Learning to Restore Multi-Degraded Images via Ingredient Decoupling and Task-Aware Path Adaptation

Hu Gao
Shanghai Jiao Tong University
gao_h@sjtu.edu.cn

Xiaoning Lei
CATL
leixn01@outlook.com

Ying Zhang
Beijing Normal University
202331081001@mail.bnu.edu.cn

Xichen Xu
Shanghai Jiao Tong University
neptune_2333@sjtu.edu.cn

Guannan Jiang
CATL
jianggn@catl.com

Lizhuang Ma*
Shanghai Jiao Tong University
lzma@sjtu.edu.cn

Abstract

Image restoration (IR) aims to recover clean images from degraded observations. Despite remarkable progress, most existing methods focus on a single degradation type, whereas real-world images often suffer from multiple coexisting degradations, such as rain, noise, and haze coexisting in a single image, which limits their practical effectiveness. In this paper, we propose an adaptive multi-degradation image restoration network that reconstructs images by leveraging decoupled representations of degradation ingredients to guide path selection. Specifically, we design a degradation ingredient decoupling block (DIDBlock) in the encoder to separate degradation ingredients statistically by integrating spatial and frequency domain information, enhancing the recognition of multiple degradation types and making their feature representations independent. In addition, we present fusion block (FBlock) to integrate degradation information across all levels using learnable matrices. In the decoder, we further introduce a task adaptation block (TABlock) that dynamically activates or fuses functional branches based on the multi-degradation representation, flexibly selecting optimal restoration paths under diverse degradation conditions. The resulting tightly integrated architecture, termed IMDNet, is extensively validated through experiments, showing superior performance on multi-degradation restoration while maintaining strong competitiveness on single-degradation tasks.

1. Introduction

Image restoration (IR) is the process of reconstructing high-quality images from degraded observations. It is a typical ill-posed problem, where multiple potential solutions exist for the same degraded input. Traditional methods [24, 44] address this by introducing task-specific priors, formulating degradation models, and applying inverse operations. Although effective in certain cases, these approaches rely on strong assumptions about degradation factors such as noise or blur kernels. In real-world scenarios, however, degradation processes are uncertain and often unknown, making accurate modeling difficult and leading to poor generalization and unstable performance.

In recent years, deep neural networks [8, 46, 55] have achieved remarkable progress in image restoration. By capturing the statistical properties of natural images, deep learning-based methods can implicitly learn a wide range of priors, leading to superior performance over traditional approaches. Based on the number of degradation types a single model can handle, image restoration methods are typically categorized into task-specific, task-aligned, and all-in-one approaches. Task-specific methods [23, 46] design specialized models for individual tasks, often tailored to the characteristics of the target data. Task-aligned methods [8, 20] aim to build a general network trained sequentially on datasets of different degradation types, which can then be applied to multiple restoration tasks. All-in-one methods [43, 55] share the goal of designing a general network but differ in that the model is trained simultaneously on multiple degradation types, enabling a single network to handle diverse degradations.

However, the above methods [8, 55] focus on single-degradation image restoration (SDIR) and generally assume that each image contains only one type of degradation, such as rain, haze, or noise. In contrast, real-world scenes are highly heterogeneous, dynamic, and uncertain, with image degradations often forming complex combinations, for example, rain, haze, and noise coexisting in a single image. The intricate nature of multi-degradation scenarios hampers the performance of existing methods.

Although some studies [18, 26, 45, 47, 60] have explored

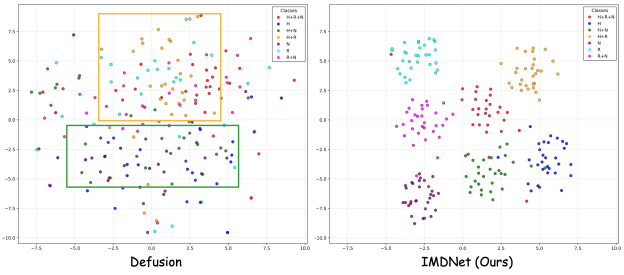

Figure 1. The figure presents t-SNE visualizations of degradation embeddings from IMDNet (ours) and Defusion [38]. Our model exhibit clearer clustering, highlighting the effectiveness of decomposing degradation ingredients into decoupled representations.

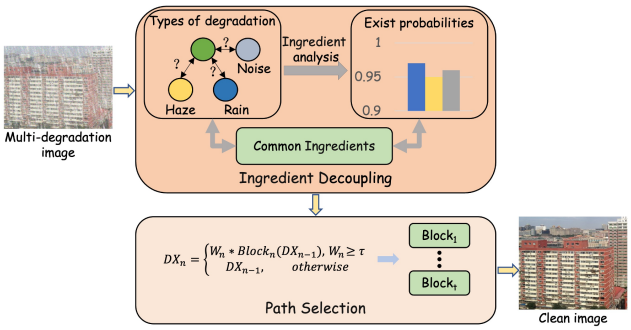

Figure 2. Mechanisms of our method. Our approach achieves multi-degradation image restoration by decomposing degradation ingredients into decoupled representations.

multi-degradation image restoration (MDIR), their performance remains limited, partly because they rely solely on simple attention mechanisms to adaptively select useful features. Ref-IRT [59] addresses this limitation with a multi-stage framework that progressively transfers similar edges and textures from reference images. However, this approach ignores the interaction between different degradation mechanisms. In addition, several All-in-one methods [38, 58] have achieved promising results on multi-degraded images by leveraging large vision models. Nevertheless, as shown in Figure 1, although the retrained Defusion [38] model can roughly distinguish degradation ingredients (highlighted by the yellow and green boxes), the separation is still weak. It struggles to learn intrinsic differences among degradation types, instead relying on its strong representational power to fit mappings from low- to high-quality images.

Given these considerations, a natural question arises: is it possible to design a network capable of adaptively restoring images affected by multiple degradations according to variations in their constituent ingredients? To address this challenge, we propose IMDNet (Figure 2), an adaptive MDIR network that reconstructs high-quality images by leveraging decoupled representations of degradation ingredients to guide path selection. In the encoder, we design a degradation ingredient decoupling block (DIDBlock) statistically separates degradation ingredients using spatial and frequency information, enhancing recognition of multiple degradation types while keeping their feature representations independent. To further consolidate information, we incorporate a fusion block (FBlock) that adaptively integrates degradation features across all hierarchical levels using learnable matrices, facilitating comprehensive representation of complex degradations. In the decoder, we introduce a task adaptation block (TABlock) to dynamically activate or fuse functional branches according to the multi-degradation representation, allowing the network to flexibly select the optimal restoration path under varying degradation conditions. Figure 1 shows that IMDNet leverages the "degradation ingredient $\rightarrow$ path selection" mechanism for accurate identification of different degradation types.

The main contributions of this work are:

1. We propose IMDNet for MDIR, leveraging decoupled representations of degradation ingredients to guide path selection. Extensive experiments show its superior performance on both MDIR and SDIR.
2. We design a degradation ingredient decoupling block (DIDBlock) to decompose degradation ingredients statistically using spatial and frequency information, improving recognition of multiple degradation types while maintaining independent feature representations.
3. We introduce a task adaptation block (TABlock) that dynamically activates or fuses branches based on the multi-degradation representation, flexibly selecting optimal restoration paths under varying conditions.
4. We incorporate a fusion block (FBlock) to combine degradation information from all levels using learnable matrices.

2. Related Works

2.1. Single-degradation Image Restoration

SDIR focuses on restoring high-quality images degraded by a single specific type of corruption. Traditional approaches [24, 44] typically address this ill-posed problem by introducing handcrafted priors to constrain the solution space. However, these priors rely heavily on expert experience and often suffer from limited adaptability and poor generalization to diverse degradation scenarios.

With the rapid advancement of deep learning in high-level vision tasks, numerous data-driven methods [8, 17, 20, 40, 55] have been developed for image restoration. Depending on the number of degradation types a single model can handle, these methods are generally classified into task-specific, task-aligned, and all-in-one approaches.

Task-specific methods [23, 46] design dedicated models for each degradation type, often customized to the characteristics of the corresponding dataset. ALGNet [16] in-

troduces a local feature extraction module to mitigate local pixel forgetting caused by excessive hidden units, achieving efficient and precise deblurring. XYScanNet [35] adopts an alternating intra- and inter-slice scanning strategy to better capture spatial dependencies. EfDeRain+ [23] redefines deraining as a predictive filtering task, avoiding the need for complex rain modeling assumptions. UPID-EDM [49] employs a diffusion-based unpaired learning framework, leveraging vision-language priors and energy-guided sampling to retain structural content while effectively removing degradations. PGH^{2}Net [46] reveals how bright/dark channel priors and histogram equalization jointly guide hierarchical feature learning for dehazing. Diff-Unmix [54] performs self-supervised denoising through spectral decomposition and conditional diffusion modeling.

Task-aligned methods [8, 20] seek to develop unified networks capable of handling multiple degradation types by sequentially training on different restoration datasets. MambaIR [21] introduces a four-directional unfolding strategy combined with channel attention to enhance spatial representation, while MambaIRV2 [22] further integrates non-causal modeling inspired by ViTs to strengthen state space expressiveness. FSNet [8] designs dynamic frequency selection modules to adaptively extract the most informative spectral components for restoration. ACL [20] leverages the mathematical equivalence between linear attention and SSM in Mamba, unlocking the potential of linear attention for image restoration. MHNet [19] adopts a mixed hierarchical design to produce restorations with richer textures and finer structural details.

All-in-one methods [43, 55] aim to build a unified network capable of handling multiple degradation types simultaneously, allowing a single model to manage diverse corruption scenarios. PromptIR [43] proposes a prompt-driven restoration framework that recovers clean images directly from inputs using lightweight, plug-and-play prompt modules. NDR [51] learns neural degradation representations to capture shared characteristics across different degradation types. VLU-Net [55] leverages vision-language model features to automatically identify degradation-aware keys, removing the need for manually defined degradation categories. AutoDIR [28] combines VLM guidance with latent diffusion to detect unknown degradations and perform structure-consistent restoration. AdaIR [12] separates degradation from clean image content by jointly exploiting spatial and frequency domain.

Despite their effectiveness in SDIR, these methods encounter difficulties in complex MDIR, where intertwined degradations severely impair restoration accuracy. Although recent all-in-one approaches [38, 58] have leveraged large vision models to enhance performance on multi-degraded images, their ability to distinguish degradation components remains limited. Instead of learning the intrinsic differences among various degradation types, these models primarily depend on strong representational power to directly map low-quality inputs to high-quality outputs.

2.2. Multi-degradation Image Restoration

MDIR aims to restore a single image affected by multiple degradation types into a clean image. AASO [47] performs multiple restoration operations in parallel and uses attention to weight each operation, selecting the most suitable restoration strategy. However, the heterogeneity of features generated by different operations limits its performance on multi-degradation images. OWAN [26] employs high-order tensor fusion to coordinate heterogeneous features and leverage high-order statistical information, but its complex network struggles to restore high-frequency details. DIGNet [45] considers sequential and spatially varying distortions, while MEPSNet [29] uses a mixture-of-experts with parameter sharing to address varying distortions across image regions.

Despite these efforts, performance remains limited, partly because they depend solely on simple attention mechanisms to adaptively select features. To overcome this, RefIRT [59] introduces a three-stage restoration framework: the first stage estimates residuals in a coarse-to-fine manner, while the second and third stages progressively transfer details from reference images. FDTANet [18] performs adaptive restoration in the frequency domain based on component differences. However, both methods overlook the interactions between different degradation mechanisms. In this paper, we first separate degradation ingredients by integrating spatial and frequency information, improving recognition of multiple degradation types and making their features more independent. Then, based on multi-degradation representations, we dynamically activate or fuse branches to flexibly select optimal restoration paths, achieving effective multi-degradation image restoration.

3. Method

In this section, we first outline the overall IMDNet framework, then detail the proposed degradation ingredient decoupling block (DIDBlock) and task adaptation block (TABlock), followed by the loss function.

3.1. Overall Pipeline

Our proposed IMDNet, illustrated in Figure 3, adopts an encoder–decoder architecture. Each encoder level comprises [4, 4, 4, 8] DIDBlocks, while the middle block includes 8 DIDBlocks. Each decoder level consists of [2, 2, 2, 2] TABlocks. Given a degraded image $\mathbf{I} \in \mathbb{R}^{H\times W\times 3}$, IMDNet first applies convolution to extract shallow features $\mathbf{F_0} \in \mathbb{R}^{H\times W\times C}$ (H, W, and C denote the height, width, and number of channels, respectively). These shallow features are then processed through a four-level encoder and a

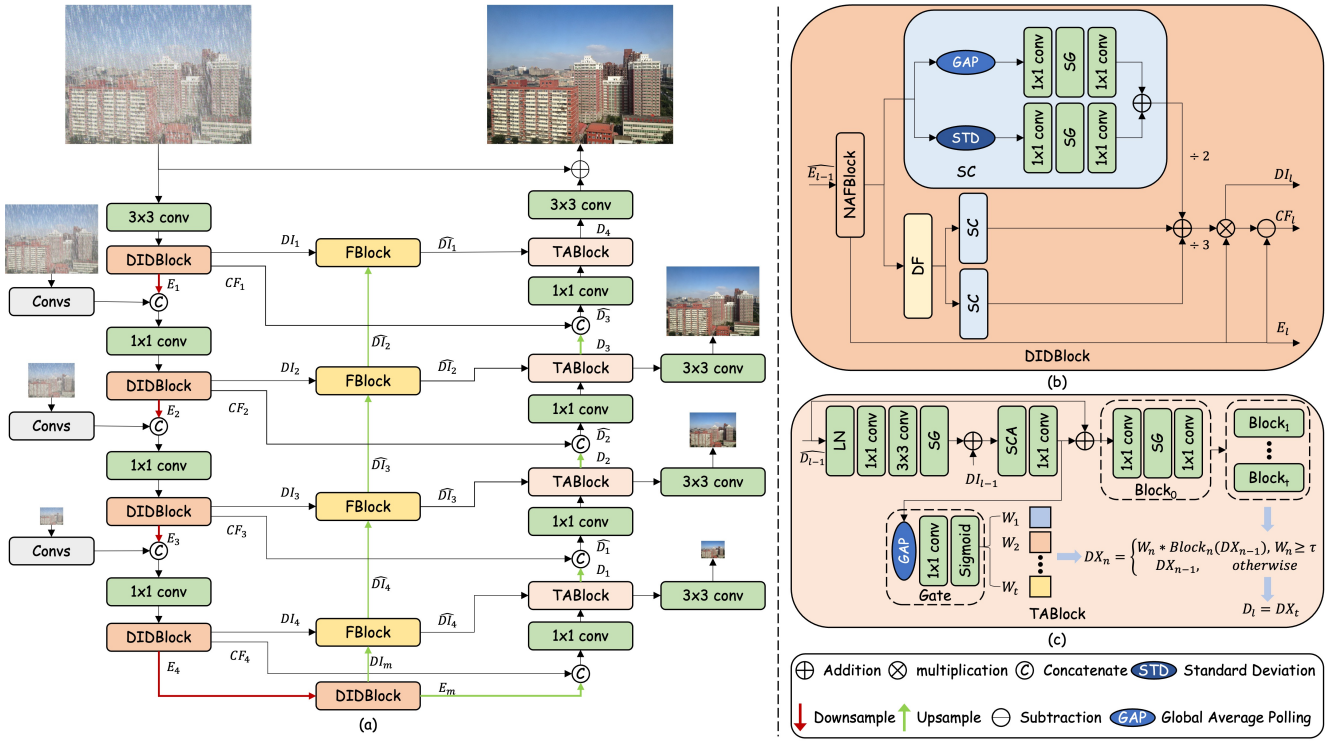

Figure 3. (a) The overall architecture of the proposed IMDNet. (b) The structure of degradation ingredient decoupling block (DIDBlock). (c) The structure of task adaptation block (TABlock).

middle block. Each encoder level and the middle block output three types of features: the encoded feature E for the next-level encoder, the decoupled degradation information DI, and the skip feature CF used to facilitate the reconstruction. To enhance training, IMDNet adopts multi-input and multi-output mechanisms. Low-resolution degraded images are introduced into the main path through a Convs consisting of multiple convolutions and concatenations, followed by a channel-adjustment convolution. The deepest features are subsequently passed into a four-scale decoder that progressively restores spatial resolution. Throughout this process, the degradation information DI is integrated across all levels using learnable matrices within the fusion block (FBlock). Finally, refined features are used to generate a 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: $\hat{\mathbf{I}} = \mathbf{I_R} + \mathbf{I}$.

3.2. Degradation Ingredient Decoupling Block

The features of degraded images entering the encoder naturally contain degradation information, which motivates us to explore whether we can decouple degraded features from clean image features during encoding. The clean features are then used for skip connections between the encoder and decoder, while the degraded features estimate the probabilities of various degradation factors and guide the decoder to perform adaptive multi-degradation image restoration.

To verify this idea, as shown in Figure 3(b), we design a degradation ingredient decoupling block (DIDBlock). First, we use the NAFBlock [3] to extract spatial-domain features. Considering the distinct characteristics of different degradation components across spatial and frequency domains, we use a learnable dynamic filter (DF) to separate frequency components and obtain high- and low-frequency features. Finally, we adopt a joint frequency–spatial analysis strategy based on statistical coefficients (SC) to achieve degradation ingredient decoupling representation, effectively isolating and characterizing degradation information from both spatial and frequency features. Specifically, for the feature E_{l-1} output by the encoder at level $l-1$, we first concatenate it with the low-resolution degraded image features and adjust its dimension using a 1×1 convolution to obtain $\widehat{E_{l-1}}$. We then feed $\widehat{E_{l-1}}$ into the NAFBlock to generate the spatial-domain feature E_l for the level-l encoder:

$$E_l = NAFBlock(f^c_{1\times 1}([E_{l-1}, Convs(I_l)])) \quad (1)$$

where $f^c_{1\times 1}$ represents the 1×1 convolution, and $[\cdot]$ is the concatenation operation. Next, we apply the dynamic filter to decompose E_l into high-frequency and low-frequency components:

$$F_H, F_L = DF(E_l) \quad (2)$$

We then perform statistical analysis of the spatial fea-

tures E_l and frequency domain features F_H, F_L to isolate the degradation information:

$$DI_l = \frac{SC(F_H) \oplus SC(F_L) \oplus SC(E_l)}{6} \otimes E_l$$
$$SC(\cdot) = f^c_{1\times1}(SG(f^c_{1\times1}(GAP(\cdot)))) \oplus f^c_{1\times1}(SG(f^c_{1\times1}(STD(\cdot)))) \tag{3}$$

where $GAP(\cdot)$ and $STD(\cdot)$ represent global average pooling and standard deviation pooling, respectively. $SG(\cdot)$ indicates the simple gate, serving as a replacement for the nonlinear activation function. Given an input $X_f \in R^{H\times W\times C}$, SG splits it into two features $X^1_f, X^2_f \in R^{H\times W\times \frac{C}{2}}$ and calculates X^1_f, X^2_f using a linear gate., Finally, we obtain the clean features CF_l by subtracting the degradation information from the encoded features:

$$CF_l = E_l \ominus DI_l \tag{4}$$

Through this process, we obtain three key outputs: the feature E_l for the next-level encoder, the feature CF_l for skip connections, and the degradation information DI_l to guide adaptive multi-degradation restoration in the decoder. This design offers several advantages. By leveraging the statistical properties of input features, the generated degradation information captures both the image's content correlation and shared ingredient characteristics, improving representation consistency and knowledge transfer across different degradation types. Moreover, unlike previous methods that use simple addition or concatenation for skip connections, our approach effectively suppresses degraded information from the encoder and minimizes interference from implicit noise.

3.3. Task Adaptation Block

To enable the model to dynamically activate or fuse multiple functional branches based on the decoupled degradation information, flexibly select the optimal restoration path, and adaptively handle various combinations of degraded images, we design a task adaptive block (TABlock) in the decoder. As shown in Figure 3(c), TABlock first captures contextual information features and then fuses the degradation information. Notable, the first branch serves as a general branch, responsible for processing common components shared across all degradation types and ensuring stable baseline restoration performance. To achieve dynamic branch fusion guided by task relevance, we employ a gated network that adaptively generates activation weight vectors for each branch. To further enhance computational efficiency, we introduce a branch sparsity constraint to achieve sparse activation, retaining only branches with significant weights for computation. Finally, the outputs of all active branches are weighted according to their activation weights to produce the final restored features.

Specifically, for the feature $\widehat{D_{l-1}}$, obtained by fusing the decoder output and skip-connection features at level $l-1$, we first extract contextual information features and then inject the degradation information $\widehat{DI_{l-1}}$ to obtain the feature DC_{l-1}:

$$DC_{l-1} = f^c_{1\times1}(SCA(SG(f^{dwc}_{3\times3}(f^c_{1\times1}(LN(\widehat{D_{l-1}})))\oplus\widehat{DI_{l-1}}))) \tag{5}$$

where $f^{dwc}_{3\times3}$ denotes the 3×3 depth-wise convolution, $SCA(\cdot)$ is the simplified channel attention [3], and $LN(\cdot)$ represents layer normalization. Notably, $\widehat{DI_{l-1}}$ is obtained by aggregating degradation information from all levels through the fusion block (FBlock) as:

$$\widehat{DI_{l-1}} = DI_l \oplus W\widehat{DI_l} \tag{6}$$

where W denotes learnable parameters that are directly optimized via backpropagation and initialized to 1. Next, for the first general branch, the computation is defined as:

$$DX_0 = Block_0(DC_{l-1} \oplus \widehat{D_{l-1}})$$
$$Block(\cdot) = f^c_{1\times1}(SG(f^c_{1\times1}(\cdot))). \tag{7}$$

Subsequently, we feed DC_{l-1} into a gated network to generate the activation weight vector for each branch:

$$W_n = Sigmoid(f^c_{1\times1}(GAP(DC_{l-1}))). \tag{8}$$

To further enhance efficiency, we apply a branch sparsity constraint to retain only branches with significant weights for computation:

$$DX_n = \begin{cases} W_n * Block_n(DX_{n-1}), W_n \geq \tau, \\ DX_{n-1}, otherwise. \end{cases} \tag{9}$$

where $Block_n(\cdot)$ represents the n-th functional branch, and τ is a threshold empirically set to 0.2.

3.4. Loss Function

Consistent with prior works, we optimize IMDNet in both spatial and frequency domains. To facilitate accurate decoupling in the encoder, the skip-connection feature CF is encouraged to contain minimal degradation information, while the branch-selection feature DI is designed to capture primarily degradation information. We enforce this separation using a cosine similarity loss, maximizing the distance between CF and DI. The overall loss function is defined

Table 1. Quantitative results of different models conducted on various combinations of degradation types, 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
Restormer [53]	23.52/0.792	25.45/0.836	26.73/0.863	25.31/0.815	29.82/0.931	28.05/0.873	27.44/0.837	26.62/0.850
AirNet [32]	27.41/0.812	28.28/0.889	27.59/0.861	26.98/0.821	30.22/0.959	28.27/0.881	27.25/0.847	28.00/0.867
U^{2}Former [13]	25.07/0.803	26.23/0.856	26.79/0.864	26.02/0.816	29.95/0.933	28.50/0.876	27.12/0.831	27.09/0.854
PromptIR [43]	27.54/0.819	28.43/0.901	27.99/0.871	27.05/0.822	30.46/0.956	28.78/0.885	27.92/0.851	28.31/0.872
VLU-Net [55]	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 [18]	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
Ref-IRT [59]	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 [58]	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 [38]	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
IMDNet(Ours)	**31.19/0.910**	**33.21/0.973**	29.01/**0.895**	32.15/0.918	31.73/0.972	40.14/**0.988**	**33.86/0.932**	**33.04/0.941**

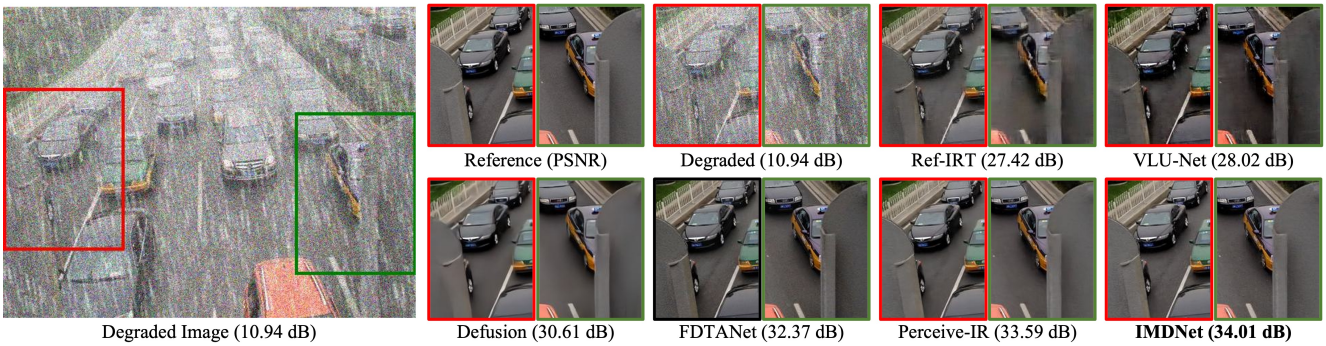

Figure 4. Qualitative results under the MDIR experimental setup. Compared to other methods, our IMDNet effectively reduces color distortion and produces images that are visually closer to the ground truth.

as:

$$
\begin{aligned}
L &= \sum_{i=1}^{4}(L_c(\hat{I}_i, \overline{I}_i) + \delta L_e(\hat{I}_i, \overline{I}_i) \\
&\quad + \lambda L_f(\hat{I}_i, \overline{I}_i) + \gamma L_d(CF_i, DI_i)) \\
L_c &= \sqrt{||\hat{I}_i - \overline{I}_i||^2 + \epsilon^2} \\
L_e &= \sqrt{||\triangle\hat{I}_i - \triangle\overline{I}_i||^2 + \epsilon^2} \\
L_f &= ||\mathcal{F}(\hat{I}_i) - \mathcal{F}(\overline{I}_i)||_1 \\
L_d &= \frac{CF_i \cdot DI_i}{\|CF_i\|_2 \, \|DI_i\|_2}
\end{aligned}
\tag{10}
$$

where i indexes input/output images at different scales, $\overline{I}_i$ denotes the target images and L_c is the Charbonnier loss with constant ϵ empirically set to 0.001 for all the experiments. L_e is the edge loss, where $\triangle$ represents the Laplacian operator. L_f denotes the frequency domains loss, where $\mathcal{F}$ represents fast Fourier transform. L_d is the decoupling loss. The weights λ, δ and γ balance the loss terms and are set to 0.1, 0.05, and 0.001, respectively, following [8, 52].

4. Experiments

In this section, we first describe the experimental setup, followed by qualitative and quantitative comparison results. Finally, we present ablation studies to validate the effectiveness of our approach. **Due to page limits, more experiments we show in the supplementary material.**

4.1. Experimental Setup

We conducted experiments under both MDIR and SDIR settings. **Datasets.** For MDIR, we use the dataset proposed by [18]. For SDIR, the model is trained on clean-rain image pairs from multiple datasets [15, 33, 50, 57] and evaluated on various test sets, including Rain100H [50], Rain100L [50], Test100 [57], and Test1200 [56] for image deraining. For image denoising, images are collected from DIV2K [1], Flickr2K [34], BSD500 [2], and WED [39], with additive white Gaussian noise applied at levels from 0 to 50. The model is evaluated on CBSD68 [41], Urban100 [25], and Kodak24 [14]. For image dehazing, we use images from the RESIDE dataset [31] and evaluate on its SOTS subset [31].

Training details. We train our models using the Adam optimizer [30] with parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The initial learning rate is set to 2×10^{-4} and gradually de-

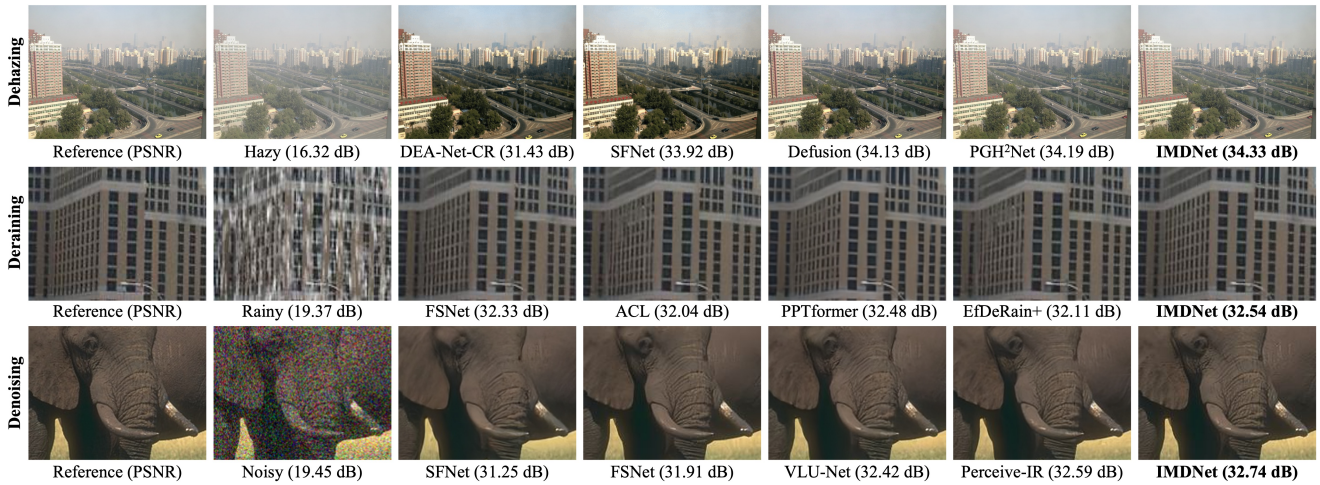

Figure 5. Qualitative results under the SDIR experimental setup. Our IMDNet recovers finer details in the reconstructed images.

cays to 1×10^{-7} following a cosine annealing schedule [37]. Training patches of size 256×256 are sampled with a batch size of 32 for 4×10^{5} iterations. Data augmentation includes horizontal and vertical flips. For fair comparisons, all deep learning-based methods are fine-tuned or retrained using the parameter settings reported in their respective papers.

4.2. Experimental Results

4.2.1. MDIR Setting

We evaluate our IMDNet across seven combinations of degradation types, encompassing all permutations of haze, rain, and noise. Table 1 reports the quantitative comparison results. On average across all tasks, our method achieves a 1.03 dB improvement over the second-best approach, Defusion [38]. Compared with previous MDIR methods, Ref-IRT [59] and FDTANet [18], our model shows substantial gains of 2.60 dB and 3.44 dB, respectively.

To further validate the adaptability of our approach to various degradation types, we conduct experiments under multiple degradation combinations using the same training dataset. The results demonstrate that IMDNet consistently achieves near-optimal or superior performance across diverse degradation scenarios. Specifically, our model outperforms competing methods by 0.97 dB on the H+R+N combination compared to Perceive-IR [58], by 0.88 dB on H+R compared to Ref-IRT [59], and by 1.14 dB on N compared to Defusion [38]. Although IMDNet does not achieve the top performance on H+N, R+N, H, and R, the differences are marginal, within 0.06 dB. As shown in Figure 4, our model produces restored images that are sharper and visually closer to the ground truth than others.

4.2.2. SDIR Setting

To verify that the proposed IMDNet is effective not only for MDIR but also performs well on SDIR, we conduct experiments on three representative image restoration tasks: image dehazing, image deraining, and image denoising. Figure 5 illustrates that the images restored by our IMDNet effectively reduce color distortion compared to other state-of-the-art methods. Moreover, our approach is able to reconstruct finer and sharper details in the degraded regions.

Table 2. Image dehazing results.

Methods	SOTS-Indoor		SOTS-Outdoor	
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
SFNet [11]	41.24	**0.996**	<u>40.05</u>	**0.996**
IRXNext [10]	41.21	**0.996**	39.18	**0.996**
DEA-Net-CR [7]	41.31	<u>0.995</u>	36.59	0.990
Defusion [38]	41.65	<u>0.995</u>	37.41	<u>0.993</u>
PGH²Net [46]	<u>41.70</u>	**0.996**	37.52	0.989
IMDNet(Ours)	**42.17**	**0.996**	**40.64**	0.992

Image Dehazing. We report the quantitative results of different image dehazing methods in Table 2. Overall, our IMDNet consistently achieves superior performance across both indoor and outdoor scenarios. Specifically, on the indoor dataset SOT-indoor, IMDNet surpasses the previous best-performing method PGH²Net [46] by 0.47 dB in PSNR. On the outdoor dataset SOT-outdoor, it further achieves a 0.59 dB improvement over the previous state-of-the-art method SFNet [11].

Image Deraining. Following the protocol in prior work [19], we evaluate PSNR and SSIM metrics on the Y channel of the YCbCr color space for the image deraining task. As shown in Table 3, our method consistently achieves comparable or superior results to existing approaches across all four benchmark datasets. In particular, IMDNet delivers an average PSNR improvement of 0.29 dB over the best-performing method EfDeRain+ [23].

Table 3. Image deraining results.

Methods	Test100 [57]		Test1200 [27]		Rain100H [50]		Rain100L [50]		Average	
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
FSNet [8]	31.05	0.919	33.08	0.916	31.77	0.906	38.00	0.972	33.48	0.928
MHNet [19]	31.25	0.901	33.45	0.925	31.08	0.899	40.04	**0.985**	33.96	0.928
PPTformer [48]	31.48	**0.922**	33.39	0.911	31.77	0.907	39.33	0.983	33.99	0.931
ACL [20]	31.51	0.914	33.27	0.928	32.22	0.920	39.18	0.983	34.05	0.936
EfDeRain+ [23]	31.10	0.911	33.12	0.925	**34.57**	**0.957**	39.03	0.972	34.46	**0.941**
IMDNet(Ours)	**32.02**	**0.922**	**34.46**	**0.939**	32.42	0.914	**40.10**	**0.985**	**34.75**	0.940

Table 4. Image denoising results.

Methods	CBSD68 [41]			Kodak24 [14]			Urban100 [25]		
	15	25	50	15	25	50	15	25	50
SFNet [11]	34.09	31.49	28.02	34.93	32.42	29.25	34.19	32.01	29.03
FSNet [8]	34.11	31.51	28.01	34.95	32.42	29.27	34.15	32.04	29.15
Perceive-IR [58]	34.38	31.74	28.53	34.84	32.50	29.16	34.86	32.55	29.42
VLU-Net [55]	34.35	31.72	28.46	-	-	-	**34.92**	**32.71**	**29.61**
IMDNet(Ours)	**34.77**	**32.14**	**29.17**	**35.55**	**33.01**	**30.09**	34.87	32.34	29.39

Table 5. Ablation study on individual components of IMDNet.

Method	PSNR	△ PSNR
Baseline	27.01	-
replace with DIDBlock	27.64	+0.63
replace with TABlock	28.55	+1.54
replace with DIDBlock and TABlock	30.91	+3.90
replace with DIDBlock and TABlock + FBlock	31.19	+4.18

Moreover, on the Test100 dataset [57], our model achieves a notable 0.52 dB gain compared to the previous leading method ACL [20], highlighting the strong deraining capability of our approach.

Image Denoising. Table 4 presents the performance of IMDNet model under different noise levels (15, 25, and 50). Our approach consistently delivers strong results across multiple datasets and noise intensities. In particular, at the challenging noise level of 50 on the CBSD68 dataset [41], IMDNet surpasses Perceive-IR [58] by 0.64 dB. Similarly, for the same noise level on the Kodak24 dataset [14], our method achieves a 0.82 dB improvement over the previous best-performing model FSNet [8]. Although IMDNet does not reach the top performance on the Urban100 dataset [25], the performance gap remains minimal.

4.3. Ablation Studies

We conduct ablation studies under the MDIR setting to demonstrate the effectiveness of our modules. We use the four-scale NAFNet [3] as the baseline network and perform a step-by-step ablation study by successively integrating the proposed modules.

4.3.1. Effectiveness of Each Module

Table 5 shows that the baseline model achieves a PSNR of 27.01 dB. When the encoder incorporates the proposed DIDBlock, the model effectively captures both spatial and frequency domain features, while reducing degradation information within the skip-connected features. This leads to more accurate information propagation and a 0.63 dB performance gain. However, the DIDBlock alone still lacks sufficient adaptability to handle diverse degradation types. When only the decoder is replaced with the TABlock, the model gains adaptive capability to distinguish and process different degradations. Yet, because its input primarily comes from the encoder output without fully decoupled degradation information, its recognition accuracy remains limited, resulting in only a 1.54 dB improvement. By jointly employing both DIDBlock and TABlock, the model achieves more precise degradation decoupling and information propagation through skip connections, leading to a significant 3.90 dB performance increase. Finally, with the integration of the FBlock to fuse degradation information across multiple scales, the model's performance is further enhanced, reaching 31.19 dB.

Table 6. Plug-and-play ablation experiments.

Method	PSNR	△ PSNR
PromptIR [43]	27.54	-
△ PromptIR [43]	30.88	+3.34
VLU-Net [55]	29.35	-
△ VLU-Net [55]	31.13	+1.78

4.3.2. Generality of the Overall Idea

To further validate the generality of our overall idea of decoupling degradation ingredient representations to guide path selection, we modify the encoder and decoder of existing methods. In the encoder, instead of changing its structure, we append a degradation information decoupling module after the original feature extraction stage. In the decoder, we adjust its feedforward network following the TABlock design. As shown in Table 6, incorporating our proposed idea leads to significant performance gains, with PromptIR [43] and VLU-Net [55] achieving improvements of 3.34 dB and 1.78 dB, respectively.

5. Conclusion

In this work, we propose IMDNet, an adaptive network for multi-degradation image restoration that addresses the challenges of images with coexisting degradations. By leveraging decoupled representations of degradation components, IMDNet effectively guides the selection of optimal restoration paths. The encoder's degradation ingredient decoupling block (DIDBlock) separates degradation factors across spatial and frequency domains, enhancing recognition and ensuring independent feature representations. The fusion block (FBlock) integrates multi-level degradation information, while the decoder's task adaptation block (TABlock) dynamically activates or fuses functional branches based on the multi-degradation representation. Extensive experiments demonstrate that IMDNet excels in MDIR while remaining highly competitive on SDIR.

References

[1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1122–1131, 2017. 6, 1

[2] Pablo Arbeláez, Michael Maire, Charless Fowlkes, and Jitendra Malik. Contour detection and hierarchical image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(5):898–916, 2011. 6, 1

[3] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. *ECCV*, 2022. 4, 5, 8, 2

[4] 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)*, pages 1–5, 2023. 2

[5] Wei-Ting Chen, Hao-Yu Fang, Jian-Jiun Ding, Cheng-Che Tsai, and Sy-Yen Kuo. Jstasr: Joint size and transparency-aware snow removal algorithm based on modified partial convolution and veiling effect removal. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXI 16*, pages 754–770. Springer, 2020. 1, 2

[6] 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 *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4196–4205, 2021. 1, 2

[7] 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. 7

[8] Yuning Cui, Wenqi Ren, Xiaochun Cao, and Alois Knoll. Image restoration via frequency selection. *TPAMI*, pages 1–16, 2023. 1, 2, 3, 6, 8

[9] 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*, pages 13001–13011, 2023. 2

[10] 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*, 2023. 7, 2

[11] 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*, 2023. 7, 8

[12] 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. 3

[13] Xin Feng, Haobo Ji, Wenjie Pei, Jinxing Li, Guangming Lu, and David Zhang. U2-former: Nested u-shaped transformer for image restoration via multi-view contrastive learning. *TCSVT*, pages 1–1, 2023. 6

[14] Rich Franzen. Kodak lossless true color image suite. *source: http://r0k. us/graphics/kodak*, 4(2):9, 1999. 6, 8, 1

[15] 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)*, pages 1715–1723, 2017. 6, 1

[16] Hu Gao, Bowen Ma, Ying Zhang, Jingfan Yang, Jing Yang, and Depeng Dang. Learning enriched features via selective state spaces model for efficient image deblurring. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 710–718, 2024. 2

[17] 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*, pages 5566–5574, 2024. 2

[18] 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, 2025. 1, 3, 6, 7

[19] Hu Gao, Ying Zhang, Jing Yang, and Depeng Dang. Mixed hierarchy network for image restoration. *Pattern Recognition*, 161:111313, 2025. 3, 7, 8

[20] 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*, pages 17913–17923, 2025. 1, 2, 3, 8

[21] 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 *European Conference on Computer Vision*, pages 222–241, 2024. 3

[22] Hang Guo, Yong Guo, Yaohua Zha, Yulun Zhang, Wenbo Li, Tao Dai, Shu-Tao Xia, and Yawei Li. Mambairv2: Attentive state space restoration. In *Proceedings of the Computer*

Vision and Pattern Recognition Conference, pages 28124–28133, 2025. 3

[23] 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, 2025. 1, 2, 3, 7, 8

[24] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *TPAMI*, 2011. 1, 2

[25] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5197–5206, 2015. 6, 8, 1

[26] 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)*, pages 1–8, 2021. 1, 3

[27] 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. 8

[28] Yitong Jiang, Zhaoyang Zhang, Tianfan Xue, and Jinwei Gu. Autodir: Automatic all-in-one image restoration with latent diffusion. In *Computer Vision – ECCV 2024: 18th European Conference, Milan, Italy, September 29–October 4, 2024, Proceedings, Part XL*, page 340–359, 2024. 3

[29] 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)*, 2020. 3

[30] D. Kingma and J. Ba. Adam: A method for stochastic optimization. *Computer Science*, 2014. 6

[31] 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, 2018. 6, 1

[32] Boyun Li, Xiao Liu, Peng Hu, Zhongqin Wu, Jiancheng Lv, and Xi Peng. All-in-one image restoration for unknown corruption. In *CVPR*, pages 17452–17462, 2022. 6

[33] 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)*, pages 2736–2744, 2016. 6, 1

[34] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 136–144, 2017. 6, 1

[35] 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)*, pages 779–789, 2025. 3

[36] Yun-Fu Liu, Da-Wei Jaw, Shih-Chia Huang, and Jenq-Neng Hwang. Desnownet: Context-aware deep network for snow removal. *IEEE Transactions on Image Processing*, 27(6): 3064–3073, 2018. 1

[37] I. Loshchilov and F. Hutter. Sgdr: Stochastic gradient descent with warm restarts. 2016. 7

[38] 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)*, pages 12764–12777, 2025. 2, 3, 6, 7

[39] Kede Ma, Zhengfang Duanmu, Qingbo Wu, Zhou Wang, Hongwei Yong, Hongliang Li, and Lei Zhang. Waterloo exploration database: New challenges for image quality assessment models. *IEEE Transactions on Image Processing*, 26 (2):1004–1016, 2016. 6, 1

[40] Xintian Mao, Qingli Li, and Yan Wang. Adarevd: Adaptive patch exiting reversible decoder pushes the limit of image deblurring. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 25681–25690, 2024. 2

[41] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, pages 416–423 vol.2, 2001. 6, 8, 1

[42] 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*, pages 416–423. IEEE, 2001. 1

[43] 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. 1, 3, 6, 8

[44] 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. 1, 2

[45] 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)*, pages 536–545, 2022. 1, 3

[46] 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*, pages 7042–7050, 2025. 1, 2, 3, 7

[47] 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. 1, 3

[48] Cong Wang, Jinshan Pan, Liyan Wang, and Wei Wang. Intra and inter parser-prompted transformers for effective image restoration. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 7609–7618, 2025. 8

[49] Yuanbo Wen, Tao Gao, and Ting Chen. Unpaired photo-realistic image deraining with energy-informed diffusion

model. In *Proceedings of the 32nd ACM International Conference on Multimedia*, page 360–369, 2024. 3

[50] 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*, pages 1685–1694, 2016. 6, 8, 1

[51] Mingde Yao, Ruikang Xu, Yuanshen Guan, Jie Huang, and Zhiwei Xiong. Neural degradation representation learning for all-in-one image restoration. *IEEE Transactions on Image Processing*, 33:5408–5423, 2024. 3

[52] 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. 6, 1

[53] 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*, 2022. 6, 1

[54] Haijin Zeng, Jiezhang Cao, Kai Zhang, Yongyong Chen, Hiep Luong, and Wilfried Philips. Unmixing diffusion for self-supervised hyperspectral image denoising. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 27820–27830, 2024. 3

[55] 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)*, pages 7524–7533, 2025. 1, 2, 3, 6, 8

[56] He Zhang and Vishal M. Patel. Density-aware single image de-raining using a multi-stream dense network. *CVPR*, pages 695–704, 2018. 6, 1

[57] He Zhang, Vishwanath A. Sindagi, and Vishal M. Patel. Image de-raining using a conditional generative adversarial network. *TCSVT*, 30:3943–3956, 2017. 6, 8, 1

[58] 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*, pages 1–1, 2025. 2, 3, 6, 7, 8

[59] 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. 2, 3, 6, 7

[60] Jingyuan Zhou, Chaktou Leong, Minyi Lin, Wantong Liao, and Congduan Li. Task adaptive network for image restoration with combined degradation factors. In *WACVW*, pages 1–8, 2022. 1

Learning to Restore Multi-Degraded Images via Ingredient Decoupling and Task-Aware Path Adaptation

Supplementary Material

6. Overview

The Appendix is composed of:

7. Dataset

7.1. MDIR

Following [18], we conduct comparative experiments on a dataset containing various combinations of degradations, including haze, rain, and noise. The dataset consists of 18,000 natural images collected from several public image restoration datasets [31, 42, 50, 57]. Among them, 13,000 images are randomly selected for training, where haze, rain, and noise degradations are synthetically introduced with intensity levels of [0,150], [0,300], and [0,50], respectively. Among the remaining, 500 images are randomly selected for testing. As summarized in Table 7, the test set covers seven distinct degradation combinations, encompassing all permutations of haze, rain, and noise. For example, in the "haze + rain + noise" setting, the same degradation processes as in training are applied, while in the "haze + rain" case, the random noise addition is omitted, with other settings kept identical. The remaining configurations are constructed in a similar fashion.

Table 7. 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]

7.2. SDIR

7.2.1. Image Deraining

Following the experimental setups of recent state-of-the-art methods for image deraining [8, 52], we train our model using 13,712 clean-rain image pairs collected from multiple datasets [15, 33, 50, 57]. Using the trained IMDNet, we conduct evaluations on various test sets, including Rain100H [50], Rain100L [50], Test100 [57], and Test1200 [56].

7.2.2. Image Denoising

In line with the methodology of [53], we train a single IMDNet model capable of handling various noise levels on a composite dataset. This dataset comprises 800 images from DIV2K [1], 2,650 images from Flickr2K [34], 400 images from BSD500 [2], and 4,744 images from WED [39]. The noisy images are generated by adding additive white Gaussian noise with a random noise level σ chosen from the set [15, 25, 50] to the clean images. Testing is conducted on the CBSD68 [41], Urban100 [25], and Kodak24 [14] benchmark datasets.

7.2.3. Image Dehazing

We assess the performance of our IMDNet on the daytime datasets encompass synthetic data from RESIDE [31]. The RESIDE dataset contains two training subsets: the indoor training set (ITS) and the outdoor training set (OTS), along with a synthetic objective testing set (SOTS). The ITS consists of 13,990 hazy images generated from 1,399 sharp images, while the OTS comprises 313,950 hazy images derived from 8,970 clean images. Our model is trained separately on the ITS and OTS datasets, and subsequently tested on their corresponding test sets, namely SOTS-Indoor and SOTS-Outdoor, each comprising 500 paired images.

7.2.4. Image Desnowing

For desnowing evaluation, we utilize three datasets: Snow100K [36], SRRS [5], and CSD [6]. Snow100K [36] comprises 50,000 image pairs for training and 50,000 for evaluation. SRRS [5] contains 15,005 image pairs for training and 15,005 for evaluation. CSD [6] consists of 8,000 image pairs for training and 2,000 for evaluation. To maintain consistency with the training strategy of the previous algorithm [8], we randomly sample 2,500 image pairs from the training set for training and 2,000 images from the testing set for evaluation.

8. More Experiments

8.1. More Experimental Results

To further validate the effectiveness of our approach, we perform image desnowing experiments under the SSIR setting.

8.1.1. Image Desnowing

We conduct a comprehensive comparison of desnowing performance across three benchmark datasets. As summarized

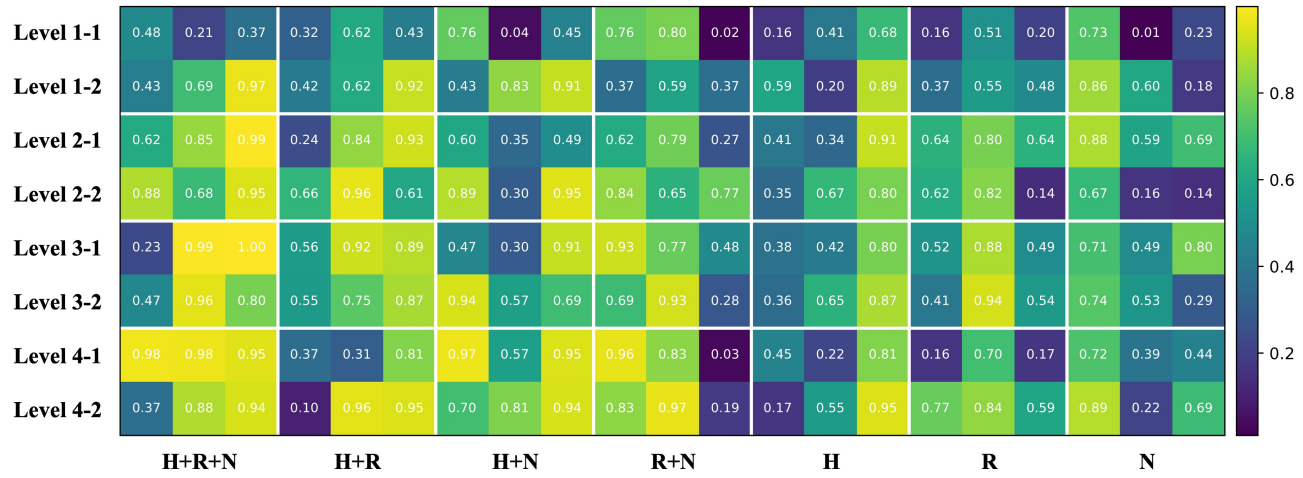

Figure 6. Heat-maps for the activation weight of TABlock.

Table 8. Quantitative comparisons with other image desnowing methods.

Methods	CSD		SRRS		Snow100K	
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
NAFNet [3]	33.13	0.96	29.72	0.94	32.41	0.95
FocalNet [9]	37.18	**0.99**	31.34	**0.98**	33.53	0.95
MSP-Former [4]	33.75	0.96	30.76	0.95	33.43	**0.96**
IRNeXt [10]	37.29	**0.99**	31.91	**0.98**	33.61	0.95
IMDNet(Ours)	**38.44**	0.98	**33.98**	0.96	**34.71**	**0.96**

in Table 8, our proposed IMDNet consistently surpasses existing state-of-the-art methods by a significant margin. In particular, IMDNet achieves an impressive 2.07 dB gain on the SRRS dataset [5] and a 1.15 dB improvement on the recently introduced CSD dataset [6] when compared with IRNeXt [10]. These results highlight the model's superior capability in handling diverse snow patterns and background complexities. Furthermore, relative to MSP-Former [4], a task-specific architecture tailored for image desnowing, IMDNet exhibits substantial quantitative gains of 4.69 dB, 3.22 dB, and 1.28 dB on the Snow100K, SRRS, and CSD datasets, respectively. This demonstrates that our adaptive multi-degradation design not only generalizes effectively to snow removal tasks but also achieves state-of-the-art restoration quality across different data distributions.

8.2. More Ablation Studies

8.2.1. Effectiveness of TABlock

To illustrate the adaptive capability of our TABlock, we show that it can dynamically activate or fuse multiple functional branches based on the decoupled degradation information, thereby flexibly selecting the optimal restoration path and handling various combinations of degraded images. We visualize the activation weight vector for each branch to provide insight into this process. As shown in Figure 6, the activation values vary across different degradation combinations. Specifically, when all degradation types are present, all branches are activated and fused. In contrast, for other combinations, certain branches are either not selected or assigned low weights. This clearly demonstrates TABlock's ability to adaptively adjust its processing according to the specific degradation characteristics of the input image.

Table 9. Design Choices of Skip Connection.

Method	PSNR	△ PSNR
Encoder feature E_i	30.98	-
Decoupled clean feature CF_i	31.19	+0.21

8.2.2. Design Choices of Skip Connection

To further demonstrate the effectiveness of employing our decoupled features CF_i for skip connections, we conduct a comparative experiment using the directly encoded features E_i from the encoder. The results, summarized in Table 9, clearly show that our approach achieves superior performance. This confirms that the proposed DIDBlock effectively separates degradation-related information from clean feature representations, ensuring that the skip connections transmit more reliable and noise-free information to the decoder, thereby enhancing the overall restoration quality.

8.2.3. Effectiveness of Decoupling Loss

To enable accurate feature decoupling in the encoder, the skip-connection feature CF is constrained to contain minimal degradation information, while the branch-selection feature DI is encouraged to primarily capture degradation-specific cues. We impose this separation through a cosine similarity loss that maximizes the distinction between CF

Figure 7. Image restoration results across various combinations of degradation types. The top row displays the degraded inputs, the middle row presents the ground truth images, and the bottom row shows the corresponding restoration results generated by IMDNet.

Table 10. Effectiveness of Decoupling Loss.

Method	PSNR	△ PSNR
w/o L_d	30.66	-
w/ L_d	31.19	+0.53

Table 11. The effect of different degradation information aggregation methods on overall performance.

Modules	Sum	Concatenation	FBlock
PSNR	31.01	31.07	31.19
FLOPs(G)	23.15	24.22	23.15

and DI. To evaluate the effectiveness of this decoupling loss, we conducted corresponding experiments, as shown in Table 10. The results demonstrate that incorporating this loss helps the model better disentangle the respective features, leading to improved restoration performance under multiple degradation conditions.

8.2.4. Design Choices for FBlock

To evaluate the effectiveness of the proposed FBlock, we compare it with alternative aggregation strategies, including simple summation and concatenation. As reported in Table 11, our FBlock achieves consistently better performance, demonstrating its ability to effectively integrate degradation features across multiple hierarchical levels. Notably, this improvement is achieved without adding any extra computational overhead.

Table 12. The evaluation of model computational complexity.

Method	Time(s)	Params(M)	PSNR	SSIM
VLU-Net [55]	0.743	35	29.35	0.842
PromptIR [43]	1.012	33	27.54	0.819
Perceive-IR [58]	0.682	45	30.22	0.894
IMDNet(Ours)	0.352	24	31.19	0.910

8.3. Resource Efficient

We further analyze the computational efficiency of IMDNet by comparing its runtime and parameter count with recent state-of-the-art methods. As shown in Table 12, IMDNet not only achieves leading restoration performance but also significantly reduces computational demand. In particular, it surpasses the previous best method, Perceive-IR [58], by 0.2 dB while requiring only 51.6% of its inference time.

9. Additional Visual Results

Figure 7 illustrates the visual results of our method when handling different combinations of degradation types. In each case, the upper row presents the input degraded images affected by multiple degradations, while the lower row displays the corresponding restored outputs produced by IMDNet. As observed, our model effectively removes various types of degradations and consistently reconstructs images with sharp details, natural textures, and visually pleasing quality across all scenarios.