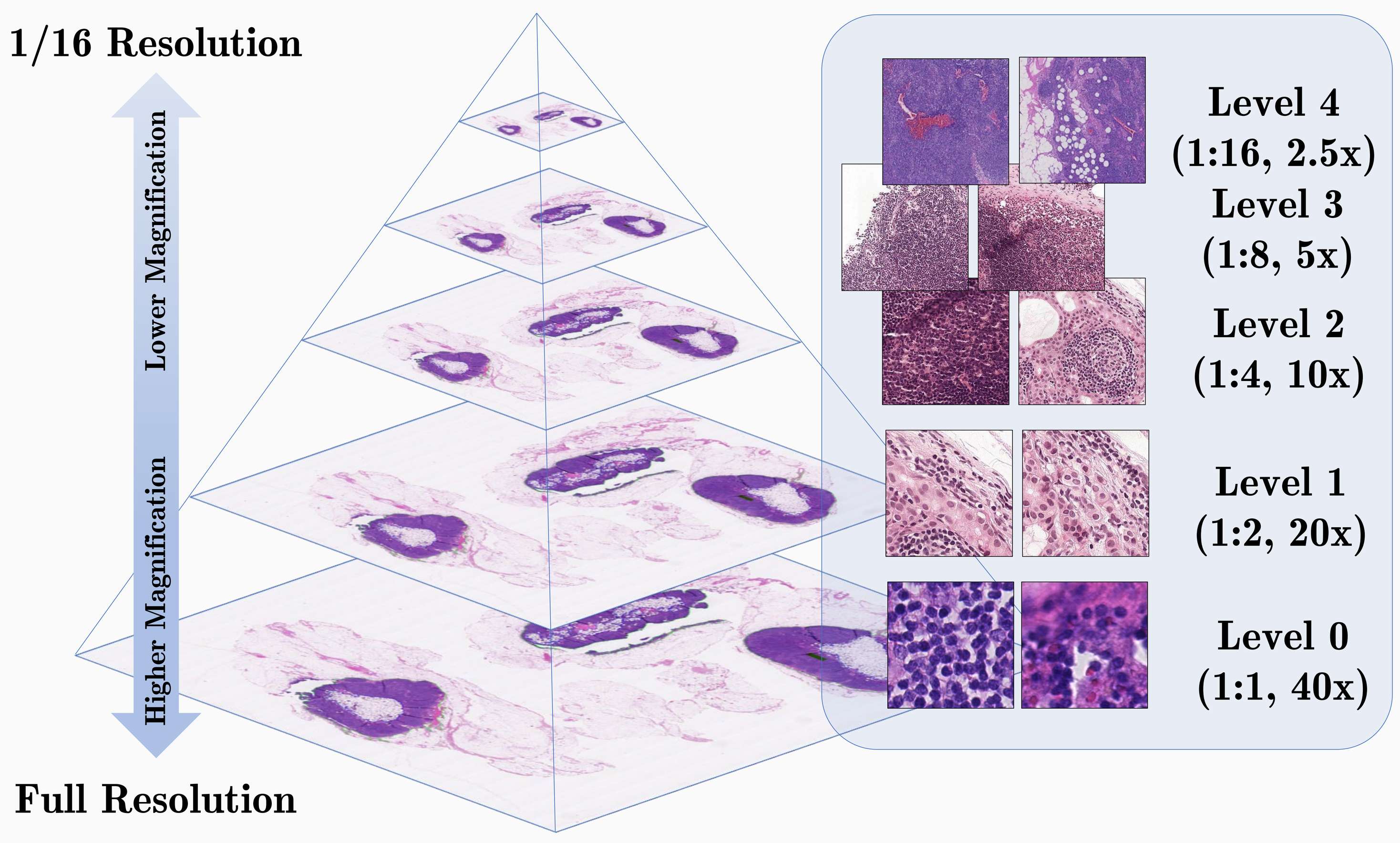
Buffer-MIL: Robust Multi-instance Learning with a Buffer-based Approach

Gianpaolo Bontempo[1,2], Luca Lumetti[1], Angelo Porrello[1], Federico Bolelli[1], Simone Calderara[1], and Elisa Ficarra[1]

[1]University of Modena and Reggio Emilia, Italy
{*name.surname*}@unimore.it
[2]University of Pisa, Italy
{*name.surname*}@phd.unipi.it

Introduction

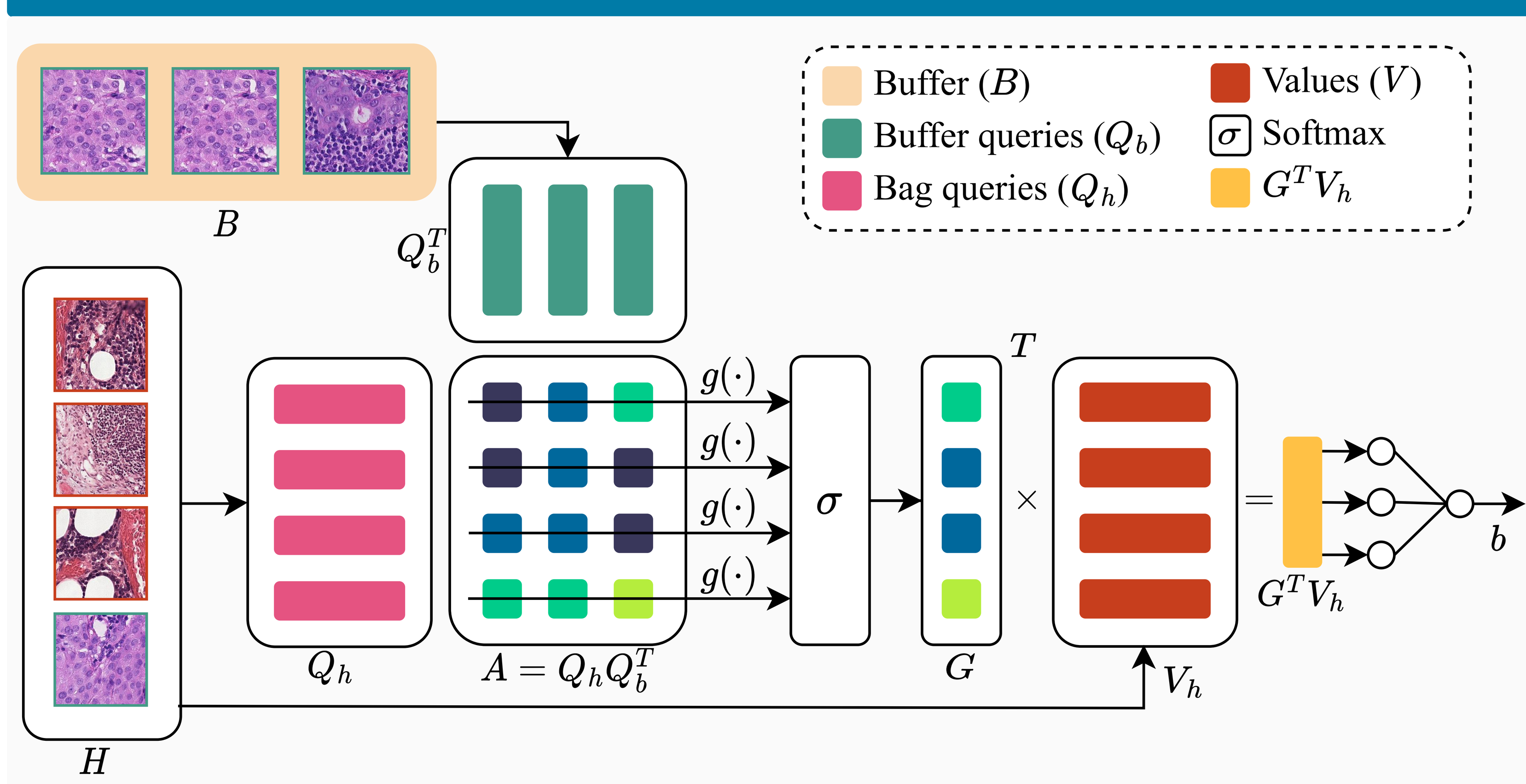


Whole-Slide Images (WSIs) present challenges for deep learning frameworks due to their large size and lack of pixel-level annotations.
Multi-Instance Learning (MIL) approaches consider the image slide as a bag composed of many patches, called instances; afterward, they weigh the instances through attention mechanisms and aggregate them into a single representation to provide a classification score for the entire bag.

Problem Statement

Class Imbalance: positive instances usually represent a low percentage of the entire set. The model will tend to overfit and might misclassify positive instances.
Covariate Shift: It occurs when the distribution of instances within positive and negative bags differs between train and test data.

Proposed Architecture



Buffer containing the most critical patches over the entire trainset is used as an anchor for the attention mechanism.

Results

	Camelyon16		TCGA Lung	
Model	Accuracy	AUC	Accuracy	AUC
mean-pooling	0.723 ± 0.004	0.672 ± 0.100	0.823 ± 0.002	0.905 ± 0.001
max-pooling	0.893 ± 0.015	0.899 ± 0.007	0.851 ± 0.008	0.909 ± 0.002
AB-MIL [1]	0.724 ± 0.015	0.744 ± 0.016	0.864 ± 0.009	0.933 ± 0.004
DSMIL [2]	0.915 ± 0.013	0.952 ± 0.005	0.888 ± 0.005	0.951 ± 0.002
BUFFER-MIL	0.935 ± 0.012	0.971 ± 0.005	0.891 ± 0.008	0.950 ± 0.002
DS-MIL [1]	0.909 ± 0.020	0.955 ± 0.010	$\mathbf{0.913 \pm 0.005}$	$\mathbf{0.966 \pm 0.002}$
BUFFER-MIL	$\mathbf{0.940 \pm 0.008}$	$\mathbf{0.969 \pm 0.005}$	0.897 ± 0.020	0.956 ± 0.010

Ablations

Mean vs Max

Agg.	N/slide	Accuracy	AUC
Mean	1	0.934 ± 0.012	0.970 ± 0.006
	2	0.932 ± 0.012	0.968 ± 0.006
	10	$\mathbf{0.935 \pm 0.012}$	$\mathbf{0.971 \pm 0.005}$
Max	1	0.925 ± 0.012	0.966 ± 0.004
	2	0.927 ± 0.020	0.967 ± 0.005
	10	0.930 ± 0.021	0.967 ± 0.003

Results reveal that producing the final attention scores by averaging critical representations in the buffer outperforms the use of a max operator.

Buffer Update

If updated too frequently, the buffer may have negative effects.

Freq.	N/slide	Accuracy	AUC
1	10	0.919 ± 0.012	0.963 ± 0.004
2	10	0.917 ± 0.009	0.967 ± 0.001
10	10	$\mathbf{0.935 \pm 0.012}$	$\mathbf{0.971 \pm 0.005}$

Sampling Strategy

	Our Method		Reservoir Sampling	
N/slide	Accuracy	AUC	Accuracy	AUC
1	0.934 ± 0.012	0.970 ± 0.006	0.922 ± 0.014	0.962 ± 0.003
2	0.932 ± 0.012	0.968 ± 0.006	0.922 ± 0.008	0.963 ± 0.004
10	$\mathbf{0.935 \pm 0.012}$	$\mathbf{0.971 \pm 0.005}$	0.925 ± 0.012	0.964 ± 0.004

Critical patches have a better effect than a Reservoir (Random) Strategy.

References

[1] Li, B., Li, Y., & Eliceiri, K. W. (2021). Dual-stream Multiple Instance Learning Network for Whole Slide Image Classification with Self-supervised Contrastive learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 14318-14328).
[2] Ilse, M., Tomczak, J., & Welling, M. (2018). Attention-based Deep Multiple Instane Learning. In International Conference on Machine Learning (pp. 2127-2136). PMLR.