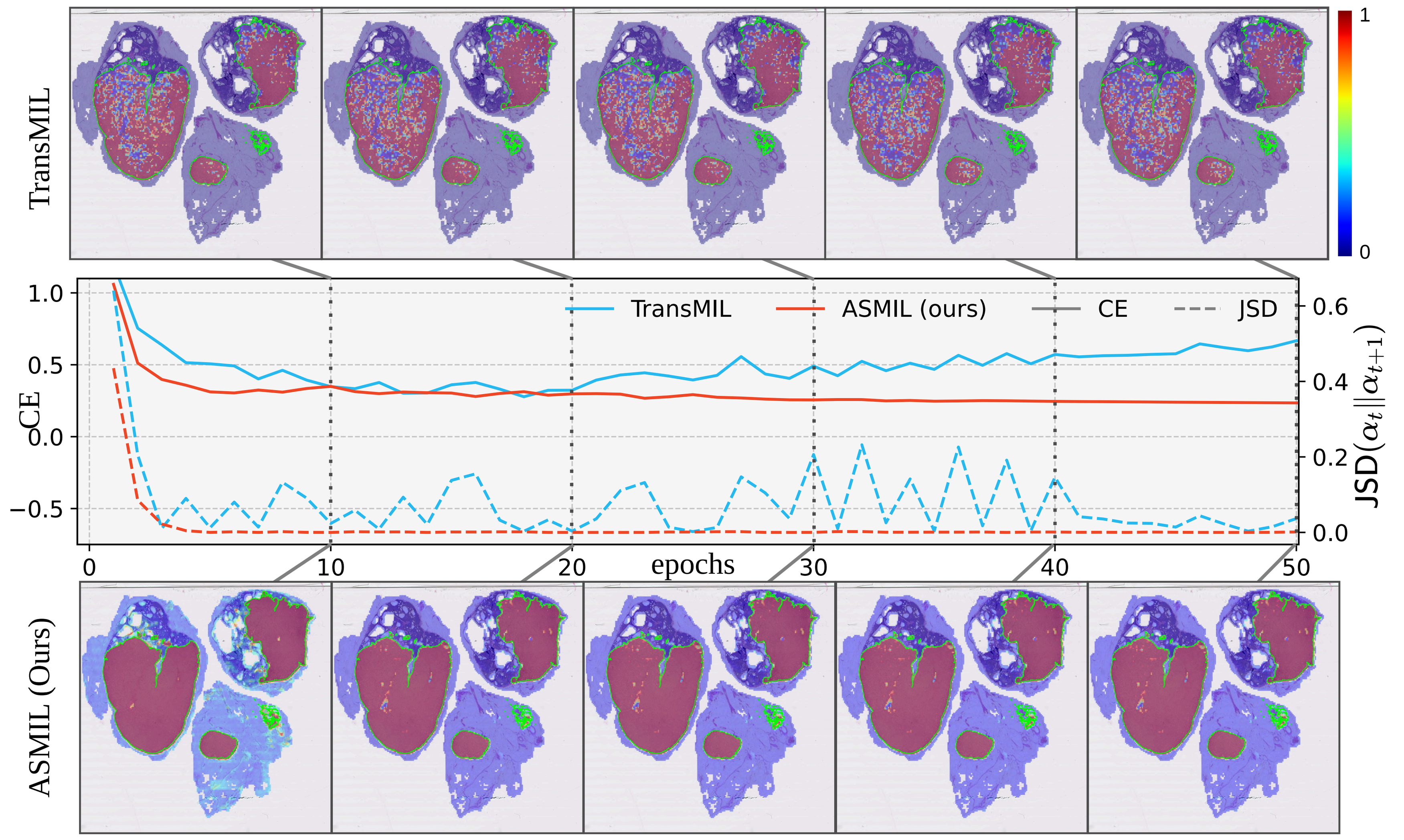
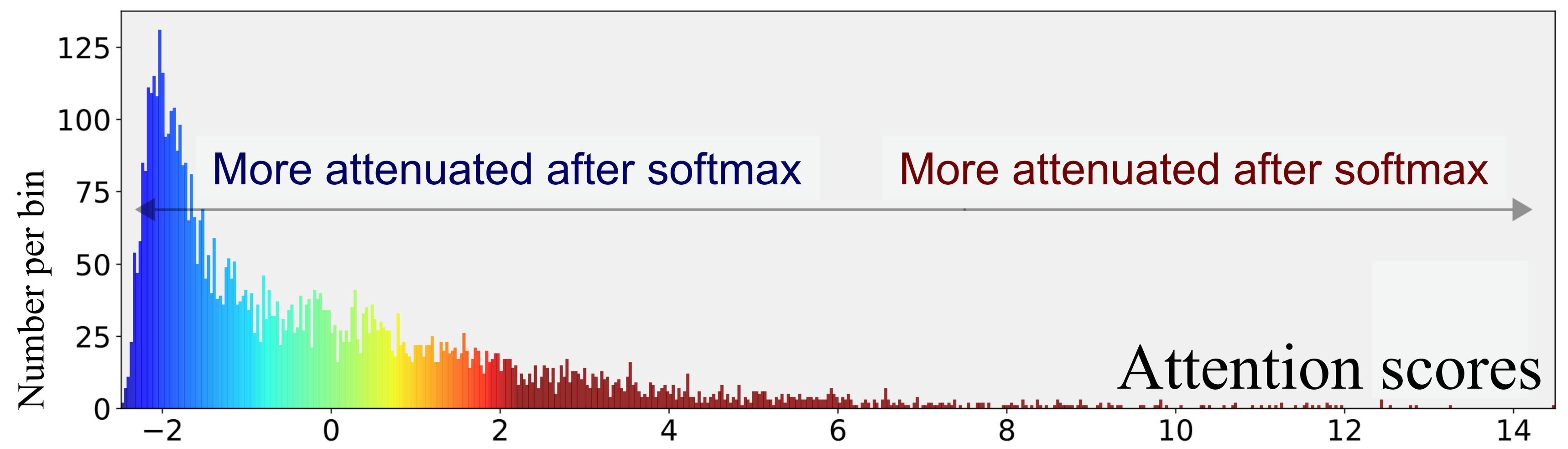
# ASMIL: Attention-Stabilized Multiple Instance Learning for Whole Slide Imaging

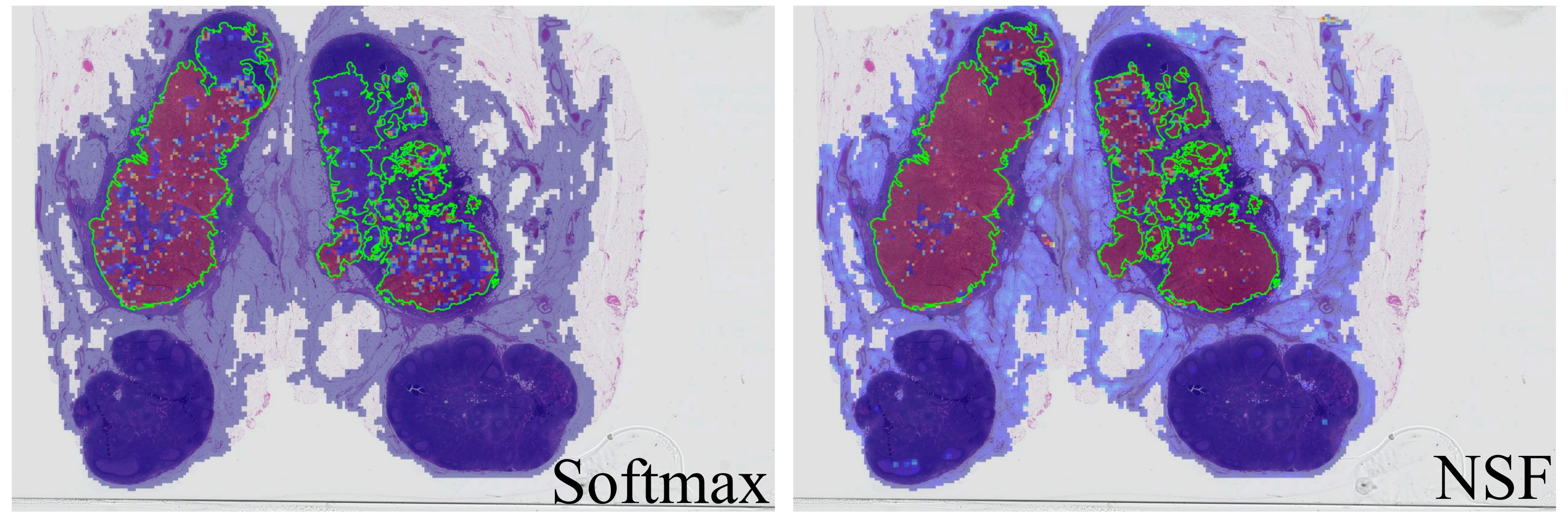
Linfeng Ye, Shayan Mohajer Hamidi, Zhixiang Chi, Guang Li, Mert Pilanci, Takahiro Ogawa, Miki Haseyama, Konstantinos N. Plataniotis

## Problem: Unstable & Over-Concentrated Attention

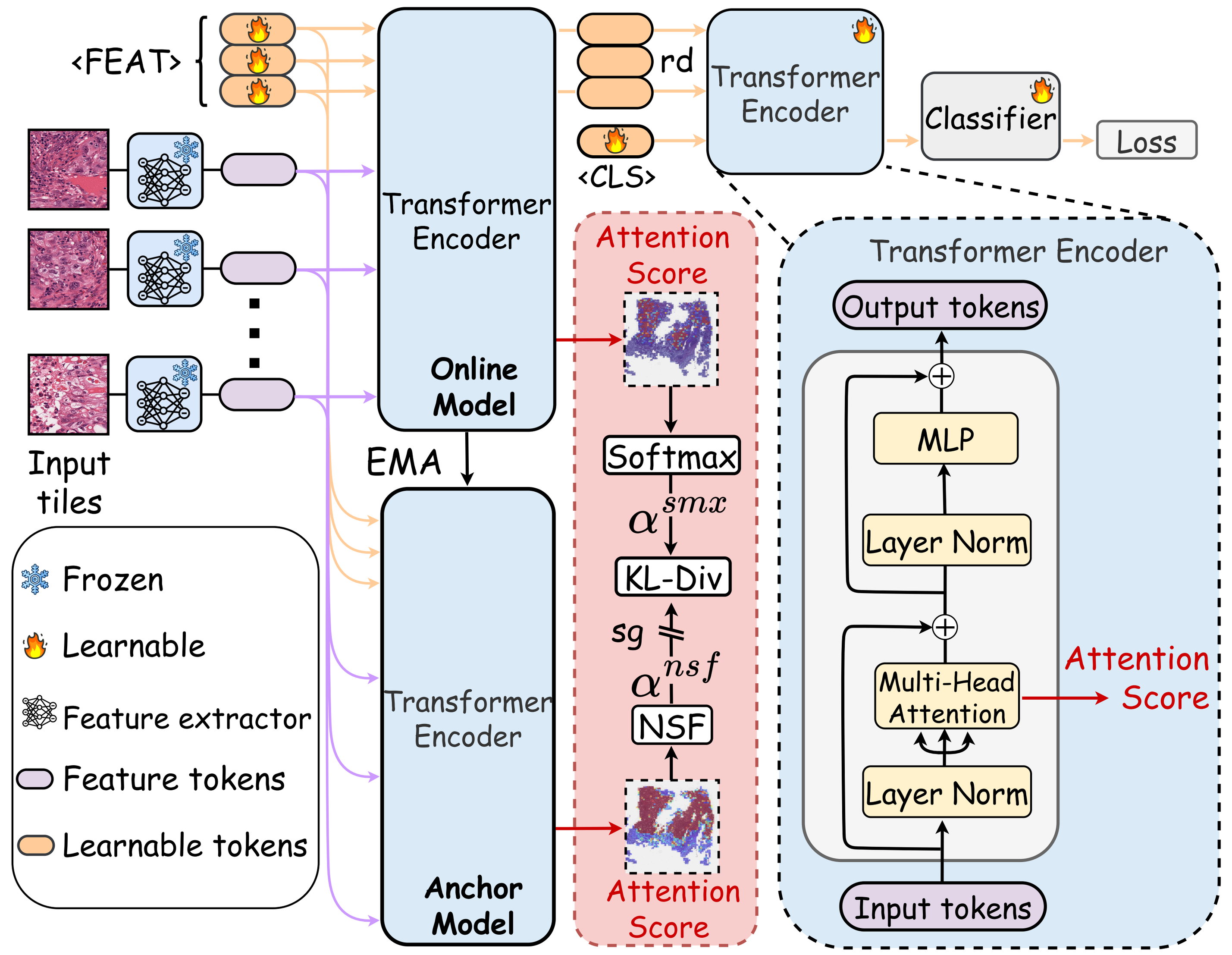
ABMIL [1] often learns **unstable** and **over-concentrated** attention over WSI tiles, hurting convergence and producing an unfaithful attention map.




## Attention-Stabilized Multiple Instance Learning




**Theorem** Normalized sigmoid function (NSF) achieves selective flattening; softmax cannot with a single $T$] Fix $\tau > 0$, $\gamma \geq 0$, and index sets $\mathcal{H}, \mathcal{L}$ with $h \geq 1$, $\ell \geq 1$. For any $\boldsymbol{z} \in \mathcal{S}(\tau, \gamma, \mathcal{H}, \mathcal{L})$:

**(A) NSF bounds.** For any $i, h' \in \mathcal{H}$ and any $j \in \mathcal{L}$,

$$\frac{\alpha_i^{\mathrm{nsf}}(\boldsymbol{z})}{\alpha_{h'}^{\mathrm{nsf}}(\boldsymbol{z})} = \frac{\sigma(z_i)}{\sigma(z_{h'})} \leq \frac{\sigma(\tau+\gamma)}{\sigma(\tau)} = \frac{1+e^{-\tau}}{1+e^{-(\tau+\gamma)}} \leq 1 + e^{-\tau}, \qquad \alpha_j^{\mathrm{nsf}}(\boldsymbol{z}) \leq \frac{\sigma(-\tau)}{h\,\sigma(\tau)} = \frac{e^{-\tau}}{h}.$$

Hence, NSF equalizes the high tokens up to a factor $1 + e^{-\tau}$ and suppresses lows to at most $e^{-\tau}/h$. As $\tau \to \infty$ with fixed $\gamma$, ratios among high tokens approach 1 and low-token weights vanish.

**(B) Softmax incompatibility with one temperature.** Suppose we desire suppression and equalization targets $(\varepsilon, \kappa)$ on $\mathcal{S}(\tau, \gamma, \mathcal{H}, \mathcal{L})$:

$$\text{(Suppression)} \quad \alpha_j^{\mathrm{smx}}(\boldsymbol{z}; T) \leq \varepsilon \;\; \forall j \in \mathcal{L}, \qquad \text{(Equalization)} \quad \frac{\max_{i \in \mathcal{H}} \alpha_i^{\mathrm{smx}}(\boldsymbol{z}; T)}{\min_{h' \in \mathcal{H}} \alpha_{h'}^{\mathrm{smx}}(\boldsymbol{z}; T)} \leq \kappa.$$

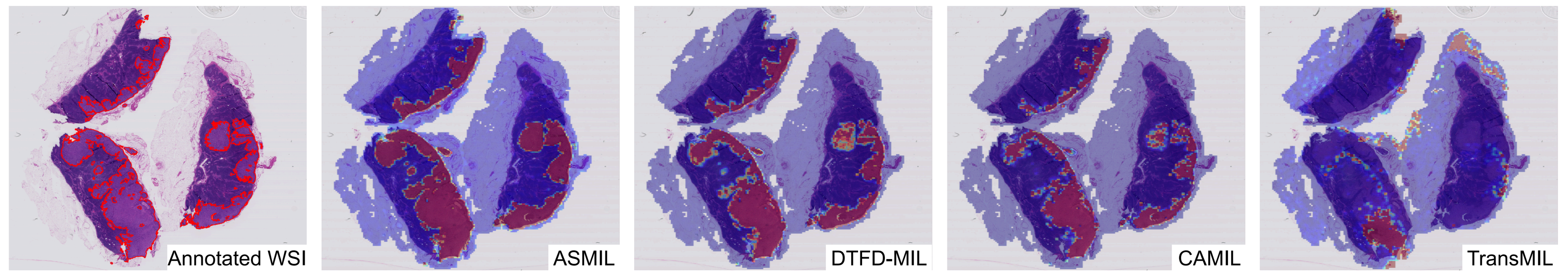
Then $T$ must satisfy $T \leq \frac{2\tau}{\log(\frac{h}{\varepsilon})}$ and $T \geq \frac{\gamma}{\log \kappa}$ simultaneously, which is impossible whenever $\frac{\gamma}{\log \kappa} > \frac{2\tau}{\log(\frac{h}{\varepsilon})}$. Thus, no single temperature achieves both targets for all $\boldsymbol{z} \in \mathcal{S}(\tau, \gamma, \mathcal{H}, \mathcal{L})$.

## Experimental Results

### Subtyping (Classification) Results

| Backbone | Dataset / Method | CAMELYON-16 F1 score ↑ | CAMELYON-16 AUC ↑ | CAMELYON-17 F1 score ↑ | CAMELYON-17 AUC ↑ | BRACS F1 score ↑ | BRACS AUC ↑ |
|---|---|---|---|---|---|---|---|
| ResNet-18 ImageNet Pretrained | ABMIL | $0.757_{\pm 0.020}$ | $0.790_{\pm 0.027}$ | $0.508_{\pm 0.032}$ | $0.779_{\pm 0.021}$ | $0.523_{\pm 0.028}$ | $0.723_{\pm 0.035}$ |
| | Clam-SB | $0.742_{\pm 0.024}$ | $0.763_{\pm 0.049}$ | $0.504_{\pm 0.012}$ | $0.778_{\pm 0.024}$ | $0.521_{\pm 0.046}$ | $0.750_{\pm 0.039}$ |
| | TransMIL | $0.643_{\pm 0.088}$ | $0.706_{\pm 0.076}$ | $0.499_{\pm 0.082}$ | $0.794_{\pm 0.053}$ | $0.444_{\pm 0.040}$ | $0.732_{\pm 0.043}$ |
| | DSMIL | $0.736_{\pm 0.025}$ | $0.773_{\pm 0.034}$ | $0.473_{\pm 0.052}$ | $0.705_{\pm 0.022}$ | $0.511_{\pm 0.052}$ | $0.751_{\pm 0.028}$ |
| | DTFD-MIL | $0.758_{\pm 0.051}$ | $0.815_{\pm 0.063}$ | $0.546_{\pm 0.010}$ | $0.735_{\pm 0.011}$ | $0.469_{\pm 0.016}$ | $0.717_{\pm 0.032}$ |
| | IBMIL | $0.777_{\pm 0.009}$ | $0.799_{\pm 0.050}$ | $0.533_{\pm 0.015}$ | $0.813_{\pm 0.092}$ | $0.510_{\pm 0.043}$ | $0.726_{\pm 0.034}$ |
| | MHIM-MIL | $0.752_{\pm 0.034}$ | $0.772_{\pm 0.026}$ | $0.56_{\pm 0.029}$ | $0.815_{\pm 0.019}$ | $0.511_{\pm 0.022}$ | $0.775_{\pm 0.021}$ |
| | ACMIL | $0.798_{\pm 0.029}$ | $0.841_{\pm 0.030}$ | $0.528_{\pm 0.053}$ | $0.789_{\pm 0.046}$ | $0.552_{\pm 0.048}$ | $0.754_{\pm 0.008}$ |
| | CAMIL | $0.778_{\pm 0.011}$ | $0.812_{\pm 0.017}$ | $0.503_{\pm 0.007}$ | $0.806_{\pm 0.006}$ | $0.569_{\pm 0.007}$ | $\underline{0.787}_{\pm 0.011}$ |
| | AEM | $\underline{0.804}_{\pm 0.022}$ | $\underline{0.859}_{\pm 0.031}$ | $0.525_{\pm 0.043}$ | $0.828_{\pm 0.054}$ | $0.554_{\pm 0.004}$ | $0.764_{\pm 0.008}$ |
| | HDMIL | $0.790_{\pm 0.023}$ | $0.856_{\pm 0.027}$ | $\underline{0.557}_{\pm 0.007}$ | $\mathbf{0.853}_{\pm 0.013}$ | $\underline{0.578}_{\pm 0.012}$ | $0.761_{\pm 0.011}$ |
| | ASMIL (Ours) | $\mathbf{0.814}_{\pm 0.052}$ | $\mathbf{0.870}_{\pm 0.064}$ | $\mathbf{0.564}_{\pm 0.020}$ | $\underline{0.851}_{\pm 0.061}$ | $\mathbf{0.601}_{\pm 0.072}$ | $\mathbf{0.810}_{\pm 0.054}$ |
| VIT-S SSL pretrained | ABMIL | $0.914_{\pm 0.031}$ | $0.945_{\pm 0.027}$ | $0.522_{\pm 0.050}$ | $0.853_{\pm 0.016}$ | $0.680_{\pm 0.051}$ | $0.866_{\pm 0.029}$ |
| | Clam-SB | $0.925_{\pm 0.085}$ | $0.969_{\pm 0.024}$ | $0.523_{\pm 0.020}$ | $0.846_{\pm 0.020}$ | $0.631_{\pm 0.034}$ | $0.863_{\pm 0.005}$ |
| | TransMIL | $0.922_{\pm 0.019}$ | $0.943_{\pm 0.009}$ | $0.554_{\pm 0.048}$ | $0.792_{\pm 0.029}$ | $0.631_{\pm 0.030}$ | $0.841_{\pm 0.006}$ |
| | DSMIL | $0.943_{\pm 0.007}$ | $0.966_{\pm 0.009}$ | $0.532_{\pm 0.064}$ | $0.804_{\pm 0.032}$ | $0.577_{\pm 0.028}$ | $0.816_{\pm 0.028}$ |
| | DTFD-MIL | $0.948_{\pm 0.007}$ | $\underline{0.980}_{\pm 0.011}$ | $0.627_{\pm 0.015}$ | $0.866_{\pm 0.012}$ | $0.612_{\pm 0.080}$ | $0.870_{\pm 0.022}$ |
| | IBMIL | $0.912_{\pm 0.034}$ | $0.954_{\pm 0.022}$ | $0.557_{\pm 0.064}$ | $0.850_{\pm 0.024}$ | $0.645_{\pm 0.041}$ | $0.871_{\pm 0.014}$ |
| | MHIM-MIL | $0.932_{\pm 0.024}$ | $0.970_{\pm 0.037}$ | $0.541_{\pm 0.022}$ | $0.845_{\pm 0.026}$ | $0.625_{\pm 0.060}$ | $0.865_{\pm 0.017}$ |
| | ACMIL | $0.954_{\pm 0.012}$ | $0.974_{\pm 0.012}$ | $0.562_{\pm 0.050}$ | $0.863_{\pm 0.004}$ | $0.722_{\pm 0.030}$ | $0.888_{\pm 0.010}$ |
| | CAMIL | $0.930_{\pm 0.009}$ | $0.963_{\pm 0.011}$ | $0.633_{\pm 0.022}$ | $0.886_{\pm 0.034}$ | $0.709_{\pm 0.011}$ | $0.836_{\pm 0.014}$ |
| | AEM | $0.947_{\pm 0.003}$ | $0.974_{\pm 0.007}$ | $\underline{0.647}_{\pm 0.007}$ | $\underline{0.887}_{\pm 0.013}$ | $\underline{0.742}_{\pm 0.030}$ | $\underline{0.905}_{\pm 0.010}$ |
| | HDMIL | $\underline{0.958}_{\pm 0.013}$ | $0.976_{\pm 0.017}$ | $0.571_{\pm 0.012}$ | $0.796_{\pm 0.022}$ | $0.717_{\pm 0.033}$ | $0.874_{\pm 0.010}$ |
| | ASMIL (Ours) | $\mathbf{0.965}_{\pm 0.020}$ | $\mathbf{0.985}_{\pm 0.017}$ | $\mathbf{0.689}_{\pm 0.005}$ | $\mathbf{0.898}_{\pm 0.010}$ | $\mathbf{0.781}_{\pm 0.042}$ | $\mathbf{0.914}_{\pm 0.014}$ |

### Localization Results




## References

[1] Maximilian Ilse, Jakub Tomczak, and Max Welling. Attention-based deep multiple instance learning. In *International conference on machine learning*, pages 2127–2136. PMLR, 2018.

{ linfeng.ye, zhixiang.chi }@mail.utoronto.ca & kostas@ece.utoronto.ca

ICLR 2026; Rio de Janeiro, Brazil

{ smohajer,pilanci } @ stanford.edu & { guang, ogawa, mhaseyama }@lmd.ist.hokudai.ac.jp