Proceedings of Machine Learning Research 1–15

StainNet: A Special Staining Self-Supervised Vision Transformer for Computational Pathology

Jiawen Li[*,1] jw-li24@mails.tsinghua.edu.cn
Jiali Hu[*,2] 72522269@cityu-dg.edu.cn
Xitong Ling[*,1] lingxt23@mails.tsinghua.edu.cn
Yongqiang Lv[4] 18146676064@163.com
Yuxuan Chen[1] chenyx23@mails.tsinghua.edu.cn
Yizhi Wang[1] 2022214454wang@gmail.com
Tian Guan[†,1] guantian@sz.tsinghua.edu.cn
Yifei Liu[†,3] 78909008@qq.com
Yonghong He[†,1] heyh@sz.tsinghua.edu.cn
[1] *Tsinghua Shenzhen International Graduate School, Tsinghua University, China*
[2] *Biomedical Engineering Programme, City University of Hong Kong (Dongguan), China*
[3] *Department of Pathology, Affiliated Hospital of Nantong University, Nantong, Jiangsu, China*
[4] *Nantong University, Nantong, Jiangsu, China*

Abstract

Foundation models trained with self-supervised learning (SSL) on large-scale histological images have significantly accelerated the development of computational pathology. These models can serve as backbones for region-of-interest (ROI) image analysis or patch-level feature extractors in whole-slide images (WSIs) based on multiple instance learning (MIL). Existing pathology foundation models (PFMs) are typically pre-trained on Hematoxylin-Eosin (H&E) stained pathology images. However, images with special stains, such as immunohistochemistry, are also frequently used in clinical practice. PFMs pre-trained mainly on H&E-stained images may be limited in clinical applications involving special stains. To address this issue, we propose StainNet, a specialized foundation model for special stains based on the vision transformer (ViT) architecture. StainNet adopts a self-distillation SSL approach and is trained on over 1.4 million patch images cropping from 20,231 publicly available special staining WSIs in the HISTAI database. To evaluate StainNet, we conduct experiments on an in-house slide-level liver malignancy classification task and two public ROI-level datasets to demonstrate its strong ability. We also perform few-ratio learning and retrieval evaluations, and compare StainNet with recently larger PFMs to further highlight its strengths. We have released the StainNet model weights at: https://huggingface.co/JWonderLand/StainNet.

Keywords: Special Staining, foundation model, computational pathology.

1. Introduction

Modern computational pathology involves the use of deep learning models, such as convolutional neural networks (He et al., 2016) and ViT (Dosovitskiy, 2020), for clinical tasks on histopathological images, including cancer screening (Bejnordi et al., 2017), tumor analysis

* Contributed equally
† Corresponding author

(Lu et al., 2021), morphological retrieval (Chen et al., 2022a), and survival prognosis (Chen et al., 2021a). In recent years, SSL methods have gained attention and have been proven to be highly effective (Kang et al., 2023). These methods design pretext tasks, such as contrastive learning (Chen et al., 2020) and image reconstruction (He et al., 2022), allowing the model to learn effective representations from unlabeled data and obtain good initialization weights, which can then be directly used as foundation models for pathology image feature extraction.

Most existing PFMs are pre-trained on H&E-stained histological images, which are routinely used and easily accessible in clinical workflows. As a result, these PFMs typically perform well with H&E-stained images. For example, by directly analyzing expert-annotated H&E-stained ROIs, PFMs can be used to identify tumor malignancy or microenvironment patterns (Lin et al., 2025). Additionally, when H&E-stained resection or biopsy WSI are cropped into patches, PFMs can extract these into high-dimensional features, which are used in weakly-supervised learning tasks based on MIL (Xu et al., 2025), such as cancer screening or biomarker detection (Campanella et al., 2025). However, there are other staining techniques in clinical practice. For example, by using antibodies that bind to proteins in tissue cells, immunohistochemistry (IHC) slides are widely used to visualize the presence, location, and expression levels of specific proteins, which are commonly utilized for precision cancer diagnosis, disease subtyping, and drug selection. Due to significant differences in nuclear color, tissue patterns, and texture compared to H&E staining, directly transferring existing PFMs to special-stained images for representation may be limited in clinical performance (Li et al., 2024b).

To address this issue, we propose StainNet, a ViT-based PFM specifically pre-trained on special-staining images. Specifically, to train StainNet, we first select 20,231 non-H&E-stained WSIs from the publicly available large-scale HISTAI database (Nechaev et al., 2025). For each WSI, we randomly crop up to 100 tissue-containing patch images to create our pre-training dataset. Next, to develop StainNet, we utilize DINO (Caron et al., 2021), a lightweight self-supervised learning framework specifically designed for ViT models, enabling StainNet to learn task-agnostic morphological information from a large collection of unlabeled special staining images. Finally, to evaluate StainNet, we construct downstream task datasets at both the ROI and WSI levels and compare its performance with existing pathology foundation models. Through this work, we aim to explore the adaptability of PFMs in clinical tasks involving special-stained images and call for the broader exploration of computational pathology in more diverse clinical pathology image datasets.

2. Related works

2.1. Self-supervised PFMs

Foundation models are accelerating the advancement of computational pathology. By performing SSL on large-scale unlabeled pathology images, models can obtain histology-specific weights, which significantly improve their performance on downstream pathology tasks. For instance, CTransPath (Wang et al., 2022) uses a MoCov3-based SSL approach (Chen et al., 2021b) and is trained on 15.6 million image patches from 25 anatomical sites to demonstrate its ability in H&E staining image retrieval. HIPT (Chen et al., 2022b) and Lunit (Kang et al., 2023) adopt DINO (Caron et al., 2021), validating the feasibility of pre-training

ViT models on histological images. Given the easy scalability of ViT parameters, recent works have aimed to train on larger datasets with larger ViT models. Examples include UNI (Chen et al., 2024), Virchow (Vorontsov et al., 2024), and Prov-Gigapath (Xu et al., 2024), which aim to diversify the pre-training dataset with more varied pathology images and leverage more computational resources to build stronger general pathology representation models. Additionally, some studies explore SSL in conjunction with special staining, such as PathoDuet (Hua et al., 2024), which utilizes registered H&E and IHC images to help the model learn cross-staining features, aiding the transfer to more advanced nuclear information.

2.2. Special staining in clinical pathology

H&E staining is the most commonly used histological staining method, providing basic structural information by staining the cell nuclei and cytoplasm. However, H&E staining does not fully reveal certain cell features or tissue details, especially in the detection of immune responses or protein expression. As a result, several special staining methods have been developed to address these limitations (Gridley, 1957). Common special staining methods, such as IHC, Masson's trichrome, and acid phosphatase staining, can highlight specific proteins, cell subtypes, or extracellular matrix components. For example, IHC utilizes antibodies to bind to target proteins, making it a valuable tool for cancer diagnosis and disease subtyping (Magaki et al., 2018); Masson's trichrome staining is employed to visualize elastic fibers in tissues, particularly in fibrosis-related research (Lefkowitch, 2006). Recently, the development of computational pathology has accelerated research into the application of computer vision models to these special staining images, such as lymphocyte detection in IHC images (Swiderska-Chadaj et al., 2019) and the use of multiple special stains for kidney biopsy evaluation (Jayapandian et al., 2021). Some studies have also explored the direct use of generative models to convert H&E images into special stains to improve diagnostic accuracy (De Haan et al., 2021; Bai et al., 2023; Pati et al., 2024; Yan et al., 2023). However, these special staining images differ significantly from H&E staining in terms of staining mechanisms, texture features, and visual appearance, making it difficult for large-scale pre-trained models based on general H&E staining images to capture the specific features of special staining images (Li et al., 2024b).

3. Methodology

We develop and evaluate our StainNet through data collection, model pre-training, and downstream testing. The entire process is illustrated in Figure 1.

3.1. Data collection

We use the publicly available large-scale WSI database HISTAI (Nechaev et al., 2025) as our pre-training data. The original HISTAI dataset contains 112,801 WSIs, of which 20,265 are special stains. We first apply the OTSU thresholding algorithm (Otsu et al., 1975) to exclude 34 special staining WSIs that can not extract foreground tissue, resulting in 20,231 WSIs. Then, for each WSI, we perform non-overlapping 224×224 pixel sliding window operations on the foreground tissue, and randomly crop 100 image patches (if a

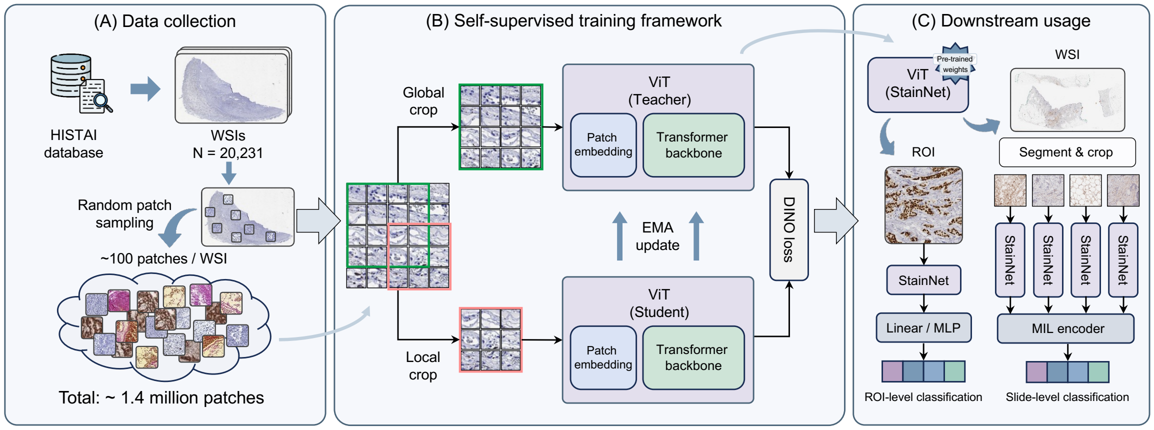

Figure 1: Overview of our proposed StainNet. We first collect approximately 1.4 million patch images with a resolution of 224×224 pixels from the HISTAI database. Then, we train StainNet using the DINO SSL method. Finally, we adapt StainNet to downstream ROI and WSI tasks to explore its performance.

WSI contains fewer than 100 patches, all the patches will be selected). In total, we obtain 1,418,938 patches for subsequent StainNet pre-training.

3.2. Model pre-training

Our proposed StainNet adopts the SSL method DINO (Caron et al., 2021) as the pre-training strategy to let the ViT model learn robust feature representations from special-staining pathology images without manual annotations. DINO consists of two networks with the same architecture: a student network optimized via gradient descent and a teacher network updated using an exponential moving average (EMA) of the student weights. We follow the original DINO approach for data augmentation and multi-crop strategies. Specifically, we generate a set of global and local views for each pathology image, and the student takes all views as input, while the teacher only receives the global views. By minimizing the cross-entropy loss between the output probability distributions of the two networks, the student is encouraged to infer global tissue context solely from local information. In addition, centering and sharpening operations are applied to the teacher outputs to prevent mode collapse. Figure 2 illustrates some examples of original images and their augmented versions. The detailed hyperparameters are provided in the implementation.

3.3. Downstream adaptation

Our pre-trained StainNet can be applied to both ROI and WSI analysis for special staining images. For ROI tasks, the extracted features from StainNet can be passed to either a single linear layer (linear probe) or a multilayer perceptron (MLP probe) consisting of two linear layers with a ReLU($\cdot$) activation function for downstream classification. For WSI tasks, we first crop a set of tissue-containing patch images and then perform offline feature extraction using StainNet. The resulting $N \times C$ feature sequence (where N is the number of image

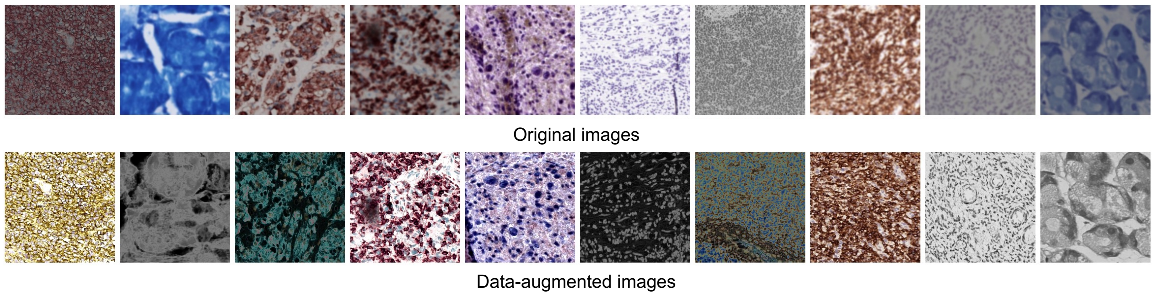

Figure 2: A set of original special staining images and their data-augmented versions.

patches and C is the feature dimension) can be fed into a series of MIL models to obtain slide-level representations and classification scores.

4. Experiments

4.1. Downstream datasets and implementation

We evaluate StainNet on one private WSI task NTUH-*Ki67*-Liver, and two publicly available ROI tasks BCI (Liu et al., 2022) and MIST (Li et al., 2023). The NTUH-*Ki67*-Liver 3-class dataset consists of 46 *Ki67* benign liver slides, 43 *Ki67* hepatocellular carcinoma slides, and 50 *Ki67* normal liver slides. All slides are collected from the Affiliated Hospital of Nantong University and scanned using the SQS-120P scanning system from Shenzhen Shengqiang Technology Co., Ltd. The annotations are performed and verified at the slide level by two pathologists. The BCI dataset is a collection of 4,870 paired H&E-IHC images showing *HER2* biomarker expression. We use the IHC images and categorize them into four *HER2* expression levels: 0, 1+, 2+, and 3+, with 240, 1,153, 2,142, and 1,355 samples, respectively. The MIST dataset is a large-scale H&E-IHC paired image dataset. We use the IHC images corresponding to four biomarkers: 5,642 *HER2*, 5,361 *Ki67*, 5,153 *ER*, and 5,139 *PR* to construct a 4-class dataset for evaluating the ability of PFMs to identify biomarkers across different staining conditions.

We use the AdamW optimizer with a learning rate of 5×10^{-4} to pre-train StainNet for 100 epochs with a batch size of 256. For multi-cropping, global crops are sampled using a random resize scale in the range of 0.4-1.0, and 8 local crops are generated with a scale in the range of 0.05-0.4. For EMA design, we increase teacher momentum linearly from 0.9995 to 1.0 throughout training. The weight decay is scheduled from 0.04 to 0.4, and the teacher temperature is set to 0.04 without warm-up and last layer freezing. We use ViT-Small, pre-trained on ImageNet datasets (Deng et al., 2009), as the initial backbone and weights for our StainNet. We also enable normalization of the last layer to stabilize the training process. The experiment is conducted on a single NVIDIA RTX PRO 6000 96GB GPU with mixed-precision and takes approximately 3.04 days to complete. We select the final epoch checkpoint as the StainNet weights used in all subsequent experiments.

For downstream fine-tuning, we use AdamW with a learning rate of 10^{-4} and a weight decay of 10^{-4}. We set 20 epochs with a batch size of 1 for the WSI task and 20 epochs with a batch size of 128 for the ROI task. Early stopping is set to 5 epochs for both func-

Patch encoder *Bal acc.*	MIL encoder ABMIL	SiMLP	TransMIL	WiKG	AMDMIL	S4MIL	Overall
ResNet-50	$0.742_{0.093}$	$0.712_{0.055}$	$0.829_{0.055}$	$0.836_{0.059}$	$0.838_{0.071}$	$0.832_{0.076}$	0.7980
CTransPath	$0.668_{0.050}$	$0.757_{0.105}$	$0.858_{0.040}$	$0.860_{0.041}$	$0.802_{0.018}$	$0.819_{0.043}$	0.7940
PathoDuet	$0.655_{0.033}$	$0.641_{0.024}$	$0.699_{0.033}$	$0.738_{0.032}$	$0.716_{0.062}$	$0.725_{0.030}$	0.6957
HIPT	$0.706_{0.086}$	$0.753_{0.053}$	$0.814_{0.050}$	$0.815_{0.058}$	$0.835_{0.020}$	$0.843_{0.017}$	0.7943
Lunit	$0.747_{0.048}$	$0.838_{0.041}$	$0.872_{0.051}$	$0.891_{0.044}$	$0.872_{0.020}$	$0.862_{0.032}$	0.8470
StainNet (Ours)	$\mathbf{0.801}_{0.092}$	$\mathbf{0.873}_{0.042}$	$\mathbf{0.896}_{0.073}$	$\mathbf{0.916}_{0.030}$	$\mathbf{0.892}_{0.051}$	$\mathbf{0.883}_{0.060}$	**0.8769**

Patch encoder *AUC*	MIL encoder ABMIL	SiMLP	TransMIL	WiKG	AMDMIL	S4MIL	Overall
ResNet-50	$0.884_{0.040}$	$0.892_{0.041}$	$0.899_{0.049}$	$0.916_{0.022}$	$0.909_{0.024}$	$0.912_{0.043}$	0.9019
CTransPath	$0.876_{0.028}$	$0.885_{0.041}$	$0.940_{0.031}$	$0.940_{0.033}$	$0.912_{0.029}$	$0.923_{0.031}$	0.9127
PathoDuet	$0.815_{0.056}$	$0.848_{0.018}$	$0.856_{0.020}$	$0.838_{0.059}$	$0.821_{0.076}$	$0.852_{0.028}$	0.8382
HIPT	$0.838_{0.113}$	$0.884_{0.035}$	$0.906_{0.042}$	$0.907_{0.012}$	$0.914_{0.014}$	$0.930_{0.038}$	0.8964
Lunit	$0.854_{0.095}$	$0.936_{0.021}$	$0.948_{0.031}$	$0.947_{0.029}$	$0.930_{0.019}$	$0.950_{0.026}$	0.9275
StainNet (Ours)	$\mathbf{0.897}_{0.053}$	$\mathbf{0.948}_{0.029}$	$\mathbf{0.960}_{0.036}$	$\mathbf{0.964}_{0.020}$	$\mathbf{0.959}_{0.030}$	$\mathbf{0.968}_{0.027}$	**0.9492**

Patch encoder *F1-score*	MIL encoder ABMIL	SiMLP	TransMIL	WiKG	AMDMIL	S4MIL	Overall
ResNet-50	$0.737_{0.093}$	$0.689_{0.059}$	$0.828_{0.052}$	$0.835_{0.060}$	$0.837_{0.066}$	$0.833_{0.077}$	0.7931
CTransPath	$0.556_{0.084}$	$0.726_{0.122}$	$0.856_{0.042}$	$0.852_{0.045}$	$0.791_{0.026}$	$0.821_{0.040}$	0.7671
PathoDuet	$0.604_{0.077}$	$0.612_{0.030}$	$0.660_{0.069}$	$0.716_{0.038}$	$0.691_{0.073}$	$0.718_{0.036}$	0.6668
HIPT	$0.657_{0.126}$	$0.731_{0.072}$	$0.813_{0.049}$	$0.810_{0.065}$	$0.836_{0.019}$	$0.834_{0.027}$	0.7801
Lunit	$0.719_{0.095}$	$0.838_{0.043}$	$0.870_{0.054}$	$0.893_{0.044}$	$0.867_{0.020}$	$0.860_{0.028}$	0.8409
StainNet (Ours)	$\mathbf{0.777}_{0.141}$	$\mathbf{0.870}_{0.038}$	$\mathbf{0.891}_{0.076}$	$\mathbf{0.914}_{0.030}$	$\mathbf{0.890}_{0.058}$	$\mathbf{0.881}_{0.061}$	**0.8706**

Table 1: Comparison results of our StainNet with five similarly sized PFMs on the NTUH-*Ki67*-Liver classification dataset. We use six different MIL methods for slide-level fine-tuning and report average balanced accuracy with standard deviation across all folds.

tions. For classification tasks, we save the fine-tuned weights and report accuracy, balanced accuracy, AUC, and F1-score from each epoch that achieves the best balanced accuracy. For retrieval tasks, we use cosine similarity for image matching and report recall@K and mAP@K, where K ranges from 1 to 20. All experiments are conducted using stratified 5-fold cross-validation and full-precision computing on a single NVIDIA RTX 3090 24GB GPU. All baseline methods are trained using the same settings.

4.2. Comparison with PFMs of similar parameter scale

We compare the proposed StainNet with five PFMs of comparable parameter size: CTransPath (Wang et al., 2022), PathoDuet (Hua et al., 2024), HIPT (Chen et al., 2022b), Lunit (Kang et al., 2023), and ResNet-50 (ImageNet pre-trained) (He et al., 2016). For the WSI task, we

Patch encoder *BCI 4 classes datasets*	Linear probe *Acc.*	 *Bal acc.*	 *AUC*	 *F1-score*	MLP probe *Acc.*	 *Bal acc.*	 *AUC*	 *F1-score*
ResNet-50	$0.587_{0.010}$	$0.407_{0.010}$	$0.774_{0.018}$	$0.541_{0.014}$	$0.709_{0.016}$	$0.614_{0.019}$	$0.900_{0.008}$	$0.703_{0.016}$
CTransPath	$0.575_{0.011}$	$0.387_{0.007}$	$0.808_{0.007}$	$0.499_{0.008}$	$0.736_{0.015}$	$0.653_{0.021}$	$0.913_{0.008}$	$0.733_{0.015}$
PathoDuet	$0.612_{0.016}$	$0.430_{0.015}$	$0.810_{0.014}$	$0.554_{0.028}$	$0.687_{0.019}$	$0.564_{0.044}$	$0.877_{0.012}$	$0.664_{0.024}$
HIPT	$0.564_{0.023}$	$0.404_{0.023}$	$0.754_{0.020}$	$0.532_{0.028}$	$0.661_{0.009}$	$0.533_{0.010}$	$0.859_{0.010}$	$0.650_{0.011}$
Lunit	$0.696_{0.015}$	$0.545_{0.018}$	$0.880_{0.008}$	$0.681_{0.017}$	$0.872_{0.016}$	$0.816_{0.033}$	$0.976_{0.002}$	$0.871_{0.017}$
StainNet (Ours)	$\mathbf{0.709}_{0.025}$	$\mathbf{0.597}_{0.036}$	$\mathbf{0.892}_{0.023}$	$\mathbf{0.701}_{0.025}$	$\mathbf{0.874}_{0.010}$	$\mathbf{0.839}_{0.020}$	$\mathbf{0.979}_{0.004}$	$\mathbf{0.873}_{0.010}$
Patch encoder *MIST 4 classes datasets*	Linear probe *Acc.*	 *Bal acc.*	 *AUC*	 *F1-score*	MLP probe *Acc.*	 *Bal acc.*	 *AUC*	 *F1-score*
ResNet-50	$0.597_{0.009}$	$0.592_{0.009}$	$0.835_{0.002}$	$0.593_{0.008}$	$0.671_{0.005}$	$0.667_{0.004}$	$0.884_{0.002}$	$0.671_{0.005}$
CTransPath	$0.663_{0.008}$	$0.658_{0.008}$	$0.878_{0.004}$	$0.659_{0.006}$	$0.732_{0.007}$	$0.728_{0.007}$	$0.920_{0.004}$	$0.731_{0.007}$
PathoDuet	$0.590_{0.010}$	$0.585_{0.011}$	$0.827_{0.006}$	$0.586_{0.013}$	$0.664_{0.013}$	$0.660_{0.012}$	$0.885_{0.004}$	$0.664_{0.011}$
HIPT	$0.587_{0.007}$	$0.581_{0.007}$	$0.830_{0.006}$	$0.577_{0.007}$	$0.681_{0.006}$	$0.676_{0.007}$	$0.896_{0.004}$	$0.676_{0.010}$
Lunit	$0.738_{0.006}$	$0.733_{0.006}$	$0.922_{0.003}$	$0.736_{0.005}$	$0.871_{0.005}$	$0.869_{0.006}$	$0.977_{0.001}$	$0.871_{0.006}$
StainNet (Ours)	$\mathbf{0.779}_{0.006}$	$\mathbf{0.775}_{0.006}$	$\mathbf{0.941}_{0.002}$	$\mathbf{0.778}_{0.007}$	$\mathbf{0.884}_{0.006}$	$\mathbf{0.882}_{0.006}$	$\mathbf{0.982}_{0.002}$	$\mathbf{0.884}_{0.006}$

Table 2: Comparison results of our StainNet with five similarly sized PFMs on the BCI and MIST classification datasets. We use linear probe and MLP probe methods for ROI-level fine-tuning and report four average evaluation metrics with standard deviation across all folds.

evaluate six MIL models: (1) ABMIL (Ilse et al., 2018), a gated-attention MIL,(2) SiMLP (Li et al., 2025), a method combining unsupervised average pooling with an MLP classifier, (3) TransMIL (Shao et al., 2021), a transformer-based MIL, (4) WiKG (Li et al., 2024a), a dynamic graph representation method, (5) AMDMIL (Ling et al., 2024), a method combining masked denosing with agent tokens, and (6) S4MIL (Fillioux et al., 2023), a state space model-based MIL. For the ROI tasks, we freeze each PFM and apply two probe settings: a linear probe and an MLP probe for fine-tuning. Table 1 and Table 2 respectively show the results on the slide-level NTUH-*Ki67*-Liver task and ROI-level BCI and MIST tasks.

For NTUH-*Ki67*-Liver, StainNet consistently achieves the best overall performance across all MIL encoders. We also observe that lightweight MIL encoders exacerbate the performance gap between H&E-stained image pre-trained PFMs, whereas more complex MIL encoders mitigate this gap. For example, StainNet outperforms the second-best PFM in terms of balanced accuracy by 5.4% with ABMIL, whereas the margin narrows to 2.5% when using WiKG. For BCI and MIST, StainNet continues to outperform significantly, especially on MIST, where StainNet surpasses the popular CTransPath by 11.7% and 15.4% in balanced accuracy under the linear and MLP probe settings.

4.3. Comparison with larger PFMs

We further compare StainNet with three recently proposed larger PFMs: PathOrchestra (Yan et al., 2025), GPFM (Ma et al., 2025), and UNI (Chen et al., 2024) as shown in Figure 3. Although StainNet has nearly 14× fewer parameters than them (22M vs 303M), it still achieves the best performance. For example, StainNet surpasses these three large PFMs by

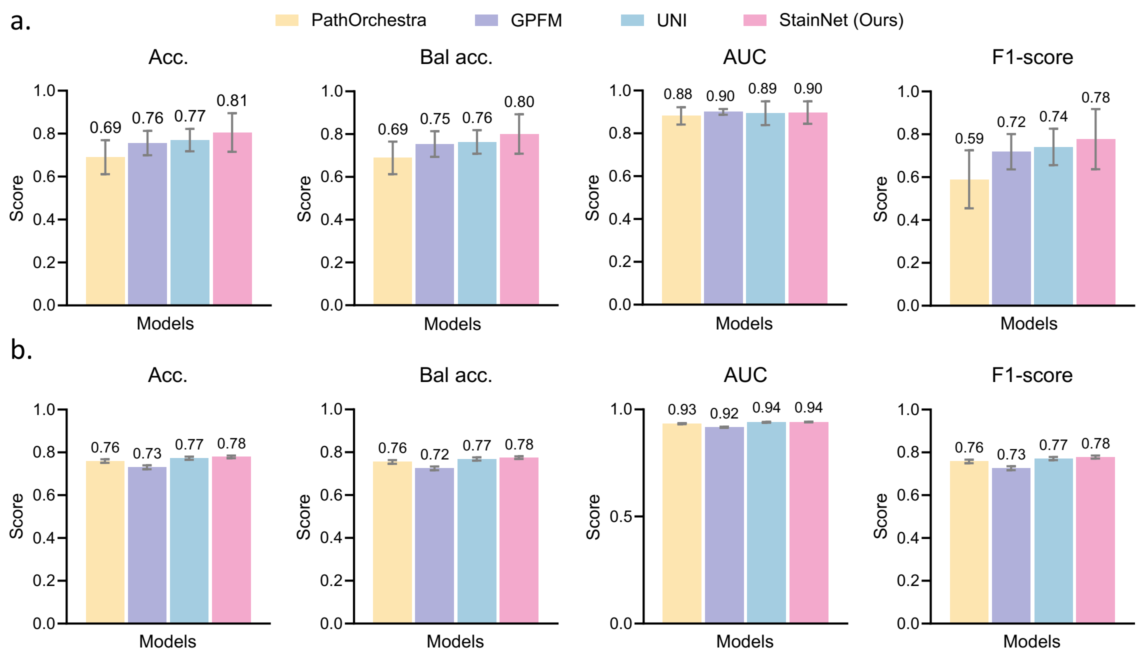

Figure 3: Comparison of StainNet with three larger PFMs across four evaluation metrics: accuracy, balanced accuracy, AUC, and F1-score. **a.** Performance on the NTUH-*Ki67*-Liver slide-level classification task with ABMIL fine-tuning. **b.** Performance on the MIST ROI-level classification task with linear probe fine-tuning.

3.72%, 4.70%, and 11.15% in terms of balanced accuracy (80.09% vs. 76.37%, 75.39%, and 68.94%). These results demonstrate that representations learned from large-scale special-stain images provide substantial advantages over H&E-based PFMs when applied to special-stain tasks.

4.4. Comparison of image retrieval

We also examine the retrieval capability of StainNet to assess its semantic representation quality on special staining histology images. We compare StainNet with CTransPath and UNI on the MIST dataset, as shown in Figure 4. We observe that StainNet consistently achieves the best Recall@K and mAP@K across all choices of K, indicating that it has learned a well-structured embedding space for special staining images.

4.5. Effectiveness of different fine-tuning data ratios

We further evaluate the fine-tuning performance of StainNet under varying amounts of downstream special-stain training data to assess its learning capability. Figure 5 presents the results of StainNet, CTransPath, and UNI on the MIST dataset using linear and MLP probes under different proportions of fine-tuning samples. We observe that although StainNet performs slightly worse than UNI under the 10% data setting with a linear probe,

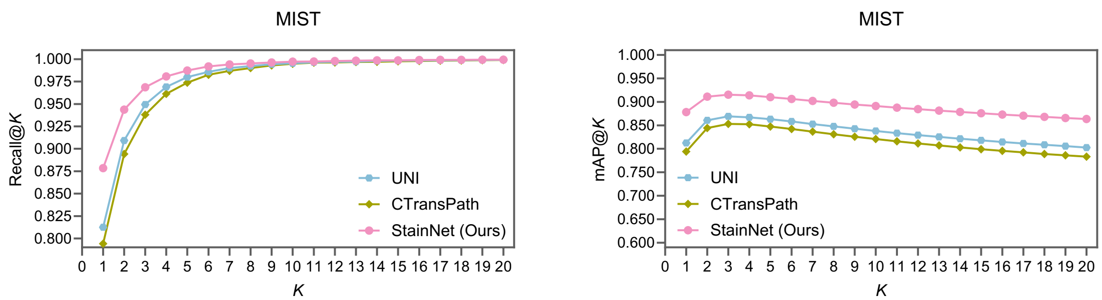

Figure 4: Comparison of image retrieval performance among CTransPath, UNI, and our StainNet on the MIST ROI-level dataset.

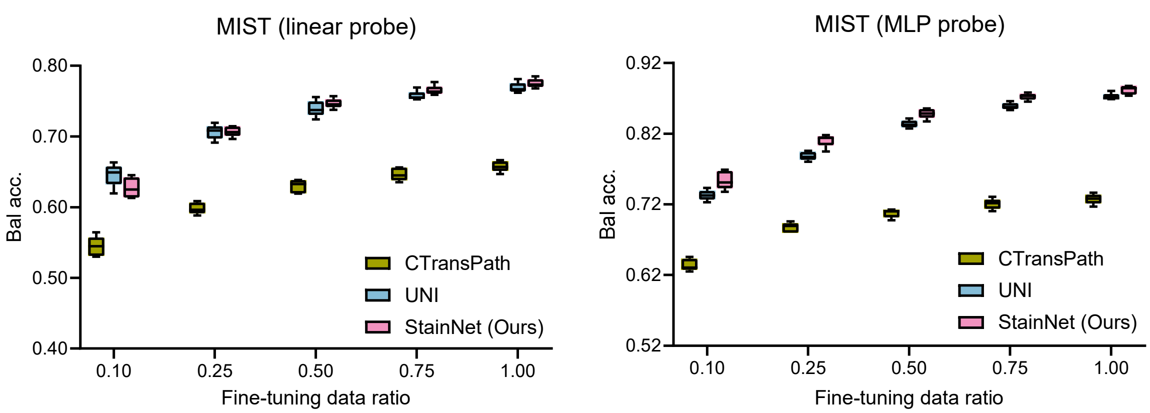

Figure 5: Linear probe and MIL probe fine-tuning results of CTransPath, UNI and our StainNet on the MIST ROI-level dataset across varying training data ratios.

it achieves the best performance in all other settings and substantially outperforms the CTransPath of similar parameter size. Notably, applying an MLP probe significantly increases the performance gap between StainNet and UNI in low-fine-tuning data scenarios.

4.6. Performance variation under different pre-training iterations

Finally, we investigate the performance variation of StainNet under different pre-training iterations and the behavior of its loss curve. We believe these observations can provide valuable insights for the development of SSL methods in computational pathology, where pre-training details such as the expected loss curve and convergence behavior are often underreported. Figure 6 illustrates the pre-training loss curve over 100 epochs and the corresponding linear probe performance on the MIST dataset at six iteration checkpoints. We observe that the loss plateaus between epoch 30 and 50, but with significant performance improvement during this period. After epoch 60, the overall loss begins to decrease again, but with slight changes in performance. Ultimately, the model reaches its optimal state at the final iteration. These findings suggest that the shift in the DINO pre-training loss

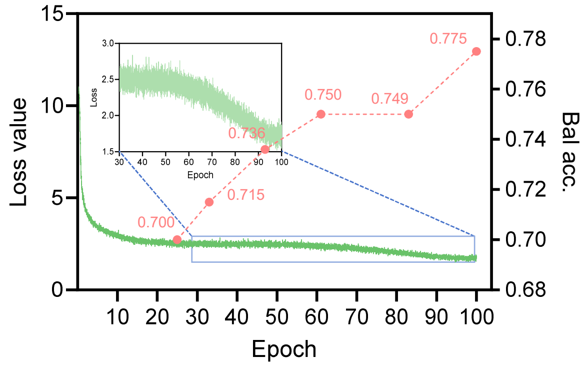

Figure 6: Pre-training loss curve of StainNet and comparison of linear probe performance on MIST under six different iterations.

may serve as a valuable indicator of representation quality. While additional exploration is required to generalize these observations to other settings, we encourage future SSL development in computational pathology to consider allocating more training epochs, which may be crucial for achieving optimal model performance.

5. Conclusion and future works

In this work, we propose StainNet, a lightweight pathology foundation model specifically for special staining histopathology image representation. Utilizing the DINO self-supervised framework, StainNet is trained from over 1.4 million unlabeled images. We evaluate StainNet on one in-house WSI task and two public ROI-level tasks, including classification, few-ratio settings, and image retrieval. The results demonstrate that StainNet learns strong and robust representations for special staining pathology images.

For future work, we plan to develop a larger StainNet by expanding the pre-training dataset, scaling the model parameter size, and evaluating it on broader downstream tasks. We will also explore integrating StainNet with H&E-based PFMs for multimodal applications such as survival prognosis and therapy response prediction. We believe that StainNet will further accelerate the research and applications in computational pathology involving diverse histological modalities.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (NSFC) (82430062), the Shenzhen Engineering Research Centre (XMHT20230115004), and the Tsinghua Shenzhen International Graduate School Cross-disciplinary Research and Innovation Fund Research Plan (JC2024002). We thank Qiehe Sun for providing GPU support.

References

Bijie Bai, Xilin Yang, Yuzhu Li, Yijie Zhang, Nir Pillar, and Aydogan Ozcan. Deep learning-enabled virtual histological staining of biological samples. *Light: Science & Applications*, 12(1):57, 2023.

Babak Ehteshami Bejnordi, Mitko Veta, Paul Johannes Van Diest, Bram Van Ginneken, Nico Karssemeijer, Geert Litjens, Jeroen AWM Van Der Laak, Meyke Hermsen, Quirine F Manson, Maschenka Balkenhol, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *Jama*, 318(22): 2199–2210, 2017.

Gabriele Campanella, Neeraj Kumar, Swaraj Nanda, Siddharth Singi, Eugene Fluder, Ricky Kwan, Silke Muehlstedt, Nicole Pfarr, Peter J Schüffler, Ida Häggström, et al. Real-world deployment of a fine-tuned pathology foundation model for lung cancer biomarker detection. *Nature Medicine*, 31(9):3002–3010, 2025.

Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9650–9660, 2021.

Chengkuan Chen, Ming Y Lu, Drew FK Williamson, Tiffany Y Chen, Andrew J Schaumberg, and Faisal Mahmood. Fast and scalable search of whole-slide images via self-supervised deep learning. *Nature Biomedical Engineering*, 6(12):1420–1434, 2022a.

Richard J Chen, Ming Y Lu, Muhammad Shaban, Chengkuan Chen, Tiffany Y Chen, Drew FK Williamson, and Faisal Mahmood. Whole slide images are 2d point clouds: Context-aware survival prediction using patch-based graph convolutional networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 339–349. Springer, 2021a.

Richard J Chen, Chengkuan Chen, Yicong Li, Tiffany Y Chen, Andrew D Trister, Rahul G Krishnan, and Faisal Mahmood. Scaling vision transformers to gigapixel images via hierarchical self-supervised learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16144–16155, 2022b.

Richard J Chen, Tong Ding, Ming Y Lu, Drew FK Williamson, Guillaume Jaume, Andrew H Song, Bowen Chen, Andrew Zhang, Daniel Shao, Muhammad Shaban, et al. Towards a general-purpose foundation model for computational pathology. *Nature medicine*, 30(3): 850–862, 2024.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PmLR, 2020.

Xinlei Chen, Saining Xie, and Kaiming He. An empirical study of training self-supervised vision transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9640–9649, 2021b.

Kevin De Haan, Yijie Zhang, Jonathan E Zuckerman, Tairan Liu, Anthony E Sisk, Miguel FP Diaz, Kuang-Yu Jen, Alexander Nobori, Sofia Liou, Sarah Zhang, et al. Deep learning-based transformation of h&e stained tissues into special stains. *Nature communications*, 12(1):4884, 2021.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

Alexey Dosovitskiy. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

Leo Fillioux, Joseph Boyd, Maria Vakalopoulou, Paul-Henry Cournède, and Stergios Christodoulidis. Structured state space models for multiple instance learning in digital pathology. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 594–604. Springer, 2023.

Mary Francis Gridley. *Manual of histologic and special staining technics*. Armed Forces Institute of Pathology, 1957.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16000–16009, 2022.

Shengyi Hua, Fang Yan, Tianle Shen, Lei Ma, and Xiaofan Zhang. Pathoduet: Foundation models for pathological slide analysis of h&e and ihc stains. *Medical Image Analysis*, 97: 103289, 2024.

Maximilian Ilse, Jakub Tomczak, and Max Welling. Attention-based deep multiple instance learning. In *International conference on machine learning*, pages 2127–2136. PMLR, 2018.

Catherine P Jayapandian, Yijiang Chen, Andrew R Janowczyk, Matthew B Palmer, Clarissa A Cassol, Miroslav Sekulic, Jeffrey B Hodgin, Jarcy Zee, Stephen M Hewitt, John O'Toole, et al. Development and evaluation of deep learning–based segmentation of histologic structures in the kidney cortex with multiple histologic stains. *Kidney international*, 99(1):86–101, 2021.

Mingu Kang, Heon Song, Seonwook Park, Donggeun Yoo, and Sérgio Pereira. Benchmarking self-supervised learning on diverse pathology datasets. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3344–3354, 2023.

Jay H Lefkowitch. Special stains in diagnostic liver pathology. In *Seminars in diagnostic pathology*, volume 23, pages 190–198. Elsevier, 2006.

Fangda Li, Zhiqiang Hu, Wen Chen, and Avinash Kak. Adaptive supervised patchnce loss for learning h&e-to-ihc stain translation with inconsistent groundtruth image pairs. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 632–641. Springer, 2023.

Jiawen Li, Yuxuan Chen, Hongbo Chu, Qiehe Sun, Tian Guan, Anjia Han, and Yonghong He. Dynamic graph representation with knowledge-aware attention for histopathology whole slide image analysis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11323–11332, 2024a.

Jiawen Li, Tian Guan, Qingxin Xia, Yizhi Wang, Xitong Ling, Jing Li, Qiang Huang, Zihan Wang, Zhiyuan Shen, Yifei Ma, et al. Unlocking adaptive digital pathology through dynamic feature learning. *arXiv preprint arXiv:2412.20430*, 2024b.

Jiawen Li, Jiali Hu, Qiehe Sun, Renao Yan, Minxi Ouyang, Tian Guan, Anjia Han, Chao He, and Yonghong He. Can we simplify slide-level fine-tuning of pathology foundation models? *arXiv preprint arXiv:2502.20823*, 2025.

Yixuan Lin, Weiping Lin, Chenxu Guo, Xinxin Yang, Hongxue Meng, and Liansheng Wang. Tumor microenvironment-guided fine-tuning of pathology foundation models for esophageal squamous cell carcinoma immunotherapy response prediction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 660–669. Springer, 2025.

Xitong Ling, Minxi Ouyang, Yizhi Wang, Xinrui Chen, Renao Yan, Hongbo Chu, Junru Cheng, Tian Guan, Sufang Tian, Xiaoping Liu, et al. Agent aggregator with mask denoise mechanism for histopathology whole slide image analysis. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 2795–2803, 2024.

Shengjie Liu, Chuang Zhu, Feng Xu, Xinyu Jia, Zhongyue Shi, and Mulan Jin. Bci: Breast cancer immunohistochemical image generation through pyramid pix2pix. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1815–1824, 2022.

Ming Y Lu, Tiffany Y Chen, Drew FK Williamson, Melissa Zhao, Maha Shady, Jana Lipkova, and Faisal Mahmood. Ai-based pathology predicts origins for cancers of unknown primary. *Nature*, 594(7861):106–110, 2021.

Jiabo Ma, Zhengrui Guo, Fengtao Zhou, Yihui Wang, Yingxue Xu, Jinbang Li, Fang Yan, Yu Cai, Zhengjie Zhu, Cheng Jin, et al. A generalizable pathology foundation model using a unified knowledge distillation pretraining framework. *Nature Biomedical Engineering*, pages 1–20, 2025.

Shino Magaki, Seyed A Hojat, Bowen Wei, Alexandra So, and William H Yong. An introduction to the performance of immunohistochemistry. *Biobanking: methods and protocols*, pages 289–298, 2018.

Dmitry Nechaev, Alexey Pchelnikov, and Ekaterina Ivanova. Histai: An open-source, large-scale whole slide image dataset for computational pathology. *arXiv preprint arXiv:2505.12120*, 2025.

Nobuyuki Otsu et al. A threshold selection method from gray-level histograms. *Automatica*, 11(285-296):23–27, 1975.

Pushpak Pati, Sofia Karkampouna, Francesco Bonollo, Eva Compérat, Martina Radić, Martin Spahn, Adriano Martinelli, Martin Wartenberg, Marianna Kruithof-de Julio, and Marianna Rapsomaniki. Accelerating histopathology workflows with generative ai-based virtually multiplexed tumour profiling. *Nature machine intelligence*, 6(9):1077–1093, 2024.

Zhuchen Shao, Hao Bian, Yang Chen, Yifeng Wang, Jian Zhang, Xiangyang Ji, et al. Transmil: Transformer based correlated multiple instance learning for whole slide image classification. *Advances in neural information processing systems*, 34:2136–2147, 2021.

Zaneta Swiderska-Chadaj, Hans Pinckaers, Mart Van Rijthoven, Maschenka Balkenhol, Margarita Melnikova, Oscar Geessink, Quirine Manson, Mark Sherman, Antonio Polonia, Jeremy Parry, et al. Learning to detect lymphocytes in immunohistochemistry with deep learning. *Medical image analysis*, 58:101547, 2019.

Eugene Vorontsov, Alican Bozkurt, Adam Casson, George Shaikovski, Michal Zelechowski, Kristen Severson, Eric Zimmermann, James Hall, Neil Tenenholtz, Nicolo Fusi, et al. A foundation model for clinical-grade computational pathology and rare cancers detection. *Nature medicine*, 30(10):2924–2935, 2024.

Xiyue Wang, Sen Yang, Jun Zhang, Minghui Wang, Jing Zhang, Wei Yang, Junzhou Huang, and Xiao Han. Transformer-based unsupervised contrastive learning for histopathological image classification. *Medical image analysis*, 81:102559, 2022.

Hanwen Xu, Naoto Usuyama, Jaspreet Bagga, Sheng Zhang, Rajesh Rao, Tristan Naumann, Cliff Wong, Zelalem Gero, Javier González, Yu Gu, et al. A whole-slide foundation model for digital pathology from real-world data. *Nature*, 630(8015):181–188, 2024.

Hongming Xu, Mingkang Wang, Duanbo Shi, Huamin Qin, Yunpeng Zhang, Zaiyi Liu, Anant Madabhushi, Peng Gao, Fengyu Cong, and Cheng Lu. When multiple instance learning meets foundation models: advancing histological whole slide image analysis. *Medical Image Analysis*, 101:103456, 2025.

Fang Yan, Jianfeng Wu, Jiawen Li, Wei Wang, Yirong Chen, Linda Wei, Jiaxuan Lu, Wen Chen, Zizhao Gao, Jianan Li, et al. Pathorchestra: A comprehensive foundation model for computational pathology with over 100 diverse clinical-grade tasks. *npj Digital Medicine*, 8(1):695, 2025.

Renao Yan, Qiming He, Yiqing Liu, Peng Ye, Lianghui Zhu, Shanshan Shi, Jizhou Gou, Yonghong He, Tian Guan, and Guangde Zhou. Unpaired virtual histological staining using prior-guided generative adversarial networks. *Computerized Medical Imaging and Graphics*, 105:102185, 2023.