

Beyond Uniform Restoration: Empowering All-in-One Restoration with Pixel-Level Multimodal Guidance

Chunxiao Liu
Xiaomi Corporation
Beijing, China
liuchunxiao@xiaomi.com

Anbin Xiong
Xiaomi Corporation
Beijing, China
xionganbin@xiaomi.com

Wei Liu
Xiaomi Corporation
Beijing, China
liuwei67@xiaomi.com

Erli Meng
Xiaomi Corporation
Beijing, China
mengerli@xiaomi.com

Abstract

All-in-one image restoration is a unified low-level vision task that aims to effectively recover high-quality images from inputs degraded by various types and levels of corruption using a single model. Recent works have achieved remarkable progress by learning degradation-adaptive prompts or network architectures. However, these methods typically apply a uniform restoration strategy across the entire image, neglecting the fact that different regions may suffer from distinct degradation types and varying degrees of severity. In contrast, we propose to perform restoration at the pixel level, thereby enabling more fine-grained and precise control over the restoration process. Specifically, we present MGN-AIR, a novel pixel-level restoration framework for all-in-one image restoration. Our approach first learns to estimate a pixel-level visual prompt. Then, it leverages both textual and visual prompts to provide global and local degradation cues, guiding the model on where to look and how to restore at each pixel. We conduct extensive experiments on multiple all-in-one image restoration benchmarks, covering a wide range of tasks including denoising, deraining, deblurring, dehazing, desnowing, and low-light enhancement. Experimental results demonstrate that our proposed method consistently and significantly outperforms existing approaches.

CCS Concepts

• **Computing methodologies** → *Reconstruction*.

Keywords

All-in-One Image Restoration, Multimodal

ACM Reference Format:

Chunxiao Liu, Wei Liu, Anbin Xiong, and Erli Meng. 2026. Beyond Uniform Restoration: Empowering All-in-One Restoration with Pixel-Level Multimodal Guidance. In *Proceedings of the 34th ACM International Conference on Multimedia (MM '26)*, November 10–14, 2026, Rio de Janeiro, Brazil. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3767308.3835303>



This work is licensed under a Creative Commons Attribution 4.0 International License. *MM '26, Rio de Janeiro, Brazil*

© 2026 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-2213-4/2026/11

<https://doi.org/10.1145/3767308.3835303>

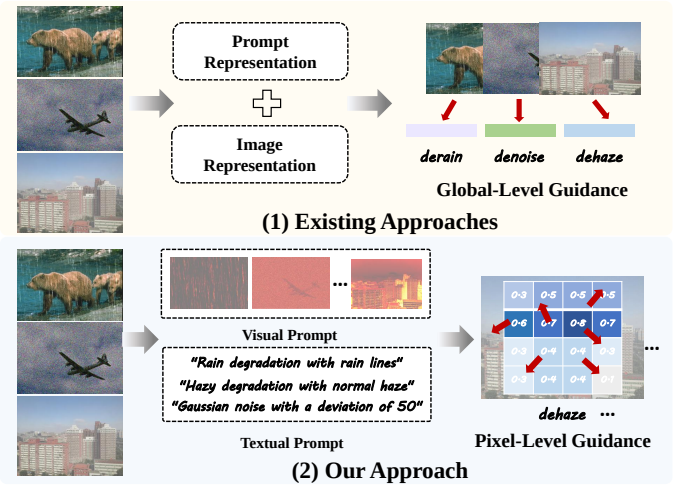


Figure 1: Existing methods typically integrate prompt to provide global-level guidance for image restoration. In contrast, our approach enables fine-grained, pixel-level restoration by leveraging both textual and visual prompts that encodes global degradation and intensity of local degradation.

1 Introduction

As a fundamental low-level vision task, image restoration focuses on recovering high-quality images from low-quality images through removing rain, haze, noise, etc. Due to the co-existence of diverse degradations in real-world images, learning a unified all-in-one restoration model to cope with diverse degradation types has drawn growing attention and achieved considerable advances.

Existing all-in-one image restoration works can be roughly categorized into prompt-based approaches and adaptive network based approaches. Prompt-based approaches mainly focus on incorporating task priors to guide the degradation process. Textual prompts [17, 65], visual prompts [36, 47] and multimodal prompts [30, 37] have been developed in image restoration networks and thus greatly improved the performance. These prompts are generally encoded by pretrained large models [49] or learnable parameters, guiding the restoration network to learn task-specific representations. Another branch of image restoration methods focuses on designing networks adaptive to various image restoration tasks, which aligns the feature space with a unified parameter space. For example, some

works [67–70] introduce mixture-of-experts (MoE) architectures to learn a task-adaptive routing strategy, assigning degradations to the most suitable experts. External tasks like degradation classification [18] and contrastive learning [17] are also employed to discriminate representations from different restorations.

However, degradation patterns exhibit non-uniform distribution due to distinct degradation types and intensity. As shown in Fig. 1, most regions in noisy images are severely corrupted, while only local regions in rainy images are rain-affected. This phenomenon indicates that image restoration requires more fine-grained guidance. For rainy images, restoration efforts should focus primarily on the rain-affected areas, while preserving the content of unaffected regions. Conversely, all regions in hazy images require restoration, but the degree of correction varies, some areas need only minor adjustments, while others demand more extensive correction. Existing methodologies typically treat the entire degraded image uniformly, neglecting the intrinsic variability in degradation and hence leading to suboptimal restoration outcomes.

To address the above issue, we propose a Multimodal Guidance Network for all-in-one image restoration, denoted as MGN-AIR, which performs more fine-grained and precise controls over the restoration process, thereby achieving more effective image restoration. Specifically, MGN-AIR first employs a visual prompt generation module to learn a visual prompt that reflects the local intensity of degradation. Next, in the multimodal guidance module, the learned visual prompt is incorporated with textual prompt to jointly predict a restoration matrix, which tells the model where to look and how to restore at each pixel. Lastly, the restoration matrix guides the image restoration in pixel-level restoration module, where heavily degraded regions primarily leverage contextual information from neighboring areas, while less-degraded regions rely more on recurring structural patterns to further enhance restoration quality. The main contributions are summarized as follows:

- We propose a method to precisely perform image restoration at the pixel level. Compared with global-level restoration strategies, our design delivers more localized guidance, thereby enabling better adaptability to diverse degradation types and varying intensities.
- We design a multimodal guidance network for pixel-level all-in-one image restoration, consisting of a Visual Prompt Generation Module (VPGM), a Multimodal Guidance Module (MGM) and a Pixel-Level Restoration Module (PLRM), to apply pixel-level image restoration through predicting local visual prompt, generating pixel-level restoration guidance and executing pixel-level restoration.
- Extensive experiments conducted on all-in-one image restoration benchmarks show that our network achieves state-of-the-art performance in all-in-one image restoration, especially for composite degradation benchmark, 1.51 dB PSNR improvement is obtained in comparison with recent works.

2 Related Work

Single Task Image Restoration. Restoring the high-quality clean image from its low-quality degraded version is an ill-posed issue. Early works primarily focus on designing task-specific networks based on hand-crafted priors [24, 29, 55, 59, 71, 76]. Recently,

deep learning-based approaches apply convolutional [3, 55, 76, 77] or Transformer [6, 35, 59, 71] architectures, achieving remarkable progress on various image restoration tasks, such as deraining [2, 27, 28, 32, 58], dehazing [5, 10, 34, 48, 53, 63], deblurring [4, 7, 22, 23, 46, 75], denoising [19, 31, 44, 45, 50, 51] and low-light enhancement [15, 39, 64]. The convolutional networks aggregate local neighboring information to restore the degraded regions. In contrast, transformer-based networks capture global dependencies via self-attention mechanism, in which similar patterns play a more important role in restoration. The typical work is Restormer [71] that leverages transformer architectures to enhance image restoration tasks in a lightweight manner. However, these works are primarily designed for single-degradation, limiting their generalization capability in all-in-one image restoration.

All-In-One Image Restoration. The typical all-in-one image restoration framework is encoder-decoder architecture. Based on this structure, various prompts have been proposed to enhance performance, including textual prompt, visual prompt and multimodal prompt. These prompts serve as an adaptive guidance module in the restoration process. For textual prompting, most works utilize the pretrained models as guidance to better understand task-specific context. InstructIR [8] is the first approach that leverages real human-written instructions to guide the overall restoration process. For visual prompting, a representative work is PromptIR [47], which enhances Restormer [71] by learning prompt representation and integrating it into the U-Net architecture [52]. For multimodal prompting, DA-CLIP [37] designs a controller that firstly employs pretrained CLIP as text and image encoder, and then integrates the prompt embedding into image restoration models via cross attention. CyclicPrompt [30] incorporates weather knowledge and conditional vectors into the visual prompt, jointly learning context-aware representations with textual prompt. Apart from prompt-based approaches, a large number of works focus on designing unified networks using adaptive learning strategies, such as multi-task collaboration [67–70], self-contrastive learning [25, 61], adding task uncertainty regularization [62], considering degradation-aware feature perturbation [56], and applying different operations for hard and easy patches [20], etc.

Unlike prior approaches designed to adapt to diverse degradation types, our work pioneers a fine-grained pixel-wise network jointly driven by degradation type and intensity, making it well-suited for realistic, complex, and non-uniform degradations.

3 Method

In this work, we present MGN-AIR, a multimodal guidance network for all-in-one image restoration. As illustrated in Fig. 2, the overall framework is a classic encoder-decoder network, in which each level is comprised of our proposed Multimodal Guidance Block (MGB), which firstly generates the visual prompt, and then integrates textual prompt and lastly performs pixel-level restoration.

3.1 Overview

Our framework builds upon a U-net like encoder-decoder architecture [3, 47, 56, 71]. Given a degraded image $I_{LQ} \in \mathbb{R}^{H \times W \times 3}$ as input, our framework aims to restore a high-quality clean image $I_{HQ} \in \mathbb{R}^{H \times W \times 3}$. Here, $H \times W$ denotes the height and width of the

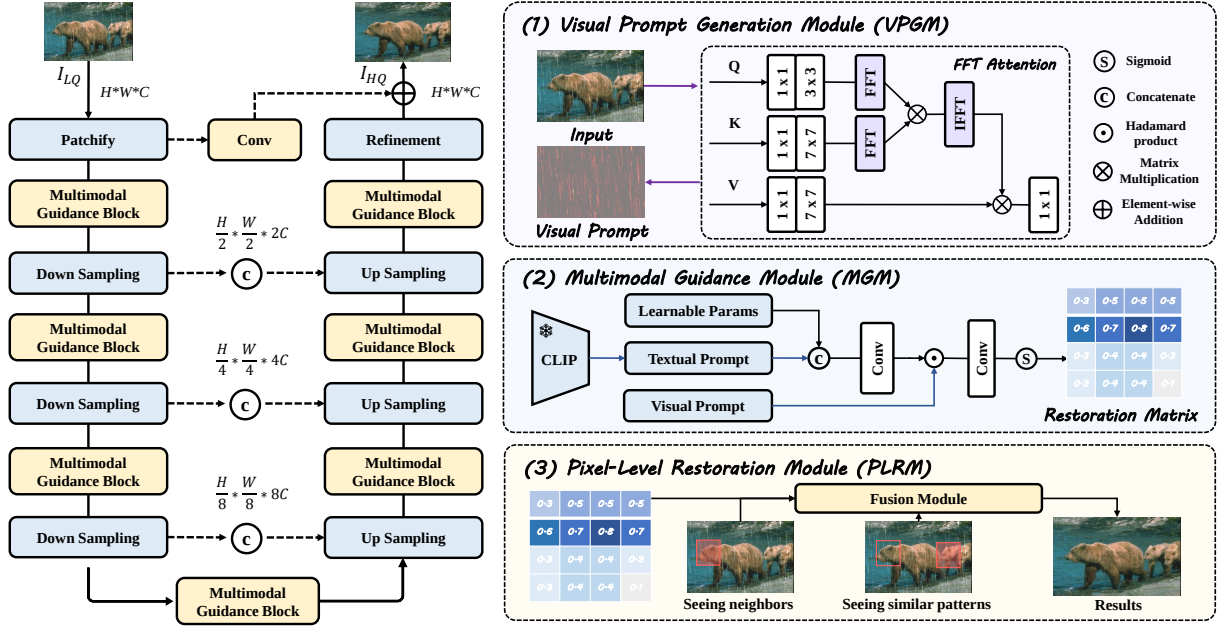


Figure 2: Overall framework of our MGN-AIR. The left part is the U-net architecture, which consists of our proposed multimodal guidance block (MGB). The right part describes the detail of MGB. (1) Generating visual prompt supervised by the difference between degraded image and clean image; (2) Estimating pixel-wise restoration matrix guided by multimodal prompts; (3) performing pixel-level restoration based on restoration matrix.

image. The encoder-decoder network comprises four levels, each equipped with a MGB. The encoder network takes a high-resolution image as input, the spatial resolution gradually decreases, while the channel increases through pixel shuffle. At the decoder stage, these changes are reversed to progressively restore the original resolution via pixel unshuffle. Encoder features at each level are fused with the decoder features at the same resolution via skip connections.

Concretely, at the i -th level, the MGB takes features F^{i-1} from previous level as input, performing pixel-level restoration spatially. The MGB mainly comprised of three modules. The first module is visual prompt generation module (VPGM) that generates pixel-level visual prompt V^i in a learnable manner. The next module is multimodal guidance module (MGM) that learns to integrate both visual and textual prompt to provide multimodal guidance P^i in a fine-grained granularity. At last, pixel-level restoration module (PLRM) executes distinct restoration at each pixel guided by P^i , producing intermediate image feature F_s^i . The overall process are:

$$\begin{aligned} V^i &= \text{VPGM}(F^{i-1}); \\ P^i &= \text{MGM}(T, V^i); \\ F_s^i &= \text{PLRM}(P^i, F^{i-1}). \end{aligned} \quad (1)$$

After that, MGB takes intermediate features as input and applies a sequence of Transformer Blocks. In this block, We sequentially perform Channel Attention and feed-forward network (FFN) to feature F_s^i , adding the output back to the original input feature via residual connection, producing latent feature F^i . The overall structure of MGB enables us to gradually and adaptively recover degraded region guided by pixel-level information. Note that in

the final decoder layer, we directly adopt the visual prompt of the preceding layer instead of applying VPGM.

$$\begin{aligned} \hat{F}^i &= \text{ChannelAttn}(F_s^i) + F_s^i, \\ F^i &= \text{FFN}(\hat{F}^i) + \hat{F}^i. \end{aligned} \quad (2)$$

3.2 Visual Prompt Generation Module (VPGM)

The goal of VPGM is to estimate the visual prompt that indicates the degradation map of each pixel. Taking the degraded image $I_{LQ} \in \mathbb{R}^{H \times W \times C}$ as input, where C denotes the channel number. For the first level, the channel number equals to 3, while other levels equals to the feature dimension. The VPGM firstly employs a 1×1 convolution, and then applies a window-based self-attention mechanism to generate visual prompt based on spatial information. To reduce the computation cost, the matrix multiplication between queries Q' and keys K' is performed in the Fourier domain [21]. Subsequently, the computed features are then transformed back into the spatial domain, performing maxtrix multiplication with values V' after Softmax:

$$\begin{aligned} Q' &= xW_Q, \quad K' = xW_K, \quad V' = xW_V; \\ \hat{Q} &= \mathcal{F}(Q'), \quad \hat{K} = \mathcal{F}(K'); \\ A &= \text{Softmax}\left(\frac{1}{\sqrt{d}}\mathcal{F}^{-1}(\hat{Q} \cdot \hat{K})\right); \\ V &= A \cdot V'. \end{aligned} \quad (3)$$

where, $x \in \mathbb{R}^{r \times r \times C}$ denotes the content within each window, the window size r is set as 32. $Q', K', V' \in \mathbb{R}^{r \times r \times d}$ represent query, key and value in self-attention mechanism, and $W_Q, W_K, W_V \in \mathbb{R}^{C \times d}$

are corresponding learnable weights. $\mathcal{F}$ and $\mathcal{F}^{-1}$ represent FFT and inverse FFT, respectively. Subsequently, the 1×1 convolution and SiLU activation function are applied, followed by Sigmoid function to predict the degradation map $V \in \mathbb{R}^{H \times W}$ of each pixel, which is our proposed visual prompt.

Notably, the visual prompt is generated explicitly only at the first level, while prompts at subsequent levels are learned implicitly. Explicit learning at the initial stage requires computing the pixel-wise difference between the degraded image and its ground-truth as supervision. This difference directly reflects the spatial intensity of degradation, thereby providing precise, pixel-level guidance for restoration. In contrast, deeper levels benefit from implicit learning, which allows the model to adaptively capture higher-order dependencies without relying on direct supervision. Specifically, we first compute the spatial pixel-wise difference and then average it across the channel dimension. After applying min-max normalization, we obtain a degradation map, which serves as the ground truth pixel-level visual prompt for the first level, formulated as follows:

$$X = \frac{1}{C} \sum_{i=1}^C \|I_{HQ}^i - I_{LQ}^i\|_1; \quad (4)$$

$$\hat{X} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}.$$

where, i denotes the channel index, I_{HQ}^i and I_{LQ}^i denote the i -th channel in ground truth image and degraded image, respectively.

3.3 Multimodal Guidance Module (MGM)

Recent works [8, 17, 65] have demonstrated the potential of textual prompts in all-in-one image restoration as it facilitates discriminating different types of degradation. However, in real-world scenarios, degradation exhibits diverse types and intensities, which demands more fine-grained, pixel-level control instead of the global guidance provided by existing textual prompts. Consequently, providing a more fine-grained guidance is crucial for this field.

To address the above issue, we propose a novel module MGM that leverages both visual and textual prompts for pixel-level image restoration. Specifically, the textual prompt conveys high-level semantic information about the degradation type (e.g., “raindrops”), thereby determining the distribution or pattern of degradation across the image. Complementarily, the visual prompt provides spatially precise guidance by indicating the severity of degradation at each pixel. For instance, for a rainy image, the textual prompt specifies that the corruption is caused by raindrops, while the visual prompt reveals where and how severely each pixel is affected.

Specifically, we employ the frozen CLIP [49] text encoder to extract a textual representation $T \in \mathbb{R}^{1 \times d}$, where $d = 512$ is the embedding dimension of CLIP. The visual prompt $V \in \mathbb{R}^{H \times W}$ is estimated by the preceding module VPGM. We first aggregate textual information by applying an average pooling on the textual prompt, producing a textual feature as global guidance. Then, a learnable scaling parameter β is concatenated with the global guidance, followed by a 1D convolution. To adapt this global feature to the spatial domain, we replicate the textual feature to match the spatial resolution $H \times W$, yielding $T^P \in \mathbb{R}^{H \times W \times C}$, where $C = 32$. Subsequently, the modulated textual prompt and the visual prompt are combined via element-wise multiplication, followed by a 1×1

convolution and Sigmoid activation. The resulting restoration matrix $P \in [0, 1]^{H \times W \times C}$ employs both local and global guidance to jointly determine the restoration at each pixel. The overall process is formulated as:

$$\begin{aligned} T^P &= \text{Concat}[\text{Avg}(T); \beta]; \\ V^P &= \text{Conv1D}(T^P) \odot V; \\ P &= \text{Sigmoid}(\text{Conv2D}(V^P)). \end{aligned} \quad (5)$$

3.4 Pixel-Level Restoration Module (PLRM)

In the preceding module, a pixel-level restoration matrix is estimated by fusing visual and textual prompts. This module, referred to as PLRM, is designed to perform image restoration adaptively. Specifically, different restoration strategies are applied depending on local degradation severity. For instance, in mildly degraded pixels, such as those affected by haze or low light, effective restoration benefits from both local contextual information and recurring structural patterns in the image. In contrast, for heavily degraded pixels, such as those corrupted by rain or noise, only neighboring pixels remain reliable, and hence require a more localized approach.

To address the above issue, we propose to utilize similar patterns and neighbor information based on restoration matrix. For heavily degraded pixels, we combine standard 3×3 2D convolutions with dilated convolutions to enhance receptive field, in case of degradation span large spatial areas. Inspired by the principle of the attention mechanism, the query feature to be restored heavily relies on similar patterns in the image. We employ a series of spatial attention blocks to address less-degraded pixels, operating within an 8×8 local window to reduce computational cost. The combination of two strategies ensures effective contextual modeling even under severe corruption.

$$\begin{aligned} \hat{F} &= W_c \cdot \text{Conv2D}(F) + W_d \cdot \text{DilaConv}(F); \\ F_s &= P \odot \hat{F} + (1 - P) \odot \text{SpatialAttn}(F). \end{aligned} \quad (6)$$

where, W_c and W_d denote learnable parameters for Conv2D and DilaConv operation, P denotes the restoration matrix learned by previous module MGM.

3.5 Optimization

Our approach takes into account both perceptual quality and pixel-level fidelity of image restoration, and hence employs pixel-wise loss $\mathcal{L}_{L_1}$, perceptual loss $\mathcal{L}_{Percep}$ and SSIM loss $\mathcal{L}_{SSIM}$ as the objective function. Here, L_1 loss measures absolute pixel differences. Additionally, we employ the L_1 loss between estimated visual prompt and the real degradation map as an auxiliary loss to supervise the visual prompt learning, denoted as $\mathcal{L}_{Aux}$. The SSIM loss accounts for structural patterns and local dependencies, which is more aligned with human visual perception. This loss is a differentiable approximation of the negative SSIM metric, which quantifies the perceived similarity between two images by comparing their luminance, contrast, and structural information. The perceptual loss measures the discrepancy between high-level feature representations of the predicted and ground-truth images. The representations are generally extracted from pre-trained VGG network [54], capturing semantic and textual similarities that better reflect human judgment.

Table 1: Comparison state-of-the-art on composited degradations dataset CDD11 [17]. PSNR (dB, $\uparrow$) and SSIM ($\uparrow$) are reported on the full RGB images. Our method consistently outperforms existing models on both single and composited degradations.

Method	CDD11-Single								CDD11-Double								CDD11-Triple				Average			
	Low(L)		Haze(H)		Rain(R)		Snow(S)		L+H		L+R		L+S		H+R		H+S		L+H+R				L+H+S	
AirNet [25]	24.83	.778	24.21	.951	26.55	.891	26.79	.919	23.23	.779	22.82	.710	23.29	.723	22.21	.868	23.29	.901	21.80	.708	22.24	.725	23.75	.814
PromptIR [47]	26.32	.805	26.10	.969	31.56	.946	31.53	.960	24.49	.789	25.05	.771	24.51	.761	24.54	.924	23.70	.925	23.74	.752	23.33	.747	25.90	.850
WGWSNet [78]	24.39	.774	27.90	.982	33.15	.964	34.43	.973	24.27	.800	25.06	.772	24.60	.765	27.23	.955	27.65	.960	23.90	.772	23.97	.771	26.96	.863
WeatherDiff [43]	23.58	.763	21.99	.904	24.85	.885	24.80	.888	21.83	.756	22.69	.730	22.12	.707	21.25	.868	21.99	.868	21.23	.716	21.04	.698	22.49	.799
OneRestore [17]	26.48	.826	32.52	.990	33.40	.964	34.31	.973	25.79	.822	25.58	.799	25.19	.789	29.99	.957	30.21	.964	24.78	.788	24.90	.791	28.47	.878
MoCE-IR [70]	27.26	.824	32.66	.990	34.31	.970	35.91	.980	26.24	.817	26.25	.800	26.04	.793	29.93	.964	30.19	.970	25.41	.789	25.39	.790	29.05	.881
MGN-AIR	27.30	.824	36.06	.994	36.53	.981	38.40	.986	27.01	.826	26.81	.808	26.06	.802	32.54	.978	33.01	.980	26.08	.801	26.33	.802	30.56	.889

Therefore, the overall loss function is:

$$\mathcal{L}_{Total} = \mathcal{L}_{L_1} + \mathcal{L}_{SSIM} + \mathcal{L}_{Percep} + \mathcal{L}_{Aux}. \quad (7)$$

4 Experiment

In this work, we closely follow previous all-in-one image restoration works [8, 12, 17, 20, 25, 47, 71], evaluating our proposed method on various experimental settings, including mixed degradations and typical all-in-one settings. In these settings, a unified restoration model is trained to address multiple degradation types. The evaluation metrics are PSNR and SSIM, which are standard and widely used for assessing image quality in prior works. Methods with best performance are highlighted in bold.

4.1 Experimental Settings

Datasets. For mixed degradation setting, we employ CDD11 dataset [17] as our benchmark to evaluate the effectiveness on composite degraded images. For typical all-in-one restoration setting, the dataset is composed of multiple task-specific datasets. The BSD400 [1] and WED [38] datasets are combined as the training dataset for image denoising, adding Gaussian noise at different levels $\sigma \in [15, 20, 50]$ to create noisy images. The SOTS [26] dataset is used for image dehazing. Rain100L [66] is adopted for image deraining. The GoPro [42] and LOLv1 [60] datasets are employed for deblurring and low-light enhancement, respectively. Evaluation is performed on standard benchmark BSD68 [40] for image denoising, and corresponding test sets for other tasks. A single unified model is trained on the union of all the aforementioned training datasets and evaluated directly across all restoration tasks without task-specific adaptation.

Implementation Details. We propose a unified framework MGN-AIR for all-in-one image restoration. Following the architectural design of PromptIR [47] and DFPIR [56], our approach adopts a 4-level encoder-decoder structure, with the number of transformer blocks at each level set to [4, 6, 6, 8] from level-1 to level-4, respectively. All experiments are implemented by PyTorch and conducted on a single NVIDIA H200 GPU. In mixed degradation setting, the model is trained for 300 epochs with an initial learning rate of 2×10^{-4} . In typical all-in-one setting, the model is trained for 80 epochs. During the training phase, we randomly crop the full degraded image into 128×128 patches, and perform the data augmentation using random horizontal and vertical flips. The batch size is set as 25. The network is optimized by Adam optimizer, with parameters β_1 equal to 0.9 and β_2 equal to 0.999.

4.2 Experimental Results

All-in-One: Mixed-degradation. To better simulate realistic degradation scenarios, recent studies [17, 70] have trained all-in-one restoration models on more challenging images, where multiple degradation types co-occur within a single image. The typical benchmark is CDD11, in which rain, haze, snow, and low illumination four degradation types are regarded as base degradations, and their combinations are also included as composite degradation scenarios, yielding a total of eleven unique restoration tasks. In this benchmark, the degradation type and degradation level of each pixel might be different. The results in Tab.1 show that our approach MGN-AIR surpasses recent work MoCE-IR [56] and OneRestore [17] on average PSNR by 1.51 dB and 2.09 dB, respectively, validating the effectiveness of our approach in handling more complex and realistic scenarios. In contrast, existing methods typically adopt a global restoration strategy, which struggles to achieve fine-grained, pixel-adaptive image restoration.

All-in-One: Five Degradations. We compare our unified image restoration approach with classic works, including image-only models like AirNet [25], PromptIR [47], multimodal models like DA-CLIP [37], InstructIR [8], and recent works like multi-expert network MoCE-IR [70], VLM-based network VLU-Net [73], and perturbation-aware network DFPIR [56]. All these models are trained on a comprehensive dataset of five degradation types, including denoising, dehazing, deraining, deblurring and low-light enhancement. The results in Tab.2 demonstrate that our proposed network MGN-AIR, achieving the best performance on average PSNR and SSIM, with 2.29 dB improvement in PSNR over baseline work PromptIR [47]. Compared with most recent work DFPIR [56], our approach significantly improves the overall performance, especially on deblurring and deraining tasks, achieving 2.36 dB and 1.85 dB improvement in PSNR, which validates the effectiveness of our approach. Notably, compared with multi-expert model MoCE-IR [70], our approach greatly outperforms it using only one expert, which demonstrates the effectiveness of fine-grained guidance.

All-in-One: Three Degradations. We also evaluate the overall average performance across three image restoration tasks: denoising at three noise levels ($\sigma = 15, 25, 50$), dehazing, and deraining. For denoising, results on BSD68 are averaged over the three noise levels and reported as BSD68_{15-50} . Experimental results in Tab.3 show that our proposed approach MGN-AIR can significantly improve the performance of dehazing and deraining. The average PSNR

Table 2: Comparison to state-of-the-art on five degradations. PSNR (dB, $\uparrow$) and SSIM ($\uparrow$) are reported on the full RGB images. Our method achieves a new state-of-the-art on average across all benchmarks. The best performances are highlighted.

Method	Venue	Dehazing		Deraining		Denoising		Deblurring		Low-Light		Average	
		SOTS		Rain100L		BSD68 $_{\sigma=25}$		GoPro		LOLv1			
NAFNet [3]	ECCV'22	25.23	.939	35.56	.967	31.02	.883	26.53	.808	20.49	.809	27.76	.881
DGUNet [41]	CVPR'22	24.78	.940	36.62	.971	31.10	.883	27.25	.837	21.87	.823	28.32	.891
SwinIR [29]	ICCV'21	21.50	.891	30.78	.923	30.59	.868	24.52	.773	17.81	.723	25.04	.835
Restormer [71]	CVPR'22	24.09	.927	34.81	.962	31.49	.884	27.22	.829	20.41	.806	27.60	.881
MambaIR [16]	ECCV'24	25.81	.944	36.55	.971	31.41	.884	28.61	.875	22.49	.832	28.97	.901
DL [13]	TPAMI'19	20.54	.826	21.96	.762	23.09	.745	19.86	.672	19.83	.712	21.05	.743
Transweather [57]	CVPR'22	21.32	.885	29.43	.905	29.00	.841	25.12	.757	21.21	.792	25.22	.836
TAPE [33]	ECCV'22	22.16	.861	29.67	.904	30.18	.855	24.47	.763	18.97	.621	25.09	.801
AirNet [25]	CVPR'22	21.04	.884	32.98	.951	30.91	.882	24.35	.781	18.18	.735	25.49	.847
IDR [74]	CVPR'23	25.24	.943	35.63	.965	31.60	.887	27.87	.846	21.34	.826	28.34	.893
PromptIR [47]	NeurIPS'23	26.54	.949	36.37	.970	31.47	.886	28.71	.881	22.68	.832	29.15	.904
InstructIR [8]	ECCV'24	27.10	.956	36.84	.973	31.40	.887	29.40	.886	23.00	.836	29.55	.908
AdaIR [9]	ICLR'25	30.53	.978	38.02	.981	31.35	.889	28.12	.858	23.00	.845	30.20	.910
MoCE-IR [70]	CVPR'25	30.48	.974	38.04	.982	31.34	.887	30.05	.899	23.00	.852	30.58	.919
VLU-Net [73]	CVPR'25	30.84	.980	38.54	.982	31.43	.891	27.46	.840	22.29	.833	30.11	.905
DFPIR [56]	CVPR'25	31.64	.979	37.62	.978	31.29	.889	28.82	.873	23.82	.843	30.64	.913
MGN-AIR (Ours)	—	31.74	.982	39.47	.986	31.12	.884	31.18	.917	23.67	.862	31.44	.926

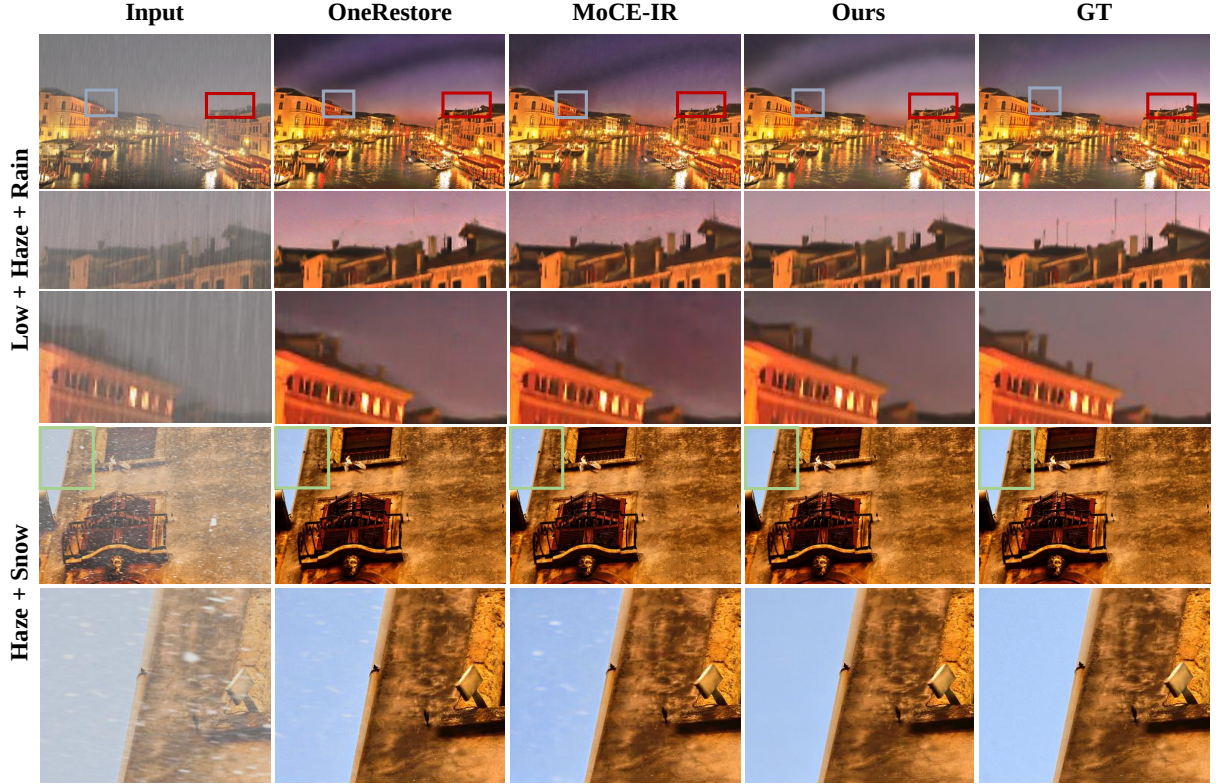


Figure 3: Visual results of MoCE-IR [56], OneRestore [17] and our approach on CDD11 dataset [17]. We provide the global image and local detail grounded by different colors for better visualization. MGN-AIR effectively removes haze, rain and snow streaks while preserving local details in the image.

of our approach also outperforms existing methods. The performance on denoising task is slightly lower than state-of-the-arts,

which is primarily due to severe disturbance of both local and global information by noise.

Table 3: Comparison to state-of-the-art on three degradations. PSNR (dB, $\uparrow$) and SSIM ($\uparrow$) are reported on the full RGB images. Our method achieves a new state-of-the-art on average across all benchmarks.

Method	Dehazing		Deraining		Denoising		Average	
	SOTS		Rain100L		BSD68 ₁₅₋₅₀			
LPNet [14]	20.84	.828	24.88	.784	24.17	.693	23.30	.768
FDGAN [11]	24.71	.929	29.89	.933	28.50	.851	27.70	.904
DL [13]	26.92	.931	32.62	.931	30.12	.838	29.89	.900
MPRNet [72]	25.28	.955	33.57	.954	30.66	.862	29.84	.924
AirNet [25]	27.94	.962	34.90	.967	31.06	.873	31.30	.934
Restormer [71]	30.43	.975	36.55	.975	30.97	.869	32.65	.940
PromptIR [47]	30.58	.974	36.37	.972	31.12	.873	32.69	.940
InstructIR [8]	30.22	.959	37.98	.978	31.32	.876	33.17	.938
AdaIR [9]	31.06	.980	38.64	.983	31.25	.876	33.65	.946
MoCE-IR [70]	31.34	.979	38.57	.984	31.25	.873	33.72	.945
DFPIR [56]	31.87	.980	38.65	.982	31.29	.878	33.94	.947
MGN-AIR	33.21	.987	39.62	.986	31.00	.868	34.61	.947

Table 4: Ablation study on the multimodal guidance block design. PSNR (dB, $\uparrow$) and SSIM ($\uparrow$) are reported.

Method	Dehazing		Deraining		Denoising		Average	
	SOTS		Rain100L		BSD68 ₁₅₋₅₀			
w/o MGB	31.23	.980	38.17	.982	31.23	.877	33.54	.946
w/o visual prompt	30.89	.978	37.34	.980	30.83	.865	33.02	.941
w/o textual prompt	31.31	.980	37.37	.980	30.80	.865	33.16	.942
w/o multi-layer block	31.91	.982	38.84	.984	30.89	.867	33.88	.944
MGN-AIR	33.21	.987	39.62	.986	31.00	.868	34.61	.947

Effectiveness of each component. We conduct extensive ablation experiments to validate the effectiveness of our proposed multimodal guidance block, reporting the results of training an all-in-one model on combined datasets across three restoration tasks. The experiment settings contain the following five variants: (1) w/o MGB: Only using U-net structure, removing multimodal guidance block; (2) w/o visual prompt: Removing the VPGM, and removing visual prompt in MGM; (3) w/o textual prompt: Removing textual prompt in MGM; (4) w/o multi-layer block: Only using multimodal guidance block at the first layer, removing this block from subsequent layers; and (5) MGN-AIR: Using all the proposed module. As shown in Tab.4, removing any component in MGN-AIR will greatly influence the performance. Concretely, removing visual prompt will largely reduce the effect, which demonstrates the fine-grained pixel-level guidance is crucial for image restoration. Moreover, removing textual prompt will also influence performance, since different degradation types indicate different distributions. Compared with w/o MGB setting, only preserving MGB at the first layer can yield a 0.34 dB PSNR improvement, and adding multiple MGB will further achieve 0.73 dB PSNR improvement, which validates the effectiveness of our proposed block.

Complexity comparisons: We compare the computational complexity of our method against standard baselines, including PromptIR [47] and AirNet [25]. To demonstrate that our performance

gains stem from the efficacy of the proposed MGB block rather than merely an increase in model capacity, we designed experiments focusing on two aspects: (a) *Capacity reduction*: We compress the number of layers within the MGB to create a smaller variant, denoted as *Ours-s*; and (b) *Structural substitution*: We replace our MGB block with alternative simple operations, such as spatial attention or gated networks, while maintaining a comparable parameter count. As shown in Tab. 5, our proposed module consistently outperforms these variants, and our lightweight version *Ours-s* outperforms baselines while keeps similar parameter counts, confirming that the improvements are attributed to the specific design of the MGB.

Table 5: Computation cost comparison.

Model	Params (M)	FLOPs (G)	Inference	
			Latency (ms)	Average
PromptIR [47]	32.97	121.08	31.09	32.69 .940
AirNet [25]	5.77	230.66	28.83	31.30 .934
Spatial	34.29	217.14	64.84	33.69 .946
Gated	33.63	208.56	61.75	29.43 .904
Ours	34.33	215.19	63.25	34.61 .947
Ours-s	26.61	133.42	35.84	34.13 .946

Visualization. To intuitively demonstrate the effectiveness of our method, we provide visualization results under the mixed degradation settings in Fig.3. In complex scenarios composited by haze, rain and low illumination, both MoCE-IR [56] and OneRestore [17] perform image restoration at global level, resulting in noticeable errors in local regions. In contrast, our approach greatly preserves the local details while removing degradations. For example, in challenging scenes composited by haze, rain and low illumination, the antenna on the right rooftop, a subtle yet important structural detail, is largely neglected in prior works. Our method recovers it with high fidelity, whereas MoCE-IR exhibits considerable blurriness. Furthermore, on the left rooftop, existing methods exhibit over-smoothing, leading to white artifacts on the roof surface, while our approach effectively addresses this issue by pixel-level restoration. In certain scenes composited by haze and snow, our approach generates clear and sharp denoised outputs, while existing methods fail to remove snow that appears in the top-left corner of the image.

5 Conclusion

This paper presents MGN-AIR, a multimodal guidance network for fine-grained all-in-one image restoration, which operates the restoration process at the pixel level for adaptive restoration control. By jointly leveraging the learned visual and textual prompts, our method tailors the restoration operation for each pixel based on local intensity and global degradation type. Extensive experiments in diverse degradation settings show that MGN-AIR can significantly outperform existing methods, especially on scenarios with multiple composited degradations.

References

- [1] Pablo Arbelaez, Michael Maire, Charless Fowlkes, and Jitendra Malik. 2010. Contour detection and hierarchical image segmentation. *IEEE transactions on pattern analysis and machine intelligence* 33, 5 (2010), 898–916.
- [2] Chenghao Chen and Hao Li. 2021. Robust representation learning with feedback for single image deraining. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 7742–7751.
- [3] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. 2022. Simple baselines for image restoration. In *European conference on computer vision*. Springer, 17–33.
- [4] Lufei Chen, Xiangpeng Tian, Shuhua Xiong, Yinjie Lei, and Chao Ren. 2024. Unsupervised blind image deblurring based on self-enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 25691–25700.
- [5] Wei-Ting Chen, Hao-Yu Fang, Jian-Jiun Ding, and Sy-Yen Kuo. 2020. PMHLD: Patch map-based hybrid learning DehazeNet for single image haze removal. *IEEE Transactions on Image Processing* 29 (2020), 6773–6788.
- [6] Zheng Chen, Yulun Zhang, Jinjin Gu, Linghe Kong, Xin Yuan, et al. 2022. Cross aggregation transformer for image restoration. *Advances in Neural Information Processing Systems* 35 (2022), 25478–25490.
- [7] Sung-Jin Cho, Seo-Won Ji, Jun-Pyo Hong, Seung-Won Jung, and Sung-Jea Ko. 2021. Rethinking coarse-to-fine approach in single image deblurring. In *Proceedings of the IEEE/CVF international conference on computer vision*. 4641–4650.
- [8] Marcos V Conde, Gregor Geigle, and Radu Timofte. 2024. Instructir: High-quality image restoration following human instructions. In *European Conference on Computer Vision*. Springer, 1–21.
- [9] Yuning Cui, Syed Waqas Zamir, Salman Khan, Alois Knoll, Mubarak Shah, and Fahad Shahbaz Khan. 2025. Adair: Adaptive all-in-one image restoration via frequency mining and modulation. In *13th International Conference on Learning Representations, ICLR 2025*. International Conference on Learning Representations, ICLR, 57335–57356.
- [10] Sourya Dipta Das and Saikat Dutta. 2020. Fast deep multi-patch hierarchical network for nonhomogeneous image dehazing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*. 482–483.
- [11] Yu Dong, Yihao Liu, He Zhang, Shifeng Chen, and Yu Qiao. 2020. FD-GAN: Generative adversarial networks with fusion-discriminator for single image dehazing. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 34. 10729–10736.
- [12] Huiyu Duan, Xiongkuo Min, Sijing Wu, Wei Shen, and Guangtao Zhai. 2024. Uniprocessor: a text-induced unified low-level image processor. In *European Conference on Computer Vision*. Springer, 180–199.
- [13] Qingnan Fan, Dongdong Chen, Lu Yuan, Gang Hua, Nenghai Yu, and Baoquan Chen. 2019. A general decoupled learning framework for parameterized image operators. *IEEE transactions on pattern analysis and machine intelligence* 43, 1 (2019), 33–47.
- [14] Hongyun Gao, Xin Tao, Xiaoyong Shen, and Jiaya Jia. 2019. Dynamic scene deblurring with parameter selective sharing and nested skip connections. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 3848–3856.
- [15] Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. 2020. Zero-reference deep curve estimation for low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 1780–1789.
- [16] Hang Guo, Jinmin Li, Tao Dai, Zhihao Ouyang, Xudong Ren, and Shu-Tao Xia. 2024. Mambair: A simple baseline for image restoration with state-space model. In *European conference on computer vision*. Springer, 222–241.
- [17] Yu Guo, Yuan Gao, Yuxu Lu, Huilin Zhu, Ryan Wen Liu, and Shengfeng He. 2024. Onerestore: A universal restoration framework for composite degradation. In *European conference on computer vision*. Springer, 255–272.
- [18] JiaKui Hu, Lujia Jin, Zhengjian Yao, and Yanye Lu. [n. d.]. Universal Image Restoration Pre-training via Degradation Classification. In *The Thirteenth International Conference on Learning Representations*.
- [19] Jie Huang, Xiao Liu, Yizhong Pan, Xiaohai He, and Chao Ren. 2022. CasaPuNet: channel affine self-attention-based progressively updated network for real image denoising. *IEEE Transactions on Industrial Informatics* 19, 8 (2022), 9145–9156.
- [20] Jiachen Jiang, Tianyu Ding, Ke Zhang, Jinxin Zhou, Tianyi Chen, Ilya Zharkov, Zhihui Zhu, and Luming Liang. 2025. Cat-AIR: Content and Task-Aware All-in-One Image Restoration. *arXiv preprint arXiv:2503.17915* (2025).
- [21] Lingshun Kong, Jiangxin Dong, Jianjun Ge, Mingqiang Li, and Jinshan Pan. 2023. Efficient frequency domain-based transformers for high-quality image deblurring. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 5886–5895.
- [22] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiri Matas. 2018. Deblurgan: Blind motion deblurring using conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 8183–8192.
- [23] Orest Kupyn, Tetiana Martyniuk, Junru Wu, and Zhangyang Wang. 2019. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In *Proceedings of the IEEE/CVF international conference on computer vision*. 8878–8887.
- [24] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. 2018. Noise2Noise: Learning image restoration without clean data. *arXiv preprint arXiv:1803.04189* (2018).
- [25] Boyun Li, Xiao Liu, Peng Hu, Zhongqin Wu, Jiancheng Lv, and Xi Peng. 2022. All-in-one image restoration for unknown corruption. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 17452–17462.
- [26] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. 2018. Benchmarking single-image dehazing and beyond. *IEEE transactions on image processing* 28, 1 (2018), 492–505.
- [27] Ruoteng Li, Loong-Fah Cheong, and Robby T Tan. 2019. Heavy rain image restoration: Integrating physics model and conditional adversarial learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 1633–1642.
- [28] Xia Li, Jianlong Wu, Zhouchen Lin, Hong Liu, and Hongbin Zha. 2018. Recurrent squeeze-and-excitation context aggregation net for single image deraining. In *Proceedings of the European conference on computer vision (ECCV)*. 254–269.
- [29] Jingyun Liang, Jie Zhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. 2021. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*. 1833–1844.
- [30] Rongxin Liao, Feng Li, Yanyan Wei, Zenglin Shi, Le Zhang, Huihui Bai, and Meng Wang. 2025. Prompt to Restore, Restore to Prompt: Cyclic Prompting for Universal Adverse Weather Removal. *arXiv preprint arXiv:2503.09013* (2025).
- [31] Xin Lin, Chao Ren, Xiao Liu, Jie Huang, and Yinjie Lei. 2023. Unsupervised image denoising in real-world scenarios via self-collaboration parallel generative adversarial branches. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 12642–12652.
- [32] Xin Lin, Jingtong Yue, Sixian Ding, Chao Ren, Lu Qi, and Ming-Hsuan Yang. 2024. Dual degradation representation for joint deraining and low-light enhancement in the dark. *IEEE transactions on circuits and systems for video technology* (2024).
- [33] Lin Liu, Lingxi Xie, Xiaopeng Zhang, Shanxin Yuan, Xiangyu Chen, Wengang Zhou, Houqiang Li, and Qi Tian. 2022. Tape: Task-agnostic prior embedding for image restoration. In *European conference on computer vision*. Springer, 447–464.
- [34] Xiaohong Liu, Yongrui Ma, Zhihao Shi, and Jun Chen. 2019. Griddehazenet: Attention-based multi-scale network for image dehazing. In *Proceedings of the IEEE/CVF international conference on computer vision*. 7314–7323.
- [35] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. 2021. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*. 10012–10022.
- [36] Wenyang Luo, Haina Qin, Zewen Chen, Libin Wang, Dandan Zheng, Yuming Li, Yufan Liu, Bing Li, and Weiming Hu. 2025. Visual-Instructed Degradation Diffusion for All-in-One Image Restoration. In *Proceedings of the Computer Vision and Pattern Recognition Conference*. 12764–12777.
- [37] Ziwei Luo, Fredrik K Gustafsson, Zheng Zhao, Jens Sjölund, and Thomas B Schön. 2024. Controlling Vision-Language Models for Multi-Task Image Restoration. In *ICLR*.
- [38] Kede Ma, Zhengfang Duanmu, Qingbo Wu, Zhou Wang, Hongwei Yong, Hongliang Li, and Lei Zhang. 2016. Waterloo exploration database: New challenges for image quality assessment models. *IEEE Transactions on Image Processing* 26, 2 (2016), 1004–1016.
- [39] Long Ma, Tengyu Ma, Risheng Liu, Xin Fan, and Zhongxuan Luo. 2022. Toward fast, flexible, and robust low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 5637–5646.
- [40] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. 2001. 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*, Vol. 2. IEEE, 416–423.
- [41] Chong Mou, Qian Wang, and Jian Zhang. 2022. Deep generalized unfolding networks for image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 17399–17410.
- [42] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. 2017. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3883–3891.
- [43] Ozan Özdenizci and Robert Legenstein. 2023. Restoring vision in adverse weather conditions with patch-based denoising diffusion models. *IEEE transactions on pattern analysis and machine intelligence* 45, 8 (2023), 10346–10357.
- [44] Yizhong Pan, Xiao Liu, Xiangyu Liao, Yuanzhouhan Cao, and Chao Ren. 2023. Random sub-samples generation for self-supervised real image denoising. In *Proceedings of the IEEE/CVF international conference on computer vision*. 12150–12159.
- [45] Yizhong Pan, Chao Ren, Xiaohong Wu, Jie Huang, and Xiaohai He. 2022. Real image denoising via guided residual estimation and noise correction. *IEEE Transactions on Circuits and Systems for Video Technology* 33, 4 (2022), 1994–2000.
- [46] Dongwon Park, Dong Un Kang, Jisoo Kim, and Se Young Chun. 2020. Multi-temporal recurrent neural networks for progressive non-uniform single image deblurring with incremental temporal training. In *European conference on computer vision*. Springer, 327–343.

- [47] Vaishnav Potlapalli, Syed Waqas Zamir, Salman Khan, and Fahad Khan. 2023. PromptIR: Prompting for All-in-One Image Restoration. In *Thirty-seventh Conference on Neural Information Processing Systems*.
- [48] Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong Xie, and Huizhu Jia. 2020. FFA-Net: Feature fusion attention network for single image dehazing. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 34. 11908–11915.
- [49] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*. PmlR, 8748–8763.
- [50] Chao Ren, Xiaohai He, Chunheng Wang, and Zhibo Zhao. 2021. Adaptive consistency prior based deep network for image denoising. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 8596–8606.
- [51] Chao Ren, Yizhong Pan, and Jie Huang. 2022. Enhanced latent space blind model for real image denoising via alternative optimization. *Advances in Neural Information Processing Systems* 35 (2022), 38386–38399.
- [52] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*. Springer, 234–241.
- [53] Yuanjie Shao, Lerenhan Li, Wenqi Ren, Changxin Gao, and Nong Sang. 2020. Domain adaptation for image dehazing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2808–2817.
- [54] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).
- [55] Ying Tai, Jian Yang, Xiaoming Liu, and Chunyan Xu. 2017. Memnet: A persistent memory network for image restoration. In *Proceedings of the IEEE international conference on computer vision*. 4539–4547.
- [56] Xiangpeng Tian, Xiangyu Liao, Xiao Liu, Meng Li, and Chao Ren. 2025. Degradation-Aware Feature Perturbation for All-in-One Image Restoration. In *Proceedings of the Computer Vision and Pattern Recognition Conference*. 28165–28175.
- [57] Jeya Maria Jose Valanarasu, Rajeev Yasarla, and Vishal M Patel. 2022. Trasnweather: Transformer-based restoration of images degraded by adverse weather conditions. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2353–2363.
- [58] Guoqing Wang, Changming Sun, and Arcot Sowmya. 2019. Erl-net: Entangled representation learning for single image de-raining. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 5644–5652.
- [59] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. 2022. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 17683–17693.
- [60] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. 2018. Deep Retinex Decomposition for Low-Light Enhancement. (2018).
- [61] Gang Wu, Junjun Jiang, Kui Jiang, Xianming Liu, and Liqiang Nie. 2025. Beyond Degradation Redundancy: Contrastive Prompt Learning for All-in-One Image Restoration. *arXiv preprint arXiv:2504.09973* (2025).
- [62] Gang Wu, Junjun Jiang, Yijun Wang, Kui Jiang, and Xianming Liu. 2025. Debaised all-in-one image restoration with task uncertainty regularization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 39. 8386–8394.
- [63] Haiyan Wu, Yanyun Qu, Shaohui Lin, Jian Zhou, Ruizhi Qiao, Zhizhong Zhang, Yuan Xie, and Lizhuang Ma. 2021. Contrastive learning for compact single image dehazing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 10551–10560.
- [64] Wenhui Wu, Jian Weng, Pingping Zhang, Xu Wang, Wenhan Yang, and Jianmin Jiang. 2022. Uretinex-net: Retinex-based deep unfolding network for low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 5901–5910.
- [65] Qiuhai Yan, Aiwen Jiang, Kang Chen, Long Peng, Qiaosi Yi, and Chunjie Zhang. 2025. Textual prompt guided image restoration. *Engineering Applications of Artificial Intelligence* 155 (2025), 110981.
- [66] Fuzhi Yang, Huan Yang, Jianlong Fu, Hongtao Lu, and Baining Guo. 2020. Learning texture transformer network for image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 5791–5800.
- [67] Hao Yang, Liyuan Pan, Yan Yang, and Wei Liang. 2024. Language-driven all-in-one adverse weather removal. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 24902–24912.
- [68] Zhiwen Yang, Haowei Chen, Ziniu Qian, Yang Yi, Hui Zhang, Dan Zhao, Bingzheng Wei, and Yan Xu. 2024. All-in-one medical image restoration via task-adaptive routing. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 67–77.
- [69] Xiaoyan Yu, Shen Zhou, Huafeng Li, and Liehuang Zhu. 2024. Multi-expert adaptive selection: Task-balancing for all-in-one image restoration. *IEEE Transactions on Circuits and Systems for Video Technology* (2024).
- [70] Eduard Zamfir, Zongwei Wu, Nancy Mehta, Yuedong Tan, Danda Pani Paudel, Yulun Zhang, and Radu Timofte. 2025. Complexity experts are task-discriminative learners for any image restoration. In *Proceedings of the Computer Vision and Pattern Recognition Conference*. 12753–12763.
- [71] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. 2022. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 5728–5739.
- [72] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. 2021. Multi-stage progressive image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 14821–14831.
- [73] Haijin Zeng, Xiangming Wang, Yongyong Chen, Jingyong Su, and Jie Liu. 2025. Vision-Language Gradient Descent-driven All-in-One Deep Unfolding Networks. In *Proceedings of the Computer Vision and Pattern Recognition Conference*. 7524–7533.
- [74] Jinghao Zhang, Jie Huang, Mingde Yao, Zizheng Yang, Hu Yu, Man Zhou, and Feng Zhao. 2023. Ingredient-oriented multi-degradation learning for image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 5825–5835.
- [75] Kaihao Zhang, Wenhan Luo, Yiran Zhong, Lin Ma, Bjorn Stenger, Wei Liu, and Hongdong Li. 2020. Deblurring by realistic blurring. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2737–2746.
- [76] Kai Zhang, Wangmeng Zuo, Shuhang Gu, and Lei Zhang. 2017. Learning deep CNN denoiser prior for image restoration. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3929–3938.
- [77] Yulun Zhang, Kunpeng Li, Kai Li, Bineng Zhong, and Yun Fu. 2019. Residual non-local attention networks for image restoration. *arXiv preprint arXiv:1903.10082* (2019).
- [78] Yurui Zhu, Tianyu Wang, Xueyang Fu, Xuanyu Yang, Xin Guo, Jifeng Dai, Yu Qiao, and Xiaowei Hu. 2023. Learning weather-general and weather-specific features for image restoration under multiple adverse weather conditions. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 21747–21758.