High-Resolution Medical Image Translation via Patch Alignment-Based Bidirectional Contrastive Learning

Wei Zhang[1], Tik Ho Hui[2], Pui Ying Tse[3], Fraser Hill[3], Condon Lau[2], and Xinyue Li[1(✉)]

[1] School of Data Science, City University of Hong Kong, Hong Kong SAR, China
xinyueli@cityu.edu.hk
[2] Department of Physics, City University of Hong Kong, Hong Kong SAR, China
[3] Veterinary Diagnostic Laboratory Medical Centre, City University of Hong Kong, Hong Kong SAR, China

Abstract. Pathology image assessment plays a crucial role in disease diagnosis and treatment. In this study, we propose a Patch alignment-based Paired medical image-to-image Translation (PPT) model that takes the Hematoxylin and Eosin (H&E) stained image as input and generates the corresponding Immunohistochemistry (IHC) stained image in seconds, which can bypass the laborious and time-consuming procedures of IHC staining and facilitate timely and accurate pathology assessment. First, our proposed PPT model introduces FocalNCE loss in patch-wise bidirectional contrastive learning to ensure high consistency between input and output images. Second, we propose a novel patch alignment loss to address the commonly observed misalignment issue in paired medical image datasets. Third, we incorporate content and frequency loss to produce IHC stained images with finer details. Extensive experiments show that our method outperforms state-of-the-art methods, demonstrates clinical utility in pathology expert evaluation using our dataset and achieves competitive performance in two public breast cancer datasets. Lastly, we release our H&E to IHC image Translation (HIT) dataset of canine lymphoma with paired H&E-CD3 and H&E-PAX5 images, which is the first paired pathological image dataset with a high resolution of 2048 × 2048. Our code and dataset are available at https://github.com/coffeeNtv/PPT.

Keywords: High-Resolution Medical Image Translation · Virtual Staining · Generative Adversarial Network

1 Introduction

Image-to-image translation is a technique that learns the mapping from the source image domain to the target image domain. It gains increasing attention

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-72083-3_17.

M. G. Linguraru et al. (Eds.): MICCAI 2024, LNCS 15004, pp. 178–188, 2024.
https://doi.org/10.1007/978-3-031-72083-3_17

in medical domains recently, especially in pathology. Pathological diagnosis plays a critical role in diagnosing and treating many diseases. Pathologists use various techniques to examine samples for more diagnostic information to identify cellular changes associated with diseases. Specifically, Hematoxylin and Eosin (H&E) staining is commonly used in pathological assessments since it is fast and cheap, but it has limited information for disease diagnosis. To obtain more accurate disease diagnosis and treatment, immunohistochemistry (IHC) staining is often required. IHC staining provides more diagnostic information by detecting specific biomarkers associated with diseases, but it is time-consuming, taking hours and even days to complete, and often expensive due to cumbersome procedures.

Virtual staining technique aims to transform H&E stained images into the corresponding IHC stained images. It can bypass the time-consuming and costly IHC staining procedures and provide doctors with virtual IHC images in seconds for timely and accurate disease diagnosis and treatment. However, this technique faces many challenges. First, medical and natural images have different natures. Natural images have rich semantic information at instance level, such as category, location, size, shape, color and texture. In comparison, medical images, such as pathological images, typically contain detailed information at the cellular level, such as clear cell structure and morphology, nuclear size and shape. Second, collecting medical imaging data poses specific difficulties. Medical images typically require specialized imaging equipment for collection and professional pathologists for annotation, making collection and annotation laborious, expensive, and time-consuming. Moreover, medical images involve privacy and ethical issues, and many valuable datasets cannot be publicly available. Therefore, the existing datasets are often limited. Third, medical images are usually high-resolution in order to provide detailed information at the cellular level, posing computational challenges when applying deep learning models. Complex models built for natural image tasks, such as models involving attention mechanism may be ineffective and computationally intensive to be applied to high-resolution medical images.

To address the challenges in virtual staining, several research studies have been conducted recently. Rivenson *et al.* [17] applied the generative adversarial network (GAN) [2] with a U-Net [18] structure to transform autofluorescence images into various histological stained images. Their results demonstrate the feasibility of virtual staining with GAN and have since inspired many researchers in this field. Pyramid Pix2pix [12] was proposed to generate IHC images from H&E images using a low-pass filter and Gaussian convolutions at multiple scales. Li *et al.* [8] introduced an adaptive supervised PatchNCE (ASP) loss into H&E-to-IHC image translation to address the inconsistency between input and label images. Meanwhile, virtual staining with unpaired images has also been investigated when paired data are unavailable. Liu *et al.* [13] introduced the pathology-consistent constraint in CycleGAN [25] to preserve microscopic structural details and pathology properties from H&E stained images. Lin *et al.* [11] proposed style-guided normalization and multiple-style encoding methods to capture the relationships among different stains for multi-domain transfer.

Existing approaches often utilized GAN-based models to learn the pathological properties of stained images and used images at a low resolution for training,

such as 256 × 256 or 512 × 512 [11,15,17], to minimize computational complexity, while virtual staining on high-resolution images still remains to be fully investigated. It is also noteworthy that existing studies rarely explored the imperfect registration of paired images. Even though perfectly aligned paired images are desirable for supervised learning, acquiring such images in clinical settings is impractical, as paired images are often obtained from adjacent tissue sections to minimize anatomical differences, and additional algorithms are required for image pair registration. Therefore, the misalignment at the pixel level is inevitable in paired images. Based on these observations, we release a paired pathological image dataset with high resolution regarding the data scarcity issue for public use, and we propose a novel image translation model named Patch alignment-based Paired medical image Translation (PPT) with bidirectional contrastive learning to alleviate the misalignment problem in paired images and generate more consistent and fine-grained virtual IHC images to effectively achieve IHC image translation for timely and accurate diagnosis.

Our contributions are summarized as: 1)We build and publish the H&E to IHC image Translation (HIT) dataset of canine lymphoma with two sets of 2048 × 2048 high-resolution paired images: H&E-CD3 and H&E-PAX5. To the best of our knowledge, this is the first paired high-resolution pathological image dataset of lymphoma. 2)We introduce the FocalNCE loss in bidirectional contrastive learning to preserve more consistent results from H&E stained images and learn a better feature representation of IHC stained images. 3)We propose a novel patch alignment loss to supervise output images at the patch level, which mitigates the negative impact of pixel mismatch on imperfectly registered pathological images. 4)We incorporate content loss and frequency loss to regulate IHC stained images at the feature and frequency levels to further improve the staining quality of IHC stained images. 5)Extensive experiments are conducted using our dataset and two public breast cancer datasets, and results show that our method outperforms state-of-the-art methods on our dataset and has competitive performance in public datasets with different IHC staining techniques.

2 Methodology

Given the input $x \in X$ from the H&E staining image domain $X \subset R^{H \times W \times C}$ and the ground truth $y \in Y$ from IHC staining domain $Y \subset R^{H \times W \times C}$, we aim to preserve the content in the domain X while keeping the staining style in the domain Y. The framework of our proposed model is demonstrated in Fig. 1.

Adversarial Learning. We apply the GAN framework [2] to generate realistic IHC image $\hat{y} = G(x)$ from H&E image x. The adversarial loss $\mathcal{L}_{\text{adv}}$ is defined as follows: $\mathcal{L}_{\text{adv}}(X, Y) = E_{y \sim Y}[\log D(y)] - E_{x \sim X}[\log(1 - D(G(x))]$.

Bidirectional Contrastive Learning. Prior to presenting our bidirectional contrastive learning, we first introduce contrastive learning, which is a framework based on noise contrastive estimation [3] implemented by the InfoNCE loss [14] to maximize the mutual information [14] within a dataset for

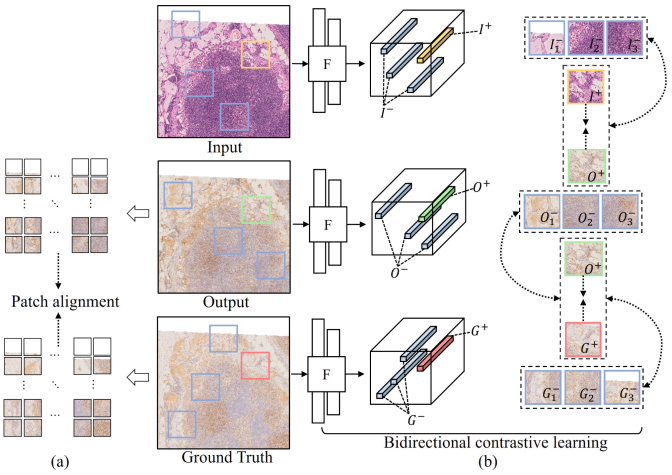

Fig. 1. Framework of proposed method.(a)Align images at patch level rather than pixel level. (b)Associate positive patch feature while push negative ones away bidirectionally.

better feature representation. The PatchNCE loss [16] applied InfoNCE loss in a patch manner within images where associated patches are brought together while disassociated patches are pushed away. Let $x_p, \hat{y}_p \subset R^{H_{\text{patch}} \times W_{\text{patch}} \times C}$ be the patches from the input x and the output $\hat{y}$ at the same spatial location. Then the patch feature representations of $\hat{y}_p$ and x_p are $v = F(\hat{y}_p)$ and $v^+ = F(x_p)$, where F is a shared encoder that projects the input patch into a K-dimensional space. $v^- \subset R^{K \times N}$ denotes the feature vector of N negative patches in x that have different spatial location from x_p. We can obtain the probability of the selected positive example over the other N negative examples: $Prob(v, v^+, v^-) = \frac{\exp(v \cdot v^+/\tau)}{\exp(v \cdot v^+/\tau) + \sum_{n=1}^{N} \exp(v \cdot v_n^-/\tau)}$, where τ is the parameter of temperature, v_n^- is the n_{th} negative patch. Moreover, x and $\hat{y}$ are extracted in a multi-layer setting in F with L layers in S_l spatial locations. The PatchNCE loss is applied using cross-entropy loss for $N+1$ patches where one patch is a positive sample and the other N patches are negative samples. This may result in a class imbalance problem, but focal loss is effective for addressing this problem [10,20]. Therefore, we introduce focal loss into our contrastive learning, where α and γ are balance and modulating factors.

$$\ell_f\left(v, v^+, v^-\right) = -\alpha\left(1 - Prob\left(v, v^+, v^-\right)\right)^{\gamma} \log Prob\left(v, v^+, v^-\right) \tag{1}$$

We denoted the PatchNCE loss combined with focal loss as $\mathcal{L}_{FocalNCE}$.

$$\mathcal{L}_{FocalNCE}(X, \hat{Y}) = E_{(x,\hat{y})\sim(X,\hat{Y})} \sum_{i=1}^{L} \sum_{s=1}^{S_l} \ell_f\left(v_{is}, v_{is}^+, v_i^-\right) \tag{2}$$

where v_{is} and v_{is}^+ are patch features extracted from the same spatial location in $\hat{y}$ and x, respectively. v_i^- is the feature of N negative patches in x that

have different spatial location from x_{is}. Inspired by previous work [1,16], we use contrastive learning between the input and the output images to force the model to generate images with more consistent details from the input. Further, we apply the FocalNCE loss between the output and the ground truth images in a bidirectional manner to learn a better feature representation of target images. Therefore, our bidirectional contrastive loss is constructed as follows:

$$\mathcal{L}_{\text{cont}}\ (X, Y, \hat{Y}) = \mathcal{L}_{FocalNCE}(X, \hat{Y}) + \mathcal{L}_{FocalNCE}(Y, \hat{Y}) + \mathcal{L}_{FocalNCE}(\hat{Y}, Y) \quad (3)$$

Consistency Learning. Paired pathological images are usually collected from nearby tissue slices to avoid anatomic differences [23] or from the same tissue by the de-stain and re-stain method [5]. Thus, perfectly aligned images may not always be available. We focus on the pixel mismatch problem in paired images and try to alleviate this problem at the patch, feature, and frequency levels. We proposed patch alignment loss, allowing for imperfect alignment within the patches in images. We first use a sliding window on the output and ground truth images without overlap and then flatten them into vectors. Then we compute the difference at the patch level instead of pixel level on the y-axis of patch vectors. The patch alignment loss is: $\ell_{\text{patch}} = \beta \sum_{i=1}^{I} \mathbb{E}_{(y,\hat{y})\sim(Y,\hat{Y})} [\|p_i(y) - p_i(\hat{y})\|_1]$, where I is the total number of all patches, p_i denotes the vector of the i_{th} patch in the corresponding image and β is the weight of loss.

Inspired by perceptual loss [7], we consider the consistency at feature level as our content loss ℓ_{content} with a pre-trained VGG-19 network [19]. Further, we also adopt a Gaussian reconstruction loss [8,9,12] to preserve consistency at frequency levels as frequency loss ℓ_{freq} . Hence, the objective of our consistency learning is: $\mathcal{L}_{\text{cons}}\ (Y, \hat{Y}) = \ell_{\text{patch}} + \ell_{\text{content}} + \ell_{\text{freq}}$.

Full Objective. The objective of PPT model is the combination of the adversarial, contrastive, and consistency loss: $\mathcal{L}_{\text{total}}\ (X, Y, \hat{Y}) = \mathcal{L}_{\text{adv}} + \mathcal{L}_{\text{cont}} + \mathcal{L}_{\text{cons}}$, where X, Y, $\hat{Y}$ are input, ground truth, and output images, respectively.

3 Experiments

Datasets. (1) Our HIT dataset includes paired H&E-CD3 stained images and paired H&E-PAX5 stained images of canine lymphoma with a resolution of 2048 × 2048, dubbed CD3 dataset and PAX5 dataset. We utilize 1652 image pairs for training and 155 image pairs for testing in CD3 dataset and 1614 image pairs for training and 163 image pairs for testing in PAX5 dataset. (2) BCI dataset [12] was collected from breast cancer tissues with a resolution of 1024 × 1024. It consists of 3896 H&E-HER2 pairs for training and 977 pairs for testing. (3) MIST dataset [8] contains four paired H&E-IHC datasets for breast cancer diagnosis, denoted as HER2, ER, Ki67 and PR dataset. Each subset includes 4000∼5000 pairs for training and 1000 pairs testing with a resolution of 1024 × 1024.

Implementation Settings. We use 9 residual blocks in generator and adopt PatchGAN discriminator [6]. The patch size in patch alignment loss is 4×4. Our PPT models are trained with 1024×1024 resolution images for 300 epochs, where the initial learning rate is set as 0.0002, and linear decay is applied after 100 epochs. We use Adam optimizer with $\beta_1 = 0.5$ and $\beta_2 = 0.999$ and set batch size as 1. Our experiments are conducted on NVIDIA RTX A6000 GPU.

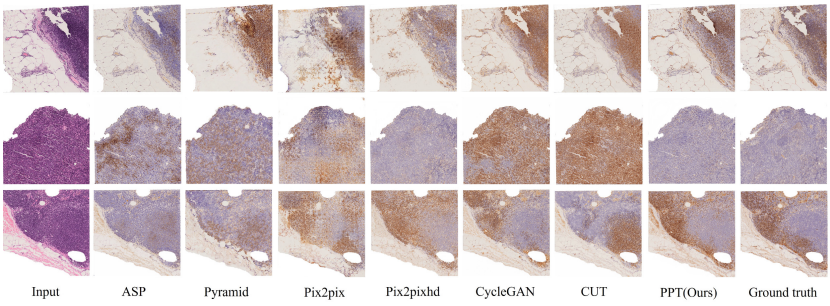

Fig. 2. Comparison between our model and benchmark models on PAX5 dataset.

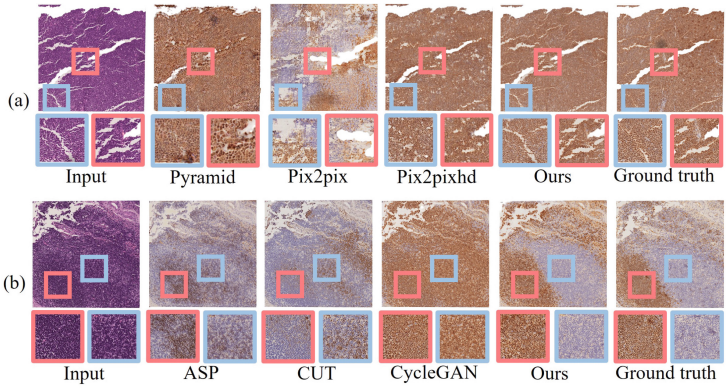

Fig. 3. Comparison using PAX5 dataset on (a)image details (b)cell type prediction.

Evaluation Metrics. We apply Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) [22] to compare the similarity at the pixel level. Further, we use Fréchet Inception Distance (FID) [4] to measure the feature distance between two image distributions and use Learned Perceptual Image Patch Similarity (LPIPS) [24] to evaluate the perceptual similarity between two images that are perceptible to the human visual system.

Qualitative Evaluation. We demonstrate our results in Figs. 2 and 3. It can be observed that ASP [8], CycleGAN [25], and CUT [16] preserve most of details

Table 1. Quantitative comparisons among benchmarks on PAX5 dataset, where the **best** and the second best results are highlighted.

Model	SSIM↑	PSNR ↑	FID ↓	LPIPS ↓
ASP	0.3531	16.8270	109.2708	0.3700
Pyramid	**0.3695**	17.3896	121.2043	0.4019
Pix2pix	0.3240	15.7357	160.3297	0.4124
Pix2pixhd	0.3484	17.3660	87.2724	0.3153
CycleGAN	0.3507	16.6412	84.7979	0.3259
CUT	0.3534	16.7610	92.2355	0.3312
PPT(Ours)	0.3548	**17.4063**	**80.7147**	**0.2988**

from H&E stained images. Still, they have poor performance in predicting cell types appearing in brown and purple colors, which are crucial for disease diagnosis. The Pyramid [12] and Pix2pixhd [21] models produce relatively accurate PAX5-stained cell types compared with other models. Nevertheless, they fail to keep as many details in H&E images as other models. The Pix2pix [6] model produces images with checkerboard artifacts and inaccurate staining prediction. Compared with these models, our model generates images with the best quality, preserving fine-grained details and predicting accurate cell types.

Table 2. Comparisons on CD3, BCI, and HER2 datasets with benchmarks, where the **best** and the second best results are highlighted.

Dataset	Model	SSIM↑	PSNR↑	FID↓	LPIPS↓
CD3	ASP	0.3958	17.0627	92.9461	0.3202
	Pyramid	**0.4127**	17.4546	119.1605	0.3778
	PPT(Ours)	0.4074	**17.7863**	**78.5448**	**0.2913**
BCI	ASP	0.3932	17.0835	199.6659	0.6087
	Pyramid	0.456	**19.2723**	84.1875	0.5105
	PPT(Ours)	**0.4968**	19.0914	**52.522**	**0.5058**
HER2	ASP	**0.2048**	14.5789	49.433	0.5500
	Pyramid	0.1718	13.7607	130.1539	0.5583
	PPT(Ours)	0.1981	**14.5811**	**49.4197**	**0.5323**

Quantitative Evaluation. We compare our model with benchmarks over multiple metrics in Table 1. Our approach has the best performance among these methods on PSNR, FID, and LPIPS, indicating that our model generates images with the best quality, which are most similar to the ground truth distribution and are closer to human visual perception than others. Pyramid seems to have higher SSIM evaluation, which may be because the Pyramid model tends to generate images with similar colors while ignoring many details, as demonstrated in

Figs. 2 and 3(a), which may lead to better performance in SSIM. Moreover, we apply our method to other virtual staining scenarios, such as CD3 staining of lymphoma and HER2 IHC of breast cancer, to further evaluate the performance of our model. Note that CD3 dataset has the same trend as PAX5 dataset on other baselines, and the ASP and Pyramid are the latest SOTAs, therefore, we only compare these two SOTAs in Table 2. As shown in Table 2, our method achieved the best performance in PSNR, FID, and LPIPS compared with the state-of-the-art methods on BCI and HER2 datasets (more results in supplementary materials). Moreover, it can be observed in Fig. 4 that our results are more consistent with the input images and have closer staining style to the label images compared to benchmarks.

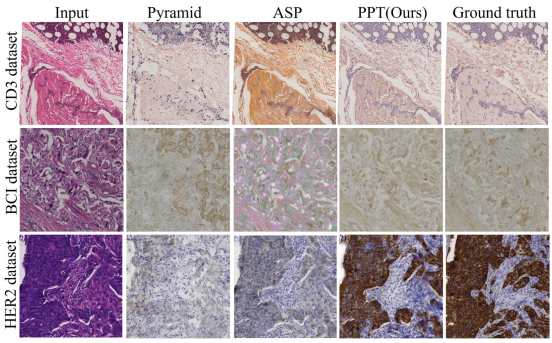

Fig. 4. Comparisons with benchmarks on different datasets.

Ablation Study. We performed an ablation study to investigate the effectiveness of the frequency loss, patch alignment loss, and content loss. The baseline is our bidirectional contrastive learning with FocalNCE loss, denoted as *base*. We conduct ablation experiments by incrementally adding the ℓ_{freq}, ℓ_{patch} and ℓ_{content} on the PAX5 dataset. The results are demonstrated in Table 3 which further indicate the feasibility of our proposed model.

Table 3. Results of the ablation study on PAX5 dataset.

Model	SSIM↑	PSNR↑	FID↓	LPIPS↓
base	0.3494	16.3711	90.9563	0.3461
base + ℓ_{freq}	0.3515	17.1360	85.9513	0.3194
base + ℓ_{freq} + ℓ_{patch}	0.3529	17.2724	82.7983	0.3095
base + ℓ_{freq} + ℓ_{patch} + ℓ_{content}	**0.3548**	**17.4063**	**80.7147**	**0.2988**

Expert Evaluation. To validate the effectiveness of our proposed model in the clinical setting, we invited two registered veterinary pathologists to evaluate our virtually stained images on the PAX5 and CD3 test sets. Pathologists were asked to rate the similarity between the real and the generated images on a scale of 1-5 for five diagnostic indicators for lymphoma, including staining intensity, cellular localization, cellular distribution, quantification, and morphological correlation. The overall clinical accuracy is the average of the five indicators in percentage. For pathologist 1, the accuracy for CD3 and PAX5 dataset is 76.23% and 68.91% respectively, and the accuracy for pathologist 2 is 66.14% and 60.32% respectively, which has 25.86% higher accuracy than existing work [12]. Expert evaluation results demonstrate the clinical utility of our proposed model in real-world clinical applications for diagnosing lymphoma using both CD3 and PAX5 stained images. We can also see there is room for further improvement. Details of the expert evaluation are provided in supplementary materials.

4 Conclusion

This study released a paired high-resolution pathological image dataset HIT of canine lymphoma and proposed a Patch alignment-based Paired medical image-to-image Translation model named PPT to transform the H&E stained images into corresponding IHC stained images for timely and accurate diagnosis. Specifically, we propose a patch alignment loss to address the pixel mismatch problem in paired pathological images and incorporate FocalNCE loss bidirectionally in contrastive learning with content and frequency loss to generate more consistent and fine-grained IHC stained images. Extensive experiments indicate our method outperforms benchmarks in our dataset and can be applied to different types of IHC staining scenarios with competitive performance. Moreover, two pathologists were invited to further validate our effectiveness in virtual IHC staining from H&E to CD3 and PAX5 images in the clinical setting. Future work will further improve the virtual staining model to enhance clinical utility and promote the application in clinical settings.

Acknowledgments. This work is supported by the Hong Kong Innovation and Technology Commission (InnoHK Project CIMDA)(9610473), City University of Hong Kong Internal Funding (7005892, 9678264) and University Grants Committee Funding (6460003).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Andonian, A., Park, T., Russell, B., Isola, P., Zhu, J.Y., Zhang, R.: Contrastive feature loss for image prediction. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 1934–1943 (2021)

2. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. Advances in neural information processing systems **27** (2014)
3. Gutmann, M., Hyvärinen, A.: Noise-contrastive estimation: A new estimation principle for unnormalized statistical models. In: Proceedings of the thirteenth international conference on artificial intelligence and statistics. pp. 297–304. JMLR Workshop and Conference Proceedings (2010)
4. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: Gans trained by a two time-scale update rule converge to a local nash equilibrium. Advances in neural information processing systems **30** (2017)
5. Ikenberg, K., Pfaltz, M., Rakozy, C., Kempf, W.: Immunohistochemical dual staining as an adjunct in assessment of mitotic activity in melanoma. Journal of cutaneous pathology **39**(3), 324–330 (2012)
6. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1125–1134 (2017)
7. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14. pp. 694–711. Springer (2016)
8. Li, F., Hu, Z., Chen, W., Kak, A.: Adaptive supervised patchnce loss for learning h&e-to-ihc stain translation with inconsistent groundtruth image pairs. In: Medical Image Computing and Computer Assisted Intervention(MICCAI). pp. 632–641 (2023)
9. Liang, J., Zeng, H., Zhang, L.: High-resolution photorealistic image translation in real-time: A laplacian pyramid translation network. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9392–9400 (2021)
10. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE international conference on computer vision. pp. 2980–2988 (2017)
11. Lin, Y., Zeng, B., Wang, Y., Chen, Y., Fang, Z., Zhang, J., Ji, X., Wang, H., Zhang, Y.: Unpaired multi-domain stain transfer for kidney histopathological images. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 36, pp. 1630–1637 (2022)
12. Liu, S., Zhu, C., Xu, F., Jia, X., Shi, Z., Jin, M.: Bci: Breast cancer immunohistochemical image generation through pyramid pix2pix. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1815–1824 (2022)
13. Liu, S., Zhang, B., Liu, Y., Han, A., Shi, H., Guan, T., He, Y.: Unpaired stain transfer using pathology-consistent constrained generative adversarial networks. IEEE Transactions on Medical Imaging **40**(8), 1977–1989 (2021)
14. Oord, A.v.d., Li, Y., Vinyals, O.: Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748 (2018)
15. Ozyoruk, K.B., Can, S., Darbaz, B., Başak, K., Demir, D., Gokceler, G.I., Serin, G., Hacisalihoglu, U.P., Kurtuluş, E., Lu, M.Y., et al.: A deep-learning model for transforming the style of tissue images from cryosectioned to formalin-fixed and paraffin-embedded. Nature Biomedical Engineering **6**(12), 1407–1419 (2022)

16. Park, T., Efros, A.A., Zhang, R., Zhu, J.Y.: Contrastive learning for unpaired image-to-image translation. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX 16. pp. 319–345. Springer (2020)
17. Rivenson, Y., Wang, H., Wei, Z., de Haan, K., Zhang, Y., Wu, Y., Günaydın, H., Zuckerman, J.E., Chong, T., Sisk, A.E., et al.: Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning. Nature biomedical engineering **3**(6), 466–477 (2019)
18. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. pp. 234–241. Springer (2015)
19. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
20. Spiegl, B.: Contrastive unpaired translation using focal loss for patch classification. arXiv preprint arXiv:2109.12431 (2021)
21. Wang, T.C., Liu, M.Y., Zhu, J.Y., Tao, A., Kautz, J., Catanzaro, B.: High-resolution image synthesis and semantic manipulation with conditional gans. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 8798–8807 (2018)
22. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing **13**(4), 600–612 (2004)
23. Zhang, R., Cao, Y., Li, Y., Liu, Z., Wang, J., He, J., Zhang, C., Sui, X., Zhang, P., Cui, L., et al.: Mvfstain: multiple virtual functional stain histopathology images generation based on specific domain mapping. Medical Image Analysis **80**, 102520 (2022)
24. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 586–595 (2018)
25. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision. pp. 2223–2232 (2017)