DARC: Distribution-Aware Re-Coloring Model for Generalizable Nucleus Segmentation

Shengcong Chen[1], Changxing Ding[1,2(✉)], Dacheng Tao[3], and Hao Chen[4]

[1] School of Electronic and Information Engineering, South China University of Technology, Guangzhou 510000, China
c.shengcong@mail.scut.edu.cn, chxding@scut.edu.cn
[2] Pazhou Lab, Guangzhou 510330, China
[3] School of Computer Science, The Faculty of Engineering, The University of Sydney, Darlington, NSW 2008, Australia
[4] Department of Computer Science and Engineering and Department of Chemical and Biological Engineering, The Hong Kong University of Science and Technology, Hong Kong, China

Abstract. Nucleus segmentation is usually the first step in pathological image analysis tasks. Generalizable nucleus segmentation refers to the problem of training a segmentation model that is robust to domain gaps between the source and target domains. The domain gaps are usually believed to be caused by the varied image acquisition conditions, e.g., different scanners, tissues, or staining protocols. In this paper, we argue that domain gaps can also be caused by different foreground (nucleus)-background ratios, as this ratio significantly affects feature statistics that are critical to normalization layers. We propose a Distribution-Aware Re-Coloring (DARC) model that handles the above challenges from two perspectives. First, we introduce a re-coloring method that relieves dramatic image color variations between different domains. Second, we propose a new instance normalization method that is robust to the variation in foreground-background ratios. We evaluate the proposed methods on two H&E stained image datasets, named CoNSeP and CPM17, and two IHC stained image datasets, called DeepLIIF and BC-DeepLIIF. Extensive experimental results justify the effectiveness of our proposed DARC model. Codes are available at https://github.com/csccsccsccsc/DARC.

Keywords: Domain Generalization · Nucleus Segmentation · Instance Normalization

1 Introduction

Automatic nucleus segmentation has captured wide research interests in recent years due to its importance in pathological image analysis [1–4]. However, as
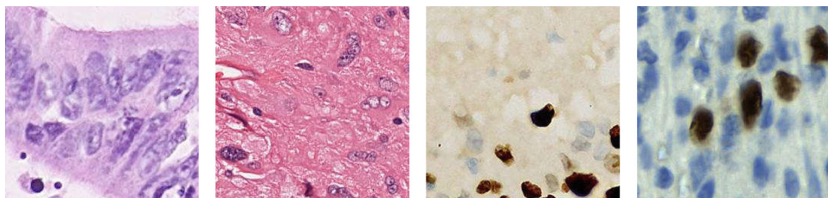
Fig. 1. Example image patches from different datasets. Their appearance differs significantly from each other due to variations in image modalities, staining protocols, scanner types, and tissues.

shown in Fig. 1, the variations in image modalities, staining protocols, scanner types, and tissues significantly affect the appearance of nucleus images, resulting in notable gap between source and target domains [5–7]. If a number of target domain samples are available before testing, one can adopt domain adaptation algorithms to transfer the knowledge learned from the source domain to the target domain [8–10]. Unfortunately, in real-world applications, it is usually expensive and time-consuming to collect new training sets for the ever changing target domains; moreover, extra computational cost is required, which is usually unrealistic for the end users. Therefore, it is highly desirable to train a robust nucleus segmentation model that is generalizable to different domains.

In recent years, the research on domain generalization (DG) has attracted wide attention. Most existing DG works are proposed for classification tasks [12,13] and they can be roughly grouped into data augmentation-, representation learning-, and optimization-based methods. The first category of methods [14–17] focus on the way to diversify training data styles and expect the enriched styles cover those appeared in target domains. The second category of methods aim to obtain domain-invariant features. This is usually achieved via improving model architectures [18–20] or introducing novel regularization terms [21,22]. The third category of methods [23–26] develop new model optimization strategies, e.g., meta-learning, that improve model robustness via artificially introducing domain shifts during training.

It is a consensus that a generalizable nucleus segmentation model should be robust to image appearance variation caused by the change in staining protocols, scanner types, and tissues, as illustrated in Fig. 1. In this paper, we argue that it is also desirable to be robust to the ratio between foreground (nucleus) and background pixel numbers. This ratio changes the statistics of each feature map channel, and affects the robustness of normalization layers, e.g., instance normalization (IN). We will empirically justify its impact in Sect. 2.3.

2   Method

2.1   Overview

In this paper, we adopt a U-Net-based model similar to that in [1] as the baseline. It performs both semantic segmentation and contour detection for nucleus
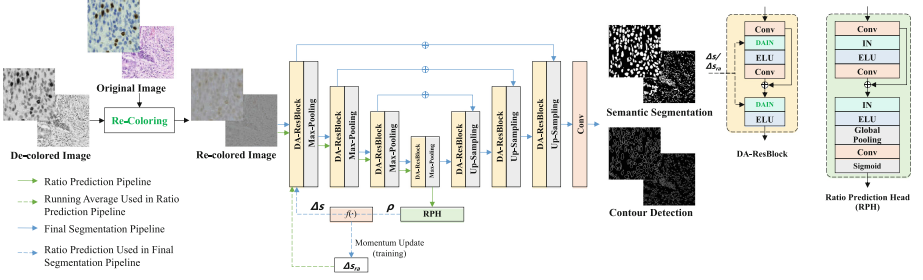
Fig. 2. Overview of the DARC model. The whole model is trained in an end-to-end manner. DARC first re-colors each image to relieve the impact caused by different image acquisition conditions. The re-colored image is then fed into the U-Net encoder and the ratio prediction head. This head predicts the foreground-background ratio ρ. Then, the re-colored image is fed into DARC again with ρ for final prediction. For simplicity, we only illustrate the data-flow of the first DA-ResBlock in details.

instances. The area of each nucleus instance is obtained via subtraction between the segmentation and contour prediction maps [1]. Details of the baseline model is provided in the supplementary material. To handle domain variations, we adopt IN rather than batch normalization (BN) in the U-Net model.

Our proposed Distribution-Aware Re-Coloring model (DARC) is illustrated in Fig. 2. Compared with the baseline, DARC replaces the IN layers with the proposed Distribution-Aware Instance Normalization (DAIN) layers. DARC first re-colors each image to relieve the influence caused by image acquisition conditions. The re-colored image is then fed into the U-Net encoder and the ratio prediction head. This head predicts the ratio between foreground and background pixel numbers. With the predicted ratio, the DAIN layers can estimate feature statistics more robustly and facilitate more accurate nucleus segmentation.

2.2 Nucleus Image Re-Coloring

We propose the Re-Coloring (RC) method to overcome the color change in different domains. Specifically, given a RGB image I, e.g., an H&E or IHC stained image, we first obtain its grayscale image I_g. We then feed I_g into a simple module T that consists of a single residual block and a 1×1 convolutional layer with output channel number of 3. In this way, we obtain an initial re-colored image I_r.

However, de-colorization results in the loss of fine-grained textures and may harm the segmentation accuracy. To handle this problem, we compensate I_r with the original semantic information contained in I. Recent works [40] show that semantic information can be reflected via the order of pixels according to their gray value. Therefore, we adopt the Sort-Matching algorithm [41] to combine the semantic information in I with the color values in I_r. Details of RC is presented in Algorithm 1, in which *Sort* and *ArgSort* denote channel-wisely sorting the values and obtaining the sorted values and indices respectively, and

---

Algorithm 1. Re-Coloring

Input:

The input RGB image $I \in R^{H \times W \times 3}$;

The module T whose input and output channel numbers are 1 and 3, respectively;

Output:

The re-colored image $I_o \in R^{H \times W \times 3}$;

1: De-colorizing I to obtain I_g;
2: $I_r \leftarrow T(I_g)$
3: Reshaping I and I_r to $R^{HW \times 3}$
4: $\boldsymbol{SortIndex} \leftarrow ArgSort(ArgSort(I))$
5: $\boldsymbol{SortValue} \leftarrow Sort(I_r)$
6: $I_o \leftarrow AssignValue(\boldsymbol{SortIndex}, \boldsymbol{SortValue})$
7: **return** I_o

Table 1. Evaluation on the impact of foreground-background ratio to model performance. Both training and testing samples are obtained from CPM17. B denotes the background expansion factor, which directly affects the foreground-background ratio.

B	1	2	4	6
AJI	65.11	59.11	53.41	54.13
Dice	86.14	84.05	80.97	79.75

$AssignValue$ denotes re-assembling the sorted values according to the provided indices. Details of the module T are included in the supplementary material.

Via RC, the original fine-grained structure information from I_g is recovered in I_r. In this way, the re-colored image is advantageous in two aspects. First, the appearance difference between pathological images caused by the change in scanners and staining protocols is eliminated. Second, the re-colored image preserves fine-grained structure information, enabling precise instance segmentation to be possible.

2.3   Distribution-Aware Instance Normalization

Due to dramatic domain gaps, feature statistics may differ significantly between domains [5–7], which means that feature statistics obtained from the source domain may not apply to the target domain. Therefore, existing DG works usually replace BN with IN for feature normalization [12,19]. However, for dense-prediction tasks like semantic segmentation or contour detection, adopting IN alone cannot fully address the feature statistics variation problem. This is because feature statistics are also relevant to the ratio between foreground and background pixel numbers. Specifically, an image with more nucleus instances produces more responses in feature maps and thus higher feature statistic values, and vice versa. The difference in this ratio causes interference to nucleus segmentation.

Algorithm 2. Distribution-Aware Instance Normalization

Input:

   Original feature maps $\boldsymbol{X} \in R^{H \times W \times C}$. The C-dimensional feature vector on its pixel (i, j) is denoted as $\boldsymbol{x}_{ij}$;

   The modules $E_{\boldsymbol{\mu}}$ and $E_{\boldsymbol{\delta}}$ that re-estimate feature statistics;

   $\boldsymbol{\Delta s}_{ra} \in R^{1 \times 1 \times C}$ that is obtained via running mean of $\boldsymbol{\Delta s}$ in the training stage;

   The momentum factor α used to update $\boldsymbol{\Delta s}_{ra}$;

   (Optional) $\boldsymbol{\Delta s} = f(\boldsymbol{\rho})$;

Output:

   Normalized feature maps $\boldsymbol{Y} \in R^{H \times W \times C}$;

1: $\boldsymbol{\mu} \leftarrow \frac{1}{HW} \sum_{i=1}^{H} \sum_{j=1}^{W} \boldsymbol{x}_{ij}$

2: $\boldsymbol{\delta^2} \leftarrow \frac{1}{HW} \sum_{i=1}^{H} \sum_{j=1}^{W} (\boldsymbol{x}_{ij} - \boldsymbol{\mu})^2$

3: **if** $\boldsymbol{\Delta s}$ is given **then**
4:    // Using $\boldsymbol{\Delta s}$ to re-estimate feature statistics for final segmentation
5:    $\boldsymbol{\mu'}, \boldsymbol{\delta'} \leftarrow E_{\boldsymbol{\mu}}(\boldsymbol{\mu}, \boldsymbol{\delta}, \boldsymbol{\Delta s}), E_{\boldsymbol{\delta}}(\boldsymbol{\mu}, \boldsymbol{\delta}, \boldsymbol{\Delta s})$
6:    **if** $Training$ **then**
7:       // Updating $\boldsymbol{\Delta s}_{ra}$ during training
8:       $\boldsymbol{\Delta s}_{ra} \leftarrow (1 - \alpha)\boldsymbol{\Delta s}_{ra} + \alpha\boldsymbol{\Delta s}$
9:    **end if**
10: **else**
11:    // Using $\boldsymbol{\Delta s}_{ra}$ to re-estimate feature statistics for ratio prediction
12:    $\boldsymbol{\mu'}, \boldsymbol{\delta'} \leftarrow E_{\boldsymbol{\mu}}(\boldsymbol{\mu}, \boldsymbol{\delta}, \boldsymbol{\Delta s}_{ra}), E_{\boldsymbol{\delta}}(\boldsymbol{\mu}, \boldsymbol{\delta}, \boldsymbol{\Delta s}_{ra})$
13: **end if**
14: $\boldsymbol{Y} \leftarrow (\boldsymbol{X} - \boldsymbol{\mu'})/\boldsymbol{\delta'}$
15: **return** $\boldsymbol{Y}$

To verify the above viewpoint, we evaluate the baseline model under different foreground-background ratios. Specifically, we first remove the foreground pixels via in-painting [27], and then pad the original testing images with the obtained background patches. We adopt B to denote the ratio between the size of the obtained new image and the original image size. Compared with the original images, the new images have the same foreground regions but more background pixels, and thus have different foreground-background ratios. Finally, we evaluate the performance of the baseline model with different B values. Experimental results are presented in Table 1. It is shown that the value of B affects the model performance significantly.

The above problem is common in nucleus segmentation because pathological images from different organs or tissues tend to have significantly different foreground-background ratios. However, this phenomenon is often ignored in existing research. To handle this problem, we propose the Distribution-Aware Instance Normalization (DAIN) method to re-estimate feature statistics that account for different ratios of foreground and background pixels. Details of DAIN is presented in Algorithm 2. The structures of $E_{\boldsymbol{\mu}}$ and $E_{\boldsymbol{\delta}}$ are included in the supplemental materials.

As shown in Fig. 2, to obtain the foreground-background ratio $\boldsymbol{\rho}$ of one input image, we first feed it to the model encoder with $\boldsymbol{\Delta s_{ra}}$ as the additional input. $\boldsymbol{\Delta s_{ra}}$ acts as pseudo residuals of feature statistics and is obtained in the training stage via averaging $\boldsymbol{\Delta s}$ in a momentum fashion. The output features by the encoder are used to predict the foreground-background ratio $\boldsymbol{\rho}$ with a Ratio-Prediction Head (RPH). $\boldsymbol{\rho}$ is then utilized to estimate the residuals of feature statistics: $\boldsymbol{\Delta s} = f(\boldsymbol{\rho})$. Here, f is a 1×1 convolutional layer that transforms $\boldsymbol{\rho}$ to a feature vector whose dimension is the same as the target layer's channel number. After that, the input image is fed into the model again with $\boldsymbol{\Delta s}$ as additional input and finally makes more accurate predictions.

The training of RPH requires an extra loss term L_{rph}, which is formulated as bellow:

$$L_{rph} = L_{BCE}(\boldsymbol{\rho}, \boldsymbol{\rho_g}) + L_{MSE}(f(\boldsymbol{\rho}), f(\boldsymbol{\rho_g})), \tag{1}$$

where $\boldsymbol{\rho_g}$ denotes the ground truth foreground-background ratio, and L_{BCE} and L_{MSE} denote the binary cross entropy loss and the mean squared error, respectively.

3   Experiments

3.1   Datasets

The proposed method is evaluated on four datasets, including two H&E stained image datasets CoNSeP [3] and CPM17 [28] and two IHC stained datasets DeepLIIF [29] and BC-DeepLIIF [29,32]. **CoNSeP** [3] contains 28 training and 14 validation images, whose sizes are 1000×1000 pixels. The images are extracted from 16 colorectal adenocarcinoma WSIs, each of which belongs to an individual patient, and scanned with an Omnyx VL120 scanner within the department of pathology at University Hospitals Coventry and Warwickshire, UK. **CPM17** [28] contains 32 training and 32 validation images, whose sizes are 500×500 pixels. The images are selected from a set of Glioblastoma Multiforme, Lower Grade Glioma, Head and Neck Squamous Cell Carcinoma, and non-small cell lung cancer whole slide tissue images. **DeepLIIF** [29] contains 575 training and 91 validation images, whose sizes are 512×512 pixels. The images are extracted from the slides of lung and bladder tissues. **BC-DeepLIIF** [29,32] contains 385 training and 66 validation Ki67 stained images of breast carcinoma, whose sizes are 512×512 pixels.

3.2   Implementation Details

In the training stage, patches of size 224×224 pixels are randomly cropped from the original samples. During training, the batch size is 4 and the total number of training iterations is 40,000. We use Adam algorithm for optimization, and the learning rate is initialized as $1e^{-3}$, which is gradually decreased to $1e^{-5}$ during training. We adopt the standard augmentation, like image color jittering and Gaussian blurring. In all experiments, the segmentation and contour detection predictions are penalized using the binary cross entropy loss.

Table 2. Comparisons in generalization performance on nucleus segmentation datasets. Results in each column are related to models trained on one domain and evaluated on the other three unseen domains. Methods marked by * are proposed in this paper. Results are in percentages.

Methods	CoNSeP		CPM17		DeepLIIF		BC-DeepLIIF		Average	
	AJI	Dice	AJI	Dice	AJI	Dice	AJI	Dice	AJI	Dice
Baseline (BN)	16.67	24.10	33.30	61.18	08.42	38.17	21.27	39.92	19.92	40.84
Baseline (IN)	32.13	48.67	33.94	65.83	41.48	67.17	21.52	37.49	32.27	54.79
BIN [19]	21.54	34.33	37.06	67.63	23.51	49.49	26.15	44.42	27.01	48.97
DSU [20]	21.42	34.66	39.12	66.55	27.21	55.10	25.09	41.83	28.21	49.53
SAN [35]	27.91	46.72	33.69	65.66	27.57	53.09	22.17	38.38	27.84	50.96
AmpNorm [36,37]	35.52	55.89	33.39	58.69	39.91	66.58	23.79	37.81	33.15	54.74
StainNorm [38]	41.06	60.81	32.75	64.68	38.55	63.95	25.41	43.81	34.44	58.11
StainMix [39]	34.22	51.07	35.05	65.49	38.48	64.92	26.88	45.62	33.66	56.78
TENT (BN) [34]	38.61	58.11	35.04	64.62	33.77	59.76	23.55	40.91	32.74	55.85
TENT (IN) [34]	32.34	48.87	33.24	65.73	42.08	66.87	22.38	38.04	32.51	54.88
EFDMix [40]	40.13	58.74	33.29	65.25	39.06	64.60	25.92	42.38	34.60	57.74
RC (IN)*	37.21	57.53	36.98	67.71	35.53	62.03	24.98	42.25	33.68	57.38
DAIN*	33.86	50.08	30.62	64.64	37.93	65.56	31.20	53.15	33.40	58.87
DAIN w/o Ratio*	27.37	40.35	33.25	65.05	40.21	66.82	29.16	48.30	32.50	55.13
$DARC_{all}$*	38.18	57.27	34.44	66.11	39.10	67.07	31.64	53.81	35.84	61.06
$DARC_{enc}$*	40.04	58.73	35.60	66.50	40.11	68.23	32.56	53.86	**37.08**	**61.83**

Table 3. Complexity comparison between the baseline model and DARC.

Models	#Parameters (M)	Inference Time (s/image)
Baseline (IN)	5.03	0.0164
$DARC_{enc}$	5.47	0.0253

3.3 Experimental Results and Analyses

In this paper, the models are compared using the AJI [33] and Dice scores. In the experiments, models trained on one of the datasets will be evaluated on the three unseen ones. To avoid the influence of the different sample numbers of the datasets, we calculate the average scores within each unseen domain respectively and then average them across domains.

In this paper, we re-implement some existing popular domain generalization algorithms for comparisons under the same training conditions. Specifically, we re-implement the TENT [34], BIN [19], DSU [20], Frequency Amplitude Normalization (AmpNorm) [36,37], SAN [35] and EFDMix [40]. We also evaluate the stain normalization [38] and stain mix-up [39] methods that are popular in pathological image analysis. Their performances are presented in Table 2. $DARC_{all}$ replaces all normalization layers with DAIN, while $DARC_{enc}$ replaces

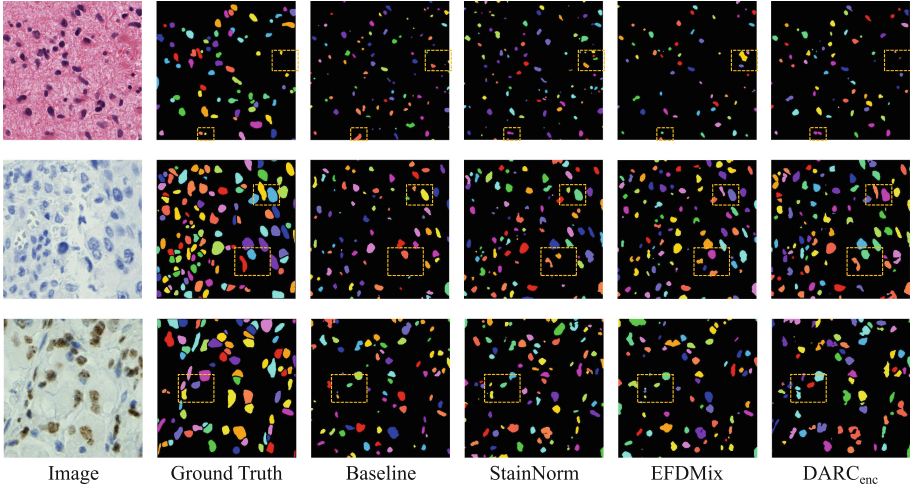|					
Image	Ground Truth	Baseline	StainNorm	EFDMix	DARC$_{enc}$

Fig. 3. Qualitative comparisons between Different Models.

the normalization layers in the encoder with DAIN and uses BN in its decoder. As shown in Table 2, DARC$_{enc}$ achieves the best average performance among all methods. Specifically, DARC$_{enc}$ improves the baseline model's average AJI and Dice scores by 4.81% and 7.04%. Compared with the other domain generalization methods, DAIN, DAIN w/o Ratio, DARC$_{all}$ and DARC$_{enc}$ achieve impressive performances on BC-DeepLIIF, which justify that re-estimating the instance-wise statistics is important for improving the domain generalization ability of models trained on BC-DeepLIIF. Qualitative comparisons are presented in Fig. 3. Moreover, the complexity analysis between the baseline model and DARC$_{enc}$ is presented in Table 3.

We separately evaluate the effectiveness of RC and DAIN, and present the results in Table 2. Also, we train a variant model without foreground-background ratio prediction, which is denoted as 'DAIN w/o Ratio' in Table 2. Compared with the baseline model, RC improves the average AJI and Dice scores by 1.41% and 2.59%, and DAIN improves the average AJI and Dice scores by 1.13% and 4.08%. Compared with the variant model without foreground-background ratio prediction, DAIN improves the average AJI and Dice scores by 0.90% and 3.74%. Finally, the combinations of RC and DAIN, i.e., DARC$_{all}$ and DARC$_{enc}$, achieve the best average scores. As shown in Table 2, DARC$_{enc}$ improves DARC$_{all}$ by 1.24% and 0.77% on AJI and Dice scores respectively. This is because after the operations by RC and DAIN in the encoder, the obtained feature maps are much more robust to the domain gaps, which enables the decoder to adopt the fixed statistics maintained during training. Moreover, using the fixed statistics is helpful to prevent the decoder from the influence of varied foreground-background ratios on feature statistics.

4   Conclusion

In this paper, we propose the DARC model for generalizable nucleus segmentation. To handle the domain gaps caused by varied image acquisition conditions, DARC first re-colors the input image while preserving its fine-grained structures as much as possible. Moreover, we find that the performance of instance normalization is sensitive to the varied ratios in foreground and background pixel numbers. This problem is well addressed by our proposed DAIN. Compared with existing works, DARC achieves significantly better performance on average across four benchmarks.

References

1. Chen, H., Qi, X., Yu, L., Dou, Q., Qin, J., Heng, P.-A.: DCAN: deep contour-aware networks for object instance segmentation from histology images. Med. Image Anal. **36**, 135–146 (2017)
2. Zhou, Y., Onder, O.F., Dou, Q., Tsougenis, E., Chen, H., Heng, P.A.: Cia-net: robust nuclei instance segmentation with contour-aware information aggregation. In: IPMI, pp. 682–693 (2019)
3. Graham, S., et al.: Hover-Net: simultaneous segmentation and classification of nuclei in multi-tissue histology images. Med. Image Anal. **58**, 101563 (2019)
4. Schmidt, U., Weigert, M., Broaddus, C., Myers, G.: Cell detection with star-convex polygons. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11071, pp. 265–273. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00934-2_30
5. Ben Taieb, A., Hamarneh, G.: Adversarial stain transfer for histopathology image analysis. IEEE Trans. Med. Imaging **37**(3), 792–802 (2018)
6. Stacke, K., Eilertsen, G., Unger, J., Lundström, C.: Measuring domain shift for deep learning in histopathology. IEEE J. Biomed. Health Inf. **25**(2), 325–336 (2021)
7. Aubreville, M., et al.: Mitosis domain generalization in histopathology images - the MIDOG challenge. Med. Image Anal. **84**, 102699 (2023)
8. Li, C., et al.: Domain adaptive nuclei instance segmentation and classification via category-aware feature alignment and pseudo-labelling. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. LNCS, vol. 13437, pp. 715–724. Springer, Heidelberg (2022). https://doi.org/10.1007/978-3-031-16449-1_68
9. Liu, D., et al.: PDAM: a panoptic-level feature alignment framework for unsupervised domain adaptive instance segmentation in microscopy images. IEEE Trans. Med. Imaging **40**(1), 154–165 (2021)
10. Yang, S., Zhang, J., Huang, J., Lovell, B. C., Han, X.: Minimizing labeling cost for nuclei instance segmentation and classification with cross-domain images and weak labels. In: AAAI, vol. 35, no. 1, pp. 697–705 (2021)
11. Gulrajani I., Lopez-Paz D.: In search of lost domain generalization. In: ICLR, (2021)

12. Zhou, K., et al.: Domain generalization: a survey. IEEE Trans. Pattern Anal. Mach. Intell. **45**, 4396–4415 (2022)
13. Wang, J., et al.: Generalizing to unseen domains: a survey on domain generalization. IEEE Trans. Knowl. Data Eng. **35**, 8052–8072 (2022)
14. Huang, J., et al.: FSDR: frequency space domain randomization for domain generalization. In: CVPR, pp. 6891–6902 (2021)
15. Shu, Y., et al.: Open domain generalization with domain-augmented meta-learning. In: CVPR, pp. 9624–9633 (2021)
16. Zhou, Z., Qi, L., Shi, Y.: Generalizable medical image segmentation via random amplitude mixup and domain-specific image restoration. In: Avidan, S., Brostow, G., Cisse, M., Farinella, G.M., Hassner, T. (eds.) ECCV, vol. 13681, pp. 420–436. Springer, Heidelberg (2022). https://doi.org/10.1007/978-3-031-19803-8_25
17. Liu, Q., Chen, C., Qin, J., Dou, Q., Heng, P.-A.: FedDG: federated domain generalization on medical image segmentation via episodic learning in continuous frequency space. In: CVPR, pp. 1013–1023 (2021)
18. Jin, X., et al.: Style normalization and restitution for generalizable person re-identification. In: CVPR, pp. 3143–3152 (2020)
19. Nam, H., Kim, H.-E.: Batch-instance normalization for adaptively style-invariant neural networks. In: NeurIPS, pp. 2563–2572 (2018)
20. Li, X., Dai, Y., Ge, Y., Liu, J., Shan, Y., Duan, L.: Uncertainty modeling for out-of-distribution generalization. In ICLR (2022)
21. Li, H., Pan, S.J., Wang, S., Kot, A.C.: Domain generalization with adversarial feature learning. In: CVPR, pp. 5400–5409 (2018)
22. Tian, C.X., Li, H., Xie, X., Liu, Y., Wang, S.: Neuron coverage guided domain generalization. IEEE Trans. Pattern Anal. Mach. Intell. **45**(1), 1302–1311 (2023)
23. Robey, A., Pappas, G.J., Hassani, H.: Model-based domain generalization. In: NeurIPS, pp. 20210–20229 (2021)
24. Qiao, F., Zhao, L., Peng, X.: Learning to learn single domain generalization. In: CVPR, pp. 12556–12565 (2020)
25. Wang, Z., Luo, Y., Qiu, R., Huang, Z., Baktashmotlagh, M.: Learning to diversify for single domain generalization. In: ICCV, pp. 834–843 (2021)
26. Shi, Y., et al.: Gradient matching for domain generalization. In: ICLR (2022)
27. Telea, A.: An image inpainting technique based on the fast marching method. J. Graph. Tools **9**(1), 23–24 (2004)
28. Vu, Q.D., et al.: Methods for segmentation and classification of digital microscopy tissue images. Front. Bioeng. Biotechnol. **53** (2019)
29. Ghahremani, P., Marino, J., Dodds, R., Nadeem, S.: DeepLIIF: an online platform for quantification of clinical pathology slides. In CVPR, pp. 21399–21405 (2022)
30. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
31. Llewellyn, B.D.: Nuclear staining with alum hematoxylin. Biotech. Histochem. **84**(4), 159–177 (2009)
32. Huang, Z., et al.: BCData: a large-scale dataset and benchmark for cell detection and counting. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12265, pp. 289–298. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59722-1_28
33. Kumar, N., Verma, R., Sharma, S., Bhargava, S., Vahadane, A., Sethi, A.: A dataset and a technique for generalizable nuclear segmentation for computational pathology. IEEE Trans. Med. Imaging **36**(7), 1550–1560 (2017)

34. Wang, D., Shelhamer, E., Liu, S., Olshausen, B., Darrell, T.: Tent: fully test-time adaptation by entropy minimization. In: ICLR (2021)
35. Peng, D., Lei, Y., Hayat, M., Guo, Y., Li, W.: Semantic-aware domain generalizable segmentation. In CVPR, pp. 2594–2605 (2022)
36. Jiang, M., Wang, Z., Dou, Q.: Harmofl: harmonizing local and global drifts in federated learning on heterogeneous medical images. In: AAAI, vol. 36, no. 1, pp. 1087–1095 (2022)
37. Zhao1, X., Liu, C., Sicilia, A., Hwang, S. J., Fu, Y.: Test-time fourier style calibration for domain generalization. In: IJCAI (2022)
38. Macenko, M., et al.: A method for normalizing histology slides for quantitative analysis. In ISBI, pp. 1107–1110 (2009)
39. Chang, J.-R., et al.: Stain mix-up: unsupervised domain generalization for histopathology images. In: de Bruijne, M., de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12903, pp. 117–126. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87199-4_11
40. Zhang, Y., Li, M., Li, R., Jia, K., Zhang L.: Exact feature distribution matching for arbitrary style transfer and domain generalization. In: CVPR, pp. 8035–8045 (2022)
41. Rolland, J.P., Vo, V., Bloss, B., Abbey, C.K.: Fast algorithms for histogram matching: application to texture synthesis. J. Electron. Imaging **9**(1), 39–45 (2000)