

Retrieve-to-Restore: Efficient All-in-One Image Restoration with a Retrieval-Based Degradation Bank

Chenxu Wang Kai Zhang* Jian Yang

School of Intelligence Science and Technology, Nanjing University, China

wangchenxu@smail.nju.edu.cn, kaizhang@nju.edu.cn, csjyang@nju.edu.cn

<https://github.com/cscxwang/R2R>

Abstract

All-in-one image restoration aims to recover clean images from heterogeneous degradations with a single model, but joint training on multiple degradations with a shared backbone often induces cross-task interference and unstable optimization, making it hard to maintain strong performance across all tasks. To address this, we propose Retrieve-to-Restore (R2R), a lightweight framework that decouples degradation adaptation from backbone computation through a retrieval-based degradation bank. Specifically, R2R externalizes degradation knowledge as unified, degradation-specific priors stored in a compact degradation bank. A degradation amalgamator aggregates intra-class features guided by ground truth into task-level clean priors during training, while degradation matching retrieves the most relevant prior at inference to modulate backbone features for restoration. This retrieval-guided design explicitly separates degradation cues from shared reconstruction capacity, enabling stable multi-degradation training and straightforward scaling to additional degradation types. Extensive comparisons on benchmarks with one, three, and five degradations show that R2R achieves PSNR on par with state-of-the-art all-in-one methods while using about 91% fewer MACs.

1. Introduction

Image restoration (IR) [44, 65, 68] is a foundational low-level vision task that reconstructs high-quality images from degraded observations. Adverse conditions such as noise [23, 65], haze [3, 39], and rain [5, 56] significantly impact the practical utility of images in downstream applications across various domains, including autonomous navigation [5, 47] and augmented reality [12, 18].

Deep learning has delivered remarkable advances, especially in task-specific IR where dedicated architectures are optimized for a single corruption type [4, 6, 19, 31, 50, 60].

*Corresponding author.

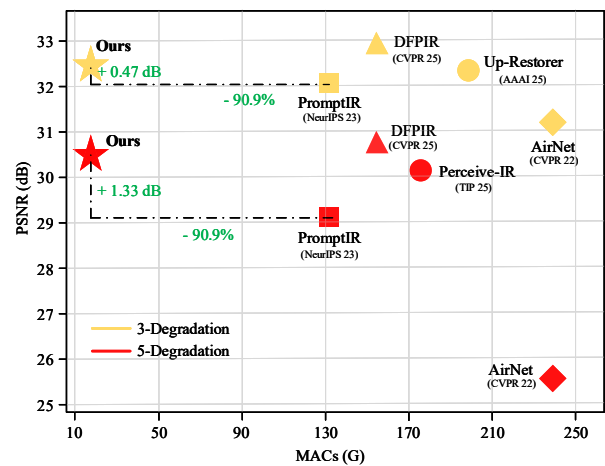


Figure 1. Comparison of average PSNR and computational cost (MACs) across different methods. Our method achieves a 0.47dB PSNR improvement while reducing MACs by approximately 91% compared to PromptIR [38].

More recently, all-in-one restoration has emerged, demonstrating that a single model can address multiple degradations and thus mitigate the need for extensive re-training as new corruption types appear [8, 14, 16, 26, 29, 38, 47, 63]. Representative approaches leverage visual [27, 38, 48] or language-based prompting [1, 8, 33], contrastive learning [26, 63], and diffusion-based models [1, 33].

Despite their success, we observe that the aforementioned models often suffer from inefficiencies: parameters specialized for one degradation can conflict with those required by others, leading to mutual interference during multi-degradation training [46]. Accordingly, we seek a single model that can stably handle multiple degradations while preserving update consistency and computational efficiency. Yet, existing approaches pursue consistency via internal modulation: some inject textual priors [46, 61, 67] or conditioning signals into a shared backbone, whereas others employ MoE routing [32, 55, 57, 58] to dispatch inputs to degradation-specific experts. Although effective, these

strategies enlarge the parameter space, introduce additional hyperparameters and routing dynamics, and can undermine the consistency of parameter updates across tasks, thereby destabilizing training. In contrast, we argue that heavy internal modulation is unnecessary, as degradations typically exhibit pronounced intra-class similarity and inter-class separability (for example, streak-like structures in rain, scene-wide low contrast and veiling in haze, and high-frequency fluctuations in noise). Building on this observation, we advocate an external prior view that summarizes each degradation by a compact and task-level descriptor.

Formalizing this perspective, let $\mathcal{T}$ denote the set of degradation families (*e.g.*, noise, haze, rain). In the multi-degradation setting, our objective is to employ a single model that: 1) explicitly separates degradation from shared reconstruction capacity to mitigate cross-task interference; 2) unifies parameters and maintains stable optimization, avoiding the instability that grows with $|\mathcal{T}|$; and 3) preserves computational efficiency while remaining extensible to new degradations. To this end, we represent each degradation as a unified, task-level prior and retrieve the prior at inference for on-demand adaptation. This separation, with adaptation handled by external priors and computation confined to a shared model, improves update consistency and scalability.

Therefore, we propose Retrieve-to-Restore (R2R), a lightweight all-in-one restoration network that achieves explicit degradation disentanglement and efficient adaptation. Specifically, we adopt a single NAF-style [4] U-shaped encoder-decoder as a task-agnostic backbone. To build a compact global degradation bank, we introduce a degradation amalgamator that aggregates clean priors from paired data into unified, task-level prototypes. At inference, a degradation matching module forms a query from encoded features and retrieves the corresponding prior from the bank; the retrieved prior is fused with the encoded features to guide reconstruction and produce a sharp restored image. The emphasis is on efficiency rather than performance stacking, achieving highly competitive results with minimal computation. Through extensive experiments across multiple image restoration tasks, we demonstrate that our approach not only preserves restoration quality but surpasses recent state-of-the-art methods. These results validate both the effectiveness of our network design and its robustness when scaling to additional tasks.

Our contributions can be summarized as follows:

- We propose Retrieve-to-Restore (R2R), a lightweight all-in-one restorer that externalizes degradation knowledge as degradation-specific unified priors and follows an Encode-Retrieve-Decode paradigm.
- A degradation amalgamator builds a compact degradation bank, and degradation matching retrieves the relevant prior at inference, clearly separating degradation cues from shared reconstruction.

- R2R achieves PSNR comparable to state-of-the-art methods while using 91% fewer MACs, and it remains stable as the number of tasks increases.

2. Related Work

Task Specific Image Restoration. Reconstructing a clean image from its degraded observation is a highly ill-posed problem, yet a large body of data-driven methods has significantly advanced image restoration beyond traditional hand-crafted priors [23, 28, 44, 50, 60, 65, 68]. Most approaches adopt CNN-based [4, 44, 65, 68] or Transformer-based architectures [6, 31, 50, 60] tailored to a single degradation type, such as denoising [6, 65, 68], dehazing [42, 43, 53], or deraining [21, 40]. While Transformer-based restorers offer strong capacity for modeling global dependencies and thus deliver impressive performance [31, 60, 69], their self-attention mechanisms introduce quadratic or window-based complexity with respect to image size, which is often prohibitive in resource-constrained or latency-sensitive scenarios. Conversely, convolutional architectures provide fast and scalable processing with favorable computational and memory costs, making them highly attractive for real-world deployment, despite their more limited global receptive field. Motivated by this trade-off, we build upon a lightweight CNN-based architecture [4] and incorporate a task-level degradation prior bank, enabling the model to efficiently capture both shared and degradation-specific representations while retaining practical efficiency.

All-in-One Image Restoration. Training a separate model for each degradation is effective but quickly becomes impractical: real-world images often exhibit mixed or unknown corruptions, while maintaining multiple task-specific networks inflates storage, scheduling complexity, and energy consumption. All-in-one image restoration addresses this by learning a single blind model that handles diverse degradations within a unified framework. Early all-in-one approaches, such as AirNet [26], learn degradation-aware representations via contrastive objectives and inject them into a shared backbone for blind restoration. IDR [63] further decomposes complex corruptions into physically meaningful ingredients and optimizes a two-stage meta-learning scheme to improve generalization across degradation types. More recently, prompt-based methods [27, 38, 46, 48] introduce learnable visual prompts to modulate a unified backbone for degradation-aware restoration. DiffUIR [70] formulates universal restoration with a diffusion-based framework, while InstructIR [8] performs high-quality restoration conditioned on human-written instructions. However, such methods typically require iterative sampling or textual guidance, resulting in higher latency and computational cost. Other extensions further explore frequency-domain prompts [10], richer context encoders [33, 61, 67], auxiliary datasets [15, 57], and

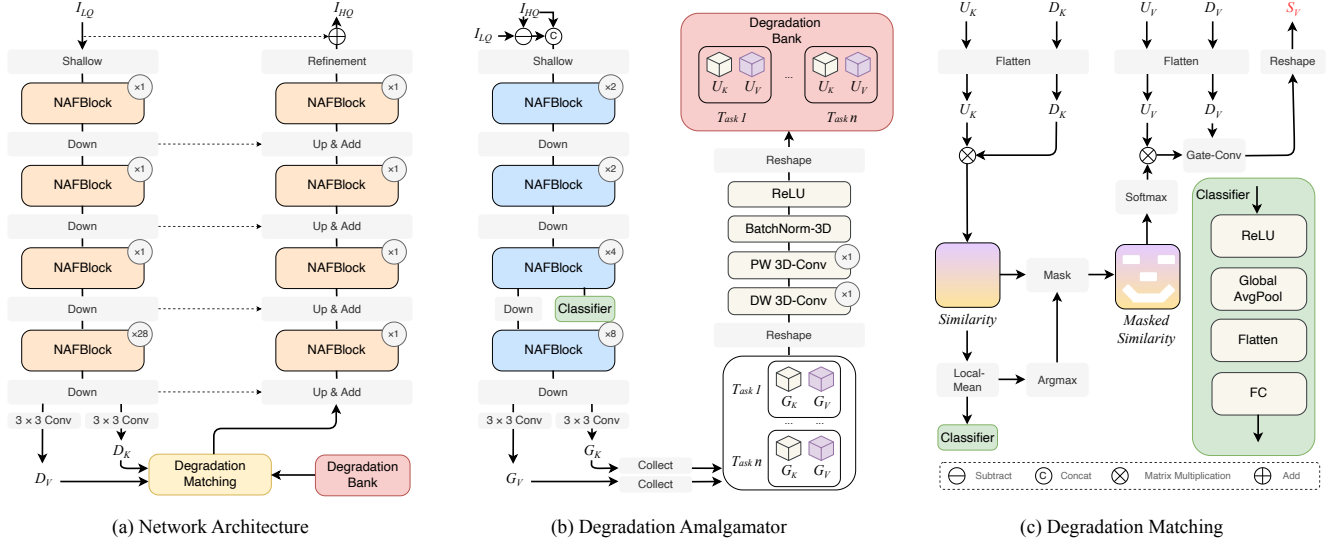


Figure 2. (a) Overall framework of the proposed R2R. R2R mainly consists of NAFBlocks [4] based U-shaped encoder–decoder, a degradation amalgamator, and a degradation matching module. (b) The degradation amalgamator identifies degradation types and encodes the corresponding clean images into a unified prior space to construct the degradation bank. (c) The degradation matching module enables the degraded input to query the bank and retrieve relevant clean priors to guide restoration.

MoE-based designs [32, 55, 57, 58]. In contrast, we pursue an explicitly disentangled framework that retains the simplicity of a single lightweight backbone while leveraging degradation-aware priors to guide adaptation. By decoupling shared reconstruction capacity from degradation-specific cues, our design enhances scalability without relying on bulky prompt stacks or large expert ensembles.

Degradation Disentanglement. A key challenge in all-in-one restoration is to decouple degradation factors from the backbone’s reconstruction capacity in order to limit cross-task interference. Prior work typically injects conditioning inside the backbone. DFPIR [46] uses CLIP-guided texts to shuffle channels, MoCE-IR [58] employs gating and routing to expert branches, and AdaIR [10] introduces prompts in the frequency domain. These choices modify parameters or enlarge the architecture. Conversely, we adopt an external-prior perspective: rather than altering network parameters or expanding the backbone with prompts or experts, we establish a compact bank of task-level priors and retrieve the relevant prior for each input. This preserves a single lightweight backbone while clearly separating shared reconstruction capacity from degradation-specific cues.

3. Network Architecture

In this section, we outline the core principles of our all-in-one image restoration method. As depicted in Fig. 2(a), we adopt a U-shaped backbone with a symmetric encoder–decoder structure. The degraded input is first mapped to shallow features via a 3×3 convolution, followed by four hierarchical encoding and decoding stages based on NAFBlocks [4].

Unlike conventional U-Net style architectures, our approach injects the proposed degradation bank (Fig. 2(b)) and degradation matching (Fig. 2(c)) solely at the final downsampling stage of the encoder. Specifically, two additional 3×3 convolutions produce feature maps $D_K \in \mathbb{R}^{B \times H \times W \times C_k}$ and $D_V \in \mathbb{R}^{B \times H \times W \times C_v}$, where B , H , and W denote the batch size, height, and width, respectively, and C_k , C_v denote channel dimensions. These features are used to retrieve degradation-aware information from the degradation bank, enabling effective representation learning without increasing architectural complexity.

3.1. Degradation Amalgamator

To mitigate interference among heterogeneous degradations in all-in-one image restoration, we propose a degradation amalgamator coupled with a compact degradation bank. The amalgamator explicitly disentangles degradation-specific information and encodes it into the bank, from which the restoration network adaptively retrieves clean priors conditioned on the input degradation, thereby enabling stable and unified restoration.

The degradation amalgamator learns degradation-specific representations that act as clean priors during inference. To achieve this, it is trained with access to paired images and is explicitly encouraged to distinguish different degradation types. Technically, as shown in Fig. 2(b), we first standardize all cases by computing the residual between the degraded image I_{LQ} and its ground-truth image I_{HQ} , and concatenating this residual with I_{HQ} along the channel dimension. The concatenated tensor is processed by a compact backbone of stacked NAFBlocks [4] with pro-

gressive downsampling to extract concise degradation-aware features. Two terminal 3×3 convolutions produce embeddings $G_K \in \mathbb{R}^{B \times H \times W \times C_k}$ and $G_V \in \mathbb{R}^{B \times H \times W \times C_v}$, which are later merged into key-value pairs stored in the degradation bank. Following [20], we insert a lightweight degradation classification head after the third NAFBlock stage to supervise degradation recognition and strengthen feature separability; this auxiliary head is used only during training and incurs no additional inference cost. In G_K and G_V , the batch dimension mixes samples from different degradation types. Guided by degradation labels, features are reorganized by task: for each i -th task, the corresponding samples are gathered and concatenated along the batch dimension, giving tensors of size $b_i \times H \times W \times C$, where C denotes C_k or C_v , $b_i \leq B$. Since b_i varies, the sequences are replicated to a fixed length M ($M > \forall b_i$, the effect of this hyperparameter M is reported in Table 5), yielding a unified shape $M \times H \times W \times C$ per task. Task-wise 3D operations, including depthwise and pointwise 3D convolutions, BatchNorm3D, and ReLU, are then applied along the batch axis to aggregate information within each degradation type, producing compact embeddings $U_K \in \mathbb{R}^{M \times H \times W \times C_k}$ and $U_V \in \mathbb{R}^{M \times H \times W \times C_v}$ that are stored in the degradation bank as task-level priors for retrieval.

3.2. Degradation Matching

For each degradation type, the corresponding clean images are explicitly encoded into its entry in the degradation bank, and a test sample is expected to retrieve a compatible degradation prior from this bank to guide restoration. However, the embeddings $D_K \in \mathbb{R}^{B \times H \times W \times C_k}$ and $D_V \in \mathbb{R}^{B \times H \times W \times C_v}$ also mix different degradations along the batch dimension, making it impossible to pre-assign which task-specific prior should be queried for each sample. To resolve this, a global matching scheme is adopted to identify, for each input, the most similar task-level prior in the bank, which is then used to condition the subsequent restoration. Specifically, each task in the degradation bank is associated with a key-value pair. For n tasks, all keys and all values are respectively concatenated along the task dimension and flattened into global vectors $U_K \in \mathbb{R}^{[n \times M, H \times W \times C_k]}$ and $U_V \in \mathbb{R}^{[n \times M, H \times W \times C_v]}$. As shown in Fig. 2(c), the flattened query feature $D_K \in \mathbb{R}^{[B, H \times W \times C_k]}$ is then matched against $U_K \in \mathbb{R}^{[n \times M, H \times W \times C_k]}$ via matrix multiplication to compute similarity S :

$$S = D_K \times U_K^T, \quad (1)$$

where $S \in \mathbb{R}^{[B, n \times M]}$ denotes the global task-level similarity matrix, with each row encoding the similarity between a degraded sample and all task embeddings. Each contiguous window of length M corresponds to one specific task.

To mitigate interference between tasks during matching, a local averaging scheme is used to infer the task assignment

of each degraded sample. For each row of S , local mean similarities are computed over windows of length M , the task with the highest averaged score is selected, and all similarity entries associated with other tasks are masked out:

$$S_{mask} = \text{Mask}(\text{Argmax}(\text{Local-Mean}(S))), \quad (2)$$

where S_{mask} is the masked similarity matrix, and entries for irrelevant tasks are set to a large negative value. Row-wise Softmax is applied to S_{mask} after masking out entries for irrelevant tasks, so that only similarities associated with candidate task priors remain active. These weights are then used to multiply the bank features U_V , yielding for each sample a retrieved degradation prior with the same shape as D_V . The retrieved prior is fused with the query feature D_V via channel-interleaved concatenation along the channel dimension and passed through a gated convolution module to produce the final sharp prior S_V :

$$S_V = \text{Gate-Conv}(U_V \times \text{Softmax}(S_{mask}), D_V), \quad (3)$$

where Gate-Conv is implemented as a 3×3 grouped convolution with C_v groups, ensuring that each degraded sample interacts only with its corresponding retrieved prior. It is worth noting that during degradation matching, the task assignment derived from the local mean operation is also fed into an auxiliary classification head, which further regularizes the network to distinguish different degradation types.

3.3. Loss Function

Beyond the network designs, the loss function helps maximize the potential of our approach. We use a combination of distortion losses and classification losses to optimize our model. First, to ensure high-fidelity (low distortion), we use L_{pixel} defined as: $L_{pixel} = \|x - \hat{x}\|_1$, where $\hat{x}$ and x are the restored and the ground-truth (clean) images, respectively.

To encourage effective disentanglement and degradation recognition, we apply two cross-entropy classification losses: L_{deg} on the degradation amalgamator backbone and L_{match} on the matching stage to guide correct prior retrieval.

To facilitate high fidelity in both spatial and frequency domains, we adopt dual-domain $\mathcal{L}_1$ loss functions following [7, 9]. For final restored/clean image pair with the same resolution, the loss functions are given by

$$\mathcal{L}_{fft} = \frac{1}{P} \|\mathcal{F}(\hat{x}) - \mathcal{F}(x)\|_1, \quad (4)$$

where P indicates total elements for normalization, $\mathcal{F}$ represents the fast Fourier transform.

The complete loss function is then given by

$$\mathcal{L} = \mathcal{L}_{pixel} + \lambda_d \cdot \mathcal{L}_{deg} + \lambda_m \cdot \mathcal{L}_{match} + \lambda_f \cdot \mathcal{L}_{fft}, \quad (5)$$

where the constants λ_d , λ_m and λ_f are loss weights empirically set to 0.1, 0.1 and 0.125, respectively.

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

Method	MACs.	<i>Dehazing</i>		<i>Deraining</i>		<i>Denoising</i>						Average	
		SOTS		Rain100L		BSD68 $_{\sigma=15}$		BSD68 $_{\sigma=25}$		BSD68 $_{\sigma=50}$			
BRDNet [45]	-	23.23	.895	27.42	.895	32.26	.898	29.76	.836	26.34	.693	27.80	.843
LPNet [17]	-	20.84	.828	24.88	.784	26.47	.778	24.77	.748	21.26	.552	23.64	.738
FDGAN [13]	-	24.71	.929	29.89	.933	30.25	.910	28.81	.868	26.43	.776	28.02	.883
DL [16]	-	26.92	.931	32.62	.931	33.05	.914	30.41	.861	26.90	.740	29.98	.876
MPRNet [59]	149G	25.28	.955	33.57	.954	33.54	.927	30.89	.880	27.56	.779	30.17	.899
AirNet [26]	238G	27.94	.962	34.90	.967	33.92	.933	31.26	.888	28.00	.797	31.20	.910
PromptIR [38]	132G	30.58	.974	36.37	.972	33.98	.933	31.31	.888	28.06	.799	32.06	.913
DA-CLIP [33]	-	29.46	.963	36.28	.968	30.02	.821	24.86	.585	22.29	.476	-	-
Gridformer [49]	251G	30.37	.970	37.15	.972	33.93	.931	31.37	.887	28.11	.801	32.19	.912
Up-Restorer [30]	-	30.68	.977	36.74	.978	33.99	.933	31.33	.888	28.07	.799	32.16	.915
R2R (<i>ours</i>)	12G	31.40	.977	37.46	.980	34.10	.936	31.45	.895	28.22	.806	32.53	.918

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

Method	MACs.	<i>Dehazing</i>		<i>Deraining</i>		<i>Denoising</i>		<i>Deblurring</i>		<i>Low-Light</i>		Average	
		SOTS		Rain100L		BSD68 $_{\sigma=25}$		GoPro		LOL-v1			
NAFNet* [4]	-	25.23	.939	35.56	.967	31.02	.883	26.53	.808	20.49	.809	27.76	.881
DGUNet* [36]	-	24.78	.940	36.62	.971	31.10	.883	27.25	.837	21.87	.823	28.32	.891
SwinIR* [28]	-	21.50	.891	30.78	.923	30.59	.868	24.52	.773	17.81	.723	25.04	.835
Restormer* [60]	141G	24.09	.927	34.81	.962	31.49	.884	27.22	.829	20.41	.806	27.60	.881
MambaIR* [19]	-	25.81	.944	36.55	.971	31.41	.884	28.61	.875	22.49	.832	28.97	.901
DL [16]	-	20.54	.826	21.96	.762	23.09	.745	19.86	.672	19.83	.712	21.05	.743
Transweather [47]	-	21.32	.885	29.43	.905	29.00	.841	25.12	.757	21.21	.792	25.22	.836
TAPE [29]	-	22.16	.861	29.67	.904	30.18	.855	24.47	.763	18.97	.621	25.09	.801
AirNet [26]	238G	21.04	.884	32.98	.951	30.91	.882	24.35	.781	18.18	.735	25.49	.847
IDR [63]	98G	25.24	.943	35.63	.965	31.60	.887	27.87	.846	21.34	.826	28.34	.893
PromptIR [38]	132G	26.54	.949	36.37	.970	31.47	.886	28.71	.881	22.68	.832	29.15	.904
Gridformer [49]	251G	26.79	.951	36.61	.971	31.45	.885	29.22	.884	22.59	.831	29.33	.904
Perceive-IR [67]	-	28.19	.964	37.25	.977	31.44	.887	29.46	.886	22.88	.833	30.11	.905
R2R (<i>ours</i>)	12G	30.64	.974	36.61	.975	31.35	.891	30.93	.911	22.88	.856	30.48	.921

4. Experiments and Comparisons

We conduct experiments adhering to the protocols of prior general image restoration works [38, 63] under two different settings: (i) *All-in-One* and (ii) *Single-task*. In the All-in-One setting, a unified model is trained to handle multiple degradation types, where we consider *three* and *five* distinct degradations. In the Single-task setting, separate models are trained for each specific restoration task.

Implementation Details. Our R2R framework is end-to-end trainable, requiring no multi-stage optimization of any components. The architecture features a symmetric, four-level encoder-decoder design. The encoder consists of four levels with varying numbers of NAFBlocks [4], arranged as [1, 1, 1, 28] from the top level to the lowest. The decoder includes four levels, containing [1, 1, 1, 1] NAFBlocks. For the degradation amalgamator backbone, we adopt a four-level NAF-style encoder with [2, 2, 4, 8] NAFBlocks, which

is used only during training to construct the degradation bank and is removed at inference. C_k and C_v are set to 64 and 512, respectively. We conduct our experiments using PyTorch on a single NVIDIA GeForce RTX 5090 GPU, requiring less than 40 hours. For training, we run for 240 epochs with an initial learning rate of 2×10^{-4} , followed by fine-tuning for 30 epochs at a learning rate of 1×10^{-6} . Initially, we set the patch size to 128^2 with the batch size of 64, and for fine-tuning, the patch size is adjusted to 224^2 . The network is optimized using Adam [22] (with parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$). We train on cropped patches and augment the dataset with random horizontal and vertical flips.

Datasets. For all-in-one, we follow existing work [26, 38] and include the following datasets. For image denoising, we combine the BSD400 [2] and WED [34] datasets, adding Gaussian noise at levels $\sigma \in [15, 25, 50]$ to create noisy images. Testing is conducted on the BSD68 [35]. For deraining, we use Rain100L [54]. The dehazing task utilizes the SOTS

Table 3. Comparison to state-of-the-art for single degradations. PSNR (dB, $\uparrow$) and SSIM ($\uparrow$) metrics are reported on the full RGB images. **Best** and **second best** performances are highlighted. Our method excels prior work on dehazing and deraining.

(a) Dehazing			(b) Deraining			(c) Denoising on CBSD68				
Method	MACs.	SOTS	Method	MACs.	Rain100L	Method	MACs.	$\sigma=15$	$\sigma=25$	$\sigma=50$
DehazeNet [3]	-	22.46 .851	DIDMDN [62]	-	23.79 .773	CBM3D [11]	-	33.50 .922	30.69 .868	27.36 .763
MSCNN [41]	-	22.06 .908	UMR [56]	-	32.39 .921	DnCNN [64]	-	33.89 .930	31.23 .883	27.92 .789
AODNet [24]	-	20.29 .877	SIRR [52]	-	32.37 .926	IRCNN [65]	-	33.87 .929	31.18 .882	27.88 .790
EPDN [39]	-	22.57 .863	MSPFN [21]	-	33.50 .948	FFDNet [66]	-	33.87 .929	31.21 .882	27.96 .789
FDGAN [13]	-	23.15 .921	LPNet [17]	-	23.15 .921	BRDNet [45]	-	34.10 .929	31.43 .885	28.16 .794
AirNet [26]	238G	23.18 .900	AirNet [26]	238G	34.90 .977	AirNet [26]	238G	34.14 .936	31.48 .893	28.23 .806
PromptIR [38]	132G	31.31 .973	PromptIR [38]	132G	37.04 .979	PromptIR [38]	132G	34.34 .938	31.71 .897	28.49 .813
R2R (ours)	12G	31.50 .978	R2R (ours)	12G	37.45 .980	R2R (ours)	12G	34.27 .938	31.64 .897	28.42 .813

[25] dataset. For deblurring and low-light enhancement, we employ the GoPro [37] and the LOL-v1 [51] datasets, respectively. To develop a unified model for all tasks, we merge these datasets in a *three* (AIO-3, $n = 3$) or *five* (AIO-5, $n = 5$) degradation setting. For single-task experiments, our method is trained on the respective training set.

4.1. Comparison to State-of-the-Art Methods

All-in-One: three degradations. We evaluate our All-in-One restorer, R2R, against other specialized methods, including BRDNet [45], LPNet [17], FDGAN [13], DL [16], MPRNet [59], AirNet [26], PromptIR [38], Gridformer [49] and Up-Restorer [30]. All methods were trained simultaneously on three types of degradation: haze, rain, and Gaussian noise. The results in Table 1 show that our proposed model, R2R, consistently outperforms previous works in both performance and efficiency. It achieves the highest scores across all benchmarks, with an average PSNR gain of 1.33dB over AirNet [26] while using 95% fewer MACs, and 0.47dB over PromptIR [38] with 90.9% fewer MACs, validating the efficiency and effectiveness of our approach.

All-in-One: five degradations. Extending the three degradation tasks to include deblurring and low-light enhancement [26, 63], we validate our method’s comprehensive performance in an All-in-One setting. As shown in Table 2, R2R effectively leverages degradation-specific features, surpassing AirNet [26], IDR [63] and Gridformer [49] by an average of 4.99dB, 2.14dB, and 1.15dB, respectively, with 95%, 88% and 95% fewer MACs. Against general image restorers, R2R outperforms Restormer [60] and MambaIR [19] on the GoPro dataset by 3.71dB and 2.32dB, while being $12\times$ more efficient than Restormer.

Single degradation. We present results in Table 3 with R2R trained on individual degradation tasks: dehazing, deraining, and denoising. For dehazing, R2R achieves the best PSNR and the highest SSIM. In deraining, our model surpasses the previous state-of-the-art, PromptIR [38], by 0.41dB in PSNR. For denoising, it ranks second overall, but with substantial efficiency gains compared to PromptIR [38] and AirNet [26].

Resilience to task interference. By explicitly disen-

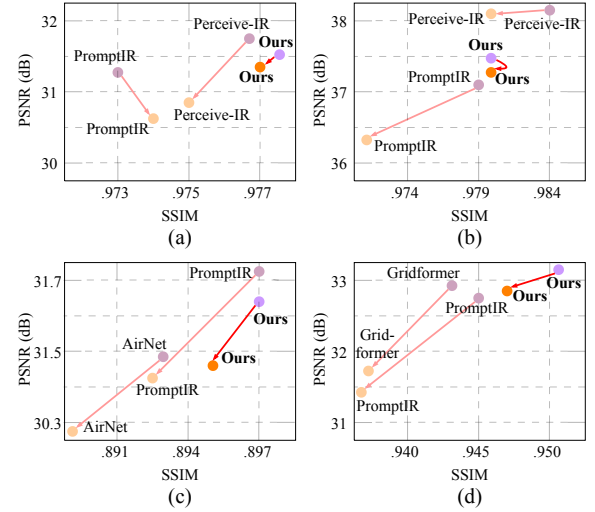


Figure 3. Task robustness of R2R and state-of-the-art methods as the number of degradation types increases. (a)–(c) denote the performance change from single-task dehazing, deraining, and denoising to the *three*-task setting; (d) denotes the extension from *three*-task to *five*-task restoration.

gling different degradation types, R2R endows the all-in-one model with strong robustness as the number of tasks increases. We analyze the behavior of different methods when scaling a unified model from single to multi-task restoration. Fig. 3(a)–(c) show the results of extending single-task dehazing, deraining, and denoising to a *three*-task setting. Although AirNet [26], PromptIR [38], Gridformer [49] and Perceive-IR [67] use more MACs than R2R, they suffer clear performance drops, while our method shows the smallest degradation. This stability comes from explicitly disentangling degradation and using task-level priors in the degradation bank to reduce cross-degradation interference. Fig. 3(d) further evaluates the extension from *three* to *five* tasks ($\sigma = 25$) and shows a consistent trend, confirming the robustness of R2R as the number of tasks increases.

Visual results. Fig. 4 shows visual comparisons under the three degradation settings with different methods, including AirNet [26], PromptIR [38], InstructIR [26], VLU-Net [61]

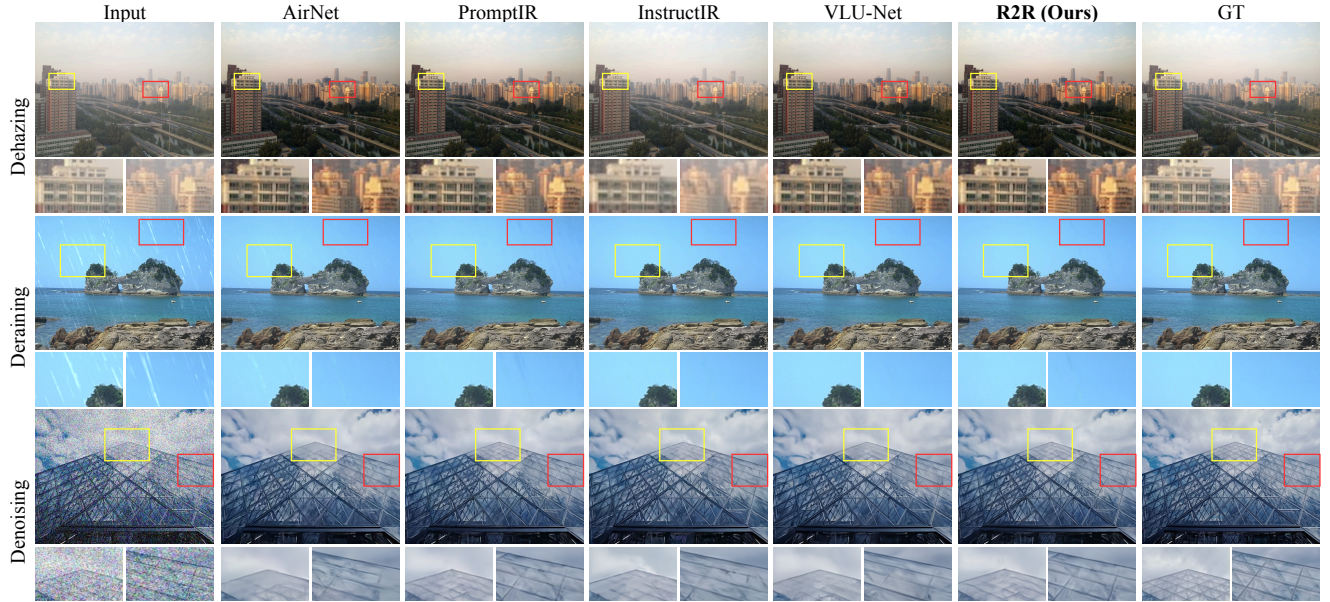


Figure 4. Visual comparison of R2R with state-of-the-art methods considering three degradations. Zoom in for a better view.

Table 4. Complexity Analysis. MACs are computed on an input image of size 224×224 using a NVIDIA RTX 5090 (32G) GPU. * denotes patch size of 512×512 .

Method	PSNR (dB, $\uparrow$)	Memory ($\downarrow$)	Params. ($\downarrow$)	MACs ($\downarrow$)
AirNet [26]	31.20	4829M	8.93M	238G
PromptIR [38]	32.06	9830M	35.59M	132G
IDR [63]	-	4905M	15.34M	98G
Restormer* [60]	-	30032M	26.11M	141G
InstructIR [8]	32.43	-	15.8M	100G
Gridformer* [49]	32.19	28462M	30.1M	251G
R2R (ours)	32.53	846M	19.7M	12G

and our R2R. In certain demanding hazy scenarios, both AirNet [26] and PromptIR [38] reveal constraints in completely eradicating haziness, resulting in noticeable color intensity discrepancies, whereas our approach ensures precise color reconstruction. Furthermore, in challenging rainy scenes, these popular approaches continue to exhibit notable remnants of rain, which our method adeptly eliminates, distinguishing itself from other approaches. Additionally, our method also generates clear and sharp denoised outputs. Note that PromptIR is approximately $10\times$ more computationally complex than ours. Despite this, our method consistently produces visually superior results, demonstrating its effectiveness. Coupled with the quantitative comparisons, these findings underscore the overall superiority of our method. *Please refer to the supplementary for more visual results.*

Model efficiency. Table 4 compares memory usage, MACs, and parameters, highlighting our framework’s efficiency over state-of-the-art all-in-one restorers. With an explicitly disentangled architecture and the proposed degradation bank

module, R2R achieves a 0.47dB PSNR improvement over PromptIR [38] while using 91.4% less memory, 44.6% fewer parameters, and only 12G MACs, making it **90.9%** more computationally efficient. This efficiency stems from our lightweight network design and the ability of degraded images to adaptively retrieve clean priors from the degradation bank. This adaptive design, along with lower memory usage and fewer parameters, gives R2R a clear advantage over existing methods.

4.2. Ablation Studies

We conduct detailed studies on the proposed components within our R2R framework. All experiments are conducted in the *All-in-One* setting with *three* degradations. *Please refer to the supplementary for additional ablations.*

Hyperparameter M . We first examine the effect of the hyperparameter M , which controls the capacity of the degradation bank, *i.e.*, the number of clean priors stored per degradation type. As shown in Table 5, we set $M \in \{24, 32, 64, 128\}$ and train R2R under each configuration. The best performance is obtained at $M = 64$, indicating a trade-off between bank capacity and restoration quality: simply enlarging M does not guarantee more useful priors and may introduce redundancy. We attribute this to the replication operation in the degradation amalgamator, which helps capture shared degradation patterns but can lead to overfitting when the bank becomes excessively large. *More detailed analysis is provided in the supplementary material.*

Visualization of degradation types. Fig. 5 presents the visual results of the degradation maps generated by R2R for rain and haze. The results indicate that the proposed method

Table 5. Ablation study of the Hyperparameter M under the 3-degradation task. The average PSNR and SSIM are reported.

Method	SOTS		Rain100L		BSD68 ₁₅₋₅₀		Avg.	
$M = 24$	30.44	.974	37.33	.980	31.25	.878	32.30	.914
$M = 32$	30.97	.976	37.32	.980	31.26	.879	32.41	.918
$M = 64$	31.40	.977	37.46	.980	31.26	.879	32.53	.918
$M = 128$	30.81	.976	37.23	.978	31.22	.877	32.34	.917

Table 6. Effectiveness of the different loss functions under the 3-degradation task. (a), (b), (c), (d) denote $\mathcal{L}_{pixel}$, $\mathcal{L}_{fft}$, $\mathcal{L}_{deg}$, $\mathcal{L}_{match}$, respectively. The average PSNR and SSIM are reported.

Method	SOTS		Rain100L		BSD68 ₁₅₋₅₀		Avg.	
(a)	30.20	.974	36.09	.973	31.04	.873	31.88	.913
(a)+(b)	30.30	.974	37.12	.978	31.24	.878	32.23	.917
(a)+(b)+(c)	30.70	.976	37.02	.977	31.25	.878	32.29	.917
(a)+(b)+(c)+(d)	31.40	.977	37.46	.980	31.26	.879	32.53	.918

Table 7. Results on HQ dependence in the Degradation Bank (DB).

Inputs	Supervision	SOTS	Rain100L	BSD68 ₁₅₋₅₀	Avg.				
{HQ*, HQ* - LQ}	HQ*	29.90	.973	36.71	.976	31.12	.877	31.99	.916
{LQ}	HQ	30.57	.977	37.33	.979	31.25	.878	32.33	.918
{HQ, HQ - LQ} (<i>ours</i>)	HQ	31.40	.977	37.46	.980	31.26	.879	32.53	.918



Figure 5. Visual degradation of our method R2R for rain and haze.

captures distinct degradation distributions across different input types, confirming its ability to accurately characterize degradation patterns.

Effects of different loss functions. As shown in Table 6, we evaluate the effectiveness of R2R under different loss functions. The base model uses only pixel-level reconstruction loss $\mathcal{L}_{pixel}$. Specifically, compared to the base, R2R incorporating $\mathcal{L}_{fft}$, $\mathcal{L}_{deg}$ and $\mathcal{L}_{match}$ simultaneously achieves a PSNR improvement of 0.65dB. By emphasizing fidelity in both spatial and frequency domains, the utilization of $\mathcal{L}_{fft}$ results in a PSNR improvement of 0.35dB, over the use of $\mathcal{L}_{pixel}$ alone. More importantly, the classification loss $\mathcal{L}_{deg}$ brings a PSNR gain of about 0.40dB on dehazing (SOTS), while $\mathcal{L}_{match}$ for supervising correct prior retrieval, yields a further average improvement of 0.24dB. We plot the t-SNE in Fig. 6 to demonstrate the effect of $\mathcal{L}_{match}$ on the degradation bank. It is evident that $\mathcal{L}_{match}$ leads to a tighter clustering of U_K and U_V , which helps the model more accurately and discriminatively retrieve effective priors. Fig. 7 illustrates the visual results of various schemes using individual loss modules. As the proposed loss modules are successively added, the restored image more closely resembles the ground truth.

Flexible DB construction without clean HQ. The proposed DB can be constructed without strict dependence on clean

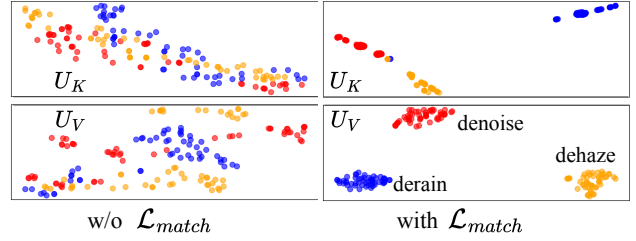


Figure 6. t-SNE visualization of U_K and U_V under the three-task setting. $\mathcal{L}_{match}$ enables each task to exhibit tighter clustering. Please refer to the supplementary for more visuals.

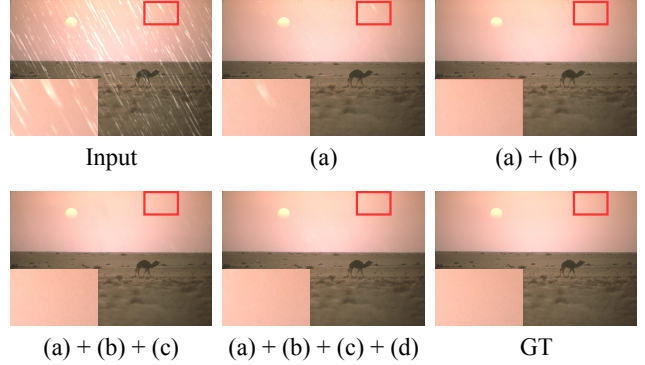


Figure 7. Visual comparison of the restored images obtained using the individual loss schemes and the proposed scheme.

HQ data. Specifically, Table 7 shows that replacing clean HQ with Gaussian-noised HQ* ($\sigma = 15$) and its residual causes only a 0.54dB PSNR drop. Moreover, constructing the DB directly from LQ also delivers performance comparable to that of using clean HQ. These results demonstrate the flexibility of the proposed DB construction strategy and highlight its potential for practical deployment under limited clean data conditions.

5. Conclusion

This paper presented R2R, an efficient all-in-one image restoration framework that unifies diverse degradations within a single lightweight backbone through explicit degradation disentanglement and retrieval-based priors. R2R introduces a degradation amalgamator to encode task-level clean priors into a compact degradation bank, and a degradation matching module to adaptively retrieve the most relevant prior for each input. This design enables degradation-aware restoration without relying on heavy prompt engineering or mixture-of-experts routing, and substantially reduces interference among heterogeneous degradations during joint training. Extensive experiments showed that R2R achieves competitive performance with significantly fewer MACs and maintains the smallest performance drop when scaling from single to multiple degradations and increasing the number of tasks, confirming it as an effective and computationally efficient solution for unified image restoration.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (Grant No. 62572234), the Gusu Innovation and Entrepreneurship Leading Talent Program (Grant No. ZXL20254324), and the Suzhou Key Technologies Project (Grant No. SGY2023136).

References

- [1] Yang Ai, Huaibo Huang, Xiaoqiang Zhou, Jiexiang Wang, and Ran He. Multimodal prompt perceiver: Empower adaptiveness generalizability and fidelity for all-in-one image restoration. In *CVPR*, pages 25432–25444, 2024. 1
- [2] Pablo Arbelaez, Michael Maire, Charles Fowlkes, and Jitendra Malik. Contour detection and hierarchical image segmentation. *IEEE TPAMI*, 33(5):898–916, 2010. 5
- [3] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. Dehazenet: An end-to-end system for single image haze removal. *IEEE TIP*, 25(11):5187–5198, 2016. 1, 6
- [4] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *ECCV*, pages 17–33, 2022. 1, 2, 3, 5
- [5] Yu-Wei Chen and Soo-Chang Pei. Always clear days: Degradation type and severity aware all-in-one adverse weather removal. *arXiv:2310.18293*, 2023. 1
- [6] Zheng Chen, Yulun Zhang, Jinjin Gu, Yongbing Zhang, Linghe Kong, and Xin Yuan. Cross aggregation transformer for image restoration. In *NeurIPS*, pages 25478–25490, 2022. 1, 2
- [7] Sung-Jin Cho, Seo-Won Ji, Jun-Pyo Hong, Seung-Won Jung, and Sung-Jea Ko. Rethinking coarse-to-fine approach in single image deblurring. In *CVPR*, pages 4641–4650, 2021. 4
- [8] Marcos V Conde, Gregor Geigle, and Radu Timofte. Instructir: High-quality image restoration following human instructions. In *ECCV*, pages 1–21, 2024. 1, 2, 7
- [9] 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. 4
- [10] Yuning Cui, Syed Waqas Zamir, Salman Khan, Alois Knoll, Mubarak Shah, and Fahad Shahbaz Khan. Adair: Adaptive all-in-one image restoration via frequency mining and modulation. In *ICLR*, pages 57335–57356, 2025. 2, 3
- [11] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Color image denoising via sparse 3d collaborative filtering with grouping constraint in luminance-chrominance space. In *ICIP*, pages 1–313, 2007. 6
- [12] Xin Dang, Hai Wang, Jie Ren, and Le Chen. An application performance optimization model of mobile augmented reality based on hd restoration. In *CBD*, pages 201–206, 2020. 1
- [13] Yu Dong, Yihao Liu, He Zhang, Shifeng Chen, and Yu Qiao. Fd-gan: Generative adversarial networks with fusion-discriminator for single image dehazing. In *AAAI*, pages 10729–10736, 2020. 5, 6
- [14] Huiyu Duan, Xiongkuo Min, Sijing Wu, Wei Shen, and Guangtao Zhai. Uniprocessor: a text-induced unified low-level image processor. In *ECCV*, pages 180–199, 2024. 1
- [15] Akshay Dudhane, Omkar Thawakar, Syed Waqas Zamir, Salman Khan, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Dynamic pre-training: Towards efficient and scalable all-in-one image restoration. *arXiv:2404.02154*, 2024. 2
- [16] Qingnan Fan, Dongdong Chen, Lu Yuan, Gang Hua, Nenghai Yu, and Baoquan Chen. A general decoupled learning framework for parameterized image operators. *IEEE TPAMI*, 43(1): 33–47, 2019. 1, 5, 6
- [17] Hongyun Gao, Xin Tao, Xiaoyong Shen, and Jiaya Jia. Dynamic scene deblurring with parameter selective sharing and nested skip connections. In *CVPR*, pages 3848–3856, 2019. 5, 6
- [18] Florin Girbacia, Silviu Butnariu, Alex Petre Orman, and Cristian Cezar Postelnicu. Virtual restoration of deteriorated religious heritage objects using augmented reality technologies. *EJST*, 9(2):223–231, 2013. 1
- [19] Hang Guo, Jinmin Li, Tao Dai, Zhihao Ouyang, Xudong Ren, and Shu-Tao Xia. Mambair: A simple baseline for image restoration with state-space model. In *ECCV*, pages 222–241, 2024. 1, 5, 6
- [20] JiaKui Hu, Lujia Jin, Zhengjian Yao, and Yanye Lu. Universal image restoration pre-training via degradation classification. *arXiv:2501.15510*, 2025. 4
- [21] Kui Jiang, Zhongyuan Wang, Peng Yi, Chen Chen, Baojin Huang, Yimin Luo, Jiayi Ma, and Junjun Jiang. Multi-scale progressive fusion network for single image deraining. In *CVPR*, pages 8346–8355, 2020. 2, 6
- [22] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv:1412.6980*, 2014. 5
- [23] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2noise: Learning image restoration without clean data. *arXiv:1803.04189*, 2018. 1, 2
- [24] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. Aod-net: All-in-one dehazing network. In *ICCV*, pages 4770–4778, 2017. 6
- [25] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE TIP*, 28(1):492–505, 2018. 6
- [26] 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. 1, 2, 5, 6, 7
- [27] Zilong Li, Yiming Lei, Chenglong Ma, Junping Zhang, and Hongming Shan. Prompt-in-prompt learning for universal image restoration. *arXiv:2312.05038*, 2023. 1, 2
- [28] Jingyun Liang, Jie Zhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. SwinIR: Image restoration using Swin transformer. In *ICCVW*, pages 1833–1844, 2021. 2, 5
- [29] Lin Liu, Lingxi Xie, Xiaopeng Zhang, Shanxin Yuan, Xianguyu Chen, Wengang Zhou, Houqiang Li, and Qi Tian. Tape: Task-agnostic prior embedding for image restoration. In *ECCV*, pages 447–464, 2022. 1, 5
- [30] Minghao Liu, Wenhan Yang, Jinyi Luo, and Jiaying Liu. Up-restorer: When unrolling meets prompts for unified image restoration. In *AAAI*, pages 5513–5522, 2025. 5, 6

- [31] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *ICCV*, pages 10012–10022, 2021. 1, 2
- [32] Yulin Luo, Rui Zhao, Xiaobao Wei, Jinwei Chen, Yijie Lu, Shenghao Xie, Tianyu Wang, Ruiqin Xiong, Ming Lu, and Shanghang Zhang. Wm-moe: Weather-aware multi-scale mixture-of-experts for blind adverse weather removal. *arXiv:2303.13739*, 2023. 1, 3
- [33] Ziwei Luo, Fredrik K Gustafsson, Zheng Zhao, Jens Sjölund, and Thomas B Schön. Controlling vision-language models for multi-task image restoration. *arXiv preprint arXiv:2310.01018*, 2023. 1, 2, 5
- [34] Kede Ma, Zhengfang Duanmu, Qingbo Wu, Zhou Wang, Hongwei Yong, Hongliang Li, and Lei Zhang. Waterloo exploration database: New challenges for image quality assessment models. *IEEE TIP*, 26(2):1004–1016, 2016. 5
- [35] 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, 2001. 5
- [36] Chong Mou, Qian Wang, and Jian Zhang. Deep generalized unfolding networks for image restoration. In *CVPR*, pages 17399–17410, 2022. 5
- [37] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *CVPR*, pages 3883–3891, 2017. 6
- [38] Vaishnav Potlapalli, Syed Waqas Zamir, Salman H Khan, and Fahad Shahbaz Khan. Promptir: Prompting for all-in-one image restoration. In *NeurIPS*, pages 71275–71293, 2023. 1, 2, 5, 6, 7
- [39] Yanyun Qu, Yizi Chen, Jingying Huang, and Yuan Xie. Enhanced pix2pix dehazing network. In *CVPR*, pages 8160–8168, 2019. 1, 6
- [40] Dongwei Ren, Wangmeng Zuo, Qinghua Hu, Pengfei Zhu, and Deyu Meng. Progressive image deraining networks: A better and simpler baseline. In *CVPR*, pages 3937–3946, 2019. 2
- [41] Wenqi Ren, Si Liu, Hua Zhang, Jinshan Pan, Xiaochun Cao, and Ming-Hsuan Yang. Single image dehazing via multi-scale convolutional neural networks. In *ECCV*, pages 154–169, 2016. 6
- [42] Wenqi Ren, Lin Ma, Jiawei Zhang, Jinshan Pan, Xiaochun Cao, Wei Liu, and Ming-Hsuan Yang. Gated fusion network for single image dehazing. In *CVPR*, pages 3253–3261, 2018. 2
- [43] Wenqi Ren, Jinshan Pan, Hua Zhang, Xiaochun Cao, and Ming-Hsuan Yang. Single image dehazing via multi-scale convolutional neural networks with holistic edges. *IJCV*, 128: 240–259, 2020. 2
- [44] Ying Tai, Jian Yang, Xiaoming Liu, and Chunyan Xu. Memnet: A persistent memory network for image restoration. In *ICCV*, pages 4539–4547, 2017. 1, 2
- [45] Chunwei Tian, Yong Xu, and Wangmeng Zuo. Image denoising using deep cnn with batch renormalization. *NN*, 121: 461–473, 2020. 5, 6
- [46] Xiangpeng Tian, Xiangyu Liao, Xiao Liu, Meng Li, and Chao Ren. Degradation-aware feature perturbation for all-in-one image restoration. In *CVPR*, pages 28165–28175, 2025. 1, 2, 3
- [47] Jeya Maria Jose Valanarasu, Rajeev Yasarla, and Vishal M Patel. TransWeather: Transformer-based restoration of images degraded by adverse weather conditions. In *CVPR*, pages 2353–2363, 2022. 1, 5
- [48] Cong Wang, Jinshan Pan, Wei Wang, Jiangxin Dong, Mengzhu Wang, Yakun Ju, and Junyang Chen. Promptrestorer: A prompting image restoration method with degradation perception. In *NeurIPS*, pages 8898–8912, 2023. 1, 2
- [49] Tao Wang, Kaihao Zhang, Ziqian Shao, Wenhan Luo, Bjorn Stenger, Tong Lu, Tae-Kyun Kim, Wei Liu, and Hongdong Li. Gridformer: Residual dense transformer with grid structure for image restoration in adverse weather conditions. *IJCV*, 132(10):4541–4563, 2024. 5, 6, 7
- [50] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *CVPR*, pages 17683–17693, 2022. 1, 2
- [51] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. *arXiv:1808.04560*, 2018. 6
- [52] Wei Wei, Deyu Meng, Qian Zhao, Zongben Xu, and Ying Wu. Semi-supervised transfer learning for image rain removal. In *CVPR*, pages 3877–3886, 2019. 6
- [53] Haiyan Wu, Yanyun Qu, Shaohui Lin, Jian Zhou, Ruizhi Qiao, Zhizhong Zhang, Yuan Xie, and Lizhuang Ma. Contrastive learning for compact single image dehazing. In *CVPR*, pages 10551–10560, 2021. 2
- [54] Fuzhi Yang, Huan Yang, Jianlong Fu, Hongtao Lu, and Baining Guo. Learning texture transformer network for image super-resolution. In *CVPR*, pages 5791–5800, 2020. 5
- [55] Hao Yang, Liyuan Pan, Yan Yang, and Wei Liang. Language-driven all-in-one adverse weather removal. In *CVPR*, pages 24902–24912, 2024. 1, 3
- [56] Rajeev Yasarla and Vishal M Patel. Uncertainty guided multi-scale residual learning-using a cycle spinning cnn for single image de-raining. In *CVPR*, pages 8405–8414, 2019. 1, 6
- [57] Eduard Zamfir, Zongwei Wu, Nancy Mehta, Danda Dani Paudel, Yulun Zhang, and Radu Timofte. Efficient degradation-aware any image restoration. *arXiv:2405.15475*, 2024. 1, 2, 3
- [58] Eduard Zamfir, Zongwei Wu, Nancy Mehta, Yuedong Tan, Danda Dani Paudel, Yulun Zhang, and Radu Timofte. Complexity experts are task-discriminative learners for any image restoration. In *CVPR*, pages 12753–12763, 2025. 1, 3
- [59] 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*, pages 14821–14831, 2021. 5, 6
- [60] 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*, pages 5728–5739, 2022. 1, 2, 5, 6, 7

- [61] Haijin Zeng, Xiangming Wang, Yongyong Chen, Jingyong Su, and Jie Liu. Vision-language gradient descent-driven all-in-one deep unfolding networks. In *CVPR*, pages 7524–7533, 2025. 1, 2, 6
- [62] He Zhang and Vishal M Patel. Density-aware single image de-raining using a multi-stream dense network. In *CVPR*, pages 695–704, 2018. 6
- [63] Jinghao Zhang, Jie Huang, Mingde Yao, Zizheng Yang, Hu Yu, Man Zhou, and Feng Zhao. Ingredient-oriented multi-degradation learning for image restoration. In *CVPR*, pages 5825–5835, 2023. 1, 2, 5, 6, 7
- [64] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE TIP*, 26(7):3142–3155, 2017. 6
- [65] Kai Zhang, Wangmeng Zuo, Shuhang Gu, and Lei Zhang. Learning deep cnn denoiser prior for image restoration. In *CVPR*, pages 3929–3938, 2017. 1, 2, 6
- [66] Kai Zhang, Wangmeng Zuo, and Lei Zhang. FFDNet: Toward a fast and flexible solution for cnn-based image denoising. *IEEE TIP*, 27(9):4608–4622, 2018. 6
- [67] 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 TIP*, 2025. 1, 2, 5, 6
- [68] Yulun Zhang, Kunpeng Li, Kai Li, Bineng Zhong, and Yun Fu. Residual non-local attention networks for image restoration. *arXiv:1903.10082*, 2019. 1, 2
- [69] Haiyu Zhao, Yuanbiao Gou, Boyun Li, Dezhong Peng, Jiancheng Lv, and Xi Peng. Comprehensive and delicate: An efficient transformer for image restoration. In *CVPR*, pages 14122–14132, 2023. 2
- [70] Dian Zheng, Xiao-Ming Wu, Shuzhou Yang, Jian Zhang, Jian-Fang Hu, and Wei-Shi Zheng. Selective hourglass mapping for universal image restoration based on diffusion model. In *CVPR*, pages 25445–25455, 2024. 2