Supplementary Materials of SlideGCD: Slide-based Graph Collaborative Training with Knowledge Distillation for Whole Slide Image Classification

Tong Shu[1], Jun Shi[2](✉), Dongdong Sun[1], Zhiguo Jiang[3], and Yushan Zheng[4](✉)

[1] School of Computer Science and Information Engineering, Hefei University of Technology, Hefei 230601, China
[2] School of Software, Hefei University of Technology, Hefei 230601, China
juns@hfut.edu.cn
[3] Image Processing Center, School of Astronautics, Beihang University, Beijing, 102206, China
[4] School of Engineering Medicine, Beijing Advanced Innovation Center on Biomedical Engineering, Beihang University, Beijing 100191, China
yszheng@buaa.edu.cn

Table A.1. The detailed experimental settings of baseline and SlideGCD, where † indicates SlideGCD specific parameters.

Configuration	Detailed setting
General Setting	
Patch size	256 × 256 (following the sliding window strategy)
Patch encoder	PLIP (patch embeddings with 512 dimensions)
Batch size	64
Optimizer	Adam optimizer with CosineAnnealingLR scheduler
Runs of training (Baseline)	100 Epochs training
Runs of training (SlideGCD)	100 Epochs training with 10 epochs of warmup
Size of hyperedge †	12
Size of node buffer †	3072
Distillation temperature †	1.5
Baseline model	ABMIL
Learning rate (Baseline)	5e-4
Learning rate (SlideGCD)	5e-4 in warmup and 1e-4 in formal training
Baseline model	PatchGCN & TransMIL
Learning rate (Baseline)	1e-4
Learning rate (SlideGCD)	1e-4 in both warmup and formal training
Baseline model	DTFDMIL (AFS with 4 pseudo-bags)
Learning rate (Baseline)	1e-4
Learning rate (SlideGCD)	1e-4 in both warmup and formal training