Any Image Restoration via Efficient Spatial-Frequency Degradation Adaptation

Bin Ren[1,2*] **Eduard Zamfir**[3] **Zongwei Wu**[3†] **Yawei Li**[4] **Yidi Li**[5] **Danda Pani Paudel**[6] **Radu Timofte**[3] **Ming-Hsuan Yang**[7] **Luc Van Gool**[6] **Nicu Sebe**[1]

[1] *University of Trento, IT* [2] *Mohamed bin Zayed University of Artificial Intelligence, UAE*
[3] *University of Würzburg, DE* [4] *ETH Zürich, CH* [5] *Taiyuan University of Technology, CN*
[6] *INSAIT, Sofia University "St. Kliment Ohridski", BG* [7] *University of California, Merced, USA*

Reviewed on OpenReview: *https://openreview.net/forum?id=RObgCpdDqr*

Abstract

Restoring multiple degradations efficiently via just one model has become increasingly significant and impactful, especially with the proliferation of mobile devices. Traditional solutions typically involve training dedicated models per degradation, resulting in inefficiency and redundancy. More recent approaches either introduce additional modules to learn visual prompts, significantly increasing the size of the model, or incorporate cross-modal transfer from large language models trained on vast datasets, adding complexity to the system architecture. In contrast, our approach, termed AnyIR, takes a unified path that leverages inherent similarity across various degradations to enable both efficient and comprehensive restoration through a joint embedding mechanism, without scaling up the model or relying on large language models. Specifically, we examine the sub-latent space of each input, identifying key components and reweighting them first in a gated manner. To unify intrinsic degradation awareness with contextualized attention, we propose a spatial–frequency parallel fusion strategy that strengthens spatially informed local–global interactions and enriches restoration fidelity from the frequency domain. Comprehensive evaluations across four all-in-one restoration benchmarks demonstrate that AnyIR attains state-of-the-art performance while reducing model parameters by 84% and FLOPs by 80% relative to the baseline. These results highlight the potential of AnyIR as an effective and lightweight solution for further all-in-one image restoration. Our code is available at: https://github.com/Amazingren/AnyIR.

1 Introduction

Image restoration (*i.e.*, IR) aims to recover a clean image from its degraded observation, and a central challenge is how to handle multiple and diverse degradations within a unified framework. This problem is inherently ill-posed since multiple solutions may explain the same output, and thus effective priors or learned representations are crucial for successful restoration. In real-world scenarios, degradations arise from heterogeneous sources (noise, blur, compression, weather artifacts, *etc.*), and practical systems must cope with them within a single pipeline rather than as isolated tasks. Meanwhile, multi-step or per-degradation pipelines introduce storage, routing, and computation overhead during both training and deployment, whereas a single-checkpoint model better matches mobile and edge constraints and simplifies system integration. This motivates the pursuit of *a single efficient model capable of addressing multiple degradation types.*

Despite significant advances, existing IR methods still struggle to efficiently handle diverse degradations while preserving essential details Li et al. (2023a); Ren et al. (2024b;a); Wu et al. (2024c); Yin et al. (2024); Zheng et al. (2024b); Ding et al. (2024); Wu et al. (2024b); Li et al. (2025). Many current solutions rely on

* This work was partially conducted during the visiting stay at INSAIT.
† Corresponding author.

per-degradation models or multi-stage pipelines, which introduce storage, routing, and computation overhead during both training and deployment, and are difficult to scale or deploy on mobile and edge platforms. In contrast, a single-checkpoint model better matches deployment constraints and simplifies system integration. This work addresses this gap by pursuing an efficient All-in-one IR model that can handle multiple degradation types within a unified representation-learning framework.

The complexity of restoration models Liang et al. (2021); Zamir et al. (2022; 2021); Wang et al. (2022); Chen et al. (2022) arises primarily from the diversity of degradations. Consequently, most methods are tailored to specific tasks with limited generalization. A versatile system often requires integrating multiple specialized models, leading to heavy frameworks. Some studies have shown that a single architecture can handle multiple degradations, but they lack parameter unification, producing multiple checkpoints for different tasks Chen et al. (2022); Ren et al. (2024a). This reduces efficiency despite simplifying the system. Recent research has sought architectural and parameter unification Li et al. (2022); Potlapalli et al. (2024); Zhang et al. (2023); Liu et al. (2022). Diffusion-based methods demonstrate strong generative capacity Ren et al. (2023b); Jiang et al. (2023); Zhao et al. (2024), while prompt-based designs guide networks with modality-specific embeddings Wang et al. (2023a); Potlapalli et al. (2024); Li et al. (2023c). Other works leverage text as an intermediate representation Luo et al. (2024); Conde et al. (2024). Despite promising results, these approaches significantly increase model size, inference time, or rely on degradation-specific supervision.

In this paper, we take a unified perspective: *although each degradation has its characteristics, all restoration tasks share underlying principles of separating nuisance degradations from structural image information*. From a learning viewpoint, this setting can be regarded as a *multi-environment learning problem*, where degradations correspond to environments and the objective is to learn a representation that is invariant to degradation-specific nuisances yet sufficient for restoration. Unlike large-scale priors in LLMs, our method builds invariance directly from single-image cues, yielding both efficiency and strong generalization. To this end, we propose AnyIR, a lightweight framework that integrates global context and local degradation-aware cues. Specifically, we introduce a gated local block that disentangles fine-grained degradation-aware details (ego, shifted, and scaled parts), adaptively reweighted via gating, and a parallel attention pathway to capture global dependencies. A spatial–frequency fusion mechanism further intertwines the two representations, balancing structural integrity with fine detail recovery. Importantly, features are processed in sub-latent partitions before aggregation, reducing computational cost while retaining rich information. This design enables AnyIR to act as an implicit degradation-invariant learner, effective and efficient across diverse restoration settings (see Fig. 4).

Our main contributions are summarized as follows:

- We propose AnyIR , a unified and efficient all-in-one IR model that achieves superior performance while reducing computational cost by **85.6%** compared to state-of-the-art counterparts.
- We design a novel local-global gated intertwining mechanism combined with a spatial–frequency fusion strategy, enabling cohesive and adaptive embeddings without degradation-specific supervision.
- Through extensive experiments on diverse restoration tasks, we demonstrate the effectiveness and efficiency of AnyIR , providing a strong and practical baseline for future research in all-in-one IR.

2 Related Work

Image Restoration (IR) Image restoration aims to solve a highly ill-posed problem by reconstructing high-quality images from their degraded counterparts. Due to its importance, IR has been applied to various applications Richardson (1972); Banham & Katsaggelos (1997); Li et al. (2023b); Zamfir et al. (2024); Miao et al. (2024); Zheng et al. (2024a). Initially, IR was addressed through model-based solutions involving the search for solutions to specific formulations. However, learning-based approaches have gained much attention with the significant advancements in deep neural networks. Numerous approaches have been developed, including regression-based Lim et al. (2017); Lai et al. (2017); Liang et al. (2021); Chen et al. (2021b); Li et al. (2023a); Zhang et al. (2024) and generative model-based pipelines Gao et al. (2023); Wang et al. (2023b); Luo et al. (2023); Yue et al. (2023); Zhao et al. (2024); Liu et al. (2023) that are based on convolutional Dong

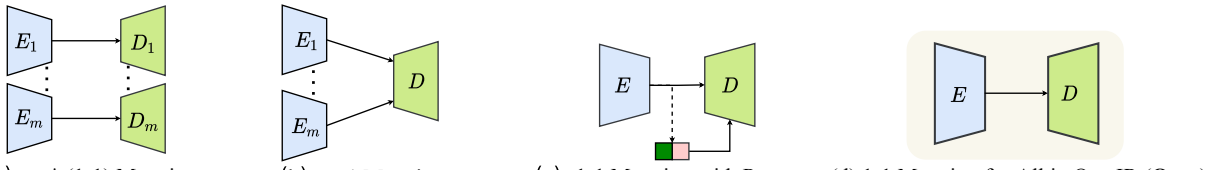


Figure 1: Mathematical formulations of mainstream paradigms for All-in-One image restoration. (a) $m \times (1\text{–}1)$ mapping: each degradation type z_i requires an independent encoder–decoder pair (E_i, D_i). (b) $m\text{–}1$ mapping with shared decoder: multiple task-specific encoders E_i share a single decoder D. (c) $1\text{–}1$ mapping with prompts: a unified encoder–decoder (E, D) is conditioned on external prompts $p(z)$. (d) Our proposed $1\text{–}1$ mapping: a pure unified framework where (E, D) directly learns degradation-invariant yet discriminative representations, without prompts or multiple encoders.

et al. (2015); Zhang et al. (2017b;a); Wang et al. (2018), MLP Tu et al. (2022), state-space mode Guo et al. (2024a); Zhu et al. (2024); Gu & Dao (2024); Dao & Gu (2024), or vision transformers-based (ViTs) architectures Liang et al. (2021); Li et al. (2023a); Zamir et al. (2022); Ren et al. (2023a); Dosovitskiy et al. (2020). Although state-of-the-art methods have achieved promising performance, mainstream IR solutions still focus on addressing single degradation tasks such as denoising Zhang et al. (2017b; 2019), dehazing Ren et al. (2020); Wu et al. (2021), deraining Jiang et al. (2020); Ren et al. (2019), and deblurring Kong et al. (2023); Ren et al. (2023b).

One for Any Image Restoration Training a task-specific model to address a single degradation is effective, but impractical for real-world deployment due to the need for separate models for each corruption. In practice, images often suffer from multiple, overlapping degradations, making it inefficient to address them individually. Such task-specific solutions require significant computing and storage resources, and their environmental cost grows with the number of degradations. To overcome these limitations, the emerging *All-in-One* IR setting develops a *single blind model* that handles multiple degradations simultaneously, without requiring task-specific specialization Tang et al. (2025); Jiang et al. (2025). Moreover, training a unified model across multiple degradations enables the learning of transferable structures, which facilitates adaptation to new or mixed degradations through lightweight fine-tuning rather than training separate models from scratch. From a learning perspective, this setting can also be interpreted as a form of multi-environment learning, where different degradations act as environments and the challenge is to learn representations that are robust across them while still sufficient for reconstruction. Several representative approaches have been proposed. AirNet Li et al. (2022) employs contrastive learning to derive degradation-aware embeddings from corrupted inputs, which are then used to guide restoration. IDR Zhang et al. (2023) follows a meta-learning perspective, decomposing degradations into fundamental physical principles and adapting via a two-stage learning process. More recently, the prompt-based paradigm Potlapalli et al. (2024); Wang et al. (2023a); Li et al. (2023c) has been introduced, where learned visual prompts condition a single model across degradations. Such prompts act as task embeddings that steer the network, with extensions including frequency-aware prompts Cui et al. (2025) or more complex designs requiring additional datasets Dudhane et al. (2024). While effective, these approaches often increase training cost and reduce efficiency. To further enhance generalization, MoCE-IR Zamfir et al. (2025) explore a mixture-of-experts design, activating specialized subnetworks for different conditions. Although this improves performance, it also increases overall architectural complexity. In contrast, our work strengthens the model's ability to capture representative degradation information without the overhead of heavy prompts or expert routing, providing a simpler and more efficient pathway for All-in-One IR.

3 AnyIR

3.1 Preliminaries

Formally, for the image restoration problem, given an observation:

$$y = \mathcal{D}_z(x) + \epsilon, \tag{1}$$

where x denotes the latent clean image, $\mathcal{D}_z$ is a degradation operator parameterized by z (*e.g.*, noise, blur, haze, rain), and ϵ is an additive perturbation, the goal of IR is to recover x from its degraded counterpart y.

In practice, this inverse mapping is inherently ill-posed, as multiple clean images may correspond to the same degraded observation. In other words, different x_1 and x_2 may both satisfy $\mathcal{D}_z(x) + \epsilon = y$. Therefore, IR does not generally admit a mathematically unique solution. Instead of explicitly enforcing the invertibility of $\mathcal{D}_z$, modern learning-based approaches aim to estimate a statistically plausible reconstruction conditioned on y.

From a probabilistic perspective, the restoration model can be interpreted as a data-driven estimator of the conditional distribution $p(x \mid y)$. During training, supervision encourages convergence toward the ground-truth mode among feasible candidates, while at inference time, the prediction reflects a statistically consistent solution favored by the learned image prior rather than an arbitrary selection among multiple possibilities. This viewpoint aligns with the standard treatment of ill-posed inverse problems in the literature and explains why learning-based IR models can produce stable and perceptually coherent reconstructions even under mixed or previously unseen degradations.

This probabilistic and ill-posed nature of IR also explains why different architectural paradigms may exhibit distinct reconstruction behaviors, depending on how degradation information and prior knowledge are encoded. To situate our design within the broader landscape of All-in-One IR, we first revisit several representative architectural paradigms that differ in their degree of parameter sharing, conditioning strategy, and task dependence. This comparison provides the conceptual motivation for our formulation and clarifies the design choices made in our framework. As summarized in Fig. 1, these approaches can be interpreted as mappings from a degraded observation y to a restored image $\hat{x}$, with different assumptions about how degradation type is modeled or encoded in the network Jiang et al. (2025).

(a) $m \times (1-1)$ **Mapping.** Each degradation type $z_i \in \{1, \ldots, m\}$ is handled by an individual encoder–decoder pair (E_i, D_i) Li et al. (2023a); Liang et al. (2021):

$$\hat{x}_i = D_i(E_i(y)), \quad i = 1, \ldots, m. \tag{2}$$

This strategy yields strong task specialization but scales linearly with the number of degradations and prevents knowledge sharing across tasks.

(b) $m-1$ **Mapping with Shared Decoder.** Multiple task-specific encoders are retained, while a shared decoder D reconstructs the output Li et al. (2020):

$$\hat{x}_i = D(E_i(y)), \quad i = 1, \ldots, m. \tag{3}$$

This partially amortizes parameters through a common reconstruction head, yet the storage and computation of m encoders still limit efficiency and scalability.

(c) $1-1$ **Mapping with Prompts.** A single encoder–decoder pair (E, D) is conditioned on an external prompt $p(z)$ describing the degradation Li et al. (2023c); Zhang et al. (2025); Li et al. (2023c):

$$\hat{x} = D(E(y), p(z)), \tag{4}$$

where $p(z)$ may correspond to learned tokens or textual embeddings. This formulation improves flexibility and controllability across degradations, but introduces auxiliary conditioning modules and requires careful prompt modeling and tuning.

(d) $1-1$ **Mapping for All-in-One IR (adopted by our solution).** We advocate a pure one-to-one mapping, where a single encoder–decoder directly captures both degradation-sensitive cues and degradation-invariant structures Cui et al. (2025); Tang et al. (2025):

$$\hat{x} = D(E(y)). \tag{5}$$

Unlike prompt-based formulations, no explicit task tokens or routing mechanisms are introduced; instead, the network architecture is designed to enable implicit disentanglement within the shared representation space, favoring parameter efficiency and generalization across degradations.

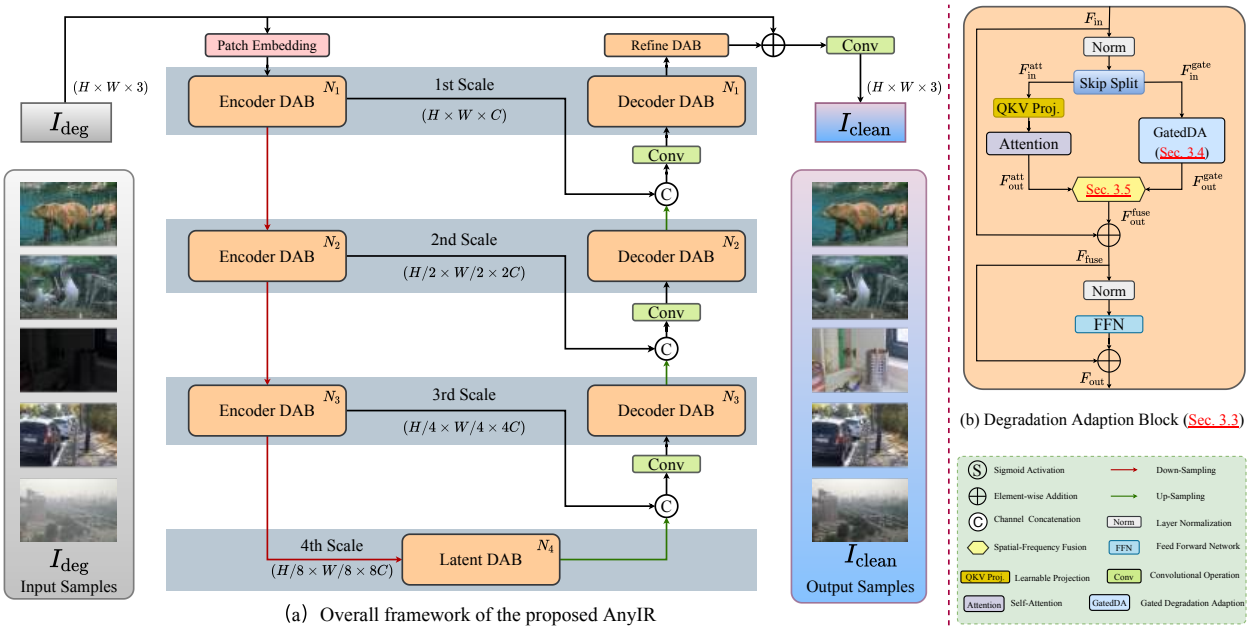


Figure 2: (a) Framework of the proposed AnyIR: *i.e.*, a convolutional patch embedding, a U-shape encoder-decoder main body, and an extra refined block. (b) Structure of degradation adaptation block (DAB).

This formulation-centric perspective highlights the trade-offs between specialization, conditioning overhead, and parameter sharing in prior All-in-One IR systems, and motivates our design choice of a unified $1-1$ mapping that seeks to balance representational robustness with computational efficiency.

3.2 Overall Framework

Building on the design space reviewed in Sec. 3.1, we now present our efficient All-in-One IR method, termed AnyIR . As motivated by Fig. 1(d), our goal is to realize a pure $1-1$ mapping that avoids multiple encoders or external prompts, while still capturing degradation-specific cues and invariant structures within a single unified framework.

The overall architecture of AnyIR is illustrated in Fig. 2. At a macro level, it adopts a U-shaped network Ronneberger et al. (2015) with four hierarchical levels. Each level incorporates $N_i, i \in [1, 2, 3, 4]$ instances of our proposed *degradation adaptation block* (DAB, Sec. 3.3), where each DAB is composed of the gated degradation adaptation module (GatedDA, Sec. 3.4) and the spatial–frequency fusion algorithm (Sec. 3.5). Initially, a convolutional layer extracts shallow features from the degraded input, creating a patch embedding of size $H \times W \times C$. As in standard U-Nets, each encoder stage doubles the embedding dimension and halves the spatial resolution, with skip connections transferring information to the corresponding decoder stage. In the decoder, features are merged with the previous decoding stage via linear projection. Finally, a global skip connection links input to output, preserving high-frequency details and producing the restored image.

3.3 Degradation Adaptation Block

The proposed degradation adaptation block (DAB) serves as the fundamental unit of AnyIR , and its structure is shown in Fig. 2(b). The design principle is to decouple global and local processing in a parameter-efficient manner, while still retaining rich feature diversity. Given an input feature $F_{\text{in}} \in \mathbb{R}^{H \times W \times C}$, we employ a selective channel-wise partitioning strategy:

$$\begin{aligned} F_{\text{in}}^{\text{att}} &= \{F_{\text{in}}^{(2k-1)} \mid k \in \mathbb{Z}^+, k \leq \tfrac{C}{2}\}, \\ F_{\text{in}}^{\text{gate}} &= \{F_{\text{in}}^{(2k)} \mid k \in \mathbb{Z}^+, k \leq \tfrac{C}{2}\}, \end{aligned} \tag{6}$$

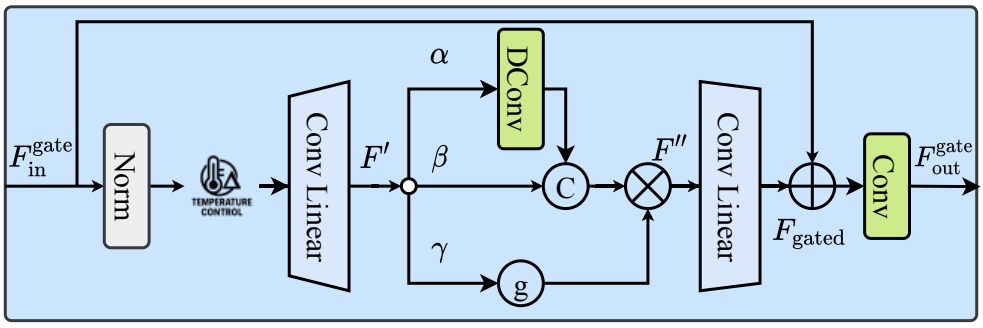


Figure 3: Structure of our GatedDA. $\oplus$, ⓒ, ⓖ, and $\otimes$ denote the element-wise addition, channel-wise concatenation, GELU Hendrycks & Gimpel (2016) activation, and element-wise multiplication, respectively.

where $F_{\text{in}}^{\text{att}}$ and $F_{\text{in}}^{\text{gate}} \in \mathbb{R}^{H \times W \times \frac{C}{2}}$ denote two interleaved channel groups. This skip-split partitioning is not intended to create two isolated feature branches; instead, it enforces controlled interaction across partial channel groups, so that degradation-variant cues can be emphasized while degradation-invariant structural information remains accessible within the same shared representation space. This yields two complementary pathways: the attention branch $F_{\text{in}}^{\text{att}}$ focuses on long-range dependency modeling, while the gated branch $F_{\text{in}}^{\text{gate}}$ specializes in degradation-sensitive local adaptation. Compared with conventional half-split operations, the skip-split design (i) reduces the effective dimensionality per branch, lowering the complexity of attention layers Dosovitskiy et al. (2020), and (ii) preserves feature diversity by uniformly sampling channels, thereby avoiding information loss. A detailed analysis is provided in Sec. 5, and the visual illustration is shown in the appendix.

To capture the complex dependencies inherent to global degradation, we employ a multi-depth convolution head attention mechanism Zamir et al. (2022); Potlapalli et al. (2024) on the feature subset $F_{\text{in}}^{\text{att}}$, resulting in $F_{\text{out}}^{\text{att}}$. This attention approach is particularly effective for image restoration, where degradation is often non-uniform and shaped by various intricate factors. Specifically, $F_{\text{in}}^{\text{att}}$ is transformed into Query (Q), Key (K), and Value (V) matrices, defined as: $Q = F_{\text{in}}^{\text{att}}\mathbf{W}_{\text{qry}}, \quad K = F_{\text{in}}^{\text{att}}\mathbf{W}_{\text{key}}, \quad V = F_{\text{in}}^{\text{att}}\mathbf{W}_{\text{val}}$, where $\mathbf{W}_{\text{qry}}$, $\mathbf{W}_{\text{key}}$, and $\mathbf{W}_{\text{val}}$ are learnable weights. The output $F_{\text{out}}^{\text{att}}$ is then calculated as:

$$F_{\text{out}}^{\text{att}} = \sum_i \texttt{Softmax}\left(\frac{QK^\top}{\sqrt{d}}\right)_i V_{i,j}, \tag{7}$$

where $\sqrt{d}$ is a scaling factor to normalize attention scores, stabilize gradients, and facilitate convergence. This attention mechanism excels at modeling long-range dependencies throughout the feature space, a vital capability for image restoration where degradation, such as noise, blur, and artifacts, exhibits spatial correlations but varies across the image Liang et al. (2021); Zamir et al. (2022); Li et al. (2023a).

In parallel, the gated branch $F_{\text{in}}^{\text{gate}}$ is processed by the GatedDA module (Sec. 3.4), yielding $F_{\text{out}}^{\text{gate}}$. This path complements attention by enhancing localized, degradation-aware features Liang et al. (2021). Thus, the DAB combines the strengths of global modeling (attention) and local selectivity (gated convolution).

To unify both pathways, we adopt a spatial–frequency fusion strategy (Alg. 2), producing the fused representation $F_{\text{out}}^{\text{fuse}}$. Finally, layer normalization, a feed-forward network (FFN), and a residual connection are applied:

$$F_{\text{out}} = \text{FFN}(\text{Norm}(F_{\text{fuse}})) + F_{\text{fuse}}. \tag{8}$$

This sequence stabilizes feature statistics and enhances expressivity, yielding a balanced representation that integrates global context with fine-grained degradation cues for high-fidelity restoration.

3.4 Gated Degradation Adaption

To capture local degradation-aware details, we leverage the selective properties of gated convolution, forming the GatedDA (Fig. 3). Given an input feature $F_{\text{in}}^{\text{gate}} \in \mathbb{R}^{H \times W \times C}$, a layer normalization is first applied

Algorithm 1 Gated Degradation Adaption

Require: Input $F_{\text{in}}^{\text{gate}} \in \mathbb{R}^{H \times W \times C}$, initial temperature τ
Ensure: Output $F_{\text{out}}^{\text{gate}} \in \mathbb{R}^{H \times W \times C}$
1: $\hat{F} \leftarrow \text{Norm}(F_{\text{in}}^{\text{gate}})$ // Normalize
2: $\mu \leftarrow \frac{1}{HW} \sum_{h=1}^{H} \sum_{w=1}^{W} \hat{F}_{:,:,h,w}$ // per-channel mean
3: $\sigma \leftarrow \sqrt{\frac{1}{HW} \sum_{h=1}^{H} \sum_{w=1}^{W} (\hat{F}_{:,:,h,w} - \mu)^2 + \delta}$ // per-channel std, δ: small constant
4: $\Delta \leftarrow \text{Sigmoid}(\mu + \sigma)$ // Temp adjust
5: $\tau_{\text{adj}} \leftarrow \tau \cdot \Delta$
6: $F' \leftarrow \hat{F}\mathbf{W}_{\text{exp}}$ // Channel expand
7: $\gamma, \beta, \alpha \leftarrow \text{split}(F')$ // Split γ, β, α
8: $\alpha' \leftarrow (\alpha \mathbf{W}_{\text{depth}}) \cdot (1 + \tau_{\text{adj}})$ // Depthwise conv
9: $F_{\text{gated}} \leftarrow \sigma(\gamma) \cdot \text{concat}(\beta, \alpha')\mathbf{W}_{\text{gate}}$ // Gate combine
10: $F_{\text{out}}^{\text{gate}} \leftarrow (F_{\text{gated}} + F_{\text{in}}^{\text{gate}})\mathbf{W}_{\text{proj}}$ // Residual, proj
11: **return** $F_{\text{out}}^{\text{gate}}$

to stabilize the feature distributions, followed by a 1×1 convolution that expands the channel dimension to hidden $= r_{\text{expan}} \cdot C$, where r_{expan} is the expansion ratio, as shown in Alg. 1. To adaptively respond to varying intensities, we introduce a temperature adjustment mechanism. Based on the input's *mean* and *standard* deviation, the initial temperature τ is modulated as $\tau_{\text{adj}} = \tau \cdot \Delta$, where Δ serves as a dynamic scaling factor (steps 2–5 of Alg. 1). This design is motivated by the fact that many degradations exhibit spatially non-uniform characteristics; therefore, rather than applying a single global transformation, the gated module performs content- and region-aware modulation on features, enabling localized adaptation to different degradation strengths. This adjustment improves the model's sensitivity to subtle degradation patterns, promoting detailed feature capture.

The expanded feature F' ($\mathbf{W}_{\text{exp}}$ denotes the 1×1 convolution weights used for channel expansion) is then split into three parts along channels: α, β, and γ (as shown in step 7, which correspond to the *scaled*, *ego*, and *shifted* features). Specifically, we take ($^{\text{hidden}}/_4$, $^{\text{hidden}}/_4$, $^{\text{hidden}}/_2$) for (α, β, γ) in this paper. Here, α undergoes depthwise convolution with τ_{adj} to capture spatial details, β retains the original information, and γ is activated with GELU Hendrycks & Gimpel (2016) to enable a non-linear gated selection mechanism. Specifically, these components are recombined and modulated, selectively emphasizing critical features (Step 9 in Alg. 1). Finally, F'' is projected back to the original channel dimension and combined with F_{in}^{gate} through a skip connection to prevent loss of information. A final 1×1 convolution fuses the degradation-adapted features with the input, resulting in the output $F_{\text{out}}^{\text{gate}}$ (Steps 10–11 in Alg. 1). Please note that $\mathbf{W}_{\text{depth}}$, $\mathbf{W}_{\text{gate}}$, and $\mathbf{W}_{\text{proj}}$ denote the learnable convolution weights for depth-wise convolution, gated convolution, and projection convolution operations.

Owing to its design, GatedDA dynamically adjusts its internal temperature and gating behavior, enabling the network to capture degradation-aware features adaptively. This makes it a natural complement to the global attention branch, providing localized detail enhancement and stronger robustness to spatially varying degradations. More analysis is provided in the appendix.

From an efficiency perspective, the proposed design improves computational economy at the architectural level rather than relying on model scaling. First, each Degradation Adaptation Block partitions the feature channels into two subsets, where only half of the channels are processed by global attention. Since our attention operator follows the Restormer-style channel-wise formulation whose complexity grows quadratically with spatial resolution, reducing the attention channels by half directly lowers the computational cost of the attention pathway while preserving its ability to model long-range dependencies. Second, the remaining channels are processed by the proposed GatedDA module, which enhances the representational capacity of the block through lightweight, convolution-based local adaptation; this branch has substantially lower complexity than applying attention on the same number of channels, further contributing to overall efficiency. Finally, because each block combines a stronger per-block representation with a more balanced global–local decomposition, the backbone can be instantiated with a smaller U-shaped hierarchy (*e.g.*, [3, 5, 5, 7] instead

Algorithm 2 Spatial-Frequency Fusion

Require: F_{att}, F_{gate}, fusion weight λ
Ensure: Fused output F_{fuse}
Main Procedure:
$F_s \leftarrow \textsc{SpatialFusion}(F_{\text{att}}, F_{\text{gate}})$ // spatial cross-enhance
$F_f \leftarrow \textsc{FrequencyFusion}(F_{\text{att}}, F_{\text{gate}})$ // detail-enhance
$F_{\text{fuse}} \leftarrow \lambda \cdot F_s + (1-\lambda) \cdot F_f$ // weighted fusion
return F_{fuse}
Function: SpatialFusion
function $\textsc{SpatialFusion}(F_{\text{att}}, F_{\text{gate}})$
    $F^{\text{ag}} \leftarrow F_{\text{att}} + \text{Sigmoid}(F_{\text{gate}})$ // gate-enhanced attention
    $F^{\text{ga}} \leftarrow F_{\text{gate}} + \text{Sigmoid}(F_{\text{att}})$ // attention-enhanced gate
    return $\text{Concat}(F^{\text{ag}}, F^{\text{ga}})$ // channel-wise merge
end function
Function: FrequencyFusion
function $\textsc{FrequencyFusion}(F_{\text{att}}, F_{\text{gate}})$
    $\hat{F}_{\text{att}} \leftarrow \text{rfft}_{2D}(F_{\text{att}})$, $\hat{F}_{\text{gate}} \leftarrow \text{rfft}_{2D}(F_{\text{gate}})$ // real FFT
    $\hat{F} \leftarrow \hat{F}_{\text{att}} + \hat{F}_{\text{gate}}$ // freq combine
    $F_{\text{freq}} \leftarrow \text{irfft}_{2D}(\hat{F})$ // inverse FFT
    return $\text{Repeat}(F_{\text{freq}})$ // match channels
end function

of the deeper [4, 6, 6, 8] configurations used in prior All-in-One IR methods such as PromptIR Potlapalli et al. (2024) and MoCE-IR Zamfir et al. (2025)), while still achieving superior restoration accuracy.

3.5 Spatial-Frequency Fusion Algorithm

To enable effective interaction between contextualized global attention and degradation-sensitive local GatedDA features, we introduce a spatial–frequency fusion algorithm (*that is*, Alg. 2). The fusion is carried out in two complementary domains: spatial and frequency.

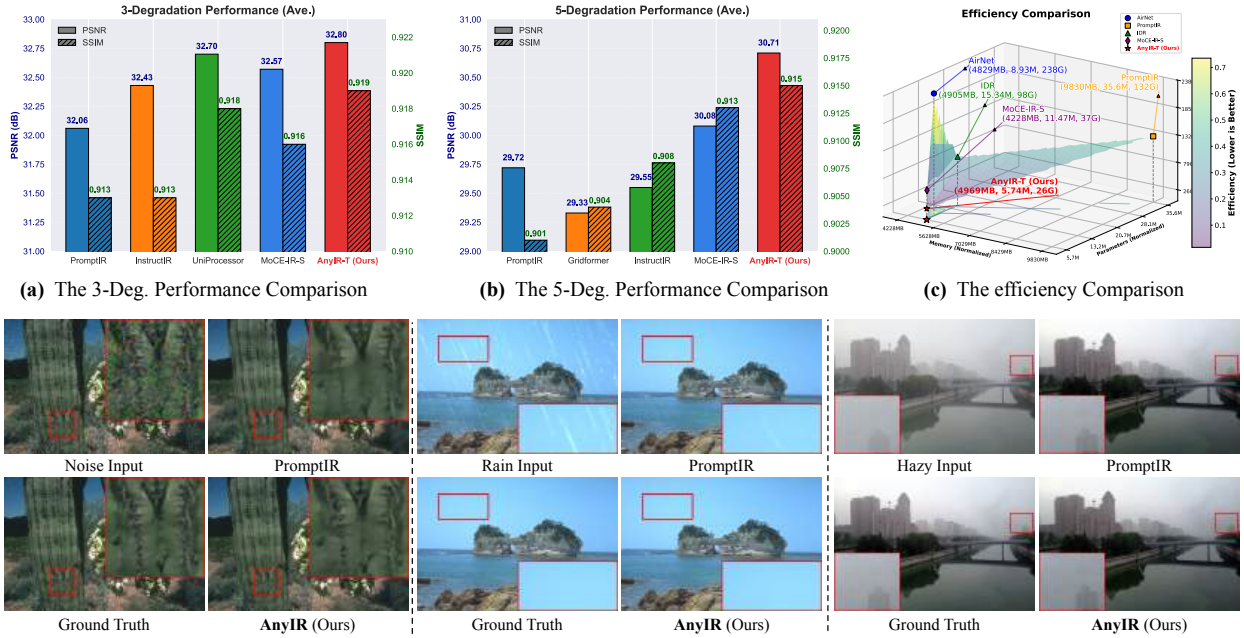
In the spatial branch, we apply a cross-enhancement mechanism (*i.e.*, SpationFusion in Alg. 2) to refine the attention feature F_{att} and gated feature F_{gate} mutually. Each branch is modulated by the sigmoid-activated signal from the other, enabling a dynamic message passing across representations. The enhanced features are then concatenated along the channel dimension to form the spatial representation.

In parallel, we perform a lightweight frequency-domain fusion to capture structural alignment. F_{att} and F_{gate} are first transformed using real-valued 2D Fast Fourier Transform (*i.e.*, $\text{rfft}_{2D}(*)$) in Step-11 of Alg. 2, additively combined in the frequency domain (Step-12), and then projected back via inverse FFT (*i.e.*, $\text{irfft}_{2D}(*)$) (Step-13). The output is duplicated (*i.e.*, $\text{Repeat}(*)$) along the channel to match the spatial counterpart (Step-14). This design leverages the complementary roles of the two domains: the spatial branch emphasizes local structural fidelity, whereas the frequency branch stabilizes global statistics and degradation patterns, which is particularly beneficial under mixed or compound degradations.

The final fused representation is obtained by a weighted summation of the two branches (Step-3 of Alg. 2), controlled by a scalar λ. A convolutional projection with learnable weights $\mathbf{W}_{\text{fuse}}$ is then applied, followed by a residual connection to the original input:

$$F_{\text{out}}^{\text{fuse}} \leftarrow \text{Conv}_{\mathbf{W}_{\text{fuse}}}(F_{\text{fuse}}) + F_{\text{in}}. \tag{9}$$

This fusion strategy leverages the complementary strengths of attention-based global context and gated local priors from both spatial and frequency domains, producing a rich and adaptive representation for downstream restoration. Notably, the design aligns with signal-domain interpretability while enhancing generalization across diverse degradation types.



(a) The 3-Deg. Performance Comparison **(b)** The 5-Deg. Performance Comparison **(c)** The efficiency Comparison

(d) Visual Comparison of the 3-Deg. setting for *denoising*, *draining*, and *dehazing* (Zoom in for a better view).

Figure 4: Overall quantitative performance (3-Degradation & 5-Degradation settings), efficiency comparison, and the qualitative comparison (Zoom in for a better view).

4 Experimental Results

We conduct experiments adhering to the protocols of previous general IR works Potlapalli et al. (2024); Zhang et al. (2023) in three main settings: *(1) All-in-One (3Degrations)*, *(2) All-in-One (5Degrations)*, *(3) Mix-Degradation Setting*, and *(4) Zero-Shot Unseen Setting* More experimental details and the introduction of the data set are provided in our *Supp. Mat.*

Before presenting the detailed results for each setting, we first provide an overview comparison to summarize the overall performance, efficiency, and visual restoration quality of AnyIR against representative All-in-one IR methods, as shown in Fig. 4.

4.1 State of the Art Comparisons

Three Degradations. We evaluate our All-in-One restorer, AnyIR , against other specialized methods listed in Tab. 1, all trained on three degradations: dehazing, deraining, and denoising. AnyIR consistently outperforms all the comparison methods, even for those with the assistance of language, multi-task, or prompts. In particular, AnyIR outperforms the baseline method PromptIR by **1.12dB**, **2.14dB** on dehazing and draining, and **0.74dB** on average, while maintaining **80%** fewer parameters.

Five Degradations. Extending the three degradation tasks to include deblurring and low-light enhancement Li et al. (2022); Zhang et al. (2023), we validate the comprehensive performance of our method in an All-in-One setting. As shown in Tab. 2, AnyIR effectively leverages degradation-specific features, surpassing AirNet Li et al. (2022) and IDR Zhang et al. (2023) by an average of 5.16 dB and 2.31 dB, respectively, with 33% and 60% fewer parameters. Compared to the most recent method, MoCE-IR Zamfir et al. (2025), besides 0.05 dB lower on deblurring, we outperform it on all the rest tasks with a 0.57dB PSNR improvement on average.

Mixed Degradation. We conduct experiments on the CDD-11 Guo et al. (2024b) dataset, a challenging benchmark for mixed degradation restoration tasks, combining real-world degradations such as low-light conditions, haze, rain, and snow. As shown in Tab. 3, our method consistently surpasses other state-of-the-art approaches like AirNet Li et al. (2022), PromptIR Potlapalli et al. (2024), WGWSNet Zhu et al. (2023),

Table 1: *Comparison to state-of-the-art on three degradations.* PSNR (dB, ↑) and SSIM (↑) metrics are reported on the full RGB images. **Best** and **Second Best** performances is highlighted. Our method sets a new state-of-the-art on average across all benchmarks while being significantly more efficient than prior work. '-' represents unreported results.

Method	Venue.	Params.	*Dehazing*		*Deraining*		*Denoising*						Average	
			SOTS		Rain100L		$BSD68_{\sigma=15}$		$BSD68_{\sigma=25}$		$BSD68_{\sigma=50}$			
BRDNet Tian et al. (2020)	NN'20	-	23.23	.895	27.42	.895	32.26	.898	29.76	.836	26.34	.693	27.80	.843
LPNet Gao et al. (2019)	CVPR'19	-	20.84	.828	24.88	.784	26.47	.778	24.77	.748	21.26	.552	23.64	.738
FDGAN Dong et al. (2020)	AAAI'20	-	24.71	.929	29.89	.933	30.25	.910	28.81	.868	26.43	.776	28.02	.883
DL Fan et al. (2019)	TPAMI'19	2M	26.92	.931	32.62	.931	33.05	.914	30.41	.861	26.90	.740	29.98	.876
MPRNet Zamir et al. (2021)	CVPR'21	16M	25.28	.955	33.57	.954	33.54	.927	30.89	.880	27.56	.779	30.17	.899
AirNet Li et al. (2022)	CVPR'22	9M	27.94	.962	34.90	.967	33.92	.933	31.26	.888	28.00	.797	31.20	.910
NDR Yao et al. (2024)	TIP'24	28.4M	25.01	.860	28.62	.848	28.72	.826	27.88	.798	26.18	.720	25.01	.810
PromptIR Potlapalli et al. (2024)	NeurIPS'23	36M	30.58	.974	36.37	.972	33.98	**.933**	31.31	**.888**	28.06	.799	32.06	.913
MoCE-IR-S Zamfir et al. (2025)	CVPR'25	11M	30.98	**.979**	38.22	**.983**	34.08	**.933**	31.42	**.888**	28.16	.798	32.57	.916
AnyIR -T(*Ours*)	2025	6M	**31.70**	**.982**	**38.51**	**.983**	**34.12**	**.936**	**31.46**	**.893**	**28.20**	**.804**	**32.80**	**.919**
AnyIR -S(*Ours*)	2025	9M	**31.85**	**.982**	**38.56**	**.983**	**34.15**	**.936**	**31.49**	**.893**	**28.24**	**.806**	**32.86**	**.920**
Methods with the assistance of vision language, multi-task learning, natural language prompts, or multi-modal control														
DA-CLIP Luo et al. (2024)	ICLR'24	125M	29.46	.963	36.28	.968	30.02	.821	24.86	.585	22.29	.476	-	-
$Art_{PromptIR}$ Wu et al. (2024a)	ACM MM'24	36M	30.83	.979	37.94	.982	34.06	.934	31.42	.891	28.14	.801	32.49	.917
InstructIR-3D Conde et al. (2024)	ECCV'24	16M	30.22	.959	37.98	.978	34.15	.933	31.52	.890	28.30	.804	32.43	.913
UniProcessor Duan et al. (2025)	ECCV'24	1002M	31.66	.979	38.17	.982	34.08	.935	31.42	.891	28.17	.803	32.70	.918

Table 2: *Comparison to state-of-the-art on five degradations.* PSNR (dB, ↑) and SSIM (↑) metrics are reported on the full RGB images with (*) denoting general image restorers, others are specialized all-in-one approaches. **Best** and **Second Best** performances is highlighted.

Method	Venue	Params.	*Dehazing*		*Deraining*		*Denoising*		*Deblurring*		*Low-Light*		Average	
			SOTS		Rain100L		$BSD68_{\sigma=25}$		GoPro		LOLv1			
NAFNet* Chen et al. (2022)	ECCV'22	17M	25.23	.939	35.56	.967	31.02	.883	26.53	.808	20.49	.809	27.76	.881
DGUNet* Mou et al. (2022)	CVPR'22	17M	24.78	.940	36.62	.971	31.10	.883	27.25	.837	21.87	.823	28.32	.891
SwinIR* Liang et al. (2021)	ICCVW'21	1M	21.50	.891	30.78	.923	30.59	.868	24.52	.773	17.81	.723	25.04	.835
Restormer* Zamir et al. (2022)	CVPR'22	26M	24.09	.927	34.81	.962	31.49	.884	27.22	.829	20.41	.806	27.60	.881
MambaIR* Guo et al. (2024a)	ECCV'24	27M	25.81	.944	36.55	.971	31.41	.884	28.61	.875	22.49	.832	28.97	.901
DL Fan et al. (2019)	TPAMI'19	2M	20.54	.826	21.96	.762	23.09	.745	19.86	.672	19.83	.712	21.05	.743
TransWeather Valanarasu et al. (2022)	CVPR'22	38M	21.32	.885	29.43	.905	29.00	.841	25.12	.757	21.21	.792	25.22	.836
TAPE Liu et al. (2022)	ECCV'22	1M	22.16	.861	29.67	.904	30.18	.855	24.47	.763	18.97	.621	25.09	.801
AirNet Li et al. (2022)	CVPR'22	9M	21.04	.884	32.98	.951	30.91	.882	24.35	.781	18.18	.735	25.49	.847
IDR Zhang et al. (2023)	CVPR'23	15M	25.24	.943	35.63	.965	31.60	**.887**	27.87	.846	21.34	.826	28.34	.893
PromptIR Potlapalli et al. (2024)	NeurIPS'23	36M	30.41	.972	36.17	.970	31.20	.885	27.93	.851	**22.89**	.829	29.72	.901
MoCE-IR-S Zamfir et al. (2025)	CVPR'25	11M	31.33	.978	37.21	.978	31.25	.884	**28.90**	**.877**	21.68	.851	30.08	.913
AnyIR -T (*ours*)	2025	6M	**31.50**	**.981**	**38.81**	**.984**	**31.40**	**.892**	28.35	.863	22.68	**.854**	**30.71**	**.915**
AnyIR -S (*ours*)	2025	9M	**31.77**	**.982**	**39.00**	**.983**	**31.44**	**.892**	**28.52**	**.867**	**23.03**	**.857**	**30.75**	**.916**
Methods with the assistance of natural language prompts or multi-task learning														
InstructIR-5D Conde et al. (2024)	ECCV'24	16M	36.84	.973	27.10	.956	31.40	.887	29.40	.886	23.00	.836	29.55	.908
$Art_{PromptIR}$ Wu et al. (2024a)	ACM MM'24	36M	29.93	.908	22.09	.891	29.43	.843	25.61	.776	21.99	.811	25.81	.846

WeatherDiff Özdenizci & Legenstein (2023), OneRestore Guo et al. (2024b), and MoCE-IR Zamfir et al. (2025). These results demonstrate its robustness in addressing complex interactions among degradations. The superior performance of AnyIR validates its advanced degradation modeling and fusion mechanisms, enabling effective restoration in various scenarios.

Zero-Shot Unseen Degradation. To assess generalization beyond training degradations, we first evaluate the model trained on the 3-degradation setting directly on the unseen desnowing task (*i.e.*, zero-shot transfer) using the CSD dataset Chen et al. (2021a). As shown in Tab. 4, AnyIR extends effectively to this novel degradation without domain-specific tuning. We further conduct a real-world zero-shot evaluation on underwater images. AnyIR -S achieves 16.78 dB PSNR and 0.770 SSIM, outperforming the best prior method MoCE-IR by +0.87 dB while being more compact (Tab. 5). Notably, our model has never seen underwater data during training, underscoring its robustness to unseen degradation scenarios.

Table 3: *Comparison to state-of-the-art on composited degradations.* PSNR (dB, ↑) and SSIM (↑) are reported on the full RGB images. Our method consistently outperforms even larger models, with favorable results in composited degradation scenarios.

Method	Params.	*CDD11-Single*								*CDD11-Double*										*CDD11-Triple*				Avg.	
		Low (L)		Haze (H)		Rain (R)		Snow (S)		L+H		L+R		L+S		H+R		H+S		L+H+R		L+H+S			
AirNet	9M	24.83	.778	24.21	.951	26.55	.891	26.79	.919	23.23	.779	22.82	.710	23.29	.723	22.21	.868	23.29	.901	21.80	.708	22.24	.725	23.75	.814
PromptIR	36M	26.32	.805	26.10	.969	31.56	.946	31.53	.960	24.49	.789	25.05	.771	24.51	.761	24.54	.924	23.70	.925	23.74	.752	23.33	.747	25.90	.850
WGWSNet	26M	24.39	.774	27.90	.982	33.15	.964	34.43	.973	24.27	.800	25.06	.772	24.60	.765	27.23	.955	27.65	.960	23.90	.772	23.97	.771	26.96	.863
WeatherDiff	83M	23.58	.763	21.99	.904	24.85	.885	24.80	.888	21.83	.756	22.69	.730	22.12	.707	21.25	.868	21.99	.868	21.23	.716	21.04	.698	22.49	.799
OneRestore	6M	26.48	.826	32.52	.990	33.40	.964	34.31	.973	25.79	.822	25.58	.799	25.19	.789	29.99	.957	30.21	.964	24.78	.788	24.90	.791	28.47	.878
MoCE-IR	11M	27.26	.824	32.66	.990	34.31	.970	35.91	.980	26.24	.817	26.25	.800	26.04	.793	29.93	.964	30.19	.970	25.41	.789	25.39	.790	29.05	.881
AnyIR-T(*ours*)	6M	27.40	.833	33.41	.991	34.53	.970	36.07	.979	26.53	.827	26.55	.810	26.36	.802	30.40	.965	30.65	.970	25.66	.800	25.86	.801	29.40	.886
AnyIR-S(*ours*)	10M	**27.47**	**.835**	**34.38**	**.992**	**34.64**	**.971**	**36.21**	**.981**	**26.54**	**.830**	**26.49**	**.812**	**26.45**	**.805**	**30.98**	**.967**	**31.55**	**.972**	**25.82**	**.804**	**26.01**	**.805**	**29.69**	**.889**

Table 4: Unseen **Desnowing** on *CSD* Chen et al. (2021a) dataset.

Method	PSNR	SSIM
AirNet Li et al. (2022)	19.32	.733
PromptIR Potlapalli et al. (2024)	20.47	.764
MoceIR-S Zamfir et al. (2025)	21.09	.771
AnyIR (Ours)	**21.64**	**.787**

Table 5: *Zero-Shot* Cross-Domain Underwater Image Enhancement Results.

Method	PSNR (dB, ↑)	SSIM (↑)
SwinIR Liang et al. (2021)	15.31	.740
NAFNet Chu et al. (2022)	15.42	.744
Restormer Zamir et al. (2022)	15.46	.745
AirNet Li et al. (2022)	15.46	.745
IDR Zhang et al. (2023)	15.58	.762
PromptIR Potlapalli et al. (2024)	15.48	.748
MoCE-IR Zamfir et al. (2025)	15.91	.765
AnyIR -S *(Ours)*	**16.78**	**.770**

Table 6: *Complexity Analysis.* FLOPs are computed on image size 224×224 via a NVIDIA A100 (40G) GPU.

Method	PSNR (dB, ↑)	Memory (↓)	Params. (↓)	FLOPs (↓)
AirNet Li et al. (2022)	31.20	4829M	8.93M	238G
PromptIR Potlapalli et al. (2024)	32.06	9830M	35.59M	132G
IDR Zhang et al. (2023)	-	4905M	15.34M	98G
MoCE-IR Zamfir et al. (2025)	32.73	5887M	25.35M	80.59 ± 5.21G
MoCE-IR-S Zamfir et al. (2025)	32.57	**4228M**	11.47M	36.93 ± 2.32G
AnyIR -T(*ours*)	**32.80**	4969M	**5.74M**	**26 ± 1.98G**
AnyIR -S(*ours*)	**32.86**	6661M	**8.51M**	**39 ± 1.87G**

Model efficiency. Tab. 6 compares memory usage, FLOPs, and parameters across recent All-in-One IR methods. With our hybrid block design and the proposed degradation adaptation module, AnyIR achieves a **0.74 dB** PSNR gain over the baseline PromptIR Potlapalli et al. (2024), while reducing parameters by **83.9%**, and FLOPs to only **26G**, making it **80%** more computationally efficient. Compared with MoCE-IR-S, AnyIR lowers FLOPs by **29.73%** (26G vs. 37G) and parameters by **49.96%** (5.74M vs. 11.47M), while maintaining comparable or superior accuracy. Such reductions not only improve efficiency in terms of model size and computation, but also translate into a smaller energy and compute footprint during inference, which is particularly relevant for resource-constrained or edge deployment. This positions AnyIR as a strong and sustainable baseline for future All-in-One IR research.

Visual results. To complement the quantitative results, we visualize the results of our method in Fig. 5. The visualizations demonstrate the efficacy of AnyIR in dehazing, denoising, and draining, and we marked out the detailed region using the red boxes. In the dehazing task, AirNet Li et al. (2022), PromptIR Potlapalli et al. (2024), and MoCE-IR Zamfir et al. (2025) exhibit limitations in fully eliminating haziness, leading

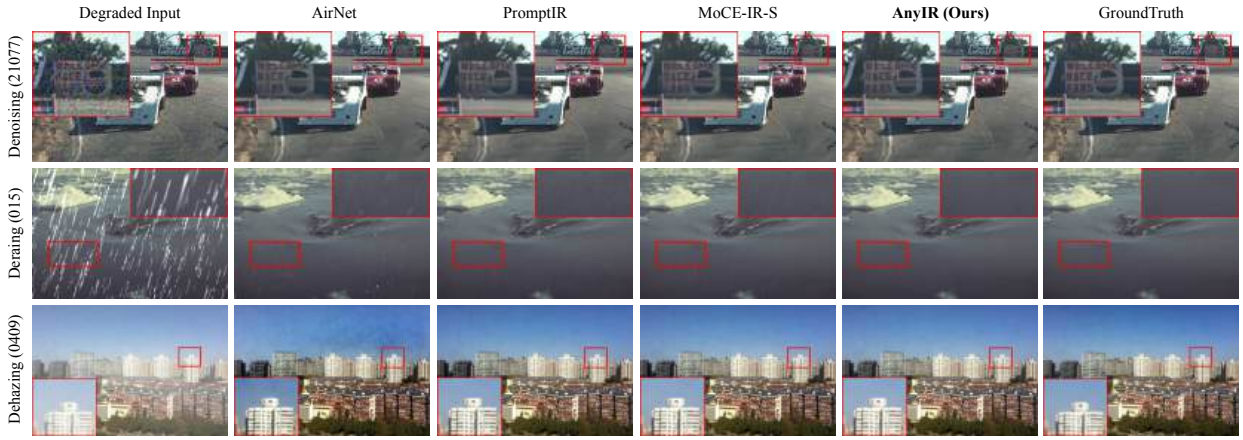


Figure 5: Visual comparison on three degradations. Zoom in for a better view.

Figure 6: Zero-Shot Underwater Image Enhancement Visual Results. Zoom in for a better view.

to noticeable color reconstruction discrepancies. In contrast, our AnyIR effectively enhances visibility and ensures a precise color reconstruction. The denoising results also show that our AnyIR can restore more detailed characters, demonstrating the rich texture edge recovery ability of our method. Meanwhile, in rainy scenes, previous methods continue to exhibit remnants of rain streaks. Please, **zoom in** for more details. In contrast, our approach excels at eliminating these artifacts and recovering underlying details, showing our superiority under adverse weather conditions. Note that the PromptIR is approximately $6\times$ larger than ours. Despite this, our method consistently produces visually superior results, demonstrating its effectiveness. Fig. 6 also shows that under the zero-shot setting, our method can also restore clear results. More detailed visual examples are given in the appendix.

4.2 Ablation Studies

Impact of different components. We conduct detailed studies on the proposed components within the framework of AnyIR . All experiments are conducted in the *All-in-One* setting with *three* degradations. We compare our simple feature-modeling block DAB against other variants. As detailed in Tab. 7, we assess the effectiveness of our key architectural contributions by removing or replacing our designed module with its counterparts. The detailed architecture overview of these considered counterparts can be found in Fig. 7.

We first examine the impact of our skip-split operation (a-b), which yields a significant improvement of **1.02 dB** over the common half-split method. This validates and supports our motivation to deeply enable intertwining within the subparts of an input feature. Introducing our proposed GatedDA in parallel to the attention layer (c) results in a substantial increase of 1.27 dB. Combining GatedDA with the skip-split operation (d) further enhances the reconstruction fidelity of our framework. This validates the effectiveness of our intention of using local gated details to reconstruct the degradation-aware output. Lastly, the introduction of cross-feature filtering (e), please refer to Fig. 2 for our plain design, improves the interconnectivity between global-local features, further benefiting overall performance.

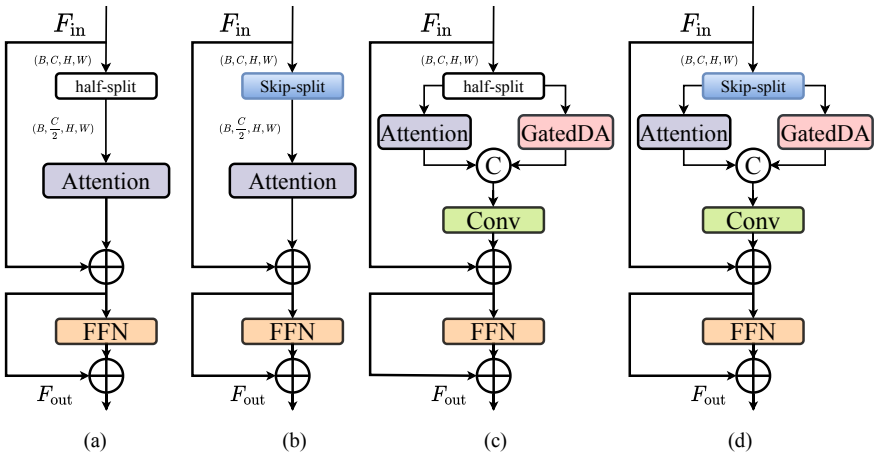


Figure 7: Structure of other different DAB variants during exploration.

Table 7: Ablation comparison (Average PSNR) of the effect of each component under the 3-degradation.

Method	Skip-Split	Fusion (Alg. 2)	GatedDA	PSNR (dB, ↑)
(a)	×	×	×	30.85
(b)	✓	×	×	31.83
(c)	×	×	✓	32.13
(d)	✓	×	✓	32.35
(e)	✓	✓	✓	**32.80**

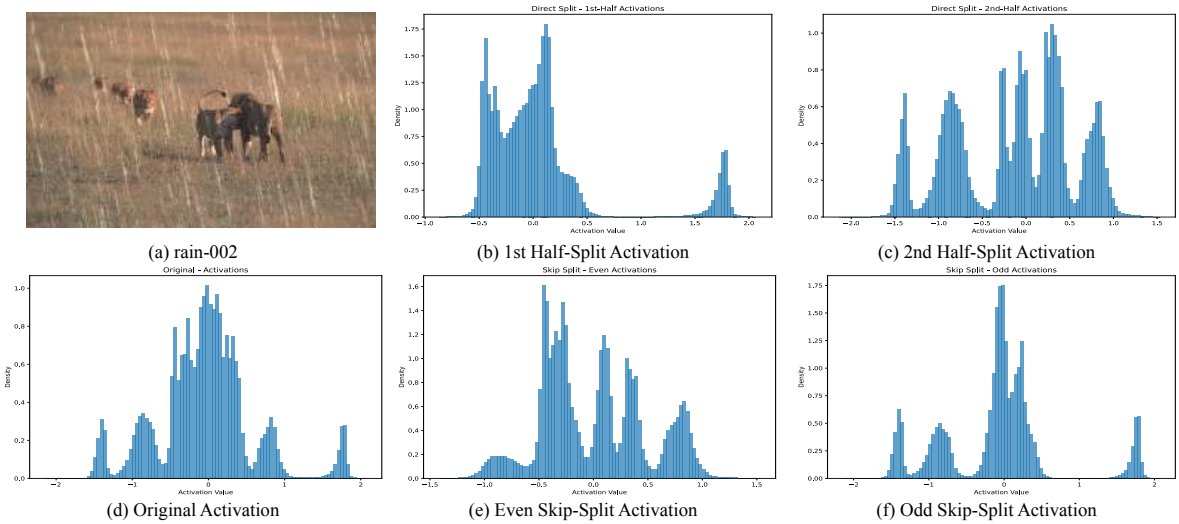
Table 8: Impact of different (α, β, γ) settings on three-degradation.

(α, β, γ)	PSNR ↑	SSIM ↑
(0, 1/2, 1/2)	32.14	0.910
(1/2, 0, 1/2)	32.21	0.912
(1/2, 1/2, 0)	31.37	0.907
(1/4, 1/4, 1/2)	**32.80**	**0.919**

Figure 8: Visual feature maps of α, β, and γ within GatedDA.

Impact of different α, β, and γ in gatedDA. Tab.8 shows the different impact of different values of (α, β, γ) in the gatedDA proposed (without Alg. 2). We noticed that when setting one of these parameters to 0, the performance decreases. When we set (α, β, γ) to ($^{hidden}/_4$, $^{hidden}/_4$, $^{hidden}/_2$), the average performance increases. This means that each part of the proposed gatedDA is necessary. The visual features shown in Fig. 8 indicate that α, β, and γ consistently focus on various aspects of the degraded regions, each specializing in different levels or types of degradation. See more analyses in our *Supp. Mat.*

Impact of the network capacity. The results in Tab. 9, derived from experiments on the draining task, provide valuable insight into the broader context of all-in-one IR. Although draining serves as a testbed, the findings reflect a key question in general-purpose restoration: *Does increasing model capacity universally*



(a) rain-002 (b) 1st Half-Split Activation (c) 2nd Half-Split Activation

(d) Original Activation (e) Even Skip-Split Activation (f) Odd Skip-Split Activation

Figure 9: Visualization of the channel activation distribution.

Table 9: *Network Capacity and Complexity Analysis on* ***Deraining****.* Comparison of PSNR, model parameters, and FLOPs, highlighting the balance between network capacity and computational efficiency. FLOPs are computed for 224×224 input images using an NVIDIA Tesla A100 (40G) GPU.

Method	Blocks	PSNR (dB, ↑)	Params. (↓)	FLOPs (↓)
PromptIR Potlapalli et al. (2024)(Base)	[4,6,6,8]	37.04	35.6M	132G
Base (No Prompt)	[4,6,6,8]	36.74	32.2M	117G
Base (Fix-Prompt)	[4,6,6,8]	36.85	-	-
Base (No Prompt)	[3,5,5,7]	36.84	29.1M	101G
Base (No Prompt)	[2,4,4,6]	36.76	26.0M	85G
Base (No Prompt)	[1,3,3,5]	36.63	22.9M	68G
AnyIR (*ours*)	[3,5,5,7]	**37.99**	**5.74M**	**26G**

lead to better performance? Our analysis suggests otherwise, emphasizing the importance of efficiency and task-aware design over brute-force scaling.

For example, the base PromptIR Potlapalli et al. (2024) model with blocks [4,6,6,8] achieves a PSNR of 37.04 dB through 35.6M parameters and 132G FLOP, and progressively reducing capacity through configurations such as [3,5,5,7], [2,4,4,6] and [1,3,3,5] results in smaller models with lower computational costs but marginal decreases in PSNR. This highlights the diminishing returns of simply reducing complexity without optimization, as smaller models lose their capacity to capture the nuances of degradation factors. In contrast, our AnyIR validates that a careful balance between network capacity and architectural innovation can achieve SOTA results. With a similar block configuration ([3,5,5,7]) but a significantly optimized architecture, AnyIR achieves a PSNR of 37.99 dB on the deraining task, outperforming larger models while using only 5.74M parameters and 26G FLOPs. This demonstrates the effectiveness of our task-specific improvements and highlights the potential to achieve better results with smaller, more efficient models. These findings underscore a critical insight for all-in-one IR: *While larger models may generalize better across diverse tasks, efficiency and tailored design can lead to both higher performance and practical utility.* The deraining results provide a compelling argument for rethinking model capacity in restoration frameworks, emphasizing that "more" is not always better, especially when thoughtful design can yield both performance gains and computational efficiency.

Exploration of λ value in Spatial-Frequency Fusion. We explore different λ settings in the fusion module: spatial-only, frequency-only, fixed, and learnable. As shown in Tab. 10, the learnable strategy yields the best results, highlighting the benefit of adaptive spatial–frequency balancing. Moreover, both fixed and learnable schemes surpass single-branch counterparts, confirming the complementarity of the two domains.

Table 10: *Ablation comparison* of different fusion strategies.

Fusion	λ	PSNR (dB, ↑)	SSIM (↑)
Spatial-only	1.0	32.37	.915
Frequency-only	0.0	32.09	.919
Fixed (0.5)	0.5	32.63	.917
Learnable	—	**32.80**	**.919**

5 Discussion

What does the proposed Skip-Split bring? Adjacent channels often contain redundant information due to spatial correlations in the data. Directly splitting the channels into two contiguous halves can lead to uneven feature distribution, with one half potentially capturing redundant features, while the other lacks important information. By using a simple skip-split method that interleaves channels between the two processing paths, we ensure a more balanced and diverse set of features in each path, enhancing the effectiveness of both the self-attention and convolutional components. These phenomena are also validated in the channel activation distribution visualization shown in Fig. 9.

Why does the proposed GatedDA & the fusion Alg. 2 work? As shown in Fig. 10, the error map visualizations reveal that degradation in an input image is often unevenly distributed, manifesting itself as global patterns and localized clusters. This highlights the need for degradation modeling to capture both widespread and fine-grained distortions. The proposed GatedDA module addresses this by selectively activating in degraded regions, closely aligning with the actual degradation distribution, and effectively enhancing localized features. When combined with global attention, which captures broader contextual dependencies, the fusion algorithm (Alg. 2) enables a more comprehensive understanding of the degradation structure of the image.

Further evidence is provided by the SVD and cumulative variance curves in Fig. 11, which demonstrate the complementary nature of the two modules. While the attention branch captures dominant global variations reflected in a steep variance accumulation and concentrated singular values, GatedDA captures more diverse and spatially distributed local signals. The fused representation achieves a better balance, integrating both local and global characteristics with faster variance accumulation than GatedDA alone. These results validate that the synergy between GatedDA and attention not only improves restoration quality but also enhances robustness across diverse types of degradation. Although some aspects of this analysis can be interpreted at a more abstract representation-learning level, our formulation and conclusions are confined to the image restoration setting, where the model is developed and evaluated specifically for multi-degradation IR.

Why is AnyIR efficient? AnyIR reduces computational costs by splitting input channels: one half is processed by self-attention, the other by a Gated CNN block. This division reduces the complexity of self-attention from $O(B \cdot \text{head} \cdot (H \cdot W)^2)$, while Gated CNN processes the remaining channels with a lower complexity of $O(B \cdot C \cdot H \cdot W)$, especially beneficial for high-resolution inputs where $(H \cdot W)^2$ dominates. Furthermore, AnyIR employs dimensionality reduction and parameter-efficient designs, collectively reducing the GFLOPs and model parameters. This balance of global context modeling and efficient local feature extraction enables AnyIR to minimize computational costs without sacrificing performance.

Is scaling down the new advantage? Although recent methods in image restoration often scale up model size and complexity, our AnyIR takes a different path: scaling down. Instead of relying on large architectures, we embrace a simple but non-trivial design, using targeted components like the GatedDA module to capture degradation without excessive parameters effectively. This simplicity is not a compromise; it is an asset - yielding high-quality restoration with minimal computational demand, faster training, and greater adaptability. The All-in-One IR framework, though new compared to degradation-specific methods, faces a key limitation: an imbalance in data distribution, with certain degradations (*e.g.* dehazing) dominating. Our experiments suggest that a more balanced dataset could significantly improve performance across degradation types, offering a constructive direction for future work. We note that these discussions and observations are made strictly within the scope of multi-degradation image restoration, and are not intended to imply a general-purpose image processing or prediction framework beyond the IR domain.

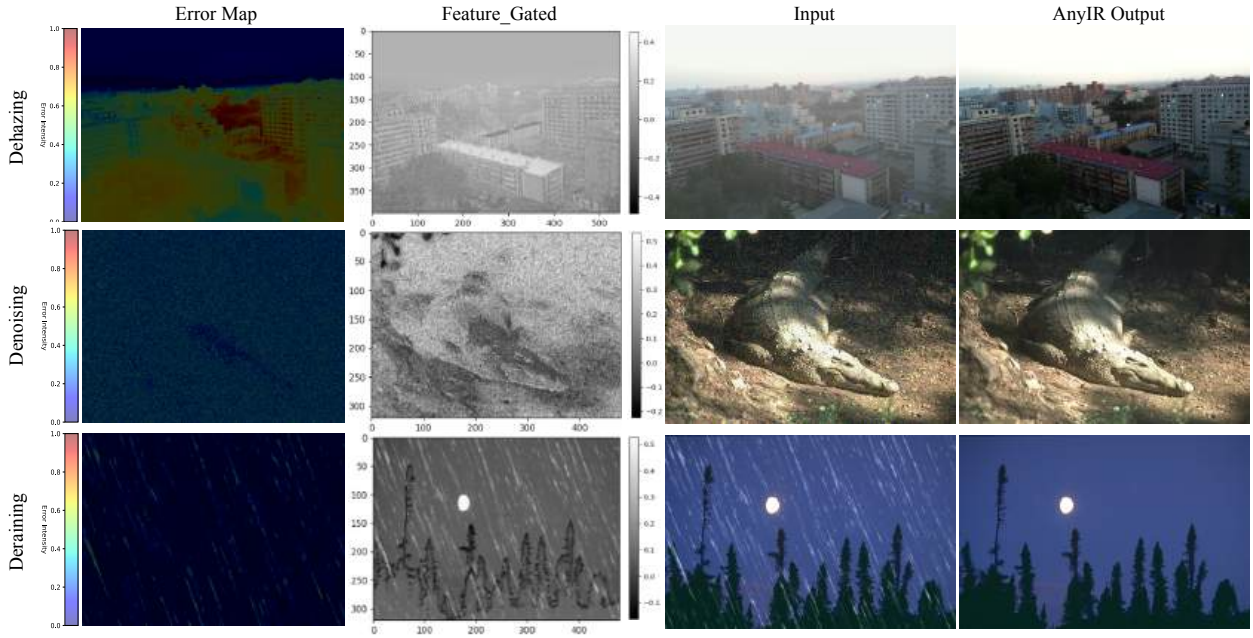


Figure 10: Error map and the output of GatedDA.

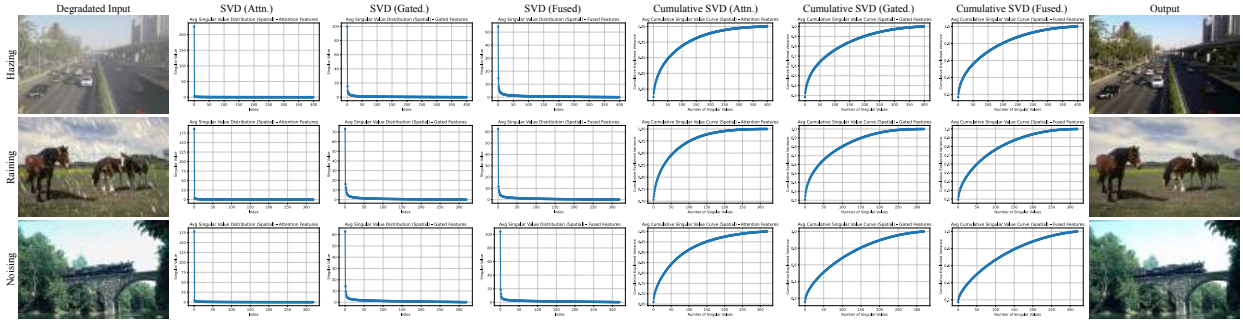


Figure 11: SVD and cumulative variance curves for attention, GatedDA, and the fused feature (Zoom in for a better view).

Effect of degradation distribution. We further observe that commonly used multi-degradation IR training sets (3 Degradation) tend to exhibit an imbalanced composition across degradation types. By making the distribution more uniform (Fig. 14) under the same 3-degradation setting—without increasing the total number of samples—we consistently obtain improved restoration performance (Tab. 12). This trend suggests that the learned reconstruction prior is influenced by the training degradation distribution, and a more balanced composition can lead to more stable and generalized restoration behavior.

6 Conclusion

We introduced AnyIR, an efficient multi-degradation image restoration model that unifies the degradations considered in current All-in-One IR settings within a single framework. By combining gated guidance with a spatial–frequency fusion strategy, the model learns embeddings that capture both degradation-specific cues and invariant structures, enabling robust restoration across tasks. Extensive experiments, including evaluations on unseen and real-world degradations, show that AnyIR achieves state-of-the-art accuracy with substantially reduced parameters and FLOPs, while maintaining strong generalization ability. We believe that AnyIR provides a strong and efficient baseline for future research, advancing both the practical deployment and the broader understanding of learning-based all-in-one image restoration.

Acknowledgments

This work was partially supported by the FIS project GUIDANCE (Debugging Computer Vision Models via Controlled Cross-modal Generation) (No. FIS2023-03251), the Alexander von Humboldt Foundation, and the National Natural Science Foundation of China (62403345).

References

Pablo Arbelaez, Michael Maire, Charless Fowlkes, and Jitendra Malik. Contour detection and hierarchical image segmentation. *IEEE TPAMI*, 33(5):898–916, 2010. 22

Mark R Banham and Aggelos K Katsaggelos. Digital image restoration. *IEEE Signal Processing Magazine*, 14(2):24–41, 1997. 2

Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *ECCV*, pp. 17–33, 2022. 2, 10

Wei-Ting Chen, Hao-Yu Fang, Cheng-Lin Hsieh, Cheng-Che Tsai, I Chen, Jian-Jiun Ding, Sy-Yen Kuo, et al. All snow removed: Single image desnowing algorithm using hierarchical dual-tree complex wavelet representation and contradict channel loss. In *ICCV*, pp. 4196–4205, 2021a. 10, 11

Yinbo Chen, Sifei Liu, and Xiaolong Wang. Learning continuous image representation with local implicit image function. In *CVPR*, pp. 8628–8638, 2021b. 2

Xiaojie Chu, Liangyu Chen, and Wenqing Yu. Nafssr: Stereo image super-resolution using nafnet. In *CVPR*, pp. 1239–1248, June 2022. 11

Marcos V Conde, Gregor Geigle, and Radu Timofte. Instructir: High-quality image restoration following human instructions. In *ECCV*, 2024. 2, 10

Yuning Cui, Syed Waqas Zamir, Salman Khan, Alois Knoll, Mubarak Shah, and Fahad Shahbaz Khan. Adair: Adaptive all-in-one image restoration via frequency mining and modulation. *ICLR*, 2025. 3, 4

Tri Dao and Albert Gu. Transformers are SSMs: Generalized models and efficient algorithms through structured state space duality. In *ICML*, 2024. 3

Xiaohuan Ding, Yangrui Gong, Tianyi Shi, Zihang Huang, Gangwei Xu, and Xin Yang. Masked snake attention for fundus image restoration with vessel preservation. In *ACM MM*, pp. 4368–4376, 2024. 1

Chao Dong, Yubin Deng, Chen Change Loy, and Xiaoou Tang. Compression artifacts reduction by a deep convolutional network. In *ICCV*, pp. 576–584, 2015. 2

Yu Dong, Yihao Liu, He Zhang, Shifeng Chen, and Yu Qiao. Fd-gan: Generative adversarial networks with fusion-discriminator for single image dehazing. In *AAAI*, 2020. 10

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. In *ICLR*, 2020. 3, 6

Huiyu Duan, Xiongkuo Min, Sijing Wu, Wei Shen, and Guangtao Zhai. Uniprocessor: a text-induced unified low-level image processor. In *ECCV*, pp. 180–199. Springer, 2025. 10

Akshay Dudhane, Omkar Thawakar, Syed Waqas Zamir, Salman Khan, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Dynamic pre-training: Towards efficient and scalable all-in-one image restoration. *arXiv preprint arXiv:2404.02154*, 2024. 3

Qingnan Fan, Dongdong Chen, Lu Yuan, Gang Hua, Nenghai Yu, and Baoquan Chen. A general decoupled learning framework for parameterized image operators. *TPAMI*, 43(1):33–47, 2019. 10

Hongyun Gao, Xin Tao, Xiaoyong Shen, and Jiaya Jia. Dynamic scene deblurring with parameter selective sharing and nested skip connections. In *CVPR*, 2019. 10

Sicheng Gao, Xuhui Liu, Bohan Zeng, Sheng Xu, Yanjing Li, Xiaoyan Luo, Jianzhuang Liu, Xiantong Zhen, and Baochang Zhang. Implicit diffusion models for continuous super-resolution. In *CVPR*, pp. 10021–10030, 2023. 2

Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. In *Language Modeling*, 2024. 3

Hang Guo, Jinmin Li, Tao Dai, Zhihao Ouyang, Xudong Ren, and Shu-Tao Xia. Mambair: A simple baseline for image restoration with state-space model. In *ECCV*, 2024a. 3, 10

Yu Guo, Yuan Gao, Yuxu Lu, Ryan Wen Liu, and Shengfeng He. Onerestore: A universal restoration framework for composite degradation. In *ECCV*, 2024b. 9, 10, 22

Dan Hendrycks and Kevin Gimpel. Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*, 2016. 6, 7

Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *CVPR*, pp. 5197–5206, 2015. 22

Hu JiaKui, Zhengjian Yao, Jin Lujia, and Lu Yanye. Universal image restoration pre-training via degradation classification. In *ICLR*, 2025. 22

Junjun Jiang, Zengyuan Zuo, Gang Wu, Kui Jiang, and Xianming Liu. A survey on all-in-one image restoration: Taxonomy, evaluation and future trends. *TPAMI*, 2025. 3, 4

Kui Jiang, Zhongyuan Wang, Peng Yi, Chen Chen, Baojin Huang, Yimin Luo, Jiayi Ma, and Junjun Jiang. Multi-scale progressive fusion network for single image deraining. In *CVPR*, pp. 8346–8355, 2020. 3

Yitong Jiang, Zhaoyang Zhang, Tianfan Xue, and Jinwei Gu. Autodir: Automatic all-in-one image restoration with latent diffusion. *arXiv preprint arXiv:2310.10123*, 2023. 2

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *ICLR*, 2015. 22

Lingshun Kong, Jiangxin Dong, Jianjun Ge, Mingqiang Li, and Jinshan Pan. Efficient frequency domain-based transformers for high-quality image deblurring. In *CVPR*, pp. 5886–5895, 2023. 3

Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. Deep laplacian pyramid networks for fast and accurate super-resolution. In *CVPR*, pp. 624–632, 2017. 2

Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE TIP*, 28(1):492–505, 2018. 22

Boyun Li, Xiao Liu, Peng Hu, Zhongqin Wu, Jiancheng Lv, and Xi Peng. All-in-one image restoration for unknown corruption. In *CVPR*, pp. 17452–17462, 2022. 2, 3, 9, 10, 11, 22

Chongyi Li, Chunle Guo, Wenqi Ren, Runmin Cong, Junhui Hou, Sam Kwong, and Dacheng Tao. An underwater image enhancement benchmark dataset and beyond. *TIP*, 29:4376–4389, 2019. 22

Ruoteng Li, Robby T Tan, and Loong-Fah Cheong. All in one bad weather removal using architectural search. In *CVPR*, pp. 3175–3185, 2020. 4

Yawei Li, Yuchen Fan, Xiaoyu Xiang, Denis Demandolx, Rakesh Ranjan, Radu Timofte, and Luc Van Gool. Efficient and explicit modelling of image hierarchies for image restoration. In *CVPR*, pp. 18278–18289, 2023a. 1, 2, 3, 4, 6

Yawei Li, Kai Zhang, Jingyun Liang, Jiezhang Cao, Ce Liu, Rui Gong, Yulun Zhang, Hao Tang, Yun Liu, Denis Demandolx, Rakesh Ranjan, Radu Timofte, and Luc Van Gool. LSDIR: A large scale dataset for image restoration. In *CVPRW*, pp. 1775–1787, 2023b. 2

Yawei Li, Bin Ren, Jingyun Liang, Rakesh Ranjan, Mengyuan Liu, Nicu Sebe, Ming-Hsuan Yang, and Luca Benini. Fractal-ir: A unified framework for efficient and scalable image restoration. *arXiv preprint arXiv:2503.17825*, 2025. 1

Zilong Li, Yiming Lei, Chenglong Ma, Junping Zhang, and Hongming Shan. Prompt-in-prompt learning for universal image restoration. *arXiv preprint arXiv:2312.05038*, 2023c. 2, 3, 4

Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. SwinIR: Image restoration using Swin transformer. In *ICCVW*, pp. 1833–1844, 2021. 2, 3, 4, 6, 10, 11

Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *CVPRW*, pp. 1132–1140, 2017. 2

Chang Liu, Mengyi Zhao, Bin Ren, Mengyuan Liu, Nicu Sebe, et al. Spatio-temporal graph diffusion for text-driven human motion generation. In *BMVC*, pp. 722–729, 2023. 2

Lin Liu, Lingxi Xie, Xiaopeng Zhang, Shanxin Yuan, Xiangyu Chen, Wengang Zhou, Houqiang Li, and Qi Tian. Tape: Task-agnostic prior embedding for image restoration. In *ECCV*, 2022. 2, 10

Ziwei Luo, Fredrik K Gustafsson, Zheng Zhao, Jens Sjölund, and Thomas B Schön. Image restoration with mean-reverting stochastic differential equations. *arXiv preprint arXiv:2301.11699*, 2023. 2

Ziwei Luo, Fredrik K Gustafsson, Zheng Zhao, Jens Sjölund, and Thomas B Schön. Controlling vision-language models for universal image restoration. In *ICLR*, 2024. 2, 10

Kede Ma, Zhengfang Duanmu, Qingbo Wu, Zhou Wang, Hongwei Yong, Hongliang Li, and Lei Zhang. Waterloo exploration database: New challenges for image quality assessment models. *IEEE TIP*, 26(2): 1004–1016, 2016. 22

David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *ICCV*, pp. 416–423, 2001. 22

Yang Miao, Francis Engelmann, Olga Vysotska, Federico Tombari, Marc Pollefeys, and Dániel Béla Baráth. Scenegraphloc: Cross-modal coarse visual localization on 3d scene graphs. In *ECCV*, pp. 127–150. Springer, 2024. 2

Chong Mou, Qian Wang, and Jian Zhang. Deep generalized unfolding networks for image restoration. In *CVPR*, pp. 17399–17410, 2022. 10

Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *CVPR*, pp. 3883–3891, 2017. 22

Ozan Özdenizci and Robert Legenstein. Restoring vision in adverse weather conditions with patch-based denoising diffusion models. *TPAMI*, 45(8):10346–10357, 2023. 10

Vaishnav Potlapalli, Syed Waqas Zamir, Salman H Khan, and Fahad Shahbaz Khan. Promptir: Prompting for all-in-one image restoration. *NeurIPS*, 36, 2024. 2, 3, 6, 8, 9, 10, 11, 14, 22, 24, 26

Bin Ren, Yahui Liu, Yue Song, Wei Bi, Rita Cucchiara, Nicu Sebe, and Wei Wang. Masked jigsaw puzzle: A versatile position embedding for vision transformers. In *CVPR*, pp. 20382–20391, 2023a. 3

Bin Ren, Yawei Li, Jingyun Liang, Rakesh Ranjan, Mengyuan Liu, Rita Cucchiara, Luc V Gool, Ming-Hsuan Yang, and Nicu Sebe. Sharing key semantics in transformer makes efficient image restoration. *NeurIPS*, 37: 7427–7463, 2024a. 1, 2

Bin Ren, Yawei Li, Nancy Mehta, Radu Timofte, Hongyuan Yu, Cheng Wan, Yuxin Hong, Bingnan Han, Zhuoyuan Wu, Yajun Zou, et al. The ninth ntire 2024 efficient super-resolution challenge report. In *CVPR*, pp. 6595–6631, 2024b. 1

Dongwei Ren, Wangmeng Zuo, Qinghua Hu, Pengfei Zhu, and Deyu Meng. Progressive image deraining networks: A better and simpler baseline. In *CVPR*, pp. 3937–3946, 2019. 3

Mengwei Ren, Mauricio Delbracio, Hossein Talebi, Guido Gerig, and Peyman Milanfar. Multiscale structure guided diffusion for image deblurring. In *ICCV*, pp. 10721–10733, 2023b. 2, 3

Wenqi Ren, Jinshan Pan, Hua Zhang, Xiaochun Cao, and Ming-Hsuan Yang. Single image dehazing via multi-scale convolutional neural networks with holistic edges. *IJCV*, 128:240–259, 2020. 3

William Hadley Richardson. Bayesian-based iterative method of image restoration. *Journal of the Optical Society of America*, 62(1):55–59, 1972. 2

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *MICCAI*, pp. 234–241, 2015. 5

Xiaole Tang, Xiang Gu, Xiaoyi He, Xin Hu, and Jian Sun. Degradation-aware residual-conditioned optimal transport for unified image restoration. *TPAMI*, 2025. 3, 4

Chunwei Tian, Yong Xu, and Wangmeng Zuo. Image denoising using deep cnn with batch renormalization. *Neural Networks*, 2020. 10

Zhengzhong Tu, Hossein Talebi, Han Zhang, Feng Yang, Peyman Milanfar, Alan Bovik, and Yinxiao Li. MAXIM: Multi-axis mlp for image processing. In *CVPR*, pp. 5769–5780, 2022. 3

Jeya Maria Jose Valanarasu, Rajeev Yasarla, and Vishal M Patel. TransWeather: Transformer-based restoration of images degraded by adverse weather conditions. In *CVPR*, pp. 2353–2363, 2022. 10

Cong Wang, Jinshan Pan, Wei Wang, Jiangxin Dong, Mengzhu Wang, Yakun Ju, and Junyang Chen. Promptrestorer: A prompting image restoration method with degradation perception. *NeurIPS*, 36: 8898–8912, 2023a. 2, 3

Xintao Wang, Ke Yu, Chao Dong, and Chen Change Loy. Recovering realistic texture in image super-resolution by deep spatial feature transform. In *CVPR*, pp. 606–615, 2018. 3

Yinhuai Wang, Jiwen Yu, and Jian Zhang. Zero-shot image restoration using denoising diffusion null-space model. *ICLR*, 2023b. 2

Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *CVPR*, pp. 17683–17693, 2022. 2

Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018. 22

Gang Wu, Junjun Jiang, Kui Jiang, and Xianming Liu. Harmony in diversity: Improving all-in-one image restoration via multi-task collaboration. In *ACM MM*, pp. 6015–6023, 2024a. 10

Haiyan Wu, Yanyun Qu, Shaohui Lin, Jian Zhou, Ruizhi Qiao, Zhizhong Zhang, Yuan Xie, and Lizhuang Ma. Contrastive learning for compact single image dehazing. In *CVPR*, pp. 10551–10560, 2021. 3

Hongjie Wu, Linchao He, Mingqin Zhang, Dongdong Chen, Kunming Luo, Mengting Luo, Ji-Zhe Zhou, Hu Chen, and Jiancheng Lv. Diffusion posterior proximal sampling for image restoration. In *ACM MM*, pp. 214–223, 2024b. 1

Zhijian Wu, Jun Li, Yang Hu, and Dingjiang Huang. Compacter: A lightweight transformer for image restoration. In *ACM MM*, pp. 3094–3103, 2024c. 1

Fuzhi Yang, Huan Yang, Jianlong Fu, Hongtao Lu, and Baining Guo. Learning texture transformer network for image super-resolution. In *CVPR*, pp. 5791–5800, 2020. 22

Mingde Yao, Ruikang Xu, Yuanshen Guan, Jie Huang, and Zhiwei Xiong. Neural degradation representation learning for all-in-one image restoration. *IEEE TIP*, 2024. 10

Zhengwei Yin, Guixu Lin, Mengshun Hu, Hao Zhang, and Yinqiang Zheng. Flexir: Towards flexible and manipulable image restoration. In *ACM MM*, pp. 6143–6152, 2024. 1

Zongsheng Yue, Jianyi Wang, and Chen Change Loy. ResShift: Efficient diffusion model for image super-resolution by residual shifting. *arXiv preprint arXiv:2307.12348*, 2023. 2

Eduard Zamfir, Zongwei Wu, Nancy Mehta, Yulun Zhang, and Radu Timofte. See more details: Efficient image super-resolution by experts mining. In *ICML*. PMLR, 2024. 2

Eduard Zamfir, Zongwei Wu, Nancy Mehta, Yuedong Tan, Danda Pani Paudel, Yulun Zhang, and Radu Timofte. Complexity experts are task-discriminative learners for any image restoration. In *CVPR*, 2025. 3, 8, 9, 10, 11, 22

Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *CVPR*, pp. 14821–14831, 2021. 2, 10

Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *CVPR*, pp. 5728–5739, 2022. 2, 3, 6, 10, 11

Jinghao Zhang, Jie Huang, Mingde Yao, Zizheng Yang, Hu Yu, Man Zhou, and Feng Zhao. Ingredient-oriented multi-degradation learning for image restoration. In *CVPR*, pp. 5825–5835, 2023. 2, 3, 9, 10, 11

Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE TIP*, 26(7):3142–3155, 2017a. 3

Kai Zhang, Wangmeng Zuo, Shuhang Gu, and Lei Zhang. Learning deep cnn denoiser prior for image restoration. In *CVPR*, pp. 3929–3938, 2017b. 3

Leheng Zhang, Yawei Li, Xingyu Zhou, Xiaorui Zhao, and Shuhang Gu. Transcending the limit of local window: Advanced super-resolution transformer with adaptive token dictionary. *arXiv preprint arXiv:2401.08209*, 2024. 2

Xu Zhang, Jiaqi Ma, Guoli Wang, Qian Zhang, Huan Zhang, and Lefei Zhang. Perceive-ir: Learning to perceive degradation better for all-in-one image restoration. *TIP*, 2025. 4

Yulun Zhang, Kunpeng Li, Kai Li, Bineng Zhong, and Yun Fu. Residual non-local attention networks for image restoration. *arXiv preprint arXiv:1903.10082*, 2019. 3

Mengyi Zhao, Mengyuan Liu, Bin Ren, Shuling Dai, and Nicu Sebe. Denoising diffusion probabilistic models for action-conditioned 3d motion generation. In *ICASSP*, pp. 4225–4229, 2024. 2

Haiyang Zheng, Nan Pu, Wenjing Li, Nicu Sebe, and Zhun Zhong. Textual knowledge matters: Cross-modality co-teaching for generalized visual class discovery. In *ECCV*, pp. 41–58. Springer, 2024a. 2

Haiyang Zheng, Ruilin Zhang, and Hongpeng Wang. Deep image clustering based on curriculum learning and density information. In *ICMR*, pp. 330–338, 2024b. 1

Lianghui Zhu, Bencheng Liao, Qian Zhang, Xinlong Wang, Wenyu Liu, and Xinggang Wang. Vision mamba: Efficient visual representation learning with bidirectional state space model. *arXiv preprint arXiv:2401.09417*, 2024. 3

Yurui Zhu, Tianyu Wang, Xueyang Fu, Xuanyu Yang, Xin Guo, Jifeng Dai, Yu Qiao, and Xiaowei Hu. Learning weather-general and weather-specific features for image restoration under multiple adverse weather conditions. In *CVPR*, pp. 21747–21758, 2023. 9

A Experimental Protocols

A.1 Datasets

3 Degradation Datasets. For both the All-in-One and single-task settings, we follow the evaluation protocols established in prior works Li et al. (2022); Potlapalli et al. (2024); Zamfir et al. (2025), utilizing the following datasets: For image denoising in the single-task setting, we combine the BSD400 Arbelaez et al. (2010) and WED Ma et al. (2016) datasets, and corrupt the images with Gaussian noise at levels $\sigma \in \{15, 25, 50\}$. BSD400 contains 400 training images, while WED includes 4,744 images. We evaluate the denoising performance on BSD68 Martin et al. (2001) and Urban100 Huang et al. (2015). For single-task deraining, we use Rain100L Yang et al. (2020), which provides 200 clean/rainy image pairs for training and 100 pairs for testing. For single-task dehazing, we adopt the SOTS dataset Li et al. (2018), consisting of 72,135 training images and 500 testing images. Under the All-in-One setting, we train a unified model on the combined set of the aforementioned training datasets for 120 epochs and directly test it across all three restoration tasks.

5 Degradation Datasets. The 5-degradation setting is built upon the 3-degradation setting, with two additional tasks included: deblurring and low-light enhancement. For deblurring, we adopt the GoPro dataset Nah et al. (2017), which contains 2,103 training images and 1,111 testing images. For low-light enhancement, we use the LOL-v1 dataset Wei et al. (2018), consisting of 485 training images and 15 testing images. Note that for the denoising task under the 5-degradation setting, we report results using Gaussian noise with $\sigma = 25$. The training takes 130 epochs.

Composited Degradation Datasets. Regarding the composite degradation setting, we use the CDD11 dataset Guo et al. (2024b). CDD11 consists of 1,183 training images for: *(i) 4 kinds of single-degradation types:* haze (H), low-light (L), rain (R), and snow (S); *(ii) 5 kinds of double-degradation types:* low-light + haze (l+h), low-light+rain (L+R), low-light + snow (L+S), haze + rain (H+R), and haze + snow (H+S). *(iii) 2 kinds of Triple-degradation type:* low-light + haze + rain (L+H+R), and low-light + haze + snow (L+H+S). We train our method for 150 epochs (significantly fewer than the 200 epochs used in MoCE-IR Zamfir et al. (2025)), and we keep all other settings unchanged.

Zero-Shot Underwater Image Enhancement Dataset. For the zero-shot underwater image enhancement setting, we follow the evaluation protocol of DCPT JiaKui et al. (2025) by directly applying our model, trained under the 5-degradation setting, on the UIEB dataset Li et al. (2019) without any finetuning. UIEB consists of two subsets: 890 raw underwater images with corresponding high-quality reference images, and 60 challenging underwater images. We evaluate our zero-shot performance on the 890-image subset with available reference images.

A.2 Implementation Details

Our AnyIR framework is designed to be end-to-end trainable, eliminating the need for multi-stage optimization of individual components. The architecture features a robust 4-level encoder-decoder structure, characterized by varying numbers of Degradation Adaptation Blocks (DAB) at each level, specifically $[3, 5, 5, 7]$ from highest to lowest level. Following established practices Potlapalli et al. (2024); Zamfir et al. (2025), we conducted training over 130 epochs via a batch size of 32 for the All-in-One and mixed settings. Optimization employed the L_1 loss and Fourier with the Adam optimizer Kingma & Ba (2015) (initial learning rate of 2×10^{-4}, $\beta_1 = 0.9$, $\beta_2 = 0.999$) and cosine decay schedule. During training, we employed random crops of size 128^2 and applied horizontal and vertical flips as augmentations. All experiments were performed using 2 NVIDIA Tesla A100 (40G) GPUs.

We also propose two scaled variants of our AnyIR , namely Tiny (AnyIR -T) and Small (AnyIR -S). As detailed in Tab. 11, these variants differ in terms of the number of Degradation Adaption Blocks (DAB) across scales, the input embedding dimension, the FFN expansion factor, and the number of refinement blocks.

Table 11: The details of the tiny and small versions of the proposed AnyIR .

	AnyIR -Tiny	AnyIR -Small
The Number of the DAB crosses 4 scales	[3, 5, 5, 7]	[4, 6, 6, 8]
The Input Embedding Dimension	28	32
The FFN Expansion Factor	2	2
The Number of the Refinement Blocks	4	4
Params. (↓)	5.74M	8.51M
FLOPs (↓)	26G	39 G

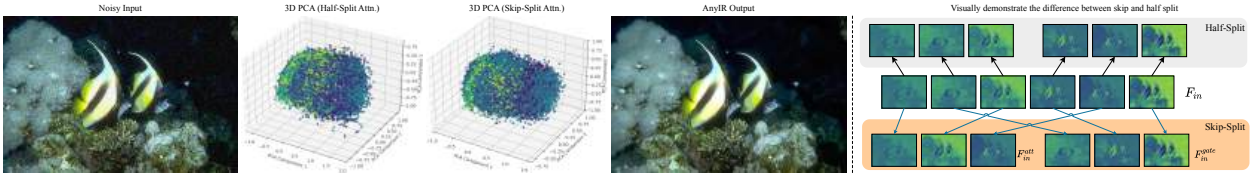


Figure 12: 3D PCA Visualization between the Half-Split and our proposed Skip-Split, and the visual illustration of Skip-Split.

B Discussion And Analysis

The visual illustration of Skip-Split and what it brings. Fig. 12 presents a 3D PCA visualization comparing Half-Split and our proposed Skip-Split. The results show that Skip-Split yields a more uniform and well-spread feature distribution, suggesting stronger and more discriminative representations. For a more intuitive understanding, a side-by-side visual comparison is also provided in the right part of Fig. 12.

What does GatedDA bring that differs from attention? As the visualization result shown in Fig. 13, the t-SNE visualizations of the feature maps reveal a key distinction between GatedDA and attention mechanisms. While attention effectively captures global relationships and context, its feature representations tend to exhibit less structural separation in the embedding space. In contrast, GatedDA introduces a localized focus on degradation-specific regions, resulting in a more distinct clustering of features in the t-SNE space. This separation highlights GatedDA's ability to emphasize degradation-specific information, complementing the broader scope of attention. Together, this synergy allows GatedDA to enhance image restoration by targeting specific degradation patterns while leveraging the global context provided by attention. As evidenced in the restored outputs, this combination leads to a more accurate recovery of both global and local details.

Attention and GatedDA Feature Maps. To investigate the effectiveness of the proposed Degradation Adaptation Block (DAB), we provide a detailed visualization of key components in Fig. 15. These include the attention mechanism, the GatedDA module, and its components—α, β, and γ—along with the fused feature map (combining attention and GatedDA), and the final restored image.

From the visualizations of the attention feature map in the second row of Fig. 15 and the GatedDA feature map in the third row, it is evident that GatedDA effectively models degradation factors such as rain, noise, and haze. This demonstrates the capability of GatedDA to capture and emphasize degradation-related information automatically.

Role of α, β, and γ in Degradation Modeling. To further analyze how α, β, and γ within GatedDA contribute to degradation modeling, we refer to the visualizations in the 4th to 6th rows of Fig. 15. These maps show that α, β, and γ consistently focus on various aspects of the degraded regions, each specializing in different levels or types of degradation. This specialization allows the model to adapt to varying degrees and types of degradation, making it more versatile and effective in the context of Image Restoration (IR) under a "One-for-Any" setting.

Fused Attention and GatedDA Features. The fused feature map, combining attention and GatedDA, offers further insights (visualized in Fig. 15). Compared to the original attention map, the fused map exhibits a richer representation of degradation, which equips the model with more comprehensive information for accurate restoration. In conclusion, the proposed DAB enhances the attention mechanism by embedding richer degradation information, thus improving restoration quality. The GatedDA module within DAB

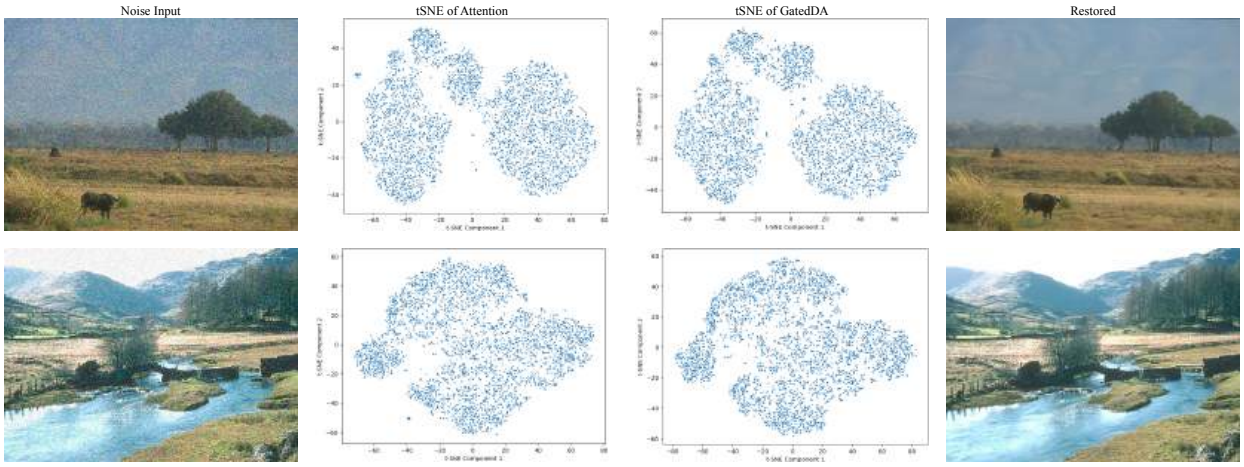


Figure 13: tSNE comparison between attention and GatedDA of the proposed method.

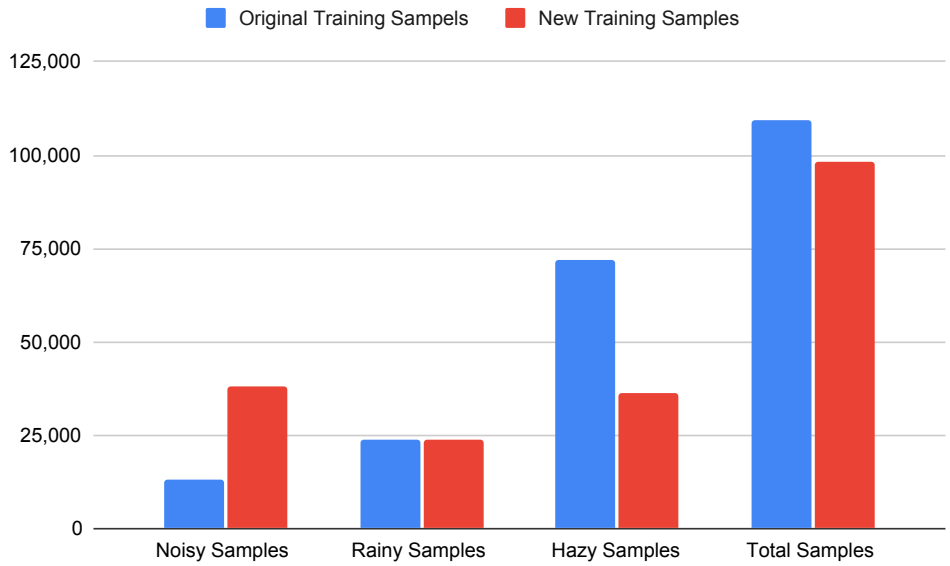


Figure 14: Training samples statistics.

introduces flexibility and diversity in handling various types and levels of degradation, contributing to the overall robustness and effectiveness of our method.

Affect of data distribution. As illustrated in Fig. 14, the original training samples exhibit significant variation across different IR tasks. Based on this observation, we propose a new set of training samples that over-represent the denoising samples and reduce the dehazing samples by half. This adjustment aims to achieve a more balanced data distribution. The results, presented in Tab. 12, demonstrate that training AnyIR with this revised set of samples, even when using only 90% of the original total training samples, yields significant performance improvements.

C Additional Results

Full Inference Time Comparison. Table 13 provides a detailed comparison of the full inference time between our method, AnyIR , and PromptIR Potlapalli et al. (2024) under the 3-degradation all-in-one IR setting. Across all tasks, our method demonstrates significantly faster inference, highlighting its computational

Figure 15: Visualization of the feature maps of each part within the proposed Degradation Adaptation Block (DAB).

Table 12: **Average** results for *denoising* and *overall* (*i.e.* dehazing, deraining, and denoising) under the 3-degradation IR setting.

Method	Denoising		Overall	
PromptIR Potlapalli et al. (2024)	31.11	.873	32.06	.913
AnyIR (Original, ours)	31.26	.878	32.80	.920
AnyIR (New, ours)	**31.31**	**.879**	**32.89**	**.921**

Table 13: Full **inference time** comparison under the 3-degradation IR setting. The lower the better.

Task	Num. Samples	PromptIR Potlapalli et al. (2024)	AnyIR (Ours)
Denoising ($\sigma = 15$)	68	41s	30s
Denoising ($\sigma = 25$)	68	41s	30s
Denoising ($\sigma = 50$)	68	42s	30s
Deraining	100	60s	43s
Dehazing	500	416s	274s
Frames Per Second	-	1.34	1.97

efficiency. For denoising tasks with varying noise levels ($\sigma = 15, 25, 50$), AnyIR reduces the inference time from 41-42 seconds to just 30 seconds, achieving over a 25% improvement. Similarly, for deraining, AnyIR processes 100 samples in 43 seconds, compared to 60 seconds for PromptIR, and for dehazing, it processes 500 samples in 274 seconds, substantially faster than PromptIR's 416 seconds. In terms of frames per second (FPS), AnyIR achieves 1.97 FPS, outperforming PromptIR's 1.34 FPS by nearly 50%. These results emphasize the efficiency of AnyIR in handling various degradation tasks, making it highly suitable for practical applications where computational speed is critical without compromising performance.

These results indicate that our AnyIR model not only maintains high performance but also offers substantial efficiency improvements, making it more suitable for real-time applications on resource-constrained devices. The reduced inference time ensures faster processing, enabling seamless deployment in scenarios requiring rapid decision-making, such as real-time video restoration or on-device image enhancement. Additionally, the improvement in frames per second (FPS) demonstrates its practicality for large-scale datasets and streaming applications. By achieving an effective balance between accuracy and speed, AnyIR provides a compelling solution for efficient, high-performance image restoration.

D Additional visual results.

The low-light enhanced visual comparison is provided in Fig. 16. The visual comparison results under the 3-degradation IR setting are shown in Fig. 17. It shows that the proposed AnyIR can effectively restore the clean image from its degraded counterparts compared to other comparison methods.

E Limitations and Future Work

Although the proposed AnyIR achieves strong performance and efficiency under the current all-in-one image restoration (IR) setting, it also presents several limitations that define the scope of its applicability.

A first limitation lies in the imbalance of degradation distribution in the benchmark training data. Certain degradations (*e.g.*, haze or rain) appear more frequently or with wider variation than others, which may bias the model toward better performance on dominant categories while providing smaller gains on under-represented ones. In future work, we plan to investigate more balanced or curriculum-style degradation scheduling, as well as data reweighting strategies, in order to improve robustness across heterogeneous degradation regimes.

In addition, while GatedDA and the spatial–frequency fusion mechanism provide consistent improvements across mixed and spatially non-uniform degradations, they are not universally optimal under all conditions.

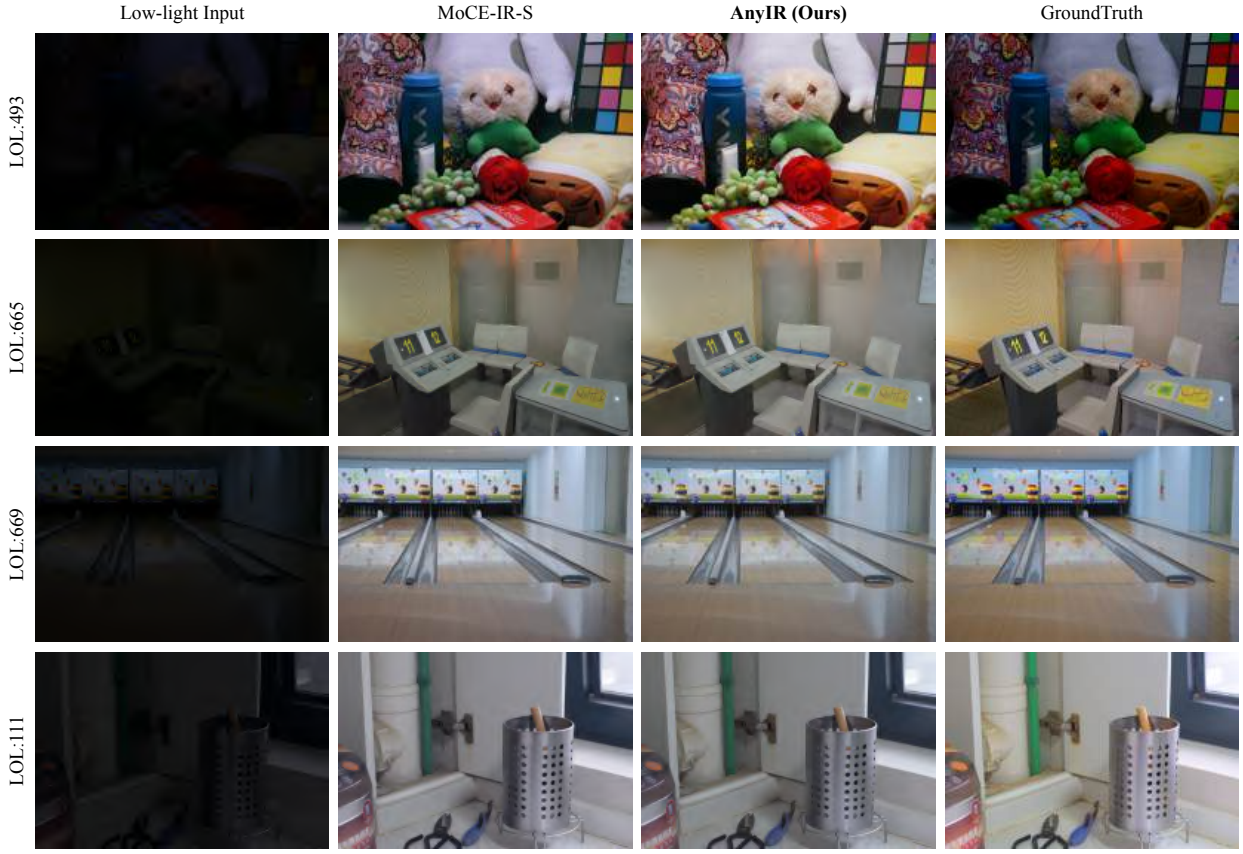


Figure 16: Additional visual comparison for low-light enhancement under the 5-degradation setting.

Since GatedDA selectively amplifies locally degraded regions, its benefit is weaker when degradations are globally uniform and largely dominated by global corruption. Likewise, our fusion design assumes complementary cues across spatial and frequency domains; when the degradation is strongly biased toward a single domain (*e.g.*, purely frequency-structured noise), the contribution of the other branch may be limited. Furthermore, the efficiency of our design arises from reducing attention channel capacity and compensating it with lightweight local modeling, which trades some global modeling flexibility for improved computational economy. Exploring adaptive routing between attention and GatedDA, as well as task-dependent fusion weighting, is a promising direction to mitigate these trade-offs.

Finally, we clarify that in this work, "All-in-One" refers to a single model jointly trained across the degradation types covered in our benchmark setting, rather than an unrestricted universal solution for arbitrary degradations beyond this scope.

F Broader Impact

The development of our unified image restoration model has significant potential, extending its impact beyond technical advancements. By reducing model complexity and computational requirements, our approach makes high-quality image restoration accessible on resource-constrained platforms like mobile devices. This enables efficient restoration in fields such as telemedicine, remote sensing, and digital archiving. Additionally, minimizing the computational footprint reduces the environmental impact of large-scale data processing, aligning with sustainable computing practices. The public availability of our code will further foster innovation and collaboration within the scientific community, setting new standards for efficiency and expanding practical applications in image restoration.

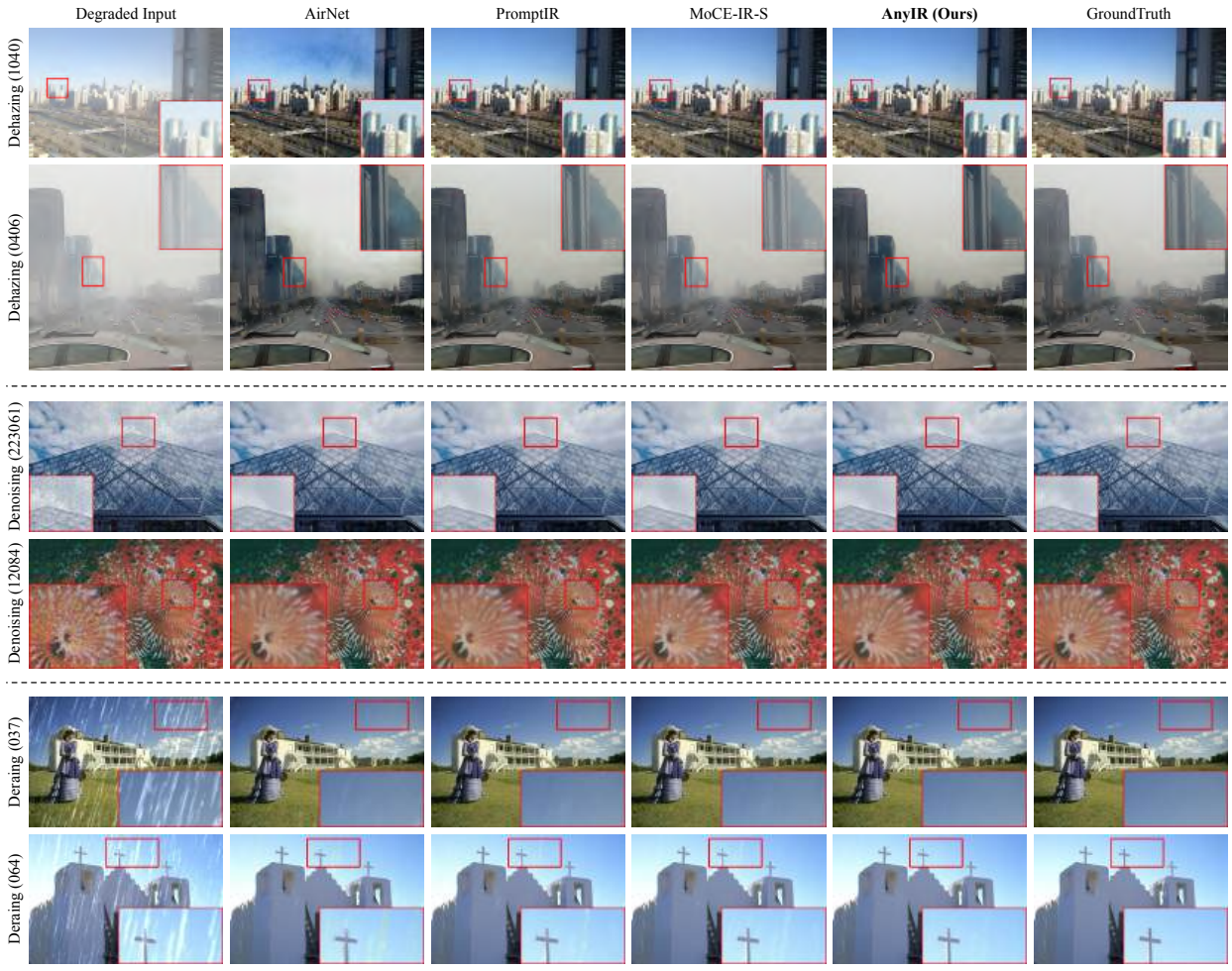


Figure 17: Additional visual comparison under the 3-degradation setting.