OMNICLEAR: SOFT EFFECTS REMOVAL FROM IMAGES WITHIN A VERSATILE MODEL

Anonymous authors
Paper under double-blind review

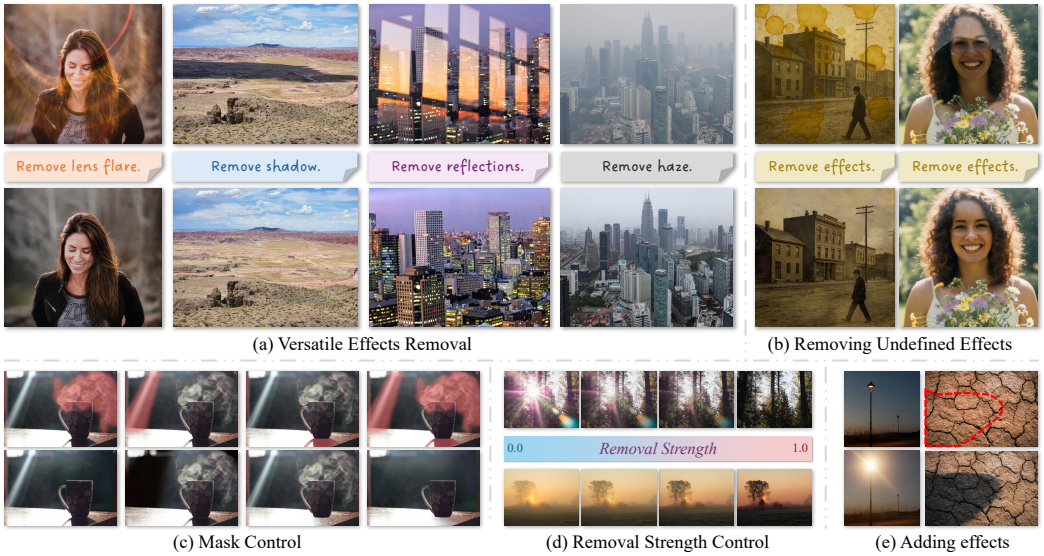

Figure 1: Our OmniClear eliminates multiple challenging (a) and even undefined (b) soft effects from in-the-wild images while preserving background identities. Besides, OmniClear supports precise pixel mask control (c), and removal strength control (d), allowing for intuitive and fine-grained restoration tailored to specific user needs. The framework is also capable of adding effects in the given region (e). Masks are global by default if not shown. **A demo video is included in the supplementary materials.**

ABSTRACT

Digital images are often degraded by soft effects such as lens flare, haze, shadows, and reflections, which reduce aesthetics even though the underlying pixels remain partially visible. The prevailing works address these degradations in isolation, developing highly specialized, specialist models that lack scalability and fail to exploit the shared underlying essences of these restoration problems. While specialist models are limited, recent large-scale pretrained generalist models offer powerful, text-driven image editing capabilities. while recent general-purpose systems (*e.g.*, GPT-4o, Flux Kontext, Nano Banana) require detailed prompts and often fail to achieve robust removal on these fine-grained tasks or preserve identity of the scene. Leveraging the common essence of soft effects, *i.e.*, semi-transparent occlusions, we introduce a foundational versatile model, capable of addressing diverse degradations caused by soft effects within a single framework. Our approach centers on fine-tuning a potent inpainting model on a large-scale, curated dataset of paired images, enabling it to learn robust restoration priors. Our method provides simple and intuitive user control, either global removal or mask-based removal with strength control, making interaction easier while ensuring higher reliability. Extensive experiments demonstrate that our unified model outperforms both prior specialist methods and popular general-purpose models, achieving robust and stable performance on in-the-wild scenarios.

1 INTRODUCTION

Images captured in real-world environments inevitably suffer from degradations. A common class of such "soft" effects includes optical phenomena (e.g., lens flare, reflections) and atmospheric conditions (e.g., haze, fog). These effects corrupt scene radiance additively or multiplicatively, degrading contrast, color fidelity, and fine details (Le & Samaras, 2019; Wan et al., 2017). Consequently, image quality and visibility are compromised, and in severe cases, occlusions cause irreversible information loss, rendering recovery fundamentally ill-posed (He et al., 2010; Wu et al., 2021; Le & Samaras, 2019).

To restore image structures, most existing works address each degradation type separately. For instance, dehazing has progressed from prior-based methods such as the Dark Channel Prior (DCP) (He et al., 2010) to deep networks estimating scattering parameters or directly predicting clean images (Li et al., 2017; Song et al., 2023; Chen et al., 2021; Engin et al., 2018; Chen et al., 2019). Similarly, shadow, flare, and reflection removal adopt task-specific designs (Le & Samaras, 2019; Dong et al., 2024; Wu et al., 2021; Xue et al., 2025; Zhu et al., 2024; Wan et al., 2017), relying on physical modeling, layer decomposition, or elaborate data and network strategies to mitigate ill-posedness. While such methods achieve strong task-specific performance, recent works (Chen et al., 2025a; Li et al., 2020b; Potlapalli et al., 2023) attempt to unify multiple degradations within one framework. Yet these models remain limited in scalability and robustness when facing extreme, diverse real-world conditions. This motivates the development of foundation models trained on large-scale data to achieve stronger generalization and resilience in the wild.

Concurrently, the rise of powerful foundation models like GPT-4o (Hurst et al., 2024) and Nano Banana (Gemini 2.5 Flash Image) (Comanici et al., 2025; Google, 2025) has introduced general-purpose, text-driven image generation/editing based on Multi-modal Large Language Models (MLLMs). These models can interpret complex prompts and perform realistic edits. However, for fine-grained tasks like soft effect removal, they exhibit significant limitations. Their performance is often unstable and heavily reliant on meticulously crafted text prompts. More critically, they lack the precise, pixel-wise control required for high-fidelity restoration and identity preservation. Treating soft effect removal as a general inpainting task can lead to the alteration of local image structures or the identity of objects within the scene, which makes them unreliable for professional photo editing and critical computer vision pipelines.

Despite their diverse appearances, effects such as lens flare, haze, reflections, and shadows share the same intrinsic property: they are all semi-transparent occlusions that preserve the identity of the underlying scenes. To this end, we define a unified and extensible task, termed Soft Effects Removal (SER). This task is highly challenging. First, these effects are typically entangled with the scene itself, rather than merely superimposed as simple overlays. Second, the local image structures, and even pixel-level identities, should be precisely preserved. Third, regions that are fully occluded or invisible (*e.g.*, overexposed areas in lens flare or areas covered by extremely dense haze) must be plausibly reconstructed.

To effectively tackle these challenges, we introduce **OmniClear** (Fig. 1 (a) & (b)), a data-centric versatile model for Soft Effects Removal. Our method is built upon two key points. First, we curated a large-scale dataset of approximately 3.8M balanced, high-quality, pixel-aligned image pairs. By unifying existing open-source datasets and augmenting them with extra real-world and synthetic data, we provide the precise supervision our model needs to learn content invariance. Second, as shown in Fig. 1 (c) & (d), we implemented fine-grained user controls, including pixel-level masks to define the removal area and strength levels to modulate the removal strength, making the process highly controllable. Beyond restoration, OmniClear can also perform aesthetic edits, such as enhancing existing effects or generating new, realistic ones on clean images (Fig. 1 (e)). Our method achieves state-of-the-art results on multiple public benchmarks and demonstrates significantly better generalization on in-the-wild testing data.

In summary, our main contributions can be summarized as follows:

- **A Versatile SER Model:** Proposed a single, versatile model OminClear for removing diverse soft effects in the wild. Our model achieves state-of-the-art performance on each task and surpasses much larger general-purpose models such as Nano Banana.
- **A Large-Scale Dataset for Generalization:** Curated a large-scale dataset of$\sim$3.8M image pairs, providing vast data distribution for strong generalization on challenging in-the-wild data.

- **Controllable Editing:** Developed fine-grained user controls for SER tasks, including spatial masks and strength levels, to enable precise and controllable effect removal.

2 Related Work

2.1 Isolated Effects Removal

Lens flare removal. Previous learning-based methods improved data synthesis by considering camera ISP to enhance realism and generalization (Zhou et al., 2023; 2025). Concurrently, architectural innovations emerged, including self-supervised methods to disentangle co-occurring flares (He et al., 2025), while others explicitly separated light source preservation from flare removal using dedicated detection modules (Ghodesawar et al., 2023), and networks leveraging both spatial and frequency domains (Vasluianu et al., 2024). More recently, large pretrained Latent Diffusion Models (LDMs) are adpated to leverage their powerful generative priors (Zhou et al., 2024). The development of these methods has also been heavily reliant on specialized datasets, from semi-synthetic ones (Wu et al., 2021), Flare7K (Dai et al., 2022), to real-world paired datasets (Dai et al., 2024).

Reflection removal. Early methods for single-image reflection removal (SIRR) focused on iterative refinement using edge maps (Fan et al., 2017) or recurrent networks (Yang et al., 2018; Li et al., 2020a). Subsequent research shifted towards improving training data realism by learning non-linear blending (Wen et al., 2019), employing physically-based rendering (Kim et al., 2020), and modeling glass absorption (Zheng et al., 2021). Architectural innovations followed, introducing location-aware modules (Dong et al., 2021) and advanced attention mechanisms (Huang et al., 2025; Zhang et al., 2025) to better distinguish between layers. More recent paradigms reduce reliance on paired data through unsupervised deep image priors (RahmaniKhezri et al., 2022) or by using Diffusion Models to generate guiding prompts (Wang et al., 2024a). This progress has been underpinned by the creation of key real-world benchmarks like SIR^2 (Wan et al., 2017) and the large-scale RRW dataset (Zhu et al., 2024).

Shadow removal. Initial approaches to shadow removal relied on traditional physical priors and optimization frameworks (Guo et al., 2011; Zhang et al., 2015). The advent of deep learning introduced end-to-end models like DeshadowNet (Qu et al., 2017) and methods that decomposed images into shadow-free and matte layers (Le & Samaras, 2019). Subsequent architectural advancements included using Generative Adversarial Networks (GANs) for joint detection and removal (Wang et al., 2018), fusing synthetic exposure pairs (Fu et al., 2021), and learning via shadow generation (Liu et al., 2021). More recent trends focus on eliminating the dependency on explicit shadow masks, utilizing mask-free transformers (Dong et al., 2024) or reformulating the problem as a dense prediction task (Lin et al., 2025). The progress in this field has been propelled by benchmarks like SRD (Qu et al., 2017), ISTD (Wang et al., 2018), and the newer high-resolution WSRD dataset (Vasluianu et al., 2023).

Haze removal. Single-image dehazing evolved from early methods based on statistical priors like the Dark Channel Prior (DCP) (He et al., 2010) to data-driven deep learning. Initial deep learning works included lightweight end-to-end networks (Li et al., 2017), hybrid models that learned priors for traditional optimization (Yang & Sun, 2018), and unpaired training with GANs to address data scarcity (Engin et al., 2018). Architectural innovations, such as gated context aggregation (Chen et al., 2019) and Vision Transformers (Song et al., 2023), were later introduced to better handle non-uniform haze. Recent efforts focus on closing the synthetic-to-real domain gap by generating more physically plausible training data (Chen et al., 2021) or leveraging diffusion models for realistic haze synthesis (Wang et al., 2025). This progress has been consistently driven by the development of comprehensive benchmarks (Li et al., 2018; Zhang et al., 2024; Islam et al., 2024).

Apart from them, some works delve into unified methods to restore image quality from multiple degradations caused by bad weathers (Li et al., 2020b; Potlapalli et al., 2023; Chen et al., 2025a). Despite the achievements from all these methods, key challenges persist including the domain gaps of synthetic datasets and limited diversity in real-world datasets, while current methods still struggle with scalable training with robust generalization abilities, as well as handling various types of challenging soft effects simultaneously.

2.2 Prompt-based Image Editing

Prompt-based image editing originated from diffusion models, enabled by deterministic inversion techniques like DDIM (Song et al., 2020) that map real images to an editable latent space. Initial methods controlled edits by manipulating internal model structures, such as altering cross-attention

maps to preserve layout (Hertz et al., 2022) or fine-tuning the entire model on a single image for complex, non-rigid changes (Kawar et al., 2023). The field has since evolved towards more direct user control, with models trained to follow natural language instructions (Brooks et al., 2023) or allow for interactive, point-based spatial adjustments (Shi et al., 2024). This shift towards more precise, semantic editing is increasingly powered by the advanced contextual understanding of Multimodal Large Language Models (MLLMs) (Hurst et al., 2024; Comanici et al., 2025; Bai et al., 2025). However, current approaches still often lack fine-grained pixel control and can struggle to perfectly preserve the subject's identity during transformation.

3 METHODOLOGY

3.1 DATA CURATION

A powerful foundation model requires large-scale, high-quality, and diverse training data. To equip OmniClear with robust generalization, we curated a comprehensive dataset by unifying pixel-aligned image pairs from four representative tasks: lens flare, shadow, haze, and reflection removal. This integration enables the model to learn a broad restoration representation while preserving content identity.

Public datasets. We incorporate multiple benchmark datasets spanning the four domains (see Table 1 and supplementary materials for details). Despite their usefulness, these datasets exhibit imbalance, such as the scarcity of large-scale flare removal data and limited diversity in haze scenarios.

Data expansion. To remedy these gaps and increase data volume, we expand training data through three complementary sources: real-world captures, 2D synthesis, and 3D rendering.

- *Lens flare.* The key bottleneck lies in insufficient data. We therefore construct 78 indoor and outdoor 3D scenes in Blender (Blender Online Community, 2018), rendering about 70K paired images, named HALO dataset. Unlike Flare7K (Dai et al., 2022), which overlays flare layers on clean images, our rendered data produce physically consistent and realistic flare effects. The dataset covers diverse flare patterns, including reflective flare, glare, shimmer, and streaks.
- *Shadow.* While public datasets cover both indoor and outdoor scenes, they contain only $\sim$5K pairs. To scale up, we add an additional 26K photo pairs. Specifically, we repurpose internal object-effect removal data: by stitching objects without shadows into background images, we synthesize corresponding shadow-free versions to form the Large Real-world Shadow Removal Dataset (LR-SRD).
- *Haze.* Existing synthetic datasets (RESIDE, HAZESPACE) often appear uniform or algorithmically simplistic. To generate more realistic and challenging cases, we use their clean ground-truth images with monocular depth (Ke et al., 2025), and apply a physically motivated atmospheric rendering pipeline. This allows precise control of parameters such as visibility, airlight color, scatter, and optical thickness. To simulate non-homogeneous haze or fog, we introduce procedural noise fields and path blurring, yielding realistic textures of haze, smoke, and fog. More synthesis details are provided in the supplementary material.

These expanded datasets extend coverage to underrepresented scenarios and complement public benchmarks, enhancing OmniClear's robustness in the wild. A detailed breakdown is given in Table 1, with representative samples in Fig. 2.

3.2 FRAMEWORK

As shown in Fig. 3, OmniClear is a unified framework designed to tackle multiple soft effect removal tasks. Inspired by UniReal (Chen et al., 2025b), the core architecture reformulates these diverse tasks as a problem of *discontinuous frame generation* within a latent diffusion model. The process begins with a Variational Autoencoder (VAE) (Kingma & Welling, 2013) encoding the input image into a compact latent space, while a text encoder processes a task-specific prompt (*e.g.*, "*remove haze*") to generate instructive embeddings. These conditional inputs (image latent and textual embeddings) are then concatenated with the noisy target latent and fed as a sequence to a Diffusion Transformer (DiT). The DiT's full attention mechanism operates on this sequence, allowing it to iteratively predict and remove noise from the target latent by conditioning on both the visual context and the textual instructions. Finally, the fully denoised latent is passed through the VAE decoder to reconstruct the final, effect-free image. The model is trained using a mean squared error (MSE)

Table 1: Summary of datasets curated for OmniClear training. "†" represents the datasets curated by us, "*" represents the datasets which we re-synthesis effects with our own algorithm.

Task	Dataset	Type	Description	Pairs
Lens flare	FlareReal600 (Dai et al., 2024)	Real-World	Nighttime flares, Streetview, Cityscapes, Outdoor	0.6k
	HALO†	3D Synthetic	Rendered, Various flares and scenes, Indoor & Outdoor	70k
Shadow	WSRD+ (Vasluianu et al., 2023)	Real-World	Object-level, Close-view, Rich texture, Complex shadows	1k
	ISTD+ (Wang et al., 2018)	Real-World	Simple-shaped shadows, Monotonous scenes, Outdoor	1.3k
	SRD (Qu et al., 2017)	Real-World	Various scenes, Outdoor	2.6k
	LR-SRD†	Real-World	Object-level, Close-view, Hard & Soft shadow, Indoor & Outdoor	26k
Haze	Haze-R (Ancuti et al., 2018b;a; 2019; 2020; 2021; 2023; 2024)	Real-World	Collection includng: I-HAZE, O-HAZE, Dense-Haze, NH-Haze, etc., Homogeneous & Non-Homogeneous, Indoor & Outdoor	0.3k
	REVIDE (Zhang et al., 2021)	Real-World	Video Frames, Indoor	1.9k
	LM-Haze (Zhang et al., 2024)	Real-World	Multi-level haze, Homogeneous, Indoor	5k
	HAZESPACE* (Islam et al., 2024)	2D Synthetic	Multi-level haze, Vast range of scenes, Outdoor	24×70k
	RESIDE* (Li et al., 2018)	2D Synthetic	Multi-level haze, Indoor & Outdoor	290k
	SYN-HAZE*	2D Synthetic	Multi-level haze, Synthetic scenes, Include extremely dense haze, Indoor & Outdoor	24×70k
Reflection	RRW (Zhu et al., 2024)	Real-World	Various scenes, Diverse glass and reflection types	14.9k
	POLAR-RR (Lei et al., 2020)	Real-World	Polarization-based, Indoor	0.8k
	RFC (Lei & Chen, 2021)	Real-World	Flash-induced reflections	5k
	BDN (Yang et al., 2018)	2D Synthetic	Linearly Blended, Public Image Sources	50k

Figure 2: Visualization of our curated data samples and synthetic haze by our method.

loss between the predicted noise and the ground truth noise, with a timestep-dependent weighting scheme to balance the contributions of different noise levels.

Random Masking Strategy. As established in the framework, a mask can be supplied as a condition to guide the denoising process toward a specific spatial region. However, most of the training sets do not contain the mask of effects. To ensure the model can robustly handle any user-provided mask shape, we adopt a random masking strategy. During training, following (Suvorov et al., 2022; Zheng et al., 2022) we synthesize a wide variety of binary masks M by randomly combining geometric primitives like rectangles with free-form, stroke-like patterns that simulate user brush strokes. Afterwards, providing pairs $\{I_{input}, I_{gt}\}$, we generate the corresponding training supervision I_{target} where the effect is removed only within the masked region via simply compositing I_{input} and I_{gt} with the mask, as shown in Equation 1. Note that the regions of effects in the input image are unavailable, hence the masks do not necessarily cover them. In this way the behaviors the model to learn is summarized as following:

- Region inside the mask w/ effects: remove effects based on the strength;
- Region inside the mask w/o effects: keep identical;
- Region outside the mask: keep identical.

Additionally, to make the supervision natural-looking, we blur the mask boundary via dilation and Gaussian blur. This strategy exposes the model to a vast distribution of possible mask shapes, enhancing its generalization capability for arbitrary user edits, and removing the sepcific effect regions.

Removal Strength Control. Beyond specifying *where* to remove an effect, OmniClear allows users to control *how much* of the effect is removed. This is achieved by training the model to interpret

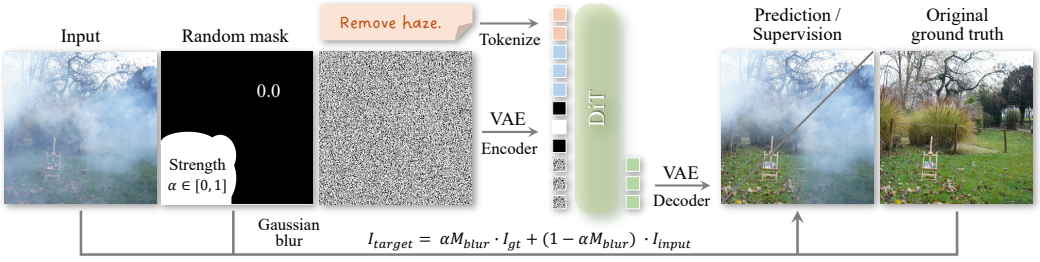

Figure 3: The architecture of OmniClear. During training, the mask is randomly synthesized along with a scalar strength, and the supervision is composed by the input image and the original ground truth via the mask and the strength.

continuous values in the conditional mask as an indicator of removal intensity. During the training process, for each sample, we uniformly sample a floating-point scalar value to represent "strength", denoted as $\alpha \in [0, 1]$. Instead of conditioning the model on a binary mask M, we provide a soft value mask αM. The model thus learns to associate a mask value of 1.0 with complete removal, 0.0 with no change, and intermediate values with partial removal. On the other hand, along with the aforementioned blurred mask, the training target is generated by linearly interpolating between the clean ground truth (I_{gt}) and the input with effects (I_{input}) using the randomly sampled α. Formally, the supervision during training is computed as following:

$$I_{target} = \alpha M_{blur} \cdot I_{gt} + (1 - \alpha M_{blur}) \cdot I_{input} \tag{1}$$

This joint strategy of conditioning on a soft mask while generating a correspondingly blended target enables the model to learn a continuous and intuitive mapping from the control signal to the desired degree of effect removal.

Handling Undefined Effects. Our framework also extends to zero-shot generalization on unseen soft effects through two complementary fine-tuning strategies. First, we randomly replace task-specific prompts with a generic prompt *"remove effects"*, encouraging the model to capture a shared notion of removal across tasks. Second, we introduce an auxiliary task using clean images: random masks are generated and overlaid with semi-transparent or opaque regions to synthesize degraded inputs, which are trained exclusively with the generic prompt. This prevents overfitting to predefined effect categories and compels the model to learn the broader concept of removing arbitrary occlusions, thereby enabling generalized restoration.

Adding & Enhancing Effects. We can easily invert the removal task to adding or enhancing effects by swapping the roles of the input and the target. Similarly, the adding or enhancing ability is controlled by the mask and strength given by users. We demonstrate this ability in Fig. 5.

4 Experiments

4.1 Benchmarks and Baselines

Benchmarks. We evaluate OmniClear across four soft-effect tasks on widely used benchmarks. For *lens flare removal*, we adopt the Flare7K real-world test set (Dai et al., 2022). For *shadow removal*, we test on SRD (Qu et al., 2017), ISTD+ (Wang et al., 2018), and the high-resolution WSRD+ (Vasluianu et al., 2023). For *haze removal*, we use the SOTS and HSTS subsets of RESIDE (Li et al., 2018). For *reflection removal*, we employ SIR^2 (Wan et al., 2017) and the Nature test set (Li et al., 2020a). OmniClear is fine-tuned on the training splits of these datasets for domain adaptation. Evaluation uses standard full-reference metrics: PSNR and SSIM.

To assess real-world robustness, we collected 39 in-the-wild images containing haze, fog, flare, reflection, and shadow. As no ground truth is available, we report reference-free metrics (LIQE (Zhang et al., 2023), contrast gain (Wang et al., 2024b)), and a reference-based evaluation with Qwen2.5-VL-72B (Bai et al., 2025), a vision-language model instructed to judge the percentage of effect removal. We will further discuss these metrics in the supplementary material.

Baselines. We compare against both generalist and specialist methods. Generalist baselines include GPT-4o (Hurst et al., 2024), FLUX Kontext (Labs et al., 2025), Nano Banana (Google, 2025), and Seedream 4.0 (ByteDance, 2024). Specialist baselines cover:

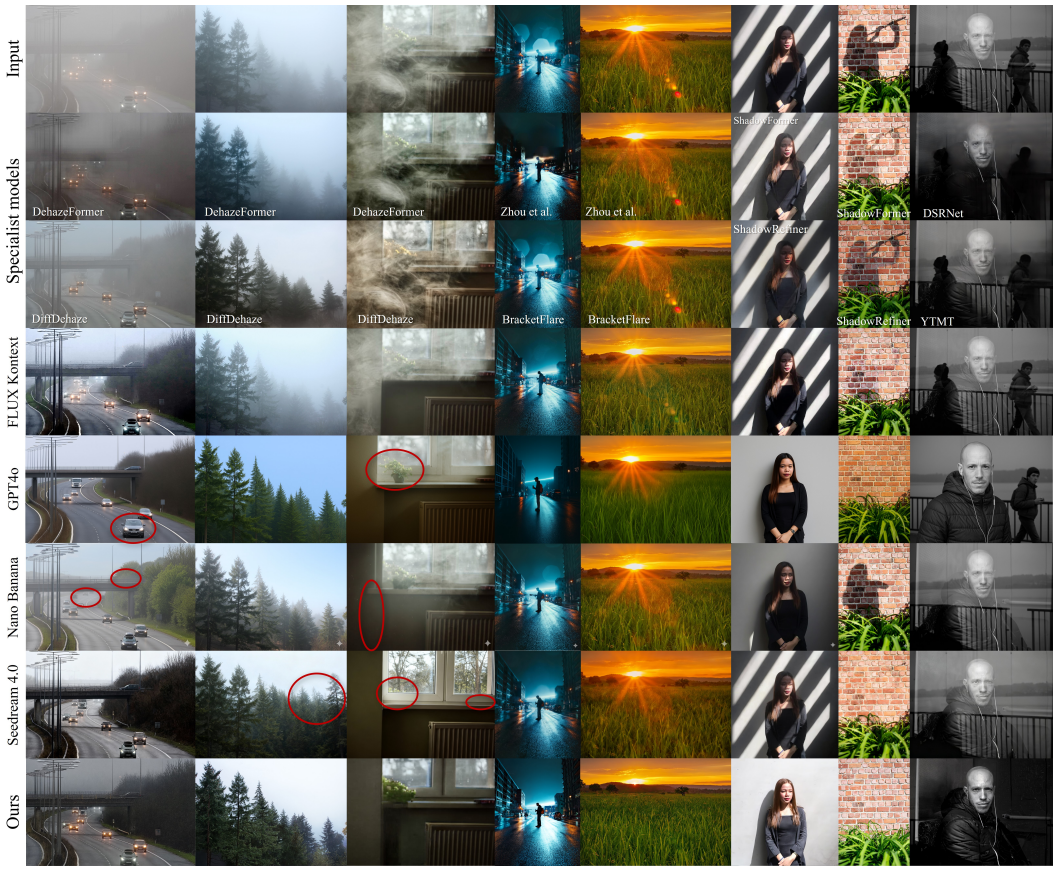

Figure 4: Comparisons with state-of-the-art specialist and generalist models on in-the-wild testing data. For effect removal, our method significantly outperforms these baselines. Moreover, generalist models fail to preserve the identity of background objects, some of the discrepancies are circled.

Lens flare: (Zhang et al., 2020; Zhou et al., 2023; Dai et al., 2022), BracketFlare (Dai et al., 2023), Difflare (Zhou et al., 2024);

Dehazing: DCP (He et al., 2010), AOD-Net (Li et al., 2017), GCANett (Chen et al., 2019), PSD (Chen et al., 2021), Dehazeformer (Song et al., 2023), MSF-Net (Zhu et al., 2021), UCL-Dehazet (Wang et al., 2024b), DiffDehaze (Wang et al., 2025);

Shadow removal: ShadowFormer (Guo et al., 2023a), ShadowRefiner (Dong et al., 2024), DCShadowNet (Jin et al., 2021), ShadowDiffusion (Guo et al., 2023b), StableShadowDiff (Xu et al., 2025);

Reflection removal: (Zhang et al., 2018), YTMT (Hu & Guo, 2021), DSRNet (Hu & Guo, 2023), PromptRR (Wang et al., 2024a), L-DiffER (Hong et al., 2024).

4.2 Comparisons with State-of-The-Art

Qualitative Comparisons. Fig. 4 visually compares OmniClear with state-of-the-art models on challenging in-the-wild images. Specialist models generalize poorly to out-of-domain data, often resulting in incomplete removal or new artifacts. Meanwhile, powerful generalist models like Nano Banana and FLUX Kontext suffer from instability and fail to preserve scene details, leading to significant content drift (highlighted by red circles). In contrast, OmniClear effectively removes a wide range of soft effects while remaining highly faithful to the original image content, producing clean and content-consistent results.

Quantitative Comparisons. To assess real-world generalization,we first conduct a comparison on a challenging in-the-wild test set using no-reference metrics, shown in Table 2. In this more difficult setting, OmniClear significantly outperforms both specialist and generalist baselines in terms of perceptual quality and removal efficacy, achieving the highest LIQE, Contrast gain, and QwenQA scores across nearly all tasks, which highlights its robust generalization. We then evaluate OmniClear against specialists on eight standard benchmarks using full-reference metrics (Table 3). The

Table 2: No-reference quantitative comparison on in-the-wild images for four SER tasks. We report results from multiple image quality assessment metrics.

Haze				**Shadow**			
Method	**LIQE↑**	**Contrast↑**	**QwenQA↑**	**Method**	**LIQE↑**	**Contrast↑**	**QwenQA↑**
Dehazeformer	1.9999	+0.74	0.0	ShadowFormer	3.3704	+3.09	18.8
DiffDehaze	1.5624	+0.03	9.1	ShadowRefiner	3.5179	-2.30	26.3
Flux Kontext	2.2584	+3.85	22.7	Flux Kontext	3.3184	+0.73	36.3
Nano Banana	2.6864	+0.26	27.3	Nano Banana	3.6399	-4.93	35.0
Seedream 4.0	2.1253	+2.60	52.7	Seedream 4.0	2.7640	-3.58	36.3
Ours	**2.8225**	**+5.57**	**60.0**	**Ours**	**3.7764**	**+3.61**	**65.0**
Lens Flares				**Reflections**			
Method	**LIQE↑**	**Contrast↑**	**QwenQA↑**	**Method**	**LIQE↑**	**Contrast↑**	**QwenQA↑**
Uformer	1.3832	-4.39	30.9	YTMT	1.1187	-2.30	14.4
BracketFlare	3.3377	-10.72	13.6	DSRNet	1.6975	-4.17	17.8
Flux Kontext	3.0574	-0.31	62.7	Flux Kontext	1.7009	+0.71	8.9
Nano Banana	3.0358	-4.05	71.8	Nano Banana	2.0935	-1.25	56.7
Seedream 4.0	2.1643	-4.41	73.6	Seedream 4.0	1.6145	-0.96	53.5
Ours	**3.5186**	**+2.33**	**92.7**	**Ours**	**2.2257**	**+1.83**	**75.6**

Table 3: Quantitative comparison with state-of-the-art methods across four soft effect removal tasks. We report PSNR (↑) and SSIM (↑) on eight benchmarks. Our unified model is compared against specialist methods in each respective category.

Lens Flares			**Haze**				
Method	Flare7k		**Method**	HSTS		SOTS	
	PSNR	SSIM		PSNR	SSIM	PSNR	SSIM
Zhang et al. (2020)	21.02	0.784	DCP	17.01	0.803	18.38	0.819
Zhou et al. (2023)	25.18	0.872	AOD-Net	19.68	0.835	20.08	0.861
UNet (Dai et al., 2022)	26.11	0.879	GCANet	21.37	0.874	21.66	0.867
Restormer (Dai et al., 2022)	26.28	0.883	PSD	19.37	0.824	20.49	0.844
Uformer (Dai et al., 2022)	26.98	0.890	MSFNet	31.03	0.931	**30.07**	0.939
Difflare	26.06	**0.898**	UCL-Dehaze	26.87	0.933	25.21	0.927
Ours	**27.34**	0.891	**Ours**	**32.17**	**0.962**	29.52	**0.955**

Shadow							**Reflections**				
Method	WSRD+		ISTD+		SRD		**Method**	SIR2		Nature20	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM		PSNR	SSIM	PSNR	SSIM
ShadowFormer	25.44	0.820	32.78	0.934	30.58	0.958	Zhang et al. (2018)	22.45	0.872	20.37	0.772
ShadowRefiner	26.04	0.827	31.03	0.928	-	-	YTMT	23.05	0.886	21.03	0.802
DCShadowNet	21.62	0.593	25.50	0.694	-	-	DSRNet	24.97	0.907	21.70	0.820
ShadowDiffusion	-	-	31.08	0.950	31.91	0.968	PromptRR	24.22	0.876	21.00	0.814
StableShadowDiff	26.26	0.827	35.19	**0.970**	33.63	0.968	L-DiffER	25.18	0.911	23.95	**0.831**
Ours	**26.91**	**0.829**	**35.59**	0.964	**34.16**	**0.971**	**Ours**	**25.98**	**0.911**	**24.17**	0.812

results show our unified model achieves state-of-the-art performance, consistently outperforming or matching specialist models by obtaining top scores across all four tasks, including the highest PSNR on multiple benchmarks.

4.3 Discussions and Applications

Ablation studies. We conduct an ablation study to validate the effectiveness of our joint-task learning strategy. As shown in Table 4, we compare our full model, trained with Joint-Task Learning (JTL), against four same models trained independently using Single-Task Learning (STL). The results clearly indicate that the JTL model consistently outperforms the STL models across all four tasks on their respective benchmarks. This superiority suggests that by learning a unified representation from diverse soft effects, OmniClear develops a more robust and generalizable feature space that benefits all individual tasks.

Strength control. As illustrated in Figure 5(a), OmniClear provides fine-grained control over the intensity of the effect removal. Users can specify a continuous strength value, allowing for a smooth transition from partial reduction to complete removal of the artifact. This feature offers greater flexibility for users to achieve their desired level of restoration.

Mask control. OmniClear supports precise, localized editing through mask-based control, as shown in Figure 5(b). By providing a binary mask, users can designate specific spatial regions for effect

Table 4: Ablation study on training strategies. JTL (Joint-Task Learning) represents our full OmniClear, while STL (Single-Task Learning) denotes models trained separately for each task.

Method	Lens Flares		Haze		Shadow		Reflections	
	Flare7k		HSTS		ISTD+		SIR2-wild	
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
STL	27.18	0.890	31.91	0.963	35.43	**0.963**	26.40	0.876
JTL	**27.34**	**0.891**	**32.17**	0.962	**35.59**	**0.964**	**27.44**	**0.918**

(a) Strength Control (c) Add / Enhance effects

(b) Mask Control (d) Zero-shot effects removal

Figure 5: (a) Illustration of Strength Control for effect removal. (b) Illustration of Mask Control for accurate user regional editing. (c) Adding realistic effects to clean image, or enhance current effects for flexible editing purpose. (d) Zero-shot generalization ability on multiple unseen degradations like rain, stain, etc.

removal while leaving the rest of the image untouched. This allows for targeted and accurate edits tailored to user needs.

Effects addition and enhancement. Beyond removal, the OmniClear framework is also capable of generative tasks. As demonstrated in Figure 5(c), by inverting the process, our model can realistically add new soft effects to clean images or enhance existing ones. This versatility makes it a valuable tool for creative photo editing and data augmentation.

Zero-shot removal. OmniClear exhibits strong generalization capabilities to novel degradations not seen during training. As shown in Figure 5(d), the model can perform zero-shot removal of unseen artifacts such as rain and stains. This ability underscores the robustness of the learned features and the model's potential to handle a wide array of real-world image restoration challenges beyond its core training tasks.

Reproducibility Statement The portion of our method that relies on public datasets is reproducible, as our implementation is based on the open-source DiT codebase.

5 Conclusion and Limitations

In this work, we introduced OmniClear, a unified foundation model for Soft Effects Removal (SER) that effectively handles diverse degradations including lens flare, haze, shadows, and reflections. Built upon a Diffusion Transformer trained on a large-scale pair-wise dataset, OmniClear overcomes the poor generalization of specialist models and the content inconsistency of generalist approaches. Extensive experiments demonstrate that our model achieves state-of-the-art performance on standard benchmarks and superior perceptual quality on in-the-wild images. Beyond high-quality removal, the framework provides fine-grained user controls, supports creative effect generation, and shows strong zero-shot capabilities on unseen degradations. Key limitations include its significant computational cost and the extensive resources required for training. Nevertheless, OmniClear represents a significant step towards a universal and controllable solution for high-fidelity image restoration.

REFERENCES

Codruta O Ancuti, Cosmin Ancuti, Radu Timofte, and Christophe De Vleeschouwer. O-haze: a dehazing benchmark with real hazy and haze-free outdoor images. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 754–762, 2018a.

Codruta O Ancuti, Cosmin Ancuti, Mateu Sbert, and Radu Timofte. Dense-haze: A benchmark for image dehazing with dense-haze and haze-free images. In *2019 IEEE international conference on image processing (ICIP)*, pp. 1014–1018. IEEE, 2019.

Codruta O Ancuti, Cosmin Ancuti, and Radu Timofte. Nh-haze: An image dehazing benchmark with non-homogeneous hazy and haze-free images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pp. 444–445, 2020.

Codruta O Ancuti, Cosmin Ancuti, Florin-Alexandru Vasluianu, and Radu Timofte. Ntire 2021 nonhomogeneous dehazing challenge report. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 627–646, 2021.

Codruta O Ancuti, Cosmin Ancuti, Florin-Alexandru Vasluianu, Radu Timofte, Han Zhou, Wei Dong, Yangyi Liu, Jun Chen, Huan Liu, Liangyan Li, et al. Ntire 2023 hr nonhomogeneous dehazing challenge report. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1808–1825, 2023.

Codruta O Ancuti, Cosmin Ancuti, Florin-Alexandru Vasluianu, Radu Timofte, Yidi Liu, Xingbo Wang, Yurui Zhu, Gege Shi, Xin Lu, Xueyang Fu, et al. Ntire 2024 dense and non-homogeneous dehazing challenge report. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6453–6468, 2024.

Cosmin Ancuti, Codruta O Ancuti, Radu Timofte, and Christophe De Vleeschouwer. I-haze: A dehazing benchmark with real hazy and haze-free indoor images. In *International conference on advanced concepts for intelligent vision systems*, pp. 620–631. Springer, 2018b.

Shuai Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, Sibo Song, Kai Dang, Peng Wang, Shijie Wang, Jun Tang, et al. Qwen2. 5-vl technical report. *arXiv preprint arXiv:2502.13923*, 2025.

D Blender Online Community. Blender—a 3d modelling and rendering package. *Blender Foundation*, 2018.

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.

ByteDance. Seedream-4: A large-scale text-to-image generation model, 2024. URL `https://seed.bytedance.com/en/seedream4_0`.

Dongdong Chen, Mingming He, Qingnan Fan, Jing Liao, Liheng Zhang, Dongdong Hou, Lu Yuan, and Gang Hua. Gated context aggregation network for image dehazing and deraining. In *2019 IEEE winter conference on applications of computer vision (WACV)*, pp. 1375–1383. IEEE, 2019.

I Chen, Wei-Ting Chen, Yu-Wei Liu, Yuan-Chun Chiang, Sy-Yen Kuo, Ming-Hsuan Yang, et al. Unirestore: Unified perceptual and task-oriented image restoration model using diffusion prior. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 17969–17979, 2025a.

Xi Chen, Zhifei Zhang, He Zhang, Yuqian Zhou, Soo Ye Kim, Qing Liu, Yijun Li, Jianming Zhang, Nanxuan Zhao, Yilin Wang, et al. Unireal: Universal image generation and editing via learning real-world dynamics. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 12501–12511, 2025b.

Zeyuan Chen, Yangchao Wang, Yang Yang, and Dong Liu. Psd: Principled synthetic-to-real dehazing guided by physical priors. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 7180–7189, 2021.

Gheorghe Comanici, Eric Bieber, Mike Schaekermann, Ice Pasupat, Noveen Sachdeva, Inderjit Dhillon, Marcel Blistein, Ori Ram, Dan Zhang, Evan Rosen, et al. Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities. *arXiv preprint arXiv:2507.06261*, 2025.

Yuekun Dai, Chongyi Li, Shangchen Zhou, Ruicheng Feng, and Chen Change Loy. Flare7k: A phenomenological nighttime flare removal dataset. *Advances in Neural Information Processing Systems*, 35:3926–3937, 2022.

Yuekun Dai, Yihang Luo, Shangchen Zhou, Chongyi Li, and Chen Change Loy. Nighttime smartphone reflective flare removal using optical center symmetry prior. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 20783–20791, 2023.

Yuekun Dai, Dafeng Zhang, Xiaoming Li, Zongsheng Yue, Chongyi Li, Shangchen Zhou, Ruicheng Feng, et al. Mipi 2024 challenge on nighttime flare removal: Methods and results. *arXiv preprint arXiv:2404.19534*, 2024.

Wei Dong, Han Zhou, Yuqiong Tian, Jingke Sun, Xiaohong Liu, Guangtao Zhai, and Jun Chen. Shadowrefiner: Towards mask-free shadow removal via fast fourier transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6208–6217, 2024.

Zheng Dong, Ke Xu, Yin Yang, Hujun Bao, Weiwei Xu, and Rynson WH Lau. Location-aware single image reflection removal. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 5017–5026, 2021.

Deniz Engin, Anil Genç, and Hazim Kemal Ekenel. Cycle-dehaze: Enhanced cyclegan for single image dehazing. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 825–833, 2018.

Qingnan Fan, Jiaolong Yang, Gang Hua, Baoquan Chen, and David Wipf. A generic deep architecture for single image reflection removal and image smoothing. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3238–3247, 2017.

Lan Fu, Changqing Zhou, Qing Guo, Felix Juefei-Xu, Hongkai Yu, Wei Feng, Yang Liu, and Song Wang. Auto-exposure fusion for single-image shadow removal. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10571–10580, 2021.

Allabakash Ghodesawar, Vinod Patil, Ankit Raichur, Swaroop Adrashyappanamath, Sampada Malagi, Nikhil Akalwadi, Chaitra Desai, Ramesh Ashok Tabib, Ujwala Patil, and Uma Mudenagudi. Deflare-net: Flare detection and removal network. In *International Conference on Pattern Recognition and Machine Intelligence*, pp. 465–472. Springer, 2023.

Google. Introducing gemini 2.5 flash image, our state-of-the-art image model, 8 2025. URL `https://developers.googleblog.com/en/introducing-gemini-2-5-flash-image/`.

Lanqing Guo, Siyu Huang, Ding Liu, Hao Cheng, and Bihan Wen. Shadowformer: Global context helps shadow removal. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pp. 710–718, 2023a.

Lanqing Guo, Chong Wang, Wenhan Yang, Siyu Huang, Yufei Wang, Hanspeter Pfister, and Bihan Wen. Shadowdiffusion: When degradation prior meets diffusion model for shadow removal. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 14049–14058, 2023b.

Ruiqi Guo, Qieyun Dai, and Derek Hoiem. Single-image shadow detection and removal using paired regions. In *CVPR 2011*, pp. 2033–2040. IEEE, 2011.

Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, 33(12):2341–2353, 2010.

Yuwen He, Wei Wang, Wanyu Wu, and Kui Jiang. Disentangle nighttime lens flares: self-supervised generation-based lens flare removal. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 3464–3472, 2025.

Amir Hertz, Ron Mokady, Jay Tenenbaum, Kfir Aberman, Yael Pritch, and Daniel Cohen-Or. Prompt-to-prompt image editing with cross attention control.(2022). *URL https://arxiv.org/abs/2208.01626*, 3, 2022.

Yuchen Hong, Haofeng Zhong, Shuchen Weng, Jinxiu Liang, and Boxin Shi. L-differ: Single image reflection removal with language-based diffusion model. In *European Conference on Computer Vision*, pp. 58–76. Springer, 2024.

Qiming Hu and Xiaojie Guo. Trash or treasure? an interactive dual-stream strategy for single image reflection separation. *Advances in Neural Information Processing Systems*, 34:24683–24694, 2021.

Qiming Hu and Xiaojie Guo. Single image reflection separation via component synergy. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 13138–13147, 2023.

Yue Huang, Zi'ang Li, Tianle Hu, Jie Wen, Guanbin Li, Jinglin Zhang, Guoxu Zhou, and Xiaozhao Fang. Single image reflection removal via inter-layer complementarity. *arXiv preprint arXiv:2505.12641*, 2025.

Aaron Hurst, Adam Lerer, Adam P Goucher, Adam Perelman, Aditya Ramesh, Aidan Clark, AJ Ostrow, Akila Welihinda, Alan Hayes, Alec Radford, et al. Gpt-4o system card. *arXiv preprint arXiv:2410.21276*, 2024.

Md Tanvir Islam, Nasir Rahim, Saeed Anwar, Muhammad Saqib, Sambit Bakshi, and Khan Muhammad. Hazespace2m: A dataset for haze aware single image dehazing. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pp. 9155–9164, 2024.

Yeying Jin, Aashish Sharma, and Robby T Tan. Dc-shadownet: Single-image hard and soft shadow removal using unsupervised domain-classifier guided network. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 5027–5036, 2021.

Bahjat Kawar, Shiran Zada, Oran Lang, Omer Tov, Huiwen Chang, Tali Dekel, Inbar Mosseri, and Michal Irani. Imagic: Text-based real image editing with diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 6007–6017, 2023.

Bingxin Ke, Kevin Qu, Tianfu Wang, Nando Metzger, Shengyu Huang, Bo Li, Anton Obukhov, and Konrad Schindler. Marigold: Affordable adaptation of diffusion-based image generators for image analysis. *arXiv preprint arXiv:2505.09358*, 2025.

Soomin Kim, Yuchi Huo, and Sung-Eui Yoon. Single image reflection removal with physically-based training images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 5164–5173, 2020.

Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

Black Forest Labs, Stephen Batifol, Andreas Blattmann, Frederic Boesel, Saksham Consul, Cyril Diagne, Tim Dockhorn, Jack English, Zion English, Patrick Esser, et al. Flux. 1 kontext: Flow matching for in-context image generation and editing in latent space. *arXiv preprint arXiv:2506.15742*, 2025.

Hieu Le and Dimitris Samaras. Shadow removal via shadow image decomposition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 8578–8587, 2019.

Chenyang Lei and Qifeng Chen. Robust reflection removal with reflection-free flash-only cues. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14811–14820, 2021.

Chenyang Lei, Xuhua Huang, Mengdi Zhang, Qiong Yan, Wenxiu Sun, and Qifeng Chen. Polarized reflection removal with perfect alignment in the wild. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 1750–1758, 2020.

Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. Aod-net: All-in-one dehazing network. In *Proceedings of the IEEE international conference on computer vision*, pp. 4770–4778, 2017.

Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE transactions on image processing*, 28 (1):492–505, 2018.

Chao Li, Yixiao Yang, Kun He, Stephen Lin, and John E Hopcroft. Single image reflection removal through cascaded refinement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 3565–3574, 2020a.

Ruoteng Li, Robby T Tan, and Loong-Fah Cheong. All in one bad weather removal using architectural search. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 3175–3185, 2020b.

Yu-Fan Lin, Chia-Ming Lee, and Chih-Chung Hsu. Densesr: Image shadow removal as dense prediction. *arXiv preprint arXiv:2507.16472*, 2025.

Zhihao Liu, Hui Yin, Xinyi Wu, Zhenyao Wu, Yang Mi, and Song Wang. From shadow generation to shadow removal. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4927–4936, 2021.

Vaishnav Potlapalli, Syed Waqas Zamir, Salman H Khan, and Fahad Shahbaz Khan. Promptir: Prompting for all-in-one image restoration. *Advances in Neural Information Processing Systems*, 36:71275–71293, 2023.

Liangqiong Qu, Jiandong Tian, Shengfeng He, Yandong Tang, and Rynson WH Lau. Deshadownet: A multi-context embedding deep network for shadow removal. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4067–4075, 2017.

Hamed RahmaniKhezri, Suhong Kim, and Mohamed Hefeeda. Unsupervised single-image reflection removal. *IEEE Transactions on Multimedia*, 25:4958–4971, 2022.

Yujun Shi, Chuhui Xue, Jun Hao Liew, Jiachun Pan, Hanshu Yan, Wenqing Zhang, Vincent YF Tan, and Song Bai. Dragdiffusion: Harnessing diffusion models for interactive point-based image editing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8839–8849, 2024.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020.

Yuda Song, Zhuqing He, Hui Qian, and Xin Du. Vision transformers for single image dehazing. *IEEE Transactions on Image Processing*, 32:1927–1941, 2023.

Roman Suvorov, Elizaveta Logacheva, Anton Mashikhin, Anastasia Remizova, Arsenii Ashukha, Aleksei Silvestrov, Naejin Kong, Harshith Goka, Kiwoong Park, and Victor Lempitsky. Resolution-robust large mask inpainting with fourier convolutions. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pp. 2149–2159, 2022.

Florin Vasluianu, Zongwei Wu, and Radu Timofte. Sfnet-a spatial-frequency domain neural network for image lens flare removal. In *2024 IEEE International Conference on Image Processing (ICIP)*, pp. 1711–1717. IEEE, 2024.

Florin-Alexandru Vasluianu, Tim Seizinger, and Radu Timofte. Wsrd: A novel benchmark for high resolution image shadow removal. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1826–1835, 2023.

Renjie Wan, Boxin Shi, Ling-Yu Duan, Ah-Hwee Tan, and Alex C Kot. Benchmarking single-image reflection removal algorithms. In *Proceedings of the IEEE international conference on computer vision*, pp. 3922–3930, 2017.

Jifeng Wang, Xiang Li, and Jian Yang. Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1788–1797, 2018.

Ruiyi Wang, Yushuo Zheng, Zicheng Zhang, Chunyi Li, Shuaicheng Liu, Guangtao Zhai, and Xiaohong Liu. Learning hazing to dehazing: Towards realistic haze generation for real-world image dehazing. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 23091–23100, 2025.

Tao Wang, Wanglong Lu, Kaihao Zhang, Wenhan Luo, Tae-Kyun Kim, Tong Lu, Hongdong Li, and Ming-Hsuan Yang. Promptrr: Diffusion models as prompt generators for single image reflection removal. *arXiv preprint arXiv:2402.02374*, 2024a.

Yongzhen Wang, Xuefeng Yan, Fu Lee Wang, Haoran Xie, Wenhan Yang, Xiao-Ping Zhang, Jing Qin, and Mingqiang Wei. Ucl-dehaze: Toward real-world image dehazing via unsupervised contrastive learning. *IEEE Transactions on Image Processing*, 33:1361–1374, 2024b.

Qiang Wen, Yinjie Tan, Jing Qin, Wenxi Liu, Guoqiang Han, and Shengfeng He. Single image reflection removal beyond linearity. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3771–3779, 2019.

Yicheng Wu, Qiurui He, Tianfan Xue, Rahul Garg, Jiawen Chen, Ashok Veeraraghavan, and Jonathan T Barron. How to train neural networks for flare removal. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 2239–2247, 2021.

Jiamin Xu, Yuxin Zheng, Zelong Li, Chi Wang, Renshu Gu, Weiwei Xu, and Gang Xu. Detail-preserving latent diffusion for stable shadow removal. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 7592–7602, 2025.

Minglong Xue, Aoxiang Ning, Shivakumara Palaiahnakote, and Mingliang Zhou. Dfdnet: Dynamic frequency-guided de-flare network. *arXiv preprint arXiv:2507.17489*, 2025.

Dong Yang and Jian Sun. Proximal dehaze-net: A prior learning-based deep network for single image dehazing. In *Proceedings of the european conference on computer vision (ECCV)*, pp. 702–717, 2018.

Jie Yang, Dong Gong, Lingqiao Liu, and Qinfeng Shi. Seeing deeply and bidirectionally: A deep learning approach for single image reflection removal. In *Proceedings of the european conference on computer vision (ECCV)*, pp. 654–669, 2018.

Jing Zhang, Yang Cao, Zheng-Jun Zha, and Dacheng Tao. Nighttime dehazing with a synthetic benchmark. In *Proceedings of the 28th ACM international conference on multimedia*, pp. 2355–2363, 2020.

Ling Zhang, Qing Zhang, and Chunxia Xiao. Shadow remover: Image shadow removal based on illumination recovering optimization. *IEEE Transactions on Image Processing*, 24(11):4623–4636, 2015.

Qing Zhang, Yizhong Zhang, Xu Kuang, Yuanbo Zhou, and Tong Tong. Pa-nafnet: An improved nonlinear activation free network with pyramid attention for single image reflection removal. *Digital Signal Processing*, pp. 105474, 2025.

Ruikun Zhang, Hao Yang, Yan Yang, Ying Fu, and Liyuan Pan. Lmhaze: intensity-aware image dehazing with a large-scale multi-intensity real haze dataset. In *Proceedings of the 6th ACM International Conference on Multimedia in Asia*, pp. 1–1, 2024.

Weixia Zhang, Guangtao Zhai, Ying Wei, Xiaokang Yang, and Kede Ma. Blind image quality assessment via vision-language correspondence: A multitask learning perspective. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 14071–14081, 2023.

Xinyi Zhang, Hang Dong, Jinshan Pan, Chao Zhu, Ying Tai, Chengjie Wang, Jilin Li, Feiyue Huang, and Fei Wang. Learning to restore hazy video: A new real-world dataset and a new method. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9239–9248, 2021.

Xuaner Zhang, Ren Ng, and Qifeng Chen. Single image reflection separation with perceptual losses. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4786–4794, 2018.

H Zheng, Z Lin, J Lu, S Cohen, E Shechtman, C Barnes, J Zhang, N Xu, S Amirghodsi, and J Luo. Cm-gan: Image inpainting with cascaded modulation gan and object-aware training. arxiv 2022. *arXiv preprint arXiv:2203.11947*, 2, 2022.

Qian Zheng, Boxin Shi, Jinnan Chen, Xudong Jiang, Ling-Yu Duan, and Alex C Kot. Single image reflection removal with absorption effect. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 13395–13404, 2021.

Tianwen Zhou, Qihao Duan, and Zitong Yu. Difflare: Removing image lens flare with latent diffusion model. *arXiv preprint arXiv:2407.14746*, 2024.

Yuyan Zhou, Dong Liang, Songcan Chen, Sheng-Jun Huang, Shuo Yang, and Chongyi Li. Improving lens flare removal with general-purpose pipeline and multiple light sources recovery. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 12969–12979, 2023.

Yuyan Zhou, Dong Liang, Songcan Chen, and Sheng-Jun Huang. Image lens flare removal using adversarial curve learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025.

Xinshan Zhu, Shuoshi Li, Yongdong Gan, Yun Zhang, and Biao Sun. Multi-stream fusion network with generalized smooth l 1 loss for single image dehazing. *IEEE Transactions on Image Processing*, 30:7620–7635, 2021.

Yurui Zhu, Xueyang Fu, Peng-Tao Jiang, Hao Zhang, Qibin Sun, Jinwei Chen, Zheng-Jun Zha, and Bo Li. Revisiting single image reflection removal in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 25468–25478, 2024.