Dual-Branch Dynamic Coupling Weakly Supervised Learning for Class-Incremental Histopathological Region Segmentation

Xiaoyan Hong[1], Jiansong Fan[1], Zhaohong Deng[1], and Xiang Pan[1](✉)

School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi, 214122, China
xiangpan@jiangnan.edu.cn

Abstract. Histopathological region segmentation faces two main challenges: catastrophic forgetting and the high cost of pixel-level annotations. Recent studies have focused on incremental learning of new categories using low-cost image-level labels. However, the limitations of multiple instance learning (MIL) in modeling instance relationships hinder further improvement in segmentation performance. To address these challenges, we propose the Dual-branch Dynamic Coupling (DDCWISS) network for weakly supervised class-incremental learning in histopathological region segmentation. Our architecture overcomes the limitations of isolated local feature computation in traditional MIL by enabling complementary feature extraction through parallel local representation and global modeling branches. Additionally, we propose a learnable coupling module to ensure effective multi-scale feature fusion, while the dual-path supervision mechanism simultaneously enhances segmentation accuracy. Experiments on the CPATH dataset demonstrate that our method significantly reduces reliance on costly pixel-level annotations for histopathological region segmentation, while effectively alleviating the catastrophic forgetting problem during incremental learning. These results highlight the potential of DDCWISS as a scalable, weakly supervised Class-Incremental paradigm for medical image analysis. The source code is publicly available at: https://github.com/XiaoyanHong24/DDCWISS.

Keywords: Incremental Learning · Histopathological region segmentation · Weakly supervised learning.

1 Introduction

Histopathological images are critical for cancer diagnosis. While intelligent models trained on public datasets have made notable progress, current methods face two fundamental limitations[18]. First, they lack mechanisms for continual learning. As cancer progresses, diagnosis requires finer-grained distinctions between tissue types—such as lymphatic infiltration or vascular invasion—that existing fixed-category models struggle to identify. Second, re-annotating entire slides and retraining from scratch is resource-intensive. Pixel-level labeling is especially costly given the complexity and variability of cellular structures. Moreover, focusing solely on new classes often leads to forgetting of previously learned categories, undermining clinical applicability. A joint incremental learning strategy is thus essential to balance knowledge retention and acquisition[15]. To address these challenges, two primary approaches have emerged. Incremental learning extends the model's capacity to recognize new classes while preserving knowledge of old ones, thereby reducing catastrophic forgetting. In parallel, weakly supervised methods—especially multiple instance learning (MIL)—aim to minimize annotation cost by using image-level labels, which are cheaper and more scalable[10]. In MIL-based frameworks, images are divided into patches (bags), with pixels treated as instances. The model learns how pixel-level patterns respond to patch-level labels, effectively reframing weakly supervised segmentation as an MIL task in settings without pixel-wise annotations.

However, traditional MIL assumes that instances (i.e., patches) are independent, introducing two major limitations. First, it ignores contextual relationships between patches. In histopathology, tumor and stroma often share similar textures, and without semantic context, the model can confuse ambiguous stroma with tumor, leading to false positives. Second, convolutional networks' limited receptive fields hinder the modeling of long-range spatial dependencies—key for understanding complex tissue architecture—thus constraining the model's global representational capacity. Recently, Wilson introduced the task of Weakly Supervised Incremental Learning for Semantic Segmentation (WISS) [2], which aims to update models progressively using only image-level labels to incorporate new classes. While centralized region-based features often suffice for semantic segmentation in natural images—where object boundaries are clear—this assumption does not hold in histopathology. In histopathological images, diagnosis depends on contextual and spatial relationships across multiple regions. For instance, tumor tissues are often discontinuous and require joint interpretation of surrounding stroma, necrosis, and lymphatic areas. In incremental learning, relying solely on localized predictions overlooks such dependencies and disrupts the feature space, worsening catastrophic forgetting as new information overwrites prior knowledge. Therefore, conventional centralized prediction strategies are ill-suited for histopathological segmentation. In this work, we emphasize the need to address both incremental model updating and annotation cost reduction in this challenging domain.

Based on the above, we propose the Dual-Branch Dynamic Coupling Network (DDCWISS), which combines local feature perception with global context modeling for accurate histopathological region segmentation. The architecture includes three core components: the Dual-Branch Feature Extraction (DBE) module, the Dynamic Feature Coupler (DFC) module, and a dual-path supervision mechanism. The DBE module addresses challenges such as catastrophic forgetting and multi-scale variation by using two parallel branches. The local representation branch improves key region recognition via an enhanced MIL mechanism[9], while the global modeling branch employs a vision Transformer to capture cross-region dependencies through adaptive weight allocation. To bridge the semantic gap caused by differences in receptive field and feature granularity, we introduce the DFC module. It uses 1×1 convolutions and learnable parameters to adaptively preserve relevant features, enabling efficient fusion of local and global information. To overcome limited pixel-level supervision, we integrate image-level labels with dynamically generated pseudo-labels in a dual-path supervision scheme. In the pseudo-labeling path, a distillation module converts image-level labels into supervisory signals to guide classification. In the optimization path, the segmentation backbone is refined using pixel-level consistency loss and pseudo-supervision, improving both feature quality and segmentation accuracy.

2 Method

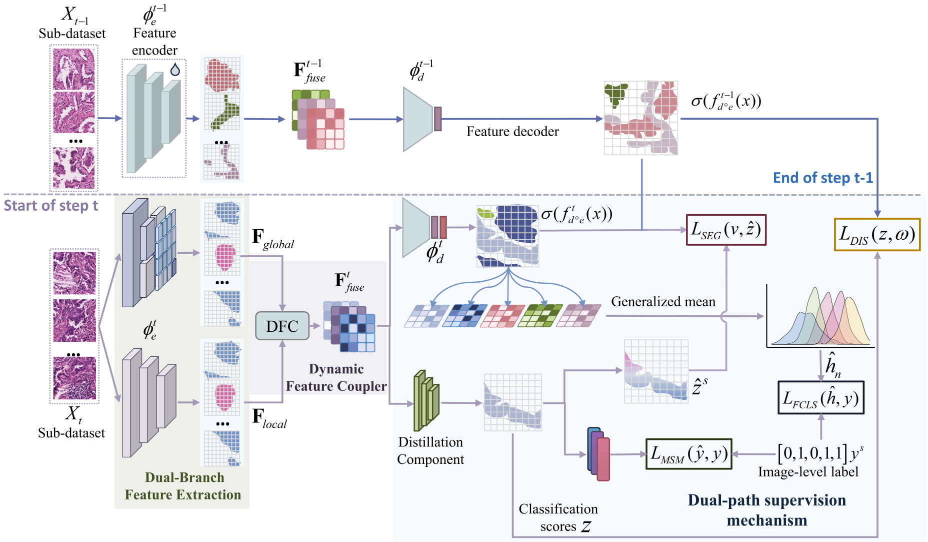

Fig. 1. Overview of the proposed Dual-branch Dynamic Coupling Weakly Supervised Continual Learning.

Our proposed model architecture, as depicted in Fig. 1, operates on input images from the space X. Each image $x \in X$ consists of a pixel set P, with cardinality $|P| = H \times W$, corresponding to the image's height and width dimensions. The label space Y expands through a multi-phase training protocol, typical of semantic segmentation in class-incremental learning. Specifically, at training stage t, the system incorporates novel classes S_t along with pixel-level annotations, progressively constructing an updated label set $Y_t = Y_{t-1} \cup S_t$.

2.1 Dual-Branch Feature Extraction with Dynamic Coupling

In contrast to the traditional class-incremental learning setup, the recently proposed WISS introduces a novel approach. In this setting, pixel-level annotations are only available during the initial training phase[3]. For subsequent steps, the training datasets consist solely of image-level annotations for the new classes, and previously used training samples are no longer accessible. The primary goal of this framework is to update and refine the model to segment newly introduced classes while preserving the knowledge of previously learned classes, thereby mitigating catastrophic forgetting.

Our mapping is realized by a model $F = \phi_d \circ \phi_e : X \mapsto \mathbf{R}^{N \times |Y|}$, e and d denote the encoder and decoder of the segmentation network, respectively. The DBE module proposed in this paper consists of two components: local representation and global modeling. The local representation is denoted as $\mathbf{F}_{local}$, while the global modeling is denoted as $\mathbf{F}_{global}$. The local representation extends MIL. By utilizing image-level annotations, we transform the multi-instance learning problem, typically associated with bag-level labels, into a histopathological segmentation problem. By treating each pixel as an instance, our model effectively learns both the spatial and semantic information of the histopathological images, leading to more precise pixel-level predictions[11].

In the global modeling module, we employ a sliding window technique to extract global features from histopathological images. This module compensates for the missing global context in the local feature representations by aggregating information from broader regions of the image. The sliding window extracts overlapping local patches, which are subsequently used to capture cross-regional dependencies and long-range semantic relationships. The dual-branch features are fused through our DFC module, as shown below:

$$\mathbf{F}_{fused} = \varPhi(\alpha \cdot \mathcal{G}(\mathbf{F}_{local}), \beta \cdot \mathcal{H}(\mathbf{F}_{global})), \tag{1}$$

where $\alpha \in [0, 1]$ and $\beta \in [0, 1]$ represent the adaptive coupling coefficients, $\mathcal{G}(\cdot)$ and $\mathcal{H}(\cdot)$ denote the size alignment (upsampling) operations performed on the dual-branch features before fusion, and $\varPhi$ represents the DFC module. The fusion process ensures alignment between the local and global representations while maintaining the consistency of the feature space.

2.2 Dual-path supervision mechanism

We introduce a distillation module trained using image-level annotations to generate pseudo-supervision signals for the segmentation model. This distiller utilizes the fused feature $\mathbf{F}_{fused}$ to predict class scores. To learn from image-level annotations, it is necessary to aggregate the pixel-level classification scores z. A commonly used approach for this is global average pooling, which assigns weights to each pixel based on its relevance to the target class[1], as expressed by the following formula:

$$\hat{y} = \sum_{i \in P} \frac{\phi(z_i) z_i}{\epsilon + u_i} + \mathcal{R}(u_i, \gamma, \lambda) \tag{2}$$

where the regularization term $\mathcal{R}(u_i, \gamma, \lambda)$ is defined as:

$$\mathcal{R}(u_i, \gamma, \lambda) = \left(1 - \frac{u_i}{|P|}\right)^{\gamma} \log\left(\lambda + \frac{u_i}{|P|}\right), \tag{3}$$

where z_i represents the classification score for pixel i, and the weight of each pixel $\phi(z_i)$ is calculated by normalizing the classification scores using the softmax operation ϕ. The relevance score u_i for pixel i is computed as $u_i = \sum_{i \in P} \phi(z_i)$, where u_i is a mask indicating the relevance of the pixel. ϵ is a small constant, γ and λ are hyper-parameters.

$$L_{MSM}(\hat{y}, y) = -\sum_{s \in S_t} y^s log(\hat{y}^s) + (1 - y^s) log(1 - \hat{y}^s) \tag{4}$$

where y^s denotes the image-level label. Considering the nature of multi-label segmentation tasks, we incorporate image-level labels to guide the supervision of the model's outputs. Specifically, for each input image $x_n \in X_t$, we compute the generalized mean of the decoder's output, which aggregates the pixel-level predictions across the entire image. This generalized mean is calculated as follows:

$$\hat{H}_n = \left(\frac{1}{|x_n|} \sum_{i \in P} z_i^r\right)^{\frac{1}{r}}, \tag{5}$$

where $|x_n|$ is the total number of pixels in the image x_n, and r is a hyperparameter that controls the influence of individual pixel probabilities in the computation. When $r = 1$, it equals the average, treating all pixels equally; when $r > 1$, high-confidence pixels have more influence while low-confidence ones are down-weighted. As r approaches infinity, only the most confident pixel dominates, improving the model's focus on salient regions under image-level supervision. The loss function for the output and image-level labels is then formulated using the L_{FCLS} loss, which can be written as:

$$L_{\text{FCLS}}(\hat{h}, y) = -\sum_{n \in X} \sum_{s \in S} \left[y^s \log(\hat{h}_n^s) + (1 - y^s) \log(1 - \hat{h}_n^s) \right], \tag{6}$$

where $\hat{h}_n^s$ is the image-level class probability prediction obtained from equation (5), and y^s is the corresponding ground truth label for class s. In the class-incremental learning framework, the WISS system utilizes a progressive knowledge integration mechanism[17]. This framework assumes that new categories are only accompanied by image-level label data (y), and under this constraint, the distillation component trains the model by optimizing a multi-class margin loss function. Since global annotations only indicate the presence of categories and lack pixel-level localization details, the proposed model, DDCWISS, leverages spatial prior knowledge embedded in historical segmentation models.

Specifically, the background confidence map is generated by the model from the previous stage $t - 1$, containing predictions only for the old classes. After new classes are introduced in the current stage t, the old classes are treated as relative background. This map is used during the boundary refinement phase to provide saliency-guided pseudo-labels for the current model, thereby improving the accuracy of boundary recognition between old and new classes. Our dual-guidance mechanism effectively directs the model's attention to potential regions for new categories. To accomplish this, DDCWISS incorporates the segmentation output from the previous iteration, applies the sigmoid activation function, and uses the classification scores z as supervisory signals. This cross-iteration collaborative training mechanism ultimately leads to the following optimization objective:

$$L_{DIS}(z,u) = -\sum_{i \in P} \sum_{s \in Y^{t-1}} u_i^s \log(\sigma(z_i^s)) + (1-\omega_i^s)\log(1-\sigma(z_i^s)), \tag{7}$$

where $u_i^s = \sigma(f^{t-1}(x))$ is the output from the previous iteration, processed by the sigmoid function, indicating the confidence of the previous model's prediction for pixel p_i belonging to class s. z_i^s represents the classification score for pixel p_i and class s from the distillation component. Existing studies have shown that pseudo-supervision signals generated by global image classifiers contain significant noise, which can easily cause learning bias in the model. To address this, we have developed a dynamic label correction mechanism [13], redefining the pixel-level pseudo-supervision signal $\hat{z}$ as follows:

$$\hat{z}^s = \min\left(u^s, b^s\right) \cdot \mathbf{1}_{\{s=\mathbf{b}\}} + b^s \cdot \mathbf{1}_{\{s \in \mathcal{S}^t\}} + u^s \cdot \mathbf{1}_{\{s \notin \{\mathbf{b}\} \cup \mathcal{S}^t\}}, \tag{8}$$

where u^s from equation 7 represents the output of the model at stage $t-1$, and $\mathbf{b}^s = \alpha \cdot \mathbb{I}\left[s = \arg\max_{k \in \mathcal{Y}_t} m_i^k\right] + (1-\alpha)m^s$, with $m = \sigma(f^{t-1}(x))$ generating a one-hot distribution for each pixel; the pixel's highest score assigns it to a class while smoothing the pseudo-labels to reduce noise. The key issue lies in the minimal value fusion of the two probability distributions. This method effectively addresses the background distribution shift in scene transfer by constructing a dynamic confidence interval for the background class. Since the weakly supervised signal b^s does not conform to the normalization constraints of a probability distribution, conventional cross-entropy loss is not directly applicable. Therefore, we adopt a multi-label soft margin loss function to optimize the model:

$$L_{\mathrm{SEG}}(v,\hat{z}) = -\sum_{i \in P} \sum_{s \in \mathcal{S}^t} \left[\hat{z}_i^s \log(v_i^s) + (1-\hat{z}_i^s)\log\left(1-\sigma(v_i^s)\right)\right], \tag{9}$$

where P denotes the set of all pixels in the image, $\mathcal{S}^t$ is the set of classes in task-t, and $v = g^t(x)$ is the pixel-level probability output of the current model iteration.

3 Experiments

3.1 Datasets and Protocols

We evaluated DDCWISS on multiple datasets, including BCSS [14], LUAD-HistoSeg [6], WSSS4LUAD [7], and a combined histopathological slide dataset

(CPATH) assembled from three sources. Following the standard approach of Augmentor, we divide the WSI images into patches of size 512 × 512 for subsequent prediction tasks. The integrated dataset comprises 19,944 images, with 15,956 used for training, 3,988 for validation, and 3,988 for testing. Five tissue categories are annotated: 0 represents the background region of the pathological slide, 1 denotes tumor tissue, 2 denotes stromal tissue, 3 denotes normal tissue, 4 denotes necrotic tissue, and 5 denotes lymphatic tissue. Prior work [16] proposed two distinct incremental learning protocols: disjoint and overlapping. In the disjoint scenario, each training step contains only pixels from previously encountered classes and those from the current phase. In contrast, the overlapping protocol includes all images at every training step, with pixels potentially belonging to any class. As a result, the overlapping protocol is both more realistic and challenging. In our experiments, we applied these protocols on the CPATH dataset, referred to as cpath 3-2, where three base classes are learned in the first phase, and two new classes are introduced in the subsequent phase.

3.2 Implementation Details

Our framework was implemented in PyTorch and trained on an A6000 GPU. We employed Deeplab V3[4] as the decoder backbone, while the encoder was built on ResNet-101[8] and the Swin Transformer[12], both pre-trained on ImageNet[5]. The Adam optimizer was used with a learning rate of 0.01, momentum parameters of 0.9 and 0.999 for the first and second moment estimates, respectively, and a weight decay of 1×10^{-8}.

Table 1. Quantitative Comparison of different methods. Sup indicates supervision type: P = pixel-level, I = image-level. Joint denotes non-incremental learning, where all classes are learned at once (upper bound). FT (Fine-Tuning) means training new tasks without incremental strategies, prone to catastrophic forgetting (lower bound). Evaluation metric: mean Intersection over Union (mIoU), averaged over all classes within each task.

Method	Sup	Overlap			Disjoint		
		1-3	4-5	All	1-3	4-5	All
Joint	P	65.88	56.21	60.06	65.07	60.03	61.13
FT	P	30.31	5.15	11.43	31.76	9.82	12.83
WILSON[3]	I	53.76	27.98	41.22	**56.10**	32.44	44.20
DDCWISS(Ours)	I	**60.50**	**43.25**	**43.06**	50.66	**44.47**	**46.48**

3.3 Comparison With Existing Methods

Since WISS is a novel framework introduced by WILSON, we compare our approach with current representative methods in supervised incremental learning and weakly supervised semantic segmentation. As shown in Table 1, our

DDCWISS outperforms methods based on image-level labels, achieving optimal performance, and even surpasses methods using pixel-level labels in some cases. Specifically, compared to pixel-level label methods, our overall performance on the Disjoint protocol reached 46.48%. On the Overlapping protocol, we achieved 43.06% overall performance. Compared to image-level label methods, our DDCWISS achieved the best results across all protocols, demonstrating an overall improvement in performance.

Under the overlap strategy, the WILSON method achieves mIoU scores of 81.28%, 63.51%, and 38.34% for the three classes in Task 1. However, after introducing new classes in Task 2, its performance declines to 75.29%, 51.38%, 30.48%, 23.64%, and 32.33%. In contrast, our method achieves mIoUs of 80.35%, 60.83%, and 31.59% in Task 1, and 76.49%, 50.78%, 20.09%, 16.03%, and 47.78% in Task 2. Compared to WILSON, our method exhibits greater resistance to forgetting on old classes such as tumor and stroma, and demonstrates improved learning ability on new classes, particularly lymphatic tissue. These results indicate that our approach achieves superior overall performance in both knowledge retention and integration of new class information.

Table 2. Ablation study of three component modules

DBE	DFC	L_{FCLS}	Overlap			Disjoint		
			1-3	4-5	All	1-3	4-5	All
			53.76	27.98	41.22	56.10	32.44	44.20
✓	✓		51.33	31.91	42.03	50.21	39.45	46.31
✓	✓	✓	**60.50**	**43.25**	**43.06**	**50.66**	**44.47**	**46.48**

3.4 Ablation Studies

We conducted a systematic ablation study on the CPATH dataset to assess the performance gains of the core components through comparative experiments. The baseline model follows the incremental learning framework proposed in WILSON, with results shown in Table 2. First, by introducing the DBE and DFC to replace the original single-branch feature encoder, the mIoU score improved by 1.46%. We hypothesize that the dual-branch architecture enables fine-grained feature representation by decoupling the feature space, while the global-local feature interaction mechanism effectively captures cross-level feature representations in pathological images. The dual-path supervision strategy provided a further 0.6% improvement by jointly optimizing image-level labels and pixel-level predictions, leveraging the hierarchical complementarity of supervisory signals to refine the segmentation results.

4 Conclusion

We propose a novel Dual-Branch Dynamic Coupling Network (DDCWISS) to address the challenges of catastrophic forgetting and the reliance on pixel-level annotations in histopathological region segmentation. Experiments on the CPATH dataset demonstrate that DDCWISS significantly reduces the dependence on pixel-level annotations and effectively alleviates the forgetting problem during incremental updates, showcasing its potential for practical applications in medical image analysis.

Acknowledgments. This work is supported in part by the National Natural Science Foundation of China under grants W2411054, U21A20521 and 62271178, the Postgraduate Research & Practice Innovation Program of Jiangsu Province KYCX23_2524, National Foreign Expert Project of China under Grant G2023144009L, Zhejiang Provincial Natural Science Foundation of China (LR23F010002), Wuxi Health Commission Precision Medicine Project (J202106), Jiangsu Provincial Six Talent Peaks Project (YY-124), and Major Projects of Wuxi Health Commission (Z202324).

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

References

1. Araslanov, N., Roth, S.: Single-Stage Semantic Segmentation From Image Labels. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 4252–4261 (Jun 2020). https://doi.org/10.1109/CVPR42600.2020.00431
2. Cermelli, F., Fontanel, D., Tavera, A., Ciccone, M., Caputo, B.: Incremental Learning in Semantic Segmentation from Image Labels. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 4361–4371. IEEE, New Orleans, LA, USA (Jun 2022). https://doi.org/10.1109/CVPR52688.2022.00433
3. Cermelli, F., Fontanel, D., Tavera, A., Ciccone, M., Caputo, B.: Incremental Learning in Semantic Segmentation from Image Labels (Mar 2022). https://doi.org/10.48550/arXiv.2112.01882
4. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. IEEE Transactions on Pattern Analysis and Machine Intelligence **40**(4), 834–848 (Apr 2018). https://doi.org/10.1109/TPAMI.2017.2699184
5. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. pp. 248–255 (Jun 2009). https://doi.org/10.1109/CVPR.2009.5206848
6. Han, C., Lin, J., Mai, J., Wang, Y., Zhang, Q., Zhao, B., Chen, X., Pan, X., Shi, Z., Xu, Z., Yao, S., Yan, L., Lin, H., Huang, X., Liang, C., Han, G., Liu, Z.: Multi-layer pseudo-supervision for histopathology tissue semantic segmentation using patch-level classification labels. Medical Image Analysis **80**, 102487 (Aug 2022). https://doi.org/10.1016/j.media.2022.102487

7. Han, C., Pan, X., Yan, L., Lin, H., Li, B., Yao, S., Lv, S., Shi, Z., Mai, J., Lin, J., Zhao, B., Xu, Z., Wang, Z., Wang, Y., Zhang, Y., Wang, H., Zhu, C., Lin, C., Mao, L., Wu, M., Duan, L., Zhu, J., Hu, D., Fang, Z., Chen, Y., Zhang, Y., Li, Y., Zou, Y., Yu, Y., Li, X., Li, H., Cui, Y., Han, G., Xu, Y., Xu, J., Yang, H., Li, C., Liu, Z., Lu, C., Chen, X., Liang, C., Zhang, Q., Liu, Z.: WSSS4LUAD: Grand Challenge on Weakly-supervised Tissue Semantic Segmentation for Lung Adenocarcinoma (Apr 2022). https://doi.org/10.48550/arXiv.2204.06455
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 770–778 (Jun 2016). https://doi.org/10.1109/CVPR.2016.90
9. Jin, C., Luo, L., Lin, H., Hou, J., Chen, H.: HMIL: Hierarchical Multi-Instance Learning for Fine-Grained Whole Slide Image Classification. IEEE Transactions on Medical Imaging pp. 1–1 (2024). https://doi.org/10.1109/TMI.2024.3520602
10. Li, K.: Weakly supervised histopathology image segmentation with self-attention. Medical Image Analysis (2023)
11. Liu, P., Ji, L., Zhang, X., Ye, F.: Pseudo-Bag Mixup Augmentation for Multiple Instance Learning-Based Whole Slide Image Classification. IEEE transactions on medical imaging **43**(5), 1841–1852 (May 2024). https://doi.org/10.1109/TMI.2024.3351213
12. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 9992–10002. IEEE, Montreal, QC, Canada (Oct 2021). https://doi.org/10.1109/ICCV48922.2021.00986
13. Lukasik, M., Bhojanapalli, S., Menon, A.K., Kumar, S.: Does label smoothing mitigate label noise? In: Proceedings of the 37th International Conference on Machine Learning. ICML'20, vol. 119, pp. 6448–6458. JMLR.org (Jul 2020)
14. M, A., H, E., H, H., La, A., Mat, E., Ls, A.E., Ra, S., Hse, S., Af, I., Am, S., J, A., Mat, E., M, R., Ia, R., Nm, E., Y, A., Mh, O., Am, A., Mm, K., Af, Y., A, A., Dm, Y., Am, G., Am, E., Sy, F., Bm, Z., J, B., Dr, C., D, M., Da, G., Lad, C.: Structured crowdsourcing enables convolutional segmentation of histology images. Bioinformatics (Oxford, England) **35**(18) (Sep 2019). https://doi.org/10.1093/bioinformatics/btz083
15. Niu, D., Wang, X., Han, X., Lian, L., Herzig, R., Darrell, T.: Unsupervised Universal Image Segmentation
16. Yu, C., Zhou, Q., Li, J., Yuan, J., Wang, Z., Wang, F.: Foundation Model Drives Weakly Incremental Learning for Semantic Segmentation. In: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 23685–23694. IEEE, Vancouver, BC, Canada (Jun 2023). https://doi.org/10.1109/CVPR52729.2023.02268
17. Yuan, B., Zhao, D.: A Survey on Continual Semantic Segmentation: Theory, Challenge, Method and Application. IEEE Transactions on Pattern Analysis and Machine Intelligence **46**(12), 10891–10910 (Dec 2024). https://doi.org/10.1109/TPAMI.2024.3446949
18. Zhang, Y., Zhang, X., Wang, J., Yang, Y., Peng, T., Tong, C.: Mamba2MIL: State Space Duality Based Multiple Instance Learning for Computational Pathology (Aug 2024). https://doi.org/10.48550/arXiv.2408.15032