Histocomponent-driven Universal Model for Virtual Immunohistochemistry Multiplex Staining via Joint Manifold Evolution

Jiajun Cen[1], Siyuan Xu[1], Lili Gao[2], and Yan Wang[1⋆]

[1] Shanghai Key Laboratory of Multidimensional Information Processing, East China Normal University, Shanghai, China
{Jjcen, syxu}@stu.ecnu.edu.cn, ywang@cee.ecnu.edu.cn
[2] Department of Pathology, Ruijin Hospital, Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China
gll12216@rjh.com.cn

Abstract. Virtual immunohistochemistry (IHC) multiplex staining has emerged as a highly promising non-destructive solution in digital pathology. However, existing universal IHC staining models typically guide the generation of specific biomarkers by injecting task-specific text prompts, which often struggle to bridge the semantic gap with complex pathological structures, leading to the "manifold drift" problem. Moreover, these models lack specificity in modeling heterogeneous tissue components. In this paper, we propose **HUSE**, a **H**istocomponent-driven **U**niversal model for virtual IHC multiplex **S**taining via joint manifold **E**volution. HUSE introduces Joint Manifold Anchoring (JMA) as a global solution to treat H&E images as persistent structural scaffolds, effectively anchoring the generation trajectory to the intrinsic image manifold. Complementing this, a localized refinement scheme named Histocomponent-driven Mixture of Experts (Hi-MoE) and Representation Conflict Gating (RCG) is introduced to precisely capture local topological features of complex manifolds. Furthermore, we pioneer an mIF-driven multi-marker IHC generation paradigm to construct large-scale $1(H\&E) : N(IHC)$ training matrices; evaluations by senior pathologists verify the feasibility of this paradigm as well as its exceptional clinical realism and structural fidelity. Extensive experiments on two distinct types of datasets demonstrate that HUSE achieves state-of-the-art performance. Code is released at https://github.com/DeepMed-Lab-ECNU/HUSE.

Keywords: Virtual Immunohistochemistry Multiplex Staining · Universal Model · mIF-driven IHC Generation

1 Introduction

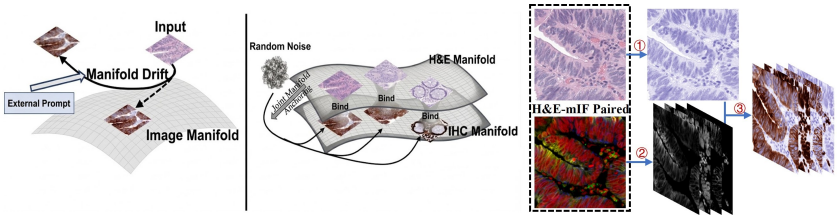
Hematoxylin and Eosin (H&E) staining is the gold standard for clinical pathological diagnosis [12], typically requiring combination with immunohistochemistry

⋆ Corresponding author.

(IHC) staining to provide auxiliary evidence at the molecular level [10]. However, traditional IHC examinations not only consume precious tissue samples but also face high time and labor costs [32]. Deep learning-based virtual staining has emerged as a highly promising non-destructive solution in the field of digital pathology [24]. Early research mostly focused on a "one-for-one" mode, which means training independent models for each biomarker. This approach is extremely inefficient when facing dozens of potential markers in clinical practice [7, 14, 25, 30]. Therefore, developing "One-to-Many" universal models that achieve the precise generation of multiple markers through a single network has become a current research frontier.

However, current universal IHC virtual staining models exhibit significant practical deficiencies. First, at the global architectural level, most models merely treat H&E images as editable base maps, using task-specific text prompts to guide biomarker generation. [6, 9, 31]. According to the manifold assumption [5], high-dimensional image data resides on structured low-dimensional manifolds; however, text-driven methods often fail to bridge the semantic gap in pathology. Unlike natural images that can be described by a general template, pathological images feature fine-grained, dense, and complex morphology, including heterogeneous nuclei and intricate stromal textures [1]. More importantly, every pathological image is unique, making it impossible to apply a universal textual template for full characterization [21]. Thus, it is evident that the distribution of external text embeddings is inherently different from the intrinsic image manifold; such an innate spatial mismatch introduces external interference that pulls the generation process away from the real morphology constraints, leading to "Manifold Drift" (see Fig. 1a (left)) [4, 5]. Furthermore, these methods are highly dependent on prompt design, suffering from poor reproducibility and unstable marker expression when facing dozens of targets. Second, at the local parameter modeling level, most models use homogeneous parameters, attempting to fit heterogeneous tissues (e.g., nuclei, stroma) with shared weights, which compromises marker specificity. While some employ independent decoders per biomarker [11, 22], these multi-head architectures represent multi-task learning rather than truly universal models. Consequently, they face prohibitive computational costs and poor scalability as the number of biomarker types increases.

To address the aforementioned deficiencies, we propose Histocomponent-driven Universal model for virtual IHC Multiplex Staining via joint manifold Evolution (HUSE), a pixel-level universal diffusion model for virtual IHC staining framework that is driven by flow matching. Different from previous methods that rely on external text prompts to guide the generation process, we propose Joint Manifold Anchoring (JMA) as a global solution, as shown in Fig. 1a (right). The core rationale behind JMA is that H&E and IHC images are complementary representations observed from the same biological substrate. Thus, their underlying data manifolds are intrinsically coupled, sharing highly aligned spatial coordinates and topological structures within the high-dimensional pixel space. By integrating deep pixel-level coupling at the initial stage of generation, we establish a robust structural constraint between the H&E scaffold and the



(a) Conceptual schematic illustration of Manifold Drift and Joint Manifold Anchoring

(b) Diagram of mIF-Driven multi-marker IHC Generation

Fig. 1: (a) Conceptual illustration of manifold evolution. Left: Conventional methods using decoupled external prompts for biomarker-specific generation often suffer from manifold drift, leading to structural blurring. Right: We integrate the H&E image with random noise at the initial stage, tightly binding the generation process to the structural scaffold of H&E via Joint Manifold Anchoring, thereby ensuring the synthesis of high-fidelity virtual staining results. (b) Diagram of mIF-driven multi-marker IHC generation.①: Color deconvolution.②: Separate and normalization.②: Physics-driven synthesis (see 3.3).

target IHC within the manifold space. This ensures that the generation process remains strictly anchored to the real morphology, effectively suppressing manifold drift. We utilize a class embedder to map marker indices into learnable tokens that are injected into the backbone to precisely control marker-specific generation, eliminating the reliance on inappropriate text.

To overcome the performance bottlenecks in modeling heterogeneous tissues, we are the first to introduce a Histocomponent-driven Mixture of Experts (Hi-MoE) mechanism within our framework. As a localized refinement scheme, Hi-MoE is designed to precisely capture the local topological features of complex manifolds. Rooted in pathological priors, Hi-MoE consists of three learnable specialized experts: the nuclear expert, the stroma expert, and the cytoplasm/background expert. Guided by histocomponent-driven router, the model matches the most suitable expert branches to different histological regions across the image, addressing the limitations of using a single set of parameters for diverse tissue types. To ensure that expert responsibilities are histocomponent-constrained from the outset, we initialize the expert prototypes using biological conceptual features extracted by pre-trained model based on their respective category. Complementing this is the Representation Conflict Gating (RCG), which monitors feature state displacement in real-time. When shared parameters encounter cognitive conflicts while processing complex structures, RCG acutely captures this representation conflict. Based on the expert category established by the router, it dynamically determines the residual contribution weights of the pathological expert for fine-grained refinement. This design endows HUSE with exceptional anatomical awareness and interpretability.

Training high-performance universal IHC staining models requires the support of large-scale, precisely aligned data [20]. However, existing H&E-IHC public datasets face the challenges of small sample sizes, poor paired alignment,

and the absence of a "one-to-many $(1 : N)$" data matrix [14, 18]. Specifically, in benchmarks like MIST, biomarkers are often partitioned into independent subsets, lacking co-localized multiplexed labels, making it difficult to comprehensively evaluate the ability of previous universal models to synthesize diverse biomarkers for the same H&E slide. In contrast, multiplex immunofluorescence (mIF) possesses a natural advantage in in-situ alignment and enjoys abundant public data resources [3, 16, 28]. Although mIF is difficult to popularize clinically due to its high cost, technical complexity, and susceptibility to quenching, its perfect alignment properties make it an ideal testbed for training universal models. To this end, we, as pioneers, propose a cross-modal data synthesis paradigm based on mIF (Fig. 1b). Through texture modulation techniques [27], we transform the in-situ aligned fluorescence signals into clinically realistic virtual IHC images, constructing a high-precision, large-scale $1(H\&E) : N(IHC)$ training matrix. We invited senior pathologists for evaluation; the IHC images synthesized by this paradigm exhibit extremely high clinical realism in terms of marker-specific expression and structural fidelity. This paradigm lays a solid foundation for the optimization of IHC universal virtual staining models.

Through the proposed HUSE framework, high-fidelity generation from a single H&E image to multiple IHC markers is achieved. Extensive experimental results on two distinct types of datasets demonstrate the superior performance of HUSE. Notably, this work represents the most extensive evaluation in the field of virtual IHC staining to date, covering a total of 19 distinct biomarkers across two datasets. Furthermore, we are the first to propose the mIF-driven multi-marker IHC generation paradigm to address the deficiency of high-quality H&E-IHC datasets. Overall, the main contributions of this paper are as follows:

- We propose HUSE, a universal virtual IHC multiplex staining framework featuring JMA, which tightly binds the generative evolution to the persistent structural scaffold of H&E. By enforcing robust pixel-level constraints within the manifold space, JMA ensures that the generation process is strictly anchored to the intrinsic tissue morphology, effectively suppressing manifold drift during high-fidelity translation across dozens of IHC markers.
- We introduce the Hi-MOE and RCG mechanism, leveraging pathological priors to build specialized expert branches for nuclei, stroma, and cytoplasm/background, and using RCG to dynamically monitor and refine representation conflicts in complex histological regions.
- We pioneer the mIF-Driven multi-marker IHC Generation paradigm, a novel cross-modal data synthesis approach that leverages the intrinsic in-situ alignment of multiplex immunofluorescence to construct large-scale, pixel-aligned $1(H\&E) : N(IHC)$ training matrices, whose high clinical realism has been rigorously verified by senior pathologists.

2 Related Work

2.1 Evolution of Virtual IHC Multiplex Staining

Early virtual IHC staining research was pioneered by "one-to-one" translation models, such as CycleGAN [33], CUT [23], and ASP [14]. These models focused on establishing independent mappings for specific markers in unpaired settings, validating the feasibility of synthesizing specific IHC markers from H&E images. However, as clinical diagnosis increasingly relies on multiplex protein expression analysis [13, 28], the "one-to-one" paradigm reveals its inherent inefficiency: training and deploying separate models for each indicator when facing dozens of target markers incurs prohibitive time costs and computational overhead.

Therefore, developing "one-to-many" universal IHC staining models, which utilize a single network to achieve precise generation of multiple markers, has become a core research focus. Recently emerged universal IHC staining models, such as VIMS [9] and PD-Unist [31], implement marker-conditioned generation control by introducing text prompts. Nevertheless, manually designed prompts have significant limitations when scaling to large-scale markers or deploying in automated workflows; furthermore, such methods typically rely on shared backbone, leading to a lack of sufficient specificity among different generated markers. Although Multi-IHC [11] attempts to alleviate this problem by assigning independent decoders for each marker, this scheme still faces severe computational overhead challenges when dealing with large-scale marker sets.

Overall, current universal models suffer from high computational costs or manual prompt engineering, complicating the balance between fidelity and specificity as the number of biomarker types increases.

2.2 High-Fidelity Synthesis via Pixel-Space Diffusion

Diffusion models have demonstrated exceptional performance in tasks such as image restoration and natural image generation [8, 17, 26]. Current mainstream research primarily utilizes conditional diffusion operating on Latent Diffusion Models (LDMs) [26], controlling the image generation process by fine-tuning denoising networks and incorporating text prompts or other auxiliary signals. However, according to manifold theory [5], traditional diffusion models predict physical quantities such as noise or velocity. These prediction targets often do not reside on the true data manifold, leading to manifold drift and potential generation failure when processing high-resolution raw pixels. Furthermore, the Variational Autoencoder (VAE) that LDMs depend on causes irreversible information loss during latent space compression, resulting in blurred or artifactual critical microscopic details in pathological images—such as nuclear boundaries and chromatin textures—making it difficult to meet the stringent high-fidelity requirements of virtual staining [26, 29].

Recently, JiT [15] and pMF [19] proposed a "back to basics" pixel-space generation paradigm by directly predicting clean data. This pixel-level diffusion model discards complex image tokenizers, enabling precise target guidance via

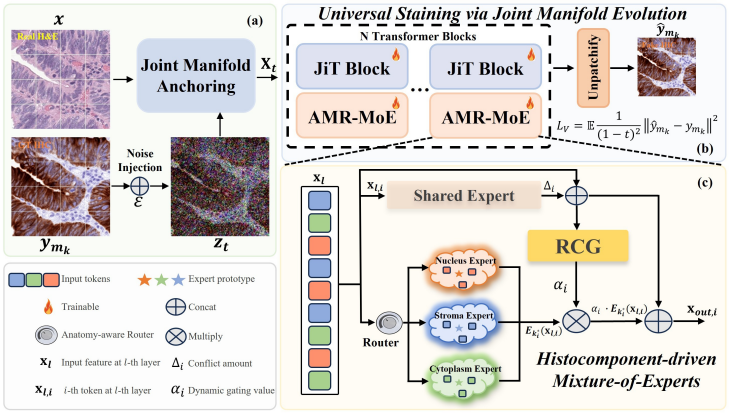


Fig. 2: Overview of our proposed HUSE. (a) Joint Manifold Anchoring treats the H&E image as a structural scaffold to anchor the evolution of noisy IHC features z_t. (b) The global architecture consists of N Transformer blocks for multiplex biomarker synthesis. (c) Histocomponent-driven Mixture-of-Experts employs specialized experts (Nucleus, Stroma, and Cytoplasm) coordinated by RCG to resolve structural ambiguities and representation conflicts.

the direct injection of indices as control signals. This architecture is inherently suited for pathology as it avoids latent-space compression loss, mapping directly to the manifold to preserve microscopic histological details. In multiplex virtual staining tasks involving various markers, this pixel-level paradigm not only improves the fidelity of synthesized images but also provides a more reliable representational foundation for processing complex biomarker features.

3 Method

In this section, we present HUSE, a universal virtual staining framework designed to synthesize high-fidelity multiplex IHC images from a single H&E slide. Our approach anchors the generation process to a pixel-level evolution path via Joint Manifold Anchoring, thereby effectively suppressing manifold drift during the sampling process. The core architecture incorporates Hi-MOE and RCG as a localized refinement scheme, working synergistically to rectify representation conflicts in complex tissue regions. The overall architecture of HUSE is depicted in Fig. 2. In the following, we first detail the network architecture and learning objectives, and finally elaborate on our pioneering mIF-driven data synthesis paradigm, which overcomes traditional alignment limitations and provides a new data pathway for large-scale universal staining research.

3.1 Universal Staining via Joint Manifold Evolution

Unlike traditional denoising diffusion models that indirectly recover images by predicting noise ϵ, we adopt the direct data prediction (x-prediction) paradigm

proposed by JiT [15]. This approach is formulated as an evolution path driven by flow matching. Specifically, we define our model as f_θ and consider a dataset $\mathcal{D} = \{(x^i, y^i_{m_1}, \ldots, y^i_{m_K})\}_{i=1}^N$, where x^i and $y^i_{m_k}$ denote the i-th H&E image and its corresponding k-th IHC marker image, respectively. Our objective is to learn a mapping that predicts the virtual staining result $\hat{y}_{m_k}$ for any target marker m_k. By directly reconstructing the clean data in the pixel space, the model effectively preserves the manifold characteristics of pathological images and retains fine subcellular structures that are often lost in discrete feature spaces.

Marker-Specific Pixel-level Reconstruction. Given a target IHC image $y_{m_k} \in \mathcal{Y}$ and the initial noise $\epsilon \sim \mathcal{N}(0, \mathbf{I})$, the noisy state $\mathbf{z}_t$ at time $t \in [0, 1]$ is defined as a linear interpolation between the target image and the noise: $\mathbf{z}_t = t y_{m_k} + (1-t)\epsilon$. To guide the model in accurately selecting the generation path within the multi-marker manifold, we utilize a class label embedder to map the marker index L_k into several learnable tokens e_k. This token is injected into specific transformer layers as a crucial semantic condition, ensuring that the generation process possesses specificity for the requested marker.

Joint Manifold Anchoring. To anchor the evolution to the underlying tissue structure, the H&E image $x \in \mathbb{R}^{H\times W\times 3}$ is treated as an invariant structural scaffold throughout the generation process. At the input stage, x is integrated with $\mathbf{z}_t \in \mathbb{R}^{H\times W\times 3}$ (obtained by adding noise to the target IHC image y_{m_k}) via channel-wise feature coupling to form a universal evolution input $\mathbf{X}_t$:

$$\mathbf{X}_t = [x \parallel \mathbf{z}_t] \in \mathbb{R}^{H\times W\times 6}, \tag{1}$$

where $\parallel$ denotes the concatenation operation along the channel dimension. The predicted virtual staining result $\hat{y}_{m_k}$ is then obtained by extracting the last three channels of the model output through a spatial slicing operation. This design ensures that the synthesis is strictly anchored within the topological framework provided by x, effectively suppressing manifold drift.

Optimization and Inference Process. During the training phase, we adopt an x-prediction-based-velocity loss to optimize the model [15]. The model directly predicts the target image $\hat{y}_{m_k} = f_\theta(\mathbf{X}_t, t, e_k)$. The corresponding predicted velocity field is parameterized as $v_{pred} = \hat{y}_{m_k} - \mathbf{z}_t / 1 - t$. The model is optimized by minimizing the mean squared error between the predicted velocity and the target velocity $v_{target} = y_{m_k} - \epsilon$:

$$\mathcal{L}_V = \mathbb{E}\left\|v_{pred} - v_{target}\right\|^2 = \mathbb{E}\frac{1}{(1-t)^2}\left\|\hat{y}_{m_k} - y_{m_k}\right\|^2. \tag{2}$$

During the inference phase, the sampling process originates from a random noise $\mathbf{z}_0 \sim \mathcal{N}(0, \mathbf{I})$. After coupling this noise with the structural scaffold x, the model predicts the target image at each time step conditioned on the biomarker label L_k. Subsequently, an ODE solver (e.g., Euler method) is employed to integrate along the predicted velocity field, ultimately synthesizing virtual IHC images with biomarker specificity within the topological framework of x.

3.2 Histocomponent-driven Mixture-of-Experts

To address the challenge where homogenized parameters in "One-to-Many" generation tasks struggle to capture the distinct characteristics of different tissue components, we introduce the Histocomponent-driven Mixture-of-Experts (HiMoE) mechanism into the virtual staining framework. This mechanism adheres to the design philosophy of "global structural alignment + local marker-specific compensation": all tokens first pass through a shared expert network to maintain the fundamental histological scaffold, followed by the Representation Conflict Gating (RCG) mechanism to identify the limitations of shared features in reconstructing specific details and dynamically incorporate specialized histocomponent experts for precise refinement.

Histocomponent-driven Router. To endow the expert assignment with explicit biological interpretability, we construct a set of learnable histocomponent prototypes $\mathbf{P} \in \mathbb{R}^{K\times D}$ ($K = 3$), corresponding to the nucleus, stroma, and cytoplasm/background, respectively. The prototypes $\mathbf{P}$ are initialized using the corresponding biological conceptual features extracted by a pre-trained CLIP model (ViT-B/32 version, trained by OpenAI on 400 million image-text pairs), ensuring that the expert responsibilities are histocomponent-constrained from the initialization stage. For the input features $\mathbf{x}_l \in \mathbb{R}^{N\times D}$ at the l-th layer, the router performs hard assignment by calculating the cosine similarity between each token and the histocomponent prototypes:

$$k_i^* = \arg\max_{k\in\{1,\dots,K\}} \frac{\mathbf{x}_{l,i}\cdot \mathbf{p}_k^\top}{\|\mathbf{x}_{l,i}\|\|\mathbf{p}_k\|\cdot\tau}, \quad i = 1,\dots,N, \tag{3}$$

where τ is the temperature factor. This histocomponent-driven dispatching ensures that each expert network can focus on modeling the marker-specific distributions of its corresponding biological structure.

Representation Conflict Gating. In universal models with shared parameters, a single set of weights often struggles to distinguish the differential expressions of various biomarkers (e.g., CD3 and panCK) at the same histocomponent location, leading to homogenized generation results that lack distinctiveness. To address this, we utilize the RCG mechanism as a localized refinement scheme to dynamically incorporate expert knowledge for enhanced marker specificity at the token level. Specifically, all tokens first enter a shared expert E_{shared} to extract general morphological features, and the resulting residual amount for each token i is defined as the conflict amount $\Delta_i = E_{shared}(\mathbf{x}_{l,i})$. Δ_i represents the state displacement of the i-th token's features after processing through the shared expert, reflecting the evolutionary intensity of the shared parameters attempting to map the current morphology to the target expression. A significant deviation between Δ_i and the original feature $\mathbf{x}_{l,i}$ indicates that general shared knowledge cannot effectively cover the specific details required for the target biomarker.

The RCG predictor calculates a token-specific dynamic gating value α_i by observing the concatenated state of $\mathbf{x}_{l,i}$ and Δ_i:

$$\alpha_i = \sigma(\mathbf{W}_{rcg}[\mathrm{sg}(\mathbf{x}_{l,i}); \mathrm{sg}(\Delta_i)] + b_{rcg}), \tag{4}$$

where σ denotes the Sigmoid activation function; $\mathbf{W}_{rcg}$ and b_{rcg} are the learnable parameters of the predictor; and $\mathrm{sg}(\cdot)$ represents the stop-gradient operation, ensuring the gate acts as an objective observer without interfering with the backbone logic. A larger conflict amount triggers a more active response from α_i, thereby introducing more specialized knowledge for compensation. Finally, the output of the i-th token in the l-th block is expressed as:

$$\mathbf{x}_{out,i} = \mathbf{x}_{l,i} + E_{shared}(\mathbf{x}_{l,i}) + \alpha_i \cdot E_{k_i^*}(\mathbf{x}_{l,i}), \tag{5}$$

where each token i is independently assigned both an optimal expert k_i^* and an adaptive gating weight α_i, ensuring that the most relevant histocomponent knowledge is applied with fine-grained intensity for detail enhancement. Through the dynamic adjustment of RCG, the model effectively mitigates the issue of insufficient distinctiveness caused by parameter sharing, significantly improving the marker specificity across different tissue components.

3.3 mIF-Driven multi-marker IHC Generation

Achieving high-performance and high-efficiency universal models necessitates the support of large-scale 1(H&E):N(IHC) training data, which existing public IHC datasets are unable to provide. To bridge this gap, we pioneeringly propose a mIF-driven multi-marker IHC generation paradigm as the foundational support for large-scale training (As shown in Figure. 1b). Specifically, given a set of pixel-wise aligned H&E-mIF paired images, where $I_{mIF} \in \mathbb{R}^{H\times W\times n}$ contains n biomarker channels. To transform I_{mIF} into multi-marker IHC images, we implemented cross-modal texture modulation. First, we extract the hematoxylin channel concentration map C_H from I_{HE} via color deconvolution to capture tissue structures [27]. Next, for the k-th biomarker channel, its normalized signal intensity $C_{IF,k}$ is coupled with the structural details from I_{HE} in a physics-driven to calculate the concentration distribution of the synthetic DAB channel:

$$C_{DAB,k} = C_{IF,k} \cdot (1 + \alpha \cdot \hat{C}_H) \cdot \beta, \tag{6}$$

where α is the modulation coefficient, $\hat{C}_H$ is the normalized texture feature of the H-channel, and β is the intensity scaling factor. We set $\alpha = 0.6$ and $\beta = 0.8$ in our implementation. Finally, we perform color space recomposition of the original C_H channel and the generated $C_{DAB,k}$ channel to reconstruct the virtual IHC image y_k that conforms to clinical visual habit [27]. Through the aforementioned process, we constructed a high-quality dataset $\mathcal{D} = \{(x^i, y^i_{m_1}, \ldots, y^i_{m_K})\}_{i=1}^N$ containing N samples, where x^i is the input H&E image and $y^i_{m_k}$ is the corresponding k-th virtual IHC target. This automated workflow facilitates the rapid generation of an extensive volume of precise pixel-level multi-marker IHC under a single H&E index.

Table 1: Quantitative comparison of various methods on Orion-CRC dataset. The best values are highlighted in bold and the suboptimal results in underline. Rows shaded in light gray represent Single-marker Staining Models.

Method	Metric	CD3e	CD4	CD8a	CD20	CD31	CD45	CD45RO	CD68	CD163	E-cad.	FOXP3	Hoechst	Ki67	Pan-CK	PD-L1	SMA	AVG
ASP	SSIM↑	0.840	0.784	0.869	0.868	0.852	**0.718**	0.869	0.841	0.835	0.678	0.889	0.752	0.875	0.664	0.869	**0.699**	0.806
	PSNR↑	22.05	19.45	26.22	27.50	25.25	**16.68**	**26.83**	20.77	20.76	**20.74**	28.80	17.98	**26.43**	19.48	**27.32**	**17.23**	22.72
	KID↓	18.19	42.36	15.98	16.33	14.61	16.76	16.70	18.26	18.79	**31.36**	13.89	38.35	18.51	25.41	18.98	71.52	24.75
PSPStain	SSIM↑	0.849	**0.809**	0.876	0.837	0.792	0.711	0.866	**0.849**	**0.858**	0.652	0.901	0.750	0.878	**0.737**	0.663	0.682	0.794
	PSNR↑	**22.16**	**19.73**	25.46	22.07	20.33	16.60	26.20	**21.39**	**21.92**	19.54	29.06	**18.02**	25.99	**20.21**	19.63	16.43	21.55
	KID↓	**18.58**	**22.42**	33.89	16.27	36.19	12.53	16.64	19.88	28.75	35.70	18.41	39.58	16.91	**22.70**	28.76	65.81	27.06
USIGAN	SSIM↑	0.847	0.781	0.883	0.860	0.879	0.716	0.873	0.837	0.853	0.632	0.899	0.747	0.887	0.661	0.877	0.686	0.807
	PSNR↑	22.12	19.29	26.68	27.24	25.64	16.47	26.17	20.43	20.79	19.46	28.76	17.17	26.14	19.55	25.82	16.28	22.38
	KID↓	26.81	26.81	42.13	14.63	26.74	18.99	16.97	33.81	33.26	34.76	14.85	38.23	25.36	30.37	22.68	81.14	30.47
Multi-IHC	SSIM↑	0.857	0.789	0.903	0.917	0.908	0.676	0.861	0.819	0.826	0.682	0.905	0.746	0.898	0.693	0.853	0.675	0.813
	PSNR↑	21.98	18.58	25.68	28.18	24.75	15.84	26.28	19.97	21.01	20.28	30.03	17.01	26.07	19.81	25.81	16.57	22.37
	KID↓	19.00	35.47	16.01	16.63	22.54	18.19	36.71	121.61	22.91	32.34	15.35	38.00	16.72	23.33	16.44	122.66	35.87
VIMs	SSIM↑	0.547	0.526	0.561	0.556	0.556	0.493	0.557	0.541	0.546	0.428	0.564	0.472	0.554	0.436	0.561	0.461	0.522
	PSNR↑	21.15	19.19	23.60	23.79	22.86	16.06	23.32	20.91	20.45	17.32	24.21	15.05	22.47	17.32	23.60	16.81	20.51
	KID↓	204.74	215.31	209.55	224.72	211.82	133.70	217.22	225.39	222.40	101.50	210.20	79.45	218.32	73.23	195.56	155.28	181.15
PDUnist	SSIM↑	0.829	0.783	0.885	0.889	0.899	0.708	0.882	0.837	0.840	0.667	0.915	0.740	0.889	0.663	0.855	0.679	0.810
	PSNR↑	21.63	19.16	26.29	27.84	25.09	16.22	25.47	20.76	20.64	20.29	29.57	16.43	25.47	19.73	25.60	16.33	22.28
	KID↓	24.79	34.04	18.01	17.62	42.52	24.25	31.37	100.76	63.83	33.69	19.66	37.71	18.51	32.93	20.20	126.01	40.37
Ours	SSIM↑	**0.866**	0.791	**0.919**	**0.920**	**0.909**	0.713	**0.890**	0.842	0.855	**0.690**	**0.936**	**0.758**	**0.905**	0.700	**0.880**	0.693	**0.829**
	PSNR↑	22.13	19.36	**26.56**	**28.65**	**25.59**	16.47	26.54	20.94	21.07	20.46	**30.17**	17.64	26.21	19.95	26.26	16.91	**22.81**
	KID↓	18.22	23.61	**15.98**	**14.32**	**14.38**	**18.04**	**16.63**	**19.31**	**18.98**	31.64	**14.01**	**36.79**	**14.37**	23.07	**15.64**	**23.94**	**19.93**

4 Experiments

4.1 Experimental Setup

Dataset. We evaluate our method on two different types of datasets: Orion-CRC [16] and MIST [14]. **(1) Orion-CRC** contains 41 Colorectal Cancer (CRC) Whole Slide Images (WSIs), achieving in-situ pixel-level alignment between H&E images and 18-channel multiplexed Immunofluorescence (mIF) through same-slide restaining technology. In this study, we use a version of the data standardized by [2]. This version selects and retains 16 core marker channels as model prediction targets, with autofluorescence subtraction and normalization applied to the original signals. All WSIs are cropped into 256×256 pixel patches, totaling approximately 320,000 patches. We perform random sampling at a 10:1 ratio from the original data to construct an experimental subset of 32,000 images, where 28,000 images are used for training, and 2,000 each for validation and testing. We apply the proposed mIF-driven multi-marker IHC generation paradigm to construct a 1 (H&E) to 16 (IHC) training matrix. **(2) MIST** is a real IHC dataset providing four biomarkers (ER, PR, Ki67, HER2) with 4,153, 4,139, 4,361, and 4,642 training pairs, and 1,000 test pairs each (all 1024×1024). In our experiments, poorly aligned MIST training samples are removed, yielding 1,000 training and 100 test pairs per biomarker (all resolutions are resized to 512×512).

Evaluation Metrics. We employ Peak Signal-to-Noise Ratio (PSNR/dB), Structural Similarity Index Measure (SSIM) and Kernel Inception Distance (KID) to

Table 2: Quantitative comparison of various methods on MIST dataset. The best values are highlighted in bold and the suboptimal results in underline. Rows shaded in light gray represent Single-marker Staining Models.

Method	SSIM↑					PSNR↑					KID↓				
	ER	HER2	Ki67	PR	AVG	ER	HER2	Ki67	PR	AVG	ER	HER2	Ki67	PR	AVG
ASP	0.201	0.126	0.201	0.147	0.169	13.23	12.85	14.58	12.91	13.39	33.50	**89.40**	69.84	31.43	56.04
PSPStain	0.201	0.125	0.212	0.158	0.174	12.77	12.58	14.76	12.92	13.26	**32.39**	90.40	79.98	**26.29**	57.27
USIGAN	0.201	0.127	0.185	0.173	0.172	12.81	13.15	14.62	13.15	13.43	33.44	91.85	77.80	37.90	60.25
Multi-IHC	0.383	0.171	0.262	0.209	0.256	16.78	14.65	15.70	14.21	15.34	177.57	137.79	281.90	211.06	202.08
VIMs	0.314	0.174	0.321	0.271	0.270	14.75	14.14	16.81	15.13	15.21	151.44	291.43	202.25	237.29	220.60
PDUnist	0.381	0.177	0.242	0.199	0.250	16.86	14.50	16.30	14.31	15.49	52.45	92.57	91.86	112.52	87.35
Ours	**0.386**	**0.186**	**0.451**	**0.419**	**0.361**	**17.94**	**14.68**	**21.12**	**19.66**	**18.35**	37.91	89.52	**58.69**	33.01	**54.78**

assess method performance. Specifically, the KID values reported in the tables are scaled by a factor of 1000.

Implementation Details. We utilize JiT-B/16 (256×256) for Orion-CRC and JiT-B/32 (512×512) for MIST [15], training for 400 epochs (5-epoch warmup) with an initial learning rate of 1×10^{-4} cosine-annealed to 1×10^{-5} after epoch 250. All experiments are implemented in PyTorch on a single NVIDIA GeForce RTX 3090 GPU. Parameter settings for comparison methods follow their original papers. Further details are provided in the **Supplementary Material**. Notably, our method requires no predefined hyperparameters.

4.2 Results

Compared with SOTA Methods. We compare our proposed HUSE with 6 SOTA methods on both datasets. These methods are categorized into: (1) Universal Staining Models, including VIMs [9], PD-UniST [31], and Multi-IHC Net [11], which, similar to HUSE, aim to simultaneously synthesize multiple IHC biomarkers via a single integrated architecture; and (2) Single-marker Staining Models, including ASP [14], PSPStain [7], and USIGAN [25], where an independent model is trained for each specific marker.

As shown in Table 1, HUSE achieves superior results on Orion-CRC, reaching SOTA average performance (PSNR: 22.81 dB, SSIM: 0.829, KID: 19.93). This superiority validates that anchoring the generative process to the intrinsic H&E scaffold via JMA is critical for maintaining morphological fidelity and suppressing manifold drift. Compared to previous universal models, HUSE surpasses them in all metrics by ensuring synthesized markers are strictly bound to the real structure. Notably, HUSE achieves comparable performance to the three specialized single-marker models. Even when certain metrics are second-best, the gap between HUSE and these top-performing models remains minimal (e.g., SSIM of -0.037 and PSNR of -0.45) . This outcome proves that the robust structural binding provided by JMA allows our single integrated network to match the professional precision of multiple independent models, while significantly reducing the computational overhead of training and deployment.

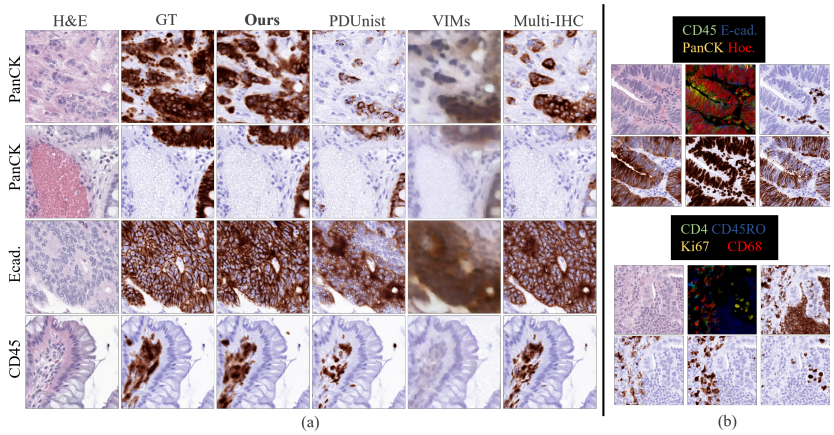


Fig. 3: (a) Qualitative comparison on Orion-CRC. (b) Transformation example of the mIF-Driven multi-marker IHC Generation paradigm on Orion-CRC.

On the MIST benchmark (Table 2), HUSE delivers SOTA results (averaging 18.35 dB PSNR, 0.361 SSIM, 54.78 KID), outperforming all existing universal models across all metrics. This further confirms the generalization of JMA's anchoring effect, which maintains a stable and strictly constrained manifold evolution across diverse datasets. Overall, HUSE showcases exceptional potential for high-fidelity virtual multiplex staining by leveraging structural coupling to bridge the gap between H&E and multi-marker IHC.

Overall, HUSE delivers superior performance on both datasets, showcasing its high-fidelity potential for virtual multiplex IHC staining.

Qualitative and Computation Comparisons. As shown in Fig. 3a, we qualitatively compare the staining translation results of HUSE with three universal models on the Orion-CRC dataset. The results demonstrate that our method achieves higher fidelity and more precise tissue morphology staining.

Furthermore, as shown in Table 6, HUSE maintains exceptional stability in terms of computational complexity (GFLOPs) and model parameters (Params). Unlike the three existing universal IHC staining models, where the computational load typically quadruples when scaling the input size from 256×256 to 512×512, HUSE benefits from the significant architectural efficiency of its JiT-based backbone. Leveraging its Transformer-based design, HUSE can adapt to various image resolutions by flexibly adjusting the input patch size. This ensures that model parameters and computational complexity remain invariant to changes in input dimensions. Such architectural efficiency enables HUSE to stably synthesize dozens of markers within a single integrated network while maintaining a consistent and manageable computational footprint.

Table 3: Ablation study of key components in HUSE: Joint Manifold Anchoring (JMA), Hi-MOE and RCG. "✓" means utilizing this blocks. Notably, in the version without JMA, the model treats the H&E image as a noisy initial input ($H\&E + \epsilon$) to initiate the flow towards the IHC target.

Module			Orion-CRC (AVG)			MIST (AVG)		
JMA	Hi-MOE	RCG	SSIM$^{\uparrow}$	PSNR$^{\uparrow}$	KID$^{\downarrow}$	SSIM$^{\uparrow}$	PSNR$^{\uparrow}$	KID$^{\downarrow}$
			0.449	17.40	90.58	0.28	16.43	165.63
	✓	✓	0.473	18.52	82.93	0.32	17.26	141.77
✓			0.818	22.41	23.33	0.31	17.20	99.84
✓	✓		0.823	22.69	20.26	0.34	18.21	97.37
✓	✓	✓	**0.829**	**22.81**	**19.33**	**0.36**	**18.35**	**54.78**

Table 4: Ablation study of Hi-MoE. "RI" denotes random initialization; "F" and "L" represent Frozen and Learnable expert parameters.

Method	Orion-CRC (AVG)		MIST (AVG)	
	PSNR$^{\uparrow}$	KID$^{\downarrow}$	PSNR$^{\uparrow}$	KID$^{\downarrow}$
Hi-MOE(RI)	20.69	23.16	14.87	155.78
Hi-MOE(F)	21.70	22.59	14.84	162.27
Hi-MOE(L)	**22.81**	**19.33**	**18.35**	**54.78**

Table 5: Ablation study of RCG with different activation functions σ to evaluate its effectiveness in regulating representation conflict resolution.

Method	Orion-CRC (AVG)		MIST (AVG)	
	PSNR$^{\uparrow}$	KID$^{\downarrow}$	PSNR$^{\uparrow}$	KID$^{\downarrow}$
Identity	20.72	32.84	18.09	80.76
Tanh	21.54	24.67	18.18	71.81
Sigmoid	**22.81**	**19.33**	**18.35**	**54.78**

4.3 Ablation Study

Key Components in HUSE. To evaluate the effectiveness of key components in the proposed ts-DC, we test different combinations and report the performance in Table 3. Specifically, the Joint Manifold Anchoring (JMA) serves as the structural foundation; by anchoring the generation process to the intrinsic H&E scaffold, it significantly suppresses manifold drift and ensures spatial consistency (+0.369 SSIM). Building upon this global scaffold, the introduced Hi-MoE mechanism specializes in heterogeneous histological regions, thereby capturing fine-grained morphological nuances and enhancing marker specificity (+1.01dB PSNR). Finally, the RCG module resolves representation conflicts in complex tissue intersections, providing a stable refinement for multi-marker synthesis (-42.59 KID). This improvement demonstrates the effectiveness of HUSE.

Expert Initialization and Update Strategy. We compare different initialization and update strategies for the histocomponent prototypes $\mathbf{P}$ in Table 4. Results demonstrate that random initialization (RI) of expert parameters causes the model to lack a clear global functional direction, leading to a blind search during training and resulting in ambiguous expert specialization. While fixed priors (F) provide a biological starting point, they also indicate that general biological concepts still require explicit alignment with the specific pathological domain. Therefore, endowing prototypes with learnability (L) is crucial, as it enables experts to adapt to the morphological distribution of specific data, achieving the most effective localized refinement.

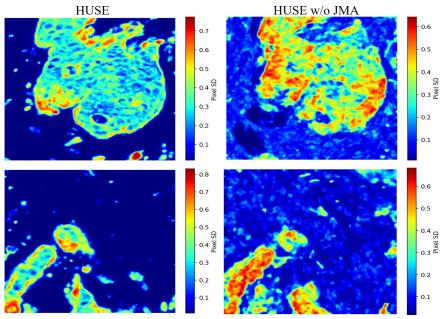


Fig. 4: Visualization of JMA Ensuring Structural Stability.

Table 6: Comparison of computational complexity (GFLOPs) and model parameters (Params) across four universal IHC staining models. B and L represent inputs at resolutions of 256×256 and 512×512, respectively.

Method	GFlops	Params(M)
PDUnist(B,L)	(114.6, 455.7)	(99.3, 99.3)
Multi-IHC(B,L)	(993.2, 3972.0)	(86.7, 86.7)
VIMs(B,L)	(555.2, 2147)	(1291, 1291)
HUSE(B,L)	(141.8, 142.9)	(301.8, 303.9)

Ablation of RCG Activation. Table 5 shows that Sigmoid outperforms Tanh and Identity. Its advantage lies in providing stable $[0, 1]$ normalized gating, which precisely regulates expert compensation while preserving backbone stability for effective localized refinement.

Pathologists Validation. To evaluate clinical realism and practical significance, three senior pathologists from a top-tire hospital of our team conducts an independent blind assessment of our mIF-driven paradigm using the Orion-CRC dataset. The study includes 16 distinct biomarkers, with at least 10 samples for each marker, totaling over 160 synthesized images. Scoring uses a 5-point scale across three clinical dimensions: 1) *Biological Localization* (signal accuracy), 2) *Texture Realism* (DAB graininess and precipitation), and 3) *Intensity Mapping* (fluorescence-to-brightfield transition). (Further details on the dimension descriptions and scoring rules are provided in the **Supplementary Material.**)

The assessment is exceptional: **100%** of the evaluated samples across all 16 markers receives scores of either 4 or 5 points, with an average score of **4.91**. The pathologists note that the synthesized IHC images are virtually indistinguishable from authentic staining in terms of both structural logic and visual texture. This outcome rigorously validates that our mIF-driven synthesis paradigm is capable of generating high-fidelity IHC data that meets clinical standards. It further underscores the substantial practical value of this approach, laying a solid foundation for the development of more robust universal IHC staining models.

Analysis of Joint Manifold Anchoring. We calculate pixel-wise standard deviation (SD) heatmaps in the Orion-CRC dataset, generated from the same H&E input across 10 random seeds. As shown in Fig.4, the left image (Ours) exhibits lower variance (deep blue), demonstrating that JMA firmly anchors the generation process to the H&E structure, achieving exceptional morphological stability. The right image (w/o JMA) shows significant high-variance regions (warm tones), indicating that the lack of JMA constraints leads to manifold drift during the generation process, causing morphological uncertainty.

5 Conclusion

In this paper, we propose the HUSE framework, which effectively addresses the challenges of manifold drift and tissue heterogeneity modeling in virtual staining through JMA, Hi-MoE, and RCG mechanisms. Furthermore, this study pioneers an mIF-driven multi-marker generation paradigm. The model achieves state-of-the-art performance on multiple datasets, and pathologists evaluations confirm its exceptional clinical realism and structural fidelity, providing a reliable solution for high-fidelity virtual staining.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 62471182), Shanghai Rising-Star Program (Grant No. 24QA2702100), Science and Technology Commission of Shanghai Municipality Basic Research Program (Grant No. 25JD1401300), and the Science and Technology Commission of Shanghai Municipality (Grant No. 22DZ2229004)

References

1. Alfasly, S., Uegami, W., Hoq, M., Alabtah, G., Tizhoosh, H.: Semantic and visual crop-guided diffusion models for heterogeneous tissue synthesis in histopathology. arXiv preprint arXiv:2509.17847 (2025)
2. Balezo, G., Trullo, R., Planas, A.P., Decenciere, E., Walter, T.: Miphei-vit: Multiplex immunofluorescence prediction from h&e images using vit foundation models. arXiv preprint arXiv:2505.10294 (2025)
3. Bian, C., Phillips, B., Cootes, T., Fergie, M.: Hemit: H&e to multiplex-immunohistochemistry image translation with dual-branch pix2pix generator. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 184–197. Springer (2024)
4. Carlsson, G.: Topology and data. Bulletin of the American Mathematical Society **46**(2), 255–308 (2009)
5. Chapelle, O., Schölkopf, B., Zien, A. (eds.): Semi-Supervised Learning. MIT Press, Cambridge, MA (2006)
6. Chen, F., Zhang, R., Hu, W., Abera, D.E., Peng, Y., Zheng, B., Sun, Y., Cai, J., Qin, W.: Pgvms: A prompt-guided unified framework for virtual multiplex ihc staining with pathological semantic learning. arXiv preprint arXiv:2602.23292 (2026)
7. Chen, F., Zhang, R., Zheng, B., Sun, Y., He, J., Qin, W.: Pathological semantics-preserving learning for h&e-to-ihc virtual staining. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 384–394. Springer (2024)
8. Dhariwal, P., Nichol, A.: Diffusion models beat gans on image synthesis. Advances in neural information processing systems **34**, 8780–8794 (2021)
9. Dubey, S., Chong, Y., Knudsen, B., Elhabian, S.Y.: Vims: virtual immunohistochemistry multiplex staining via text-to-stain diffusion trained on uniplex stains. In: International Workshop on Machine Learning in Medical Imaging. pp. 143–155. Springer (2024)

10. Fedchenko, N., Reifenrath, J.: Different approaches for interpretation and reporting of immunohistochemistry analysis results in the bone tissue–a review. Diagnostic pathology **9**(1), 221 (2014)
11. Guan, X., Zhang, Z., Wang, Y., Zhang, Y., Jiang, D., Zhang, Y.: Correlated multiple ihc virtual staining for breast histopathological images. In: ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 1–5. IEEE (2025)
12. de Haan, K., Zhang, Y., Zuckerman, J.E., Liu, T., Sisk, A.E., Diaz, M.F., Jen, K.Y., Nobori, A., Liou, S., Zhang, S., et al.: Deep learning-based transformation of h&e stained tissues into special stains. Nature Communications **12**(1), 4884 (2021)
13. Harms, P.W., Frankel, T.L., Moutafi, M., Rao, A., Rimm, D.L., Taube, J.M., Thomas, D., Chan, M.P., Pantanowitz, L.: Multiplex immunohistochemistry and immunofluorescence: a practical update for pathologists. Modern Pathology **36**(7), 100197 (2023)
14. 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: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 632–641. Springer (2023)
15. Li, T., He, K.: Back to basics: Let denoising generative models denoise. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 36115–36125 (2026)
16. Lin, J.R., Chen, Y.A., Campton, D., Cooper, J., Coy, S., Yapp, C., Tefft, J.B., McCarty, E., Ligon, K.L., Rodig, S.J., et al.: High-plex immunofluorescence imaging and traditional histology of the same tissue section for discovering image-based biomarkers. Nature cancer **4**(7), 1036–1052 (2023)
17. Liu, J., Xu, S., Yang, Q., Wang, Y., Chen, X., Ji, Z.: Fape-ir: Frequency-aware planning and execution framework for all-in-one image restoration. arXiv preprint arXiv:2511.14099 (2025)
18. 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)
19. Lu, Y., Lu, S., Sun, Q., Zhao, H., Jiang, Z., Wang, X., Li, T., Geng, Z., He, K.: One-step latent-free image generation with pixel mean flows. arXiv preprint arXiv:2601.22158 (2026)
20. Nasir, E.S., Elhaminia, B., Eastwood, M., King, C., Cain, O., Harper, L., Moss, P., Chanouzas, D., Snead, D., Rajpoot, N., et al.: Core–a cell-level coarse-to-fine image registration engine for multi-stain image alignment. arXiv preprint arXiv:2511.03826 (2025)
21. Nawaz, S., Heindl, A., Koelble, K., Yuan, Y.: Beyond immune density: critical role of spatial heterogeneity in estrogen receptor-negative breast cancer. Modern Pathology **28**(6), 766–777 (2015)
22. Ounissi, M., Sarbout, I., Hugot, J.P., Martinez-Vinson, C., Berrebi, D., Racoceanu, D.: Scalable, trustworthy generative model for virtual multi-staining from h&e whole slide images. PLOS Computational Biology **21**(10), e1013516 (2025)
23. Park, T., Efros, A.A., Zhang, R., Zhu, J.Y.: Contrastive learning for unpaired image-to-image translation. In: European conference on computer vision. pp. 319–345. Springer (2020)

24. Pati, P., Karkampouna, S., Bonollo, F., Compérat, E., Radić, M., Spahn, M., Martinelli, A., Wartenberg, M., Kruithof-de Julio, M., Rapsomaniki, M.: Accelerating histopathology workflows with generative ai-based virtually multiplexed tumour profiling. Nature machine intelligence **6**(9), 1077–1093 (2024)
25. Peng, Y., Xiong, B., Chen, F., Eybo, D., Zhang, R., Hu, W., Cai, J., Qin, W.: Usigan: Unbalanced self-information feature transport for weakly paired image ihc virtual staining. arXiv preprint arXiv:2507.05843 (2025)
26. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 10684–10695 (2022)
27. Ruifrok, A.C., Johnston, D.A., et al.: Quantification of histochemical staining by color deconvolution. Analytical and Quantitative Cytology and Histology **23**(4), 291–299 (2001)
28. Tan, W.C.C., Nerurkar, S.N., Cai, H.Y., Ng, H.H.M., Wu, D., Wee, Y.T.F., Lim, J.C.T., Yeong, J., Lim, T.K.H.: Overview of multiplex immunohistochemistry/immunofluorescence techniques in the era of cancer immunotherapy. Cancer Communications **40**(4), 135–153 (2020)
29. Wei, D., Hu, Y., Peng, Q., Liu, H., Huang, Y., Wu, X., Zheng, Y., Wang, L., et al.: D-vst: Diffusion transformer for pathology-correct tone-controllable cross-dye virtual staining of whole slide images. Advances in Neural Information Processing Systems **38**, 3128–3162 (2026)
30. Xu, S., Song, H., Deng, Y., Wang, J., Wang, Y., Li, Q.: Advancing stain transfer for multi-biomarkers: A human annotation-free method based on auxiliary task supervision. In: Proceedings of the 34th International Joint Conference on Artificial Intelligence, IJCAI 2025. pp. 2116–2124. International Joint Conferences on Artificial Intelligence (2025)
31. Zhang, C., Xie, Y., Li, Y., Liang, X., Lin, L., Chen, Y.W.: Pd-unist: Prompt-driven universal model for unpaired h&e-to-ihc stain translation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 463–472. Springer (2025)
32. 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)
33. 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)