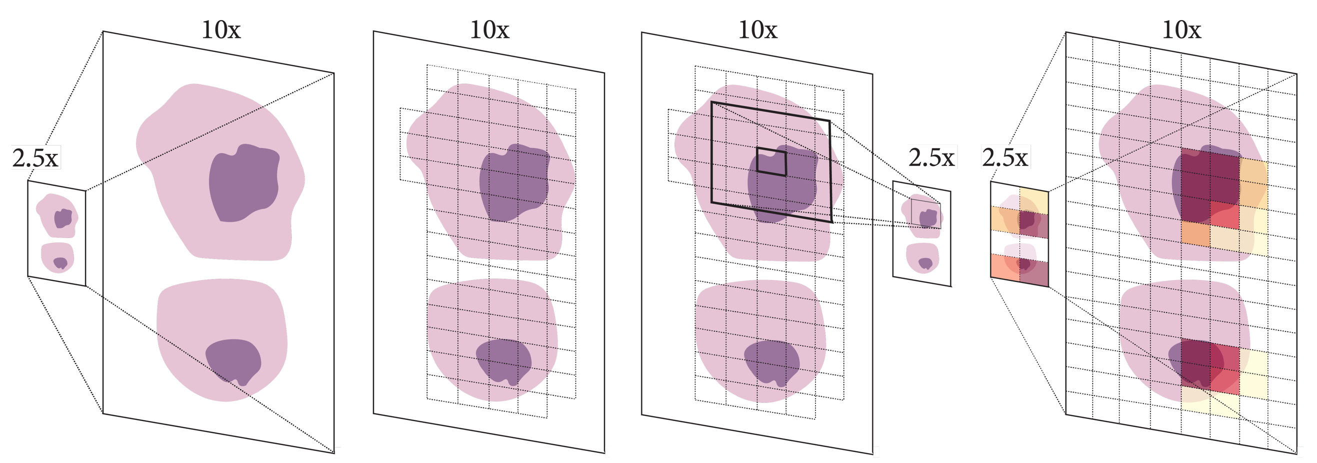
Differentiable Zooming for Multiple Instance Learning on Whole-Slide Images

Kevin Thandiackal[1,2]*, Boqi Chen[1,2]*, Pushpak Pati[1], Guillaume Jaume[3], Drew F. K. Williamson[3], Maria Gabrani[1], and Orcun Goksel[2,4]
[1]Accelerated Discovery and AI, IBM Research Europe; [2]Computer-assisted Applications in Medicine, ETH Zurich;
[3]Dept. of Pathology, Brigham and Women's Hospital, Harvard Medical School; [4]Dept. of Information Technology, Uppsala University
* Contributed equally
github.com/histocartography/zoommil
kth@zurich.ibm.com

1 Background

- Pathologists examine whole-slide images (WSIs) in a hierarchical manner, detecting diagnostically informative regions at a low magnification and examining these regions at a high magnification
- Most existing WSI classification methods build on multiple instance learning (MIL) to process all image patches in a WSI, either at a single magnification (**lack of context**) or across multiple magnifications (**computationally expensive**)
- We propose **ZoomMIL**: a multi-scale, context-aware MIL method that learns to zoom for highly efficient (up to 40x faster computation time) WSI classification



(a) Pathologist (b) Single-scale MIL (c) Multi-scale MIL (d) ZoomMIL (ours)

2 Idea

- First identify informative patches at a low magnification, then zoom in on the selected patches at high magnification
- Instead of relying on a handcrafted patch selection strategy, employ a differentiable Top-K module[1] to learn identifying the most relevant patches
- Aggregate information across multiple magnifications to obtain a context-aware WSI representation tailored to the downstream task

References
1. Cordonnier, J. et al.: Differentiable patch selection for image recognition. CVPR (2021)
2. Ilse, M. et al.: Attention-based deep multiple instance learning. ICML (2018)
3. Lerousseau, M. et al.: Sparseconvmil: Sparse convolutional context-aware multiple instance learning for whole slide image classification. MICCAI Workshop on Computational Pathology (2021)
4. Lu, M. et al.: Data efficient and weakly supervised computational pathology on whole slide images. Nature Biomedical Engineering 5 (2021)
5. Shao, Z. et al.: TransMIL: Transformer based correlated multiple instance learning for whole slide image classification. NeurIPS (2021)
6. Hashimoto, N. et al.: Multi-scale domain-adversarial multiple-instance cnn for cancer subtype classification with unannotated histopathological images. CVPR (2020)
7. Li, B. et al.: Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning. CVPR (2021)
8. Brancati, N. et al.: BRACS: A dataset for breast carcinoma subtyping in H&E histology images. arXiv (2021)
9. Bejnordi, B. et al.: Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. JAMA 318 (2017)

3 Method: ZoomMIL

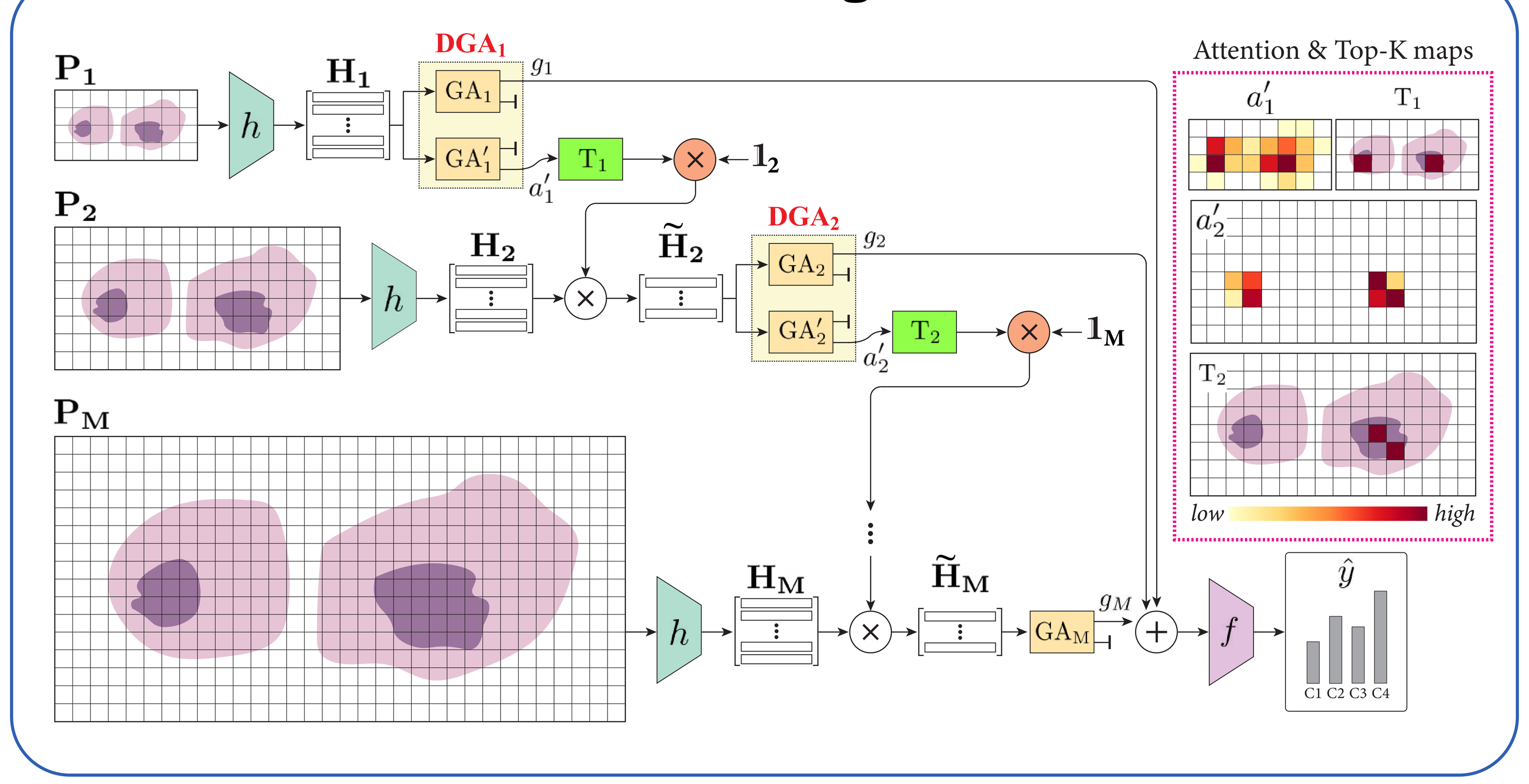
- A WSI is split into a bag of patch features $\mathbf{H} = [\mathbf{h}_1 \ \dots \mathbf{h}_N]^\mathrm{T} \in \mathbb{R}^{N\times D}$
- The WSI-level representation is obtained through gated attention pooling[2]: $\mathbf{g}(\mathbf{H}) = \sum_{i=1}^{N} a_i \mathbf{h}_i$, with an attention weight a_i per patch
- Using a differentiable Top-K module, compute the indices of the K patches with the highest attention weights

$$\mathbf{T}_m = \mathbb{E}\left[\underset{\hat{\mathbf{T}}}{\operatorname{argmax}} \langle\, \hat{\mathbf{T}}\,, (\mathbf{a}_m + \sigma\,\mathbf{Z})\,\mathbf{1}^\mathrm{T}\,\rangle\right]$$

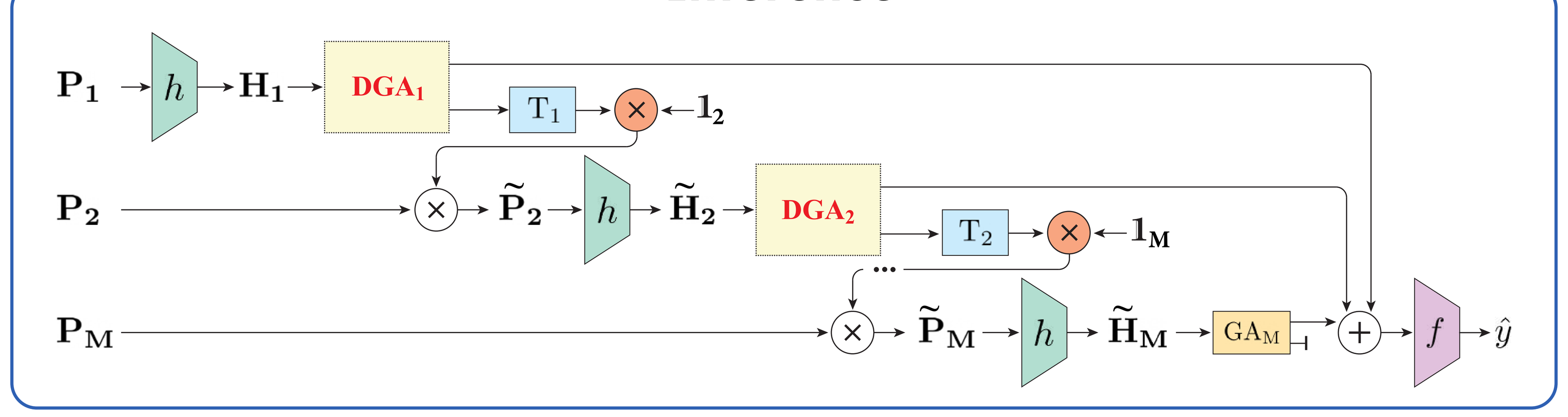
- The resulting selector matrix $\mathbf{T}_m$ then selects the K most informative patches at magnification m: $\widetilde{\mathbf{H}}_m = \mathbf{T}_m^\mathrm{T}\,\mathbf{H}_m$
- The final classifier $f(\cdot)$ computes the predicted label $\hat{\mathbf{y}}$ by sum-pooling WSI-level representations over all magnifications:

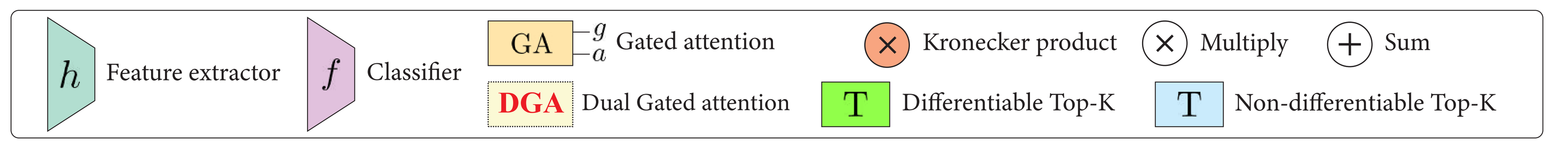
$$\hat{\mathbf{y}} = f(\mathbf{g}_1(\mathbf{H}_1) + \mathbf{g}_2(\widetilde{\mathbf{H}}_2) + \ \dots + \mathbf{g}_M(\widetilde{\mathbf{H}}_M))$$

Training



Inference



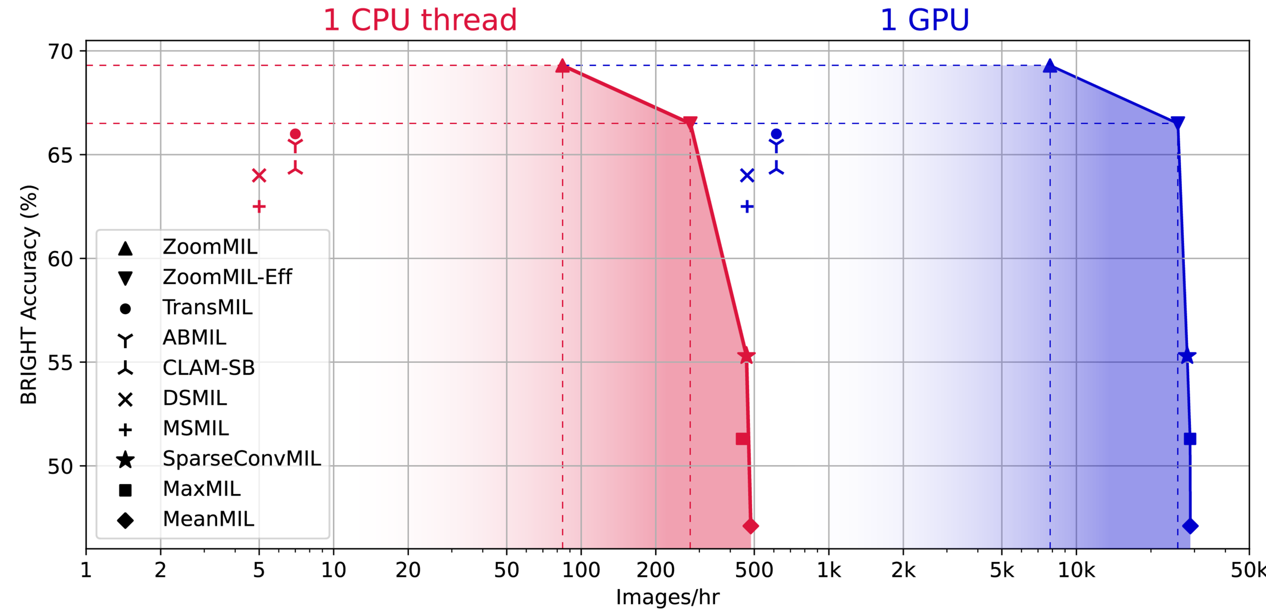


4 Results

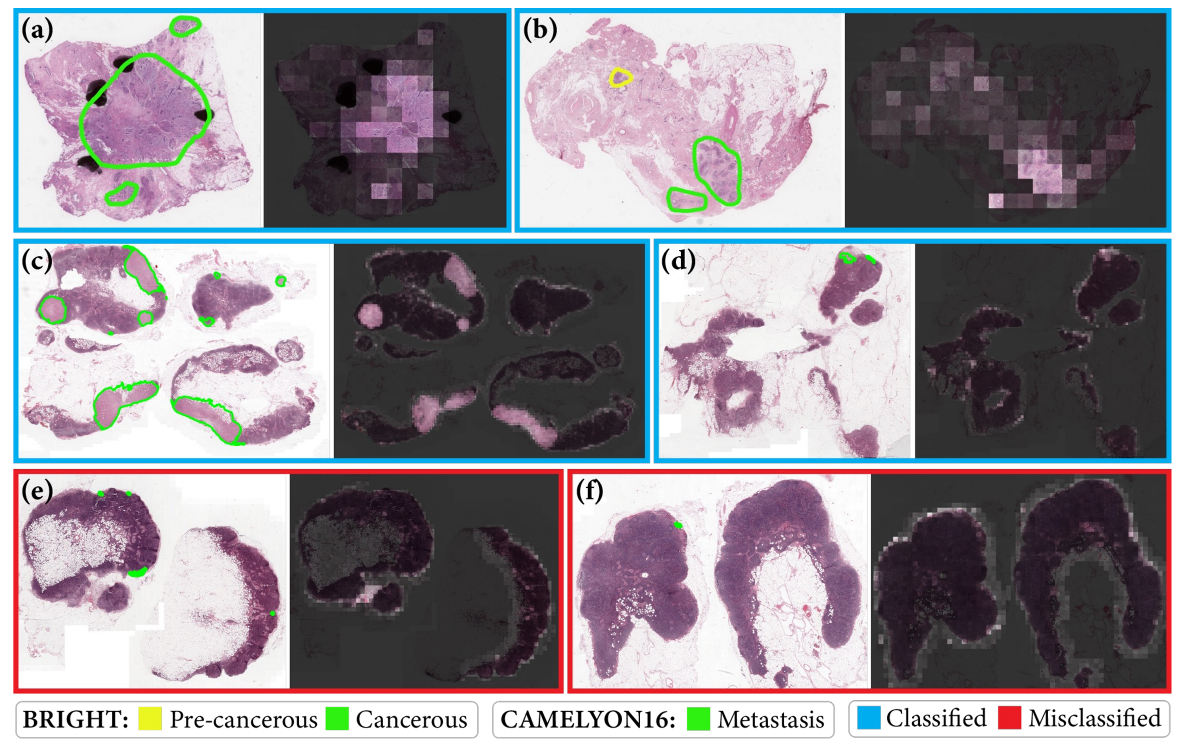
- SOTA results for breast cancer subtyping on the BRIGHT[8] dataset as well as two other WSI classification datasets

Methods	Classification		Computation	
	Weighted F1	Accuracy	TFLOPs	Time (s)
SparseConvMIL[3] (10x)	53.2 ± 3.6	55.3 ± 3.7	0.96	0.13
ABMIL[2] (10x)	63.5 ± 2.7	65.5 ± 1.9	16.45	5.86
CLAM-SB[4] (10x)	63.1 ± 1.7	64.3 ± 1.7	16.45	5.86
TransMIL[5] (10x)	65.5 ± 2.8	66.0 ± 2.7	16.46	5.86
MSMIL[6] (1.25x + 2.5x + 10x)	61.7 ± 0.6	62.5 ± 1.1	21.59	7.69
DSMIL[7] (1.25x + 2.5x + 10x)	63.1 ± 1.6	64.0 ± 1.1	21.66	7.69
ZoomMIL-Eff (1.25x → 2.5x)	66.0 ± 1.9	66.5 ± 1.5	0.40	0.14
ZoomMIL (1.25x → 2.5x → 10x)	**68.3 ± 1.1**	**69.3 ± 1.0**	1.29	0.46

- Best throughput-accuracy trade-off on CPU and GPU



- Annotated tumor regions and ZoomMIL's attention maps for WSIs from the (a, b) BRIGHT and (c-f) CAMELYON16[9] datasets



October 23-27, 2022, Tel Aviv