

UniDemoiré: Towards Universal Image Demoiréing with Data Generation and Synthesis

Zemin Yang^{1,*}, Yujing Sun^{2,*}, Xidong Peng¹, Siu Ming Yiu², Yuexin Ma^{1,†}

¹ShanghaiTech University

²The University of Hong Kong

{csyangzm, mayuxin}@shanghaitech.edu.cn, {yjsun, smyi}@cs.hku.uk

Abstract

Image demoiréing poses one of the most formidable challenges in image restoration, primarily due to the unpredictable and anisotropic nature of moiré patterns. Limited by the quantity and diversity of training data, current methods tend to overfit to a single moiré domain, resulting in performance degradation for new domains and restricting their robustness in real-world applications. In this paper, we propose a universal image demoiréing solution, **UniDemoiré**, which has superior generalization capability. Notably, we propose innovative and effective data generation and synthesis methods that can automatically provide vast high-quality moiré images to train a universal demoiréing model. Our extensive experiments demonstrate the cutting-edge performance and broad potential of our approach for generalized image demoiréing.

Code — <https://github.com/4DVLab/UniDemoire>

Introduction

Digital screens have become essential devices for displaying information in our daily work and life. However, images captured from screens frequently suffer from frustrating moiré patterns, significantly degrading image quality and hindering content extraction. Therefore, it becomes crucial to effectively remove such moiré artifacts to help users obtain high-quality images from their digital imaging devices and to support industries in maintaining high-standard product visual presentation and digital archiving. However, moiré patterns are characterized as anisotropic and multi-scale, as well as involving considerable shape variations and color distortions (Amidror 2009). Such traits are seldom seen in other types of artifacts, like noise, rain streaks, fog, blurring, etc., posing a significant challenge for even the most advanced image restoration methods (Luo et al. 2023; Zhu et al. 2023; Fei et al. 2023).

Hence, many methods have been proposed to tackle the problem of demoiréing in recent years (Sun, Yu, and Wang 2018; Liu et al. 2020; Luo et al. 2020; He et al. 2019, 2020;

Wang et al. 2023a; Yue et al. 2022; Yu et al. 2022). Nevertheless, the effectiveness of such supervised methods heavily depends on the volume of training data, consisting of pairs of moiré images and their clean counterparts. As we know, collecting such data is a daunting task and it requires precise calibration between natural images and moiré patterns. The limitations of the data lead to the limitations of the methods, resulting in poor generalization of the network model, which performs poorly on the data containing new moiré patterns or new natural images. In order to expand the quantity and diversity of the training data in a convenient way, some methods have started to explore the synthesis of moiré patterns. LCDMoiré (Yuan et al. 2019) designs hand-craft mathematical models. However, it could not represent complex features of moiré patterns and leads to a substantial discrepancy between the synthetic data and actual moiré images. To enhance realism, recent studies (Park et al. 2022; Zhong et al. 2024) extract moiré patterns from existing real images and combine them with clean images for data synthesis. Nevertheless, these methods do not escape from the moiré domains of the existing training data, bringing limited performance improvement on new moiré domains. To develop a universal model for image demoiréing with greater generalization capability and practicality, two critical challenges emerge: *how to generate a vast amount of diverse data, and how to ensure the authenticity of the data?*

To address the above challenges, we propose a universal image demoiréing solution, **UniDemoiré**, capable of generating a vast amount of realistic-looking training data to enhance the generalization capabilities of the image demoiréing model, as Figure 1 shows. First, inspired by the fact that the moiré pattern is unrelated to the content of the image, we introduce a novel, large-scale **Moiré Pattern Dataset** by capturing moiré patterns against a plain white background. Unlike previous moiré datasets that capture nature images with moiré, our pure moiré patterns can be applied to arbitrary nature images to scale up the data domain automatically. Moreover, our dataset does not need calibrations between the moiré image and the clean image, which can avoid the effect of calibration errors and facilitate the learning process of the model. In particular, our dataset introduces more pattern diversity by considering various previously overlooked factors (Yang et al. 2023), including zooming rate, CMOS technology, pixel size, and

*These authors contributed equally.

†Corresponding author.

Copyright © 2025, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

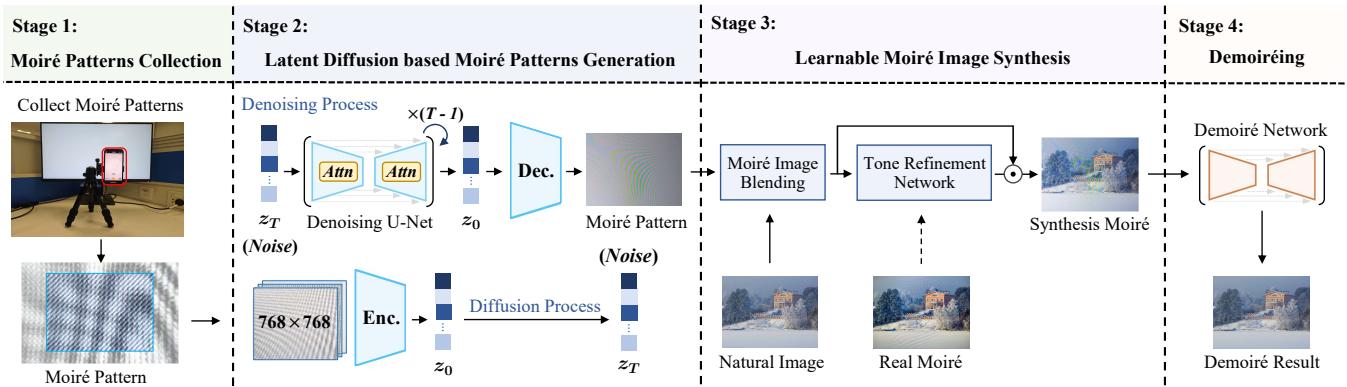


Figure 1: The workflow of our proposed UniDemoiré.

panel types. Second, building on this real-captured moiré pattern dataset, we propose a diffusion model-based **Moiré Pattern Generation** method to further increase the diversity of moiré patterns. Specifically, we implement a multi-scale cropping strategy to accommodate different input image sizes and an effective data filtering strategy to ensure the quality of training data for the diffusion model. Third, we propose a **Moiré Image Synthesis** method to create a sufficient amount of diverse and realistic-looking moiré images by blending the generated moiré patterns with clean natural images. In particular, to improve the authenticity of our synthesized data, we develop an effective learnable network and three effective losses to closely mimic the real captured moiré images in terms of color and brightness. Finally, our synthesized abundant moiré images serve to train an **Image Demoiréing Model** that achieves superior performance and promising generalization capabilities for zero-shot image demoiréing and cross-domain evaluations.

Our contribution can be summarized as follows:

- We propose a universal demoiréing solution, which substantially enlarges the knowledge domain and improves the generalization capability of demoiréing models.
- We collect a large-scale and high-resolution moiré pattern dataset and develop an effective moiré pattern generator to further increase the diversity of moiré patterns.
- We present a novel moiré image synthesis approach, providing a large amount of realistic-looking and high-quality moiré image samples, facilitating the training of a universal image demoiréing model.

Related Work

Image Restoration and Demoiréing

The inherent complexity of moiré patterns presents a unique challenge compared to other artifacts such as noise (Xing and Egiazarian 2021), haze (Li et al. 2021), blur (Lee et al. 2021), multiple artifacts in one go (Luo et al. 2023; Zhu et al. 2023; Fei et al. 2023; Zhang et al. 2023), etc. Consequently, these methods may not effectively solve the moiré issue. Current mainstream methods for image demoreing are learning based (Sun, Yu, and Wang 2018; Liu et al. 2020; Luo et al. 2020; He et al. 2019, 2020; Niu et al. 2023; Wang et al.

2023a; Yue et al. 2022; Liu et al. 2024; Zheng et al. 2020, 2021; Yu et al. 2022), greatly outperforming early handcraft feature based approaches (Sun, Li, and Sun 2014; Liu, Yang, and Yue 2015; Yang et al. 2017a,b). However, they exhibit poor generalization capability due to insufficient diverse and realistic training data, and researchers have thus begun exploring the potential of synthesized data.

Moiré Image Synthesis

An important category focuses on extracting moiré patterns from existing moiré images. Cyclic (Park et al. 2022) and UnDeM (Zhong et al. 2024) utilized GAN-based networks to generate moiré images from unpaired real moiré image datasets, resembling moiré patterns found in moiré images while retaining details from moiré-free images. However, they are unstable and constrained by the moiré patterns present in the real image datasets. Another category directly simulates moiré patterns on natural images. Shooting (Niu, Guo, and Wang 2021) simulated the interference of image processing to produce moiré patterns on natural images while Yang et al. (2023) collected background-independent moiré patterns and then superimposes the natural image with the collected pattern to synthesize moiré images. Unfortunately, due to the real-to-synthetic discrepancy, their model performance is limited in real-world applications. In contrast, our solution can produce realistic-looking and diverse data to greatly improve demoiréing models' performance.

Moiré Dataset

TIP18 (Sun, Yu, and Wang 2018), FHDMi (He et al. 2020), UHDM (Yu et al. 2022) are the most widely-used real-world moiré image dataset with increased resolutions 256, 1080P, and 4K, respectively. To lessen the burden of huge human efforts, a synthetic moiré image dataset LCDMoiré (2019) has been generated through shooting simulation. However, synthetic datasets often fail to accurately replicate real imaging processes, making it difficult for demoiréing models trained on them to perform well in real-world situations. More recently, MoireSpace (Yang et al. 2023) collects background-independent moiré pattern data for a different task, moiré detection. Inspired by it, we propose to collect a real moiré pattern dataset for image demoiréing. Taking inspiration from

Datasets		Avg. Resolution	Size	Capture settings				
Type	Name			Phone	Screen	Multi-zooming rate	Multi-camera / CMOS	Screen Panel
Moiré Image Dataset	TIP2018(R)	256 × 256	135000	3	3	X(1x-only)	X(Main-only)	IPS-only
	FHDMi(R)	1024 × 1024	12000	3	2	X(1x-only)	X(Main-only)	IPS-only
	UHDMi(R)	4328 × 3248	5000	3	3	X(1x-only)	X(Main-only)	IPS-only
	LCDMoiré(S)	1024 × 1024	10200	-	-	-	-	-
Moiré Pattern Dataset	MoireSpace(R)	2160 × 1286	18147	3	3	X(1x-only)	X(Main-only)	IPS-only
	Ours(R)	3840 × 2160	150000	6	6	✓(1x,2x,3x)	✓(Main,Telephoto)	IPS, SVA

Table 1: Comparisons of different moiré datasets. The “R” denotes the real dataset, and the “S” denotes the synthetic dataset.

this effort, we introduce a real moiré pattern dataset specifically tailored for image demoiréing. Comparatively, our dataset boasts a larger volume and greater diversity of data.

Method

Overview

The generalization ability of SOTA demoiréing models is greatly limited by the scarcity of data. Therefore, we mainly face two challenges to obtain a universal model with improved generalization capability: To obtain a vast amount of 1) diverse and 2) realistic-looking moiré data. Notice that traditional moiré image datasets contain real data, but continuously expanding their size to involve more diversity is extremely time-consuming and impractical. While current synthesized datasets/methods struggle to synthesize realistic-looking moiré images. Hence, to tackle these challenges, we introduce a universal solution, UniDemosoiré (Figure 1). The data diversity challenge is solved by collecting a more diverse moiré pattern dataset and presenting a moiré pattern generator to increase further pattern variations. Meanwhile, the data realistic-looking challenge is undertaken by a moiré image synthesis module. Finally, our solution can produce realistic-looking moiré images of sufficient diversity, substantially enhancing the zero-shot and cross-domain performance of demoiréing models.

Moiré Pattern Dataset

The traditional demoiréing datasets (Sun, Yu, and Wang 2018; He et al. 2020; Yu et al. 2022) typically exhibit a 1-1 correspondence, 1 clean image corresponds to only 1 moiré-contaminated image. However, in the real world, an image may be affected by various moiré patterns. Meanwhile, aligning moiré images with clean images often introduces errors because of the non-linear distortions and moiré artifacts within cameras. Therefore, we propose to collect a moiré pattern dataset rather than a moiré image dataset, with no need for image alignment and can easily synthesize multiple moiré counterparts of a single natural image. The collection of such a dataset is inspired by MoireSpace, which is designed to address the problem of detecting the presence of moiré rather than to eliminate moiré artifacts.

Capturing Process We capture videos of real-world moiré patterns on a pure white screen with a mobile phone to minimize color distortion in the moiré patterns. After record-

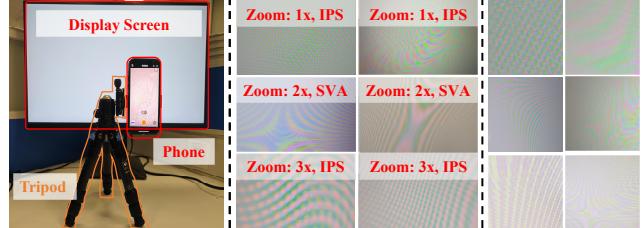


Figure 2: Data collection setup (left), and examples of moiré patterns in our dataset captured at different zoom rates and screen panel (middle), and our generated patterns (right).

ing, frames are uniformly extracted from each video to constitute our dataset. The setup is shown in Figure 2-left.

Data Diversity To enhance pattern diversity, we build our dataset by considering additional factors that influence moiré formation, which were overlooked in previous moiré datasets, including zooming rate, camera types, CMOS, and screen panel types. Besides, we doubled the number of mobile devices and display screens compared to existing datasets. A detailed comparison of ours and others is shown in Table 1. In summary, our dataset showcases an expanded size, 150000 moiré patterns, in standard 4K resolution with increased diversity. More dataset details are in the appendix.

Moiré Pattern Generation

Although we have collected a large scale of diverse data, it cannot encompass all conceivable moiré patterns. Inspired by recent diffusion models, which have been successfully trained towards diverse image generation in many tasks (Dhariwal and Nichol 2021), we propose to use diffusion models to further sample more diverse moiré patterns by sufficiently learning the structural, textural, and color representations of real moiré patterns. In this stage, we propose a multi-scale cropping strategy and a colorfulness-sharpness selection strategy to filter high-quality real data. Then we learn the distribution of real moiré patterns in the latent space to generate diverse patterns (Figure 2-right).

Multi-Scale Cropping Demoiréing models typically employ image patches cropped from the entire image for training. However, given the significant variation of image size in different demoiréing datasets, the scale of content in cropped image patches of the same size also varies greatly. Hence, to

simulate this process and enhance the diversity of the training data, we perform multi-scale cropping (Figure 3 up). In particular, 4k images are resized to different sizes, from which we extract and randomly select image patches of uniform size as training data. In this way, the patches extracted from low and high-resolution images emphasize overall patterns and finer details, respectively.

Colorfulness-Sharpness Selection We notice that certain patches involve visually invisible patterns (with a “X” mark in Figure 3). They potentially confuse the generator during training, aiming to generate moiré pattern images rather than to reproduce plain white images. Hence, we filter out such patches based on colorfulness and sharpness. As depicted in Figure 3 lower-right, an increased sharpness value indicates more visible moiré patterns, while an increased colorfulness value signifies patterns with richer colors. The sharpness metric is calculated as the standard deviation of grayscaled input image processed with an edge filter, while the colorfulness metric is calculated as the average standard deviations of A and B channels in image LAB color space.

Learning Moiré Patterns in the Latent Space As shown in Figure 2(middle), plenty of pixels in the moiré pattern appear pure white. This leads to a polarization in the pixel distribution of the moiré pattern images, where informative data is concentrated in a few pixels with high values while the rest contains little information. Based on this observation, we choose to compress the moiré pattern into the latent space through an autoencoder for a more compact and efficient representation of its structural, textural, and color information. For better stability and controllability, we utilize the Latent Diffusion Model (Rombach et al. 2022) to effectively model the complex distribution of the moiré pattern in the latent space. Examples of generated moiré patterns are shown in Figure 2 right. More examples are in the appendix.

Moiré Image Synthesis

Via data collection and generation, we obtain a vast number of diverse moiré patterns. Then, we need to composite moiré patterns with clean images I_n to form moiré images. To make the synthesized images realistic-looking, We first create handcraft rules to produce initial moiré images in the Moiré Image Blending (MIB) module, then design a Tone Refinement Network (TRN) to further faithfully replicate the color and brightness variations observed in real scenes that cannot be fully formulated in those handcraft rules. The proposed synthesis process is illustrated in Figure 4.

Moiré Image Blending We blend the clean natural image I_n (background layer) with the moiré pattern I_{mp} (foreground layer) to form our initial moiré image I_{mib} . Notice that MoireSpace (Yang et al. 2023) synthesized their moiré image I'_{sm} via a Multiply Strategy $M(\cdot, \cdot)$,

$$I'_{sm} = M(I_{mp}, I_n) = I_{mp} \odot I_n, \quad (1)$$

where “ \odot ” denotes element-wise multiplication. However, the result produced by MoireSpace (Yang et al. 2023) tends to be dark and cannot replicate the desired contrast and color distortion, as shown in Figure 5. Therefore, we design the

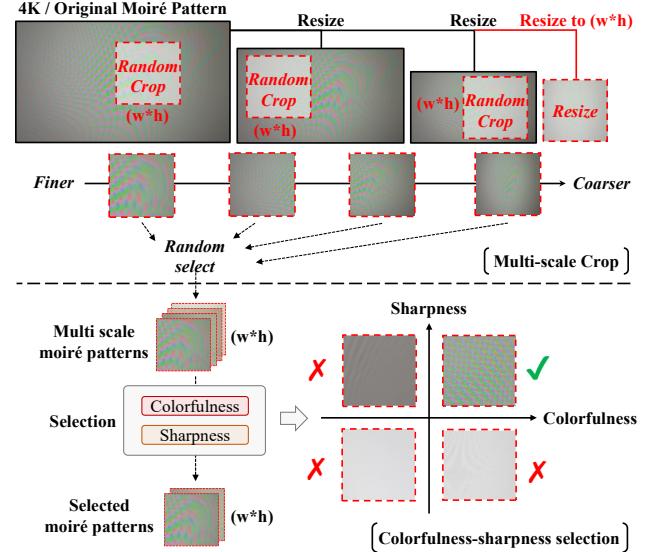


Figure 3: Data preprocessing for moiré pattern generation.

following handcraft rules to make the blending more realistic (Figure 4b). we first incorporate an additional blending strategy, Grain Merge (GIMP 2023) $G(\cdot, \cdot)$. Such a brighter strategy can balance the darker result from $M(\cdot, \cdot)$:

$$G(I_{mp}, I_n) = I_{mp} + I_n - 0.5. \quad (2)$$

Then, we incorporate transparency of the layers using alpha blending (Porter and Duff 1984) to obtain I_{comp}^M and I_{comp}^G :

$$I_{comp}^M = r_m \cdot M(I_{mp}, I_n) + [1 - r_m] \cdot I_n, \quad (3)$$

$$I_{comp}^G = r_g \cdot G(I_{mp}, I_n) + [1 - r_g] \cdot I_n. \quad (4)$$

where r_m and r_g represent the composition ratio parameter of foreground layer $M(I_{mp}, I_n)$ and $G(I_{mp}, I_n)$:

$$r_x = \frac{op_x}{op_x + (1 - op_x) \cdot op_n}, x \in \{m, g\}. \quad (5)$$

where op_m , op_g represent the opacity of the output layers from the Multiply and the Grain Merge strategies, and op_n represent the opacity of the background layer I_n . Finally, we perform a weighted (weight ω_m, ω_g) combination of I_{comp}^M and I_{comp}^G to obtain I_{mib} :

$$I_{mib} = \omega_m \cdot I_{comp}^M + \omega_g \cdot I_{comp}^G. \quad (6)$$

A visual comparison of MoireSpace result I'_{sm} , and our I_{mib} is shown in Figure 5, showing the superior of I_{mib} over I'_{sm} . Please refer to the appendix for more visual results.

Tone Refinement Network Though the moiré image blending module creates a preliminary moiré image I_{mib} , such a synthesized result based on handcraft rules still struggles to replicate accurate color and brightness changes. Comparatively, networks are more powerful in capturing such unknown changes and distortion by progressive learning. Hence, we present a learnable refinement network to synthesize more realistic results.

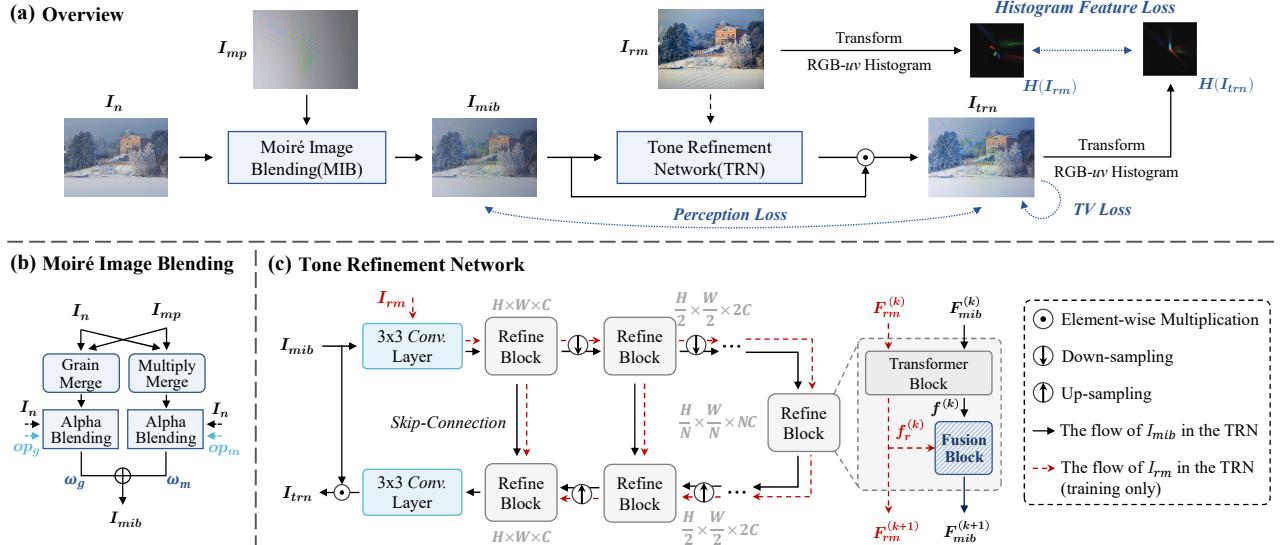


Figure 4: Overview of the Moiré Image Synthesis stage (a). It involves a Moiré Image Blending module (b) for initial moiré image synthesis and a Tone Refinement Network (c) to refine for more realistic results.

The Tone Refinement Network (TRN) proposed here is built on a U-shaped transformer backbone (Wang et al. 2022) incorporating multiple refine blocks, illustrated in Figure 4 (c). It takes I_{mib} as input, applies pixel-wise tone adjustment to I_{mib} , and minimizes the tone gap between the output I_{trn} and the given real moiré images I_{rm} . To be clear, TRN firstly applies a 3×3 convolutional layer with LeakyReLU to extract tone features $F_{mib}^{(0)}$, $F_{rm}^{(0)}$. Next, the feature maps $F_{mib}^{(0)}$ and $F_{rm}^{(0)}$ are passed through N encoder phases and N decoder phases with skip connections. Each phase contains a refine block to capture long-range dependencies, benefiting from the self-attention in Transformer.

Inspired by research in style transfer and domain generalization (Ulyanov, Vedaldi, and Lempitsky 2016; Huang and Belongie 2017; Zhou et al. 2021), we design a tone feature fusion block within each refine block to better fuse the tone feature statistics between I_{mib} and corresponding I_{rm} . It mixes the feature statistics of two instances with a random convex weight. As illustrated in Figure 4 (c), the computations inside a fusion block module in the k -th refine block can be summarized into two steps. First, given two sets of feature maps $f^{(k)}$ and $f_r^{(k)}$ for I_{mib} and I_{rm} , the fusion block generates a mixture of feature statistics,

$$\gamma_{mix} = \lambda \cdot \sigma(f^{(k)}) + (1 - \lambda) \cdot \sigma(f_r^{(k)}), \quad (7)$$

$$\beta_{mix} = \lambda \cdot \mu(f^{(k)}) + (1 - \lambda) \cdot \mu(f_r^{(k)}). \quad (8)$$

where μ and σ represent the mean and variance of feature maps, while λ is a random weight sampled from the beta distribution, $\lambda \in \text{Beta}(\alpha, \alpha)$ with $\alpha \in (0, \infty)$ being a hyper-parameter. Then, the mixture of feature statistics is applied to the tone-normalized $F_{mib}^{(k+1)}$:

$$F_{mib}^{(k+1)} = \gamma_{mix} \odot \frac{f^{(k)} - \mu(f^{(k)})}{\sigma(f^{(k)})} + \beta_{mix}. \quad (9)$$

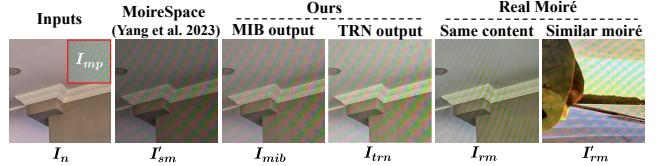


Figure 5: Visualization of our intermediate synthetic results.

The fusion block can effectively utilize the moiré feature information of I_{rm} and greatly helps reduce the moiré domain gap between the final synthesized image I_{trn} and real moiré image I_{rm} , which is one significant innovation. After the N decoder stages, we apply a 3×3 convolution layer on feature maps $F_{mib}^{(2N)}$ to obtain a tone refinement matrix M_{trn} . Finally, the synthesized image is obtained by $I_{trn} = I_{mib} \odot M_{trn}$ after color normalization, where “ \odot ” represents element-wise multiplication. Notice that the fusion block is solely utilized in the training phase, and I_{rm} is exclusively fed into the network during training. Figure 5 compares the initial blending result I_{mib} with the final synthesized result I_{trn} . Please refer to the appendix for more results.

Loss Functions The tone adjustment network aims to adjust the overall color tone and contrast of I_{trn} in a way that it resembles I_{rm} without affecting moiré pattern I_{mp} .

First, moiré patterns can disrupt image structures by generating strip-shaped artifacts (Yu et al. 2022). Therefore, comparing two moiré images directly in pixel space is less effective. Thus, we adopt the perceptual loss (Johnson, Alahi, and Fei-Fei 2016) \mathcal{L}_{per} to optimize the L_1 distance between the extracted content features of I_{mib} and I_{trn} :

$$\mathcal{L}_{per}(I_{trn}, I_{mib}) = \sum_{j=1}^{N_L} \frac{\|\phi_j(I_{trn}) - \phi_j(I_{mib})\|_1}{C_j H_j W_j}, \quad (10)$$

where $\phi_j(I)$ is the activations of the j -th layer of the

Test Dataset	Metric	Demoiréing Network: MBCNN					Demoiréing Network: ESDNet-L				
		Shooting	UnDeM [†]	UnDeM [‡]	MoireSpace	Ours	Shooting	UnDeM [†]	UnDeM [‡]	MoireSpace	Ours
UHDM	PSNR \uparrow	9.2284	13.4256	14.5237	14.7826	17.9162	10.2568	15.2269	15.2947	14.7989	17.2524
	SSIM \uparrow	0.5180	0.3973	0.4425	0.4724	0.6280	0.5664	0.5873	0.5777	0.4859	0.6454
	LPIPS \downarrow	0.6664	0.6489	0.6332	0.5568	0.4162	0.5130	0.4190	0.4241	0.5254	0.3238
FHDMI	PSNR \uparrow	10.6750	17.8355	18.1652	18.5523	19.0094	11.6022	18.4335	18.5390	18.0763	19.8128
	SSIM \uparrow	0.4478	0.6802	0.6999	0.7094	0.7137	0.5425	0.6900	0.6812	0.7189	0.7319
	LPIPS \downarrow	0.5978	0.2606	0.2472	0.2742	0.2390	0.4515	0.2877	0.2986	0.2616	0.2134

Table 2: Quantitative results of zero-shot demoiréing trained with synthesized data only. “ \dagger ” indicates UnDem uses moiré patterns retrieved from real data in TIP for inference. “ \ddagger ” indicates UnDem uses our generated moiré pattern for inference.

Cross Dataset		Metric	Demoiréing Network: MBCNN					Demoiréing Network: ESDNet-L				
Source	Target		Baseline	Shooting	UnDeM	MoireSpace	Ours	Baseline	Shooting	UnDeM	MoireSpace	Ours
UHDM	FHDMI	PSNR \uparrow	19.3848	19.2032	19.4676	19.4531	19.8625	20.3422	20.2407	20.4014	20.2806	20.7543
		SSIM \uparrow	0.7436	0.7459	0.7455	0.7496	0.7525	0.7599	0.7579	0.7510	0.7603	0.7653
		LPIPS \downarrow	0.3002	0.2975	0.2964	0.2993	0.2842	0.2525	0.2632	0.2509	0.2324	0.2136
	TIP	PSNR \uparrow	17.8107	18.3730	18.6674	18.9214	19.3922	18.8040	18.4543	19.3545	19.3964	19.5009
		SSIM \uparrow	0.6627	0.6888	0.6911	0.6996	0.7022	0.6921	0.6930	0.6998	0.7111	0.7149
		LPIPS \downarrow	0.3580	0.3886	0.3909	0.3829	0.3781	0.3524	0.3849	0.3601	0.3522	0.3495
FHDMI	UHDM	PSNR \uparrow	17.1331	17.5326	17.4870	17.6050	18.7931	18.0049	18.4189	17.9574	17.9751	18.9240
		SSIM \uparrow	0.6159	0.6334	0.6331	0.6642	0.7186	0.5755	0.5780	0.5857	0.5548	0.6658
		LPIPS \downarrow	0.4470	0.4350	0.4285	0.4020	0.3508	0.4420	0.4279	0.4460	0.4579	0.3405
	TIP	PSNR \uparrow	20.2161	20.7793	20.8261	20.1194	21.0694	20.6647	20.8678	20.4663	20.8107	21.5786
		SSIM \uparrow	0.7340	0.7304	0.7381	0.7347	0.7494	0.7504	0.7606	0.7278	0.7582	0.7668
		LPIPS \downarrow	0.2979	0.2884	0.2891	0.2961	0.2832	0.2459	0.2450	0.2998	0.2468	0.2310
TIP	UHDM	PSNR \uparrow	17.3409	17.4011	17.4407	17.4987	18.2937	17.4332	16.1836	16.8402	16.6296	18.4978
		SSIM \uparrow	0.6144	0.6062	0.6066	0.6059	0.6913	0.5523	0.5511	0.5692	0.5748	0.6866
		LPIPS \downarrow	0.4726	0.4487	0.4473	0.4412	0.3990	0.4987	0.4723	0.4532	0.4387	0.3231
	FHDMI	PSNR \uparrow	18.9458	19.2731	19.0336	19.1101	20.1053	19.2368	18.1936	19.2112	18.8385	19.9971
		SSIM \uparrow	0.7369	0.7399	0.7215	0.7321	0.7725	0.7354	0.7297	0.7499	0.7389	0.7580
		LPIPS \downarrow	0.2494	0.2447	0.2452	0.2382	0.2315	0.2316	0.2320	0.2130	0.2228	0.1915

Table 3: Quantitative results of cross-dataset evaluations.

VGG16 network (Simonyan and Zisserman 2014), and N_L represents the number of convolutional layers in VGG16.

In addition, to effectively tune the tone of I_{trn} , we take advantage of color differentiable RGB-uv histogram features $H(I_{trn})$ and $H(I_{rm})$ in log chromaticity space, inspired by color constancy method (Barron 2015; Afifi, Brubaker, and Brown 2021), as shown in Figure 4 (a). Such RGB-uv histograms have proven efficient in color transfer tasks (Eibenberger and Angelopoulou 2012). We optimize color loss using the differentiable Hellinger distances

$$\mathcal{L}_{color}(I_{trn}, I_{rm}) = \left\| H(I_{trn})^{1/2} - H(I_{rm})^{1/2} \right\|_2, \quad (11)$$

where $\|\cdot\|_2$ is the standard Euclidean norm and $.^{1/2}$ is an element-wise square root.

Finally, we use total variation regularizer \mathcal{L}_{tv} to remove unwanted details while encouraging spatial smoothness:

$$\mathcal{L}_{tv}(I_{trn}) = \sum_{i=1}^H \sum_{j=1}^W \left| I_{trn}^{i+1,j} - I_{trn}^{i,j} \right| + \left| I_{trn}^{i,j+1} - I_{trn}^{i,j} \right| \quad (12)$$

Total loss \mathcal{L} is then defined as a weighted compound of \mathcal{L}_{per} , \mathcal{L}_{color} and \mathcal{L}_{tv} :

$$\mathcal{L} = \lambda_{per} \mathcal{L}_{per} + \lambda_{color} \mathcal{L}_{color} + \lambda_{tv} \mathcal{L}_{tv}. \quad (13)$$

Image Demoiréing

Our contributions mainly lie in the above three stages. Then, diverse and realistic-looking data synthesized by our solution can be seamlessly integrated with demoiréing models to improve their performance.

Experiments

Experimental Setups

For all compared methods, we used their released code. Thorough implementation details are in the appendix.

Datasets and Metrics. 1) Moiré Pattern Dataset is used to train our moiré pattern generator. 2) Real Moiré Image Dataset, TIP (Sun, Yu, and Wang 2018), FHDMI (He et al. 2020), and UHDM (Yu et al. 2022), are used to demonstrate our ability in restoring real moiré images. 3) Evaluation Metrics.

We evaluate demoiréing performance on the Peak-Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM) (Wang et al. 2004), and LPIPS (Zhang et al. 2018).

Comparison Methods We compare UniDemoiré to the SOTA synthesis methods in 3 current modalities: the simulation method “Shooting” (Niu, Guo, and Wang 2021), the

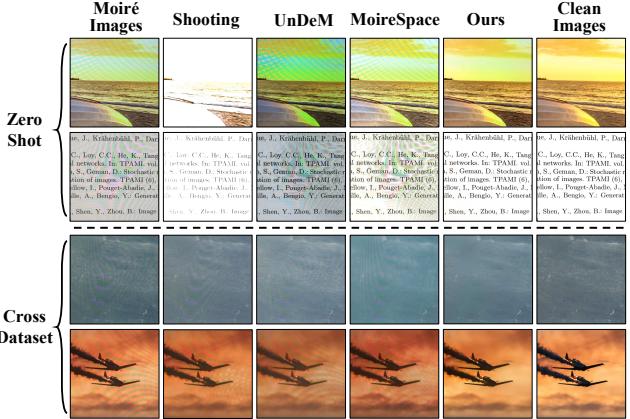


Figure 6: Comparisons of demoiréing results.

implicit moiré synthesis approach ‘‘UnDeM’’ (Zhong et al. 2024), which employs a neural network, and the explicit synthesis method termed ‘‘MoireSpace’’ (Yang et al. 2023), which utilizes its moiré pattern dataset.

Demoiréing Models We test on the most effective SOTA demoiréing models, MBCNN (Zheng et al. 2020) and ESDNet-L (Yu et al. 2022).

Zero-Shot Demoiréing with Synthesized Data Only

We first demonstrate demoiréing results on real moiré images trained on purely synthesized data by SOTA moiré synthesis methods. To avoid data overlap in training sets and test sets, we have collected a comprehensive Mixed High-Resolution Natural Image Dataset (MHRNID), based on which, moiré images are synthesized for training demoiréing models. Quantitative comparisons can be found in Table 2. Visual comparisons on demoiréing real data in UHDM are illustrated in Figure 6. Due to that UnDeM relies on existing moiré images in both the training (fusion networks) and inference phase, we trained their networks on the TIP dataset and showed the result of UnDeM using the real moiré in the TIP dataset (“†” in Table 2) and our sampled moiré pattern (“‡” in Table 2) during inference, respectively. For a fair comparison, we also use real moiré images from TIP dataset to train our TRN. Notice that UnDeM and our method only use real moiré images to guide the synthesis, and neither of us uses such real data to train demoiréing models directly.

From the quantitative perspective (Table 2), our method substantially outperforms all other approaches, particularly by more than 3.2 dB and 2.0 dB for MBCNN and ESDNet-L on the UHDM dataset, respectively. Besides, UnDeM[‡] using our generated moiré patterns outperforms UnDeM[†] using real moiré patterns in all experiments, proving our effectiveness further. From the qualitative perspective (Figure 6), our method demonstrates strong capability even when images in the target domain are contaminated by severe moiré patterns, which other synthesis methods fail to address. We attribute our superiority to the diversity and realism of our synthetic data. Such high-quality data by our UniDemoiré enables the demoiréing model to learn moiré characteristics

Components	PSNR↑	SSIM↑	LPIPS↓
ALL	20.7543	0.7653	0.2136
w/o MPG	20.1607	0.7326	0.2456
w/o TRN	20.1691	0.7372	0.2544
TRN (w/o \mathcal{L}_{per})	20.3076	0.7508	0.2558
TRN (w/o \mathcal{L}_{color})	20.2692	0.7406	0.2301
TRN (w/o \mathcal{L}_{tv})	20.3961	0.7451	0.2324
TRN (w/o fusion block)	20.2868	0.7370	0.2311

Table 4: Ablation studies. Source: UHDM, Target: FHDMD.

better, improving performance in removing unseen moiré artifacts. More visual results are in the appendix.

Cross-Dataset Evaluation

We then demonstrate our ability to improve the performance of demoiréing models across domains. Quantitative results are shown in Table 3. Note that ‘‘Baseline’’ means that the demoiréing models (MBCNN and ESDNet-L) are trained with the original source real moiré datasets and tested on the target dataset. For each synthesis approach, a demoiréing model is trained with combined original real data in the source dataset and corresponding synthesized data.

As shown, the Shooting method struggles with real data due to differences between synthetic and real moiré. UnDeM relies on a GAN network but can be inconsistent depending on the dataset and quality. The MoireSpace method performs better than UnDeM but has inferior moiré patterns and synthesis quality, resulting in lower experimental metrics. Thanks to the realistic and diverse synthesized data, our method outperforms all previous methods across every experiment. Visual comparisons in Figure 6 (lower, Source: UHDM, Target: FHDMD) demonstrate our effectiveness.

Ablation Study

We individually ablate submodules in our proposed method to analyze their contribution. All these experiments are trained with the UHDM dataset and validated on the FHDMD dataset. Experimental results in Table 4 verify that all components in our UniDemoiré solution are crucial for achieving the desired demoiréing performance. Removing any component such as the Moiré Pattern Generator (MPG), Tone Refinement Network (TRN), loss functions, and feature fusion block leads to a significant performance decline. More ablation studies are provided in the appendix.

Conclusion

By addressing the issue of data diversity and realism, our universal solution, UniDemoiré, tackles one of the most important bottlenecks in image demoiréing problems. It showcases significant performance in zero-shot demoiréing and demonstrates a strong capability of enhancing the cross-domain performance of existing demoiréing models. More importantly, our method holds the potential to generate billions of moiré data and to significantly expand demoiréing models with a vast increase in parameters. Our limitations are discussed in the appendix.

Acknowledgments

This work is supported by NSFC (No.62206173), Shanghai Frontiers Science Center of Human-centered Artificial Intelligence (ShangHAI), MoE Key Laboratory of Intelligent Perception and Human-Machine Collaboration (KLIP-HuMaCo). This work is also partially supported by HKU-SCF FinTech Academy, HKRGC Theme-based research scheme project T35-710/20-R, and SZ-HK-Macau Technology Research Programme #SGDX20210823103537030.

References

- Afifi, M.; and Brown, M. S. 2019. Sensor-Independent Illumination Estimation for DNN Models. *arXiv:1912.06888*.
- Afifi, M.; Brubaker, M. A.; and Brown, M. S. 2021. HistogramGAN: Controlling Colors of GAN-Generated and Real Images via Color Histograms. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Afifi, M.; Price, B.; Cohen, S.; and Brown, M. S. 2019. When Color Constancy Goes Wrong: Correcting Improperly White-Balanced Images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Amidror, I. 2009. *The Theory of the Moiré Phenomenon: Volume I: Periodic Layers*, volume 38. Springer Science & Business Media.
- Barron, J. T. 2015. Convolutional Color Constancy. *arXiv:1507.00410*.
- Dhariwal, P.; and Nichol, A. 2021. Diffusion Models Beat GANs on Image Synthesis. *arXiv:2105.05233*.
- Dosovitskiy, A.; and Brox, T. 2016. Generating Images with Perceptual Similarity Metrics based on Deep Networks. *Neural Information Processing Systems, Neural Information Processing Systems*.
- Eibenberger, E.; and Angelopoulou, E. 2012. The importance of the normalizing channel in log-chromaticity space. In *2012 19th IEEE International Conference on Image Processing*.
- Esser, P.; Rombach, R.; and Ommer, B. 2021. Taming Transformers for High-Resolution Image Synthesis. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Fei, B.; Lyu, Z.; Pan, L.; Zhang, J.; Yang, W.; Luo, T.; Zhang, B.; and Dai, B. 2023. Generative Diffusion Prior for Unified Image Restoration and Enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9935–9946.
- GIMP. 2023. Layer Blending Modes. [Online; accessed 10-April-2024].
- He, B.; Wang, C.; Shi, B.; and Duan, L.-Y. 2019. Mop moire patterns using mopenet. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2424–2432.
- He, B.; Wang, C.; Shi, B.; and Duan, L.-Y. 2020. FHDe 2 Net: Full High Definition Demoiréing Network. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXII 16*, 713–729. Springer.
- Hosu, V.; Agnolucci, L.; Wiedemann, O.; and Iso, D. 2024. UHD-IQA Benchmark Database: Pushing the Boundaries of Blind Photo Quality Assessment. *arXiv:2406.17472*.
- Huang, X.; and Belongie, S. 2017. Arbitrary Style Transfer in Real-time with Adaptive Instance Normalization. In *2017 IEEE International Conference on Computer Vision (ICCV)*.
- Johnson, J.; Alahi, A.; and Fei-Fei, L. 2016. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, 694–711. Springer.
- Kingma, D. P.; and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Lee, J.; Son, H.; Rim, J.; Cho, S.; and Lee, S. 2021. Iterative Filter Adaptive Network for Single Image Defocus Deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2034–2042.
- Li, H.; Li, J.; Zhao, D.; and Xu, L. 2021. Dehazeflow: Multi-scale conditional flow network for single image dehazing. In *Proceedings of the 29th ACM International Conference on Multimedia*, 2577–2585.
- Liu, F.; Yang, J.; and Yue, H. 2015. Moiré pattern removal from texture images via low-rank and sparse matrix decomposition. In *2015 Visual Communications and Image Processing (VCIP)*, 1–4. IEEE.
- Liu, L.; An, J.; Yuan, S.; Zhou, W.; Li, H.; Wang, Y.; and Tian, Q. 2024. Video Demoiréing with Deep Temporal Color Embedding and Video-Image Invertible Consistency. *IEEE Transactions on Multimedia*.
- Liu, L.; Liu, J.; Yuan, S.; Slabaugh, G.; Leonardis, A.; Zhou, W.; and Tian, Q. 2020. Wavelet-based dual-branch network for image demoiréing. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIII 16*, 86–102. Springer.
- Loshchilov, I.; and Hutter, F. 2016. Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*.
- Loshchilov, I.; and Hutter, F. 2019. Decoupled Weight Decay Regularization. *arXiv:1711.05101*.
- Luo, X.; Zhang, J.; Hong, M.; Qu, Y.; Xie, Y.; and Li, C. 2020. Deep wavelet network with domain adaptation for single image demoiréing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 420–421.
- Luo, Z.; Gustafsson, F. K.; Zhao, Z.; Sjölund, J.; and Schön, T. B. 2023. Refusion: Enabling large-size realistic image restoration with latent-space diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1680–1691.
- Niu, D.; Guo, R.; and Wang, Y. 2021. Morié attack (ma): A new potential risk of screen photos. *Advances in Neural Information Processing Systems*, 34: 26117–26129.
- Niu, Y.; Lin, Z.; Liu, W.; and Guo, W. 2023. Progressive Moire Removal and Texture Complementation for Image Demoiréing. *IEEE Transactions on Circuits and Systems for Video Technology*.

- Odena, A.; Dumoulin, V.; and Olah, C. 2016. Deconvolution and Checkerboard Artifacts. *Distill*.
- Park, H.; Vien, A. G.; Kim, H.; Koh, Y. J.; and Lee, C. 2022. Unpaired screen-shot image demoiréing with cyclic moiré learning. *IEEE Access*, 10: 16254–16268.
- Porter, T.; and Duff, T. 1984. Compositing digital images. *ACM SIGGRAPH Computer Graphics*, 253–259.
- Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; and Ommer, B. 2022. High-Resolution Image Synthesis with Latent Diffusion Models. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, 234–241. Springer.
- Simonyan, K.; and Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Song, J.; Meng, C.; and Ermon, S. 2022. Denoising Diffusion Implicit Models. *arXiv:2010.02502*.
- Sun, B.; Li, S.; and Sun, J. 2014. Scanned image descreening with image redundancy and adaptive filtering. *IEEE transactions on image processing*, 23(8): 3698–3710.
- Sun, Y.; Yu, Y.; and Wang, W. 2018. Moiré photo restoration using multiresolution convolutional neural networks. *IEEE Transactions on Image Processing*, 27(8): 4160–4172.
- Ulyanov, D.; Vedaldi, A.; and Lempitsky, V. 2016. Instance Normalization: The Missing Ingredient for Fast Stylization. *arXiv: Computer Vision and Pattern Recognition,arXiv: Computer Vision and Pattern Recognition*.
- Wang, C.; He, B.; Wu, S.; Wan, R.; Shi, B.; and Duan, L.-Y. 2023a. Coarse-to-fine Disentangling Demoiréing Framework for Recaptured Screen Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Wang, J.; Chen, K.; Xu, R.; Liu, Z.; Loy, C. C.; and Lin, D. 2019. CARAFE: Content-Aware ReAssembly of FFEatures. *arXiv:1905.02188*.
- Wang, T.; Zhang, K.; Shen, T.; Luo, W.; Stenger, B.; and Lu, T. 2023b. Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 2654–2662.
- Wang, X.; Xie, L.; Dong, C.; and Shan, Y. 2021. Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 1905–1914.
- Wang, Z.; Bovik, A. C.; Sheikh, H. R.; and Simoncelli, E. P. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4): 600–612.
- Wang, Z.; Cun, X.; Bao, J.; Zhou, W.; Liu, J.; and Li, H. 2022. Uformer: A General U-Shaped Transformer for Image Restoration. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Xing, W.; and Egiazarian, K. 2021. End-to-End Learning for Joint Image Demosaicing, Denoising and Super-Resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3507–3516.
- Yang, C.; Yang, Z.; Ke, Y.; Chen, T.; Grzegorzek, M.; and See, J. 2023. Doing More With Moiré Pattern Detection in Digital Photos. *IEEE Transactions on Image Processing*, 32: 694–708.
- Yang, J.; Liu, F.; Yue, H.; Fu, X.; Hou, C.; and Wu, F. 2017a. Textured image demoiréing via signal decomposition and guided filtering. *IEEE Transactions on Image Processing*, 26(7): 3528–3541.
- Yang, J.; Zhang, X.; Cai, C.; and Li, K. 2017b. Demoiréing for screen-shot images with multi-channel layer decomposition. In *2017 IEEE Visual Communications and Image Processing (VCIP)*, 1–4. IEEE.
- Yu, J.; Li, X.; Koh, J.; Zhang, H.; Pang, R.; Qin, J.; Ku, A.; Xu, Y.; Baldridge, J.; and Wu, Y. 2021. Vector-quantized Image Modeling with Improved VQGAN. *Cornell University - arXiv,Cornell University - arXiv*.
- Yu, X.; Dai, P.; Li, W.; Ma, L.; Shen, J.; Li, J.; and Qi, X. 2022. Towards efficient and scale-robust ultra-high-definition image demoiréing. In *European Conference on Computer Vision*, 646–662. Springer.
- Yuan, S.; Timofte, R.; Slabaugh, G.; Leonardis, A.; Zheng, B.; Ye, X.; Tian, X.; Chen, Y.; Cheng, X.; Fu, Z.; et al. 2019. Aim 2019 challenge on image demoireing: Methods and results. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, 3534–3545. IEEE.
- Yue, H.; Cheng, Y.; Mao, Y.; Cao, C.; and Yang, J. 2022. Recaptured screen image demoiréing in raw domain. *IEEE Transactions on Multimedia*.
- Zhang, C.; Zhu, Y.; Yan, Q.; Sun, J.; and Zhang, Y. 2023. All-in-one multi-degradation image restoration network via hierarchical degradation representation. In *Proceedings of the 31st ACM International Conference on Multimedia*, 2285–2293.
- Zhang, R.; Isola, P.; Efros, A. A.; Shechtman, E.; and Wang, O. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 586–595.
- Zheng, B.; Yuan, S.; Slabaugh, G.; and Leonardis, A. 2020. Image demoireing with learnable bandpass filters. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3636–3645.
- Zheng, B.; Yuan, S.; Yan, C.; Tian, X.; Zhang, J.; Sun, Y.; Liu, L.; Leonardis, A.; and Slabaugh, G. 2021. Learning frequency domain priors for image demoireing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11): 7705–7717.
- Zhong, Y.; Zhou, Y.; Zhang, Y.; Chao, F.; and Ji, R. 2024. Learning Image Demoireing from Unpaired Real Data. *arXiv preprint arXiv:2401.02719 (AAAI2024)*.
- Zhou, K.; Yang, Y.; Qiao, Y.; and Xiang, T. 2021. MixStyle Neural Networks for Domain Generalization and Adaptation. *Cornell University - arXiv,Cornell University - arXiv*.

Zhu, Y.; Zhang, K.; Liang, J.; Cao, J.; Wen, B.; Timofte, R.; and Van Gool, L. 2023. Denoising Diffusion Models for Plug-and-Play Image Restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1219–1229.

Technical Appendix

This document supplements the main body of our paper with additional details, discussions, and results. In Section A, we present more details of the Moiré Pattern Dataset collection, including a brief analysis of various previously overlooked factors affecting moiré pattern diversity. In Section B, we will provide a detailed explanation of the two stages involved in implementing UniDemoiré: Moiré Pattern Generator and Moiré Image Synthesis. In Section C, we provide more implementation details of experiments and show more qualitative results. Furthermore, as shown in Section C.4, we performed additional ablation experiments on the blending strategy in the Moiré Image Blending (MIB) module and the design of the upsampling block and the loss function in the Tone Refinement Network (TRN).

A Dataset Capture and Analysis

In this section, we first present a brief introduction of various previously overlooked factors of devices that affect moiré pattern diversity. Then, we provide more details about our capture settings.

A.1 The Impact of Device on Moiré Pattern Diversity

Previous studies (Yu et al. 2022; Yang et al. 2023) have indicated that the geometric correlation between the screen and the camera significantly influences the features of the moiré pattern. However, such studies have overlooked that some aspects of the camera and the screen can also impact the moiré pattern.

For cameras, the two most critical factors affecting the moiré pattern are the CMOS and the lens used. The pixel density of a CMOS sensor (i.e., the number of pixels per unit area) determines its maximum sampling frequency, also known as the Nyquist frequency. The higher the pixel density, the higher the sampling frequency of the sensor and the higher the frequency of the signal that can be sampled, resulting in a higher frequency of moiré produced by the aliasing effect, which impacts the moiré pattern. In addition, the lens’s focal length also affects the formation of moiré. In cell phone photography, lenses with shorter focal lengths (e.g., wide lenses/main camera lenses) usually have wider angles of view and can capture more of the scene content. Lenses with longer focal lengths (such as telephoto or telescopic lenses), on the other hand, offer a narrower angle of view and greater magnification for capturing distant details. When the screen is photographed with lenses of different focal lengths, the relative positional relationship between the pixels on the sensor and the pixels on the screen changes, which may cause the moiré pattern to appear or disappear.

Furthermore, the layout and the distance of pixels dots in the panel used can also significantly impact the formation of moiré on the display screen. The frequency of detail that a screen can display depends on how the pixel dots are arranged. Various arrangements result in distinct frequencies of detail, which impacts the formation of moiré patterns. The distance between pixel dots on the screen then affects the shooting distance. Larger pixel dot spacing will make the

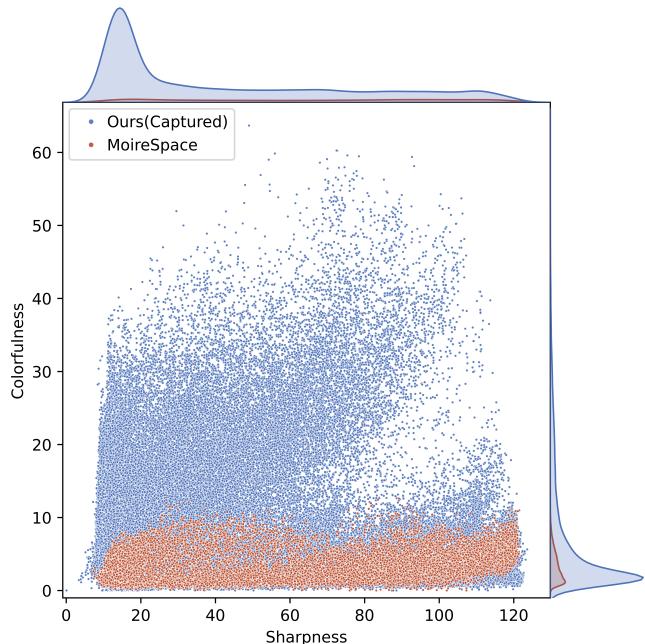


Figure 7: Comparison of sharpness and colorfulness between our Moiré Pattern Dataset and MoireSpace.

distance at which the moiré is formed to be photographed farther away. Conversely, the smaller the distance between pixel dots, the closer the distance needed to photograph the molded moiré.

A.2 More Details about Capture Settings

Based on the above analysis, we take screen images through different camera viewpoints to generate diverse moiré patterns. Specifically, we apply six mobile phones and six digital screens, as shown in Table 5 and 6 ($6 \times 6 = 36$ combinations). Figure 7 presents a scatter plot comparison between our 4K Moiré Pattern Dataset and the MoireSpace dataset in terms of texture definition and color vibrancy, where each data point represents an individual sample. As illustrated in the visualization, our captured moiré patterns demonstrate a more comprehensive coverage in both the sharpness of textural details and the chromatic saturation range. This comparative analysis is further substantiated by the representative samples juxtaposed in Figure 8, where side-by-side comparisons visually confirm that our dataset encompasses a broader diversity of morphological textures and a wider gamut of color expressions.

Mobile Phones We chose six mobile phones with varying camera specifications to capture diverse moiré patterns, as shown in Table 5. Our selection criteria included the camera type, CMOS category, and number of megapixels. For the regular main camera with moderate resolution, we picked the iPhone 12 and iPhone 13. For electronic zooming at 2x and 3x, we selected the Honor 90 and Xiaomi 10s, which have high pixels. Additionally, we picked two iPhone 12 Pro and iPhone 15 Pro models with different CMOS specifically

Samples from MoireSpace (Yang et al. 2023)



Samples from Our 4K Moiré Pattern Dataset

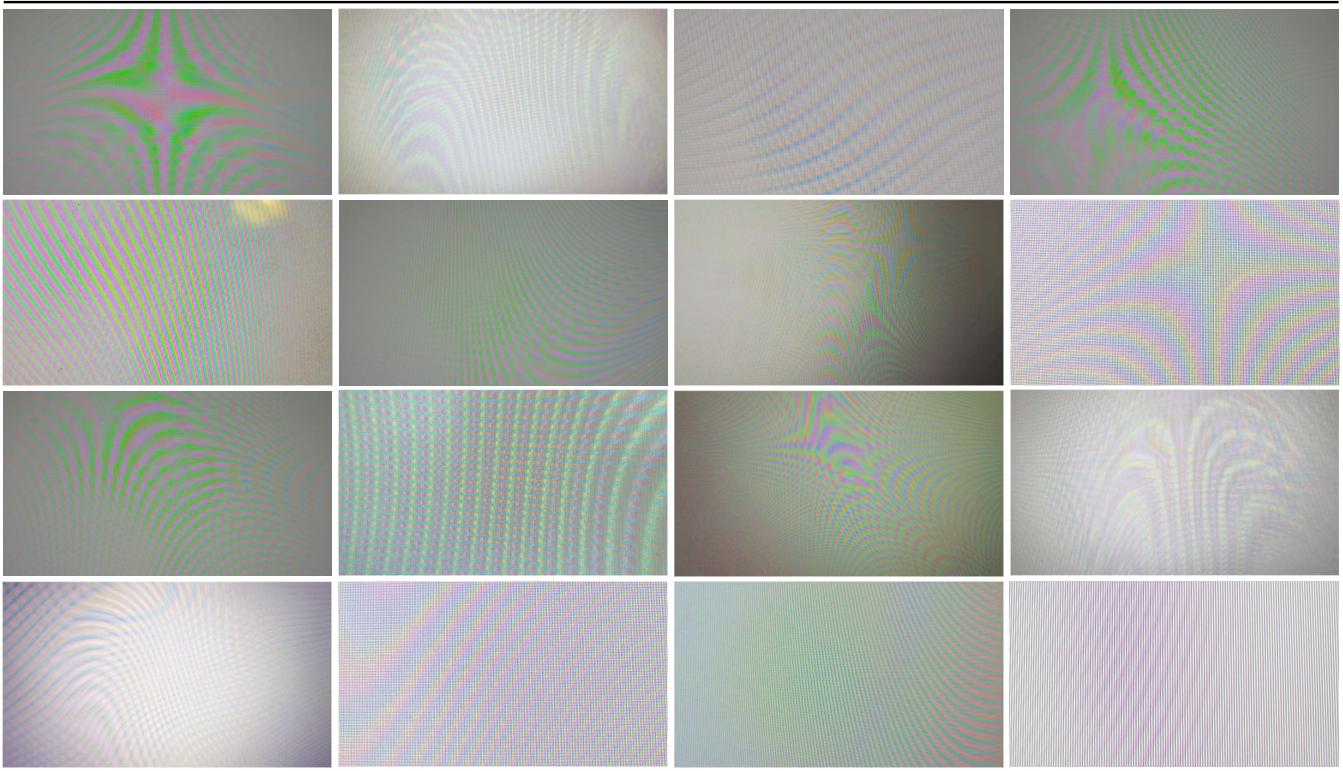


Figure 8: Samples from MoireSpace (Yang et al. 2023) and our 4K Moiré Pattern Dataset.

Mobile Phone	Camera	CMOS	MegaPixel (MP)	Optical format (Inches)	Pixel Size (μm)
iPhone 12	Main	SONY IMX503	12	1/2.55	1.40
iPhone 13	Main	SONY IMX603	12	1/1.88	1.70
Honor 90	Main	ISOCELL HP3	200	1/1.40	0.56
Xiaomi 10s	Main	ISOCELL HMX	108	1/1.33	0.80
iPhone 12 Pro	Main Telephoto	SONY IMX503 SONY IMX613 (2x zoom)	12 12.2	1/2.55 1/3.40	1.40 1.00
iPhone 15 Pro	Main Telephoto	SONY IMX803 SONY IMX713 (2x/4x zoom)	48 12	1/1.28 1/3.40	1.22 1.00

Table 5: The mobile phone we apply to get the moiré patterns

Digital Screen	Size (Inches)	Panel type	Resolution	Brightness (cd/m^2)	Coating
DELL D2720DS	27	IPS(LED)	2560×1440	350	Matte
Macbook Air 2022	13.3	IPS(LED)	2560×1600	500	Glossy
AOC 27G2G8	27	IPS(W-LED)	2560×1440	250	Matte
Philips 27E1N5500	27	IPS(LED)	2560×1440	300	Matte
Xiaomi C34WQBA-RG	34	Curved SVA(W-LED)	3440×1440	300	Matte
ViewSonic VX2771-4K-HD	27	IPS(LED)	3840×2160	350	Matte

Table 6: The screen we apply to get the moiré patterns

for telephoto lenses. These models use the telephoto lens for optical zoom at fixed magnifications of 2x and 3x.

Display Screens To capture a wider variety of moiré patterns in different forms, we selected display screens based on size, panel type, and resolution guidelines to maximize pixel point layouts and spacing on the screen. As shown in Table 6, we have selected three 27-inch IPS panel LED matte screen monitors with a 2K resolution - DELL D2720DS, AOC 27G2G8, and Philips 27E1N5500. This specification is the most common among the available options. The AOC 27G2G8 is a W-LED monitor with an RGBW pixel layout. This IPS screen has white sub-pixels in addition to the standard RGB arrangement, creating a more varied pixel point layout. To capture the moiré pattern on the glossy display, we opted for a 13.3-inch IPS panel with a 2K resolution MacBook Air notebook. Finally, we selected two high-resolution displays: the Xiaomi C34WQBA-RG and the ViewSonic VX2771-4K-HD. These displays were explicitly chosen to capture moiré patterns with smaller pixel dot spacing. The Xiaomi C34WQBA-RG is a 34-inch curved display with an SVA panel and W-LED technology. It boasts a 3K resolution. On the other hand, the ViewSonic VX2771-4K-HD is a 27-inch matte screen display with an IPS panel and LED technology. It offers a standard 4K resolution.

B Further details of our Method

This section will showcase the details of the implementation of our UniDemoiré’s Moiré Pattern Generator stage and Moiré Image Synthesis stage.

B.1 Moiré Pattern Generator

The visualization of moiré pattern patches generated using the Moiré Pattern Generator(MPG) is shown in Figure 9.

Algorithm 1: Data Preprocessing in MPG.

```

Input: 4K Moiré Pattern Dataset  $\mathcal{D}_{mp}$ , patch size
         $(w, h)$ .
Output: Selected moiré pattern patch  $I_{mp}$ .
1 while True do
2   Randomly select a 4K moiré pattern  $\mathcal{I} \in \mathcal{D}_{mp}$ .
3   for  $i = 1$  to  $n$  do
4     1. Multi-Scale Cropping:
5       Randomly select probability  $p_1, p_2$ .
6       if  $p_1 \leq 50\%$  then
7          $I_{mp} \leftarrow$  Random Crop( $\mathcal{I}, w, h$ ).
8       else if  $p_2 \leq 33.33\%$  then
9          $\mathcal{I} \leftarrow$  Resize( $\mathcal{I}, 2560, 1440$ ),
10         $I_{mp} \leftarrow$  Random Crop( $\mathcal{I}, w, h$ ).
11       else if  $33.33\% < p_2 \leq 66.66\%$  then
12          $\mathcal{I} \leftarrow$  Resize( $\mathcal{I}, 1920, 1080$ ),
13          $I_{mp} \leftarrow$  Random Crop( $\mathcal{I}, w, h$ ).
14       else  $I_{mp} \leftarrow$  Resize( $\mathcal{I}, w, h$ ).
15     2. Sharpness-Colorfulness selection:
16      $G_{mp} \leftarrow$  RGB_to_Gray( $I_{mp}$ ),
17      $L_{mp}, A_{mp}, B_{mp} \leftarrow$  RGB_to_LAB( $I_{mp}$ ),
18     Sharpness  $\leftarrow \sigma(\mathcal{F} * G_{mp})$ ,
19     Colorfulness  $\leftarrow \sqrt{\sigma(A_{mp})^2 + \sigma(B_{mp})^2}$ ,
20     if Sharpness  $\geq \delta_s$  and Colorfulness  $\geq \delta_c$ 
21     then
22       return  $I_{mp}$ .
23   end
24 end

```

The details of the data preprocessing and networking implementations of MPG are described below.

The implementation details of data preprocessing The details of our data preprocessing method in the Moiré Pattern Generator(MPG) are described in Algorithm 1. In Multi-Scale Cropping, “Random Crop(\mathcal{I}, w, h)” means to randomly crop a patch of size $w \times h$ from \mathcal{I} , and “Resize(\mathcal{I}, w, h)” means to resize the width and height of \mathcal{I} to w and h directly. In Sharpness-Colorfulness selection, “RGB_to_Gray(I_{mp})” refers to convert I_{mp} to grayscale image G_{mp} , while “RGB_to_LAB(I_{mp})” refers to convert I_{mp} to LAB space and retrieve the corresponding channel matrices L_{mp} , A_{mp} , and B_{mp} respectively. Moreover, “ $\mathcal{F} * G_{mp}$ ” denotes the convolution operation on the grayscale image G_{mp} using the Laplace edge detection operator \mathcal{F} . In the actual training process of MPG, we specify set n to 3 and (w, h) to $(768, 768)$, while δ_s and δ_c are set to 15 and 2, respectively.

The implementation details of Latent Diffusion Model We utilize the Latent Diffusion Model(LDM) (Rombach et al. 2022) as the network component of the Moiré Pattern Generator.

Firstly, we have trained our autoencoder model for moiré patterns according to the method described in (Rombach et al. 2022). Specifically, given a moiré pattern patch $I_{mp} \in \mathbb{R}^{w \times h \times 3}$ that has gone through Multi-Scale Cropping and Sharpness-Colorfulness selection, we utilize an Encoder \mathcal{E} to convert I_{mp} to the latent space $z = \mathcal{E}(I_{mp})$ through multiple downsampling blocks. Simultaneously, we expect the corresponding Decoder \mathcal{D} to reconstruct the moiré pattern from the latent space variable $z : I_{mp} = \mathcal{D}(z) = \mathcal{D}(\mathcal{E}(I_{mp}))$ by using the same upsampling factor. Note that the overall downsampling factor is denoted as $f = h/h_0 = w/w_0$, where h_0 and w_0 are hyperparameters chosen to ensure that f is precisely 2^m , with $m \in \mathbb{N}$. Our loss function is a combination of a perceptual loss function \mathcal{L}_{rec} (Zhang et al. 2018) and patch-based adversarial targets \mathcal{L}_{adv} (Dosovitskiy and Brox 2016; Esser, Rombach, and Ommer 2021; Yu et al. 2021), along with a KL-reg regularization term \mathcal{L}_{reg} where the patch-based discriminator D_ψ we used is optimized to differentiate between the original moiré pattern I_{mp} and the reconstructed moiré pattern $\mathcal{D}(\mathcal{E}(I_{mp}))$. The full objective to train the autoencoder (\mathcal{E}, \mathcal{D}) is:

$$\begin{aligned} \mathcal{L} = & \min_{\mathcal{E}, \mathcal{D}} \max_{\psi} (\mathcal{L}_{rec}(I_{mp}, \mathcal{D}(\mathcal{E}(I_{mp}))) + \log D_\psi(I_{mp}) \\ & - \mathcal{L}_{adv}(\mathcal{D}(\mathcal{E}(I_{mp}))) + \mathcal{L}_{reg}(I_{mp}; \mathcal{E}, \mathcal{D})) \end{aligned} \quad (14)$$

The network structure of our \mathcal{E} and \mathcal{D} are the same as the autoencoder in (Rombach et al. 2022). To compress the moiré pattern as much as possible, we use the downsampling factor $f = 32$ and the number of hidden-space channels 64, which gives the latent variable z a dimension of $64 \times 64 \times 24$. We adopted 6 downsampling/upsampling blocks in \mathcal{E} and \mathcal{D} . Each downsampling/upsampling block contains two layers of ResBlock as well as one layer of multi-head self-attention block, and the list of channel scaling multipliers is [1,1,2,2,4,4]. We trained the autoencoder on 8 NVIDIA A40 GPUs, with a batch size of 2 on each GPU and a learning rate of 4.5e-6. Both the autoencoder (\mathcal{E} and \mathcal{D}) and the

discriminator (D_ψ) in the loss functions are optimized by Adam (Kingma and Ba 2014) with $\beta_1 = 0.5$ and $\beta_2 = 0.9$. In total, we trained the autoencoder for 35 Epochs.

Subsequently, we adopt the diffusion model to modify the complex distribution $p(z)$ obeyed by the latent variable z after the \mathcal{E} transformation with the objective function:

$$\mathcal{L} = \mathbb{E}_{\mathcal{E}(I_{mp}), \epsilon, t} [\|\epsilon - \epsilon_\theta(\alpha_t \mathcal{E}(I_{mp}) + \sigma_t \epsilon, t)\|_1]. \quad (15)$$

where $\epsilon \sim \mathcal{N}(0, \mathbb{I})$ is the variable sampled by the standard Gaussian distribution, and ϵ_θ is the noisy prediction network parameterized by θ , where we implemented it by using UNet (Ronneberger, Fischer, and Brox 2015) which integrates the time-step conditioning variable t . The α_t is the value at step t of the signal-to-noise ratio.

The network structures of the diffusion model in latent space are the same as those of the unconditional model in LDM (Rombach et al. 2022). For the UNet model ϵ_θ , we set the number of channels to 192, and the encoder and decoder of the UNet contain 4 downsampling/upsampling blocks, and the structure of those blocks are kept the same in the autoencoder. The list of channel scaling multipliers is [1,2,4,8]. The model is trained on 8 NVIDIA A40 GPUs for 50 epochs and optimized by AdamW (Loshchilov and Hutter 2019) with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The batch size on each GPU is set to 2, and the learning rate is initially set to 1e-4 and scheduled by linear warmup on the first 10,000 steps. The total learning rate complies with the linear multiplication in the LDM (Rombach et al. 2022) based on the number of GPUs and the batch size. In the training stage, we set the diffusion steps to 1000 and utilized the linear noise schedule to add noise to the latent variable z . We utilize the DDIM sampler (Song, Meng, and Ermon 2022) to accelerate sampling after training, using 200 sampling steps. We sampled 100,000 moiré patterns using a single NVIDIA A40 GPU, and some of the samples are shown in Figure 9.

B.2 Moiré Image Synthesis

In this section, we will demonstrate the implementation details of the Moiré Image Synthesis stage that were omitted in our main paper. Additionally, we include more visualizations of the synthesis results in Figure 10.

Implementations of the Moiré Image Blending For the MIB module, ω_m in Eq. (5) is randomly selected from [0.65, 0.75], while $\omega_g = 1 - \omega_m$. The op_m and op_g in Eq. (6) are set to 1.0 and 0.8, respectively. Performance changes resulting from the use of both the Multiply and Grain Merge strategy are detailed in the additional ablation study in Section C.4.

Implementations of the Tone Refinement Network We implement the backbone of our Tone Refinement Network(TRN) using Uformer-T(Tiny) (Wang et al. 2022), where the Transformer Block uses the Locally-enhanced Window (LeWin) Transformer block proposed by Uformer and sets the window size to 8×8 . At the same time, we change the encoder depth from {2,2,2,2} to {1,1,1,1}. Performance changes resulting from the use of the Uformer are detailed in the additional ablation study in Section C.4.

In the context of TRN, utilizing the transposed convolutional upsampling block similar to Uformer may lead to the

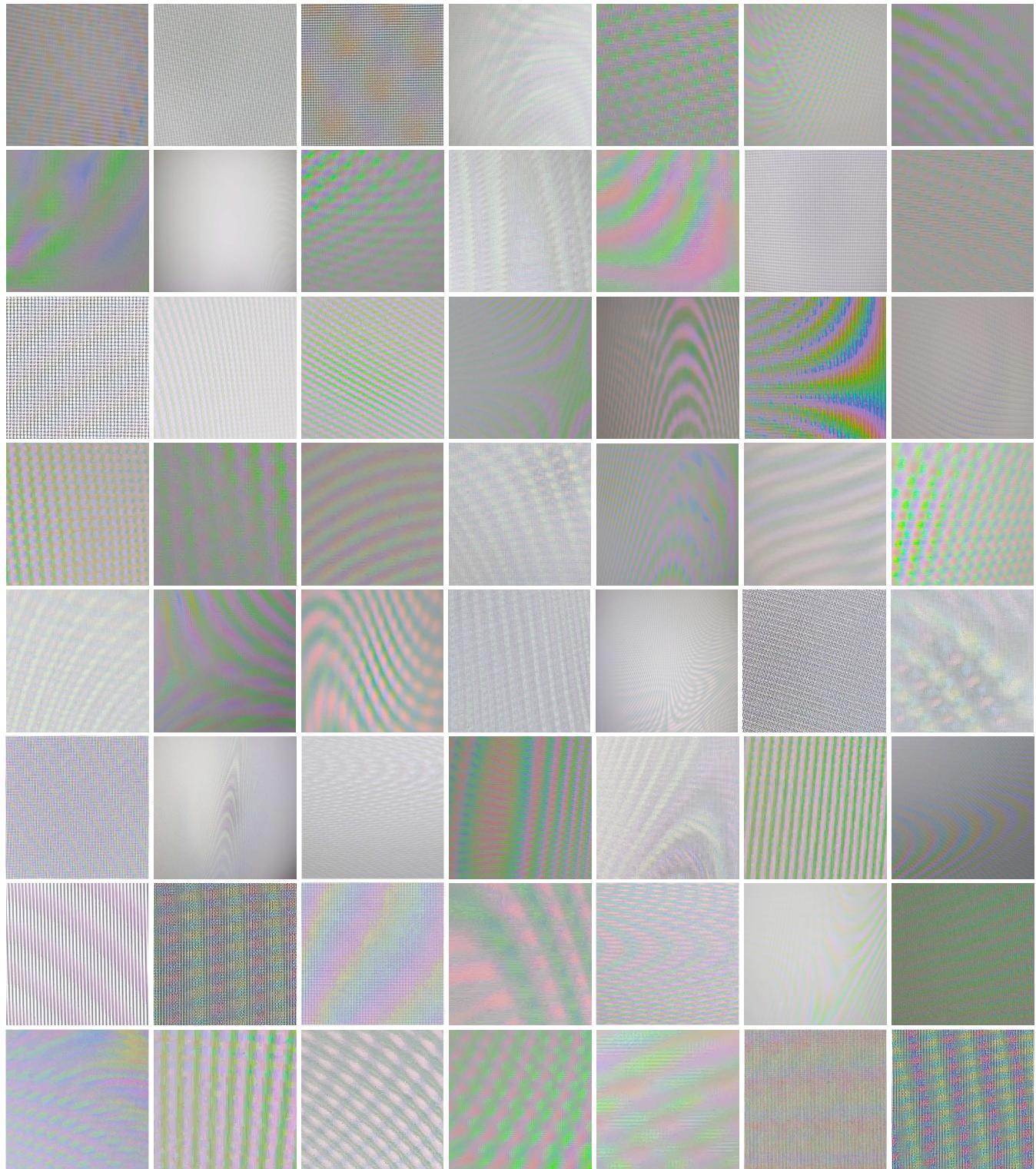


Figure 9: Visualization of sampled patches using our Moiré Pattern Generator.

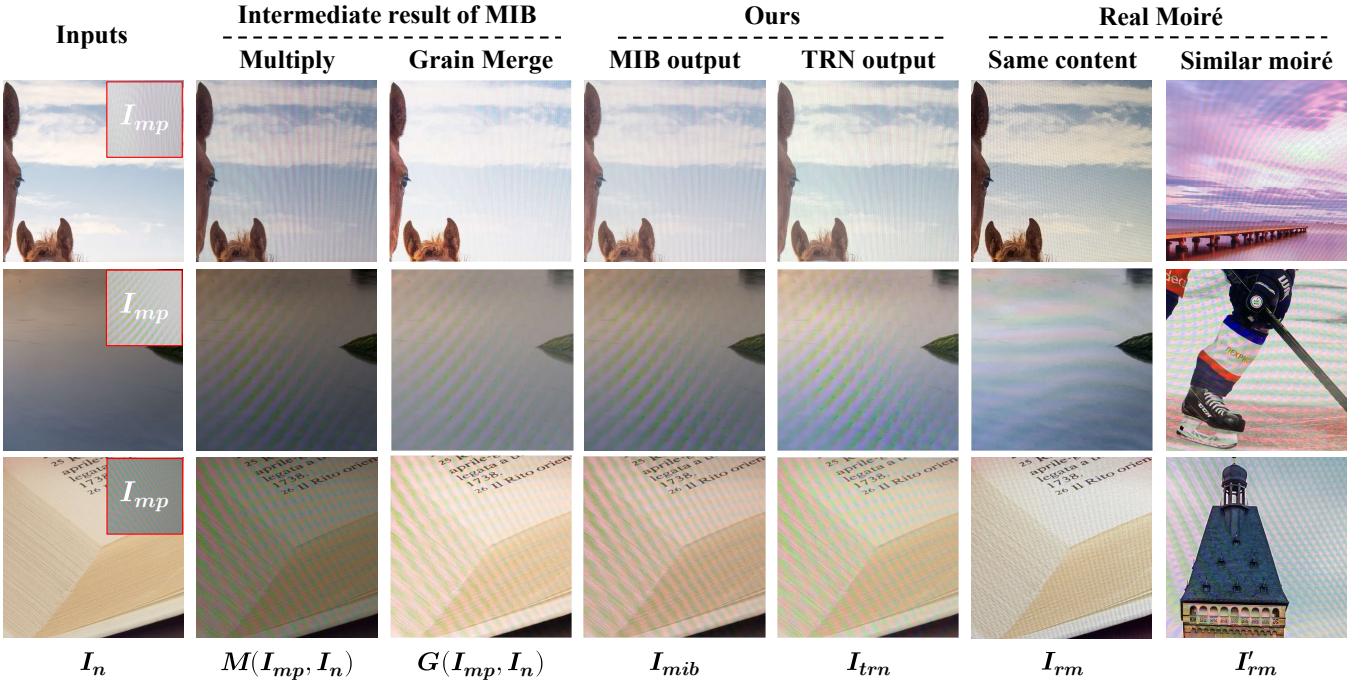


Figure 10: Visualization of our intermediate synthetic results. The final synthesis of I_{trn} best resembles the real moiré images in contrast and brightness distortions.

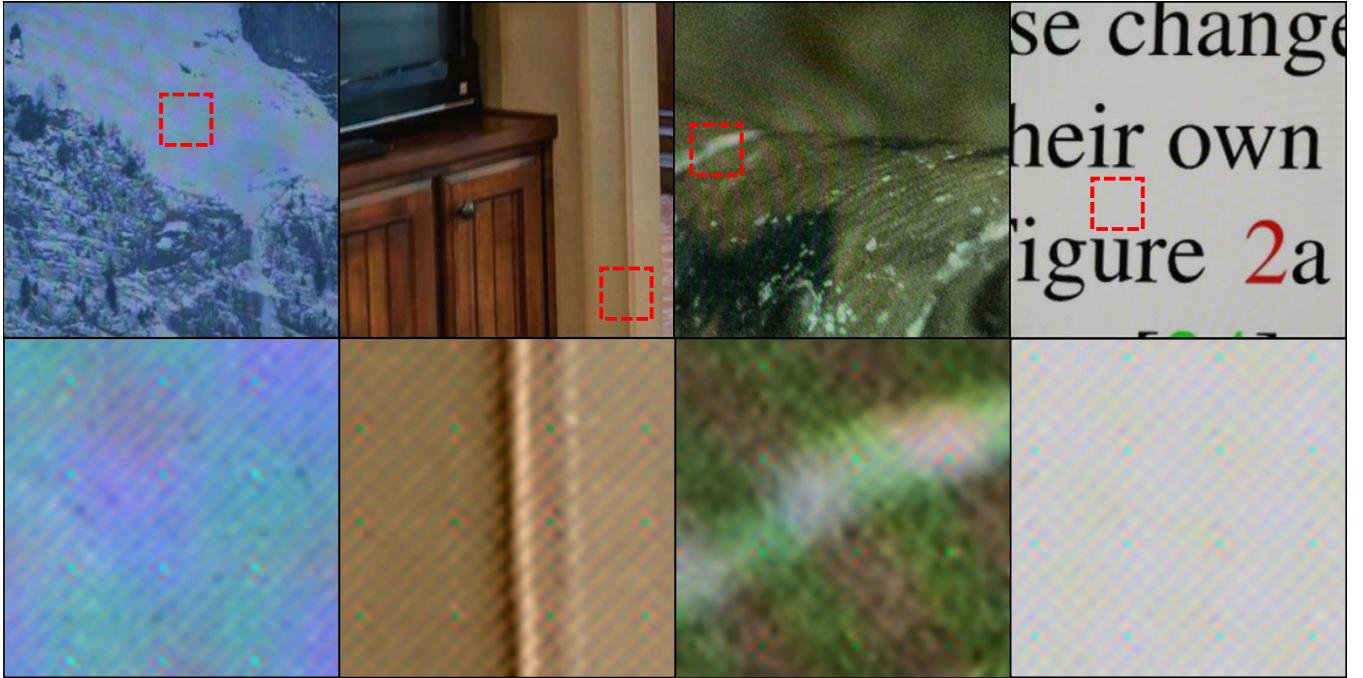


Figure 11: Examples of the “Checkerboard Artifacts” that occur in the I_{trn} when upsampling with Uformer’s transpose convolution (Wang et al. 2022).

emergence of “Checkerboard Artifacts” in the output I_{trn} , as illustrated in Figure 11. This issue stems from the uneven overlap when transposed convolution is employed dur-

ing the upsampling process (Odena, Dumoulin, and Olah 2016). As a solution, we utilize the CARAFE upsampling operator (Wang et al. 2019) to replace transposed convolu-

tion in Uformer for upsampling. CARAFE effectively addresses the "Checkerboard Artifacts" by predicting diverse up-sampling kernels based on the semantic information of the input feature maps (Wang et al. 2019), thereby contributing to improved feature reorganization within TRN. Performance changes resulting from the use of CARAFE are detailed in the additional ablation study in Section C.4.

We utilize 2 NVIDIA A40 GPUs to train our Tone Refinement Network on UHDM (Yu et al. 2022) for 50 epochs, FHDMi (He et al. 2020) for 25 epochs, and TIP (Sun, Yu, and Wang 2018) for 2 epochs. The learning rate is initially set to 1e-5 and scheduled by cyclic cosine annealing (Loshchilov and Hutter 2016), and models are optimized by Adam (Kingma and Ba 2014) with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. For the input clean natural images, we set the random crop size to 384×384 , and for the moiré patterns sampled by the Moiré Pattern Generator, we resized their resolution to 384×384 as well.

Implementations of loss functions For the loss function Eq. (13) in the main paper, we simply set $\lambda_{per} = \lambda_{color} = 1.0$ and $\lambda_{tv} = 0.1$ to balance the scale of the values during training.

For the perception loss L_{per} , to further validate our assumption in the main paper whether computing the content loss between I_{trn} and I_{mib} directly in the pixel space is less effective, we utilize the \mathcal{L}_1 loss instead of the \mathcal{L}_{per} loss for our synthesis network in the additional ablation study in Section C.4.

For the color loss \mathcal{L}_{color} , we convert I_{trn} and I_{rm} into RGB-uv histogram feature $H(I_{trn})$ and $H(I_{rm})$ from the log-chrominance space followed by prior work on color constancy (Afifi and Brown 2019; Afifi et al. 2019), which represents the color distribution of those two images. In particular, u and v are used to control the contribution of each color channel in the generated histogram and the smoothness of the histogram bin. Specifically, given an RGB image $I(x)$ where x denotes the pixel point index, we first convert it to YUV color space:

$$I_y(x) = \sqrt{I_r^2(x) + I_g^2(x) + I_b^2(x)}. \quad (16)$$

and:

$$I_{ur}(i) = \log \frac{I_r(i) + \epsilon}{I_g(i) + \epsilon}; I_{vr}(i) = \log \frac{I_r(i) + \epsilon}{I_b(i) + \epsilon} \quad (17)$$

$$I_{ug}(i) = \log \frac{I_g(i) + \epsilon}{I_r(i) + \epsilon}; I_{vg}(i) = \log \frac{I_g(i) + \epsilon}{I_b(i) + \epsilon} \quad (18)$$

$$I_{ub}(i) = \log \frac{I_b(i) + \epsilon}{I_r(i) + \epsilon}; I_{vb}(i) = \log \frac{I_b(i) + \epsilon}{I_g(i) + \epsilon} \quad (19)$$

where " I_r ", " I_g ", and " I_b " subscripts refer to the color channels of the image I , $\epsilon = 10^{-6}$ is a small constant added for numerical stability, and (I_{ur}, I_{vr}) , (I_{ug}, I_{vg}) and (I_{ub}, I_{vb}) are the uv coordinates of the I_r , I_g , and I_b .

We then generated the unnormalized histogram $H(u, v, c)$ of each color channel $c \in \{r, g, b\}$ according to the HistoGAN (Afifi, Brubaker, and Brown 2021), computed as follows:

$$H(u, v, c) \propto \sum_x k(I_{uc}(x), I_{vc}(x), u, v) I_y(x), \quad (20)$$

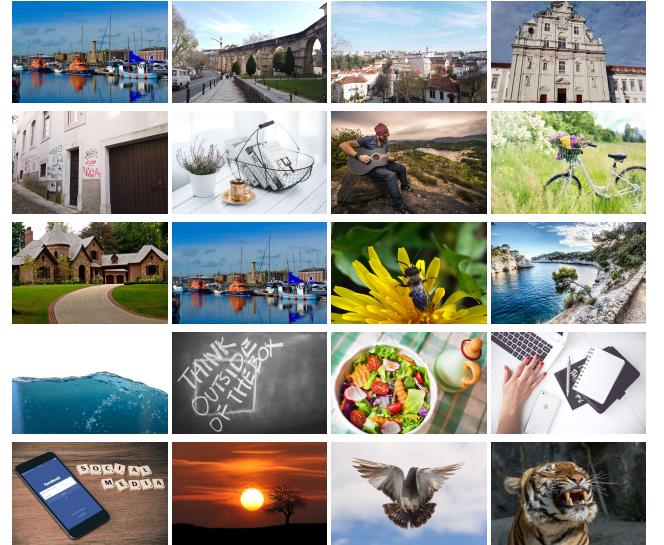


Figure 12: Examples of the MHRNID dataset.

where $k(\cdot)$ is the inverse-quadratic kernel:

$$k(I_{uc}, I_{vc}, u, v) = \left(1 + (|I_{uc} - u|/\tau)^2\right)^{-1} \times \left(1 + (|I_{vc} - v|/\tau)^2\right)^{-1} \quad (21)$$

where τ is a fall-off parameter to control the smoothness of the histogram's bins. Finally, the histogram features $H(I) \in R^{h \times h \times 3}$ stacked by $H(u, v, c)$ of 3 color channels is normalized to sum to one:

$$H(I) = \frac{[H(u, v, r), H(u, v, g), H(u, v, b)]}{\sum_{u,v,c} H(u, v, c)}. \quad (22)$$

Following HistoGAN (Afifi, Brubaker, and Brown 2021), we set the histogram bin, h , to 64 and set the fall-off parameter of our histogram's bins, τ , to 0.02.

C Experiments

In this section, we will provide a more detailed overview of the experimental setups, present additional visualization results and runtime comparisons, carry out further ablation studies, and address the limitations of our proposed method.

C.1 Experimental Setups

We implement all the experiments using PyTorch Lightning on multiple NVIDIA A40 GPUs. All experiments were conducted once after setting the seed to the same values as (Yu et al. 2022) and (Zhong et al. 2024).

Implementation Details of other Comparison Methods

For Shooting, we migrated their implementation code from OpenCV to PyTorch based on the implementation idea provided by (Niu, Guo, and Wang 2021). Note that the Shooting method produces a distorted composite image after random projective transformation. We maintain the transformation parameter and adjust the clean image accordingly to ensure that the moiré image aligns with the clean image during the subsequent demoiréing stage. For UnDeM (Zhong

Methods	Params(M)	Runtime(s)
Shooting	-	0.03
UnDeM	2.4	0.56
MoireSpace	-	0.01
Ours(MIB)	-	0.02
Ours(MIB & TRN)	3.0	0.04

Table 7: Runtime comparisons.

et al. 2024), we directly use their 384×384 moiré image synthesis network trained on UHDM (Yu et al. 2022) and FHDMi (He et al. 2020) and also train their synthesis network on TIP (Sun, Yu, and Wang 2018) in their code framework (Zhong et al. 2024). For MoireSpace (Yang et al. 2023), we utilize the moiré patterns provided by their dataset to obtain the synthesis result by deploying their multiply blending strategy. We resize their moiré patterns to 384×384 for a fair comparison.

Mixed High-Resolution Natural Image Dataset In the Zero-Shot experiments, we collected a comprehensive Mixed High-Resolution Natural Image Dataset (MHRNID) to avoid data overlap between the training and test sets. The MHRNID dataset consists of the super-resolution datasets DF2K-OST (Wang et al. 2021), the natural image datasets UHD-LOL4K (Wang et al. 2023b), and UHD-IQA (Hosu et al. 2024) collated and incorporated, which contains 26,000 high-definition images. We also provide several visual examples of MHRNID, as shown in Figure 12.

Implementation Details of Demoiréing Models For MBCNN (Zheng et al. 2020) and ESDNet-L (Yu et al. 2022), we followed the experimental settings from (Yu et al. 2022) and (Zhong et al. 2024). We trained for 150 epochs on UHDM (Yu et al. 2022) and FHDMi (He et al. 2020) and 70 epochs on TIP (Sun, Yu, and Wang 2018). Additionally, we trained for 50 epochs on the MHRNID dataset.

C.2 More Qualitative Comparisons

Moiré Image Synthesis The visualization results of synthesis moiré images on the MHRNID dataset using Shooting (Niu, Guo, and Wang 2021), UnDeM (Zhong et al. 2024), and our UniDemoiré are shown in Figure 14. The moiré image produced by our UniDemoiré is notably superior to other synthesis methods in terms of diversity and realism. In comparison, the moiré image generated by the Shooting (Niu, Guo, and Wang 2021) method is excessively distorted, UnDeM’s network (Zhong et al. 2024) is susceptible to anomalies during image generation, and the moiré pattern dataset provided by MoireSpace (Yang et al. 2023) is of subpar quality. Additionally, the multiplication strategy results in a darker synthesized image.

Demoiréing Figure 15 shows the visualization results of zero-shot demoiréing on UHDM (Yu et al. 2022). Additionally, Figures 16 and 17 illustrate the demoiréing results on FHDMi (He et al. 2020) and TIP (Sun, Yu, and Wang 2018) using ESDNet-L (Yu et al. 2022) trained on UHDM (Yu

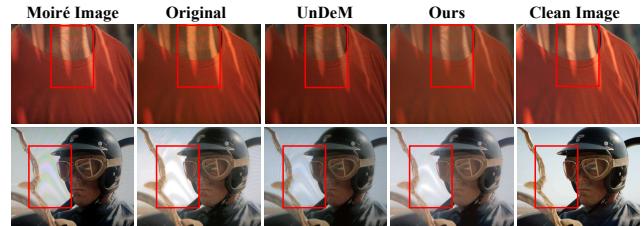


Figure 13: Failure Examples.

et al. 2022). Our method’s model effectively removes moiré artifacts and retains high-frequency details, indicating the strong generalization ability of our proposed UniDemoiré.

C.3 Runtime Comparisons

Table 7 shows the comparison of the parameters and the running time of our synthesis module with other methods. To ensure fair comparisons, our method and UnDeM use torchinfo for parameter counting, with all methods utilizing 256×256 input images. Our experimental results indicate that our method, slightly exceeding UnDeM in parameters, achieves a runtime comparable to non-learning algorithms like Shooting and MoireSpace, demonstrating the efficiency of our MIB and TRN. Furthermore, our model’s FLOPs are 5.6266G, significantly lower than UnDeM’s 26.7576G, indicating high performance and reduced computational cost.

C.4 Additional Ablation Study

The results of the additional ablation experiments are in Table 8, where “ $\mathcal{L}_{per} \rightarrow \mathcal{L}_1$ ” denotes replacing the perception loss \mathcal{L}_{per} in the synthesis network with the L1 loss \mathcal{L}_1 . “Uformer \rightarrow UNet” denotes switching the entire backbone network of the TRN from Uformer to UNet (Ronneberger, Fischer, and Brox 2015). For a fair comparison, we kept the number of upsampling/downsampling blocks and the base channel in UNet consistent with Uformer, while removing the attention block.

Components	PSNR↑	SSIM↑	LPIPS↓
ALL	20.7543	0.7653	0.2136
MIB (w/o Multiply)	20.3158	0.7598	0.2328
MIB (w/o Grain Merge)	20.3930	0.7587	0.2414
TRN (w/o CARAFE)	20.4414	0.7408	0.2256
TRN ($\mathcal{L}_{per} \rightarrow \mathcal{L}_1$)	20.1404	0.7447	0.2495
TRN (Uformer \rightarrow UNet)	20.3899	0.7476	0.2413

Table 8: Additional ablation studies. Source: UHDM, Target: FHDMi.

The results of two sets of ablation experiments on layer blending strategies also show that using only one of them leads to distortion of the synthesis results, which in turn affects the model’s generalization ability. The results of the “ $\mathcal{L}_{per} \rightarrow \mathcal{L}_1$ ” show that computing the loss function in this way leads to a degradation of the model per-

formance because moiré patterns can disrupt image structures by generating strip-shaped artifacts. The results of the “*w/o CARAFE*” indicate that using the CARAFE upsampling operator (Wang et al. 2019) yields better fusion performance than the transposed convolution originally employed by Uformer (Wang et al. 2022). Furthermore, the results from the “Uformer → UNet” demonstrate that the LeWin Transformer Block within Uformer is more effective at extracting color features from moiré patterns compared to the original UNet architecture.

C.5 Limitations

In some cases, particularly when the moiré artifacts in the target domain significantly differ from those in the source domain, our solution may struggle to completely remove all artifacts, as Figure 13 shows. However, even in these challenging scenarios, our method tends to perform better at artifact removal compared to the baselines. Our performance can be further refined by generating more diverse moiré patterns and synthesized training data. In Figure 13, we show a failure case. When the moiré artifacts in the target domain are too different from those in the source domain, our solution still struggles to produce a completely moiré-free result. However, we still remove the artifacts comparatively better than baselines.

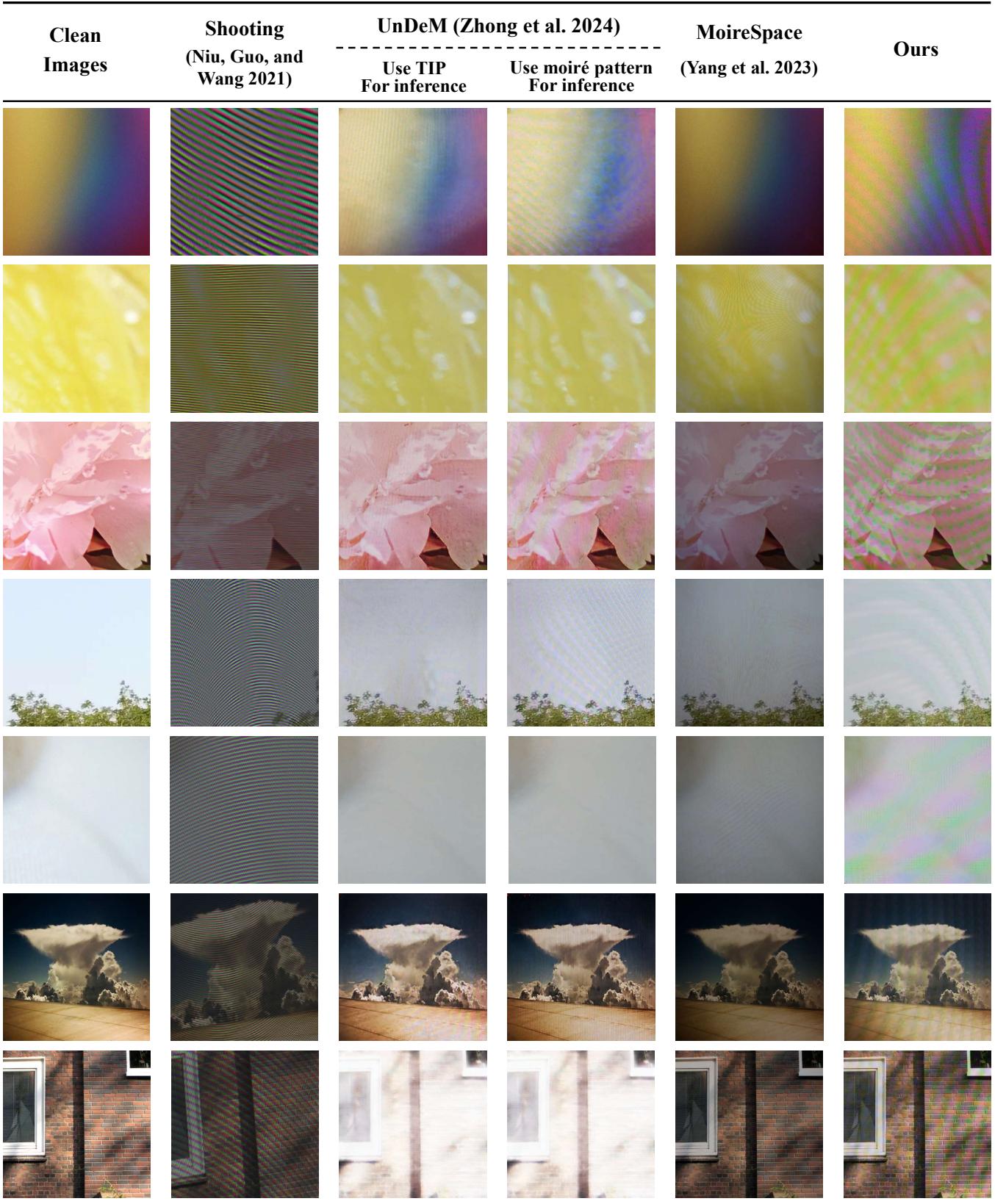


Figure 14: Qualitative comparisons of synthesized moire images were obtained using the shooting method, UnDeM, MoireSpace, and our UniDemoiré.

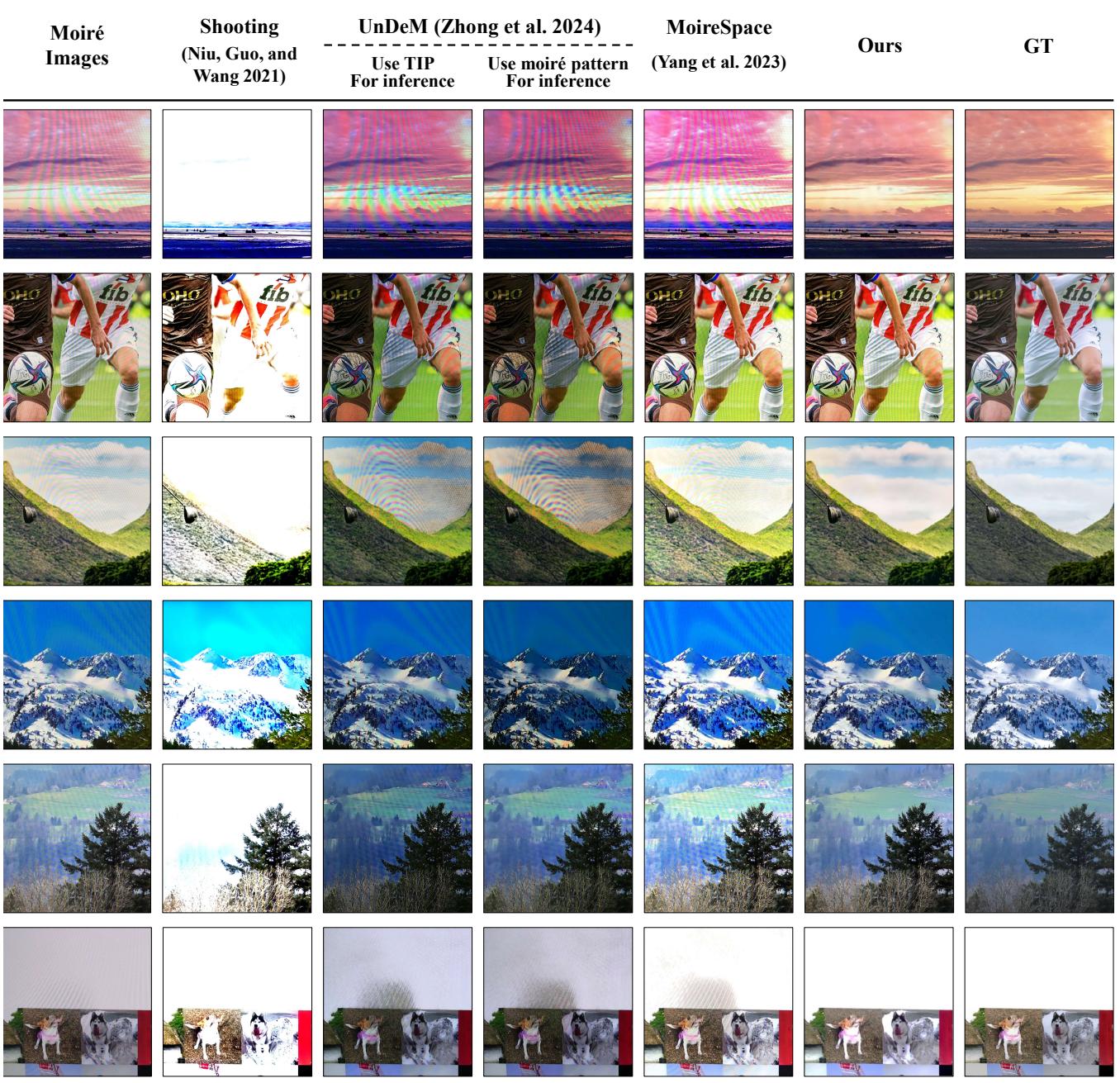


Figure 15: Qualitative comparisons of zero-shot evaluation on the UHDM dataset.

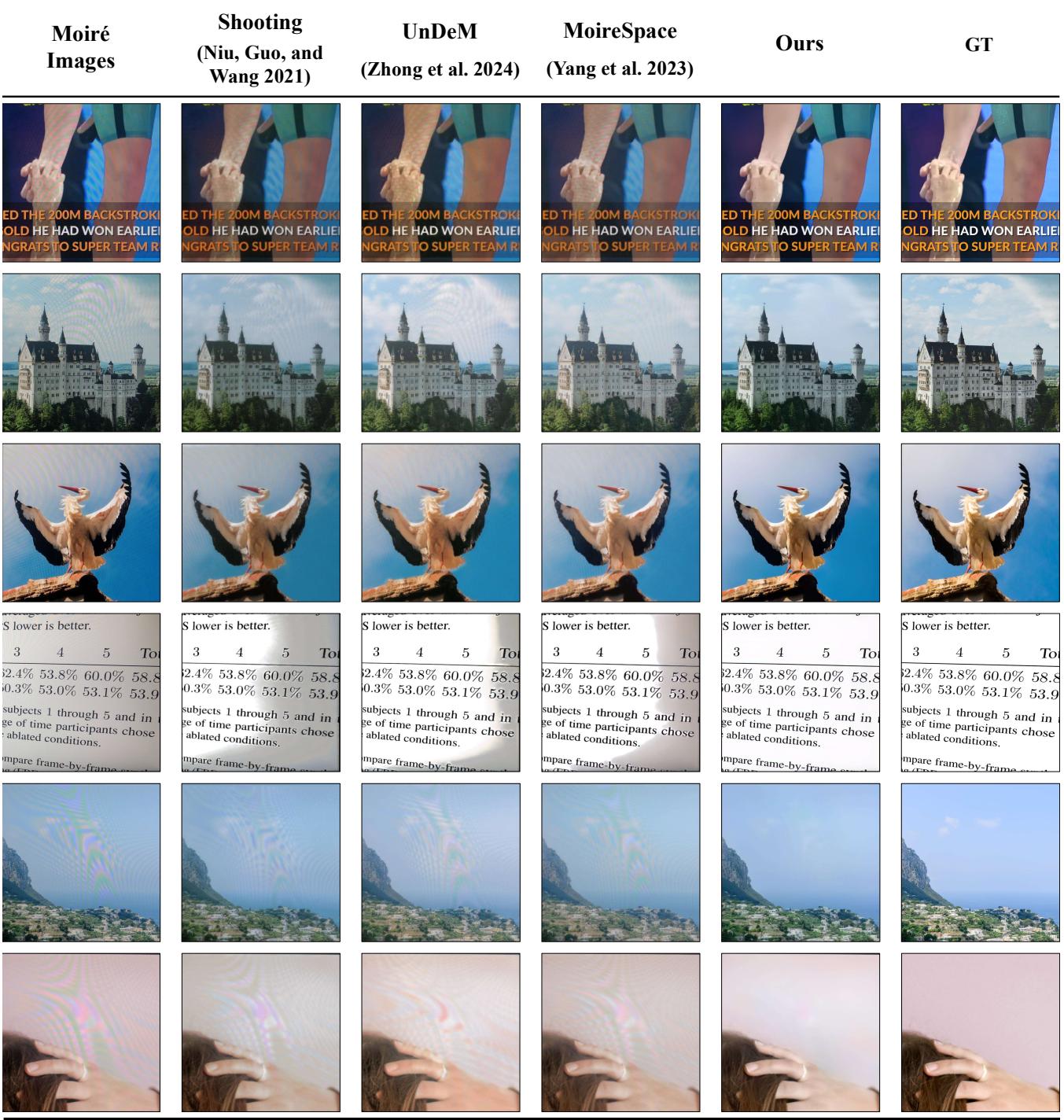


Figure 16: Qualitative comparisons of our models with other state-of-the-art methods on the FHDMD dataset.

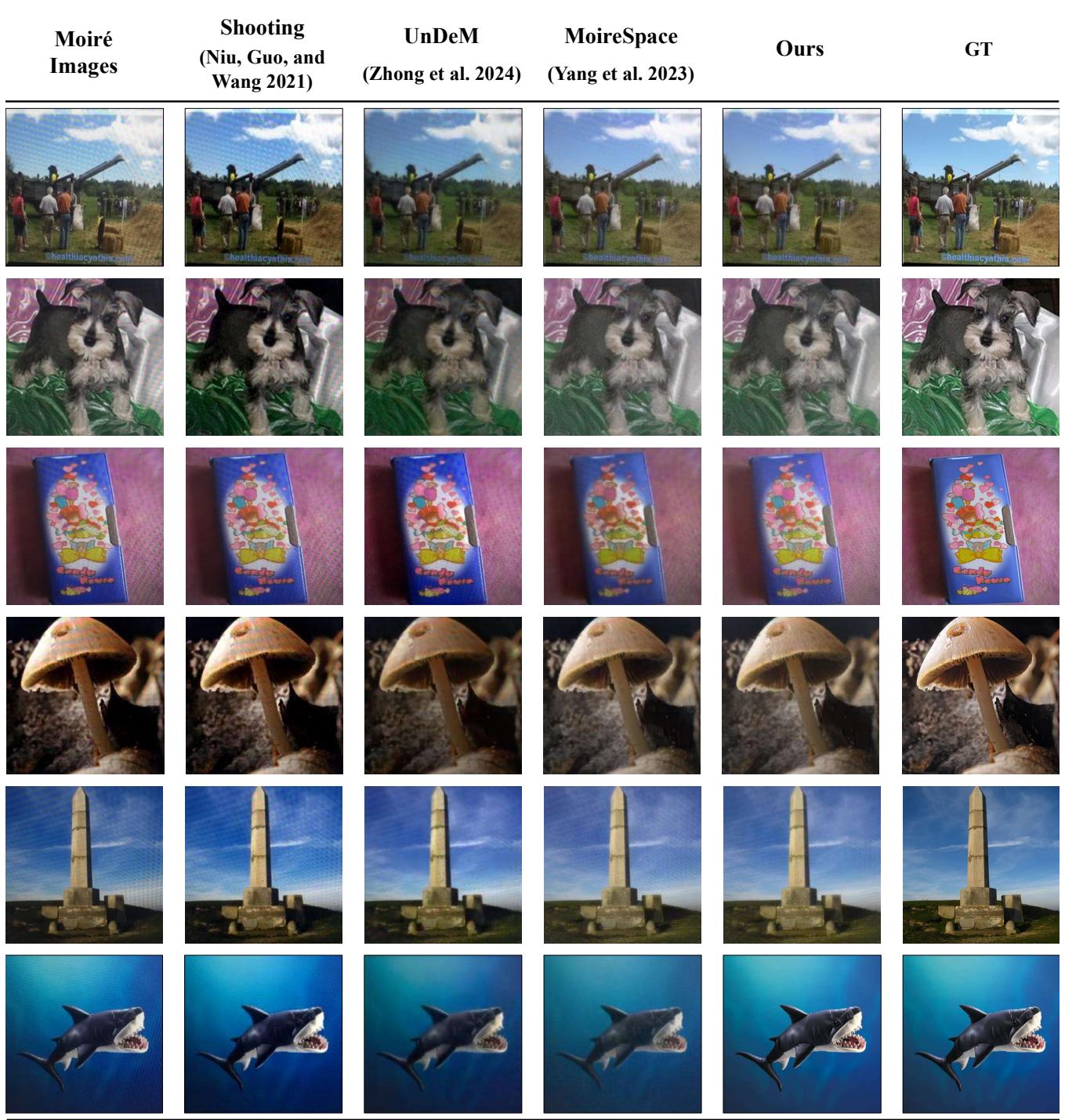


Figure 17: Qualitative comparisons of our models with other state-of-the-art methods on the TIP dataset.