TOWARDS ALL-IN-ONE IMAGE RESTORATION WITH UNPAIRED TRAINING

Anonymous authors
Paper under double-blind review

ABSTRACT

All-in-one (AIO) image restoration (IR) is crucial for addressing the diverse and complex degradations in real-world scenarios. However, existing AIO IR models are primarily trained with synthetic paired supervision and therefore suffer from severe domain limitations when applied to real-world data. Unpaired learning offers a promising alternative but introduces a significant challenge: removing degradations while faithfully preserving image content without pixel-level paired supervision. This challenge is further amplified in the AIO IR setting. To address it, we propose UADiff, an unpaired AIO IR framework that fine-tunes a pre-trained diffusion model into a one-step restoration backbone. Specifically, UADiff resolves two fundamental issues: degradation adaptation and unpaired constraint. For degradation adaptation, we design a Degradation-Aware Prompt Learning (DAPL) module leveraging SigLIP to learn a set of textual prompts representing degradation and clean concepts. These prompts identify degradations and guide a Degradation-Conditional Low-Rank Adaptation (LoRA) module, achieving degradation-adaptive restoration. For the unpaired constraint, we propose Content-Preserving Regularization (CPR), which promotes contrastive learning on degradation-invariant content features and guides the restored results away from degradation concepts. Extensive experiments demonstrate that UADiff achieves superior restoration performance and generalization, establishing a competitive baseline for unpaired AIO IR in real-world scenarios.

1 INTRODUCTION

In real-world scenarios, images are often corrupted by multiple unknown and complex degradations, such as blur, haze, or rain, making the image restoration (IR) process highly challenging (Jiang et al., 2025; 2024; Ai et al., 2024a; Vaishnav et al., 2023). To address this, all-in-one (AIO) IR has emerged as a critical research direction that seeks to handle diverse degradation types within a single unified framework. As shown in Figure 1 (I), most existing AIO IR models are trained in a fully supervised manner (Jiang et al., 2025; Zamfir et al., 2025), requiring paired degraded–clean images for optimization. However, acquiring large-scale, real-world paired training data is often infeasible or extremely challenging (Safayani et al., 2025; Chen et al., 2022). Consequently, these models mainly rely on large-scale synthetic paired datasets. This reliance inevitably leads to a severe domain gap: models tend to overfit to simple synthetic degradations and exhibit a substantial drop in performance when applied to real-world data (Dong et al., 2025a; Wang et al., 2024b).

Unpaired learning (Han et al., 2021; Lee et al., 2018; Park et al., 2020; Zhu et al., 2017) offers a promising alternative as it enables training on real-world data by only requiring separate domains of degraded and clean images. However, unpaired IR methods are typically designed for single-task restoration (Safayani et al., 2025; Dong et al., 2025a; Ye et al., 2022; Wang et al., 2024c), resulting in considerable model redundancy and poor scalability in the AIO setting. In addition, unpaired IR itself is an inherently ill-posed problem: without pixel-level supervision, the model must remove degradations while faithfully preserving image content (Safayani et al., 2025; Lan et al., 2025; Dong et al., 2025a). This difficulty is further amplified under the AIO configuration, where the model must simultaneously distinguish and restore multiple heterogeneous degradations without ground-truth reference. Therefore, fulfilling the goal of AIO IR under unpaired conditions requires addressing two key challenges. The first is degradation adaptation, which involves equipping the model with the ability to adaptively process diverse degradations. The second is the unpaired constraint, which

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

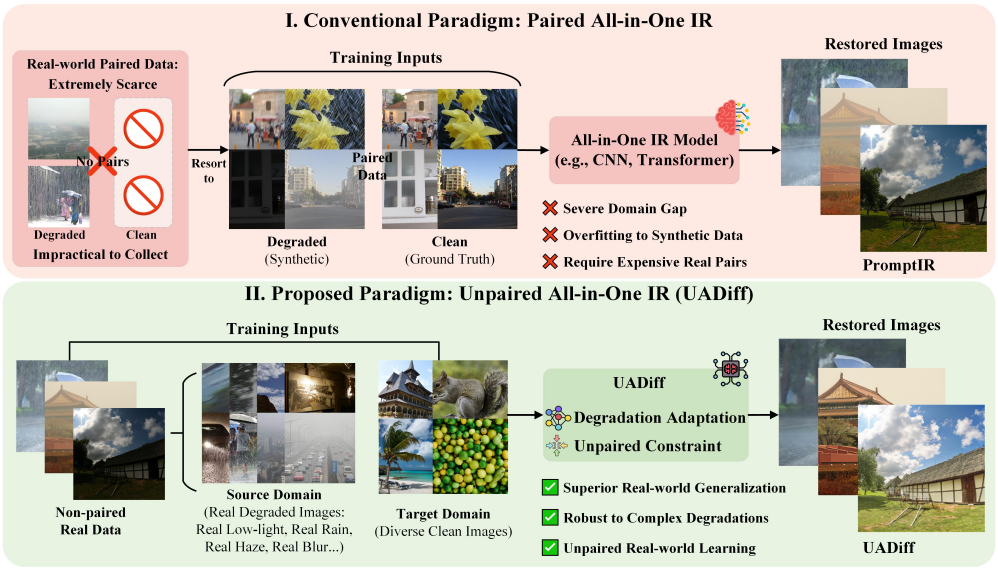

Figure 1: Comparison of different AIO IR paradigms. (I) Conventional AIO IR models rely heavily on synthetic paired data due to the scarcity of real-world pairs, resulting in a significant domain gap and limited generalization to complex real-world degradations. (II) In contrast, our UADiff eliminates the need for paired supervision. By leveraging unpaired real data with degradation adaptation and unpaired constraints, it achieves stronger real-world generalization and robust restoration.

involves establishing a robust regularization strategy that effectively resolves the inherent conflict between content preservation and degradation removal.

To address the challenges of unpaired AIO IR, we propose UADiff, a unified framework that leverages the prior from a pre-trained text-to-image (T2I) diffusion model (Rombach et al., 2022), fine-tuning it into a one-step restoration network. Specifically, UADiff integrates degradation-adaptive restoration and unpaired constraint within its framework. To achieve degradation adaptation, we introduce Degradation-Aware Prompt Learning (DAPL) to capture degradation semantics and provide adaptive guidance, together with Degradation-Conditional Restoration (DCR), which injects these cues into the denoising process via a conditional Low-Rank Adaptation (LoRA) (Hu et al., 2022) mechanism for degradation-aware restoration. To tackle the unpaired constraint, we propose Content-Preserving Regularization (CPR), a objective that encourages contrastive learning on degradation-invariant content features and guides results away from degradation semantics.

Our main contributions can be summarized as follows:

- We propose UADiff, an unpaired AIO IR framework that fine-tunes a pre-trained diffusion model into a one-step restoration network. Without pixel-aligned training pairs, UADiff achieves superior restoration performance and generalization to real-world degradations.
- We introduce a degradation-adaptive restoration strategy that integrates DAPL and DCR to accurately perceive degradation semantics and effectively adapt the restoration process to diverse degradation types and severity levels.
- We present CPR, a novel objective that promotes contrastive learning on degradation-invariant content features, preserving content while removing degradations.

2 Related Work

Image Restoration. IR focuses on recovering clean images from degraded observations (Jiang et al., 2025; Luo et al., 2024b; Tang et al., 2025). Traditional methods (Liang et al., 2021; Zamir et al., 2022; Wang et al., 2022) often target a single degradation, such as denoising or deblurring, limiting generalization to complex real-world degradations (Jiang et al., 2025; Vaishnav et al., 2023). To overcome this, AIO IR handles multiple degradations in a unified framework (Jiang et al., 2025; Zheng et al., 2024; Cui et al., 2025). Early work like AirNet (Li et al., 2022a) used contrastive

learning to map degradations to a shared feature space. Later studies, including PromptIR (Vaishnav et al., 2023) and ProRes (Ma et al., 2023), used prompt learning to represent degradation types and improve restoration. Methods such as InstructIR (Conde et al., 2024), AutoDIR (Jiang et al., 2024), DA-CLIP (Luo et al., 2024a), and MPerceiver (Ai et al., 2024a) incorporated language or multimodal priors for semantic understanding and improved generalization. However, most AIO IR approaches still rely on large-scale paired datasets, limiting robustness to real-world degradations. In contrast, unpaired learning (Zhu et al., 2017; Lee et al., 2018; Park et al., 2020; Han et al., 2021) enables training on real-world data with separate degraded and clean domains. Therefore, developing a unified degradation-adaptive model under unpaired supervision remains a critical challenge.

Unpaired Image Restoration. Unpaired IR reconstructs realistic images without pixel-level supervision when paired data are unavailable (Safayani et al., 2025). The challenge is removing degradations while preserving structural integrity without correspondence. Early work like CycleGAN (Zhu et al., 2017) introduced cycle-consistency loss to preserve content, but its one-to-one mapping constraint conflicted with the ill-posed nature of restoration. Later approaches such as MUNIT (Huang et al., 2018), DRIT++ (Lee et al., 2018), and CinCGAN (Yuan et al., 2018) improved training stability through latent-space disentanglement or weakly aligned pseudo-supervision. CUT (Park et al., 2020) further used a contrastive framework with PatchNCE loss for single-direction mapping, while DCLGAN (Han et al., 2021) and ODCR (Wang et al., 2024d) refined feature alignment with dual-encoder architectures and orthogonal decoupling regularization. Recent methods, including UNSB (Kim et al., 2024), FrDiff (Liu et al., 2025a), and TP-Diff (Liu et al., 2025b), applied probabilistic or diffusion-based models for unpaired restoration, improving stability and perceptual consistency. However, most of these approaches (Wang et al., 2024c; 2021; 2024b; Lan et al., 2025) focus on single degradation types. Achieving AIO, degradation-adaptive restoration under unpaired supervision remains a crucial challenge that has yet to be fully addressed.

Diffusion Prior-Based Image Restoration. Latent Diffusion Models (LDMs) (Rombach et al., 2022) demonstrate strong generative capability and have been widely adopted as priors in low-level vision tasks (Chen et al., 2025; Ai et al., 2024b; Wang et al., 2024a). Methods like StableSR (Wang et al., 2024a) and DiffBIR (Lin et al., 2024) employed time-aware encoders and two-stage frameworks for super-resolution and blind restoration, while SeeSR (Wu et al., 2024b) and SUPIR (Yu et al., 2024) integrated multimodal prompts for degradation-aware reconstruction. However, these approaches rely on multi-step sampling, leading to high computational costs. To improve efficiency, researchers have explored accelerated diffusion priors. OSEDiff (Wu et al., 2024a) demonstrated that low-quality images could be used as effective diffusion inputs, fine-tuning latent networks with LoRA (Hu et al., 2022) and aligning one-step and multi-step distributions using Variational Score Distillation (VSD). Building on this, OSDFace (Wang et al., 2025) applies the one-step paradigm for face restoration, while PiSA-SR (Sun et al., 2025) and TSD-SR (Dong et al., 2025b) enhance perceptual quality through dual-LoRA and score distillation. While these methods improve efficiency and perceptual fidelity, they still depend on large paired datasets and remain task-specific. An unpaired, and degradation-aware diffusion framework remains a crucial direction for future research.

3 METHODS

To address the challenges of unpaired AIO IR, we propose a unified framework that integrates degradation adaptation and unpaired constraint within a diffusion-based restoration paradigm. As illustrated in Figure 2, our method builds upon a pre-trained T2I diffusion model (Rombach et al., 2022) and adapts it into a one-step restoration backbone. Specifically, to achieve degradation-adaptive restoration, we introduce two modules: Degradation-Aware Prompt Learning (DAPL), which captures degradation semantics and provides adaptive guidance, and Degradation-Conditional Restoration (DCR), which injects these cues into the denoising process through degradation-conditional LoRA for flexible restoration. Furthermore, to ensure faithful reconstruction under unpaired supervision, we propose Content-Preserving Regularization (CPR), which employs adversarial and semantic constraints to balance degradation removal and content consistency.

3.1 DEGRADATION-AWARE PROMPT LEARNING

To handle AIO IR task, the model needs to develop degradation awareness to recognize and understand diverse degradation types such as blur, noise and haze. Motivated by the strong cross-modal

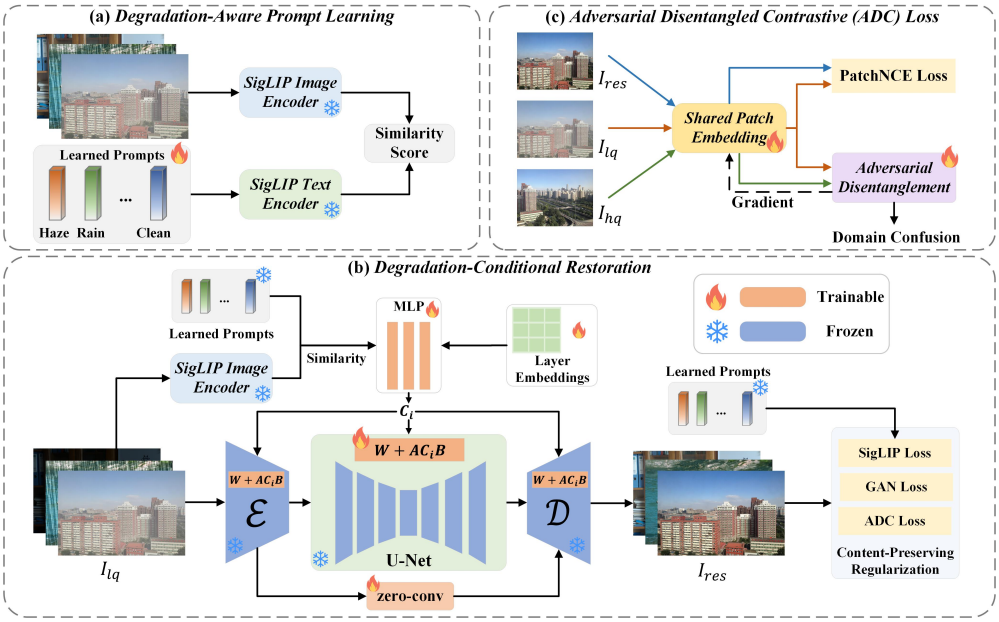

Figure 2: Overview of the proposed UADiff framework. (a) Degradation-Aware Prompt Learning (DAPL) leverages a frozen SigLIP encoder with learnable textual prompts to represent degradation and clean concepts. (b) Degradation-Conditional Restoration (DCR) computes similarity score between the input and the learned prompts, then fuses it with layer embeddings via an MLP to generate modulation coefficients C_i. These coefficients modulate the LoRA layers under the supervision of Content-Preserving Regularization (CPR). (c) Adversarial Disentangled Contrastive Loss ($\mathcal{L}_{adc}$) purifies the shared patch embedding through adversarial disentanglement so that the PatchNCE loss operates on degradation-invariant content features, ensuring robust content preservation.

representation and perceptual reasoning capabilities of vision-language models (Liang et al., 2023; Morawski et al., 2024; Lan et al., 2025), we explore leveraging SigLIP (Zhai et al., 2023) to identify and characterize different degradation patterns. However, as shown in Figure 3, using fixed textual prompts such as "hazy image" proves suboptimal, as handcrafted descriptions fail to capture the fine-grained and continuous nature of real-world degradations that vary in both severity and type.

To address this limitation, we propose a Degradation-Aware Prompt Learning (DAPL) module that replaces fixed textual descriptions with a set of learnable textual prompts. The module consists of $N + 1$ prompt embeddings: N degradation prompts (e.g., $T_{blur}, T_{noise}, \dots$) corresponding to the primary degradation categories, and one clean prompt T_{clean} representing the target domain. As illustrated in Figure 2 (a), we leverage the pre-trained SigLIP image encoder Φ_{image} and text encoder Φ_{text} to separately extract features from degraded images or clean images and the initialized prompt set T. Unlike softmax-based contrastive learning (Radford et al., 2021) that enforces mutual exclusivity, we optimize the prompts using a pairwise Sigmoid cross-entropy loss, enabling independent modeling of multiple co-existing degradations:

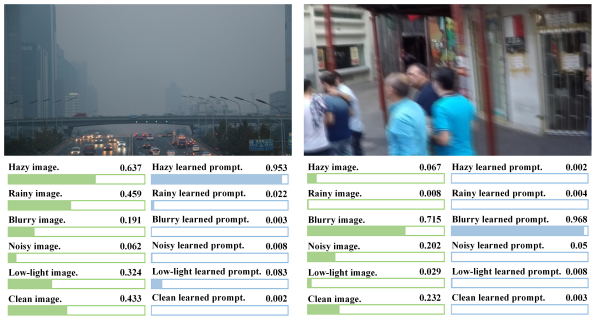

Figure 3: Comparison of scores produced by fixed and learned prompts. While fixed prompts struggle to capture complex real-world degradations, learned prompts provide robust perception of diverse degradation types.

$$\mathcal{L}_{prompt} = -\frac{1}{M}\sum_{i=1}^{M}\sum_{k=1}^{N+1}\log\frac{1}{1+e^{-z_{i,k}(t\cdot(\Phi_{image}(I_i)\cdot\Phi_{text}(T_k))+b)}}, \quad (1)$$

where I_i is the i-th input image and T_k represents the k-th prompt embedding. $z_{i,k} \in \{1, -1\}$ indicates whether concept T_k is present in image I_i. t and b are learnable scaling and bias terms in SigLIP, and M is the batch size.

After optimization, the learned prompts serve two purposes in the framework. For restoration guidance, the similarity score between image features and prompt embeddings is computed as

$$S_{i,k} = \frac{1}{1 + e^{-(t \cdot (\Phi_{image}(I_i) \cdot \Phi_{text}(T_k)) + b)}}, \tag{2}$$

and the resulting scores collectively define a degradation condition c_{deg} that guides adaptive restoration. Furthermore, under unpaired supervision, the prompts act as semantic regularizers. A SigLIP-based loss $\mathcal{L}_{siglip}$ encourages the restored image to approach the clean concept T_{clean} while being pushed away from all identified degradation concepts captured in c_{deg}, promoting perceptually clean outputs without pixel-level supervision.

3.2 Degradation-Conditional Restoration

To leverage the strong generative prior of large-scale diffusion models and enhance restoration stability, our framework builds upon a pre-trained T2I diffusion model (Rombach et al., 2022). Unlike conventional diffusion models (Lin et al., 2024; Wang et al., 2024a; Wu et al., 2024b), our method completes restoration through a single forward pass. Similar to recent one-step diffusion studies (Wu et al., 2024a; Zhang et al., 2024; Wang et al., 2025), we also regard the low-quality (LQ) image as an effective starting point of the reverse diffusion process. Consequently, the restoration process is formulated as a single-step denoising operation.

Given an LQ image I_{lq}, VAE encoder $\mathcal{E}$ produces a latent representation z_{lq}, which is fed into denoising network ϵ_θ with timestep t and degradation condition c_{deg} to predict noise $\epsilon = \epsilon_\theta(z_{lq}, t, c_{deg})$. We then apply one-step reverse diffusion to estimate clean latent representation $\hat{z}_0$:

$$\hat{z}_0 = \frac{z_{lq} - \sqrt{1 - \overline{\alpha}_t}\epsilon}{\sqrt{\overline{\alpha}_t}}, \tag{3}$$

where $\overline{\alpha}_t$ denotes the noise scheduler of the forward diffusion process. Finally, VAE decoder $\mathcal{D}$ reconstructs the restored image $I_{res} = \mathcal{D}(\hat{z}_0)$. However, a single fixed-step network cannot effectively handle the diversity of real-world degradations. To enable adaptive restoration, we introduce Degradation-Conditional Restoration (DCR), as shown in Figure 2 (b). This process is achieved by incorporating a degradation-conditional LoRA mechanism into the pre-trained diffusion model.

In vanilla LoRA, weights are updated as $W' = W + AB$ with low-rank matrices A and B, resulting in applying the same parameters to different types of degradations and making it difficult to adapt to real-world scenarios. To enable degradation-aware adaptation, we condition the LoRA update on the degradation representation obtained from the DAPL module. Specifically, the degradation condition c_{deg} is first projected into an embedding vector l_{deg}. Meanwhile, each network layer is associated with a learnable layer embedding $\mathcal{L} = \{l_i\}_{i=1}^{K}$. For the i-th layer, we concatenate its layer embedding l_i with the degradation embedding l_{deg} and feed the result into a shared MLP to generate a modulation matrix C_i:

$$C_i = \text{MLP}(\text{concat}(l_{deg}, l_i)), \tag{4}$$

the conditional LoRA update is formulated as:

$$W' = W + A(C_i B), \tag{5}$$

where C_i serves as a modulation coefficient that dynamically transforms the LoRA layers according to the predicted degradation condition. This design enables each network layer to adapt its transformation behavior to varying degradation types, achieving degradation-adaptive restoration.

3.3 Content-Preserving Regularization

To effectively train our framework under the unpaired and AIO settings, we design a composite loss function that resolves the intrinsic conflict between content preservation and degradation removal.

Unlike previous approaches (Liu et al., 2025b; Lan et al., 2025; Dong et al., 2025a) that rely on degradation-specific priors, our goal is to establish a unified training objective that can generalize to diverse degradation types.

The fundamental challenge in unpaired IR is maintaining content consistency while avoiding the preservation of degradation cues. CUT-based contrastive learning (Park et al., 2020) maximizes the mutual information between the degraded input I_{lq} and the restored output I_{res} through a PatchNCE loss ($\mathcal{L}_{patchnce}$). However, degradation patterns such as rain or noise are inherently embedded in this mutual information, which causes a contradiction: the network aims to remove degradations, whereas the contrastive loss implicitly encourages their retention. To address this issue, we propose the Adversarial Disentangled Contrastive Loss ($\mathcal{L}_{adc}$), which purifies the feature space before contrastive learning. As shown in Figure 2 (c), an adversarial disentanglement discriminator D_{cls} is trained to distinguish between features extracted by the shared patch embedding (SPE) from clean images I_{hq} and degraded images I_{lq}. In parallel, the SPE is optimized adversarially to prevent D_{cls} from distinguishing the domains, thereby producing degradation-invariant representations. After this purification process, $\mathcal{L}_{patchnce}$ is applied between I_{lq} and I_{res}, ensuring that the contrastive constraint focuses exclusively on structural consistency rather than degradation patterns. This adversarial disentanglement effectively balances degradation removal and content preservation, enabling stable optimization under unpaired supervision.

In addition to $\mathcal{L}_{adc}$, we introduce two auxiliary losses to enhance realism and stability. The adversarial loss ($\mathcal{L}_{gan}$) uses a standard GAN formulation to ensure that the restored image I_{res} matches the distribution of real clean images, thereby improving photorealism and perceptual fidelity. The semantic loss $\mathcal{L}_{siglip}$ leverages the learned prompts to guide the restored image toward the clean concept T_{clean} while pushing it away from the degradation concept T_{deg} in the SigLIP feature space. This non-pixel-level constraint reinforces semantic consistency and facilitates degradation removal. The complete objective function is defined as:

$$\mathcal{L} = \lambda_{gan}\mathcal{L}_{gan} + \lambda_{adc}\mathcal{L}_{adc} + \lambda_{siglip}\mathcal{L}_{siglip}. \tag{6}$$

4 EXPERIMENTS

To comprehensively evaluate the performance of the proposed framework, we conduct experiments on several publicly available benchmark datasets, covering standard AIO IR benchmarks and real-world evaluation. Specifically, we test on five representative restoration tasks: denoising, dehazing, deraining, motion deblurring, and low-light enhancement.

Datasets. For unpaired training, the degradation domain combines datasets across five degradation types: Rain800 (Zhang et al., 2019), DDN-Data (Fu et al., 2017), and RealRain (Li et al., 2022b) for deraining; GoPro (Nah et al., 2017), RealBlur (Rim et al., 2020), and HIDE (Shen et al., 2019) for deblurring; RESIDE (Li et al., 2018) for dehazing; LOL-v2 (Yang et al., 2021) for low-light enhancement; and WED (Ma et al., 2016), BSD400 (Arbelaez et al., 2010), Urban100 (Huang et al., 2015), and SIDD (Abdelhamed et al., 2018) for denoising. The clean domain consists of corresponding clean images from these datasets along with images from DIV2K (Agustsson & Timofte, 2017). For standard benchmark evaluation, we adopt paired test sets from Test100 (Zhang et al., 2019) (deraining), SOTS (Li et al., 2018) (dehazing), BSD68 (Martin et al., 2001) (denoising), LOL-v2 (Yang et al., 2021) (low-light), and GoPro (Nah et al., 2017) (deblurring). For real-world evaluation, we further assess model robustness on datasets captured in real-world scenarios, including SSID (Huang et al., 2022) (deraining), RENOIR (Anaya & Barbu, 2018) (denoising), Haze2020 (Shao et al., 2020) and Fattal's dataset (Fattal, 2014) (dehazing), RB2V-Street (Pham et al., 2023) (deblurring), and DICM (Guo et al., 2016) and MEF (Ma et al., 2015) (low-light enhancement).

Metrics. For standard AIO IR benchmark, we employ PSNR and SSIM (Wang et al., 2004) to measure pixel-level fidelity, LPIPS (Zhang et al., 2018) to evaluate perceptual similarity. For real-world evaluation, where ground truth is unavailable, we adopt FID (Heusel et al., 2017) and MUSIQ (Ke et al., 2021) to evaluate visual realism and perceptual quality under real-scene conditions.

Implementation Details. Our framework is built upon the Stable Diffusion Turbo (Sauer et al., 2024) and trained using the AdamW optimizer with a learning rate of $1e^{-5}$. The degradation-conditional LoRA modules are configured with ranks of 32 for network layers and 16 for VAE layers. More details are presented in the supplementary material.

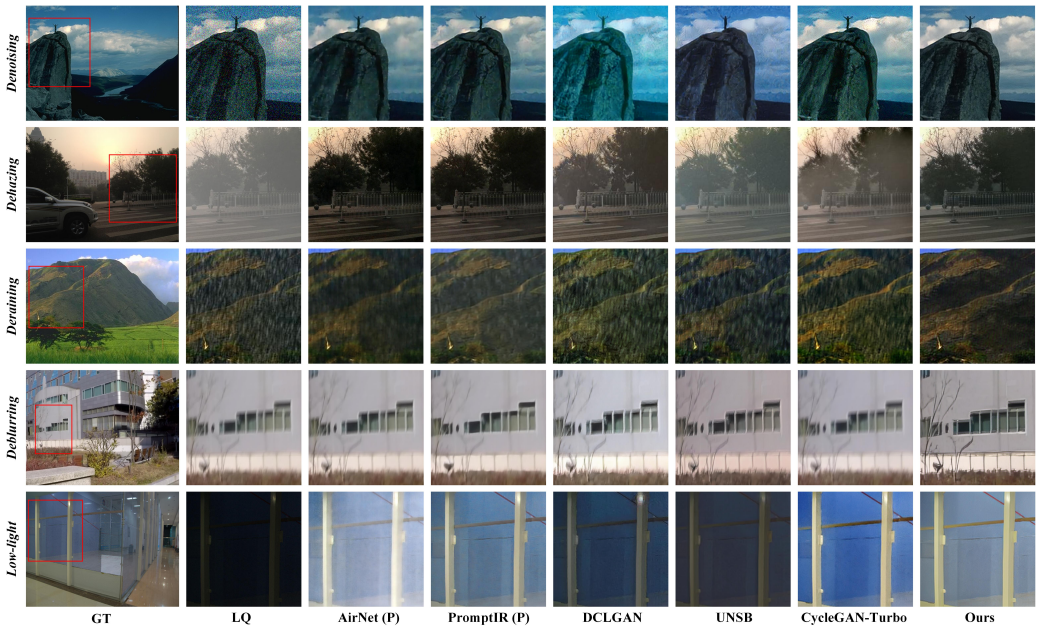

Figure 4: Visual comparison on standard AIO IR benchmarks. P denotes paired-training methods.

Table 1: Quantitative results on standard AIO IR benchmarks. Paired and Unpaired indicate training settings. Results are reported as PSNR↑ / SSIM↑ / LPIPS↓. In the Unpaired group, **bold** denotes the best and underline denotes the second-best performance.

	Method	*Dehazing*	*Deraining*	*Denoising*	*Low-light*	*Deblurring*
		SOTS	Test100	BSD68	LOL-v2	GoPro
Paired	AirNet	25.67 / 0.968 / 0.044	24.11 / 0.796 / 0.252	32.11 / 0.880 / 0.145	21.95 / 0.888 / 0.198	26.99 / 0.815 / 0.289
	PromptIR	29.57 / 0.976 / 0.041	27.31 / 0.857 / 0.169	31.85 / 0.884 / 0.136	24.34 / 0.917 / 0.115	27.77 / 0.830 / 0.275
	DA-RCOT	26.17 / 0.971 / 0.045	23.49 / 0.763 / 0.242	28.36 / 0.807 / 0.166	22.94 / 0.889 / 0.175	27.13 / 0.815 / 0.285
	DA-CLIP	27.73 / 0.975 / 0.022	24.75 / 0.814 / 0.159	29.76 / 0.828 / 0.117	20.95 / 0.843 / 0.206	25.26 / 0.773 / 0.304
	AdaIR	29.55 / 0.974 / 0.042	26.98 / 0.855 / 0.173	31.92 / 0.886 / 0.136	24.54 / 0.919 / 0.108	27.78 / 0.830 / 0.272
Unpaired	CUT	20.84 / 0.852 / 0.167	22.56 / 0.732 / 0.295	23.81 / 0.786 / 0.248	17.43 / 0.695 / 0.373	24.22 / 0.742 / 0.249
	DCLGAN	24.13 / 0.875 / 0.158	21.64 / 0.687 / 0.340	24.42 / 0.768 / 0.269	16.47 / 0.686 / 0.327	24.09 / 0.741 / 0.257
	DA-RCOT	20.48 / 0.842 / 0.188	21.98 / 0.651 / 0.340	24.79 / 0.622 / 0.397	14.31 / 0.644 / 0.411	23.71 / 0.753 / 0.307
	NLCL	23.59 / 0.881 / 0.133	21.61 / 0.643 / 0.356	21.58 / 0.677 / 0.398	15.74 / 0.700 / 0.328	19.69 / 0.730 / 0.294
	Santa	22.23 / 0.805 / 0.200	20.85 / 0.636 / 0.355	23.31 / 0.693 / 0.295	15.19 / 0.605 / 0.430	23.06 / 0.739 / 0.287
	UNSB	23.67 / 0.853 / 0.164	21.62 / 0.674 / 0.349	24.60 / 0.783 / 0.236	14.62 / 0.637 / 0.358	23.33 / 0.755 / 0.297
	FrDiff	23.80 / 0.867 / 0.160	21.85 / 0.723 / 0.306	23.79 / 0.776 / 0.265	17.59 / 0.715 / 0.322	23.88 / 0.747 / 0.259
	TP-Diff	23.18 / 0.857 / 0.179	22.74 / 0.765 / 0.273	25.04 / 0.785 / 0.226	16.33 / 0.690 / 0.346	24.26 / 0.760 / 0.269
	CycleGAN-Turbo	23.12 / 0.894 / 0.124	22.62 / 0.754 / 0.266	26.07 / 0.795 / 0.208	18.87 / 0.808 / 0.265	24.39 / 0.758 / 0.269
	Ours	**25.49 / 0.920 / 0.102**	**23.42 / 0.814 / 0.182**	**27.86 / 0.824 / 0.171**	**21.42 / 0.877 / 0.182**	**24.48 / 0.789 / 0.218**

4.1 Experimental Results

In this section, we conduct extensive quantitative and qualitative experiments to validate the effectiveness of the proposed method. We compare UADiff with two categories of approaches: (1) Paired AIO IR methods, including AirNet (Li et al., 2022a), PromptIR (Vaishnav et al., 2023), DA-CLIP (Luo et al., 2024a), and AdaIR (Cui et al., 2025); and (2) Unpaired methods, including CUT (Park et al., 2020), DCLGAN (Han et al., 2021), NLCL (Ye et al., 2022), Santa (Xie et al., 2023), UNSB (Kim et al., 2024), FrDiff (Liu et al., 2025a), TP-Diff (Liu et al., 2025b) and CycleGAN-Turbo (Parmar et al., 2024). Notably, DA-RCOT (Tang et al., 2025) is evaluated under both settings, as it supports both paired and unpaired training.

Results on AIO IR Benchmarks. We first evaluate our method on standard AIO IR benchmarks. The quantitative results are shown in Table 1. Compared with unpaired methods, our framework demonstrates significant superiority, achieving competitive results across all metrics and tasks. Compared to diffusion prior-based methods like CycleGAN-Turbo, our approach also achieves competitive results, primarily due to its degradation adaptation and unpaired constraint mechanisms. In comparison to paired AIO IR Methods, UADiff achieves sub-optimal results on reference-based metrics, as these supervised models are optimized with L2 or perceptual losses that directly rely

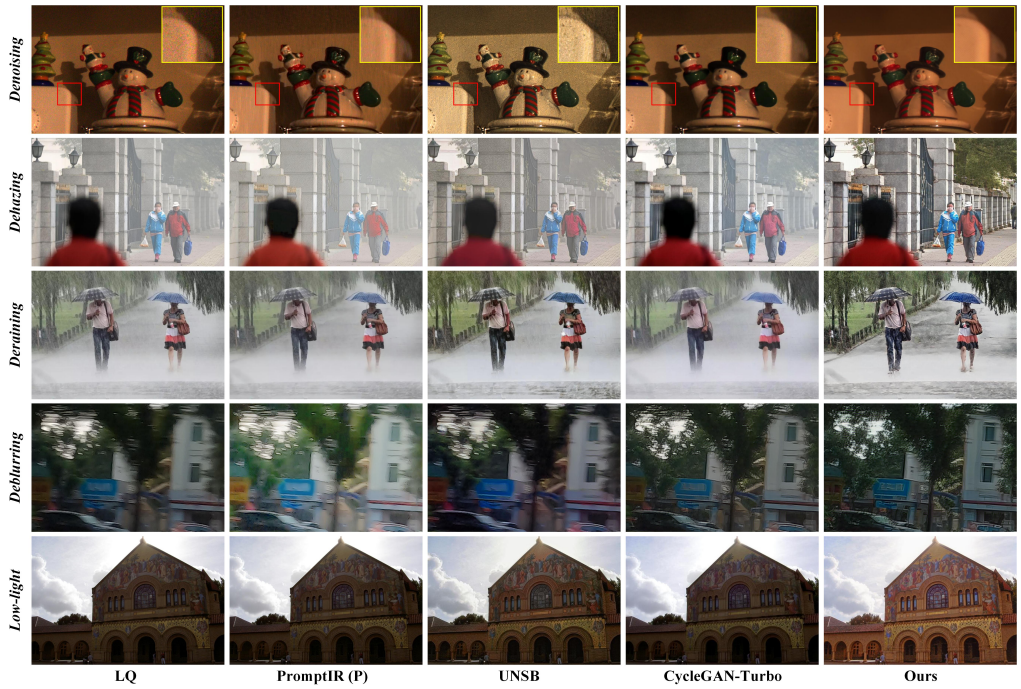

Figure 5: Visual comparison on real-world datasets. P denotes paired-training methods.

Table 2: Quantitative results on real-world datasets. **Bold** denotes the best results and underline denotes the second-best performance.

	Method	*Dehazing*	*Deraining*	*Denoising*	*Low-light*	*Deblurring*
		FID↓ / MUSIQ↑	FID↓ / MUSIQ↑	FID↓ / MUSIQ↑	FID↓ / MUSIQ↑	FID↓ / MUSIQ↑
Paired	AirNet	82.55 / 52.48	172.91 / 56.81	143.08 / 26.83	155.06 / 62.81	179.42 / 36.83
	PromptIR	75.61 / 56.96	164.04 / 58.31	136.48 / 27.95	143.55 / 63.55	153.75 / 38.07
	DA-RCOT	79.19 / 53.78	163.85 / 54.31	134.11 / 26.91	139.26 / 60.34	156.89 / 33.13
	DA-CLIP	76.22 / 56.64	168.46 / 56.36	136.15 / 30.56	145.14 / 62.65	136.33 / 39.61
	AdaIR	76.51 / 57.84	164.37 / 58.69	131.43 / 30.70	142.03 / 63.64	154.28 / 35.88
Unpaired	CUT	75.66 / 56.55	153.63 / 60.55	172.57 / 33.67	138.68 / 58.02	122.36 / 34.76
	DCLGAN	76.72 / 58.51	176.77 / 60.83	141.54 / 31.42	137.83 / 60.63	125.61 / 36.18
	DA-RCOT	79.17 / 53.89	163.00 / 54.26	134.25 / 26.99	134.62 / 60.37	159.97 / 33.08
	NLCL	75.54 / 57.62	162.60 / 56.89	140.02 / 27.21	146.73 / 55.15	136.09 / 37.37
	Santa	90.47 / 54.17	169.30 / 58.36	172.77 / 29.63	152.45 / 59.34	140.45 / 32.32
	UNSB	73.50 / 56.12	155.86 / 59.01	131.93 / 30.31	135.35 / 59.03	124.49 / 34.62
	FrDiff	73.11 / 56.24	158.78 / 59.63	139.91 / 28.87	134.28 / 59.27	126.41 / 35.83
	TP-Diff	74.01 / 55.91	151.44 / 59.94	132.12 / 32.88	136.72 / 60.41	120.22 / 36.71
	CycleGAN-Turbo	72.95 / 58.05	147.41 / 60.41	128.40 / 34.33	127.52 / 64.71	119.19 / 41.34
	Ours	**70.31 / 60.81**	**140.71 / 62.78**	**121.32 / 35.64**	**117.68 / 66.79**	**115.34 / 41.92**

on pixel-aligned, paired supervision. However, this optimization encourages the model to focus on synthetic degradation patterns, leading to overly smoothed outputs rather than faithfully restoring the intrinsic content (Yu et al., 2024; Wang et al., 2024a; Wu et al., 2024b). Qualitative comparisons in Figure 4 further validate these findings. Although trained without paired supervision, UADiff produces restoration results that are visually comparable or even superior to supervised methods, while delivering more stable and perceptually coherent results than existing unpaired methods.

Results on Real-World Datasets. To validate the generalization capability of our approach and demonstrate the advantage of unpaired frameworks in solving domain limitation, we evaluate all methods on real-world datasets. As summarized in Table 2, UADiff achieves superior FID and MUSIQ scores over both paired and unpaired baselines. As Shown in Figure 5, the qualitative results demonstrate that PromptIR, trained on synthetic paired data, fails to adapt to real-world degradations, leaving heavy rain streaks and exhibiting insufficient dehazing. In contrast, unpaired methods show better adaptability but still struggle with incomplete degradation removal or color distortions, reflecting their limited ability to balance fidelity and realism. Benefiting from degradation-adaptive restoration, UADiff consistently produces visually realistic, detail-preserving, and perceptually coherent restorations, delivering clean outputs with superior detail and color fidelity.

4.2 Ablation Study

We conduct comprehensive ablation studies on standard AIO IR benchmarks to evaluate each core component of our framework. The analysis focuses on three aspects: Degradation-Aware Prompt Learning (DAPL), Degradation-Conditional Restoration (DCR), and Content-Preserving Regularization (CPR). The metrics are averaged over the standard AIO IR benchmarks.

Table 3: Ablation studies on the effectiveness of DAPL and DCR.

(a) Degradation-Aware Prompt Learning (DAPL)

Setting	PSNR↑	SSIM↑	LPIPS↓
w/o Prompts	22.27	0.768	0.247
Fixed Prompts	23.31	0.804	0.204
UADiff (Learned Prompts)	**24.53**	**0.845**	**0.171**

(b) Degradation-Conditional Restoration (DCR)

Setting	PSNR↑	SSIM↑	LPIPS↓
Textual Prompt	23.13	0.802	0.212
Conditional LoRA	23.79	0.817	0.193
UADiff (*DC-LoRA*)	**24.53**	**0.845**	**0.171**

Effectiveness of DAPL. We further evaluate the necessity of the proposed DAPL module. As shown in Table 3(a), removing prompts entirely or replacing learned prompts with fixed textual prompts both result in noticeable performance degradation. In contrast, the learned prompts achieve the best performance, demonstrating that accurate degradation awareness provides effective guidance for subsequent adaptive restoration.

Effectiveness of DCR. As shown in Table 3(b), we evaluate the effectiveness of DCR by comparing the degradation-conditional LoRA (*DC-LoRA*) module with two simplified variants: (1) *Textual Prompt*, where only textual prompts are used to provide basic degradation awareness without conditional parameter adjustment; and (2) *Conditional LoRA*, where LoRA incorporates degradation information but applies identical modulation across layers. Results show that *Textual Prompt* performs the worst, indicating that semantic cues alone are insufficient to handle diverse degradation conditions in AIO restoration. *Conditional LoRA* improves performance by introducing degradation-aware conditioning, yet its adaptation remains coarse due to the absence of layer-wise modulation. In contrast, *DC-LoRA* achieves superior overall performance, demonstrating that layer-specific conditional modulation enables flexible and precise adaptation across varying degradation types.

Table 4: Ablation study on the effectiveness of CPR.

Setting	PSNR↑	SSIM↑	LPIPS↓
w/o $\mathcal{L}_{gan}$	19.92	0.689	0.324
w/o $\mathcal{L}_{siglip}$	23.36	0.798	0.209
w/o $\mathcal{L}_{adc}$	18.13	0.607	0.396
w/ $\mathcal{L}_{patchnce}$	23.46	0.807	0.209
UADiff (w/ $\mathcal{L}_{adc}$)	**24.53**	**0.845**	**0.171**

Effectiveness of CPR. As shown in Table 4, removing $\mathcal{L}_{adc}$ leads to substantial degradation in image fidelity, indicating that adversarial and semantic losses alone are insufficient for effective content preservation. Without explicit content preserving objective, the model lacks constraints to maintain structural information during restoration. Replacing $\mathcal{L}_{adc}$ with PatchNCE loss partially improves performance but still underperforms the full model, as PatchNCE tends to retain degradation artifacts. In contrast, $\mathcal{L}_{adc}$ disentangles content from degradation through adversarial feature purification, enabling reliable content preservation. Removing adversarial loss $\mathcal{L}_{gan}$ reduces perceptual realism, while eliminating semantic loss $\mathcal{L}_{siglip}$ eads to consistent performance degradation across all three metrics, with more analyses and qualitative results provided in the supplementary material. These results confirm that $\mathcal{L}_{adc}$, $\mathcal{L}_{siglip}$, and $\mathcal{L}_{gan}$ jointly contribute to balancing content fidelity, perceptual realism, and degradation removal.

5 Conclusion

In this paper, we address unpaired AIO IR by identifying two key challenges: achieving effective degradation adaptation and establishing reliable unpaired constraint. To tackle these issues, we propose UADiff, a unified framework that fine-tunes a pre-trained diffusion model into one-step restoration backbone. For degradation adaptation, we introduce DAPL and DCR, which enable model to perceive diverse degradations and adapt the restoration process. For the unpaired constraint, we present CPR, which balances degradation removal and content fidelity through adversarial feature purification and semantic guidance. Extensive experiments on standard AIO IR benchmarks and real-world datasets demonstrate that UADiff achieves superior perceptual quality and generalization, establishing a strong baseline for unpaired AIO restoration in real-world scenarios.

REFERENCES

Abdelrahman Abdelhamed, Stephen Lin, and Michael S Brown. A high-quality denoising dataset for smartphone cameras. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1692–1700, 2018.

Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition workshops*, pp. 126–135, 2017.

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*, pp. 25432–25444, 2024a.

Yuang Ai, Xiaoqiang Zhou, Huaibo Huang, Xiaotian Han, Zhengyu Chen, Quanzeng You, and Hongxia Yang. Dreamclear: High-capacity real-world image restoration with privacy-safe dataset curation. *Advances in Neural Information Processing Systems*, 37:55443–55469, 2024b.

Josue Anaya and Adrian Barbu. Renoir–a dataset for real low-light image noise reduction. *Journal of Visual Communication and Image Representation*, 51:144–154, 2018.

Pablo Arbelaez, Michael Maire, Charless Fowlkes, and Jitendra Malik. Contour detection and hierarchical image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 33(5):898–916, 2010.

Junyang Chen, Jinshan Pan, and Jiangxin Dong. Faithdiff: Unleashing diffusion priors for faithful image super-resolution. In *Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition*, pp. 28188–28197, 2025.

Xiang Chen, Jinshan Pan, Kui Jiang, Yufeng Li, Yufeng Huang, Caihua Kong, Longgang Dai, and Zhentao Fan. Unpaired deep image deraining using dual contrastive learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2017–2026, 2022.

Marcos V Conde, Gregor Geigle, and Radu Timofte. Instructir: High-quality image restoration following human instructions. In *Proceedings of the European Conference on Computer Vision*, pp. 1–21. Springer, 2024.

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. In *International Conference on Learning Representations*, pp. 57335–57356, 2025.

Guanglu Dong, Tianheng Zheng, Yuanzhouhan Cao, Linbo Qing, and Chao Ren. Channel consistency prior and self-reconstruction strategy based unsupervised image deraining. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7469–7479, 2025a.

Linwei Dong, Qingnan Fan, Yihong Guo, Zhonghao Wang, Qi Zhang, Jinwei Chen, Yawei Luo, and Changqing Zou. Tsd-sr: One-step diffusion with target score distillation for real-world image super-resolution. In *Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition*, pp. 23174–23184, 2025b.

Raanan Fattal. Dehazing using color-lines. *ACM transactions on graphics*, 34(1):1–14, 2014.

Xueyang Fu, Jiabin Huang, Delu Zeng, Yue Huang, Xinghao Ding, and John Paisley. Removing rain from single images via a deep detail network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3855–3863, 2017.

Xiaojie Guo, Yu Li, and Haibin Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on image processing*, 26(2):982–993, 2016.

Yu Guo, Yuan Gao, Yuxu Lu, Huilin Zhu, Ryan Wen Liu, and Shengfeng He. Onerestore: A universal restoration framework for composite degradation. In *Proceedings of the European Conference on Computer Vision*, pp. 255–272. Springer, 2024.

Junlin Han, Mehrdad Shoeiby, Lars Petersson, and Mohammad Ali Armin. Dual contrastive learning for unsupervised image-to-image translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 746–755, 2021.

Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017.

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

Huaibo Huang, Mandi Luo, and Ran He. Memory uncertainty learning for real-world single image deraining. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3):3446–3460, 2022.

Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5197–5206, 2015.

Xun Huang, Ming-Yu Liu, Serge Belongie, and Jan Kautz. Multimodal unsupervised image-to-image translation. In *Proceedings of the European Conference on Computer Vision*, pp. 172–189, 2018.

Junjun Jiang, Zengyuan Zuo, Gang Wu, Kui Jiang, and Xianming Liu. A survey on all-in-one image restoration: Taxonomy, evaluation and future trends. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025.

Yitong Jiang, Zhaoyang Zhang, Tianfan Xue, and Jinwei Gu. Autodir: Automatic all-in-one image restoration with latent diffusion. In *Proceedings of the European Conference on Computer Vision*, pp. 340–359. Springer, 2024.

Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. Musiq: Multi-scale image quality transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 5148–5157, 2021.

Beomsu Kim, Gihyun Kwon, Kwanyoung Kim, and Jong Chul Ye. Unpaired image-to-image translation via neural schrödinger bridge. In *International Conference on Learning Representations*, 2024.

Yunwei Lan, Zhigao Cui, Xin Luo, Chang Liu, Nian Wang, Menglin Zhang, Yanzhao Su, and Dong Liu. When schrodinger bridge meets real-world image dehazing with unpaired training. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 8756–8765, 2025.

Hsin-Ying Lee, Hung-Yu Tseng, Jia-Bin Huang, Maneesh Singh, and Ming-Hsuan Yang. Diverse image-to-image translation via disentangled representations. In *Proceedings of the European Conference on Computer Vision*, pp. 35–51, 2018.

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.

Boyun Li, Xiao Liu, Peng Hu, Zhongqin Wu, Jiancheng Lv, and Xi Peng. All-in-one image restoration for unknown corruption. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17452–17462, 2022a.

Wei Li, Qiming Zhang, Jing Zhang, Zhen Huang, Xinmei Tian, and Dacheng Tao. Toward real-world single image deraining: A new benchmark and beyond. *arXiv preprint arXiv:2206.05514*, 2022b.

Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1833–1844, 2021.

Zhexin Liang, Chongyi Li, Shangchen Zhou, Ruicheng Feng, and Chen Change Loy. Iterative prompt learning for unsupervised backlit image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 8094–8103, 2023.

Xinqi Lin, Jingwen He, Ziyan Chen, Zhaoyang Lyu, Bo Dai, Fanghua Yu, Yu Qiao, Wanli Ouyang, and Chao Dong. Diffbir: Toward blind image restoration with generative diffusion prior. In *Proceedings of the European Conference on Computer Vision*, pp. 430–448. Springer, 2024.

Chengxu Liu, Lu Qi, Jinshan Pan, Xueming Qian, and Ming-Hsuan Yang. Frequency domain-based diffusion model for unpaired image dehazing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2025a.

Chengxu Liu, Lu Qi, Jinshan Pan, Xueming Qian, and Ming-Hsuan Yang. Learning deblurring texture prior from unpaired data with diffusion model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 14195–14204, 2025b.

Ziwei Luo, Fredrik K Gustafsson, Zheng Zhao, Jens Sjölund, and Thomas B Schön. Controlling vision-language models for multi-task image restoration. In *International Conference on Learning Representations*, 2024a.

Ziwei Luo, Fredrik K Gustafsson, Zheng Zhao, Jens Sjölund, and Thomas B Schön. Photo-realistic image restoration in the wild with controlled vision-language models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6641–6651, 2024b.

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.

Kede Ma, Kai Zeng, and Zhou Wang. Perceptual quality assessment for multi-exposure image fusion. *IEEE Transactions on Image Processing*, 24(11):3345–3356, 2015.

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.

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 of the IEEE/CVF International Conference on Computer Vision*, volume 2, pp. 416–423. IEEE, 2001.

Igor Morawski, Kai He, Shusil Dangi, and Winston H Hsu. Unsupervised image prior via prompt learning and clip semantic guidance for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5971–5981, 2024.

Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3883–3891, 2017.

Taesung Park, Alexei A Efros, Richard Zhang, and Jun-Yan Zhu. Contrastive learning for unpaired image-to-image translation. In *Proceedings of the European Conference on Computer Vision*, pp. 319–345. Springer, 2020.

Gaurav Parmar, Taesung Park, Srinivasa Narasimhan, and Jun-Yan Zhu. One-step image translation with text-to-image models. *arXiv preprint arXiv:2403.12036*, 2024.

Bang-Dang Pham, Phong Tran, Anh Tran, Cuong Pham, Rang Nguyen, and Minh Hoai. Hypercut: Video sequence from a single blurry image using unsupervised ordering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9843–9852, 2023.

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*, pp. 8748–8763, 2021.

Jaesung Rim, Haeyun Lee, Jucheol Won, and Sunghyun Cho. Real-world blur dataset for learning and benchmarking deblurring algorithms. In *Proceedings of the European Conference on Computer Vision*, pp. 184–201. Springer, 2020.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10684–10695, 2022.

Mehran Safayani, Behnaz Mirzapour, Nasrin Salehi, Hamid Ravaee, et al. Unpaired image-to-image translation with content preserving perspective: A review. *arXiv preprint arXiv:2502.08667*, 2025.

Axel Sauer, Dominik Lorenz, Andreas Blattmann, and Robin Rombach. Adversarial diffusion distillation. In *Proceedings of the European Conference on Computer Vision*, pp. 87–103. Springer, 2024.

Yuanjie Shao, Lerenhan Li, Wenqi Ren, Changxin Gao, and Nong Sang. Domain adaptation for image dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2808–2817, 2020.

Ziyi Shen, Wenguan Wang, Xiankai Lu, Jianbing Shen, Haibin Ling, Tingfa Xu, and Ling Shao. Human-aware motion deblurring. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 5572–5581, 2019.

Lingchen Sun, Rongyuan Wu, Zhiyuan Ma, Shuaizheng Liu, Qiaosi Yi, and Lei Zhang. Pixel-level and semantic-level adjustable super-resolution: A dual-lora approach. In *Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition*, pp. 2333–2343, 2025.

Xiaole Tang, Xiang Gu, Xiaoyi He, Xin Hu, and Jian Sun. Degradation-aware residual-conditioned optimal transport for unified image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025.

P Vaishnav, Z Syed Waqas, K Salman, and K Fahad Shahbaz. Promptir: Prompting for all-in-one blind image restoration. *Advances in Neural Information Processing Systems*, 2023.

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, 2024a.

Jingkai Wang, Jue Gong, Lin Zhang, Zheng Chen, Xing Liu, Hong Gu, Yutong Liu, Yulun Zhang, and Xiaokang Yang. Osdface: One-step diffusion model for face restoration. In *Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition*, pp. 12626–12636, 2025.

Longguang Wang, Yingqian Wang, Xiaoyu Dong, Qingyu Xu, Jungang Yang, Wei An, and Yulan Guo. Unsupervised degradation representation learning for blind super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10581–10590, 2021.

Longguang Wang, Juncheng Li, Yingqian Wang, Qingyong Hu, and Yulan Guo. Learning coupled dictionaries from unpaired data for image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 25712–25721, 2024b.

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

Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17683–17693, 2022.

Zhongze Wang, Haitao Zhao, Jingchao Peng, Lujian Yao, and Kaijie Zhao. Odcr: Orthogonal decoupling contrastive regularization for unpaired image dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 25479–25489, 2024d.

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

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, 2024a.

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*, pp. 25456–25467, 2024b.

Shaoan Xie, Yanwu Xu, Mingming Gong, and Kun Zhang. Unpaired image-to-image translation with shortest path regularization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10177–10187, 2023.

Xiaogang Xu, Shu Kong, Tao Hu, Zhe Liu, and Hujun Bao. Boosting image restoration via priors from pre-trained models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2900–2909, 2024.

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.

Yuntong Ye, Changfeng Yu, Yi Chang, Lin Zhu, Xi-Le Zhao, Luxin Yan, and Yonghong Tian. Unsupervised deraining: Where contrastive learning meets self-similarity. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5821–5830, 2022.

Fanghua Yu, Jinjin Gu, Zheyuan Li, Jinfan Hu, Xiangtao Kong, Xintao Wang, Jingwen He, Yu Qiao, and Chao Dong. Scaling up to excellence: Practicing model scaling for photo-realistic image restoration in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 25669–25680, 2024.

Yuan Yuan, Siyuan Liu, Jiawei Zhang, Yongbing Zhang, Chao Dong, and Liang Lin. Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition workshops*, pp. 701–710, 2018.

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 IEEE/CVF conference on Computer Vision and Pattern Recognition*, pp. 12753–12763, 2025.

Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5728–5739, 2022.

Xiaohua Zhai, Basil Mustafa, Alexander Kolesnikov, and Lucas Beyer. Sigmoid loss for language image pre-training. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 11975–11986, 2023.

Aiping Zhang, Zongsheng Yue, Renjing Pei, Wenqi Ren, and Xiaochun Cao. Degradation-guided one-step image super-resolution with diffusion priors. *arXiv preprint arXiv:2409.17058*, 2024.

He Zhang, Vishwanath Sindagi, and Vishal M Patel. Image de-raining using a conditional generative adversarial network. *IEEE transactions on circuits and systems for video technology*, 30(11): 3943–3956, 2019.

Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 586–595, 2018.

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*, pp. 25445–25455, 2024.

Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 2223–2232, 2017.

SUPPLEMENTARY OVERVIEW

In this supplementary material, we provide experiments and analyses omitted from the main paper due to space constraints, together with additional technical details that support reproducibility. We first report mixed-degradation experiments and qualitative ablations because they directly complement the main experimental findings. We then present efficiency analysis, detailed formulations of Content-Preserving Regularization (CPR), further analysis of Degradation-Aware Prompt Learning (DAPL), implementation details, limitations, additional visual comparisons, and a statement on the use of large language models. Specifically, the supplementary material contains:

- A statement of large language model usage in manuscript preparation, and a statement of reproducibility;
- Quantitative and qualitative evaluation under mixed degradations;
- Qualitative ablations of the content-preserving objectives;
- Comprehensive analysis of computational cost, parameter efficiency, and practical deployment;
- More detailed explanations and mathematical formulations of the proposed Content-Preserving Regularization (CPR);
- Quantitative evaluations of the Degradation-Aware Prompt Learning (DAPL) mechanism on real-world datasets;
- Further implementation details regarding the two-stage training configurations and unpaired dataset construction;
- In-depth discussion of current limitations and potential directions for future work;
- Ablation studies on LoRA rank settings and additional qualitative results on standard AIO benchmarks, real-world images, and mixed-degradation scenarios.

STATEMENT ON THE USE OF LARGE LANGUAGE MODELS

In the preparation of this paper, we used large language models (LLMs) as an auxiliary tool. The primary role of LLMs was to assist with language polishing, phrasing improvements, grammar checking, and enhancing the clarity and fluency of certain sentences, ensuring that our research ideas and technical details are expressed more accurately and professionally. Importantly, LLMs were not involved in the core aspects of this work, including research conception, experimental design, result generation, or data analysis.

REPRODUCIBILITY STATEMENT

We state that UADiff is highly reproducible. Complete implementation details and training hyperparameters are thoroughly documented in the Method section and Appendix. In addition, the unpaired dataset construction pipeline and mathematical formulations of all training objectives are detailed in the supplementary material to support full reproducibility of our work.

A MIXED-DEGRADATION EXPERIMENTS AND ANALYSIS

We further evaluate UADiff on the CDD11 dataset (Guo et al., 2024), which contains images corrupted by multiple co-existing degradations. This setting is particularly challenging because a restoration model must identify and remove several heterogeneous degradations while preserving image content.

As shown in Table 5, UADiff consistently outperforms existing methods across all mixed-degradation combinations. The gains become particularly clear when low light co-exists with haze and rain, where independently designed restoration operations can interfere with one another. DAPL can represent multiple degradation cues simultaneously, while Degradation-Conditional Restoration (DCR) translates these cues into adaptive layer-wise modulation. Together with CPR, this design

removes co-existing degradations without sacrificing structural fidelity. The qualitative results in Figure 6 further show that UADiff removes compound artifacts more completely while retaining natural colors and fine details.

B Qualitative Ablation Analysis

The main paper quantitatively verifies the contributions of the CPR objectives. Figures 7 and 8 provide the corresponding qualitative comparisons for the adversarial disentangled contrastive loss $\mathcal{L}_{adc}$ and semantic regularization loss $\mathcal{L}_{siglip}$, respectively.

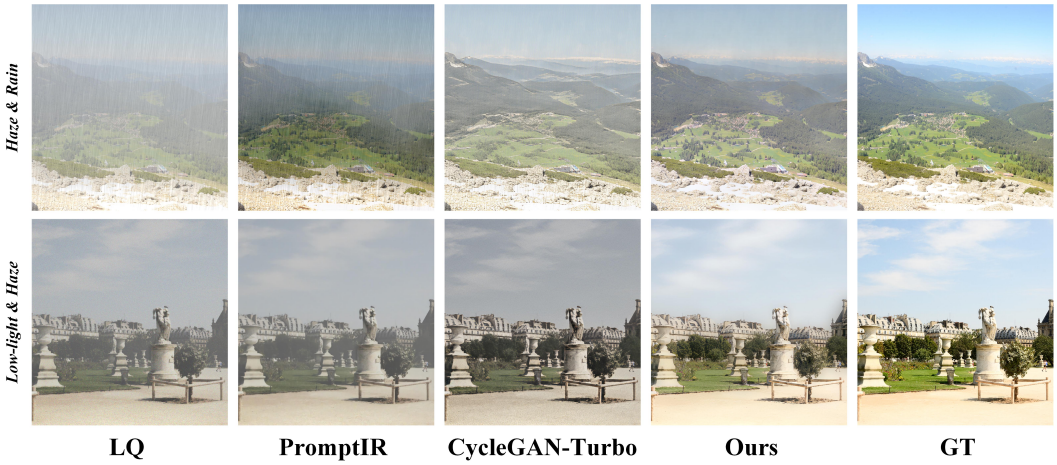

Figure 6: Visual comparison on CDD11 mixed degradation datasetsGuo et al. (2024).

Table 5: Quantitative results on CDD11 under mixed-degradation settings. Results are reported as PSNR↑ / SSIM↑ / LPIPS↓. The best and second-best results are highlighted in red and blue, respectively.

Method	Haze+Rain	Low-light+Haze	Low-light+Rain	Low-light+Haze+Rain
AirNet	18.89 / 0.796 / 0.342	17.97 / 0.703 / 0.375	17.56 / 0.689 / 0.405	17.42 / 0.673 / 0.411
PromptIR	19.52 / 0.813 / 0.328	18.39 / 0.728 / 0.364	18.21 / 0.713 / 0.371	18.03 / 0.701 / 0.382
DCLGAN	19.13 / 0.737 / 0.376	17.64 / 0.693 / 0.386	16.89 / 0.673 / 0.412	16.92 / 0.669 / 0.423
UNSB	18.32 / 0.743 / 0.358	18.25 / 0.711 / 0.380	17.83 / 0.698 / 0.392	17.51 / 0.677 / 0.408
CycleGAN-Turbo	19.31 / 0.801 / 0.337	18.71 / 0.737 / 0.352	18.52 / 0.723 / 0.364	18.24 / 0.718 / 0.378
Ours	20.03 / 0.824 / 0.312	19.44 / 0.761 / 0.334	19.12 / 0.741 / 0.349	18.86 / 0.732 / 0.362

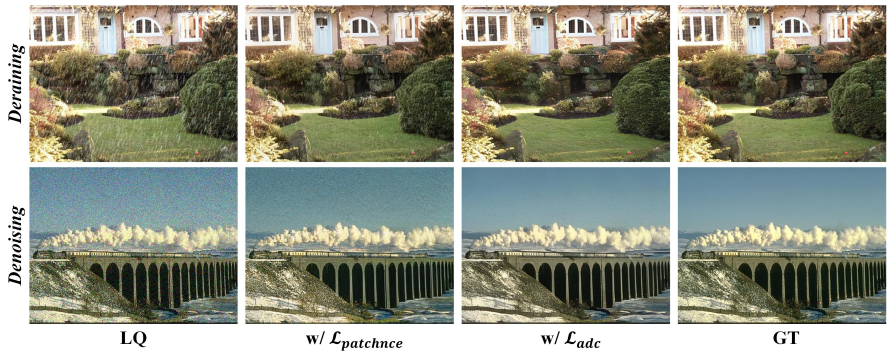

Figure 7: Qualitative ablation of $\mathcal{L}_{adc}$. Removing $\mathcal{L}_{adc}$ weakens content preservation and leaves structural artifacts.

Without $\mathcal{L}_{adc}$, the restored images exhibit less reliable structural preservation because the representation still entangles content with degradation-specific features. Replacing this objective with a conventional contrastive constraint can also retain unwanted degradation cues. Without $\mathcal{L}_{siglip}$, the model lacks explicit semantic guidance away from degradation concepts, resulting in incomplete restoration and reduced perceptual quality. These visual observations agree with the quantitative ablations in the main paper and demonstrate that the two objectives play complementary roles: $\mathcal{L}_{adc}$ protects degradation-invariant content, whereas $\mathcal{L}_{siglip}$ encourages effective degradation removal.

C Computational Cost and Application Analysis

To comprehensively evaluate our proposed UADiff, we analyze the computational cost across three key dimensions: inference speed, parameter efficiency, and application scenarios. The detailed comparison is presented in Table 6.

Inference Speed and Computational FLOPs. Although diffusion models are typically computationally expensive, UADiff alleviates this issue through two key designs: one-step diffusion acceleration and latent-space denoising. By adopting a one-step generation paradigm, UADiff significantly reduces the inference cost of conventional multi-step diffusion models. As shown in Table 6, the vanilla Latent Diffusion Model (LDM) Rombach et al. (2022) requires 3.14 s and 18769.78 G FLOPs for iterative denoising, while UADiff reduces this to 0.13 s and 2816.42 G FLOPs. Moreover, unlike diffusion-based methods operating in the high-resolution pixel space (e.g., DA-CLIP with 46462.29 G FLOPs and 5.79 s latency), our method performs denoising in a compressed latent space, further improving efficiency. Consequently, compared with non-diffusion methods such as PromptIR (0.27 s) and DA-RCOT (0.35 s), our framework does not introduce significant computational overhead while achieving superior restoration performance.

Parameter Efficiency. Compared to non-diffusion models, UADiff has a large parameter count due to the pre-trained diffusion backbone. However, due to the inherently ill-posed nature of image restoration tasks, research has shown that modifying network architectures or blindly increasing parameters often causes the model to overfit to the training data and synthetic degradations Xu et al. (2024). In contrast, effectively leveraging robust generative priors offers a more promising solution. Driven by this insight, we use the diffusion model prior as the backbone for fine-tuning. By freezing the large base model and introducing Degradation-Conditional LoRA (DC-LoRA), UADiff requires only 29.68 M trainable parameters. This is even smaller than some lightweight non-diffusion models and much more efficient than approaches like CycleGAN-Turbo (126.49 M). This demonstrates that our framework harnesses the generative power of large models without excessive parameter tuning.

Practicality and Application Scenarios. In practice, the large total parameter count of UADiff does not present a significant deployment bottleneck. Since the computationally intensive denoising process is executed within a spatially compressed latent space, VRAM consumption during inference remains manageable, and latency is kept competitive. Given this computational profile, the practical deployment of restoration models should be tailored to specific application scenarios:



Figure 8: Qualitative ablation of $\mathcal{L}_{siglip}$. Removing $\mathcal{L}_{siglip}$ reduces semantic guidance for degradation removal.

Table 6: Computational cost comparison. We report the FLOPs, inference latency, number of trainable parameters, and total parameters. Latency is measured on 512×512 images using an RTX 3090 GPU.

Method	FLOPs (G)	Latency (s)	Trainable Params (M)	Total Params (M)	Type
AirNet	1205.09	0.43	8.93	8.93	Non-diffusion-based
PromptIR	632.56	0.27	35.59	35.59	Non-diffusion-based
DA-RCOT	1332.21	0.35	52.34	52.34	Non-diffusion-based
DA-CLIP	46462.29	5.79	143.92	295.19	Diffusion-based
AdaIR	588.73	0.29	28.78	28.78	Non-diffusion-based
CUT	256.52	0.03	11.38	11.38	Non-diffusion-based
DCLGAN	268.42	0.11	22.72	22.72	Non-diffusion-based
NLCL	176.30	0.06	1.72	1.72	Non-diffusion-based
Santa	118.63	0.07	11.43	11.43	Non-diffusion-based
UNSB	1269.15	0.57	14.68	14.68	Diffusion-based
FrDiff	234.34	0.04	8.76	8.76	Diffusion-based
TP-Diff	681.51	0.12	11.89	11.89	Diffusion-based
LDM	18769.78	3.14	1243.19	1243.19	Diffusion-based
CycleGAN-Turbo	2753.86	0.11	126.49	1436.38	Diffusion-based
UADiff (Ours)	2816.42	0.13	29.68	1272.87	Diffusion-based

- **Quality-Centric Scenarios:** For applications where perceptual fidelity, generalization, and structural accuracy are critical (e.g., high-end computational photography, remote sensing, and archival image restoration), UADiff serves as the optimal choice. By leveraging generative prior, it successfully resolves highly ill-posed, complex, and unpaired real-world degradations that traditional non-diffusion models inherently struggle to recover.
- **Speed-Centric Scenarios:** In latency-sensitive or resource-constrained environments (e.g., real-time video streaming on mobile devices, low-power drones), lightweight non-diffusion models are more suitable, especially when performance is less critical. In future work, we plan to explore pruning, quantization and knowledge distillation techniques to compress the model, aiming to develop a lightweight restoration network that balances restoration quality and inference efficiency.

D Details of Content-Preserving Regularization

In this section, we provide the mathematical formulations of the objective functions used in UADiff. We specifically elaborate on the **Adversarial Disentangled Contrastive Loss** ($\mathcal{L}_{adc}$), discussing its role as a *feature purification mechanism* that resolves the inherent conflict between degradation removal and content preservation in unpaired learning.

D.1 Total Objective Function

The overall training objective of UADiff is a weighted sum of three components:

$$\mathcal{L}_{total} = \lambda_{gan}\mathcal{L}_{gan} + \lambda_{adc}\mathcal{L}_{adc} + \lambda_{siglip}\mathcal{L}_{siglip}, \tag{7}$$

where λ_{gan}, λ_{adc}, and λ_{siglip} are hyperparameters balancing perceptual fidelity, content preservation, and semantic correctness. In our experiments, we set $\lambda_{gan} = 1.0$, $\lambda_{adc} = 1.0$. The weight λ_{siglip} is not fixed; instead, it is scheduled to decay from an initial value of 1.0 to 0.1 over the course of training, allowing strong semantic guidance in the early stages while preventing over-constraining the optimization in later iterations.

D.2 Feature Purification via Adversarial Disentangled Contrastive Loss

A core challenge in unpaired image restoration is that standard contrastive learning (e.g., CUTPark et al. (2020)) maximizes the mutual information between the input I_{lq} and the output I_{res}. Since I_{lq} contains degradations (e.g., rain streaks, noise), standard contrastive loss inadvertently encourages the model to preserve these degradation patterns in I_{res}.

To address this, we introduce a **Feature Purification** mechanism. Rather than applying the contrastive loss directly to raw features that still carry degradation cues, we first transform the feature space so that it encodes only structural and semantic information. This purification step effectively removes degradation-related components, allowing contrastive learning to reinforce content consistency without reinforcing the degradations themselves.

Adversarial Feature Disentanglement. We employ a Shared Patch Embedding (SPE) encoder, denoted as E_{patch}, and a domain discriminator D_{cls}. The goal is to force the distribution of features extracted from degraded images (I_{lq}) to be indistinguishable from those extracted from high-quality images (I_{hq}).

Let $z_{lq} = E_{patch}(I_{lq})$ and $z_{hq} = E_{patch}(I_{hq})$. We formulate a min-max adversarial optimization process where E_{patch} functions as a *purifier*:

$$\begin{aligned}\mathcal{L}_{dis}(E_{patch}, D_{cls}) = \mathbb{E}_{I_{hq}}[\log D_{cls}(E_{patch}(I_{hq}))] \\ +\mathbb{E}_{I_{lq}}[\log(1 - D_{cls}(E_{patch}(I_{lq})))].\end{aligned} \tag{8}$$

Here, the discriminator D_{cls} tries to identify whether a feature patch comes from the clean or degraded domain. The encoder E_{patch} aims to maximize this loss (fool the discriminator).

$$\min_{E_{patch}} \max_{D_{cls}} \mathcal{L}_{dis}(E_{patch}, D_{cls}). \tag{9}$$

Mathematical Interpretation of Disentanglement. To understand the effect of the adversarial constraint, we consider the feature representation z produced by the encoder, which can be conceptually decomposed into a content component z_c and a degradation-related component z_d, i.e., $z = \{z_c, z_d\}$. In the clean domain, samples I_{hq} contain only the content component and inherently lack degradation information. Therefore, for the discriminator to be unable to distinguish $E_{\text{patch}}(I_{lq})$ from $E_{\text{patch}}(I_{hq})$, the encoder must suppress or eliminate the degradation component z_d. Through this adversarial pressure, the encoder gradually removes degradation-specific signals and produces a purified representation $\hat{z}$ that retains content structure while discarding degradation cues. Thus, we obtain the purified representation $\hat{z}$:

$$\hat{z} = \text{Purify}(z) \approx z_c. \tag{10}$$

Contrastive Learning on Purified Features. Once the features have been purified, we apply the PatchNCE loss. Since the purified representation $\hat{z}$ is expected to retain only degradation-invariant content, maximizing the mutual information between the input and the output in this space enforces structural consistency without inadvertently preserving degradation artifacts. The PatchNCE loss is computed using the purified features of the degraded input I_{lq} and the restored image I_{res}:

$$\begin{aligned}\mathcal{L}_{\text{PatchNCE}}(G, E_{\text{patch}}, I_{lq}) = \mathbb{E}_{I_{lq}} \sum_{l=1}^{L} \sum_{s=1}^{S_l} \\ -\log \frac{\exp(\hat{z}_l^{\,s} \cdot \hat{v}_l^{\,s}/\tau)}{\sum_{j \in Neg} \exp(\hat{z}_l^{\,s} \cdot \hat{v}_l^{\,j}/\tau)},\end{aligned} \tag{11}$$

where $\hat{z}_l^s = E_{patch}(I_{lq})_l^s$ denotes the purified feature of the input at layer l, spatial location s, while $\hat{v}_l^s = E_{patch}(G(I_{lq}))_l^s$ represents the corresponding feature extracted from the restored output. The negative set Neg consists of other spatial locations within the same layer that serve as negative samples for contrastive learning. $\tau = 0.07$ is the temperature parameter. The total $\mathcal{L}_{adc}$ is the combination of the adversarial purification objective and the contrastive objective:

$$\mathcal{L}_{adc} = \mathcal{L}_{PatchNCE} + \lambda_{dis}\mathcal{L}_{dis} \tag{12}$$

where λ_{dis} is set 0.2.

D.3 Semantic Regularization via Prompt Guidance

To further guide the restoration towards the target domain distribution without paired data, the learned prompts act as semantic regularizers. We utilize a SigLIP-based loss $\mathcal{L}_{siglip}$ that encourages the restored image I_{res} to approach the clean concept T_{clean} while being pushed away from

the degradation concept T_{deg} in the SigLIP feature space.

$$\mathcal{L}_{siglip} = \lambda_{\text{clean}} \left(1 - S(I_{res}, T_{clean})\right) + \frac{S(I_{res}, T_{deg})}{S(I_{res}, T_{deg}) + S(I_{lq}, T_{deg})}, \tag{13}$$

where $S(I, T)$ denotes the SigLIP similarity score formulated as $\sigma(t(\Phi_{image}(I) \cdot \Phi_{text}(T)) + b)$, and Φ_{image} and Φ_{text} represent the frozen SigLIP image and text encoders, respectively. This non-pixel-level constraint operates in the semantic space, ensuring the global visual style aligns with high-quality images.

D.4 Adversarial Loss

We employ a standard LSGAN loss to ensure perceptual realism. The discriminator D learns to distinguish between real clean images and restored images:

$$\mathcal{L}_{gan} = \mathbb{E}_{I_{hq}}[(D(I_{hq}) - 1)^2] + \mathbb{E}_{I_{lq}}[(D(G(I_{lq})))^2], \tag{14}$$

This term ensures that the restored images lie on the manifold of natural, high-quality images.

E Analysis of Degradation-Aware Prompt Learning

In the main paper, we introduced Degradation-Aware Prompt Learning (DAPL) to capture degradation semantics and provide adaptive guidance. To further assess its effectiveness, we provide here a quantitative evaluation that examines its classification mechanism and its capability to distinguish diverse degradation types. Specifically, we design a classification experiment on real-world datasets to ensure that the learned prompts are tested under authentic and naturally occurring degradations rather than synthetic or idealized distortions.

In this experiment, fixed prompts are constructed using common handcrafted textual descriptions such as "rainy image", "low-light image", and "noisy image", which are encoded using the frozen SigLIP text encoder. In contrast, the learned prompts correspond to our optimized degradation prompt embeddings. We compute the zero-shot classification accuracy using the SigLIP similarity score. This score is defined as the sigmoid activation of the scaled dot-product between the image and text features. Because DAPL enables independent modeling of multiple co-existing degradations via a pairwise Sigmoid cross-entropy loss, a prediction is regarded as correct only when the similarity score with its ground-truth degradation prompt exceeds 0.5.

Table 7: Quantitative comparison of classification accuracy between Fixed Textual Prompts and our Learned Prompts on real-world datasets.

Degradation	Fixed Prompts	Learned Prompts	Improvement
Low-light	68.44%	**99.13%**	+30.69%
Hazy	73.17%	**98.33%**	+25.16%
Blurry	67.26%	**98.22%**	+30.96%
Rainy	70.13%	**96.62%**	+26.49%
Noisy	77.40%	**99.43%**	+22.03%
Average	71.28%	**98.35%**	**+27.07%**

Table 7 presents a comparison between fixed textual prompts and those generated by the DAPL module. The results indicate that the learned prompts significantly improve recognition accuracy across all degradation types. Fixed prompts often fail to capture the fine-grained nature of real-world degradations, whereas the learned prompts provide much more robust and accurate representations. For instance, the fixed prompt "blurry" struggles to distinguish between different types of blur, while the learned prompts achieve much higher accuracy after optimization. This improvement enables the Degradation-Conditional Restoration (DCR) module to receive more precise degradation cues, leading to more effective restoration.

F Implementation Details

We provide additional clarification on the implementation of our two-stage training framework, which consists of (1) Degradation-Aware Prompt Learning (DAPL) and (2) Degradation-Conditional Restoration (DCR).

Stage 1: DAPL. In the first stage, we freeze both the CLIP image encoder and text encoder while optimizing only the learnable prompt embeddings. Each prompt is parameterized as a sequence of 16 trainable token embeddings, initialized from a neutral placeholder template composed of repeated "X" symbols. This design ensures that the prompts begin from an unbiased semantic state and evolve purely based on the data-driven optimization process. During training, we use a batch size of 512 and optimize the prompt embeddings with the AdamW optimizer at a learning rate of $1e^{-5}$. The DAPL stage is trained for a total of 40k iterations, providing the learned prompts with sufficient capacity to capture the semantic signatures of different degradations before entering the second-stage restoration training.

Stage 2: DCR. In the computation of $\mathcal{L}_{adc}$, the term shared patch embedding refers to the intermediate feature representations extracted from both the VAE encoder and the U-Net encoder. These modules provide complementary information, with the VAE offering low-level structural cues and the U-Net contributing higher-level semantic features. Feature maps are extracted from block 2 of the VAE encoder and blocks 1, 3, 4, and 6 of the U-Net encoder. Multi-scale patches are sampled from these feature maps and used together within the contrastive learning framework. For each anchor patch, the positive sample is the corresponding patch at the same spatial location in the target domain, while negative samples are drawn from different spatial locations within the feature map. For the second-stage restoration training, we adopt a lightweight configuration to ensure stable optimization. The model is trained for 150k steps using the AdamW optimizer and a global batch size of 4. LoRA modules are updated with a learning rate of $1e^{-5}$, using rank 32 for the U-Net layers and rank 16 for the VAE layers. The temperature parameter for $\mathcal{L}_{adc}$ is set to 0.07.

To maintain a strictly unpaired training regime, we enforce complete independence between the degraded and clean domains throughout the learning process. The source domain D_{src} is constructed from collections of degraded images such as Rain800 Zhang et al. (2019), GoPro Nah et al. (2017), and RESIDE Li et al. (2018). These datasets are thoroughly shuffled, and any clean counterparts that might exist in the original datasets are explicitly excluded from training. The target domain D_{tgt} consists solely of high-quality images drawn from DIV2K Agustsson & Timofte (2017) and other clean subsets corresponding to the aforementioned datasets. During each training iteration, a batch of degraded samples B_{src} and a batch of clean samples B_{tgt} are sampled independently from their respective domains. No pixel-level alignment, metadata linkage, or scene overlap exists between the two batches. This ensures that the model learns restoration behavior without relying on any form of paired supervision or implicit correspondence between degraded and clean images.

G Limitations and Future Work

While the proposed UADiff framework demonstrates superior restoration performance and robust generalization for unpaired All-in-One image restoration, it still has certain limitations that point to promising directions for future research.

Computational Cost and Edge Deployment. Although our framework achieves highly efficient parameter tuning, it fundamentally relies on a large-scale pre-trained diffusion backbone. This inherently incurs substantial computational overhead, posing challenges for real-time deployment on resource-constrained edge devices such as mobile phones or low-power drones. To address this limitation, our future work will focus on advanced model compression techniques explicitly tailored for one-step diffusion models. Specifically, we plan to explore structural pruning to eliminate redundant attention heads and channel dimensions. Furthermore, we will investigate low-bit post-training quantization to significantly reduce the memory footprint, and leverage knowledge distillation to transfer the rich generative priors from our heavy foundation model into a compact, lightweight student network. These efforts aim to develop a cutting-edge, lightweight image restoration model that strikes an optimal balance between high-fidelity restoration quality and real-time inference efficiency.

Table 8: Ablation study on the rank settings of the Degradation-Conditional LoRA (DC-LoRA). We evaluate the impact of different rank combinations for the VAE (r_{vae}) and U-Net (r_{unet}) layers.

r_{vae}	r_{unet}	**PSNR** ↑	**SSIM** ↑	**LPIPS** ↓
8	16	24.17	0.820	0.183
16	16	24.31	0.831	0.180
16	32	24.53	0.845	0.171
32	32	24.56	0.839	0.173
32	64	24.49	0.843	0.169

Generalization to Unseen Degradations. UADiff demonstrates strong adaptability within the degradation types seen during training, but its ability to generalize to entirely unseen degradations remains limited. When encountering artifact types that fall outside the predefined prompt set, such as pixelation, glare, or compression noise, the DAPL module tends to map them to the closest known degradation prompt, which may not represent the underlying corruption accurately and can lead to suboptimal restoration. To tackle this, future research will explore open-vocabulary restoration frameworks. By synergizing our architecture with Multi-modal Large Language Models (MLLMs), we aim to dynamically query complex, zero-shot degradation contexts during inference, enabling adaptive restoration for extreme real-world data corruptions without retraining the entire system.

Benchmark and Data Limitations. A key limitation of unpaired AIO restoration lies in the model's strong dependence on the target data distribution. Without strict pixel-level supervision, the network tends to forcefully align degraded input features with the global statistical distribution of the target domain during the generation process. This makes the model highly vulnerable to target domain bias, causing it to generate unnatural textures or semantic shifts in an attempt to match the prior distribution of the training set. Additionally, existing AIO benchmarks typically focus on isolated degradation types or use simple linear combinations to artificially stack degradations, which fail to capture the complex, deeply coupled real-world degradations caused by physical environments and sensor limitations. To address this, our future work will focus on improving both the data and evaluation frameworks. We aim to collect and build a large-scale real-world degradation dataset covering more extreme scenarios (e.g., night-time fog and rain, complex light artifacts) to bridge the domain gap between synthetic data and real-world applications. We will also explore advanced self-supervised learning paradigms and reference-free evaluation metrics, enabling the model to learn robust, decoupled representations directly from naturally degraded images, thereby facilitating more scientific and comprehensive training and testing for AIO restoration models.

H Additional Ablations and Qualitative Results

Sensitivity Analysis of LoRA Ranks. To determine the optimal configuration for our Degradation-Conditional LoRA (DC-LoRA), we conduct an ablation study on the rank settings for the VAE (r_{vae}) and U-Net (r_{unet}) layers, as shown in Table 8. The results reveal that setting the ranks too low (e.g., $r_{vae} = 8$, $r_{unet} = 16$) limits the model's ability to adapt to complex degradations, leading to suboptimal restoration performance. As the ranks are increased to $r_{vae} = 16$ and $r_{unet} = 32$, the framework achieves the best balance, providing the highest SSIM alongside competitive PSNR and LPIPS scores. However, further increasing the ranks (e.g., to $32, 32$ or $32, 64$) yields marginal or inconsistent improvements, with slight drops in SSIM and PSNR. Therefore, we adopt $r_{vae} = 16$ and $r_{unet} = 32$ as the default configuration, achieving an optimal trade-off between parameter efficiency and restoration performance.

Additional Qualitative Results. To further demonstrate the restoration effectiveness and generalization capability of UADiff, we present additional qualitative comparisons on both standard benchmarks and real-world images. Specifically, Figures 9, 10, 11, 12, and 13 show representative results for the five core restoration tasks, where our unpaired framework effectively removes degradations while preserving fine structural details. Moreover, Figures 14 and 15 demonstrate robust generalization on real-world images. Finally, Figure 16 presents the restoration results under complex mixed degradations.

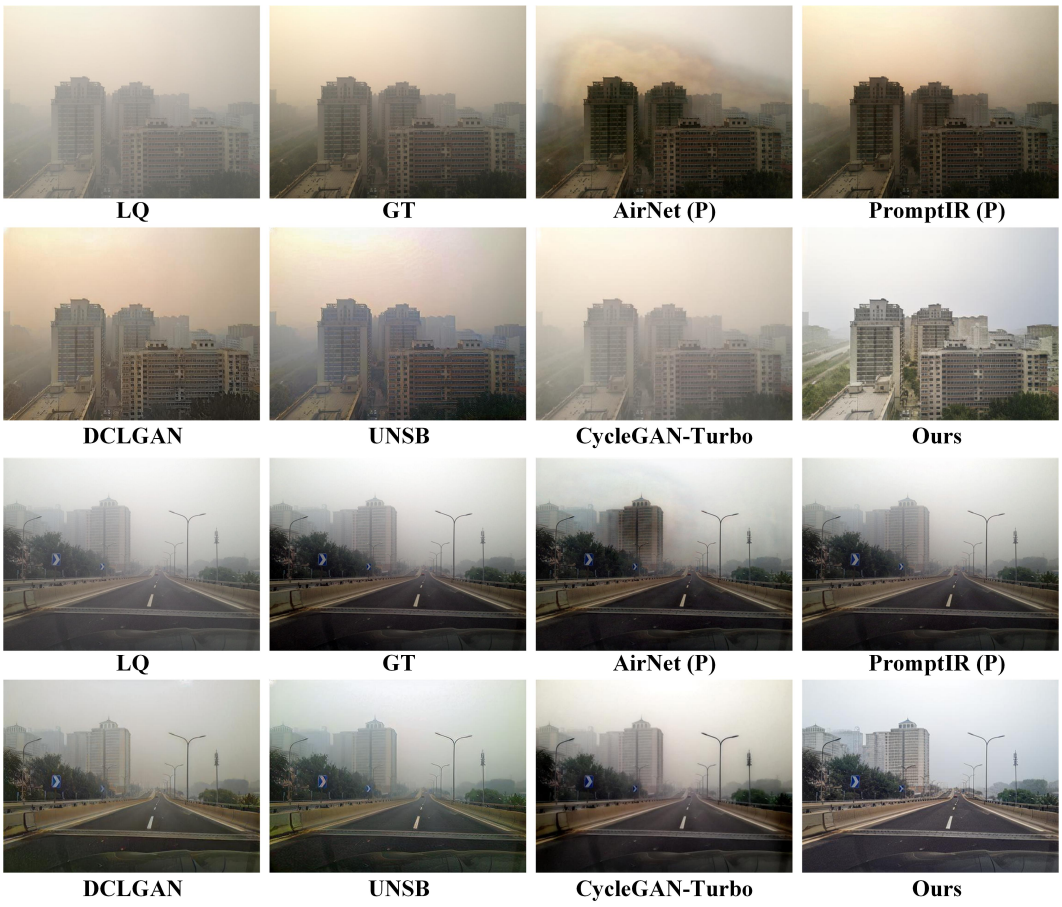

Figure 9: Visual comparison of dehazing results on standard AIO benchmark.

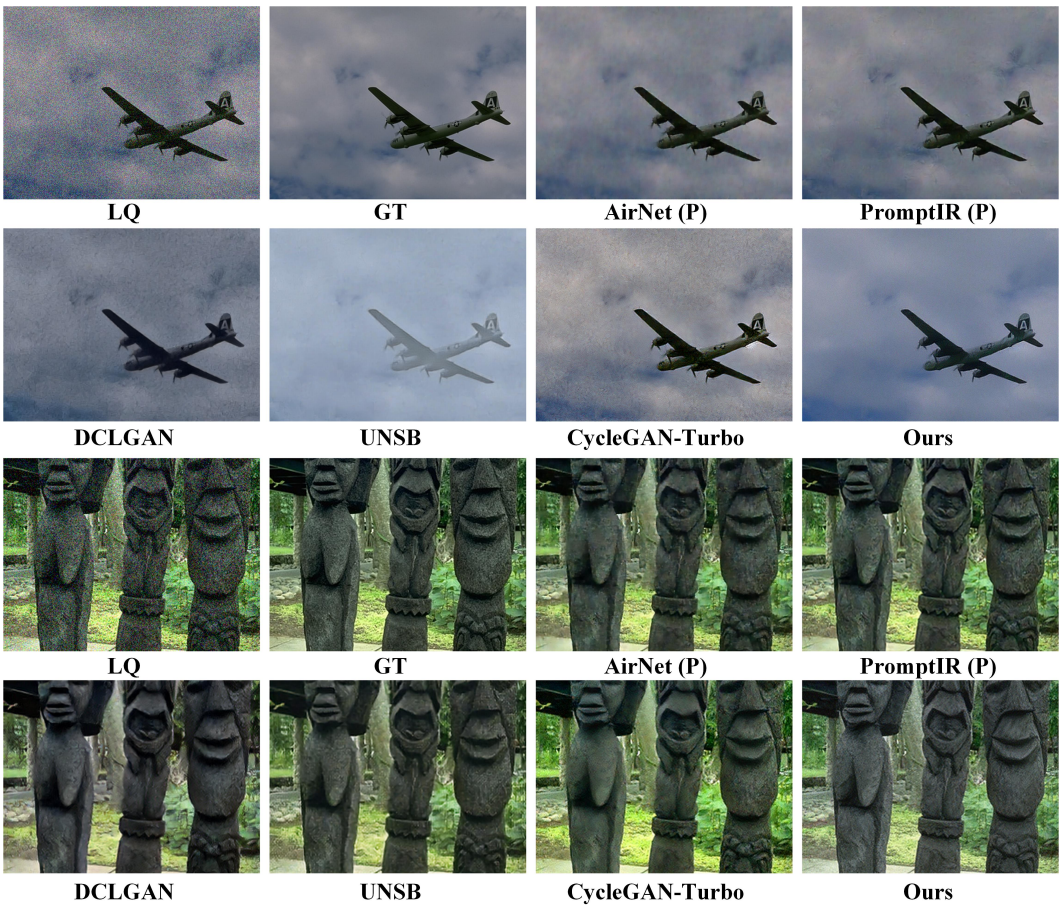

Figure 10: Visual comparison of denoising results on standard AIO benchmark.

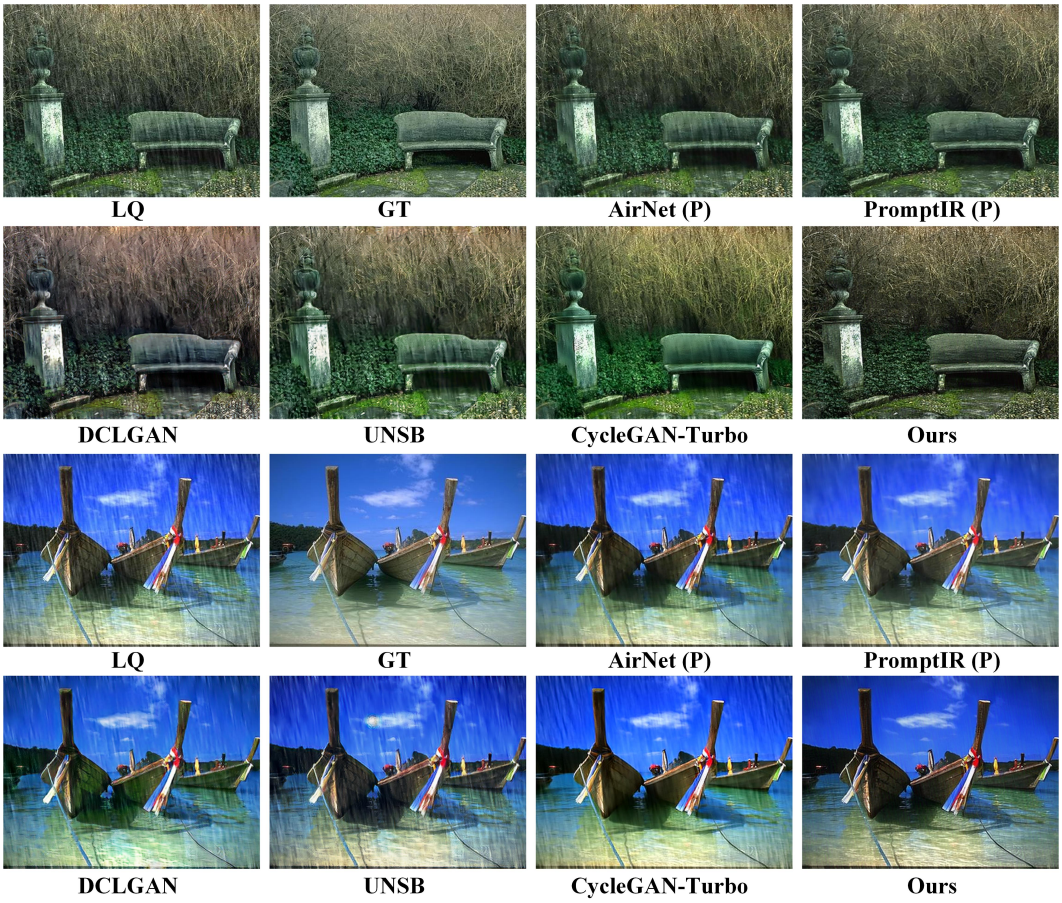

Figure 11: Visual comparison of deraining results on standard AIO benchmark.

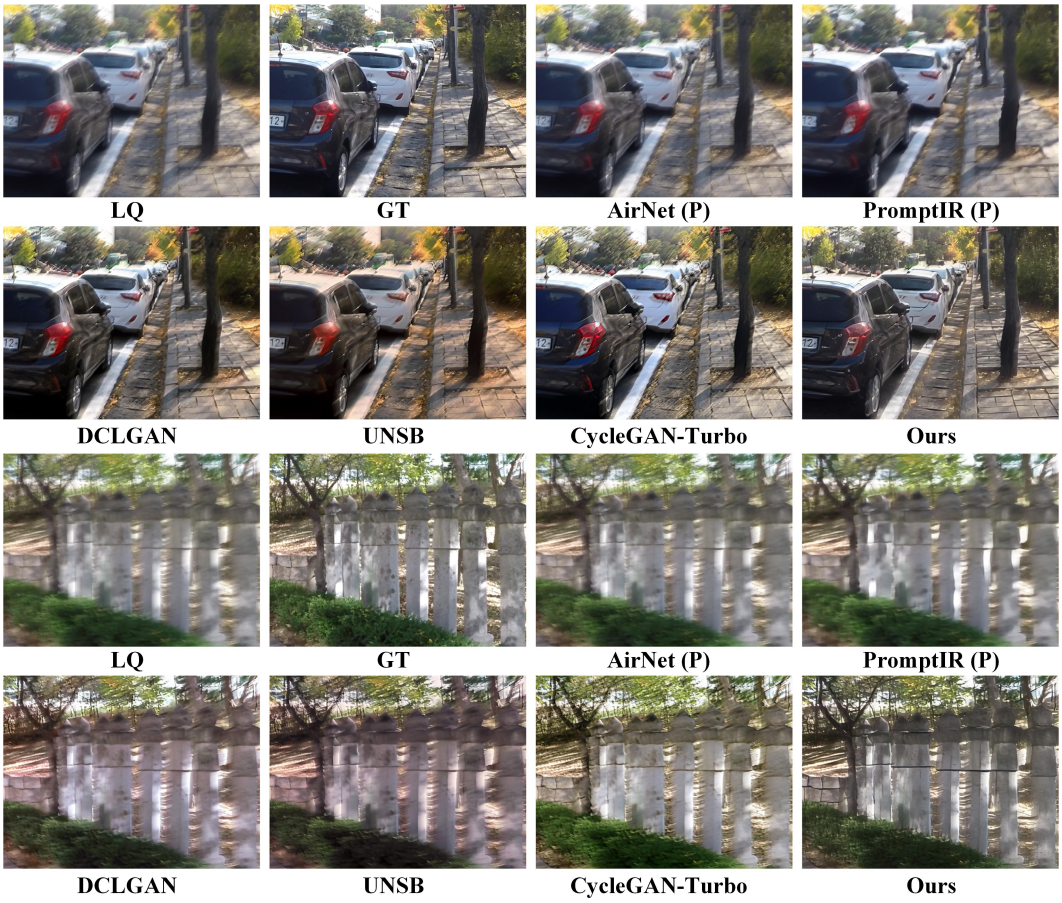

Figure 12: Visual comparison of deblurring results on standard AIO benchmark.

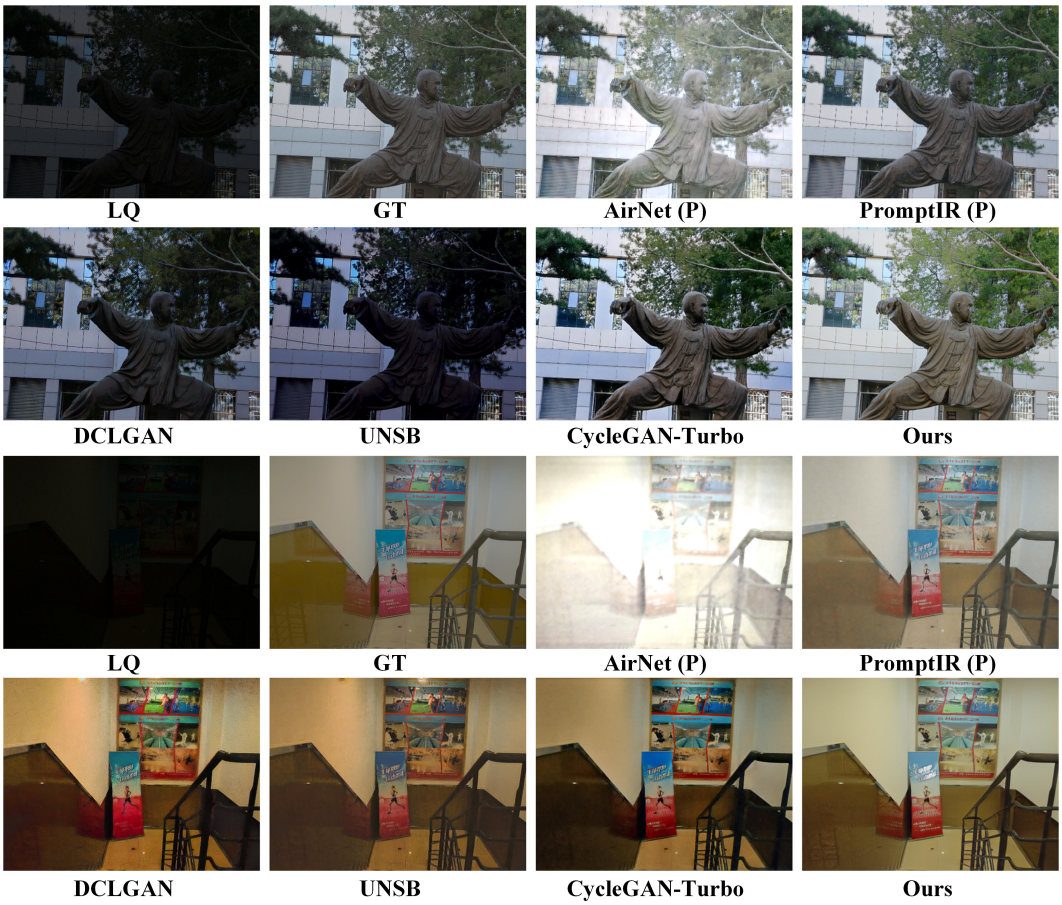

Figure 13: Visual comparison of low-light results on standard AIO benchmark.

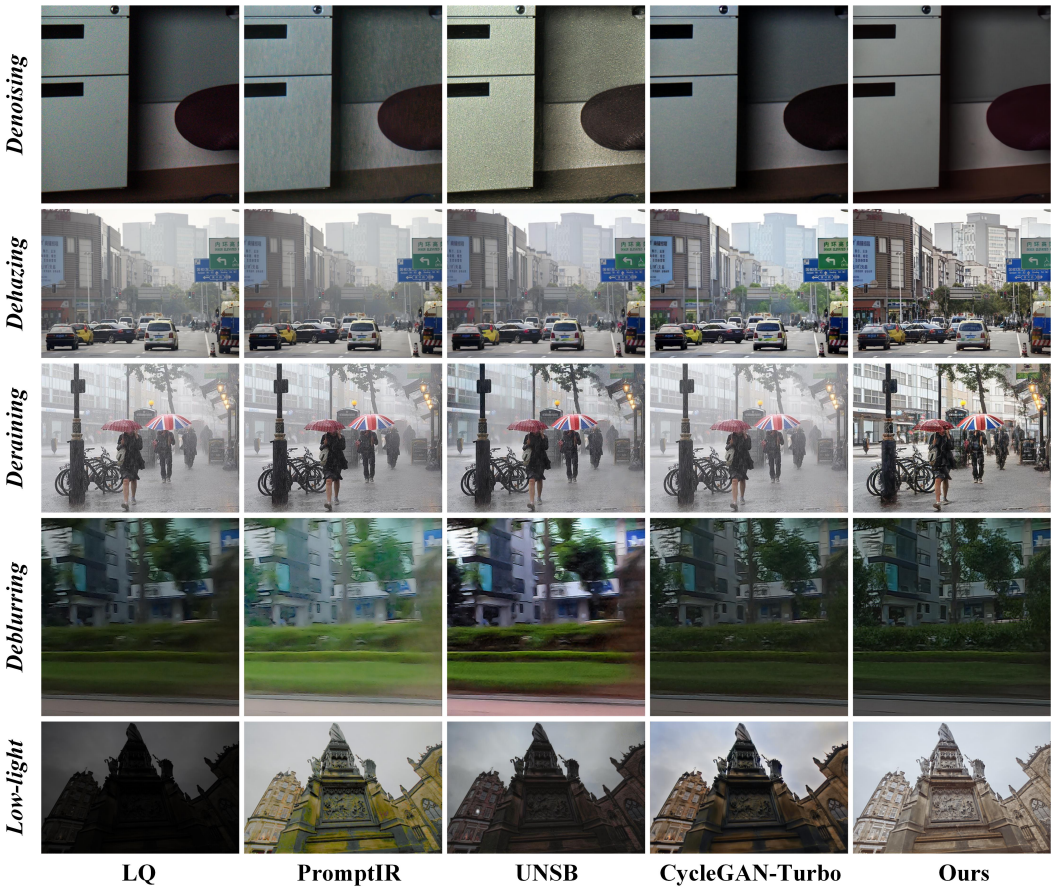

Figure 14: Visual comparison on real-world datasets.

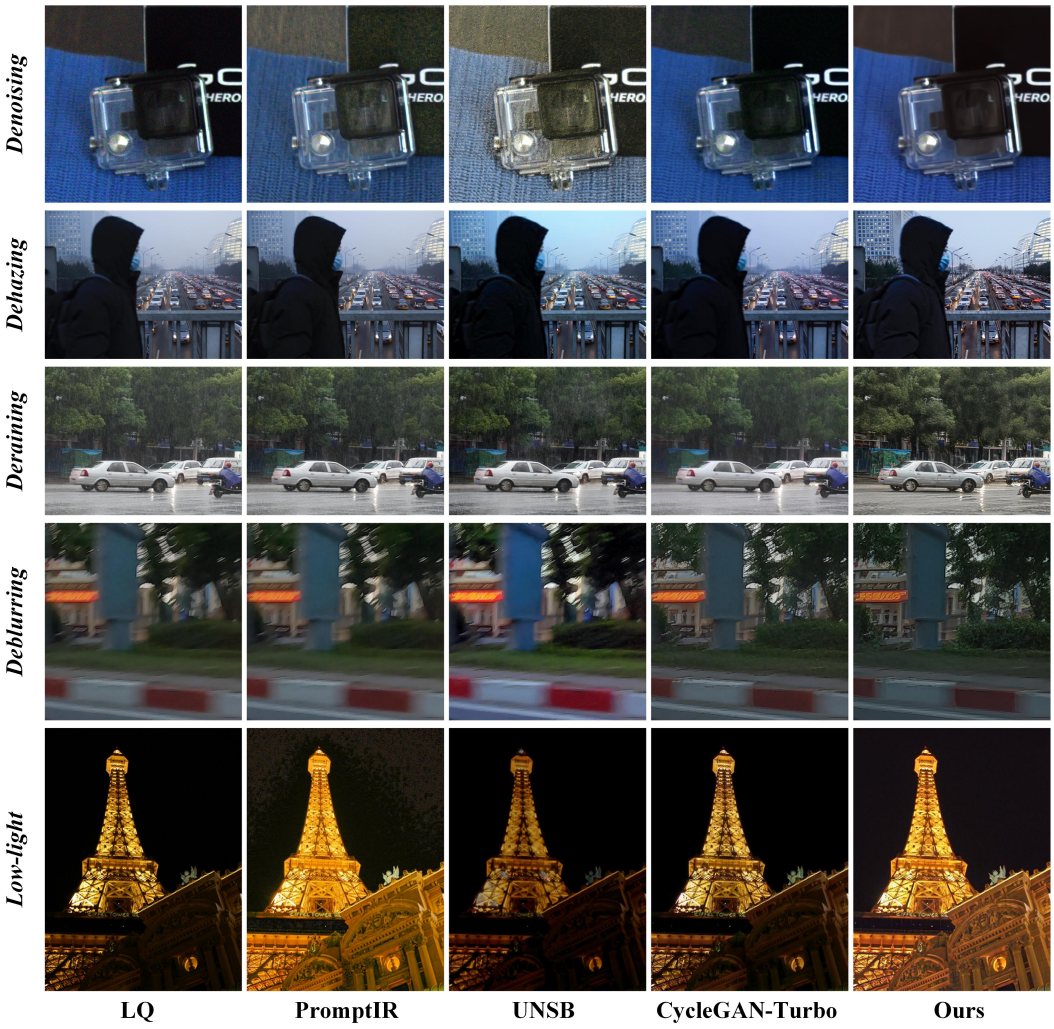

Figure 15: Visual comparison on real-world datasets.

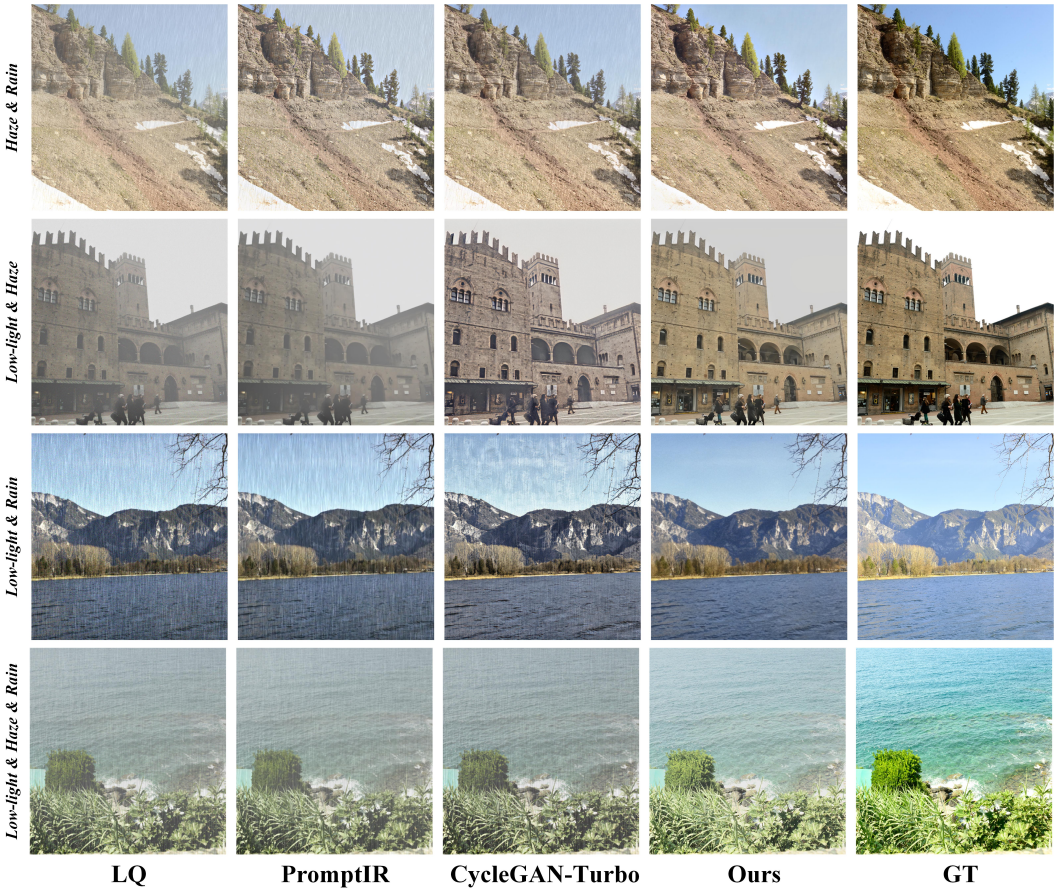

Figure 16: Visual comparison on mixed degradation datasetsGuo et al. (2024).