Learning to Blur is Learning to Deblur: Realistic Synthetic UHD Blurred Image via Diffusion

Xin Su[*] Xiuyi Jia[†] Chen Wu[‡] Dianjie Lu[§] Guijuan Zhang[¶] Yang Wen[‖]
Zhuoran Zheng[**]

Abstract

Generating large-scale, diverse, and realistic paired data for ultra-high-definition (UHD) image deblurring is challenging due to the complex textures and information contained in UHD images. Existing synthetic methods often fail to replicate the complex, spatially-varying blurs present in real-world 4K imagery, limiting model performance. To address this gap, we introduce two diffusion-centric contributions: First, ***UHD-RealBlur****, a large-scale 4K dataset produced by our novel PhysicsGuided-BlurSynth framework. PhysicsGuided-BlurSynth leverages a pre-trained Stable Diffusion model controlled using both content guidance from a clean input image and explicit conditioning on* ***real-world camera settings*** *(ISO, aperture, shutter speed, focus mode, etc.). Futhermore, we collected a set of real-world blurred images (with 4K resolution) and adopted unpaired training to fine-tune the distribution of generated blurred images to make it closer to real-world distributions. Second, we develope a FreqDiff, which incorporates essential frequency information from blurred inputs into the diffusion process and is specifically engineered for UHD image deblurring. Extensive experiments demonstrate that FreqDiff trained solely on UHD-RealBlur exhibits outstanding performance on real-world 4K blurred images.*

1. Introduction

Image blur remains a persistent challenge that significantly compromises visual quality in high-resolution imagery, particularly at 4K resolution (3840×2160), where even subtle degradations become visually apparent. This impairs human viewing experience and severely hinders downstream computer vision tasks such as object recognition, scene understanding, and autonomous navigation [3, 24]. Despite recent advances in image deblurring algorithms, their effectiveness is fundamentally constrained by a critical data bottleneck: the scarcity of large-scale, diverse, and accurately paired real-world training data that captures identical scene content in both sharp and naturally blurred states [14].

[*]Corresponding author: zhengzr@njust.edu.cn
[*]Fuzhou University
[†]Nanjing University of Science and Technology
[‡]University of Science and Technology of China
[§]Shandong Normal University
[¶]Shandong Normal University
[‖]Shenzhen University
[**]Sun Yat-sen University

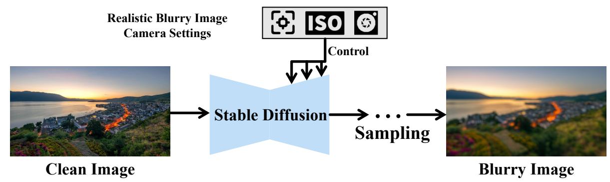

Figure 1. Overview of our proposed realistic blur synthesis framework. A pre-trained Stable Diffusion is controlled using both the content guidance from a clean input image and explicit conditioning on real-world camera settings (ISO, aperture, shutter speed, focus mode, etc.). The distribution alignment of diffusion-generated blurred images with unpaired real-world 4K blurred images is achieved using GAN loss.

The conventional approach to addressing this data scarcity relies on synthetic blur generation, primarily through averaging adjacent video frames [1] or convolving clean images with predefined mathematical kernels [8]. While computationally efficient, these methods struggle to capture the intricate complexity of real-world blur phenomena. Real-world blur exhibits subtle spatial variations and physical characteristics visible at high resolutions that mathematical models often fail to replicate. This creates a substantial domain gap between synthetic training data and real-world conditions, severely limiting the generalization capabilities of deblurring models in practical applications [22]. Despite architectural innovations in deblurring networks, this fundamental data limitation creates a performance bottleneck that persists across the field. Furthermore, the task of recovering fine details from severely blurred 4K images presents unique challenges beyond data limitations. Recent diffusion-based restoration approaches [20, 25] demonstrate considerable promise for generative image restoration but encounter difficulties when scaled to UHD deblurring. The reverse diffusion process requires sophisticated conditioning mechanisms to recover high-frequency details obliterated by complex blurs—information that current approaches fail to effectively leverage during restoration. The frequency characteristics of blur, which contain crucial information about the degradation process, remain underutilized in existing frameworks.

To address these issues, using our PhysicsGuided-BlurSynth framework, we create UHD-RealBlur, a comprehensive dataset of high-quality 4K resolution image pairs that replicates the complexity and diversity of real-world blur. We further introduce FreqDiff, a diffusion-based deblurring framework specifically engineered for UHD restoration. The distinctive feature of FreqDiff is its Frequency-Domain Conditioning mechanism that incorporates essential frequency information from blurred inputs into the diffusion process. Our experiments demonstrate that FreqDiff, trained solely on UHD-RealBlur, consistently surpasses existing methods across multiple benchmarks, with particularly impressive results on challenging real-world blurred images. The model preserves intricate details and suppresses artifacts effectively, especially when handling complex blur patterns in 4K resolution. We also confirm practical relevance through downstream vision tasks, where our restored images substantially boost performance in object detection and semantic segmentation applications. Our main **contributions** include:

- We propose a physics-aware UHD blur synthesis paradigm that bridges the gap between synthetic and real-world blurs by leveraging camera metadata, diffusion models, and limited real-world 4K blurred images to generate realistic 4K training data.
- We propose a novel frequency-aware diffusion method for UHD deblurring. This method uses frequency-domain information to direct the restoration process. Abundant experimental results show that our approach can efficiently generalize to real-world blurry situations, setting new state-of-the-art performance for high-resolution deblurring.

2. Related Work

Image deblurring has progressed from optimization-based approaches [13, 26] to CNN architectures leveraging multi-scale processing [11], recurrent structures [19], and attention mechanisms [2, 27]. Recent diffusion-based methods [20, 25] show promise but remain limited by training data quality. The scarcity of paired sharp-blurred images has led to various synthetic data generation strategies. Convolution with predefined kernels [6, 11] offers computational efficiency but produces overly uniform blurs that poorly represent complex real-world degradations. Frame-averaging from high-framerate videos [12] better simulates motion blur but inadequately captures other blur types. Generative approaches using GANs [5] and diffusion models [8, 22] show potential but typically lack explicit control over physical blur formation processes. Existing methods fail to incorporate camera parameters (aperture, shutter speed, ISO, etc.) into the parameters that fundamentally control blur characteristics in real photographs. Our approach addresses this limitation by explicitly conditioning the diffusion process on camera metadata from real-world photos, generating physically accurate blur patterns under different capture conditions with the help of GAN loss, thereby significantly reducing the domain gap between synthesis and reality.

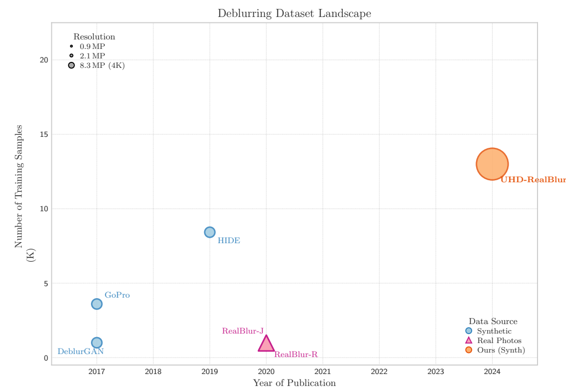

Figure 2. Comparison of training sample sizes across various deblurring datasets published over the years. Our dataset has significantly improved in terms of quantity, resolution, and degradation quality.

3. UHD-RealBlur Dataset of Synthesis Method

Figure 3. Qualitative comparison of blur synthesis methods. (a) Original sharp 4K image (I_{sharp}). (b) Blurred image generated using a simple Gaussian blur kernel. (c) Blurred image (I_{blur}) synthesized by GAN. (d) Blurred image (I_{blur}) synthesized by our method using target parameters.

To address the critical scarcity of realistic training data for UHD image deblurring, we develop **PhysicsGuided-BlurSynth**, a novel framework for synthesizing authentic, high-resolution blur. This framework generates **UHD-RealBlur**, a large-scale dataset comprising paired sharp

Algorithm 1 PhysicsGuided-BlurSynth Process

Require: Clean image I_{sharp}, Camera metadata $M = \{\text{ISO, aperture, ...}\}$
Ensure: Realistically blurred 4K image I_{blur}
Initialize pre-trained Stable Diffusion D_θ and its VAE Decoder.
Prepare content conditioning $c_{content} \leftarrow \text{ControlNet}(I_{sharp})$. ▷ Spatial conditioning
Prepare physics conditioning $c_{physics} \leftarrow$ CLIP. ▷ Text prompt
Fusion conditioning: $c \leftarrow \text{Combine}(c_{content}, c_{physics})$.
Sample initial noise latent $z_T \sim \mathcal{N}(0, I)$.
for $t = T \dots 1$ **do** ▷ Reverse diffusion process
    Predict noise $\epsilon_\theta \leftarrow D_\theta(z_t, t, c)$ using Stable Diffusion guided by c.
    Update latent state $z_{t-1} \leftarrow \text{DDIMStep}(z_t, \epsilon_\theta, t)$. ▷ Using DDIM sampler
end for
Decode final latent state z_0 using VAE: $I_{blur} \leftarrow \text{VAEDecoder}(z_0)$.
return I_{blur}

and realistically blurred 4K images, significantly bridging the synthetic-to-real domain gap prevalent in conventional datasets reliant on kernel convolution [8] or frame averaging [1]. Unlike physical multi-capture approaches, our method leverages generative modeling conditioned explicitly on photographic parameters.

The PhysicsGuided-BlurSynth method employs the Stable Diffusion [15] as its generative engine. Firstly, **ControlNet** [29] integrates robust spatial guidance derived from the clean 4K source image I_{sharp}, preserving the underlying scene structure throughout the synthesis. Secondly, physical realism is instilled by conditioning on target camera metadata M, encompassing parameters like ISO, aperture, shutter speed, and focus settings. This metadata is translated into descriptive **text prompts** and processed by Stable Diffusion's integrated text encoder, directly influencing the characteristics of the generated blur. The synthesis, procedurally outlined in Algorithm 1, utilizes the **DDIM sampler** [18] to iteratively refine a latent representation z_t under the joint influence of content ($c_{content}$) and physics ($c_{physics}$) conditions. Stable Diffusion's VAE then decodes the final latent z_0 into the resulting blurred image I_{blur}. Here, I_{blur} is constrained by GAN loss to match the distribution of 1,000 real-world 4K blurred images. The UHD-RealBlur of high-quality, sharp 4K source images (I_{sharp}) spanning various scene categories (including both indoor and outdoor scenes, featuring not only natural objects but also text and icons) is curated.

4. FreqDiff for UHD Deblurring

Leveraging the high-fidelity UHD-RealBlur dataset generated by PhysicsGuided-BlurSynth, we propose **FreqDiff**, a novel diffusion-based deblurring framework tailored for the challenges of UHD image restoration, illustrated in Figure 4. Standard diffusion models for image restoration often rely solely on spatial conditioning from the degraded input [17]. While existing methods may fail to fully capture nuanced frequency characteristics altered by complex blur (notably the attenuation of high-frequency details critical for UHD perceptual quality), FreqDiff directly incorporates

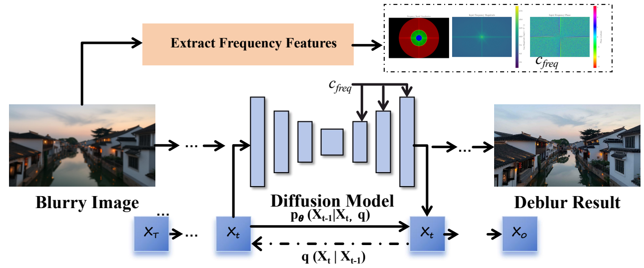

Figure 4. Overview of the FreqDiff framework. Frequency features c_{freq} are extracted from the input Blurry Image I_{blur}. These features, along with the noisy image x_t and timestep t, serve as conditions for the Diffusion Model (U-Net, ϵ_θ). The model is trained (Eq. 2) to predict the noise component $\epsilon_\theta(x_t, t, c_{freq})$. During inference, starting from noise x_T, the model iteratively applies the reverse diffusion step, guided by the predicted noise and the constant frequency condition c_{freq}, to generate the Deblur Result $\hat{I}_{sharp}$.

frequency-domain information from blurred inputs into the diffusion model's reverse process to guide accurate high-frequency detail recovery.

The core of FreqDiff is a time-conditional U-Net architecture [4, 16], common in diffusion models. This network, denoted as ϵ_θ, is trained to predict the noise component ϵ added to a sharp image I_{sharp} during the forward diffusion process at timestep t. The forward process gradually adds Gaussian noise according to a variance schedule $\beta_1, \dots, \beta_T$:

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t\mathbf{I}) \quad (1)$$

where $x_0 = I_{sharp}$. The reverse process, learned by ϵ_θ, aims to denoise a sample x_t drawn from $q(x_t|x_0)$ to iteratively predict x_{t-1}, ultimately recovering $x_0 \approx \hat{I}_{sharp}$.

To enhance high-frequency restoration, FreqDiff introduces a novel **Frequency-Domain Conditioning** mechanism. As shown in Figure 4, given the blurred input image I_{blur}, we first compute its frequency representation via FFT, obtaining magnitude $|\mathcal{F}(I_{blur})|$ and phase $\angle\mathcal{F}(I_{blur})$. Recognizing that blur predominantly affects magnitude while phase carries structural information, we extract relevant frequency features c_{freq} from these components. These features c_{freq} are then injected as conditional information into the U-Net backbone ϵ_θ, potentially at multiple resolutions analogous to spatial conditioning [29], allowing the network to leverage frequency information pertinent to different spatial scales.

The model is trained to predict the noise ϵ based on the noisy image x_t, the timestep t, and the crucial frequency condition c_{freq}, minimizing the loss:

$$\mathcal{L}_{FreqDiff} = \mathbb{E}_{t, I_{sharp}, I_{blur}, \epsilon} \left\| \epsilon - \epsilon_\theta(x_t, t, c_{freq}) \right\|^2 \quad (2)$$

where $x_t = \sqrt{\bar{\alpha}_t} I_{sharp} + \sqrt{1-\bar{\alpha}_t}\epsilon$, $\bar{\alpha}_t = \prod_{i=1}^{t}(1-\beta_i)$, and $c_{freq} = \text{ExtractFreqFeatures}(I_{blur})$.

By explicitly conditioning on c_{freq}, FreqDiff encourages the reverse diffusion process to prioritize the reconstruction of frequency components attenuated by the blur. This is particularly advantageous for UHD images where fine textures and sharp edges (high frequencies) are perceptually crucial. The model learns to correlate patterns in the blurred image's frequency spectrum (c_{freq}) with the noise (ϵ) required to reverse the degradation. Training on our UHD-RealBlur dataset ensures the model learns realistic blur-frequency relationships. During inference (Figure 4), c_{freq} is extracted from the input I_{blur} and guides the iterative denoising from $x_T \sim \mathcal{N}(0, \mathbf{I})$ to yield the restored image $\hat{I}_{sharp} \approx x_0$.

5. Experiments

5.1. Experimental Setup

Datasets. We train our FreqDiff models primarily on the **UHD-RealBlur** dataset generated using PhysicsGuided-BlurSynth, ensuring exposure to physically realistic blur characteristics common in UHD imagery. For evaluation and comparison with state-of-the-art methods, we test on a widely-used deblurring benchmark **GoPro** [11]. **Implementation Details.** Our FreqDiff framework is implemented using PyTorch. We train using the AdamW optimizer [9] with a learning rate of 1.5×10^{-4} decayed using a cosine schedule. Training is performed with a batch size of 8 for 600k iterations on 4$\times$ NVIDIA A100 GPUs.

5.2. Comparison with State-of-the-Art Methods

Quantitative Comparisons. Table 1 presents the quantitative results (PSNR / SSIM) on the GoPro test datasets. Both FreqDiff variants demonstrate highly competitive performance. Notably, FreqDiff-Adv consistently achieves the best or second-best results across all datasets, particularly showing significant gains on the challenging RealBlur benchmarks, which contain complex, real-world blur patterns. This highlights the effectiveness of incorporating explicit frequency-domain conditioning into the diffusion process for high-quality deblurring.

Table 1. Quantitative comparison (PSNR / SSIM ↑) with state-of-the-art and other restoration methods on the GoPro [11] and our UHD-RealBlur datasets. **Bold** indicates the best performance, <u>underline</u> indicates the second best for each dataset.

Method	GoPro [11]		UHD-RealBlur (Ours)	
	PSNR (↑)	SSIM (↑)	PSNR (↑)	SSIM (↑)
Real-ESRGAN [21]	29.55	0.925	21.80	0.695
AirNet [7]	30.10	0.930	22.15	0.710
DGUNet [10]	30.50	0.938	22.50	0.725
MPRNet [27]	32.66	0.959	23.55	0.760
NAFNet [2]	32.72	0.960	23.68	0.765
Uformer [23]	32.88	0.961	23.80	0.770
Restormer [28]	32.92	0.961	23.95	0.775
FreqDiff-Base (Ours)	<u>32.95</u>	<u>0.962</u>	<u>24.30</u>	<u>0.785</u>
FreqDiff-Adv (Ours)	**33.05**	**0.963**	**24.85**	**0.795**

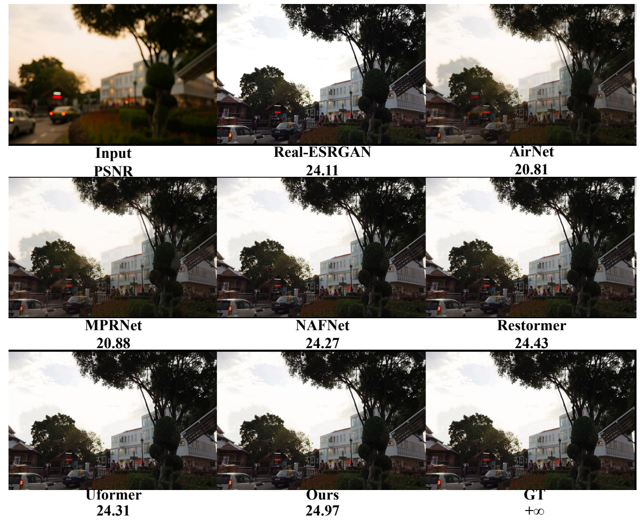

Figure 5. Qualitative comparison of deblurring results on a sample image. Our FreqDiff method demonstrates superior performance in restoring sharp details and reducing blur compared to state-of-the-art methods.

Figure 6. This figure demonstrates the restoration results of real-world 4K blurred images.

Qualitative Comparisons. Figure 5 showcases the visual results of our FreqDiff method compared to several state-of-the-art deblurring techniques on a representative sample. The visualization highlights FreqDiff's ability to recover finer details and textures while effectively suppressing blur artifacts, aligning with the superior quantitative metrics presented in Table 1. Our method produces visually sharper and more faithful reconstructions compared to other approaches.

Real-world deblurring. As shown in Figure 6, we also present a case of real-world deblurred image restoration using our method.

6. Conclusion

We introduce PhysicsGuided-BlurSynth, a framework using camera parameter conditioning to generate physically realistic UHD blurs for the UHD-RealBlur dataset, addressing synthetic-to-real domain gaps. Complementing this, FreqDiff incorporates frequency-domain conditioning into diffusion restoration. Joint experiments demonstrate their synergy advances UHD deblurring, setting new state-of-the-art performance, particularly on complex real-world benchmarks.

References

[1] Yogesh Balaji, Seungjun Nah, Xun Huang, Arash Vahdat, Jiaming Song, Karsten Kreis, Miika Aittala, Timo Aila, Samuli Laine, Bryan Catanzaro, and Tero Karras. ediff-i: Text-to-image diffusion models with ensemble of expert denoisers. *arXiv preprint arXiv:2211.01324*, 2022. 1, 3

[2] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022. 2, 4

[3] Z Chen, H Sun, L Zhang, and F Zhang. Survey on visual signal coding and processing with generative models: Technologies, standards and optimization. *IEEE Journal on Emerging Topics in Signal Processing*, 2024. 1

[4] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 6840–6851, 2020. 3

[5] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiri Matas. Deblurgan: Blind motion deblurring using conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8183–8192, 2018. 2

[6] Anat Levin, Yair Weiss, Frédo Durand, and William T Freeman. Understanding and evaluating blind deconvolution algorithms. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1964–1971. IEEE, 2009. 2

[7] Boyun Li, Xiao Liu, Peng Hu, Zhongqin Wu, Jiancheng Lv, and Xi Peng. All-in-one image restoration for unknown corruption. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 4

[8] Zhendong Li, Chen Wang, Jian Liu, Weimian Wang, and Qing Hou. Diffusion models for image deblurring. *arXiv preprint arXiv:2311.17201*, 2023. 1, 2, 3

[9] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations (ICLR)*, 2019. 4

[10] Chong Mou, Qian Wang, and Jian Zhang. Deep generalized unfolding networks for image restoration. In *CVPR*, 2022. 4

[11] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3883–3891, 2017. 2, 4

[12] Seungjun Nah, Sanghyun Baik, Seokil Hong, Gyeongsik Moon, Sanghyun Son, Radu Timofte, and Kyoung Mu Lee. Ntire 2019 challenge on video deblurring: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019. 2

[13] Jinshan Pan, Deqing Sun, Hanspeter Pfister, and Ming-Hsuan Yang. Blind image deblurring using dark channel prior. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1628–1636, 2016. 2

[14] C Peng, Y Tang, D Li, Y Zhou, and N Wang. Bags: Blur agnostic gaussian splatting through multi-scale kernel modeling. *Springer*, 2024. 1

[15] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10684–10695, 2022. 3

[16] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention (MICCAI)*, pages 234–241. Springer, 2015. 3

[17] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022. 3

[18] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations (ICLR)*, 2021. 3

[19] Xin Tao, Hongyun Gao, Xiaoyong Shen, Jue Wang, and Jiaya Jia. Scale-recurrent network for deep image deblurring. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8174–8182, 2018. 2

[20] Jianyi Wang, Zongsheng Yue, Shangchen Zhou, Kelvin CK Chan, and Chen Change Loy. Exploiting diffusion prior for real-world image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14149–14159, 2023. 1, 2

[21] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV Workshops)*, 2021. 4

[22] Xin Wang, Xin Tao, Zhengfang Lin, Chen Change Loy, and Ming-Hsuan Yang Ziwei Li. Diffusion models for image restoration and enhancement–a comprehensive survey. *International Journal of Computer Vision*, pages 1–29, 2024. 1, 2

[23] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 17683–17693, 2022. 4

[24] Z Wang, Z Zhang, X Zhang, and H Zheng. Dr2: Diffusion-based robust degradation remover for blind face restoration. In *Proceedings of the IEEE*, 2023. 1

[25] Jay Whang, Mauricio Delbracio, Hossein Tian, Nuno Vasconcelos, Peyman Milanfar, and Sunghyun Cho. Deblurring via stochastic refinement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16293–16303, 2022. 1, 2

[26] Li Xu, Shicheng Zheng, and Jiaya Jia. Unnatural l0 sparse representation for natural image deblurring. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1107–1114, 2013. 2

[27] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling

Shao. Multi-stage progressive image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14821–14831, 2021. 2, 4

[28] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *CVPR*, 2022. 4

[29] Lvmin Zhang and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 3836–3847, 2023. 3