This CVPR paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

Restore, Assess, Repeat: A Unified Framework for Iterative Image Restoration

I-Hsiang Chen[1,2,*] Isma Hadji[1] Enrique Sanchez[1] Adrian Bulat[1,3] Sy-Yen Kuo[2,4]
Radu Timofte[5] Georgios Tzimiropoulos[1,6] Brais Martinez[1]

[1]Samsung AI Center Cambridge [2]National Taiwan University [3]Technical University of Iasi
[4] Chang Gung University [5]University of Wurzburg [6]Queen Mary University of London

Abstract

Image restoration aims to recover high quality images from inputs degraded by various factors, such as adverse weather, blur, or low light. While recent studies have shown remarkable progress across individual or unified restoration tasks, they still suffer from limited generalization and inefficiency when handling unknown or composite degradations. To address these limitations, we propose RAR, a Restore, Assess and Repeat process, that integrates Image Quality Assessment (IQA) and Image Restoration (IR) into a unified framework to iteratively and efficiently achieve high quality image restoration. Specifically, we introduce a restoration process that operates entirely in the latent domain to jointly perform degradation identification, image restoration, and quality verification. The resulting model is fully trainable end-to-end and allows for an all-in-one assess and restore approach that dynamically adapts the restoration process. Also, the tight integration of IQA and IR into a unified model minimizes the latency and information loss that typically arises from keeping the two modules disjoint, (e.g. during image and/or text decoding). Extensive experiments show that our approach offers consistent improvements under single, unknown and composite degradations, thereby establishing a new state-of-the-art.

1. Introduction

Image Restoration is a challenging task that aims at recovering a clean image from one or several degradations such as blur, haze, rain, or lighting. Most early works focused on tackling one kind of degradation only [7, 31, 40, 44]. However, in real world settings, degradations are complex and rarely known a priori. In here, we focus on image restoration under unknown composite degradations, so images can present one or more unknown degradations at a time.

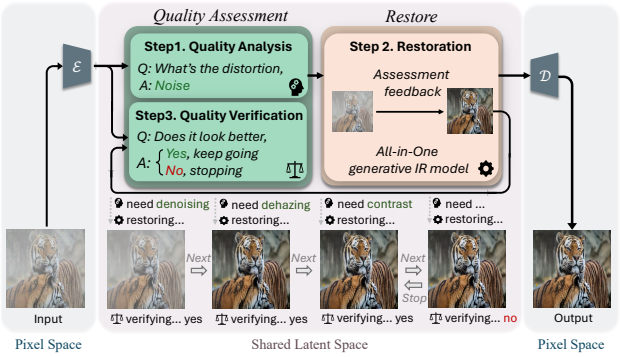

Figure 1. **Overview of RAR**. Given an input image with unknown composite degradations, RAR operates entirely in latent by iteratively assessing current image quality, restoring the image based on the assessment, and verifying the quality improvement.

Recent methods tackling composite degradations can be broadly categorized into two paradigms. The first, *all-in-one* models [14, 22, 30, 46] utilize a single, unified generative model for restoration. The second, *agentic* models [3, 42, 48], employ an agent that iteratively selects and applies tools from a suite of specialized, single-degradation restorers. Tool selection is guided by an IQA step, which identifies the main degradation. While agentic models allow for progressive restoration, they are cumbersome and slow. Conversely, all-in-one models offer a more efficient and unified alternative, but currently lag in performance.

AutoDIR [14], one of the best-performing all-in-one models, also uses an assessment step in line with agentic models. However, it relies on CLIP [29], which treats IQA as n-way classifier over a predefined set of degradation labels. This limits the generalization capability of the restoration model. Furthermore, zero-shot CLIP performs poorly for IQA. Thus, supervised finetuning is required to make it focus on degradations rather than semantics. Instead, recent agentic models (e.g AgenticIR [48]) rely on a description-based IQA model [41], which provides a rich free-form assessment of the image and covers multiple assessment tasks.

Project: https://restore-assess-repeat.github.io/
* Work conducted during I-Hsiang's internship at Samsung AI Center, Cambridge, UK.

However, given the lack of integration with the restoration step, even when the IQA explicitly identifies image degradations, final performance is still limited by the availability of a corresponding restoration tool and proper execution.

In this work, we take the best of both worlds by integrating a more descriptive IQA model, capable of assessing image quality through free form text, into an all-in-one generative IR model. Importantly, we propose a deep integration of the two modules such that they share a common latent space, thereby yielding a single end-to-end trainable model that is capable of multiple rounds of degradation identification, restoration and quality verification in latent space.

To achieve this, we start from an IQA model that can describe the degradations in an image using free form text but enforce the IQA model to operate in the same latent space as the IR module. For the IR we use a generative model, which we condition on the feedback obtained from the IQA model directly using its output logits rather than the decoded text, thereby tightly combining the assessment and restoration steps. The resulting model is fully trainable in an end-to-end fashion to dynamically adapt the IR on the IQA feedback. At inference, the IQA model is also used to compare images before and after each restoration step to decide that (a) the restoration process is successful or (b) repeat the process with a new conditioning based on trailing issues in the restored image. The entire process is illustrated in Fig. 1. In summary, our contributions are threefold:

- We propose to combine free-form IQA with an all-in-one restoration model.
- We tightly integrate the IQA and restoration into a single end-to-end trainable model.
- We enable multiple assessment, restoration, validation rounds, all operating in latent space, thereby improving both training process and inference performance.

Together, these contribution yield state-of-the-art performance on unknown, composite and single degradation settings. For example, in the challenging composite degradations setup, our method achieves improvements ranging from +2.71 dB (PSNR) to +72.4 (MANIQA), while being 11.27× faster than the state-of-the-art.

2. Related Work

Image restoration is a challenging inverse imaging task. Early works focused on single-degradation settings such as denoising [7, 44], deblurring [24, 31], dehazing [6, 10, 44], deraining [35, 39] or low-light enhancement [40, 45]. To better align with real-world conditions, recent works address multiple degradations, which is our main focus in this work. We categorize methods into two main tacks: all-in-one models and agentic models.

All-in-one models rely on a single model to tackle multiple degradations. These methods are either non-generative [5, 19, 20, 27, 34, 43, 44], or rely on a generative model [4, 22, 26, 30, 38, 46, 49]. Most such methods use known degradation priors or a classifier to recognize the degradation category present in the image to guide the restoration model. Therefore, these methods are constrained to handle a limited set of degradations. While we also rely on generative priors to perform restoration, we do not assume knowledge of the degradation. Instead, we enlist an IQA model to condition the restoration process. Most closely related work to ours is AutoDIR [14], which attempts to handle unknown degradations also using an IQA model. However, they rely on CLIP, which by default performs poorly for degradation identification, thus needing finetuning to operate on a predefined closed set of degradations. Therefore, it inherits the shortcoming of prior methods. Instead, we start from DepictQA [41], a VLM-based IQA, which not only can describe image quality with free-form text, but also can compare pairs of images. We then adapt it to operate in the latent space of a restoration module, to yield an end-to-end trainable model, which enables an iterative assess and restore process.

Agentic Models. instead consider a large set of models (tools) that specialize on different single degradations [3, 42, 48]. They then either rely on reinforcement learning to define a policy that select appropriate tools [42], or finetune LLM/VLM models to iteratively select and apply the different tools considered [3, 48]. We also use a VLM-based IQA, however, agentic model do not fully integrate the IQA and IR modules into a unified model. Hence, they are typically slow and need more steps to guide the restoration. For example, AgenticIR [48] starts from a similar IQA to ours. However, given the disconnect between the IQA and IR tools, it requires the restoration models to first decode their output into actual images, which are then encoded in the latent space of the IQA to get an assessment, which is then passed to a powerful LLM [25] to devise a restoration plan that is tested in a trial-and-error manner, thereby making the entire process extremely inefficient. In addition, even when the IQA describes the correct degradations, the restoration process is limited by the availability of a restoration tool in the predefined set available to the agent. Our work takes the best of both worlds using a strong VLM-based IQA and fully integrating it into a strong all-in-one generative IR model, attaining superior performance and better efficiency.

3. Proposed Method

3.1. Preliminaries

Given a degraded image, $I_{deg} \in \mathbb{R}^{H\times W\times C}$, our goal is to automatically identify image degradations using an IQA module, and to progressively restore the image into a high quality version, $I_{hq} \in \mathbb{R}^{H\times W\times C}$, by relying on the IQA's

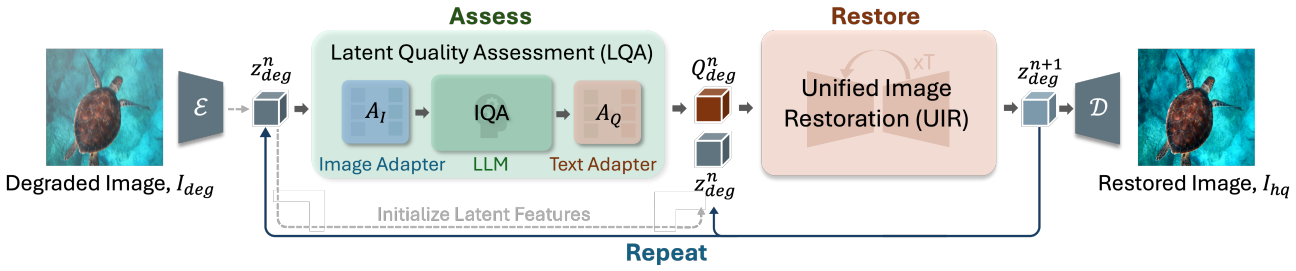

Figure 2. To enable our RAR process, we define an LQA that fully integrates the input and out of the IQA into the latent space of the restoration module, using adapters $\mathcal{A}_I$ and $\mathcal{A}_Q$, respectively. Then, we instore a feedback loop between the two modules to iteratively recover best image quality. The dashed-line is only used for the very first iteration to feed $\mathbf{z}^0_{deg}$ to the restoration model.

assessment for conditioning. To this end, we consider a framework consisting of two main modules; namely, (1) an Image Quality Assessment (IQA) module that can describe image quality in free-form text and (2) an Image Restoration (IR) module that can be conditioned on such text.

We have the following key requirements for the IQA module: (i) we want it to generalize rather than focus on a closed set of degradation labels, (ii) handle composite degradations and (iii) provide rich conditioning information for the restoration module. To this end, we use the recently introduced DepictQA [41], which we further finetune for our purposes. Note that previous work built upon the same IQA model to activate different restoration agents [48]. For the IR module, we can take advantage of the strong priors learned by any state-of-the-art generative model, such as a standard diffusion model [32]. We term this component as Unified Image Restoration (UIR) model.

A naive implementation would perform IQA on the degraded image to obtain a free-form description of the degradations present, and condition the restoration process on the output of the IQA by using the text-conditioning branch of diffusion models. More formally:

$$Q_{deg} = IQA(I_{deg}) \quad (1)$$

$$I_{hr} = Restore(I_{deg}|Q_{deg}) \quad (2)$$

However, we have two main extra desiderata in respect to this naive integration: (a) we would like the IQA and the restoration modules to be tightly integrated to better exploit their synergetic relationship, both at inference and training time. (b) we would like an iterative process where the IQA assessment is updated throughout the restoration process to better adapt to the current restoration status. To this end, we introduce Latent Quality Assessment (LQA) in Sec. 3.2, which results in an end-to-end trainable unified model. Then we describe our iterative restoration process and adaptive stopping criteria, which rely on intermediate quality assessments in Sec. 3.3.

3.2. Latent Quality Assessment

The naive integration in the previous section treats the IQA and UIR as independent models. In this section, we unify them into a single end-to-end trainable model. This requires two main changes: (i) aligning the input latent spaces of the IQA and the restoration models, (ii) aligning the output IQA latents and the restoration conditioning latents.

Latent space alignment: In the naive integration, the restoration model first projects the degraded image I_{deg} into latent space using an auto-encoder $\mathcal{E}_{restore}$, obtaining $\mathbf{z}_{deg} = \mathcal{E}_{restore}(I_{deg})$. Instead, the IQA operates from pixel space and uses a different auto-encoder $\mathcal{E}_{IQA}$. We instead modify the input to the IQA to obtain a Latent Quality Assessment (LQA) model,

$$LQA(\mathbf{z}_{deg}) = IQA(\mathcal{A}_I(\mathbf{z}_{deg})) \quad (3)$$

where $\mathcal{A}_I$ is an adapter module that we train to bridge the gap between the two representations. We follow a two-stage finetuning process. First, we only finetune the image adapter, $\mathcal{A}_I$, while keeping the IQA weights frozen. We then unfreeze all layers to further finetune the full LQA.

Conditioning alignment: The LQA module described above produces an assessment text form. This is then encoded by the text-conditioning branch of the diffusion model to produce the conditioning embeddings. The decoding step is undesirable because it is non-differentiable and also loses information compared to the IQA output latents. Furthermore, conditioning based on the IQA text output relies on the parameter and latency-heavy text-conditioning branch of the restoration model. We propose instead to directly align the latent output of the IQA, $\tilde{Q}_{deg} = LQA(\mathbf{z}_{deg})$, to the output embeddings of the text-conditioning branch of the restoration model, $\mathcal{T}$. Once again, we use an adapter A_Q to align the two representations such that

$$Q_{deg} = \mathcal{T}(\mathcal{A}_Q(\tilde{Q}_{deg})) \quad (4)$$

We train in two stages, first finetuning the text adapter, $\mathcal{A}_Q$,

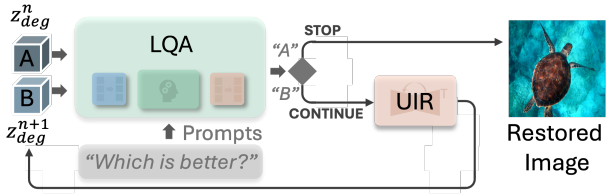

Figure 3. We illustrate our quality verification step. This is used at inference time to define a stopping criterion of the RAR process.

before unfreezing the UIR weights in the second stage. Notably, with this step, we have removed the need for decoding between the assessment and restoration stage, thereby yielding an end-to-end trainable model. Furthermore, the text-conditioning branch on the restoration model can now be dropped entirely, saving both the latency and the parameter count. The details architecture and training are provided in the supplemental material.

3.3. Restore, Assess, Repeat

Up until now, our approach has been agnostic of the restorer module employed, as long as it can be conditioned on the output of the IQA. However, now we want to dynamically update the conditioning signal provided by the LQA. This requires re-assessing the quality using intermediate latents, $\mathbf{z}^n_{deg}$, as input (e.g. output of denoising/flow step n). If we adopt the diffusion-based denoising paradigm, intermediate representations will contain added noise, leading to inaccurate readings of the LQA. We thus require a model that does not involve a noising process with every update to the input condition. We therefore rely on Stable Diffusion 3.5 (SD3.5) [11] as it uses Flow Matching (FM).

While FM learns a direct mapping between input and output distributions, it is common for image editing to learn a mapping from noise, $\mathcal{N} \sim (0, I)$, to the target distribution, and include the degraded image as part of the conditioning context. Instead, we remove the noise term and learn a direct mapping from the degraded image distribution, ρ_{deg}, to the high quality image distribution, ρ_{hq}, and remove the degraded image from the conditioning context.

More specifically, we define a mapping between the degraded input, $\mathbf{z}_{deg} \sim \rho_{deg}$, and the target output, $\mathbf{z}_{hq} \sim \rho_{hq}$, where $\mathbf{z}$ is the representation of images in the latent space of the UIR model. A model v_θ, is trained to predict the velocity vector between the two distribution such that

$$\mathcal{L}_v = \mathbb{E}_{\mathbf{z}_{deg}, \mathbf{z}_{hq}, Q_{deg}, t} ||v_\theta(\mathbf{z}_t, Q_{deg}, t) - (\mathbf{z}_{hq} - \mathbf{z}_{deg})||^2 \tag{5}$$

where $\mathbf{z}_t$ is a linear interpolation between the two distributions, defined as $\mathbf{z}_t = (1-t)\mathbf{z}_{deg} + t\mathbf{z}_{hq}$, and $v = v_\theta(\mathbf{z}_t, Q_{deg}, t)$ is the vector pointing from $\mathbf{z}_{deg}$ to $\mathbf{z}_{hq}$.

In Table. 5 we show, quantitatively, the importance of this formulation, which allows us to iteratively update the input, $\mathbf{z}_{deg}$, without being corrupted by added noise.

Iterative assessment: We now take full advantage of the multi-step nature of the adopted flow-based model to iteratively assess the restoration progress by adding a feedback loop between the LQA and the UIR modules. In particular, the UIR model starts the restoration process by taking a degraded image latent $\mathbf{z}^0_{deg}$ and its corresponding LQA assessment Q^0_{deg} as input. The input conditioning is then updated dynamically after n iterations. Specifically, the training objective in (5) is updated such that

$$\mathcal{L}_v = \mathbb{E}_{\mathbf{z}^n_{deg}, \mathbf{z}_{hq}, Q^n_{deg}, t} ||v_\theta(\mathbf{z}^n_t, Q^n_{deg}, t) - (\mathbf{z}_{hq} - \mathbf{z}^n_{deg})||^2 \tag{6}$$

with $\mathbf{z}^n_t = (1-t)\mathbf{z}^n_{deg} + t\mathbf{z}_{hq}$, and $v^n = v_\theta(\mathbf{z}^n_t, Q^n_{deg}, t)$. The predicted velocity is used periodically to update the input such that $\mathbf{z}^{n+1}_{deg} = \mathbf{z}^n_{deg} + v^n$. Similarly, the conditioning term is updated using the integrated LQA module such that $Q^{n+1}_{deg} = LQA(\mathbf{z}^{n+1}_{deg})$. Thanks to our formulation with direct mapping from degraded distribution to target distribution, the LQA can provide meaningful feedback without being affected by noise, as shown in Tab. 5.

Fig. 1 exemplifies this process, showing an input image (leftmost side) that contains multiple degradations. The restoration model removes the noise first. Rather than insisting on denoising, the resulting latent is then fed back to the LQA, that now identifies haze as the remaining degradation. The restorer takes this updated feedback and removes haze in the next iteration.

During training, our RAR process is seamlessly integrated into standard flow-matching training, where the model is typically trained with a predefined number of timesteps and the LQA can be called for any random timestep. At inference time, we perform an assessment every T steps and stop the iterations upon reaching a stopping criterion, as described next. The fully integrated model is illustrated in Fig. 2.

RAR stopping criterion: To devise a stopping criterion for the proposed RAR process, we again use the LQA model. Given that our LQA model is based on DepictQA [41], which can compare pairs of images, we re-enlist it to define a stopping criterion. Specifically, after every T steps through our model, we use the LQA to compare the quality of the images before and after restoration. As illustrated in Fig. 3, the LQA takes latents $\mathbf{z}^n_{deg}$ and $\mathbf{z}^{n+T}_{deg}$ as input and is tasked to compare their quality and yield a binary decision to either `CONTINUE` if the quality of the new latent is better than the previous, or `STOP`, otherwise. In the latter case, we use $\mathbf{z}^n_{deg}$ as our final model prediction.

4. Experiments

We now thoroughly evaluate the effectiveness of the proposed RAR process. We first define our experimental setup in Sec. 4.1. We then provide extensive quantitative and

qualitative comparisons to state-of-the-art in Sec. 4.2. Finally, we support each component in our framework with a thorough ablation study in Sec. 4.3. We provide implementation details in the supplementary material.

4.1. Experimental Setup

Training Dataset. We consider the same restoration tasks used in previous work [14, 23]; namely, denoising, deblurring, super-resolution, low-light enhancement, dehazing, deraining, and deraindrop. For super-resolution, we use DIV2K [2] and Flickr2K [21] training sets, following RealESRGAN [36] for degraded image generation. For denoising, we use SIDD [1] and a synthetic Gaussian and Poisson noise dataset with DIV2K [2] and Flickr2K [21]. We also use the training set from GoPro [24], LOL [37], RESIDE [17], Rain200L [39], and Raindrop [28] for deblurring, low-light enhancement, dehazing and deraining. For the Latent Quality Assessment, we followed the DepictQA [41] pipeline and extended its training corpus to include the aforementioned restoration datasets.

Testing Dataset. During inference, we categorize image restoration into three tasks: composite degradations, single degradation and unknown degradations. For the *composite degradations*, we followed the AgenticIR [48] setting, which constructs 16 combinations of mixed degradations based on the MiO100 [15] dataset. These combinations are divided into three subsets (A,B, and C), where subsets A and B involve two degradations and subset C involves three, simulating progressively complex degradation scenarios. For the *single degradation*, we followed the AutoDIR [14] setting, where single degradation restoration is evaluated on standard benchmarks, including SIDD [1], Kodak24 [12], DIV2K [2], GoPro [24], LOL [37], SOTS-Outdoor [18], Rain100 [39], and Raindrop [28] dataset. For the *unknown degradations*, we used UDC [47] and EUVP [13] datasets to assess generalization to unseen degradation types.

Baselines: To evaluate the effectiveness of our method, we compare it with multiple All-in-One restoration approaches, including AirNet [16], PromptIR [27], MioIR [15], DA-CLIP [23], InstructIR [8], AdaIR [9], DARCOT [33] and AutoDIR [14]. For completeness, we also compare to AgenticIR [48], one of the most recent agent-based restoration frameworks.

Metrics: We consider both fidelity (e.g., PSNR, SSIM) and perceptual metrics (e.g., LPIPS, MANIQA, CLIP-IQA, MUSIQ, NIQE and NIMA), providing balanced assessment between restoration accuracy and perceptual performance.

4.2. Comparison to state-of-the-art

Composite Degradations. We first evaluate our method in the composite degradation setup using the same test set introduced by AgenticIR [48] and compare to main methods reporting results on this data. Table 1, speaks decisively in favor of our approach, which outperforms existing methods in almost all settings with *significant* margins, where we are on average about $\times 2$ better than state-of-the-art across perceptual metrics. Importantly, the superiority of our method is especially striking on the most challenging setting, which includes 3 degradations (i.e. Group C). Similarly, Fig. 4 shows that our method can generate richer details, making the images more realistic

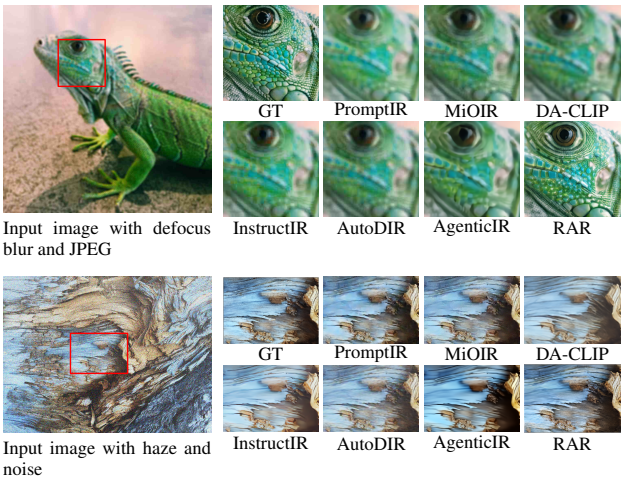

Figure 4. **Qualitative comparison on Composite Degradation.**

Table 1. Quantitative results for **Composite Degradations**. Results taken from AgenticIR [48]

Method	PSNR ↑	SSIM ↑	LPIPS ↓	MANIQA↑	CLIP-IQA↑	MUSIQ↑
Group A						
AirNet	19.13	0.6019	0.4283	0.2581	0.3930	42.46
PromptIR	20.06	0.6088	0.4127	0.2633	0.4013	42.62
MiOIR	20.84	0.6558	0.3715	0.2451	0.3933	47.82
DA-CLIP	19.58	0.6032	0.4266	0.2418	0.4139	42.51
InstructIR	18.03	0.5751	0.4429	0.2660	0.3528	45.77
AdaIR	19.06	0.6061	0.4170	0.2718	0.4117	44.66
DARCOT	18.82	0.5960	0.4303	0.2641	0.4058	43.42
AutoDIR	19.64	0.6286	0.3967	0.2500	0.3767	47.01
AgenticIR	**21.04**	0.6818	0.3148	0.3071	0.4474	56.88
RAR	20.46	**0.7144**	**0.1299**	**0.4659**	**0.6566**	**57.19**
Group B						
AirNet	19.31	0.6567	0.3670	0.2882	0.4274	47.88
PromptIR	20.47	0.6704	0.3370	0.2893	0.4289	48.10
MiOIR	20.56	0.6905	0.3243	0.2638	0.4330	51.87
DA-CLIP	18.56	0.5946	0.4405	0.2435	0.4154	43.70
InstructIR	18.34	0.6235	0.4072	0.3022	0.3790	50.94
AdaIR	19.24	0.6637	0.3467	0.3001	0.4380	49.16
DARCOT	19.08	0.6372	0.3910	0.2845	0.4321	46.37
AutoDIR	19.90	0.6643	0.3542	0.2534	0.3986	49.64
AgenticIR	20.55	0.7009	0.3072	0.3204	0.4648	**57.57**
RAR	**21.04**	**0.7326**	**0.1269**	**0.4582**	**0.6483**	56.91
Group C						
AirNet	17.95	0.5145	0.5782	0.1854	0.3113	30.12
PromptIR	18.51	0.5166	0.5756	0.1906	0.3104	29.71
MiOIR	15.63	0.4896	0.5376	0.1717	0.2891	37.95
DA-CLIP	18.53	0.5320	0.5335	0.1916	0.3476	33.87
InstructIR	17.09	0.5135	0.5582	0.1732	0.2537	33.69
AdaIR	17.88	0.5209	0.5802	0.2036	0.3276	30.83
DARCOT	17.45	0.4993	0.6097	0.1937	0.3172	29.50
AutoDIR	18.61	0.5443	0.5019	0.2045	0.2939	37.86
AgenticIR	18.82	0.5474	0.4493	0.2698	0.3948	48.68
RAR	**19.33**	**0.6579**	**0.1489**	**0.4653**	**0.6554**	**56.56**

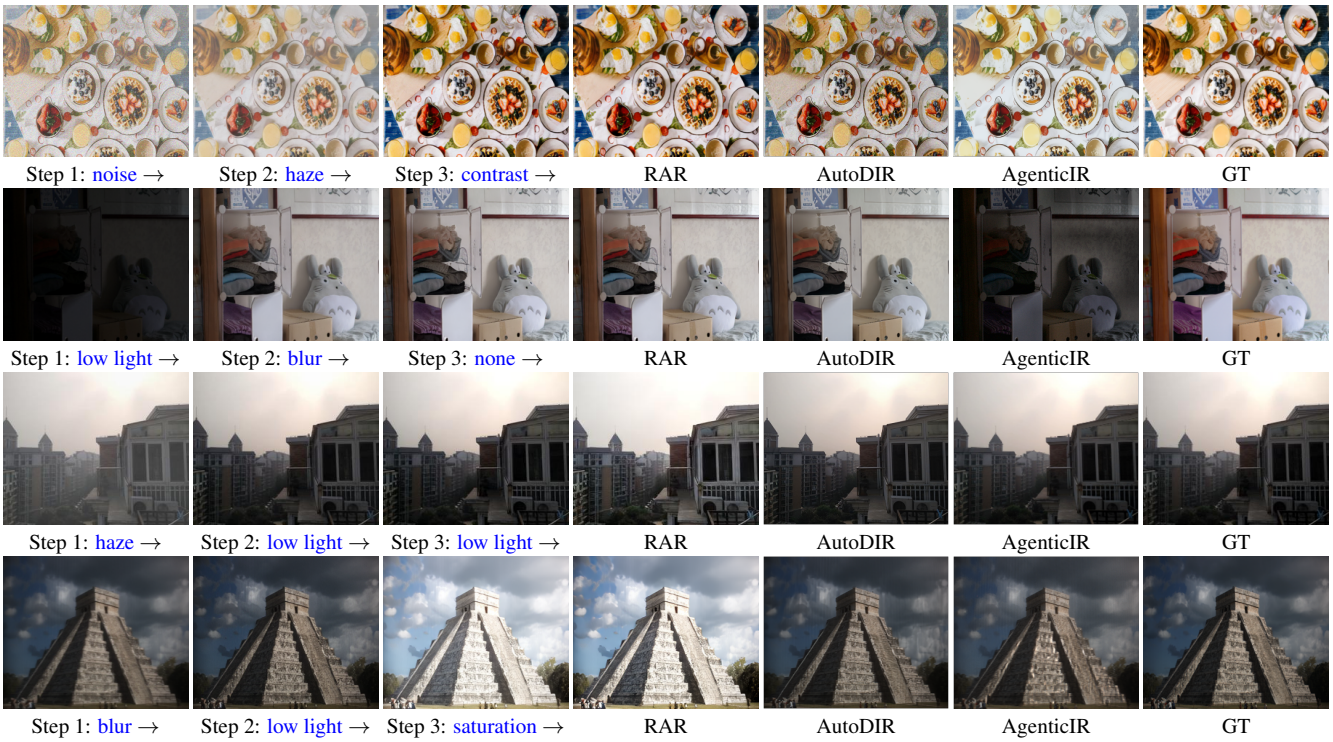

Figure 5. **Qualitative analysis:** Comparison with other multi-round methods (AutoDIR [14], AgenticIR [48]). Intermediate steps for our method are shown for illustrative purposes. **First row:** Composite degradations (haze + noise). RAR identifies and corrects them, also enhances contrast, making the result visually clearer. **Second row:** Single degradation (low light). AgenticIR fails due to inadequate tools for very dark scenes. **Third row:** Single degradation (haze). RAR successfully dehazes, but also improves low light, lowering the fidelity to ground truth. **Fourth row:** Unknown degradations. RAR demonstrates stronger generalization (also improving over ground truth).

Table 2. Results for **Unknown Degradations** on the UDC dataset.

Method	MUSIQ ↑	CLIP-IQA ↑	NIQE ↓	NIMA ↑
AirNet	41.20	0.265	7.681	4.048
PromptIR	45.40	0.303	8.294	4.344
AdaIR	21.14	0.225	7.688	4.023
DARCOT	20.38	0.211	8.224	4.039
LD	43.41	0.288	7.983	4.371
AutoDIR	49.27	0.311	6.392	4.427
AgenticIR	52.76	0.358	6.670	4.456
RAR	**55.97**	**0.602**	**6.162**	**4.579**

Table 3. Quantitative comparison on **Single Degradation** tasks.

Method	PSNR ↑	SSIM ↑	LPIPS ↓	CLIP-IQA ↑	MUSIQ ↑	MANIQA ↑
AirNet	25.65	0.8230	0.1820	0.4516	48.69	0.3163
PromptIR	27.04	0.8445	0.1515	0.4196	48.77	0.3169
AdaIR	26.48	0.7825	0.2604	0.4368	53.51	0.3456
DARCOT	26.08	0.7772	0.2579	0.4409	54.34	0.3461
LD	22.17	0.6922	0.1935	0.4253	54.94	0.3438
AutoDIR	**27.81**	**0.8703**	0.1283	0.4119	54.37	0.3053
AgenticIR	24.41	0.7794	0.1392	0.3763	55.95	0.2978
RAR	25.88	0.8378	**0.0699**	**0.5566**	**56.00**	**0.4125**

and better aligned with human visual perception. Also, the first row in Fig. 5, rolls out our iterative process on an example from this dataset. Clearly, our iterative process successfully identifies and fixes degradations to finally arrive at the cleanest image compared to strong baselines.

Unknown Degradations. We follow previous work [14] and evaluate our method in the unknown degradation setup on the UDC [47] dataset (e.g., TOLED) and use the perceptual metrics for this setup. As the results summarized in Table 2, show that our method outperforms all baselines. Importantly, margins are especially sizable on the MUSIQ and CLIP-IQA metrics, which have been previously reported to align better with human judgment [14]. The last row in Fig. 5, uses an image from this dataset. It provides an example of how our method can successfully handle unknown degradations, by iteratively removing degradations, yielding a final output that is even cleaner than the provided ground truth.

Single Degradation. Finally, following AutoDIR [14], we evaluate our performance on single degradation tasks. In particular, we evaluate on SIDD [1], Kodak24 [12], DIV2K [2], GoPro [24], LOL [37], SOTS-Outdoor [18], Rain100 [39], and Raindrop [28] datasets and report average scores across all tasks in Table 3. We perform particu-

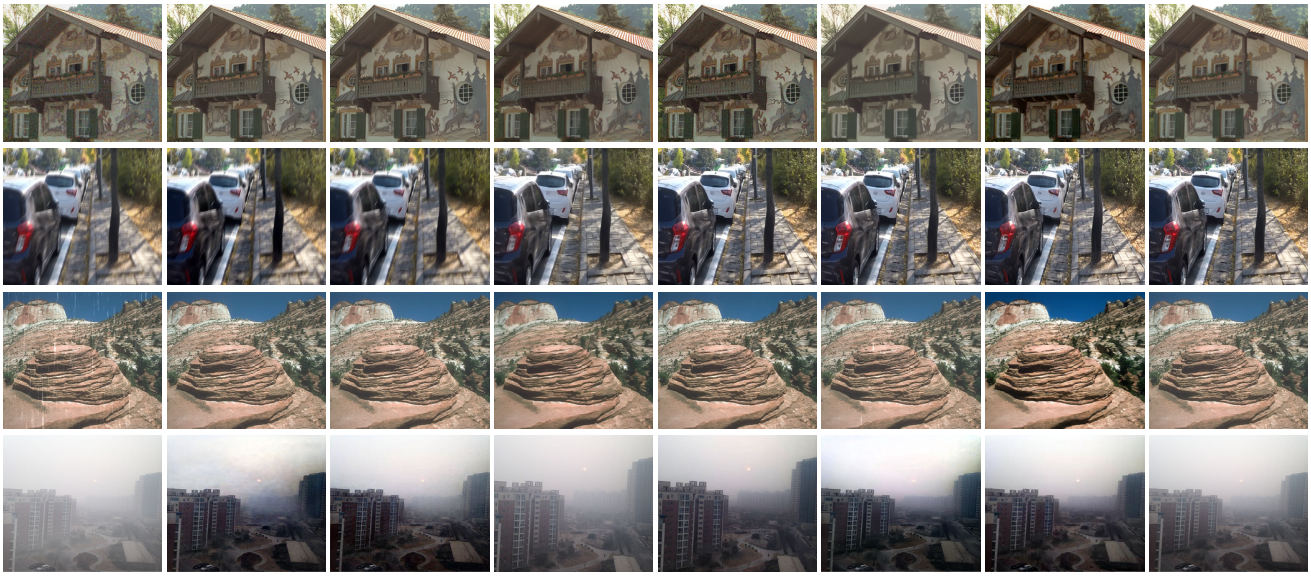

Figure 6. **Qualitative comparison on single degradation:** Row1: Kodak24 (noise), Row2: GoPro (blur), Row3: Rain100L (rain), Row4: SOTS (haze), In the single-degradation setting, our method produces clearer images that better align with human visual perception. This results lower fidelity scores but higher perceptual scores.

larly well on perceptual metrics in this setup. However, we note that our method is slightly behind on fidelity metrics. We attribute this behaviour to two main reasons. First, our iterative approach is particularly well suited for challenging multi-degradation setup. More importantly, we note that the ground truth only fixes one issue in the degraded image, even while there remains clear degradations (e.g. See GT in Fig.6). Therefore, our method, which tries to remove all degradations, goes beyond the GT, affecting fidelity metrics. For example, the middle two rows in Fig. 5 show examples taken from single degradation test sets. Our model, can identify more degradations than advertised in GT and yield more visually appealing images. We provide more examples of this effect in the supplementary material.

4.3. Ablation Study

We ablate the components of our method in the challenging setup of unknown and composite degradations, using both fidelity (PSNR, SSIM) and perceptual metrics (LPIPS).

Integration of the IQA and UIR modules. In Tab. 4 we ablate all steps in our method to tightly integrate the assessment and restoration modules. First, we focus on the role of the IQA model and integration of both image and text into the latent and embeddings space, respectively, of the restoration model. We follow the same setup as AutoDIR [14] in the first two rows to directly measure the impact of using a more descriptive IQA model. We use our flow-based formulation in the last four rows and show the great benefit of integrating the IQA into the restoration model latent space. Notably, replacing text with corresponding embeddings brings significant gains. We attribute this to the embeddings containing richer information than text, which is obtained from the embeddings but going through a lossy decoding step. This is illustrated in Fig. 8.

Switching attention to the role of the UIR backbone, we can see that while relying on the stronger SD3.5 brings a sizable performance boost for unknown degradations, its effect is minimal in the case of composite degradations, which shows that the task needs more than a higher capacity model. Importantly, removing the noise-based conditioning in favor of directly mapping from degraded to high quality image distributions does not adversely affect performance and yields better perceptual results as illustrated in Fig. 7.

Role of the RAR iterative process. We now show the benefit the iterative process, enabled by our tight integration, both at training and inference time. In Table 5 we can see that using iterative conditioning at training time benefits the flow-based formulation, while it actually is detrimental to the diffusion-based model. We attribute this behaviour to the utilization of a noising process in the diffusion paradigm, which alters the latents consumed by the LQA, thereby making it less useful. In contrast, our noise-free formulation using flow-matching allows us to directly use the degraded latents during training without noise contamination, thereby taking better advantage of the LQA feedback at training time. Importantly, we note that this iterative conditioning benefits more the composite degradation setup as it is best aligned with our iterative setup.

Table 4. **Ablation** of the integration process between the IQA and the UIR modules.

Backbone	Configuration				Unknown Degradations (UDC)			Composite Degradations (Group-C)		
	IQA	Text Space	Image Space	noise-free cond.	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
SD1.5	CLIP	Text	Pixel	✗	20.67	0.6217	0.2211	16.70	0.4848	0.2997
SD1.5	DepictQA	Text	Pixel	✗	22.34	0.6596	0.2016	17.67	0.5260	0.2267
SD3.5	DepictQA	Text	Pixel	✗	24.41	0.7583	0.1652	17.88	0.5696	0.2306
SD3.5	DepictQA	Text	Pixel	✓	24.70	0.7766	0.1589	17.89	0.5776	0.2289
SD3.5	DepictQA	Text	Latent	✓	24.90	0.7952	0.1131	18.72	0.6278	0.1576
SD3.5	DepictQA	Embedding	Latent	✓	28.49	0.8494	0.1007	18.76	0.6254	0.1730

(a) Input (b) SD3.5 (c) Noise-free SD3.5 (d) GT

Figure 7. **Noise vs. noise-free flow matching formulation.** Notice how noise affects the eye of the parrot (zoom required).

(a) Input (b) Text (c) Embeddings (d) GT

Figure 8. **Text vs. embeddings as conditioning inputs.** Embeddings provide more efficient and informative conditioning.

Effectiveness of the RAR stopping criterion. Finally, we analyze the effectiveness of the proposed method for stopping the RAR iterations. To this end we run the RAR process for varying number of *predefined* fixed iterations and compare results to directly using our stopping criterion. Given the time consuming nature of this experimental setup, we only report results on the UDC [47] dataset in Tab. 6. Using our stopping criterion, we average 2.4 RAR rounds, which strikes a good balance between fidelity and perceptual metrics, demonstrating the efficacy of RAR.

Computational Efficiency. In addition to qualitative and quantitative comparisons, we also wanted to highlight the efficiency of the proposed iterative process in terms of latency compared to other iterative methods. In particular, we compare to AgenticIR [48] and AutoDIR [14]. In Table 7 we can see that our RAR process is a lot more efficient both in terms of latency and average number of iterations required, which shows that our all-in-one model can readily compete, both in terms of performance and efficiency, with methods using a suite of tools and other all-in-one iterative models, even while they have lower parameter count.

Table 5. **Role of the RAR iterative process.** Intermediate assessments are not accurate for SD1.5 due to the noised latents. Instead, our noise-free FM variant improves across all metrics and datasets.

Backbone	Training approach Iterative	Unknown (UDC)			Composite (Group-C)		
		PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
SD1.5	-	25.57	0.747	0.132	18.37	0.555	0.175
SD1.5	✓	23.05	0.697	0.154	18.17	0.553	0.177
SD3.5	-	28.49	0.849	0.101	18.76	0.625	0.173
SD3.5	✓	**28.60**	**0.856**	**0.083**	**19.16**	**0.649**	**0.149**

Table 6. **Stopping Criterion Analysis.** Results for unknown composite degradations on the UDC dataset.

# RAR rounds	PSNR ↑	SSIM ↑	LPIPS ↓	CLIPIQA ↑	MUSIQ ↑	MINIQA ↑
1	30.69	0.901	0.0583	0.591	54.87	0.417
2	29.55	0.873	0.073	0.601	56.29	0.428
3	29.12	0.866	0.077	0.608	56.80	0.438
4	28.60	0.856	0.083	0.614	57.19	0.446
2.4	29.69	0.874	0.071	0.603	55.97	0.423

Table 7. **Analysis of Computational Efficiency.** Comparison with iterative approaches in terms of latency and performance.

Degradations	Method	Wall clock time	# of Round	PSNR ↑	SSIM ↑	LPIPS ↓
Group A	AutoDIR	14.30	2.92	19.64	0.6286	0.3967
	AgenticIR	48.00	3.37	**21.04**	0.6818	0.3148
	RAR	**6.29**	**2.82**	20.46	**0.7144**	**0.1299**
Group B	AutoDIR	10.59	**1.98**	19.90	0.6643	0.3542
	AgenticIR	54.00	3.63	20.55	0.7009	0.3072
	RAR	**6.71**	2.84	**21.04**	**0.7326**	**0.1269**
Group C	AutoDIR	18.12	4.42	18.61	0.5443	0.5019
	AgenticIR	78.00	4.77	18.82	0.5474	0.4493
	RAR	**6.92**	**2.94**	**19.33**	**0.6579**	**0.1489**

5. Conclusion

We introduce a new methodology for image restoration that can seamlessly handle unknown composite degradations. Our method starts with aligning the tasks of free-form IQA and all-in-one restoration to form a unified model that can both assess and restore. We further improve upon this design by tightly integrating both components into a single end-to-end trainable model, where the inter-dependencies of both tasks can be fully exploited. Finally, we enable adaptive assessment, both at inference and training time, leading to our restore, assess, repeat model. Our method establishes a new state-of-the-art across all datasets considered, including single, composite and unknown degradations, all under one single test-time configuration.

References

[1] Abdelrahman Abdelhamed, Stephen Lin, and Michael S Brown. A high-quality denoising dataset for smartphone cameras. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 5, 6

[2] Eirikur Agustsson and Radu Timofte. NTIRE 2017 challenge on single image super-resolution: Dataset and study. In *IEEE Conference on Computer Vision and Pattern Recognition - Workshops*, 2017. 5, 6

[3] Haoyu Chen, Wenbo Li, Jinjin Gu, Jingjing Ren, Sixiang Chen, Tian Ye, Renjing Pei, Kaiwen Zhou, Fenglong Song, and Lei Zhu. RestoreAgent: Autonomous image restoration agent via multimodal large language models. In *Neural Information Processing Systems*, 2024. 1, 2

[4] I Chen, Wei-Ting Chen, Yu-Wei Liu, Yuan-Chun Chiang, Sy-Yen Kuo, Ming-Hsuan Yang, et al. UniRestore: Unified perceptual and task-oriented image restoration model using diffusion prior. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2025. 2

[5] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *European Conference on Computer Vision*, 2022. 2

[6] Wei-Ting Chen, Jian-Jiun Ding, and Sy-Yen Kuo. Pms-net: Robust haze removal based on patch map for single images. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2019. 2

[7] Hamadi Chihaoui and Paolo Favaro. When self-supervised pre-training meets single image denoising. In *International Conference on Image Processing*, 2024. 1, 2

[8] Marcos V Conde, Gregor Geigle, and Radu Timofte. Instructir: High-quality image restoration following human instructions. In *European Conference on Computer Vision*, 2024. 5

[9] 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 *International Conference on Learning Representations*, 2025. 5

[10] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *European Conference on Computer Vision*, 2014. 2

[11] Patrick Esser, Sumith Kulal, Andreas Blattmann, Rahim Entezari, Jonas Müller, Harry Saini, Yam Levi, Dominik Lorenz, Axel Sauer, Frederic Boesel, Dustin Podell, Tim Dockhorn, Zion English, and Robin Rombach. Scaling rectified flow transformers for high-resolution image synthesis. In *International Conference on Machine Learning*, 2024. 4

[12] Rich Franzen. Kodak lossless true color image suite, 1999. 5, 6

[13] Md Jahidul Islam, Youya Xia, and Junaed Sattar. Fast underwater image enhancement for improved visual perception. *IEEE Robotics and Automation Letters*, 2020. 5

[14] Yitong Jiang, Zhaoyang Zhang, Tianfan Xue, and Jinwei Gu. AutoDIR: Automatic all-in-one image restoration with latent diffusion. In *European Conference on Computer Vision*, 2024. 1, 2, 5, 6, 7, 8

[15] Xiangtao Kong, Chao Dong, and Lei Zhang. Towards effective multiple-in-one image restoration: A sequential and prompt learning strategy. *arXiv preprint arXiv:2401.03379*, 2024. 5

[16] Boyun Li, Xiao Liu, Peng Hu, Zhongqin Wu, Jiancheng Lv, and Xi Peng. All-In-One Image Restoration for Unknown Corruption. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2022. 5

[17] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Reside: A benchmark for single image dehazing. *arXiv preprint arXiv:1712.04143*, 2017. 5

[18] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 2018. 5, 6

[19] Ruoteng Li, Robby T Tan, and Loong-Fah Cheong. All in one bad weather removal using architectural search. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2020. 2

[20] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. SwinIR: Image restoration using swin transformer. In *IEEE International Conference on Computer Vision - Workshops*, 2021. 2

[21] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition - Workshops*, 2017. 5

[22] Xinqi Lin, Jingwen He, Ziyan Chen, Zhaoyang Lyu, Bo Dai, Fanghua Yu, Wanli Ouyang, Yu Qiao, and Chao Dong. Diffbir: Towards blind image restoration with generative diffusion prior. *arXiv preprint arXiv:2308.15070*, 2023. 1, 2

[23] Ziwei Luo, Fredrik K. Gustafsson, Zheng Zhao, Jens Sjölund, and Thomas B. Schön. Controlling vision-language models for multi-task image restoration. *International Conference on Learning Representations*, 2024. 5

[24] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2017. 2, 5, 6

[25] OpenAI. Gpt-4 technical report, 2023. 2

[26] Ozan Özdenizci and Robert Legenstein. Restoring vision in adverse weather conditions with patch-based denoising diffusion models. *IEEE Transactions of Pattern Analysis and Machine Intelligence*, 2023. 2

[27] Vaishnav Potlapalli, Syed Waqas Zamir, Salman H Khan, and Fahad Shahbaz Khan. PromptIR: Prompting for all-in-one image restoration. In *Neural Information Processing Systems*, 2024. 2, 5

[28] Rui Qian, Robby T Tan, Wenhan Yang, Jiajun Su, and Jiaying Liu. Attentive generative adversarial network for raindrop removal from a single image. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 5, 6

[29] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*, 2021. 1

[30] Sudarshan Rajagopalan, Kartik Narayan, and Vishal M. Patel. RestoreVAR: Visual autoregressive generation for all-

in-one image restoration. *arXiv preprint arXiv:2505.18047*, 2025. 1, 2

[31] Mengwei Ren, Mauricio Delbracio, Hossein Talebi, Guido Gerig, and Peyman Milanfar. Multiscale structure guided diffusion for image deblurring. In *IEEE International Conference on Computer Vision*, 2023. 1, 2

[32] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2022. 3

[33] Xiaole Tang, Xiang Gu, Xiaoyi He, Xin Hu, and Jian Sun. Degradation-aware residual-conditioned optimal transport for unified image restoration. *IEEE Transactions of Pattern Analysis and Machine Intelligence*, 2025. 5

[34] Jeya Maria Jose Valanarasu, Rajeev Yasarla, and Vishal M Patel. Transweather: Transformer-based restoration of images degraded by adverse weather conditions. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2022. 2

[35] Hong Wang, Qi Xie, Qian Zhao, and Deyu Meng. A model-driven deep neural network for single image rain removal. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2020. 2

[36] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data. In *IEEE International Conference on Computer Vision*, 2021. 5

[37] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. *British Machine Vision Conference*, 2018. 5, 6

[38] Bin Xia, Yulun Zhang, Shiyin Wang, Yitong Wang, Xinglong Wu, Yapeng Tian, Wenming Yang, and Luc Van Gool. DiffIR: Efficient diffusion model for image restoration. In *IEEE International Conference on Computer Vision*, 2023. 2

[39] Wenhan Yang, Robby T Tan, Jiashi Feng, Jiaying Liu, Zongming Guo, and Shuicheng Yan. Deep joint rain detection and removal from a single image. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2017. 2, 5, 6

[40] Xunpeng Yi, Han Xu, Hao Zhang, Linfeng Tang, and Jiayi Ma. Diff-retinex: Rethinking low-light image enhancement with a generative diffusion model. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2023. 1, 2

[41] Zhiyuan You, Zheyuan Li, Jinjin Gu, Zhenfei Yin, Tianfan Xue, and Chao Dong. Depicting beyond scores: Advancing image quality assessment through multi-modal language models. In *European Conference on Computer Vision*, 2024. 1, 2, 3, 4, 5

[42] Ke Yu, Chao Dong, Liang Lin, and Chen Change Loy. Crafting a toolchain for image restoration by deep reinforcement learning. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 1, 2

[43] 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 *IEEE Conference on Computer Vision and Pattern Recognition*, 2022. 2

[44] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2021. 1, 2

[45] Yi Zhang, Xiaoyu Shi, Dasong Li, Xiaogang Wang, Jian Wang, and Hongsheng Li. A unified conditional framework for diffusion-based image restoration. In *Neural Information Processing Systems*, 2023. 2

[46] 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 *IEEE Conference on Computer Vision and Pattern Recognition*, 2024. 1, 2

[47] Yuqian Zhou, David Ren, Neil Emerton, Sehoon Lim, and Timothy Large. Image restoration for under-display camera. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2021. 5, 6, 8

[48] Kaiwen Zhu, Jinjin Gu, Zhiyuan You, Yu Qiao, and Chao Dong. An intelligent agentic system for complex image restoration problems. *International Conference on Learning Representations*, 2025. 1, 2, 3, 5, 6, 8

[49] Yuanzhi Zhu, Kai Zhang, Jingyun Liang, Jiezhang Cao, Bihan Wen, Radu Timofte, and Luc Van Gool. Denoising diffusion models for plug-and-play image restoration. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2023. 2