = \mathcal{F}_{\text{spp}}\left(\text{Concat}\left[X, \bigcup_{k\in\{1,2,3,6\}} \text{Upsample}\left(\text{AdaptiveGAP}_k(X)\right)\right]\right), \tag{4}$$

where $\mathcal{F}_{\text{Spp}}$ includes 1×1 convolution and batch normalization.

An adaptive fusion gating mechanism $m = \sigma(\mathcal{F}_{\text{gate}}(X))$ balances contributions from edge and semantic features based on a Hybrid Edge-Aware Transformer and a Context Channel Amplifier. $\mathcal{F}_{\text{gate}}$ stands for global averaging pooling and the convolution of 1×1.

$$\Psi_{\text{HEAT}} = m \otimes \Psi_{\text{edge}} + \Psi_{\text{edge}}, \quad \Psi_{\text{COCA}} = (1 - m) \otimes \Psi_{\text{semantic}} + \Psi_{\text{semantic}}. \tag{5}$$

2.3 Adaptive Noise-aware Mechanism

The segmentation task faces inherent challenges such as feature ambiguity, noisy pseudo-labels. Traditional denoising loss functions, with fixed weights, lack adaptability across training phases. Moreover, existing methods often trade boundary precision for noise reduction. To address these issues, we propose an ANM consisting of: (1) Adaptive cross-entropy loss to emphasize reliable regions, (2) Uncertainty-aware regularization to reduce ambiguous predictions, and (3) Spatial smoothing to enhance boundary coherence. A dynamic weighting mechanism adjusts component contributions based on training progress, ensuring both noise suppression and boundary refinement. This unified approach significantly improves segmentation accuracy and robustness without the need for manual parameter tuning.

To enhance the model's robustness against low-confidence pixels, we propose an adaptive weighted cross-entropy loss. Given the model's raw output $p = \text{softmax}(f)$ (before softmax transformation) and the target label y, we calculate the maximum confidence for each pixel $\text{conf}(x) = \max(p(x))$, representing the highest probability among predicted classes. A confidence threshold τ is set to adjust the weight of each pixel in the loss function. Pixels with confidence higher than τ are assigned higher weights, while those below the threshold are weighted less. To ensure the network learns sufficient features in the early stage, τ is set as an adaptive transformation ranging from 0.2 to 0.5. This design allows the network to focus on feature learning initially, while imposing a greater penalty on misclassifications in the later stage. The formula is as follows:

$$\text{weight}(x) = \begin{cases} 1 & \text{if } \text{conf}(x) > \tau, \\ 0 & \text{if } \text{conf}(x) \leq \tau. \end{cases} \tag{6}$$

The weighted cross-entropy loss is then computed as:

$$\mathcal{L}_{ce} = -\frac{1}{N} \sum_{x} \text{weight}(x) \cdot \left(y(x) \log(p(x)) + (1 - y(x)) \log(1 - p(x)) \right), \tag{7}$$

where N is the total number of pixels, $y(x)$ is the true label, and $p(x)$ is the predicted probability of pixel x.

To account for model uncertainty, especially in noisy or low-confidence areas, we introduce an uncertainty loss that measures pixel-wise confidence. The uncertainty is derived from the maximum confidence of each pixel, and the loss is defined as:

$$\mathcal{L}_{\text{u}} = -\frac{1}{N} \sum_{x} \left((1 - \text{conf}(x)) \log(1 - \text{conf}(x)) \right). \tag{8}$$

This loss prioritizes more accurate predictions for confident regions while suppressing predictions in uncertain areas, thereby enhancing model reliability.

To improve spatial consistency and reduce sensitivity to adjacent pixel predictions, we design a smoothing loss based on prediction probability gradients.

Table 1: Comparison with state-of-the-art weakly supervised methods. The results are reported in mean ± std. * means statistical significant with other methods($p < 0.05$ under two-tailed T-test).

LUAD-HistoSeg							
Method	Backbone	TE	NEC	LYM	TAS	mIoU	Dice
OEEM [8]	ResNet38	76.96±0.27*	74.71±0.37*	72.30±2.20*	71.30±1.52*	73.92±0.44*	84.99±0.30*
MLPS [7]	ResNet101	77.68±0.51*	76.95±0.58*	72.40±0.84*	71.81±0.77*	74.71±0.41*	85.50±0.30*
HAMIL [17]	ResNet50	76.63±0.37*	77.43±0.51*	74.59±0.22*	70.97±0.95*	74.91±0.60*	85.73±0.25*
TPRO [15]	MixTransformer	75.85±0.47*	81.94±0.95*	74.66±1.25*	71.27±0.46*	75.90±0.44*	86.16±0.28*
ARML [6]	ResNest101	78.54±0.28*	80.34±0.27*	75.29±0.45*	74.29±0.48*	77.11±0.33*	87.01±0.22*
ESFAN(phase1)	ResNet38	77.61±0.27	81.41±0.38	78.07±0.17	73.13±0.13	77.56±0.24	87.33±0.14
ESFAN(phase2)	ResNest200	**79.67±0.16**	**84.47±0.51**	**78.27±0.25**	**74.74±0.11**	**79.29±0.35**	**88.41±0.23**
BCSS							
Method	Backbone	TUM	STR	LYM	NEC	mIoU	Dice
OEEM [8]	ResNet38	79.11±0.37*	72.88±0.79*	54.34±1.61*	63.07±2.43*	67.35±0.58*	80.11±0.42*
MLPS [7]	ResNet101	78.53±0.60*	71.74±0.69*	60.71±0.52*	60.51±1.11*	67.87±0.45*	80.62±0.31*
HAMIL [17]	ResNet50	79.11±0.34*	72.04±0.50*	54.39±1.08*	60.21±0.47*	66.44±0.38*	78.92±0.37*
TPRO [15]	MixTransformer	**80.10±0.59**	73.34±0.21*	56.26±1.06*	64.26±1.39*	68.49±0.24*	80.95±0.19*
ARML [6]	ResNest101	79.20±0.30*	73.83±0.17*	60.25±0.14*	**68.96±0.44**	70.56±0.19*	82.48±0.10*
ESFAN(phase1)	ResNet38	72.84±0.18	61.39±0.11	53.10±0.22	56.08±0.65	60.85±0.18	75.40±0.17
ESFAN(phase2)	ResNest200	80.07±0.11	**74.11±0.17**	**62.71±0.30**	68.75±0.37	**71.41±0.26**	**83.16±0.15**

This loss minimizes the gradient differences in horizontal and vertical directions, formulated as:

$$\mathcal{L}_{\mathrm{s}} = \frac{1}{N} \sum_{x} \left(|p(x) - p(x+1)| + |p(x) - p(x+W)| \right), \tag{9}$$

where W represents the width of the image, $p(x)$ is the predicted probability of pixel x, and $|p(x) - p(x+1)|$ and $|p(x) - p(x+W)|$ denote the absolute differences of first-order spatial gradients in the horizontal and vertical directions, respectively.

The final loss combines these three components through a weighted sum:

$$\mathcal{L} = \alpha \cdot \mathcal{L}_{ce} + \beta \cdot \mathcal{L}_{\mathrm{u}} + \gamma \cdot \mathcal{L}_{\mathrm{s}}. \tag{10}$$

To ensure the weights α, β, and γ adapt dynamically throughout training, an Adaptive Weight Scheduler is implemented:

$$w(i) = w_{\mathrm{start}} + (w_{\mathrm{end}} - w_{\mathrm{start}}) \cdot \frac{i}{t}, \tag{11}$$

where w stands α, β, and γ initial weights are set to 0.4, 0.4, and 0.2, respectively. And the target weights are set to 0.8, 0.1, and 0.1, respectively. i represents the current epoch, and t is the total number of training epochs.

3 Experiments

3.1 Datasets

LUAD-HistoSeg: This dataset comprises whole slide images (WSIs) of lung adenocarcinoma from 54 patients, classified into four tissue types: tumor epithe-

lium (TE), tumor-associated stroma (TAS), necrosis (NEC), and lymphocyte (LYM). It includes 16,678 patches for training, 300 patches with pixel-level annotations for validation, and 307 pixel-level annotated patches for testing [7]. **BCSS**: Derived from breast cancer WSIs of 151 patients, this dataset features four distinct categories: tumor (TUM), stroma (STR), lymphocyte infiltration (LYM), and necrosis (NEC). It provides 23,422 patches for training, 3,418 pixel-level annotated patches for validation, and 4,986 pixel-level annotated patches for testing [1].

Table 2: Performance comparisons of each component of our method on BCSS dataset.

Baseline	ESF	ANM	TUM	STR	LYM	NEC	mIoU	Dice
✓			77.93±0.24	71.73±0.65	54.62±1.41	65.49±1.22	67.44±0.45	80.23±0.36
✓	✓		78.93±0.37	73.43±0.11	61.32±1.36	67.35±0.36	70.26±0.34	82.35±0.25
✓		✓	79.12±0.26	73.87±0.27	62.41±0.14	66.57±0.44	70.49±0.12	82.52±0.07
✓	✓	✓	**80.07±0.11**	**74.11±0.17**	**62.71±0.30**	**68.75±0.37**	**71.41±0.26**	**83.16±0.15**

3.2 Experimental Details

All experiments were conducted using PyTorch 1.10.2 on an NVIDIA RTX 3080Ti GPU. In the pseudo-masks generation phase, ResNet38d [12] is used as the backbone with a learning rate of 1×10^{-2}, following a polynomial decay over 20 epochs. For segmentation, the PSPNet model [16] with a pre-trained ResNest-200 [14] encoder is employed, utilizing the SGD optimizer with a learning rate of 1×10^{-2}, momentum of 0.9, and weight decay of 5×10^{-4}, also for 20 epochs. During the inference stage, performance is evaluated using three metrics: per-class Intersection-over-Union (IoU), mean IoU (mIoU), and mean Dice coefficient.

3.3 Comparison Study

We compared our method with five other state-of-the-art weakly supervised methods under the same experimental conditions. As shown in Table 1, our method outperforms existing approaches by 2.18% mIoU on LUAD-HistoSeg and 0.85% on BCSS when using ResNeSt-200 as the backbone ($p < 0.05$ under two-tailed T-test).

Notably, our method achieves superior performance across all categories on LUAD-HistoSeg, surpassing the results of existing methods. This improvement underscores the synergistic effect of the two modules: ESF refines the semantic features and boundaries, while ANM mitigates noise and improves the robustness of the masks. This combined effect leads to significant improvements in overall segmentation accuracy, validating the complementary nature of these components in our framework. A visual comparison is shown in Fig. 2.

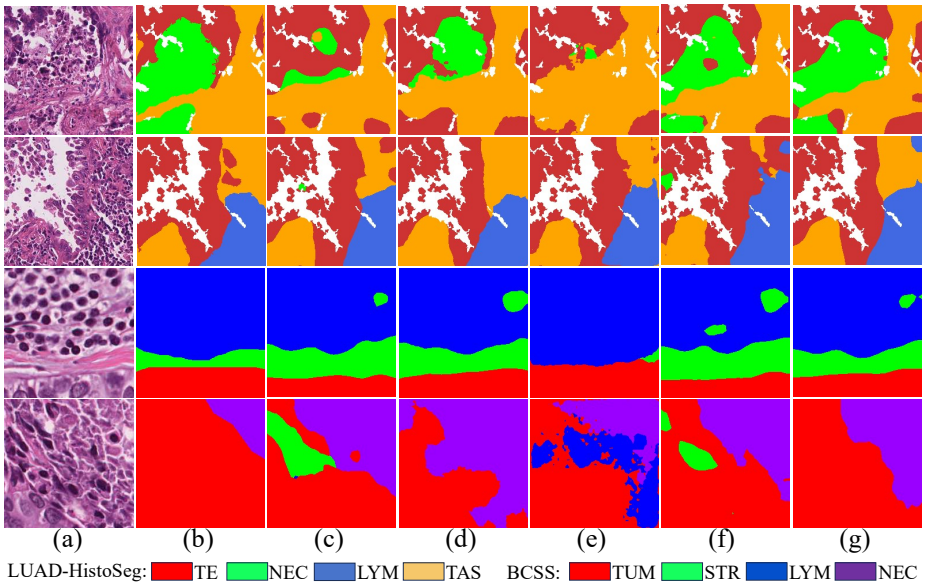

Fig. 2: Visualization of tissue segmentation on LUAD-HistoSeg (first two rows) and BCSS (last two rows) datasets. (a) Test image. (b) Ground truth. (c) OEEM. (d) HAMIL. (e) TPRO. (f) ARML. (g) Our method.

Table 3: Performance comparisons of each component of ESF on BCSS dataset.

PCI	HEAT	COCA	AFG	mIoU
✓				70.40±0.38
✓	✓			70.44±0.32
✓	✓	✓		71.01±0.17
✓	✓	✓	✓	**71.41±0.26**

Table 4: Performance comparisons of each component of ANM on BCSS dataset.

L_{CE}	L_u	L_s	mIoU
✓			69.20±0.49
✓	✓		69.39±0.27
✓		✓	69.77±0.35
✓	✓	✓	**71.41±0.26**

3.4 Ablation Study

To validate the effectiveness of the proposed method, we conducted extensive ablation experiments on the BCSS dataset. Specifically, the ESF and ANM modules were incorporated into the baseline model separately. As shown in Table 2, integrating the ESF and ANM modules improved the mIoU by 2.82% and 3.05%, respectively. When both modules were combined, the mIoU reached the current highest value of 71.41%. These results demonstrate that both components contribute significantly to segmentation performance, and their synergy further enhances the overall efficacy of our approach.

At the same time, we conducted ablation experiments on ESF and ANM components.The results in Table 3 and Table 4 indicate that each component contributes positively to the final segmentation performance.

4 Conclusion

Weakly supervised histological tissue segmentation faces two critical challenges: the generation of low-quality pseudo-masks and the impact of noisy labels during segmentation training. To address these issues, we propose the Edge-semantic Synergy Fusion method, which enhances both tissue segmentation accuracy and boundary precision by generating high-quality pseudo-masks. Additionally, we introduce an Adaptive Noise-aware Mechanism to mitigate the effects of noisy labels, effectively suppressing noise while preserving essential information from other regions. Extensive experiments on two publicly available datasets demonstrate the superior performance of our method, achieving state-of-the-art results and contributing significantly to the field of computational pathology.

Acknowledgments. This work was supported in part by Guangxi Natural Science Foundation (Nos. 2024GXNSFFA010014 and AB25069496), the National Natural Science Foundation of China (Nos. 82360356, 62172120, 82272075), and Innovation Project of Guangxi Graduate Education (No. YCSW2025362).

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

References

1. 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)
2. Cai, W., Li, Y., Chen, Y., Lin, J., Huang, Z., Gao, P., Gadekallu, T.R., Wang, W., Gao, Y.: Enhancing weakly supervised semantic segmentation with multi-label contrastive learning and llm features guidance. IEEE Journal of Biomedical and Health Informatics (2024)
3. El Bairi, K., Haynes, H., Blackley, E., Fineberg, S., Shear, J., Turner, S., et al.: The tale of tils in breast cancer: A report from the international immuno-oncology biomarker working group. npj breast cancer 2021; 7 (1): 150
4. Fang, Z., Chen, Y., Wang, Y., Wang, Z., Ji, X., Zhang, Y.: Weakly-supervised semantic segmentation for histopathology images based on dataset synthesis and feature consistency constraint. In: Proceedings of the AAAI conference on artificial intelligence. vol. 37, pp. 606–613 (2023)
5. Fang, Z., Wang, Y., Xie, P., Wang, Z., Zhang, Y.: Hisynseg: Weakly-supervised histopathological image segmentation via image-mixing synthesis and consistency regularization. IEEE Transactions on Medical Imaging (2024)
6. 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)
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. 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)
9. Song, A.H., Jaume, G., Williamson, D.F., Lu, M.Y., Vaidya, A., Miller, T.R., Mahmood, F.: Artificial intelligence for digital and computational pathology. Nature Reviews Bioengineering **1**(12), 930–949 (2023)
10. Tang, F., Xu, Z., Qu, Z., Feng, W., Jiang, X., Ge, Z.: Hunting attributes: Context prototype-aware learning for weakly supervised semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3324–3334 (2024)
11. Wang, Y., Lin, H., Yao, N., Chen, X., Qiu, B., Cui, Y., Liu, Y., Li, B., Han, C., Li, Z., et al.: Computerized tertiary lymphoid structures density on h&e-images is a prognostic biomarker in resectable lung adenocarcinoma. Iscience **26**(9) (2023)
12. Wu, Z., Shen, C., Van Den Hengel, A.: Wider or deeper: Revisiting the resnet model for visual recognition. Pattern recognition **90**, 119–133 (2019)
13. Zhang, D., Zhao, J., Chen, J., Zhou, Y., Shi, B., Yao, R.: Edge-aware and spectral–spatial information aggregation network for multispectral image semantic segmentation. Engineering Applications of Artificial Intelligence **114**, 105070 (2022)
14. 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)
15. 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)
16. Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J.: Pyramid scene parsing network. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2881–2890 (2017)
17. 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)
18. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A.: Learning deep features for discriminative localization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2921–2929 (2016)