# Supplementary material: Dual-Query Multiple Instance Learning for Dynamic Meta-Embedding based Tumor Classification

Simon Holdenried-Krafft[1]
simon.krafft@uni-tuebingen.de
Peter Somers[3]
somers@isys.uni-stuttgart.de
Ivonne A. Montes-Majarro[2]
ivonne.montes@med.uni-tuebingen.de
Diana Silimon[2]
dianasilimon@icloud.de
Cristina Tarín[3]
cristina.tarin-sauer@isys.uni-stuttgart.de
Falko Fend[2]
falko.fend@med.uni-tuebingen.de
Hendrik P. A. Lensch[1]
hendrik.lensch@uni-tuebingen.de

[1] Institute for Computer Graphics,
University of Tübingen,
Tübingen, Germany
[2] Institute of Pathology and
Neuropathology,
University Hospital of Tübingen,
Tübingen, Germany
[3] Institute for System Dynamics,
University of Stuttgart,
Stuttgart, Germany

## 1 Datasets

The experimental setup in this work utilizes three publicly available histopathological datasets: Camelyon16 [1], The Cancer Genome Atlas (TCGA) Breast Invasive Carcinoma (BRCA) [13], and the TCGA Urothelial Bladder Carcinoma (BLCA) [11]. This section highlights the purposes of each dataset, the curation, and the pre-processing procedure. As mentioned in the main part of this work, all patches are extracted at 20× magnification in a non-overlapping manner with a size of 256×256.

### 1.1 Camelyon16

The Camelyon16 dataset [1] consists of 399 hematoxylin and eosin (H&E) stained lymph node sections, scanned and stored as whole-slide images (WSIs). Each slide is fully annotated and permits pixel-wise detection of breast cancer metastasis. We focus on slide-level cancer classification in our weakly supervised setup and ignore the pixel-wise annotations. The WSIs are labeled as "tumor" as soon as they incorporate annotated cancerous regions, otherwise they are "normal". We follow the official dataset split with 270 training samples (110 tumor, 160 normal) and 129 test samples (49 tumor, 80 normal). During pre-processing, we combine threshold-based filtering [9] with a pre-trained U-Net [10] for tissue segmentation, yielding about 11,500 patches per slide.

### 1.2 TCGA-BRCA

The TCGA-BRCA [13] contains 1,133 diagnostics digital H&E slides of invasive breast cancer and is made available by the National Cancer Institute (NCI) Genomic Data Commons (GDC) [4]. The dataset covers 15 histological types and can be augmented with additional modalities such as genomic data. Following the experimental design of Chen et al. [3], we focus on classifying the two most frequent histological types of breast cancer: invasive ductal carcinoma (IDC) and invasive lobular carcinoma (ILC). We apply a stratified data split with a ratio of 80:20 (training:test) on the patient-level, which leads to 698 training samples (578 IDC, 120 ILC) and 177 test samples (148 IDC, 29 ILC). As the WSIs do not contain a 20× magnification, we extract patches of size 512 at magnification 40× and apply a downsampling operation of factor 2 to acquire patches of size 256×256. The remaining steps during pre-processing are the same as described in Section 1.1, leading to roughly 11,000 patches per slide.

### 1.3 TCGA-BLCA

The TCGA-BLCA dataset [11] is also published by the NCI GDC [4] and comprises 449 labeled diagnostic H&E WSIs of muscle-invasive bladder cancer (MIBC). In our experiments, we intend to classify the slides into two histological types: papillary MIBC and non-papillary MIBC. We exploit the same procedure as in Section 1.2 and apply a patient-level data split with 80% training cases (351 WSIs) and 20% test cases (98 WSIs). After pre-processing, we acquire approximately 16,500 patches per slide.

## 2 Implementation Details

To train the DQ-MIL architecture, a self-distillation loss $\mathcal{L}_{SD}$, inspired by Zhang et al. [14, 15], is utilized and combined with a Lookahead RAdam optimizer [8, 16]. For all experiments, a learning rate of $2 \times 10^{-4}$ and a weight decay of $10^{-5}$ is used [12]. The mini-batch during training is set to one bag-of-instances (1 WSI). Following Jaegle et al. [6], a truncated normal distribution with $\mu = 0$, $\sigma = 0.02$, and truncation bounds of [-2, 2] is used to randomly initialize the latent representations ($\mathbf{Q_1}$, $\mathbf{Q_2}$). The hyper-parameter setting of the DQ-MIL architecture used for the experiments, results in a computational complexity of 25 GFLOPS, which is decreased compared to TransMil [12] with 40 GFLOPS and DS MIL [7] with 45 GFLOPS.

## 3 Ablation Study

### 3.1 Temperature-Based Instance Masking

Motivated by the results of the Iterative Patch Selection (IPS) module [2], which condenses a bag into its M most salient instances, we conduct an ablation study to explore the potential of temperature $\tau$ for implicit instance masking. As shown in the main section of this work, the general attention operation, based on queries $\mathbf{Q}$, keys $\mathbf{K}$, values $\mathbf{V}$, and temperature $\tau$, can be expressed as:

$$Attention(\mathbf{Q},\mathbf{K},\mathbf{V}) = softmax\left(\frac{\mathbf{QK}^T}{\tau}\right)\mathbf{V}. \tag{1}$$

In standard self-attention, $\tau$ serves to decouple the attention scores from the inner channel dimension $d_k$. Therefore, parameter $\tau$ is given by $\tau = \sqrt{d_k}$. In contrast to Bergner et al. [2], our idea is not to reduce the computational burden. We aim to sharpen the training signal by implicitly masking out less significant instances. Therefore, we decrease the temperature $\tau$ to collapse the probability distribution to the most essential instances. To explore the effect of this approach, we conducted experiments with various values for $\tau$. The results are shown in Table 1.

| Temperature | Camelyon16 | | TCGA-BRCA | | TCGA-BLCA | |
|---|---|---|---|---|---|---|
| | AUC | Accuracy | AUC | Accuracy | AUC | Accuracy |
| $\tau = \sqrt{d_k} = 8$ | $\underline{0.9594}$ | **0.9457** | **0.9441** | **0.9266** | **0.8462** | **0.9184** |
| $\tau = 1$ | 0.9487 | **0.9457** | $\underline{0.9369}$ | $\underline{0.9039}$ | $\underline{0.8452}$ | 0.8061 |
| $\tau = 1/8$ | 0.9556 | 0.9380 | 0.9306 | 0.8249 | 0.8081 | 0.7959 |
| $\tau = 1/16$ | **0.9651** | **0.9457** | 0.9359 | 0.8531 | 0.8027 | **0.9184** |

Table 1: Comparison of different temperature values, evaluated with a fixed DQ-MIL-SD aggregation model.

Although we achieve an improvement of the AUC metric on Camelyon16, which resonates with the insights from Bergner et al. [2], the potential of implicit instance selection using temperature $\tau$ is limited. Collapsing the probability distributions by decreasing $\tau$ seems only beneficial for unbalanced bags-of-instances, given in the Camelyon16 dataset. For other tasks, such as histological subtyping, temperature-based instance masking may even be detrimental to the overall performance. An alternative approach could be to convert the hyperparameter $\tau$ into a trainable parameter [5].

# References

[1] Babak Ehteshami Bejnordi, Mitko Veta, Paul Johannes Van Diest, Bram Van Ginneken, Nico Karssemeijer, Geert Litjens, Jeroen A.W.M. Van Der Laak, et al. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. *JAMA*, 318(22):2199–2210, dec 2017. ISSN 0098-7484. doi: 10.1001/JAMA.2017.14585.

[2] Benjamin Bergner, Christoph Lippert, and Aravindh Mahendran. Iterative patch selection for high-resolution image recognition. In *The Eleventh International Conference on Learning Representations*, 2 2023.

[3] Richard J. Chen, Chengkuan Chen, Yicong Li, Tiffany Y. Chen, Andrew D. Trister, Rahul G. Krishnan, and Faisal Mahmood. Scaling vision transformers to gigapixel im-

ages via hierarchical self-supervised learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16144–16155, June 2022.

[4] Robert L. Grossman, Allison P. Heath, Vincent Ferretti, Harold E. Varmus, Douglas R. Lowy, Warren A. Kibbe, and Louis M. Staudt. Toward a Shared Vision for Cancer Genomic Data. *The New England journal of medicine*, 375(12):1109–1112, sep 2016. ISSN 1533-4406. doi: 10.1056/NEJMP1607591.

[5] Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger. On Calibration of Modern Neural Networks. *34th International Conference on Machine Learning, ICML 2017*, 3:2130–2143, jun 2017.

[6] Andrew Jaegle, Felix Gimeno, Andy Brock, Oriol Vinyals, Andrew Zisserman, and Joao Carreira. Perceiver: General perception with iterative attention. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 4651–4664. PMLR, 18–24 Jul 2021.

[7] Bin Li, Yin Li, and Kevin W. Eliceiri. Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 14313–14323, 11 2020. ISSN 10636919. doi: 10.48550/arxiv.2011.08939.

[8] Liyuan Liu, Haoming Jiang, Pengcheng He, Weizhu Chen, Xiaodong Liu, Jianfeng Gao, and Jiawei Han. On the Variance of the Adaptive Learning Rate and Beyond. aug 2019.

[9] 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. *Nature Biomedical Engineering*, 5:555–570, 4 2020. ISSN 2157846X. doi: 10.48550/arxiv.2004.09666.

[10] Abtin Riasatian, Maral Rasoolijaberi, Morteza Babaei, and H. R. Tizhoosh. A comparative study of u-net topologies for background removal in histopathology images. *Proceedings of the International Joint Conference on Neural Networks*, 6 2020. doi: 10.1109/IJCNN48605.2020.9207018.

[11] A. Gordon Robertson, Jaegil Kim, Hikmat Al-Ahmadie, Joaquim Bellmunt, Guangwu Guo, Andrew D. Cherniack, Toshinori Hinoue, et al. Comprehensive Molecular Characterization of Muscle-Invasive Bladder Cancer. *Cell*, 171(3):540–556.e25, oct 2017. ISSN 10974172. doi: 10.1016/J.CELL.2017.09.007/ATTACHMENT/FAD0F071-6C7E-4BA4-A5F1-E8F84BF6BB7C/MMC3.XLSX.

[12] 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. *Advances in Neural Information Processing Systems*, 3:2136–2147, 6 2021. ISSN 10495258. doi: 10.48550/arxiv.2106.00908.

[13] Aatish Thennavan, Francisco Beca, Youli Xia, Susana Garcia-Recio, Kimberly Allison, Laura C. Collins, Gary M. Tse, et al. Molecular analysis of TCGA breast cancer histologic types. *Cell Genomics*, 1(3):100067, dec 2021. ISSN 2666-979X. doi: 10.1016/J.XGEN.2021.100067.

[14] Linfeng Zhang, Jiebo Song, Anni Gao, Jingwei Chen, Chenglong Bao, and Kaisheng Ma. Be your own teacher: Improve the performance of convolutional neural networks via self distillation. In *Proceedings of the IEEE International Conference on Computer Vision*, volume 2019-Octob, pages 3712–3721, 2019. ISBN 9781728148038. doi: 10.1109/ICCV.2019.00381.

[15] Linfeng Zhang, Chenglong Bao, and Kaisheng Ma. Self-Distillation: Towards Efficient and Compact Neural Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(8):4388–4403, aug 2022. ISSN 19393539. doi: 10.1109/TPAMI.2021.3067100.

[16] Michael R. Zhang, James Lucas, Geoffrey Hinton, and Jimmy Ba. Lookahead Optimizer: k steps forward, 1 step back. *Advances in Neural Information Processing Systems*, 32, jul 2019. ISSN 10495258.