Test-Time Stain Adaptation with Diffusion Models for Histopathology Image Classification: Appendix

Cheng-Chang Tsai[1], Yuan-Chih Chen[2], and Chun-Shien Lu[1,2]

[1] Research Center for Info. Technology Innovation, Academia Sinica, Taiwan, ROC
[2] Institute of Information Science, Academia Sinica, Taiwan, ROC
{cctsai, willpower057, lcs}@iis.sinica.edu.tw

A.1 Visual Realization for Stain Generalization Methods

The different stain generalization methods are illustrated in Fig. 4 for visual realization.

A.2 Maximum A Posteriori (MAP) Estimation & Pseudo Code

MAP To make TT-SaD easier to follow, we provide an additional framework to understand the proposed method. Eq. (11) is the inverse problem that we aim to solve. The solution $\hat{x}$ can be obtained by solving a maximum a posteriori estimation problem as follows:

$$\hat{x} = \underset{x}{\operatorname{argmax}} \ \log p(x^t|x) + \log p(x), \tag{17}$$

where $\log p(x^t|x)$ is the log-likelihood term of x^t and $\log p(x)$ is the prior term of source data.

TT-SaD solves Eq. (17) in an iterative fashion. In each iteration, TT-SaD first updates the prior term via Eq. (15) and then the likelihood term via Eq. (16).

Pseudo Code The Pseudo code of the proposed test-time stain adaption method is described in Algorithm 1.

A.3 Datasets and Sample Images

We show the sample images of each hospital in Fig. 5 and specify the numbers of whole-slide images and histopathology images in each hospital in Table 7.

A.4 Comparison with training-time methods under different scenarios of testing data

Here, we demonstrate more comparison scenarios of tumor classification between TT-SaD (with domain tumor center) and other "training-time" stain augmentation methods in Table 8.

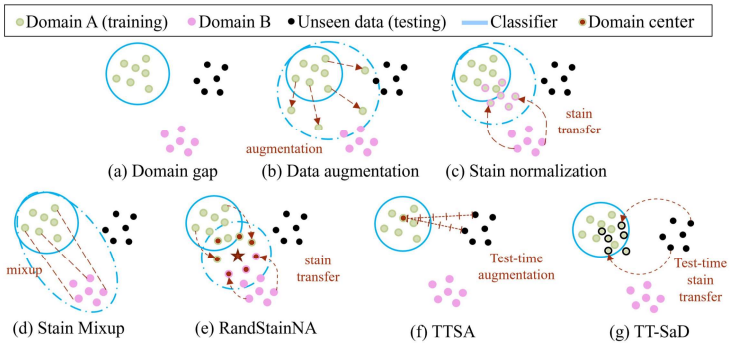

Fig. 4: Illustration of different methods for stain generalization of a classifier (blue circle), where the samples (green dots) from domain A represent training data, the samples (dark dots) denote the unseen data for testing, the samples (pink dots) from domain B represent the additional data for training, if necessary. Stain Mixup [6] and RandStainNA [40] are training-time methods, while TTSA [47] and our TT-SaD method belong to test-time methods. (a) Domain gaps exist between different domains and unseen data. (b) Data augmentation increases the classifier's generalization by stain variation. (c) Stain normalization keeps stain consistency by converting domain A and domain B to the same stain pattern. (d) Stain Mixup [6] mixes stain matrices of samples from domain A and domain B to fill the space between the two domains. (e) RandStainNA [40] normalizes all samples to the mean and standard deviation of each channel of domain A and domain B in the color space. (f) TTSA [47] is the test-time augmentation version of (d) in that the domain center is the nearest image to the average stain matrix of the entire domain A, and augmentations are the six interpolation points from the testing data to the domain center (*i.e.*, mixup two stain matrices with different parameters). (g) Our TT-SaD method transfers the testing data by diffusion model with the stain matrix from domain A.

Table 7: Numbers of slides and patches/images in each hospital.

Dataset	MitosAtypia14		CAMELYON17				
Hospital	Aperio	Hamamatsu	1	2	3	4	5
Slides	1,136	1,136	10	10	10	10	10
Images	35,996	35,993	105,111	279,525	91,573	154,031	51,781

A.5 Visualization of Stain Adaption

Here, we illustrate the visualizations of stain matrix distributions from all hospitals and those after stain shift in Figs. 6 and 7, respectively. We can observe from Fig. 6 that the points from all hospitals, excluding Hospital 1, before stain shift do not fully overlap with those from Hospital 1. Nevertheless, after stain

Algorithm 1 Test-Time Stain Adaptation Model

Input: Input image x^t, Stain matrix of the desired stain W_s
Output: Generated image x_0
W_t: stain matrix of x^t ▷ by Equations (1) and (2)
A: stain shift matrix ▷ by Equation (10)
Sample $x_N \sim q\left(x_N \mid x^t\right)$ ▷ by Equation (5)
for $t \leftarrow N \ldots 1$ **do**
    $x_{0;t} = \frac{x_t - \sqrt{1-\bar{\alpha}_t}\epsilon_\theta(x_t,t)}{\sqrt{\bar{\alpha}_t}}$
    $\hat{x}_{0;t} = \mathrm{BL}^{-1}\left(A^\dagger(\mathrm{BL}(x^t)) + (\mathbf{I} - A^\dagger A)\mathrm{BL}(x_{0;t})\right)$
    $x_{t-1} = \sqrt{\bar{\alpha}_{t-1}}\hat{x}_{0;t} + \sqrt{1-\bar{\alpha}_{t-1}-\sigma_t^2}\epsilon_\theta(x_t,t) + \ \sigma_t\epsilon$
end for
return x_0

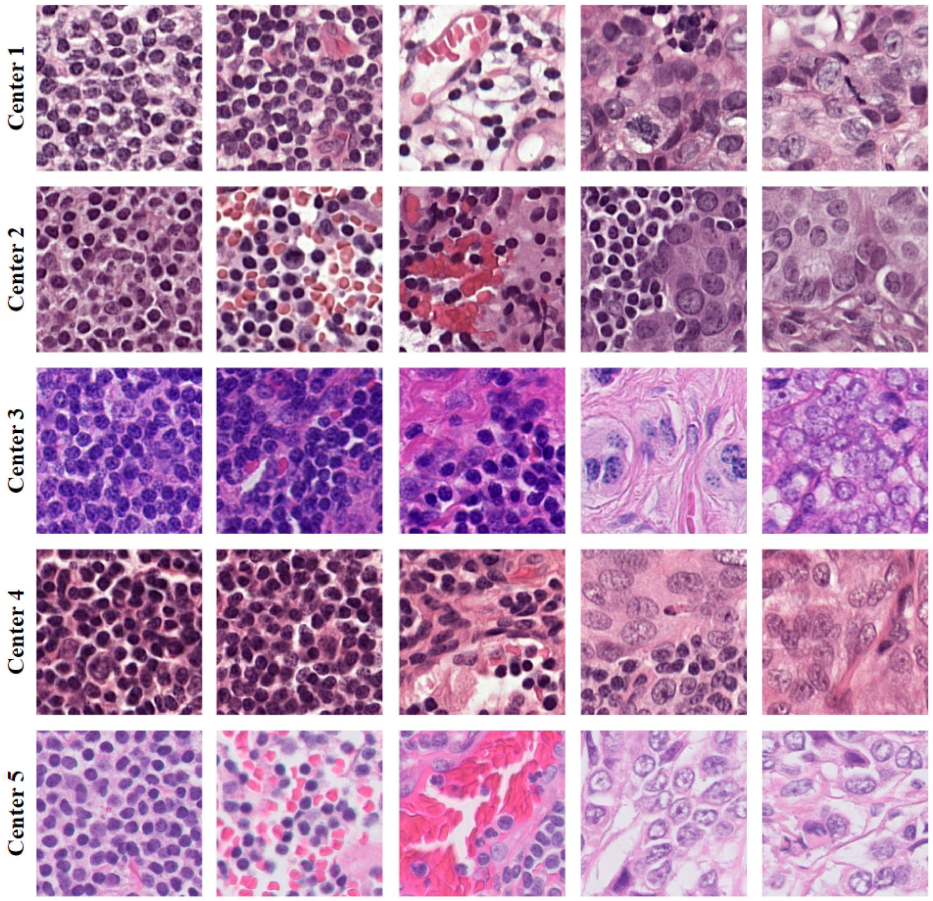

Fig. 5: Sample images from all hospitals of the CAMELYON17 dataset [5].

Table 8: Comparison of tumor classification between TT-SaD (with domain tumor center) and other training-time stain augmentation methods. "v" denotes the use of a specific dataset. "w/o any augmentation" denotes "Training data from H1" and "Testing data from the combination of other hospitals.

	Training data					Testing data					Result		
Methods \ Hospital	H1	H2	H3	H4	H5	H1	H2	H3	H4	H5	ACC	AUC	AUPRC
w/o any augmentation	v							v	v	v	58.71	77.93	38.70
Stain Mix-Up [6]	v	v						v	v	v	73.47	77.87	43.89
RandStainNA [40]	v	v						v	v	v	75.23	15.86	2.94
TT-SaD (Ours)	v							v	v	v	**81.54**	**89.89**	**44.20**
w/o any augmentation	v						v	v	v	v	62.57	81.92	30.13
Stain Mix-Up [6]	v	v					v	v	v	v	77.35	76.71	33.88
RandStainNA [40]	v	v					v	v	v	v	80.53	24.73	2.33
TT-SaD (Ours)	v						v	v	v	v	**85.87**	**90.73**	**35.71**
w/o any augmentation	v					v		v	v	v	68.80	83.44	53.85
Stain Mix-Up [6]	v	v				v		v	v	v	77.20	82.76	52.86
RandStainNA [40]	v	v				v		v	v	v	80.39	23.62	2.95
TT-SaD (Ours)	v					v		v	v	v	**85.92**	**92.42**	**57.97**
w/o any augmentation	v					v	v	v	v	v	69.86	85.53	43.96
Stain Mix-Up [6]	v	v				v	v	v	v	v	79.56	80.85	43.06
RandStainNA [40]	v	v				v	v	v	v	v	83.60	29.16	2.46
TT-SaD (Ours)	v					v	v	v	v	v	**88.50**	**92.58**	**48.43**

shift, we do observe from Fig. 7 that the points of all hospitals other than Hospital 1 move toward Hospital 1 so that the dispersion extent of points in each cluster becomes much more concentrated. Therefore, it is empirically observed that the stain matrix distributions can be shifted to that of Hospital 1 quite well.

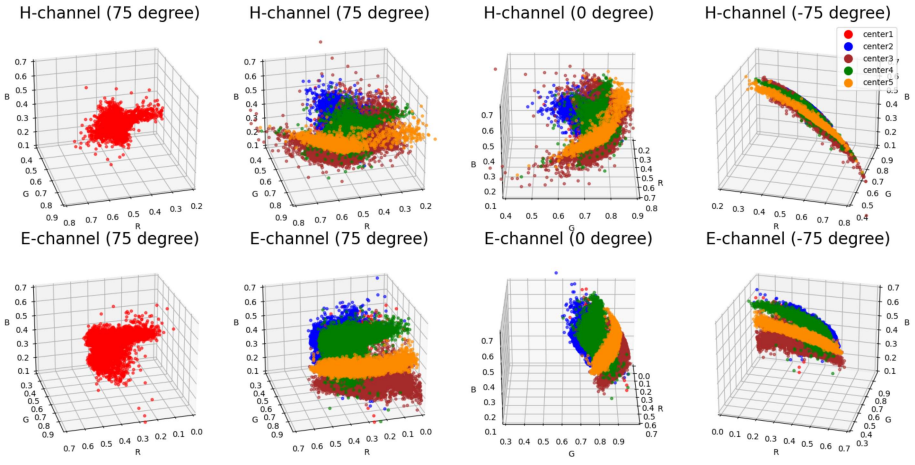

Fig. 6: The 3D plot (in RGB space) of stain matrices from five hospitals before stain shift. "H-channel" and "E-channel" denote the two channels of H&E staining, and "degree" indicates the degree of rotation along the B axis. The leftmost column illustrates the single stain matrix distribution of Hospital 1, without being obscured by data from other hospitals. The remaining columns show that the stain matrix distributions of all hospitals are separated well in three different views.

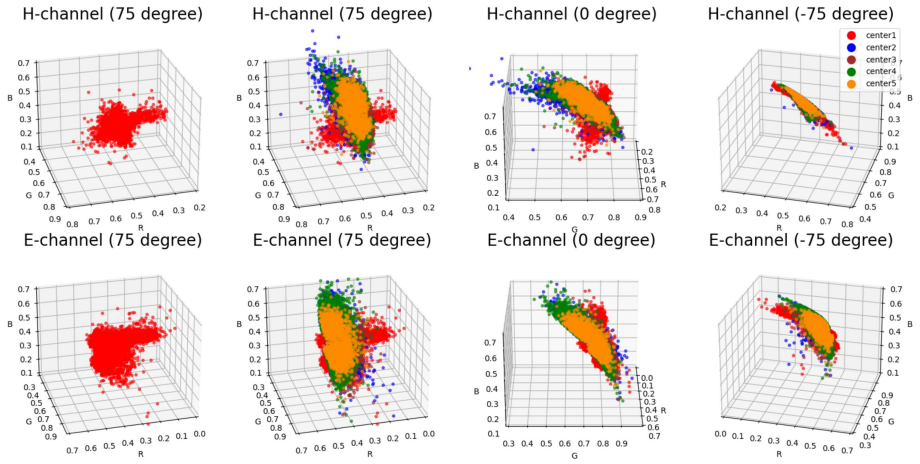

Fig. 7: The 3D plot (in RGB space) of stain matrices from five hospitals after being shifted to Hospital 1. "H-channel" and "E-channel" denote the two channels of H&E staining, and "degree" indicates the degree of rotation along the B axis. The leftmost column illustrates the single stain matrix distribution of Hospital 1, without being obscured by data from other hospitals. The remaining columns show that the stain matrix distributions of all hospitals are greatly overlapped together in three different views.