

# FAPE-IR: Frequency-Aware Planning and Execution Framework for All-in-One Image Restoration

Jingren Liu<sup>1,4\*</sup>, Shuning Xu<sup>2,4\*</sup>, Qirui Yang<sup>1\*</sup>, Yun Wang<sup>3,4</sup>, Xiangyu Chen<sup>4†</sup>, Zhong Ji<sup>1†</sup>

<sup>1</sup>Tianjin University, <sup>2</sup>University of Macau, <sup>3</sup>City University of Hong Kong

<sup>4</sup>Institute of Artificial Intelligence (TeleAI), China Telecom

## Abstract

*All-in-One Image Restoration (AIO-IR) aims to develop a unified model that can handle multiple degradations under complex conditions. However, existing methods often rely on task-specific designs or latent routing strategies, making it hard to adapt to real-world scenarios with various degradations. We propose FAPE-IR, a Frequency-Aware Planning and Execution framework for image restoration. It uses a frozen Multimodal Large Language Model (MLLM) as a planner to analyze degraded images and generate concise, frequency-aware restoration plans. These plans guide a LoRA-based Mixture-of-Experts (LoRA-MoE) module within a diffusion-based executor, which dynamically selects high- or low-frequency experts, complemented by frequency features of the input image. To further improve restoration quality and reduce artifacts, we introduce adversarial training and a frequency regularization loss. By coupling semantic planning with frequency-based restoration, FAPE-IR offers a unified and interpretable solution for all-in-one image restoration. Extensive experiments show that FAPE-IR achieves state-of-the-art performance across seven restoration tasks and exhibits strong zero-shot generalization under mixed degradations. Code is available at <https://github.com/Programmergg/FAPE-IR>.*

## 1. Introduction

Image restoration seeks to recover clean images from degradations such as rain, snow, fog, blur, low-light, and noise, and is fundamental to safety- and quality-critical applications in medical imaging [7, 84], consumer photography [101], and autonomous driving [25, 42, 87]. Traditional pipelines typically employ specialized architectures or optimization strategies for each type of degradation [28, 63, 107, 112]. Although recent single-degradation models report strong results on curated benchmarks [20,

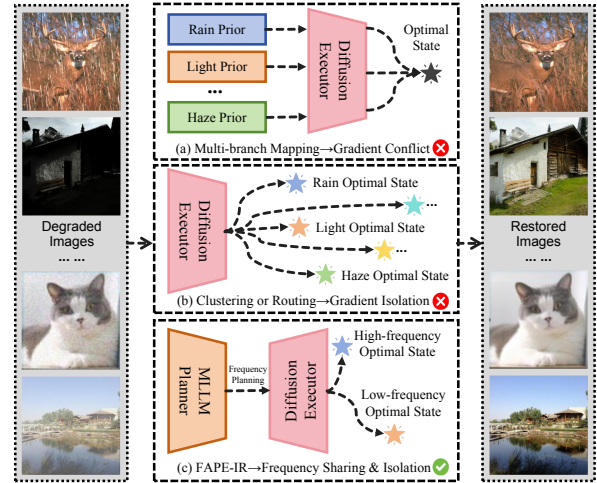


**Figure 1.** Comparison of AIO-IR methods: (a) multi-branch mappings with task-level priors; (b) task-specific routing/clustering; (c) our FAPE-IR, unifying understanding and restoration.

[31, 85, 100, 115], real-world images rarely present a single known corruption; degradations are often unknown, co-occurring, and subject to distribution shift, which undermines stability and controllability. This gap motivates All-in-One Image Restoration (AIO-IR), which aims to handle multiple degradations within a single model [8, 10, 11, 49, 59, 68, 78, 106]. Yet, most current efforts still assume task separability [1, 49] or tether restoration to reconstruction pipelines driven by task-level priors [86, 96], preventing models from autonomously perceiving and disentangling degradation attributes. As a result, model generalization to real-world scenarios remains inadequate.

Within current AIO-IR methods, a prevailing strategy is to employ multi-branch mappings (see Figure 1(a)), injecting task-level conditions via explicit priors (e.g., prompts, embeddings, or specialized encoders) and learning the restoration paths accordingly [8, 39, 50, 51, 57, 114]. This design is prone to severe cross-task interference. In a unified model with shared parameters, gradients from different degradations conflict, making it difficult to reach task-specific optima simultaneously. Moreover, both training and inference depend on prior labels or textual inputs, incur-

\*Equal contribution.

†Corresponding author.

Work done during an internship at TeleAI.

ring substantial annotation and prompt-engineering costs. A complementary strategy clusters features or employs specialized routing mechanisms within the model (see Figure 1(b)) [13, 15, 78, 90, 109, 113]. Because tasks are strongly separated in latent space, their learning processes become nearly independent, which hampers the discovery of shared structures and compositionality across degradations. Consequently, robustness to compound degradations is limited, and generalization suffers. More critically, both lines lack semantic understanding and rely on opaque degradation pipelines, hindering content-aware adaptation and interpretability. These limitations call for a unified, semantics-aware AIO-IR framework in which degradation understanding and restoration are coupled, and task knowledge is both disentangled and shared in an adaptive way.

To instantiate such a framework, we adopt diffusion-based image restoration, which has recently become a popular choice for AIO-IR due to its strong generative priors [8, 13, 15, 57, 90, 113, 114]. This formulation naturally models diverse degradations, yet prior work typically treats diffusion as a task-conditioned generator driven by labels or hand-crafted priors, leaving degradation modeling opaque and under-exploiting cross-degradation interactions. We instead regard diffusion as the execution engine in a unified multimodal understanding-generation paradigm. In this paradigm, degradations are first parsed and organized, and then restored. Inspired by recent unified models [9, 17, 40, 95, 99, 102], we instantiate FAPE-IR, a multimodal large language model (MLLM) + diffusion framework for low-level vision (see Figure 1(c)). An MLLM explicitly parses image content and degradation semantics and outputs understanding and planning tokens that condition a diffusion executor to perform content-adaptive restoration in the frequency domain, grouping tasks into high- and low-frequency regimes, thereby encouraging sharing within same-frequency tasks and isolating conflicting ones.

Concretely, FAPE-IR couples a multimodal planner with a diffusion-based executor to mitigate cross-task gradient conflict and isolation from a frequency perspective. In the planning stage, we employ a label-free procedure to extract as many degradation-related low-level features as possible, and introduce several well-designed instructions that constrain the model to output: the degradation types present in the image, the primary frequency bands requiring restoration, the underlying causes, and the proposed restoration pipeline. This design yields effective, per-image understanding that more accurately guides the restoration process. In the execution stage, tokens produced by the planner, augmented with high-/low-frequency visual features from SigLIP-v2 [79] and a VAE [30], guide the restoration of the entire image. Furthermore, we propose a Frequency-Aware LoRA-MoE architecture, where text tokens from the planner and the high-/low-frequency signals extracted from

intermediate executor’s features drive the MoE gating to select among high-/low-frequency experts dynamically and interpretably. For the overall training recipe, we provide theoretical and empirical evidence that adversarial training on diffusion-pretrained weights can yield higher fidelity and fewer artifacts than other fine-tuning methods under our setting. Building on this, we further introduce an energy-based frequency regularization loss so that each frequency expert is optimized in its most suitable band.

Our main contributions are summarized as follows: (i) We introduce FAPE-IR, a unified understanding-generation AIO-IR framework that couples a multimodal planner with a diffusion-based executor, enabling semantics-driven, content-adaptive restoration across diverse degradations. (ii) We develop a frequency-aware modeling strategy that dynamically routes information between high- and low-frequency experts, mitigating cross-task interference while promoting shared representations among related tasks. (iii) We conduct comprehensive experiments on a wide range of single-degradation and mixed-degradation benchmarks, demonstrating that FAPE-IR achieves state-of-the-art or competitive performance and exhibits strong zero-shot generalization to unseen compound degradations.

## 2. Related Works

### 2.1. Diffusion-based AIO-IR

In recent years, diffusion-based AIO-IR methods have advanced rapidly. Existing approaches fall into two families: (i) multi-branch mappings and (ii) clustering or routing. The former relies on injecting task-level priors into a shared backbone via task encoders or textual/latent prompts. PromptIR [58], Prompt-in-Prompt [39], and InstructIR [14] unify degradations through prompts or instructions. ProRes [51] and DA-CLIP [50] supply content/degradation embeddings via visual prompts or CLIP [61] controllers and fuse them with cross-attention. UniRestore [8] injects task cues into diffusion, whereas UniRes [114] composes task specialists at sampling time. The latter adapts internal representations by clustering features or routing to experts. AdaIR [15] and DFPIR [78] mine frequency cues and reshape separable, degradation-aware features. AMIR-Net [110] performs label-free, layer-wise clustering with domain alignment. Within diffusion, DiffUIR [113] disentangles shared/private factors for selective routing. MoCE-IR [109] and M2Restore [90] use complexity-aware MoE to activate only necessary experts. Overall, these strategies improve efficiency but lack sample-level planning, relying on fixed task-level designs and ignoring whether tasks should share or isolate knowledge. This leads to under-shared representations, misrouting, and artifacts under unknown or composite degradations, limiting open-world robustness. To address this, we propose FAPE-IR, a unified understanding-generation framework.

## 2.2. Unified Models

A central goal of multimodal AI is to develop general-purpose foundation models that can jointly understand and generate diverse modalities within a single framework. Recent unified models largely follow three paradigms. Multimodal autoregression methods, such as Chameleon [77], Emu3 [89], and UGen [76], tokenize images into a shared discrete space and train a single decoder-only Transformer, but this tokenization degrades fine details and incurs long generation latency, limiting their suitability for precise low-level restoration. Multimodal diffusion approaches instead unify modalities at the denoising stage: MMaDA [102] employs a modality-agnostic diffusion backbone with unified post-training; LaViDa [36] uses elastic masked diffusion for understanding, editing, and generation; Dimple [108] couples an autoregression warm-up with discrete diffusion and parallel confident decoding; and UniDisc [75] casts joint text-image modeling as discrete diffusion in a shared token space. However, this paradigm remains nascent, with unresolved issues in consistency and inference efficiency. A third line, MLLM+Diffusion frameworks, decouple high-level multimodal understanding from generation: UniWorld-V1 [40] conditions diffusion on high-resolution semantic features from MLLMs and contrastive encoders; BAGEL [17] integrates decoder-only MLLMs with diffusion heads in an “integrated Transformer” for both understanding and image generation; Janus [95] decouples visual encoding for understanding versus generation under a unified backbone; and BLIP3-o [9] uses CLIP-based semantic spaces with diffusion transformers and staged pre-training to balance image understanding and generation. Yet these conditioning interfaces are primarily designed for high-level semantic editing or creation, rather than pixel-accurate, artifact-free restoration. Our FAPE-IR framework follows the MLLM+Diffusion paradigm but is specialized for low-level image restoration, where fine-grained controllability and artifact suppression are central.

## 3. Methodology

As shown in Figure 2, FAPE-IR adopts a planning and execution paradigm that decomposes image restoration into a frequency-aware planner and a band-specialized restoration executor. In the remainder of this section, we first describe the frequency-aware planner, including the construction of a label-free low-level feature pool, instructions and planning, and the encoding of frequency-aligned understanding tokens (Section 3.1). We then present the proposed Frequency-Aware LoRA-MoE architecture in executor (Section 3.2). Finally, we introduce the adversarial training strategy and the overall loss formulation, which couples frequency-aware routing with perceptual and adversarial supervision (Section 3.3).

### 3.1. Frequency-aware Planner

**Label-free Low-level Feature Pool.** Given a degraded input image  $c$ , we first construct a label-free low-level feature pool that characterizes its degradation pattern from a frequency perspective (details in Appendix). Specifically, we compute a vector of simple image statistics  $P_{\text{hints}}$  directly from pixel values, without relying on degradation labels or auxiliary metadata. Each component of  $P_{\text{hints}}$  corresponds to one of seven representative degradation types: for **rain**, the strength of oriented streaks; for **snow**, the density of small bright blobs; for **noise**, luminance and chroma variations in flat regions; for **blur**, the responses of Laplacian and gradient operators; for **haze**, dark-channel and saturation statistics; for **low light**, global luminance; and for **super-resolution**, the spatial dimensions  $H \times W$  of the input. These lightweight global statistics can be computed in a single pass and require no supervision. The resulting vector  $P_{\text{hints}}$  is then provided to the planner as frequency-aware visual information describing the input degradation.

**Instructions and Planning.** We consider a low-level restoration task space  $\mathcal{T}$  that covers seven representative degradations, including low-light enhancement, dehazing, desnowing, deraining, deblurring, denoising, and super-resolution. In FAPE-IR, Qwen2.5-VL [4] serves as a multimodal planner that turns frequency-aware visual cues and textual instructions into a structured, frequency-explicit restoration plan. Concretely, the planner is conditioned on three types of inputs: a universal restoration instruction  $r$  that describes the general goal of recovering a clean and sharp image (details in Appendix); an expert rule  $P_{\text{expert}}$  that encodes the task taxonomy and prior knowledge about which degradations affect which frequency bands (also detailed in Appendix); and the concise visual hints  $P_{\text{hints}}$  derived from the label-free low-level feature pool. Based on these inputs, Qwen2.5-VL produces a parse-friendly textual output, which we then parse into a tuple  $FP = (\hat{t}, \hat{f}, \mathcal{R}, \mathcal{E})$ , where  $\hat{t}$  is the selected task from  $\mathcal{T}$ ,  $\hat{f}$  indicates whether the restoration should primarily focus on high- or low-frequency content,  $\mathcal{R}$  summarizes the intended restoration pipeline in natural language, and  $\mathcal{E}$  explains the reasoning behind this choice. This human-readable plan acts as a routing signal for downstream high- and low-frequency experts, making the decision process transparent and interpretable even under composite degradations.

**Encoding.** To ensure the planner’s decisions effectively guide the diffusion-based executor, we use Qwen2.5-VL’s encoding capability to compile the planner decisions  $FP$  into compact, frequency-aligned understanding tokens  $h$ . To mitigate the gap between understanding tokens and the executor’s generation tokens, we enrich the conditioning tokens with high- and low-frequency visual information. Concretely, we use Qwen2.5-VL’s visual placeholders to inject tokens produced by SigLIP-v2  $\mathcal{E}_{\text{sig}}(c)$  and the VAE

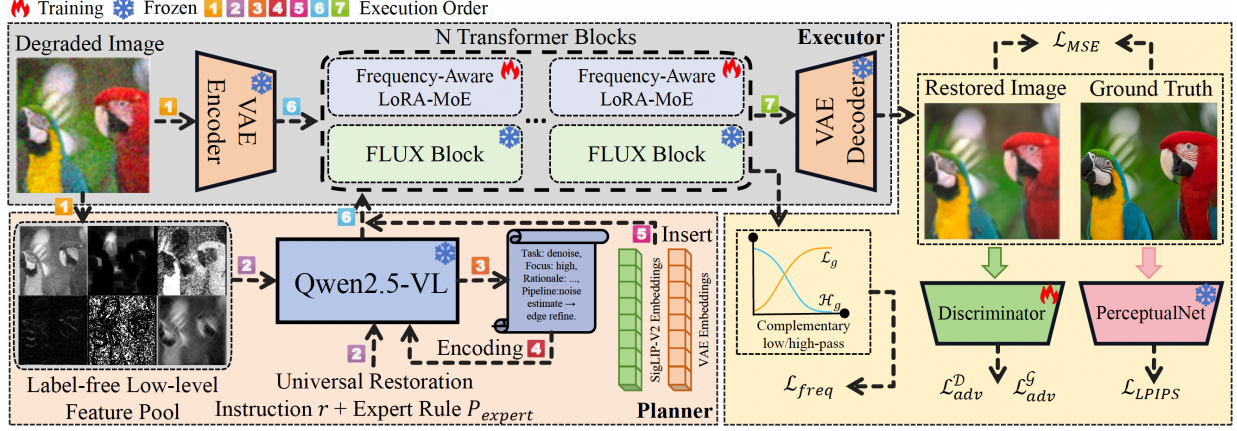


Figure 2. Overview of the proposed FAPE-IR framework.

$\mathcal{E}_{\text{vae}}(c)$ , replacing the executor’s original T5 encoding for conditioning  $h_{\text{cond}} = \text{insert}(h, \mathcal{E}_{\text{sig}}(c), \mathcal{E}_{\text{vae}}(c))$ . In addition, to retain precise, human-interpretable semantics for the subsequent Frequency-Aware LoRA-MoE module, we similarly use text placeholders to extract the textual tokens  $h_{\text{text}} = h[:, \text{slot}_{\text{text}}, :]$ .

### 3.2. Restoration Executor

Given a degraded image  $c$ , we employ a restoration executor based on the FLUX transformer [30] to refine the VAE latent  $z$  into a restored latent  $\hat{z}$ , guided jointly by the textual tokens  $h_{\text{text}}$  and the conditioning tokens  $h_{\text{cond}}$ . The VAE decoder then maps  $\hat{z}$  to the restored image  $\hat{x}$ .

To meet constraints on model size and optimization time, FAPE-IR adopts a parameter-efficient LoRA-MoE executor with two experts specialized for high- and low-frequency bands. Building on this design, we introduce a Frequency-Aware LoRA-MoE module with a dual-end gating mechanism: a planner-side gate uses the frequency-aligned understanding tokens  $h_{\text{text}}$  to propose an expert selection, while an executor-side gate inspects intermediate high- and low-frequency generation tokens to refine that choice. As illustrated in Figure 3, this design conditions the MoE router jointly on semantic and spectral cues, leading to more precise expert assignment and improved robustness to frequency-band drift during optimization.

#### 3.2.1. Frequency-Aware LoRA-MoE

**Frequency-Aware Text Router.** Since the planner yields  $h_{\text{text}} \in \mathbb{R}^{B \times K \times D}$ , where  $B$  denotes the batch size,  $K$  denotes the planner token length, and  $D$  denotes the embedding dimension, while the LoRA-MoE gate operates on representations in  $\mathbb{R}^{B \times L \times D}$ , where  $L$  denotes the LoRA-MoE gate token length, we right-pad  $h_{\text{text}}$  along the token axis to length  $L$  ( $K \leq L$ ), leaving  $D$  unchanged, so that the gating network can be applied token-wise over a sequence of consistent length. Then, we implement the text gating function as a token-wise fully connected layer on the padded text to-

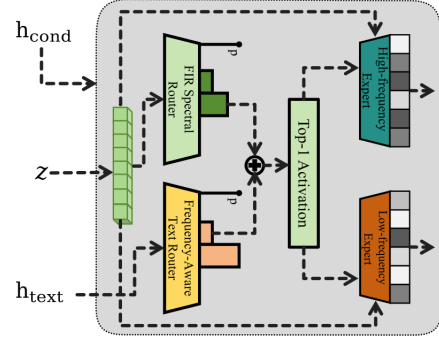


Figure 3. Frequency-Aware LoRA-MoE architecture.

kens, followed by a softmax over experts to obtain the text gating weights  $w_{\text{text}} = \text{Softmax}(W_t \cdot \text{Padding}(h_{\text{text}}))$ .

**FIR Spectral Router.** To compensate for the limitations of high-level semantic tokens in gating, we add a visual-frequency path that separates the executor’s generation tokens into low/high-frequency components using a depth-wise FIR low-pass filter [12, 53]. Let  $h_{\text{gen}} \in \mathbb{R}^{B \times L \times D}$  denote the per-block input. For an odd kernel size  $K$  and a symmetric, normalized 1D Gaussian  $g \in \mathbb{R}^K$  with  $\sum_k g_k = 1$ , we define a depthwise 1D convolution along the token axis as  $h_{\text{low}} = \mathcal{L}_g(h_{\text{gen}})$  and  $h_{\text{high}} = h_{\text{gen}} - h_{\text{low}}$ .

Next, to allocate the proportion of energy between the low- and high-frequency generation tokens, we compute their relative energy and derive the high-low gating weights. Specifically, for each token we set  $e_{\text{low}} = \|h_{\text{low}}\|_2^2$  and  $e_{\text{high}} = \|h_{\text{high}}\|_2^2$ , then  $p_{\text{low}} = \frac{e_{\text{low}}}{e_{\text{low}} + e_{\text{high}}}$  and  $p_{\text{high}} = 1 - p_{\text{low}}$ ; a temperature-scaled softmax over  $[p_{\text{low}}, p_{\text{high}}]$  yields the high-low gating weights  $w_{\text{visual}}$ .

**Softmax-gated Merged Weights.** Subsequently, we fuse the textual and spectral gating weights with non-negative, learnable scalars  $\lambda_s$ :  $\tilde{\alpha} = \lambda_s w_{\text{text}} + (1 - \lambda_s) w_{\text{visual}}$ . Finally, we apply top-1 routing to obtain the final frequency-band selection, i.e.,  $\alpha = \text{Top1}(\tilde{\alpha})$ , where  $\alpha$  is a one-hot vector indicating the chosen band.

With per-expert coefficients  $\alpha_i$  produced by the fused and sparsified router, the FLUX projection is updated as:

$$W' = W + \sum_{i=1}^N \alpha_i A_i B_i, \quad (1)$$

where  $(A_i, B_i)$  is a rank- $r_i$  LoRA adapter for the  $i$ -th frequency expert and the backbone  $W$  is frozen.

### 3.3. Adversarial Training

To improve efficiency at both training and inference, our FAPE-IR replaces the flow-matching fine-tuning objective [45] commonly used in unified models with adversarial training. In Theorem 6 (Appendix), we show that, compared with auto-regression [74, 77, 89] and flow-matching [9, 17, 40, 95] losses, adversarial training attains fewer artifacts. Next, we detail FAPE-IR’s training paradigm in Figure 2.

**Discriminator.** Given a ground truth image  $x$  (restored image  $\hat{x}$ ), the frozen SigLIP-v2 discriminator  $\mathcal{F}_{\text{sig}}$  yields (i) a sequence of hidden states from which we take  $L$  token maps at increasing depths and reshape them into spatial feature maps  $\{\mathbf{f}^{(l)} \in \mathbb{R}^{C_l \times H_l \times W_l}\}_{l=1}^L$ , and (ii) a pooled representation  $\mathbf{p} \in \mathbb{R}^D$  from the final layer. To more comprehensively judge the input, we attach a multi-level discriminator head that mines spatial evidence across scales.

**Multi-level Discriminator Head.** The discriminator head  $\mathcal{H}_\psi$  processes each spatial map with a shallow, spectrally normalized convolutional path that includes anti-aliasing BlurPool downsampling, and processes the pooled token:

$$\mathbf{s}^{(l)} = \mathcal{H}_\psi^{(l)}(\mathbf{f}^{(l)}) \in \mathbb{R}^{H'_l \times W'_l}, \mathbf{s}^{\text{pool}} = \mathcal{H}_\psi^{\text{pool}}(\mathbf{p}). \quad (2)$$

Each spatial score map  $\mathbf{s}^{(l)}$  is spatially averaged to a scalar  $\bar{s}^{(l)} = \frac{1}{H'_l W'_l} \sum_{i,j} \mathbf{s}_{ij}^{(l)}$ ; we then aggregate levelwise with uniform weights:  $D(x) = \frac{1}{L+1} \left( \sum_{l=1}^L \bar{s}^{(l)} + \mathbf{s}^{\text{pool}} \right)$ . This design encourages consistency between local structures (via multi-scale spatial paths) and global semantics (via the pooled path), while remaining lightweight.

**Discriminator Loss.** Following the adversarial training, we optimize the critic with the standard discriminator loss:

$$\mathcal{L}_{\text{adv}}^{\mathcal{D}} = -\mathbb{E}_x[\log D(x)] - \mathbb{E}_{\hat{x}}[\log(1 - D(\hat{x}))]. \quad (3)$$

**Generator Loss.** To minimize over-generation and distortion, we do not concatenate additional noise channels into the inputs of the FLUX transformer. Finally, we train  $\mathcal{G}_\theta$  under a composite adversarial training loss.

$$\mathcal{L}_{\text{adv}} = \underbrace{\alpha \|\hat{x} - x\|_2^2}_{\mathcal{L}_{\text{MSE}}} + \underbrace{\beta \|\phi(\hat{x}) - \phi(x)\|_2^2}_{\mathcal{L}_{\text{LPIPS}}} - \underbrace{\lambda \mathbb{E}[D(\hat{x})]}_{\mathcal{L}_{\text{adv}}^{\mathcal{G}}}. \quad (4)$$

While  $\mathcal{L}_{\text{adv}}$  promotes pixel fidelity, perceptual similarity, and adversarial realism, it remains agnostic to how spectral content is allocated across experts. To align training

with the frequency-aware routing in Subsection 3.1–3.2, we add a single frequency regularizer on adapter outputs that penalizes out-of-band synthesis. This preserves objective parsimony, only one additional scalar term, while explicitly enforcing low/high-band specialization.

**Frequency Regularizer.** To make the two LoRA–MoE experts specialize in complementary bands, we penalize out-of-band energy on their adapter outputs. Let  $\mathcal{L}_g$  be a depthwise FIR low-pass along the token axis, and define its complementary high-pass  $\mathcal{H}_g \triangleq I - \mathcal{L}_g$ . For the low/high-frequency experts’ outputs  $y_{\text{low}}$  and  $y_{\text{high}}$ , we minimize  $\mathcal{L}_{\text{freq}} = \text{mean}[\|\mathcal{H}_g(y_{\text{low}})\|_2^2 + \|\mathcal{L}_g(y_{\text{high}})\|_2^2]$ . In practice,  $g$  is a fixed Gaussian kernel stored as a buffer (non-trainable), and the loss is summed across layers.

At this point, the overall FAPE-IR objective is:

$$\mathcal{L}_{\text{Total}} = \mathcal{L}_{\text{adv}} + \gamma \mathcal{L}_{\text{freq}}. \quad (5)$$

In general, by marrying a frequency-aware planner with band-specialized LoRA–MoE and an adversarial training objective augmented by a single frequency regularizer, our FAPE-IR reconciles interpretability with fidelity.

## 4. Experiments

### 4.1. Experimental Setup

**Training Data.** We construct a training corpus covering seven restoration tasks, including deblurring, dehazing, low-light enhancement, deraining, desnowing, denoising, and super-resolution. Weather-related degradations (deraining, desnowing, dehazing) are sourced from Snow100K-Train [47], Rain100-L/H-Train [104], and the NTIRE 2025 Challenge dataset [37]. For dehazing, we use a curated 10K subset of OTS/ITS-Train [33] combined with URHI-Train [18]. Deblurring utilizes RealBlur-Train (R/J) [64], GoPro-Train, and its gamma-corrected variant [56]. Low-light enhancement is trained on LOL-v2-Train [105]. For denoising, BSD400 [54] and WaterlooED [52] are corrupted with Gaussian noise at  $\sigma \in \{15, 25, 50\}$ . For super-resolution, we follow prior work [92, 97, 98] and adopt LS-DIR [38] plus a 10K-image FFHQ subset [27], with degradations synthesized via Real-ESRGAN [88].

**Testing Data.** We evaluate our method on a comprehensive collection of benchmarks that span both synthetic and real-world degradations across deblurring, dehazing, low-light enhancement, deraining, desnowing, denoising, and super-resolution. For deblurring, we adopt the official test splits of RealBlur-J/R [64], GoPro, and GoPro\_gamma [56]. Dehazing evaluation is conducted using the ITS validation set and the URHI test split [33]. Low-light enhancement is assessed on the test sets of LOL-v1 [91] and LOL-v2 [105]. Deraining is benchmarked on Rain100-L/H [104], OutDoor [35], and RainDrop [60]. For desnowing, we use the Snow100K-L and Snow100K-S test subsets [47] for large-scale quali-

**Table 1.** Unified comparison across six AIO-IR task series. Each series shows five metrics (PSNR $\uparrow$ , SSIM $\uparrow$ , LPIPS $\downarrow$ , FID $\downarrow$ , DISTS $\downarrow$ ). The best results are highlighted in **red**, and the second-best results are shown in **blue**.

| Method         | Deraining       |                 |                    |                  |                    | Denoising       |                 |                    |                  |                    | Deblurring      |                 |                    |                  |                    |
|----------------|-----------------|-----------------|--------------------|------------------|--------------------|-----------------|-----------------|--------------------|------------------|--------------------|-----------------|-----------------|--------------------|------------------|--------------------|
|                | PSNR $\uparrow$ | SSIM $\uparrow$ | LPIPS $\downarrow$ | FID $\downarrow$ | DISTS $\downarrow$ | PSNR $\uparrow$ | SSIM $\uparrow$ | LPIPS $\downarrow$ | FID $\downarrow$ | DISTS $\downarrow$ | PSNR $\uparrow$ | SSIM $\uparrow$ | LPIPS $\downarrow$ | FID $\downarrow$ | DISTS $\downarrow$ |
| PromptIR [58]  | <b>21.94</b>    | <b>0.73</b>     | <b>0.30</b>        | 100.17           | <b>0.21</b>        | 31.27           | <b>0.84</b>     | <b>0.16</b>        | 57.82            | <b>0.14</b>        | —               | —               | —                  | —                | —                  |
| FoundIR [34]   | 19.68           | 0.68            | 0.37               | 118.29           | 0.24               | 30.85           | 0.80            | 0.21               | 67.08            | 0.22               | 28.35           | 0.85            | 0.21               | 42.03            | 0.16               |
| DFPIR [78]     | 19.27           | 0.68            | 0.37               | 104.13           | 0.27               | 31.31           | 0.83            | 0.17               | 61.80            | 0.15               | <b>30.82</b>    | <b>0.90</b>     | <b>0.15</b>        | <b>28.23</b>     | <b>0.12</b>        |
| MoCE-IR [109]  | 21.42           | 0.71            | 0.32               | 101.21           | 0.23               | <b>31.63</b>    | <b>0.84</b>     | <b>0.16</b>        | <b>56.74</b>     | <b>0.14</b>        | 21.67           | 0.74            | 0.28               | 48.85            | 0.19               |
| AdaIR [15]     | 21.64           | 0.71            | 0.32               | <b>100.07</b>    | 0.23               | <b>31.45</b>    | <b>0.84</b>     | 0.17               | 58.05            | <b>0.14</b>        | 22.23           | 0.75            | 0.28               | 49.64            | 0.19               |
| FAPE-IR (Ours) | <b>28.30</b>    | <b>0.84</b>     | <b>0.09</b>        | <b>21.55</b>     | <b>0.07</b>        | 30.34           | <b>0.87</b>     | <b>0.10</b>        | <b>30.33</b>     | <b>0.09</b>        | <b>30.91</b>    | <b>0.88</b>     | <b>0.10</b>        | <b>16.53</b>     | <b>0.07</b>        |

| Method         | Desnowing       |                 |                    |                  |                    | Dehazing        |                 |                    |                  |                    | Low-light enhancement |                 |                    |                  |                    |
|----------------|-----------------|-----------------|--------------------|------------------|--------------------|-----------------|-----------------|--------------------|------------------|--------------------|-----------------------|-----------------|--------------------|------------------|--------------------|
|                | PSNR $\uparrow$ | SSIM $\uparrow$ | LPIPS $\downarrow$ | FID $\downarrow$ | DISTS $\downarrow$ | PSNR $\uparrow$ | SSIM $\uparrow$ | LPIPS $\downarrow$ | FID $\downarrow$ | DISTS $\downarrow$ | PSNR $\uparrow$       | SSIM $\uparrow$ | LPIPS $\downarrow$ | FID $\downarrow$ | DISTS $\downarrow$ |
| PromptIR [58]  | —               | —               | —                  | —                | —                  | <b>21.94</b>    | <b>0.90</b>     | <b>0.14</b>        | <b>24.26</b>     | <b>0.11</b>        | —                     | —               | —                  | —                | —                  |
| FoundIR [34]   | 23.79           | <b>0.78</b>     | 0.28               | 29.31            | 0.17               | 13.65           | 0.75            | 0.31               | 34.41            | 0.21               | 14.44                 | 0.67            | 0.32               | 86.46            | 0.22               |
| DFPIR [78]     | 20.78           | 0.74            | 0.30               | 34.86            | 0.18               | 13.54           | 0.73            | 0.30               | 29.43            | 0.21               | <b>25.92</b>          | <b>0.90</b>     | <b>0.16</b>        | 52.15            | 0.13               |
| MoCE-IR [109]  | 23.82           | 0.76            | 0.28               | 30.47            | <b>0.16</b>        | 15.54           | 0.80            | 0.22               | 27.98            | 0.15               | 23.36                 | <b>0.89</b>     | <b>0.16</b>        | <b>44.61</b>     | <b>0.12</b>        |
| AdaIR [15]     | <b>24.19</b>    | 0.77            | <b>0.27</b>        | <b>26.82</b>     | <b>0.16</b>        | 19.89           | 0.88            | 0.16               | 25.71            | 0.12               | 24.51                 | 0.86            | 0.17               | 50.78            | 0.13               |
| FAPE-IR (Ours) | <b>30.29</b>    | <b>0.88</b>     | <b>0.08</b>        | <b>1.49</b>      | <b>0.06</b>        | <b>33.85</b>    | <b>0.97</b>     | <b>0.04</b>        | <b>7.21</b>      | <b>0.04</b>        | <b>25.34</b>          | <b>0.90</b>     | <b>0.11</b>        | <b>40.30</b>     | <b>0.09</b>        |

**Table 2.** Unified comparison across SR task series. Each series shows five metrics. The best results are marked in **red**, and the second-best results are shown in **blue**.

| Method         | SR              |                 |                    |                  |                    |
|----------------|-----------------|-----------------|--------------------|------------------|--------------------|
|                | PSNR $\uparrow$ | SSIM $\uparrow$ | LPIPS $\downarrow$ | FID $\downarrow$ | DISTS $\downarrow$ |
| StableSR [83]  | 25.10           | 0.73            | 0.29               | <b>110.18</b>    | 0.22               |
| DiffBIR [41]   | 25.65           | 0.65            | 0.39               | 124.53           | 0.25               |
| SeeSR [98]     | 26.60           | 0.74            | 0.31               | 122.10           | 0.23               |
| PASD [103]     | <b>26.87</b>    | 0.75            | 0.29               | 120.46           | <b>0.21</b>        |
| OSDiff [97]    | 26.49           | <b>0.76</b>     | <b>0.28</b>        | 117.55           | <b>0.21</b>        |
| PURE [92]      | 24.47           | 0.64            | 0.38               | 121.07           | 0.24               |
| FAPE-IR (Ours) | <b>28.53</b>    | <b>0.85</b>     | <b>0.19</b>        | <b>85.82</b>     | <b>0.15</b>        |

tative comparison. Denoising performance is measured on BSD68 [54] and Urban100 [22], each corrupted with additive Gaussian noise at  $\sigma \in \{15, 25, 50\}$ . Super-resolution is evaluated on RealSR [26] and DRealSR [93] at  $\times 2$  and  $\times 4$  scales. We follow standard evaluation protocols: RealSR is tested using the center-crop strategy [92], while images from other benchmarks are resized to  $512 \times 512$ .

**Optimization.** We train our FAPE-IR framework using the Prodigy optimizer [55] for the main model and AdamW [48] for the discriminator head. The base learning rate for AdamW is initialized at  $1e-4$  and cosine-annealed during training. All models are trained for 200K steps with a batch size of 1 on  $512 \times 512$  inputs across all datasets. Training is conducted on  $8 \times$  NVIDIA H200 GPUs. In the one-step setting, we use  $\alpha = 50.0$ ,  $\beta = 5.0$ ,  $\lambda = 0.5$ ,  $\gamma = 1 \times 10^{-3}$ , and a diffusion timestep  $t = 300$ .

## 4.2. Quantitative Results

We conduct a quantitative comparison between FAPE-IR and recent AIO-IR methods, including PromptIR [58] (NeurIPS’23), FoundIR [34] (ICCV’25), AdaIR [15] (ICLR’25), DFPIR [78] (CVPR’25), and MoCE-IR [109] (CVPR’25), as well as some SR methods: StableSR [83] (IJCV’24), DiffBIR [41] (ECCV’24), SeeSR [98]

**Table 3.** Complexity comparison. All methods are evaluated on  $512 \times 512$  inputs, and inference is measured on an H200 GPU. Param (G) denotes the inference memory footprint.

| Metric    | Methods      |            |               |            |           |       |
|-----------|--------------|------------|---------------|------------|-----------|-------|
|           | FoundIR [34] | DFPIR [78] | MoCE-IR [109] | AdaIR [15] | PURE [92] | Ours  |
| Time (s)  | 0.18         | 0.08       | 0.10          | 0.10       | 201.67    | 1.57  |
| Param (G) | 3.36         | 3.92       | 2.11          | 3.49       | 26.22     | 38.92 |

(CVPR’24), PASD [103] (ECCV’24), OSDiff [97] (NeurIPS’24), and PURE [92] (ICCV’25). We evaluate all tasks on a single trained model rather than training for each task and testing them individually.

As summarized in Table 1 and Table 2 (with per-benchmark scores and perceptual metrics detailed in Table 5 and Table 6 in the Appendix), FAPE-IR consistently secures either the best or second-best results across all six AIO-IR task series. On weather-related benchmarks (deraining, desnowing, dehazing), FAPE-IR brings substantial gains over the strongest baselines, typically improving PSNR by about 6–8 dB and reducing perceptual distances (LPIPS, FID, DISTS) by several times (e.g., deraining improves PSNR from 21.94 dB to 28.30 dB and FID from around 100.07 to 21.55). These improvements are most evident on degradations with rich semantic layouts and complex frequency patterns, where the frequency-aware planner guides the LoRA-MoE executor to apply band-specialized experts, yielding cleaner structures and fewer artifacts. For denoising and low-light enhancement, FAPE-IR attains slightly lower PSNR than the best methods but achieves the best SSIM and consistently lower LPIPS/FID/DISTS, suggesting that the combination of frequency-aware planning and adversarial training favors perceptually faithful details. On SR benchmarks, FAPE-IR again outperforms all competitors on all five metrics, for instance increasing PSNR from 26.87 dB to 28.53 dB and SSIM from 0.76 to 0.85 while notably reducing FID and DISTS, indicating that our FAPE-IR performs well on mixed degradations. Overall, these re-



**Figure 4.** Comparison among unified models, including BAGEL [17], Nexus-Gen [111], Uniworld-V1 [40], and Emu3.5 [16].

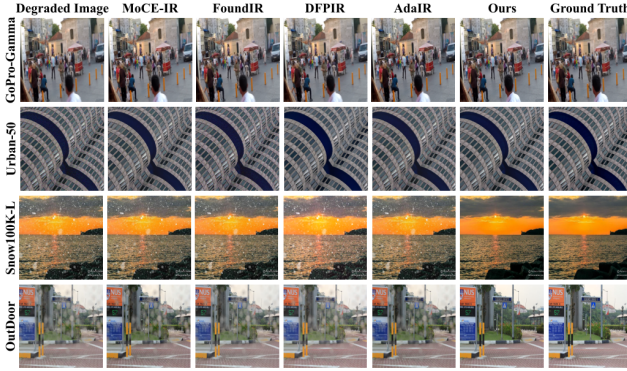


**Figure 5.** High-frequency-dominant tasks (derain/desnow/deblur/denoise): FAPE-IR preserves fine structures and textures with less ringing/oversharpening. Please zoom in for details.

sults show that explicitly modeling both semantic structure and frequency content allows FAPE-IR to strike a strong balance between distortion and perceptual quality across a broad range of degradations.

To meet real-world latency requirements, Table 3 further shows that, despite its larger number of parameters, FAPE-IR maintains reasonable inference time and runs notably faster than the unified counterpart PURE [92], suggesting that the framework is feasible for interactive applications.

### 4.3. Visual Comparison with Unified Models

As shown in Figure 4, we benchmark FAPE-IR against recent unified models on low-level restoration. On Rain-Drop, most unified models lack explicit low-level modeling and thus fail to complete restoration, either removing only a subset of droplets or causing noticeable color shifts. Emu 3.5 [16] shows a marked jump over earlier Emu variants [73, 74, 89] and contemporaries like BAGEL [17], Uniworld-V1 [40], and Nexus-Gen [111], likely due to its large corpus and 34.1B parameters. However, its multi-modal autoregression design still introduces fine-detail artifacts. By contrast, FAPE-IR reliably restores structures without color drift, underscoring the benefits of frequency-aware planning and band-specialized execution, and pointing toward a path to transfer it to future unified models.

### 4.4. Visual Comparison with AIO-IR Models

As shown in Figure 5 and Figure 6, we provide visual comparisons between FAPE-IR and prior AIO-IR methods. Relative to contemporaries, FAPE-IR exhibits stronger vi-

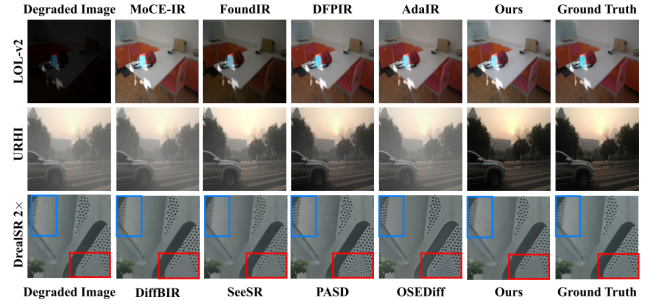


**Figure 6.** Low-frequency & SR: FAPE-IR cleans haze and balances illumination, preserving color; for SR it yields higher fidelity and fewer artifacts. Please zoom in for details.

sual robustness. In high-frequency tasks, on the large-scale Snow100K-L/S benchmarks, competing methods often fail to recover fine structures, whereas FAPE-IR reconstructs them faithfully. On Urban-50, although a few methods achieve slightly higher scores, FAPE-IR produces visibly less noise and cleaner edges. In low-frequency cases, such as URHI, most baselines restore only portions of the scene, while FAPE-IR more completely removes the haze veil and corrects illumination. For super-resolution, FAPE-IR maintains the highest fidelity to GT with fewer artifacts, preserving textures and text without over-sharpening. These observations are consistent with the quantitative trends. Overall, FAPE-IR suppresses artifacts robustly, maintains strong pixel anchoring, and improves the realism of fine details.

### 4.5. Ablation Study

Under a controlled setting where we fix all hyperparameters and train for only 10K steps, we perform a minimal ablation of the frequency-aware pipeline. As shown in Table 4, on URHI (low-frequency / haze-dominant), the baseline without Qwen, Freq-U, or Freq-G reaches only 25.03 dB / 0.92 SSIM. Simply adding Qwen2.5-VL without any routing control improves performance to 27.95 dB / 0.94 SSIM, indicating that semantic planning is beneficial but insufficient to yield effective compute allocation. Coupling planner outputs to the MoE gate (Freq-U) further stabilizes performance at around 28.9 dB / 0.94–0.95 SSIM, showing that the model can already autonomously perceive and select frequency bands. Injecting token-derived spectral priors into the gate (Freq-G) yields the best result of 29.71 dB / 0.95 SSIM, corresponding to a +4.68 dB PSNR gain over the no-planner baseline and noticeably more consistent

**Table 4.** Ablation on URHI benchmark. Qwen: remove Qwen2.5-VL; Freq-U: remove frequency-aware text router; Freq-G: remove FIR spectral router;  $r$ : LoRA rank in expert combos.

|                  | Qwen | Freq-U | Freq-G | $r=4$ | $r=8$ | $r=16$ | PSNR $\uparrow$ | SSIM $\uparrow$ |
|------------------|------|--------|--------|-------|-------|--------|-----------------|-----------------|
|                  | ✗    | ✗      | ✗      |       | ✓     |        | 25.03           | 0.92            |
|                  | ✓    | ✗      | ✗      |       | ✓     |        | 27.95           | 0.94            |
| + Freq-U enabled |      |        |        |       |       |        |                 |                 |
|                  | ✓    | ✓      | ✗      |       | ✓     |        | 28.92           | 0.94            |
|                  | ✓    | ✓      | ✗      |       | ✓✓    |        | 28.89           | 0.95            |
|                  | ✓    | ✓      | ✗      | ✓     |       | ✓      | 24.75           | 0.91            |
|                  | ✓    | ✓      | ✗      |       | ✓     | ✓      | 24.77           | 0.92            |
| + Freq-G enabled |      |        |        |       |       |        |                 |                 |
|                  | ✓    | ✓      | ✓      |       | ✓✓    |        | 29.71           | 0.95            |
|                  | ✓    | ✓      | ✓      | ✓     |       | ✓      | 25.93           | 0.93            |
|                  | ✓    | ✓      | ✓      |       | ✓     | ✓      | 29.60           | 0.95            |

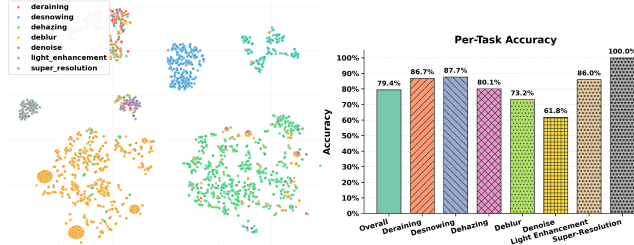


**Figure 7.** Low-level task planning. (a) The frequency-aware planner yields separable, spectrum-aligned manifolds in feature space using only its own hidden states. (b) Text-based task classification accuracy of Qwen2.5-VL on the corresponding tasks.

behavior across configurations.

We also study the impact of LoRA–MoE capacity ( $r$ ) and expert allocation. With no frequency priors, a balanced capacity (8/8) is the most stable. Under the Freq-G constraint, a mildly asymmetric allocation (8/16), aligned with the planner-indicated dominant band, maintains near-optimal performance (29.60 dB). In contrast, extremely skewed settings (4/16) degrade performance both with and without priors (down to 25.93 / 24.75 dB), suggesting that simply reshaping the LoRA structure is insufficient to disentangle frequency bands. Instead, the coupling of frequency priors with semantic planning is key to achieving stable routing and overall performance gains.

#### 4.6. Low-Level Task Planning

We assess whether the planner’s decisions form a task- and frequency-aware representation that is both separable and useful. Specifically, we freeze the Qwen2.5-VL planner, extract the decision token sequence  $FP$  for each input (BSD68-50, URHI, GoPro, LOL-v2, RainDrop, Snow100K-S, RealSR $\times 2$ ), and mean-pool their final-layer states. The resulting t-SNE embedding (Figure 7) exhibits clean, task-dependent manifolds. Nevertheless, the text-based task readout achieves 79.4% accuracy overall (all images in BSD68-50 are grayscale, which can spuriously trigger low-light flags). Moreover, the exemplars (Figure 8) show that the planner’s outputs not only describe but also instantiate a causal frequency control path for restoration.

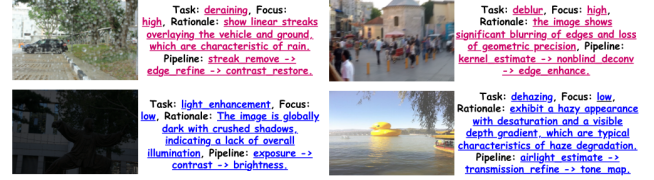


**Figure 8.** Frequency-aware planner outputs. We visualize results on representative corrupted inputs covering four tasks.

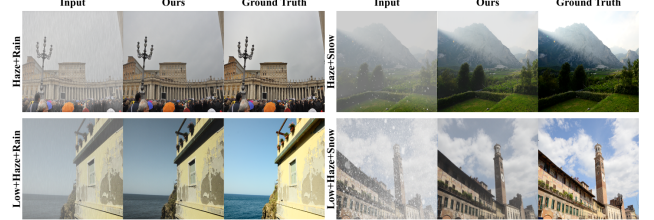


**Figure 9.** Compound degradations (haze+rain/snow; low-light mixtures): FAPE-IR removes low-frequency artifacts, preserves details, and reduces cross-artifacts. Please zoom in for details.

Collectively, these analyses demonstrate that the planner provides interpretable, frequency-aligned, and causally effective signals that steer band-specialized restoration, explaining the sharper, artifact-averse behavior.

#### 4.7. Generalization to Multi-Degradation Tasks

To assess generalization beyond single degradations, we evaluate FAPE-IR on the recent CDD-11 benchmark [21], which contains real-world images with multi-degradation tasks. As shown in Figure 9, although FAPE-IR is trained exclusively on single-degradation data, it exhibits strong out-of-distribution performance on these compound mixtures, indicating a non-trivial level of degradation-agnostic generalization. We attribute this to the frequency-aware LoRA–MoE module, which in a single forward pass dispatches complementary experts, allocating low-frequency capacity to illumination correction and veil removal while activating high-frequency capacity around edges and fine textures, consistent with the layered nature of degradations. Overall, the results substantiate FAPE-IR’s robustness and its ability to generalize to unseen composite degradations.

### 5. Conclusion

We introduce FAPE-IR, a unified AIO-IR framework that couples an interpretable, frequency-aware planner with a diffusion-based executor equipped with the frequency-aware LoRA–MoE module, trained under the adversarial training objective with a single frequency regularizer. By making the semantics–spectrum link explicit, FAPE-IR improves cross-task disentanglement and sharing, stabilizes optimization, and improves open-world robustness. It achieves state-of-the-art or competitive performance across seven AIO-IR tasks and shows strong zero-shot generalization on unseen composite degradations.

## 6. Acknowledgements

This work was supported by National Natural Science Foundation of China (NSFC) under Grants 62441235 and 62176178.

## References

- [1] Yuang Ai, Huaibo Huang, Xiaoqiang Zhou, Jiexiang Wang, and Ran He. Multimodal prompt perceiver: Empower adaptiveness generalizability and fidelity for all-in-one image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25432–25444, 2024. 1
- [2] Luigi Ambrosio, Elia Brué, Daniele Semola, et al. *Lectures on optimal transport*. Springer, 2021. 1, 6, 7, 8
- [3] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *International conference on machine learning*, pages 214–223. PMLR, 2017. 1, 6, 8
- [4] Shuai Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, Sibao Song, Kai Dang, Peng Wang, Shijie Wang, Jun Tang, et al. Qwen2. 5-vl technical report. *arXiv preprint arXiv:2502.13923*, 2025. 3
- [5] Samy Bengio, Oriol Vinyals, Navdeep Jaitly, and Noam Shazeer. Scheduled sampling for sequence prediction with recurrent neural networks. *Advances in neural information processing systems*, 28, 2015. 4, 8
- [6] Ashish Bora, Ajil Jalal, Eric Price, and Alexandros G Dimakis. Compressed sensing using generative models. In *International conference on machine learning*, pages 537–546. PMLR, 2017. 1, 6
- [7] Haowei Chen, Zhiwen Yang, Haotian Hou, Hui Zhang, Bingzheng Wei, Gang Zhou, and Yan Xu. All-in-one medical image restoration with latent diffusion-enhanced vector-quantized codebook prior. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 67–77. Springer, 2025. 1
- [8] 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*, pages 17969–17979, 2025. 1, 2
- [9] Jiuhai Chen, Zhiyang Xu, Xichen Pan, Yushi Hu, Can Qin, Tom Goldstein, Lifu Huang, Tianyi Zhou, Saining Xie, Silvio Savarese, et al. Blip3-o: A family of fully open unified multimodal models-architecture, training and dataset. *arXiv preprint arXiv:2505.09568*, 2025. 2, 3, 5
- [10] Xiangyu Chen, Yihao Liu, Yuandong Pu, Wenlong Zhang, Jiantao Zhou, Yu Qiao, and Chao Dong. Learning a low-level vision generalist via visual task prompt. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 2671–2680, 2024. 1
- [11] Xiangyu Chen, Kaiwen Zhu, Yuandong Pu, Shuo Cao, Xiaohui Li, Wenlong Zhang, Yihao Liu, Yu Qiao, Jiantao Zhou, and Chao Dong. Exploring scalable unified modeling for general low-level vision. *arXiv preprint arXiv:2507.14801*, 2025. 1
- [12] Yunpeng Chen, Haoqi Fan, Bing Xu, Zhicheng Yan, Yan-nis Kalantidis, Marcus Rohrbach, Shuicheng Yan, and Jiashi Feng. Drop an octave: Reducing spatial redundancy in convolutional neural networks with octave convolution. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3435–3444, 2019. 4
- [13] Zihan Cheng, Liangtai Zhou, Dian Chen, Ni Tang, Xiaotong Luo, and Yanyun Qu. Unildiff: Unlocking the power of diffusion priors for all-in-one image restoration. *arXiv preprint arXiv:2507.23685*, 2025. 2
- [14] Marcos V Conde, Gregor Geigle, and Radu Timofte. Instructir: High-quality image restoration following human instructions. In *European Conference on Computer Vision*, pages 1–21. Springer, 2024. 2
- [15] Yuning Cui, Syed Waqas Zamir, Salman Khan, Alois Knoll, Mubarak Shah, and Fahad Shahbaz Khan. Adair: Adaptive all-in-one image restoration via frequency mining and modulation. *arXiv preprint arXiv:2403.14614*, 2024. 2, 6, 9, 10, 11
- [16] Yufeng Cui, Honghao Chen, Haoge Deng, Xu Huang, Xinghang Li, Jirong Liu, Yang Liu, Zhuoyan Luo, Jinsheng Wang, Wenxuan Wang, et al. Emu3. 5: Native multimodal models are world learners. *arXiv preprint arXiv:2510.26583*, 2025. 7, 12
- [17] Chaorui Deng, Deyao Zhu, Kunchang Li, Chenhui Gou, Feng Li, Zeyu Wang, Shu Zhong, Weihao Yu, Xiaonan Nie, Ziang Song, et al. Emerging properties in unified multimodal pretraining. *arXiv preprint arXiv:2505.14683*, 2025. 2, 3, 5, 7, 12
- [18] Chengyu Fang, Chunming He, Fengyang Xiao, Yulun Zhang, Longxiang Tang, Yuelin Zhang, Kai Li, and Xiu Li. Real-world image dehazing with coherence-based pseudo labeling and cooperative unfolding network. *Advances in Neural Information Processing Systems*, 37:97859–97883, 2024. 5
- [19] David J Field. Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A*, 4(12):2379–2394, 1987. 1
- [20] Jinpei Guo, Zheng Chen, Wenbo Li, Yong Guo, and Yulun Zhang. Compression-aware one-step diffusion model for jpeg artifact removal. *arXiv preprint arXiv:2502.09873*, 2025. 1
- [21] Yu Guo, Yuan Gao, Yuxu Lu, Huilin Zhu, Ryan Wen Liu, and Shengfeng He. Onerestore: A universal restoration framework for composite degradation. In *European conference on computer vision*, pages 255–272. Springer, 2024. 8
- [22] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5197–5206, 2015. 6
- [23] Aapo Hyvärinen and Peter Dayan. Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6(4), 2005. 1, 6, 8

- [24] Aapo Hyvärinen, Jarmo Hurri, and Patrick O Hoyer. *Natural image statistics: A probabilistic approach to early computational vision*. Springer Science & Business Media, 2009. 1
- [25] Yuhwan Jeong, Yunseo Yang, Youngho Yoon, and Kuk-Jin Yoon. Robust adverse weather removal via spectral-based spatial grouping. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 11872–11883, 2025. 1
- [26] Xiaozhong Ji, Yun Cao, Ying Tai, Chengjie Wang, Jilin Li, and Feiyue Huang. Real-world super-resolution via kernel estimation and noise injection. In *proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 466–467, 2020. 6
- [27] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4401–4410, 2019. 5
- [28] Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. *Advances in neural information processing systems*, 35:23593–23606, 2022. 1
- [29] Dilip Krishnan and Rob Fergus. Fast image deconvolution using hyper-laplacian priors. *Advances in neural information processing systems*, 22, 2009. 1
- [30] Black Forest Labs. Flux. <https://github.com/black-forest-labs/flux>, 2024. 2, 4
- [31] Guanzhou Lan, Qianli Ma, Yuqi Yang, Zhigang Wang, Dong Wang, Xuelong Li, and Bin Zhao. Efficient diffusion as low light enhancer. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 21277–21286, 2025. 1
- [32] Anat Levin, Yair Weiss, Fredo Durand, and William T Freeman. Understanding and evaluating blind deconvolution algorithms. In *2009 IEEE conference on computer vision and pattern recognition*, pages 1964–1971. IEEE, 2009. 1
- [33] 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. 5
- [34] Hao Li, Xiang Chen, Jiangxin Dong, Jinhui Tang, and Jinshan Pan. Foundir: Unleashing million-scale training data to advance foundation models for image restoration. In *ICCV*, 2025. 6, 9, 10, 11
- [35] Ruoteng Li, Loong-Fah Cheong, and Robby T Tan. Heavy rain image restoration: Integrating physics model and conditional adversarial learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1633–1642, 2019. 5
- [36] Shufan Li, Konstantinos Kallidromitis, Hritik Bansal, Akash Gokul, Yusuke Kato, Kazuki Kozuka, Jason Kuen, Zhe Lin, Kai-Wei Chang, and Aditya Grover. Lavidia: A large diffusion language model for multimodal understanding. *arXiv preprint arXiv:2505.16839*, 2025. 3
- [37] Xin Li, Yeying Jin, Xin Jin, Zongwei Wu, Bingchen Li, Yufei Wang, Wenhan Yang, Yu Li, Zhibo Chen, Bihan Wen, et al. Ntire 2025 challenge on day and night raindrop removal for dual-focused images: Methods and results. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 1172–1183, 2025. 5
- [38] Yawei Li, Kai Zhang, Jingyun Liang, Jie Zhang Cao, Ce Liu, Rui Gong, Yulun Zhang, Hao Tang, Yun Liu, Denis Demandolx, et al. Lsdir: A large scale dataset for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1775–1787, 2023. 5
- [39] Zilong Li, Yiming Lei, Chenglong Ma, Junping Zhang, and Hongming Shan. Prompt-in-prompt learning for universal image restoration. *arXiv preprint arXiv:2312.05038*, 2023. 1, 2
- [40] Bin Lin, Zongjian Li, Xinhua Cheng, Yuwei Niu, Yang Ye, Xianyi He, Shenghai Yuan, Wangbo Yu, Shaodong Wang, Yuniang Ge, et al. Uniworld: High-resolution semantic encoders for unified visual understanding and generation. *arXiv preprint arXiv:2506.03147*, 2025. 2, 3, 5, 7, 12
- [41] Xinqi Lin, Jingwen He, Ziyang Chen, Zhaoyang Lyu, Bo Dai, Fanghua Yu, Yu Qiao, Wanli Ouyang, and Chao Dong. Diffbir: Toward blind image restoration with generative diffusion prior. In *European conference on computer vision*, pages 430–448. Springer, 2024. 6, 9, 10, 11
- [42] Yunlong Lin, Zixu Lin, Haoyu Chen, Panwang Pan, Chenxin Li, Sixiang Chen, Kairun Wen, Yeying Jin, Wenbo Li, and Xinghao Ding. Jarvisir: Elevating autonomous driving perception with intelligent image restoration. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 22369–22380, 2025. 1
- [43] Tony Lindeberg. Feature detection with automatic scale selection. *International journal of computer vision*, 30(2): 79–116, 1998. 1
- [44] Torngny Lindvall. *Lectures on the coupling method*. Courier Corporation, 2002. 8
- [45] Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022. 5, 1, 8, 13
- [46] Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. *arXiv preprint arXiv:2209.03003*, 2022. 1, 8
- [47] Yun-Fu Liu, Da-Wei Jaw, Shih-Chia Huang, and Jenq-Neng Hwang. Desnownet: Context-aware deep network for snow removal. *IEEE Transactions on Image Processing*, 27(6): 3064–3073, 2018. 5
- [48] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017. 6
- [49] Wenyang Luo, Haina Qin, Zewen Chen, Libin Wang, Dandan Zheng, Yuming Li, Yufan Liu, Bing Li, and Weiming Hu. Visual-instructed degradation diffusion for all-in-one image restoration. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 12764–12777, 2025. 1
- [50] Ziwei Luo, Fredrik K Gustafsson, Zheng Zhao, Jens Sjölund, and Thomas B Schön. Controlling vision-language models for universal image restoration. *arXiv preprint arXiv:2310.01018*, 3(8), 2023. 1, 2

- [51] Jiaqi Ma, Tianheng Cheng, Guoli Wang, Qian Zhang, Xinggang Wang, and Lefei Zhang. Prores: Exploring degradation-aware visual prompt for universal image restoration. *arXiv preprint arXiv:2306.13653*, 2023. 1, 2
- [52] Kede Ma, Zhengfang Duanmu, Qingbo Wu, Zhou Wang, Hongwei Yong, Hongliang Li, and Lei Zhang. Waterloo exploration database: New challenges for image quality assessment models. *IEEE Transactions on Image Processing*, 26(2):1004–1016, 2016. 5
- [53] Aziz Makandar and Bhagirathi Halalli. Image enhancement techniques using highpass and lowpass filters. *International Journal of Computer Applications*, 109(14), 2015. 4
- [54] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings eighth IEEE international conference on computer vision. ICCV 2001*, pages 416–423. IEEE, 2001. 5, 6
- [55] Konstantin Mishchenko and Aaron Defazio. Prodigy: An expeditiously adaptive parameter-free learner. *arXiv preprint arXiv:2306.06101*, 2023. 6
- [56] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3883–3891, 2017. 5
- [57] Vaishnav Potlapalli, Syed Waqas Zamir, Salman Khan, and Fahad Khan. Promptir: Prompting for all-in-one image restoration. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. 1, 2
- [58] Vaishnav Potlapalli, Syed Waqas Zamir, Salman Khan, and Fahad Khan. Promptir: Prompting for all-in-one image restoration. In *Advances in Neural Information Processing Systems*, 2023. 2, 6, 9, 10, 11
- [59] Yuandong Pu, Le Zhuo, Kaiwen Zhu, Liangbin Xie, Wenlong Zhang, Xiangyu Chen, Peng Gao, Yu Qiao, Chao Dong, and Yihao Liu. Lumina-omniv: A unified multimodal framework for general low-level vision. *arXiv preprint arXiv:2504.04903*, 2025. 1
- [60] Rui Qian, Robby T Tan, Wenhan Yang, Jiajun Su, and Jiaying Liu. Attentive generative adversarial network for rain-drop removal from a single image. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2482–2491, 2018. 5
- [61] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR, 2021. 2
- [62] Marc’Aurelio Ranzato, Sumit Chopra, Michael Auli, and Wojciech Zaremba. Sequence level training with recurrent neural networks. *arXiv preprint arXiv:1511.06732*, 2015. 4, 8
- [63] Mengwei Ren, Mauricio Delbracio, Hossein Talebi, Guido Gerig, and Peyman Milanfar. Multiscale structure guided diffusion for image deblurring. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10721–10733, 2023. 1
- [64] Jaesung Rim, Haeyun Lee, Jucheol Won, and Sunghyun Cho. Real-world blur dataset for learning and benchmarking deblurring algorithms. In *European conference on computer vision*, pages 184–201. Springer, 2020. 5
- [65] Yaniv Romano, Michael Elad, and Peyman Milanfar. The little engine that could: Regularization by denoising (red). *SIAM journal on imaging sciences*, 10(4):1804–1844, 2017. 1, 6
- [66] Stéphane Ross, Geoffrey Gordon, and Drew Bagnell. A reduction of imitation learning and structured prediction to no-regret online learning. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 627–635. JMLR Workshop and Conference Proceedings, 2011. 4, 8
- [67] Daniel Ruderman and William Bialek. Statistics of natural images: Scaling in the woods. *Advances in neural information processing systems*, 6, 1993. 1
- [68] Zhenning Shi, Chen Xu, Changsheng Dong, Bin Pan, Along He, Tao Li, Huazhu Fu, et al. Resfusion: Denoising diffusion probabilistic models for image restoration based on prior residual noise. *Advances in Neural Information Processing Systems*, 37:130664–130693, 2024. 1
- [69] Eero P Simoncelli and Bruno A Olshausen. Natural image statistics and neural representation. *Annual review of neuroscience*, 24(1):1193–1216, 2001. 1
- [70] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020. 1, 6, 8
- [71] Jon Sparring, Mads Nielsen, Luc Florack, and Peter Johansen. *Gaussian scale-space theory*. Springer Science & Business Media, 2013. 1
- [72] Bharath K Sriperumbudur, Kenji Fukumizu, Arthur Gretton, Bernhard Schölkopf, and Gert RG Lanckriet. On integral probability metrics,  $\phi$ -divergences and binary classification. *arXiv preprint arXiv:0901.2698*, 2009. 1, 6, 8
- [73] Quan Sun, Qiyang Yu, Yufeng Cui, Fan Zhang, Xiaosong Zhang, Yueze Wang, Hongcheng Gao, Jingjing Liu, Tiejun Huang, and Xinlong Wang. Emu: Generative pretraining in multimodality. *arXiv preprint arXiv:2307.05222*, 2023. 7
- [74] Quan Sun, Yufeng Cui, Xiaosong Zhang, Fan Zhang, Qiyang Yu, Yueze Wang, Yongming Rao, Jingjing Liu, Tiejun Huang, and Xinlong Wang. Generative multimodal models are in-context learners. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14398–14409, 2024. 5, 7
- [75] Alexander Swerdlow, Mihir Prabhudesai, Siddharth Gandhi, Deepak Pathak, and Katerina Fragkiadaki. Unified multimodal discrete diffusion. *arXiv preprint arXiv:2503.20853*, 2025. 3
- [76] Hongxuan Tang, Hao Liu, and Xinyan Xiao. Ugen: Unified autoregressive multimodal model with progressive vocabulary learning. *arXiv preprint arXiv:2503.21193*, 2025. 3
- [77] Chameleon Team. Chameleon: Mixed-modal early-fusion foundation models. *arXiv preprint arXiv:2405.09818*, 2024. 3, 5

- [78] Xiangpeng Tian, Xiangyu Liao, Xiao Liu, Meng Li, and Chao Ren. Degradation-aware feature perturbation for all-in-one image restoration. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 28165–28175, 2025. 1, 2, 6, 9, 10, 11
- [79] Michael Tschanen, Alexey Gritsenko, Xiao Wang, Muhammad Ferjad Naeem, Ibrahim Alabdulmohsin, Nikhil Parthasarathy, Talfan Evans, Lucas Beyer, Ye Xia, Basil Mustafa, et al. Siglip 2: Multilingual vision-language encoders with improved semantic understanding, localization, and dense features. *arXiv preprint arXiv:2502.14786*, 2025. 2
- [80] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 9446–9454, 2018. 1, 4
- [81] van A Van der Schaaf and JH van van Hateren. Modelling the power spectra of natural images: statistics and information. *Vision research*, 36(17):2759–2770, 1996. 1
- [82] Singanallur V Venkatakrishnan, Charles A Bouman, and Brendt Wohlberg. Plug-and-play priors for model based reconstruction. In *2013 IEEE global conference on signal and information processing*, pages 945–948. IEEE, 2013. 1, 6
- [83] Jianyi Wang, Zongsheng Yue, Shangchen Zhou, Kelvin CK Chan, and Chen Change Loy. Exploiting diffusion prior for real-world image super-resolution. *International Journal of Computer Vision*, 132(12):5929–5949, 2024. 6, 9, 10, 11
- [84] Runze Wang, Zeli Chen, Zhiyun Song, Wei Fang, Jiajin Zhang, Danyang Tu, Yuxing Tang, Minfeng Xu, Xianghua Ye, Le Lu, et al. Anatomy-aware low-dose ct denoising via pretrained vision models and semantic-guided contrastive learning. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 13–23. Springer, 2025. 1
- [85] 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*, pages 23091–23100, 2025. 1
- [86] Sen Wang, Shao Zeng, Tianjun Gu, Zhizhong Zhang, Ruixin Zhang, Shouhong Ding, Jingyun Zhang, Jun Wang, Xin Tan, Yuan Xie, et al. From enhancement to understanding: Build a generalized bridge for low-light vision via semantically consistent unsupervised fine-tuning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 13804–13814, 2025. 1
- [87] Tao Wang, Peiwen Xia, Bo Li, Peng-Tao Jiang, Zhe Kong, Kaihao Zhang, Tong Lu, and Wenhan Luo. Moerl: When mixture-of-experts meet reinforcement learning for adverse weather image restoration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 13673–13683, 2025. 1
- [88] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1905–1914, 2021. 5
- [89] Xinlong Wang, Xiaosong Zhang, Zhengxiong Luo, Quan Sun, Yufeng Cui, Jinsheng Wang, Fan Zhang, Yueze Wang, Zhen Li, Qiyang Yu, et al. Emu3: Next-token prediction is all you need. *arXiv preprint arXiv:2409.18869*, 2024. 3, 5, 7
- [90] Yongzhen Wang, Yongjun Li, Zhuoran Zheng, Xiao-Ping Zhang, and Mingqiang Wei. M2restore: Mixture-of-experts-based mamba-cnn fusion framework for all-in-one image restoration. *arXiv preprint arXiv:2506.07814*, 2025. 2
- [91] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018. 5
- [92] Hongyang Wei, Shuaizheng Liu, Chun Yuan, and Lei Zhang. Perceive, understand and restore: Real-world image super-resolution with autoregressive multimodal generative models. *arXiv preprint arXiv:2503.11073*, 2025. 5, 6, 7, 9, 10, 11
- [93] Pengxu Wei, Ziwei Xie, Hannan Lu, Zongyuan Zhan, Qixiang Ye, Wangmeng Zuo, and Liang Lin. Component divide-and-conquer for real-world image super-resolution. In *European conference on computer vision*, pages 101–117. Springer, 2020. 6
- [94] Andrew P Witkin. Scale-space filtering. In *Readings in computer vision*, pages 329–332. Elsevier, 1987. 1
- [95] Chengyue Wu, Xiaokang Chen, Zhiyu Wu, Yiyang Ma, Xingchao Liu, Zizheng Pan, Wen Liu, Zhenda Xie, Xingkai Yu, Chong Ruan, et al. Janus: Decoupling visual encoding for unified multimodal understanding and generation. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 12966–12977, 2025. 2, 3, 5
- [96] Gang Wu, Junjun Jiang, Yijun Wang, Kui Jiang, and Xianming Liu. Debiased all-in-one image restoration with task uncertainty regularization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 8386–8394, 2025. 1
- [97] Rongyuan Wu, Lingchen Sun, Zhiyuan Ma, and Lei Zhang. One-step effective diffusion network for real-world image super-resolution. *Advances in Neural Information Processing Systems*, 37:92529–92553, 2024. 5, 6, 9, 10, 11
- [98] Rongyuan Wu, Tao Yang, Lingchen Sun, Zhengqiang Zhang, Shuai Li, and Lei Zhang. Seesr: Towards semantics-aware real-world image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 25456–25467, 2024. 5, 6, 9, 10, 11
- [99] Jinheng Xie, Weijia Mao, Zechen Bai, David Junhao Zhang, Weihao Wang, Kevin Qinghong Lin, Yuchao Gu, Zhijie Chen, Zhenheng Yang, and Mike Zheng Shou. Show-o: One single transformer to unify multimodal understanding and generation. *arXiv preprint arXiv:2408.12528*, 2024. 2
- [100] Xinan Xie, Qing Zhang, and Wei-Shi Zheng. Diffusion-based event generation for high-quality image deblurring. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 2194–2203, 2025. 1

- [101] Shuning Xu, Binbin Song, Xiangyu Chen, Xina Liu, and Jiantao Zhou. Image demoiring in raw and srgb domains. In *European Conference on Computer Vision*, pages 108–124. Springer, 2024. [1](#)
- [102] Ling Yang, Ye Tian, Bowen Li, Xinchun Zhang, Ke Shen, Yunhai Tong, and Mengdi Wang. Mmada: Multimodal large diffusion language models. *arXiv preprint arXiv:2505.15809*, 2025. [2](#), [3](#)
- [103] Tao Yang, Rongyuan Wu, Peiran Ren, Xuansong Xie, and Lei Zhang. Pixel-aware stable diffusion for realistic image super-resolution and personalized stylization. In *European conference on computer vision*, pages 74–91. Springer, 2024. [6](#), [9](#), [10](#), [11](#)
- [104] Wenhan Yang, Robby T Tan, Jiashi Feng, Zongming Guo, Shuicheng Yan, and Jiaying Liu. Joint rain detection and removal from a single image with contextualized deep networks. *IEEE transactions on pattern analysis and machine intelligence*, 42(6):1377–1393, 2019. [5](#)
- [105] Wenhan Yang, Wenjing Wang, Haofeng Huang, Shiqi Wang, and Jiaying Liu. Sparse gradient regularized deep retinex network for robust low-light image enhancement. *IEEE Transactions on Image Processing*, 30:2072–2086, 2021. [5](#)
- [106] Tian Ye, Sixiang Chen, Wenhao Chai, Zhaohu Xing, Jing Qin, Ge Lin, and Lei Zhu. Learning diffusion texture priors for image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2524–2534, 2024. [1](#)
- [107] Xunpeng Yi, Han Xu, Hao Zhang, Linfeng Tang, and Jiayi Ma. Diff-retinex: Rethinking low-light image enhancement with a generative diffusion model. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 12302–12311, 2023. [1](#)
- [108] Runpeng Yu, Xinyin Ma, and Xinchao Wang. Dimple: Discrete diffusion multimodal large language model with parallel decoding. *arXiv preprint arXiv:2505.16990*, 2025. [3](#)
- [109] Eduard Zamfir, Zongwei Wu, Nancy Mehta, Yuedong Tan, Danda Pani Paudel, Yulun Zhang, and Radu Timofte. Complexity experts are task-discriminative learners for any image restoration. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 12753–12763, 2025. [2](#), [6](#), [9](#), [10](#), [11](#)
- [110] Cheng Zhang, Yu Zhu, Qingsen Yan, Jinqiu Sun, and Yanming Zhang. All-in-one multi-degradation image restoration network via hierarchical degradation representation. In *Proceedings of the 31st ACM international conference on multimedia*, pages 2285–2293, 2023. [2](#)
- [111] H Zhang, Z Duan, X Wang, Y Zhao, W Lu, Z Di, Y Xu, Y Chen, and Y Zhang. Nexus-gen: Unified image understanding, generation, and editing via prefilled autoregression in shared embedding space. *arXiv preprint arXiv:2504.21356*, 2025. [7](#), [12](#)
- [112] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE transactions on image processing*, 26(7):3142–3155, 2017. [1](#)
- [113] Dian Zheng, Xiao-Ming Wu, Shuzhou Yang, Jian Zhang, Jian-Fang Hu, and Wei-Shi Zheng. Selective hourglass mapping for universal image restoration based on diffusion model. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 25445–25455, 2024. [2](#)
- [114] Mo Zhou, Keren Ye, Mauricio Delbracio, Peyman Milanfar, Vishal M Patel, and Hossein Talebi. Unires: Universal image restoration for complex degradations. *arXiv preprint arXiv:2506.05599*, 2025. [1](#), [2](#)
- [115] Tianwen Zhou, Jing Wang, Songtao Wu, and Kuanhong Xu. Prodehaze: Prompting diffusion models toward faithful image dehazing. *arXiv preprint arXiv:2503.17488*, 2025. [1](#)