IMPROVING COMPLEX MOIRÉ REMOVAL WITH GENERATIVE SUPERVISION

Anonymous authors
Paper under double-blind review

ABSTRACT

The availability of high-quality paired data is essential for training learning-based image demoiréing models. However, it remains challenging for existing datasets to encompass the complex moiré patterns captured in uncontrolled real-world scenarios. Such degradations typically manifest as large-scale, multicolored moiré patterns. Moreover, these patterns frequently occur in images for which clean counterparts are difficult to obtain, such as photographs acquired from public displays and existing online resources. In this work, we propose a novel data engine designed to improve the removal of these challenging moiré patterns by generating training supervision. Specifically, we initially collect real-world images containing complex moiré patterns and localize the corresponding screen regions. Multiple image-conditioned generative foundation models are subsequently deployed to produce candidate references. To establish reliable supervision, these candidates are subjected to patch-level quality control to filter and select the optimal results. Based on this systematic paradigm, we construct the WildMoiré dataset, which contains 6.8K moiré-GT training pairs. For evaluation, we additionally build an independent test set comprising $\sim$250 pairs with captured clean ground truth. Extensive experiments on ESDNet, SDXL, and Qwen-Image-Edit demonstrate that the proposed generative supervision consistently improves the performance of complex moiré removal. The code and dataset will be made publicly available.

1 INTRODUCTION

Using smartphones and digital cameras to record information from electronic screens has become increasingly common in daily life. However, captured screen images often contain moiré artifacts that are absent from the original displayed content. These artifacts may appear as fine colored grids, curved waves, broad low-frequency color bands, and mixtures of several patterns. They arise because both the display and the camera sensor sample visual signals on discrete grids, and interference between the display pixel grid and the camera sensor array can produce objectionable aliasing patterns (Liu et al., 2018). Moiré artifacts not only degrade visual quality, but may also obscure text, alter colors, and destroy image textures and structural details. Therefore, removing moiré from camera-captured screen images is an important problem for mobile photography, digital-signage capture, document recording, and visual-content sharing.

Facilitated by the collection of large-scale paired data (Sun et al., 2018; He et al., 2020; Yu et al., 2022; Mei et al., 2025), learning-based image demoiréing methods have achieved substantial progress through multi-scale processing strategies (Zheng et al., 2020; Cheng et al., 2023), the integration of attention mechanisms (He et al., 2025), and the introduction of diffusion models (Zhu et al., 2026; Yang et al., 2026). Nevertheless, these trained demoiréing models yield satisfactory results only within the distribution of the training data, exhibiting limited generalizability across different screens and complex scenes, especially for large-scale, multicolored moiré patterns (see Fig. 1(b)). A straightforward approach is to continuously collect and scale up paired data for training. However, this solution is impractical and expensive in terms of both manpower and computational resources, owing to the diverse range of existing displays and the continuous emergence of new devices.

Fortunately, recent foundation models for image generation and editing, *e.g.*, Nano-Banana-2 (Google DeepMind, 2026) and GPT-Image-2 (OpenAI, 2026), demonstrate remarkable generalization capabilities for moiré removal when provided with carefully crafted prompts, which stems from their

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

rich and strong visual priors. But simultaneously, these models are difficult to adopt directly as professional-grade demoiréing tools because they are prone to generating hallucinations, *e.g.*, inconsistencies in content and color. To address this issue, we propose to post-process and filter these outputs to serve as supervision for demoiréing models, which functions as a more cost-effective and convenient solution for scaling up paired data. Moreover, instead of relying on the outputs of a single generative model, we compare the results from multiple generative models and select the optimal output for supervision. This strategy aims to leverage the complementary strengths across different models, such as moiré removal and texture preservation capabilities.

Specifically, we initially collect a substantial number of complex moiré images through on-site photography and existing online resources. Subsequently, we localize the screen regions and process them using foundation models for image generation and editing. Given that different models exhibit distinct trade-offs between moiré removal and content preservation, we deploy multiple models. These include closed-source models (*i.e.*, GPT-Image-2 (OpenAI, 2026) and Nano-Banana-2 (Google DeepMind, 2026)) and open-source models (FLUX.2 (Black Forest Labs, 2025), SDXL (Podell et al., 2023), and Qwen-Image-Edit (Wu et al., 2025)) that have been fine-tuned on existing demoiréing datasets. The corresponding outputs are aligned with the input images in terms of spatial positions and colors, and are subjected to patch-based quality control. Finally, the optimal output among them is selected as the ground truth (GT). In total, we construct 6.8K moiré-GT pairs with a resolution of 1024×1024, forming the WildMoiré dataset. Furthermore, we introduce a scale transformation strategy for data augmentation. By rescaling the moiré images, the scale of the moiré patterns is correspondingly altered, thereby further enhancing the generalization capability of demoiréing models.



Figure 1: Results on moiré images. The (a) column presents three moiré inputs with large-scale chromatic bands and overlapping interference patterns. The (b) and (c) columns show Qwen-Image-Edit trained without and with our constructed WildMoiré dataset, respectively. Training only on the previous dataset leaves artifacts and suppresses details, while incorporating WildMoiré removes the broad interference and preserves original contents better.

We additionally collect ∼250 pairs with captured clean ground truth images for the evaluation of complex moiré removal. Utilizing the proposed WildMoiré dataset and data augmentation strategy, we fine-tune the CNN-based ESDNet (Yu et al., 2022), the diffusion-UNet-based SDXL (Podell et al., 2023), and the diffusion-Transformer-based Qwen-Image-Edit (Wu et al., 2025) for experimental validation. The results demonstrate significant improvements in both quantitative metrics and visual quality (see Fig. 1). Moreover, comprehensive ablation studies confirm the effectiveness of the various components in the proposed pipeline.

Our contributions are summarized as follows:

- We propose generating supervision to extend the training data for the improvement of removing complex and challenging moiré. Specifically, we introduce a systematic paradigm that filters and selects the optimal outputs from multiple generative models to serve as high-quality supervision.
- Using the proposed data engine, we construct WildMoiré dataset, which contains 6.8K moiré-GT pairs. Furthermore, we collect ∼250 pairs with captured clean ground truth as an evaluation benchmark.
- Extensive experiments on ESDNet, SDXL, and Qwen-Image-Edit demonstrate that WildMoiré dataset consistently improves the demoiréing performance.

2 Related Work

2.1 Image Demoiréing

Early image demoiréing methods relied on handcrafted signal decomposition and frequency-domain filtering. With the development of deep learning, convolutional networks have become the dominant solution for camera-captured screen images. DMCNN introduced a multi-resolution architecture to account for the large variation in moiré scales (Sun et al., 2018). MopNet classified moiré patterns and processed different categories in a divide-and-conquer manner (He et al., 2019), while MBCNN explicitly modeled frequency-selective decomposition using learnable bandpass filters (Zheng et al., 2020). To support high-resolution restoration, FHDe2Net employed a coarse-to-fine framework for full-HD images (He et al., 2020), and ESDNet introduced semantic-aligned scale-aware processing for efficient 4K demoiréing (Yu et al., 2022). More recently, DCID explored dual-camera fusion for mobile image demoiréing, using complementary wide- and ultra-wide-camera observations to remove severe moiré while preserving high-resolution image details (Mei et al., 2025).These approaches have substantially advanced model architecture and computational efficiency. Nevertheless, their restoration capability remains closely related to the moiré pattern distribution in the training data.

2.2 Demoiréing Data Construction

Real paired datasets have been essential to the development of learning-based demoiréing. Sun *et al.* introduced an early benchmark of real camera-captured screen images (Sun et al., 2018). FHDMi extended paired acquisition to full-HD resolution (He et al., 2020), while UHDM collected 5,000 ultra-high-definition pairs using multiple smartphones, displays, and shooting configurations (Yu et al., 2022). DCID further constructed 8,959 samples containing wide-angle images, ultra-wide-angle observations, and digital source supervisions, with particular attention to severe moiré patterns (Mei et al., 2025). Although these datasets provide reliable supervision, their acquisition generally requires displaying known clean images on accessible screens and carefully aligning the captured images with their digital sources. Such controlled pipelines are costly to scale and cannot exhaustively cover complex moiré patterns encountered in uncontrolled environments.

To alleviate the dependence on costly paired acquisition, another line of research synthesizes moiré-degraded inputs from clean images. Early work simulated the camera–display imaging process to construct synthetic training pairs (Liu et al., 2018), while LCDMoiré adopted a handcrafted synthesis process for the AIM image demoiréing challenge (Yuan et al., 2019). Cyclic moiré learning jointly trained a moiréing network and a demoiréing network using unpaired clean and moiré images (Park et al., 2022). UnDeM divided real moiré images into patches according to their complexity and learned to synthesize diverse pseudo-moiré images for supervised demoiréing training (Zhong et al., 2024). More recently, UniDemoiré collected background-independent real moiré patterns, employed a diffusion model to generate additional pattern variations, and developed a learnable synthesis module to reproduce the color and brightness characteristics of captured moiré images (Yang et al., 2025). These methods improve data scalability by generating the degraded side of a training pair.

2.3 Generative Supervision for Image Restoration

Recent studies have explored generative models as offline producers of training supervision rather than only as restoration models used at inference time. HGGT generated multiple enhanced high-resolution targets for image super-resolution and employed human annotations to identify regions with beneficial or harmful perceptual changes (Chen et al., 2023). GMD used a frozen generative oracle to refine restoration results from unlabeled target-domain images and mixed the resulting pseudo-pairs with reliable source-domain supervision for model adaptation (Hu et al., 2025). Additionally, we take note of the concurrent work GGT-100K, which systematically evaluates multimodal foundation models and implements multi-stage quality control to construct a large-scale dataset for general real-world image restoration (Kong et al., 2026). These studies demonstrate the potential of generative supervision, while also highlighting the risks of hallucination, content drift, and inconsistent restoration quality. In this work, we utilize multiple generative models and implement carefully designed quality control measures to construct more reliable supervision signals.

Our work specializes this paradigm for complex screen demoiréing. Instead of relying on a single generator, we use five generative priors with complementary restoration behaviors. Their outputs are

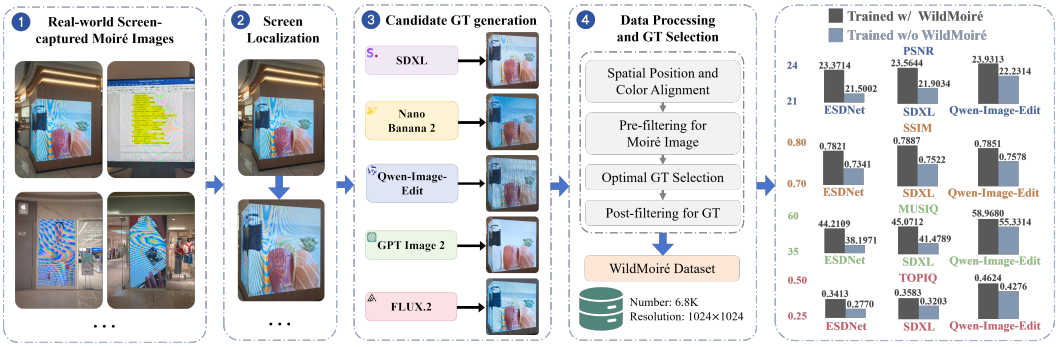

Figure 2: Overview of WildMoiré dataset construction. We first collect real screen-captured images containing complex moiré and localize the display regions. Five image-conditioned generative models then produce candidate GT images. The candidates are processed at matched spatial locations through spatial and color alignment, pre-filtering, optimal GT selection, and post-filtering. This offline procedure yields 6.8K moiré-GT pairs at 1024×1024 resolution. The plots on the right summarize the improvements on the collected challenging moiré images.

spatially and photometrically aligned with the real moiré inputs and evaluated locally before being retained as training supervisions. This design preserves authentic complex moiré degradations on the input side, reduces the influence of unreliable generated content, and produces reusable training data without modifying the deployed restoration architectures or adding test-time generative computation.

3 WildMoiré Dataset

3.1 Motivation and Overview

Moiré patterns arise from interference between the sensor array of a camera and the pixel grid of a displayed screen. Existing moiré datasets primarily cover simple and small-scale patterns. This limited coverage makes demoiréing models less effective in handling complex, large-scale, and multi-color moiré patterns, which are particularly common on large public screens.

It is time-consuming, labor-intensive, and sometimes even impossible to obtain corresponding clean images for such complex moiré images. Firstly, large public displays often present dynamic videos or frequently changing content, making it difficult to recover the exact displayed frame after the moiré image has been captured. Secondly, for images collected from online sources, the original clean content is generally unavailable. Fortunately, recent image generation and editing foundation models (OpenAI, 2026; Google DeepMind, 2026; Black Forest Labs, 2025; Podell et al., 2023; Wu et al., 2025) are developing rapidly, and their powerful generative capabilities can provide effective supervision for improving the removal of complex moiré patterns.

Although these image generation models exhibit strong restoration capabilities, no single model consistently performs well on all complex moiré images. These models show complementarity in moiré removal and content preservation. Some effectively remove moiré patterns but redraw the image, reducing fidelity, while others retain the original content but leave residual moiré artifacts. Therefore, we suggest identifying the highest-quality one from multiple model results as the final supervision. At the same time, inappropriate samples are filtered out through strict quality control. With this data engine, we construct WildMoiré, a dataset designed to improve demoiréing models' ability on complex moiré patterns. The dataset construction pipeline is illustrated in Fig. 2.

3.2 Source Images Collection

To support robust demoiréing of complex moiré images, we collect images from two sources. First, we capture 4K-resolution images across diverse scenes using mobile phones. Second, we collect additional images through web crawling to increase diversity of displayed content, screen types, and moiré patterns. We intentionally retained challenging examples with broad chromatic bands, overlapping waves, fine repetitive interference, oblique viewing angles, text-rich content, and human

subjects, as shown in **Step 1** in Fig. 2. After collection, we used Sa2VA (Yuan et al., 2025) to detect screen boundaries and crop the corresponding screen regions, as shown in **Step 2** in Fig. 2.

3.3 Candidate GT Generation

To obtain corresponding clean supervisions for the collected complex moiré images, we employ SDXL (Podell et al., 2023), Nano-Banana-2 (Google DeepMind, 2026), Qwen-Image-Edit (Wu et al., 2025), GPT-Image-2 (OpenAI, 2026), and FLUX.2 (Black Forest Labs, 2025) to generate candidate GTs, as shown in **Step 3** of Fig. 2.

Different image generation and editing models exhibit complementary behaviors in restoration capability and content fidelity. GPT-Image-2 and Nano-Banana-2 generally provide stronger removal of severe interference but can introduce content or color changes, while the other three open-source models are more conservative after fine-tuning on existing demoiréing datasets. Their complementary behaviors motivate selecting supervision locally rather than relying on a single generator on the full image. The open-source generators are adapted on existing demoiréing datasets before candidate generation, and detailed discussion and setup are provided in Appendix B.1.

For candidate generation, the prompts are adapted to each model while preserving a consistent intent, *i.e.*, removing moiré and abnormal color interference while retaining text, identities, objects, layout, geometry, and the overall appearance of the photographed display. The exact prompts used for all five generators are provided in Appendix B.2.

3.4 Data Processing and GT Selection

Spatial and Photometric Alignment. During candidate GT generation, we observe that each input and its candidate GTs may exhibit spatial and photometric misalignments. To address these issues, we first correct the spatial misalignments and subsequently perform photometric alignment. Specifically, for spatial alignment, we first employ GlueStick (Pautrat et al., 2023) to estimate a robust global transformation for initial alignment. We then use FlowFormer (Huang et al., 2022) to estimate the dense optical flow between each initially aligned candidate GT and its input for refining the alignment. Finally, we center-crop each input and its aligned candidate GT to remove the misaligned areas around the edges. For photometric alignment, we first apply Gaussian blurring to the input and its aligned candidate GTs to suppress the influence of high-frequency moiré patterns. We then fit a linear 3×3 RGB transformation for each candidate GT to align its color with the input.

Pre-filtering for Moiré Image. This pre-filtering step aims to retain collected input patches with complex moiré and remove clean patches. We treat each input and its aligned candidate GTs as an image group. For each image group, we sample n spatial locations and crop 1024×1024 patches from every image at each location, yielding n patch groups per image group.Then we calculate two texture-complexity statistics for moiré patches, *i.e.*, the standard deviation σ_{img} (Moulden et al., 1990) and the mean Laplacian-variance pyramid σ_{lap} (Burt & Adelson, 1983). The patches whose lack texture (*i.e.*, $\sigma_{img} < \tau_{img}$ or $\sigma_{lap} < \tau_{lap}$, τ_{img} and τ_{lap} are thresholds) discarded. However, these statistics measure texture and moiré richness at the same time. Some moiré-free patches with rich textures may also be retained. To identify such patches, we process each remaining input patch with a pre-trained demoiréing model (*i.e.*, ESDNet) and compute the SSIM between its output and the original input patch. The demoiréing model can remove moiré patterns to a certain extent but preserve clean patches. We therefore classify the input patch with an SSIM greater than τ_{ssim} as a clean patch with rich textures.

Optimal GT Selection. We use an image quality assessment (IQA) metric to select a high-quality supervision from the candidate GTs. Full-reference metric is unavailable due to the need for a perfect reference. No-reference metrics such as MUSIQ (Ke et al., 2021) and Q-Align (Wu et al., 2024) assess visual quality without a reference. Their scores can reflect different quality preferences and do not directly check content consistency with the input. A visually pleasing candidate may be selected even when its content has changed. Therefore, we adopt A-FINE (Chen et al., 2025), denoted by $\mathcal{D}$, which supports imperfect references and considers both fidelity and naturalness.

Each retained patch group comprises 1 input patch P_i and 5 candidate GT patches $\{Q_i^k\}_{k=1}^5$. For the i-th group, we compute $\mathcal{D}(P_i, Q_i^k)$ for each candidate, using the moiré input as the reference and the

Table 1: Distribution of final selected GT source in WildMoiré dataset.

Model	GPT-Image-2	Nano-Banana-2	FLUX.2	Qwen-Image-Edit	SDXL
Ratio (%)	62.73	17.97	9.10	5.96	4.23
Count	4,286	1,228	622	407	289

Table 2: Datasets used in the experiment.

Dataset	Environment	Supervision	Resolution	Size
UHDM	Controlled	Digital source	∼4328×3248	4,271
DCID	Controlled	Digital source	∼4096×3072	7,176
WildMoiré	In the wild	Generation	1024×1024	6,832

candidate as the evaluated image. The GT is selected by $k^* = \arg\min_{k\in\{1,\dots,5\}} \mathcal{D}(P_i, Q_i^k)$, where k^* is the index of the selected GT patch. The lower values of $\mathcal{D}$ indicate better quality.

Post-filtering for GT. Selecting the best candidate within a group does not ensure that it is suitable enough for supervision. All five candidates may still contain residual moiré or content changes. We therefore apply a post-filter that requires the selected candidate to provide sufficient predicted quality improvement over the moiré input. For the i-th group, we first compute $\mathcal{D}(P_i, P_i)$, the score obtained by evaluating the input against itself. We then compare this score with $\mathcal{D}(P_i, Q_i^{k^*})$. Their difference measures the predicted improvement over the input. We discard the entire patch group when this improvement is below a threshold τ_{post}, *i.e.*, $\mathcal{D}(P_i, P_i) - \mathcal{D}(P_i, Q_i^{k^*}) < \tau_{post}$. Otherwise, the selected pair $(P_i, Q_i^{k^*})$ is retained.

3.5 Dataset Statistics

Finally, WildMoiré contains 6,832 moiré-GT pairs at 1024×1024 resolution. Table 1 reports the number and proportion of selected supervisions from each model. Table 2 shows the comparison between our WildMoiré and existing demoiréing datasets. The filtering stages reduce 13,615 synchronized patch groups to 10,889 after pre-filtering, and 6,832 after post-filtering. Retained ratios are reported in Appendix C.1.

4 Data Augmentation and Mixing

4.1 Scale Transformation Data Augmentation

We observe that real-world demoiréing datasets often exhibit imbalanced distributions of moiré scales, which will limit the generalization of the demoiréing models trained on them. To address this issue, we suggest a scale transformation data augmentation strategy, which simply rescales input images during training to diversify the scales of moiré patterns. Specifically, during training, we randomly apply $s\times$ upsampling to moiré-GT pairs, and use the rescaled moiré images as inputs and the corresponding rescaled GTs as supervisions. As the diversity of moiré scales in the training data increases, the generalization ability of the demoiréing model is improved to handle moiré patterns at different scales.

4.2 Multiple Dataset Mixing

Compared with existing demoiréing datasets, WildMoiré contains images with more complex moiré patterns. To enable the demoiréing model to handle both such complex patterns and the conventional moiré patterns represented in existing datasets, we additionally incorporate UHDM (Yu et al., 2022) and DCID (Mei et al., 2025) into the training set, thereby further improving its generalization ability. The dataset details are provided in Table 2.

5 Experiments

5.1 Experimental Setup

Evaluation Data. To evaluate demoiréing performance, we collected a separate test set of 247 additional moiré images. We obtain their ground truth images through camera capture rather than image generation to ensure the authenticity. All pairs were captured with an OPPO Find X5. Each moiré image and its clean reference show exactly the same static content. We change only the camera

Table 3: Comparison of quantitative results on self-collected 247 challenging moiré images. The optimal results in every part are highlighted in bold.

	Models	Training Datasets	Full-reference Fidelity Metrics			No-reference Perceptual Metrics		
			PSNR ↑	SSIM ↑	LPIPS ↓	MUSIQ ↑	TOPIQ ↑	Q-Align ↑
Closed Source	GPT-Image-2	–	15.1965	0.5265	**0.4636**	**65.4375**	**0.5831**	4.3665
	Nano-Banana-2	–	**16.8793**	**0.5679**	0.4696	60.9621	0.5002	**4.4743**
Open Source	ESDNet (CNN-based)	UHDM+DCID	21.5002	0.7341	0.3471	38.1971	0.2770	3.6343
		UHDM+DCID+WildMoiré	**23.3714**	**0.7821**	**0.2614**	**44.2109**	**0.3413**	**3.9607**
Open Source	SDXL (Diffusion-UNet-based)	UHDM+DCID	21.9034	0.7522	0.2951	41.4789	0.3203	3.8128
		UHDM+DCID+WildMoiré	**23.5644**	**0.7887**	**0.2621**	**45.0712**	**0.3583**	**4.0811**
Open Source	Qwen-Image-Edit (Diffusion-Transformer-based)	UHDM+DCID	22.2314	0.7578	0.3101	55.3314	0.4276	4.1005
		UHDM+DCID+WildMoiré	**23.9313**	**0.7851**	**0.2528**	**58.9680**	**0.4624**	**4.2982**

distance and viewing angle, while keeping the other acquisition settings unchanged. The pairs are aligned using GlueStick (Pautrat et al., 2023), FlowFormer (Huang et al., 2022), and the same color alignment described above. We center-crop misaligned boundary regions and inspect the aligned pairs manually.

Metrics. We evaluate content fidelity using the full-reference metrics PSNR, SSIM (Wang et al., 2004), and LPIPS (Zhang et al., 2018). PSNR measures pixel-level fidelity, SSIM assesses structural similarity, and LPIPS captures perceptual similarity. We additionally report MUSIQ (Ke et al., 2021), TOPIQ (Chen et al., 2024), and Q-Align (Wu et al., 2024) as no-reference image-quality metrics to assess perceptual restoration quality.

Compared Settings. We train ESDNet, Qwen-Image-Edit, and SDXL on the combined UHDM and DCID datasets without scale transformation data augmentation, and use the resulting models as baselines (Yu et al., 2022; Wu et al., 2025; Podell et al., 2023; Mei et al., 2025). We then augment the training set with WildMoiré and train both without and with scale transformation data augmentation. GPT-Image-2 (OpenAI, 2026) and Nano-Banana-2 (Google DeepMind, 2026) are closed-source models that cannot be fine-tuned; we therefore evaluate their official direct-editing outputs.

Implementation Details. All models are implemented in PyTorch (Paszke et al., 2019). For each architecture, both the baseline and the WildMoiré setting are trained for 150 epochs with 10,000 sampled pairs per epoch and the same batch size. Adding WildMoiré changes only the sample composition, not the total number of sampled pairs or optimization steps. Each dataset is sampled cyclically across epoch boundaries, so later samples are visited in subsequent epochs rather than repeatedly restarting from the beginning. We set $\tau_{img} = 30$, $\tau_{lap} = 30$, $\tau_{ssim} = 0.9$, and $\tau_{post} = 15$. During training, the upsampling scale is $s = 4$. Full optimization and hardware settings are provided in Appendix B.3.

5.2 Experimental Results

Quantitative Results. Table 3 shows that incorporating WildMoiré consistently improves the overall fidelity and perceptual quality of a compact CNN restoration network and two generative editors. Compared with the corresponding baselines, the complete training setting with WildMoiré achieves clear improvements across all three architectures and the overall metric suite. The consistency across distinct model families indicates that WildMoiré provides broadly useful supervision for complex moiré removal rather than benefiting a particular training mechanism or model architecture. The independent contribution of scale transformation augmentation is further analyzed in Table 6.

The commercial editors occupy a different fidelity–perception operating point. Their strong no-reference scores reflect clean and visually appealing outputs, but the substantially weaker full-reference metrics expose content mismatch. Nano-Banana-2 generally preserves the displayed layout and local content more faithfully than GPT-Image-2, yet it can leave visible moiré in difficult cases. GPT-Image-2 removes interference more aggressively but may introduce severe content or structural inconsistency; its outputs can also suppress physical screen-capture characteristics and resemble



Figure 3: Comparison of qualitative results. The first example emphasizes logo and text fidelity under broad color interference, while the second contains severe mixed chromatic patterns over fine scene textures. Nano-Banana-2 is relatively content-faithful but may retain moiré, whereas GPT-Image-2 removes artifacts aggressively while replacing scene content and producing an overly digital-image appearance. Using WildMoiré suppresses residual bands more effectively across all 3 models.

Table 4: Quantitative results on challenging cases from the UHDM and DCID test sets. The optimal results in every part are highlighted in bold.

	Models	Training Datasets	Full-reference Fidelity Metrics			No-reference Perceptual Metrics		
			PSNR ↑	SSIM ↑	LPIPS ↓	MUSIQ ↑	TOPIQ ↑	Q-Align ↑
Closed Source	GPT-Image-2	–	16.4847	0.6288	0.3597	**47.6108**	**0.4396**	**4.4022**
	Nano-Banana-2	–	**19.5363**	**0.7320**	**0.2859**	45.6903	0.4112	4.2289
Open Source	ESDNet (CNN-based)	UHDM+DCID	26.5372	0.8722	0.2483	34.1202	0.3006	3.9790
		UHDM+DCID+WildMoiré	**26.9181**	**0.8784**	**0.2441**	**34.9631**	**0.3109**	**4.0698**
Open Source	SDXL (Diffusion-UNet-based)	UHDM+DCID	26.5779	0.8706	0.2467	36.5014	0.3082	4.0967
		UHDM+DCID+WildMoiré	**26.8603**	**0.8746**	**0.2458**	**37.1075**	**0.3187**	**4.1343**
Open Source	Qwen-Image-Edit (Diffusion-Transformer-based)	UHDM+DCID	26.9703	0.8801	**0.2360**	44.7053	0.4118	4.2721
		UHDM+DCID+WildMoiré	**27.3071**	**0.8870**	0.2367	**46.8816**	**0.4174**	**4.3152**

clean digital source images. These observations motivate using large generators as offline supervision producers instead of unrestricted test-time restorers.

Qualitative Results. Figure 3 supports the quantitative findings. The UHDM+DCID baseline models reduce part of the corruption but retain colored bands, local waves, or interference over fine texture. After adding WildMoiré, the same architectures recover cleaner logos, object boundaries, wall patterns, and furniture while retaining the photographic appearance of the display. The contrast with the two direct commercial editors also illustrates why perceptual quality alone is insufficient: an output can look clean while no longer representing the captured scene faithfully.

Results on UHDM and DCID Datasets. We also evaluate challenging samples from the UHDM and DCID test sets. We use the same pre-filtering strategy to select a subset with complex moiré. This yields 11 UHDM and 378 DCID samples. Table 4 reports the quantitative results on these selected samples. Adding WildMoiré improves PSNR, SSIM, and all three no-reference metrics for each architecture. LPIPS also improves for ESDNet and SDXL, with a small increase for Qwen-Image-Edit. These results show that WildMoiré complements existing data on challenging cases beyond our captured test set. Additional visual comparisons are provided in Appendix C.3.

Table 5: Effect of candidate GT sources.

GT Source	PSNR↑	SSIM↑	LPIPS↓	MUSIQ↑	TOPIQ↑	Q-A.↑
Nano-Banana-2	22.5640	0.7647	0.2921	40.9345	0.3087	3.8122
GPT-Image-2	23.0013	0.7721	0.2742	42.0071	0.3236	3.8834
Optimal of 5 Models	**23.3714**	**0.7821**	**0.2614**	**44.2109**	**0.3413**	**3.9607**

Table 6: Effect of scale transformation augmentation.

Training Datasets	Scale Aug.	PSNR ↑	SSIM ↑	LPIPS ↓	MUSIQ ↑	TOPIQ ↑	Q-ALIGN ↑
UHDM+DCID	✗	21.5002	0.7341	0.3471	38.1971	0.2770	3.6343
UHDM+DCID	✓	**21.6414**	**0.7468**	**0.3230**	**39.3647**	**0.2941**	**3.7478**
UHDM+DCID+WildMoiré	✗	23.3164	0.7710	**0.2611**	43.1129	0.3224	3.9091
UHDM+DCID+WildMoiré	✓	**23.3714**	**0.7821**	0.2614	**44.2109**	**0.3413**	**3.9607**

5.3 Ablation Study

Unless otherwise specified, all ablation experiments are conducted using ESDNet.

Effect of Candidate GT Sources. A central hypothesis of our framework is that different generators provide complementary candidate supervisions. We construct variants that restrict the candidate GT source to Nano-Banana-2 or GPT-Image-2 and train ESDNet with the same source data, per-epoch WildMoiré sampling, and augmentation recipe. Table 5 shows that GPT-Image-2 is the stronger individual source, but the five-prior construction achieves the optimal result on every metric. The method does not average candidate outputs; instead, it selects a locally suitable supervision. Therefore, a prior that is weaker on average can still contribute useful supervision at locations where its balance between artifact removal and content preservation is preferable.

Effect of Scale Transformation Augmentation. Table 6 isolates the effect of scale transformation augmentation on ESDNet under two training-data settings. When trained on UHDM and DCID, enabling the augmentation consistently improves all six metrics, showing that the transformed samples help the model accommodate variations in the apparent spatial scale of moiré patterns. A similar trend is observed after WildMoiré is introduced: scale augmentation improves PSNR, SSIM, MUSIQ, TOPIQ, and Q-ALIGN. These results indicate that the augmentation is beneficial both for existing paired data and for the mixed training setting. Meanwhile, the substantial improvement obtained by adding WildMoiré remains evident even without scale augmentation, confirming that the proposed dataset and the augmentation strategy provide complementary gains.

6 Conclusion

We presented a multi-prior framework for constructing reliable generative supervision from unpaired photographs containing complex screen moiré, and built WildMoiré with 6,832 training pairs. For evaluation, we additionally built an independent test set comprising 247 pairs with captured clean ground truth. The framework leverages multiple generators as complementary visual priors and applies spatial and color alignment, patch filtering, and local optimal-supervision selection to obtain reliable paired training data. Mixing WildMoiré with UHDM and DCID consistently improves ESDNet, Qwen-Image-Edit, and SDXL without modifying their deployed architectures or introducing additional inference-time cost. Additional experiments on challenging subsets of the UHDM and DCID test sets further show that WildMoiré complements existing paired datasets and improves complex moiré removal beyond our captured test set. The comprehensive ablations further validate the effectiveness of the proposed components, demonstrating that generative supervision improves robustness to challenging complex moiré patterns across different model architectures. These results also confirm the practical value of combining complementary generative priors with quality-controlled supervision for complex image demoiréing.

AI USE STATEMENT

Generative AI tools were used solely to assist with language polishing and were not used to generate figures, analyses, conclusions, or references. All AI-assisted text was carefully reviewed and verified by the authors. The authors take responsibility for the final content of this work.

REPRODUCIBILITY STATEMENT

The main paper describes data construction, the paired test-set protocol, and the reported comparisons. Appendix B.1 and Appendix B.3 provide generator adaptation and full training settings. Exact prompts, and filtering statistics are provided in Appendix B.2, and Appendix C.1, respectively.

REFERENCES

Black Forest Labs. Flux.2: Frontier visual intelligence, 2025. Model documentation.

Tim Brooks, Aleksander Holynski, and Alexei A. Efros. InstructPix2Pix: Learning to follow image editing instructions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 18392–18402, 2023. doi: 10.1109/CVPR52729.2023.01764.

Peter J. Burt and Edward H. Adelson. The laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, 31(4):532–540, 1983. doi: 10.1109/TCOM.1983.1095851.

Chaofeng Chen, Jiadi Mo, Jingwen Hou, Haoning Wu, Liang Liao, Wenxiu Sun, Qiong Yan, and Weisi Lin. Topiq: A top-down approach from semantics to distortions for image quality assessment. *IEEE Transactions on Image Processing*, 33:2404–2418, 2024.

Du Chen, Jie Liang, Xindong Zhang, Ming Liu, Hao Zeng, and Lei Zhang. Human guided ground-truth generation for realistic image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.

Du Chen, Tianhe Wu, Kede Ma, and Lei Zhang. Toward generalized image quality assessment: Relaxing the perfect reference quality assumption. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025.

Yijia Cheng, Xin Liu, and Jingyu Yang. Recaptured raw screen image and video demoiréing via channel and spatial modulations. *Advances in Neural Information Processing Systems*, 36: 40414–40425, 2023.

Google DeepMind. Nano banana 2, 2026. Model documentation.

Bin He, Ce Wang, Boxin Shi, and Ling-Yu Duan. Mop moiré patterns using MopNet. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 2424–2432, 2019.

Bin He, Ce Wang, Boxin Shi, and Ling-Yu Duan. Fhde2net: Full high definition demoireing network. In *European Conference on Computer Vision (ECCV)*, 2020.

Xuyi He, Yuhui Quan, Ruotao Xu, and Hui Ji. A universal scale-adaptive deformable transformer for image restoration across diverse artifacts. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12731–12741, 2025.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022.

Yuyang Hu, Mojtaba Sahraee-Ardakan, Aayush Bansal, Kangfu Mei, Chenyang Qi, Peyman Milanfar, and Mauricio Delbracio. Learning from a generative oracle: Domain adaptation for restoration. *arXiv preprint arXiv:2512.11121*, 2025.

Zhaoyang Huang, Xiaoyu Shi, Chao Zhang, Qiang Wang, Ka Chun Cheung, Hongwei Qin, Jifeng Dai, and Hongsheng Li. Flowformer: A transformer architecture for optical flow. In *European Conference on Computer Vision (ECCV)*, 2022.

Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. Musiq: Multi-scale image quality transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*, 2015.

Xiang Kong, Jun Zhao, Lei Sun, Rui Wu, and Lei Zhang. Ggt-100k: Generative ground truth for generalizable real-world image restoration. *arXiv preprint arXiv:2605.31039*, 2026.

Bolin Liu, Xiao Shu, and Xiaolin Wu. Demoir\'eing of camera-captured screen images using deep convolutional neural network. *arXiv preprint arXiv:1804.03809*, 2018.

Ilya Loshchilov and Frank Hutter. SGDR: Stochastic gradient descent with warm restarts. In *International Conference on Learning Representations*, 2017.

Yanting Mei, Zhilu Zhang, Xiaohe Wu, and Wangmeng Zuo. Image demoiréing using dual camera fusion on mobile phones. In *2025 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6. IEEE, 2025.

Bernard Moulden, Fred Kingdom, and Linda F. Gatley. The standard deviation of luminance as a metric for contrast in random-dot images. *Perception*, 19(1):79–101, 1990. doi: 10.1068/p190079.

OpenAI. Gpt image 2, 2026. Model documentation.

Hyunkook Park, An Gia Vien, Hanul Kim, Yeong Jun Koh, and Chul Lee. Unpaired screenshot image demoiréing with cyclic moiré learning. *IEEE Access*, 10:16254–16268, 2022. doi: 10.1109/ACCESS.2022.3149478.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. PyTorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, volume 32, pp. 8024–8035, 2019.

Remi Pautrat, Iago Suarez, Yifan Yu, Marc Pollefeys, and Viktor Larsson. Gluestick: Robust image matching by sticking points and lines together. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023.

Dustin Podell, Zion English, Kyle Lacey, Andreas Blattmann, Tim Dockhorn, Jonas Muller, Joe Penna, and Robin Rombach. Sdxl: Improving latent diffusion models for high-resolution image synthesis. *arXiv preprint arXiv:2307.01952*, 2023.

Yujing Sun, Yizhou Yu, and Wenping Wang. Moire photo restoration using multiresolution convolutional neural networks. *IEEE Transactions on Image Processing*, 27(8):4160–4172, 2018.

Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4): 600–612, 2004.

Chenfei Wu, Jiahao Li, Jingren Zhou, et al. Qwen-image technical report. *arXiv preprint arXiv:2508.02324*, 2025.

Haoning Wu, Zicheng Zhang, Weixia Zhang, Chaofeng Chen, Liang Liao, Chunyi Li, Yixuan Gao, Annan Wang, Erli Zhang, Wenxiu Sun, et al. Q-align: Teaching lmms for visual scoring via discrete text-defined levels. In *Proceedings of the 41st International Conference on Machine Learning*, 2024.

Yufeng Yang, Xianfang Zeng, Zhangqi Jiang, Fukun Yin, Jianzhuang Liu, Wei Cheng, Shiyu Liu, Yuqi Peng, Gang YU, Shifeng Chen, et al. Realrestorer: Towards generalizable real-world image restoration with large-scale image editing models. *arXiv preprint arXiv:2603.25502*, 2026.

Zemin Yang, Yujing Sun, Xidong Peng, Siu Ming Yiu, and Yuexin Ma. UniDemoiré: Towards universal image demoiréing with data generation and synthesis. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(9):9354–9362, 2025. doi: 10.1609/aaai.v39i9.33013.

Xin Yu, Peng Dai, Wenbo Li, Lan Ma, Jiali Shen, Jia Li, and Xiaojuan Qi. Towards efficient and scale-robust ultra-high-definition image demoireing. In *European Conference on Computer Vision (ECCV)*, 2022.

Haobo Yuan, Xiang Li, Tao Zhang, Zhe Huang, Shilin Xu, Shun Ji, Yuhang Tong, Lu Qi, Jiashi Feng, and Ming-Hsuan Yang. Sa2va: Marrying sam2 with llava for dense grounded understanding of images and videos. *arXiv preprint arXiv:2501.04001*, 2025.

Shanxin Yuan, Radu Timofte, Gregory Slabaugh, and Ales Leonardis. Aim 2019 challenge on image demoireing: Dataset and study. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, 2019.

Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

Bolun Zheng, Shanxin Yuan, Gregory Slabaugh, and Ales Leonardis. Image demoireing with learnable bandpass filters. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

Yunshan Zhong, Yuyao Zhou, Yuxin Zhang, Fei Chao, and Rongrong Ji. Learning image demoiréing from unpaired real data. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(7): 7623–7631, 2024. doi: 10.1609/aaai.v38i7.28595.

Libo Zhu, Zihan Zhou, Zhiyi Zhou, Yiyang Qu, Weihang Zhang, Keyu Shi, Yifan Fu, and Yulun Zhang. Combined flicker-banding and moire removal for screen-captured images. *arXiv preprint arXiv:2602.01559*, 2026.

APPENDIX

The content of the appendix involves:

- A discussion of the reliability of generative supervision in Sec. A.
- Additional details in Sec. B, including generator adaptation and behavior, prompts of candidate GT generation, and training settings.
- Additional experiments in Sec. C, including data-filtering statistics, effect of post-filtering threshold, and additional visual comparisons.
- Limitations and future work in Sec. D.

A RELIABILITY OF GENERATIVE SUPERVISION

Complex moiré images collected from public displays or online resources often lack exact clean source frames. Directly treating an unrestricted generative output as ground truth is nevertheless unreliable because a generator may remove interference while modifying text, objects, colors, or local structures. WildMoiré therefore uses generative models as offline candidate producers rather than as unrestricted restorers.

The resulting supervision should be regarded as a quality-controlled approximation of the unavailable clean reference. This is addressed in three ways. First, multiple priors increase the chance that at least one candidate provides a favorable balance between artifact removal and content preservation. Second, spatial and color alignment, synchronized local filtering, and image-quality metric gating prevent unqualified outputs from being admitted directly. Third, WildMoiré is mixed with reliable digital-source supervision from UHDM and DCID rather than replacing existing paired data. The real paired datasets continue to anchor content fidelity, while WildMoiré expands the degradation distribution toward complex real-world patterns.

The downstream evidence in the main paper provides an additional practical validation. Multi-prior supervision outperforms supervision restricted to either GPT-Image-2 or Nano-Banana-2, and improvements are observed across a CNN, a diffusion UNet, and a diffusion Transformer. These results do not imply that every selected patch is error-free, but they indicate that the quality-controlled set contains useful supervision that is not tied to a particular restoration architecture.

B ADDITIONAL DETAILS

B.1 GENERATOR ADAPTATION AND BEHAVIOR

As discussed above, different image generation and editing models exhibit complementary behaviors in restoration capability and content fidelity. The five selected models provide representative examples of such complementarity. GPT-Image-2 (OpenAI, 2026) is a powerful commercial model with strong semantic understanding and editing capability. It can remove severe moiré patterns effectively, but its aggressive editing may introduce noticeable changes in image content and color, as well as occasional local hallucinations. Nano-Banana-2 (Google DeepMind, 2026) is also a closed-source model and generally preserves the original content more faithfully than GPT-Image-2 (OpenAI, 2026). Nevertheless, it may still introduce content inconsistencies and leave conspicuous residual moiré in some challenging cases.

In comparison, SDXL (Podell et al., 2023), FLUX.2 (Black Forest Labs, 2025), and Qwen-Image-Edit (Wu et al., 2025) are open-source diffusion-based models whose generation behaviors are generally more conservative after fine-tuning with previous demoiréing datasets (Yu et al., 2022; Mei et al., 2025). Although their moiré-removal capability may be weaker than that of the commercial models, they often preserve the original structures and colors more faithfully. They therefore provide useful complementary candidates, especially in regions where GPT-Image-2 (OpenAI, 2026) and Nano-Banana-2 (Google DeepMind, 2026) produces severe hallucinations. Specifically, to produce

Table A: Prompts used for candidate GT generation.

Generative Models	Prompt
GPT-Image-2	Remove the moiré patterns from the image. Strictly keep all other content unchanged. Preserve the original structure, colors, and fine details exactly. Ensure that all text on the screen remains clear, accurate, and completely unmodified. Preserve the original aspect ratio.
Nano-Banana-2	Remove the moiré patterns, together with any abnormal stripes, color bands, or color blocks caused by the moiré effect. Strictly keep all other content unchanged, including the original structure, colors, and fine details. Ensure that all text on the screen remains clear, accurate, and completely unmodified. Preserve the original aspect ratio.
SDXL, FLUX.2, and Qwen-Image-Edit	Remove only the moiré patterns from the image. Keep all other content unchanged, including the colors, text, and fine details.

reasonable results of moiré removal, SDXL is fine-tuned for 80 epochs using the InstructPix2Pix training framework (Brooks et al., 2023), while FLUX.2 and Qwen-Image-Edit are fine-tuned for 100 epochs using DiffSynth-Studio. We use UHDM (Yu et al., 2022) and DCID (Mei et al., 2025) as the training set. During training, we also apply the proposed Scale Transformation Data Augmentation strategy described in the main paper to enhance the model's generalization ability.

B.2 Prompts of Candidate GT Generation

We employ GPT-Image-2 (OpenAI, 2026), Nano-Banana-2 (Google DeepMind, 2026), FLUX.2 (Black Forest Labs, 2025), SDXL (Podell et al., 2023), and Qwen-Image-Edit (Wu et al., 2025) to generate candidate GTs. Table A lists the exact English instructions used during data construction. GPT-Image-2 is given stronger content-preservation constraints because of its relatively aggressive editing behavior. Nano-Banana-2 receives a more explicit description of abnormal stripes and color regions to encourage more complete artifact removal. The three open-source priors use the same concise instruction. All three open-source priors are adapted with LoRA (Hu et al., 2022) before candidate generation.

B.3 Training and Implementation Settings

All models are implemented in PyTorch (Paszke et al., 2019). For ESDNet, we use a batch size of 8 and optimize the model for 150 epochs using Adam (Kingma & Ba, 2015) with an initial learning rate of 2×10^{-4}, $\beta_1 = 0.9$, and $\beta_2 = 0.999$. The learning rate is adjusted using cosine annealing with periodic warm restarts (Loshchilov & Hutter, 2017). SDXL is fine-tuned following the InstructPix2Pix training formulation (Brooks et al., 2023), with a batch size of 4 and a learning rate of 5×10^{-5} for 150 epochs. Qwen-Image-Edit-2511 is fine-tuned using DiffSynth-Studio with LoRA (Hu et al., 2022) rank 8, a batch size of 4, and a learning rate of 1×10^{-4} for 150 epochs. Both training and inference for ESDNet are conducted on an NVIDIA RTX A6000 GPU, while SDXL and Qwen-Image-Edit are trained and evaluated on an NVIDIA RTX PRO 6000 GPU. During pre-filtering, we set the τ_{img} and the τ_{lap} to 30, the τ_{ssim} to 0.9. During training, we set the upsampling scale s in scale transformation data augmentation strategy to 4.

C More Experiments

C.1 Statistics of Data Filtering

The candidate-generation stage yields synchronized patch groups, each containing one real moiré patch and five aligned candidate restorations. In the Pre-filtering stage, patches with relatively smooth textures or without moiré are filtered out. In the Post-filtering stage, patch groups are discarded if the selected candidate does not provide sufficient quality improvement over the input, as described in the main paper.

Table B: Statistics of data filtering on constructing WildMoiré.

Stage	Groups	Retained
Initial synchronized groups	13,615	100.00%
Pre-filtering	10,889	79.98%
Post-filtering	6,832	50.18%

Table C: Effect of the post-filtering threshold.

τ	#Pairs	PSNR↑	SSIM↑	LPIPS↓	MUSIQ↑	TOPIQ↑	Q-ALIGN↑
5	8,991	23.2863	0.7750	0.2720	43.0117	0.3309	3.8603
10	8,143	23.3411	0.7807	0.2643	43.7824	0.3365	3.9431
15	6,832	**23.3714**	**0.7821**	**0.2614**	**44.2109**	**0.3413**	**3.9607**
20	5,284	23.3647	0.7786	0.2684	43.7306	0.3374	3.9513
25	2,818	23.3021	0.7707	0.2925	42.9205	0.3315	3.8341

Table B reports the complete filtering funnel. Pre-filtering stage removes 2,726 groups. The final Post-filtering gate retains 6,832 groups, corresponding to 50.18% of the initial patch groups. This substantial reduction reflects the conservative objective of retaining only candidates that provide a clear quality improvement over the imperfect moiré observation.

C.2 Effect of the Post-filtering Threshold

The post-filtering threshold controls whether the selected candidate provides sufficient predicted quality improvement over the input. We evaluate the effect on downstream performance. For each $\tau \in \{5, 10, 15, 20, 25\}$, we mix the retained WildMoiré pairs with UHDM and DCID and train ESDNet using the same training setup. All results are evaluated on our 247-pair captured test set.

Table C shows that the tested thresholds 10, 15, and 20 give similar PSNR, ranging from 23.3411 to 23.3714 dB. A low threshold retains more pairs with smaller predicted quality gains, while a high threshold removes more training data. These results suggest that 10–20 is a practical range among the settings considered, rather than a single required value. We use $\tau = 15$ in this work to balance the quality requirement and downstream performance while retaining 6,832 training pairs.

C.3 More Visual Comparisons

Figure A presents additional qualitative comparisons on the independently captured test set, which contains challenging complex moiré patterns such as broad chromatic bands, mixed-frequency interference, text-rich regions, and fine scene structures. Across ESDNet, SDXL, and Qwen-Image-Edit, models trained with WildMoiré consistently reduce residual colored bands and local wave patterns more effectively than their corresponding baselines, while preserving displayed text, object boundaries, and underlying image details.

Figure B further presents challenging cases from the UHDM and DCID test sets. Although these samples are drawn from existing demoiréing benchmarks, they still contain complex moiré patterns that are difficult to remove using models trained only on the original UHDM and DCID data. Incorporating WildMoiré supervision improves restoration across all three architectures, indicating that the proposed generative supervision can complement existing training data and enhance robustness to challenging moiré patterns beyond our own captured test set.

D Limitations and Future Work

Our WildMoiré provides quality-controlled generative supervision for complex moiré images whose exact clean counterparts are unavailable, but it should not be regarded as a perfect substitute for digital-source ground truth. Although the proposed filtering and local selection procedures reduce unreliable candidates, subtle hallucinations, residual moiré, text modifications, or color inconsistencies may still

remain. Complementary fidelity checks or human verification could further reduce metric-specific selection bias.

Future work will explore complementary quality-assessment models, manual or vision-language-model-assisted verification for ambiguous samples, and stronger constraints for preserving text and fine structures. Expanding the collection to more cameras, display types, venues, and shooting conditions may further improve degradation coverage.



Figure A: **Additional qualitative comparisons on our test set.** Incorporating WildMoiré consistently improves complex moiré suppression across ESDNet, SDXL, and Qwen-Image-Edit while preserving the captured content.



Figure B: **Additional qualitative comparisons on challenging cases from the UHDM and DCID test sets.** Models trained only on UHDM and DCID may retain broad color interference, local waves, or mixed high- and low-frequency artifacts. Incorporating WildMoiré improves complex moiré suppression across all three trainable architectures while maintaining the captured content.