Diffusion-based Virtual Staining from Polarimetric Mueller Matrix Imaging

Xiaoyu ZHENG[1], Jing Wen[2], Jiaxin Zhuang[1], Yao Du[1], Jing Cong[3], Limei Guo[4], Lin Luo[2] ✉, Chao He[5] ✉, and Hao Chen[1]

[1] Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Hong Kong SAR, China
[2] College of Engineering, Peking University, Beijing, China
luol@pku.edu.cn
[3] Beijing Institute of Collaborative Innovation, Beijing, China
[4] Department of Pathology, School of Basic Medical Science, Peking University Health Science Center, Peking University Third Hospital, Beijing, China
[5] Department of Engineering Science, University of Oxford, Oxford X1 3PJ, UK
chao.he@eng.ox.ac.uk

Abstract. Polarization, as a new optical imaging tool, has been explored to assist in the diagnosis of pathology. Moreover, converting the polarimetric Mueller Matrix (MM) to standardized stained images becomes a promising approach to help pathologists interpret the results. However, existing methods for polarization-based virtual staining are still in the early stage, and the diffusion-based model, which has shown great potential in enhancing the fidelity of the generated images, has not been studied yet. In this paper, a Regulated Bridge Diffusion Model (RBDM) for polarization-based virtual staining is proposed. RBDM utilizes the bidirectional bridge diffusion process to learn the mapping from polarization images to other modalities such as H&E and fluorescence. And to demonstrate the effectiveness of our model, we conduct the experiment on our manually collected dataset, which consists of 18,000 paired polarization, fluorescence and H&E images, due to the unavailability of the public dataset. The experiment results show that our model greatly outperforms other benchmark methods. Our data and code are available at https://github.com/xiaoyu-z/RBDM/.

Keywords: Virtual Staining · Polarization · Image-to-image Translation.

1 Introduction

Polarization imaging technology enables non-invasive acquisition of birefringence characteristics and surface microstructure information at sub-resolution scales of tissue structures by analyzing the modulation properties of biological tissues on the polarization state of incident light waves (such as Stokes vectors, Mueller matrices, and other parameters) [1]. This unique imaging mechanism imparts high sensitivity in detecting pathological features such as collagen fiber alignment and

cell membrane integrity, demonstrating significant potential for clinical diagnosis applications in recent years, including tumor boundary delineation and tissue fibrosis evaluation[2, 3]. However, due to the human visual system's lack of direct perception of polarization states, the rich physical information contained in polarization images needs to be converted into grayscale or pseudo-color images through complex post-processing algorithms for visualization. This secondary encoding process results in significant differences between the imaging results and the cognitive paradigms of pathologists, severely limiting the interpretability and acceptability of polarization imaging technology in clinical diagnosis [4].

In traditional pathological practice, fluorescence imaging and Hematoxylin and Eosin (H&E) staining techniques have become the gold standard due to their clear biological relevance. Fluorescence labeling enables functional imaging at the molecular level through specific antibody-antigen reactions [5], while H&E staining provides intuitive histomorphological information through differential staining of cell nuclei and cytoplasm [6]. The emergence of virtual staining technology has offered novel solutions to improve the efficiency of computational pathology. This technology establishes a non-linear mapping relationship between unstained tissue images and target staining modalities through deep learning models, achieving significant advances in the field of computational pathology in recent years [7]. Mainstream methods primarily rely on Generative Adversarial Networks (GANs) and their variants, realizing cross-modal image translation via Cycle-Consistency constraints [8–10]. Recent research has further explored the application of Diffusion Models in virtual staining, leveraging their progressive denoising mechanism to enhance the morphological fidelity of staining boundaries [11]. However, diffusion methods for polarization data has not been investigated in previous works.

In this study, we integrate the Mueller matrix full-polarization imaging system with a diffusion-based virtual staining framework to establish an image-to-image translation model from raw polarization features to standard Hematoxylin and Eosin (H&E) as well as fluorescence images, namely the Regulated Bridge Diffusion Model (RBDM). By bridging polarization data to diagnostically familiar stains, our method unlocks a staining-free, low-cost, high-fidelity paradigm for precision pathology. The contributions are summarized in the following aspects:

(1) First Public Multi-modality Polarization Pathology Dataset: We construct and release a benchmark dataset that contains 18,000 strictly registered image patches with polarization, H&E, and fluorescence modalities.

(2) First Diffusion Model from Polarization to Other Modalities: We design a diffusion model to achieve stable mapping from 16-channel Mueller matrix imaging to the stained image domain via a regulated diffusion process.

(3) State-of-the-art Performance Validation: Our model significantly outperforms other benchmark methods on our dataset in both polarization-to-fluorescence and polarization-to-H&E tasks.

2   Related Work

2.1   Image-to-image Translation

Image-to-image Translation (I2I) is an important research direction in the field of computer vision, which aims to transform images between different domains while maintaining the semantic consistency of the content. Recent advances in deep learning, particularly GANs and diffusion models, have significantly propelled progress in this field. Isola et al. proposed Pix2Pix, a Conditional-GAN (cGAN) framework for paired image translation, achieving success in tasks such as semantic segmentation and colorization [12]. Wang et al. extended this with Pix2PixHD, improving resolution and detail through a multi-scale architecture [13]. To realize structural fidelity, PyramidPix2pix [14] introduced a pyramid network architecture that progressively refines translations on multiple scales, effectively preserving hierarchical features in high-resolution outputs.

Despite their success, GAN-based methods often suffer from training instability and mode collapse, limiting their ability to generate high-quality complex images. Recent advances in diffusion models have addressed these limitations. Li et al. introduced BBDM [15], modeling I2I as a stochastic Brownian bridge process to directly learn domain transformations without iterative refinement. Xia et al. developed DiffI2I [16], which introduces a compact I2I prior extraction network and dynamic I2I transformer to guide denoising, enabling fewer iterations and lighter models for high-quality translation.

2.2   Polarization Imaging in Virtual Staining

Polarization imaging has emerged as a powerful tool in biomedical imaging due to its sensitivity to tissue micro-structural and anisotropic properties [17]. Recent advancements in polarization imaging have shown its potential in virtual staining [18]. Si et al. [19] proposed a deep learning-based approach using a cGAN structure to translate MM images into bright-field microscopy equivalents, addressing low throughput issues while preserving structural insights. Fan et al. [20] further simplified data acquisition by introducing an unsupervised CycleGAN framework trained on unpaired MM and H&E-stained images, eliminating the need for spatial registration.

These studies highlight the potential of polarization imaging in virtual staining, providing new auxiliary indicators for pathological diagnosis, and improving the value of multi-modality data through deep learning. But in those studies, there is a lack of model comparisons and Principal Component Analysis (PCA) is relied on for dimension reduction, which discards critical MM information.

3   Multi-modality Polarization Pathology Dataset

We present the Multi-modality Polarization Pathology Dataset (MPPD), the first breast cancer pathology dataset integrating Mueller matrix polarimetric

imaging, H&E staining, and fluorescence staining. The MPPD comprises 7 breast cancer patient samples, consisting of 18,000 pairs of 600×600 patches with precisely spatially aligned samples. This multi-modality framework enables the comprehensive characterization of pathological samples in structural, molecular, and physical domains.

3.1 Data Collection

To achieve high-efficiency Mueller matrix imaging, we developed an adaptive full-vector imaging system that integrates hardware control, image acquisition, and data analysis modules. The acquisition process begins with the use of specialized slide scanners to obtain Whole-Slide Images (WSIs) of H&E staining and fluorescence staining. Subsequently, the samples are placed on a stage for polarimetric imaging. After autofocus determines the optimal focal plane, the system scans the sample region-by-region with a preset step size, recording the Mueller matrix data for each region. Given that samples are large and a single shot cannot cover the entire slice, an ITKMontage [21] module is employed for multi-modality image registration and stitching. To address issues such as size consistency and missing data in multi-modality image stitching, the system automatically generates configuration files to record tile position information. It also utilizes boundary detection and virtual tile generation techniques. Additionally, grayscale uniformity correction and brightness normalization are applied. Ultimately, this process results in the formation of complete panoramic WSI images under a 4× objective lens.

3.2 Data Processing

We designed a multi-stage registration pipeline to achieve precise spatial alignment for pixel-level correlation across modalities.Initial pre-processing standardizes the image scales and removes extraneous blank regions to mitigate feature interference. The polarization-derived gray modality is designated as the reference framework due to its consistent correspondence with other modalities. Feature
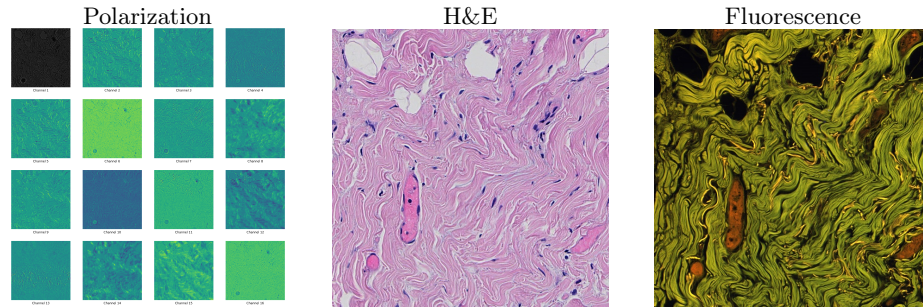

Fig. 1: Polarization-H&E-Fluorescence Dataset Visualization

matching is performed using the SuperPoint algorithm [22], a deep learning-based detector that identifies cross-modal keypoints for robust correspondence estimation. Rigid registration optimizes global alignment via affine transformations. Subsequent non-rigid registration refines local deformations using B-spline modeling [23] combined with curvature regularization [24] to prevent artifacts. After registration, all modalities are resampled to 2304×1296 pixels and aligned with the polarization field-of-view (FOV). Then we applied adaptive thresholding to exclude tissue-free or artifact-dominated regions. The validated FOVs are partitioned into 600×600 pixel patches using a sliding window approach, preserving multi-modality alignment through strict spatial correspondence. This process produced 18,000 high-quality patches for downstream analysis.

4 Methods

We extend the Brownian Bridge Diffusion Models (BBDM) [15] to the polarization domain with additional regulations. Fig. 2 shows the general structure of our Regulated Bridge Diffusion Model (RBDM).
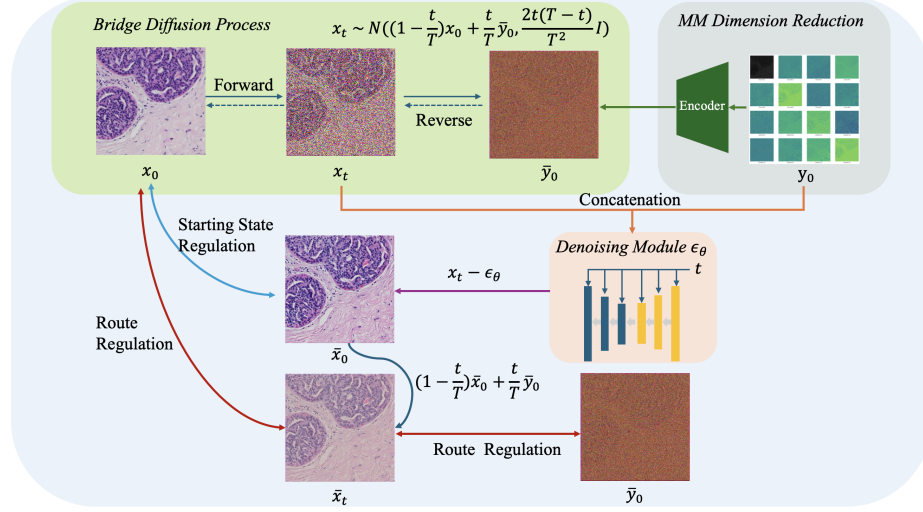

Fig. 2: Overall Framework of Regulated Bridge Diffusion Model

4.1 Bridge Diffusion Model

Given an image x_0 from domain A and image y_0 from domain B, the Bridge Diffusion Model aims to perform successful image-to-image translation from y_0 to x_0. During training, translation will go through the forward and reverse processes. In the forward process, the state distribution at each time step t can be formulated as

$$q(x_t|x_0, y_0) = N(x_t; (1 - \frac{t}{T})x_0 + \frac{t}{T}y_0, \sigma_t I) \tag{1}$$

Here T is the total number of time steps, σ_t stands for the variance, which is $\frac{2t(T-t)}{T^2}$, and it is easy to observe that at staring state and ending state, the σ_t is 0 and the mean value equals to x_0 and y_0 respectively, which makes it a fixed mapping process. And the intermediate state x_t can therefore be represented as

$$x_t = (1 - \frac{t}{T})x_0 + \frac{t}{T}y_0 + \sqrt{\sigma_t}\epsilon \tag{2}$$

where $\epsilon \sim N(0, I)$.

The reverse process starts with y_0 and aims to predict x_{t-1} given x_t. It can be formulated as

$$p_\theta(x_{t-1}|x_t, y_0) = N(x_{t-1}; \mu_\theta(x_t, t), \tilde{\sigma}_t I) \tag{3}$$

where $\tilde{\sigma}_t$ is the variance of the noise and $\mu_\theta(x_t, t)$ is the predicted mean value of the noise.

The training objective of the bridge diffusion model is to optimize the Evidence Lower Bound (ELBO) and is proven equivalent to training a network ϵ_θ to predict noise, which can be formulated as

$$L_{BDM} = \left\| \frac{t}{T}(y_0 - x_0) + \sqrt{\sigma_t}\epsilon - \epsilon_\theta(x_t, t, y_0) \right\| \tag{4}$$

Once we know the y_0 and noise at each time step, x_0 can be reconstructed.

4.2 Regulated Bridge Diffusion Model for Polarization

One main difference between polarization and other images is about the dimension, the traditional BBDM model cannot handle data with 16-dimensionality, so we design a simple auto-encoder structure E_p for dimension reduction ra. The auto-encoder contains two convolutional layers, the first one expands the dimension from 16 to 64 and the second layer reduces it to 3, a tanh layer is appended to ensure the output scale is between -1 and 1, in this setting, the output of $E_p(y_0)$ can be represented as $\bar{y}_0$, and the first training objective can be written as

$$L_1 = \left\| \frac{t}{T}(\bar{y}_0 - x_0) + \sqrt{\sigma_t}\epsilon - \epsilon_\theta(x_t, t, y_0) \right\| \tag{5}$$

However, since the auto-encoder is trainable, intermediate states generated by the bridge diffusion process may lack consistency during training, leading to suboptimal results. Therefore, we design two other regulation modules to enforce stability across diffusion timesteps to enhance performance, denoted as Starting State Regulation (SSR) and Route Regulation (RR).

Starting State Regulation Given any time step t, x_0 can actually be estimated by the following formula during training

$$\begin{aligned} \bar{x}_0 &= x_t - \epsilon_\theta(x_t, t, y_0) \\ &= (1 - \frac{t}{T})x_0 + \frac{t}{T}\bar{y}_0 + \sqrt{\sigma_t}\epsilon - \epsilon_\theta(x_t, t, y_0) \end{aligned} \tag{6}$$

To ensure the estimated quality, we apply the perceptual loss [25] to narrow the difference between feature maps of x_0 and $\bar{x}_0$ using the pre-trained VGG [26] network. The starting state regulation loss can be expressed as

$$L_2 = \|F(x_0) - F(\bar{x}_0)\| \tag{7}$$

where F stands for feature extraction network.

Route Regulation The route regulation is according to the attribute of the Bridge Diffusion Model. When the time step t is close to T, the intermediate image $\bar{x}_t$ generated should be more similar to y_0 than x_0, and when the time step t is close to 0, the image $\bar{x}_t$ should be more like x_0 without considering noise. MS-SSIM loss [27] is applied to measure the multi-scale structural similarity between the ground truth and the generated image, and the higher the value of the MS-SSIM is, the more similar the two images are. So the route regulation loss is then represented as

$$L_3 = max(0, \frac{T - 2t}{T} \left[M(\bar{x}_t, \bar{y}_0) - M(\bar{x}_t, x_0) \right]) \tag{8}$$

Where M means MS-SSIM value, and in order to train the whole network, we use $\bar{x}_0$ to generate new $\bar{x}_t$. The equation is

$$\bar{x}_t = (1 - \frac{t}{T})\bar{x}_0 + \frac{t}{T}\bar{y}_0 \tag{9}$$

Thus the overall training objective for our model becomes

$$L = L_1 + L_2 + L_3 \tag{10}$$

5   Experiments

5.1   Implementation Detail

We use our original dataset for performance evaluation. 70% patches are selected for training, and the rest are reserved for testing. All patches are resized to
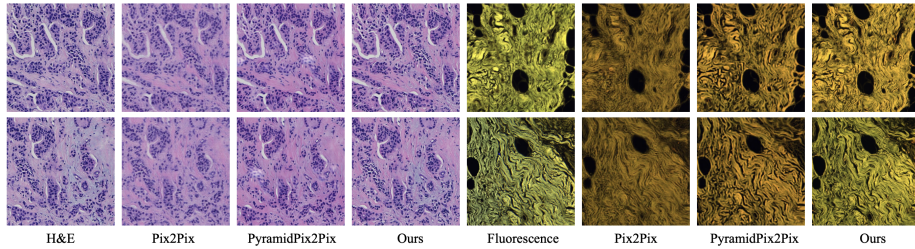

Fig. 3: Visualization Results of Our Model and Other Baselines

Table 1: Model Performance from Polarization to H&E and Fluorescence

Task	Model	PSNR↑	SSIM↑	FID↓	LPIPS↓
Polar->H&E	Pix2Pix	15.42±3.50	0.19±0.10	215.93	0.60
	Pyramid Pix2pix	19.68±2.68	0.41±0.13	62.02	0.38
	Ours	**20.93±3.25**	**0.53±0.18**	**19.67**	**0.26**
Polar->Fluo	Pix2Pix	16.13±1.28	0.26±0.05	116.73	0.62
	Pyramid Pix2pix	18.09±2.69	0.32±0.13	65.79	0.52
	Ours	**18.52±2.32**	**0.44±0.12**	**43.03**	**0.39**

256×256 and normalized to $[-1, 1]$. For training, the total number of epochs is set to 40, and the Adam optimizer with a learning rate of $1e^{-4}$ is used. The batch size is 2 and the total number of time steps is set to 500 with a sample step of 100.

5.2   Result

To make a fair comparison, we select multiple evaluation metrics and benchmark methods. The widely used metrics including Structural Similarity Index Measure (SSIM) [28], Peak Signal-to-Noise Ratio (PSNR) [29], Fréchet Inception Distance (FID) [30] and Learned Perceptual Image Patch Similarity (LPIPS) [31] are calculated to measure the image quality and similarity. The baseline methods are GAN-based model for paired data, including Pix2Pix [12] and Pyramid-Pix2pix [14]. The results of polarization-to-H&E and polarization-to-fluorescence are shown in Table 1. The task of transforming polarization to fluorescence is more difficult as the fluorescence is more complex. The best results are highlighted in bold. And it is obvious that our model significantly outperforms other methods in terms of all the evaluation indicators. We also picked four patches for the visualization of the image generation outcomes, the results are shown in Fig. 3, the results from our model are more accurate and detailed compared to other baseline approaches.

5.3   Ablation Study

To verify the effectiveness of the regulation modules, we also conduct the ablation study by adding the Starting State Regulation (SSR) and the Route Regulation (RR). The result is shown in Table 2, which indicates that the diffusion-based

Table 2: Ablation Result of Starting State Regulation and Route Regulation

Module		Polar->H&E				Polar->Fluo			
SSR	RR	PSNR↑	SSIM↑	FID↓	LPIPS↓	PSNR↑	SSIM↑	FID↓	LPIPS↓
×	×	20.66±3.37	0.51±0.19	21.21	0.27	18.13±2.44	0.42±0.13	55.89	0.42
✓	×	20.71±3.31	0.52±0.20	20.80	0.26	18.34±2.41	0.43±0.13	49.12	0.40
✓	✓	20.93±3.25	0.53±0.18	19.67	0.26	18.52±2.32	0.44±0.12	43.03	0.39

model achieves better performance compared to previous baselines, and both SSR and RR help improve the performance of the model.

6   Conclusion

In this paper, we extend the bridge diffusion model to the polarization domain and propose first public multi-modality polarization dataset named MPPD and a novel diffusion-based model named RBDM, which realizes the mapping from polarimetric Mueller matrix to standardized stained domain. We also conduct experiments on our MPPD datasets and the RBDM model achieves the state-of-the-art performance, which indicates a promising direction for polarization-based virtual staining in the future.

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

References

1. He, C., He, H., Chang, J., Chen, B., Ma, H., Booth, M.J.: Polarisation optics for biomedical and clinical applications: a review. Light: Science & Applications **10**(1), 194 (2021)
2. Layden, D., Ghosh, N., Vitkin, A.: Quantitative polarimetry for tissue characterization and diagnosis. Advanced Biophotonics: Tissue Optical Sectioning **5**, 73–108 (2013)
3. Yao, Y., Wan, J., Zhang, F., Dong, Y., Chen, L., Ma, H.: Correlation of image textures of a polarization feature parameter and the microstructures of liver fibrosis tissues. Journal of Innovative Optical Health Sciences **16**(05), 2241004 (2023)
4. Gao, W.: Parameterizing depolarizing mueller matrices of anisotropic media–a review. Waves in Random and Complex Media pp. 1–37 (2023)
5. Bancroft, J.D., Gamble, M.: Theory and Practice of Histological Techniques, 7th ed. Elsevier (2019)
6. Deng, L., Chen, C., Yu, W., Shao, C., Shen, Z., Wang, Y., He, C., Li, H., Liu, Z., He, H., et al.: Influence of hematoxylin and eosin staining on linear birefringence measurement of fibrous tissue structures in polarization microscopy. Journal of Biomedical Optics **28**(10), 102909–102909 (2023)
7. Bai, B., Yang, X., Li, Y., Zhang, Y., Pillar, N., Ozcan, A.: Deep learning-enabled virtual histological staining of biological samples. Light: Science & Applications **12**(1), 57 (2023)
8. 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)

9. Ma, J., Chen, H.: Efficient supervised pretraining of swin-transformer for virtual staining of microscopy images. IEEE Transactions on Medical Imaging **43**(4), 1388-1399 (2023)

10. Zhang, W., Yang, D., Che, H., Ran, A.R., Cheung, C.Y., Chen, H.: Unpaired optical coherence tomography angiography image super-resolution via frequency-aware inverse-consistency gan. IEEE Journal of Biomedical and Health Informatics (2024)

11. Shen, Y., Ke, J.: Staindiff: Transfer stain styles of histology images with denoising diffusion probabilistic models and self-ensemble. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 549–559. Springer (2023)

12. 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)

13. Wang, T.C., Liu, M.Y., Zhu, J.Y., Tao, A., Kautz, J., Catanzaro, B.: Highresolution image synthesis and semantic manipulation with conditional gans. In: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. pp. 8798–8807 (2018)

14. 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)

15. Li, B., Xue, K., Liu, B., Lai, Y.K.: Bbdm: Image-to-image translation with brownian bridge diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1952–1961 (2023)

16. Xia, B., Zhang, Y., Wang, S., Wang, Y., Wu, X., Tian, Y., Yang, W., Timotfe, R., Van Gool, L.: Diffi2i: efficient diffusion model for image-to-image translation. IEEE Transactions on Pattern Analysis and Machine Intelligence (2024)

17. Guan, C., Zeng, N., He, H.: Review of polarization-based technology for biomedical applications. Journal of Innovative Optical Health Sciences **18**(02), 2430002 (2025)

18. Wang, W., Lim, L.G., Srivastava, S., Bok-Yan So, J., Shabbir, A., Liu, Q.: Investigation on the potential of mueller matrix imaging for digital staining. Journal of Biophotonics **9**(4), 364–375 (2016)

19. Si, L., Li, N., Huang, T., Du, S., Dong, Y., Yao, Y., Ma, H.: Computational image translation from mueller matrix polarimetry to bright-field microscopy. Journal of Biophotonics **15**(3), e202100242 (2022)

20. Fan, J., Zeng, N., He, H., He, C., Liu, S., Ma, H.: Generating bright-field images of stained tissue slices from mueller matrix polarimetric images with cyclegan using unpaired dataset. Journal of Innovative Optical Health Sciences **18**(02), 2343003 (2025)

21. Zukić, D., Jackson, M., Dimiduk, D., Donegan, S., Groeber, M., McCormick, M.: Itkmontage: a software module for image stitching. Integrating Materials and Manufacturing Innovation **10**, 115–124 (2021)

22. DeTone, D., Malisiewicz, T., Rabinovich, A.: Superpoint: Self-supervised interest point detection and description. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 224–236 (2018)

23. Unser, M., Aldroubi, A., Eden, M.: B-spline signal processing. i. theory. IEEE Transactions on Signal Processing **41**(2), 821–833 (1993)

24. Myronenko, A., Song, X., Carreira-Perpinan, M.: Non-rigid point set registration: Coherent point drift. Advances in Neural Information Processing Systems **19** (2006)

25. 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)
26. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
27. Wang, Z., Simoncelli, E.P., Bovik, A.C.: Multiscale structural similarity for image quality assessment. In: the Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003. vol. 2, pp. 1398–1402. IEEE (2003)
28. 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)
29. Hore, A., Ziou, D.: Image quality metrics: Psnr vs. ssim. In: 20th International Conference on Pattern Recognition. pp. 2366–2369. IEEE (2010)
30. 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)
31. 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)