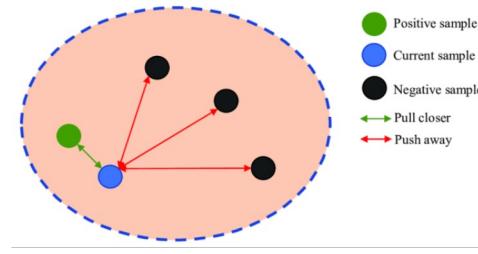


# SRA: A Novel Method to Improve Feature Embedding in Self-supervised Learning for Histopathological Images

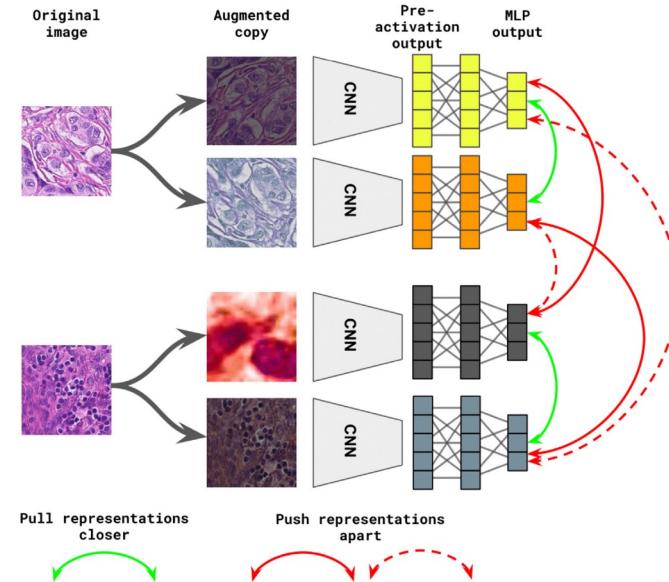
Presenter: Bodong Zhang

# Background - Self-supervised contrastive learning

- Contrastive learning: a machine learning technique focused on learning representations by encouraging similar data points to be closer together in the embedding space while pushing dissimilar points farther apart.
- Self-supervised contrastive learning enables the model to learn meaningful representations without requiring explicit labels.



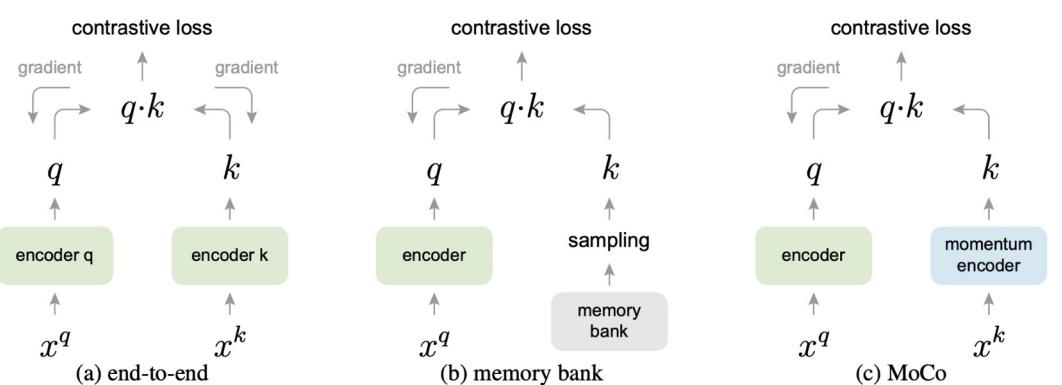
Yu, J., Yuan, Y., Zhang, Q., Zhang, W., Fan, Z., & Jin, F. (2022). OFDM emitter identification method based on data augmentation and contrastive learning. *Applied Sciences*, 13(1), 91.



Ciga, O., Xu, T., & Martel, A. L. (2022). Self supervised contrastive learning for digital histopathology. *Machine Learning with Applications*, 7, 100198.

# Background - Momentum Contrast (MoCo)

- The hypothesis is that good features can be learned by a large dictionary that covers a rich set of negative samples, while the encoder for the dictionary keys is kept as consistent as possible despite its evolution.
- The core of the approach is maintaining the dictionary as a queue of data samples. The introduction of a queue decouples the dictionary size from the mini-batch size.



$$\mathcal{L}_q = -\log \frac{\exp(q \cdot k_+ / \tau)}{\sum_{i=0}^K \exp(q \cdot k_i / \tau)}$$

$$\theta_k \leftarrow m\theta_k + (1 - m)\theta_q$$

# Background - Image augmentations

Image augmentation is important in contrastive learning:

- Creation of positive pairs
- Learning invariant representations
- Creating a diversity of views for better contrast



(a) Original



(b) Crop and resize



(c) Crop, resize (and flip)



(d) Color distort. (drop)



(e) Color distort. (jitter)



(f) Rotate  $\{90^\circ, 180^\circ, 270^\circ\}$



(g) Cutout



(h) Gaussian noise



(i) Gaussian blur

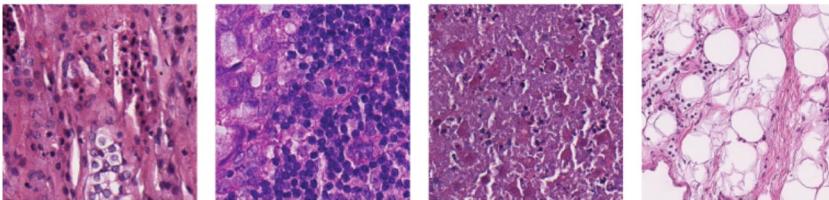


(j) Sobel filtering

Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020, November). A simple framework for contrastive learning of visual representations. In *International conference on machine learning* (pp. 1597-1607). PMLR.

# Motivation

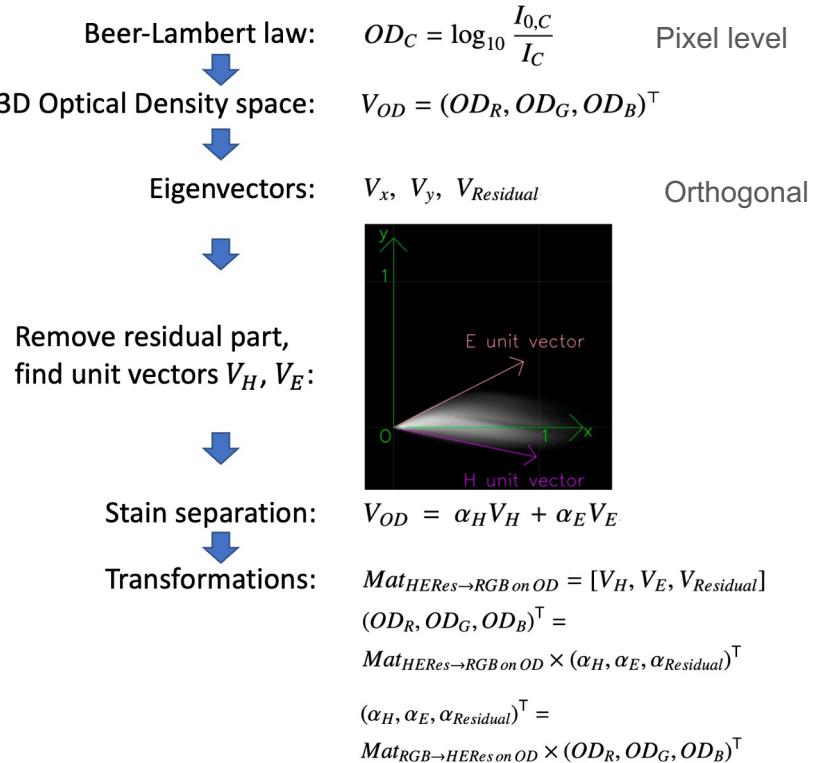
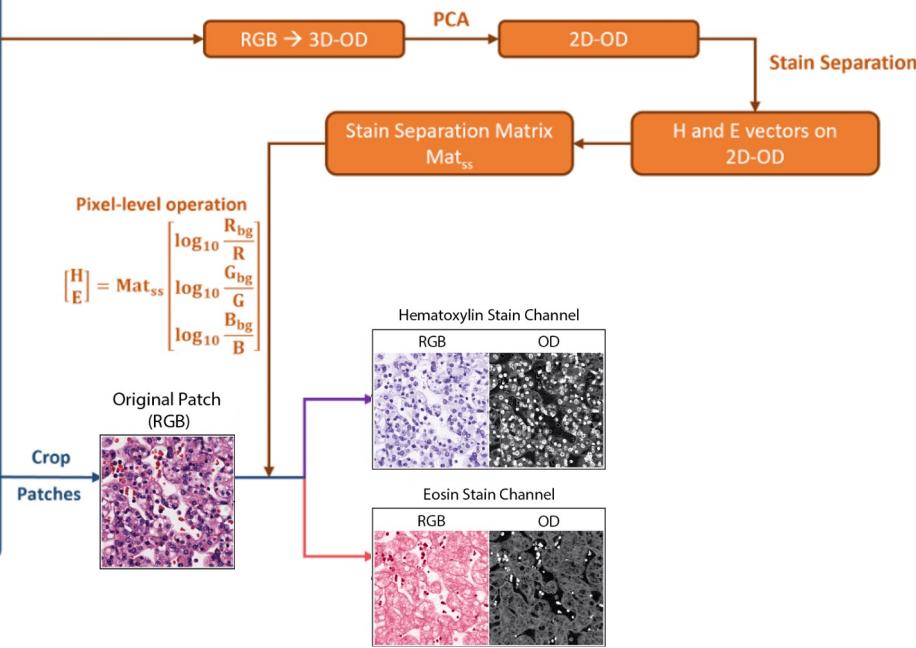
- The standard image augmentation techniques overlook the unique characteristics of histopathological images.
- Hematoxylin & Eosin (H&E) stained histopathological image is the dominant image type in histopathology. Hematoxylin stains the cell nuclei a deep blue or purple, while Eosin stains the cytoplasm and extracellular matrix pink.
- The staining process introduces a lot of variations on stain strengths.



- Idea: In augmentation, we try to randomly set stain strength for each stain channel independently.
- With this idea, the augmented images are also clinically reasonable, and try to mimic the variations on stain strength in real world.

# Stain separation

Whole Slide Image (WSI)



Zhang, B., Manoochehri, H., Ho, M. M., Fooladgar, F., Chong, Y., Knudsen, B. S., ... & Tasdizen, T. (2023). CLASS-M: Adaptive stain separation-based contrastive learning with pseudo-labeling for histopathological image classification. *arXiv preprint arXiv:2312.06978*.

Macenko, M., Niethammer, M., Marron, J. S., Borland, D., Woosley, J. T., Guan, X., ... & Thomas, N. E. (2009, June). A method for normalizing histology slides for quantitative analysis. In *2009 IEEE international symposium on biomedical imaging: from nano to macro* (pp. 1107-1110). IEEE.

# Stain reconstruction augmentation (SRA)

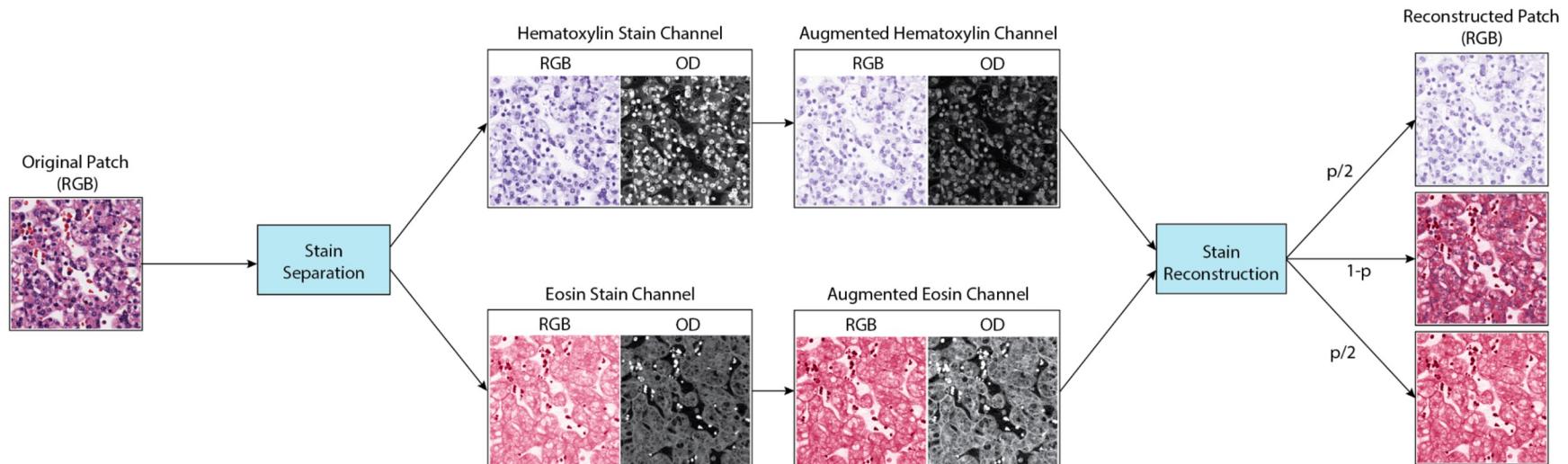


Figure 4. Demonstration of stain reconstruction augmentation (SRA). Single stain images are shown in both RGB space and OD space. The augmentations are performed on each stain channel independently. There is a probability of  $p$  that only single channel is adopted.

# Stain reconstruction augmentation (SRA)

- The original stain strengths vary a lot in different slides.
- Instead of multiplying a random value to original strengths (e.g. [0.95, 1.05]), we directly define the target range of stain strength for each stain channel.

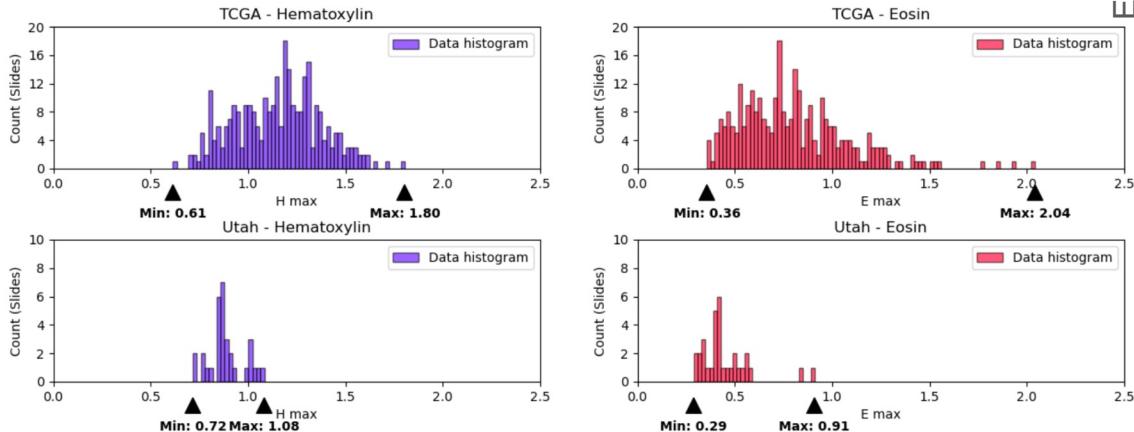


Figure 6. Distributions of strengths of Hematoxylin stain and Eosin stain in Optical Density (OD) space on TCGA training set and Utah training set.

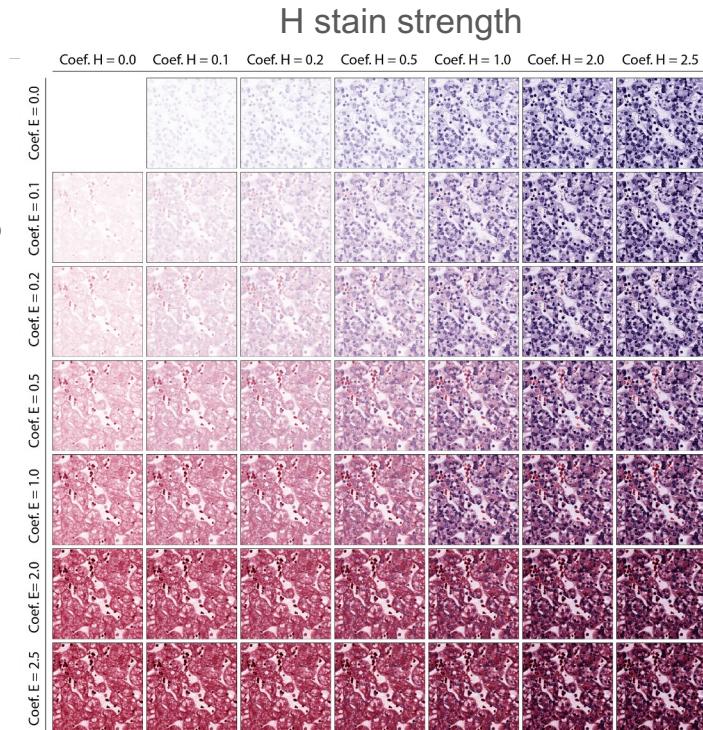
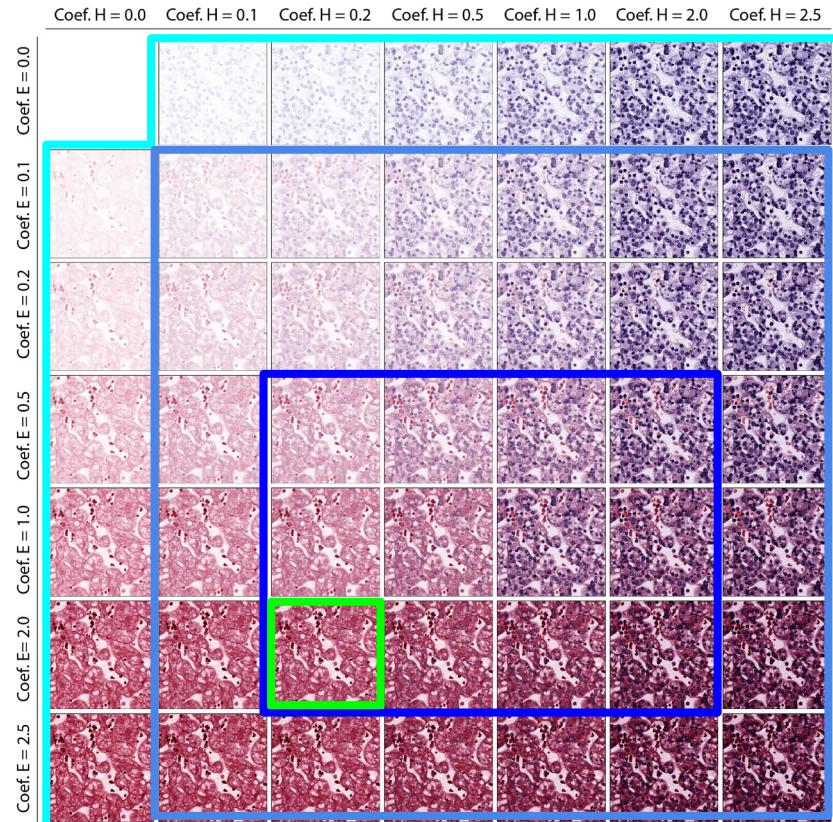
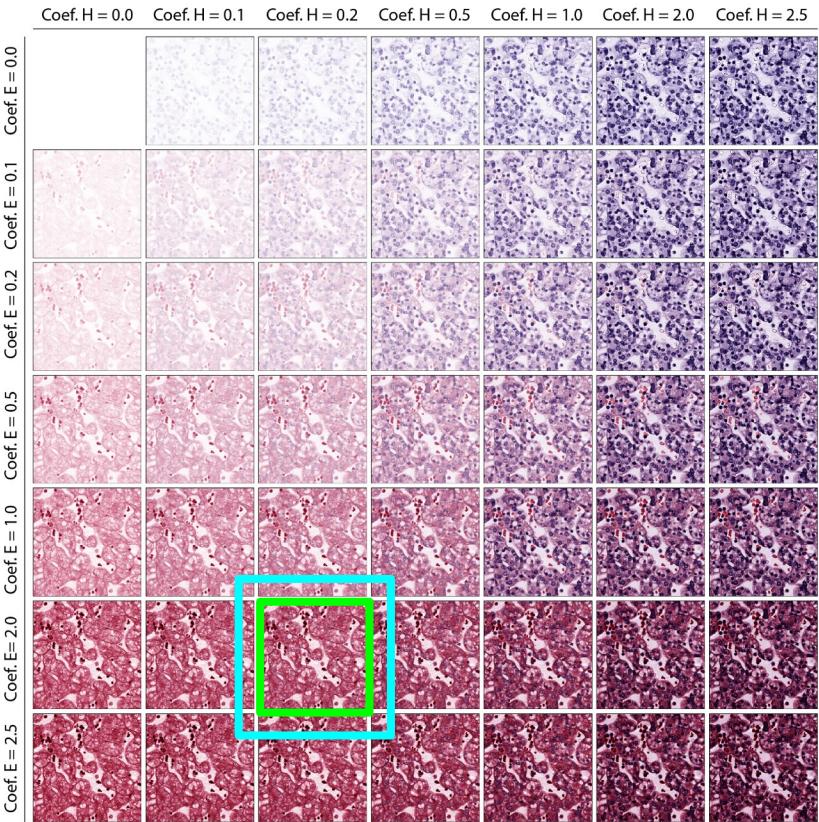


Figure 3. Examples of augmentations by SRA with different target strengths of H channel and E channel.

# Traditional stain augmentation

vs

# Stain reconstruction augmentation (SRA)



Green: original patch

Blue: range of augmentations

# SRA-MoCo v3

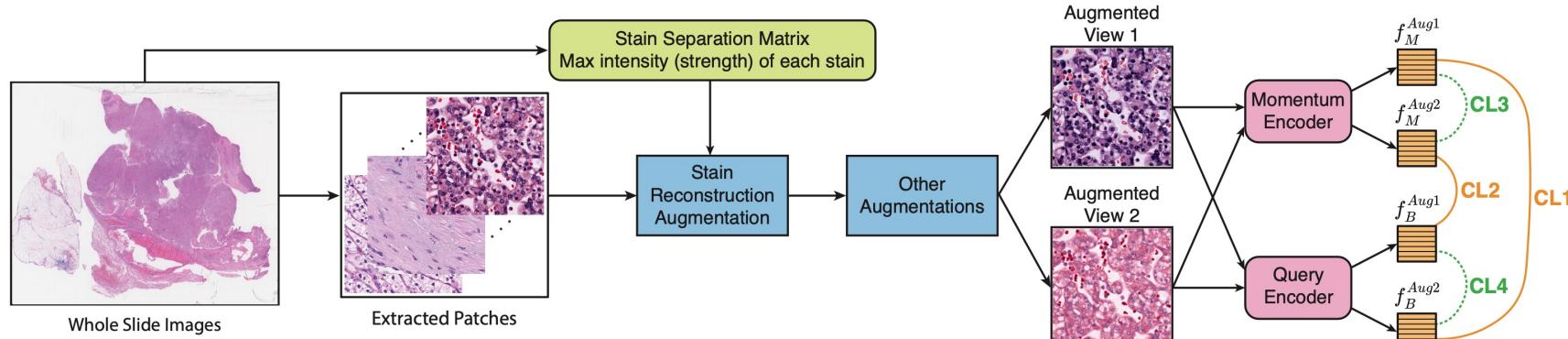


Figure 2: The pipeline of our SRA-MoCo v3. We integrate our Stain Reconstruction Augmentation (SRA) as well as additional contrastive loss terms (CL3 and CL4) into MoCo v3.

# Loss function

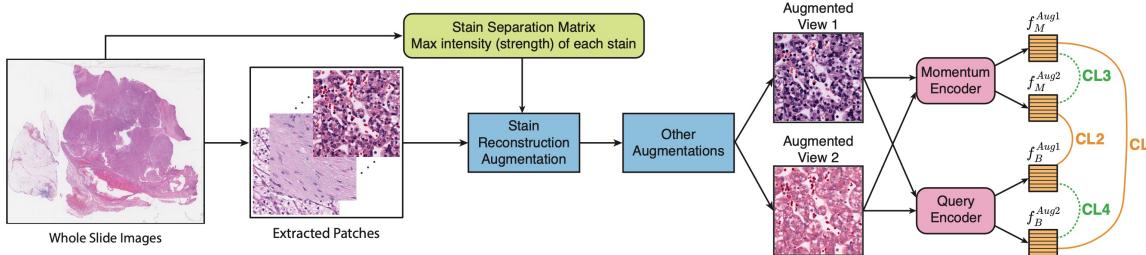


Figure 2: The pipeline of our SRA-MoCo v3. We integrate our Stain Reconstruction Augmentation (SRA) as well as additional contrastive loss terms (CL3 and CL4) into MoCo v3.

$$CL(k, q) = -\log \left( \frac{\exp(q \cdot k^+ / \tau)}{\exp(q \cdot k^+ / \tau) + \sum_{k^-} \exp(q \cdot k^- / \tau)} \right)$$

$$CL_{ori} = CL_1(f_M^{Aug1}, f_B^{Aug2}) + CL_2(f_M^{Aug2}, f_B^{Aug1})$$

---


$$CL_{aug} = CL_3(f_B^{Aug1}, f_B^{Aug2}) + CL_4(f_M^{Aug1}, f_M^{Aug2})$$

$$CL_{ori} + CL_{aug}$$

# Datasets

Utah KIRC as internal dataset, TCGA KIRC and Camelyon16 as external datasets

**Utah KIRC: 49 Slides: 32 train, 10 val, 7 test**

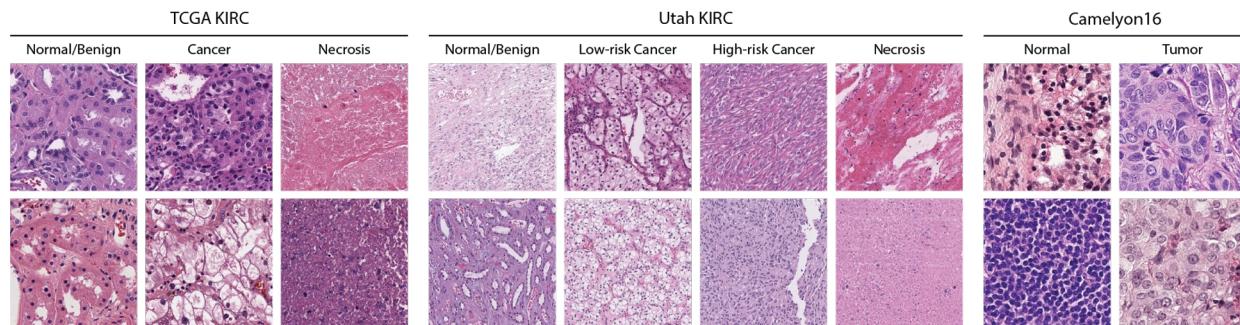
Magnification: 10X

Set	Normal/ Benign	Low-risk Cancer	High-risk Cancer	Necrosis	Total
171,113 Unlabeled					
Train	28,497	2,044	2,522	4,115	208,291
Val.	5,472	416	334	2,495	8,117
Test	7,263	598	389	924	9,174

**Camelyon16:**

Magnification: 20X

Set	Normal	Tumor	Total
Train	144	100	244
Validation	16	11	27
Test	80	50	130



# Experiments Establishing

Pre-training: Self-supervised Contrastive Feature Representation Learning

Classification: Linear Classifier (Tile-based or MIL)

- A. **Scenario 1:** Pre-training and classification on **same** dataset.
- B. **Scenario 2:** Pre-training on TCGA KIRC and classification on Utah KIRC and Camelyon16

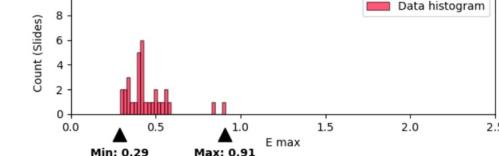
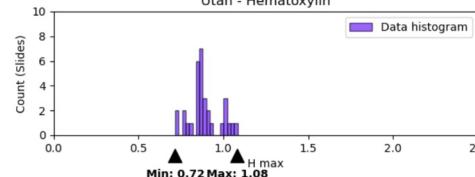
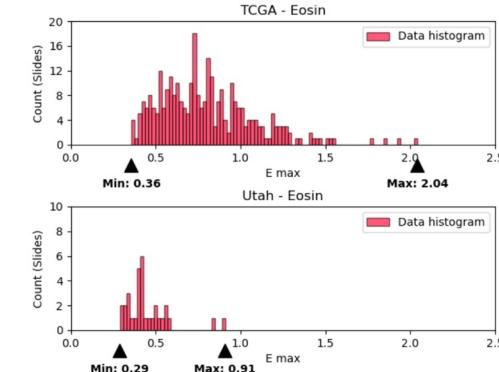
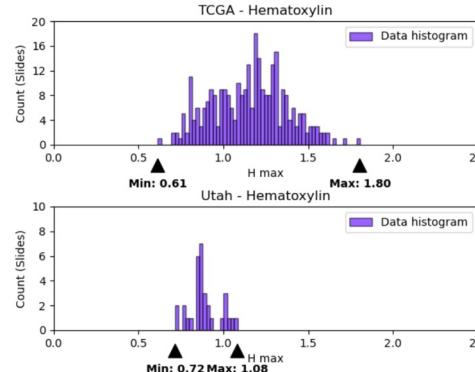
Comparing with:

- Foundational Models:
  - a. Barlow Twins
  - b. MoCo v2
  - c. SwAV
  - d. PathDino
- Similar model: TSA (David Tellez et. al.)

# Experiments Establishing

## Stain reconstruction augmentation ranges:

- **H** max dataset range: TCGA: [0.61-1.80], Utah: [0.72, 1.08]
- **E** max dataset range: TCGA: [0.36-2.04], Utah: [0.29, 0.91]
- Selected Ranges:
  - Setting 1: New H max: [0.5, 2.0], New E max: [0.2, 2.0]
  - Setting 2: New H max: [0.1, 2.5], New E max: [0.1, 2.5]
  - Added option: 5% Pure H + 5% Pure E



# Scenario 1: Pre-training and classification on same datasets

## Compared with foundational models

TCGA KIRC			Utah KIRC		
Pre-trained Dataset	Model	Balanced Accuracy (TCGA KIRC)	Pre-trained Dataset	Model	Balanced Accuracy (Utah KIRC)
ImageNet	ResNet50	$69.97 \pm 5.59$	ImageNet	ResNet50	$87.76 \pm 0.10$
TCGA + TULIP 32.6M patches	Barlow Twins	$81.59 \pm 2.65$	TCGA + TULIP 32.6M patches	Barlow Twins	$90.23 \pm 1.81$
	MoCo v2	$79.04 \pm 0.11$		MoCo v2	$91.45 \pm 0.44$
	SwAV	$77.43 \pm 1.15$		SwAV	$94.96 \pm 1.04$
11,765 TCGA Slides 6.1M patches	PathDino	$76.92 \pm 6.22$	11,765 TCGA Slides 6.1M patches	PathDino	$92.14 \pm 1.65$
TCGA KIRC 300 Slides 1.6M patches	MoCo v3	$79.37 \pm 1.18$	Utah KIRC 49 Slides 0.2M patches	MoCo v3	$93.77 \pm 0.86$
	MoCo v3 + TSA	$81.50 \pm 0.23$		MoCo v3 + TSA	$94.00 \pm 0.26$
	SRA-MoCo v3	$83.62 \pm 0.28$		SRA-MoCo v3	$95.85 \pm 0.34$

## Scenario 2: Pre-training on TCGA and classification on Utah and Camelyon16, Compared with foundational models

		Utah KIRC (Tile-level)		Camelyon16 (MIL-based)	
Pre-trained Dataset	Model	Balanced Accuracy Utah KIRC	F1-score Camelyon16	Accuracy Camelyon16	Balanced Accuracy Camelyon16
ImageNet	ResNet50	$87.76 \pm 0.10$	$0.8372 \pm 0.0149$	$88.37 \pm 0.78$	$86.54 \pm 1.35$
TCGA + TULIP 32.6M patches	Barlow Twins	$90.23 \pm 1.81$	$0.9019 \pm 0.0100$	$93.02 \pm 0.78$	$91.35 \pm 0.67$
	MoCo v2	$91.45 \pm 0.44$	$0.9291 \pm 0.0049$	$94.83 \pm 0.45$	$93.73 \pm 0.13$
	SwAV	$94.96 \pm 1.04$	$0.9264 \pm 0.0098$	$94.57 \pm 0.78$	$93.65 \pm 0.63$
11,765 TCGA Slides 6.1M patches	PathDino	$92.14 \pm 1.65$	$0.9176 \pm 0.0252$	$93.80 \pm 2.05$	$93.15 \pm 1.74$
TCGA KIRC 300 Slides 0.4M/1.6M patches	MoCo v3	$95.32 \pm 0.30$	$0.8075 \pm 0.0135$	$85.79 \pm 1.79$	$84.33 \pm 0.89$
	MoCo v3 + TSA	$94.17 \pm 0.82$	$0.8268 \pm 0.0163$	$87.60 \pm 2.05$	$85.65 \pm 0.94$
	SRA-MoCo v3	$98.12 \pm 0.15$	$0.9207 \pm 0.0084$	$94.31 \pm 0.44$	$92.91 \pm 0.95$

# Ablation Study: Pre-training on TCGA

				Utah KIRC (Tile-level)	Camelyon16 (MIL-based)		
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 $\pm$ 0.30	0.8075 $\pm$ 0.0135	85.79 $\pm$ 1.79	84.33 $\pm$ 0.89
N/A	N/A	0	$CL_{aug}$	95.34 $\pm$ 0.41	0.8457 $\pm$ 0.0253	88.11 $\pm$ 1.19	87.78 $\pm$ 2.44
[0.2, 2.0]	[0.5, 2.0]	0	—	96.51 $\pm$ 0.37	0.8244 $\pm$ 0.0047	87.34 $\pm$ 0.45	85.58 $\pm$ 0.42
[0.1, 2.5]	[0.1, 2.5]	0	—	96.95 $\pm$ 0.76	0.8494 $\pm$ 0.0154	89.67 $\pm$ 0.89	87.19 $\pm$ 1.25
[0.2, 2.0]	[0.5, 2.0]	0	$CL_{aug}$	96.86 $\pm$ 0.17	0.8341 $\pm$ 0.0060	88.37 $\pm$ 0.78	86.15 $\pm$ 0.33
[0.1, 2.5]	[0.1, 2.5]	0	$CL_{aug}$	98.09 $\pm$ 0.12	0.8596 $\pm$ 0.0213	89.15 $\pm$ 2.05	88.75 $\pm$ 1.40
[0.2, 2.0]	[0.5, 2.0]	10%	$CL_{aug}$	97.41 $\pm$ 0.08	0.9079 $\pm$ 0.0150	93.28 $\pm$ 1.18	92.08 $\pm$ 1.02
[0.1, 2.5]	[0.1, 2.5]	10%	$CL_{aug}$	98.12 $\pm$ 0.15	0.9207 $\pm$ 0.0084	94.31 $\pm$ 0.44	92.91 $\pm$ 0.95