Low-Shot Prompt Tuning for Multiple Instance Learning based Histology Classification - Supplementary Material

Philip Chikontwe[1], Myeongkyun Kang[2], Miguel Luna[2], Siwoo Nam[2], and Sang Hyun Park[2,3*]

[1] Department of Biomedical Informatics, Harvard Medical School, MA, USA
philip_chikontwe@hms.harvard.edu

[2] Department of Robotics and Mechatronics Engineering, Daegu Gyeongbuk Institute of Science and Technology, Daegu, Republic of Korea
shpark13135@dgist.ac.kr

[3] Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, PA, USA
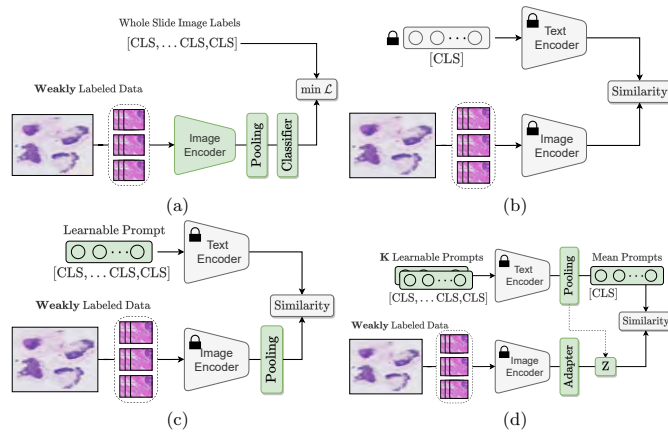
Fig. S1: Illustration of different learning paradigms. (a) standard model fine-tuning with labels only. (b) zero-shot inference per patch with pre-defined prompts. (c) prompt tuning in the weakly supervised setting. (d) our proposed method using visual adaptation and K learnable prompts-driven visual feature pooling for optimization.

---
[*] Corresponding author.

Positive Whole Slide Instances

	Adipose↓	Backg.↓	Normal↓	Tumor↑	Tumor↑
AbMIL	0.70	0.58	0.34	0.53	0.40
CoOp	0.38	0.62	0.12	0.83	0.80
Ours	0.10	0.23	0.05	0.92	0.91

Negative Whole Slide Instances

	Normal↓	Adipose↓	Backg.↓	Debris↓	Debris↓
AbMIL	0.29	0.67	0.55	0.52	0.34
CoOp	0.19	0.29	0.26	0.32	0.21
Ours	0.06	0.08	0.06	0.20	0.06

Fig. S2: Illustration of instance-level scoring on the NCT test dataset sample bags (*i.e.*, positive and negative) for different methods trained in the 16-shot setting. For each tissue type in the set, compared methods assign high scores to non-cancerous tissues (*e.g.*, *Adipose*), whereas the proposed was more consistent in each bag.
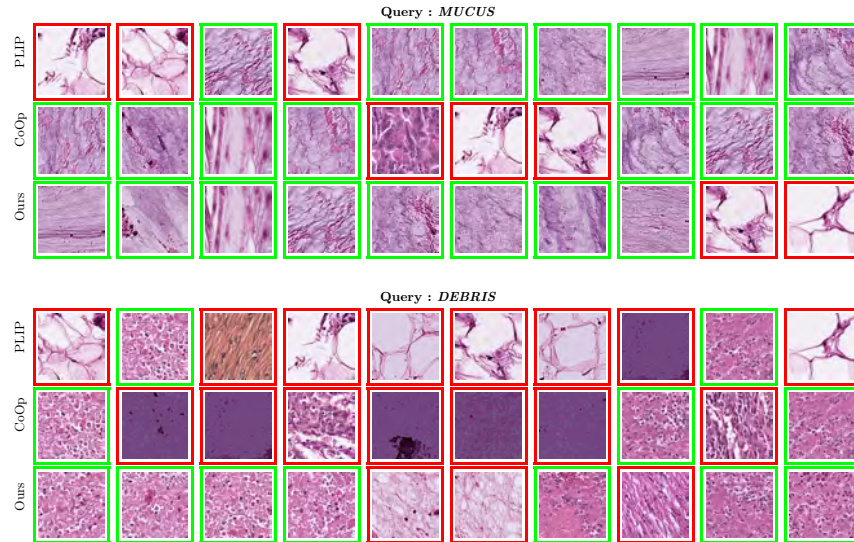

Fig. S3: Illustration of top retrieved images when prompting the tuned models with different classes on NCT dataset (16-shot models) in descending order *i.e.*, left (high) to right (low) scores. Correctly retrieved image class are highlighted by *green* whereas *red* denotes wrong classes. For the query "*debris*" (second section), CoOp failed to retrieve relevant images and showed high similarity scores on others. In contrast, the proposed method showed better retrieval especially for the "*mucus*" query compared to both zero-shot PLIP and CoOp.