Supplementary Material
Unsupervised Latent Stain Adaptation for Computational Pathology

Daniel Reisenbüchler[1,⋆,@], Lucas Luttner[1,⋆], Nadine S. Schaadt[2],
Friedrich Feuerhake[2], and Dorit Merhof[1,3]

[1] Institute of Image Analysis and Computer Vision, University of Regensburg, Regensburg, Germany
[2] Institute of Pathology, Hannover Medical School, Hannover, Germany
[3] Fraunhofer Institute for Digital Medicine MEVIS, Bremen, Germany
[⋆]These authors contributed equally to this work.
[@]Correspondence: daniel.reisenbuechler@informatik.uni-regensburg.de

1 Test Datasets

The following tables serve as overview of the number patches used for each staining and tissue combination in the segmentation task.

Table 1. Overview of the NEPTUNE dataset used for the segmentation experiments

Staining	Tissue class	#Images
PAS	Glomerulus	329
	Glomerular Tuft	352
	Tubule	231
	Artery	264
SIL	Glomerulus	248
	Glomerular Tuft	217
	Artery	223
TRI	Glomerulus	260
	Glomerular Tuft	233
	Artery	324
H&E	Artery	402

Table 2. Overview of the HuBMAP dataset used for the segmentation experiments

Staining	Tissue class	#Images
PAS	Glomerulus	2670

2 Hyperparameter search

All hyperparameter searches were performed via grid search on validation sets only. In the following we detail hyperparameter selections.

cGAN. We performed a careful selection of hyperparameters to ensure that the images were perfectly translated into the target stainings. The number of epochs for the adversarial model was searched within the interval $[200, 500]$ and set to 300, while the learning rate was adjusted to $1.5e - 4$, within the search range $[1e-3, 1e-5]$. The momentum term of Adam was set to 0.5, within the interval $[0.01, 1]$. The buffer for storing artifical images was fixed at 50 $[10, 200]$, and the batch size was fixed at 2, which achieved the best results within the interval of $[1, 4]$. Finally, the number of unlabeled training data was set to 10000, within the interval $[1000, 50000]$.

ULSA. We performed grid search for finding the weighting λ between $\mathcal{L}_{\mathcal{S}}$ and $\mathcal{L}_{\mathcal{U}}$ in the range of $[0.3, 1.5]$ with step size $\Delta = 0.1$. We used the overall batch sizes between labeled b_L and unlabeled samples b_U with $b_{overall} = b_L + b_U = 128$ where we tried $b_U = \lambda b_L$ for different $\lambda = 1, 2, 3$. We tested several noise injection approaches including salt and pepper, gaussian blurring and gaussian noise. Best results were archived with gaussian blurring (kernel size: $(3, 5)$, intensity: $(0.01, 0.4)$). Other augmentation methods like color jitter and random sharpness adjustments were tried, but did not show promising results. We further tried to replace Reinhard with Macenko, which was not possible, due to computational overload (ULSA with Reinhard: 7-10h, ULSA with Macenko: at least 48-60h). Also translating the stains offline and storing them locally would not be possible, because of the huge amount of images needed to store: $x^U = 1.749.458^{|T|}, |T| = [3, 4]$.

Comparable methods. *Reinhard and Macenko.* For each image in the minibatch we used a random target stained image as reference for transformation. Thus each image was translated multiple times into different target stains during training. *UDA.* We used various combinations such as color jitter and gaussian blurring (also see augmentations for ULSA) for data augmentation in the semi-supervised part. Other augmentations lead to worst results. The batch size factor λ for unlabeled data in the unsupervised data augmentation procedure was set to 3 as proposed by the authors within the interval $[3, 5]$. *FixMatch.* We used a confidence threshold of 0.95, which was the same the authors used for their implementations. All other parameters were obtained as in UDA.