PaM-MIL: Proliferation and Metastasis Enhanced Localization for Multiple Instance Learning on Pathology Images

Pengyu Guo[1] Jiachuan Wang[2,*] Zhao Chen[1] Caleb Chen Cao[2]
Liping Wang[1] Tingyi Jiang[1] Lei Chen[1,2]
[1] HKUST(GZ), China [2] HKUST, Hong Kong SAR, China
pguo657@connect.hkust-gz.edu.cn, jwangey@cse.ust.hk
chenzhao@hkust-gz.edu.cn, caochen.hkust@gmail.com
lwang347@connect.hkust-gz.edu.cn, tjiang622@connect.hkust-gz.edu.cn
leichen@cse.ust.hk

Abstract

In multiple instance learning (MIL)-based whole-slide image (WSI) analysis, positive instance localization is a crucial downstream task. However, current MIL approaches prioritize WSI-level classification accuracy, resulting in insufficient localization of positive instances: (i) they produce sparse localization results that focus only on the most discriminative instances; and (ii) they overlook spatially scattered tumor regions. We introduce PaM-MIL, a biologically inspired framework that explicitly enhances positive-instance localization by mirroring the biological tumor growth process. PaM-MIL comprises two complementary modules: a Tumor Proliferation Module (TPM) that mitigates localization sparsity, and a Tumor Metastasis Module (TMM) that recovers small and scattered regions overlooked by MIL models. Extensive experiments on three public pathology datasets (CAMELYON-16, CAMELYON-17 and PANDA) demonstrate that PaM-MIL achieves new state-of-the-art instance localization performance while consistently improving WSI-level classification accuracy. Comprehensive ablations further validate the distinct and indispensable contributions of TPM and TMM. Code is available at https://github.com/hkustgary/PaM-MIL.git.

1. Introduction

The pathology biopsy represents the *gold standard* for cancer diagnosis [23]. With the development of modern digital scanning devices, pathology slides can be digitized into *whole-slide images* (WSIs) [25]. These images are multi-scale and extremely large - often on the order of a billion pixels [30]. As a result, pixel-level annotation on WSIs is prohibitively expensive and labor-intensive. To solve this problem, a promising approach is to leverage weakly supervised learning techniques to bypass costly pixel-level annotations [8, 19, 27, 36]. Among these, *multiple instance learning* (MIL) [16, 17, 21, 28, 29, 38] stands out for its ability to enable model training using only slide-level labels (e.g., tumor vs normal) that are easy to obtain.

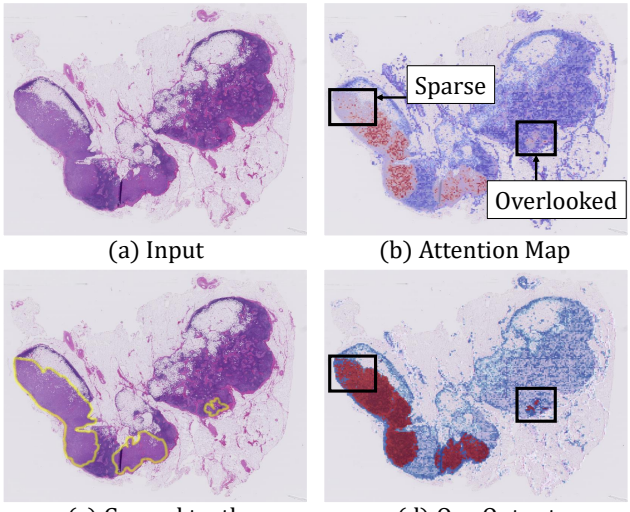


Figure 1. Positive instance localization in WSI. (a) A raw WSI image from the CAMELYON-16 testing set. (b) ABMIL [16] attention map, which gives sparse localization and overlooks a key tumor region. (c) Official localization ground truth. (d) Localization given by PaM-MIL.

MIL methods treat each WSI as a labeled *bag* containing numerous unlabeled *instances* (patches). Given a WSI, conventional MIL methods can make predictions on both the bag-level (i.e., classification) and instance-level (i.e., localization). The bag-level prediction is generated by aggregating instance features to form a bag representation. The

*Corresponding author.

instance-level localization is typically derived from the post hoc attention weights assigned to each instance during this aggregation process [16, 17, 21, 38]. Recently, the importance of the instance localization task has been increasingly recognized, as it is critical for enabling quantitative clinical analysis. For instance, assessing a patient's eligibility for cancer immunotherapy requires the Tumor Proportion Score (TPS) [26], which in turn requires accurate and comprehensive localization of positive instances.

However, existing MIL methods [16, 28, 38] prioritize bag-level classification accuracy while sacrificing the completeness of positive instance localization [6]. As illustrated in Figure 1(b), relying on the attention map used for bag classification, existing methods often fail to maintain good coverage of positive instances. Specifically, existing methods have two limitations: (1) focusing only on the most discriminative positive instances, resulting in sparse localization; (2) overlooking small and scattered tumor regions when multiple tumor regions are present. Overlooking these regions skews metrics related to tumor proportion scores, leading to erroneous patient stratification [9].

To tackle these challenges, we delve into the tumor growth process: cancer typically originates from a *primary point*, where it first *proliferates* and grows before gradually acquiring the ability to invade surrounding tissues [10]. Eventually, it may *metastasize* through the lymphatic system or bloodstream [24]. Motivated by this process, we propose a novel MIL method called **P**roliferation **a**nd **M**etastasis Multiple Instance Learning (PaM-MIL), which consists of two modules—the Tumor Proliferation Module (TPM) and the Tumor Metastasis Module (TMM). Specifically, the TPM module addresses limitation (1), namely sparse localization, by reformulating the tumor proliferation process as a label propagation problem on a spatial k-NN graph. TPM first identifies primary points using a density-based primary point identification strategy, where tumor density increases as one moves closer to the primary point and decreases toward the boundary of the tumor region [31]. Then, to mimic the tumor proliferation process, TPM introduces an efficient iterative propagation solver that scales label propagation to WSIs, avoiding the high computational complexity induced by the massive number of instances. Further, the TMM module addresses limitation (2): overlooking small tumor regions. TMM is based on the pathological prior [35] that these small cancerous regions should share similar representations with the primary tumor. To exploit this similarity while robustly separating true tumors from feature-similar false negatives, TMM introduces a Bi-correlation Measurement (BcM) strategy that jointly computes the correlation and anti-correlation between instances and tumor regions to purify the signal. Then, TMM applies a Self-adaptive Thresholding (SaT) mechanism to dynamically accommodate slide-specific variations in discriminative signals. Together, PaM-MIL enables effective instance localization from slide-level supervision while preserving strong bag-level classification performance. Our key contributions are summarized as follows:

- We propose a novel attention-based MIL method, PaM-MIL, which explicitly models the distinct pathological processes of tumor proliferation and metastasis. It converts slide-level supervision into dense instance localization maps without requiring pixel-level labels.
- The proposed TPM module formulates tumor proliferation as a computationally efficient label propagation problem on a spatial graph to densify and regularize sparse attention. The proposed TMM module couples a Bi-correlation Measurement (BcM) strategy with a GMM-based Self-adaptive Thresholding (SaT) mechanism to recover overlooked regions.
- We conduct extensive experiments on three datasets to validate the effectiveness of PaM-MIL, which achieves state-of-the-art performance, outperforms the best per-dataset baseline by 9.4% in F_1 score on average.

2. Related Work

Deep MIL approaches on WSI generally include two main categories: instance-level methods and bag-level methods. Instance-level methods [5, 32, 37] assign slide-level labels to patches, then train a patch-level classifier, and subsequently aggregate the instance predictions for bag-level prediction. In contrast, bag-level methods [16, 17, 21, 28] focus on combining instance-level features to form high-level bag representations, which are then used to train bag-level classifiers based on bag labels. Among the two categories, bag-level methods have been shown to be more effective. Approaches in this category mainly differ in how they determine attention values. For example, in AB-MIL [16], attention values are learned through a side-branch network; in DSMIL [17], they are derived from the cosine distance between the features of an instance and those of a critical instance. Additionally, CLAM [21] introduces multiple attention branches, providing separate slide-level representations for each class. DTFD-MIL [38] addresses overfitting by resampling instances to create a variety of sub-bags for augmentation. However, these methods disregard the structural dependencies shared among instances within the same bag. To solve this problem, TransMIL [28] uses the transformer architecture to capture mutual correlations among instances. Recent advancements in MIL have integrated spatial information using diverse methods, including structured graphs [18], neighbor-constrained attention modules [11], and superpatch-based spatial quantification [12]. In addition, vision-language models have been employed in recent works [13, 14, 29] to enhance WSI classification in few-shot settings.

Despite strong slide-level accuracy, most methods offer

limited localization ability. To address this issue, methods such as MHIM-MIL [33] and ACMIL [39] employ instance masking to mine broader discriminative patterns. However, these score-driven masking strategies lack explicit pathological structure modeling. To achieve more contiguous localization, recent work has concentrated on incorporating local spatial priors [6, 11]. A prominent example is Sm [6], which assumes that neighboring instances may share the same label and therefore introduces a smoothing operator to enforce local consistency. However, the smoothness assumption conflicts with key pathological phenomena such as metastasis, which often manifests as small tumor regions. As a result, the Sm smoothing operator tends to misinterpret these critical non-smooth signals as noise and suppress them. Other approaches like CAMIL [11] explore neighborhood-constrained attention, but they still hinge on simplistic local-neighborhood assumptions.

3. Methodology

3.1. Overview

Problem Formulation. We consider a binary classification scenario where a WSI $\mathbf{X}$ is treated as a bag containing b non-overlapping patches (instances) $\{x_i\}_{i=1}^{b}$ with the number of instances b varying across different WSIs (bags). Our goal is to define a function $f : \mathbf{X} \to \{0,1\}^b$ such that, for a given WSI $\mathbf{X}$, $f(\mathbf{X})$ produces a vector of patch-level predictions $\{y_i\}_{i=1}^{b}$ where each $y_i \in \{0,1\}$ corresponds to the prediction of whether patch x_i is part of a tumor region. Crucially, while this goal requires a patch-level classification for each WSI $\mathbf{X}$, only bag-level label $Y \in \{0,1\}$ is available during training. This task of patch-level predictions has been termed positive instance localization in recent literature [6, 17]. We adopt this terminology throughout the paper.

Revisit Attention-based MIL. Attention-based MIL methods [16, 17, 21, 38] focus on learning bag representation $\mathbf{E}$ from features extracted from instances $\{x_i\}_{i=1}^{b}$. This approach transforms the problem into a conventional classification task where a classifier is trained using the bag representations. Specifically, most bag-level methods obtain the bag representation $\mathbf{E}$ as follows:

$$\mathbf{E} = \sum_{i=1}^{b} a_i \mathbf{f}_i \in \mathbb{R}^d, \tag{1}$$

where $\mathbf{f}_i$ denotes the embedding of x_i, which is generated by projecting x_i through a pretrained feature extractor, and a_i is the learnable weight, which is referred to as the attention value [16] for $\mathbf{f}_i$. Eq. (1) implies that the attention map $\mathbf{a} = \{a_i\}_{i=1}^{b}$ generated by bag-level methods has the potential to localize positive instances within WSIs. One intuitive method is to select instances with high attention values from the attention map to form a positive instance set:

$$\mathcal{I} = \{x_i | a_i \geqslant \tau_a\}, \tag{2}$$

where τ_a is an attention value threshold used to determine the label of each instance.

Framework Overview. In attention-based MIL, the primary focus is on achieving correct bag-level classification. Consequently, the resulting attention maps tend to capture only the most discriminative instances rather than the full extent of the positive region (as illustrated in Figure 1(b)). To address this problem, we propose the PaM-MIL framework, which is designed as a lightweight framework that integrates directly into the training loop of any attention-based MIL backbone. An overview of PaM-MIL is given in Figure 2. In the forward pass, the MIL backbone first produces its sparse attention map $\mathbf{a}$. This map, together with instance features $\mathbf{f}$ and spatial coordinates, is processed by our TPM (Figure 2(b)) and TMM (Figure 2(c)) modules to yield a refined attention map $\mathbf{a}'$. This refined map $\mathbf{a}'$ is then used to compute the final bag representation $\mathbf{E} = \sum_{i=1}^{b} a'_i \mathbf{f}_i$. The training loss is then computed from $\mathbf{E}$ using the bag-level labels. This design allows PaM-MIL to enhance localization while being trained using only slide-level supervision.

3.2. TPM: Tumor Proliferation Module

The core idea of TPM is to mimic the tumor proliferation process. It first identifies primary points and then simulates the outward spread of tumor cells from these seeds into the surrounding tissue [31]. To model spatial proliferation, we first capture the spatial structure of instances in the WSI. Given a set of b instances $V = \{x_i\}_{i=1}^{b}$, we construct a spatial proximity graph $G = (V, E)$. An undirected edge $(i, j) \in E$ is added between x_i and x_j if x_j is among the k nearest spatial neighbors of x_i in the WSI's 2D coordinate system. Let $\mathbf{W} \in \{0,1\}^{b \times b}$ denote the binary adjacency matrix of G, where $\mathbf{W}_{ij} = 1$ if $(i, j) \in E$ and 0 otherwise.

Density-based primary point identification. We begin by querying the MIL backbone for an attention map $\mathbf{a}$ that yields non-negative scores a_i per instance x_i. Since larger a_i contribute more to the WSI-level prediction, we treat a_i as a surrogate measure of the likelihood that x_i is positive.

Given G, the neighborhood of x_i is defined as: $\mathcal{N}_i = \{x_j | (i, j) \in E\}$. The attention density ρ_i is calculated by aggregating local context via neighbor averaging:

$$\rho_i = \frac{1}{|\mathcal{N}_i|} \sum_{x_j \in \mathcal{N}_i} a_j, \quad \text{if } |\mathcal{N}_i| > 0 \tag{3}$$

Intuitively, ρ_i measures whether an instance is spatially surrounded by high-attention instances. We therefore define the set of primary points as the instances attaining the maximum attention density:

$$\mathcal{P} = \{x_i \in V \mid \rho_i = \max_j \{\rho_j\}_{j=1}^{b}\} \tag{4}$$

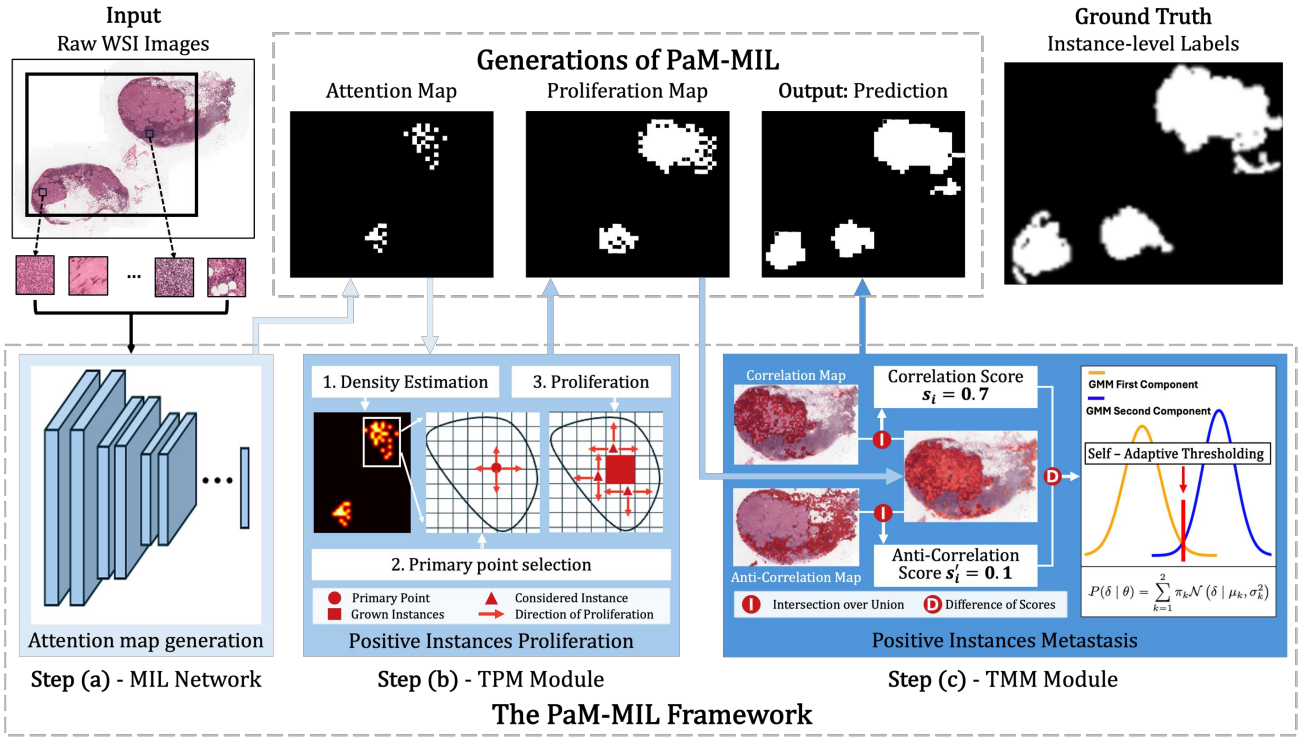


Figure 2. Illustration of our PaM-MIL framework for the positive instance localization. In step (a), a MIL model is used to generate attention maps for WSIs. Then, in step (b), TPM module is used to identify primary points and proliferate positive instances, resulting in a fine-grained positive instance localization map. Finally, in step (c), TMM module is employed to identify small and scattered tumor regions that may be overlooked during the proliferation process.

Positive Instance Proliferation. A fundamental process in pathology, known as proliferation, involves a tumor growing from its primary points and gradually expanding to invade surrounding healthy tissues [10]. To simulate this process, we cast proliferation from $\mathcal{P}$ as a label-propagation problem on the spatial graph G. We first define a binary source vector $\mathbf{y} \in \{0,1\}^b$:

$$y_i = \begin{cases} 1 & \text{if } x_i \in \mathcal{P} \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

Our goal is to compute a proliferation map $\mathbf{r} \in \mathbb{R}^b$ that (i) is semantically consistent on G (i.e., spatially adjacent instances should have similar scores) and (ii) remains faithful to its source $\mathbf{y}$. We achieve this by solving the following convex optimization:

$$\mathbf{r}^* = \arg\min_{\mathbf{r}} \left\{ \mu ||\mathbf{r} - \mathbf{y}||_2^2 + \mathbf{r}^T \mathbf{L} \mathbf{r} \right\} \quad , \tag{6}$$

where $\mathbf{L} = \mathbf{I} - \hat{\mathbf{S}}$ is the symmetrically normalized graph Laplacian of G, $\hat{\mathbf{S}} = \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$ is the normalized adjacency, $\mathbf{D} = \mathrm{diag}\,(\mathbf{W}\mathbf{1}_N)$ is the degree matrix, and $\mathbf{1}_N$ is an all-ones vector. $\mu > 0$ is a regularization weight.

Although the optimization in Eq. (6) admits a closed-form solution, directly inverting a $b \times b$ matrix is impractical for WSIs with large b. We therefore adopt a computationally efficient solution [6] to solve it:

$$\mathbf{r}^{(t+1)} = \alpha \hat{\mathbf{S}} \mathbf{r}^{(t)} + (1-\alpha)\mathbf{y}, \tag{7}$$

where $\mathbf{r}^{(t)}$ is the proliferation map at iteration t, and $\alpha \in (0,1)$ is a propagation parameter. Intuitively, the propagation term $\alpha \hat{\mathbf{S}} \mathbf{r}^{(t)}$ simulates tumor growth, propagating the current scores to spatial neighbors. The source-injection term $(1-\alpha)\mathbf{y}$ repeatedly re-injects the primary point signal at each iteration to ensure the process remains anchored to its source. As shown in Proposition 1, the iteration is guaranteed to converge to a unique fixed point $\mathbf{r}$. We define this stationary vector $\mathbf{r}$ as the final proliferation map. This map yields a dense distribution of scores over positive instances that mitigates the sparsity of the initial attention map.

Proposition 1 *For $\alpha \in (0,1)$, the iterative process in Eq. (7) is guaranteed to converge from any initialization $\mathbf{r}^{(0)}$ to a unique fixed point $\mathbf{r}^*$, which coincides with the unique minimizer of the convex objective in Eq. (6) and satisfies $\mathbf{r} = \alpha \hat{\mathbf{S}} \mathbf{r} + (1-\alpha)\mathbf{y}$. As a result, $\mathbf{r}$ covers a larger portion of the tumor region than the sparse attention map $\mathbf{a}$ (see Supplementary Material A for proof).*

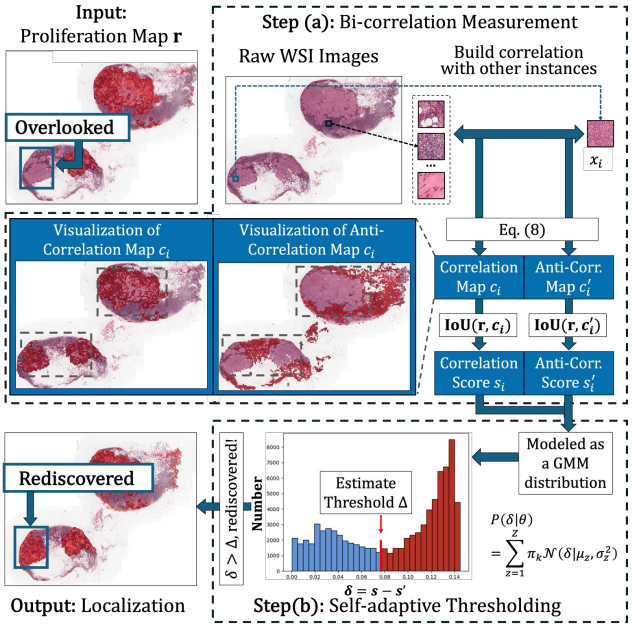


Figure 3. Overview of the TMM. In step (a), we compute the correlation and anti-correlation score for each instance. In step (b), a self-adaptive thresholding mechanism is employed to refine the predictions of localization.

3.3. TMM: Tumor Metastasis Module

While the proliferation map $\mathbf{r}$ effectively mitigates the sparse localization issue of the raw attention map, it introduces the risk of failing to capture small and scattered tumor regions. This limitation arises because the MIL attention map is designed to focus primarily on the most discriminative regions, often overlooking small and scattered tumor areas, as illustrated in Figure 1(b). Relying solely on the TPM module is insufficient to fully address this issue. Therefore, we propose a two-stage TMM module, as illustrated in Figure 3. First, a Bi-correlation Measurement (BcM) strategy leverages inter-instance correlations to rediscover positive instances that the TPM module might have overlooked. Then, a Self-adaptive Thresholding (SaT) mechanism further refines the localization results.

Bi-correlation Measurement. In the pathological process, metastatic tumors often share morphological traits with the original tumor region. Based on this principle, it is reasonable to assume that the small cancer regions overlooked by the TPM are likely to exhibit similar features to those captured in the proliferation map. Therefore, we aim to identify instances whose features are similar to those of the potential tumor regions highlighted by $\mathbf{r}$.

Specifically, to identify instances that might share similar features with $\mathbf{r}$, we calculate two scores for each instance: the correlation score and the anti-correlation score. The correlation score measures the similarity between an instance and $\mathbf{r}$, while the anti-correlation score measures their dissimilarity. Specifically, the correlation score of instance x_i is computed by evaluating the Intersection over Union (IoU) between a correlation map $\mathbf{c}_i$ and $\mathbf{r}$, where $\mathbf{c}_i$ is constructed by calculating the cosine similarity between the feature vector $\mathbf{f}_i$ of instance x_i and the feature vectors of all other instances in the WSI, mapped to the range $[0, 1]$. Mathematically, this is defined as:

$$\mathbf{c}_i = \left[\frac{1}{2} \left(\frac{\mathbf{f}_i \cdot \mathbf{f}_j}{\|\mathbf{f}_i\|_2 \|\mathbf{f}_j\|_2} + 1 \right) \right]_{j=1}^{b}, \tag{8}$$

where the feature vector $\mathbf{f}_i$ is obtained by passing instance x_i through a pretrained encoder, and $\mathbf{c}_i^j$ denotes the similarity between the features of instances x_i and x_j. The correlation score s_i for instance x_i is then calculated as the soft IoU between its correlation map $\mathbf{c}_i$ and $\mathbf{r}$:

$$s_i = \frac{\sum_{j=1}^{b} \mathbf{r}_j \mathbf{c}_i^j}{\sum_{j=1}^{b} \left(\mathbf{r}_j + \mathbf{c}_i^j - \mathbf{r}_j \mathbf{c}_i^j \right)}. \tag{9}$$

The anti-correlation score is computed similarly using the anti-correlation map, which is obtained by:

$$s_i' = \frac{\sum_{j=1}^{b} \mathbf{r}_j (1 - \mathbf{c}_i^j)}{\sum_{j=1}^{b} \left(\mathbf{r}_j + (1 - \mathbf{c}_i^j) - \mathbf{r}_j (1 - \mathbf{c}_i^j) \right)}. \tag{10}$$

Self-adaptive Thresholding. Intuitively, if the correlation score of a patch is greater than its anti-correlation score, it means that the patch is more similar to the potential tumor region than to the background. However, in many cases, the difference between the two scores may be marginal, for instance, $s_i = 0.05$ and $s_i' = 0.049$. Although the former is slightly greater, this small difference might not reflect a significant distinction in the alignment with $\mathbf{r}$, and it could lead to incorrect labeling of an instance as a tumor patch.

Therefore, the major challenge becomes determining a threshold Δ for each WSI that indicates whether the difference $\delta_i = s_i - s_i'$ is large enough for a given instance x_i to be assigned a positive label. For this purpose, we propose a self-adaptive thresholding mechanism. By plotting a histogram of δ for all instances in WSIs from the CAMELYON-16 dataset, two distinct peaks corresponding to different types of δ values emerge, as illustrated in Step (b) of Figure 3. The left peak comprises instances that the scores exhibit a marginal difference, whereas the right peak consists of instances with large δ values, making them eligible for a positive label. Since the two modes can be reasonably approximated by Gaussian distributions, we model them with a Gaussian probability density function as:

$$\mathcal{N}(\delta \,|\, \theta_k) = \frac{1}{\sigma_k \sqrt{2\pi}} e^{-\frac{(\delta - \mu_k)^2}{2(\sigma_k)^2}}, \quad (k = 1, 2) \tag{11}$$

where θ_k denotes the parameter of the k-th Gaussian component. Consequently, a two-component Gaussian Mixture

Model, $P(\delta \mid \theta)$, can be employed to characterize the distribution of the score differences.

$$P(\delta \mid \theta) = \sum_{k=1}^{2} \pi_k \mathcal{N}\left(\delta \mid \mu_k, \sigma_k^2\right) \tag{12}$$

The parameters of the GMM in Eq. (12) can be estimated using the EM algorithm [2]. Since the distribution of δ is bimodal, with a clear separation between the two components, we can determine the threshold Δ by letting $\mathcal{N}(\delta \mid \theta_1) = \mathcal{N}(\delta \mid \theta_2)$. Then, the x-axis value of the intersection point between the two Gaussian distributions is computed and adopted as the self-adaptive threshold.

Thereby, by fitting a GMM to each WSI image, the threshold Δ can be adapted dynamically according to the corresponding score-difference distribution. Finally, only instances with $\delta_i > \Delta$ are assigned a positive label, effectively eliminating the false positives introduced by instances with smaller or negligible differences between s_i and s'_i.

Overall, the TMM effectively detects positive instances that may have been missed by the TPM, leading to more comprehensive and accurate localization. This improvement is evident in Figure 6 in the Supplementary Material and Table 3 in Section 4.3.

4. Experiments and Results

To assess the effectiveness of the PaM-MIL framework, we conduct a comprehensive suite of experiments aimed at demonstrating that: (1) PaM-MIL substantially outperforms state-of-the-art baselines on positive-instance localization; (2) these localization gains are achieved without compromising, and in many cases improving the WSI-level classification; and (3) each biologically inspired module in the framework, namely TPM and TMM, provides a critical and complementary contribution to the final performance.

Datasets and Pre-processing. We evaluate PaM-MIL on three public pathology datasets with pixel-level annotations: CAMELYON-16 [3], CAMELYON-17 [1], and PANDA [4]. The CAMELYON-16 dataset targets lymph-node metastasis detection. Following the official split, we use 271 training WSIs and 129 test WSIs. The CAMELYON-17 [1] dataset contains 500 WSIs and targets pN staging. Following prior setups [34], we formulate a binary task (pN0 vs. pN+) with a test/train split of 0.2:0.8. The PANDA [4] dataset includes prostate cancer Gleason grading. For all three datasets, 25% of the training set is reserved for validation.

All WSIs are processed at 20× magnification. We first remove background via the Otsu thresholding algorithm, and then tile tissue regions into non-overlapping 512×512 patches. Although the test sets provide instance-level labels, we strictly adhere to a weak-supervision protocol during the training of PaM-MIL and use only slide-level labels.

Implementation Details. All baseline models and PaM-MIL are trained with the Adam optimizer with a learning rate of 1e-4. Experiments are run on NVIDIA A800 GPUs. By default, we adopt an ImageNet-pretrained ResNet-50 [15] as the feature extractor. To construct the graph G, we apply k-NN with $k = 8$. In label propagation, the diffusion parameter α in Eq. (7) is set to 0.2. The iterative update in Eq. (7) is run until convergence, with a maximum of $T = 20$ iterations.

Baselines. We compare PaM-MIL against a strong set of baselines. **Traditional attention-based MIL**: These methods primarily target bag-level classification, where instance attention weights arise as a by-product of feature aggregation. Representatives include ABMIL [16], CLAM [21], DSMIL [17], and DTFD-MIL [38]. **Global-context MIL**: Beyond simple attention, these approaches explicitly model inter-instance relations and long-range dependencies, e.g., via Transformers or graph structures. We include TransMIL [28] and IBMIL [20] in this category. **Localization-enhanced MIL**: These methods introduce modules aimed at improving localization by injecting spatial priors. We include CAMIL [11] (neighborhood-constrained) and Sm [6] (a smoothing operator to regularize sparse attention).

4.1. Quantitative performance evaluation.

Positive Instance Localization Performance. As summarized in Table 1, PaM-MIL establishes a new state-of-the-art performance across all three datasets and all metrics. First, it consistently surpasses both traditional attention-based and global-context MIL baselines. For example, on CAMELYON-16, PaM-MIL improves the F_1 score from 0.3694 (ABMIL) and 0.2541 (TransMIL) to 0.6651, corroborating our core premise that raw attention maps derived from MIL methods are inherently sparse and incomplete for localization. Moreover, even against localization-enhanced methods, PaM-MIL maintains a clear advantage. For example, compared to Sm, it improves the mean F_1 score across datasets by 12.76%. These results demonstrate the effectiveness of PaM-MIL for localization on WSIs.

Bag-level Classification Performance. In this subsection, we answer the question: Does the enhancement of localization hurt WSI classification? Table 2 reports WSI-level classification results. When built on top of ABMIL, PaM-MIL yields consistent gains across all three datasets. For instance, on PANDA it improves AUROC from 0.9214 (ABMIL) to 0.9394. We posit that this is because standard MIL methods produce suboptimal bag representations (Eq. (1)) due to their incomplete understanding of positive instances. By contrast, PaM-MIL constructs a more complete map of positive regions, which in turn enables stronger re-aggregation and a more robust bag-level representation, ultimately improving classification accuracy.

Table 1. Instance localization performance comparison, reported as the mean ± standard deviation over five independent runs with different seeds. **best** result in bold, second-best result with an underline

Method	CAMELYON-16 [3]			CAMELYON-17 [1]			PANDA [4]		
	F_1 ↑	AUROC ↑	ACC ↑	F_1 ↑	AUROC ↑	ACC ↑	F_1 ↑	AUROC ↑	ACC ↑
ABMIL [16]	$0.3694_{\pm 0.087}$	$0.5208_{\pm 0.119}$	$0.9601_{\pm 0.025}$	$0.4783_{\pm 0.074}$	$0.5668_{\pm 0.098}$	$0.9895_{\pm 0.012}$	$0.4506_{\pm 0.012}$	$0.7450_{\pm 0.007}$	$0.8156_{\pm 0.008}$
CLAM [21]	$0.4002_{\pm 0.065}$	$0.4927_{\pm 0.079}$	$0.9642_{\pm 0.021}$	$0.3879_{\pm 0.053}$	$0.4075_{\pm 0.092}$	$0.9851_{\pm 0.008}$	$0.5119_{\pm 0.011}$	$0.8036_{\pm 0.005}$	$0.8396_{\pm 0.007}$
DSMIL [17]	$0.2052_{\pm 0.195}$	$0.4661_{\pm 0.125}$	$0.9562_{\pm 0.029}$	$0.2393_{\pm 0.159}$	$0.4306_{\pm 0.132}$	$0.9741_{\pm 0.009}$	$0.4788_{\pm 0.008}$	$0.7534_{\pm 0.005}$	$0.8179_{\pm 0.008}$
DTFD-MIL [38]	$0.4969_{\pm 0.107}$	$\underline{0.6824}_{\pm 0.053}$	$0.9609_{\pm 0.041}$	$\underline{0.5401}_{\pm 0.091}$	$0.5945_{\pm 0.072}$	$\underline{0.9912}_{\pm 0.007}$	$0.5166_{\pm 0.009}$	$0.7865_{\pm 0.009}$	$0.8397_{\pm 0.006}$
TransMIL [28]	$0.2541_{\pm 0.136}$	$0.5918_{\pm 0.068}$	$0.9597_{\pm 0.039}$	$0.3614_{\pm 0.129}$	$0.6916_{\pm 0.053}$	$0.9847_{\pm 0.009}$	$0.5150_{\pm 0.034}$	$0.7614_{\pm 0.026}$	$0.8407_{\pm 0.009}$
IBMIL [20]	$0.1951_{\pm 0.149}$	$0.4439_{\pm 0.081}$	$0.9532_{\pm 0.031}$	$0.3383_{\pm 0.102}$	$0.4415_{\pm 0.085}$	$0.9872_{\pm 0.008}$	$0.5238_{\pm 0.024}$	$0.7341_{\pm 0.024}$	$0.8152_{\pm 0.021}$
CAMIL [11]	$0.3801_{\pm 0.098}$	$0.6671_{\pm 0.075}$	$0.9624_{\pm 0.028}$	$0.4735_{\pm 0.079}$	$\underline{0.7092}_{\pm 0.067}$	$0.9813_{\pm 0.009}$	$0.5095_{\pm 0.019}$	$0.7723_{\pm 0.031}$	$0.8249_{\pm 0.009}$
Sm [6]	$\underline{0.5309}_{\pm 0.093}$	$0.6579_{\pm 0.097}$	$\underline{0.9671}_{\pm 0.020}$	$0.4394_{\pm 0.011}$	$0.5383_{\pm 0.114}$	$0.9890_{\pm 0.007}$	$\underline{0.5736}_{\pm 0.011}$	$\underline{0.8195}_{\pm 0.005}$	$\underline{0.8432}_{\pm 0.010}$
PaM-MIL	$\mathbf{0.6651}_{\pm 0.067}$	$\mathbf{0.8108}_{\pm 0.041}$	$\mathbf{0.9677}_{\pm 0.021}$	$\mathbf{0.6137}_{\pm 0.091}$	$\mathbf{0.7792}_{\pm 0.059}$	$\mathbf{0.9931}_{\pm 0.006}$	$\mathbf{0.6479}_{\pm 0.011}$	$\mathbf{0.8345}_{\pm 0.006}$	$\mathbf{0.8543}_{\pm 0.012}$

Table 2. Bag classification performance comparison, reported as the mean ± standard deviation over five independent runs with different seeds. **best** result in bold, second-best result with an underline

Method	CAMELYON-16 [3]			CAMELYON-17 [1]			PANDA [4]		
	F_1 ↑	AUROC ↑	ACC ↑	F_1 ↑	AUROC ↑	ACC ↑	F_1 ↑	AUROC ↑	ACC ↑
ABMIL [16]	$0.6383_{\pm 0.021}$	$0.7967_{\pm 0.035}$	$0.7364_{\pm 0.009}$	$0.5917_{\pm 0.082}$	$0.7112_{\pm 0.079}$	$0.7251_{\pm 0.012}$	$0.8901_{\pm 0.006}$	$0.9214_{\pm 0.004}$	$0.8439_{\pm 0.003}$
CLAM [21]	$0.6827_{\pm 0.062}$	$\underline{0.8217}_{\pm 0.051}$	$0.7685_{\pm 0.007}$	$\mathbf{0.6411}_{\pm 0.092}$	$\underline{0.7391}_{\pm 0.094}$	$\mathbf{0.7682}_{\pm 0.008}$	$0.8735_{\pm 0.007}$	$0.9104_{\pm 0.004}$	$0.8326_{\pm 0.006}$
DSMIL [17]	$0.4792_{\pm 0.157}$	$0.7355_{\pm 0.048}$	$0.6958_{\pm 0.015}$	$0.4017_{\pm 0.185}$	$0.6589_{\pm 0.087}$	$0.6753_{\pm 0.039}$	$0.8837_{\pm 0.009}$	$0.9063_{\pm 0.005}$	$0.8347_{\pm 0.009}$
DTFD-MIL [38]	$0.7064_{\pm 0.049}$	$0.7608_{\pm 0.047}$	$0.7892_{\pm 0.005}$	$0.5342_{\pm 0.097}$	$0.7013_{\pm 0.046}$	$0.7041_{\pm 0.009}$	$\underline{0.8936}_{\pm 0.005}$	$\underline{0.9249}_{\pm 0.003}$	$0.8425_{\pm 0.006}$
TransMIL [28]	$0.6631_{\pm 0.084}$	$0.7605_{\pm 0.024}$	$0.7569_{\pm 0.007}$	$0.5789_{\pm 0.113}$	$0.6421_{\pm 0.069}$	$0.7247_{\pm 0.105}$	$0.8820_{\pm 0.012}$	$0.9223_{\pm 0.004}$	$0.8417_{\pm 0.008}$
IBMIL [20]	$0.6497_{\pm 0.192}$	$0.7849_{\pm 0.124}$	$0.7523_{\pm 0.026}$	$0.6018_{\pm 0.157}$	$0.6652_{\pm 0.142}$	$0.7349_{\pm 0.046}$	$0.8661_{\pm 0.018}$	$0.9192_{\pm 0.017}$	$0.8251_{\pm 0.012}$
CAMIL [11]	$0.6417_{\pm 0.058}$	$0.7417_{\pm 0.041}$	$0.7678_{\pm 0.007}$	$0.5476_{\pm 0.079}$	$0.6550_{\pm 0.097}$	$0.7069_{\pm 0.095}$	$0.8892_{\pm 0.004}$	$0.9211_{\pm 0.002}$	$0.8449_{\pm 0.005}$
Sm [6]	$\underline{0.7442}_{\pm 0.052}$	$0.7992_{\pm 0.057}$	$\underline{0.8295}_{\pm 0.004}$	$0.6053_{\pm 0.118}$	$0.7261_{\pm 0.112}$	$0.7482_{\pm 0.007}$	$0.8931_{\pm 0.007}$	$0.9241_{\pm 0.004}$	$\underline{0.8489}_{\pm 0.008}$
PaM-MIL	$\mathbf{0.7562}_{\pm 0.091}$	$\mathbf{0.8235}_{\pm 0.075}$	$\mathbf{0.8298}_{\pm 0.024}$	$\underline{0.6292}_{\pm 0.122}$	$\mathbf{0.7471}_{\pm 0.129}$	$\underline{0.7584}_{\pm 0.014}$	$\mathbf{0.8954}_{\pm 0.009}$	$\mathbf{0.9394}_{\pm 0.009}$	$\mathbf{0.8519}_{\pm 0.018}$

4.2. Qualitative Positive Instance Localization Analysis

To further illustrate PaM-MIL's ability for positive instance localization, we visualize the localization results of PaM-MIL on the CAMELYON-16 test set. This experiment compares (i) the raw, highly sparse attention maps from the ABMIL baseline, (ii) the Sm-smoothed attention maps, and (iii) our PaM-MIL output. As shown in Figure 4, ABMIL highlights only a few of the most discriminative instances, yielding extremely sparse maps. Sm improves continuity over the primary tumor regions; however, as illustrated in (a) and (b), it overlooks smaller, scattered tumor regions. Moreover, in (c) it produces numerous false positives. In contrast, PaM-MIL captures nearly the entire positive region, offering a more comprehensive and robust localization of positive instances. These visual results support our claim that PaM-MIL's advantage stems from its biologically inspired modeling, where TPM mitigates sparsity (proliferation) and TMM recovers missed regions (metastasis). Furthermore, even in challenging scenarios where the initial attention maps are extremely noisy, as depicted in row (c), PaM-MIL demonstrates strong robustness and consistently achieves excellent localization.

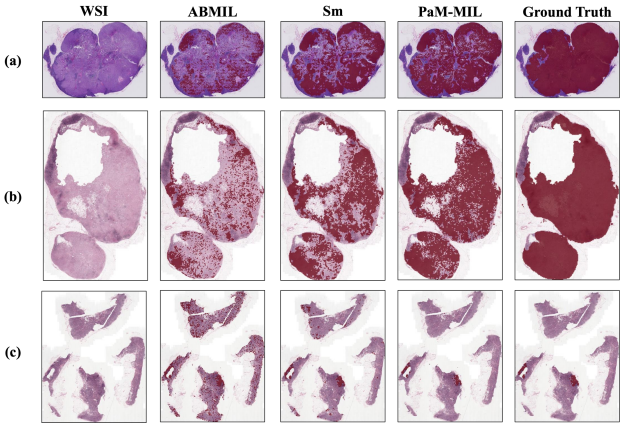


Figure 4. Visualization of the localization results on the CAMELYON-16 dataset. The red color denotes positive instances.

4.3. Ablation Studies

In this subsection, we conduct ablation studies on the three datasets to justify key design choices in PaM-MIL.

Effects of Key Modules in PaM-MIL. We first compare the performance of PaM-MIL with the following variations: (1) Raw attention map produced by ABMIL; (2) PaM-MIL with TPM module only (i.e., the proliferation map); (3) PaM-MIL with TPM and TMM, but where the TMM employs either the bi-correlation measurement or the self-adaptive thresholding mechanism; and (4) PaM-MIL with TPM and TMM, where the TMM utilizes both the bi-correlation measurement and the self-adaptive thresholding mechanism. The results are presented in Table 3. Notably, an ablation setting evaluating the TMM module in isola-

Table 3. Ablation study on CAMELYON-16 and PANDA datasets. Metrics are reported as the mean of AUROC and F_1 over five independent runs for both bag-level classification and instance localization.(↑ indicates higher is better).

Method	CAMELYON-16				PANDA			
	AUROC$_{loc}$ ↑	F1$_{loc}$ ↑	AUROC$_{bag}$ ↑	F1$_{bag}$ ↑	AUROC$_{loc}$ ↑	F1$_{loc}$ ↑	AUROC$_{bag}$ ↑	F1$_{bag}$ ↑
Map$_{Att}$	0.5208	0.3694	0.7967	0.6383	0.7450	0.4506	0.9214	0.8901
Map$_{Att}$ + TPM	0.8065	0.6247	0.7923	0.6667	0.8121	0.5318	0.9217	0.8807
Map$_{Att}$ + TPM + TMM$_{BcM}$	0.8042	0.6620	0.7982	0.6367	0.8256	0.5933	0.9194	0.8836
Map$_{Att}$ + TPM + TMM$_{SaT}$	0.8082	0.6549	0.8182	0.6947	0.8172	0.6192	0.9289	0.8839
Map$_{Att}$ + TPM + TMM$_{BcM+SaT}$	**0.8108**	**0.6651**	**0.8235**	**0.7562**	**0.8345**	**0.6479**	**0.9394**	**0.8954**

* **Map$_{Att}$** denotes the attention map; **TMM$_{BcM}$** and **TMM$_{SaT}$** represent bi-correlation measurement and self-adaptive thresholding in the TMM module.

tion is omitted, as TMM is structurally designed as a downstream component that strictly relies on the TPM. A visualization of each module's specific role is also provided in the Supplementary Materials C.

As shown in Table 3, on CAMELYON-16, the raw attention map from ABMIL (Map$_{Att}$) attains an F_1 = 0.3694. Adding TPM (Map$_{Att}$ + TPM) further improves it to 0.6247. This +25.53% gain confirms that TPM's proliferation model is central to mitigating attention sparsity. Building on TPM, introducing the full TMM further boosts F_1 to 0.6651, while AUROC rises from 0.8065 to 0.8108. This indicates that TMM successfully recovers regions missed by TPM. Moreover, Table 3 highlights a critical synergy between BcM and SaT, and their combination enables TMM to precisely detect omitted metastatic areas while effectively suppressing TMM-induced false positives.

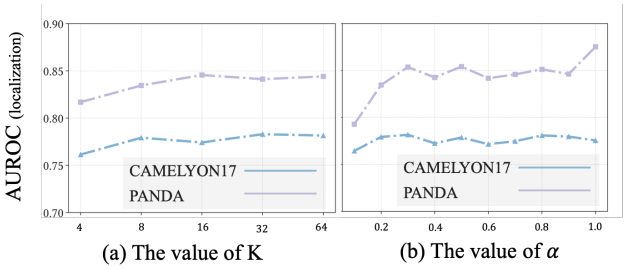
Effects of Primary Point Identification Strategies. Recall that in Section 3.2, we select high-density instances as primary points and conduct proliferation based on them. In this ablation study, we test different primary point identification strategies on CAMELYON-16, and the results are shown in Table 4. Strategies include: (1) randomly selecting $k\%$ positive instances from the attention map as primary points, (2) selecting the top-$k\%$ instances with the highest attention values as primary points, and (3) selecting the instances with the highest density as primary points.

We observe that the random strategy performs the worst, underscoring the necessity of deliberate primary point selection. The attention strategy is clearly better, yet still inferior to our density strategy. This advantage substantiates TPM's biological hypothesis: TPM is designed to mimic tumor proliferation outward from the primary point. In contrast, attention-based seeding often picks the most discriminative patches that may lie near tumor boundaries. However, proliferating from boundaries is biologically implausible and yields suboptimal localization. Our density strategy instead seeks the centroid of high-attention tumor regions, providing a more faithful starting point and, consequently, stronger localization.

Sensitivity to Hyper Parameters. We further examine how the TPM parameter α in Eq. (7) and the k-NN parameter k in the density-based primary point selection strategy affects localization. Figure 5 plots the localization performance on the CAMELYON-17 and PANDA datasets, the AUROC metric remains stable across a wide range of α and k values, demonstrating that our method is of low sensitivity on hyperparameters.

Table 4. Performance comparison of different primary point identification strategies.

Strategy	Setting	AUROC$_{loc}$ ↑	F1$_{loc}$ ↑	ACC$_{loc}$ ↑
Random	$k = 1\%$	0.6105	0.3267	0.9471
	$k = 5\%$	0.6116	0.3346	0.9482
Attention	$k = 1\%$	0.7142	0.5621	0.9506
	$k = 5\%$	0.7491	0.6115	0.9571
Density-based strategy	–	**0.8108**	**0.6651**	**0.9677**



(a) The value of K (b) The value of α

Figure 5. Hyperparameter sensitivity analysis.

5. Conclusion

In this paper, we presented PaM-MIL, a lightweight module for enhancing the localization performance of pathology MIL models. Inspired by the biological processes of tumor proliferation and metastasis, we introduce two modules: the Tumor Proliferation Module (TPM) and the Tumor Metastasis Module (TMM). Starting from the primary points identified by our density-based primary point strategy, TPM reformulates localization as a label propagation task on a spatial k-NN graph. Complementing this, TMM utilizes a bi-correlation measurement strategy along with a self-adaptive thresholding mechanism to identify small tumor regions that MIL methods might overlook. Extensive experimental results demonstrate that our method achieves new state-of-the-art performance in positive instance localization on WSIs.

Acknowledgments

Lei Chen is supported by National Key Research and Development Program of China Grant No. 2023YFF0725100, National Science Foundation of China under Grant No. U22B2060, Guangdong-Hong Kong Technology Innovation Joint Funding Scheme Project No. 2024A0505040012, AOE Project AoE/E-603/18, Theme-based project TRS T41-603/20R, CRF Project C2004-21G, Key Areas Special Project of Guangdong Provincial Universities 2024ZDZX1006, Guangdong Province Science and Technology Plan Project 2023A0505030011, HKUST(GZ) CMCC (Guangzhou Branch) Metaverse Joint Innovation Lab under Grant No. P00659, Hong Kong ITC TC-SKLCRCC26EG01, ITF grant PRP/004/22FX, Zhujiang scholar program 2021JC02X170, HKUST Webank joint research lab. Also, this work is supported by the HKUST(GZ)-LBP Medical Data Intelligence Joint-Lab. Jiachuan Wang's work is supported in part by JST CREST (JPMJCR22M2).

References

[1] Peter Bandi, Oscar Geessink, Quirine Manson, Marcory Van Dijk, Maschenka Balkenhol, Meyke Hermsen, Babak Ehteshami Bejnordi, Byungjae Lee, Kyunghyun Paeng, Aoxiao Zhong, et al. From detection of individual metastases to classification of lymph node status at the patient level: the camelyon17 challenge. *IEEE transactions on medical imaging*, 38(2):550–560, 2018. 6, 7

[2] Leonard E Baum, Ted Petrie, George Soules, and Norman Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. *The annals of mathematical statistics*, 41(1):164–171, 1970. 6

[3] Babak Ehteshami Bejnordi, Mitko Veta, Paul Johannes Van Diest, Bram Van Ginneken, Nico Karssemeijer, Geert Litjens, Jeroen AWM Van Der Laak, Meyke Hermsen, Quirine F Manson, Maschenka Balkenhol, 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. 6, 7

[4] Wouter Bulten, Kimmo Kartasalo, Po-Hsuan Cameron Chen, Peter Ström, Hans Pinckaers, Kunal Nagpal, Yuannan Cai, David F Steiner, Hester Van Boven, Robert Vink, et al. Artificial intelligence for diagnosis and gleason grading of prostate cancer: the panda challenge. *Nature medicine*, 28(1):154–163, 2022. 6, 7

[5] Gabriele Campanella, Matthew G Hanna, Luke Geneslaw, Allen Miraflor, Vitor Werneck Krauss Silva, Klaus J Busam, Edi Brogi, Victor E Reuter, David S Klimstra, and Thomas J Fuchs. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nature medicine*, 25(8):1301–1309, 2019. 2

[6] Francisco M. Castro-Macías, Pablo Morales-Álvarez, Yunan Wu, Rafael Molina, and Aggelos K. Katsaggelos. Sm: enhanced localization in multiple instance learning for medical imaging classification. *CoRR*, abs/2410.03276, 2024. 2, 3, 4, 6, 7

[7] Richard J Chen, Tong Ding, Ming Y Lu, Drew FK Williamson, Guillaume Jaume, Andrew H Song, Bowen Chen, Andrew Zhang, Daniel Shao, Muhammad Shaban, et al. Towards a general-purpose foundation model for computational pathology. *Nature medicine*, 30(3):850–862, 2024. 2

[8] Yuan-Chih Chen and Chun-Shien Lu. Rankmix: Data augmentation for weakly supervised learning of classifying whole slide images with diverse sizes and imbalanced categories. In *CVPR*, pages 23936–23945. IEEE, 2023. 1

[9] Deborah B Doroshow, Wei Wei, Swati Gupta, Jon Zugazagoitia, Charles Robbins, Blythe Adamson, and David L Rimm. Programmed death-ligand 1 tumor proportion score and overall survival from first-line pembrolizumab in patients with nonsquamous versus squamous nsclc. *Journal of Thoracic Oncology*, 16(12):2139–2143, 2021. 2

[10] Mark A Feitelson, Alla Arzumanyan, Rob J Kulathinal, Stacy W Blain, Randall F Holcombe, Jamal Mahajna, Maria Marino, Maria L Martinez-Chantar, Roman Nawroth, Isidro Sanchez-Garcia, et al. Sustained proliferation in cancer: Mechanisms and novel therapeutic targets. In *Seminars in cancer biology*, pages S25–S54. Elsevier, 2015. 2, 4

[11] Olga Fourkioti, Matt De Vries, Chen Jin, Daniel C Alexander, and Chris Bakal. Camil: Context-aware multiple instance learning for cancer detection and subtyping in whole slide images. *arXiv preprint arXiv:2305.05314*, 2023. 2, 3, 6, 7

[12] Zeyu Gao, Anyu Mao, Yuxing Dong, Hannah Clayton, Jialun Wu, Jiashuai Liu, ChunBao Wang, Kai He, Tieliang Gong, Chen Li, et al. Smmile enables accurate spatial quantification in digital pathology using multiple-instance learning. *Nature Cancer*, pages 1–17, 2025. 2

[13] Zhengrui Guo, Conghao Xiong, Jiabo Ma, Qichen Sun, Lishuang Feng, Jinzhuo Wang, and Hao Chen. Focus: Knowledge-enhanced adaptive visual compression for few-shot whole slide image classification. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 15590–15600, 2025. 2

[14] Minghao Han, Linhao Qu, Dingkang Yang, Xukun Zhang, Xiaoying Wang, and Lihua Zhang. Mscpt: Few-shot whole slide image classification with multi-scale and context-focused prompt tuning. *IEEE Transactions on Medical Imaging*, 2025. 2

[15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778. IEEE Computer Society, 2016. 6

[16] Maximilian Ilse, Jakub M. Tomczak, and Max Welling. Attention-based deep multiple instance learning. In *ICML*, pages 2132–2141. PMLR, 2018. 1, 2, 3, 6, 7

[17] Bin Li, Yin Li, and Kevin W. Eliceiri. Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning. In *CVPR*, pages 14318–14328. Computer Vision Foundation / IEEE, 2021. 1, 2, 3, 6, 7

[18] Jiawen Li, Yuxuan Chen, Hongbo Chu, Qiehe Sun, Tian Guan, Anjia Han, and Yonghong He. Dynamic graph representation with knowledge-aware attention for histopathology

whole slide image analysis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11323–11332, 2024. 2

[19] Jiahan Li, Jiuyang Dong, Shenjin Huang, Xi Li, Junjun Jiang, Xiaopeng Fan, and Yongbing Zhang. Virtual immunohistochemistry staining for histological images assisted by weakly-supervised learning. In *CVPR*, pages 11259–11268. IEEE, 2024. 1

[20] Tiancheng Lin, Zhimiao Yu, Hongyu Hu, Yi Xu, and Chang-Wen Chen. Interventional bag multi-instance learning on whole-slide pathological images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19830–19839, 2023. 6, 7

[21] Ming Y. Lu, Drew F. K. Williamson, Tiffany Y. Chen, Richard J. Chen, Matteo Barbieri, and Faisal Mahmood. Data efficient and weakly supervised computational pathology on whole slide images. *CoRR*, abs/2004.09666, 2020. 1, 2, 3, 6, 7

[22] Ming Y Lu, Bowen Chen, Drew FK Williamson, Richard J Chen, Ivy Liang, Tong Ding, Guillaume Jaume, Igor Odintsov, Long Phi Le, Georg Gerber, et al. A visual-language foundation model for computational pathology. *Nature medicine*, 30(3):863–874, 2024. 2

[23] Joseph A Ludwig and John N Weinstein. Biomarkers in cancer staging, prognosis and treatment selection. *Nature Reviews Cancer*, 5(11):845–856, 2005. 1

[24] Tracey A Martin, Lin Ye, Andrew J Sanders, Jane Lane, and Wen G Jiang. Cancer invasion and metastasis: molecular and cellular perspective. In *Madame Curie Bioscience Database [Internet]*. Landes Bioscience, 2013. 2

[25] Muhammad Khalid Khan Niazi, Anil V Parwani, and Metin N Gurcan. Digital pathology and artificial intelligence. *The lancet oncology*, 20(5):e253–e261, 2019. 1

[26] Sushant Patkar, Alex Chen, Alina Basnet, Amber Bixby, Rahul Rajendran, Rachel Chernet, Susan Faso, Prashanth Ashok Kumar, Devashish Desai, Ola El-Zammar, et al. Predicting the tumor microenvironment composition and immunotherapy response in non-small cell lung cancer from digital histopathology images. *NPJ Precision Oncology*, 8(1):280, 2024. 2

[27] Linhao Qu, Xiaoyuan Luo, Kexue Fu, Manning Wang, and Zhijian Song. The rise of AI language pathologists: Exploring two-level prompt learning for few-shot weakly-supervised whole slide image classification. In *NeurIPS*, 2023. 1

[28] Zhuchen Shao, Hao Bian, Yang Chen, Yifeng Wang, Jian Zhang, Xiangyang Ji, and Yongbing Zhang. Transmil: Transformer based correlated multiple instance learning for whole slide image classification. In *NeurIPS*, pages 2136–2147, 2021. 1, 2, 6, 7

[29] Jiangbo Shi, Chen Li, Tieliang Gong, Yefeng Zheng, and Huazhu Fu. Vila-mil: Dual-scale vision-language multiple instance learning for whole slide image classification. In *CVPR*, pages 11248–11258. IEEE, 2024. 1, 2

[30] Andrew H Song, Guillaume Jaume, Drew FK Williamson, Ming Y Lu, Anurag Vaidya, Tiffany R Miller, and Faisal Mahmood. Artificial intelligence for digital and computational pathology. *Nature Reviews Bioengineering*, 1(12): 930–949, 2023. 1

[31] Andreas Stadlbauer, Oliver Ganslandt, Rolf Buslei, Thilo Hammen, Stephan Gruber, Ewald Moser, Michael Buchfelder, Erich Salomonowitz, and Christopher Nimsky. Gliomas: histopathologic evaluation of changes in directionality and magnitude of water diffusion at diffusion-tensor mr imaging. *Radiology*, 240(3):803–810, 2006. 2, 3

[32] Ziyu Su, Thomas E. Tavolara, Gabriel Carreno-Galeano, Sang Jin Lee, Metin N. Gurcan, and M. Khalid Khan Niazi. Attention2majority: Weak multiple instance learning for regenerative kidney grading on whole slide images. *Medical Image Anal.*, 79:102462, 2022. 2

[33] Wenhao Tang, Sheng Huang, Xiaoxian Zhang, Fengtao Zhou, Yi Zhang, and Bo Liu. Multiple instance learning framework with masked hard instance mining for whole slide image classification. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4078–4087, 2023. 3

[34] Paul Tourniaire, Marius Ilie, Paul Hofman, Nicholas Ayache, and Hervé Delingette. Ms-clam: Mixed supervision for the classification and localization of tumors in whole slide images. *Medical Image Analysis*, 85:102763, 2023. 6

[35] Liling Wan, Klaus Pantel, and Yibin Kang. Tumor metastasis: moving new biological insights into the clinic. *Nature medicine*, 19(11):1450–1464, 2013. 2

[36] Xiyue Wang, Jinxi Xiang, Jun Zhang, Sen Yang, Zhongyi Yang, Ming-Hui Wang, Jing Zhang, Wei Yang, Junzhou Huang, and Xiao Han. SCL-WC: cross-slide contrastive learning for weakly-supervised whole-slide image classification. In *NeurIPS*, 2022. 1

[37] Zhihua Wang, Lequan Yu, Xin Ding, Xuehong Liao, and Liansheng Wang. Lymph node metastasis prediction from whole slide images with transformer-guided multiinstance learning and knowledge transfer. *IEEE Trans. Medical Imaging*, 41(10):2777–2787, 2022. 2

[38] Hongrun Zhang, Yanda Meng, Yitian Zhao, Yihong Qiao, Xiaoyun Yang, Sarah E. Coupland, and Yalin Zheng. DTFD-MIL: double-tier feature distillation multiple instance learning for histopathology whole slide image classification. In *CVPR*, pages 18780–18790. IEEE, 2022. 1, 2, 3, 6, 7

[39] Yunlong Zhang, Honglin Li, Yunxuan Sun, Sunyi Zheng, Chenglu Zhu, and Lin Yang. Attention-challenging multiple instance learning for whole slide image classification. In *European conference on computer vision*, pages 125–143. Springer, 2024. 3