

Seeing Through the Rain: Resolving High-Frequency Conflicts in Deraining and Super-Resolution via Diffusion Guidance

Wenjie Li¹ Jinglei Shi², Jin Han³, Heng Guo^{1*}, Zhanyu Ma¹

¹Beijing University of Posts and Telecommunications

²Nankai University

³The University of Tokyo

{cswjli, guoheng, mazhanyu}@bupt.edu.cn

Abstract

Clean images are crucial for visual tasks such as small object detection, especially at high resolutions. However, real-world images are often degraded by adverse weather, and weather restoration methods may sacrifice high-frequency details critical for analyzing small objects. A natural solution is to apply super-resolution (SR) after weather removal to recover both clarity and fine structures. However, simply cascading restoration and SR struggle to bridge their inherent conflict: removal aims to remove high-frequency weather-induced noise, while SR aims to hallucinate high-frequency textures from existing details, leading to inconsistent restoration contents. In this paper, we take deraining as a case study and propose DHGM, a Diffusion-based High-frequency Guided Model for generating clean and high-resolution images. DHGM integrates pre-trained diffusion priors with high-pass filters to simultaneously remove rain artifacts and enhance structural details. Extensive experiments demonstrate that DHGM achieves superior performance over existing methods, with lower costs.

Code — <https://github.com/PRIS-CV/DHGM>

Introduction

Adverse weather commonly degrades visual quality, impacting downstream tasks such as object detection (Varghese and Sambath 2024). To mitigate these degradations, weather restoration is employed to remove weather-induced noise and recover clean images. However, as pointed out in recent studies (Yang et al. 2017; Jin, Chen, and Li 2020), existing methods inevitably sacrifice high-frequency details, resulting in excessive smoothing. This observation is also confirmed by our analysis at the top of Fig. 1, showing that high-frequency parts in weather removal images are attenuated compared to ground truth (GT). This indicates that existing methods remove not only weather noise but also high-frequency textures. Such an issue severely impairs detection of small targets, such as distant vehicles, since even at 2K resolutions, these objects are low-resolution, typically occupy only a few dozen pixels, making accurate detection particularly sensitive to texture or edge loss introduced by weather removal.

To preserve and reconstruct essential high-frequency textures for reliable downstream detection, an intuitive solution

*Corresponding Author.

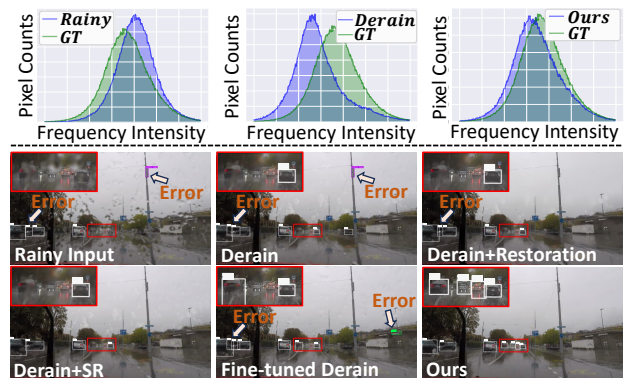


Figure 1: **(Top)** Frequency-domain analysis shows that weather removal methods eliminate not only rain streaks but also valuable high-frequency textures. **(Bottom)** Compared with existing methods, our method better preserves details and improves small-object detection under rainy conditions.

is to apply super-resolution (SR) (Zhou et al. 2023) after weather removal. Unlike general image restoration methods, which primarily focus on noise removal or texture enhancement without explicitly upsampling images, SR methods directly upscale image resolution, thus effectively enlarging small targets to reduce hallucination, as shown in Fig. 1. However, simply cascading weather removal and SR methods fails to address their inherent conflicts: weather removal aims to suppress high-frequency noise, whereas SR attempts to infer high-frequency details from existing textures. Errors introduced in weather removal propagate through SR, resulting in amplified artifacts and inconsistent texture reconstruction. Similarly, fine-tuning deraining models (Sun et al. 2024) on paired low-resolution (LR) rainy and high-resolution (HR) clean images faces similar issues, as these models inherently struggle to balance high-frequency noise removal and texture recovery. Consequently, there remains a critical need for a method that can simultaneously achieve effective weather noise suppression and accurate SR texture restoration. *In this paper, we focus on image deraining as a representative example to address this challenging scenario.*

Inspired by guided filters (He, Sun, and Tang 2012) and high-pass filters (Khan et al. 2016), we try to utilize these

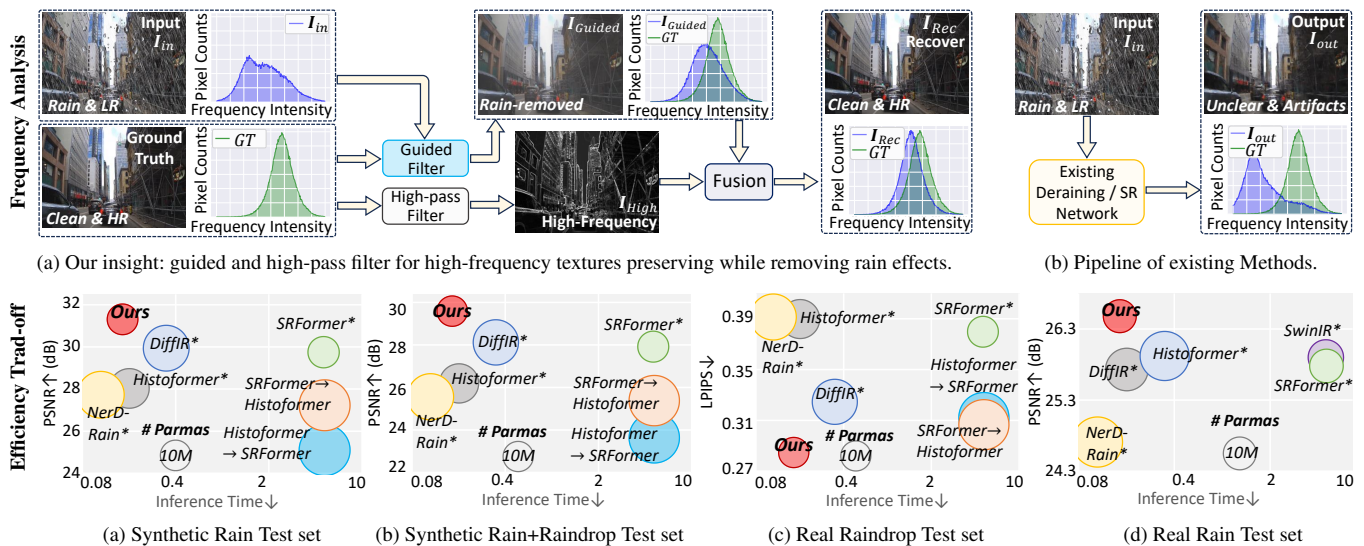


Figure 2: **(Top)** Given a clean and HR image, guided filters can remove high-frequency noises while preserving most high-frequency textures, and high-pass filters further enhance blurred high-frequency edges. **(Bottom)** Our method achieves the best performance while requiring less cost in terms of speed and model size. (* denotes results of fine-tuning on our training dataset.)

to reconstruct rainy LR images. Specifically, as shown in Fig. 2, with the help of clean and HR images, we observe guided filter (He, Sun, and Tang 2012) can smooth out messy high-frequency noise (*e.g.*, rain and raindrop) while preserving image edge (*e.g.*, high-frequency textures of LR images), thereby aligning the frequency distribution is close to the ground truth (GT). However, high-frequency texture details from restored images remain missing, as shown in the spatial results and frequency distribution. Therefore, a high-pass filter can be further applied to clean HR images to compensate for high-frequency texture features, leading to rain-free and HR outputs. Compared to existing deraining and SR methods, this strategy achieves cleaner and clearer results. Therefore, combining priors with guided and high-pass filters can be a promising solution for SR in rainy weather to address the balance in high-frequency reconstruction.

However, clean and HR images are required by guided and high-pass filters. To achieve this chicken-egg problem, we try to learn content priors close to GT distributions. Specifically, we propose a Diffusion-based High-frequency Guided Model (DHGM) with two phases. In the first phase, we employ encoders to compress contents reflecting true distributions into latent spaces as priors. To leverage latent priors for rain removal and texture reconstruction, we propose a Media Remover (MR) based on guided filters and a Texture Compensator (TC) based on high-pass filters. Recognizing the potential of diffusion models (Ho, Jain, and Abbeel 2020) (DM) to achieve high-quality mappings from randomly sampled Gaussian noise to latent distributions (Rombach et al. 2022; Xia et al. 2023), we proceed to the second phase by freezing encoder weights and using DM to learn content distributions within pre-trained latent priors. Simultaneously, we fine-tune our MR and TC from the first phase, training them alongside DM to reconstruct images jointly. As observed at

the bottom of Fig. 1 and Fig. 2, our method can reconstruct clean and HR images from rainy LR images while maintaining efficiency on speed and model size, further improving the accuracy of downstream tasks.

To summarize, our contributions are as follows:

- We focus on deraining as an example, propose DHGM that explores the challenge of recovering clean and HR images from potential LR images captured in rainy conditions;
- We propose MR and TC modules based on guided filter and high-pass filter, which direct pre-trained diffusion priors to remove rain-induced noise and recover textures;
- Experiments on extensive datasets show our method can recover clean and HR images for downstream tasks while requiring less computational cost than existing methods.

Related Work

Restoration in Rainy Weather

Image Deraining. To handle rainy images that obstruct the view and are not conducive to downstream tasks (Peng et al. 2024), a series of specially designed networks (Peng et al. 2025b) employ strategies like multi-branch (Jiang et al. 2020), multi-scale (Chen et al. 2023), and multi-stage (Wang et al. 2020) to achieve end-to-end deraining. Uddin *et al.* (Uddin 2022) focuses on SR and adjusts rainy tones rather than joint deraining and SR. Subsequently, to enhance the global representation of models, IDT (Xiao et al. 2022) proposes a window-based Transformer, while NeRD-Rain (Chen, Pan, and Dong 2024) combines multi-scale implicit neural representations. Unlike previous studies, we focus on joint deraining and SR that may occur in rainy weather.

All-in-One Weather Restoration. Recent attempts have been made to unify complex weather recovery efforts into one network. The all-in-one restoration network (Li,

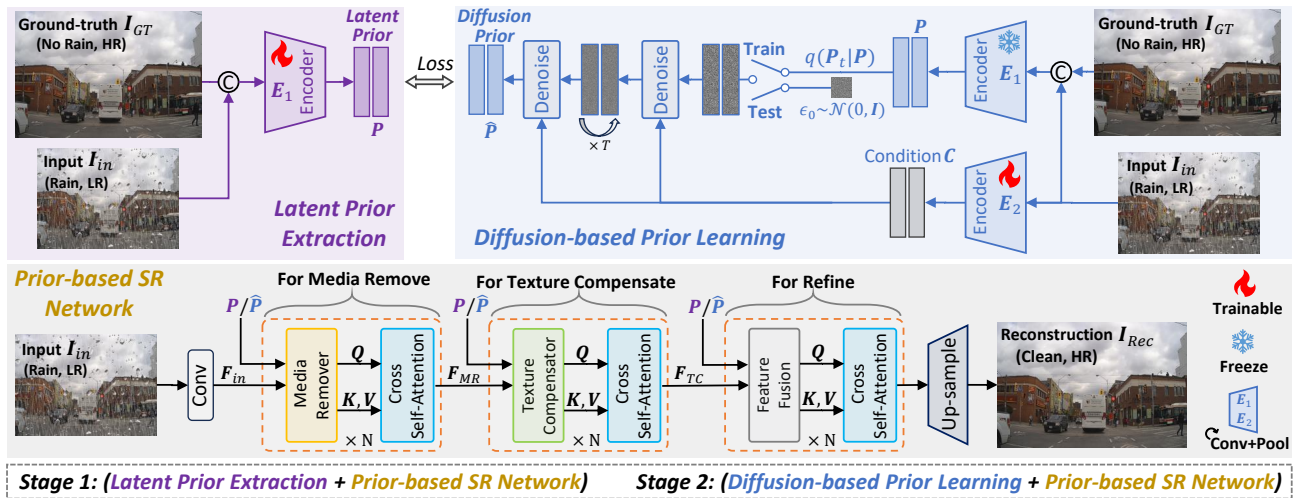


Figure 3: Overview of our method, which utilizes our Media Remover (MR) and Texture Compensator (TC) to guide learned diffusion priors in latent spaces to complete high-frequency rain-induced media removal and high-frequency texture reconstruction.

Tan, and Cheong 2020) is the first try with multiple task-specific encoders and a shared decoder. TransWeather (Valanarasu, Yasarla, and Patel 2022) improves performance in various rainy conditions through Transformer-based encoder-decoders. WeatherStream (Zhang et al. 2023), DTMWR (Patil et al. 2023), and WGWS-Net (Zhu et al. 2023) improve existing models by learning weather-specific degradation data. WeatherDiff (Özdenizci and Legenstein 2023) proposes patch-based denoising diffusion models (Ho, Jain, and Abbeel 2020) to achieve size-agnostic restoration. OneRestore (Guo et al. 2024) proposes a versatile imaging model to simulate possible weather degradation in the environment. Additionally, recent studies also address real data scarcity, fidelity, model lightweight, or modal fusion through techniques such as adaptive filters (Park, Lee, and Chun 2023), codebooks (Ye et al. 2023), knowledge distillation (Chen et al. 2022), and pre-trained language models (Tan et al. 2024). However, texture loss and edge distortion caused by direct weather restoration may seriously affect the contours of small targets at long distances, which is detrimental to downstream tasks, especially including small objects.

Image Super-resolution

Numerous SR algorithms (Li et al. 2023b) have been developed to improve image resolution. Specifically, SRCNN (Dong et al. 2015) first uses a 3-layer convolutional neural network for SR. RCAN (Zhang et al. 2018b), SAN (Dai et al. 2019), and NLSA (Mei, Fan, and Zhou 2021) enhance SR performance by introducing attention mechanisms. WDSR (Yu et al. 2018) and FDIWN (Gao et al. 2022) reduce the feature loss of SR processes caused by activation functions. With the development of ViT (Yuan et al. 2021), IPT (Chen et al. 2021) tries to improve SR by utilizing the Transformer. Then, SwinIR (Liang et al. 2021), FIWHN (Li et al. 2024b), and SRFormer (Zhou et al. 2023) utilize the Windows-based strategy to reduce the enormous costs caused by the Transformer. OmniSR (Wang et al. 2023), ATD (Zhang

et al. 2024), and DMNet (Li et al. 2025a) expand the receptive fields of self-attention in the Transformer to improve SR further. These methods promote the advancement of SR, but they default to images without the interference of weathers.

Methods

As shown in Fig. 3, our method with two training phases consists of three components: latent Prior Extraction, Diffusion-based Prior Learning, and Prior-based SR Network. In the first phase, we jointly train the Latent Prior Extraction module for extracting latent priors P and the Prior-based SR Network for utilizing P . In the second phase, we train the Diffusion-based Prior Learning module for learning diffusion priors \hat{P} from P while fine-tuning the pre-trained Prior-based SR Network for improved reconstruction. After that, rainy LR inputs $I_{in} \in \mathbb{R}^{H \times W \times 3}$ can be recovered to clean and HR outputs $I_{Rec} \in \mathbb{R}^{sH \times sW \times 3}$, s is a scale factor.

Pre-training Priors Learning (Stage I)

In the first stage of training, as shown in Fig. 3, we focus on jointly training the latent prior extraction module and prior-based SR Network to obtain a compact representation P from a hybrid of inputs and ground truth, serving as latent priors. Mixing inputs with ground truth mitigates the distribution gap, preventing excessive divergence that could degrade network performance. For inputs I_{in} , our components remove rain media, reconstruct edge textures, and refine outputs to get final results I_{Rec} via up-sampling.

Media Remover (MR). As shown in Fig. 4 (a), we incorporate our GFM into MR for rain-induced media removal and incorporate cross-attention (Rombach et al. 2022) for feature fusion. Specifically, for input priors $P \in \mathbb{R}^{4C}$ and input rainy LR features $F_{in} \in \mathbb{R}^{H \times W \times C}$, we first embed P to obtain a set of vectors $\{\alpha_0, \alpha_1 \dots \alpha_n\} \in \mathbb{R}^{C \times 1 \times 1}$. To obtain different interaction patterns of priors with F_{in} , we fuse prior vectors with features using multiplication and addition,

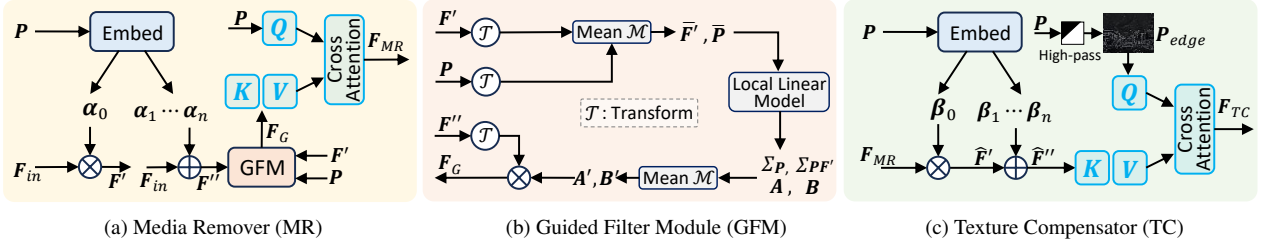


Figure 4: Detailed structure of (a) Media Remover, (b) Guided Filter Module, and (c) Texture Compensator.

respectively. After that, we obtain two coarse fusion features $F', F'' \in \mathbb{R}^{H \times W \times C}$ with different patterns. Then, F', F'' are fed into our GFM \mathcal{G} together with priors P to obtain outputs $F_G \in \mathbb{R}^{H \times W \times C}$ without rain-induced media:

$$F', F'' = \alpha_0 \times F_{in}, \{\alpha_1 \dots \alpha_n\} + F_{in}, \quad (1)$$

$$F_G = \mathcal{G}(F', F'', P). \quad (2)$$

Next, inspired by the performance of cross-attention (Romach et al. 2022; Li et al. 2023a) for feature fusion, we introduce a cross-attention to fuse P and F_G to find feature similarity within both. Specifically, we first embed F_G and P to project into vectors $\{Q, K, V\} \in \mathbb{R}^{H \times W \times C}$:

$$Q, K, V = \mathcal{W}_Q F_G \times P, \mathcal{W}_K F_G, \mathcal{W}_V F_G, \quad (3)$$

where $\mathcal{W}_Q, \mathcal{W}_K$, and \mathcal{W}_V are convolution operations. Next, we utilize cross-attention to achieve information fusion and explore the relationship between P and F_G :

$$\text{CrossAttention}(Q, K, V) = V \text{Softmax}(QK^T / \gamma), \quad (4)$$

where γ is a learnable factor. Finally, we reconstruct features F_{MR} without rain-induced media using cross-attention. Details of GFM responsible for guiding prior removal of rain-related media are described in the following paragraph.

Guided Filter Module (GFM). Inspired by the concept of guided filter (He, Sun, and Tang 2012), we propose a GFM to bootstrap priors for removing rain-induced media while preserving structural details. As shown in Fig. 4 (b), the process begins with two coarse features F' and F'' , modulated by priors P . First, a transformation function \mathcal{T} is applied to F' and P , followed by a mean filtering \mathcal{M} with a radius r , producing the smoothed representations \bar{F}' and \bar{P} :

$$\bar{F}', \bar{P} = \mathcal{M}(\mathcal{T}(F'), r), \mathcal{M}(\mathcal{T}(P), r). \quad (5)$$

This step captures local dependencies between features and priors, enabling the network to model interactions across different spatial scales. Next, we refine the interaction between input features F' and priors P by computing their filtered correlation terms, PF' and P^2 , which help to characterize the relationship between features and priors, providing an accurate representation of low-frequency structures:

$$\overline{PF'}, \overline{P^2} = \mathcal{M}(F' \cdot P, r), \mathcal{M}(P \cdot P, r). \quad (6)$$

We then calculate coefficients A and B , which quantify priors' influence on inputs and low-frequency details, respectively, helping accurately extract background details:

$$\sum P, \sum PF' = \overline{P^2} - \bar{P} \cdot \bar{P}, \overline{PF'} - \bar{P} \cdot \bar{F}', \quad (7)$$

$$A = \frac{\sum PF'}{(\sum P + \epsilon)}, B = \bar{F}' - A \cdot \bar{P}. \quad (8)$$

To ensure consistency in smoothing, we apply mean filtering, yielding learned guidance coefficients \bar{A} and \bar{B} . They are then fused with F'' through element-wise modulation to generate guided features F_G , which is free from rain:

$$\bar{A} = \mathcal{M}(A), \bar{B} = \mathcal{M}(B), \quad (9)$$

$$F_G = \bar{A} \cdot F'' + \bar{B}. \quad (10)$$

Through the guide of \bar{A} , which reflects priors' influence, and \bar{B} , which reflects background details, GFM removes rain-induced media while preserving critical structural details.

Texture Compensator (TC). As shown in Fig.4 (c), we propose a TC module to refine high-frequency parts of features after rain media removal. Inputs consist of features $F_{MR} \in \mathbb{R}^{H \times W \times C}$ from the MR module, which has undergone rain-related media removal, along with priors $P \in \mathbb{R}^{4C}$. We begin by embedding prior P into a set of sub-priors $\{\beta_0, \beta_1 \dots \beta_n\} \in \mathbb{R}^{C \times 1 \times 1}$. These sub-priors are then fused with F_{MR} to obtain coarse features \hat{F}'' :

$$\hat{F}'' = \beta_0 \times F_{MR} + \{\beta_1 \dots \beta_n\} + F_{MR}. \quad (11)$$

To inject missed textures, we decide to apply a high-pass filter for extracting high-frequency parts from priors P . This is achieved by performing a 1D Discrete Cosine (Khayam 2003) Transform (DCT) along channels, transforming P from the spatial domain to the frequency domain. After this, we perform an inverse transform (IDCT) to recover spatial domain features, thus extracting edge features P_{edge} :

$$X_k = \text{DCT}(P) = \sum_{n=0}^{N-1} P_n \cdot \cos\left(\frac{\pi}{N} \left(n + \frac{1}{2}\right) k\right), \quad (12)$$

$$X_h = \begin{cases} 0, & k < k_{cutoff} \\ X_k, & k \geq k_{cutoff}, \end{cases} \quad (13)$$

$$P_{edge} = \text{IDCT}(X_h) = \sum_{k=0}^{N-1} X_h^k \cdot \cos\left(\frac{\pi}{N} \left(n + \frac{1}{2}\right) k\right), \quad (14)$$

where P_n is n -th element of P , N is channel counts, $k = 0, 1, \dots, N - 1$, X_k represents DCT coefficients, X_h represents high-frequency components of P , and k_{cutoff} is the cut-off frequency. To get matched dimensions, we reshape P_{edge} to $\mathbb{R}^{1 \times 1 \times C}$ and feed it with coarse features \hat{F}'' into

Methods	Params	Speed	RainDS-Syn-Rain			RainDS-Syn-RD-Rain			Raindrop		
			PSNR/SSIM	LPIPS↓	DISTS↓	PSNR/SSIM	LPIPS↓	DISTS↓	PSNR/SSIM	LPIPS↓	DISTS↓
Bicubic (LR+Rainy)	-	-	22.25/0.6525	0.4210	0.2513	19.41/0.5571	0.4892	0.2698	22.69/0.7027	0.3674	0.2002
Histoformer→SwinIR	28.5M	5.29s	25.42/0.7404	0.4305	0.2523	23.62/0.6804	0.4749	0.2829	24.52/0.7355	0.3130	0.1424
Histoformer→SRFormer	27.1M	5.24s	25.10/0.7210	0.4311	0.2525	23.63/0.6804	0.4753	0.2830	24.53/0.7364	0.3122	0.1422
SwinIR→Histoformer	28.5M	5.29s	27.20/0.7998	0.3195	0.1783	25.35/0.7553	0.3762	0.2125	24.40/0.7359	0.3066	<u>0.1385</u>
SRFormer→Histoformer	27.1M	5.24s	27.23/0.8005	0.3190	0.1786	25.38/0.7563	0.3758	0.2122	24.45/0.7374	<u>0.3047</u>	0.1387
Fine-tuned Histoformer	16.6M	0.18s	28.03/0.8274	0.2818	0.1584	26.19/0.7883	0.3373	0.1913	24.26/0.7064	0.3898	0.1982
Fine-tuned NeRD-Rain	22.9M	0.11s	27.73/0.8206	0.2962	0.1642	25.56/0.7705	0.3642	0.1945	22.66/0.6825	0.3914	0.2016
Fine-tuned SRFormer	<u>10.5M</u>	5.15s	29.75/0.8751	0.2658	0.1031	27.92/0.8444	0.2465	<u>0.1222</u>	24.45/0.7314	0.3798	0.1905
Fine-tuned DiffIR	22.0M	0.34s	<u>29.94/0.8795</u>	<u>0.1950</u>	<u>0.1000</u>	<u>28.12/0.8517</u>	<u>0.2375</u>	0.1225	<u>25.42/0.7401</u>	0.3246	0.1597
Ours	10.1M	0.16s	31.28/0.9067	0.1575	0.0787	29.63/0.8863	0.1916	0.0959	26.21/0.7709	0.2862	0.1378

Table 1: Quantitative results of ours with sequentially performed deraining and SR methods, fine-tuned single deraining, all-in-one weather restoration, SR, and general restoration methods on deraining, deraining & raindrop removal, and raindrop removal test sets at scale of $\times 2$ (with the resolution of 720×480). Best and second-best results are emphasized in **bold** and underlined.

a cross-attention to further explore high-frequency components and enhance edges (Li et al. 2024a). This enables the reconstruction of sharp edges in outputs:

$$\mathbf{Q}, \mathbf{K}, \mathbf{V} = \mathcal{W}_Q \hat{\mathbf{F}}'' \times \mathbf{P}_{edge}, \mathcal{W}_K \hat{\mathbf{F}}'', \mathcal{W}_V \hat{\mathbf{F}}'', \quad (15)$$

$$\text{CrossAttention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \mathbf{V} \text{Softmax}(\mathbf{Q}\mathbf{K}^T/\gamma). \quad (16)$$

Through these operators, we compensate for edges missed in inputs. Next, as shown in Fig.3, we use product to fuse \mathbf{F}_{TC} with priors \mathbf{P} , followed by cross-attention for feature refinement. Final results, \mathbf{I}_{rec} are obtained through up-sampling, producing texture-compensated and rain-free images.

Prior-based Guided Restoration (Stage II)

In the second stage, as shown in Fig. 3, we freeze pre-trained encoder \mathbf{E}_1 weights and output priors $\hat{\mathbf{P}}$ extracted from GT. We hope to accurately estimate $\hat{\mathbf{P}}$ close to \mathbf{P} without GT for fine-tuning our prior-based SR network. Inspired by the ability of DDIM (Ho, Jain, and Abbeel 2020) to generate high-quality images from random noises, surpassing StyleGAN (Karras, Laine, and Aila 2019) and VQGAN (Esser, Rombach, and Ommer 2021), we adopt it for prior learning. Furthermore, instead of conventional DDIM, we follow the latent-space diffusion denoising approach (Xia et al. 2023), reducing result randomness and avoiding the inefficiency of pixel-wise generation (Li et al. 2025b; Peng et al. 2025a).

Diffusion and Denoising Process. As shown in Fig. 3, for the diffusion process, we obtain latent prior \mathbf{P} from froze encoder \mathbf{E}_1 , and progressively add Gaussian noise on \mathbf{P} :

$$q(\mathbf{P}_t | \mathbf{P}) = \mathcal{N}(\mathbf{P}_t; \sqrt{\bar{\alpha}_t} \mathbf{P}, (1 - \bar{\alpha}_t) \mathbf{I}), \quad (17)$$

where \mathbf{P}_t is a noised prior at time-step t , \mathcal{N} is a Gaussian distribution, $\alpha_t = 1 - \beta_t$, $\bar{\alpha}_t = \prod_{i=0}^t \alpha_i$, β_t is a scale factor to control the variance of noises, and \mathbf{I} is an identity matrix. For the denoising process, following the Markov chain, the reverse process from \mathbf{P}_t to \mathbf{P}_{t-1} can be formulated as:

$$p(\mathbf{P}_{t-1} | \mathbf{P}_t, \mathbf{P}_0) = \mathcal{N}(\mathbf{P}_{t-1}; \boldsymbol{\mu}_t(\mathbf{P}_t, \mathbf{P}_0), \sigma_t^2 \mathbf{I}), \quad (18)$$

$$\boldsymbol{\mu}_t(\mathbf{P}_t, \mathbf{P}_0) = \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{P}_t - \epsilon \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}} \right), \sigma_t^2 = \frac{1 - \bar{\alpha}_{t-1}}{1 - \bar{\alpha}_t} \beta_t, \quad (19)$$

where ϵ is noises in \mathbf{P}_t . In our denoising phase, we use encoder \mathbf{E}_2 to encode input images \mathbf{I}_{in} to output conditional features \mathbf{C} to control the range of noise predicted by the denoising network and denoise \mathbf{P}_t stepwise:

$$\mathbf{P}_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{P}_t - \frac{(1 - \alpha_t) \epsilon_\theta}{\sqrt{1 - \bar{\alpha}_t}} (\mathbf{P}_t, \mathbf{C}, t) \right) + \sqrt{1 - \alpha_t} \epsilon_t, \quad (20)$$

where ϵ_t is estimated noise ϵ of each step, $\epsilon_t \sim \mathcal{N}(0, \mathbf{I})$. With T iterations of the above sampling, predicted priors $\hat{\mathbf{P}}$ can be generated to fine-tune our restoration network.

Inference. Without access to the ground truth during inference, we follow DDIM (Song, Meng, and Ermon 2020) and randomly sample Gaussian noise $\epsilon_0 \sim \mathcal{N}(0, \mathbf{I})$ as noised inputs. The noises, along with the condition \mathbf{C} extracted from \mathbf{I}_{in} , are fed into the denoising process. After T iterations, we get generated $\hat{\mathbf{P}}$, which serves as guidance for our prior-based SR network, enabling plausible inference without requiring the ground truth.

Loss Function

Stage I. We jointly train the latent prior extraction module and prior-based SR network. Our training loss \mathcal{L}_{S1} is:

$$\mathcal{L}_{S1} = \|\mathbf{I}_{Rec} - \mathbf{I}_{GT}\|_1, \quad (21)$$

where \mathbf{I}_{Rec} is recovered image, \mathbf{I}_{GT} is the ground truth.

Stage II. Following previous works (Rombach et al. 2022; Xia et al. 2023), we conduct denoising in latent spaces. Unlike the time-consuming mode of traditional DM, which denoises full images. This strategy allows DM to run denoising iterations to obtain denoising results, which are then sent to prior-based SR networks for joint training. \mathcal{L}_{S2} is:

$$\mathcal{L}_{S2} = \|\mathbf{I}_{Rec} - \mathbf{I}_{GT}\|_1 + \left\| \hat{\mathbf{P}} - \mathbf{P} \right\|_1, \quad (22)$$

where \mathbf{P} is prior extracted by encoder \mathbf{E}_1 in the first stage, $\hat{\mathbf{P}}$ is prior estimated by our diffusion model.

Experiments

Datasets and Evaluation Metrics

We train our model on datasets containing multiple rainy conditions, which consider raindrops on the camera sensor,

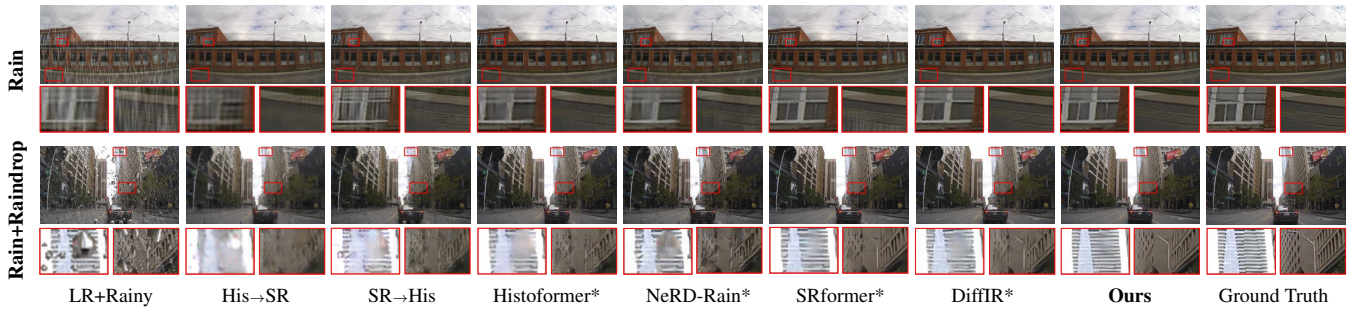


Figure 5: Qualitative comparisons with existing methods on synthesized test sets at the scale of $\times 2$. *: Method after fine-tuning.

heavy rain, and heavy rain with raindrops, respectively. RainDrop (Qian et al. 2018) consists of 861 training images and 307 test images with on-camera raindrops. RainDS (Quan et al. 2021) includes synthetic parts and real parts, which the synthetic part and real part including 3000 training images and 600 test images with rain, on-camera raindrops, and rain with raindrops, and 450 training images and 294 test images with rain, on-camera raindrops, and rain with raindrops, respectively. For evaluation metrics, we calculate PSNR (Wang and Bovik 2002) and SSIM (Wang and Bovik 2002), LPIPS (Zhang et al. 2018a), DISTS (Ding et al. 2020), and NIQE (Mittal, Soundararajan, and Bovik 2012).

Implementation Details

We implement all experiments in the Pytorch framework with one NVIDIA RTX4090 GPU. During training, we set the batch size to 8, the learning rate to 2×10^{-4} , and the patch size to 64×64 . We use Adam optimizer with $\beta_1=0.9$, $\beta_2=0.99$ to train 500k iterations. Our model sets channel counts to 64, N to 12, and time-steps to 4 in our diffusion model. *Since directly performing deraining and SR on original rainy images is impractical due to the absence of corresponding clean HR ground truth, for experimental validation, we construct paired data rainy LR by downsampling, including both synthetic and real adverse weather.* For fine-tuned SR (Liang et al. 2021; Zhou et al. 2023), restoration (Xia et al. 2023), deraining (Valanarasu, Yasarla, and Patel 2022; Sun et al. 2024) methods, we load official pre-trained weights and fine-tune 500k iterations with same hyperparameters in their paper on our training sets. For downstream tasks, we utilize YOLOv8 (Varghese and Sambath 2024) for object detection.

Comparison of Deraining under LR Scenes

We select fine-tuned deraining method (e.g. NerD-Rain (Chen, Pan, and Dong 2024)), all-in-one weather restoration method (e.g. Histoformer (Sun et al. 2024)), fine-tuned SR methods (e.g. SwinIR (Liang et al. 2021), SRFormer (Zhou et al. 2023)), and fine-tuned general image restoration methods (e.g. DiffIR (Xia et al. 2023)). Additionally, we alternate the order of deraining and SR methods for fair comparison. For fine-tuned deraining and restoration methods without up-sampling modules, we first up-sample inputs to match the size of the ground truth before passing them to networks.

Comparison on Rainy Datasets. As present in Table 1, fine-tuned and alternating methods consistently lag behind

Methods	RainDS-(RD+Rain)		RainDS-(Rain)	
	PSNR/SSIM	NIQE↓	PSNR/SSIM	NIQE↓
Fine-tuned SwinIR	22.84/0.6061	7.3883	25.90/0.6796	6.9523
Fine-tuned SRFormer	22.82/0.6040	5.1953	25.78/0.6758	4.9914
Fine-tuned NeRD-Rain	20.54/0.5764	6.3025	24.71/0.6620	5.8472
Fine-tuned Histoformer	22.77/0.6173	5.1382	25.73/0.6923	5.7635
Fine-tuned DiffIR	22.83/0.6256	7.3596	25.92/0.7107	7.0178
Ours	23.21/0.6522	4.9628	26.48/0.7273	4.9401

Table 2: Quantitative comparison at a scale of $\times 2$ (with the resolution of 1296×728), where derain & raindrop removal and derain evaluations are shown on the left and right sides.

our method across all metrics, including visual perception and structural metrics. Furthermore, our method show efficiency, with fewer Params and faster inference, especially compared to alternating methods. Besides, as shown in Fig. 5, visual comparisons reveal that existing methods fail to remove rain-induced media or introduce artifacts. In contrast, our method effectively handles this scene, producing high-quality results.

In Table 1 and Table 2, we conduct experiments on real rainy test sets (LR is synthetic), including deraining, removing raindrops, and deraining & removing raindrops. Since fine-tuned methods generally outperform alternating methods. In Table 1, we only show fine-tuned methods for simplicity. Our method achieves superior results in evaluations.

Extension to More Weather Conditions. As shown in Fig. 6, we further validate the ability of our method to handle potential LR scenes in more adverse weather, including snow and rain with haze weather. Visual comparisons show that our method can handle multiple weather LR conditions, and reconstructed images have sharper textures and cleaner backgrounds, with the potential to be extended to more weather.

Ablation Study

We focus on two aspects in ablations: (i) whether priors are needed and how to learn priors. (ii) Are guided filters, high-pass filters, and cross-attention in our Media Remover (MR) and Texture Compensator (TC) effective?

Analysis on Prior Learning. As shown in Table 3, we show the importance of prior for restoration and the approach of prior learning. First, without the support of priors, reconstruction accuracy drops significantly, resulting in a PSNR loss of approximately 0.64 dB. Secondly, our strategy of

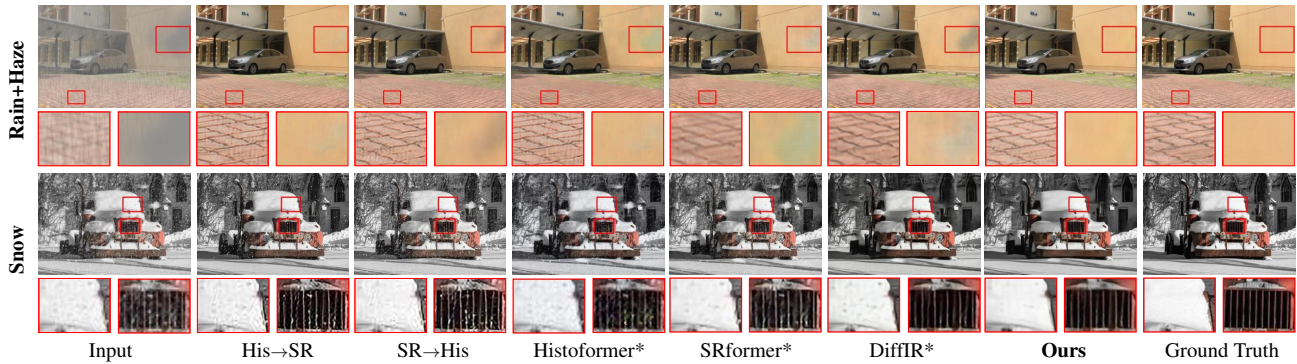


Figure 6: We show our method’s capability to resolve possible LR images under more weather, like rain & haze or snow conditions. See *supplementary material* for results of more weather condition and downstream tasks.

Params	Prior	Diffusion	Diffusion Space		Raindrop	
			Latent Space	Feature Maps	PSNR↑	SSIM↑
9.8M	✗	✗	✗	✗	25.41	0.7397
9.7M	✓	✗	✗	✗	25.63	0.7587
10.1M	✓	✓	✗	✓	24.73	0.7459
10.1M	✓	✓	✓	✗	26.05	0.7674

Table 3: Ablation on prior learning, where “Diffusion Space” indicates whether diffusion models learn priors in the latent space or directly from feature maps. Gray cells indicate ours.

Params	Cross-attn	Guided Filter	High-pass Filter	Raindrop	
				PSNR↑	SSIM↑
8.52M	✗	✓	✓	26.73	0.7688
8.10M	✓	✗	✓	26.48	0.7650
8.10M	✓	✓	✗	26.51	0.7660
7.26M	✓	✗	✗	26.11	0.7617
8.94M	✓	✓	✓	26.82	0.7701

Table 4: Ablation on different modules in our method.

using diffusion to learn priors outperforms CNN-based encoders for prior learning by 0.42 dB in PSNR. Finally, unlike traditional diffusion models that estimate feature maps or full images, our method estimates vectors in a one-dimensional latent space. It enables the joint training of diffusion models for prior estimation alongside the prior-based SR network, leading to more accurate results with a PSNR gain of over 1.3 dB compared to diffusion models that estimate full images.

Effectiveness of Filters in Our Method. As shown in Table 4, we analyze the effect of guided filter (*e.g.*, MR), high-pass filter (*e.g.*, TC), and cross-attention in our method. When priors are selected using either guided or high-pass filters, PSNR increases by an average of 0.33 dB with only a small parameter increase, while removing both results in a performance drop of over 0.7 dB. Cross-attention facilitates fusing post-filter features with input features, bringing a PSNR gain of 0.11 dB. As shown in Fig.7, we further show the role of different modules: guided filters guide priors to remove rain-related media, and high-pass filters help recover sharp textures from high-frequency priors. Fig.8 also shows that guided filters aid raindrop removal and high-pass filters

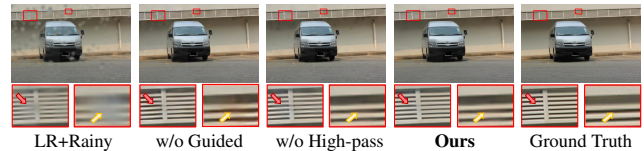


Figure 7: Effects of guided and high-pass filters. Yellow and red arrows indicate artifacts from raindrop removal and blurry edges. Without guided filters, raindrop removal appears as artifacts, but edges remain sharp. Without high-pass filters, edges are blurred, but raindrop removal is relatively clean.

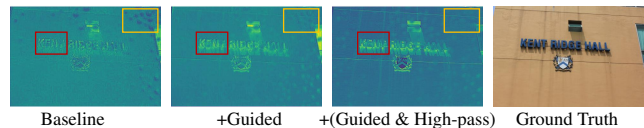


Figure 8: Feature visualization across different modules shows that guided and high-pass filters still play a role in rain media removal and edge recovery in the feature domain. Yellow and red boxes highlight raindrops and edges.

aid texture recovery. Combining these visualizations with the results in Table 4, we conclude that these effects appear not only in the pixel domain but also in the feature domain.

Conclusion

We introduce DHGM, a diffusion-based high-frequency guided model, which can effectively compensate for the loss of fine textures caused by rain removal and recover potential LR objects under rainy conditions through a unified framework of joint deraining and SR. By fully leveraging pre-trained latent diffusion priors together with guided and high-pass filters, DHGM simultaneously removes complex weather-reduced noise and restores missing high-frequency details. Comprehensive experiments on multiple deraining benchmarks demonstrate that our approach can produce cleaner and HR images than existing deraining, restoration, or cascaded SR pipelines, and significantly improves the perception and detection accuracy of small objects in downstream tasks with less computational cost.

Acknowledgments

This work was supported by Beijing Municipal Science & Technology Program No. Z251100007125021, Hebei Natural Science Foundation Project No. 242Q0101Z, Beijing-Tianjin-Hebei Basic Research Funding Program No. F2024502017, National Natural Science Foundation of China (Grant No. 62472044, U24B20155, U23B2052).

References

- Chen, H.; Wang, Y.; Guo, T.; Xu, C.; Deng, Y.; Liu, Z.; Ma, S.; Xu, C.; Xu, C.; and Gao, W. 2021. Pre-trained image processing transformer. In *CVPR*, 12299–12310.
- Chen, S.; Ye, T.; Bai, J.; Chen, E.; Shi, J.; and Zhu, L. 2023. Sparse sampling transformer with uncertainty-driven ranking for unified removal of raindrops and rain streaks. In *ICCV*, 13106–13117.
- Chen, W.-T.; Huang, Z.-K.; Tsai, C.-C.; Yang, H.-H.; Ding, J.-J.; and Kuo, S.-Y. 2022. Learning multiple adverse weather removal via two-stage knowledge learning and multi-contrastive regularization: Toward a unified model. In *CVPR*, 17653–17662.
- Chen, X.; Pan, J.; and Dong, J. 2024. Bidirectional multi-scale implicit neural representations for image deraining. In *CVPR*, 25627–25636.
- Dai, T.; Cai, J.; Zhang, Y.; Xia, S.-T.; and Zhang, L. 2019. Second-order attention network for single image super-resolution. In *CVPR*, 11065–11074.
- Ding, K.; Ma, K.; Wang, S.; and Simoncelli, E. P. 2020. Image quality assessment: Unifying structure and texture similarity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(5): 2567–2581.
- Dong, C.; Loy, C. C.; He, K.; and Tang, X. 2015. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2): 295–307.
- Esser, P.; Rombach, R.; and Ommer, B. 2021. Taming transformers for high-resolution image synthesis. In *CVPR*, 12873–12883.
- Gao, G.; Li, W.; Li, J.; Wu, F.; Lu, H.; and Yu, Y. 2022. Feature distillation interaction weighting network for lightweight image super-resolution. In *AAAI*, volume 36, 661–669.
- Guo, Y.; Gao, Y.; Lu, Y.; Zhu, H.; Liu, R. W.; and He, S. 2024. Onerestore: A universal restoration framework for composite degradation. In *ECCV*, 255–272. Springer.
- He, K.; Sun, J.; and Tang, X. 2012. Guided image filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(6): 1397–1409.
- Ho, J.; Jain, A.; and Abbeel, P. 2020. Denoising diffusion probabilistic models. In *NeurIPS*, volume 33, 6840–6851.
- Jiang, K.; Wang, Z.; Yi, P.; Chen, C.; Huang, B.; Luo, Y.; Ma, J.; and Jiang, J. 2020. Multi-scale progressive fusion network for single image deraining. In *CVPR*, 8346–8355.
- Jin, X.; Chen, Z.; and Li, W. 2020. AI-GAN: Asynchronous interactive generative adversarial network for single image rain removal. *Pattern Recognition*, 100: 107143.
- Karras, T.; Laine, S.; and Aila, T. 2019. A style-based generator architecture for generative adversarial networks. In *CVPR*, 4401–4410.
- Khan, M.; Alam, M.; Masud, M.; and Amin, A. 2016. Importance of high order high pass and low pass filters. *World Applied Sciences Journal*, 34(9): 1261–1268.
- Khayam, S. A. 2003. The discrete cosine transform (DCT): theory and application. *Michigan State University*, 114(1): 31.
- Li, R.; Tan, R. T.; and Cheong, L.-F. 2020. All in one bad weather removal using architectural search. In *CVPR*, 3175–3185.
- Li, W.; Guo, H.; Hou, Y.; Gao, G.; and Ma, Z. 2025a. Dual-domain modulation network for lightweight image super-resolution. *IEEE Transactions on Multimedia*.
- Li, W.; Guo, H.; Liu, X.; Liang, K.; Hu, J.; Ma, Z.; and Guo, J. 2024a. Efficient face super-resolution via wavelet-based feature enhancement network. In *ACM MM*, 4515–4523.
- Li, W.; Li, J.; Gao, G.; Deng, W.; Yang, J.; Qi, G.-J.; and Lin, C.-W. 2024b. Efficient image super-resolution with feature interaction weighted hybrid network. *IEEE Transactions on Multimedia*.
- Li, W.; Li, J.; Gao, G.; Deng, W.; Zhou, J.; Yang, J.; and Qi, G.-J. 2023a. Cross-receptive focused inference network for lightweight image super-resolution. *IEEE Transactions on Multimedia*, 26: 864–877.
- Li, W.; Wang, M.; Zhang, K.; Li, J.; Li, X.; Zhang, Y.; Gao, G.; Deng, W.; and Lin, C.-W. 2023b. Survey on deep face restoration: From non-blind to blind and beyond. *arXiv preprint arXiv:2309.15490*.
- Li, W.; Wang, X.; Guo, H.; Gao, G.; and Ma, Z. 2025b. Self-Supervised Selective-Guided Diffusion Model for Old-Photo Face Restoration. In *NeurIPS*.
- Liang, J.; Cao, J.; Sun, G.; Zhang, K.; Van Gool, L.; and Timofte, R. 2021. Swinir: Image restoration using swin transformer. In *ICCVW*, 1833–1844.
- Mei, Y.; Fan, Y.; and Zhou, Y. 2021. Image super-resolution with non-local sparse attention. In *CVPR*, 3517–3526.
- Mittal, A.; Soundararajan, R.; and Bovik, A. C. 2012. Making a “completely blind” image quality analyzer. *IEEE Signal Processing Letters*, 20(3): 209–212.
- Özdenizci, O.; and Legenstein, R. 2023. Restoring vision in adverse weather conditions with patch-based denoising diffusion models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(8): 10346–10357.
- Park, D.; Lee, B. H.; and Chun, S. Y. 2023. All-in-one image restoration for unknown degradations using adaptive discriminative filters for specific degradations. In *CVPR*, 5815–5824.
- Patil, P. W.; Gupta, S.; Rana, S.; Venkatesh, S.; and Murala, S. 2023. Multi-weather image restoration via domain translation. In *ICCV*, 21696–21705.
- Peng, L.; Cao, Y.; Sun, Y.; and Wang, Y. 2024. Lightweight adaptive feature de-drifting for compressed image classification. *IEEE Transactions on Multimedia*, 26: 6424–6436.

- Peng, L.; Li, W.; Pei, R.; Ren, J.; Xu, J.; Wang, Y.; Cao, Y.; and Zha, Z.-J. 2025a. Towards Realistic Data Generation for Real-World Super-Resolution. In *ICLR*.
- Peng, L.; Wang, Y.; Di, X.; Fu, X.; Cao, Y.; Zha, Z.-J.; et al. 2025b. Boosting image de-raining via central-surrounding synergistic convolution. In *AAAI*, volume 39, 6470–6478.
- Qian, R.; Tan, R. T.; Yang, W.; Su, J.; and Liu, J. 2018. Attentive generative adversarial network for raindrop removal from a single image. In *CVPR*, 2482–2491.
- Quan, R.; Yu, X.; Liang, Y.; and Yang, Y. 2021. Removing raindrops and rain streaks in one go. In *CVPR*, 9147–9156.
- Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; and Ommer, B. 2022. High-resolution image synthesis with latent diffusion models. In *CVPR*, 10684–10695.
- Song, J.; Meng, C.; and Ermon, S. 2020. Denoising diffusion implicit models. In *ICLR*.
- Sun, S.; Ren, W.; Gao, X.; Wang, R.; and Cao, X. 2024. Restoring Images in Adverse Weather Conditions via Histogram Transformer. In *ECCV*.
- Tan, Z.; Wu, Y.; Liu, Q.; Chu, Q.; Lu, L.; Ye, J.; and Yu, N. 2024. Exploring the Application of Large-Scale Pre-Trained Models on Adverse Weather Removal. *IEEE Transactions on Image Processing*.
- Uddin, M. S. 2022. Real-World Single Image Super-Resolution Under Rainy Condition. *arXiv preprint arXiv:2206.08345*.
- Valanarasu, J. M. J.; Yasarla, R.; and Patel, V. M. 2022. Transweather: Transformer-based restoration of images degraded by adverse weather conditions. In *CVPR*, 2353–2363.
- Varghese, R.; and Sambath, M. 2024. Yolov8: A novel object detection algorithm with enhanced performance and robustness. In *ADICS*, 1–6. IEEE.
- Wang, H.; Chen, X.; Ni, B.; Liu, Y.; and Liu, J. 2023. Omni aggregation networks for lightweight image super-resolution. In *CVPR*, 22378–22387.
- Wang, Y.; Song, Y.; Ma, C.; and Zeng, B. 2020. Rethinking image deraining via rain streaks and vapors. In *ECCV*, 367–382. Springer.
- Wang, Z.; and Bovik, A. C. 2002. A universal image quality index. *IEEE Signal Processing Letters*, 9(3): 81–84.
- Xia, B.; Zhang, Y.; Wang, S.; Wang, Y.; Wu, X.; Tian, Y.; Yang, W.; and Van Gool, L. 2023. Diffir: Efficient diffusion model for image restoration. In *ICCV*, 13095–13105.
- Xiao, J.; Fu, X.; Liu, A.; Wu, F.; and Zha, Z.-J. 2022. Image de-raining transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(11): 12978–12995.
- Yang, W.; Tan, R. T.; Feng, J.; Liu, J.; Guo, Z.; and Yan, S. 2017. Deep joint rain detection and removal from a single image. In *CVPR*, 1357–1366.
- Ye, T.; Chen, S.; Bai, J.; Shi, J.; Xue, C.; Jiang, J.; Yin, J.; Chen, E.; and Liu, Y. 2023. Adverse weather removal with codebook priors. In *ICCV*, 12653–12664.
- Yu, J.; Fan, Y.; Yang, J.; Xu, N.; Wang, Z.; Wang, X.; and Huang, T. 2018. Wide activation for efficient and accurate image super-resolution. In *CVPR NTIRE*.
- Yuan, L.; Chen, Y.; Wang, T.; Yu, W.; Shi, Y.; Jiang, Z.-H.; Tay, F. E.; Feng, J.; and Yan, S. 2021. Tokens-to-token vit: Training vision transformers from scratch on imagenet. In *ICCV*, 558–567.
- Zhang, H.; Ba, Y.; Yang, E.; Mehra, V.; Gella, B.; Suzuki, A.; Pfahnl, A.; Chandrappa, C. C.; Wong, A.; and Kadambi, A. 2023. Weatherstream: Light transport automation of single image deweathering. In *CVPR*, 13499–13509.
- Zhang, L.; Li, Y.; Zhou, X.; Zhao, X.; and Gu, S. 2024. Transcending the limit of local window: Advanced super-resolution transformer with adaptive token dictionary. In *CVPR*, 2856–2865.
- Zhang, R.; Isola, P.; Efros, A. A.; Shechtman, E.; and Wang, O. 2018a. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 586–595.
- Zhang, Y.; Li, K.; Li, K.; Wang, L.; Zhong, B.; and Fu, Y. 2018b. Image super-resolution using very deep residual channel attention networks. In *ECCV*, 286–301.
- Zhou, Y.; Li, Z.; Guo, C.-L.; Bai, S.; Cheng, M.-M.; and Hou, Q. 2023. Srformer: Permuted self-attention for single image super-resolution. In *ICCV*, 12780–12791.
- Zhu, Y.; Wang, T.; Fu, X.; Yang, X.; Guo, X.; Dai, J.; Qiao, Y.; and Hu, X. 2023. Learning weather-general and weather-specific features for image restoration under multiple adverse weather conditions. In *CVPR*, 21747–21758.