PHYSICALLY INTERPRETABLE MULTI-DEGRADATION IMAGE RESTORATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Although image restoration has advanced significantly, most existing methods target only a single type of degradation. In real-world scenarios, images often contain multiple degradations simultaneously, such as rain, noise, and haze, requiring models capable of handling diverse degradation types. Moreover, methods that improve performance through module stacking often suffer from limited interpretability. In this paper, we propose a novel interpretability-driven approach for multi-degradation image restoration, built upon a deep unfolding network that maps the iterative process of a mathematical optimization algorithm into a learnable network structure. Specifically, we employ an improved second-order semi-smooth Newton algorithm to ensure that each module maintains clear physical interpretability. To further enhance interpretability and adaptability, we design an explainable convolution module inspired by the human brain's flexible information processing and the intrinsic characteristics of images, allowing the network to flexibly leverage learned knowledge and autonomously adjust parameters for different input. The resulting tightly integrated architecture, named InterIR, demonstrates excellent performance in multi-degradation restoration while remaining highly competitive on single-degradation tasks.

1 INTRODUCTION

Image restoration (IR) aims to reconstruct a high-quality image $\overline{I}$ from its degraded observation D. The degradation process can be expressed as:

$$D = P\overline{I} + N \tag{1}$$

where P denotes the degradation operator and N represents additive noise. Different choices of P correspond to various image restoration tasks. In practice, however, the degradation process is often difficult to define precisely, leaving both P and N unknown. As a result, IR becomes a typical ill-posed problem, where multiple plausible solutions may correspond to the same degraded image. Traditional methods He et al. (2011); Rong et al. (2024) tackle this issue by incorporating task-specific priors, constructing explicit degradation models, and performing inverse operations. Although these approaches are physically interpretable and sometimes effective, they depend heavily on strong assumptions about degradation factors such as noise and blur kernels. In real-world conditions, however, the degradation is typically uncertain and complex, making accurate modeling challenging and limiting the generalization ability of such methods.

In recent years, the field of image restoration has experienced a paradigm shift driven by advances in deep learning Su et al. (2025); Gao et al. (2026); Cui et al. (2026). By leveraging the statistical characteristics of natural images, deep learning-based methods can implicitly learn diverse priors, achieving significantly better performance than traditional approaches. However, these data-driven networks, typically built through module stacking, often suffer from excessive parameter redundancy, limited interpretability. To improve model interpretability, some image restoration methods Fu et al. (2024); Luo et al. (2025b); Guo et al. (2025c) based on deep unfolding networks map the iterative solutions of mathematical optimization algorithms into learnable network structures, fully integrating model priors. This significantly enhances both the interpretability of network inference and the robustness of network learning. However, most deep unfolding-based methods are tailored to specific tasks or impose strict constraints on the degradation process, making it challenging to design a robust network capable of handling multiple restoration tasks under complex conditions.

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

Some approaches Deng et al. (2025); Zeng et al. (2025) attempt to address this by reformulating the original nonlinear coupled system into a network-friendly convex optimization problem.

While these methods Deng et al. (2025); Zeng et al. (2025) allow a single model to handle various degradation types, they are mainly focused on single-degradation image restoration (SDIR) and typically assume that each image contains only one type of degradation. In real-world scenarios, however, images are often heterogeneous, dynamic, and uncertain, with multiple degradations such as rain, haze, and noise occurring simultaneously. This complexity limits the effectiveness of current approaches in multi-degradation restoration. Some works Suganuma et al. (2019); Gao et al. (2025a) have investigated multi-degradation image restoration (MDIR), but their performance remains restricted, partly because they rely solely on simple attention mechanisms to select relevant features. Ref-IRT Zhang et al. (2024) attempts to overcome this by using a multi-stage framework that progressively transfers similar edges and textures from reference images. Nevertheless, it still depends heavily on the model's representational power and does not address MDIR from an interpretability standpoint.

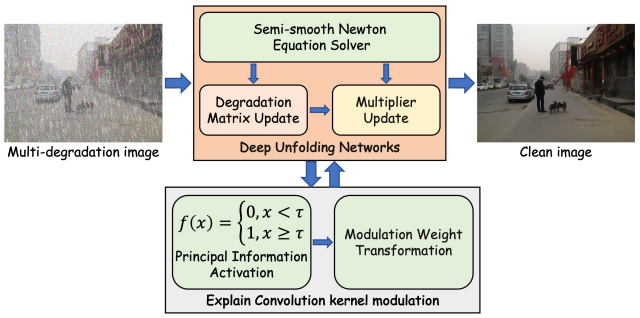

Figure 1: Mechanisms of our method.

Given these considerations, a natural question arises: can we design a network that adaptively restores images with multiple degradations by explicitly representing and dynamically regulating its internal decision-making process? To address this challenge, we propose InterIR (Figure 1), an adaptive MDIR network by guided by the improved semi-smooth Newton algorithm in its modules and connections, providing strong physical interpretability and transparency, while integrating an explainable convolution module to achieve image-specific adaptive modeling and iterative parameter optimization. Specifically, we reformulate the image restoration task as an improved second-order semi-smooth Newton optimization problem. The parameters of this optimization are learned through the network, ensuring that each module operates with clear physical interpretability. Furthermore, inspired by the human brain's flexible information processing and the intrinsic characteristics of images, we design an explainable convolution module. This module enables principal information activation and modulation weight transformation, allowing the network to dynamically adjust convolution kernel parameters based on multi-degradation image features, thereby achieving image-specific adaptive modeling. Extensive experiments show that our method delivers outstanding performance on multi-degradation restoration and remains highly competitive on single-degradation tasks.

The main contributions of this work are:

1. We propose InterIR for MDIR, which follows the improved semi-smooth Newton algorithm in its modules and connections for strong physical interpretability and transparency, and incorporates an explainable convolution module for image-specific adaptive modeling and iterative parameter optimization.

2. We reformulate the image restoration task as an improved second-order semi-smooth Newton optimization problem, with parameters learned through the network to ensure that each module operates with clear physical interpretability.

3. We design an explainable convolution module that flexibly leverages learned knowledge and autonomously adjusts parameters based on multi-degradation image features, enabling image-specific adaptive modeling.

4. Extensive experiments demonstrate that our method achieves outstanding performance in MDIR, effectively handling complex combinations of degradations, while also maintaining strong competitiveness on SDIR.

2 Related Works

2.1 Image Restoration

SDIR aims to recover high-quality images affected by a specific type of corruption. Traditional methods He et al. (2011); Rong et al. (2024) tackle this ill-posed problem by introducing hand-crafted priors to constrain the solution space. While effective in certain scenarios, these priors are highly dependent on expert knowledge and often lack adaptability. With the rapid advancement of deep learning, numerous data-driven approaches Gao et al. (2024b); Cui et al. (2026); Tang et al. (2026); Zhang et al. (2026); Zeng et al. (2025) have been developed to enhance image restoration performance. LSSR Gao et al. (2024b) introduces an adaptive module combination strategy for lightweight stereo image super-resolution, while XYScanNet Liu et al. (2025) employs an alternating intra- and inter-slice scanning strategy to better capture spatial dependencies. SFNet Cui et al. (2023d) and FSNet Cui et al. (2023a) propose dynamic frequency selection modules that adaptively identify the most informative components for restoration. AdaRevD Mao et al. (2024a) employs a classifier to estimate the degradation level of image patches, though it lacks adaptive flexibility. LoFormer Mao et al. (2024b) utilizes local channel-wise self-attention in the frequency domain to capture cross-covariance patterns within both low- and high-frequency local regions. However, the above methods generally train a single model dedicated to one specific degradation type. CAPT-Net Gao et al. (2024a) leverages changes in data components through prompt learning to achieve integrated image restoration. IDR Zhang et al. (2023) introduces an ingredient-oriented perspective to model degradation components, enhancing the framework's scalability across diverse conditions. AdaIR Cui et al. (2025) disentangles degradation information from clean image content by jointly exploiting both spatial and frequency representations. Despite their promising results in SDIR, these methods encounter significant challenges in MDIR, where complex, intertwined degradations severely hinder restoration accuracy and generalization.

MDIR aims to recover a clean image from one affected by multiple types of degradation. AASO Suganuma et al. (2019) executes parallel restoration operations with attention-based weighting to select the optimal strategy, but feature heterogeneity limits its performance on multi-degraded images. OWAN Huang et al. (2021) fuses heterogeneous features via high-order tensors to exploit richer statistics, yet its complex design hampers fine-detail reconstruction. DIGNet Shin et al. (2022) models sequential and spatially varying degradations, while MEPSNet Kim et al. (2020) adopts a mixture-of-experts framework with parameter sharing to handle region-specific distortions. Despite these advancements, their performance remains constrained, as most focus on task-adaptive attention designs that execute multiple parallel operations weighted by attention to select the most suitable restoration strategy. To address this, Ref-IRT Zhang et al. (2024) adopts a three-stage framework: the first stage estimates residuals in a coarse-to-fine manner, while the subsequent stages progressively transfer fine details from reference images. FDTANet Gao et al. (2025a) performs adaptive restoration in the frequency domain by analyzing component differences. Although some approaches Luo et al. (2025a); Zhang et al. (2025) are primarily tailored for SDIR, they also demonstrate strong MDIR performance through enhanced representational power from large vision models. However, these methods still overlook the intricate interactions among different degradation mechanisms. In this work, we address this limitation by modeling from the perspective of degradation mechanisms, mapping the iterative process of a mathematical optimization algorithm into a learnable network structure. Based on image-specific characteristics, we realize MDIR through an interpretable and adaptive framework.

2.2 Deep Unfolding Networks

Deep unfolding network is a neural architecture that integrates model-based optimization and data-driven learning by unfolding the iterative steps of a traditional mathematical optimization algorithm into a sequence of learnable network layers. Each layer corresponds to one iteration of the optimization process, where the algorithmic parameters are learned from data, enabling the network to retain physical interpretability. DIVA Dutta et al. (2022) unfolds a baseline adaptive denoising algorithm within a deep neural network framework, drawing inspiration from quantum many-body physics. D^3U-Net Han et al. (2024) introduces a dual-domain collaborative optimization strategy that fuses visual representations from the image domain with multi-resolution information from the wavelet domain. PEUNet Guo et al. (2025c) incorporates a snow shape prior to generate super-

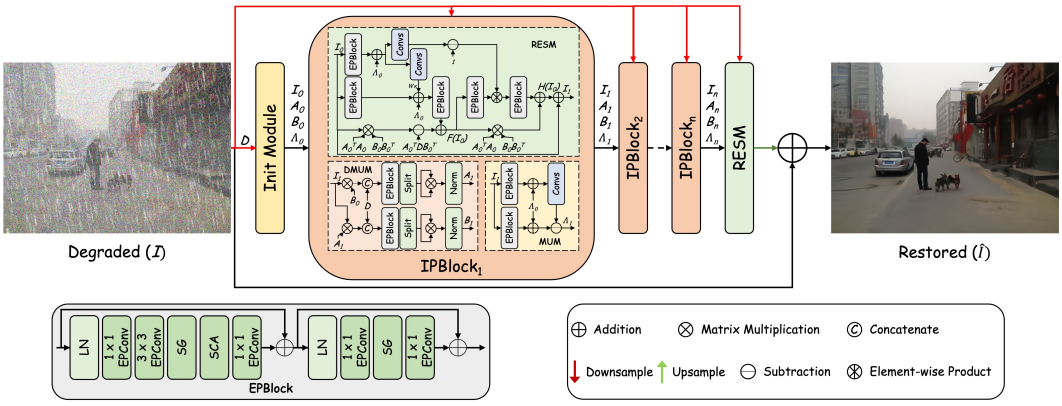

Figure 2: The overall architecture of the proposed InterIR consists of an initial module followed by n interpretability blocks (IPBlocks). Each IPBlock contains a restoration equation solver module (RESM), a degradation matrix update module (DMUM), and a multiplier update module (MUM).

vision signals, addressing the lack of real snow mask annotations. VLU-Net Zeng et al. (2025) leverages vision-language model features to automatically extract degradation-aware cues, eliminating the need for manually defined categories. REPO Fu et al. (2024) proposes a high-accuracy rotation-equivariant proximal network that embeds rotation symmetry priors into the deep unfolding framework. DeepSN-Net Deng et al. (2025) reformulates an improved second-order semi-smooth Newton (ISN) algorithm, transforming the original nonlinear equations into an optimization problem amenable to network implementation. Although the above methods assign physical meaning to each iteration, they still rely on neural networks to update matrices and solve operators, leaving the feature transformation and decision-making process largely a "black box." To address this limitation, we introduce an explainable convolution module inspired by the human brain's flexible information processing and the intrinsic properties of images, enabling the network to adaptively utilize learned knowledge and autonomously adjust parameters for different inputs.

3 METHOD

In this section, we first present an overview of the InterIR framework, followed by the formulation of the restoration problem and a detailed description of the solving process. Finally, we introduce the explainable convolution module.

3.1 OVERALL PIPELINE

Our proposed InterIR, illustrated in Figure 2. Given a degraded image $D \in \mathbb{R}^{H\times W\times 3}$, InterIR first applies $4\times$ downsampling on it to obtain a downsampled tensor $\mathbf{F_0} \in \mathbb{R}^{H\times W\times C}$ (H, W, and C denote the height, width, and number of channels, respectively). The downsampled tensor is then processed through an init module to initialize degradation matrix, multiplier and restored image. These initialized variables then go through n interpretability blocks (IPBlocks) and a restoration equation solver module (RESM) to obtain the final residual image $\mathbf{I_R} \in \mathbb{R}^{H\times W\times 3}$, which is added to the degraded input to produce the restored image: $\hat{\mathbf{I}} = \mathbf{I_R} + D$.

3.2 PROBLEM STATEMENT

The image degradation process is typically formulated as Eq. 1. However, the degradation matrix P grows quadratically with spatial resolution, leading to substantial computational overhead. To alleviate this issue, inspired by LoRA Hu et al. (2022), we decompose P into two smaller degradation matrices via a Kronecker product, effectively reducing computational complexity and allowing the degradation operation to scale linearly with image resolution. Specifically, we factorize $P \in \mathbb{R}^{HW\times HW\times C}$ into $A \in \mathbb{R}^{H\times H\times C}$ and $B \in \mathbb{R}^{W\times W\times C}$, enabling Eq. 1 to be reformulated as $D = A\bar{I}B + N$, where $B^T \otimes A \simeq P$. Following Deng et al. (2025), to enable adaptive restoration

under complex and diverse real-world degradations, we impose multiple constraints on the model and define the following optimization objective:

$$\min_{\overline{I},A,B} |A\overline{I}B - D|_2^2 + \alpha g(V(\overline{I})) + \beta\phi(A) + \gamma\varphi(B) \tag{2}$$

where $g(\cdot)$, $\phi(\cdot)$, and $\varphi(\cdot)$ are regularization functions, and α, β, and γ denote the corresponding weight coefficients. The operator $V(\cdot)$ extracts shallow image features. To facilitate optimization, we introduce an auxiliary variable C to substitute $V(\overline{I})$ and reformulate the problem using the augmented Lagrangian method as follows:

$$\mathcal{L}_{I,A,B,V(\overline{I}),\Lambda} := \frac{1}{2}\|AIB - D\|_2^2 + \alpha g(C) + \beta\phi(A) + \gamma\varphi(B) + \langle\Lambda, V(\overline{I}) - C\rangle + \frac{\epsilon}{2}\|V(\overline{I}) - C\|_2^2 \tag{3}$$

where Λ is the multiplier and ϵ is the penalty coefficient.

3.3 Solving Process

To solve the optimization objective in Eq. 3, we adopt the improved second-order semi-smooth Newton (ISN) algorithm, which decomposes the problem into three subproblems to alternately update $(\hat{I}_n, C_n)$, (A_n, B_n), and Λ_n:

$$\begin{aligned}(\hat{I}_n, C_n) &= \arg\min_{\hat{I},C} \mathcal{L}(\hat{I}, C, A_{n-1}, B_{n-1}, \Lambda_{n-1}) \\ (A_n, B_n) &= \arg\min_{A,B} \mathcal{L}(\hat{I}_n, C_n, A, B, \Lambda_{n-1}) \\ \Lambda_n &= \Lambda_{n-1} + \epsilon\left(V(\hat{I}_n) - C_n\right)\end{aligned} \tag{4}$$

We now focus on solving the core of the ISN algorithm in Eq. 4. Given the updated parameters $\Lambda = \Lambda_{n-1}$, $A = A_{n-1}$, and $B = B_{n-1}$, the optimal solution can be obtained by setting:

$$\begin{aligned}\alpha\,\partial g(C) - \Lambda + \epsilon(C - V(\hat{I})) &\ni 0 \\ A^T(A\hat{I}B - D)B^T + V^T[\Lambda + \epsilon(V(\hat{I}) - C)] &= 0\end{aligned} \tag{5}$$

where $\partial g(C)$ denotes the subderivative of $g(C)$. Then, it can be inferred that:

$$\begin{aligned}C &= \frac{\Lambda}{\epsilon} + V(\hat{I}) - \frac{\partial g(C)}{\epsilon} = S(\frac{\Lambda}{\epsilon} + V(\hat{I})) \\ A^T(A\hat{I}B - D)B^T + V^T[\Lambda + \epsilon(V(\hat{I}) - S(\frac{\Lambda}{\epsilon} + V(\hat{I})))] &= 0\end{aligned} \tag{6}$$

where $S(\cdot)$ is a nonlinear function determined by $\partial g(\cdot)$ and the ratio α/ϵ. For simplicity, we denote the resulting function as $F(\hat{I})$ and solve it iteratively using gradient descent:

$$\begin{aligned}F(\hat{I}_{n-1}) &= A^T(A\hat{I}_{n-1}B - D)B^T + V^T[\Lambda + \epsilon(V(\hat{I}_{n-1}) - S(\frac{\Lambda}{\epsilon} + V(\hat{I}_{n-1})))] \\ H(\hat{I}_{n-1}) &= A^TAF(\hat{I}_{n-1})BB^T + \epsilon V^T[V(F(\hat{I}_{n-1}) \odot (\mathbf{1} - S_d(\frac{\Lambda}{\epsilon} + V(\hat{I}_{n-1}))))] \\ \hat{I}_n &\leftarrow \hat{I}_{n-1} - \eta H(\hat{I}_{n-1})\end{aligned} \tag{7}$$

where $H(\hat{I}_{n-1}) \in \partial\|F(\hat{I}_{n-1})\|_2^2$ is the subderivative with respect to $\hat{I}_{n-1}$, $S_d(\cdot)$ is the subderivative of $S(\cdot)$, and η is the step size. The solution of the sub-problem can be obtained as:

$$\begin{aligned}A_n = D_\phi(D, A_{n-1}\hat{I}_n), B_n &= D_\varphi(D, \hat{I}_n A_n) \\ \Lambda_n &= \Lambda_{n-1} + \epsilon[V(\hat{I}_n) - S(\frac{\Lambda_{n-1}}{\epsilon} + V(\hat{I}_n))]\end{aligned} \tag{8}$$

where $D_\phi(\cdot)$ and $D_\varphi(\cdot)$ are determined by the regularization functions $\phi(\cdot)$ and $\varphi(\cdot)$, respectively.

The above procedure follows the derivation presented in DeepSN-Net Deng et al. (2025) to introduce the solving process of the ISN algorithm. Based on Eq. 8, we design a corresponding network structure that ensures each module maintains strong physical interpretability. As illustrated in Figure 2, given a degraded image $\mathbf{D}$, the network first downsamples it and initializes the degradation matrices, multipliers, and restored image parameters as:

$$\hat{I}_0 = D_{\downarrow_4}, \quad \Lambda_0 = f_1(\hat{I}_0) + \hat{I}_0, \quad A_0 = \frac{\hat{I}_0 \hat{I}_0{}^T}{||\hat{I}_0 \hat{I}_0{}^T||_2}, \quad B_0 = \frac{\hat{I}_0{}^T \hat{I}_0}{||\hat{I}_0{}^T \hat{I}_0||_2} \tag{9}$$

These initialized variables are subsequently fed through n interpretability blocks (IPBlocks). Each IPBlock comprises a restoration equation solver module (RESM) constructed according to Eq. 7, a degradation matrix update module (DMUM), and a multiplier update module (MUM) following Eq. 8. In particular, given the initial estimate I_0, the RESM computes I_1 as follows:

$$\begin{aligned} F(\hat{I}_0) =& A_0^T(D \ominus A_0 \hat{I}_0 B_0) B_0^T \oplus EP(EP(\hat{I}_0) \oplus \Lambda_0 \oplus W_\kappa \text{Convs}(\Lambda_0 \oplus EP(\hat{I}_0))) \\ H(\hat{I}_0) =& A_0^T A_0 F(\hat{I}_0) B_0 B_0^T \oplus EP(1 \ominus \text{Convs}(\Lambda_0 \oplus EP(\hat{I}_0))) \circledast EP(F(\hat{I}_0)) \\ \hat{I}_1 =& \hat{I}_0 \otimes -\eta H(\hat{I}_0) \end{aligned} \tag{10}$$

where $EP(\cdot)$ denotes the explainable block (EPBlock), W_κ is a learnable weight, and Convs consists of multiple convolutions followed by a ReLU. Subsequently, the DMUM is employed to update the degradation matrix as follows:

$$\begin{aligned} A_1^0, A_1^1 = Split(EP(Concat(\hat{I}_1 \otimes B_0, D))), \quad A_1 &= Norm(A_1^0 \otimes A_1^1) \\ B_1^0, B_1^1 = Split(EP(Concat(A_1 \otimes \hat{I}_1, D))), \quad B_1 &= Norm(B_1^0 \otimes B_1^1) \end{aligned} \tag{11}$$

Finally, we use the MUM to update the multiplier as:

$$\Lambda_1 = \Lambda_0 \oplus EP(\hat{I}_1) \ominus Convs(\Lambda_0 \oplus EP(\hat{I}_1)) \tag{12}$$

It is worth noting that, unlike DeepSN-Net Deng et al. (2025), in Eq. 10 we do not explicitly solve $F(\cdot)$ and $H(\cdot)$ as in Eq. 7. Instead, we introduce additional networks within both the RESM and DMUM to learn the weighting parameters between variables, making our formulation more consistent with Eq. 2. Ablation studies further confirm that this design leads to improved performance.

3.4 Explainable Block

While the operations above provide interpretability for each restoration step, we find that relying solely on the constraints in Eq. 2 is insufficient for fully adaptive multi-degradation restoration. To enhance both interpretability and input adaptability in convolutional operations, we design an explainable block (EPBlock), which not only improves MDIR performance but also mitigates the black-box limitations of DeepSN-Net Deng et al. (2025) in learning matrices and multiplier. As shown in Figure 2, the EPBlock follows the structure of NAFBlock Chen et al. (2022), with the key distinction that we replace the standard convolution with an explainable convolution. Inspired by principles from brain science and visual perception, this convolution combines the human brain's flexible information processing with the intrinsic characteristics of images. It dynamically modulates the convolution kernel for each input sample without modifying the original weights. Unlike conventional convolutions that use a fixed kernel for all samples, our design introduces an input-specific attention mask to adaptively adjust the effective receptive weights according to the input content, enabling more transparent and adaptive multi-degradation image restoration. Given an input feature map $x \in \mathbb{R}^{B \times C \times H \times W}$, a standard convolution computes as:

$$y = \text{Conv}(x) = W * x + \text{bias}, \tag{13}$$

where $W \in \mathbb{R}^{C_{\text{out}} \times C_{\text{in}} \times K_h \times K_w}$ is the kernel. In contrast, the explainable convolution introduces an input-specific modulation. For each input sample, we first generate a binary attention mask M based on a learnable threshold τ to highlight the most informative features, helping the model identify the types of multiple degradations in the image:

$$M_{b,c,h,w} = \begin{cases} 1, & x_{b,c,h,w} \geq \tau, \\ 0, & \text{otherwise.} \end{cases} \tag{14}$$

Table 1: Quantitative results on various combinations of degradation, where H, R, and N denote haze, rain, and noise, respectively.

Methods	H + R + N	H + R	H + N	R + N	H	R	N	Average
VLU-Net Zeng et al. (2025)	29.35/0.842	30.38/0.935	28.79/0.869	28.21/0.843	31.37/0.959	33.82/0.968	27.88/0.853	29.97/0.896
FDTANet Gao et al. (2025a)	29.68/0.846	30.91/0.937	**29.06**/0.874	27.63/0.827	**31.91/0.981**	29.13/0.892	28.88/0.857	29.60/0.888
DeepSN-Net Deng et al. (2025)	29.11/0.841	30.01/0.921	28.43/0.864	27.99/0.842	31.42/0.959	33.82/0.969	28.01/0.852	29.83/0.893
Ref-IRT Zhang et al. (2024)	28.95/0.823	32.33/0.961	28.62/0.872	31.22/0.831	30.69/0.961	31.89/0.902	29.39/0.899	30.44/0.893
Perceive-IR Zhang et al. (2025)	**30.22/0.894**	31.55/0.945	28.32/0.871	30.52/0.849	31.90/0.980	38.95/0.983	31.88/0.923	31.90/0.921
Defusion Luo et al. (2025a)	29.66/0.849	29.93/0.937	28.11/0.866	**32.17/0.920**	31.29/0.969	**40.15/0.986**	32.72/0.925	32.01/0.922
StarIR Cui et al. (2026)	29.55/0.832	29.76/0.931	27.95/0.871	28.63/0.842	30.65/0.965	32.89/0.971	28.19/0.854	29.66/0.895
ClearAIR Zhang et al. (2026)	30.08/0.857	30.92/0.922	28.66/0.879	30.99/0.859	31.11/0.959	33.52/0.960	29.53/0.897	30.69/0.905
InterIR(Ours)	**30.22**/0.880	**32.46/0.963**	**29.06/0.883**	31.85/0.885	31.88/0.968	39.33/0.984	**32.75/0.926**	**32.51/0.927**

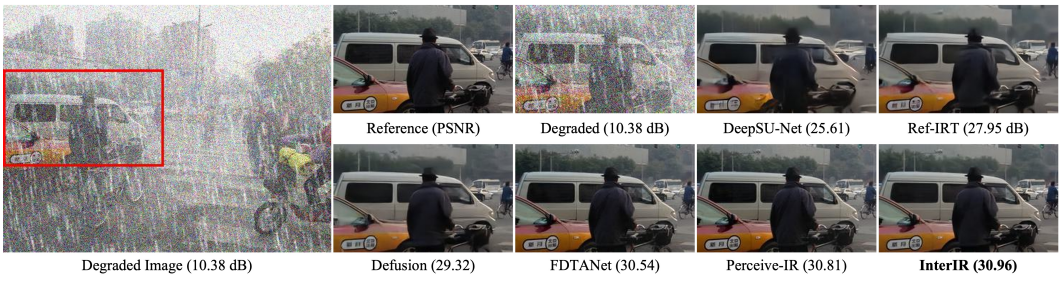

Figure 3: Qualitative results under the MDIR experimental setup.

To align with the convolution kernel size, the mask is spatially pooled $\tilde{M}_{b,c} = \text{Pool}(M_{b,c}) \in \mathbb{R}^{K_h \times K_w}$. Next, a soft attention distribution is obtained via softmax normalization:

$$A_{b,c,i,j} = \frac{\exp(\tilde{M}_{b,c,i,j})}{\sum_{i',j'} \exp(\tilde{M}_{b,c,i',j'})}, \quad \text{s.t.} \sum_{i,j} A_{b,c,i,j} = 1. \tag{15}$$

The attention map A acts as a modulation factor, producing an adaptive kernel for each input $W_b = W \odot A_b$, where $\odot$ denotes element-wise multiplication and W_b is the dynamically modulated kernel for the b-th input. Finally, the output is computed via batch-independent convolution $y_b = W_b * x_b + \text{bias}, b = 1, 2, \ldots, B$. This design offers several advantages. First, each input sample generates its own kernel modulation A_b, enabling the network to adaptively focus on the most relevant spatial regions. Second, the learned mask M explicitly highlights which regions activate the convolution kernel, providing an interpretable view of the model's attention. Third, since W_b is produced through modulation rather than replacing W, the original kernel remains trainable and shared across the dataset, ensuring stability during learning.

4 Experiments

In this section, we describe the comparison results, and present ablation studies. The experimental setup and more experiments are provided in Appendix A.2 and A.3.

4.1 Experimental Results

4.1.1 MDIR Setting

We evaluate InterIR on seven different combinations of degradation types, covering all permutations of haze, rain, and noise in Table 1. Averaged across all tasks, our method outperforms the second-best approach, Defusion Luo et al. (2025a), by 0.5 dB. Compared with other MDIR methods, Ref-IRT Zhang et al. (2024) and FDTANet Gao et al. (2025a), our model achieves significant improvements of 2.07 dB and 2.91 dB, respectively. Compared with the baseline model DeepSN-Net Deng et al. (2025), our InterIR achieves consistently better performance across all evaluations.

To further assess the robustness of our approach across diverse degradation types, we conduct additional experiments using the same training dataset. The results indicate that while InterIR may

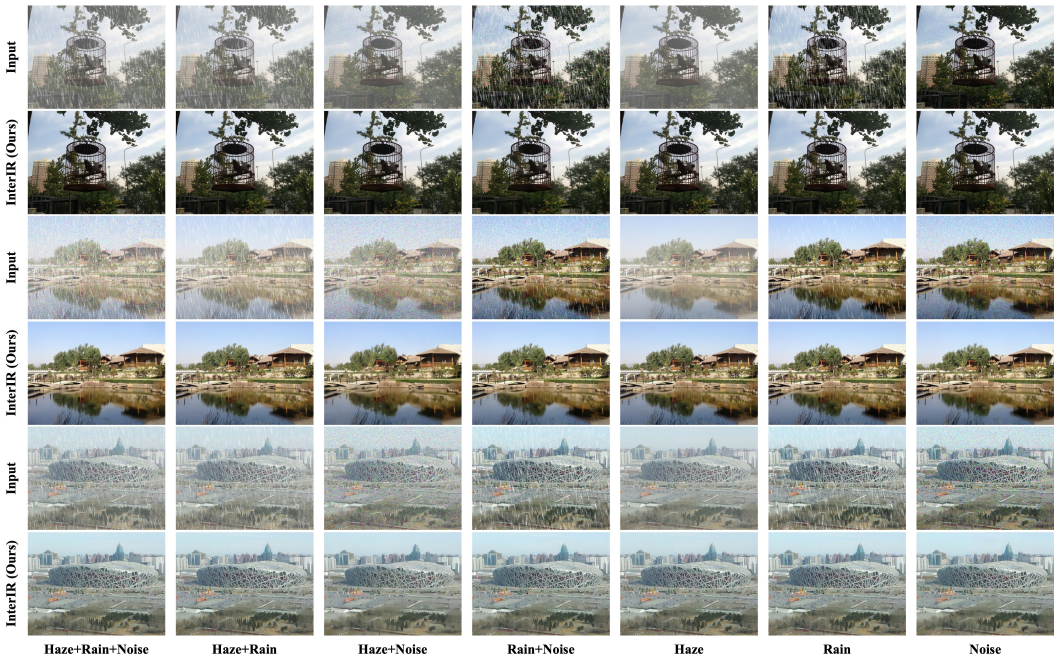

Figure 4: Image restoration results across various combinations of degradation types.

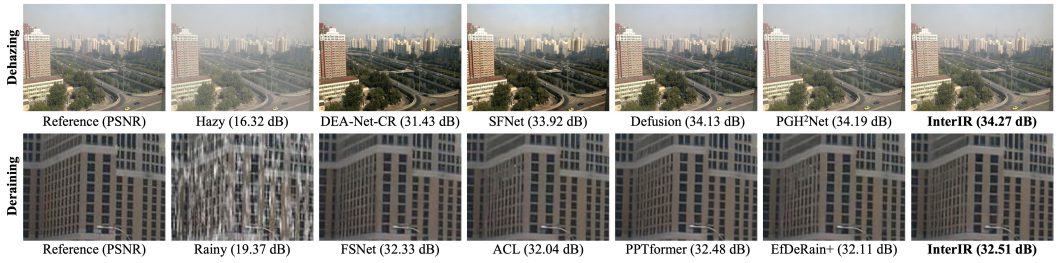

Figure 5: Qualitative results under the SDIR experimental setup.

not always produce the largest gain in every specific scenario, it consistently delivers stable performance across an arbitrary number of degradation combinations. As illustrated in Figure 3, our model generates restored images that are sharper and visually closer to the ground truth compared to competing methods. Figure 4 presents the visual results of our method under different combinations of degradation types. For each case, the top row shows the input images suffering from multiple degradations, while the bottom row displays the corresponding outputs restored by InterIR. It can be observed that our model effectively eliminates diverse degradations and consistently reconstructs images with sharp details, natural textures, and high visual fidelity across all scenarios.

4.1.2 SDIR Setting

Image Dehazing. Table 2 reports the quantitative results of different image dehazing methods. InterIR consistently achieves superior performance, demonstrating its effectiveness and robustness. Specifically, on the indoor dataset, InterIR surpasses the previous best-performing method, PGH^{2}Net Su et al. (2025), by 0.28 dB in PSNR. On the outdoor dataset, it further achieves a 0.59 dB improvement over the previous SOTA method, IRNext Cui et al. (2023c).

Image Deraining. Following the evaluation protocol in prior work Gao et al. (2025b), we report PSNR and SSIM metrics on the Y channel of the YCbCr color space. As presented in Table 2, InterIR consistently matches or surpasses existing methods across Test100 Zhang et al. (2017) and

Table 2: Quantitative results in the SDIR setting.

Methods	Dehazing		Deraining		Methods	Dehazing		Deraining	
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑		PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
IRNext Cui et al. (2023c)	41.21	**0.996**	39.18	**0.996**	EfDeRain+ Guo et al. (2025a)	31.10	0.911	33.12	0.925
DEA-Net-CR Chen et al. (2024)	41.31	0.995	36.59	0.990	MHNet Gao et al. (2025b)	31.25	0.901	33.45	0.925
Defusion Luo et al. (2025a)	41.65	0.995	37.41	0.993	DeepSN-Net Deng et al. (2025)	31.60	0.920	33.45	0.931
PGH^{2}Net Su et al. (2025)	41.70	**0.996**	37.52	0.989	StarIR Cui et al. (2026)	31.58	**0.922**	33.39	0.928
InterIR(Ours)	**41.98**	**0.996**	**39.77**	0.989	**InterIR(Ours)**	**32.65**	0.919	**35.02**	**0.940**

Table 3: Ablation study on individual components of InterIR.

Method	Baseline	replace with EPConv	replace with RESM	replace with DMUM and MUM	Ours
PSNR	29.11	29.83	29.44	29.35	30.22

Test1200 Jiang et al. (2020) benchmark datasets. Notably, it achieves an average PSNR improvement of 1.31 dB gain over our baseline DeepSN-Net Deng et al. (2025).

As shown in Figure 5, InterIR produces visually superior results, significantly mitigating color distortion and restoring natural tones compared with state-of-the-art methods. Furthermore, our model reconstructs sharper textures and finer structural details, highlighting its robust restoration capability.

4.2 Ablation Studies

We perform ablation studies in the MDIR setting. Using DeepSN-Net Deng et al. (2025) as the baseline, we conduct a stepwise ablation by progressively integrating each module.

Table 3 presents the results of ablation experiments evaluating the contribution of each component in InterIR. The baseline network achieves a PSNR of 29.11 dB. Replacing the standard convolution with the proposed explainable convolution module raises the PSNR to 29.83 dB, a gain of 0.72 dB, demonstrating its effectiveness in adaptively leveraging image-specific features for restoration. Adding the RESM module increases the PSNR to 29.44 dB, a 0.33 dB improvement over the baseline, while integrating the DMUM and MUM modules yields a PSNR of 29.35 dB, corresponding to a 0.24 dB gain. Finally, the complete InterIR model, incorporating all proposed modules, achieves the highest PSNR of 30.22 dB, an overall improvement of 1.11 dB. These results indicate that each module contributes positively to the network's performance, and their combination produces the most significant enhancement in multi-degradation image restoration.

To further assess the generality and effectiveness of the explainable convolution module, we integrate it into the convolution layers of existing image restoration methods. As reported in Table 4, this modification results in substantial performance improvements. Specifically, FDTANet Gao et al. (2025a) and DeepSN-Net Deng et al. (2025) achieve PSNR gains of 0.63 dB and 0.72 dB, respectively, demonstrating that our explainable convolution can consistently enhance both existing architectures and their restoration capabilities.

5 Conclusion

In this paper, we present InterIR, an interpretability-driven framework for multi-degradation image restoration. Based on a deep unfolding architecture, our approach transforms the iterative process of a second-order semi-smooth Newton optimization algorithm into a learnable and physically interpretable network structure. In addition, the inclusion of an explainable convolution module, allows InterIR to dynamically adjust its parameters according to the characteristics of each input image, thereby enhancing both interpretability and adaptability. Extensive experiments show that InterIR achieves SOTA performance in MDIR tasks and remains highly competitive in SDIR scenarios.

Table 4: Plug-and-play ablation experiments.

Method	FDTANet Gao et al. (2025a)	replace with EPConv	DeepSN-Net Deng et al. (2025)	replace with EPConv
PSNR	29.68	30.31	29.11	29.83

AI USE STATEMENT

In this work, we used generative AI tools to assist with manuscript editing. GPT-based tools were used to improve the language, grammar, and readability of the manuscript without changing the underlying scientific content. The underlying data and numerical results in the figure were not generated or modified by AI.

All AI-assisted outputs were manually reviewed and verified by the authors. The authors take full responsibility for the final manuscript, including all text, figures, experimental results, scientific claims, and conclusions.

REFERENCES

Abdelrahman Abdelhamed, Stephen Lin, and Michael S Brown. A high-quality denoising dataset for smartphone cameras. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1692–1700, 2018.

Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. *ECCV*, 2022.

Sixiang Chen, Tian Ye, Yun Liu, Taodong Liao, Jingxia Jiang, Erkang Chen, and Peng Chen. Msp-former: Multi-scale projection transformer for single image desnowing. In *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5, 2023. doi: 10.1109/ICASSP49357.2023.10095605.

Zixuan Chen, Zewei He, and Zhe-Ming Lu. Dea-net: Single image dehazing based on detail-enhanced convolution and content-guided attention. *IEEE Transactions on Image Processing*, 2024.

Yuning Cui, Wenqi Ren, Xiaochun Cao, and Alois Knoll. Image restoration via frequency selection. *TPAMI*, pp. 1–16, 2023a.

Yuning Cui, Wenqi Ren, Xiaochun Cao, and Alois Knoll. Focal network for image restoration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 13001–13011, 2023b.

Yuning Cui, Wenqi Ren, Sining Yang, Xiaochun Cao, and Alois Knoll. Irnext: Rethinking convolutional network design for image restoration. In *Proceedings of the 40th International Conference on Machine Learning*, 2023c.

Yuning Cui, Yi Tao, Zhenshan Bing, Wenqi Ren, Xinwei Gao, Xiaochun Cao, Kai Huang, and Alois Knoll. Selective frequency network for image restoration. In *ICLR*, 2023d.

Yuning Cui, Syed Waqas Zamir, Salman Khan, Alois Knoll, Mubarak Shah, and Fahad Shahbaz Khan. AdaIR: Adaptive all-in-one image restoration via frequency mining and modulation. In *The Thirteenth International Conference on Learning Representations*, 2025.

Yuning Cui, Syed Waqas Zamir, Ming-Hsuan Yang, Alois Knoll, Fahad Shahbaz Khan, and Salman Khan. Starir: Convolutional image restoration with spatial-frequency fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–18, 2026.

Xin Deng, Chenxiao Zhang, Lai Jiang, Jingyuan Xia, and Mai Xu. Deepsn-net: Deep semi-smooth newton driven network for blind image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(4):2632–2646, 2025.

S. Dutta, Adrian Basarab, Bertrand Georgeot, and Denis Kouam'e. Diva: Deep unfolded network from quantum interactive patches for image restoration. *Pattern Recognit.*, 155:110676, 2022.

Jiahong Fu, Qi Xie, Deyu Meng, and Zongben Xu. Rotation equivariant proximal operator for deep unfolding methods in image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(10):6577–6593, 2024.

Xueyang Fu, Jiabin Huang, Delu Zeng, Yue Huang, Xinghao Ding, and John Paisley. Removing rain from single images via a deep detail network. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1715–1723, 2017. doi: 10.1109/CVPR.2017.186.

Hu Gao, Jing Yang, Ying Zhang, Ning Wang, Jingfan Yang, and Depeng Dang. Prompt-based ingredient-oriented all-in-one image restoration. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(10):9458–9471, 2024a.

Hu Gao, Jing Yang, Ying Zhang, Jingfan Yang, Bowen Ma, and Depeng Dang. Learning optimal combination patterns for lightweight stereo image super-resolution. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pp. 5566–5574, 2024b.

Hu Gao, Bowen Ma, Ying Zhang, Jingfan Yang, Jing Yang, and Depeng Dang. Frequency domain task-adaptive network for restoring images with combined degradations. *Pattern Recognition*, 158:111057, 2025a.

Hu Gao, Ying Zhang, Jing Yang, and Depeng Dang. Mixed hierarchy network for image restoration. *Pattern Recognition*, 161:111313, 2025b.

Hu Gao, Bowen Ma, Ying Zhang, Jingfan Yang, Jing Yang, Xingjian Wang, and Depeng Dang. Emphasizing crucial features for efficient image restoration. *Pattern Recognition*, pp. 113575, 2026.

Yubin Gu, Yuan Meng, Jiayi Ji, and Xiaoshuai Sun. Acl: Activating capability of linear attention for image restoration. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 17913–17923, 2025.

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. *arXiv preprint arXiv:2402.15648*, 2024.

Qing Guo, Hua Qi, Jingyang Sun, Felix Juefei-Xu, Lei Ma, Di Lin, Wei Feng, and Song Wang. Efficientderain+: Learning uncertainty-aware filtering via rainmix augmentation for high-efficiency deraining. *International Journal of Computer Vision*, 133(4):2111–2135, 2025a.

Xin Guo, Xi Wang, Xueyang Fu, and Zheng-Jun Zha. Deep unfolding network for image desnowing with snow shape prior. *IEEE Transactions on Circuits and Systems for Video Technology*, 35(5): 4740–4752, 2025b.

Xin Guo, Xi Wang, Xueyang Fu, and Zheng-Jun Zha. Deep unfolding network for image desnowing with snow shape prior. *IEEE Transactions on Circuits and Systems for Video Technology*, 35(5): 4740–4752, 2025c.

Kai Han, Jin Wang, Yunhui Shi, Nam Ling, and Baocai Yin. D3u-net: Dual-domain collaborative optimization deep unfolding network for image compressive sensing. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pp. 9952–9960, 2024.

Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *TPAMI*, 2011.

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

Yuanfei Huang, Jie Li, Yanting Hu, Hua Huang, and Xinbo Gao. Deep convolution modulation for image super-resolution. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(5): 3647–3662, 2024.

Zihao Huang, Chao Li, Feng Duan, and Qibin Zhao. Multi-distorted image restoration with tensor 1×1 convolutional layer. In *2021 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, 2021.

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. *CVPR*, 2020.

Xingyu Jiang, Ning Gao, Hongkun Dou, Xiuhui Zhang, Xiaoqing Zhong, Yue Deng, and Hongjue Li. Global modeling matters: A fast, lightweight, and effective baseline for efficient image restoration. *IEEE Transactions on Image Processing*, 35:2740–2754, 2026.

Sijin Kim, Namhyuk Ahn, and Kyung-Ah Sohn. Restoring spatially-heterogeneous distortions using mixture of experts network. In *Proceedings of the Asian Conference on Computer Vision (ACCV)*, November 2020.

D. Kingma and J. Ba. Adam: A method for stochastic optimization. *Computer Science*, 2014.

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

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

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

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

Yaowei Li, Hang An, Tong Zhang, Xiaoxuan Chen, Bo Jiang, and Jinshan Pan. Omni-deblurring: Capturing omni-range context for image deblurring. *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1–1, 2025.

Yu Li, Robby T. Tan, Xiaojie Guo, Jiangbo Lu, and Michael S. Brown. Rain streak removal using layer priors. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2736–2744, 2016. doi: 10.1109/CVPR.2016.299.

Hanzhou Liu, Chengkai Liu, Jiacong Xu, Peng Jiang, and Mi Lu. Xyscannet: An interpretable state space model for perceptual image deblurring. In *Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR)*, pp. 779–789, 2025.

I. Loshchilov and F. Hutter. Sgdr: Stochastic gradient descent with warm restarts. 2016.

Wenyang Luo, Haina Qin, Zewen Chen, Libin Wang, Dandan Zheng, Yuming Li, Yufan Liu, Bing Li, and Weiming Hu. Visual-instructed degradation diffusion for all-in-one image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12764–12777, June 2025a.

Yu Luo, Yue Cai, Jie Ling, Yingdan Ji, Yanmei Tie, and Shun Yao. Joint edge optimization deep unfolding network for accelerated mri reconstruction. *IEEE Transactions on Computational Imaging*, 11:11–23, 2025b.

Xintian Mao, Qingli Li, and Yan Wang. Adarevd: Adaptive patch exiting reversible decoder pushes the limit of image deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 25681–25690, June 2024a.

Xintian Mao, Jiansheng Wang, Xingran Xie, Qingli Li, and Yan Wang. Loformer: Local frequency transformer for image deblurring. In *Proceedings of the 32nd ACM International Conference on Multimedia*, 2024b.

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*, volume 2, pp. 416–423. IEEE, 2001.

Vaishnav Potlapalli, Syed Waqas Zamir, Salman Khan, and Fahad Shahbaz Khan. Promptir: Prompting for all-in-one blind image restoration. *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

Jianxiang Rong, Hua Huang, and Jia Li. Imu-assisted accurate blur kernel re-estimation in non-uniform camera shake deblurring. *IEEE Transactions on Image Processing*, 33:3823–3838, 2024.

Wooksu Shin, Namhyuk Ahn, Jeong-Hyeon Moon, and Kyung-Ah Sohn. Exploiting distortion information for multi-degraded image restoration. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 536–545, 2022.

Xiongfei Su, Siyuan Li, Yuning Cui, Miao Cao, Yulun Zhang, Zheng Chen, Zongliang Wu, Zedong Wang, Yuanlong Zhang, and Xin Yuan. Prior-guided hierarchical harmonization network for efficient image dehazing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 7042–7050, 2025.

M. Suganuma, X. Liu, and T. Okatani. Attention-based adaptive selection of operations for image restoration in the presence of unknown combined distortions. In *CVPR*, 2019.

Xiaole Tang, Xiaoyi He, Jiayi Xu, Xiang Gu, and Jian Sun. Learning continuous wasserstein barycenter space for generalized all-in-one image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–16, 2026. doi: 10.1109/TPAMI.2026.3669121.

Xueyang Wang, Zhixin Zheng, Jiandong Shao, Yule Duan, and Liang-Jian Deng. Adaptive rectangular convolution for remote sensing pansharpening. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 17872–17881, 2025.

Wenhan Yang, Robby T. Tan, Jiashi Feng, Jiaying Liu, Zongming Guo, and Shuicheng Yan. Deep joint rain detection and removal from a single image. *CVPR*, pp. 1685–1694, 2016.

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*, 2021.

Haijin Zeng, Xiangming Wang, Yongyong Chen, Jingyong Su, and Jie Liu. Vision-language gradient descent-driven all-in-one deep unfolding networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7524–7533, June 2025.

He Zhang and Vishal M. Patel. Density-aware single image de-raining using a multi-stream dense network. *CVPR*, pp. 695–704, 2018.

He Zhang, Vishwanath A. Sindagi, and Vishal M. Patel. Image de-raining using a conditional generative adversarial network. *TCSVT*, 30:3943–3956, 2017.

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.

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. *IEEE Transactions on Image Processing*, pp. 1–1, 2025.

Xu Zhang, Huan Zhang, Guoli Wang, Qian Zhang, and Lefei Zhang. Clearair: A human-visual-perception-inspired all-in-one image restoration. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 12861–12869, 2026.

Yi Zhang, Qixue Yang, Damon M. Chandler, and Xuanqin Mou. Reference-based multi-stage progressive restoration for multi-degraded images. *IEEE Transactions on Image Processing*, 33: 4982–4997, 2024.

A Appendix

A.1 Loss

To enable effective restoration in both the spatial and frequency domains, we optimize the proposed InterIR network using the following loss function:

$$L = L_1(\hat{I}, \overline{I}) + \lambda |\mathcal{F}(\hat{I}) - \mathcal{F}(\overline{I})|_1 \tag{16}$$

where $\overline{I}$ denotes the target images, and L_1 is the L1 loss. $\mathcal{F}$ denoting the fast Fourier transform. λ balances the contributions of the two loss terms and is set to 0.1, following Cui et al. (2023a).

Table 5: Test set generation by multi-degradation types.

Multi-degradation	Haze level	Rain level	Noise level
Haze + Rain + Noise	[0,150]	[0,300]	[0,50]
Haze + Rain	[0,150]	[0,300]	[0]
Haze + Noise	[0,150]	[0]	[0,50]
Rain + Noise	[0]	[0,300]	[0,50]
Haze	[0,150]	[0]	[0]
Rain	[0]	[0,300]	[0]
Noise	[0]	[0]	[0,50]

A.2 Experimental Setup

A.2.1 Dataset

MDIR. Following Gao et al. (2025a), we perform comparative experiments on a dataset that includes various combinations of degradations such as haze, rain, and noise. The dataset contains 18,000 natural images collected from multiple public image restoration benchmarks Li et al. (2018a); Yang et al. (2016); Zhang et al. (2017); Martin et al. (2001). Among them, 13,000 images are randomly selected for training, where synthetic degradations are applied with intensity ranges of [0,150] for haze, [0,300] for rain, and [0,50] for noise. From the remaining images, 500 are randomly chosen for testing. As summarized in Table 5, the test set includes seven distinct degradation combinations, covering all possible permutations of haze, rain, and noise. For instance, in the "haze + rain + noise" scenario, all three degradation types are applied as in training, whereas in the "haze + rain" case, only the noise component is omitted while the other settings remain unchanged. The remaining configurations are constructed following the same principle.

SDIR. Following the experimental settings of recent state-of-the-art image deraining methods Zamir et al. (2021); Cui et al. (2023a), we train our model on 13,712 clean–rain image pairs sourced from multiple benchmarks Yang et al. (2016); Zhang et al. (2017); Fu et al. (2017); Li et al. (2016). The trained InterIR is then evaluated on four widely used test sets: Rain100H Yang et al. (2016), Rain100L Yang et al. (2016), Test100 Zhang et al. (2017), and Test1200 Zhang & Patel (2018).

To evaluate the image dehazing performance, we conduct experiments on the daytime subset of the RESIDE dataset Li et al. (2018a), which includes both synthetic indoor and outdoor scenes. RESIDE provides two training sets: the indoor training set (ITS) and the outdoor training set (OTS), as well as a synthetic objective testing set (SOTS). The ITS contains 13,990 hazy images generated from 1,399 clean images, while the OTS includes 313,950 hazy images synthesized from 8,970 clean images. We train our model separately on the ITS and OTS datasets and evaluate it on their corresponding test sets, SOTS-Indoor and SOTS-Outdoor, each containing 500 paired samples.

A.2.2 Training details

Our models are optimized using Adam Kingma & Ba (2014) with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The learning rate is initialized at 2×10^{-4} and gradually reduced to 1×10^{-7} according to the cosine annealing strategy Loshchilov & Hutter (2016). During training, image patches of size 256×256 are randomly sampled with a batch size of 32 for a total of 4×10^5 iterations. Data augmentation is performed through horizontal and vertical flipping. The n in Figure 2 is 16. We set the $\eta = 0.01$ in Eq. 10. To ensure fairness, all compared deep learning-based methods are either fine-tuned or retrained following the parameter configurations specified in their original papers.

A.3 More Experiments

A.3.1 Additional SDIR Results

Image Desnowing. Table 6 reports the results on CSD, SRRS, and Snow100K. InterIR achieves the highest PSNR on all three datasets, outperforming the strongest competing methods by 0.26 dB, 0.23 dB, and 0.09 dB, respectively. Its SSIM results are also competitive across the three bench-

Table 6: Image desnowing results.

Methods	CSD		SRRS		Snow100K	
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
NAFNet Chen et al. (2022)	33.13	0.96	29.72	0.94	32.41	0.95
FocalNet Cui et al. (2023b)	37.18	**0.99**	31.34	**0.98**	33.53	0.95
MSP-Former Chen et al. (2023)	33.75	0.96	30.76	0.95	33.43	**0.96**
IRNeXt Cui et al. (2023c)	37.29	**0.99**	31.91	**0.98**	33.61	0.95
PEUNet Guo et al. (2025b)	37.28	0.97	31.89	**0.98**	34.11	**0.96**
PW-FNet Jiang et al. (2026)	-	-	32.74	**0.98**	34.50	0.95
InterIR(Ours)	**37.55**	0.98	**32.97**	**0.98**	**34.59**	**0.96**

Table 7: Image deblurring results.

Methods	GoPro		HIDE	
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
MPRNet Zamir et al. (2021)	32.66	0.959	30.96	0.939
FSNet Cui et al. (2023a)	33.29	0.963	31.05	0.941
MambaIR Guo et al. (2024)	33.21	0.962	31.01	0.939
ACL Gu et al. (2025)	33.25	**0.964**	-	-
Omni-Deblurring Li et al. (2025)	33.29	0.963	**31.65**	**0.947**
InterIR(Ours)	**33.53**	**0.964**	31.61	**0.947**

marks. These results indicate that InterIR performs consistently across different snow distributions and scene contents.

Image Deblurring. Table 7 compares different methods on GoPro and HIDE. On GoPro, InterIR achieves the best PSNR with a 0.24 dB improvement over the strongest previous result, while matching the best SSIM. On HIDE, it achieves comparable performance to Omni-Deblurring, with only a small PSNR difference and the same SSIM. Overall, InterIR maintains competitive performance across both synthetic motion blur and more complex human-aware blur scenarios.

Image Denoising. Table 8 reports denoising results on CBSD68, Kodak24, and Urban100 under noise levels $\epsilon = 15$, 25, and 50. InterIR achieves the best results in most evaluated settings. The improvements are particularly evident on Kodak24, where gains of up to 0.33 dB are obtained over previous methods. On Urban100, InterIR achieves the best performance at noise levels 15 and 50, while remaining competitive at $\epsilon = 25$. These results show that InterIR maintains stable denoising performance across different image contents and noise intensities.

A.3.2 Generalization

Zero-Shot Generalization in Real-World Scenes. We further evaluate models trained under the MDIR setting directly on real-world degradation datasets without additional adaptation. As shown in Table 9, InterIR achieves the best performance on RealRain-1k-L Li et al. (2022) across PSNR, SSIM, and LPIPS, improving the strongest competing method by 0.17 dB in PSNR. On RTTS Li et al. (2018b), InterIR obtains the lowest FADE and the highest NIMA, although its BRISQUE score remains below the best competing results. On SIDD Abdelhamed et al. (2018), InterIR achieves the

Table 8: Image denoising results.

Methods	CBSD68			Kodak24			Urban100		
	15	25	50	15	25	50	15	25	50
SFNet Cui et al. (2023d)	34.09	31.49	28.02	34.93	32.42	29.25	34.19	32.01	29.03
FSNet Cui et al. (2023a)	34.11	31.51	28.01	34.95	32.42	29.27	34.15	32.04	29.15
Perceive-IR Zhang et al. (2025)	34.38	31.74	28.53	34.84	32.50	29.16	34.86	32.55	29.42
VLU-Net Zeng et al. (2025)	34.35	31.72	28.46	-	-	-	34.92	**32.71**	29.61
InterIR(Ours)	**34.59**	**31.87**	**28.54**	**35.01**	**32.83**	**29.59**	**34.95**	32.47	**29.68**

Table 9: Real-world results in the MDIR setting.

Methods	RealRain-1k-L			RTTS			SIDD	
	PSNR ↑	SSIM ↑	LPIPS ↓	FADE ↓	BRISQUE ↓	NIMA ↑	PSNR ↑	SSIM ↑
EfDeRain+ Guo et al. (2025a)	24.05	0.762	0.398	1.393	28.865	4.457	24.07	0.475
PGH^{2}Net Su et al. (2025)	21.65	0.747	0.419	1.352	28.501	<u>4.625</u>	24.00	0.466
ACL Gu et al. (2025)	23.96	0.757	0.408	1.360	28.962	4.319	<u>24.66</u>	0.473
VLU-Net Zeng et al. (2025)	27.22	0.899	0.375	<u>1.242</u>	**23.490**	4.564	24.31	0.471
Defusion Luo et al. (2025a)	<u>27.37</u>	<u>0.903</u>	<u>0.371</u>	1.277	23.520	4.615	24.52	**0.495**
InterIR(Ours)	**27.54**	**0.904**	**0.365**	**1.234**	<u>24.508</u>	**4.631**	**24.67**	<u>0.484</u>

Table 10: Results on the training-unseen degradation type in the MDIR setting .

Methods	UIEB			C60	
	PSNR ↑	SSIM ↑	LPIPS ↓	UCIQE↑	UIQM↑
EfDeRain+ Guo et al. (2025a)	17.66	0.741	0.377	0.480	1.866
ECFNet Gao et al. (2026)	20.49	0.856	0.223	0.539	2.517
ACL Gu et al. (2025)	20.94	0.867	0.201	0.552	2.485
VLU-Net Zeng et al. (2025)	21.58	0.889	<u>0.172</u>	<u>0.556</u>	2.493
Perceive-IR Zhang et al. (2025)	<u>21.77</u>	<u>0.891</u>	**0.169**	**0.557**	<u>2.555</u>
InterIR(Ours)	**21.82**	**0.895**	**0.169**	**0.557**	**2.565**

highest PSNR while maintaining competitive SSIM. These results indicate that the learned restoration representation transfers reasonably well from synthetic or training degradation distributions to real-world corruptions.

Generalization to an Unseen Degradation Type We further evaluate the model trained under the MDIR setting on underwater image restoration, which is not included in the training degradation set. As shown in Table 10, InterIR achieves the highest PSNR and SSIM on UIEB Li et al. (2019), while matching the best LPIPS result. On C60 Li et al. (2019), it achieves the highest UIQM and matches the best UCIQE. Although the improvements over the strongest existing methods are relatively modest, the consistent performance across both datasets suggests that InterIR can transfer to degradation characteristics not observed during training.

A.3.3 Additional Ablation Studies

To further validate the effectiveness of our design, Figure 6 presents t-SNE visualizations of the degradation embeddings from InterIR (ours) and the baseline model DeepSN-Net Deng et al. (2025). Different colors correspond to different degradation types. In our model, the embeddings for each task are more distinctly clustered, demonstrating the ability of image-specific adaptive modeling to capture discriminative degradation contexts that facilitate restoration. In contrast, the baseline model fails to clearly separate different degradation types, which aligns with the unsatisfactory MDIR performance reported in Table 1.

A.3.4 Alternative of Explain Convolution

To further assess the effectiveness of the proposed explain convolution, we substitute it with two representative convolutional variants, CoMo Huang et al. (2024) and ARConv Wang et al. (2025), and conduct comparative experiments under identical settings. As summarized in Table 11, replacing our explain convolution with either alternative results in a noticeable performance drop. Specifically, the model using CoMo decreases the PSNR by 0.38 dB, while the ARConv-based model suffers a reduction of 0.29 dB. Moreover, both alternatives lead to a significant increase in model parameters, with ARConv introducing an additional 10.0M parameters. These results clearly demonstrate that our explain convolution achieves a better balance between restoration accuracy and model compactness, validating its design advantage in effectively extracting image-specific features while maintaining computational efficiency.

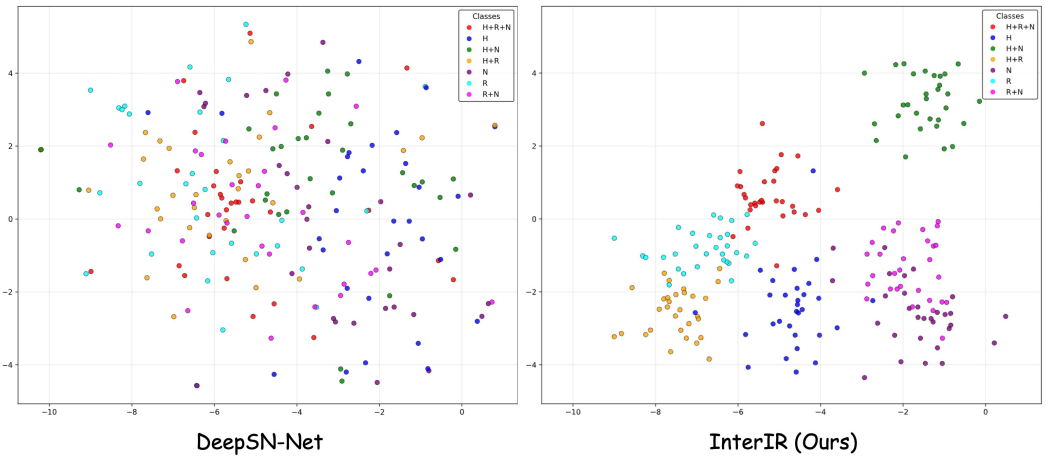

Figure 6: The figure presents t-SNE visualizations of degradation embeddings.

Table 11: Alternative of Explain Convolution.

Net	PSNR	△ PSNR	Params(M)
Explain Convolution	32.51	-	-
CoMo Huang et al. (2024)	32.13	- 0.38	+0.5
ARConv Wang et al. (2025)	32.22	- 0.29	+10.0

Table 12: The evaluation of model computational complexity.

Method	Time(s)	Params(M)	Flops(G)	PSNR	SSIM
VLU-Net Zeng et al. (2025)	0.743	35	143	29.97	0.896
PromptIR Potlapalli et al. (2023)	1.012	33	134	28.31	0.872
Perceive-IR Zhang et al. (2025)	0.682	45	144	31.90	0.921
DeepSN-Net Deng et al. (2025)	0.331	24	86	29.83	0.893
InterIR(Ours)	0.334	27	22	32.51	0.927

A.3.5 RESOURCE EFFICIENT

We further evaluate the computational efficiency of InterIR by comparing its runtime, parameter count, and FLOPs with recent state-of-the-art methods. As shown in Table 12, InterIR not only delivers superior restoration performance but also substantially reduces computational cost. Specifically, it surpasses the previous best method, Perceive-IR Zhang et al. (2025), by 0.61 dB while using only 50.2% of its inference time. Compared with the baseline model DeepSN-Net Deng et al. (2025), although our method introduces additional modules for learning weighting parameters—resulting in slightly higher inference time and parameter count—it avoids a multi-scale architecture, leading to a significant reduction in FLOPs.