ISPRS Journal of Photogrammetry and Remote Sensing 240 (2026) 410–429

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

CoRE-UIR: Prior-guided common and residual experts for efficient all-in-one remote sensing image restoration

Zaiyan Zhang [a], Qiangqiang Yuan [a,*], Jie Li [a], Ziyang Lihe [a], Yu Wan [a], Yuzeng Chen [a], Xin Su [b], Liangpei Zhang [c]

[a] School of Geodesy and Geomatics, Wuhan University, Wuhan 430079, Hubei, China
[b] School of Artificial Intelligence, Wuhan University, Wuhan 430072, Hubei, China
[c] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, Hubei, China

ARTICLE INFO

Keywords:
Remote sensing image restoration
All-in-one restoration
Common and residual experts
Vision-language prior
Mixture of experts

ABSTRACT

Remote sensing images acquired by unmanned aerial vehicles (UAVs) and satellites are often degraded by adverse weather, illumination variation, and imaging artifacts, which may co-occur and jointly induce global distribution shifts and local structural corruption. Although All-in-One image restoration offers an appealing unified alternative to task-specific pipelines, existing methods still suffer from weak or implicit degradation cues and parameter redundancy caused by full-rank multi-expert designs with overlapping restoration behaviors. We propose CoRE-UIR (Common and Residual Experts for Universal Image Restoration), a prior-guided global–local framework centered on the Common-and-Residual Expert Block (CoRE). CoRE explicitly decomposes restoration capacity into a common dense expert for degradation-invariant restoration and low-rank residual experts for degradation-specific compensation, enabling adaptive specialization without redundant expert replication. Built on this design, Degradation Prior Embedding (DPE) adapts frozen CLIP features into an explicit restoration-oriented prior, while Global Feature Modulation (GFM) aligns global feature statistics before local residual compensation. We also construct MDVD-108K (Multi-Degradation VisDrone), a large-scale UAV restoration dataset covering both single and compound degradations, together with a real-world test set. Extensive experiments on multiple datasets show that CoRE-UIR improves the overall average PSNR by 1.05 dB while running 11.83× faster and reducing peak memory by 85.3% relative to the strongest baseline, BaryIR, thereby maintaining a favorable quality-efficiency trade-off.Evaluations on downstream tasks and unseen degradation also validate the generalizability of CoRE-UIR. The code and dataset will be released at https://github.com/zzaiyan/CoRE-UIR.

1. Introduction

Remote sensing imagery acquired from unmanned aerial vehicles (UAVs) and satellites plays a pivotal role in applications such as disaster response (J. Wang et al., 2025), environmental monitoring, urban planning, and intelligent target recognition (R. Liu et al., 2024; Chen et al., 2025; Zhang et al., 2026). In practice, however, these images are frequently degraded by adverse weather, illumination variation, and imaging artifacts during acquisition (Shen et al., 2015). Such degradations are diverse, spanning fog, dust, rain, low-light, and blur. They also often appear in compound forms, jointly reducing visibility, shifting appearance statistics, corrupting local structures, and undermining downstream perception and interpretation tasks (Yuan et al., 2020).

Task-specific restoration models remain the dominant solution for individual degradations, such as dehazing (Song et al., 2023; Li et al., 2019), deraining (Zamir et al., 2021), and low-light enhancement (Cai et al., 2023). While effective for single tasks, maintaining one model per degradation is poorly suited to remote sensing pipelines that must process heterogeneous or overlapping degradations within a unified workflow. These limitations motivate All-in-One image restoration (AiOIR), which seeks a single restoration model for multiple degradation types (Jiang et al., 2025).

Despite recent progress, existing AiOIR methods still face three coupled limitations in remote sensing scenarios. First, degradation cues are often weak, implicit, or insufficiently coupled to the restoration backbone, making it difficult to organize degradation-aware processing in a

* Corresponding author.
E-mail addresses: zzaiyan@whu.edu.cn (Z. Zhang), qqyuan@sgg.whu.edu.cn (Q. Yuan), jli89@sgg.whu.edu.cn (J. Li), Ziyang_Lihe@whu.edu.cn (Z. Lihe), 2024282140091@whu.edu.cn (Y. Wan), yuzeng_chen@whu.edu.cn (Y. Chen), xinsu.rs@whu.edu.cn (X. Su), zlp62@whu.edu.cn (L. Zhang).

https://doi.org/10.1016/j.isprsjprs.2026.07.028
Received 12 May 2026; Received in revised form 17 June 2026; Accepted 28 July 2026
Available online 8 August 2026

stable manner. Even when pre-trained vision-language models such as CLIP (Radford et al., 2021) are introduced, their representations are often used only as auxiliary hints rather than restoration-oriented priors. Since CLIP is trained on billion-scale vision-language data, parameter-efficient adaptation provides a practical way to preserve its strong generalization while translating semantic features into restoration-oriented degradation representations. Second, many adaptive restorers expand capacity through Mixture-of-Experts (MoE) (Shazeer et al., 2017) or multi-branch designs that replicate full-rank experts, leading to parameter redundancy and repeated learning of similar restoration behaviors. Third, remote sensing degradations often mix global statistical shifts with local structural corruption. Fog and dust may alter scene-level color, contrast, and visibility, yet their intensity can still vary with depth and layout, while rain introduces sparse local streaks. Likewise, most degradations require shared restoration abilities such as contrast enhancement, texture refinement, and brightness normalization, but some also need specialized operations such as rain-streak removal, motion compensation, or low-light denoising.

These observations motivate two complementary decompositions for universal restoration. We decouple degradation adaptation into global modulation and local compensation: the former handles scene-wide shifts in visibility, contrast, and color statistics, whereas the latter addresses spatially varying degradation patterns and structural corruption. We further decouple restoration capacity into a common dense expert and specialized residual experts: the common expert captures restoration behavior shared across degradations, while the residual experts model degradation-specific corrections without replicating full-rank branches.

Based on this view, we propose **CoRE-UIR** (*Common and Residual Experts for Universal Image Restoration*), a prior-guided global–local framework for efficient AiOIR. CoRE-UIR first uses Degradation Prior Embedding (DPE) to adapt frozen CLIP features from global and local image views into a restoration-oriented degradation prior. Conditioned on this prior, Global Feature Modulation (GFM) performs prior-state global modulation at each stage entrance to organize intermediate features into a compact, prior-consistent, and more routable space. We further introduce the **Common-and-Residual Expert Block (CoRE)**, which couples the original dense backbone block, treated as a common dense expert, with low-rank residual experts for degradation-specific local compensation. The residual experts are sparsely selected by a lightweight Top-k router, allowing multiple residual experts to be activated for compound degradations. These components are instantiated in a U-shaped global–local restoration backbone, with CoRE serving as the primary structural innovation.

To evaluate these design choices under remote sensing settings, we construct MDVD-108K (*Multi-Degradation VisDrone Dataset*), a large-scale UAV multi-degradation dataset built on VisDrone (Zhu et al., 2021), containing six single degradations, six compound degradations, and real-world degraded UAV images for qualitative evaluation. We further perform satellite-domain validation on MDRS-Landsat.

Our main contributions are summarized as follows:

- A prior-guided global–local restoration framework, CoRE-UIR, is introduced to adapt frozen CLIP features into restoration-oriented degradation priors and inject them into a unified backbone through global modulation and local expert compensation.
- We propose the Common-and-Residual Expert Block (CoRE), which decomposes restoration capacity into a common dense expert and low-rank residual experts, reducing the redundancy of full-rank MoE branches while retaining degradation-adaptive compensation.
- We construct MDVD-108K, a large-scale UAV restoration benchmark comprising 108K paired synthetic samples with object detection annotations, covering six single degradations, six compound degradations, and real-world degraded UAV images.
- Extensive experiments on UAV and satellite benchmarks show that CoRE-UIR consistently improves restoration quality and achieves a stronger quality-efficiency trade-off than strong universal restoration baselines across single and compound degradations.

The remaining sections of the paper are organized as follows. Section 2 presents related work. Section 3 describes the proposed method. Section 4 reports the experiments and analysis. Finally, Section 5 concludes the paper.

2. Related work

2.1. Single-task image restoration

Single-task image restoration is still dominated by task-specific or fixed-degradation models. In natural images, representative progress spans early CNN-based restoration and super-resolution methods such as SRCNN (Dong et al., 2015) and RCAN (Y. Zhang et al., 2018), generic transformer restorers such as IPT (Chen et al., 2021), SwinIR (Liang et al., 2021), CAT (Z. Chen et al., 2022), and HAT (Chen et al., 2023), and degradation-specific models for dehazing (Song et al., 2023; Liu et al., 2026), low-light enhancement (Cai et al., 2023), dust removal (Wei et al., 2025), together with generic backbones such as MPRNet (Zamir et al., 2021), Restormer (Zamir et al., 2022), DGUNet (Mou et al., 2022), NAFNet (L. Chen et al., 2022), and MambaIR (Guo et al., 2024). More recent progress also includes newer transformer variants, task-specific weather restoration models, and deblurring benchmarks (Jin et al., 2025; Zhang et al., 2022, 2021; Wen et al., 2026; K. Zhang et al., 2023). These methods are effective when the degradation type is known and fixed, but they do not address the unified setting in which one model must adapt to heterogeneous degradations.

The same paradigm persists in remote sensing restoration. Earlier studies often relied on handcrafted low-rank and total-variation priors for hyperspectral denoising and image completion (Cheng et al., 2014; He et al., 2015). More recent deep models focus on task-specific cloud removal, missing-data reconstruction, and multimodal restoration, including multitemporal sequence modeling (Stucker et al., 2023; Z. Zhang et al., 2025; Shu et al., 2025), transformer or diffusion-based cloud removal (Li et al., 2020, 2025; Jing et al., 2023; Sui et al., 2024), remote-sensing dehazing (Wen et al., 2023, 2025a), panshapening (Cui et al., 2025a,b) and SAR-optical or multimodal fusion strategies (Pan et al., 2024; P. Wang et al., 2025; Z. Zhang et al., 2026a,b; Chen et al., 2026). Although these methods leverage temporal, sensor, or generative priors, they are still specialized to isolated restoration families such as cloud removal or missing-data recovery. As a result, they provide limited guidance for remote sensing scenarios where multiple degradations may co-exist within a unified deployment pipeline.

2.2. All-in-one image restoration

AiOIR relaxes the fixed-degradation assumption by training a single model for multiple degradations. Early representative methods include AirNet (Li et al., 2022), TransWeather (Valanarasu et al., 2022), WeatherDiff (Özdenizci et al., 2023), ProRes (Ma et al., 2023), PromptIR (Potlapalli et al., 2024), and IDR (J. Zhang et al., 2023), which explore degradation encoding, weather-oriented modeling, diffusion-based reconstruction, prompt conditioning, and ingredient-oriented learning. Later methods strengthen universal restoration through improved prompts, stronger degradation representations, or modular adaptation, such as sequential/prompt learning (Kong et al., 2024), EvoIR (Ma et al., 2025), DACLIP-UIR (Luo et al., 2024), and BaryIR (Tang et al., 2026). Related extensions further explore weather-oriented (Gao et al., 2023; Wen et al., 2025b) and perception-oriented formulations (X. Zhang et al., 2026a; Hu et al., 2025; X. Zhang et al., 2026b; Wang et al., 2026), as well as low-rank and domain-specialized variants (Ai et al., 2024; C. Zhang et al., 2024; X. Zhang et al., 2025).

This universal-restoration paradigm has only recently been extended to remote sensing, partly enabled by multi-degradation benchmarks such as MDRS-Landsat introduced in Ada4DIR (Lihe et al., 2025). Related prompt-based universal restoration has also been explored for hyperspectral imagery (Wu et al., 2025) and multi-modal imagery (Cui and Liu, 2026). Ada4DIR (Lihe et al., 2025) enhances multi-level degradation extraction and adaptive degradation recognition through prompt-injection-fusion and model-driven prompt blocks, while Phy-DAE (Dong et al., 2026) introduces physics-guided degradation-adaptive experts with progressive degradation mining and sparse activation. These methods highlight the value of explicit degradation modeling and remote-sensing-specific priors, yet they still do not explicitly decompose restoration capacity into a common dense path and lightweight low-rank residual experts, nor do they fully address the redundancy of heavy expert branches under compound degradations.

2.3. Mixture of experts

Mixture of Experts (MoE) offers an effective mechanism for scaling model capacity while keeping per-sample computation sparse (Shazeer et al., 2017; Riquelme et al., 2021). The key idea is to maintain multiple experts and use a gating function to activate only a subset of them for each input, thereby improving specialization without evaluating the entire expert pool. This design has been widely explored in large-scale representation learning and provides a natural tool for modeling heterogeneous degradations in image restoration.

In AiOIR, MoE-style designs have been used to allocate degradation-dependent capacity through routed expert branches or adaptive expert modules. Representative examples include multi-expert adaptive selection (Yu et al., 2024), feature-modulated experts (R. Zhang et al., 2024), CLIP-guided MoE gating in M2Restore (Y. Wang et al., 2025), complexity-aware experts in MoCE-IR (Zamfir et al., 2025), and degradation-adaptive experts in remote sensing restoration (Dong et al., 2026). These studies show that routed experts are effective for handling heterogeneous degradations and compound corruptions. However, most existing MoE-based restorers allocate adaptation capacity through full-rank or heavily overlapping experts. In universal restoration, many operations such as visibility enhancement, denoising, and structure recovery are broadly shared across degradations, so expert replication can introduce redundancy in parameters, computation, and learned behaviors. This limitation is especially noticeable under compound degradations, where multiple routed branches may still repeat common restoration operations.

These observations motivate our common-residual expert design, where the dense backbone preserves common restoration capacity and the routed low-rank residual branch focuses on degradation-specific residual compensation. CoRE can also be viewed as an asymmetric sparse MoE, combining an always-active common expert with selectively activated low-rank residual experts to reduce parameter redundancy relative to full-rank expert replication. In this way, expert routing remains adaptive, while expert allocation becomes more redundancy-aware for universal remote sensing restoration.

3. Methodology

In this section, we first introduce the problem formulation and the overall framework, then describe the key components of the proposed CoRE-UIR, and finally present the training strategy (see Fig. 1).

3.1. Overview

3.1.1. Problem formulation

Let $\mathcal{D} = \{d_m\}_{m=1}^{M}$ denote the set of M base degradation types in a given restoration setting, and let $\mathcal{C} \subseteq \{\mathcal{S} \mid \emptyset \neq \mathcal{S} \subseteq \mathcal{D}\}$ denote the degradation configurations, where each $\mathcal{S} \in \mathcal{C}$ specifies the active degradation subset. For a clean remote sensing image $I_{\mathrm{gt}} \in \mathbb{R}^{H\times W\times 3}$, the degraded observation is modeled as:

$$I_{\mathrm{deg}} = \mathcal{G}_{\mathcal{S}}\left(I_{\mathrm{gt}}\right), \quad \mathcal{S} \in \mathcal{C}, \tag{1}$$

where $\mathcal{G}_{\mathcal{S}}(\cdot)$ denotes the structured degradation operator induced by the active subset. Single degradation corresponds to $|\mathcal{S}| = 1$, where the observation is generated by one elemental operator, while compound degradation corresponds to $|\mathcal{S}| > 1$, where multiple operators are composed in sequence.

The goal is to learn one universal restoration model for all degradation configurations in the same setting:

$$\hat{I} = f_{\theta}\left(I_{\mathrm{deg}}\right), \quad \forall\, \mathcal{S} \in \mathcal{C}, \tag{2}$$

where f_{θ} shares one set of parameters across all base and compound degradations.

3.1.2. Framework overview

As summarized in Eqs. (3) and (4), CoRE-UIR consists of Degradation Prior Embedding (DPE) and the Global–Local Adaptive Network (GLA-Net). DPE converts one global view and one local view of the degraded image into a restoration-oriented prior embedding z_{deg} using a frozen CLIP image encoder and a lightweight adapter, so that the prior space remains generalizable while becoming more suitable for restoration.

$$z_{\mathrm{deg}} = \mathrm{DPE}\left(I_{\mathrm{deg}}\right) \in \mathbb{R}^{d_z}, \tag{3}$$

$$\hat{I} = \text{GLA-Net}\left(I_{\mathrm{deg}}, z_{\mathrm{deg}}\right) \in \mathbb{R}^{H\times W\times 3}. \tag{4}$$

The Global–Local Adaptive Network (GLA-Net) restores the image conditioned on z_{deg}. It adopts a shared U-shaped backbone with two complementary modules: Global Feature Modulation (GFM), which injects stage-level global conditioning, and the Common-and-Residual Expert block (CoRE), which performs block-level local restoration. Within CoRE, the original dense block serves as the common expert, while routed low-rank residual experts provide degradation-specific compensation.

3.2. Degradation prior embedding

Frozen CLIP features provide a strong starting point for degradation reasoning, but they are optimized for semantic alignment rather than restoration control and therefore cannot be used directly as restoration priors. Benefiting from CLIP pretraining on billion-scale vision-language data, parameter-efficient adaptation can preserve its strong generalization while translating semantic representations into restoration-oriented degradation embeddings. We design Degradation Prior Embedding (DPE) to adapt CLIP features into a compact degradation space with minimal trainable overhead. By combining a global view and a local crop, DPE preserves both scene-level degradation context and local degradation textures before lightweight adaptation.

3.2.1. Multi-scale prior extraction

Raw UAV aerial images typically exhibit high resolution and a spatially uniform scale, making direct full-resolution encoding computationally expensive and insensitive to both global degradation patterns (*e.g.*, overall haze opacity) and local degradation details (*e.g.*, fine rain streaks). We therefore construct two complementary views from I_{deg}:

$$I_{\mathrm{g}} = \mathrm{Resize}\left(I_{\mathrm{deg}}\right), \quad I_{\mathrm{l}} = \mathrm{Resize}\left(\mathrm{Crop}\left(I_{\mathrm{deg}}\right)\right), \tag{5}$$

where both views are scaled to the CLIP encoder's native input resolution. The global view I_{g} preserves overall degradation context, while the local view I_{l} captures regional degradation textures. Their encoder outputs are fused by concatenation:

$$z_{\mathrm{clip}} = \mathrm{Concat}\left(\mathrm{CLIP}\left(I_{\mathrm{g}}\right), \mathrm{CLIP}\left(I_{\mathrm{l}}\right)\right) \in \mathbb{R}^{2d_{\mathrm{clip}}}, \tag{6}$$

where d_{clip} is the CLIP feature dimension of each individual view. This multi-scale fusion combines complementary degradation cues from global context and local textures before projection.

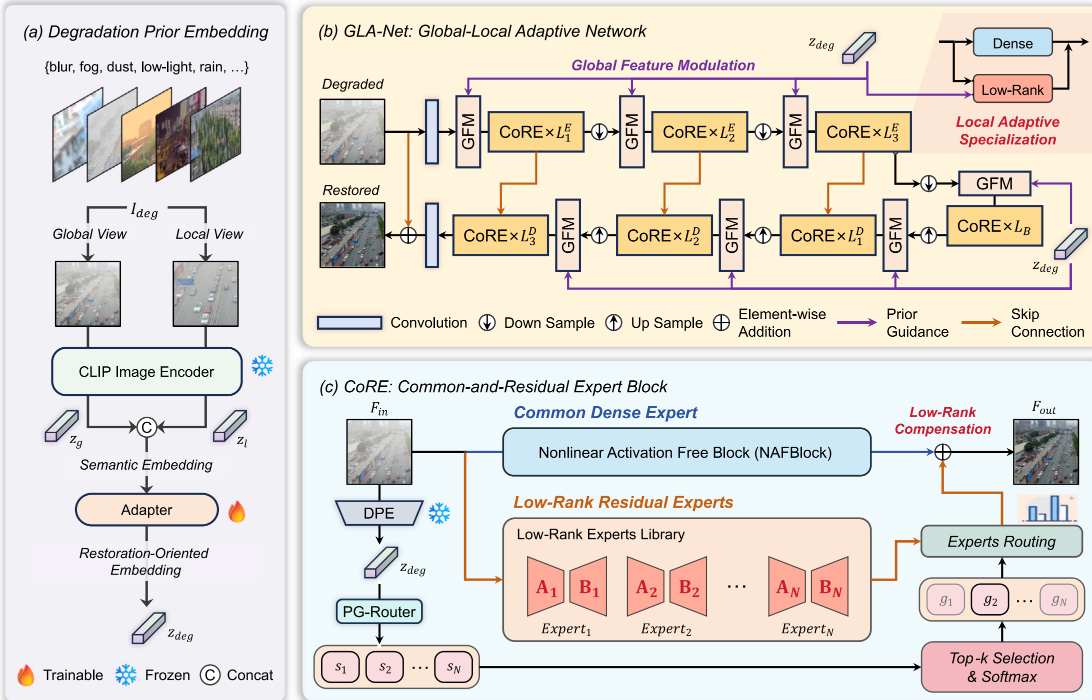

Fig. 1. The overall framework of the proposed CoRE-UIR. (a) Degradation Prior Embedding (DPE): a frozen CLIP image encoder processes multi-scale views of the input, and a lightweight adapter maps the features into restoration-oriented degradation embeddings. (b) Global–Local Adaptive Network: the unified restoration backbone integrates Global Feature Modulation (GFM) and Common-and-Residual Expert Blocks (CoRE). (c) The CoRE block decomposes restoration into a common dense expert for degradation-invariant restoration and low-rank residual experts for degradation-specific compensation.

3.2.2. Parameter-efficient adaptation

Although the frozen CLIP encoder offers strong generalization, its pretraining objective targets high-level semantic alignment rather than low-level restoration guidance. We therefore introduce a lightweight learnable adapter to bridge this gap:

$$z_{\text{deg}} = \text{Adapter}\left(z_{\text{clip}}\right) \in \mathbb{R}^{d_z}, \tag{7}$$

where Adapter$(\cdot)$ consists of two linear layers with a GELU activation, mapping z_{clip} to a compact restoration-oriented embedding z_{deg} of dimension d_z. To make this embedding discriminative for mixed degradation configurations, we attach a linear classification head to z_{deg} and optimize a multi-label degradation classification objective, while keeping the CLIP encoder frozen and updating only the adapter and classifier.

This parameter-efficient adaptation rapidly aligns the semantic feature space with the remote-sensing restoration domain. After convergence, the classification head is discarded and the resulting DPE is used to condition both GFM and CoRE during restoration learning.

3.3. Global feature modulation (GFM)

An explicit degradation prior alone is insufficient for feature modulation, because the same degradation may appear with different severity and intermediate feature responses across images and network stages. GFM therefore combines the degradation prior with the current feature state to perform lightweight degradation-aware channel recalibration, as illustrated in Fig. 2.

Given a feature map $F \in \mathbb{R}^{H' \times W' \times C}$ from the backbone and the degradation prior $z_{\text{deg}} \in \mathbb{R}^{d_z}$, GFM extracts a state descriptor from the current feature response and a prior descriptor from the degradation embedding:

$$f_{\text{state}} = \text{Linear}\left(\text{GAP}\left(F\right)\right) \in \mathbb{R}^{C}, \tag{8}$$

$$f_{\text{prior}} = \text{Linear}\left(z_{\text{deg}}\right) \in \mathbb{R}^{C}, \tag{9}$$

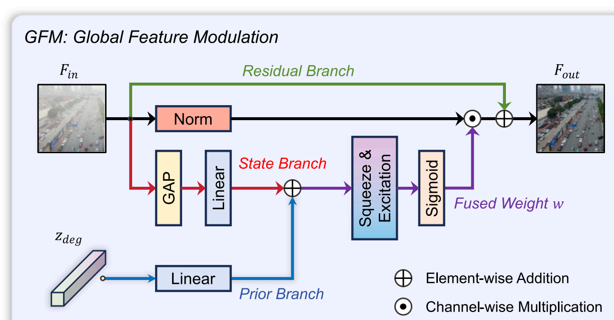

Fig. 2. Architecture of the proposed **Global Feature Modulation (GFM)** module. GFM combines the degradation prior with the current feature state to perform lightweight degradation-aware channel recalibration before CoRE processing.

where GAP$(\cdot)$ denotes global average pooling across spatial dimensions. The two descriptors are then fused and passed through a bottleneck excitation head:

$$f_{\text{fused}} = f_{\text{state}} + f_{\text{prior}}, \tag{10}$$

$$w = \sigma\left(\text{MLP}_{\text{SE}}\left(f_{\text{fused}}\right)\right) \in \mathbb{R}^{C}, \tag{11}$$

where $\text{MLP}_{\text{SE}}(\cdot)$ denotes a two-layer bottleneck network with GELU activation and reduction ratio ρ, and $\sigma(\cdot)$ denotes the Sigmoid function. The resulting gate modulates the feature map as:

$$\text{GFM}\left(F\right) = F + \text{LN}\left(F\right) \odot w, \tag{12}$$

where LN$(\cdot)$ denotes LayerNorm, and $\odot$ denotes broadcast channel-wise multiplication over spatial dimensions.

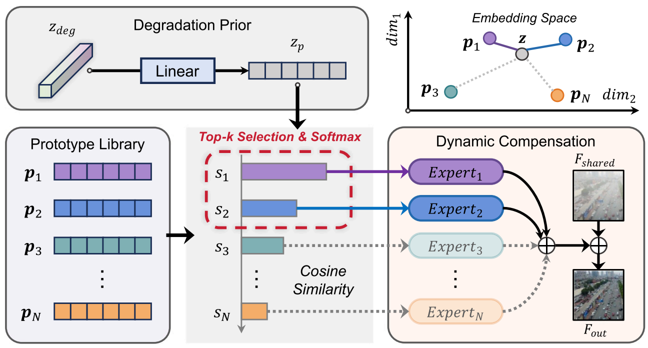

Fig. 3. Architecture of the proposed **Prototype-Guided Router (PG-Router)**, which selects the most relevant low-rank residual experts for each input sample based on degradation archetypes.

GFM thus extends the squeeze-and-excitation paradigm (Hu et al., 2018) with an explicit degradation prior branch, providing compact global modulation before downstream CoRE processing.

3.4. Common-and-residual expert block (CoRE)

After GFM organizes global feature statistics, the restoration network still requires degradation-specific residual transformations to correct locally varying structural corruption. Since much of the restoration process is shared across degradations, replicating multiple full-rank experts is unnecessarily costly. We therefore introduce the Common-and-Residual Expert Block (CoRE), which is inserted in parallel with each backbone block and decomposes restoration capacity into a **common dense expert** and **low-rank residual experts**. Borrowing the low-rank bottleneck from LoRA-style adaptation, CoRE instead couples an always-active common expert with routed low-rank residual experts, yielding an asymmetric sparse-MoE form rather than fixed low-rank updates to a frozen backbone.

Specifically, the dense backbone block captures restoration knowledge shared across degradations, such as visibility enhancement, texture recovery, and structure reconstruction, while the routed low-rank branch focuses on degradation-specific residual compensation. The low-rank constraint limits the capacity of each residual expert, discouraging repeated learning of common operations and instead driving the experts to model degradation-dependent differences. This common-residual decomposition improves parameter efficiency while preserving adaptive capacity.

3.4.1. Low-rank residual expert library

Each CoRE block maintains a library of N low-rank expert pairs $\{(A_n, B_n)\}_{n=1}^{N}$, where $A_n \in \mathbb{R}^{r\times C}$ and $B_n \in \mathbb{R}^{C\times r}$ are learnable projection matrices with bottleneck rank $r \ll C$, applying to the channel dimension at each spatial location. Each pair, therefore, forms a compact low-rank residual expert. In implementation, A_n and B_n are implemented by two successive 3×3 convolutions that map $C \to r$ and $r \to C$, respectively. This preserves a low-rank bottleneck along the channel dimension while enlarging the local receptive field.

3.4.2. Prototype-guided router (PG-router)

We adopt a PG-Router to select low-rank residual experts, as shown in Fig. 3. We maintain N prototype vectors $\{p_n\}_{n=1}^{N}$, where each $p_n \in \mathbb{R}^{d_z}$ represents a learnable degradation archetype in the prior embedding space. Given the original degradation prior embedding z_{deg}, the router first maps it to a routing query:

$$z_{\text{p}} = \text{Linear}\left(z_{\text{deg}}\right) \in \mathbb{R}^{d_z}, \tag{13}$$

The routing logit for the nth expert is then computed as the temperature-scaled cosine similarity between the normalized routing query and the corresponding prototype:

$$s_n = \tau \cdot \hat{z}_{\text{p}}^{\top} \hat{p}_n, \quad n = 1, \dots, N, \tag{14}$$

where $\hat{z}_{\text{p}} = z_{\text{p}}/\|z_{\text{p}}\|_2$ and $\hat{p}_n = p_n/\|p_n\|_2$ are the ℓ_2-normalized embeddings, and $\tau = \exp(t)$ is a learnable temperature controlling the sharpness of the routing distribution. The routing weights are then obtained by retaining the k highest logits and normalizing:

$$g = \text{Softmax}\big(\text{TopK}\,(s,\, k)\big) \in \mathbb{R}^{N}, \tag{15}$$

where $\text{TopK}(\cdot, k)$ masks the remaining $N - k$ values to $-\infty$ before Softmax. This cosine-similarity-based design offers a natural interpretation: each prototype encodes a degradation archetype, and the routing score reflects how closely the current sample's degradation style aligns with that archetype. Sparse Top-k activation is especially suitable for compound degradations because multiple low-rank experts can be activated simultaneously, while irrelevant experts remain suppressed.

3.4.3. Low-rank residual compensation

The activated experts are aggregated to form a sample-adaptive low-rank residual branch. Meanwhile, the original backbone block serves as the common dense expert:

$$E_{\text{com}}\left(F_{\text{in}}\right) = \text{BasicBlock}\left(F_{\text{in}}\right). \tag{16}$$

This dense expert preserves the full channel capacity of the backbone and is responsible for degradation-invariant restoration behavior. For an input feature $F \in \mathbb{R}^{H'\times W'\times C}$, the nth low-rank expert first produces an individual residual response:

$$E_{\text{res},n}^{\text{LR}}(F) = B_n\left(A_n(F)\right), \tag{17}$$

$$E_{\text{res}}^{\text{LR}}\left(F; z_{\text{deg}}\right) = \sum_{n\in\mathcal{K}} g_n \cdot E_{\text{res},n}^{\text{LR}}(F), \tag{18}$$

where $\mathcal{K}$ denotes the set of Top-k activated experts and g_n is the corresponding routing weight produced by the PG-Router conditioned on z_{deg}.

The final CoRE output combines the common dense expert and the routed low-rank residual experts through residual addition:

$$F_{\text{out}} = \underbrace{E_{\text{com}}\left(F_{\text{in}}\right)}_{\text{common dense expert}} + \underbrace{E_{\text{res}}^{\text{LR}}\left(F_{\text{in}}; z_{\text{deg}}\right)}_{\text{low-rank residual experts}}, \tag{19}$$

This formulation makes the common-residual division explicit: the common dense expert provides shared restoration capability, while the routed low-rank residual experts provide degradation-specific residual compensation. The efficiency gain comes from avoiding full-rank expert replication rather than sacrificing adaptive capacity.

3.5. Global–local adaptive network (GLA-Net)

We instantiate the universal restoration framework as a shared U-shaped encoder–decoder backbone with skip connections. The backbone contains several encoder stages, a bottleneck, and several decoder stages connected by strided downsampling and PixelShuffle upsampling (Shi et al., 2016). Each basic block is implemented by NAFBlock (L. Chen et al., 2022). In implementation, GFM is applied once at the entrance of each stage, while CoRE is applied inside the stage for block-wise refinement.

Concretely, let F_{in}^{i} and F_{out}^{i} denote the input and output features of the ith stage, respectively, and let L_i be the number of blocks in that stage. The backbone first applies GFM once to the stage input and then recursively updates the feature through the stacked blocks:

$$F_0^{i} = \text{GFM}\left(F_{\text{in}}^{i}; z_{\text{deg}}\right), \tag{20}$$

$$F_j^{i} = \text{CoRE}_j^{i}\left(F_{j-1}^{i}; z_{\text{deg}}\right), \tag{21}$$

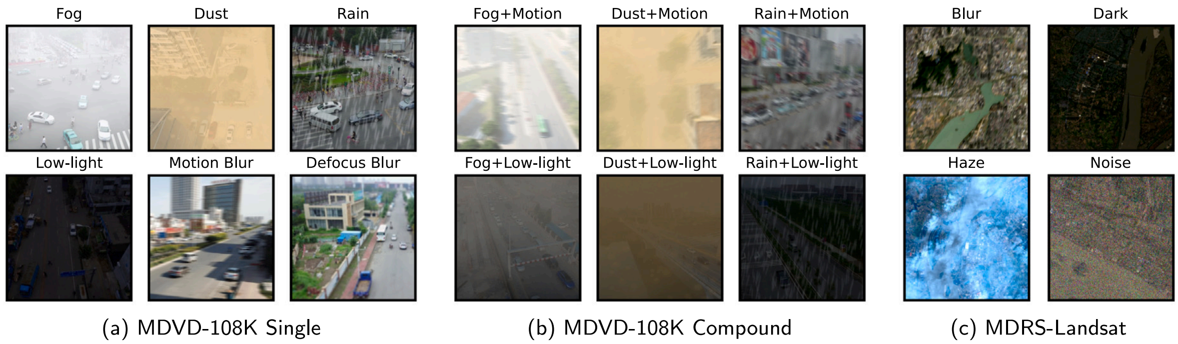

Fig. 4. Representative samples of the MDVD-108K and MDRS-Landsat datasets, showing (a) single degradations UAV samples, (b) compound degradations UAV samples, and (c) satellite samples from MDRS-Landsat.

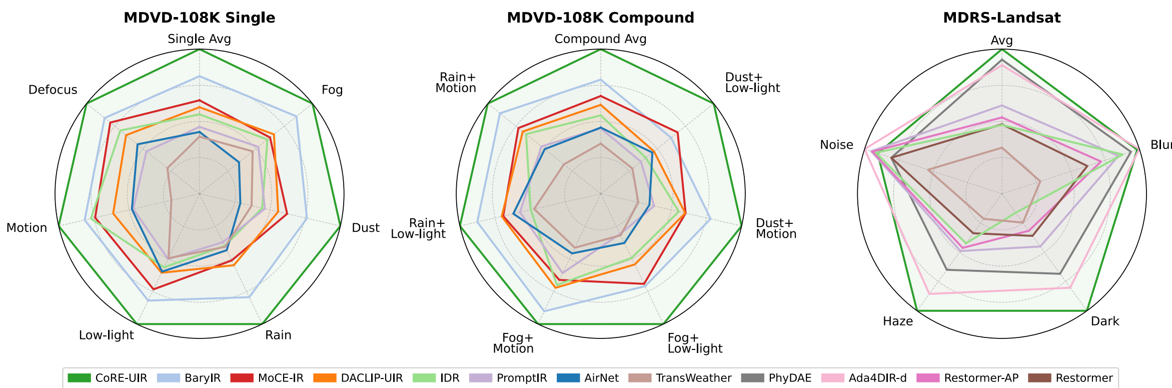

Fig. 5. Radar-chart comparison of PSNR for representative algorithms. From left to right, the three panels summarize MDVD-108K single degradations, MDVD-108K compound degradations, and MDRS-Landsat. Larger radius indicates higher PSNR.

$$F^i_{\text{out}} = F^i_{L_i}. \tag{22}$$

Thus, each stage performs one global modulation at its entrance, followed by recursive CoRE-based block updates. Repeating this pattern across the encoder, bottleneck, and decoder stages yields a compact implementation of GLA-Net, where GFM handles stage-level conditioning and CoRE handles block-level residual refinement.

3.6. Training strategy

CoRE-UIR adopts a two-phase training strategy.

Phase I: Degradation Prior Adaptation. We first optimize DPE with a multi-label degradation classification task over the M base degradation types. The CLIP image encoder is frozen, and only the lightweight adapter and a linear classification head are trainable. The Phase-I objective is the binary cross-entropy loss:

$$\mathcal{L}_{\text{cls}} = \text{BCE}\left(\text{Head}\left(z_{\text{deg}}\right), y\right), \tag{23}$$

where $\text{Head}(\cdot)$ denotes the linear classification head, $y \in \{0,1\}^M$ is the multi-hot degradation label vector, and $\text{BCE}(\cdot)$ denotes the binary cross-entropy loss. Compound degradations are naturally represented by activating multiple entries in y. Since this phase involves very few trainable parameters, it converges quickly while adapting the pre-trained semantic representation model toward restoration-oriented embeddings for remote-sensing images.

Phase II: Restoration Network Training. After Phase I, the classification head is discarded and the entire DPE, including the CLIP encoder and adapter, is frozen. We then train only the instantiated restoration backbone for image restoration. The overall training objective combines pixel, structural, and perceptual losses:

$$\mathcal{L}_{\text{total}} = \lambda_{\text{pix}}\mathcal{L}_{\text{pix}} + \lambda_{\text{str}}\mathcal{L}_{\text{str}} + \lambda_{\text{per}}\mathcal{L}_{\text{per}}, \tag{24}$$

where $\mathcal{L}_{\text{pix}} = \|\hat{I} - I_{\text{gt}}\|_1$ denotes the pixel loss, $\mathcal{L}_{\text{str}} = 1 - \text{SSIM}(\hat{I}, I_{\text{gt}})$ denotes the structural loss (Wang et al., 2004), and $\mathcal{L}_{\text{per}}$ denotes the perceptual loss implemented by LPIPS (R. Zhang et al., 2018). The concrete training hyperparameters are deferred to the implementation details in Section 4.1.4.

Although the overall framework is trained in two phases, the additional cost of Phase I is small because only the lightweight adapter and classification head are optimized. More importantly, the pretrained DPE provides a stable and discriminative degradation embedding space, which makes the subsequent optimization of the CoRE-UIR restoration network more stable in Phase II.

4. Experiments

4.1. Settings

4.1.1. Datasets

We evaluate CoRE-UIR on two public remote sensing datasets covering UAV and satellite imagery, respectively. Representative samples from these datasets are shown in Fig. 4.

MDVD-108K. We construct a large-scale multi-degradation UAV aerial dataset based on the VisDrone (Zhu et al., 2021) static detection imagery.[1] Degraded images are synthesized using Depth Anything V2 (Yang et al., 2024) for scene depth estimation and physically-grounded degradation models. The dataset contains six single degradations grouped into two categories: **weather-induced** degradations

[1] The original VisDrone labels are mapped into three coarse categories: person (pedestrian/people), micro_vehicle (bicycle/tricycle/awning-tricycle/motor), and vehicle (car/van/truck/bus).

Table 1
Quantitative comparison on **weather-induced degradation** types (fog, dust, rain) from the MDVD-108K dataset. Best results are highlighted in **bold** and second-best are underlined. ↑ indicates higher is better and ↓ indicates lower is better.

Method	Venue	Fog			Dust			Rain			Weather Average		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Single-task methods													
MPRNet	CVPR'21	24.72	0.9277	0.0622	25.91	0.9043	0.1185	29.32	0.9590	0.0450	26.65	0.9303	0.0752
Restormer	CVPR'22	28.33	0.9553	0.0404	27.96	0.9208	0.0868	31.27	0.9718	0.0276	29.19	0.9493	0.0516
NAFNet	ECCV'22	26.77	0.9476	0.0382	26.54	0.9084	0.0696	30.95	0.9687	0.0281	28.09	0.9416	0.0453
DGUNet	CVPR'22	27.63	0.9452	0.0439	27.30	0.9028	0.1010	30.84	0.9601	0.0402	28.59	0.9360	0.0617
All-in-One methods													
TransWeather	CVPR'22	28.77	0.9533	0.0359	28.02	0.9128	0.0888	31.78	0.9656	0.0317	29.52	0.9439	0.0521
AirNet	CVPR'22	27.73	0.9537	0.0324	27.57	0.9179	0.0561	31.94	0.9728	0.0222	29.08	0.9482	0.0369
PromptIR	NeurIPS'23	29.20	0.9575	0.0266	28.56	0.9208	0.0494	31.56	0.9718	0.0215	29.78	0.9500	0.0325
IDR	CVPR'23	29.90	0.9610	0.0329	28.45	0.9254	0.0762	31.74	0.9746	0.0221	30.03	0.9537	0.0438
DACLIP-UIR	ICLR'24	30.40	0.9619	0.0224	29.10	0.9255	0.0446	32.63	0.9755	0.0174	30.71	0.9543	0.0281
MoCE-IR	CVPR'25	30.10	0.9611	0.0234	29.46	0.9271	0.0440	32.41	0.9772	0.0162	30.66	0.9551	0.0279
BaryIR	TPAMI'26	32.11	0.9657	0.0184	30.27	0.9304	0.0372	34.11	0.9795	0.0137	32.16	0.9585	0.0231
CoRE-UIR	*Ours*	**33.33**	**0.9673**	**0.0161**	**31.62**	**0.9336**	**0.0330**	**35.35**	**0.9817**	**0.0114**	**33.44**	**0.9609**	**0.0201**

Table 2
Quantitative comparison on **imaging-induced degradation** types (low-light, motion blur, defocus blur) from the MDVD-108K dataset. Best results are highlighted in **bold** and second-best are underlined. ↑ indicates higher is better and ↓ indicates lower is better.

Method	Venue	Low-light			Motion Blur			Defocus Blur			Imaging Average		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Single-task methods													
MPRNet	CVPR'21	22.15	0.8195	0.2042	25.64	0.7400	0.3351	28.43	0.8346	0.2190	25.41	0.7980	0.2528
Restormer	CVPR'22	24.41	0.8465	0.1642	27.43	0.8053	0.2278	29.66	0.8653	0.1495	27.16	0.8390	0.1805
NAFNet	ECCV'22	24.12	0.8394	0.1035	26.42	0.7602	0.1237	29.38	0.8495	0.0747	26.64	0.8164	0.1006
DGUNet	CVPR'22	24.57	0.8103	0.1847	25.09	0.6831	0.3793	28.51	0.8239	0.1899	26.06	0.7724	0.2513
All-in-One methods													
TransWeather	CVPR'22	25.82	0.8255	0.1682	25.83	0.7216	0.3281	29.09	0.8385	0.1755	26.91	0.7952	0.2239
AirNet	CVPR'22	26.59	0.8517	0.0929	26.86	0.7860	0.1003	29.71	0.8612	0.0594	27.72	0.8330	0.0842
PromptIR	NeurIPS'23	25.88	0.8496	0.0892	26.83	0.7785	0.0972	29.52	0.8575	0.0602	27.41	0.8285	0.0822
IDR	CVPR'23	26.35	0.8592	0.1528	27.94	0.8233	0.1886	30.06	0.8738	0.1329	28.11	0.8521	0.1581
DACLIP-UIR	ICLR'24	26.66	0.8560	0.0817	27.35	0.7994	0.0839	29.94	0.8660	0.0516	27.98	0.8405	0.0724
MoCE-IR	CVPR'25	27.60	0.8589	0.0782	27.83	0.8137	0.0789	30.27	0.8727	0.0510	28.57	0.8484	0.0694
BaryIR	TPAMI'26	28.24	0.8653	0.0706	28.10	0.8180	0.0710	30.38	0.8747	0.0451	28.91	0.8527	0.0622
CoRE-UIR	*Ours*	**29.58**	**0.8702**	**0.0645**	**28.78**	**0.8376**	**0.0600**	**30.75**	**0.8812**	**0.0411**	**29.71**	**0.8630**	**0.0552**

(fog, dust, rain), and **imaging-induced** degradations (low-light, motion blur, defocus blur). It further contains six selected cross-category compound degradations constructed by pairing the three weather-induced degradations with two imaging-induced degradations, namely motion blur and low-light. To avoid data leakage, the clean source images strictly follow the original VisDrone train/val/test partition before degradation synthesis, and all synthetic pairs are generated within their respective splits. In total, MDVD-108K comprises 108,500 images (512 × 512): 108,000 synthetic paired samples (86,400/10,800/10,800 for training/validation/testing) and 500 real-world degraded UAV images collected via manual screening, which serve as a **real-world test set** for qualitative evaluation. Detailed synthesis protocols are provided in Appendix.

MDRS-Landsat. For satellite-domain evaluation, we employ MDRS-Landsat introduced by Lihe et al. (2025). This dataset is designed for four restoration tasks, including deblurring, denoising, dehazing, and dedarkening. It is constructed from 5400 clean Landsat-8 images (512 × 512) from RSHaze (Song et al., 2023), and each degradation subset is split into 5130/270 samples for training/testing. Following Lihe et al. (2025), degradations are synthesized with task-specific protocols, including anisotropic Gaussian blur, Gaussian noise, band-varied non-uniform haze based on Landsat-8 B9 (cirrus) information and atmospheric scattering, and power-law darkening transformation. In summary, MDRS-Landsat contains 20,520 training pairs and 1080 testing pairs.

4.1.2. Comparison methods

We compare CoRE-UIR against two categories of methods:

Single-task methods: MPRNet (Zamir et al., 2021), Restormer (Zamir et al., 2022), NAFNet (L. Chen et al., 2022), and DGUNet (Mou et al., 2022). For fair comparison, each single-task method is trained separately for every restoration subtask, and the reported averages aggregate the corresponding dedicated checkpoints.

All-in-One methods: TransWeather (Valanarasu et al., 2022), AirNet (Li et al., 2022), PromptIR (Potlapalli et al., 2024), IDR (J. Zhang et al., 2023), DACLIP-UIR[2] (Luo et al., 2024), Restormer-AP (Kong et al., 2024), MoCE-IR (Zamfir et al., 2025), and BaryIR (Tang et al., 2026).

In addition, we compared two physics-based models developed specifically for the four-degradation scenario in MDRS-Landsat: Ada4DIR (Lihe et al., 2025) and PhyDAE (Dong et al., 2026).

4.1.3. Evaluation metrics

For restoration quality, PSNR measures pixel-level reconstruction fidelity in decibels, SSIM (Wang et al., 2004)[3] evaluates structural consistency in terms of luminance, contrast, and local pattern similarity, and LPIPS (R. Zhang et al., 2018) measures perceptual distance in a deep feature space and better reflects visual realism. Higher PSNR and SSIM indicate better restoration quality, while lower LPIPS indicates better perceptual quality.

[2] DACLIP-UIR follows its official 100-step inference setting for both restoration evaluation and efficiency measurement.

[3] PSNR and SSIM are computed on full RGB images without border cropping, per channel and then averaged.

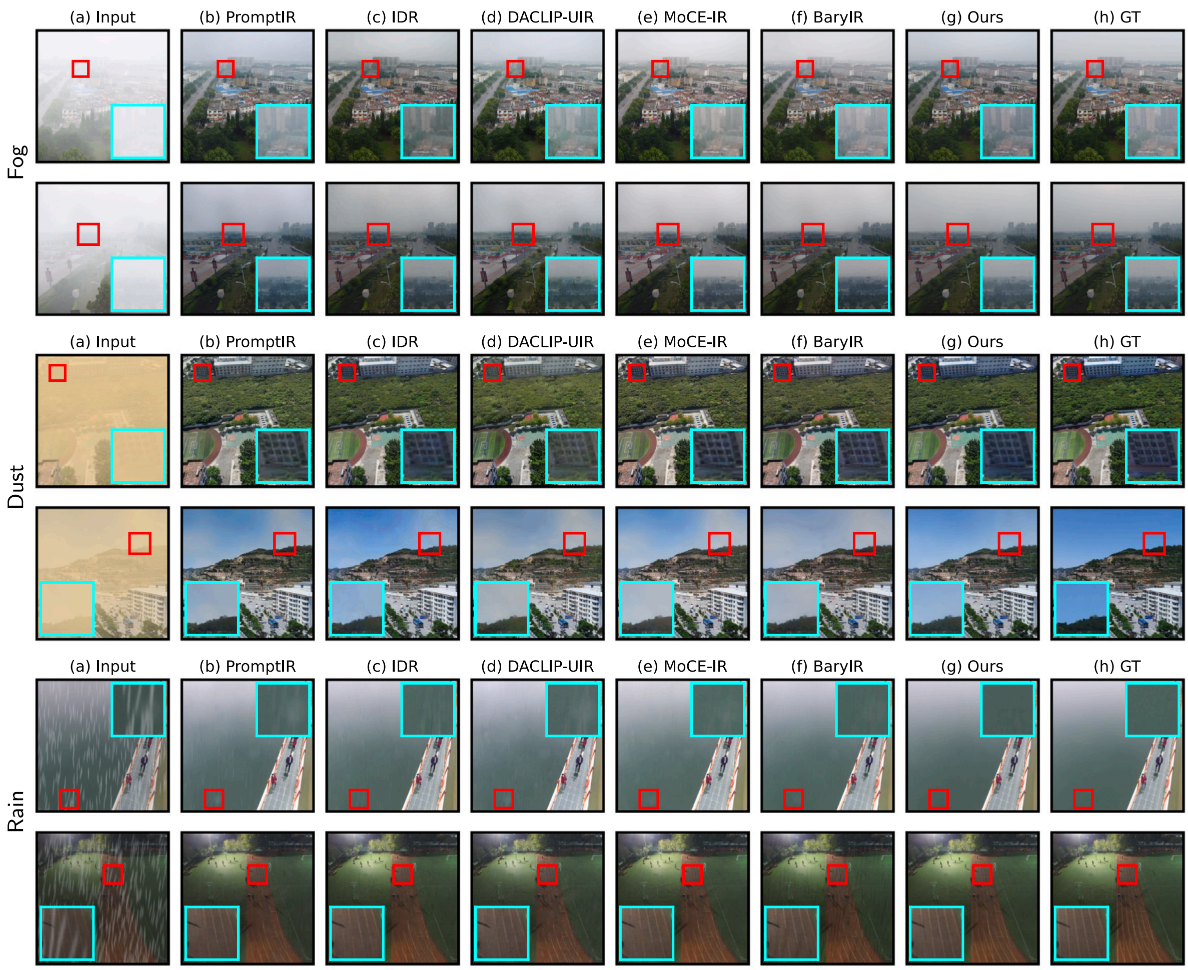

Fig. 6. Visual comparison on **weather-induced degradation** cases from MDVD-108K. CoRE-UIR produces more natural visibility, contrast, and color than prior universal restoration baselines, which is consistent with the role of DPE and GFM in organizing global appearance modulation under weather-induced degradations.

For the Phase-I proxy task in the DPE ablations, we use multi-label degradation classification over the $M = 6$ base degradation types, so compound samples can activate multiple labels. Precision, Recall, and F1 are macro-averaged.

4.1.4. Implementation details

The NAFNet backbone is configured with encoder stages of $[1, 1, 1, 28]$ NAFBlocks and a base channel width of $C = 32$. The degradation prior is extracted by a frozen CLIP ViT-B/32 (224×224 input) image encoder and mapped to a degradation embedding of dimension $d_z = 384$ via a lightweight two-layer adapter ($d_{\text{hidden}} = 384$). For CoRE, we set the number of low-rank expert pairs to $N = 6$, matching the $M = 6$ base degradation types, the bottleneck rank to $r = 4$, and $k = 3$. The GFM uses a bottleneck reduction ratio of $\rho = 16$. The code and dataset will be released at https://github.com/zzaiyan/CoRE-UIR.

The framework is trained in two phases. In Phase I (Degradation Prior Adaptation), DPE is optimized with a multi-label degradation classification task over the $M = 6$ base degradation types by freezing the CLIP encoder and updating only the lightweight adapter together with a linear classification head. We use the binary cross-entropy loss, the AdamW optimizer (Loshchilov and Hutter, 2017a) ($\beta_1 = 0.9$, $\beta_2 = 0.9$, and weight decay 10^{-3}), a batch size of 16, and an initial learning rate of 2×10^{-4} decayed to 1×10^{-5} by cosine annealing. Phase I is trained for 30 epochs on both single and compound training samples, and the latest checkpoint is used as the DPE checkpoint for Phase II. For DPE, the local view is randomly cropped during training and replaced by a fixed center crop at inference. In Phase II (Restoration Network Training), the classification head is discarded, the entire DPE is frozen, and the instantiated restoration backbone (GLA-Net) is trained for 700,000 iters with the AdamW optimizer using the same settings, a batch size of 8, and cosine annealing (Loshchilov and Hutter, 2017b) that decays the learning rate from 1×10^{-3} to 1×10^{-6}. During this phase, we jointly optimize a pixel loss (implemented as L1), a structural loss ($1 - \text{SSIM}$), and a perceptual loss (implemented by LPIPS[4]), with weights 0.6, 0.2, and 0.2, respectively, in Eq. (24). Training images are randomly cropped to 256×256 patches. Data augmentation includes random horizontal/vertical flips and rotations. For CoRE-UIR, the latest Phase-II checkpoint is used for evaluation. Baseline methods follow the authors' official training and inference settings. All experiments are conducted on a single NVIDIA RTX 4090 24 GB GPU using PyTorch 2.5 framework.

4.2. Comparison with state-of-the-art

Fig. 5 first provides a compact PSNR overview of representative algorithms on MDVD-108K single degradations, MDVD-108K compound degradations, and MDRS-Landsat. The three radar panels show that CoRE-UIR forms the strongest overall envelope on the two MDVD-108K settings and preserves the best average with a competitive outer frontier on MDRS-Landsat, which is consistent with the detailed quantitative comparisons below.

[4] https://github.com/richzhang/PerceptualSimilarity

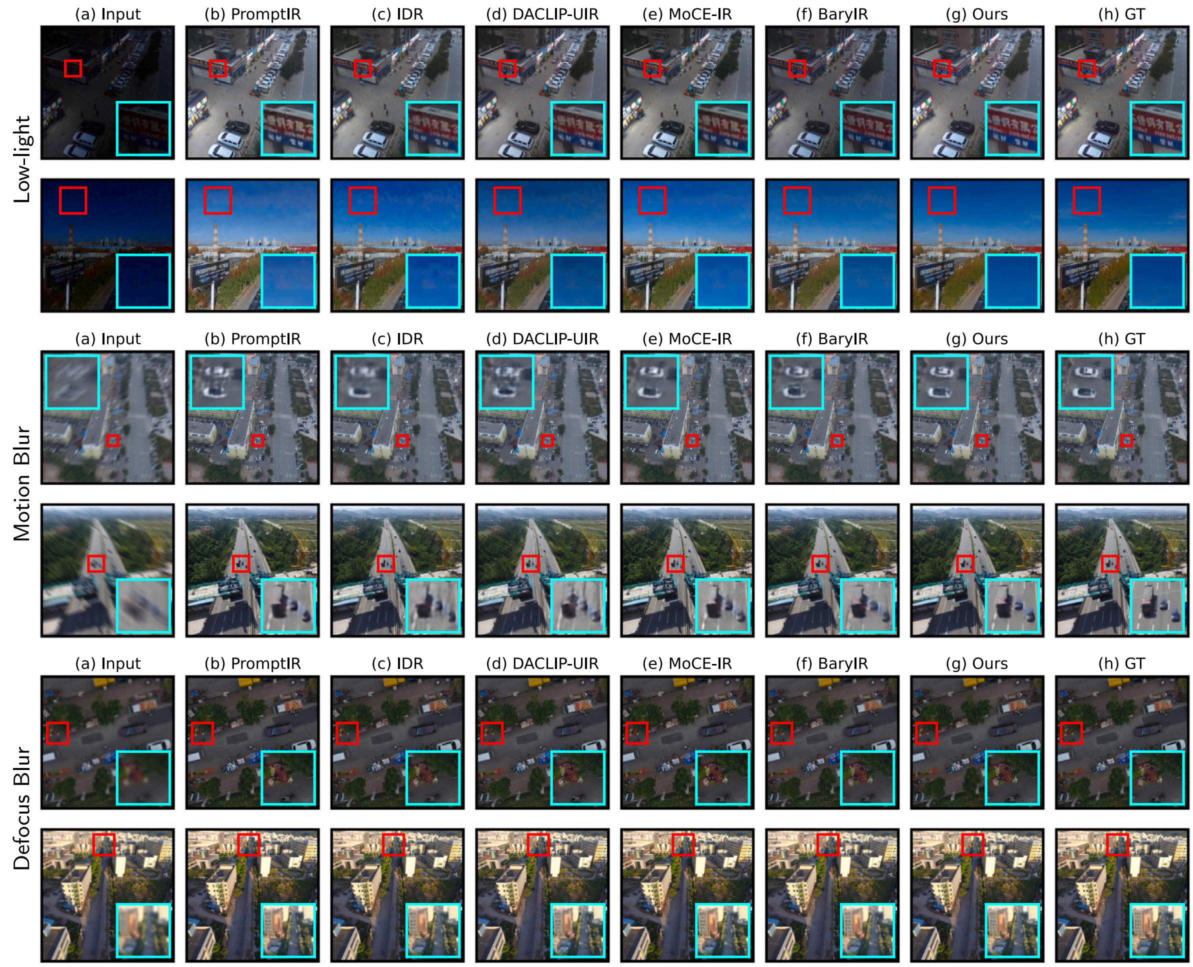

Fig. 7. Visual comparison on **imaging-induced degradation** cases from MDVD-108K. CoRE-UIR produces more uniform brightness correction and sharper structural recovery than prior universal restoration baselines, reflecting the benefit of low-rank residual specialization for local degradation compensation.

Table 3
Quantitative comparison on **compound degradation** types from the MDVD-108K dataset. Three representative pairwise combinations are shown, together with the average over all six compound types. Best results are highlighted in **bold** and second-best are underlined. ↑ indicates higher is better and ↓ indicates lower is better.

Method	Venue	Rain+Low-light			Dust+Motion			Fog+Motion			Compound Average		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Single-task methods													
MPRNet	CVPR'21	21.31	0.7219	0.3336	22.09	0.6897	0.4364	20.26	0.6756	0.4229	20.86	0.7067	0.3846
Restormer	CVPR'22	23.47	0.7644	0.2820	23.46	0.7498	0.3289	22.97	0.7568	0.3007	22.55	0.7591	0.2905
NAFNet	ECCV'22	21.15	0.7384	0.2126	22.53	0.7042	0.2104	22.13	0.7118	0.1810	21.78	0.7269	0.1974
DGUNet	CVPR'22	22.93	0.6979	0.3272	22.63	0.6308	0.4585	21.20	0.6240	0.4563	21.65	0.6640	0.3966
All-in-One methods													
TransWeather	CVPR'22	23.72	0.7180	0.3034	23.18	0.6708	0.4153	22.72	0.6762	0.4060	22.69	0.6977	0.3583
AirNet	CVPR'22	24.53	0.7646	0.1871	23.50	0.7271	0.1738	22.98	0.7375	0.1516	23.27	0.7493	0.1712
PromptIR	NeurIPS'23	24.27	0.7654	0.1788	23.64	0.7306	0.1616	23.88	0.7415	0.1376	23.25	0.7523	0.1607
IDR	CVPR'23	23.86	0.7764	0.2604	24.35	0.7689	0.2929	24.42	0.7826	0.2571	23.72	0.7790	0.2603
DACLIP-UIR	ICLR'24	24.98	0.7745	0.1638	24.53	0.7509	0.1465	24.55	0.7623	0.1218	24.10	0.7686	0.1433
MoCE-IR	CVPR'25	24.92	0.7743	0.1628	24.54	0.7552	0.1461	24.19	0.7663	0.1247	24.43	0.7729	0.1420
BaryIR	TPAMI'26	25.92	0.7866	0.1482	25.26	0.7677	0.1269	25.62	0.7828	0.1037	25.02	0.7840	0.1262
CoRE-UIR	*Ours*	**26.65**	**0.7937**	**0.1362**	**26.05**	**0.7847**	**0.1148**	**25.92**	**0.8003**	**0.0937**	**26.17**	**0.7989**	**0.1136**

4.2.1. *Single degradation*

Tables 1 and 2 show that CoRE-UIR attains the best average performance across both weather-induced and imaging-induced degradations, and ranks first across all six single-degradation subsets. Relative to BaryIR, the strongest recent universal restoration baseline by average score, CoRE-UIR gains 1.28 dB on the weather average and 0.80 dB on the imaging average, while also improving SSIM and LPIPS in both groups. Other recent universal restoration models remain competitive on individual cases: DACLIP-UIR is strong on weather-related perceptual quality, and MoCE-IR narrows the gap on several imaging subsets, but CoRE-UIR is more consistent across both groups. The visual comparisons in Fig. 6 and Fig. 7 are broadly aligned with these margins::

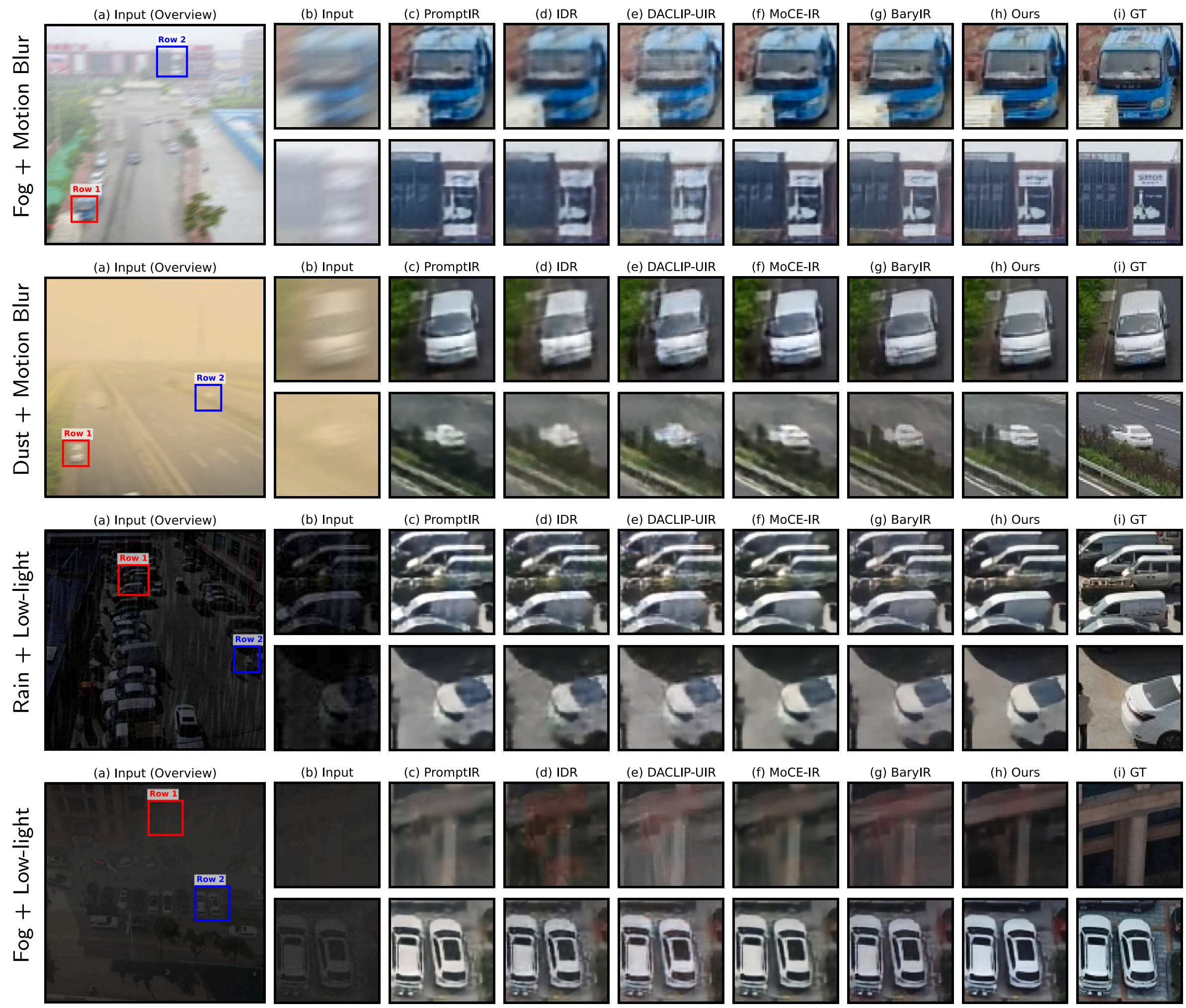

Fig. 8. Qualitative comparison for **compound degradation** cases. CoRE-UIR produces cleaner structures and fewer residual compound artifacts than prior universal restoration baselines under overlapping degradations, where the common dense path handles common restoration while low-rank residual experts compensate degradation-specific residuals.

Table 4
Satellite-domain evaluation on the **MDRS-Landsat** dataset. Best results are highlighted in **bold** and second-best are <u>underlined</u>. ↑ indicates higher is better and ↓ indicates lower is better. * denotes remote-sensing-specific methods.

Method	Venue	Blur			Dark			Haze			Noise			Average		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Single-task methods																
MPRNet	CVPR'21	32.54	0.8404	0.3514	31.52	0.9666	0.0523	29.15	0.9673	0.0575	33.02	0.8271	0.2515	31.56	0.9003	0.1782
Restormer	CVPR'22	35.23	0.8559	0.2099	37.86	0.9872	0.0110	36.18	0.9867	0.0124	34.53	0.8589	0.1356	35.95	0.9222	0.0922
NAFNet	ECCV'22	33.10	0.8120	0.3194	30.40	0.9516	0.0542	31.56	0.9642	0.0417	33.08	0.8263	0.1872	32.03	0.8885	0.1506
DGUNet	CVPR'22	29.64	0.7822	0.3405	27.15	0.9010	0.1339	27.45	0.9338	0.0713	30.31	0.7314	0.2491	28.64	0.8371	0.1987
All-in-One methods																
TransWeather	CVPR'22	33.45	0.8159	0.2868	36.33	0.9705	0.0193	35.02	0.9689	0.0241	33.69	0.8428	0.1530	34.62	0.8995	0.1208
AirNet	CVPR'22	28.27	0.7887	0.3244	28.38	0.9472	0.0569	24.39	0.9331	0.0641	30.30	0.7446	0.1918	27.84	0.8534	0.1593
PromptIR	NeurIPS'23	36.41	0.8861	0.1557	39.09	0.9900	0.0084	37.61	0.9897	0.0084	<u>34.99</u>	<u>0.8729</u>	0.1029	37.02	0.9347	0.0689
IDR	CVPR'23	36.57	0.8902	0.1498	35.19	0.9865	0.0096	36.99	0.9892	0.0087	34.88	0.8681	0.1091	35.91	0.9335	0.0693
Restormer-AP	CVPR'24	35.75	0.8732	0.1800	37.27	0.9885	0.0102	37.36	0.9888	0.0098	34.96	0.8697	0.1239	36.34	0.9301	0.0810
Ada4DIR-d*	INFFUS'25	**37.20**	<u>0.9004</u>	<u>0.1308</u>	<u>43.85</u>	<u>0.9954</u>	**0.0023**	<u>41.06</u>	<u>0.9938</u>	**0.0038**	**35.14**	0.8724	0.1268	39.31	<u>0.9405</u>	0.0659
PhyDAE*	TGRS'26	36.88	0.8824	0.1487	42.24	0.9949	0.0121	39.12	0.9928	0.0227	34.53	0.8651	<u>0.0991</u>	<u>39.62</u>	0.9390	<u>0.0600</u>
CoRE-UIR	*Ours*	<u>37.12</u>	**0.9039**	**0.0517**	**46.51**	**0.9964**	**0.0023**	**42.42**	**0.9947**	<u>0.0059</u>	34.84	**0.8754**	**0.0910**	**40.22**	**0.9426**	**0.0377**

CoRE-UIR recovers clearer distant facades in fog, removes the yellow cast more thoroughly in dust, and preserves bridge, pedestrian, shop-sign, vehicle, and roadside boundaries more cleanly in rain, low-light, motion-blur, and defocus scenes.

4.2.2. Compound degradation

Table 3 shows that the advantage is retained under overlapping degradations. CoRE-UIR achieves the best result on all three representative combinations and improves the compound average by 1.15 dB

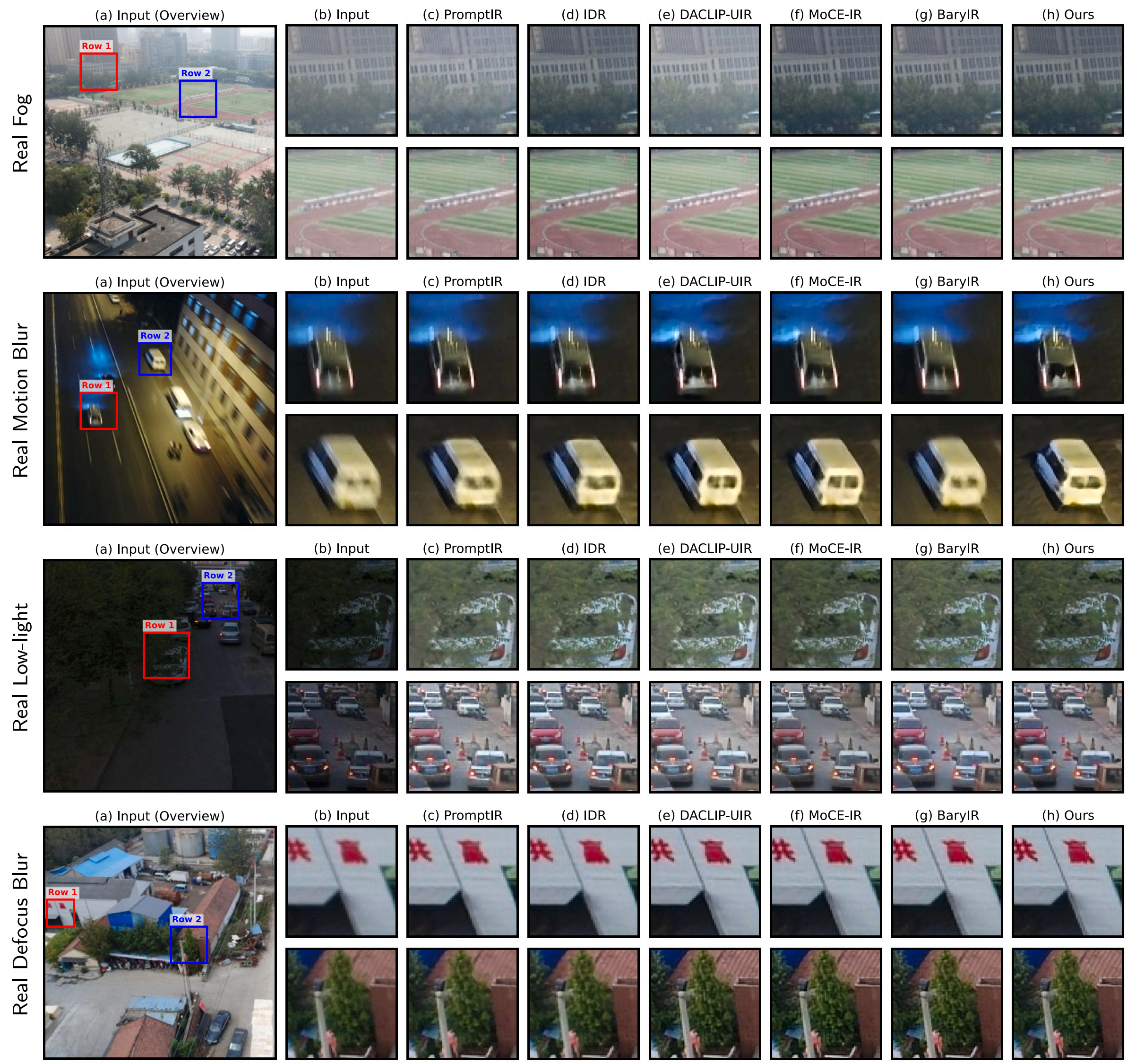

Fig. 9. Qualitative comparison for **real-world degradation** results. CoRE-UIR preserves land-cover structures while suppressing authentic degradation artifacts more reliably than competing universal restoration methods.

over BaryIR and by 1.74 dB over MoCE-IR, while also delivering the best SSIM and LPIPS. Although visual differences can be subtle in some cases, the gains are consistent across PSNR, SSIM, and LPIPS in this harder compound-degradation setting. In the visual examples shown in Fig. 8, CoRE-UIR restores clearer bus contours and sharper sign edges in fog+motion, and better balances streak suppression with dark-region recovery in rain+low-light. Competing methods remain visually reasonable, but they tend to leave slightly more residual artifacts or softer boundaries in these examples.

4.2.3. Real-world degradation

We further examine the real-world degradation test set of MDVD-108K, where ground-truth references are unavailable. The qualitative comparisons in Fig. 9 suggest that CoRE-UIR generalizes stably beyond the synthetic pipeline: it recovers building windows and track boundaries more clearly in fog, preserves vehicle contours and highlight structure more faithfully under night motion blur, reveals dark-region objects without obvious over-brightening in low light, and restores character strokes and roof boundaries more cleanly in defocus scenes. Several competing baselines also produce appealing results on individual images, but CoRE-UIR appears comparatively balanced between detail recovery and artifact suppression across the four scenes.

4.2.4. Satellite-domain evaluation

We further use MDRS-Landsat for satellite-domain evaluation, with the quantitative results summarized in Table 4. The comparison is stronger because Ada4DIR-d and PhyDAE are both tailored to remote sensing restoration with physical or domain-specific priors. Ada4DIR-d remains highly competitive and obtains the best blur PSNR (37.20) and noise PSNR (35.14), while PhyDAE provides the strongest non-ours overall average (39.62) together with low LPIPS. CoRE-UIR nevertheless achieves the best overall average of 40.22 and the best overall SSIM/LPIPS of 0.9426/0.0377, mainly due to clear gains on dark restoration (46.51 vs. 43.85) and haze removal (42.42 vs. 41.06). The qualitative comparisons in Fig. 10 show a similar tendency: CoRE-UIR preserves coastlines and sandbars more cleanly in blur, lifts dark regions with less loss of land-cover contrast, removes the haze veil while keeping ridge textures, and maintains cleaner shoreline boundaries

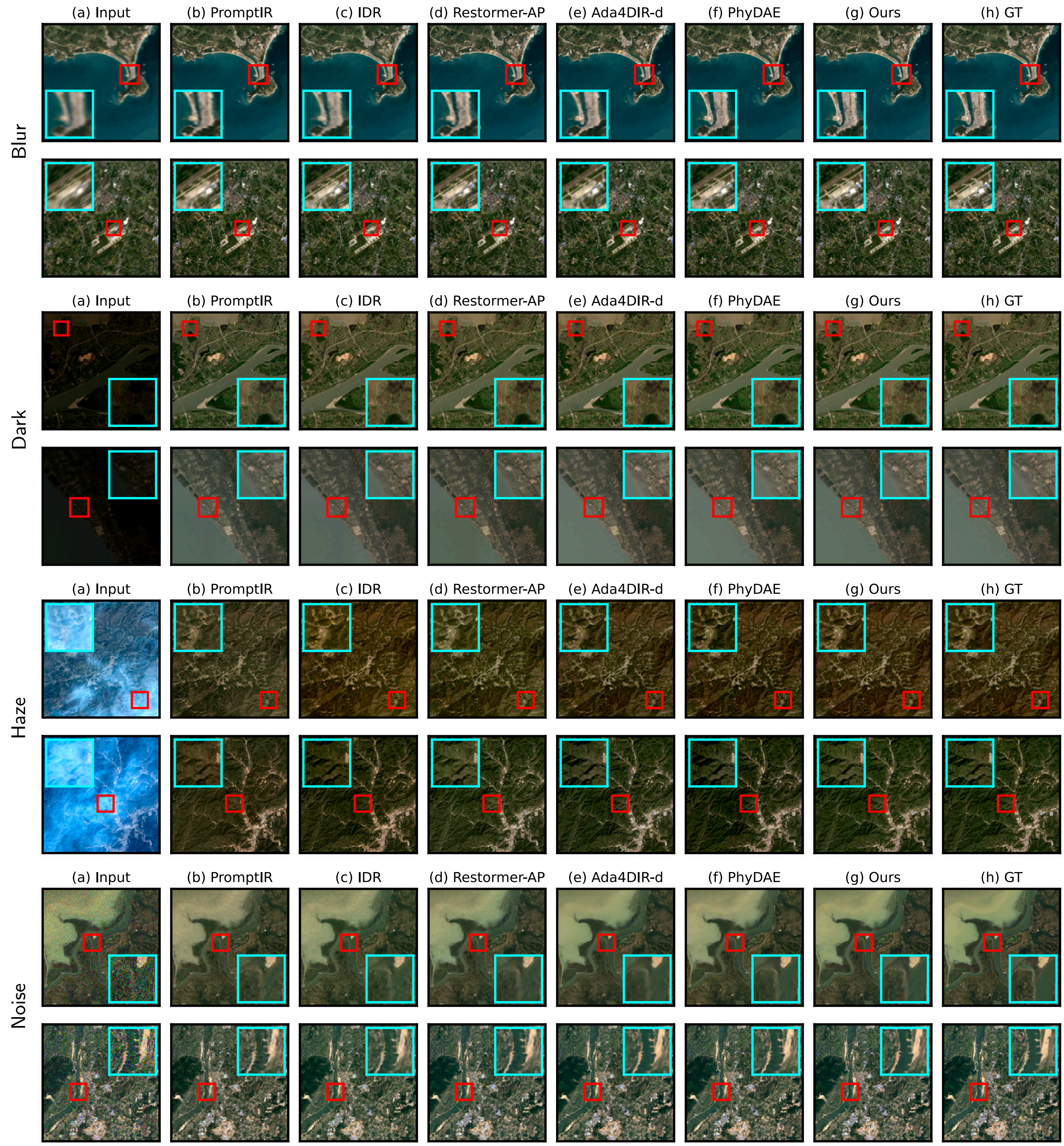

Fig. 10. Qualitative comparison on MDRS-Landsat. The panel illustrates that CoRE-UIR preserves large-scale structures while suppressing domain-shifted degradation artifacts more reliably than competing universal restoration methods.

under noise. These results suggest that the prior-guided global–local design adapts well to the satellite benchmark, demonstrating its versatility across different remote sensing scenarios.

4.3. Ablation studies

We conduct ablation experiments on MDVD-108K to evaluate DPE, GFM, and CoRE. Quantitative restoration results are reported by averaging the metrics over all single and compound degradation samples. Restoration and efficiency measurements use 512 × 512 inputs. Unless otherwise specified, Table 5 reports only DPE-branch complexity, while Tables 6 and 7 report only GLA-Net complexity. End-to-end efficiency is analyzed separately in Section 4.4.

4.3.1. Degradation prior embedding

We first examine whether DPE provides priors that are both discriminative and useful for restoration. Table 5 studies this question from three aspects: training strategy, prior encoder, and input view type. For the Phase-I proxy task, we additionally report macro Precision, Recall, and F1.

Training strategy. Comparing the first two rows shows that the proposed two-phase optimization improves restoration under identical DPE complexity, improving PSNR by 0.80 dB, raising SSIM from 0.8804 to 0.8928, and lowering LPIPS from 0.0647 to 0.0503. This indicates that Phase I is not merely an auxiliary objective, but reshapes the degradation embedding into a more restoration-oriented prior before the backbone is trained. Fig. 11(a)–(b) provides a direct visualization of this effect: the raw frozen CLIP features are heavily entangled

Table 5
Ablation study on **Degradation Prior Embedding (DPE)**. The table is organized into three groups: (a) training strategy, (b) prior encoder, and (c) input view type. Complexity is counted on the entire DPE branch, including the encoder, adapter, and classification head. "–" means the metric is not applicable for that variant.

Variant	Classification			Restoration			Complexity		
	Pre. (%)↑	Rec. (%) ↑	F1 (%) ↑	PSNR ↑	SSIM ↑	LPIPS ↓	Total Params	Trainable	FLOPs
(a) Training strategy									
Joint training (no Phase-I)	–	–	–	29.87	0.8804	0.0647	88.57M	0.72M	8.83G
Two-phase training (ours)	**99.4**	99.2	**99.3**	**30.67**	**0.8928**	**0.0503**	88.57M	0.72M	8.83G
(b) Prior encoder									
ResNet-50 (trainable)	98.1	98.1	98.1	29.51	0.8824	0.0612	26.95M	26.95M	8.22G
DACLIP (w/ adapter)	98.9	99.1	99.0	30.49	0.8911	0.0536	183.87M	1.09M	9.09G
RemoteCLIP (w/ adapter)	98.7	99.0	98.8	30.44	0.8913	0.0546	88.57M	0.72M	8.83G
CLIP (w/ adapter, ours)	**99.4**	**99.2**	**99.3**	**30.67**	**0.8928**	**0.0503**	88.57M	0.72M	8.83G
(c) Input view type									
Global view only	99.3	98.8	99.0	30.45	0.8894	0.0533	88.31M	0.46M	4.41G
Local view only	96.9	94.0	95.4	30.11	0.8806	0.0635	88.31M	0.46M	4.41G
Global + local views (ours)	**99.4**	**99.2**	**99.3**	**30.67**	**0.8928**	**0.0503**	88.57M	0.72M	8.83G

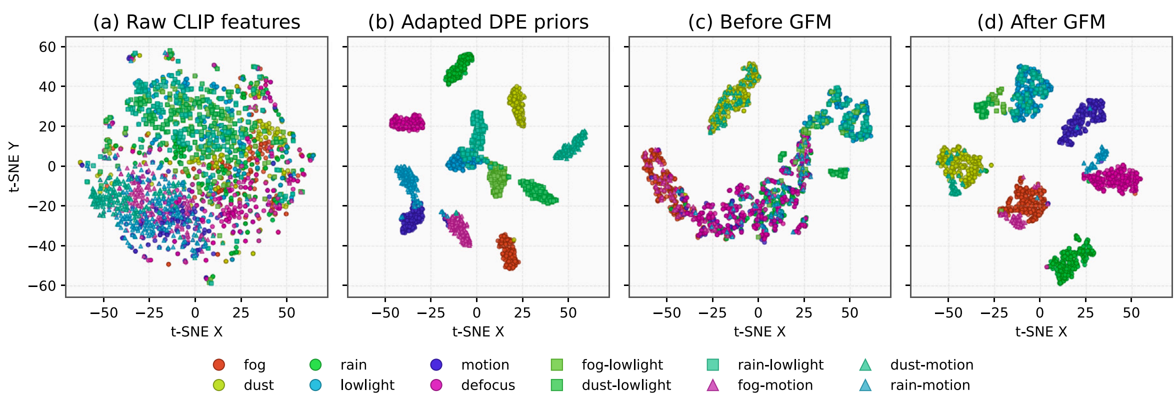

Fig. 11. Feature visualization for **DPE** and **GFM**, showing (a) Raw CLIP features, (b) Adapted DPE priors, (c) Features before GFM, and (d) Features after GFM. (a)–(b) demonstrate that DPE improves the degradation discriminativeness of the raw CLIP feature, while (c)–(d) show that GFM reshapes the restoration feature space into a more compact and routable manifold.

across single and compound degradations, whereas the adapted DPE priors form much tighter and cleaner clusters with larger inter-cluster margins.[5]

Prior encoder. Table 5(b) shows a clear hierarchy among four representative prior encoders: the fully trainable ResNet-50 (He et al., 2016) is weakest, DACLIP (Luo et al., 2024) and RemoteCLIP (F. Liu et al., 2024) bring clear gains as degradation-aware and remote-sensing priors, and the proposed CLIP with adapter tuning achieves the best classification and restoration results. Notably, although the trainable ResNet-50 already attains respectable classification performance, the restoration model guided by its priors still trails the CLIP-based design by 1.16 dB PSNR. This indicates that the gain comes not simply from fitting the proxy classification task or enlarging the trainable encoder, but from adapting foundation priors efficiently, with the general CLIP prior performing best with only 0.72M trainable parameters. The feature distribution in Fig. 11(b) is also more structured than the raw CLIP space in Fig. 11(a), which is consistent with the quantitative improvement.

Input view type. Using only the global or local view weakens both classification and restoration, with the local-only variant degrading the most. The combined global–local representation achieves the best overall result, confirming that broad scene context and local degradation textures are complementary for prior learning.

[5] The t-SNE projection uses 2000 randomly sampled features with perplexity 30 and seed 42.

4.3.2. Global feature modulation (GFM)

We next examine whether GFM indeed performs effective prior-state global feature modulation before CoRE specializes local residuals. Table 6 compares removing GFM entirely, simplifying it to a single branch, and using the full dual-branch design while keeping the rest of the pipeline unchanged. Removing GFM causes the largest performance drop, reducing PSNR from 30.67 to 28.87. Using only the prior branch or only the state branch recovers part of the gain, but both remain inferior to the full module, indicating that effective modulation requires both explicit degradation conditioning and the current feature-state response.

Fig. 11(c)–(d) visualizes pooled image features before and after the first GFM module. Before GFM, the restoration features largely lie on an elongated and partially entangled manifold, with several degradation types still overlapping around the central transition region. After modulation, the features reorganize into tighter and more separated groups, while related degradations still preserve meaningful neighborhood relations rather than collapsing into overly rigid class partitions. This behavior matches the role of GFM: it does not simply classify degradations, but reshapes intermediate features into a more compact, prior-consistent, and more routable space for downstream CoRE specialization.

4.3.3. Common-and-residual expert block (CoRE)

Finally, we examine whether CoRE provides efficient degradation-specific residual specialization. We study the common-residual expert design, expert sparsity, and routing strategy.

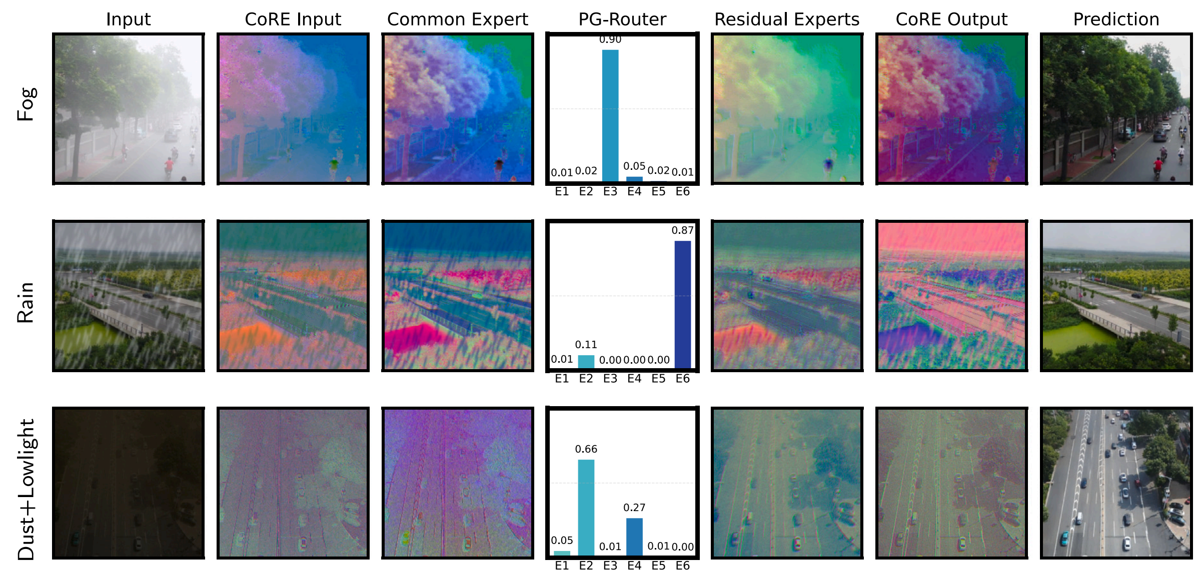

Fig. 12. Feature visualization for the first **CoRE** module at stage 2. From left to right, we show Input, CoRE Input Feature, Common Expert Feature, Router Weights, Residual Experts Feature, CoRE Output Feature, and Prediction. The common expert exhibits similar responses across degradations and builds a shared restoration basis, while the residual experts focus on degradation-specific compensation such as fog depth contrast, rain background brightening, and distant denoising in dust-dominated regions.

Table 6
Ablation study on **GFM**. The table compares removing GFM, single-branch simplifications, and the full dual-branch module.

Configuration	Params	FLOPs	PSNR ↑	SSIM ↑
No GFM	21.06M	76.90G	28.87	0.8729
Prior branch only	21.69M	76.90G	30.54	0.8914
State branch only	21.56M	76.93G	30.27	0.8896
Full GFM (Ours)	**22.12M**	**76.93G**	**30.67**	**0.8928**

Table 7
Ablation study on **CoRE**. The table is organized into four groups: (a) common-residual expert design, (b) expert sparsity, (c) routing granularity, and (d) routing method.

Configuration	Params	FLOPs	PSNR ↑	SSIM ↑
(a) Common-residual expert design				
Common expert only	18.47M	63.91G	30.36	0.8895
Full-rank FFN MoE	51.71M	121.89G	30.51	0.8918
CoRE (ours)	22.12M	76.93G	**30.67**	**0.8928**
(b) Expert sparsity				
$k = 1,\ N = 6$	22.12M	68.25G	30.40	0.8901
$k = 3,\ N = 6$ (ours)	22.12M	76.93G	**30.67**	**0.8928**
$k = 6,\ N = 6$	22.12M	89.96G	30.58	0.8924
(c) Routing granularity				
sample-level routing	22.11M	76.93G	30.52	0.8913
stage-level routing (ours)	22.12M	76.93G	**30.67**	**0.8928**
block-level routing	22.19M	76.93G	30.47	0.8887
(d) Routing method				
MLP router	23.45M	76.94G	30.50	0.8916
PG-Router (ours)	22.12M	76.93G	**30.67**	**0.8928**

Common-residual expert design. Table 7 consolidates the quantitative study of CoRE into four groups, where the first two focus on expert design and the last two focus on routing. Table 7(a) compares the common dense expert only variant, a full-rank FFN MoE branch, and the proposed CoRE branch to validate the common-residual expert design. The common dense expert only variant already reaches 30.36 PSNR and 0.8895 SSIM, indicating that a large portion of restoration behavior is shared across degradations. The full-rank MoE slightly improves adaptive capacity but expands the model to 51.71M parameters and 121.89G FLOPs. CoRE further improves the result to 30.67/0.8928 while remaining much lighter, validating the proposed common dense plus low-rank residual decomposition.

Expert sparsity. Table 7(b) fixes the expert-pair number to $N = 6$, matching the $M = 6$ base degradation types, and studies only the effect of k. When $k = 1$, routing is overly restrictive and cannot combine complementary experts for compound degradations. When $k = 6$, the combination becomes nearly dense, increases FLOPs from 76.93G to 89.96G, and yields no further gain. $k = 3$ therefore provides the most balanced trade-off between flexibility and sparsity.

Common dense *vs.* low-rank residual experts. Fig. 12 visualizes the first CoRE module in Block 2. Across fog, rain, and dust+low-light scenes, the common dense expert exhibits highly similar responses and mainly strengthens the dominant scene layout, large structural boundaries, and overall contrast, thereby providing a stable restoration basis. The low-rank residual experts are more localized and degradation-dependent: in fog they emphasize the haze veil and distant depth transition, in rain they focus more strongly on streak-corrupted structures and background recovery, and in dust+low-light they concentrate on illumination lifting together with lane and vehicle details. The PG-Router activations are also sparse and input-dependent: different single degradations activate different dominant experts, whereas the compound degradation activates multiple experts jointly. Taken together, these responses visually support the intended common-residual decomposition and the sparse adaptive routing mechanism.

Routing granularity and router type. Table 7(c) and Table 7(d) show that stage-level routing achieves the best result and is therefore adopted in CoRE-UIR. The sample-level variant is more stable but too coarse to exploit local degradation diversity, whereas the block-level variant is more flexible but less stable to optimize. Stage-level routing therefore provides the best balance between adaptability and stability. The PG-Router also slightly improves accuracy while reducing parameters relative to the generic MLP alternative, so it is kept as the default router.

Table 8
Efficiency and restoration-quality comparison among AiOIR methods on MDVD-108K. Best results are highlighted in **bold** and second-best are underlined. CoRE-UIR achieves the best restoration quality while remaining substantially efficient.

Method	Venue	Complexity		Measured Cost			Restoration		
		Params↓	FLOPs↓	Latency (ms)↓	Throughput↑	Memory↓	PSNR↑	SSIM↑	LPIPS↓
TransWeather	CVPR'22	38.1M	**24.2G**	**3.80±0.99**	**263.04 img/s**	**0.74 GB**	27.30	0.8409	0.1747
AirNet	CVPR'22	**8.93M**	1174.8G	403.17±0.52	2.48 img/s	5.14 GB	27.54	0.8670	0.0790
PromptIR	NeurIPS'23	35.6M	690.9G	231.47±0.38	4.32 img/s	9.32 GB	27.70	0.8665	0.0746
IDR	CVPR'23	12.3M	363.1G	127.24±0.30	7.86 img/s	12.92 GB	28.18	0.8823	0.1275
DACLIP-UIR	ICLR'24	295.2M	56.5T	7226.26±6.56	0.14 img/s	10.68 GB	28.47	0.8759	0.0658
MoCE-IR	CVPR'25	25.4M	382.4G	200.33±0.93	4.99 img/s	3.58 GB	28.75	0.8803	0.0642
BaryIR	TPAMI'26	53.5M	851.1G	265.58±0.74	3.77 img/s	9.85 GB	29.62	0.8853	0.0566
CoRE-UIR	*Ours*	110.7M	85.76G	22.44±1.06	44.56 img/s	1.45 GB	**30.67**	**0.8928**	**0.0503**

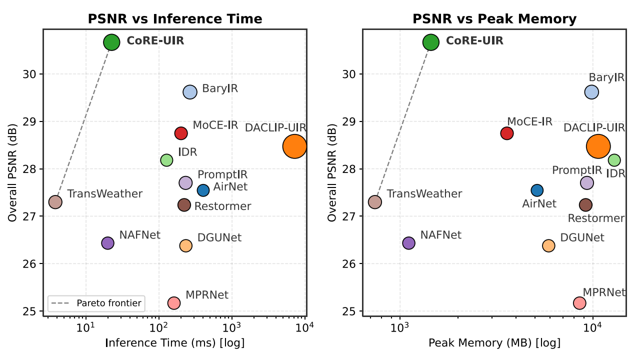

Fig. 13. Quality-efficiency trade-off on MDVD-108K. Left: PSNR *vs.* inference time. Right: PSNR *vs.* peak memory. Both horizontal axes are logarithmic, and CoRE-UIR stays near the upper-left frontier in both views.

4.4. Further analysis

We further examine CoRE-UIR from four practical perspectives: model efficiency, prior quality, downstream perception utility, and robustness to unseen compound degradations. These analyses evaluate deployment cost, the effect of DPE prediction reliability, whether restoration benefits target perception, and whether the model remains stable under unseen degradation combinations.

4.4.1. Model efficiency

Following the benchmark protocol, all efficiency measurements are collected on a single NVIDIA RTX 4090 GPU with batch size 4. DPE's CLIP encoder is run with 16-bit AMP, while other methods follow their official settings. We first warm up 5 batches, then time 20 consecutive batches, and finally divide the batch latency by 4 to report per-image mean and standard deviation. Table 8 summarizes complexity, measured cost, and restoration quality under this unified setting. Although CoRE-UIR includes a frozen CLIP ViT-B/32 prior encoder, its full inference path remains efficient: after including DPE, the total complexity is 85.76G FLOPs, with 22.44 ± 1.06 ms latency, 44.56 img/s throughput, and 1.45 GB peak memory. Relative to BaryIR, it still gains 1.05 dB PSNR, runs 11.83× faster, and uses 85.3% less peak memory while preserving the best overall quality. By contrast, DACLIP-UIR is constrained by iterative sampling: one reverse step costs 72.00 ± 0.33 ms, but 100 denoising steps expand end-to-end latency to 7226.26 ± 6.56 ms. Fig. 13 shows the same trend: TransWeather occupies the extreme low-latency corner at lower quality, whereas CoRE-UIR remains near the low-latency and low-memory side while exceeding 30 dB. Overall, CoRE-UIR preserves a favorable quality-efficiency trade-off for practical deployment.

Table 9
Prior-quality stratified restoration performance on MDVD-108K.

DPE prediction	Samples	Ratio	PSNR ↑	SSIM ↑	LPIPS ↓
Exact match	10,643	98.55%	30.76	0.8939	0.0493
Partial match	143	1.32%	25.17	0.8144	0.1066
Wrong	14	0.13%	17.73	0.8262	0.1732
All	10,800	100.00%	30.67	0.8928	0.0503

Table 10
Downstream object detection with a fixed YOLO11s detector. Metrics are reported in percentage. Best and second-best restored-image results are highlighted in **bold** and underlined.

Input	Precision ↑	Recall ↑	mAP_{50} ↑	$mAP_{50:95}$ ↑
(a) Synthetic test set – 10,800 samples				
Clean	72.40	52.81	57.40	31.49
Degraded	54.47	28.27	28.38	15.01
MoCE-IR	65.65	43.84	46.67	25.05
BaryIR	66.52	45.33	48.66	25.65
CoRE-UIR	**67.84**	**47.08**	**50.36**	**27.10**
(b) Real test set – 500 samples				
Real degraded	55.06	35.91	37.16	20.65
MoCE-IR	55.29	37.64	38.70	22.30
BaryIR	58.70	39.15	40.86	23.89
CoRE-UIR	**59.12**	**40.73**	**41.73**	**25.20**

4.4.2. DPE prior reliability

To examine how prior reliability affects restoration, we stratify the test samples according to the DPE multi-label prediction. Exact match means that the predicted degradation-label set is identical to the ground truth; partial match means that at least one degradation label is correct but the set is incomplete or contains extra labels; wrong means that no ground-truth degradation label is recovered. Table 9 shows that 98.55% of test samples fall into the exact-match bucket, where CoRE-UIR reaches 30.76 dB PSNR and 0.0493 LPIPS. The small partial and wrong buckets exhibit clearly lower restoration quality, especially in PSNR and LPIPS, indicating that inaccurate priors tend to coincide with difficult samples and residual artifacts. Nevertheless, these cases account for only 1.45% of the test set, so the overall result remains close to the exact-match bucket.

4.4.3. Downstream object detection

To evaluate whether restoration benefits downstream perception, we train a YOLO11s (Khanam and Hussain, 2024) detector for 50 epochs[6] using only clean GT images from the MDVD-108K training split. The detector weights are then fixed and evaluated on clean GT, degraded input, and restored images from both the paired synthetic test

[6] We use the official implementation of YOLO11s by Ultralytics at https://github.com/ultralytics/ultralytics and keep all default training and inference settings.

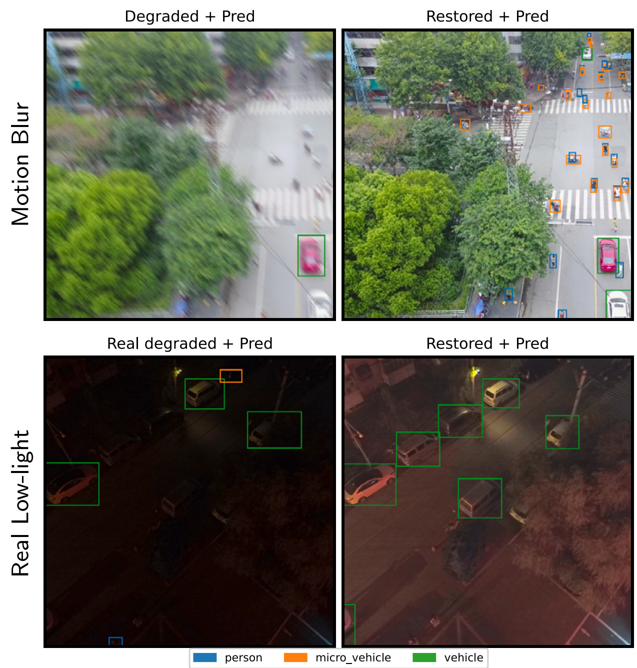

Fig. 14. Qualitative downstream detection examples on synthetic motion blur and real low-light scenes. Compared with degraded inputs, CoRE-UIR restores clearer object boundaries and illumination, yielding more complete detector responses.

set and the real degraded test set with detection labels. As shown in Table 10, degraded inputs sharply weaken detection, whereas restored images recover a large portion of the lost accuracy. CoRE-UIR gives the best restored-image results on the synthetic test set, recovering 75.7% and 73.4% of the degradation-induced gaps in mAP_{50} and $\mathrm{mAP}_{50:95}$. The same trend holds on the real test set without clean GT images, where CoRE-UIR also outperforms BaryIR and the degraded input. Fig. 14 further shows that restoration sharpens motion-blurred object boundaries and brightens real low-light scenes, leading to more complete detections of vehicles, micro-vehicles, and pedestrians. These results indicate that CoRE-UIR preserves object-level cues useful for a detector that is never trained on restored images.

4.4.4. Unseen compound-degradation robustness

We further evaluate whether the learned restoration behavior transfers to unseen compound degradations. Here, seen compound types refer to the six compound degradations defined in the dataset, while unseen ones are previously unused combinations of a weather degradation with defocus blur. Table 11 compares three representative universal restoration baselines under this setting. Although unseen cases are more challenging for all methods, CoRE-UIR remains best across the three metrics. Relative to BaryIR, it gains 0.73 dB PSNR and also preserves clearer structural and perceptual quality. These results suggest that the common-residual expert design does not merely memorize observed compound pairs, but preserves useful degradation-adaptive behavior for new combinations of known degradation factors.

5. Conclusion

In this paper, we present CoRE-UIR, a prior-guided global–local framework for efficient all-in-one remote sensing image restoration. By coupling restoration-oriented degradation priors from DPE, prior-state global alignment from GFM, and common-residual specialization from CoRE, the model first establishes a reliable global restoration basis and then injects degradation-specific corrections through low-rank residual experts. This design reduces redundant expert replication, yields more interpretable specialization behavior, and remains efficient in deployment. We also construct MDVD-108K, a large-scale UAV multi-degradation restoration dataset covering both synthetic and real-world degraded images.

Extensive experiments on MDVD-108K and MDRS-Landsat validate the effectiveness of this design across single-degradation, compound-degradation, real-world qualitative, satellite-domain, downstream detection, and unseen compound-degradation evaluations. The results show that CoRE-UIR not only improves restoration quality, but also advances the Pareto frontier of universal restoration by achieving a better balance between accuracy, latency, and memory cost than existing baselines. More broadly, these findings suggest that universal restoration benefits from keeping shared restoration behaviors in a common dense expert and modeling degradation-specific differences as lightweight residual corrections.

Despite these advancements, CoRE-UIR still mainly addresses known degradation factors and evaluates downstream utility through object detection. Future work will focus on extending CoRE-UIR to broader unseen degradation categories, more complex real-world mixtures, and additional downstream remote sensing tasks.

Table 11
Robustness to unseen compound degradations on MDVD-108K. Best and second-best results are highlighted in **bold** and <u>underlined</u>.

Method	Seen Compound Avg.			Unseen Compound Avg.		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
MoCE-IR	24.43	0.7729	0.1420	22.21	<u>0.7341</u>	0.2397
BaryIR	<u>25.02</u>	<u>0.7840</u>	<u>0.1262</u>	<u>22.64</u>	0.7312	<u>0.2011</u>
CoRE-UIR	**26.17**	**0.7989**	**0.1136**	**23.37**	**0.7742**	**0.1502**

CRediT authorship contribution statement

Zaiyan Zhang: Writing – original draft, Visualization, Software, Methodology, Data curation. **Qiangqiang Yuan:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization. **Jie Li:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization. **Ziyang Lihe:** Validation, Software, Investigation. **Yu Wan:** Validation, Software, Investigation. **Yuzeng Chen:** Validation, Software, Investigation. **Xin Su:** Validation, Software, Investigation. **Liangpei Zhang:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grants 42471414 and 42471504 and the Fundamental Research Funds for the Central Universities under Grant No. 2042026kf0001.

Appendix. Dataset synthesis protocol

This appendix instantiates the degradation operator $\mathcal{G}_S$ and the selected configuration set C in Section 3.1.1 for the **MDVD-108K** benchmark. The goal is to generate controllable degradations while preserving a reasonable degree of physical realism for UAV imagery.

For each base degradation $d \in \mathcal{D}$ and severity level s, we predefine a parameter space $\Omega_{d,s}$ and randomly sample

$$\boldsymbol{\theta}_{d,s} \sim p(\boldsymbol{\theta} \mid d, s). \tag{25}$$

Let $T_d(\cdot; \boldsymbol{\theta}_{d,s})$ denote the corresponding synthesis operator. For a singleton active set $\mathcal{S} = \{d\}$, the operator in Eq. (1) reduces to

$$\mathcal{G}_{\{d\}}(\mathbf{I}) = T_d(\mathbf{I}; \boldsymbol{\theta}_{d,s}). \tag{26}$$

Starting from a clean image $\mathbf{I} \in [0,1]^{H\times W\times 3}$, we estimate a monocular depth map and convert it into a normalized distance-like representation $\mathbf{D} \in [0,1]^{H\times W}$, where distant regions are close to 1 and near regions are close to 0. All depth-aware degradation models are defined on this normalized map. Compound cases use the ordered cascade in Appendix A.2. To avoid data leakage, train/val/test samples are synthesized independently from disjoint clean source images that follow the original VisDrone partition, and real degraded images are never reused as synthesis sources.

A.1. Single-degradation models

The current release instantiates six base operators, namely defocus blur, motion blur, fog, rain, dust, and low-light. Each operator corresponds to one singleton case $\mathcal{S} = \{d\}$ in Section 3.1.1.

Defocus blur. Defocus blur is modeled in a scene-adaptive manner. Instead of directly amplifying normalized depth differences, we first estimate the global depth span of the scene and use it to modulate the effective separation strength between depth layers. The image is partitioned into several depth bins, and each bin is assigned a disk-blur radius

$$r_b = r_g + r_{\max}\left(\lambda\, s_d\, |c_b - f|\right)^p, \tag{27}$$

where r_g is a global blur floor, f is the sampled focal plane, and λ is determined by the scene depth span. The final result is obtained by softly blending the responses of all depth bins. This design avoids unrealistically large foreground–background differences in scenes with limited depth variation.

Motion blur. Motion blur is modeled as a trajectory-induced point spread function (PSF) during exposure. Let $\boldsymbol{\Gamma}(t)$ denote the image-plane motion trajectory and $w(t)$ the exposure weighting function. The degraded image is generated by convolving the clean image with the corresponding motion PSF,

$$\mathbf{Y} = h_{\text{motion}} * \mathbf{I}. \tag{28}$$

This formulation covers common motion patterns in UAV imagery, including camera shake and platform motion.

Fog. Fog is synthesized using a depth-aware atmospheric scattering model. We first map the normalized distance map to an effective depth with a positive lower bound and then compute a transmission map

$$t(x) = \exp\left(-\beta(x)\,\tilde{\mathbf{D}}(x)^{\gamma}\right), \tag{29}$$

where $\beta(x)$ is either constant or modulated by a low-frequency random field to produce non-uniform fog. The final foggy image is written as

$$\mathbf{Y}(x) = t(x)\mathbf{I}(x) + \left(1 - t(x)\right)\mathbf{A}, \tag{30}$$

with $\mathbf{A}$ denoting atmospheric light. This model reproduces depth-dependent contrast attenuation and spatially varying haze density.

Rain. Rain degradation is composed of three components: a streak layer, a low-frequency rain veil, and optional sparse lens droplets at severe levels. Rain streaks are synthesized by sampling line primitives with random length, orientation, and opacity, followed by anisotropic blur and contrast shaping. The streak layer is first blended with the clean image as

$$\mathbf{I}_r = (1 - \alpha_r R)\mathbf{I} + \alpha_r R\,\mathbf{C}_r + \eta_r R, \tag{31}$$

where R denotes the processed streak map, $\mathbf{C}_r$ is the rain color, and α_r and η_r control opacity and brightness enhancement, respectively. A low-frequency veil field is then introduced to mimic the global curtain-like appearance of rainfall, yielding

$$\mathbf{I}_v(x) = (1 - m_v(x))\, t_v(x)\mathbf{I}_r(x) + m_v(x)\mathbf{A}_v, \tag{32}$$

where $t_v(x)$ and $m_v(x)$ denote veil-related attenuation and mixing terms, and $\mathbf{A}_v$ is the corresponding airlight color. Heavy-rain cases may additionally include localized droplet distortions. This formulation allows the synthesized rain to exhibit both local streak structures and global visibility degradation.

Dust. Dust degradation combines warm-color atmospheric scattering, spatially non-uniform attenuation, and sparse particulate occlusion. The base layer follows a haze-like scattering process with yellow-brown airlight:

$$\mathbf{I}_b(x) = t_d(x)\mathbf{I}(x) + \left(1 - t_d(x)\right)\mathbf{A}_d, \tag{33}$$

where $t_d(x)$ is the dust transmission map and $\mathbf{A}_d$ denotes warm atmospheric light. A sparse set of elliptical particles is then blended onto the base image by

$$\mathbf{I}_p(x) = \left(1 - M_p(x)\right)\mathbf{I}_b(x) + M_p(x)\mathbf{C}_p, \tag{34}$$

where $M_p(x)$ is the aggregate particle mask and $\mathbf{C}_p$ is a slightly brighter particle color. Finally, saturation and contrast are moderately compressed to mimic the desaturated appearance of dusty environments. Compared with standard haze synthesis, this model better reflects the color cast and particulate interference frequently observed in UAV scenes.

Low-light. Low-light degradation primarily simulates under-exposure and sensor noise. We first attenuate exposure and apply a gamma transform,

$$\mathbf{I}_0 = (\alpha_e \mathbf{I})^{\gamma_e}, \tag{35}$$

and then inject both signal-dependent shot noise and signal-independent read noise. To approximate practical sensor artifacts, the noise field is lightly smoothed to form spatially correlated noise. Mild color shifts and compression-like distortions can also be introduced at stronger settings.

A.2. Compound degradation protocol

The compound benchmark instantiates a selected subset of the $|\mathcal{S}| > 1$ cases in Eq. (1) through cascaded processing rather than direct pixel-wise mixing. For an ordered active degradation tuple $\mathcal{S} = (d^{(1)}, \ldots, d^{(K)})$, we define

$$\mathcal{G}_{\mathcal{S}} = T_{d^{(K)}}\left(\cdot; \boldsymbol{\theta}_{d^{(K)},s_K}\right) \circ \cdots \circ T_{d^{(1)}}\left(\cdot; \boldsymbol{\theta}_{d^{(1)},s_1}\right), \tag{36}$$

so that the degraded sample is generated by

$$\mathbf{Y} = \mathcal{G}_{\mathcal{S}}(\mathbf{I}) \tag{37}$$

$$= T_{d^{(K)}}\left(\cdots T_{d^{(2)}}\left(T_{d^{(1)}}(\mathbf{I}; \boldsymbol{\theta}_{d^{(1)},s_1}); \boldsymbol{\theta}_{d^{(2)},s_2}\right) \cdots ; \boldsymbol{\theta}_{d^{(K)},s_K}\right). \tag{38}$$

In MDVD-108K we use $K = 2$ and a fixed cross-category order: a weather operator $d^{(1)} \in \{\text{fog}, \text{rain}, \text{dust}\}$ is applied first, followed by an imaging operator $d^{(2)} \in \{\text{motion}, \text{low-light}\}$. This yields six compound settings: fog+motion, fog+low-light, dust+motion, dust+low-light, rain+motion, and rain+low-light. Sequential composition preserves the physical interpretation of each stage. For example, fog+motion first attenuates scene radiance through scattering and then applies directional blur, while rain+low-light couples visible rain structures with subsequent under-exposure and sensor noise. Compared with direct pixel-wise mixing, this cascade better matches the progressive accumulation of multiple degradations in practical imaging pipelines.

Table 12
Stratified sample counts of MDVD-108K.

Split	Single degradation	Compound degradation	Total
Training	$12,000 \times 6$	$2,400 \times 6$	86,400
Validation	$1,500 \times 6$	300×6	10,800
Testing	$1,500 \times 6$	300×6	10,800
Real	–	–	500
Total	$15,000 \times 6$	$3,000 \times 6$	108,500

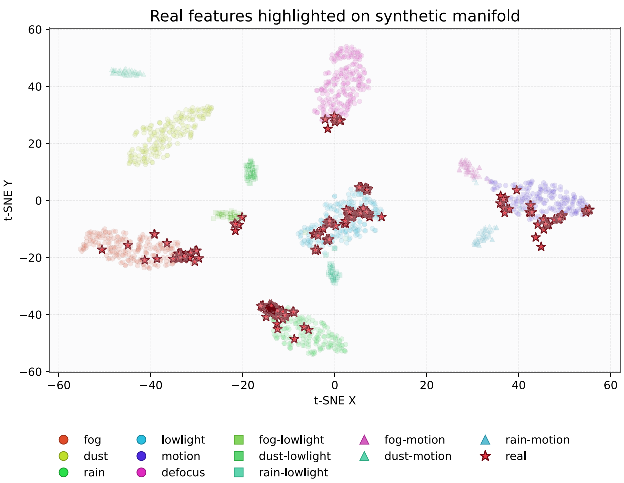

Fig. 15. Synthetic-real comparison in the DPE feature space. Colored points are synthetic degradations and red stars are real UAV samples. The broader synthetic manifold covers the real samples.

A.3. Dataset statistics and split policy

Table 12 shows that MDVD-108K contains 15,000 samples for each single degradation type and 3000 samples for each compound setting, yielding $(15,000 + 3000) \times 6 = 108,000$ synthesized images in total. The split ratio is fixed to $8:1:1$ for training, validation, and testing. Each split is synthesized only from its own clean source images, so the same undegraded image never contributes samples across train/validation/test partitions. The real-world degraded subset is reserved for evaluation only and does not participate in degradation synthesis. Overall, MDVD-108K combines physically motivated operator design with strict split separation, making it controllable, reproducible, and suitable for systematic evaluation.

A.4. Synthetic-real distribution analysis

To examine the distribution gap between synthesized and real degradations, we compare synthetic degraded samples and real-world degraded UAV images in the DPE feature space. Fig. 15 overlays real samples on the synthetic manifold. The synthetic samples span a broader set of degradation regions, while the real samples are distributed inside or near these synthetic clusters rather than forming an isolated out-of-distribution group. This distributional relationship indicates that MDVD-108K covers the main real-world degradation appearances observed in UAV imagery within the learned degradation-prior space.

References

Ai, Y., Huang, H., He, R., 2024. LoRA-IR: Taming low-rank experts for efficient all-in-one image restoration. ArXiv preprint arXiv:2410.15385.

Cai, Y., Bian, H., Lin, J., Wang, H., Timofte, R., Zhang, Y., 2023. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 12504–12513.

Chen, L., Chu, X., Zhang, X., Sun, J., 2022. Simple baselines for image restoration. In: Proceedings of the European Conference on Computer Vision. Springer, pp. 17–33.

Chen, H., Wang, Y., Guo, T., Xu, C., Deng, Y., Liu, Z., Ma, S., Xu, C., Xu, C., Gao, W., 2021. Pre-trained image processing transformer. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12299–12310.

Chen, X., Wang, X., Zhou, J., Qiao, Y., Dong, C., 2023. Activating more pixels in image super-resolution transformer. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 22367–22377.

Chen, Y., Yuan, Q., Xie, H., Tang, Y., Xiao, Y., He, J., Guan, R., Liu, X., Zhang, L., 2025. Hyperspectral video tracking with spectral-spatial fusion and memory enhancement. IEEE Trans. Image Process.

Chen, Z., Zhang, Y., Gu, J., Kong, L., Yuan, X., et al., 2022. Cross aggregation transformer for image restoration. Adv. Neural Inf. Process. Syst. 35, 25478–25490.

Chen, H., Zhang, J., Wang, H., Wang, S., Huang, P., Li, J., Guo, H., Wang, D., Wang, Z., Du, B., 2026. Any2any: Unified arbitrary modality translation for remote sensing. ArXiv preprint arXiv:2603.04114.

Cheng, Q., Shen, H., Zhang, L., Yuan, Q., Zeng, C., 2014. Cloud removal for remotely sensed images by similar pixel replacement guided with a spatio-temporal MRF model. ISPRS J. Photogramm. Remote Sens. 92, 54–68.

Cui, Y., Liu, P., 2026. A unified foundation model for all-in-one multi-modal remote sensing image restoration and fusion with language prompting. ArXiv preprint arXiv:2604.05629.

Cui, Y., Liu, P., Ma, Y., Chen, L., Xu, M., Guo, X., 2025a. Pansharpening via predictive filtering with element-wise feature mixing. ISPRS J. Photogramm. Remote Sens. 219, 22–37.

Cui, Y., Liu, P., Zhang, H., 2025b. Empowering your pansharpening models with generalizability: Unified distribution is all you need. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 11850–11860.

Dong, C., Loy, C.C., He, K., Tang, X., 2015. Image super-resolution using deep convolutional networks. IEEE Trans. Pattern Anal. Mach. Intell. 38 (2), 295–307.

Dong, Z., Zhang, Z., Sun, Y., Jiang, H., Liu, T., Gu, Y., 2026. PhyDAE: Physics-guided degradation-adaptive experts for all-in-one remote sensing image restoration. IEEE Trans. Geosci. Remote Sens.

Gao, T., Wen, Y., Zhang, K., Zhang, J., Chen, T., Liu, L., Luo, W., 2023. Frequency-oriented efficient transformer for all-in-one weather-degraded image restoration. IEEE Trans. Circuits Syst. Video Technol. 34 (3), 1886–1899.

Guo, H., Li, J., Dai, T., Ouyang, Z., Ren, X., Xia, S.-T., 2024. Mambair: A simple baseline for image restoration with state-space model. In: European Conference on Computer Vision. Springer, pp. 222–241.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778.

He, W., Zhang, H., Zhang, L., Shen, H., 2015. Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration. IEEE Trans. Geosci. Remote Sens. 54 (1), 178–188.

Hu, S., Ma, J., Wan, J., Min, W., Jing, Y., Zhang, L., Tao, D., 2025. ClusIR: Towards cluster-guided all-in-one image restoration. ArXiv preprint arXiv:2512.10948.

Hu, J., Shen, L., Sun, G., 2018. Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 7132–7141.

Jiang, J., Zuo, Z., Wu, G., Jiang, K., Liu, X., 2025. A survey on all-in-one image restoration: Taxonomy, evaluation and future trends. IEEE Trans. Pattern Anal. Mach. Intell. 47 (12), 11892–11911.

Jin, Z., Qiu, Y., Zhang, K., Li, H., Luo, W., 2025. MB-TaylorFormer V2: Improved multi-branch linear transformer expanded by taylor formula for image restoration. IEEE Trans. Pattern Anal. Mach. Intell.

Jing, R., Duan, F., Lu, F., Zhang, M., Zhao, W., 2023. Denoising diffusion probabilistic feature-based network for cloud removal in sentinel-2 imagery. Remote. Sens. 15 (9), 2217.

Khanam, R., Hussain, M., 2024. Yolov11: An overview of the key architectural enhancements. ArXiv preprint arXiv:2410.17725.

Kong, X., Dong, C., Zhang, L., 2024. Towards effective multiple-in-one image restoration: A sequential and prompt learning strategy. ArXiv preprint arXiv:2401.03379.

Li, B., Liu, X., Hu, P., Wu, Z., Lv, J., Peng, X., 2022. All-in-one image restoration for unknown corruption. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 17452–17462.

Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., Wang, Z., 2019. Benchmarking single-image dehazing and beyond. IEEE Trans. Image Process. 28 (1), 492–505.

Li, J., Wang, Y., Sheng, Q., Wu, Z., Wang, B., Ling, X., Liu, X., Du, Y., Gao, F., Camps-Valls, G., et al., 2025. CloudRuler: Rule-based transformer for cloud removal in Landsat images. Remote Sens. Environ. 328, 114913.

Li, J., Wu, Z., Hu, Z., Zhang, J., Li, M., Mo, L., Molinier, M., 2020. Thin cloud removal in optical remote sensing images based on generative adversarial networks and physical model of cloud distortion. ISPRS J. Photogramm. Remote Sens. 166, 373–389.

Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., Timofte, R., 2021. SwinIR: Image restoration using swin transformer. In: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. pp. 1833–1844.

Lihe, Z., Yuan, Q., He, J., Jin, X., Xiao, Y., Chen, Y., Shen, H., Zhang, L., 2025. Ada4DIR: An adaptive model-driven all-in-one image restoration network for remote sensing images. Inf. Fusion 118, 102930.

Liu, F., Chen, D., Guan, Z., Zhou, X., Zhu, J., Ye, Q., Fu, L., Zhou, J., 2024. RemoteCLIP: A vision language foundation model for remote sensing. IEEE Trans. Geosci. Remote Sens. 62, 1–16.

Liu, Y., Li, T., Tan, C., Ren, W., Ancuti, C., Lin, W., 2026. IHDCP: Single image dehazing using inverted haze density correction prior. IEEE Trans. Image Process.

Liu, R., Luo, T., Huang, S., Wu, Y., Jiang, Z., Zhang, H., 2024. CrossMatch: Cross-view matching for semi-supervised remote sensing image segmentation. IEEE Trans. Geosci. Remote Sens. 62, 1–15. http://dx.doi.org/10.1109/TGRS.2024.3507050.

Loshchilov, I., Hutter, F., 2017a. Decoupled weight decay regularization. In: International Conference on Learning Representations.

Loshchilov, I., Hutter, F., 2017b. SGDR: Stochastic gradient descent with warm restarts. In: International Conference on Learning Representations.

Luo, Z., Gustafsson, F.K., Zhao, Z., Sjölund, J., Schön, T.B., 2024. Controlling vision-language models for universal image restoration. In: International Conference on Learning Representations. pp. 1–13.

Ma, J., Cheng, T., Wang, G., Zhang, Q., Wang, X., Zhang, L., 2023. Prores: Exploring degradation-aware visual prompt for universal image restoration. ArXiv preprint arXiv:2306.13653.

Ma, J., Hu, S., Zhang, X., Wan, J., Huang, J., Zhang, L., Khan, S., 2025. EvoIR: Towards all-in-one image restoration via evolutionary frequency modulation. ArXiv preprint arXiv:2512.05104.

Mou, C., Wang, Q., Zhang, J., 2022. Deep generalized unfolding networks for image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 17399–17410.

Özdenizci, O., Bhatt, M., Legenstein, R., 2023. Restoring vision in adverse weather conditions with patch-based denoising diffusion models. IEEE Trans. Pattern Anal. Mach. Intell. 45 (8), 10346–10357.

Pan, J., Xu, J., Yu, X., Ye, G., Wang, M., Chen, Y., Ma, J., 2024. HDRSA-Net: Hybrid dynamic residual self-attention network for SAR-assisted optical image cloud and shadow removal. ISPRS J. Photogramm. Remote Sens. 218, 258–275.

Potlapalli, V., Zamir, S.W., Khan, S., Khan, F.S., 2024. PromptIR: Prompting for all-in-one image restoration. In: Advances in Neural Information Processing Systems, vol. 36. pp. 71275–71293.

Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al., 2021. Learning transferable visual models from natural language supervision. In: Proceedings of the International Conference on Machine Learning. pp. 8748–8763.

Riquelme, C., Puigcerver, J., Mustafa, B., Neumann, M., Jenatton, R., Susano Pinto, A., Keysers, D., Houlsby, N., 2021. Scaling vision with sparse mixture of experts. Adv. Neural Inf. Process. Syst. 34, 8583–8595.

Shazeer, N., Mirhoseini, A., Maziarz, K., Davis, A., Le, Q., Hinton, G., Dean, J., 2017. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. Int. Conf. Learn. Represent.

Shen, H., Li, X., Cheng, Q., Zeng, C., Yang, G., Li, H., Zhang, L., 2015. Missing information reconstruction of remote sensing data: A technical review. IEEE Geosci. Remote. Sens. Mag. 3 (3), 61–85.

Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A.P., Bishop, R., Rueckert, D., Wang, Z., 2016. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1874–1883.

Shu, Q., Zhu, X., Xu, S., Wang, Y., Liu, D., 2025. RESTORE-DiT: Reliable satellite image time series reconstruction by multimodal sequential diffusion transformer. Remote Sens. Environ. 328, 114872.

Song, Y., He, Z., Qian, H., Du, X., 2023. Vision transformers for single image dehazing. IEEE Trans. Image Process. 32, 1927–1941.

Stucker, C., Garnot, V.S.F., Schindler, K., 2023. U-TILISE: A sequence-to-sequence model for cloud removal in optical satellite time series. IEEE Trans. Geosci. Remote Sens. 61, 1–16.

Sui, J., Ma, Y., Yang, W., Zhang, X., Pun, M.-O., Liu, J., 2024. Diffusion enhancement for cloud removal in ultra-resolution remote sensing imagery. IEEE Trans. Geosci. Remote Sens. 62, 1–14.

Tang, X., He, X., Xu, J., Gu, X., Sun, J., 2026. Learning continuous wasserstein barycenter space for generalized all-in-one image restoration. IEEE Trans. Pattern Anal. Mach. Intell.

Valanarasu, J.M.J., Yasarla, R., Patel, V.M., 2022. Transweather: Transformer-based restoration of images degraded by adverse weather conditions. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2353–2363.

Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., 2004. Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. 13 (4), 600–612.

Wang, P., Chen, Y., Huang, B., Zhu, D., Lu, T., Dalla Mura, M., Chanussot, J., 2025. MT_GAN: A SAR-to-optical image translation method for cloud removal. ISPRS J. Photogramm. Remote Sens. 225, 180–195.

Wang, Y., Li, Y., Zheng, Z., Zhang, X.-P., Wei, M., 2025. M2Restore: Mixture-of-experts-based mamba-cnn fusion framework for all-in-one image restoration. IEEE Trans. Image Process. 34, 8086–8100.

Wang, J., Xuan, W., Qi, H., Liu, Z., Liu, K., Wu, Y., Chen, H., Song, J., Xia, J., Zheng, Z., et al., 2025. Disasterm3: A remote sensing vision-language dataset for disaster damage assessment and response. ArXiv preprint arXiv:2505.21089.

Wang, H., Zhang, J., Chen, H., Guo, H., Wang, D., Ma, J., Du, B., 2026. Residual diffusion bridge model for image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8375–8386.

Wei, B., Liu, H., Qian, C., Li, Z., Wu, W., Meng, Z., 2025. Robust single image sand removal by leveraging uncertainty-aware SAM priors and prompt learning with refined perceptual loss. In: Proceedings of the 33rd ACM International Conference on Multimedia. pp. 4932–4941.

Wen, Y., Gao, T., Chen, T., Li, Z., Liu, M., Liu, L., 2025a. Cross-level interaction and intra-level fusion network for remote sensing image dehazing. IEEE Trans. Geosci. Remote Sens.

Wen, Y., Gao, T., Li, Z., Zhang, J., Zhang, K., Chen, T., 2025b. All-in-one weather-degraded image restoration via adaptive degradation-aware self-prompting model. IEEE Trans. Multimed.

Wen, Y., Gao, T., Li, Z., Zhang, Q., Zhang, J., Chen, T., Liu, L., 2026. Structure-preserving frequency-regularized text-guided optimal transport for unpaired rain streaks and raindrops removal. IEEE Trans. Multimed.

Wen, Y., Gao, T., Zhang, J., Li, Z., Chen, T., 2023. Encoder-free multiaxis physics-aware fusion network for remote sensing image dehazing. IEEE Trans. Geosci. Remote Sens. 61, 1–15.

Wu, Z., Chen, Y., Yokoya, N., He, W., 2025. MP-HSIR: A multi-prompt framework for universal hyperspectral image restoration. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 13009–13020.

Yang, L., Kang, B., Huang, Z., Zhao, Z., Xu, X., Feng, J., Zhao, H., 2024. Depth anything v2. Adv. Neural Inf. Process. Syst. 37, 21875–21911.

Yu, X., Zhou, S., Li, H., Zhu, L., 2024. Multi-expert adaptive selection: Task-balancing for all-in-one image restoration. IEEE Trans. Circuits Syst. Video Technol. 35 (5), 4619–4634.

Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., Wang, J., et al., 2020. Deep learning in environmental remote sensing: Achievements and challenges. Remote Sens. Environ. 241, 111716.

Zamfir, E., Wu, Z., Mehta, N., Tan, Y., Paudel, D.P., Zhang, Y., Timofte, R., 2025. Complexity experts are task-discriminative learners for any image restoration. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 12753–12763.

Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.-H., 2022. Restormer: Efficient transformer for high-resolution image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5728–5739.

Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.-H., Shao, L., 2021. Multi-stage progressive image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14821–14831.

Zhang, C., Gong, D., He, J., Zhu, Y., Sun, J., Zhang, Y., 2024. UIR-LoRA: achieving universal image restoration through multiple low-rank adaptation. ArXiv preprint arXiv:2409.20197.

Zhang, J., Huang, J., Yao, M., Yang, Z., Yu, H., Zhou, M., Zhao, F., 2023. Ingredient-oriented multi-degradation learning for image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5825–5835.

Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O., 2018. The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 586–595.

Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., Fu, Y., 2018. Image super-resolution using very deep residual channel attention networks. In: Proceedings of the European Conference on Computer Vision. ECCV, pp. 286–301.

Zhang, Z., Li, J., Liang, Y., Yan, J., Xiao, Y., Su, X., Yuan, Q., 2026a. ECRformer: An efficient cloud removal transformer with semantic-decoupled learning for multimodal satellite imagery. ISPRS J. Photogramm. Remote Sens. 237, 323–338.

Zhang, K., Li, D., Luo, W., Ren, W., Liu, W., 2022. Enhanced spatio-temporal interaction learning for video deraining: Faster and better. IEEE Trans. Pattern Anal. Mach. Intell. 45 (1), 1287–1293.

Zhang, Z., Li, J., Shi, S., Yuan, Q., 2026b. Task-driven prompt learning: A joint framework for multi-modal cloud removal and segmentation. ArXiv preprint arXiv:2601.12052.

Zhang, K., Li, R., Yu, Y., Luo, W., Li, C., 2021. Deep dense multi-scale network for snow removal using semantic and depth priors. IEEE Trans. Image Process. 30, 7419–7431.

Zhang, R., Luo, Y., Liu, J., Yang, H., Dong, Z., Gudovskiy, D., Okuno, T., Nakata, Y., Keutzer, K., Du, Y., et al., 2024. Efficient deweather mixture-of-experts with uncertainty-aware feature-wise linear modulation. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 38, no. 15. pp. 16812–16820.

Zhang, X., Ma, J., Wang, G., Zhang, Q., Zhang, H., Zhang, L., 2026a. Perceive-IR: Learning to perceive degradation better for all-in-one image restoration. IEEE Trans. Image Process. 35, 2018–2033.

Zhang, K., Wang, T., Luo, W., Ren, W., Stenger, B., Liu, W., Li, H., Yang, M.-H., 2023. MC-Blur: A comprehensive benchmark for image deblurring. IEEE Trans. Circuits Syst. Video Technol. 34 (5), 3755–3767.

Zhang, Z., Yan, J., Liang, Y., Feng, J., He, H., Cao, L., 2025. Multiscale restoration of missing data in optical time-series images with masked spatial-temporal attention network. IEEE Trans. Geosci. Remote Sens. 63, 1–15.

Zhang, X., Zhang, H., Wang, G., Zhang, Q., Zhang, L., 2026b. ClearAIR: A human-visual-perception-inspired all-in-one image restoration. In: Proceedings of the AAAI Conference on Artificial Intelligence. pp. 12861–12869.

Zhang, X., Zhang, H., Wang, G., Zhang, Q., Zhang, L., Du, B., 2025. UniUIR: Considering underwater image restoration as an all-in-one learner. IEEE Trans. Image Process. 34, 6963–6977.

Zhang, Y., Zhou, Q., Gao, J., Wang, Q., 2026. STAR-IOD: Scale-decoupled topology alignment with pseudo-label refinement for remote sensing incremental object detection. ISPRS J. Photogramm. Remote Sens. 238, 650–663. http://dx.doi.org/10.1016/j.isprsjprs.2026.05.036.

Zhu, P., Wen, L., Du, D., Bian, X., Fan, H., Hu, Q., Ling, H., 2021. Detection and tracking meet drones challenge. IEEE Trans. Pattern Anal. Mach. Intell. 44 (11), 7380–7399.