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

Abstract. Existing WSI analysis methods lie on the consensus that histopathological characteristics of tumors are significant guidance for cancer diagnostics. Particularly, as the evolution of cancers is a continuous process, the correlations and differences across various stages, anatomical locations and patients should be taken into account. However, recent research mainly focuses on the inner-contextual information in a single WSI, ignoring the correlations between slides. To verify whether introducing the slide inter-correlations can bring improvements to WSI representation learning, we propose a generic WSI analysis pipeline SlideGCD that considers the existing multi-instance learning (MIL) methods as the backbone and forge the WSI classification task as a node classification problem. More specifically, SlideGCD declares a node buffer that stores previous slide embeddings for subsequent extensive slide-based graph construction and conducts graph learning to explore the inter-correlations implied in the slide-based graph. Moreover, we frame the MIL classifier and graph learning into two parallel workflows and deploy the knowledge distillation to transfer the differentiable information to the graph neural network. The consistent performance boosting, brought by SlideGCD, of four previous state-of-the-art MIL methods is observed on two TCGA benchmark datasets. The code will be available at https://github.com/HFUT-miaLab/SlideGCD.

Keywords: Computational pathology · Whole slide classification · Graph learning · Knowledge distillation.

1 Introduction

Histopathology slides contain significant diagnostic information, which leads the slide screening to an indispensable process for cancer diagnoses [8,25]. With

the development of computational pathology, computer-aided whole slide image (WSI) analysis is greatly achieved for assisting pathologists in making accurate and reproducible diagnoses efficiently. Owing to hardware limitations and the difficulty of obtaining fine-grained annotation [17], the direct process of gigapixel WSIs is nearly impossible or highly expensive. Therefore, computer-aided WSI analysis is usually formulated into the multi-instance learning (MIL) task which considers a WSI as a collection of thousands of patches, analogous to a bag of thousands of instances in the standard multiple instance description. Current patch-based WSI analysis methods focus on how to represent the relationships between patches more comprehensively and efficiently.

The graph-based method is one of the key branches of WSI analysis. Existing graph-based WSI analysis methods follow the same pattern including three stages: 1) patch feature extraction, 2) patch-based graph construction (where one graph describes one slide and nodes represent patches), and 3) graph message passing and graph classification. Compared with the classical MIL method and the sequence-based method, the most significant advantage of the graph-based method lies in the flexibility of the graph construction strategy. Early explorations [4,14,18] directly modeled WSIs as patch-based graphs based on patch similarity in the different spaces. Further research began to involve detailed considerations such as the category of instances [24] and the hierarchical structure of instances [2,7]. Recent works began to combine non-simple graphs (e.g. heterogeneous graph [3] and hypergraph [5,16,22]) with morphological characteristics, e.g. lesion state of cell [3] or tissue topology [5], for histopathology WSI analysis. Although there are some works [6,13,21] have acknowledged the existence of complementary information between the slides from the same patient, the unrestricted slide-level inter-correlations exploration still has not drawn much attention. Since the development of cancers is a continuous process, the correlations and differences across various stages, anatomical locations and patients might also contribute to the slide representation.

Motivated by that consideration, we propose the generic **Slide**-based **G**raph **C**ollaborative training pipeline with knowledge **D**istillation (SlideGCD) for WSI analysis. The intuitive differences between the patch-based graph and the slide-based graph are shown in Fig. 1. More specifically, we take existing MIL methods as the backbone of the proposed SlideGCD for obtaining the initial slide embeddings. Then, SlideGCD is used to explore the slide correlations via the slide-based graph. Finally, the slide-level predictions are obtained by node classification. The main contributions of our work are summarized as follows:

- A generic histopathology WSI analysis pipeline SlideGCD is proposed which can be adapted to any existing MIL methods. This work exploringly frames the WSI dataset into a slide-based graph where WSIs are nodes and mines the slide correlations through graph learning.
- A rehearsal-based graph construction strategy is deployed to describe the inter-correlations between slides adaptively. Besides, collaborative training for the specially designed GCN with knowledge distillation is applied which aims to fully utilize the well-learned knowledge implied in the MIL classifier.
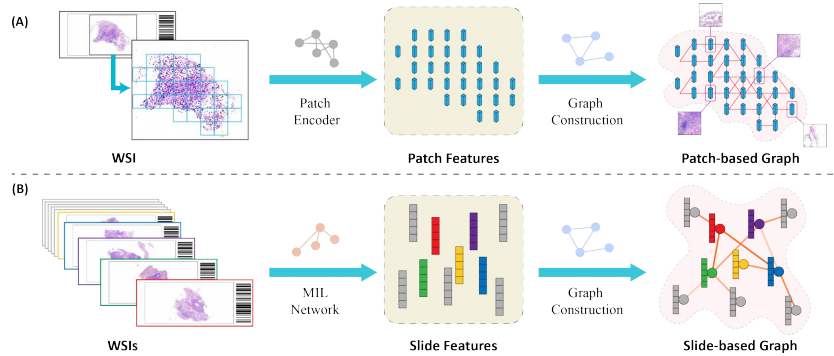
Fig. 1. Diagrams of different types of graph in WSI analysis: (A) the Patch-based Graph & (B) the Slide-based Graph.

2   Method

2.1   Overview

The overall workflow of the proposed SlideGCD is shown in Fig. 2. Similar to the standard MIL framework, SlideGCD first transforms gigapixel WSIs to a series of patch embeddings using a frozen pre-trained patch encoder f. Then, a MIL network so-called backbone is deployed to generate the slide embeddings. After that, the slide embeddings are fed into two branches. One branch is actually the rest of the backbone with only the classifier to ensure the convergence and stability of the network during the early training stage. The other graph-based branch is the specific workflow of SlideGCD which explicitly explores the inter-correlations on the extensive slide-based graph. During the inference stage, the slide-level predictions from the graph-based branch are used as the final predictions.

2.2   Problem Formulation

Assuming there is a dataset with N WSIs denoted by $\mathcal{D} = \{(S_i, y_i)\}_{i=1}^{N}$. Each WSI $S_i = \{p_{i,j}\}_{j=1}^{M_i}$ is annotated with a label $y_i \in \{0, ..., C-1\}$, where $p_{i,j}$ is tiled patch without patch-level label, M_i is the number of patches and C represents the number of categories. Then, there is a pre-trained patch encoder $f(\cdot)$ to transform the patch p_k into D_p dimensional patch embeddings for reducing the computational cost. The goal of WSI classification is to make prediction $\hat{y}_i = g(f(p_{i,1}), ..., f(p_{i,M_i}))$ as close as possible to the ground-truth y_i, where $g(\cdot)$ is an aggregator that aggregates the instances information and makes final predictions.

2.3   Backbone: MIL Network

The proposed SlideGCD aims to explore the inter-correlations between WSIs, and a necessary process is to obtain slide-level representations. We leave this
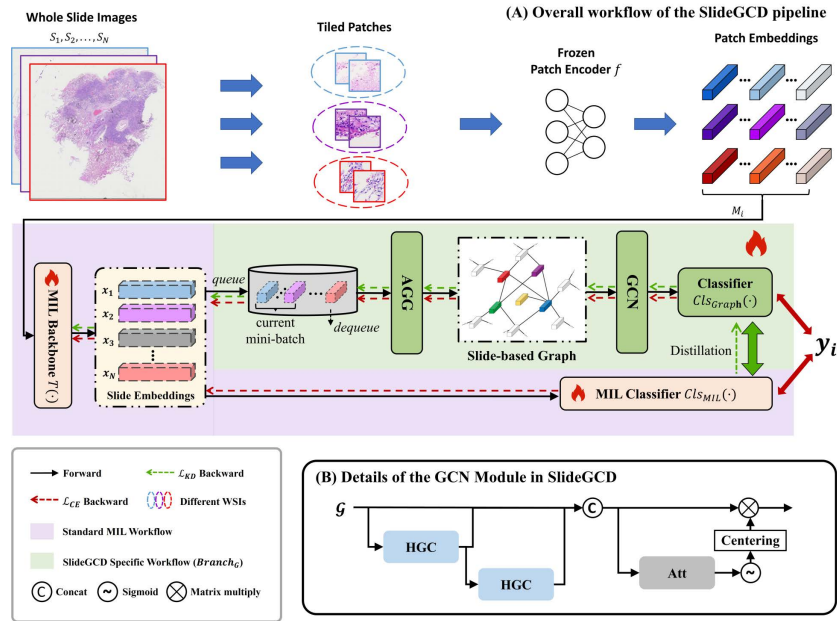
Fig. 2. Illustration of the proposed SlideGCD.

work to the so-called backbone. In an ideal implementation, the backbone $T(\cdot)$: $\mathbb{R}^{N_i \times D_P} \to \mathbb{R}^{D_S}$ can be any MIL methods with any architecture as long as it can generate fixed dimensional (D_S) slide-level embeddings that will be considered the initial node (vertice) embeddings in the subsequent graph-based branch $Branch_G$. Notably, we separate the classifier head from the backbone MIL network and denote it as Cls_{MIL} in the following description. The prediction from Cls_{MIL} is denoted as $\hat{y}_{MIL}$.

For generating stable slide embeddings at the training of SlideGCD, there are a few warmup epochs that only optimize the backbone for pre-convergence. Then, the formal training will start with a smaller learning rate, e.g. 1e-4.

2.4   SlideGCD

In this section, we will introduce and formulate the details of the specific workflow of SlideGCD. Firstly, a rehearsal-based graph construction module is exploited to generate and update a slide-based graph during training adaptively. Then, a specially designed GCN is deployed to exploit the contextual information from the slide-based graph and refine the slide embeddings. Additionally, a collaborative training strategy with knowledge distillation is employed to solve the problem of knowledge misaligned between Cls_{MIL} and $Branch_G$.

Slide-based Graph Construction. Inspired by the idea of Memory Bank [9] and the rehearsal buffer [11], a **Node Buffer** with a length of L $(L > N)$ is designed to store the slide embeddings from the recent training epochs. In each mini-batch training, the current batch of slide embeddings will be pushed into the node buffer and several outdated slide embeddings at the end of the node buffer will be popped out simultaneously (First-in-first-out, like MoCo [9]). With the node buffer, the involved nodes can be expanded to exceed the capacity of datasets. Specifically, when $L > N$, the node buffer will contain at least two different node embeddings related to the same input slide. The extensive node collection will potentially alleviate the graph heterophily [15] since similar nodes are more likely to connect.

Considering a general situation where the clinical information of patients is not thoroughly available, we conduct the Adaptive Graph Generation (AGG) strategy to automatically infer the inter-dependencies from the embedding space. Our AGG module consists of two fully connected layers that transform the slide embeddings into an intermediate hidden space. Then, the k-NN clustering is conducted and the slides belonging to the same cluster are connected by a hyperedge $\mathcal{E}$, which formulates the slides as a hypergraph $\mathcal{G}$. The adaptive graph generation process can be formulated as:

$$\mathcal{G} = (X^{(0)}, \mathcal{E}), \quad \mathcal{E} = KNN(AGG(X^{(0)}), k), \tag{1}$$

where $X^{(0)} \in \mathbb{R}^{L \times D_S}$ represents the node embedding sequence in the Node Buffer that contains the current mini-batch data at the head of the sequence. Notably, the nodes retrieved from the buffer are considered static data with no gradients, and gradient propagation only comes from the nodes from the current mini-batch.

Graph Learning on Slide-based Graph. For the slide-based graph $\mathcal{G}$, a designed GCN composed of two hypergraph convolutional layers [1] and a Centering-Attention module is applied to explore the context implied in the slide-based graph, as:

$$X^{(i+1)} = \text{Leaky\_ReLU}(HGC(X^{(i)}, \mathcal{E})), \tag{2}$$

$$H = Concat(X^{(0)}, X^{(1)}, X^{(2)}), H \in \mathbb{R}^{L \times 3D_S} \tag{3}$$

where $HGC(\cdot)$ denotes the hypergraph convolution [1] and $X^{(i)}$ contains the information accumulated from the node itself to its i-hop neighbors.

To alleviate the graph heterophily, the participation of information from k-hop neighbors should be reweighed. Thus, we applied a channel-wise attention module with $Centering$ to rebalance the message and prevent the attention score from always being positive even when facing defective partial information. The computations can be formulated as:

$$H' = H \cdot Centering(A) = H \cdot (A - Mean(A)), \tag{4}$$

$$A = \text{Sigmoid}(\text{ReLU}(H^{\text{T}}W_0)W_1), \tag{5}$$

where $W_0, W_1 \in \mathbb{R}^{L \times L}$ are the learnable weights. Finally, an MLP classifier Cls_{Graph} is used to make final predictions for current mini-batch:

$$\hat{y}_G = \text{Softmax}(\text{MLP}(H')), \tag{6}$$

Collaborative Training with Knowledge Distillation. The network can exploit the slide inter-correlations on the extensive slide-based graph with the above modifications. However, the well-learned intrinsic knowledge of slides implied in the Cls_{MIL} may be neglected. To associate it with the slide inter-correlations and constrain both branches, we involved the knowledge distillation [10] to transfer the knowledge learned by Cls_{MIL} to the $Branch_G$.

We treat the Cls_{MIL} and the $Branch_G$ as the teacher and student model separately, letting $Branch_G$ draw on the beneficial information learned by Cls_{MIL}. Specifically, a response-based knowledge distillation loss is adopted as:

$$L_{KD} = L_{JS}(\hat{y}_G, \hat{y}_{MIL}, \hat{t}), \tag{7}$$

where L_{JS} denotes the JS divergence [19] loss, and $\hat{t}$ is the temperature coefficient.

Then, the final loss of SlideGCD can be written as below, $L_{CE}(\cdot)$ represents the Cross-Entropy loss function, and $w = 1$ is the weight for knowledge distillation:

$$L = L_{CE}(\hat{y}_{MIL}, Y) + L_{CE}(\hat{y}_G, Y) + w \cdot L_{KD}. \tag{8}$$

3   Experiments

3.1   Experimental Setting

We conducted experiments on two publicly available cancer cohorts (BRCA & NSCLC) derived from The Cancer Genome Atlas (TCGA) for evaluating the performance of SlideGCD in the WSI classification task. **TCGA-BRCA** contains 998 diagnostic digital slides of two breast cancer subtypes. Specifically, 794 WSIs of invasive ductal carcinoma (IDC) and 204 WSIs of invasive lobular carcinoma (ILC). **TCGA-NSCLC** is a collection of two subtype projects for lung cancer, i.e. Lung Squamous Cell Carcinoma (TCGA-LUSC) and Lung Adenocarcinoma (TCGA-LUAD), for a total of 995 diagnostic WSIs, including 496 WSIs of LUSC and 499 WSIs of LUAD. The average accuracy (ACC), macro-average F1 score (F1), and macro-average area under the receiver operating characteristic curve (AUC) are calculated for evaluating the classification performance.

It is worth mentioning that we directly adopted the default setting of hyperparameters to the baseline from its public repository and remained fixed when applying SlideGCD for the fairness of comparison. More experimental setting details are in Supplementary. During inference, all parameters and the Node Buffer are frozen. When a WSI is inputted, 1) its initial embedding will be

made with $T(\cdot)$, 2) the AGG module will insert it into the slide-based graph by connecting it with its k-nearest buffer nodes, 3) the trained GCN will make message passing to refine its embeddings for final classification. We implemented all the models in $PyTorch$ 1.8 and PyG libraries and ran the experiments on a computer with an NVIDIA RTX 3090 GPU.

3.2   Effectiveness for WSI Classification

Table 1 shows the results for the four previous state-of-the-art MIL methods, namely ABMIL [12], PatchGCN [4], TransMIL [20] and DTFDMIL [23], with or without the collaboration of SlideGCD. Each of them represents a typical branch in WSI analysis. ABMIL [12] speaks for the classical lightweight attention-based MIL methods without considering patch relationships. PatchGCN [4] is a typical graph-based MIL method that involves the patch correlation via the patch-based graph. TransMIL [20] is a powerful transformer-based MIL method that exploits the patch correlations by utilizing the self-attention mechanism. DTFDMIL [23] is a novel pseudo-bags-based MIL method that derives the instance probabilities as the attention score of each instance.

Overall, SlideGCN is capable of bringing stable improvement to all four types of baseline. Specifically, on the TCGA-BRCA dataset, SlideGCD-DTFDMIL promotes F1-score from 0.8317 to 0.8519 and improves AUC from 0.9008 to 0.9123. On TCGA-NSCLC dataset, SlideGCD improves DTFDMIL to achieve ACC of 0.8983 (↑3.41%), AUC of 0.9636 (↑0.77%) and F1-score of 0.8982 (↑3.42%). Considering the characteristics of datasets and baselines, SlideGCD prefers pseudo-bag-based DTFDMIL most and the performance boosting of lightweight mod-

Table 1. The performance of SlideGCD on different baselines. All reported results are the means and standard deviations with five-fold cross-validation (CV). †: The discrepancies in reported results from the original paper are mainly caused by the difference in patch encoder and CV settings where the original TransMIL conducts a four-fold CV.

Method\Metrics	TCGA-BRCA Dataset			TCGA-NSCLC Dataset		
	ACC(%)	AUC(%)	F1(%)	ACC(%)	AUC(%)	F1(%)
ABMIL [12] (Baseline)	88.97±0.85	89.87±0.89	81.61±2.27	86.96±1.16	95.14±0.51	86.89±1.20
SlideGCD-ABMIL	89.04±1.11	90.22±1.31	85.53±1.53	89.57±0.77	95.68±0.56	89.56±0.78
Improvement Δ	+0.07	+0.45	**+3.92**	**+2.61**	**+0.54**	**+2.67**
PatchGCN [4] (Baseline)	84.80±1.77	87.18±1.55	75.11±4.57	86.62±2.38	94.81±1.82	86.59±2.43
SlideGCD-PatchGCN	85.13±0.96	87.36±0.96	76.08±1.59	88.29±1.34	95.21±1.08	88.26±1.35
Improvement Δ	+0.33	+0.18	**+0.97**	**+1.67**	**+0.40**	**+1.67**
TransMIL† [20] (Baseline)	88.17±1.00	90.99±0.91	82.09±1.75	85.82±1.67	94.82±1.17	85.77±1.67
SlideGCD-TransMIL	89.37±2.23	91.14±1.48	82.68±3.22	86.82±1.41	95.59±0.27	86.78±1.45
Improvement Δ	**+1.20**	+0.15	**+0.59**	**+1.00**	**+0.77**	**+1.01**
DTFDMIL [23] (Baseline)	89.30±0.44	90.08±0.86	83.17±1.43	86.42±1.07	95.59±0.66	86.40±1.06
SlideGCD-DTFDMIL	90.43±1.65	91.23±1.26	85.19±2.59	89.83±1.80	96.36±0.51	89.82±1.82
Improvement Δ	**+1.13**	**+1.15**	**+2.02**	**+3.41**	**+0.77**	**+3.42**
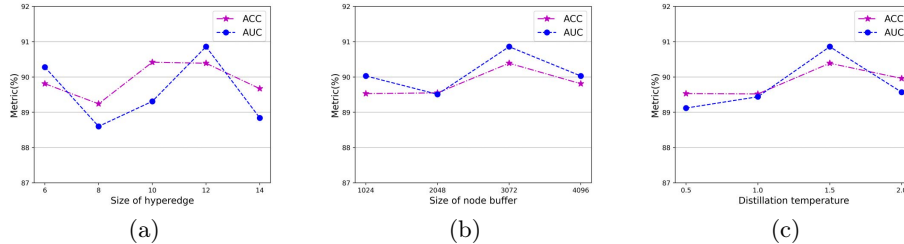
Fig. 3. Hyperparameter studies of (a) Size of hyperedge k, (b) Size of node buffer L and (c) Distillation temperature $\hat{t}$.

els (i.e. ABMIL&PatchGCN) on imbalanced dataset TCGA-BRCA is relatively small. Compared with previous methods that only inspect the slide innateness, SlideGCD additionally looks at the potential connection between the WSIs with similar patterns in tumors. Such consideration exploits the WSI dataset more comprehensively and therefore achieves better performance.

As for the extra computations, they mainly come from several additional linear layers and two graph convolutional layers. Concretely, our method has GFLOPs of 404.37 with a mini-batch consisting of 64 WSIs with 5000 patches, which is only 1.01% higher than the 398.93 GFLOPs of its baseline TransMIL. This increase is minimal.

3.3   Hyperparameter Studies

Ablation experiments on hyper-parameters are conducted on the TCGA-BRCA dataset with the backbone of DTFDMIL. As presented in Fig. 3, the reported results are the means of ACC\AUC on the validation set with five-fold CV.

Size of hyperedge k. k controls the topology of the slide-based graph, and as k increases, there is more overlap between hyperedges. From Fig. 3a, the topology of the slide-based graph could affect the performance of SlideGCD. It does not improve monotonically with the increased size of hyperedge k and seems saturated at $k = 12$. Thus, we chose a size of 12 as the default configuration.

Size of node buffer L. L affects the timeliness of node features in the constructed slide-based graph. When it becomes too large, it can be foreseen that there will be a decrease in performance as the node buffer contains many outdated slide features. As it becomes too small, the improvement shall descend. According to Fig. 3b, we empirically set $L = 3072$.

Distillation temperature $\hat{t}$. In the higher temperature situation (i.e. $\hat{t} > 1$), distillation focuses on transferring knowledge from the teacher model. At lower temperatures (i.e. $\hat{t} < 1$), distillation tends to alleviate the impact of noise in negative samples [10]. Our motivation is to transfer the well-learned knowledge in Cls_{MIL} to $Branch_G$, thus a relatively large temperature coefficient should have a better effect. Following the results in Fig. 3c, we set $\hat{t} = 1.5$.

4   Conclusion

In this paper, we proposed a generic pipeline SlideGCD for histopathology WSI classification and verified the possible improvements of introducing slide correlations via the slide-based graph. SlideGCD exploringly takes the potential connections between slides with similar pathological patterns into account and eventually achieves better performance in WSI classification in a more comprehensive way of utilizing the WSI dataset. Comprehensive experiments have been conducted on various MIL benchmarks and the results show that SlideGCD can boost existing methods consistently.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Bai, S., Zhang, F., Torr, P.H.: Hypergraph convolution and hypergraph attention. Pattern Recognition **110**, 107637 (2021)
2. Bontempo, G., Bolelli, F., Porrello, A., Calderara, S., Ficarra, E.: A graph-based multi-scale approach with knowledge distillation for wsi classification. IEEE Transactions on Medical Imaging (2023)
3. Chan, T.H., Cendra, F.J., Ma, L., Yin, G., Yu, L.: Histopathology whole slide image analysis with heterogeneous graph representation learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 15661–15670 (2023)
4. Chen, R.J., Lu, M.Y., Shaban, M., Chen, C., Chen, T.Y., Williamson, D.F.K., Mahmood, F.: Whole slide images are 2d point clouds: Context-aware survival prediction using patch-based graph convolutional networks. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2021. pp. 339–349. Springer (2021)
5. Di, D., Zou, C., Feng, Y., Zhou, H., Ji, R., Dai, Q., Gao, Y.: Generating hypergraph-based high-order representations of whole-slide histopathological images for survival prediction. IEEE Transactions on Pattern Analysis and Machine Intelligence **45**(5), 5800–5815 (2022)
6. Fan, L., Sowmya, A., Meijering, E., Song, Y.: Cancer survival prediction from whole slide images with self-supervised learning and slide consistency. IEEE Transactions on Medical Imaging (2022)
7. Guan, Y., Zhang, J., Tian, K., Yang, S., Dong, P., Xiang, J., Yang, W., Huang, J., Zhang, Y., Han, X.: Node-aligned graph convolutional network for whole-slide image representation and classification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 18813–18823 (2022)

8. Hartman, D.J., Van Der Laak, J.A., Gurcan, M.N., Pantanowitz, L.: Value of public challenges for the development of pathology deep learning algorithms. Journal of Pathology Informatics **11**(1), 7 (2020)

9. He, K., Fan, H., Wu, Y., Xie, S., Girshick, R.: Momentum contrast for unsupervised visual representation learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9729–9738 (2020)

10. Hinton, G., Vinyals, O., Dean, J.: Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531 (2015)

11. Huang, Y., Zhao, W., Wang, S., Fu, Y., Jiang, Y., Yu, L.: Conslide: Asynchronous hierarchical interaction transformer with breakup-reorganize rehearsal for continual whole slide image analysis. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 21349–21360 (2023)

12. Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: Proceedings of the 35th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 80, pp. 2127–2136. PMLR (2018)

13. Li, F., Wang, M., Huang, B., Duan, X., Zhang, Z., Ye, Z., Huang, B.: Patients and slides are equal: A multi-level multi-instance learning framework for pathological image analysis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 63–71. Springer (2023)

14. Li, R., Yao, J., Zhu, X., Li, Y., Huang, J.: Graph cnn for survival analysis on whole slide pathological images. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. pp. 174–182. Springer (2018)

15. Li, X., Zhu, R., Cheng, Y., Shan, C., Luo, S., Li, D., Qian, W.: Finding global homophily in graph neural networks when meeting heterophily. In: International Conference on Machine Learning. pp. 13242–13256. PMLR (2022)

16. Liang, M., Jiang, X., Cao, J., Li, B., Wang, L., Chen, Q., Zhang, C., Zhao, Y.: Caf-ahgcn: context-aware attention fusion adaptive hypergraph convolutional network for human-interpretable prediction of gigapixel whole-slide image. The Visual Computer pp. 1–19 (2024)

17. Lu, M.Y., Williamson, D.F., Chen, T.Y., Chen, R.J., Barbieri, M., Mahmood, F.: Data-efficient and weakly supervised computational pathology on whole-slide images. Nature Biomedical Engineering **5**(6), 555–570 (2021)

18. Lu, W., Toss, M., Dawood, M., Rakha, E., Rajpoot, N., Minhas, F.: Slidegraph+: Whole slide image level graphs to predict her2 status in breast cancer. Medical Image Analysis **80**, 102486 (2022)

19. Menéndez, M., Pardo, J., Pardo, L., Pardo, M.: The jensen-shannon divergence. Journal of the Franklin Institute **334**(2), 307–318 (1997)

20. Shao, Z., Bian, H., Chen, Y., Wang, Y., Zhang, J., Ji, X., et al.: Transmil: Transformer based correlated multiple instance learning for whole slide image classification. Advances in Neural Information Processing Systems **34**, 2136–2147 (2021)

21. Shao, Z., Chen, Y., Bian, H., Zhang, J., Liu, G., Zhang, Y.: Hvtsurv: Hierarchical vision transformer for patient-level survival prediction from whole slide image. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 37, pp. 2209–2217 (2023)

22. Shi, J., Shu, T., Wu, K., Jiang, Z., Zheng, L., Wang, W., Wu, H., Zheng, Y.: Masked hypergraph learning for weakly supervised histopathology whole slide image classification. Computer Methods and Programs in Biomedicine p. 108237 (2024)

23. Zhang, H., Meng, Y., Zhao, Y., Qiao, Y., Yang, X., Coupland, S.E., Zheng, Y.: Dtfd-mil: Double-tier feature distillation multiple instance learning for histopathology whole slide image classification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 18802–18812 (2022)

24. Zhang, X., Cao, M., Wang, S., Sun, J., Fan, X., Wang, Q., Zhang, L.: Whole slide cervical cancer screening using graph attention network and supervised contrastive learning. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2022. pp. 202–211. Springer (2022)
25. Zhang, Z., Chen, P., McGough, M., Xing, F., Wang, C., Bui, M., Xie, Y., Sapkota, M., Cui, L., Dhillon, J., et al.: Pathologist-level interpretable whole-slide cancer diagnosis with deep learning. Nature Machine Intelligence **1**(5), 236–245 (2019)