Advancing Stain Transfer for Multi-Biomarkers: A Human Annotation-Free Method Based on Auxiliary Task Supervision

Siyuan Xu, **Haofei Song**, **Yingjiao Deng**, **Jiansheng Wang**, **Yan Wang** and **Qingli Li***
Shanghai Key Laboratory of Multidimensional Information Processing, East China Normal University, Shanghai, China
{syxu, hfsong}@stu.ecnu.edu.cn, ywang@cee.ecnu.edu.cn, qlli@cs.ecnu.edu.cn

Abstract

Histopathological examination primarily relies on hematoxylin and eosin (H&E) and immunohistochemical (IHC) staining. Though IHC provides more crucial molecular information for diagnosis, it is more costly than H&E staining. Stain transfer technology seeks to efficiently generate virtual IHC images from H&E images. While current deep learning-based methods have made progress, they still struggle to maintain pathological and structural consistency across biomarkers without pixel-level aligned reference. To address the problem, we propose an **A**uxiliary **T**ask supervision-based **S**tain **T**ransfer method for multi-biomarkers (ATST-Net), which pioneeringly employs human annotation-free masks as ground truth (GT). ATST-Net ensures pathological consistency, structural preservation and style transfer. It automatically annotates H&E masks in a cost-effective manner by utilizing consecutive IHC sections. Multiple auxiliary tasks provide diverse supervisory information on the location and intensity of biomarker expression, ensuring model accuracy and interpretability. We design a pretrained model-based generator to extract deep feature in H&E images, improving generalization performance. Extensive experiments demonstrate the effectiveness of ATST-Net's components. Compared to existing methods, ATST-Net achieves state-of-the-art (SOTA) accuracy on datasets with multiple biomarkers and intensity levels, while also reflecting high practical value. Code is available at https://github.com/SikangSHU/ATST-Net.

1 Introduction

Cancer remains one of the leading causes of death worldwide, posing a significant threat to human health [Chhikara *et al.*, 2023; Li *et al.*, 2025]. Histopathological examination is the gold standard for cancer diagnosis and treatment.

Hematoxylin and eosin (H&E) staining is widely used in clinical practice to enhance the visualization of tissues and cells. Hematoxylin (Hema) stains nuclei blue or dark purple, while eosin stains the cytoplasm and extracellular matrix pink. Though H&E staining reveals tissue structure and cellular morphology, it cannot differentiate cancerous cells from normal ones due to its lack of specific protein expression. This limitation is addressed by molecular staining techniques like immunohistochemical (IHC) staining, which uses antigen-antibody specificity to visualize protein (positive) expression in tissues and cells. IHC staining typically combines Hema with diaminobenzidine (DAB) chromogen, employing different staining configurations to highlight biomarkers such as ER, PR, Ki67 and HER2, which are crucial for breast cancer analysis. However, IHC examination is more expensive, time-consuming and requires more specialized equipments and techniques, limiting its accessibility and hindering progress in pathological diagnostics. Currently, while no research directly establishes a molecular biological connection between H&E and IHC images, progress in related prediction indirectly suggests a relationship between stains worth exploring via artificial intelligence (AI). For example, [Zeng *et al.*, 2022] achieves semi-supervised H&E-to-PR stain transfer, [Liu *et al.*, 2020] predicts Ki67 positive cells, and [Farahmand *et al.*, 2022] infers HER2 status from H&E images.

We aim to develop a deep learning-based stain transfer method that models the relationship between H&E and IHC staining, as illustrated in Figure 1. It generates virtual IHC

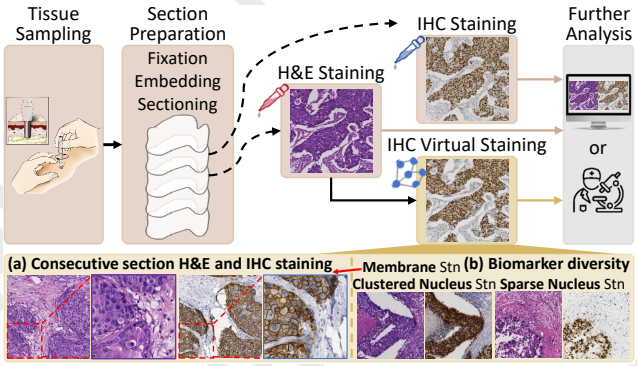

Figure 1: Diagram of stain transfer. After tissue sampling, fixation, embedding and sectioning to obtain sections, we aim to generate virtual IHC images that replace real staining for further analysis. "Stn" in (b) stands for "Staining". Best viewed at a zoomed-in level.

*Corresponding author.

images that align with H&E images in both pathology and structure, enabling immediate diagnosis of multiple biomarkers from the same tissue. Despite significant progress in stain transfer, existing methods still struggle to overcome several key challenges: (a) **Absence of pixel-level aligned ground truth (GT)**. Staining is irreversible in clinical practice, meaning tissue sections stained with one dye cannot be restored to their pre-staining state for re-staining. Pathologists often use consecutive tissue sections (3 to 5μm apart) for different stains. Due to misalignment and staining-induced variability, pixel-level aligned GT is unavailable (Figure 1a). Several studies employ expert annotations to enhance pathological consistency supervision, such as patch-level [Liu *et al.*, 2021; Boyd *et al.*, 2022; Pati *et al.*, 2024] and cell-level [Liu *et al.*, 2020] annotations. However, manual annotations are time-consuming, labor-intensive and impractical for large-scale fine-grained labeling, often resulting in coarse, region-level annotations. (b) **Difficulty in maintaining pathological and structural consistency**. Tissue and biomarker diversity (Figure 1b) complicates the preservation of pathological and structural details during style transfer, including consistency in protein (positive) expression regions, tissue structures and specific expression sites (e.g., clustered nuclei, sparse nuclei, or membrane staining). (c) **Limited practical applicability**. The lack of pixel-level aligned GT for supervision impedes the interpretability of existing methods, limiting their clinical practical applicability.

In this paper, we propose a novel supervised generative stain transfer method that utilizes automatically generated GT masks. Biomarker expression masks, derived from IHC images of consecutive sections, serve as region-level annotations for corresponding H&E images. This approach, achieved through stain unmixing and morphological operations, provides finer-grained annotations than manual labeling in a *cost-effective* way and allows parameter adjustments to account for varying region differences between consecutive sections. To fully leverage region-level alignment, we introduce multiple auxiliary tasks that exploit the DAB and Hema channels, offering supervisory information on positive expression location, intensity and nucleus number. This improves both model *accuracy* and *interpretability*. Furthermore, to enhance the model's ability to extract pathological and structural information across biomarkers, we design a generator architecture incorporating a pretrained encoder trained on large-scale H&E datasets. By robustly capturing stain-invariant feature of various biomarkers from H&E images, this generator boosts model's *generalization ability*.

The main contributions of this paper are as follows:

(1) This paper presents ATST-Net, a generative stain transfer method that converts H&E to IHC images, designed for high interpretability and generalization. It novelly employs non-pixel-aligned IHC-stained consecutive sections as GT, eliminating the need for any manual annotation or prior information, thus markedly reducing annotation costs and errors.

(2) Multiple specialized auxiliary tasks are proposed to fully supervise the generation process, addressing the lack of precise annotations. These tasks include global and local positive expression location matching, intensity matching, and nucleus number matching in positive regions.

(3) We propose a robust stain transfer generator architecture that enhances H&E image interpretation and ensures accurate feature extraction of biomarkers. This is the first work to demonstrate the effectiveness of a specialized large-scale pretrained model for stain transfer.

(4) We conduct experiments with the latest methods on diverse datasets, including the public dataset MIST [Li *et al.*, 2023] with multiple biomarkers, and BCI [Liu *et al.*, 2022], which classifies positive expression intensity. Our method, ATST-Net, achieves SOTA accuracy in terms of pathology, structure and style. It also enables bidirectional transfer.

2 Related Work

Traditional methods mainly focus on color mapping. [Reinhard *et al.*, 2001] proposed a statistical transfer technique adjusting each channel's values. [Macenko *et al.*, 2009] introduced a histology stain normalization method using stain vector determination and deconvolution. These methods only achieve partial color transfer and lack pathological consistency constraints, failing to capture the full stain relationship.

Deep learning-based methods adopt a generative approach. They better capture pathological semantic relationships between stained images, leading to more accurate and reliable transfer. Methods that directly use pixel-level supervision typically rely on Pix2pix [Isola *et al.*, 2017] as the backbone. For instance, [Liu *et al.*, 2022] introduced a pyramid Pix2pix method that supervises feature at multiple scales during image generation. However, calculating absolute errors between generated images and pixel-level unaligned GT distorts the image structure, resulting in inaccurate stain transfer. More recent models are based on CycleGAN [Zhu *et al.*, 2017; Shaban *et al.*, 2019] and CUT [Park *et al.*, 2020], which apply fully unpaired or non-pixel-aligned H&E and IHC image pairs. For example, UMDST [Lin *et al.*, 2022] enables the simultaneous generation of multiple stains from a single stain using unpaired training data. [Li *et al.*, 2023] proposed an Adaptive Supervised PatchNCE loss to address misalignment between the source and target domains. PSPStain [Chen *et al.*, 2024] focuses on improving pathological semantic mining and spatial misalignment. These specifically designed methods significantly improve the performance of transfer. However, annotation-free methods suffer from limited interpretability, while manual annotation is costly, incomplete and subject to bias. [He *et al.*, 2024] proposed PST-Diff, a pilot study applying diffusion models to this task, though its ability to preserve tissue structure requires further validation.

3 Method

The architecture of ATST-Net is depicted in Figure 2 and 3. This section starts with data preprocessing, which maximizes the utility of annotation information in consecutive sections. For limited annotations, we introduce multiple auxiliary tasks to guide biomarker generation. A specialized generator is designed to deeply mine information, ensuring the preservation of pathological and structural feature during stain transfer. Finally, the whole training and inference process is detailed.

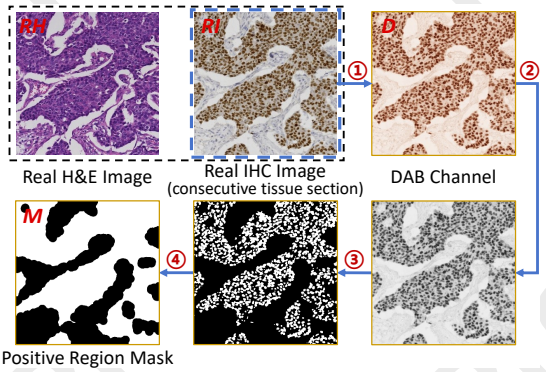

Figure 2: Diagram of data preprocessing. (1) Stain unmixing. (2) Grayscale conversion & Gaussian filtering. (3) Otsu thresholding & Erosion & Dilation. (4) Closing.

3.1 Data Preprocessing

Though pixel-level alignment is unavailable between H&E and corresponding IHC images from consecutive sections, high-quality dataset preparation ensures region-level alignment. As shown in Figure 2, data preprocessing extracts the positive region mask $M \in \mathbb{R}^{H\times W\times C}$ with height H, width W and channel C for each H&E image $RH \in \mathbb{R}^{H\times W\times C}$ from its corresponding real IHC image (real IHC) $RI \in \mathbb{R}^{H\times W\times C}$. ATST-Net then adopts the automated annotated masks to capture biomarker information in H&E images, generating virtual IHC images (fake IHC) $FI \in \mathbb{R}^{H\times W\times C}$.

During preprocessing, the color deconvolution stain unmixing method $ColorDeconv$ [Ruifrok *et al.*, 2001] is first applied to extract the DAB channel from the real IHC RI. The extracted channel is converted to grayscale and smoothed by Gaussian filtering G_σ. Otsu thresholding $Otsu$ then separates the foreground from the background. After the regular erosion and dilation operations, morphological closing $Closing$ fills gaps to produce a continuous automated annotated mask M. The entire process is summarized as:

$$M = Closing(Otsu(G_\sigma(ColorDeconv(RI)))), \quad (1)$$

where the filter's standard deviation σ is empirically set to 1. The disk radius in $Closing$ is adjustable based on alignment quality, with larger radii accommodating greater shift.

3.2 Auxiliary Tasks for Supervision

The absence of pixel-level aligned guidance makes it difficult for the model to learn feature of various biomarkers. To address this, we exploit the approximately aligned positive region masks from preprocessing, which are fully utilized via multiple auxiliary tasks. They guide the model in capturing the location and intensity of positive expression, as shown in Figure 3. Positive nucleus number matching is incorporated to improve the precision of pathological details further.

To globally constrain positive regions, the fake IHC FI undergoes stain unmixing and morphological processing to produce the DAB channel D', Hema channel H' and extracted mask $M' \in \mathbb{R}^{H\times W\times C}$, as detailed in Section 3.1. The location loss for global positive expression between M' and its GT M is:

$$L_1 = \frac{1}{HW}\sum_{x=1}^{H}\sum_{y=1}^{W}\big(M'(x,y) - M(x,y)\big)^2, \quad (2)$$

where x and y represent the vertical and horizontal coordinates of M' and M respectively. To further enhance feature learning for both positive and negative regions at a more localized level, M is divided into n equal-sized patches, denoted as P_k, with k as the patch index. The positive region area within each patch is $S(P_k)$. The top $\frac{n}{4}$ patches with the highest $S(P_k)$ form set $\mathcal{H}$, and the bottom $\frac{n}{4}$ patches with the lowest $S(P_k)$ form set $\mathcal{L}$. The local positive expression supervision losses L_2 and L_3 are then computed as:

$$L_{i=2,3} = \frac{4}{n}\sum_{P_k\in\mathcal{S}_i}\frac{1}{HW}\sum_{x,y\in P_k}\big(M'(x,y) - M(x,y)\big)^2, \quad (3)$$

in which $\mathcal{S}_i$ denotes the patch set, with $\mathcal{S}_2 = \mathcal{H}$ and $\mathcal{S}_3 = \mathcal{L}$.

Given the relatively small size of focused regions and similar intensity of positive expression within them, average staining intensity of positive regions in masks M' and M is used for intensity supervision. The intensity loss L_4 is defined as:

$$L_4 = \Big(\frac{\sum(D'\circ M')}{N_{M'}} - \frac{\sum(D\circ M)}{N_M}\Big)^2, \quad (4)$$

where $N_{M'}$ and N_M are the number of foreground pixels, and $D\circ M$ represents extracting M's foreground region from D.

Finally, a positive nucleus number matching task is introduced to further refine the model's ability to capture fine-grained details. The convergence of the total loss, with this term incorporated, offers stronger evidence of pathological consistency and structural preservation. The widely recognized nucleus segmentation method CPP-Net [Chen *et al.*, 2023] is used for segmentation. Loss L_5 is expressed as:

$$L_5 = \big|N_{\text{CPP}}\big(Seg(H'\circ M')\big) - N_{\text{CPP}}\big(Seg(H\circ M)\big)\big|, \quad (5)$$

where Seg represents segmentation and N_{CPP} is the nucleus number predicted by CPP-Net. $|\cdot|$ denotes the absolute value.

3.3 Deep Pathology Mining Generator

In this study, we improve the stain transfer model's generalization ability by leveraging the feature extraction capability of a specialized large-scale pretrained model. While re-staining slides after cleaning them is impractical in clinical workflows, AI enables us to simulate this ideal scenario.

ATST-Net employs an encoder-decoder architecture. The encoder, based on Vision Transformer (ViT) [Dosovitskiy, 2020], focuses on extracting essential stain-invariant feature from the source image, which corresponds to the de-staining phase. The decoder, composed of multiple upsampling layers, reconstructs key features to the original image size while integrating the style characteristics of new staining during the re-staining phase. Skip connections are incorporated between encoder and decoder layers to share multi-scale details.

To be specific, the encoder integrates PathoDuet [Hua *et al.*, 2024], a cutting-edge SOTA pretrained model for H&E images based on ViT. PathoDuet is trained within a self-supervised framework built on the contrastive learning

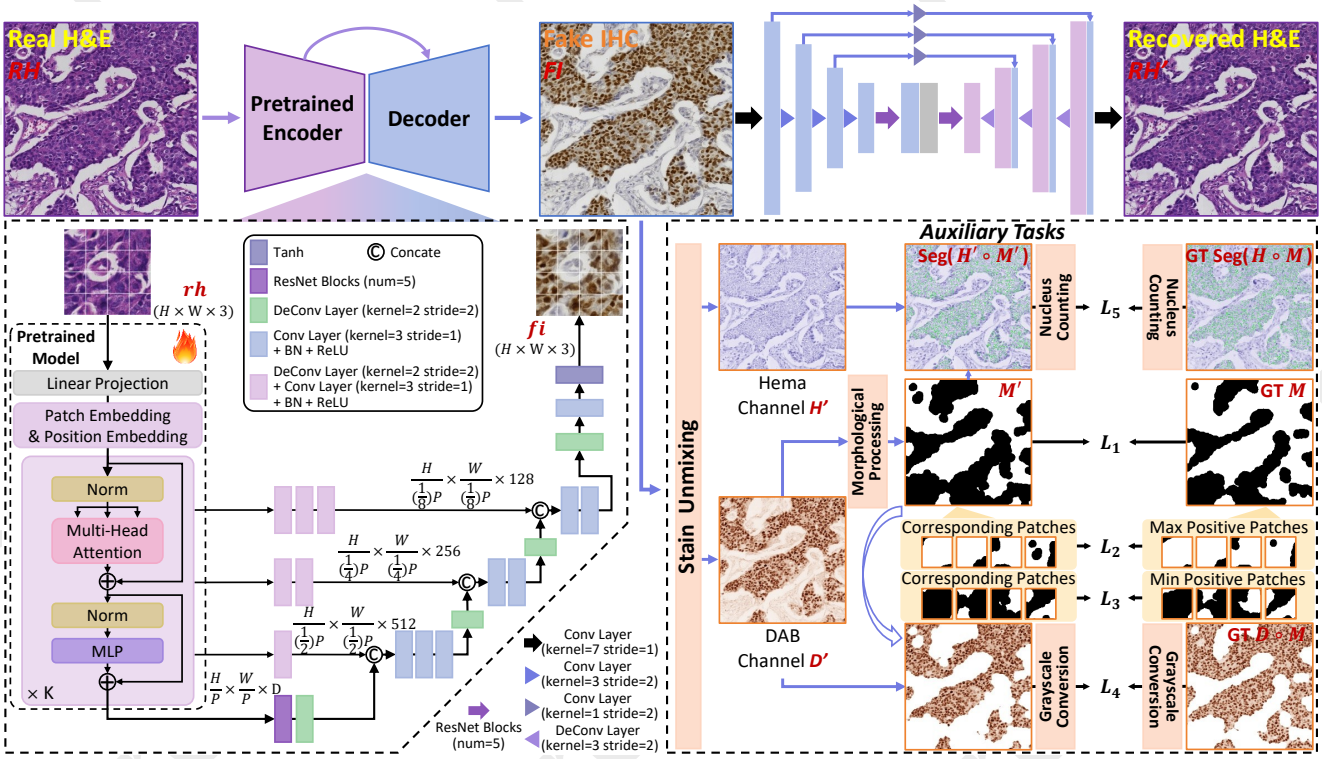

Figure 3: The overview of our proposed ATST-Net. Best viewed at a zoomed-in level.

method MoCo v3 [Chen *et al.*, 2021] using pretext tasks. In our model, the input image $rh \in \mathbb{R}^{H\times W\times C}$ is initially divided into a series of flattened tokens $rh_p \in \mathbb{R}^{L\times(P^2\cdot C)}$, where P is the token size and $L = \frac{HW}{P^2}$ is the total number of tokens. These tokens are linearly projected into dimensions $L \times D$, where D is the latent space dimension, producing rh_{pat}. Next, position embedding rh_{pos} is added, along with a learnable class token rh_{cls} and a placeholder parameter ϵ to align with the PathoDuet structure. The process generates the final input v for the transformer encoder, which consists of K transformer blocks. At the network bottleneck, five ResNet blocks are introduced to fully integrate the extracted feature. Upsampling operations then restore feature maps to the original image size. To further preserve structural details, skip connections transmit low-level information to the decoder layers. Following the convolutional process in CellViT [Hörst *et al.*, 2024], outputs from selected transformer blocks $v_k, k \in \frac{N}{4}, \frac{2N}{4}, \frac{3N}{4}, N$ are concatenated and fused with the corresponding decoder layers after convolutional operations adjust their dimensions.

3.4 Training and Inference

In ATST-Net, CycleGAN [Zhu *et al.*, 2017] serves as the backbone for style transformation. The deep pathology mining generator, along with multiple auxiliary task constraints, ensures the accuracy of pathological consistency and structural preservation during stain transfer. The proposed deep pathology mining architecture is applied for H&E-to-IHC conversion. For IHC-to-H&E conversion, a simpler U-Net structure [Liu *et al.*, 2021] based on convolutional operations is employed. This choice improves computational efficiency, given the lower color richness of H&E images. The discriminator structure follows the design in [Zhu *et al.*, 2017].

During training, two primary loss functions are applied. The first is the adversarial loss $L_{adv}(G_1, D_{IHC})$ and $L_{adv}(G_2, D_{HE})$, where G_1 is the generator maps images from the real H&E domain HE to the IHC domain IHC, and G_2 maps images from the IHC domain to H&E. D_{IHC} and D_{HE} are discriminators that enforce the generated images to conform to the target domain's distribution and staining properties. To constrain the mapping space and avoid infinite possibilities, a cycle consistency loss $L_{cycle}(G_1, G_2)$ ensures that the output fake IHC can be mapped back to the H&E domain. Additionally, structural similarity (SSIM) is incorporated into $L_{cycle}(G_1, G_2)$ to enhance brightness, contrast and structural fidelity in generated images. Combined with multiple auxiliary task supervision losses applied solely during the H&E-IHC-H&E transformation, the total loss function is:

$$\begin{aligned} L_{total} = {} & L_{adv}(G_1, D_{IHC}) + L_{adv}(G_2, D_{HE}) \\ & + \alpha L_{cycle}(G_1, G_2) + \beta_1 L_1 + \cdots + \beta_n L_n, \end{aligned} \tag{6}$$

where $n = 5$, and α and β control the weight of each term.

4 Experimental Setup

4.1 Datasets

Experiments are conducted on two high-quality public datasets: the Breast Cancer Immunohistochemical (BCI) challenge dataset [Liu *et al.*, 2022] and the Multi-IHC Stain

Translation (MIST) dataset [Li *et al.*, 2023]. BCI contains 3896 H&E-HER2 training pairs and 977 test pairs (all 1024×1024) from 51 whole slide images (WSIs), with HER2 expression categorized into four intensity levels: 0, 1+, 2+ and 3+. MIST provides four biomarkers (ER, PR, Ki67, HER2) with 4153, 4139, 4361 and 4642 training pairs, and 1000 test pairs each (all 1024×1024). In our experiments, poorly aligned MIST training samples are removed, yielding 4000 training and 1000 test pairs per biomarker. All BCI data is used without filtering. The BCI test set contains 38, 235, 446 and 258 image pairs for HER2 level 0, 1+, 2+ and 3+.

4.2 Evaluation Metrics

We employ six evaluation metrics to comprehensively assess method performance, categorized into paired and unpaired metrics. The paired metrics [Liu *et al.*, 2021] include: (1) Peak Signal-to-Noise Ratio (PSNR). (2) SSIM: Evaluates similarity in brightness, contrast and structure. (3) Contrast-Structure Similarity (CSS): A SSIM variant that focuses on contrast and structure rather than intensity. (4) Perceptual Hash Value (PHV): Assesses perceptual similarity between fake and real images via feature maps. The unpaired metrics [Li *et al.*, 2023] are: (1) Fréchet Inception Distance (FID) and (2) Kernel Inception Distance (KID): Measure feature-space distribution similarity between fake and real IHC image sets.

4.3 Implementation Details

All methods are implemented in PyTorch 1.12.1 on a single NVIDIA GeForce RTX 3090 GPU. During training and testing, images are cropped to 512×512 and stitched after processing. To retain pathological details in H&E and IHC images, no normalization is applied. CPP-Net is trained on a systematic IHC nucleus dataset [Xu *et al.*, 2024]. The proposed deep pathology mining generator uses 224×224 inputs to match the pretrained model. Training adopts a batch size of 1, an initial learning rate of 0.0002 with linear decay after half the epochs, and the Adam optimizer. Hyper-parameters in L_{total} are: $\alpha = 5$, $\beta_1 = 60$, $\beta_2 = \beta_3 = 10$, $\beta_4 = 20$, $\beta_5 = 1$. The model is trained end-to-end. Parameter settings for comparison methods follow their original papers.

5 Experiments and Analysis

5.1 Comparisons with the SOTA Methods

To assess stain transfer accuracy, ATST-Net is evaluated on four MIST biomarkers and one BCI biomarker with four levels. We compare it with the baseline style transfer method CycleGAN and SOTA stain transfer methods, i.e., PC-StainGAN [Liu *et al.*, 2021], UMDST [Lin *et al.*, 2022], PyramidP2P [Liu *et al.*, 2022], ASP [Li *et al.*, 2023] and PSPStain [Chen *et al.*, 2024]. Following the protocol in Section 4.1, we adopt the same data volume and splits, with image sizes as reported. For PC-StainGAN, H&E labels are generated using the method in Section 3.1 due to unavailable expert annotations. UMDST is trained separately biomarkers.

Quantitative results for MIST (ER, HER2) and BCI are shown in Table 1, with MIST (PR, Ki67) in the supplementary materials. The quality of generated images (fake IHC) is first assessed using PSNR and SSIM, though these metrics are suboptimal due to structural deviations caused by section misalignment and preparation variations. Thus, they serve as reference metrics. This is further reflected in CSS, which mitigates per-pixel intensity impact but still suffers from misalignment, resulting in lower scores across all methods.

While CycleGAN is effective for style transfer, it lacks pathological consistency constraints, leading to incorrect stain transfer and poor metric performance. UMDST, with its dynamic style adjustment ability, adaptively represents the relationship between stains but struggles with complex IHC images involving multiple biomarkers, where positive region accuracy decreases. PyramidP2P aligns high-dimensional feature maps of fake and real IHC. It maintains positive region consistency through aggregated feature constraints, but loses the original tissue structure totally. PC-StainGAN improves pathological feature differentiation while preserving structure. However, it relies on precise manual annotations, with any decrease in annotation accuracy greatly affects its performance. Notably, ASP and PSPStain excel in PHV, FID and KID by using task-specific networks to preserve pathological consistency and tissue structure in H&E-to-IHC style transfer. PHV shows higher stain transfer accuracy across different layer levels between fake and corresponding real IHC, while FID and KID further confirm consistency in feature distribution across a larger set. Our ATST-Net outperforms existing methods on most key metrics. On $\text{MIST}_{\text{HER2}}$, it improves CSS, PHV(avg), FID and KID by 0.019, 0.027, 11.2 and 4.6, showcasing superior accuracy. Quantitative analysis reveals ATST-Net shows reduced improvements on BCI compared to MIST, mainly due to MIST's higher alignment precision, which offers more accurate guidance for model learning.

We further evaluate methods for differentiating positive expression intensity in IHC (0, 1+, 2+ and 3+) when learning pathological information. Results for level 0 and 3+ in the BCI test data are in Table 1, with level 1+ and 2+ in the supplementary materials. PSPStain, which aligns average protein expression intensity during training, outperforms other methods in most metrics. Our proposed ATST-Net constrains positive expression location while focusing on intensity. Its generator excels at identifying varying expression degrees across different locations, surpassing PSPStain on multiple metrics.

To assess ATST-Net's generalization performance, we design cross-dataset validation. The MIST dataset has clearer tissue structures, greater variation in expression intensity and better alignment between H&E and IHC images compared to BCI. We train on $\text{MIST}_{\text{HER2}}$ and test on BCI, which exhibits great difference. Results for this MIST→BCI (training dataset→testing dataset) experiment are presented in the $\text{MIST}_{\text{HER2}}$ section of Table 1 (values in parentheses). ATST-Net achieves improvements of 0.015, 0.027, 14.4 and 6.3 in CSS, PHV(avg), FID and KID, respectively, demonstrating stronger generalization. However, the metrics are lower than when training on BCI, highlighting how variations in tissue sample, staining and imaging setup, even for the same tissue and biomarker type, can cause significant image variations. In practice, stain transfer models perform optimally only when test data closely matches training data in tissue and biomarker type, sample batch, staining and imaging condition.

In addition, compared to other methods, ATST-Net stands

Dataset	Method	PSNR↑	SSIM↑	CSS↑	PHV$_{T=0.01}$ ↓					FID↓	KID↓
					layer1	layer2	layer3	layer4	avg.		
Comparisons on MIST and BCI.											
MIST$_{ER}$	CycleGAN	11.859	0.183	0.152	0.567	0.486	0.311	0.856	0.555	132.1	92.5
	PC-StainGAN	12.260	0.174	0.167	0.484	0.431	**0.267**	0.845	0.507	62.5	15.4
	UMDST	**12.348**	0.188	0.163	0.513	0.430	0.294	0.851	0.522	76.2	19.7
	PyramidP2P	11.759	**0.204**	0.178	0.461	0.448	0.343	0.868	0.530	101.1	82.3
	ASP	11.550	0.166	0.177	0.451	0.405	0.271	0.845	0.493	53.7	6.2
	PSPStain	11.782	0.184	0.181	0.436	0.396	0.273	0.840	0.486	52.8	6.4
	ATST-Net	11.989	0.201	**0.190**	**0.407**	**0.374**	0.274	**0.817**	**0.468**	**46.3**	**5.0**
MIST$_{HER2}$ (MIST→BCI)	CycleGAN	13.240	0.186	0.109	0.589	0.530	0.329	0.857	0.576	235.2	323.9
	CycleGAN	(15.270)	(0.270)	(0.087)	(0.684)	(0.586)	(0.392)	(0.873)	(0.634)	(242.9)	(356.2)
	PC-StainGAN	13.216	**0.195**	0.116	0.486	0.472	0.301	0.846	0.526	84.4	16.5
	PC-StainGAN	(14.311)	(**0.355**)	(0.081)	(0.671)	(0.602)	(0.378)	(0.874)	(0.631)	(196.9)	(89.0)
	UMDST	13.216	0.169	0.123	0.504	0.445	0.273	0.842	0.516	86.1	18.4
	UMDST	(15.396)	(0.261)	(0.072)	(0.634)	(0.555)	(0.371)	(0.856)	(0.604)	(117.1)	(39.2)
	PyramidP2P	**13.636**	**0.195**	0.113	0.460	0.458	0.360	0.869	0.537	106.2	74.7
	PyramidP2P	(**17.207**)	(0.305)	(0.044)	(0.670)	(0.649)	(0.386)	(0.865)	(0.643)	(198.7)	(97.7)
	ASP	13.452	0.192	0.118	0.454	0.423	0.267	0.839	0.496	84.3	15.3
	ASP	(16.152)	(0.288)	(0.077)	(0.582)	(0.514)	(0.366)	(0.859)	(0.580)	(122.0)	(47.2)
	PSPStain	13.514	0.178	0.137	0.424	0.398	0.259	0.838	0.480	79.3	13.2
	PSPStain	(15.517)	(0.299)	(0.083)	(0.650)	(0.573)	(0.364)	(**0.843**)	(0.608)	(128.3)	(54.5)
	ATST-Net	13.619	0.187	**0.156**	**0.391**	**0.364**	**0.238**	**0.819**	**0.453**	**68.1**	**8.6**
	ATST-Net	(15.223)	(0.304)	(**0.102**)	(**0.541**)	(**0.477**)	(**0.345**)	(0.848)	(**0.553**)	(**102.7**)	(**32.9**)
BCI$_{HER2}$	CycleGAN	20.029	0.437	0.089	0.427	0.505	0.324	0.798	0.514	60.3	10.4
	PC-StainGAN	20.224	0.446	0.096	0.422	0.375	0.244	0.752	0.448	69.0	19.5
	UMDST	19.003	0.458	0.093	0.510	0.422	0.276	0.778	0.497	79.7	27.0
	PyramidP2P	**20.978**	0.463	0.083	0.611	0.419	0.263	0.740	0.508	117.5	76.3
	ASP	20.173	0.484	0.107	0.520	0.373	0.259	0.736	0.472	64.8	10.2
	PSPStain	20.484	0.444	0.126	0.434	0.367	0.234	**0.715**	0.438	49.3	8.5
	ATST-Net	20.655	**0.491**	**0.135**	**0.408**	**0.351**	**0.218**	0.729	**0.427**	**43.6**	**6.5**
Comparisons on different expression levels of HER2 in BCI.											
BCI$_{HER2}$ (level 0)	CycleGAN	21.226	0.505	0.118	**0.421**	0.510	0.312	0.789	0.508	164.9	7.3
	PC-StainGAN	21.742	0.523	0.126	0.453	0.386	0.240	0.755	0.459	182.7	15.7
	UMDST	20.683	**0.541**	0.112	0.499	0.394	0.253	0.762	0.477	198.7	27.1
	PyramidP2P	**22.409**	0.529	0.110	0.623	0.380	0.241	0.730	0.494	175.8	60.8
	ASP	21.344	0.538	0.145	0.505	0.359	0.229	0.727	0.455	177.0	7.1
	PSPStain	22.078	0.513	0.171	0.426	0.345	0.214	0.714	0.425	163.2	6.8
	ATST-Net	22.215	0.535	**0.182**	0.430	**0.333**	**0.199**	**0.711**	**0.418**	**148.9**	**6.2**
BCI$_{HER2}$ (level 3+)	CycleGAN	17.546	0.411	0.079	0.470	0.560	0.359	0.837	0.557	135.7	44.5
	PC-StainGAN	17.560	0.420	0.094	0.481	0.430	0.277	0.800	0.497	149.7	52.4
	UMDST	16.909	0.433	0.088	0.558	0.470	0.302	0.812	0.536	153.1	51.2
	PyramidP2P	**17.753**	0.426	0.080	0.642	0.485	0.332	0.791	0.563	208.5	126.9
	ASP	17.410	0.440	0.104	0.562	0.462	0.301	**0.788**	0.528	152.8	45.8
	PSPStain	17.633	0.418	0.111	0.491	0.416	0.269	0.793	0.492	133.9	40.5
	ATST-Net	17.437	**0.451**	**0.118**	**0.450**	**0.395**	**0.245**	0.815	**0.476**	**112.7**	**30.4**

Table 1: Comparisons of various methods on MIST and BCI. The KID values in the table are scaled by a factor of 1000.

out by enabling bidirectional transfer, including IHC-to-H&E transfer via its backbone. As shown in Figure 3, a simpler network performs the reverse transfer, leveraging the lower color complexity. Figure 5 shows H&E images recovered from IHC inputs. ATST-Net preserves IHC tissue structures while restoring the H&E stain style, enhancing practical value.

5.2 Qualitative Comparisons

We qualitatively compare all methods on the four MIST biomarkers, as shown in Figure 4. PyramidP2P distorts tissue structure in H&E images, while other methods maintain structural integrity. Notably, ATST-Net shows superior pathological consistency, as highlighted by the red boxes in the first two rows. The fake IHC images more closely match real IHC images in both positive expression location and intensity. Moreover, ATST-Net transfers sparse Ki67 expression more accurately (third row). Patches generated by ATST-Net exhibit minimal color discontinuity when stitched (fourth row), ensuring better prediction consistency and practical value.

6 Ablation Study

We conduct ablation experiments on all ER and HER2 test images in MIST. Firstly, we evaluate the proposed generator and multiple loss functions, then analyze the generator's structure. Evaluation metrics are shown in Table 2 and 3, where "Gen", "U", "Skip" and "Res" represent "Generator", "U-Net-based", "Skip connections" and "ResNet blocks".

When solely employing the U-Net-based generator for H&E-to-IHC transfer, identical to that used for IHC-to-H&E transfer, fake IHC images lack pathological consistency, resulting in poor performance, with evaluation metrics comparable to traditional CycleGAN. In contrast, the proposed deep pathology mining generator, based on a pretrained model, effectively learns pathological features of various biomarkers from H&E images. However, its potential is not fully realized without guidance from biomarker positive expression reference. As shown in the second row of Table 2, the model's performance on CSS, PHV, FID and KID in MIST$_{ER}$ shows

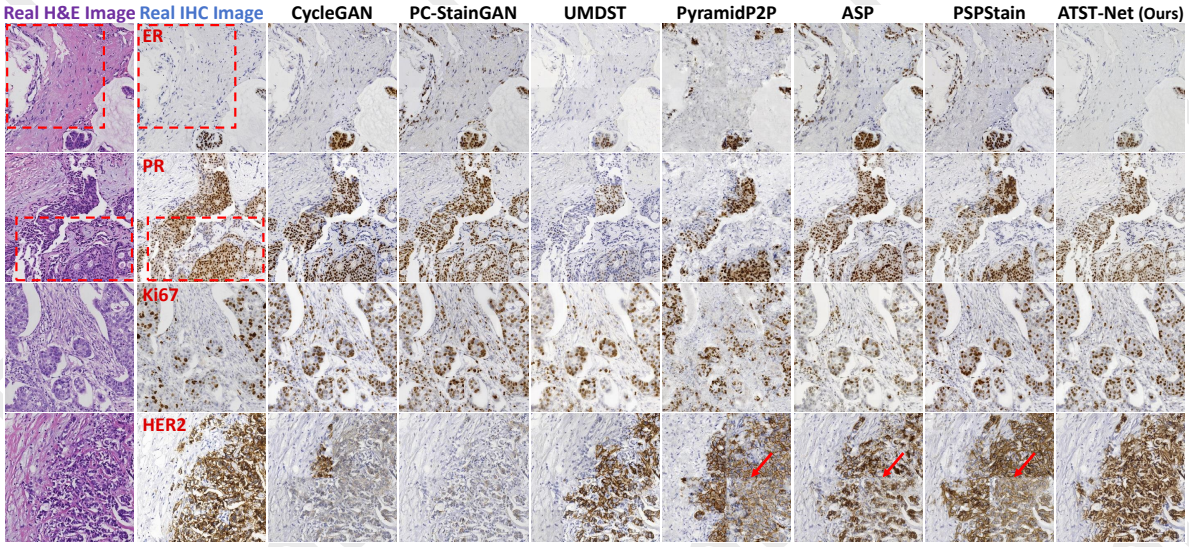

Figure 4: Qualitative comparison on MIST. The four rows correspond to ER, PR, Ki67 and HER2. All images are 1024×1024.

Dataset	Gen		Auxiliary Task				CSS↑	PHV↓ (avg.)	FID↓	KID↓
	U	Ours	L_1	L_2&L_3	L_4	L_5				
$MIST_{ER}$	✓						0.155	0.541	119.0	83.3
		✓					0.165	0.529	92.3	62.0
		✓	✓				0.161	0.505	69.6	18.8
		✓	✓	✓			0.177	0.491	58.5	6.5
		✓	✓	✓	✓		0.184	0.477	49.8	5.4
	✓		✓	✓	✓	✓	0.175	0.494	57.0	7.2
		✓	✓	✓	✓	✓	**0.190**	**0.468**	**46.3**	**5.0**
$MIST_{HER2}$	✓						0.115	0.562	190.6	262.6
		✓					0.120	0.554	164.8	233.9
		✓	✓				0.127	0.531	90.2	23.7
		✓	✓	✓			0.146	0.476	78.4	11.0
		✓	✓	✓	✓		0.151	0.465	72.9	12.2
	✓		✓	✓	✓	✓	0.153	0.470	81.7	11.3
		✓	✓	✓	✓	✓	**0.156**	**0.453**	**68.1**	**8.6**

Table 2: Ablation study of components in ATST-Net.

Dataset	Module		Auxiliary Task	CSS↑	PHV↓ (avg.)	FID↓	KID↓
	Skip	Res	$L_1, L_2, ..., L_5$				
$MIST_{ER}$			✓	0.160	0.505	55.0	7.9
	✓		✓	0.194	0.486	50.1	5.1
		✓	✓	**0.196**	0.479	53.9	7.1
	✓	✓	✓	0.190	**0.468**	**46.3**	**5.0**
$MIST_{HER2}$			✓	0.063	0.500	83.0	25.7
	✓		✓	**0.161**	0.484	81.6	23.4
		✓	✓	0.105	0.461	75.7	13.9
	✓	✓	✓	0.156	**0.453**	**68.1**	**8.6**

Table 3: Ablation study on the generator.

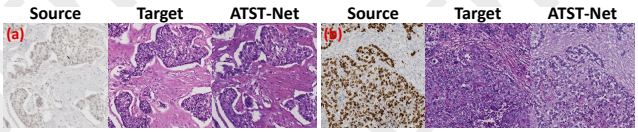

Figure 5: IHC-to-H&E stain transfer. (a) $MIST_{ER}$, (b) $MIST_{Ki67}$.

marginal improvement over the U-Net-based generator.

Applying global positive expression location loss (L_1) improves accuracy in tissue regions with clustered positive expression, but the model struggles in sparse expression areas. Local expression supervision losses (L_2&L_3) mitigate this by guiding the model to focus on both positive and negative feature, substantially improving all metrics. While L_1, L_2 and L_3 enable the model to identify positive expression locations, generated intensities tend to be uniform and biased toward dominant intensities in the training data. To address this, intensity supervision loss (L_4) ensures precise expression intensity. Additionally, nucleus number matching loss (L_5) refines expression location while aligning the pathological distribution of all fake IHC images with real ones from consecutive sections. These improvements are reflected in FID and KID, which increase by 3.5 and 0.4 in $MIST_{ER}$. Integrating the proposed generator with auxiliary task supervision losses yields optimal results across all metrics: CSS, PHV, FID and KID in $MIST_{ER}$ reach 0.190, 0.468, 46.3 and 5.0, respectively, and in $MIST_{HER2}$, these values are 0.156, 0.453, 68.1 and 8.6. The last two rows for each biomarker in Table 2 further confirm our generator's higher accuracy and generalization ability.

We demonstrate the role of each module in the proposed generator. Skip connections transmit multi-scale shallow and deep features to the decoder, helping preserve textural and structural details. Multiple ResNet blocks in the network bottleneck integrate and stabilize deep stain-invariant features. As shown in Table 3, the combination of skip connections and ResNet blocks yields superior performance across all metrics.

7 Conclusion

This study proposes a cost-effective H&E-to-IHC stain transfer method with high accuracy, generalization and interpretability. It integrates automatic annotation, auxiliary tasks and a generator architecture. Comparisons on public datasets with diverse biomarkers and expression levels demonstrate its SOTA accuracy and strong clinical applicability.

Acknowledgments

This work is supported by the Fundamental Research Funds for the Central Universities, the National Natural Science Foundation of China (Grant No. 62475072, 62471182), the Science and Technology Commission of Shanghai Municipality (Grant No. 22S31905800, 22DZ2229004), Shanghai Rising-Star Program (Grant No. 24QA2702100) and the Open Research Fund of Key Laboratory of Advanced Theory and Application in Statistics and Data Science–MOE, ECNU (Grant No. KLATASDS2406).

References

[Boyd *et al.*, 2022] Joseph Boyd, Irène Villa, Marie-Christine Mathieu, Eric Deutsch, Nikos Paragios, Maria Vakalopoulou, and Stergios Christodoulidis. Region-guided cyclegans for stain transfer in whole slide images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 356–365. Springer, 2022.

[Chen *et al.*, 2021] Xinlei Chen, Saining Xie, and Kaiming He. An empirical study of training self-supervised vision transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9640–9649, 2021.

[Chen *et al.*, 2023] Shengcong Chen, Changxing Ding, Minfeng Liu, Jun Cheng, and Dacheng Tao. Cpp-net: Context-aware polygon proposal network for nucleus segmentation. *IEEE Transactions on Image Processing*, 32:980–994, 2023.

[Chen *et al.*, 2024] Fuqiang Chen, Ranran Zhang, Boyun Zheng, Yiwen Sun, Jiahui He, and Wenjian Qin. Pathological semantics-preserving learning for h&e-to-ihc virtual staining. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 384–394. Springer, 2024.

[Chhikara *et al.*, 2023] Bhupender S Chhikara, Keykavous Parang, et al. Global cancer statistics 2022: The trends projection analysis. *Chemical Biology Letters*, 10(1):451–451, 2023.

[Dosovitskiy, 2020] Alexey Dosovitskiy. An image is worth 16x16 words: Transformers for image recognition at scale. *ArXiv Preprint ArXiv:2010.11929*, 2020.

[Farahmand *et al.*, 2022] Saman Farahmand, Aileen I Fernandez, Fahad Shabbir Ahmed, David L Rimm, Jeffrey H Chuang, Emily Reisenbichler, and Kourosh Zarringhalam. Deep learning trained on hematoxylin and eosin tumor region of interest predicts her2 status and trastuzumab treatment response in her2+ breast cancer. *Modern Pathology*, 35(1):44–51, 2022.

[He *et al.*, 2024] Yufang He, Zeyu Liu, Mingxin Qi, Shengwei Ding, Peng Zhang, Fan Song, Chenbin Ma, Huijie Wu, Ruxin Cai, Youdan Feng, et al. Pst-diff: Achieving high-consistency stain transfer by diffusion models with pathological and structural constraints. *IEEE Transactions on Medical Imaging*, 2024.

[Hörst *et al.*, 2024] Fabian Hörst, Moritz Rempe, Lukas Heine, Constantin Seibold, Julius Keyl, Giulia Baldini, Selma Ugurel, Jens Siveke, Barbara Grünwald, Jan Egger, et al. Cellvit: Vision transformers for precise cell segmentation and classification. *Medical Image Analysis*, 94:103143, 2024.

[Hua *et al.*, 2024] Shengyi Hua, Fang Yan, Tianle Shen, Lei Ma, and Xiaofan Zhang. Pathoduet: Foundation models for pathological slide analysis of h&e and ihc stains. *Medical Image Analysis*, 97:103289, 2024.

[Isola *et al.*, 2017] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1125–1134, 2017.

[Li *et al.*, 2023] Fangda Li, Zhiqiang Hu, Wen Chen, and Avinash Kak. Adaptive supervised patchnce loss for learning h&e-to-ihc stain translation with inconsistent groundtruth image pairs. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 632–641. Springer, 2023.

[Li *et al.*, 2025] Tianye Li, Haoxiang Zhang, Mengyi Lian, Qionghua He, Mingwei Lv, Lingyun Zhai, Jianwei Zhou, Kongming Wu, and Ming Yi. Global status and attributable risk factors of breast, cervical, ovarian, and uterine cancers from 1990 to 2021. *Journal of Hematology & Oncology*, 18(1):1–25, 2025.

[Lin *et al.*, 2022] Yiyang Lin, Bowei Zeng, Yifeng Wang, Yang Chen, Zijie Fang, Jian Zhang, Xiangyang Ji, Haoqian Wang, and Yongbing Zhang. Unpaired multi-domain stain transfer for kidney histopathological images. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 1630–1637, 2022.

[Liu *et al.*, 2020] Yiqing Liu, Xi Li, Aiping Zheng, Xihan Zhu, Shuting Liu, Mengying Hu, Qianjiang Luo, Huina Liao, Mubiao Liu, Yonghong He, et al. Predict ki-67 positive cells in h&e-stained images using deep learning independently from ihc-stained images. *Frontiers in Molecular Biosciences*, 7:183, 2020.

[Liu *et al.*, 2021] Shuting Liu, Baochang Zhang, Yiqing Liu, Anjia Han, Huijuan Shi, Tian Guan, and Yonghong He. Unpaired stain transfer using pathology-consistent constrained generative adversarial networks. *IEEE Transactions on Medical Imaging*, 40(8):1977–1989, 2021.

[Liu *et al.*, 2022] Shengjie Liu, Chuang Zhu, Feng Xu, Xinyu Jia, Zhongyue Shi, and Mulan Jin. Bci: Breast cancer immunohistochemical image generation through pyramid pix2pix. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1815–1824, 2022.

[Macenko *et al.*, 2009] Marc Macenko, Marc Niethammer, James S Marron, David Borland, John T Woosley, Xiaojun Guan, Charles Schmitt, and Nancy E Thomas. A method for normalizing histology slides for quantitative analysis. In *IEEE International Symposium on Biomedical Imaging*, pages 1107–1110. IEEE, 2009.

[Park *et al.*, 2020] Taesung Park, Alexei A Efros, Richard Zhang, and Jun-Yan Zhu. 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*, pages 319–345. Springer, 2020.

[Pati *et al.*, 2024] Pushpak Pati, Sofia Karkampouna, Francesco Bonollo, Eva Compérat, Martina Radić, Martin Spahn, Adriano Martinelli, Martin Wartenberg, Marianna Kruithof-de Julio, and Marianna Rapsomaniki. Accelerating histopathology workflows with generative ai-based virtually multiplexed tumour profiling. *Nature Machine Intelligence*, 6(9):1077–1093, 2024.

[Reinhard *et al.*, 2001] Erik Reinhard, Michael Adhikhmin, Bruce Gooch, and Peter Shirley. Color transfer between images. *IEEE Computer Graphics and Applications*, 21(5):34–41, 2001.

[Ruifrok *et al.*, 2001] Arnout C Ruifrok, Dennis A Johnston, et al. Quantification of histochemical staining by color deconvolution. *Analytical and Quantitative Cytology and Histology*, 23(4):291–299, 2001.

[Shaban *et al.*, 2019] M Tarek Shaban, Christoph Baur, Nassir Navab, and Shadi Albarqouni. Staingan: Stain style transfer for digital histological images. In *IEEE International Symposium on Biomedical Imaging*, pages 953–956. IEEE, 2019.

[Xu *et al.*, 2024] Siyuan Xu, Guannan Li, Haofei Song, Jiansheng Wang, Yan Wang, and Qingli Li. Genseg-net: A general segmentation framework for any nucleus in immunohistochemistry images. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 4475–4484, 2024.

[Zeng *et al.*, 2022] Bowei Zeng, Yiyang Lin, Yifeng Wang, Yang Chen, Jiuyang Dong, Xi Li, and Yongbing Zhang. Semi-supervised pr virtual staining for breast histopathological images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 232–241. Springer, 2022.

[Zhu *et al.*, 2017] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2223–2232, 2017.