



This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

Image Distillation for Safe Data Sharing in Histopathology

Zhe Li^[0009-0003-3101-718X] and Bernhard Kainz^[0000-0002-7813-5023]

Image Data Exploration and Analysis Lab, Friedrich-Alexander-Universität
Erlangen-Nürnberg
zhe.li@fau.de

Abstract. Histopathology can help clinicians make accurate diagnoses, determine disease prognosis, and plan appropriate treatment strategies. As deep learning techniques prove successful in the medical domain, the primary challenges become limited data availability and concerns about data sharing and privacy. Federated learning has addressed this challenge by training models locally and updating parameters on a server. However, issues, such as domain shift and bias, persist and impact overall performance. Dataset distillation presents an alternative approach to overcoming these challenges. It involves creating a small synthetic dataset that encapsulates essential information, which can be shared without constraints. At present, this paradigm is not practicable as current distillation approaches only generate non human readable representations and exhibit insufficient performance for downstream learning tasks. We train a latent diffusion model and construct a new distilled synthetic dataset with a small number of human readable synthetic images. Selection of maximally informative synthetic images is done via graph community analysis of the representation space. We compare downstream classification models trained on our synthetic distillation data to models trained on real data and reach performances suitable for practical application. Codes are available at <https://github.com/ZheLi2020/InfoDist>.

Keywords: Dataset Distillation · Image Generation · Privacy.

1 Introduction

In histopathology, pathologists examine thin sections of tissues derived from biopsies. These tissue samples are typically stained to enhance the visibility of cellular structures. Subsequently, the stained slides undergo microscopic analysis to identify abnormal changes or patterns that may indicate the presence of a disease. Histopathology plays a crucial role in understanding various medical conditions, help clinicians make accurate diagnoses, determine disease prognosis, and plan appropriate treatment strategies.

Deep learning and computer vision methods have demonstrated success in the analysis of histopathology images, encompassing tasks such as disease detection, tumor classification, and cell segmentation. Some approaches use multiple

instance learning (MIL) [32] or hierarchical approaches [5,12,14]. These achieve commendable classification performance. Nevertheless, the challenges of data scarcity and data sharing is hard to tackle due to concerns related to privacy protection, diverse dataset formats, and legal and regulatory frameworks. Federated learning has addressed this issue through a distributed approach, where individuals train their data locally, and model parameters are updated on a central server. While this method maintains data privacy, it introduces new challenges such as domain shift and bias.

Some researchers have explored knowledge distillation [1,35,29,36,37] or the generation of synthetic images [26,34,16] as data enrichment methods to address the data shortage problem and enhance performance. Concerns about data privacy still exist because these methods only assist the training and models are trained on real data as before. In this context, dataset distillation emerges as a more appropriate solution. It learns from a large real dataset and generating a small synthetic dataset that maximally encapsulates essential information. Subsequently, downstream tasks are trained on the dataset that solely consists of synthetic images. Effective dataset distillation would allow sharing of a small synthetic dataset that can a) guarantee to not contain any privacy concerning identifiable information, b) be used as a direct representation of the underlying data distribution characteristics for downstream applications and bias mitigation, and c) allow resource efficient training and refinement of, *e.g.*, general foundation models. Leveraging data distillation to work with compact, synthesized images across clinical sites could effectively anonymize the training process by removing all patient-specific information on the image level.

We would expect that approaches for dataset distillation can achieve comparable performance to those trained on a large real dataset. However, the performance achieved by training solely on a small synthetic dataset is sub-optimal. One reason is the majority of approaches only generate a limited number of synthetic images, typically around 1 or 10 for each class. Hence, its practical applicability in the real world is currently limited. Our goal is to extract maximum information from the original large dataset with real images, and generate a minimal synthetic dataset that can achieve comparable performance when training a task-specific model. To achieve this, we train a class conditional latent diffusion model and generate 1000 synthetic images for each class. We then utilize the Infomap algorithm [3] to select 100 images from it in the representation space of convolutional networks. Afterwards, we train classifiers on this distilled subset to evaluate the set’s representational power. This approach enables us to achieve comparable classification accuracy to real images while efficiently reducing storage costs and training efforts simultaneously.

To the best of our knowledge, there are no works that have specifically investigated dataset distillation for histopathology. Our contribution consists of:

1. We propose an approach to use synthetic images for data sharing, which mitigates privacy concerns – InfoDist.
2. We explore an efficient method to select images with essential information. The process involves projecting images to the embedding using a pre-trained

- convolutional network, followed by the utilization of the map equation and Infomap algorithm [3] to select images with high modular centrality.
3. We design a contrasting learning loss to further improve performance.
 4. In our experiments, we report the average results of test accuracy, F1 score, and AUC score. We demonstrate that competitive performance can be achieved when training solely on a distilled set of synthetic images.

Related Work. Dataset distillation is initially explored on natural images, such as CIFAR-10 or ImageNet to reduce the storage and computation cost. Recently, such frameworks have been applied on gastric X-ray images [18] and a COVID-19 chest X-ray dataset [19]. There is no work on histopathology images so far. Most researchers focus on approaches [27,25,13,23,28,15] using the MIL method, multi-modal data combination [21] or data augmentation [9] for the classification of whole slide images or segmentation [6].

2 Method

Our idea for InfoDist is outlined in Fig. 1. The overall goal is to select the most informative subset of images X_s from a set of all images X , so that X_s will provide enough information to train another downstream classifier, ideally yielding the same classification performance on an unseen test set X_t as if trained on the full X . X could not readily be shared with third parties, since this may be samples from a set of real patient data. Thus we propose to transform X to a newly generated synthetic set $X_g = \mathcal{F}_\theta(X)$ with infinite extent and start the data distillation process from there.

Latent diffusion models recently became a viable option to model $\mathcal{F}_\theta(X)$ without too much loss of information about the underlying distribution in X . We utilize a class conditional latent diffusion model U-ViT [2] to model $X_g = \mathcal{F}_\theta(X)$. It’s a backbone latent diffusion model in a U-Net shape, where each block is a vision transformer [8]. The diffusion model is trained on real data and learns the data distribution. It can be trained locally once without sharing it publicly. Then we can generate synthetic images, but these images are different from any real images, even though they contain some realistic features. Therefore, there is no personal information in synthetic images. After training on the pathological dataset, it can generate synthetic images $X_g \in \mathbb{R}^{3 \times W \times H}$ that are comparable to real images. Our goal is to capture the input data distribution as comprehensively as possible. To achieve this, we generate a large number of synthetic images for X_g and then select a small subset of representative images.

X_g can be further embedded into the embedding space \mathbb{R}^n of a classifier if the size of elements $\in X_g$ is computationally prohibitive. The embedding space \mathbb{R}^n is the output of the penultimate layer of a classifier pretrained on all real images X .

From the embedding space $X_g \in \mathbb{R}^n$ we construct a weighted graph from the set $X_g = x_1, \dots, x_N$ with a metric $d : X \times X \rightarrow \mathbb{R}_{w \geq \eta}$, where w is the inversed Euclidean distance and η is a threshold. This leads to a weighted directed graph $\tilde{G} = (V, E, w)$, where the nodes $V = X_g$ and the edges $E = (x_i, x_{i_j}) | w \geq \eta$.

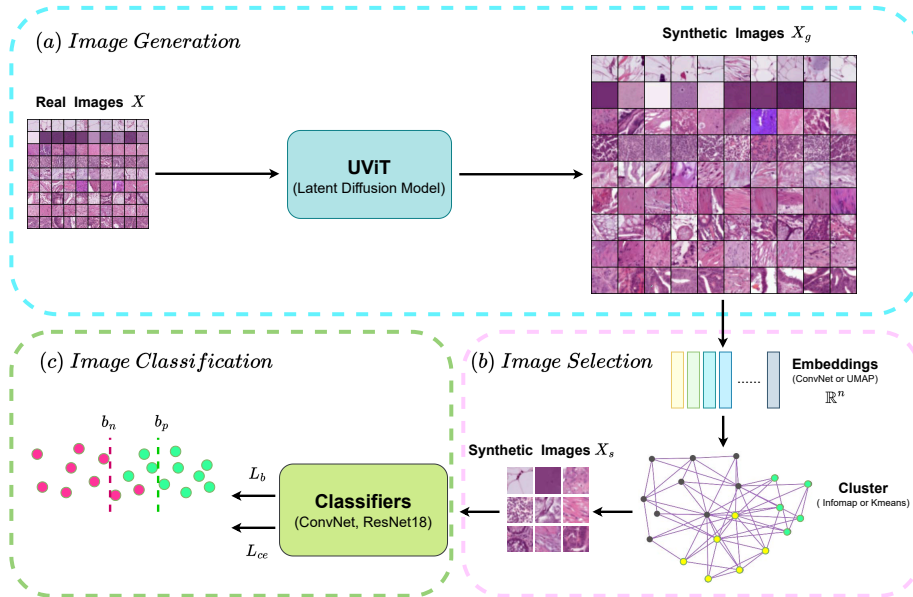


Fig. 1: Overview of our InfoDist approach. (a) We train a latent diffusion model UViT [2] and generate a synthetic dataset. (b) We extract the image embeddings by a pre-trained convolutional network or UMAP [22], then use the modified infomap algorithm to detect communities. We select a small synthetic dataset in which images have high modular centrality in each community. (c) We train the classifiers only on the small selected synthetic dataset and apply both cross entropy loss \mathcal{L}_{ce} and contrastive learning loss \mathcal{L}_{con} in training.

An alternative embedding method is UMAP [22]. We flatten X_g and reduce the dimension by UMAP which also constructs a graph by Euclidean distance. In this graph, each node x_i connects its top- k neighbors and the edges become $E = (x_i, x_{i_j}) | 1 \leq j \leq k, 1 \leq i \leq N$. We also use this approach in our experiments for comparison.

We hypothesize that \tilde{G} contains essential information about the relevance among individual images $I \in X_g$ for dataset distillation and their effectiveness to shape decision boundaries in downstream classifiers. In exploring the nuances of nodes and the identification of weak boundaries among clusters, we introduce an unsupervised method to unearth samples that encode intricate details, even within a single class. Here, community detection emerges as an important factor; specifically, the nodes with the high modular centrality in each community could be deemed the most relevant. The scalar score of modular centrality combines two scores to quantify both the intra-community and inter-community influence [10]. This allows community detection not merely as a tool for group identification but as a method of compression and dataset distillation. Via the Infomap algorithm [3], we can identify those key samples that, despite their limited number, encapsulate the dataset’s complexity and facilitate the formation

of robust decision boundaries, thereby enhancing the efficiency and effectiveness of the distillation process.

Community detection can leverage the Map Equation [24], an unsupervised method grounded in information theory, which aims to optimize community identification based on the principle of minimum description length [11]. This approach seeks to encapsulate the behavior of a random walker within a network in the most concise way, by reducing the expected per-step codelength. It achieves this by organizing the network into clusters within which the random walker is likely to remain for extended periods. The Map Equation method does not require simulating random walks to achieve its objectives; rather, it analytically computes the codelength:

$$L(M) = q_{\curvearrowright} H(Q) + \sum_{i=1}^m p_{\circlearrowleft}^i H(P^i) \quad (1)$$

where $L(M)$ is the total codelength for a given partition M of the network into m clusters, q_{\curvearrowright} is the probability of the random walker transitioning between clusters, $H(Q)$ is the entropy of the module exit probabilities, quantifying the uncertainty in module exits, p_{\circlearrowleft}^i is the probability of the random walker staying within module i , $H(P^i)$ is the entropy of the visitation probabilities within module i , and m is the number of clusters. Algorithm 1, which is a modified version of InfoMap [3] can then be used to distill nodes that have high modular centrality from each community uniformly. Here, we generate graphs for each class separately, so all synthetic images with same class label are nodes in the graph. We calculate the inverse Euclidean distance and apply *softmax* operation as weights of links between nodes. To ensure that higher weight represents a stronger relation between nodes, we set a threshold η to remove the links with low weights.

Contrastive Learning Loss. To further the performance of downstream classifiers that are trained on very limited data like distilled datasets, we propose the use of a contrastive learning loss [33]. We set two boundaries and process output probabilities of the convolutional model during training to generate the boundaries with the assistance of masks defined by the ground truth. Specifically, we slice the output to obtain the probabilities for one class and mask out the positive and negative samples within the current batch. The mask is generated based on the ground truth label. If the ground truth corresponds to the current processing class label, the value is set to 1. Otherwise, it is set to 0. After applying the mask to probabilities, we obtain the probabilities of samples associated with the current class label. We assume that there are positive samples with low probabilities which are easily confused with negative samples. We set a threshold such that a percentage ρ of the positive samples are predicted correctly. Thus, we sort the probabilities of positive samples in descending order and set the sample probability at the last position of $B \times \rho$ samples as the positive boundary b_p . The remaining $B \times (1 - \rho)$ positive samples with lower probabilities are contributed to the loss calculation as the first term of Eq. 2. The negative boundary b_n for those probabilities with a mask value 0 is calculated by

Algorithm 1 Our proposed InfoDist algorithm for Community Detection

- 1: **Input:** The real images X^k or synthetic images X_g^k with class label k ;
 - 2: N , the number of selected images.
 - 3: Initialize each node as its own community C .
 - 4: Initialize empty distillation set X_s .
 - 5: **repeat**
 - 6: **for** each node i in the network **do**
 - 7: **for** each community C that i 's neighbors belong to **do**
 - 8: Calculate the change in $L(M)$ if i is moved to C .
 - 9: **end for**
 - 10: Move node i to the community that results in the greatest decrease in $L(M)$.
 - 11: **end for**
 - 12: Update the community structure based on node movements.
 - 13: **until** no further reduction in $L(M)$ is possible
 - 14: Calculate N_i , the number of images that needs to be selected from each community.
 - 15: **for** each community C_i in the network **do**
 - 16: $X_s \leftarrow I \in X^k$ or $I \in X_g^k$ nodes that have high modular centrality.
 - 17: **end for**
 - 18: Apply the algorithm recursively to images in other classes to construct the small real or synthetic dataset.
-

subtracting a hyperparameter τ from b_p . The probabilities that are higher than the negative boundary are calculated in the loss as the second term of Eq. 2.

$$\mathcal{L}_c = \sum_{i=0}^{Pos} |\min((p_i - b_p), 0)| + \sum_{j=0}^{Neg} |\max((p_j - b_n + \tau), 0)|, \quad (2)$$

where Pos denotes the number of positive images for current class label in the ground truth and Neg denotes the number of the remaining images with other class labels. p_i, p_j is the output probability of the synthetic images s_i, s_j after *softmax* operation. We compute two boundaries and the contrastive learning loss for each class separately and calculate the contrastive loss $\mathcal{L}_b = \sum_{c=0}^C \mathcal{L}_c$. Finally, the total loss is a combination with cross entropy $\mathcal{L} = \mathcal{L}_b + \mathcal{L}_{ce}$.

3 Experiments

Datasets. We evaluate our distillation approach on the public MedMNIST datasets [31] which comprises MNIST-like datasets featuring standardized biomedical images that consists of a total of 12 datasets for 2D and 6 datasets for 3D. Our method is specifically applied to one of these datasets, PathMNIST, which includes histopathology images for colon biopsies. This dataset consists of 107,180 image samples which are divided into training (89,996), validation (10,004), and test (7,180) sets. All images are resampled into 28 x 28 (2D) resolution. The PathMNIST dataset has 9 types of tissues: ADI, adipose tissue; BACK, background; CRC, colorectal cancer; DEB, debris; HE, hematoxylin-eosin; LYM,

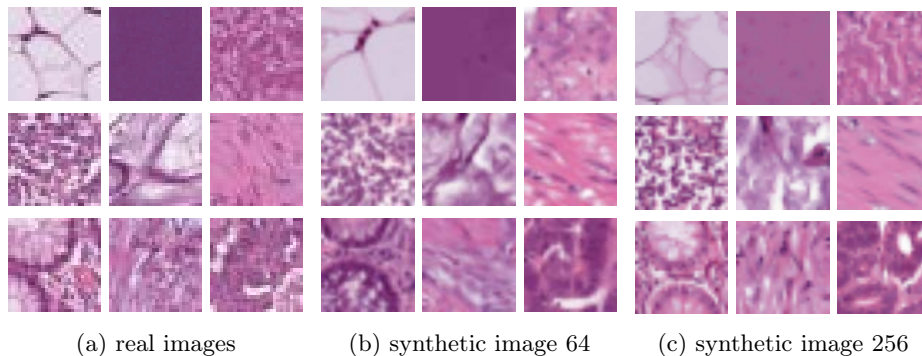


Fig. 2: The real and synthetic samples at different resolutions.

lymphocytes; MUC, mucus; MUS, smooth muscle; NCT, National Center for Tumor Diseases; NORM, normal colon mucosa; STR, cancer-associated stroma; TUM, colorectal adenocarcinoma epithelium [17].

Metrics. We report 3 evaluation metrics, the classification accuracy, the F1 score, and the AUC on the test set of PathMNIST.

Implementation Details. In the embedding process, images are projected into features with dimensions of either 2048 or 8192, corresponding to resolutions of 64×64 or 256×256 , using a pre-trained ConvNet or ResNet18. The UMAP projects the images into features with dimension 10, with a specified number of neighbors set to 10 and a minimum distance parameter of 0.05. In Infomap, we can either utilize the graph generated by UMAP or calculate the distance matrix using the Euclidean distance where the weight threshold η is set to 0.001 or 0.004. For computation cost, training ConvNet model for 5 times needs about 20 minutes and ResNet18 needs about 40 minutes on a A100 GPU.

Synthetic Images. We train a class conditional latent diffusion model U-ViT [2] on the training set of the PathMNIST dataset which consists of 89,996 images. We resize input images from resolutions 28×28 to 64×64 or 256×256 respectively before training. After training the U-ViT-L/4 on images with resolution 64×64 and U-ViT-L/2 on 256×256 respectively, we generate 1000 synthetic images for each class, resulting in a total of 9000 synthetic images for each size. Fig. 2(a) shows the samples of real images. Fig. 2(b) exhibits samples of synthetic images generated at a resolution of 64×64 , while Fig. 2(c) showcases samples of synthetic images generated at a resolution of 256×256 .

Distilled data. To create a distilled dataset, we select 100 images for each class from the generated synthetic dataset with our InfoDist approach. Therefore, the condensed synthetic dataset comprises 900 images for training a downstream classifier, considering there are 9 classes in the PathMNIST dataset. We employ two classifiers: ConvNet and ResNet18. The reported results are the average of 5 runs on the entire real test set of the PathMNIST dataset and the corresponding standard deviation. In each run, the selected training images are updated.

Table 1 shows our results of InfoDist on a small distilled dataset compared to the state-of-the-art. We train two classifiers on the whole real training set and

Table 1: Results compared to state-of-the-art. The top part shows the upper bound for the performance when the private training data is available, following results of distilled real dataset. The bottom part shows the results on all synthetic images or distilled synthetic images. Our performance are better than the two baselines and comparable to results of distilled real dataset in top part.

	Real	res.	#img/class	ConvNet			ResNet18			
				ACC	F1	AUC	ACC	F1	AUC	
upper bound	[30,31,7,20]	64 ²	~ 10k	-	-	-	85.28 \pm 5.99	-	97.33 \pm 1.68	
	Reproduced	64 ²	~ 10k	91.44 \pm 0.20	88.36 \pm 0.24	99.23 \pm 0.05	91.25 \pm 0.71	87.95 \pm 0.97	99.00 \pm 0.24	
	[30,31]	224 ²	~ 10k	-	-	-	88.45 \pm 3.47	-	98.35 \pm 0.78	
	Reproduced	256 ²	~ 10k	92.02 \pm 0.48	88.76 \pm 0.67	99.25 \pm 0.06	90.96 \pm 0.35	87.27 \pm 0.46	98.92 \pm 0.09	
	real distilled									
	InfoDist	64 ²	100	77.34 \pm 0.42	69.45 \pm 0.51	94.98 \pm 0.23	78.50 \pm 0.62	71.77 \pm 0.21	96.47 \pm 0.57	
InfoDist	256 ²	100	81.17 \pm 0.53	73.81 \pm 0.64	96.73 \pm 0.24	83.90 \pm 0.50	77.47 \pm 0.77	97.77 \pm 0.11		
Synthetic										
Reproduced	64 ²	1k	79.87 \pm 0.86	73.59 \pm 0.90	96.74 \pm 0.22	86.18 \pm 0.46	80.80 \pm 0.80	98.15 \pm 0.12		
syn distilled										
results	GLaD [4]	64 ²	1	40.49 \pm 1.26	28.46 \pm 1.55	74.45 \pm 0.74	46.68 \pm 1.16	34.63 \pm 0.87	75.97 \pm 0.41	
	Random	64 ²	100	61.52 \pm 0.60	54.29 \pm 0.65	89.31 \pm 0.94	72.82 \pm 2.13	64.74 \pm 3.10	94.33 \pm 1.03	
	InfoDist	64 ²	100	69.79\pm1.28	63.01\pm1.10	91.15\pm0.35	77.48\pm1.52	70.80\pm1.03	96.84\pm0.33	
	GLaD [4]	256 ²	1	38.81 \pm 0.91	29.44 \pm 0.78	70.26 \pm 0.60	45.70 \pm 2.59	35.59 \pm 1.43	82.24 \pm 0.55	
	Random	256 ²	100	58.29 \pm 2.58	50.95 \pm 2.29	89.97 \pm 0.97	71.82 \pm 3.21	63.88 \pm 3.14	94.99 \pm 0.86	
	InfoDist	256 ²	100	65.45\pm1.40	57.47\pm1.40	91.45\pm0.43	77.79\pm1.18	72.53\pm1.53	94.59\pm0.37	

indicate results as *Reproduced*. We also apply InfoDist on the real dataset and select a small dataset with 900 real images. We report the results in *Real/real distilled/InfoDist*. In the bottom part, *Synthetic*, we train two classifiers on all synthetic images and report results in the rows *Reproduced*. For the distilled synthetic dataset with 900 images, we compare our approach with two baselines, GLAD [4] and random image selection. With resolution 64×64 , a ConvNet achieves 69.79 test accuracy which is comparable to the performance 77.34 of the distilled real dataset. A ResNet18 can achieve better performance 77.48 in test accuracy which is on par to the 78.50 of distilled real images. At a resolution of 256×256 , we achieve competitive results as well. A ResNet18 can achieve performance 77.79 in test accuracy compared to the 83.90 of distilled real images.

Ablation Study. We provide an extensive ablation study including different embeddings (ConvNet, RestNet, UMAP) with different clustering methods in several configurations for the node selection in the Appendix. For Table 1 we fixed the Infomap node selection metric to *modular centrality*. In the Appendix we also explore the effect of using enter flow or exit flow as alternatives. We also conducted an ablation study on hard code hyperparameters and selected the best combination to report the results.

4 Conclusion

In this paper, our goal is to make dataset distillation applicable in the real world because its various advantages. In our setting, privacy information is removed

and data security concerns are alleviated. We initiate the process by training a latent diffusion model and generating a synthetic dataset. Subsequently, we employ cluster methods to select a smaller dataset. Classifiers are then trained on this reduced synthetic dataset, and we report the test accuracy, F1 score, and AUC score. The incorporation of a contrastive learning loss contributes to the enhancement of performance. We also select an equivalent number of real images for comparison. Our results on a distilled synthetic dataset are comparable with those on a small real dataset. Furthermore, our AUC score is competitive with that of the entire real dataset.

Acknowledgments. This work was supported by the State of Bavaria, the High-Tech Agenda (HTA) Bavaria and HPC resources provided by the Erlangen National High Performance Computing Center (NHR@FAU) of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) under the NHR project b180dc. NHR@FAU hardware is partially funded by the German Research Foundation (DFG) - 440719683. Support was also received from the ERC - project MIA-NORMAL 101083647 and DFG KA 5801/2-1, INST 90/1351-1.

Disclosure of Interests. The authors have no competing interests for this work.

References

1. Azadi, P., Suderman, J., Nakhli, R., Rich, K., Asadi, M., Kung, S., Oo, H., Keyes, M., Farahani, H., MacAulay, C., et al.: All-in: A local global graph-based distillation model for representation learning of gigapixel histopathology images with application in cancer risk assessment. In: MICCAI'23. pp. 765–775. Springer (2023)
2. Bao, F., Nie, S., Xue, K., Cao, Y., Li, C., Su, H., Zhu, J.: All are worth words: A vit backbone for diffusion models. In: CVPR'23. pp. 22669–22679 (2023)
3. Blöcker, C., Tan, C., Scholtes, I.: The map equation goes neural. preprint arXiv:2310.01144 (2023)
4. Cazenavette, G., Wang, T., Torralba, A., Efros, A.A., Zhu, J.Y.: Generalizing dataset distillation via deep generative prior. In: CVPR'23. pp. 3739–3748 (2023)
5. Chen, R.J., Chen, C., Li, Y., Chen, T.Y., Trister, A.D., Krishnan, R.G., Mahmood, F.: Scaling vision transformers to gigapixel images via hierarchical self-supervised learning. In: CVPR'22. pp. 16144–16155 (2022)
6. Deng, R., Li, Y., Li, P., Wang, J., Remedios, L.W., Agzamkhodjaev, S., Asad, Z., Liu, Q., Cui, C., Wang, Y., et al.: Democratizing pathological image segmentation with lay annotators via molecular-empowered learning. In: MICCAI'23. pp. 497–507. Springer (2023)
7. Derakhshani, M.M., Najdenkoska, I., van Sonsbeek, T., Zhen, X., Mahapatra, D., Worring, M., Snoek, C.G.: Lifelong: A benchmark for continual disease classification. In: MICCAI'22. pp. 314–324. Springer (2022)
8. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. preprint arXiv:2010.11929 (2020)

9. Gadermayr, M., Koller, L., Tschuchnig, M., Stangassinger, L.M., Kreutzer, C., Couillard-Despres, S., Oostingh, G.J., Hittmair, A.: Mixup-mil: Novel data augmentation for multiple instance learning and a study on thyroid cancer diagnosis. In: MICCAI'23. pp. 477–486. Springer (2023)
10. Ghalmane, Z., El Hassouni, M., Cherifi, C., Cherifi, H.: Centrality in modular networks. *epj data sci* 8 (1): 15 (2019)
11. Grünwald, P.D., Myung, I.J., Pitt, M.A.: Advances in minimum description length: Theory and applications. MIT press (2005)
12. 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: CVPR'22. pp. 18813–18823 (2022)
13. 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: ICCV'23. pp. 21349–21360 (2023)
14. Jiang, C., Hou, X., Kondepudi, A., Chowdury, A., Freudiger, C.W., Orringer, D.A., Lee, H., Hollon, T.C.: Hierarchical discriminative learning improves visual representations of biomedical microscopy. In: CVPR'23. pp. 19798–19808 (2023)
15. Jin, T., Xie, X., Wan, R., Li, Q., Wang, Y.: Gene-induced multimodal pre-training for image-omic classification. In: MICCAI'23. pp. 508–517. Springer (2023)
16. Kang, M., Chikontwe, P., Kim, S., Jin, K.H., Adeli, E., Pohl, K.M., Park, S.H.: One-shot federated learning on medical data using knowledge distillation with image synthesis and client model adaptation. In: MICCAI'23. pp. 521–531. Springer (2023)
17. Kather, J.N., Krisam, J., Charoentong, P., Luedde, T., Herpel, E., Weis, C.A., Gaiser, T., Marx, A., Valous, N.A., Ferber, D., et al.: Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study. *PLoS medicine* 16(1), e1002730 (2019)
18. Li, G., Togo, R., Ogawa, T., Haseyama, M.: Compressed gastric image generation based on soft-label dataset distillation for medical data sharing. *Computer Methods and Programs in Biomedicine* 227, 107189 (2022)
19. Li, G., Togo, R., Ogawa, T., Haseyama, M.: Dataset distillation for medical dataset sharing. preprint arXiv:2209.14603 (2022)
20. Liu, J., Li, Y., Cao, G., Liu, Y., Cao, W.: Feature pyramid vision transformer for medmnist classification decathlon. In: IJCNN'22. pp. 1–8. IEEE (2022)
21. Lu, M., Wang, T., Xia, Y.: Multi-modal pathological pre-training via masked autoencoders for breast cancer diagnosis. In: MICCAI'23. pp. 457–466 (2023)
22. McInnes, L., Healy, J., Melville, J.: Umap: Uniform manifold approximation and projection for dimension reduction. preprint arXiv:1802.03426 (2018)
23. Qu, L., Yang, Z., Duan, M., Ma, Y., Wang, S., Wang, M., Song, Z.: Boosting whole slide image classification from the perspectives of distribution, correlation and magnification. In: ICCV'23. pp. 21463–21473 (2023)
24. Rosvall, M., Axelsson, D., Bergstrom, C.T.: The map equation. *The European Physical Journal Special Topics* 178(1), 13–23 (2009)
25. Shao, Z., Wang, Y., Chen, Y., Bian, H., Liu, S., Wang, H., Zhang, Y.: Lnpl-mil: Learning from noisy pseudo labels for promoting multiple instance learning in whole slide image. In: ICCV'23. pp. 21495–21505 (2023)
26. Shrivastava, A., Fletcher, P.T.: Nasdm: Nuclei-aware semantic histopathology image generation using diffusion models. MICCAI'23 (2023)
27. Tang, W., Huang, S., Zhang, X., Zhou, F., Zhang, Y., Liu, B.: Multiple instance learning framework with masked hard instance mining for whole slide image classification. In: ICCV'23. pp. 4078–4087 (2023)

28. Wang, H., Luo, L., Wang, F., Tong, R., Chen, Y.W., Hu, H., Lin, L., Chen, H.: Iteratively coupled multiple instance learning from instance to bag classifier for whole slide image classification pp. 467–476 (2023)
29. Wang, X., Li, Z., Luo, X., Wan, J., Zhu, J., Yang, Z., Yang, M., Xu, C.: Black-box domain adaptative cell segmentation via multi-source distillation. In: MICCAI'23. pp. 749–758. Springer (2023)
30. Yang, J., Shi, R., Ni, B.: Medmnist classification decathlon: A lightweight automl benchmark for medical image analysis. In: ISBI'21. pp. 191–195. IEEE (2021)
31. Yang, J., Shi, R., Wei, D., Liu, Z., Zhao, L., Ke, B., Pfister, H., Ni, B.: Medmnist v2- a large-scale lightweight benchmark for 2d and 3d biomedical image classification. *Scientific Data* **10**(1), 41 (2023)
32. Yang, J., Chen, H., Zhao, Y., Yang, F., Zhang, Y., He, L., Yao, J.: Remix: A general and efficient framework for multiple instance learning based whole slide image classification. In: MICCAI'22. pp. 35–45. Springer Nature Switzerland, Cham (2022)
33. Yao, X., Li, R., Zhang, J., Sun, J., Zhang, C.: Explicit boundary guided semi-push-pull contrastive learning for supervised anomaly detection. In: CVPR'23. pp. 24490–24499 (2023)
34. Ye, J., Ni, H., Jin, P., Huang, S.X., Xue, Y.: Synthetic augmentation with large-scale unconditional pre-training. In: MICCAI'23. pp. 754–764. Springer (2023)
35. Yu, Z., Lin, T., Xu, Y.: SLPD: Slide-level prototypical distillation for WSIs. In: Greenspan, H., Madabhushi, A., Mousavi, P., Salcudean, S., Duncan, J., Syeda-Mahmood, T., Taylor, R. (eds.) MICCAI'23. pp. 259–269. Springer Nature Switzerland, Cham (2023)
36. Yu, Z., Lin, T., Xu, Y.: Slpd: slide-level prototypical distillation for wsis. In: MICCAI'23. pp. 259–269. Springer (2023)
37. Zhong, L., Liao, X., Zhang, S., Wang, G.: Semi-supervised pathological image segmentation via cross distillation of multiple attentions. MICCAI'23 (2023)