Stain Mix-Up: Unsupervised Domain Generalization for Histopathology Images

Jia-Ren Chang[1(✉)], Min-Sheng Wu[1], Wei-Hsiang Yu[1], Chi-Chung Chen[1], Cheng-Kung Yang[1], Yen-Yu Lin[2], and Chao-Yuan Yeh[1(✉)]

[1] aetherAI, Taipei, Taiwan
{JiaRenChang,vincentwu,seanyu,chenchc,jimmy,joeyeh}@aetherai.com
[2] National Yang Ming Chiao Tung University, Hsinchu, Taiwan
lin@cs.nctu.edu.tw

Abstract. Computational histopathology studies have shown that stain color variations considerably hamper the performance. Stain color variations indicate the slides exhibit greatly different color appearance due to the diversity of chemical stains, staining procedures, and slide scanners. Previous approaches tend to improve model robustness via data augmentation or stain color normalization. However, they still suffer from generalization to new domains with unseen stain colors. In this study, we address the issue of unseen color domain generalization in histopathology images by encouraging the model to adapt varied stain colors. To this end, we propose a novel data augmentation method, **stain mix-up**, which incorporates the stain colors of unseen domains into training data. Unlike previous mix-up methods employed in computer vision, the proposed method constructs the combination of stain colors without using any label information, hence enabling unsupervised domain generalization. Extensive experiments are conducted and demonstrate that our method is general enough to different tasks and stain methods, including H&E stains for tumor classification and hematological stains for bone marrow cell instance segmentation. The results validate that the proposed stain mix-up can significantly improves the performance on the unseen domains.

Keywords: Domain generalization · Mix-up · Stain color

1 Introduction

Computer-aided diagnosis based on histopathology images, such as whole slide images (WSIs) and field of views (FoVs) of tissue sections, gains significant progress owing to the great success of machine learning algorithms in digital pathology. Tissue sections are typically stained with various stains to make tissues visible under the microscope. However, tissue manipulation, staining,

---
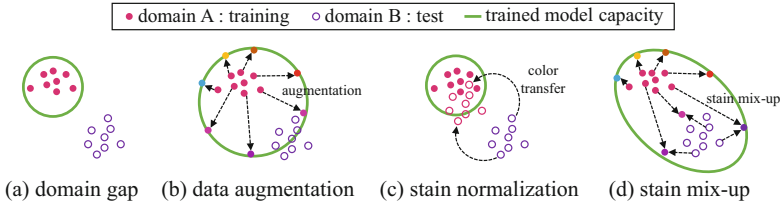
Fig. 1. Comparisons among different methods for model capacity generalization for histological images analysis. (a) There exists the domain gap between domain A (red dots) and domain B (purple cycles) due to the color variation; (b) Data augmentation is adopted to increase the color variations of domain A; (c) Color normalization transfers domain B to the color patterns of domain A to avoid color variation; (d) In this paper, we propose stain mix-up for randomly augmenting the domain A according to the stain color matrices of domain B and thus generalize the model to domain B. Moreover, the proposed domain generalization technique is unsupervised where data labels on domain B are not required. (Color figure online)

and even scanning often result in substantial color appearance variations in histopathology images, and degrade machine learning algorithms due to the domain gap of colors. Thus, it is crucial to take color appearance variations into account when developing machine learning algorithms for histopathology image analysis. Specifically, two strategies are widely used, including 1) augmenting color patterns of training data to enhance model robustness; and 2) normalizing all histopathology images to a single color pattern so that the unfavorable impact of color variations in the subsequent process can be alleviated.

To augment color patterns of training data, most techniques conduct brightness, contrast and hue perturbations [7,19]. Specifically, Bug *et al.* [3] utilize principle component analysis (PCA) to decompose images into a low-dimensional space spanned by a few principle components where augmentation of H&E images is carried out by perturbing the main components. Tellez *et al.* [17] divide H&E images into the hematoxylin, eosin, and residual color channels by color deconvolution [14]. Then, stain-specific transformation by perturbing each channel is used to complete augmentation.

Another way for addressing color variations is to normalize all images to have similar color patterns. Several color normalization methods have been proposed to achieve it. Given a source image and a target image, Reinhard *et al.* [13] convert images to the $l\alpha\beta$ space and normalize the source image by aligning the mean and standard deviation of the source and target images. Other methods decompose stain colors into a stain color matrix and a stain density map, and apply the stain color matrix of the target image onto the source image. Macenko *et al.* [10] use singular value decomposition while Vahadane *et al.* [18] employ sparse non-negative matrix factorization (SNMF) to decompose the stain color matrix. Nadeem *et al.* [12] adopt Wasserstein barycenter to normalize images.

We observe that highly effective methods [17,18] eliminate color variations by decomposing stain color matrices for further transferring or augmenting. It implies that stain color matrices can reliably encode the color patterns. However,

current studies [13,17,18] consider stain color matrices decomposed from only a single domain and thus may restrict their generalization abilities.

In this paper, we propose a novel method **stain mix-up** for data augmentation. It randomly interpolates a new stain color matrix between different domains during training and can improve the generalization performance accordingly. The Mix-up technique [20] has become essential to data augmentation for recognition [6] or domain adaption [9,11] in computer vision. In contrast to Mix-up [20] which mixes images and labels, the proposed method is *label-free*. It mixes stain color matrices between different domains and can synthesize various types of stain colors for learning color-invariant representations. In Fig. 1, we describe the concept of different approaches for generalizing histology image analysis, including data augmentation, stain normalization and the proposed stain mix-up. Extensive experiments are conducted on two kinds of stains and tasks, *i.e.*, tumor classification on H&E stained images and bone marrow cell instance segmentation on hematological stained images. Since both tasks have multiple sources (domains) of images, we train the model on one domain where data were collected from one medical center, and test it on others. To this end, the training center is denoted as the source domain which consists of images and labels whereas other centers are denoted as the target domain which only have images but no labels. The stain color matrices of the target domain are mixed with those of the source domain to synthesize new training samples in the source domain for on-the-fly augmentation. The results show that the proposed stain mix-up achieves the-state-of-art generalization performance on both tasks.

The main contributions of this work are summarized as follows. First, we propose a novel data augmentation approach, namely stain mix-up, to achieve unsupervised domain generalization for histology image analysis. Second, we perform extensive experiments to demonstrate the effectiveness of the proposed method. It turns out that our method consistently achieves the state-of-the-art performance on different tasks and stains. To the best of our knowledge, the proposed method is the first work for unsupervised domain generalization in histology image analysis.

2   Method

This section describes the proposed method, which is composed of two stages: stain separation and stain mix-up augmentation. The former is conducted to extract color characteristics from histology images of different domains. It estimates stain color matrices that represent chromatic components and stain density maps of each domain. The latter uses the estimated matrices of different domains to augment training images on-the-fly through the proposed stain mix-up, enabling unsupervised domain generalization. Details of the two stages are elaborated as follows. Figure 2 illustrates the pipeline of our method.

2.1   Stain Separation via SNMF

Stains are optical absorption materials that occlude certain spectra of light, making tissues visible in the complementary colors. They help visualize tissues for
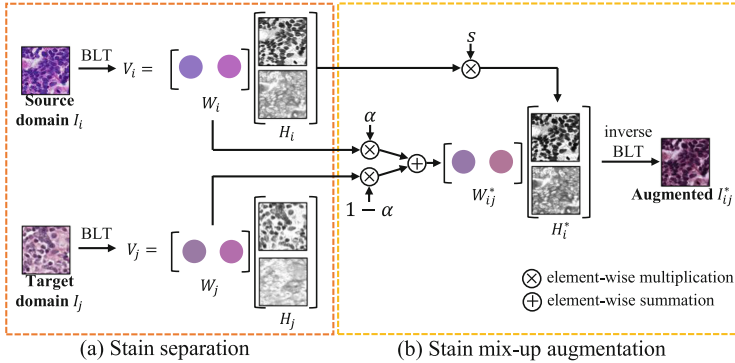
Fig. 2. Pipeline of the proposed stain mix-up augmentation. (a) Given a labeled image I_i from the source domain and an unlabeled image I_j from the target domain, stain separation decomposes the optical density of I_i derived via Beer-Lambert transformation (BLT), *i.e.* V_i, into its stain color matrix W_i and density map H_i. Similarly, we have W_j and H_j from V_j. (b) Stain mix-up augmentation is carried out by applying inverse BLT to a mixed stain color matrix W_{ij}^* and a perturbed density map H_i^*.

medical diagnosis. Stained tissue colors result from light attenuation, depending on the type and amount of stains that tissues have absorbed. This property can be mathematically formulated by Beer-Lambert law [2] defined as follows:

$$V = -\log \frac{I}{I_0} = WH \ , \tag{1}$$

where $I \in \mathbb{R}^{3 \times n}$ is a histology image in the RGB color space, I_0 is the illuminating light intensity of sample with $I_0 = 255$ for 8-bit images in our cases, $W \in \mathbb{R}^{3 \times m}$ is the stain color matrix to encode the color appearance of each stain, and $H \in \mathbb{R}^{m \times n}$ is the density map of each stain, for an m-stained n-pixel image. Given a histology image I, its optical density V can be derived via Beer-Lambert transformation (BLT) in Eq. 1. Stain separation aims to estimate the corresponding stain color matrix W and density map H. In this work, we follow [18] and solve W and H of a histology image I through SNMF in the experiments.

2.2   Stain Mix-Up Augmentation

The proposed stain mix-up for data augmentation aims to reduce the domain gaps caused by color appearance variations between different sources. It synthesizes diversified images for augmentation and can increase the potential data coverage. As shown in Fig. 2, we carry out this task by mixing the stain color characteristics of both the source and target domains. Specifically, we randomly sample a pair of histological images I_i and I_j from the source domain and the target domain respectively, followed by decomposing them into the stain color matrices W_i and W_j and the stain density maps H_i and H_j through BLT and
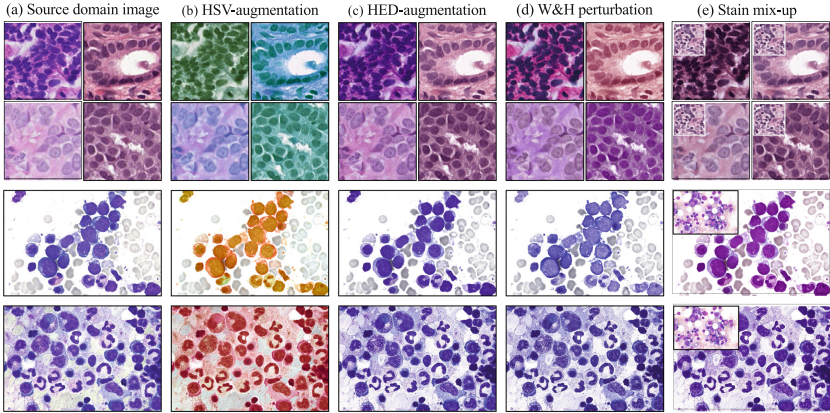
Fig. 3. Comparisons among different augmentation methods. The **top two rows** show images from CAMELYON17 [1] and the **bottom two rows** show images from Hema. (a) shows the training images from the source domain. (b), (c), and (d) are the augmented training images generated via three baseline methods which cannot include target domain information. The results of our proposed stain mix-up are demonstrated in (e). By mixing the stain matrices from (a) source domain image and target domain images (upper-left corners in (e)), the stain mix-up yields more realistic stain color compared with the other augmentations. Therefore, our generated images can effectively help accommodate the model to the target domain.

SNMF. A mixed stain color matrix W_{ij}^* is a linear interpolation between W_i and W_j with a coefficient α randomly sampled from a uniform distribution, namely,

$$W_{ij}^* = \alpha W_i + (1 - \alpha)W_j, \quad \text{where } \alpha \sim U(0, 1). \tag{2}$$

Random interpolation between stain color matrices increases the diversity of stain color appearance while keeping the mixed stain color matrices realistic, thus improving the generalization ability to the target domain.

Similar to [17], we perturb the stain density map H_i to simulate the extents of stain concentrations and color fading,

$$H_i^* = sH_i, \quad \text{where } s \sim U(1 - \delta, 1 + \delta), \tag{3}$$

where s is a scaling factor randomly drawn from a uniform distribution controlled by $\delta \in [0, 1]$. By referring to the interpolated stain color matrix W_{ij}^* in Eq. 2 and the perturbed map H_i^* in Eq. 3, the resulting augmented image I_{ij}^* is generated by the inverse BLT,

$$I_{ij}^* = I_0 \exp\left(-W_{ij}^* H_i^*\right). \tag{4}$$

Figure 3 shows several examples of the augmented images.

In this study, the histological images are collected from multiple medical centers, and their stained color appearances may vary considerably due to different staining processes. We aim to generalize the model trained on labeled data of one center (source domain) to unlabeled data of other centers (target domains). To

this end, the proposed stain mix-up is applied for augmentation. Unlike existing augmentation methods such as [13,17,18] where only source domain data are considered, our method leverages data from both source and target domains to synthesize augmented data that are more consistent with the target domain. Compared with existing domain generalization methods such as [20], our method make use of image data without labels in the target domain, and hence enables unsupervised domain generalization. The properties described above make the stain mix-up a simple yet efficient module that can augment images on-the-fly to achieve a state-of-the-art performances on various tasks.

3   Experiments

In this section, after describing materials, implementation details, and evaluation metrics of our proposed method, we present and elaborate the results of the experiments. Two datasets, namely CAMELYON17 and Hema, are adopted to experiment with different augmentation methods on different types of computer vision tasks in histology image analysis. To better understand how stain matrix augmenting affects the model generalization, we also perform an ablation study on CAMELYON17 for validating the effect from perturbing W and H.

3.1   Datasets

CAMELYON17. We use the CAMELYON17 [1] dataset to evaluate the performance of the proposed method on tumor/normal classification. In this dataset, a total of 500 H&E stained WSIs are collected from *five* medical centers (denoted by C_1, C_2, ... C_5 respectively), 50 of which include lesion-level annotations. All positive and negative WSIs are randomly split into training/validation/test sets with the following distributions: $C_1 : 37/22/15$, $C_2 : 34/20/14$, $C_3 : 43/24/18$, $C_4 : 35/20/15$, $C_5 : 36/20/15$. We extract image tiles in a size of 256×256 pixels from the annotated tumors for positive patches and from tissue regions of WSIs without tumors for negative patches.

Hema. We evaluate the proposed method on a custom hematology dataset for bone marrow cell instance segmentation. In the Hema dataset, a total of 595 WSIs of hematological stained bone marrow smears are collected from *two* medical centers, denoted by M_1 and M_2 respectively. We sample 21,048 FoVs from M_1 as training data and 311 FoVs from M_2 as testing data. All FoVs are in a size of $1,149 \times 1,724$. This dataset has a total of 662,988 blood cell annotations, which are annotated by a cohort of ten annotators, consisting of senior hematologists and medical technicians with an average of over ten years of clinical experience.

3.2   Implementation Details

For the CAMELYON17 dataset, we train five ResNet-50 [5] classifiers on each center individually and test each classifier on the test data of all centers to

Table 1. (**Top**) Mean AUC of different methods for tumor classification on the CAME-LYON17 [1] dataset. (**Bottom**) Ablation studies on the components of our method.

Method	C_1	C_2	C_3	C_4	C_5	Average (95% CI)
No manipulation	0.947	0.753	0.854	0.777	0.861	0.838 (0.824, 0.852)
Vahadane normalization [18]	0.943	0.826	0.898	0.853	0.707	0.846 (0.829, 0.861)
HSV-aug.	0.984	0.844	0.954	0.937	0.845	0.913 (0.902, 0.924)
HED-aug. [17]	0.984	0.925	0.903	0.861	0.908	0.916 (0.909, 0.924)
Stain mix-up (ours)	0.979	0.948	0.946	0.965	0.942	**0.956** (0.953, 0.959)
Ablation study						
H perturb	0.984	0.925	0.903	0.861	0.908	0.916 (0.909, 0.924)
W, H perturb	0.979	0.948	0.911	0.965	0.939	0.948 (0.944, 0.953)
W mix-up, H perturb	0.979	0.948	0.946	0.965	0.942	**0.956** (0.953, 0.959)

evaluate the effectiveness of generalization. Since the CAMELYON17 dataset contains H&E stained images, we decompose each image into a stain color matrix $W \in \mathbb{R}^{3 \times 2}$ and a density map $H \in \mathbb{R}^{2 \times n}$. The parameter δ in Eq. 3 is set to 0.2. All models are trained with AdamW [8], a learning rate of 0.001, and a batch size of 32 for 100,000 iterations with an Nvidia Quadro RTX8000.

For the Hema dataset, we adopt Mask R-CNN [4] with the ResNet-50 backbone pre-trained on ImageNet [15] for instance segmentation. The stain of WSIs in Hema is composed of three chemicals, namely methylene blue, eosin, and azur. Thereby, we factorize each image into a stain color matrix $W \in \mathbb{R}^{3 \times 3}$ and a density map $H \in \mathbb{R}^{3 \times n}$. The parameter δ is set to 0.5. The model is trained on M_1 with SGD, a learning rate of 0.02, a momentum of 0.9, a batch size of 4, and weight decay 0.0001 for 12 epochs with an Nvidia V100. After data augmentation and model fine-tuning, we evaluate the generalization performance on M_2.

Please note that the stain matrices are calculated using SNMF before training for saving computational time. That is, we only compute stain matrices once and use them repeatedly during training. The computational time of SNMF decomposition for a single image in CAMELYON17 and Hema takes 1.14 and 2.40 s, respectively, measured on an Intel Xeon CPU E5-2697 v3.

3.3   Results on the CAMELYON17 Dataset

In Table 1, we compare the proposed stain mix-up with existing augmentation methods for tumor classification on the CAMELYON17 dataset. It is consistent with previous findings [16,21] that models trained without color augmentations result in weaker performance and a larger performance fluctuation when testing on images from other centers (AUC = 0.838, 95% CI 0.824 − 0.852), which reveals the domain gaps among different centers. The models trained with data augmented by the proposed stain mix-up achieves significant performance gains

Table 2. Performance in mAP of bone marrow cell instance segmentation using different augmentation methods on the Hema dataset.

Method	Box			Mask		
	AP	AP_{50}	AP_{75}	AP	AP_{50}	AP_{75}
No manipulation	48.3	52.5	51.2	49.6	52.5	52.0
HSV-augmentation	48.7	53.4	52.1	50.2	53.4	52.9
HED-augmentation [17]	48.1	52.4	51.5	49.5	52.5	52.0
Stain mix-up (ours)	**50.3**	**55.4**	**54.5**	**52.1**	**55.5**	**55.1**

over HSV-augmentation and HED-augmentation [17]. In addition, the stain mix-up method helps reach stable performance when evaluated on images of different centers, while other competing methods show larger performance variations. We attribute these advantages to the cross-domain interpolation of the stain color matrices in the proposed stain mix-up, while the competing methods such as HSV-augmentation and HED-augmentation refer to only images of the source domain. The augmented images by our method are realistic and more consistent with those in the target domain, leading to a better generalization ability.

Cross-domain interpolation is the key component of the proposed stain mix-up for unsupervised and diversified stain color matrix generation. While the stain color matrix W can be interpolated between the source and target domains, it can be self-perturbed with random degrees sampled from a uniform distribution. The implementation details of the self-perturbed W is described in the supplementary material. In the ablation study, we explore how different perturbation methods contribute to the model generalization. Some example patches generated by using random W, H perturbation are visualized in Fig. 3d. As shown in the bottom of Table 1, stochastic fluctuations in W achieve the AUC of 0.948 (95% CI $0.944 - 0.953$), which is inferior to models trained with the stain mix-up. This result suggests that: 1) Models can benefit from perturbing interaction between color channels, and 2) with the identified stain matrix of centers in advance, interpolating combinations of matrices can be more effective for model adaptation across different centers.

3.4   Results on the Hema Dataset

In addition to tumor classification, we evaluate the proposed stain mix-up for cell instance segmentation on the Hema dataset. As shown in Table 2, our method consistently outperforms baseline methods by substantial margins, more than 2.0% in box mAP and mask mAP in most cases. We observe that the baseline methods, HSV-augmentation and HED-augmentation, make no improvement on this dataset. The major reason is that the large domain gap makes augmentation based on only the source domain, which is irrelevant to the target domain. By taking the stain color matrices of unlabeled data of the target domain into consideration, our method can effectively accomplish domain generalization in

an unsupervised manner. The results validate that our model can alleviate the domain gaps between different histology image collections even on the challenging instance segmentation task.

4   Conclusions

We have presented stain mix-up, a simple yet effective data augmentation method for unsupervised domain generalization in histological image analysis. Our stain mix-up constructs various virtual color patterns by random linear interpolation of two stain color matrices, one from the source domain and one from the target domain. Cross-domain interpolation refers to color distributions of both domains, and color patterns that are realistic and more consistent to the target patterns can be synthesized, facilitating model adaptation to the target domain. Since accessing only stain color matrices is label-free, the proposed method carries out unsupervised domain generalization. Through extensive experiments, we have shown that the proposed stain mix-up significantly improves the generalization ability on diverse tasks and stains, such as tumor classification on the H&E stained images and bone marrow cell segmentation on the hematological stained images. We believe the proposed stain mix-up can advance the community of digit pathology for practical usage.

References

1. Bandi, P., et al.: From detection of individual metastases to classification of lymph node status at the patient level: the camelyon17 challenge. IEEE Trans. Med. Imaging **38**(2), 550–560 (2018)
2. Beer, A., Beer, P.: Determination of the absorption of red light in colored liquids. Annalen der Physik und Chemie **86**(5), 78–88 (1852)
3. Bug, D., et al.: Context-based normalization of histological stains using deep convolutional features. In: Cardoso, M.J., et al. (eds.) DLMIA/ML-CDS -2017. LNCS, vol. 10553, pp. 135–142. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67558-9_16
4. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 2961–2969 (2017)
5. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
6. He, T., Zhang, Z., Zhang, H., Zhang, Z., Xie, J., Li, M.: Bag of tricks for image classification with convolutional neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 558–567 (2019)

7. Liu, Y., et al.: Detecting cancer metastases on gigapixel pathology images. arXiv preprint arXiv:1703.02442 (2017)

8. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. In: International Conference on Learning Representations (2019)

9. Luo, C., Song, C., Zhang, Z.: Generalizing person re-identification by camera-aware invariance learning and cross-domain mixup. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12360, pp. 224–241. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58555-6_14

10. Macenko, M., et al.: 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 (2009)

11. Mao, X., Ma, Y., Yang, Z., Chen, Y., Li, Q.: Virtual mixup training for unsupervised domain adaptation. arXiv preprint arXiv:1905.04215 (2019)

12. Nadeem, S., Hollmann, T., Tannenbaum, A.: Multimarginal Wasserstein barycenter for stain normalization and augmentation. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12265, pp. 362–371. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59722-1_35

13. Reinhard, E., Adhikhmin, M., Gooch, B., Shirley, P.: Color transfer between images. IEEE Comput. Gr. Appl. **21**(5), 34–41 (2001)

14. Ruifrok, A.C., Johnston, D.A., et al.: Quantification of histochemical staining by color deconvolution. Anal. Quant. Cytol. Histol. **23**(4), 291–299 (2001)

15. Russakovsky, O., et al.: ImageNet large scale visual recognition challenge. Int. J. Comput. Vis. **115**(3), 211–252 (2015). https://doi.org/10.1007/s11263-015-0816-y

16. Shaban, M.T., Baur, C., Navab, N., Albarqouni, S.: Staingan: stain style transfer for digital histological images. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (Isbi 2019), pp. 953–956. IEEE (2019)

17. Tellez, D., et al.: 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 (2018)

18. Vahadane, A., et al.: Structure-preserving color normalization and sparse stain separation for histological images. IEEE Trans. Med. Imaging **35**(8), 1962–1971 (2016)

19. Veta, M., Van Diest, P.J., Jiwa, M., Al-Janabi, S., Pluim, J.P.: Mitosis counting in breast cancer: object-level interobserver agreement and comparison to an automatic method. PloS one **11**(8), e0161286 (2016)

20. Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D.: Mixup: beyond empirical risk minimization. In: International Conference on Learning Representations (2018)

21. Zhou, N., Cai, D., Han, X., Yao, J.: Enhanced cycle-consistent generative adversarial network for color normalization of h&e stained images. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11764, pp. 694–702. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32239-7_77