Journal of Pathology Informatics 14 (2023) 100320

Contents lists available at ScienceDirect

Journal of Pathology Informatics

journal homepage: www.elsevier.com/locate/jpi

HistoPerm: A permutation-based view generation approach for improving histopathologic feature representation learning

Joseph DiPalma [a], Lorenzo Torresani [a], Saeed Hassanpour [a,b,c,*]

[a] Department of Computer Science, Dartmouth College, Hanover, NH 03755, USA
[b] Department of Biomedical Data Science, Geisel School of Medicine at Dartmouth, Hanover, NH 03755, USA
[c] Department of Epidemiology, Geisel School of Medicine at Dartmouth, Hanover, NH 03755, USA

ARTICLE INFO

Keywords:
Representation learning
Joint embedding architectures
Digital pathology

ABSTRACT

Deep learning has been effective for histology image analysis in digital pathology. However, many current deep learning approaches require large, strongly- or weakly labeled images and regions of interest, which can be time-consuming and resource-intensive to obtain. To address this challenge, we present HistoPerm, a view generation method for representation learning using joint embedding architectures that enhances representation learning for histology images. HistoPerm permutes augmented views of patches extracted from whole-slide histology images to improve classification performance. We evaluated the effectiveness of HistoPerm on 2 histology image datasets for Celiac disease and Renal Cell Carcinoma, using 3 widely used joint embedding architecture-based representation learning methods: BYOL, SimCLR, and VICReg. Our results show that HistoPerm consistently improves patch- and slide-level classification performance in terms of accuracy, F1-score, and AUC. Specifically, for patch-level classification accuracy on the Celiac disease dataset, HistoPerm boosts BYOL and VICReg by 8% and SimCLR by 3%. On the Renal Cell Carcinoma dataset, patch-level classification accuracy is increased by 2% for BYOL and VICReg, and by 1% for SimCLR. In addition, on the Celiac disease dataset, models with HistoPerm outperform the fully supervised baseline model by 6%, 5%, and 2% for BYOL, SimCLR, and VICReg, respectively. For the Renal Cell Carcinoma dataset, HistoPerm lowers the classification accuracy gap for the models up to 10% relative to the fully supervised baseline. These findings suggest that HistoPerm can be a valuable tool for improving representation learning of histopathology features when access to labeled data is limited and can lead to whole-slide classification results that are comparable to or superior to fully supervised methods.

Introduction

Digital pathology involves the visualization and analysis of whole-slide images (WSIs) to assist pathologists in the diagnosis and prognosis of various diseases. These WSIs are digitized at high resolutions and can be analyzed manually, using computer vision models, or a combination. However, the large size of these images, up to 150 000 × 150 000 pixels, can present challenges for typical computer vision-based image analysis tools.

In recent years, various computer vision-based methods and solutions have been proposed and developed to handle the gigapixel size of WSIs and address other unique challenges of digital pathology.[1–8] In terms of label and annotation requirements, these methods differ from those used on natural images in several ways. Firstly, the labeling process for WSIs requires highly trained experts, while natural images often require minimal or no prerequisites for labeling. Secondly, labels are typically provided at the slide level rather than at the patch level. Finally, the class label may only be determined by a small portion of the WSI. These characteristics present major challenges for the application of standard computer vision methods in digital pathology.

Among these 3 annotation bottlenecks, the last 2 are most unique to digital pathology. Due to the large size of the WSIs, it is infeasible for pathologists to label all regions of interest on a slide. Instead, the label is usually provided at the slide level, which also applies to class-negative regions of a slide. Moreover, an object in the average image from the ImageNet natural image dataset occupies 25% of the area,[9] while a typical region-of-interest annotation in a WSI can occupy as little as 5% of the image.[4,10] The combination of weak labeling and low object scale poses a unique challenge and makes applying standard computer vision methods a suboptimal solution in digital pathology.

In the last decade, deep learning models have been highly successful in numerous classic computer vision tasks.[11–13] To make these standard deep learning models more feasible and effective for WSIs, it is common to preprocess the images into smaller patches, typically 224 × 224 pixels.

* Corresponding author at: One Medical Center Drive, HB 7261, Lebanon, NH 03756, USA.
E-mail address: Saeed.Hassanpour@dartmouth.edu (S. Hassanpour).

http://dx.doi.org/10.1016/j.jpi.2023.100320
Received 5 April 2023; Received in revised form 23 June 2023; Accepted 28 June 2023
Available online 4 July 2023

However, this can lead to further issues if the weakly labeled nature of the slides is not considered. A common approach involves the use of a convolutional neural network (CNN) on smaller patches extracted from large whole-slide images, with the patch classification results being aggregated for whole-slide inferencing.[14–21] However, this approach can have suboptimal performance if the signal-to-noise ratio is low among the extracted patches. More advanced methods involving attention[22,23] or multiple-instance learning[24–32] have been developed to use the weak-labels, but these still require large, labeled datasets.

In recent years, self-supervised representation learning techniques have gained significant traction for their ability to solve difficult problems in computer vision without relying on labor-intensive, manually labeled datasets. These methods utilize a pretext task to learn a latent representation of an unlabeled dataset, which is often readily available in the medical domain. To address the labeling challenge, self-supervised approaches have been successfully applied to histology images using existing computer vision techniques.[33–37] These approaches aim to exploit the unique characteristics of these images, such as rotation invariance or local-to-global consistency. In addition, contrastive learning-based methods have gained popularity in histology feature representation, [38–45] with techniques such as Contrastive Predictive Coding,[38,42] and DSMIL[46] proving effective for incorporating multiscale information into contrastive models. However, all these approaches still require all input data to be labeled, whether weakly or strongly.

To address this shortcoming, we propose a model-agnostic view generation method called HistoPerm for representation learning in histology image classification. Unlike prior methods, HistoPerm is flexible and incorporates both labeled and unlabeled data into the learning process. In contrast to prior view generation approaches for histology images, which produce views at random from the same instance, we perform a permutation on a portion of the mini-batch such that the view comes from the same class but a different instance of the class. By taking advantage of the large pool of both class-positive and class-negative patches, our approach can derive stronger representations for histological features. Our experiments show that adding HistoPerm to an existing state-of-the-art representation learning image analysis pipeline improves the histology image classification performance.

Representation learning with joint embedding architectures

Several paradigms for representation learning have been proposed, including contrastive learning,[47–53] non-contrastive learning,[54–56] and information preservation.[57–60] These paradigms all employ joint embedding architectures, where 2 models are trained to produce similar outputs when given augmented views of the same source image.

Contrastive learning relies on positive and negative samples to guide the network through learning unique identifiers for each class in the downstream task. However, these approaches often require large mini-batch sizes and significant computational resources, making them impractical for many studies and applications. Smaller mini-batch sizes can be used by implementing "tricks" such as momentum encoders,[49,50] but in general, these approaches are still resource-intensive and require massive computational power that is inaccessible to most researchers.

Non-contrastive approaches, which utilize only positive instances, also require fewer resources but may result in a slight decrease in the downstream classification performance.[54–56] The underlying principle that prevents convergence to trivial, constant (i.e., collapsed) embeddings in these methods is unknown, but prior works have shown that implementation details do play some part in their success.[61–63]

Information preservation methods, such as Barlow Twins,[64,65] Whitening-MSE,[65] and VICReg,[66,67] aim to decorrelate variables in the learned representations and explicitly prevent collapse. These methods are effective at avoiding trivial embeddings and have shown promise in natural image tasks.

Rationale for our work

Prior work has shown that representation learning methods rely on building representations that are invariant to irrelevant variations in the input.[68] For histopathology, many patches share similar histologic features and visual attributes, independent of the class. Given this, many of the patches sampled from WSIs are unsuitable as negative samples for learning unlike natural images. Hence, we utilize the large pool of *both* class-positive and class-negative patches to build stronger representations for histologic features by allowing permutation at the mini-batch scale. While we may encounter instances where class-positive and class-negative instances are paired and these instances are *not* morphologically similar, these hard cases should not be common enough in a typical WSI classification task to impact feature learning adversely, and such hard cases may even be beneficial to learning according to previous research.[69] Notably, HistoPerm is model-agnostic and can be integrated into any joint-embedding architecture-based representation learning framework operating on 2 input views to improve histologic feature representation learning.

Method

In this section, we introduce our proposed method, HistoPerm, a permutation-based view generation technique to improve capturing histologic features in representation learning frameworks. A high-level overview of our approach is shown in Fig. 1.

WSI to patch conversion

Let $\mathcal{D}$ be a dataset comprised of WSIs. Disjoint labeled and unlabeled subsets $\mathcal{D}_l$ and $\mathcal{D}_u$ partition $\mathcal{D}$. For each WSI $\mathcal{S}_i \in \mathcal{D}_l$, we have an associated ground-truth label y_i corresponding to the pathologist-provided slide-level classification label. Given a slide $\mathcal{S}_i$, we produce a set of patches p_i and assign them the slide-level label y_i. In this weakly labeled setting, we can have anywhere between 1 and $|p_i|$ class-positive patches per set p_i, where class-positive means the patch has the same class label as the slide. As discussed earlier, in histopathological classification, we can assume that the majority of patches in p_i will be negative relative to slide-level class y_i. After this step, we have labeled and unlabeled sets of patches $\mathcal{P}_l$ and $\mathcal{P}_u$ produced from $\mathcal{D}_l$ and $\mathcal{D}_u$, respectively. In the next section, we explain how we generate the input views for our model, given $\mathcal{P}_l$ and $\mathcal{P}_u$.

View generation

View augmentation. Given a weakly labeled patch dataset $\mathcal{P}_l$ and unlabeled patch dataset $\mathcal{P}_u$, we sample without replacement mini-batches $\mathcal{X}_l$ and $\mathcal{X}_u$ such that $\mathcal{X}_l \sim \mathcal{P}_l$ and $\mathcal{X}_u \sim \mathcal{P}_u$. Furthermore, let $\alpha \in [0, 1]$ be a hyperparameter representing the fraction of the mini-batch sampled from $\mathcal{P}_l$. Given a mini-batch size of N, we have $|\mathcal{X}_l| = \lfloor \alpha \cdot N \rfloor$ and $|\mathcal{X}_u| = N - |\mathcal{X}_l|$. When $\alpha = 0$, this reduces to the default view generation scheme. Starting with data transformation sets $\mathcal{T}_1$ and $\mathcal{T}_2$, we compute $x_{u,i}^{(1)} = t_1(x_{u,i})$ and $x_{u,i}^{(2)} = t_2(x_{u,i})$ with $t_1 \sim \mathcal{T}_1$ and $t_2 \sim \mathcal{T}_2$ for all $x_{u,i} \in \mathcal{X}_\sqcap$. These augmented instances are combined to form views $v_{u,1} = \left\{x_{u,1}^{(1)}, x_{u,2}^{(1)}, \ldots, x_{u,|\mathcal{X}_u|}^{(1)}\right\}$ and $v_{u,2} = \left\{x_{u,1}^{(2)}, x_{u,2}^{(2)}, \ldots, x_{u,|\mathcal{X}_u|}^{(2)}\right\}$. Analogously, we produce augmented views of the labeled mini-batch $\mathcal{X}_l$, denoted as $v_{l,1} = \left\{x_{l,1}^{(1)}, x_{l,2}^{(1)}, \ldots, x_{l,|\mathcal{X}_l|}^{(1)}\right\}$ and $v_{l,2} = \left\{x_{l,1}^{(2)}, x_{l,2}^{(2)}, \ldots, x_{l,|\mathcal{X}_l|}^{(2)}\right\}$. Next, we describe the view permutation process on labeled views $v_{l,1}$ and $v_{l,2}$.

View permutation. Given labeled augmented views, $v_{l,1}$ and $v_{l,2}$, we define a bijective permutation function $\pi : \{1, \ldots, |\mathcal{X}_l|\} \to \{1, \ldots, |\mathcal{X}_l|\}$ to generate a random permutation of $v_{l,2}$ denoted as $\tilde{v}_{l,2}$. Our permuted view, $\tilde{v}_{l,2}$, is defined as $\tilde{v}_{l,2} = \{x_{l,\pi(i)}^{(2)} : x_{l,i}^{(2)} \in v_{l,2}, y_i = y_{\pi(i)}\}$. Now, $\tilde{v}_{l,2}$ is a permutation of $v_{l,2}$ where the original image differs, but the ground-truth class

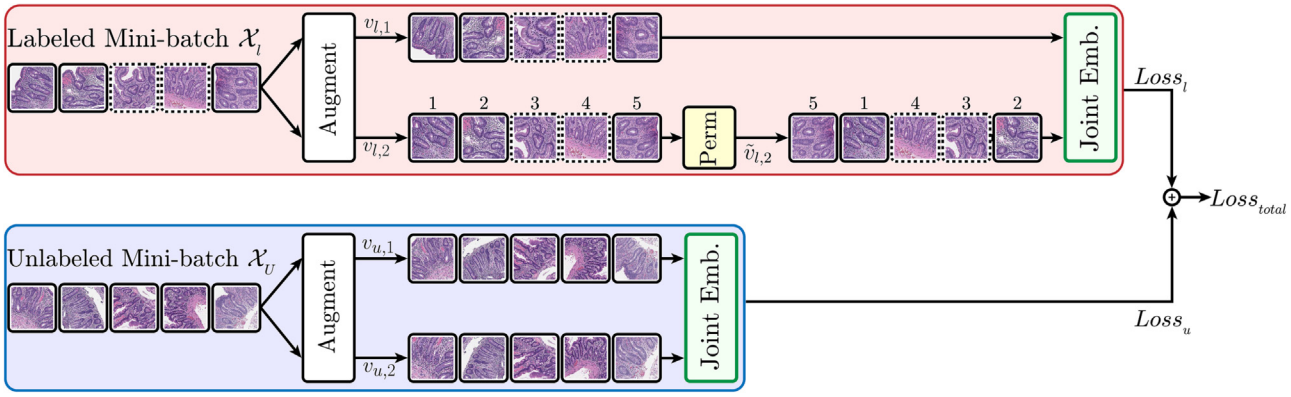

Fig. 1. Overview of our HistoPerm method. The joint embedding networks are fed randomly augmented views $v_{*,1}$ and $v_{*,2}$. For the labeled mini-batch of patches, $\mathcal{X}_l$, the solid or dashed patch outlines represent the labels. The numbers for labeled views $v_{l,2}$ and $\tilde{v}_{l,2}$ show the change in patch order before and after the permutation operation, respectively. The unlabeled mini-batch of views, $\mathcal{X}_u$, is fed to joint embedding networks without permutation as in the standard architectures.

is the same. Through this permutation, we augment the size of possible view pairings, enabling the model to learn richer representations. Note that it was an arbitrary choice to permute $v_{l,2}$, and due to symmetry either view could be permuted without loss of generality. Moreover, we only shuffle $v_{l,2}$ and not both $v_{l,1}$ and $v_{l,2}$ as shuffling both does not change the desired outcome.

Experimental setup

Datasets

We applied and evaluated our approach on 2 datasets from the Dartmouth-Hitchcock Medical Center (DHMC), a tertiary academic medical center in New Hampshire, USA. This study, and the usage of human participant data in this project, were approved by the Dartmouth-Hitchcock Medical Center Institutional Review Board (IRB) with a waiver of informed consent. Our datasets are representative of Celiac Disease (CD) and Renal Cell Carcinoma (RCC). Each dataset consists of hematoxylin-eosin-stained, formalin-fixed, paraffin-embedded slides scanned at either 20× (0.5 μm/pixel) or 40× (0.25 μm/pixel) magnification. For run-time purposes, we downsampled the slides to 5× (2 μm/pixel) magnification using the Lanczos filter.[70] We divided the slides into overlapping 224×224-pixel patches for use with the PyTorch deep learning framework.[71] A different overlapping factor was used across each class in the training set to produce approximately 80 000 patches per class. For the development and testing sets, we used a constant overlap factor of 112 pixels among patches. We provide dataset statistics in the supplementary material. Although these datasets are labeled in their original form, we have ignored the labels for a portion of the dataset used in the unlabeled section of the architecture in each epoch according to the formulation provided in the Method Section to simulate the intended use of our approach. This means that $\mathcal{D}_l$ and $\mathcal{D}_u$ are varying and built dynamically in each epoch where $|\mathcal{D}_l| = \lfloor \alpha \cdot |\mathcal{D}| \rfloor$ and $|\mathcal{D}_u| = |\mathcal{D}| - |\mathcal{D}_l|$. When the desired α does not match the proportion of unlabeled to labeled data, either under- or oversampling can be employed. Alternatively, the value of α could be modified to match the ratio of unlabeled to labeled data.

Implementation details

Image augmentation. We used a typical set of image augmentations in our experiments according to common joint embedding architecture-based representation learning methods. A crop from each image is randomly selected and resized to 224×224 pixels with bilinear interpolation. Next, we randomly flip the patches over both the horizontal and vertical axes, as histology patches are rotation invariant. Finally, we performed random Gaussian blurring on the augmented images. Empirical justification, as well as exact implementation details for these transformations, are provided in the supplementary material.

Pretraining. In the pretraining phase, we used the LARS optimizer[74] for 50 epochs of training the networks with a 5-epoch warm-up and cosine learning rate decay[75] thereafter. The initial learning rate was 0.45 with a mini-batch size of 256 and weight decay of 10^{-6}. We choose $\alpha = 0.75$ (i.e., 64 unlabeled and 192 labeled examples) as the optimal balance between the unlabeled and labeled portions of the mini-batch. We provide details of how we selected $\alpha = 0.75$ in the supplementary material. For experiments without HistoPerm, all 256 examples in the mini-batch are considered unlabeled.

Linear evaluation. Linear training uses the SGD optimizer with Nesterov momentum[76] for 80 epochs of training a linear layer on top of the frozen encoders with cosine learning rate decay.[75] Moreover, we use the cross-entropy loss for our objective as we are solving a multi-class classification problem for both datasets. We used an initial learning rate of 0.2 and a mini-batch size of 256. Unlike the pretraining step, we only performed affine transformations to the input data. In this phase, we utilized all data in the respective training set.

Results

We now present the performance of HistoPerm in the patch- and slide-level classification scenarios. All presented experimental configurations were run 3 times to account for run-to-run variation from the stochastic nature of the optimization process. For each configuration, we report the mean and standard deviation of the metrics across the 3 runs.

Patch-level results

First, we investigated the effect of HistoPerm on model patch-level classification performance. Table 1 shows that the use of HistoPerm consistently resulted in improved accuracy compared to baseline approaches across all datasets. Specifically, BYOL with HistoPerm outperformed standard BYOL by 8% and 2% on the CD and RCC datasets, respectively, in terms of classification accuracy. SimCLR with HistoPerm also demonstrated improved accuracy by 3% and 1% on the CD and RCC datasets, respectively. Additionally, the incorporation of HistoPerm into VICReg led to an increase in accuracy by 8% and 2% on the CD and RCC datasets, respectively.

Slide-level results

In Table 2, we present the effects of HistoPerm on slide-level classification performance. For slide-level classification, we utilized average-pooling to aggregate the patch-level predictions. This slide-level aggregation

Table 1
Patch-level linear performance results on the respective test sets. All reported values are the mean of 3 different runs with standard deviation in parentheses. The top results for each architecture are presented in boldface.

	Celiac disease			Renal cell carcinoma		
Method	Accuracy	F1-score	AUC	Accuracy	F1-score	AUC
BYOL	0.7958 (0.0205)	0.7750 (0.0286)	0.9427 (0.0092)	0.5802 (0.0072)	0.5334 (0.0151)	0.8390 (0.0073)
BYOL + HistoPerm	**0.8770 (0.0049)**	**0.8773 (0.0062)**	**0.9721 (0.0018)**	**0.6084 (0.0101)**	**0.5604 (0.0055)**	**0.8530 (0.0066)**
SimCLR	0.8507 (0.0031)	0.8473 (0.0038)	0.9583 (0.0016)	0.5920 (0.0069)	0.5359 (0.0029)	0.8452 (0.0104)
SimCLR + HistoPerm	**0.8855 (0.0057)**	**0.8832 (0.0061)**	**0.9767 (0.0023)**	**0.6033 (0.0121)**	**0.5433 (0.0109)**	**0.8634 (0.0041)**
VICReg	0.7717 (0.0164)	0.7394 (0.0266)	0.9218 (0.0073)	0.5621 (0.0033)	0.4940 (0.0075)	0.8074 (0.0068)
VICReg + HistoPerm	**0.8501 (0.0092)**	**0.8442 (0.0109)**	**0.9602 (0.0039)**	**0.5890 (0.0005)**	**0.5336 (0.0033)**	**0.8263 (0.0022)**

Table 2
Slide-level linear performance results on the respective test sets. All reported values are the mean of 3 different runs with standard deviation in parentheses. The top results for each architecture are presented in boldface. We provide the supervised results on the top row for comparison.

	Celiac disease			Renal cell carcinoma		
Method	Accuracy	F1-score	AUC	Accuracy	F1-score	AUC
Fully supervised	0.9167 (0.0111)	0.9168 (0.0109)	0.9856 (0.0005)	0.7393 (0.0074)	0.6716 (0.0122)	0.9608 (0.0071)
BYOL	0.8077 (0.0333)	0.7918 (0.0477)	0.9823 (0.0035)	0.6068 (0.0074)	0.5274 (0.0190)	0.9409 (0.0058)
BYOL + HistoPerm	**0.9808 (0.0000)**	**0.9804 (0.0000)**	**0.9967 (0.0015)**	**0.6410 (0.0128)**	**0.5661 (0.0111)**	**0.9477 (0.0097)**
SimCLR	0.9423 (0.0192)	0.9421 (0.0185)	0.9928 (0.0044)	**0.6410 (0.0256)**	**0.5695 (0.0333)**	0.9422 (0.0060)
SimCLR + HistoPerm	**0.9679 (0.0111)**	**0.9672 (0.0114)**	**0.9961 (0.0028)**	0.6282 (0.0222)	0.5331 (0.0319)	**0.9536 (0.0043)**
VICReg	0.7821 (0.0111)	0.7669 (0.0169)	0.9823 (0.0030)	0.5726 (0.0196)	0.4430 (0.0336)	0.9196 (0.0094)
VICReg + HistoPerm	**0.9423 (0.0192)**	**0.9424 (0.0204)**	**0.9978 (0.0014)**	**0.5897 (0.0000)**	**0.4888 (0.0063)**	**0.9287 (0.0032)**

approach is straightforward and keeps the evaluation focus on the impact of HistoPerm. Our results showed that incorporating HistoPerm improved performance for all cases of the CD dataset compared to the fully supervised baseline. On the RCC dataset, the models with HistoPerm showed improved performance for BYOL and VICReg, although all models fell short of the fully supervised baseline.

Discussion

In this study, we presented HistoPerm, an approach for generating views of histology images to improve representation learning. HistoPerm leverages the weakly labeled nature of histology images to expand the available pool of views. By expanding the available view pool, we improved the learned representation quality and observed enhanced downstream performance. Our results suggest that HistoPerm is a promising approach for medical image analysis in digital pathology when access to labeled data is limited.

We incorporated HistoPerm into BYOL, SimCLR, and VICReg, and showed improvement in classification performance on 2 histology datasets. At the patch level, adding HistoPerm to BYOL, SimCLR, and VICReg improved accuracy by 8%, 3%, and 8% on the CD dataset. Similarly, on the RCC dataset, models with HistoPerm outperformed on accuracy by 2%, 1%, and 2% for BYOL, SimCLR, and VICReg, respectively. For CD, we see that models with HistoPerm at the slide-level increase accuracy by 18%, 2%, and 22% for BYOL, SimCLR, and VICReg, respectively. On the RCC data, HistoPerm increases slide-level accuracy by 4% on BYOL and 1% on VICReg, but decreases performance by 2% for SimCLR. Critically, HistoPerm was able to outperform fully supervised models at the slide-level without patch-level annotations. These findings have important implications for using unlabeled histology images in clinical settings, as image annotation can be a labor-intensive and highly skilled process. Reducing the need for labeled data using HistoPerm, would increase the utility of existing representation learning approaches.

We demonstrated that the addition of HistoPerm can lead to a notable performance improvement on the CD dataset compared to the fully supervised baseline. However, this trend was not observed on the RCC dataset, where all models performed worse than the fully supervised baseline. For whole-slide classification, we used an average-pooling approach to aggregate the patch-level predictions. We expect that as we utilize more sophisticated approaches, like multi-head attention or self-attention, our slide-level classification results will outperform the presented results, including the fully supervised ones.

Of note, the results on the RCC dataset did not show as much improvement as those on the CD dataset. It is possible that this difference is due to the higher morphological complexity and variability of the RCC samples, as indicated by the original study on this dataset.[14] Despite the smaller improvements on the RCC dataset, the use of HistoPerm on both datasets showed clear benefits over standard representation learning approaches. In future work, we plan to investigate the relationship between histological pattern complexity and learned representation quality to enhance the ability of the model to generate more representative features. Furthermore, we intend on expanding our work to explore the biological underpinnings in depth.

While HistoPerm requires less labeled data than fully supervised approaches, it still requires some labeled data. In future work, we aim to reduce the labeled data requirements further for HistoPerm to enable use in labeled data-constrained settings. We also plan to examine the impact of incorporating unlabeled data from diverse data sources to explore the generalizability of HistoPerm across histology datasets with varied preparation and scanning procedures. In addition, we intend to utilize datasets from multiple disease types and evaluate the effectiveness of the learned histologic representations for transfer learning. This is particularly relevant as data for certain disease types may be scarce, and pretrained representations could provide a solution for building effective image analysis models. Due to the infeasible compute required, we did not perform thorough experimentation on the impact of mini-batch sizes on HistoPerm. In future work, we will explore the effects of the mini-batch size in order to determine the robustness of HistoPerm. Finally, we will explore datasets for tasks like survival prediction in future work.

Conclusion

The presented study showed that the proposed permutation-based view generation method, HistoPerm, offered improved histologic feature representations and resulted in enhanced classification accuracy compared to

current representation learning techniques. In some cases, HistoPerm even outperformed the fully supervised model. This approach allows for the incorporation of unlabeled histology data alongside labeled data for representation learning, resulting in overall higher classification performance. Additionally, the use of HistoPerm may reduce the annotation workload for pathologists, making it a viable option for various digital pathology applications.

Funding

This research was supported in part by grants from the US National Library of Medicine (R01LM012837 & R01LM013833) and the US National Cancer Institute (R01CA249758).

Declaration of generative AI and AI-assisted technologies in the writing process

The authors used AI-powered tools to proofread the manuscript. Any potential edits were solely editorial and were carefully reviewed by the authors. Otherwise, none of the presented content was generated by an AI model.

Conflict of interest statement

None declared.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jpi.2023.100320.

References

1. Zhu M, Ren B, Richards R, Suriawinata M, Tomita N, Hassanpour S. Development and evaluation of a deep neural network for histologic classification of renal cell carcinoma on biopsy and surgical resection slides. Sci Rep 2021;11(1):7080. https://doi.org/10.1038/s41598-021-86540-4.
2. Tomita N, Abdollahi B, Wei J, Ren B, Suriawinata A, Hassanpour S. Attention-based deep neural networks for detection of cancerous and precancerous esophagus tissue on histopathological slides. JAMA Netw Open 2019;2(11):e1914645. https://doi.org/10.1001/jamanetworkopen.2019.14645.
3. Wei JW, Wei JW, Jackson CR, Ren B, Suriawinata AA, Hassanpour S. Automated detection of celiac disease on duodenal biopsy slides: a deep learning approach. J Pathol Inform 2019;10(1):7. https://doi.org/10.4103/jpi.jpi_87_18.
4. Wei JW, Tafe LJ, Linnik YA, Vaickus LJ, Tomita N, Hassanpour S. Pathologist-level classification of histologic patterns on resected lung adenocarcinoma slides with deep neural networks. Sci Rep 2019;9(1):3358. https://doi.org/10.1038/s41598-019-40041-7.
5. Korbar B, Olofson A, Miraflor A, et al. Deep learning for classification of colorectal polyps on whole-slide images. J Pathol Inform 2017;8(1):30. https://doi.org/10.4103/jpi.jpi_34_17.
6. Jiang S, Zanazzi GJ, Hassanpour S. Predicting prognosis and IDH mutation status for patients with lower-grade gliomas using whole slide images. Sci Rep 2021;11(1):16849. https://doi.org/10.1038/s41598-021-95948-x.
7. Nasir-Moin M, Suriawinata AA, Ren B, et al. Evaluation of an artificial intelligence–augmented digital system for histologic classification of colorectal polyps. JAMA Netw Open 2021;4(11):e2135271. https://doi.org/10.1001/jamanetworkopen.2021.35271.
8. Tomita N, Tafe LJ, Suriawinata AA, et al. Predicting oncogene mutations of lung cancer using deep learning and histopathologic features on whole-slide images. Transl Oncol 2022;24, 101494. https://doi.org/10.1016/J.TRANON.2022.101494.
9. Russakovsky O, Deng J, Huang Z, Berg AC, Fei-Fei L. Detecting avocados to zucchinis: what have we done, and where are we going?. 2013 IEEE International Conference on Computer Vision; 2013. p. 2064–2071.
10. Bird B, Bedrossian K, Laver N, Miljković M, Romeo MJ, Diem M. Detection of breast micro-metastases in axillary lymph nodes by infrared micro-spectral imaging. Analyst 2009;134(6):1067–1076.
11. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In: *Pereira F, Burges CJ, Bottou L, Weinberger KQ, eds.* Advances in Neural Information Processing Systems. Curran Associates, Inc; 2012. https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf.
12. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2015. p. 3431–3440. https://doi.org/10.1109/CVPR.2015.7298965.
13. Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society; 2014. p. 580–587. https://doi.org/10.1109/CVPR.2014.81.
14. Zhu M, Ren B, Richards R, Suriawinata M, Tomita N, Hassanpour S. Development and evaluation of a deep neural network for histologic classification of renal cell carcinoma on biopsy and surgical resection slides. Sci Rep 2021;11(1):7080.
15. Wei JW, Wei JW, Jackson CR, Ren B, Suriawinata AA, Hassanpour S. Automated detection of celiac disease on duodenal biopsy slides: a deep learning approach. J Path Inform 2019;10(1):7. https://doi.org/10.4103/jpi.jpi_87_18.
16. Wei JW, Tafe LJ, Linnik YA, Vaickus LJ, Tomita N, Hassanpour S. Pathologist-level classification of histologic patterns on resected lung adenocarcinoma slides with deep neural networks. Sci Rep 2019;9(1):3358. https://doi.org/10.1038/s41598-019-40041-7.
17. Korbar B, Olofson A, Miraflor A, et al. Deep learning for classification of colorectal polyps on whole-slide images. J Path Inform 2017;8(1):30. https://doi.org/10.4103/jpi.jpi_34_17.
18. Liu Y, Gadepalli K, Norouzi M, et al. Detecting Cancer Metastases on Gigapixel Pathology ImagesPublished online. 2017. https://doi.org/10.48550/ARXIV.1703.02442.
19. Swiderska-Chadaj Z, Nurzynska K, Grala B, et al. A deep learning approach to assess the predominant tumor growth pattern in whole-slide images of lung adenocarcinoma. In: *Tomaszewski JE, Ward AD, eds.* Medical Imaging 2020: Digital Pathology. SPIE; 2020. p. 84–91. https://doi.org/10.1117/12.2549742.
20. Graham S, Shaban M, Qaiser T, Koohbanani NA, Khurram SA, Rajpoot N. Classification of lung cancer histology images using patch-level summary statistics. In: *Tomaszewski JE, Gurcan MN, eds.* Medical Imaging 2018: Digital Pathology. SPIE; 2018. p. 327–334. https://doi.org/10.1117/12.2293855.
21. Cruz-Roa A, Basavanhally A, González F, et al. Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks. In: *Gurcan MN, Madabhushi A, eds.* Medical Imaging 2014: Digital Pathology. SPIE; 2014. p. 1-15. https://doi.org/10.1117/12.2043872.
22. Tomita N, Abdollahi B, Wei J, Ren B, Suriawinata A, Hassanpour S. Attention-based deep neural networks for detection of cancerous and precancerous esophagus tissue on histopathological slides. JAMA Netw Open 2019;2(11):e1914645. https://doi.org/10.1001/jamanetworkopen.2019.14645.
23. BenTaieb A, Hamarneh G. Predicting cancer with a recurrent visual attention model for histopathology images. In: *Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G, eds.* MICCAI. Springer International Publishing; 2018. p. 129–137.
24. Ilse M, Tomczak J, Welling M. Attention-based deep multiple instance learning. In: *Dy J, Krause A, eds.* ICML. PMLR; 2018. p. 2127–2136.
25. Lerousseau M, Vakalopoulou M, Classe M, et al. Weakly supervised multiple instance learning histopathological tumor segmentation. In: *Martel AL, Abolmaesumi P, Stoyanov D, et al, eds.* MICCAI 2020. Springer International Publishing; 2020. p. 470–479.
26. Mercan C, Aksoy S, Mercan E, Shapiro LG, Weaver DL, Elmore JG. Multi-instance multi-label learning for multi-class classification of whole slide breast histopathology images. IEEE Trans Med Imag 2018;37(1):316–325.
27. Patil A, Tamboli D, Meena S, Anand D, Sethi A. Breast cancer histopathology image classification and localization using multiple instance learning. WIECON-ECE; 2019. p. 1–4. https://doi.org/10.1109/WIECON-ECE48653.2019.9019916.
28. Campanella G, Hanna MG, Geneslaw L, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nat Med 2019;25(8):1301–1309. https://doi.org/10.1038/s41591-019-0508-1.
29. Zhao Y, Yang F, Fang Y, et al. Predicting lymph node metastasis using histopathological images based on multiple instance learning with deep graph convolution. CVPR; 2020. p. 4836–4845. https://doi.org/10.1109/CVPR42600.2020.00489.
30. Katharopoulos A, Fleuret F. Processing megapixel images with deep attention-sampling models. In: *Chaudhuri K, Salakhutdinov R, eds.* ICML. PMLR; 2019. p. 3282–3291.
31. Tarkhan A, Nguyen TK, Simon N, Bengtsson T, Ocampo P, Dai J. Attention-based deep multiple instance learning with adaptive instance sampling. ISBI; 2022. p. 1–5. https://doi.org/10.1109/ISBI52829.2022.9761661.
32. Chikontwe P, Kim M, Nam SJ, Go H, Park SH. Multiple instance learning with center embeddings for histopathology classification. In: *Martel AL, Abolmaesumi P, Stoyanov D, et al, eds.* MICCAI. Springer International Publishing; 2020. p. 519–528.
33. Koohbanani NA, Unnikrishnan B, Khurram SA, Krishnaswamy P, Rajpoot N. Self-path: self-supervision for classification of pathology images with limited annotations. IEEE Trans Med Imag 2021;40(10):2845–2856. https://doi.org/10.1109/TMI.2021.3056023.
34. Sahasrabudhe M, Christodoulidis S, Salgado R, et al. Self-supervised nuclei segmentation in histopathological images using attention. In: *Martel AL, Abolmaesumi P, Stoyanov D, et al, eds.* MICCAI. Springer International Publishing; 2020. p. 393–402.
35. Chen RJ, Krishnan RG. *Self-supervised vision transformers learn visual concepts in histopathology. Learning meaningful representations of life. NeurIPS*. 2021.Published online.
36. Vu QD, Rajpoot K, Raza SEA, Rajpoot N. Handcrafted Histological Transformer (H2T): Unsupervised Representation of Whole Slide Images. Published Online. 2022. https://doi.org/10.48550/ARXIV.2202.07001.
37. Srinidhi CL, Martel AL. Improving self-supervised learning with hardness-aware dynamic curriculum learning: an application to digital pathology. ICCVW. IEEE Computer Society; 2021. p. 562–571. https://doi.org/10.1109/ICCVW54120.2021.00069.
38. Lu MY, Chen RJ, Wang J, Dillon D, Mahmood F. Semi-Supervised Histology Classification Using Deep Multiple Instance Learning and Contrastive Predictive CodingPublished online. . 2019. https://doi.org/10.48550/ARXIV.1910.10825.
39. Li B, Li Y, Eliceiri KW. Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning. CVPR; 2021. p. 14318–14328.
40. Stacke K, Unger J, Lundström C, Eilertsen G. Learning Representations with Contrastive Self-Supervised Learning for Histopathology ApplicationsPublished online 2021. https://doi.org/10.48550/ARXIV.2112.05760.
41. Dehaene O, Camara A, Moindrot O, de Lavergne A, Courtiol P. Self-Supervision Closes the Gap Between Weak and Strong Supervision in HistologyPublished online 2020. https://doi.org/10.48550/ARXIV.2012.03583.

42. Stacke K, Lundström C, Unger J, Eilertsen G. Evaluation of contrastive predictive coding for histopathology applications. In: *Alsentzer E, McDermott MBA, Falck F, Sarkar SK, Roy S, Hyland SL, eds.* ML4H NeurIPS Workshop. PMLR; 2020. p. 328–340.
43. Ke J, Shen Y, Liang X, Shen D. Contrastive learning based stain normalization across multiple tumor in histopathology. In: *de Bruijne M, Cattin PC, Cotin S, et al, eds.* MICCAI. Springer International Publishing; 2021. p. 571–580.
44. Ciga O, Xu T, Martel AL. Self supervised contrastive learning for digital histopathology. Machine Learning with Applications, 7. 2022, 100198. https://doi.org/10.1016/j.mlwa.2021.100198.
45. Gildenblat J, Klaiman E. Self-supervised similarity learning for digital pathology. MICCAI COMPAY Workshop. arXiv; 2019. https://doi.org/10.48550/ARXIV.1905.08139.
46. Li B, Li Y, Eliceiri KW. Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2021. p. 14318–14328.
47. Chen T, Kornblith S, Norouzi M, Hinton G. A simple framework for contrastive learning of visual representations. In: *III HD, Singh A, eds.* Proceedings of the 37th International Conference on Machine Learning. Proceedings of Machine Learning Research. PMLR; 2020. p. 1597–1607. https://proceedings.mlr.press/v119/chen20j.html.
48. Chen T, Kornblith S, Swersky K, Norouzi M, Hinton GE. Big self-supervised models are strong semi-supervised learners. In: *Larochelle H, Ranzato M, Hadsell R, Balcan MF, Lin H, eds.* Advances in Neural Information Processing Systems. Curran Associates, Inc; 2020. p. 22243–22255. https://proceedings.neurips.cc/paper/2020/file/fcbc95ccdd551da181207c0c1400c655-Paper.pdf.
49. He K, Fan H, Wu Y, Xie S, Girshick R. Momentum contrast for unsupervised visual representation learning. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2020. p. 9726–9735. https://doi.org/10.1109/CVPR42600.2020.00975.
50. Chen X, Fan H, Girshick R, He K. *Improved Baselines with Momentum Contrastive Learning.* 2020.
51. Dwibedi D, Aytar Y, Tompson J, Sermanet P, Zisserman A. With a little help from my friends: nearest-neighbor contrastive learning of visual representations. 2021 IEEE/CVF International Conference on Computer Vision (ICCV); 2021. p. 9588–9597.
52. van den Oord A, Li Y, Vinyals O. Representation Learning with Contrastive Predictive CodingPublished online. 2018. https://doi.org/10.48550/ARXIV.1807.03748.
53. Hénaff OJ, Srinivas A, De Fauw J, et al. Data-efficient image recognition with contrastive predictive coding. Proceedings of the 37th International Conference on Machine Learning. arXiv; 2020. p. 4182–4192. https://doi.org/10.48550/ARXIV.1905.09272.
54. Grill JB, Strub F, Altché F, et al. Bootstrap your own latent - a new approach to self-supervised learning. In: *Larochelle H, Ranzato M, Hadsell R, Balcan MF, Lin H, eds.* Advances in Neural Information Processing Systems. Curran Associates, Inc; 2020. p. 21271–21284-https://proceedings.neurips.cc/paper/2020/file/f3ada80d5c4ee70142b17b8192b2958e-Paper.pdf.
55. Chen X, He K. Exploring simple siamese representation learning. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2021. p. 15750–15758.
56. Azabou M, Azar MG, Liu R, et al. Mine Your Own vieW: Self-Supervised Learning Through Across-Sample PredictionPublished online. 2021.
57. Bardes A, Ponce J, LeCun Y. VICReg: variance-invariance-covariance regularization for self-supervised learning. International Conference on Learning Representations; 2022. https://openreview.net/forum?id=xm6YD62D1Ub.
58. Bardes A, Ponce J, LeCun Y. VICRegL: self-supervised learning of local visual features. In: *Oh AH, Aga .rwal A, Belgrave D, Cho K, eds.* Advances in Neural Information Processing Systems; 2022. https://openreview.net/forum?id=ePZsWeGJXyp.
59. Zbontar J, Jing L, Misra I, LeCun Y, Deny S. Barlow twins: self-supervised learning via redundancy reduction. In: *Meila M, Zhang T, eds.* Proceedings of the 38th International Conference on Machine Learning. PMLR; 2021. p. 12310–12320.
60. Ermolov A, Siarohin A, Sangineto E, Sebe N. Whitening for self-supervised representation learning. In: *Meila M, Zhang T, eds.* Proceedings of the 38th International Conference on Machine Learning. PMLR; 2020. p. 3015–3024. https://doi.org/10.48550/ARXIV.2007.06346.
61. Chen X, He K. Exploring simple Siamese representation learning. CVPR; 2021. p. 15750–15758.
62. Richemond PH, Grill JB, Altché F, et al. BYOL Works Even Without Batch StatisticsPublished online. 2020. https://doi.org/10.48550/ARXIV.2010.10241.
63. Fetterman A, Albrecht J. Understanding Self-Supervised and Contrastive Learning with "Bootstrap Your Own Latent" (BYOL). Published online: https://untitled-ai.github.io/understanding-self-supervised-contrastive-learning.html April 2020.
64. Zbontar J, Jing L, Misra I, LeCun Y, Deny S. Barlow twins: self-supervised learning via redundancy reduction. In: *Meila M, Zhang T, eds.* ICML. PMLR; 2021. p. 12310–12320.
65. Ermolov A, Siarohin A, Sangineto E, Sebe N. Whitening for self-supervised representation learning. In: *Meila M, Zhang T, eds.* ICML. PMLR; 2020. p. 3015–3024. https://doi.org/10.48550/ARXIV.2007.06346.
66. Bardes A, Ponce J, LeCun Y. *VICReg: variance-invariance-covariance regularization for self-supervised learning.* ICLR. 2022. https://openreview.net/forum?id=xm6YD62D1Ub.
67. Bardes A, Ponce J, LeCun Y. VICRegL: self-supervised learning of local visual features. In: *Oh AH, Aga .rwal A, Belgrave D, Cho K, eds.* Advances in Neural Information Processing Systems; 2022. https://openreview.net/forum?id=ePZsWeGJXyp.
68. Tian Y, Sun C, Poole B, Krishnan D, Schmid C, Isola P. What makes for good views for contrastive learning?. In: *Larochelle H, Ranzato M, Hadsell R, Balcan MF, Lin H, eds.* NeurIPS. Curran Associates, Inc; 2020. p. 6827–6839.
69. Khosla P, Teterwak P, Wang C, et al. Supervised contrastive learning. In: *Larochelle H, Ranzato M, Hadsell R, Balcan MF, Lin H, eds.* NeurIPS. Curran Associates, Inc; 2020. p. 18661–18673.
70. Turkowski K. *Filters for Common Resampling Tasks.* Academic Press Professional, Inc. 1990.
71. PyTorch: an imperative style, high-performance deep learning library. In: *Paszke A, Gross S, Massa F, et al, eds.* NeurIPS. Curran Associates, Inc; 2019. p. 8024–8035.
72. You Y, Gitman I, Ginsburg B. Large Batch Training of Convolutional NetworksPublished online. 2017. https://doi.org/10.48550/ARXIV.1708.03888.
73. Loshchilov I, Hutter F. SGDR: stochastic gradient descent with warm restarts. ICLR; 2017.
74. Sutskever I, Martens J, Dahl G, Hinton G. On the importance of initialization and momentum in deep learning. In: *Dasgupta S, McAllester D, eds.* ICML. PMLR; 2013. p. 1139–1147.