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

Supplementary Material

A. Proposition 1 Proof

Proposition 1. *For $\alpha \in (0,1)$, the iterative process defined 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 $\mathbf{r} = \alpha\hat{\mathbf{S}}\mathbf{r} + (1-\alpha)\mathbf{y}$.*

Convergence Analysis We begin by showing that starting from an arbitrary initial point $\mathbf{r}^{(0)}$, Eq. (7) always converges to a unique fixed point $\mathbf{r}^*$. To this end, we rewrite the update rule in Eq. (7) as a mapping $\mathcal{T} : \mathbb{R}^b \to \mathbb{R}^b$:

$$\mathcal{T}(\mathbf{r}) = \alpha\hat{\mathbf{S}}\mathbf{r} + (1-\alpha)y \tag{13}$$

With this notation, we can derive:

$$\begin{aligned} &||\mathcal{T}(\mathbf{r}_1) - \mathcal{T}(\mathbf{r}_2)||_2 \\ &= ||(\alpha\hat{\mathbf{S}}\mathbf{r}_1 + (1-\alpha)y) - (\alpha\hat{\mathbf{S}}\mathbf{r}_2 + (1-\alpha)y)||_2 \\ &= ||\alpha\hat{\mathbf{S}}(\mathbf{r}_1 - \mathbf{r}_2)||_2 \\ &\leqslant \alpha||\hat{\mathbf{S}}||_2||\mathbf{r}_1 - \mathbf{r}_2||_2. \end{aligned} \tag{14}$$

From Eq. (14), we observe that $\mathcal{T}$ is a contraction mapping. In other words, there exists a Lipschitz constant $\kappa \in [0,1)$ satisfying: $||\mathcal{T}(\mathbf{r}_1) - \mathcal{T}(\mathbf{r}_2)|| \leqslant \kappa||\mathbf{r}_1 - \mathbf{r}_2||$ for any vectors $\mathbf{r}_1, \mathbf{r}_2$. Then according to the **Banach Fixed Point Theorem**, the iteration $\mathbf{r}^{(t+1)} = \mathcal{T}(\mathbf{r}^{(t)})$ converges to a unique fixed point $\mathbf{r}^*$ if $\mathcal{T}$ is a contraction mapping. Therefore, the sequence $\{\mathbf{r}^{(t)}\}$ converges linearly to a unique fixed point $\mathbf{r}^*$ regardless of the initialization $\mathbf{r}^{(0)}$. At convergence, $\mathbf{r}^*$ satisfies:

$$\mathbf{r}^* = \alpha\hat{\mathbf{S}}\mathbf{r}^* + (1-\alpha)y \tag{15}$$

Rearranging terms yields the closed-form solution:

$$\mathbf{r}^* = (1-\alpha)(I - \alpha\hat{\mathbf{S}})^{-1}y, \tag{16}$$

where $(I - \alpha\hat{\mathbf{S}})$ is strictly positive definite and invertible because the eigenvalues of $\alpha\hat{\mathbf{S}}$ are strictly less than 1.

Optimality Proof We now demonstrate that the iteration process presented in Eq. (7) minimizes the objective function in Eq. (6). We reformulate Eq. (6) as the following generalized energy function $\mathcal{J}(\mathbf{r})$

$$\mathcal{J}(\mathbf{r}) = \mu||\mathbf{r} - y||_2^2 + \mathbf{r}^T\mathbf{L}\mathbf{r} \tag{17}$$

where $L = I - \hat{\mathbf{S}}$ is the normalized Laplacian, and $\mu > 0$ is a regularization weight. Since L is positive semi-definite and the squared error term is strictly convex, $\mathcal{J}(\mathbf{r})$ is strictly convex. The global minimizer is characterized by $\nabla_r\mathcal{J}(\mathbf{r}) = 0$:

$$\nabla_r\mathcal{J}(\mathbf{r}) = 2\mu(\mathbf{r} - y) + 2\mathbf{L}\mathbf{r} = 0 \tag{18}$$

$$\Rightarrow \mu(\mathbf{r} - y) + (I - \hat{\mathbf{S}})\mathbf{r} = 0 \tag{19}$$

$$\Rightarrow (\mu I + I - \hat{\mathbf{S}})\mathbf{r} = \mu y \tag{20}$$

$$\Rightarrow ((\mu+1)I - \hat{\mathbf{S}})\mathbf{r} = \mu y \tag{21}$$

$$\Rightarrow (I - \frac{1}{\mu+1}\hat{\mathbf{S}})\mathbf{r} = \frac{\mu}{\mu+1}y \tag{22}$$

Letting $\alpha = \frac{1}{\mu+1}$, the optimality condition becomes:

$$(I - \alpha\hat{\mathbf{S}})\mathbf{r} = (1-\alpha)y, \tag{23}$$

which is identical to the fixed point equation of the iterative process described in Eq. (7).

B. Complexity Analysis

We analyze the computational complexity of PaM-MIL with respect to the number of instances per bag N and the feature dimension D. The TPM module (Sec. 3.2) is computationally efficient. It consists of two main steps: (i) Building a k-NN graph (with k=8) in the 2D coordinate space and (ii) running label propagation, which converges within a fixed number of iterations $T = 20$. Since both k and T are small constants, the cost of graph construction and label propagation grows linearly with the number of instances N. Therefore, the time complexity of TPM is $O(N)$. Then, the main computational bottleneck lies in the bi-correlated measurement of the TMM module (Sec. 3.3). A straightforward implementation requires computing pairwise similarities between all instances, resulting in an $N \times N$ similarity matrix with complexity $O(N^2D)$. However, this similarity matrix is in fact low-rank and can be efficiently approximated. In our implementation, we use the Nyström method, a standard technique for approximating large kernel matrices. By sampling m landmark instances (with $m \ll N$), we reduce the complexity of approximating the $N \times N$ similarity matrix from $O(N^2D)$ to $O(NmD)$. Putting everything together, since TPM is linear in N and $m \ll N$, the overall practical complexity of PaM-MIL is dominated by the Nyström-based approximation in TMM, i.e., $O(NmD)$.

C. Visualization Ablation Study

We evaluate PaM-MIL under different settings (as shown in Fig. 6): the raw attention map generated by ABMIL, the re-

sults with only the TPM, and the complete PaM-MIL with both TPM and TMM. As observed, each component performs as expected. The TPM, by simulating proliferation, successfully addresses the sparsity issue present in the original attention map, while the TMM effectively finds the instances that the proliferation map alone fails to capture.

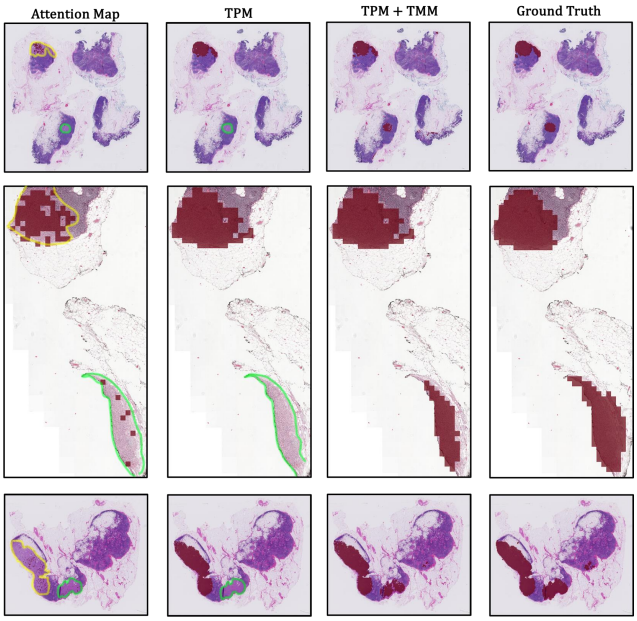

Figure 6. Visualization of the localization results with/without TPM and TMM. Sparse and overlooked predictions are highlighted in yellow and green, respectively.

D. Limitations and Future Work

Because our biologically inspired modules (TPM and TMM) operate at the attention and spatial-graph levels, they are theoretically orthogonal to the choice of the underlying patch encoder. Recent pathology-specific foundation models, such as UNI [7] and CONCH [22], provide highly discriminative representations. Exploring the integration of our method with these advanced models, and examining whether our proliferation and metastasis priors can further refine their high-quality initial localizations, is a promising direction for future research. Furthermore, akin to most existing localization-enhanced MIL approaches, our current framework focuses primarily on binary instance localization. Adapting PaM-MIL to handle multi-class scenarios, such as distinguishing various tumor subtypes or grading levels is another valuable avenue for future exploration.