SRA: A Novel Method to Improve Feature Embedding in Self-supervised Learning for Histopathological Images

Hamid Manoochehri[*1,2] HAMID.MANOOCHEHRI@UTAH.EDU
[1] *Scientific Computing and Imaging Institute, University of Utah, Salt Lake City, UT, USA*
[2] *Dept. of Electrical and Computing Engineering, University of Utah, Salt Lake City, UT, USA*

Bodong Zhang[*†1,2] BODONG.ZHANG@UTAH.EDU
Beatrice S. Knudsen[3] BEATRICE.KNUDSEN@PATHOLOGY.UTAH.EDU
[3] *Department of Pathology, University of Utah, Salt Lake City, UT, USA*

Tolga Tasdizen[1,2] TOLGA.TASDIZEN@UTAH.EDU

Editors: Accepted for publication at MIDL 2025

Abstract

Self-supervised learning has become a cornerstone in various areas, particularly histopathological image analysis. Image augmentation plays a crucial role in self-supervised learning, as it generates variations in image samples. However, traditional image augmentation techniques often overlook the unique characteristics of histopathological images. In this paper, we propose a new histopathology-specific image augmentation method called stain reconstruction augmentation (SRA). We integrate our SRA into various self-supervised learning models. We demonstrate that our SRA always improves the standard models across various downstream tasks and achieves superior performance to a state-of-the-art foundation model pre-trained on significantly larger histopathology datasets.

Keywords: image augmentation, histopathological image, self-supervised learning, contrastive learning

1. Introduction

In this paper, we propose a novel image augmentation method on H&E stained histopathological images called stain reconstruction augmentation (SRA). Through experiments, we demonstrate that our SRA further improves the original self-supervised learning (SSL)(Chen et al., 2021; Caron et al., 2021) models based on performances in various downstream tasks.

2. Stain Reconstruction Augmentation (SRA)

As shown in Figure 1, for an H&E image, we first perform stain separation using the algorithm from (Macenko et al., 2009) to obtain single-stain images. For each whole slide image (WSI), the RGB pixel values are mapped into Optical Density (OD) space (OD_R, OD_G, OD_B) according to the Beer-Lambert law (Beer, 1852; Lambert, 1760), where higher OD values indicate stronger stains:

$$OD_C = \log_{10}(I_{0,C}/I_C) \tag{1}$$

[*] Contributed equally
[†] Corresponding author

© 2025 CC-BY 4.0, H. Manoochehri, B. Zhang, B.S. Knudsen & T. Tasdizen.

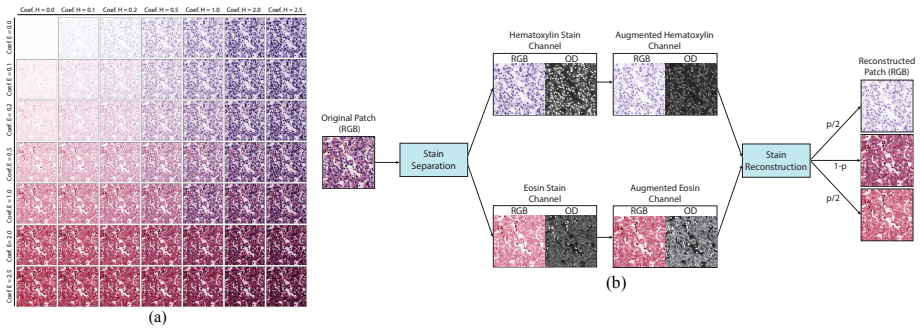

Figure 1: (a) Examples of augmentations by SRA with different target strengths of H channel and E channel (b) Pipeline of stain reconstruction augmentation (SRA).

The channel C is red, green, or blue channel. The $I_{0,C}$ denotes background intensity, which is usually 255. I_C is the intensity in channel C in current pixel. All pixels are mapped to the same OD space. Based on the distribution of these pixels in OD space, three unit vectors, V_H, V_E, and $V_{Residual}$, are derived, which allow for the decomposition of OD values:

$$(OD_R, OD_G, OD_B) = \alpha V_H + \beta V_E + \gamma V_{Residual} \tag{2}$$

In each slide, we calculate the values of α and β for each tissue pixel, where α represents the proportion of Hematoxylin stain and β represents the proportion of Eosin stain for each pixel. For each WSI, we define H_{max} as the maximum intensity of all α values in the tissue pixels of the slide and similarly define E_{max} for the maximum intensity for β.

In the first step of implementing our SRA, we predefine global target ranges for each stain after augmentation. For each training image, following the stain separation process, we independently and randomly select coefficients $coef_H$ and $coef_E$ from within the target ranges. We then multiply α by $coef_H/H_{max}$ and β by $coef_E/E_{max}$ to randomly adjust the stain strength. Furthermore, inspired by the previous work of contrastive learning between pure Hematoxylin images and pure Eosin images, (Zhang et al., 2022a, 2023) we introduce a hyperparameter p, which defines the probability of randomly setting either $coef_H$ or $coef_E$ to zero during stain reconstruction augmentation, thereby creating additional variations. Finally, after all processes are complete, we reconstruct the images back into RGB space from OD space, based on the new proportions of Hematoxylin and Eosin stains. Our SRA is an unsupervised method that does not use any labels to apply the algorithm.

3. Experiments

We conducted our experiments on three datasets: The Cancer Genome Atlas Kidney Renal Clear Cell Carcinoma (TCGA KIRC) dataset (National Cancer Institute, 2023), the Utah KIRC dataset (Zhang et al., 2023), and the Utah Renal Vein Thrombus (RVT) dataset (Zhang et al., 2025). The TCGA KIRC dataset and Utah KIRC dataset provide 300 slides (cropped to 1,646,665 patches) and 32 slides (cropped to 208,291 patches), separately, for self-supervised learning, as well as additional patches with patch-level labels for downstream patch classification tasks. The Utah RVT dataset is only used for multiple instance learning

in downstream tasks. We ran each setting 3 times to calculate the mean and standard deviation. (More details about datasets can be found in Appendix C.1)

We first evaluated SRA by performing self-supervised pre-training and downstream classification on the same datasets (TCGA KIRC or Utah KIRC). Results in Table 1 show that our SRA greatly enhances original MoCo v3 (Chen et al., 2021) during pre-training. We call the new model SRA-MoCo v3. For reference, we also compared the classification results with a state-of-the-art foundation encoder model - Prov-GigaPath (Xu et al., 2024).

Model	**TCGA KIRC (20X)**		**Utah KIRC (10X)**	
	Pre-trained Dataset	Balanced Accuracy (%)	Pre-trained Dataset	Balanced Accuracy (%)
ResNet50	ImageNet	69.97 ± 5.59	ImageNet	87.76 ± 0.10
Prov-GigaPath	>170,000 Slides >1,000,000,000 patches	77.88 ± 1.22	>170,000 Slides >1,000,000,000 patches	95.28 ± 1.02
MoCo v3	TCGA KIRC	79.37 ± 1.18	Utah KIRC	93.77 ± 0.86
MoCo v3 + TSA	300 Slides	81.50 ± 0.23	49 Slides	94.00 ± 0.26
SRA-MoCo v3	1,646,665 patches	**83.62** ± 0.28	208,291 patches	**95.85** ± 0.34

Table 1: Comparison of pre-trained models' performance on TCGA KIRC (20X magnification) and Utah KIRC (10X magnification). TSA means traditional stain augmentation methods (Tellez et al., 2018a,b). (Differences are shown in Appendix B.)

To evaluate in a transfer learning setting, we pre-trained encoders on the TCGA KIRC dataset and subsequently evaluated the encoders on the Utah KIRC and the Utah RVT. Based on Table 2, we observed 2.8% improvement on Utah KIRC and 3.7% improvement on Utah RVT with SRA-MoCo v3 compared to MoCo v3 on balanced accuracy.

Model	Utah KIRC Balanced Acc.	Utah RVT Balanced Acc.	Utah RVT F1-score	Utah RVT AUC
Prov-GigaPath	95.28 ± 1.02	70.96 ± 1.40	64.77 ± 1.66	0.7576 ± 0.0101
DINO (TCGA KIRC)	95.49 ± 0.24	67.51 ± 1.96	62.54 ± 1.29	0.7206 ± 0.0154
SRA-DINO (TCGA KIRC)	96.37 ± 0.21	69.70 ± 1.01	63.98 ± 0.33	0.7323 ± 0.0152
MoCo v3 (TCGA KIRC)	95.32 ± 0.30	71.80 ± 2.12	66.51 ± 2.80	0.7677 ± 0.0307
MoCo v3 + TSA (TCGA KIRC)	94.17 ± 0.82	73.40 ± 3.36	67.74 ± 3.26	0.7609 ± 0.0776
SRA-MoCo v3 (TCGA KIRC)	**98.12** ±0.15	**75.50** ±7.08	**71.11** ± 7.70	**0.8013** ± 0.0337

Table 2: Performance of pre-trained models on Utah KIRC and Utah RVT dataset.

Lastly, we conducted ablation studies to analyze the impact of hyperparameter settings in SRA. Based on the experimental results (in Appendix C.3), the best performance was achieved with target ranges of [0.1, 2.5]. Both narrower and wider ranges negatively affected the performance of the SRA model. Even though effective, if the possibility p that only a single channel is adopted is too large, it produces a harmful effect on the training. In experiments, we abandoned the residual part during stain reconstruction as keeping residual part does not provide benefits. We also found that both SRA and standard color augmentations are beneficial, so we kept both of them in SRA-MoCo v3. We made our SRA-MoCo v3 code publicly available at `github.com/hamidmanoochehri/Paper_SRA`

References

Saghir Alfasly, Abubakr Shafique, Peyman Nejat, Jibran Khan, Areej Alsaafin, Ghazal Alabtah, and Hamid R Tizhoosh. Rotation-agnostic image representation learning for digital pathology. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11683–11693, 2024.

Beer. Bestimmung der absorption des rothen lichts in farbigen flüssigkeiten. *Ann. Phys.*, 162(5):78–88, January 1852. ISSN 0003-3804,1521-3889. doi: 10.1002/andp.18521620505.

Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. Unsupervised learning of visual features by contrasting cluster assignments. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS '20, Red Hook, NY, USA, 2020. Curran Associates Inc. ISBN 9781713829546.

Mathilde Caron, Hugo Touvron, Ishan Misra, Herve Jegou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, October 2021. doi: 10.1109/iccv48922.2021.00951.

Wenhao Chai and Gaoang Wang. Deep vision multimodal learning: Methodology, benchmark, and trend. *Appl. Sci. (Basel)*, 12(13):6588, June 2022. ISSN 2076-3417,2076-3417. doi: 10.3390/app12136588.

Jia-Ren Chang, Min-Sheng Wu, Wei-Hsiang Yu, Chi-Chung Chen, Cheng-Kung Yang, Yen-Yu Lin, and Chao-Yuan Yeh. Stain mix-up: Unsupervised domain generalization for histopathology images. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, Lecture notes in computer science, pages 117–126. Springer International Publishing, Cham, 2021. doi: 10.1007/978-3-030-87199-4_11.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In Hal Daumé Iii and Aarti Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 1597–1607. PMLR, 2020a.

Ting Chen, Simon Kornblith, Kevin Swersky, Mohammad Norouzi, and Geoffrey E Hinton. Big self-supervised models are strong semi-supervised learners. *Advances in neural information processing systems*, 33:22243–22255, 2020b.

Xinlei Chen, Saining Xie, and Kaiming He. An empirical study of training self-supervised vision transformers. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, October 2021. doi: 10.1109/iccv48922.2021.00950.

Timothée Darcet, Maxime Oquab, Julien Mairal, and Piotr Bojanowski. Vision transformers need registers. In *The Twelfth International Conference on Learning Representations*, 2024.

Babak Ehteshami Bejnordi, Mitko Veta, Paul Johannes van Diest, Bram van Ginneken, Nico Karssemeijer, Geert Litjens, Jeroen A W M van der Laak, the CAMELYON16

Consortium, Meyke Hermsen, Quirine F Manson, Maschenka Balkenhol, Oscar Geessink, Nikolaos Stathonikos, Marcory Crf van Dijk, Peter Bult, Francisco Beca, Andrew H Beck, Dayong Wang, Aditya Khosla, Rishab Gargeya, Humayun Irshad, Aoxiao Zhong, Qi Dou, Quanzheng Li, Hao Chen, Huang-Jing Lin, Pheng-Ann Heng, Christian Haß, Elia Bruni, Quincy Wong, Ugur Halici, Mustafa Ümit Öner, Rengul Cetin-Atalay, Matt Berseth, Vitali Khvatkov, Alexei Vylegzhanin, Oren Kraus, Muhammad Shaban, Nasir Rajpoot, Ruqayya Awan, Korsuk Sirinukunwattana, Talha Qaiser, Yee-Wah Tsang, David Tellez, Jonas Annuscheit, Peter Hufnagl, Mira Valkonen, Kimmo Kartasalo, Leena Latonen, Pekka Ruusuvuori, Kaisa Liimatainen, Shadi Albarqouni, Bharti Mungal, Ami George, Stefanie Demirci, Nassir Navab, Seiryo Watanabe, Shigeto Seno, Yoichi Takenaka, Hideo Matsuda, Hady Ahmady Phoulady, Vassili Kovalev, Alexander Kalinovsky, Vitali Liauchuk, Gloria Bueno, M Milagro Fernandez-Carrobles, Ismael Serrano, Oscar Deniz, Daniel Racoceanu, and Rui Venâncio. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA*, 318(22): 2199–2210, December 2017. ISSN 0098-7484,1538-3598. doi: 10.1001/jama.2017.14585.

Babak Ehteshami Bejnordi, Oscar Geessink, Meyke Hermsen, Geert Litjens, Jeroen van der Laak, Quirine Manson, Mitko Veta, and Nikolaos Stathonikos. CAMELYON16 - Grand Challenge. https://camelyon16.grand-challenge.org/Data/, 2023. Accessed: 2023-12-15.

Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre H. Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Daniel Guo, Mohammad Gheshlaghi Azar, Bilal Piot, Koray Kavukcuoglu, Rémi Munos, and Michal Valko. Bootstrap your own latent a new approach to self-supervised learning. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS '20, Red Hook, NY, USA, 2020. Curran Associates Inc. ISBN 9781713829546.

Sai Chowdary Gullapally, Yibo Zhang, Nitin Kumar Mittal, Deeksha Kartik, Sandhya Srinivasan, Kevin Rose, Daniel Shenker, Dinkar Juyal, Harshith Padigela, Raymond Biju, Victor Minden, Chirag Maheshwari, Marc Thibault, Zvi Goldstein, Luke Novak, Nidhi Chandra, Justin Lee, Aaditya Prakash, Chintan Shah, John Abel, Darren Fahy, Amaro Taylor-Weiner, and Anand Sampat. Synthetic DOmain-targeted augmentation (S-DOTA) improves model generalization in digital pathology. *arXiv [cs.CV]*, May 2023.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *arXiv [cs.CV]*, pages 770–778, December 2015.

J H Lambert. *Photometria sive de mensura et gradibus luminis, colorum et umbrae.* sumptibus vidvae E. Klett, typis C.P. Detleffsen, 1760.

Marc Macenko, Marc Niethammer, J S Marron, David Borland, John T Woosley, Xiaojun Guan, Charles Schmitt, and Nancy E Thomas. A method for normalizing histology slides for quantitative analysis. In *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. IEEE, June 2009. ISBN 9781424439317. doi: 10.1109/isbi.2009.5193250.

National Cancer Institute. National cancer institute GDC data portal. https://portal.gdc.cancer.gov/, 2023. Accessed: 2023-11-2.

Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy V. Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel HAZIZA, Francisco Massa, Alaaeldin El-Nouby, Mido Assran, Nicolas Ballas, Wojciech Galuba, Russell Howes, Po-Yao Huang, Shang-Wen Li, Ishan Misra, Michael Rabbat, Vasu Sharma, Gabriel Synnaeve, Hu Xu, Herve Jegou, Julien Mairal, Patrick Labatut, Armand Joulin, and Piotr Bojanowski. DINOv2: Learning robust visual features without supervision. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856.

Zekun Qi, Runpei Dong, Guofan Fan, Zheng Ge, Xiangyu Zhang, Kaisheng Ma, and Li Yi. Contrast with reconstruct: contrastive 3d representation learning guided by generative pretraining. In *Proceedings of the 40th International Conference on Machine Learning*, ICML'23. JMLR.org, 2023.

A C Ruifrok and D A Johnston. Quantification of histochemical staining by color deconvolution. *Anal. Quant. Cytol. Histol.*, 23(4):291–299, August 2001. ISSN 0884-6812.

Yiqing Shen, Yulin Luo, Dinggang Shen, and Jing Ke. RandStainNA: Learning stain-agnostic features from histology slides by bridging stain augmentation and normalization. In *Lecture Notes in Computer Science*, Lecture notes in computer science, pages 212–221. Springer Nature Switzerland, Cham, 2022. ISBN 9783031164330,9783031164347. doi: 10.1007/978-3-031-16434-7_21.

David Tellez, Maschenka Balkenhol, Nico Karssemeijer, Geert Litjens, Jeroen van der Laak, and Francesco Ciompi. H and E stain augmentation improves generalization of convolutional networks for histopathological mitosis detection. In John E. Tomaszewski and Metin N. Gurcan, editors, *Medical Imaging 2018: Digital Pathology*, volume 10581, page 105810Z. International Society for Optics and Photonics, SPIE, 2018a. doi: 10.1117/12.2293048. URL https://doi.org/10.1117/12.2293048.

David Tellez, Maschenka Balkenhol, Irene Otte-Holler, Rob van de Loo, Rob Vogels, Peter Bult, Carla Wauters, Willem Vreuls, Suzanne Mol, Nico Karssemeijer, Geert Litjens, Jeroen van der Laak, and Francesco Ciompi. Whole-slide mitosis detection in H&E breast histology using PHH3 as a reference to train distilled stain-invariant convolutional networks. *IEEE Trans. Med. Imaging*, 37(9):2126–2136, September 2018b. ISSN 0278-0062,1558-254X. doi: 10.1109/TMI.2018.2820199.

Luyang Wang, Feng Liang, Yangguang Li, Honggang Zhang, Wanli Ouyang, and Jing Shao. Repre: Improving self-supervised vision transformer with reconstructive pre-training. In *International Joint Conference on Artificial Intelligence*, 2022.

Hanwen Xu, Naoto Usuyama, Jaspreet Bagga, Sheng Zhang, Rajesh Rao, Tristan Naumann, Cliff Wong, Zelalem Gero, Javier González, Yu Gu, Yanbo Xu, Mu Wei, Wenhui Wang, Shuming Ma, Furu Wei, Jianwei Yang, Chunyuan Li, Jianfeng Gao, Jaylen Rosemon, Tucker Bower, Soohee Lee, Roshanthi Weerasinghe, Bill J Wright, Ari Robicsek, Brian Piening, Carlo Bifulco, Sheng Wang, and Hoifung Poon. A whole-slide foundation model

for digital pathology from real-world data. *Nature*, 630(8015):181–188, June 2024. ISSN 0028-0836,1476-4687. doi: 10.1038/s41586-024-07441-w.

Pengshuai Yang, Xiaoxu Yin, Haiming Lu, Zhongliang Hu, Xuegong Zhang, Rui Jiang, and Hairong Lv. CS-CO: A hybrid self-supervised visual representation learning method for H&E-stained histopathological images. *Med. Image Anal.*, 81(102539):102539, October 2022. ISSN 1361-8415,1361-8423. doi: 10.1016/j.media.2022.102539.

Jure Zbontar, Li Jing, Ishan Misra, Yann LeCun, and Stéphane Deny. Barlow twins: Self-supervised learning via redundancy reduction. *ArXiv*, abs/2103.03230, 2021.

Bodong Zhang, Beatrice Knudsen, Deepika Sirohi, Alessandro Ferrero, and Tolga Tasdizen. Stain based contrastive co-training for histopathological image analysis. In *Medical Image Learning with Limited and Noisy Data*, pages 106–116. Springer Nature Switzerland, 2022a. doi: 10.1007/978-3-031-16760-7_11.

Bodong Zhang, Hamid Manoochehri, Man Minh Ho, Fahimeh Fooladgar, Yosep Chong, Beatrice S Knudsen, Deepika Sirohi, and Tolga Tasdizen. CLASS-M: Adaptive stain separation-based contrastive learning with pseudo-labeling for histopathological image classification. *arXiv [cs.CV]*, December 2023.

Bodong Zhang, Hamid Manoochehri, Beatrice S Knudsen, and Tolga Tasdizen. WeakSupCon: Weakly supervised contrastive learning for encoder pre-training. *arXiv [cs.CV]*, March 2025.

Hongrun Zhang, Yanda Meng, Yitian Zhao, Yihong Qiao, Xiaoyun Yang, Sarah E. Coupland, and Yalin Zheng. Dtfd-mil: Double-tier feature distillation multiple instance learning for histopathology whole slide image classification. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 18780–18790, 2022b. doi: 10.1109/CVPR52688.2022.01824.

Jinghao Zhou, Chen Wei, Huiyu Wang, Wei Shen, Cihang Xie, Alan Yuille, and Tao Kong. ibot: Image bert pre-training with online tokenizer. *International Conference on Learning Representations (ICLR)*, 2022.

Appendix A. Introduction of related models

A.1. Self-supervised learning and contrastive loss

Recent advancements in self-supervised learning (Zbontar et al., 2021; Grill et al., 2020) have introduced novel frameworks for learning robust and accurate features across various datasets. Barlow Twins (Zbontar et al., 2021) encourages two augmented views of the same input to produce similar but decorrelated representations by minimizing the cross-correlation between them. DINO (Caron et al., 2021; Oquab et al., 2024; Darcet et al., 2024) employs a student-teacher framework within a vision transformer architecture, without requiring labeled data. In DINO, the student encoder attempts to mimic the teacher encoder, which is updated based on an exponential moving average (EMA). Unlike the

student encoder, which processes both global and local views of the images, the teacher encoder only receives global views. Building on the DINO framework, PathDino (Alfasly et al., 2024) combines lightweight transformers with a novel 360° rotation augmentation (HistoRotate), achieving robust performance across 12 diverse pathology datasets.

Contrastive learning is one of the most widely used and fundamental approaches in self-supervised learning pipelines. For instance, the SimCLR framework (Chen et al., 2020b,a) utilizes NT-Xent loss on strongly augmented views of images, aiming to minimize the distance between different views of the same image while maximizing the distance between views of different images. In contrast, SwAV (Caron et al., 2020) employs a cluster-based contrastive learning approach rather than a pairwise one, using a swapped prediction mechanism to encourage the features of the same cluster to be as invariant as possible. Other popular contrastive learning methods include iBOT (Zhou et al., 2022), RePre (Wang et al., 2022) and RECON (Qi et al., 2023).

A.2. Pathology-specific augmentation

In the context of pathology-specific augmentations, various methods have been proposed to address domain-specific challenges and variations in histopathological images. Shen et al. (Shen et al., 2022) introduced RandStainNA, which generates random template slides for color normalization and augmentation in HSV, LAB, and HED color spaces to tackle variations in staining and colors across different slides and datasets. Additionally, Gullapally et al. (Gullapally et al., 2023) addressed inter-laboratory and scanner variability through Scanner Transform (ST) and Stain Vector Augmentation (SVA), enhancing out-of-distribution performance on tasks such as tissue segmentation.

A fundamental operation for many pathology-specific augmentations is stain separation, which isolates single-channel images in Optical Density (OD) space based on the Beer-Lambert law (Lambert, 1760; Beer, 1852). In (Yang et al., 2022), perturbations are applied to the stain separation matrix to deal with the errors in separation matrix calculation. In (Tellez et al., 2018a,b), each channel is randomly scaled and biased within a narrow range before converting back to RGB space. However, the maximum possible intensity after augmentation is still influenced by the original image's maximum intensity. (Chang et al., 2021) also utilizes this method, along with random stain matrix interpolation, to handle domain variations across datasets by incorporating information from both source and target data.

Appendix B. Integration of SRA into MoCo v3 (An overview of SRA-MoCo v3)

Figure 2 shows the overall workflow of SRA-MoCo v3. First, all pixels within the tissue regions of an H&E whole slide image (WSI) are collected to analyze the max intensity (strength) of each stain, where the intensity is measured on Optical Density (OD) space after stain separation (Ruifrok and Johnston, 2001) process. To perform stain reconstruction augmentation, we predefine an absolute range for the target strength of each stain and map the real strength of each stain to a random value in this target range. Unlike the approach in (Tellez et al., 2018a,b; Chang et al., 2021), which only slightly adjusts the strength of each stain within a relative range by multiplying a random factor between 95% and 105%,

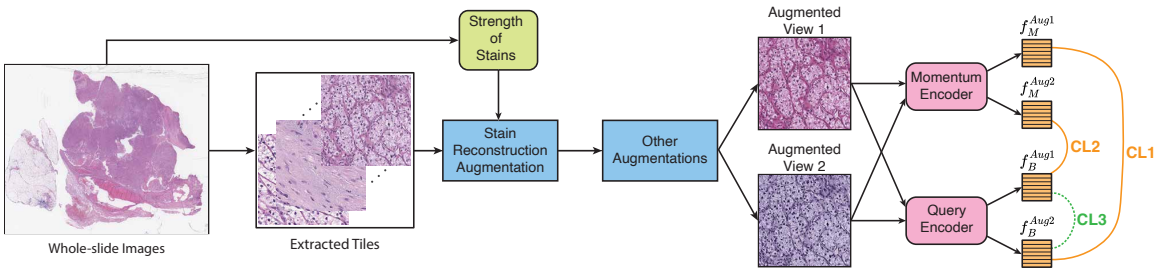

Figure 2: The pipeline of our SRA-MoCo v3. We integrate our Stain Reconstruction Augmentation (SRA) as well as additional contrastive loss term (CL3) into MoCo v3. The f in the figure shows the output features from the encoders.

where the augmented images are statistically affected by the intensities of original stain channel images, our SRA directly defines a much broader absolute range for target strength of each stain channel. For instance, if the target range is set to [0.1, 2.5], and the original strength of a particular stain in a WSI is 2.5, which indicates a deeply stained image. Augmenting this stain channel only makes the new maximum intensity fall inside [0.1, 2.5], without surpassing the original strength since original image is already a deepest stained image in our defined range. While the traditional augmentation makes the new maximum intensity fall between 2.375 and 2.625. Our method allows for more extensive and stronger augmentations while ensuring the strength remains within an appropriate range. Moreover, inspired by multi-modal contrastive learning (Chai and Wang, 2022; Zhang et al., 2022a), we create probability p for excluding one stain channel, allowing only the other stain channel to remain after augmentation. Figure 3 shows the differences between traditional stain augmentation and our SRA.

After stain reconstruction augmentation, additional general image augmentation methods are applied to introduce further variations. MoCo v3 is used as the backbone for histopathology image representation learning. In MoCo v3, the contrastive loss is computed between queries from the query encoder and keys from the momentum encoder using different augmentations. However, there are no loss term that specifically focus on contrastive learning between different augmentations from the same query encoder. Given the substantial variations introduced by stain reconstruction augmentation, we further explored the addition of this contrastive loss term that focuses solely on augmentations. This additional loss was only designed for query encoder on MoCo framework, as there is no gradient's backpropagation on momentum encoder.

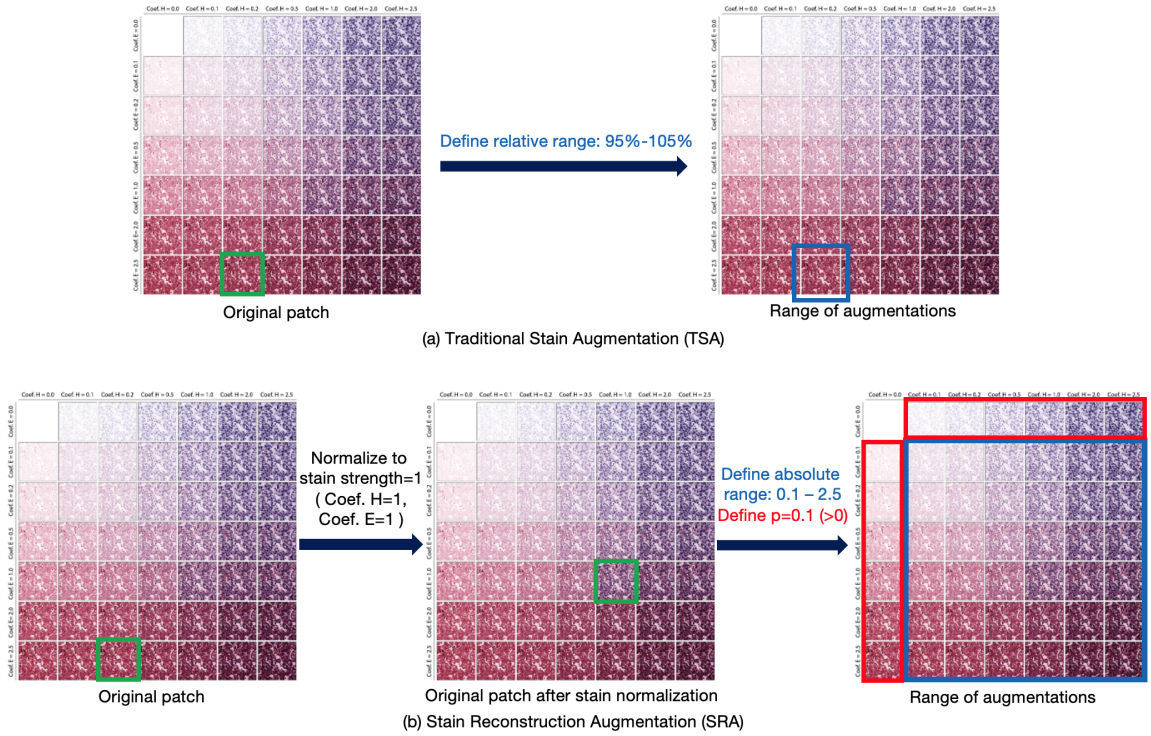

Figure 3: Comparisons between traditional stain augmentation (TSA) and our stain reconstruction augmentation (SRA). The blue rectangles show the range of augmentations. If the probability p for excluding one stain channel is not zero, then SRA further introduces new augmentation results shown in red rectangles.

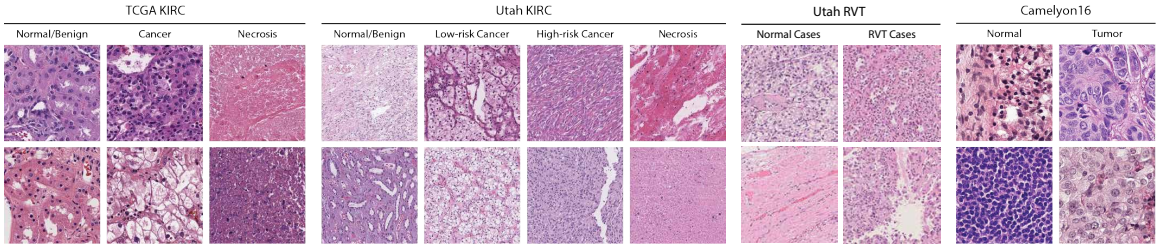

Figure 4: Patch examples from different classes and different datasets.

Appendix C. Additional information in experiments

C.1. Datasets details

We conducted all main experiments [1] on three datasets: The Cancer Genome Atlas Kidney Renal Clear Cell Carcinoma (TCGA KIRC) dataset (National Cancer Institute, 2023),

1. The experiments were funded by NIH/NCI R21CA277381, DoD HT94252410186, and Department of Veterans Affairs I01CS002622. Part of the results presented in this paper are based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga

the Utah KIRC dataset (Zhang et al., 2023), the Utah Renal Vein Thrombus (RVT) dataset(Zhang et al., 2025), all of which consist of H&E stained kidney WSIs. In ablation studies, we also include the Camelyon 16 breast cancer dataset (Ehteshami Bejnordi et al., 2017, 2023). Patch examples are shown in Figure 4.

The TCGA KIRC dataset includes 420 WSIs in total, with 300 slides used for self-supervised pre-training. The 300 slides provide 1,646,665 tissue patches of size 400x400 if cropped at 20X resolution. Patch-level labels are also provided in the remaining slides for downstream patch-level 3-class classification tasks. More details are provided at Table 3.

In the Utah KIRC dataset, there are 49 slides from different patients, with 32 slides for self-supervised pre-training. In total, the 32 slides provide 208,291 tissue patches of size 400x400 at 10X resolution. During downstream 4-class patch classification tasks, the 32 slides also provide a small portion of patches that have patch-level labels. The remaining 17 slides provide patches with labels for validation and test. More details can be found at Table 3.

Task	**TCGA KIRC**					**Utah KIRC**				
SSL	Total Unlabeled Patches 1,646,665					Total Unlabeled Patches 208,291				
Downstream Classification (Evaluation)	Set	Normal/ Benign	Cancer	Necrosis	Total	Normal/ Benign	Low-Risk Cancer	High-Risk Cancer	Necrosis	Total
	Train	84,578	180,471	7,932	272,981	28,497	2,044	2,522	4,115	37,178
	Val.	19,638	79,382	1,301	100,321	5,472	416	334	2,495	8,117
	Test	15,323	62,565	6,168	84,056	7,263	598	389	924	9,174

Table 3: Summary of the number of patches for each category in each set on the TCGA and Utah KIRC datasets. The training sets include labeled and unlabeled patches.

The Utah RVT dataset is for multiple instance learning (MIL) tasks. Instead of slide-level labels in most of MIL tasks, only case-level (patient-level) labels are provided, while a case could contain multiple slides. In summary, there are 74 negative cases and 31 positive cases of Renal Vein Thrombus in the training set, with 12 negative cases and 8 positive cases in the validation set, as well as 18 negative cases and 11 positive cases in the test set. The total number of slides in training set is 862. The total number of patches in the training set is 2,214,311. In experiments, the patch features are generated by encoders pre-trained on TCGA KIRC dataset to evaluate different self-supervised learning methods in a transfer learning setting.

Camelyon 16 is another public dataset consisting of breast cancer slides for multi-instance learning. It provides only slide-level labels for binary classification between Normal and Tumor. The training set contains 159 Normal slides and 111 Tumor slides, while the test set includes 80 Normal slides and 49 Tumor slides. We randomly selected 10% of the training slides to create a validation set. We followed the standard procedure as other papers (Zhang et al., 2022b) to crop patches in the 20X resolution. Similar to Utah RVT dataset, the patch features are also generated by encoders pre-trained on TCGA KIRC dataset in experiments.

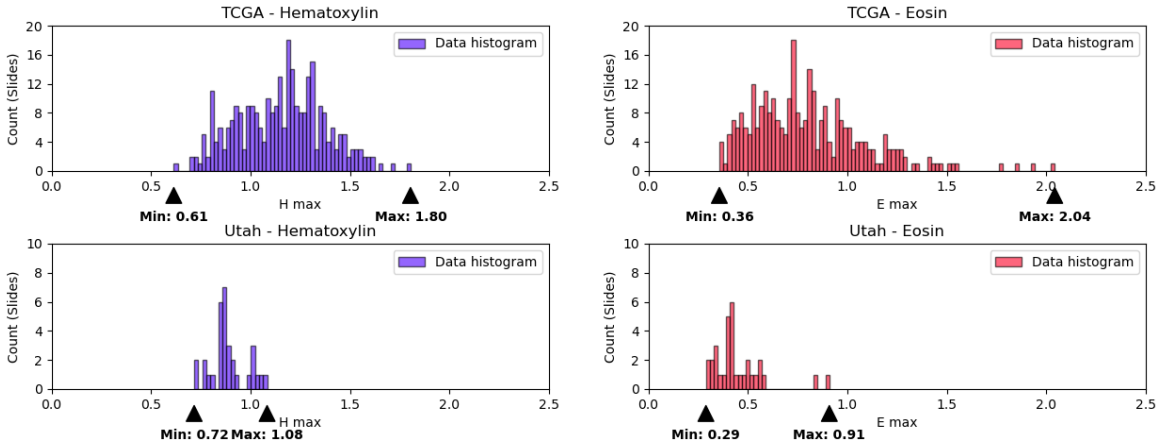

Figure 5: Distributions of strengths of Hematoxylin stain and Eosin stain in Optical Density (OD) space on TCGA training set and Utah training set.

C.2. Experiment settings

In experiments, we began by pre-training encoders using self-supervised feature representation learning, followed by testing the pre-trained encoders on downstream tasks. Before pre-training, we performed a one-time operation to calculate stain separation matrix for each slide. The computational overhead of SRA during pre-training is negligible.

For stain reconstruction augmentation, we first predefined a range for the target strengths of the Hematoxylin and Eosin stains. We calculated the distributions of H_{max} and E_{max} across all training slides in both TCGA KIRC and Utah KIRC datasets. As shown in Figure 5, the stain strengths H_{max} and E_{max} varies across different slides in both datasets. In our experiments, we mainly tested two sets of target ranges. The first set marginally covers the distribution of stain intensities in TCGA KIRC, with a target range of $[0.5, 2.0]$ for new H_{max} and $[0.2, 2.0]$ for new E_{max}. The second set has wider target ranges to introduce stronger augmentations, with both new H_{max} and new E_{max} set to $[0.1, 2.5]$ after discussion with pathologists. We selected this range because further widening would cause saturation when reconstructing images from Optical Density (OD) space back to RGB space.

After applying stain reconstruction augmentation, we followed the standard MoCo v3 training procedure with batch size as 512, including additional general augmentations, with ResNet50 (He et al., 2015) as backbone. We also included an option in our code to add the augmentation contrastive loss, CL_{aug}, as mentioned before. For the downstream tasks on the Utah RVT dataset and Camelyon 16 dataset, we employed DTFD-MIL (Zhang et al., 2022b). Stain reconstruction augmentation was only applied during pre-training, not in any downstream tasks. All experiments were conducted using Python 3.11.4, PyTorch 2.0.1, torchvision 0.15.2, and CUDA 11.8 on NVIDIA RTX A6000 GPUs.

C.3. Ablation studies

We also conducted ablation studies to analyze the impact of components in SRA-MoCo v3.

Range $coef_H$	Range $coef_E$	p(only H or E)	Extra Loss	Balanced Acc. Utah KIRC	F1-score Camelyon16	Accuracy Camelyon16	Balanced Acc. Camelyon16
N/A	N/A	0	–	95.32 ± 0.30	80.75 ± 1.35	85.79 ± 1.79	84.33 ± 0.89
N/A	N/A	0	CL_{aug}	95.34 ± 0.41	84.57 ± 2.53	88.11 ± 1.19	87.78 ± 2.44
[0.5, 2.0]	[0.2, 2.0]	0	–	96.51 ± 0.37	82.44 ± 0.47	87.34 ± 0.45	85.58 ± 0.42
[0.1, 2.5]	[0.1, 2.5]	0	–	96.95 ± 0.76	84.94 ± 1.54	89.67 ± 0.89	87.19 ± 1.25
[0.5, 2.0]	[0.2, 2.0]	0	CL_{aug}	96.86 ± 0.17	83.41 ± 0.60	88.37 ± 0.78	86.15 ± 0.33
[0.1, 2.5]	[0.1, 2.5]	0	CL_{aug}	98.09 ± 0.12	85.96 ± 2.13	89.15 ± 2.05	88.75 ± 1.40
[0.5, 2.0]	[0.2, 2.0]	10%	CL_{aug}	97.41 ± 0.08	90.79 ± 1.50	93.28 ± 1.18	92.08 ± 1.02
[0.1, 2.5]	[0.1, 2.5]	10%	CL_{aug}	**98.12** ± 0.15	**92.07** ± 0.84	**94.31** ± 0.44	**92.91** ± 0.95

Table 4: Ablation study results showing the impact of each component of SRA-MoCo v3 on Utah KIRC dataset and Camelyon 16 dataset.

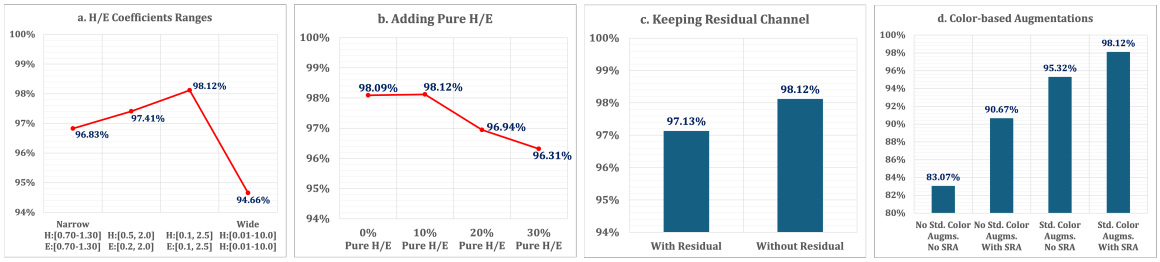

Figure 6: Systematic analysis of hyperparameters of SRA on Utah KIRC dataset.

The ablation studies for transfer learning from the TCGA KIRC dataset to the Utah KIRC dataset are presented in Table 4. From the results, we found that simply adding the extra augmentation contrastive loss to MoCo v3 does not yield any improvement. However, this loss becomes effective when combined with stain reconstruction augmentation. Using a wider target range for stain strength and incorporating the possibility of generating pure Hematoxylin or pure Eosin images in stain reconstruction augmentation also proved beneficial. In summary, SRA-MoCo v3 addresses domain shift by generating highly augmented images that are also clinically meaningful. The Figure 6 more systematically analyzed the impact of hyperparameters' selections to the performance. Figure 6(a) shows that target ranges that are too narrow or too wide both have negative impacts to the SRA model. According to Figure 6(b), if the possibility p that only a single channel is adopted is too large, it produces a harmful effect on the training. In experiments, we abandoned the residual part during stain reconstruction. As shown in Figure 6(c), keeping residual part does not provide benefits. In Figure 6(d), we tried to remove standard color-related augmentations like color jitterring. We found that both SRA and standard color augmentations are beneficial, so we kept both of them in SRA-MoCo v3.

Lastly, we evaluated the contribution of each component in SRA-MoCo v3 through downstream tasks on the Camelyon 16 dataset. We observed that both the augmentation contrastive loss and stain reconstruction augmentation independently improve performance. When combining the augmentation contrastive loss with stain reconstruction augmentation, we achieved a 4.42% improvement in balanced accuracy. Also, we observed a 4.16% increase

in balanced accuracy on Camelyon 16 by simply adjusting p from 0 to 0.1, which implies that stronger augmentations are more beneficial on cases with stronger domain shift.