

VERSATILE STAIN TRANSFER IN HISTOPATHOLOGY USING A UNIFIED DIFFUSION FRAMEWORK

Xudong Yan¹ Mingze Yuan¹ Yao Lu² Ying Zhang¹ Zifan Chen¹ Peng Bao¹
Zehua Li² Bin Dong¹ Li Yang² Li Zhang^{1*} Fangxu Zhou^{1*}

¹ Peking University, Beijing, China

² Peking University First Hospital, Beijing, China

ABSTRACT

Histological staining is vital in clinical pathology for visualizing tissue structures. However, these techniques are laborious and time-consuming. Digital virtual staining offers a promising solution, but existing methods typically rely on Generative Adversarial Networks (GANs), which may suffer from artifacts and mode collapse. Motivated by the success of diffusion models, we present DUST, a novel **Diffusion-based Unified** framework for versatile **Stain Transfer** in histopathology. To enhance domain awareness and task-specific performance, we propose a dual encoding strategy that integrates the stain types of both the source and target domains. Additionally, we introduce a dynamic dual-output head to address the unstable intensity issue encountered with conventional DDPM implementations. Validated on a curated four-stain kidney histopathological dataset (H&E, MT, PAS, and PASM), DUST demonstrates superior versatile stain transfer capabilities. Our research highlights the potential of diffusion models to advance virtual staining, paving the way for more efficient digital pathology analyses.

Index Terms— Virtual staining, Versatile stain transfer, Diffusion models, Kidney histopathology

1. INTRODUCTION

Histological staining is crucial for visualizing tissue structures and cellular morphology in clinical pathology. Common stains like hematoxylin and eosin (H&E) provide essential contrast between nuclei and extracellular matrices. Special stains such as Masson’s trichrome (MT), Periodic Acid-Schiff (PAS), and Periodic Acid-silver Methenamine (PASM) are used to highlight collagen fibers, glycoproteins, and basement membranes, respectively. Despite their significance, these staining processes are time-consuming and labor-intensive, involving extensive sample preparation and manual execution by skilled histotechnicians. Moreover, given that the same section cannot be repeatedly stained, the necessity for multiple tissue sections for different stains poses challenges, especially in resource-constrained environments, hindering accessibility and increasing costs.

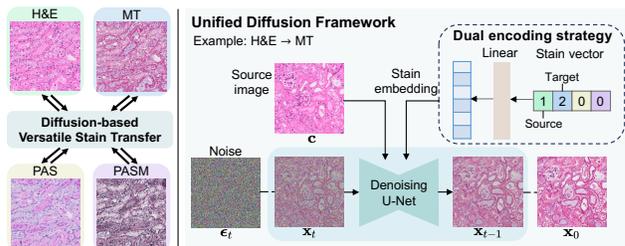


Fig. 1. Framework Overview. Left: The versatile stain transfer problem, *i.e.*, conversion across four stain domains: H&E, MT, PAS, and PASM. Right: Our proposed unified diffusion-based framework, which incorporates the source image and stain information as additional conditions. For stain conditioning, we introduce a dual encoding strategy that integrates information from both the source and target domains.

Recently, the advent of digital virtual staining has emerged as a transformative approach, enabling the conversion of one stain to another through deep learning [1]. Previous research has predominantly focused on developing specialized models to transfer H&E stains to one [2] or a few [1, 3] special stains, largely due to the prevalence of H&E staining. Yet, the concept of a unified framework that facilitates versatile stain conversion, *i.e.*, translating any given stain to any desired stain (Figure 1 left), remains relatively unexplored. Developing such a framework is advantageous for two key reasons. First, despite H&E’s affordability, it has its limitations in providing detailed surface and color contrast, making the translation from special stains beneficial for capturing complementary details [1]. Second, the proven synergy of generalist models in both natural [4] and medical imaging [5] underscores the potential of a unified framework to surpass the capabilities of specialized models. By enabling efficient broad-spectrum stain conversion, this framework allows pathologists to instantly switch between different stain types within their existing workflows, thus supporting swift, effective, and high-throughput pathological analysis.

Furthermore, traditional stain transfer techniques [3, 6, 1, 7, 2] often rely on Generative Adversarial Networks

(GANs) [8, 9, 10, 11], which can suffer from issues like artifacts and mode collapse. Additionally, unpaired translation methods [6, 3] lead to suboptimal results due to their inability to leverage pixel-wise correspondence, indispensable for detailed pathological images [7]. Fortunately, emerging diffusion models [12] have beat GANs in terms of stability and image quality across natural image generation [13, 14, 15], denoising, and translation [16] tasks, providing a promising alternative.

In this paper, we propose DUST, a novel **D**iffusion-based **U**nified framework designed for versatile **S**tain **T**ransfer in histopathology to address these limitations. Our method integrates the source image and the stain types of both the source and target domains as additional conditions to enable multi-task learning. Unlike the commonly used one-hot encoding [10, 6, 3] or adaptive normalization strategies [17, 18], we propose a dual encoding strategy to integrate the stain types of both source and target domains, thereby enhancing its task-specific performance. Furthermore, we addressed the issue of unstable intensity encountered with conventional Denoising Diffusion Probabilistic Models (DDPM) implementations [12] by introducing a dynamic dual-output head [19]. Validated on our curated kidney histopathological four-stain (H&E, MT, PAS, and PASM) dataset, our method demonstrates superior performance on versatile stain transfer in the proof-of-concept transfer cycle. Our study underscores the potential of diffusion models in advancing the field of virtual staining.

2. METHOD

2.1. Unified Diffusion Framework

Problem Formulation. Versatile stain transfer can be formulated as a multi-domain image-to-image translation problem. Denote the set of histopathological stain domains with K stains as $\mathcal{S} = \{1, 2, \dots, K\}$. Given any source-target pair of stains $\{s_0, s_1\} \subset \mathcal{S}$ and a source image \mathbf{x} with stain s_0 , we aim to apply a unified framework that can transfer the style of \mathbf{x} from domain s_0 to the desired domain s_1 , while preserving the content of \mathbf{x} . The entire dataset is denoted as $\mathcal{D} = \{(\mathbf{x}_1^{(i)}, \mathbf{x}_2^{(i)}, \dots, \mathbf{x}_K^{(i)}) \mid i = 1, 2, \dots, N\}$, where $\mathbf{x}_j^{(i)} \in \mathbb{R}^{3 \times H \times W}$ represents the i -th pathological slice with stain j . Here, we have $K = 4$ corresponding to the stains H&E, MT, PAS, and PASM.

Preliminaries. We briefly review key concepts of diffusion models [12]. Diffusion models initiate with a forward diffusion process that gradually adds Gaussian noise to an initial image \mathbf{x}_0 , i.e., $q(\mathbf{x}_t | \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \sqrt{\bar{\alpha}_t} \mathbf{x}_0, (1 - \bar{\alpha}_t) \mathbf{I})$, with $\sqrt{\bar{\alpha}_t}$ being hyperparameters. This process is reparameterized as $\mathbf{x}_t = \sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon_t$, where $\epsilon_t \sim \mathcal{N}(0, \mathbf{I})$ represents the sampled noise. Then \mathbf{x}_0 can be easily backtraced via:

$$\mathbf{x}_0 = \frac{1}{\sqrt{\bar{\alpha}_t}} (\mathbf{x}_t - \sqrt{1 - \bar{\alpha}_t} \epsilon_t). \quad (1)$$

The reverse diffusion process, which reconstructs the original data by reversing the noise, is approximated by $p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t))$. Training optimizes the variational lower bound on the log-likelihood of the initial data, simplified to $\mathcal{L}_\theta = -\log p_\theta(\mathbf{x}_0 | \mathbf{x}_1) + \sum_t \mathcal{D}_{KL}(q(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{x}_0) \| p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t))$. By characterizing the mean $\boldsymbol{\mu}_\theta$ as a noise prediction network ϵ_θ and fixing the variance [12], the training process simplifies to minimizing the mean-squared error between the predicted and actual noise, i.e., $\mathcal{L}_{simple}(\theta) = \|\epsilon_\theta(\mathbf{x}_t, t) - \epsilon_t\|_2^2$.

Unified Diffusion Framework. To enable versatile stain transfer between any source-target stain pair, our framework accommodates additional inputs: the source image \mathbf{c} from stain $s_0 \in \mathcal{S}$, the desired target stain $s_1 \in \mathcal{S}$ (where $s_0 \neq s_1$), the noise timestep t , and the noised target image $\tilde{\mathbf{x}}_t$. Given the sampled noise ϵ_t , our unified noise prediction network ϵ_θ [12] denoises $\tilde{\mathbf{x}}_t$ conditioned on these inputs, formalized by the optimization objective:

$$\min_\theta \mathbb{E}_{\{s_0, s_1\} \subset \mathcal{S}} \mathbb{E}_{(\mathbf{c}, \mathbf{x})} \mathbb{E}_{\epsilon_t \sim \mathcal{N}(0, \mathbf{I})} \mathbb{E}_t \|\epsilon_\theta(\tilde{\mathbf{x}}_t, \mathbf{c}, s_0, s_1, t) - \epsilon_t\|_2^2.$$

Here the source image \mathbf{c} is conditioned via concatenation, and time features t are conditioned via sinusoidal positional encodings and spatial addition [15].

2.2. Advanced Functionality Design

Stain Conditioning. To achieve a unified framework, a crucial aspect is the conditioning on stain domains. We propose a dual-encoding approach to differentiate between various tasks. Specifically, we use an integer vector to denote both the source and target domains. We encode the source domain as 1 and the target domain as 2, i.e., the stain vector $\mathbf{r} = (r_1, r_2, \dots, r_K)$, where $r_{s_0} = 1, r_{s_1} = 2$, and $r_i = 0$ for each $i \in \mathcal{S} \setminus \{s_0, s_1\}$ (Figure 1 right). This vector is embedded and added with the time embedding to feed to the model through spatial addition, mirroring the class label conditioning approach in LDM [14]. This method incorporates both the source domain and the target domain into the model, enhancing its ability to distinguish between different tasks. The design can easily be extended to scenarios with multiple input domains.

Learning Objective. In a preliminary experiment, we observed that a conventional DDPM implementation [12] tends to produce unstable intensity shifts in the generated images, an issue evident in Figure 2(c). Closer inspection of the diffusion process revealed errors at larger timesteps t . This phenomenon is associated with the learning objective for noise prediction, particularly challenging at greater timesteps. Referring to Equation 1, when t is large, $\sqrt{\bar{\alpha}_t}$ approaches zero, then even minor errors in ϵ can lead to substantial deviations in the final output, which causes the unstable intensity

Table 1. Quantitative results on versatile stain transfer, using a proof-of-concept translation cycle: H&E→MT→PASM→PAS→H&E. (FID: Fréchet Inception Distance, KID: Kernel Inception Distance)

Method	H&E→MT		MT→PASM		PASM→PAS		PAS→H&E	
	FID↓	KID↓	FID↓	KID↓	FID↓	KID↓	FID↓	KID↓
pix2pix [8]	182.75	0.195	225.42	0.211	142.30	0.137	73.78	0.058
CycleGAN [9]	54.03	0.032	77.17	0.045	45.59	0.025	189.65	0.208
Palette [16]	87.35	0.058	51.88	0.012	75.20	0.064	78.10	0.052
StarGAN [10]	84.45	0.068	144.80	0.129	99.30	0.083	87.65	0.072
StarGANv2 [11]	104.51	0.099	151.68	0.190	87.71	0.083	89.30	0.079
Ours	41.76	0.019	42.66	0.012	41.92	0.022	41.46	0.023

Table 2. Ablation studies comparing different strategies of stain conditioning.

Stain conditioning	H&E→MT		MT→PASM		PASM→PAS		PAS→H&E	
	FID↓	KID↓	FID↓	KID↓	FID↓	KID↓	FID↓	KID↓
1) One-hot enc.	42.28	0.020	47.86	0.020	52.19	0.035	49.05	0.033
2) Dual enc.	41.76	0.019	42.66	0.012	41.92	0.022	41.46	0.023

issue. To mitigate this issue, we incorporate the dynamic dual-output diffusion head [19] into our framework, which leverages the complementary strengths of noise prediction and original image reconstruction, resulting in more stable outputs (Figure 2(d)). Specifically, we modify the output channels of our model from C to $3 \times C$, where C represents the number of channels in the target images. For RGB images in our experiments, $C = 3$. The generate model f_θ computes:

$$\epsilon_\theta, \mathbf{x}_\theta, \mathbf{r}_\theta = f_\theta(\tilde{\mathbf{x}}_t, \mathbf{c}, s_0, s_1, t).$$

Here ϵ_θ is the prediction of noise, \mathbf{x}_θ is the prediction of the original image, and \mathbf{r}_θ is a weighting factor. All of ϵ_θ , \mathbf{x}_θ , and \mathbf{r}_θ have C channels. Subsequently, both \mathbf{x}_θ and ϵ_θ can estimate the forward process posterior mean $\tilde{\mu}_t$. Specifically, $\mu_{\mathbf{x}}(\mathbf{x}_\theta)$ represents the mean obtained through the prediction of the original image \mathbf{x}_θ , while $\mu_\epsilon(\epsilon_\theta)$ represents the mean obtained through the prediction of noise ϵ_θ . We then use \mathbf{r}_θ as a weight to combine these two estimates:

$$\mu_\theta = \mathbf{r}_\theta \cdot \mu_{\mathbf{x}}(\mathbf{x}_\theta) + (1 - \mathbf{r}_\theta) \cdot \mu_\epsilon(\epsilon_\theta).$$

Finally, we define our loss function as:

$$\mathcal{L}_\theta = \|\epsilon_\theta - \epsilon_t\|_2^2 + \|\mathbf{x}_\theta - \mathbf{x}_0\|_2^2 + \|\mu_\theta - \tilde{\mu}_t\|_2^2.$$

3. EXPERIMENTS

3.1. Experimental Settings

Datasets. We used Whole Slide Imaging (WSI) to obtain 40 slices of mouse kidney tissue for each staining method, including H&E, PAS, MT, and PASM. High-resolution scanned images with approximately $\sim 17,000 \times 27,000$ pixels were acquired using an Olympus VS200 microscope. The data were collected within the same batch, thus eliminating potential batch effects. To minimize pixel misalignment across various stains, we collected each eight slices in sequence

and applied a four-stain cycle: H&E, PAS, MT, and PASM. Further, we employed VALIS [20], an advanced registration algorithm, for serial registration of multiple consecutive slices. The images were $4 \times$ downsampled and divided into 256×256 patches, following the approach described in [3]. Only patches containing a minimum of 20% foreground pixels were retained. After manual examination by expert pathologists, we created a pixel-level, well aligned paired dataset. For model training and evaluation, we randomly partitioned the dataset into two sets, with 29,566 patches for training and 3,532 for testing.

Implementation Details. We adopted a linear variance scheduler setting the maximum timestep T at 1,000, with a noise level interval ranging from 0.002 to 0.02. Our model architecture is designed upon the efficient-UNet described in [13], with modifications tailored to our specific task. The architecture encompasses six stages, with channel numbers per stage being 32, 64, 96, 192, 256, and 512 respectively. Following [19], we adjusted the output head’s channel count from 3 to 9 to incorporate the dynamic dual-output head strategy. The dimensions for time and stain embedding were at 1024, incorporating a linear transformation for the stain vector embedding. We employed the AdamW optimizer with a constant learning rate of 1×10^{-4} . The training was conducted over 300 epochs for a unified model, accommodating all $A_4^2 = 12$ potential combinations of source-target domains. The batch size was 48. Our model was developed using PyTorch and MONAI frameworks and was trained on three NVIDIA A800 80G GPUs. For sampling, we used an advanced sampler, DPM-Solver++ [21], employing a single-step mode with 20 steps, translating a 256×256 image using ~ 0.7 s on a single A800 GPU.

Evaluation Metrics. Given the challenges in obtaining exact ground truth matches, even after registration [7], and the variability in stains produced by the same technician [1], we employed unpaired metrics in our experiments. Specifically, we utilized the Fréchet Inception Distance (FID) and the Kernel Inception Distance (KID). It is important to note that methods like SSIM, which evaluate structural similarity, are not well-suited for consecutive tissue slices. This is because consecutive slices inherently exhibit structural differences. Therefore, we opted for widely-used metrics in generative tasks, such as FID and KID, which are commonly accepted as standard evaluation metrics in the field. These metrics provide a robust assessment of the quality and realism of generated images, aligning with the current best practices in the community.

Baselines. Our comparative analysis includes a variety of baseline methods, encompassing both GAN-based strategies (such as pix2pix [8], CycleGAN [9], StarGAN [10], and StarGANv2 [11]) and a diffusion-based approach (Palette [16]). For methods not inherently equipped for multi-domain tasks (specifically pix2pix, CycleGAN, and Palette), we conducted individual training for each domain pair. Each model was developed using official source code and trained for 300 epochs.

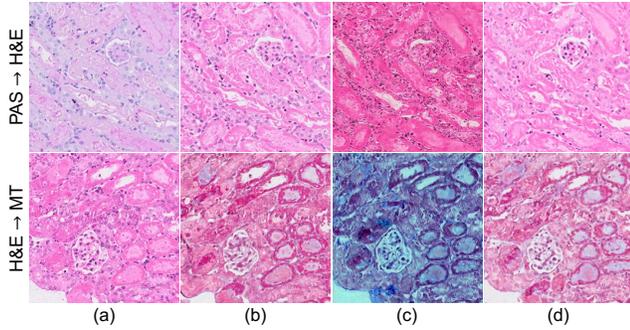


Fig. 2. Image translation comparison: (a) Source, (b) Target, (c) DDPM [12], (d) Our method. The conventional DDPM implementation using a learning objective for noise prediction yields unstable intensity shifts, while our approach ensures stability.

3.2. Quantitative & Qualitative Results

Our Framework’s Proficiency in Versatile Stain Transfer. Utilizing a conceptual conversion cycle (H&E→ MT→ PASM→ PAS→ H&E), we demonstrated our framework’s superiority in versatile stain transfer. As indicated in Table 1, our method surpasses competing techniques significantly in FID and KID metrics, recording an average FID of around 42 and a KID of approximately 0.02.

Our method, in contrast to Palette [16], which is another diffusion-based approach that trains separate networks for each task, exhibits superior performance across all tasks. Notably, it outperforms Palette in the H&E to MT translation task, achieving an FID improvement of approximately 45.59. Additionally, our method demonstrates enhanced detail capture, possibly due to the introduced dynamic dual-output head. For instance, while Palette struggles with accurately representing cell nuclei and glomerulus boundaries, as observed within the red box in the fourth row of Figure 3(e), our framework achieves precise translations. This underscores the synergistic effect of our all-in-one pathological translation framework over task-specific models, resonating with the emerging trends in generalist and foundation models [5, 4, 22].

In comparison to other multi-domain methods like StarGAN [10] and StarGANv2 [11], our framework maintains consistent performance across all tasks, unlike the others which may falter in specific conversions, such as MT→PASM, and struggle with style capture (see Figure 3(d)). This demonstrates that our method can maintain stability across multiple domains while using a unified framework.

3.3. Ablation Studies

We conducted an ablation study focusing on the design of stain conditioning. Besides our dual-encoding strategy, we

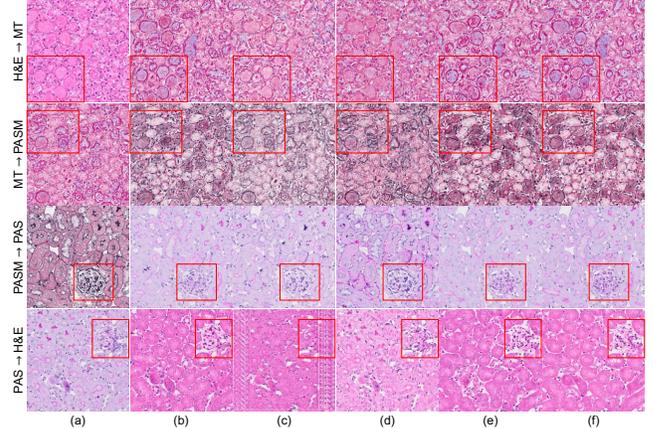


Fig. 3. Visualization of image translation results using various methods in the proof-of-concept cycle: H&E→ MT→ PASM→ PAS→ H&E. From left to right: (a) Source, (b) Target, (c) CycleGAN [9], (d) StarGAN [10], (e) Palette [16], (f) Our method. Our approach effectively captures the target domain’s style while retaining the source image’s content. (Best viewed when zoomed in.)

employ a one-hot encoding scheme. In this method of stain conditioning, the source domain is disregarded, and a single number is used to represent the target domain s_1 . Our results demonstrate that dual-encoding yields superior performance compared to one-hot encoding. As detailed in Table 2, our “dual-encoding” strategy outperforms “one-hot encoding” strategy in all the four tasks. Our dual-encoding strategy, by integrating source stain information, enhances task differentiation within the framework. Consequently, we have selected the dual-encoding approach for our framework.

4. CONCLUSION

We propose a novel diffusion-based unified framework for versatile stain transfer (DUST) in histopathology. By integrating dual encoding strategies and a dynamic dual-output head, it achieves superior performance in transferring images across multiple stain types. Our extensive experiments on a four-stain kidney histopathological dataset showcase its potential to advance the field of virtual staining towards efficient, accessible and high-throughput pathological analysis.

5. COMPLIANCE WITH ETHICAL STANDARDS

This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Ethics Committee of Peking University First Hospital (approval number: J2022138).

6. REFERENCES

- [1] Kevin de Haan, Yijie Zhang, Jonathan E Zuckerman, Tairan Liu, Anthony E Sisk, Miguel FP Diaz, Kuang-Yu Jen, Alexander Nobori, Sofia Liou, Sarah Zhang, et al., “Deep learning-based transformation of h&e stained tissues into special stains,” *Nature communications*, vol. 12, no. 1, pp. 4884, 2021.
- [2] Xilin Yang, Bijie Bai, Yijie Zhang, Yuzhu Li, Kevin de Haan, Tairan Liu, and Aydogan Ozcan, “Virtual stain transfer in histology via cascaded deep neural networks,” *ACS Photonics*, vol. 9, no. 9, pp. 3134–3143, 2022.
- [3] 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 *AAAI*, 2022, vol. 36, pp. 1630–1637.
- [4] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al., “Segment anything,” in *ICCV*, 2023.
- [5] Jun Ma, Yuting He, Feifei Li, Lin Han, Chenyu You, and Bo Wang, “Segment anything in medical images,” *Nature Communications*, vol. 15, no. 1, pp. 654, 2024.
- [6] Lulin Shi, Yan Zhang, Ivy HM Wong, Claudia TK Lo, and Terence TW Wong, “Mulhist: Multiple histological staining for thick biological samples via unsupervised image-to-image translation,” in *MICCAI*. Springer, 2023, pp. 735–744.
- [7] 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 *MICCAI*, 2023, pp. 632–641.
- [8] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros, “Image-to-image translation with conditional adversarial networks,” in *CVPR*, 2017, pp. 1125–1134.
- [9] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” in *ICCV*, 2017, pp. 2223–2232.
- [10] Yunjey Choi, Minje Choi, Munyoung Kim, Jung-Woo Ha, Sunghun Kim, and Jaegul Choo, “Stargan: Unified generative adversarial networks for multi-domain image-to-image translation,” in *CVPR*, 2018, pp. 8789–8797.
- [11] Yunjey Choi, Youngjung Uh, Jaejun Yoo, and Jung-Woo Ha, “Stargan v2: Diverse image synthesis for multiple domains,” in *CVPR*, 2020, pp. 8188–8197.
- [12] Jonathan Ho, Ajay Jain, and Pieter Abbeel, “Denoising diffusion probabilistic models,” *NeurIPS*, vol. 33, pp. 6840–6851, 2020.
- [13] Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily L Denton, Kamyar Ghasemipour, Raphael Gontijo Lopes, Burcu Karagol Ayan, Tim Salimans, et al., “Photorealistic text-to-image diffusion models with deep language understanding,” *NeurIPS*, vol. 35, pp. 36479–36494, 2022.
- [14] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer, “High-resolution image synthesis with latent diffusion models,” in *CVPR*, 2022, pp. 10684–10695.
- [15] Prafulla Dhariwal and Alexander Nichol, “Diffusion models beat gans on image synthesis,” *NeurIPS*, vol. 34, pp. 8780–8794, 2021.
- [16] Chitwan Saharia, William Chan, Huiwen Chang, Chris Lee, Jonathan Ho, Tim Salimans, David Fleet, and Mohammad Norouzi, “Palette: Image-to-image diffusion models,” in *ACM SIGGRAPH 2022 Conference Proceedings*, 2022, pp. 1–10.
- [17] Xun Huang and Serge Belongie, “Arbitrary style transfer in real-time with adaptive instance normalization,” in *ICCV*, 2017, pp. 1501–1510.
- [18] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu, “Semantic image synthesis with spatially-adaptive normalization,” in *CVPR*, 2019, pp. 2337–2346.
- [19] Yaniv Benny and Lior Wolf, “Dynamic dual-output diffusion models,” in *CVPR*, 2022, pp. 11482–11491.
- [20] Chandler D Gatenbee, Ann-Marie Baker, Sandhya Prabhakaran, Otilie Swinyard, Robbert JC Slebos, Gunjan Mandal, Eoghan Mulholland, Noemi Andor, Andriy Marusyk, Simon Leedham, et al., “Virtual alignment of pathology image series for multi-gigapixel whole slide images,” *Nature communications*, vol. 14, no. 1, pp. 4502, 2023.
- [21] Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu, “Dpm-solver++: Fast solver for guided sampling of diffusion probabilistic models,” *arXiv:2211.01095*, 2022.
- [22] Mingze Yuan, Peng Bao, Jiajia Yuan, Yunhao Shen, Zifan Chen, Yi Xie, Jie Zhao, Yang Chen, Li Zhang, Lin Shen, et al., “Large language models illuminate a progressive pathway to artificial healthcare assistant: A review,” *arXiv preprint arXiv:2311.01918*, 2023.