

Multi-Agent Image Restoration

Xu Jiang* Gehui Li* Bin Chen* Jian Zhang†
 School of Electronic and Computer Engineering, Peking University

xujiang2642081986@gmail.com

{chenbin, lighui921}@stu.pku.edu.cn

zhangjian.sz@pku.edu.cn

Abstract

Image restoration (IR) is challenging due to the complexity of real-world degradations. While many specialized and all-in-one IR models have been developed, they fail to effectively handle complex, mixed degradations. Recent agentic methods RestoreAgent and AgenticIR leverage intelligent, autonomous workflows to alleviate this issue, yet they suffer from suboptimal results and inefficiency due to their resource-intensive finetunings, and ineffective searches and tool execution trials for satisfactory outputs. In this paper, we propose **MAIR**, a novel **Multi-Agent** approach for complex **IR** problems. We introduce a real-world degradation prior, categorizing degradations into three types: (1) scene, (2) imaging, and (3) compression, which are observed to occur sequentially in real world, and reverse them in the opposite order. Built upon this three-stage restoration framework, MAIR emulates a team of collaborative human specialists, including a “scheduler” for overall planning and multiple “experts” dedicated to specific degradations. This design minimizes search space and trial efforts, improving image quality while reducing inference costs. In addition, a registry mechanism is introduced to enable easy integration of new tools. Experiments on both synthetic and real-world datasets show that proposed MAIR achieves competitive performance and improved efficiency over the previous agentic IR system. Code and models will be made available.

1. Introduction

Image restoration (IR) is a long-standing, challenging problem in computer vision. It aims to reconstruct high-quality (HQ) original images from low-quality (LQ) degraded ones. Traditional deep IR networks [8, 37, 41, 59, 65, 67, 73, 76, 83] are typically designed for specific IR tasks, focusing on single degradations such as rain, haze, noise, blur, and JPEG compression. However, in real-world scenarios, images often suffer from multiple degradations that can interact with each other, significantly increasing the complexity of IR.

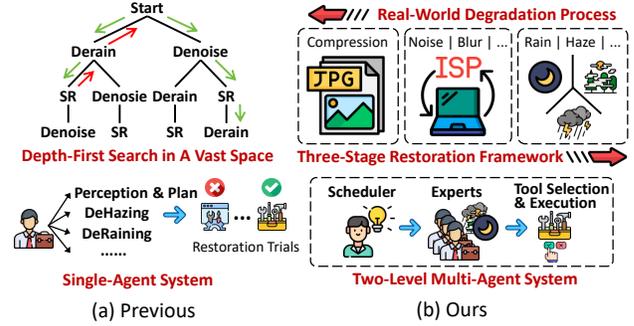


Figure 1. **Comparison between our proposed MAIR and typical agentic IR approaches.** (a) The state-of-the-art method AgenticIR [85] employs a single agent for perception, planning, restoration, etc., suffering from resource-intensive searches and trials for degradation removal. (b) We decompose the complex IR problem into manageable sub-tasks and address them using multiple collaborative agents, under our proposed three-stage restoration framework, achieving improved performance, efficiency, and flexibility.

To address this, researchers have developed All-in-One (AiO) IR methods [3, 14, 29, 32, 34, 42, 46, 48, 69] that can handle multiple degradations simultaneously using one unified model. While current AiO IR networks are more effective than traditional ones designed for single degradations, training them is more challenging due to the potential conflicts among different optimization objectives [10, 28, 85]. Moreover, these models are typically limited to 3-5 specific tasks seen during training and struggle to generalize to unseen ones. When encountering degradations not included in training, they either require retraining or suffer from a large drop in the quality of recovered images. This hampers their practical application in real-world scenarios, where images are often affected by a wider range of degradations.

The success of large language model (LLM) [2, 54, 68]-based autonomous AI agents [47, 51, 63] in handling complex tasks has inspired researchers [10, 85] to develop intelligent systems to improve the practical applicability of IR methods. In general, agentic systems intelligently perceive degradations in the given LQ image and invoke a series of off-the-shelf pretrained IR networks (referred to as “tools”) to reverse multiple degradations. They reflect on each step’s output and can roll back to previous results to explore more effective tool execution plans. This methodology expands

*Equal Contribution. †Corresponding Author.

the potential for IR performance improvement while offering greater flexibility than most existing AiO networks. For instance, RestoreAgent [10] finetunes a multi-modal LLM (MLLM) [54] to serve as the perception and planning model of an agent, enabling it to solve complex IR problems step-by-step. AgenticIR [85] incorporates statistical experience from pre-collected effective tool execution sequences into LLM [26]’s text prompts to guide agent in planning, leading to improved quality and consistency of recovered results.

Despite holding potential for autonomous and intelligent IR, existing agentic approaches [10, 85] still suffer from issues in performance and efficiency, as shown in Fig. 1 (a). **First**, given an LQ input, they search for an effective execution plan in a vast space of tool sequences without considering the characteristics of real-world degradation processes, leading to high resource consumption due to excessive trials and rollbacks with a large number of tool/LLM invocations. For example, AgenticIR can require up to 200 seconds and 20 invocations to restore a 256×256 input image with two NVIDIA 3090 GPUs. **Second**, they rely on a single agent to handle all of perception, planning, invoking tools for degradations, *etc.* This often results in suboptimal tool execution plans due to the limited capability of one single agent, constraining their effectiveness and practical applicability.

To improve performance while reducing resource consumption, we propose MAIR, a novel Multi-Agent system for complex IR problems. Moving beyond the single-agent designs in existing methods, MAIR establishes a real-world degradation prior-augmented multi-agent system. The prior is based upon our induction of real-world degradation processes and statistics of effective tool execution plans. As illustrated in Fig. 1 (b), we categorize degradations into three types, and assume that they occur sequentially in most real-world cases: (1) degradations in the scene (*e.g.*, low light and rain) [45, 53, 57], (2) degradations introduced by the imaging process (*e.g.*, noise and blur) [17, 44, 80], and (3) degradations caused by post-processings (*e.g.*, JPEG compression) [43, 56]. Based on this, we propose a three-stage framework that reverses these degradations in the order opposite to their occurrences, effectively reducing the search space and accelerating execution compared to previous approaches that lack the prior. To overcome the performance limitations of single agent, our design incorporates multiple collaborative agents at two levels for more effective IR problem-solving: a “scheduler” agent at the first level controls the overall IR process, while multiple “expert” agents at the second level leverage tools to address specific single degradations. All tools are registered in our MAIR system using textual descriptions, allowing users to easily add or modify them, and flexibly control the execution process and recovered result using instructions. This agentic system design enables MAIR to perform IR in an autonomous manner more effectively and efficiently, particularly when handling

real-world LQ inputs. In summary, our contributions are:

- (1) We propose a novel agentic IR system that consists of a three-stage framework and a two-level multi-agent design.
- (2) We develop a three-stage restoration framework based on our proposed real-world image degradation prior.
- (3) We develop a two-level multi-agent design that consists of a “scheduler” agent for overall perception and planning, and multiple “expert” agents for degradation removal.
- (4) Experiments manifest that MAIR achieves competitive image quality and superior efficiency than the previous agentic method [85]. In addition, it enjoys flexibility in following instruction and extensibility in adding new tools.

2. Related Work

Image Restoration (IR) aims to reconstruct original HQ images from their LQ observations. In the past, a large number of methods have focused on solving single-degradation problems, such as denoising [67, 83], deraining [20, 21], dehazing [18, 23, 52], and super-resolution [8, 59, 65, 76]. These methods have generally achieved state-of-the-art performance for specific types of degradations. However, real-world images can often suffer from multiple mixed degradations, causing these methods to perform poorly in the scenarios beyond their intended scope. Recent research has explored AiO IR methods [3, 14, 29, 32, 34, 42, 46, 48, 69], aiming to develop a unified framework capable of handling multiple degradations. For example, AirNet [34] employs contrastive learning to help the network distinguish image features between different IR tasks and apply the most appropriate processing. PromptIR [48] and InstructIR [14] introduce additional degradation context to guide the restoration model. MiOIR [32] incorporates sequential and prompt learning strategies to enable the network to incrementally learn individual IR tasks in an organized manner. AutoDIR [29] automatically detects and removes degradations step-by-step. DA-CLIP [42] integrates a pre-trained CLIP model [49] within a restoration network to enhance image quality. Despite these advancements, existing AiO approaches still struggle with multi-task learning, making it difficult to balance generalization ability and reconstruction performance.

Autonomous Agents are systems developed to perceive environment, make decisions, and execute actions independently. In recent years, a growing body of research has explored the use of LLMs as core controllers in autonomous agents [51, 63, 70, 71]. To enhance the ability of AI systems to solve complex problems, researchers have designed various multi-agent frameworks [22, 25, 31, 61, 64] that enable agents to specialize, coordinate, and collaborate. Some works [47, 74, 86] simulate sociological dynamics utilizing multiple agents, further expanding their capabilities.

A series of notable approaches have emerged in this field. For instance, MetaGPT [25] introduces human role struc-

tures into multi-agent systems, assigning different responsibilities to agents in software development tasks. CAMEL [36] proposes a role-playing framework, demonstrating the power of structured agent interactions in problem-solving. AutoGen [64] enhances collaborative agentic AI systems by automating agent communications and coordination, while AutoAgents [9] dynamically generates and organizes specialized agents into teams tailored for various specific tasks.

This evolution of agents has inspired new approaches for IR. RestoreAgent [10] finetunes an MLLM [54] as the perception and planning model of agent on synthetic datasets, enabling autonomous evaluation and tool execution. AgenticIR [85] develops a system to mimic a human user, leveraging statistical experience for complex IR tasks. However, the performance and efficiency of existing agentic IR methods are constrained due to the limited capability of single-agent systems and their resource-intensive search for effective tool execution plans. To tackle these problems, we propose a three-stage restoration framework based on our real-world degradation prior to reduce the search space of plans. Additionally, a two-level multi-agent system is introduced, with specialized agents for perception, planning, and reflection, as well as handling individual degradations, improving performance while suppressing total resource consumption.

3. Method

3.1. Problem Definition

The task of IR addressed in this work is to recover the original HQ image \mathbf{x}_{HQ} from its LQ input image \mathbf{x}_{LQ} , which undergoes a complex degradation process \mathcal{D} , assumed to be a composition of n single degradations $\{\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_n\}$ (such as rain, haze, noise, and blur) applied sequentially. To be formal, this degradation process can be formulated as:

$$\begin{aligned} \mathbf{x}_{\text{LQ}} &= \mathcal{D}(\mathbf{x}_{\text{HQ}}) = (\mathcal{D}_n \circ \dots \circ \mathcal{D}_2 \circ \mathcal{D}_1)(\mathbf{x}_{\text{HQ}}) \\ &= \mathcal{D}_n(\dots \mathcal{D}_2(\mathcal{D}_1(\mathbf{x}_{\text{HQ}}))). \end{aligned} \quad (1)$$

The goal is to produce a prediction $\hat{\mathbf{x}}_{\text{HQ}}$ from \mathbf{x}_{LQ} that closely approximates \mathbf{x}_{HQ} by applying n tools (*i.e.*, IR models) $\{\mathcal{R}_1, \mathcal{R}_2, \dots, \mathcal{R}_n\}$, assumed to counteract the effects of $\{\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_n\}$ in the reverse order of their application. Formally, this restoration process can be expressed as:

$$\begin{aligned} \hat{\mathbf{x}}_{\text{HQ}} &= \mathcal{R}(\mathbf{x}_{\text{LQ}}) = (\mathcal{R}_1 \circ \mathcal{R}_2 \circ \dots \circ \mathcal{R}_n)(\mathbf{x}_{\text{LQ}}) \\ &= \mathcal{R}_1(\mathcal{R}_2(\dots \mathcal{R}_n(\mathbf{x}_{\text{LQ}}))). \end{aligned} \quad (2)$$

Under the two assumptions about the degradations and restoration tools outlined above, and building on previous works [10, 85], the problem in the context of agentic IR is to determine the optimal selection of tools and their execution order (*i.e.*, a plan for their applications). Given the constraints of limited resources, the challenge is to devise

and execute this plan in a way that achieves the highest possible image quality in the recovered output, while minimizing computational cost. In the following, we will elaborate on our induced prior regarding real-world degradations and explain how we use this prior and MLLM-based agents to design an autonomous system that simulates a group of humans to collaboratively implement the process in Eq. (2).

3.2. Three-Stage Restoration Framework

Real-World Degradation Prior. In most real-world cases, image degradation does not occur in all permutations of single degradations but instead generally follows a structured process with an inherent order. This process can be broadly divided into three phases. *First*, the scene itself introduces inherent degradations due to environmental factors such as low light, rain, and haze [20, 23, 32, 33, 35]. *Second*, during the imaging process, additional degradations arise due to the imperfect propagation of light from the scene to the image sensors. These can include noise, blur, and low resolution, which result from sensor limitations, physical disturbances during capture, and signal processing constraints [17, 44, 80]. *Finally*, once captured, the image is often post-processed by several information-lossy digital compression techniques like JPEG [56] to reduce its storage requirement.

Although the exact ordering of degradations within each phase could vary and some exceptions may exist, the overall degradation often follows a sequence of degradations which can be classified into three types: inherent scene degradation $\mathcal{D}_{\text{scene}}$, degradation introduced during imaging $\mathcal{D}_{\text{imaging}}$, and degradation caused by storage compression $\mathcal{D}_{\text{compression}}$. Later-stage degradations can alter the feature distribution of earlier ones. Formally, this process can be expressed as:

$$\mathcal{D} = \mathcal{D}_{\text{compression}} \circ \mathcal{D}_{\text{imaging}} \circ \mathcal{D}_{\text{scene}}. \quad (3)$$

Each of $\mathcal{D}_{\text{scene}}$, $\mathcal{D}_{\text{imaging}}$, and $\mathcal{D}_{\text{compression}}$ can be the identity or compositions of single degradations within its respective category. We refer to this structured degradation sequence as *real-world degradation prior*, which extends the previous synthesis pipelines [59, 79] by incorporating scene degradations and broadening the scope from blind super-resolution to general IR tasks. It serves as a simplified guideline for reducing the space of degradation orderings in agentic IR.

Three-Stage Framework. Based on our assumption in Sec. 3.1 that the restoration tools are sufficiently powerful to reverse their corresponding single degradations, the ideal restoration process should follow the inverse of the degradation sequence in Eq. (3). To be more specific, the restoration should sequentially counteract compression, imaging, and scene degradations. We refer to this approach as *three-stage restoration framework*, which can be formally expressed as:

$$\mathcal{R} = \mathcal{R}_{\text{scene}} \circ \mathcal{R}_{\text{imaging}} \circ \mathcal{R}_{\text{compression}}, \quad (4)$$

where $\mathcal{R}_{\text{scene}}$, $\mathcal{R}_{\text{imaging}}$, and $\mathcal{R}_{\text{compression}}$ represent either the identity mapping, or the compositions of tool applications

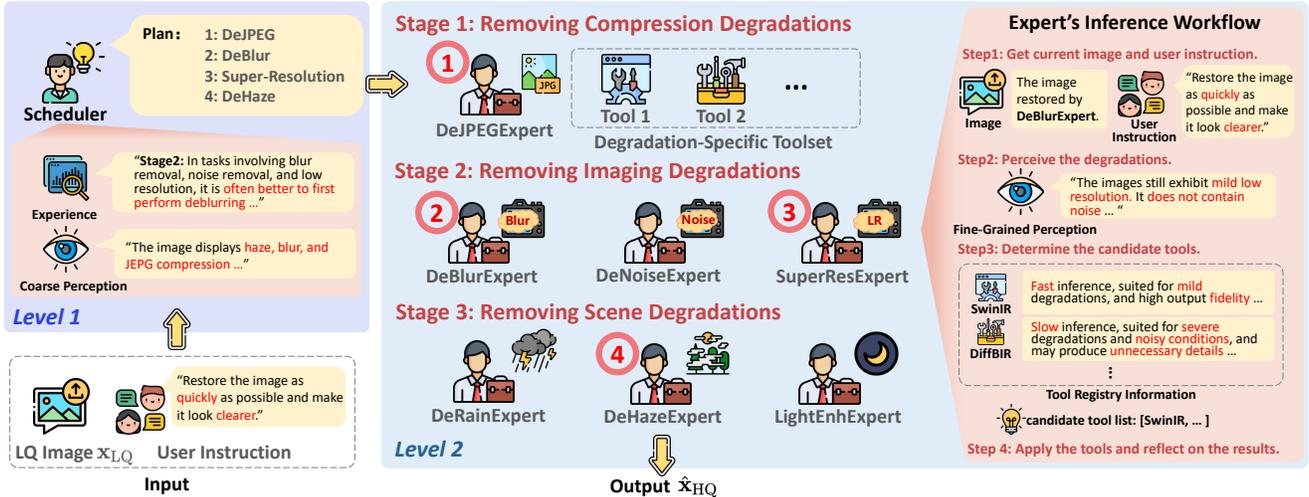


Figure 2. **Illustration of the inference workflow of MAIR.** (Left) Given an LQ image and a user instruction, a “scheduler” agent first obtains the coarse perception results of degradation types using DepictQA. It then inputs the experience, perception results, and user instruction into GPT-4o to formulate an overall restoration plan, following our three-stage framework. (Right) A group of “expert” agents sequentially removes degradations and outputs the reconstructed result, adhering to scheduler’s plan. Each expert specializes in a single degradation and uses GPT-4o to intelligently select and apply a list of candidate tools to current image, effectively removing degradation based on the image, instruction, DepictQA’s fine-grained perception results of degradation levels, tool registry information, and reflection.

Table 1. **Verification of our degradation prior and three-stage restoration framework** on five real-world validation sets, reporting the probability of the best plans adhering to our framework.

Set	T-OLED-Val	RealSR-Val	DRealSR-Val	LHP-/Real-Rain-Val	SIDD-Val
Top-1	98%	88%	93%	87%	91%
Top-3	100%	100%	100%	95%	100%

to reverse the corresponding three types of degradation, *i.e.*, $\mathcal{D}_{\text{scene}}$, $\mathcal{D}_{\text{imaging}}$, and $\mathcal{D}_{\text{compression}}$, respectively.

To verify the effectiveness of proposed prior and framework, inspired by the approach in [85] for exploring effective plans, we conduct exhaustive restoration attempts using over 13,000 tool execution sequences as plans on our pre-collected five real-world validation image sets. Specifically, we employ the MLLM DepictQA [75] of AgenticIR as the perception model to identify the degradations present in LQ inputs. Based on these perception results, we apply all permutations and combinations of corresponding tools in [85] to reverse the degradations, select the best plans, and check if they align with proposed three-stage framework. To cover a wide range of degradations, the validation datasets include 50 LQ-HQ pairs captured with under-display cameras in T-OLED [84] (T-OLED-Val), 85 and 85 low-resolution LQ-HQ pairs from RealSR [6] and DRealSR [62] (RealSR-Val and DRealSR-Val), 75 and 25 rain-degraded LQ-HQ pairs from LHP-Rain [21] and RealRain [38], merged into a single set (LHP-/Real-Rain-Val), and 100 noisy LQ-HQ pairs from SIDD [1] (SIDD-Val). To evaluate plans, we employ an extended version of the scoring function in [10], which aggregates multiple image quality assessment (IQA) metrics, including PSNR, SSIM [60], LPIPS [82], DISTS [16], MANIQA [72], CLIP-IQA [58], and MUSIQ [30]. These metrics are standardized and summed as in [10] to compute

an overall score for each recovered image, reflecting the effectiveness of a plan in restoring each given LQ input, with higher scores indicating better restoration performance.

Tab. 1 presents the statistical results of best plans aligning with our framework across the five sets. We can observe that the top-1 plans follow our framework with a probability of at least 87%, while the top-3 plans exhibit an even higher probability exceeding 95%. These findings provide strong empirical support for our prior and three-stage framework, demonstrating their applicability to real-world scenarios.

3.3. Two-Level Multi-Agent System Design

Overview. As illustrated in Fig. 2, our MAIR adopts a two-level multi-agent design. The restoration process begins at the first level, where a “scheduler” agent coarsely perceives single degradations present in the LQ input, and formulates an overall restoration plan based on user instruction to counteract them. This plan is then executed at the second level of MAIR system by a group of “expert” agents, each specializing in removing a specific type of degradation. During its turn, each expert first conducts a fine-grained perception of degradation level, selects the most appropriate tools from its toolset according to the user instruction, and applies them to the current image while reflecting on their results. Once its restoration is completed, the updated image is passed to the next expert for further processing. This iterative process continues until all relevant expert agents have finished their restorations, ultimately producing the final recovered result.

“Scheduler” for Overall Perception and Planning. At the first level, a scheduler employs a perception model DepictQA [75] (a finetuned MLLM Vicuna-v1.5-7B [13]), as in [85], to perceive the degradations present in the LQ im-

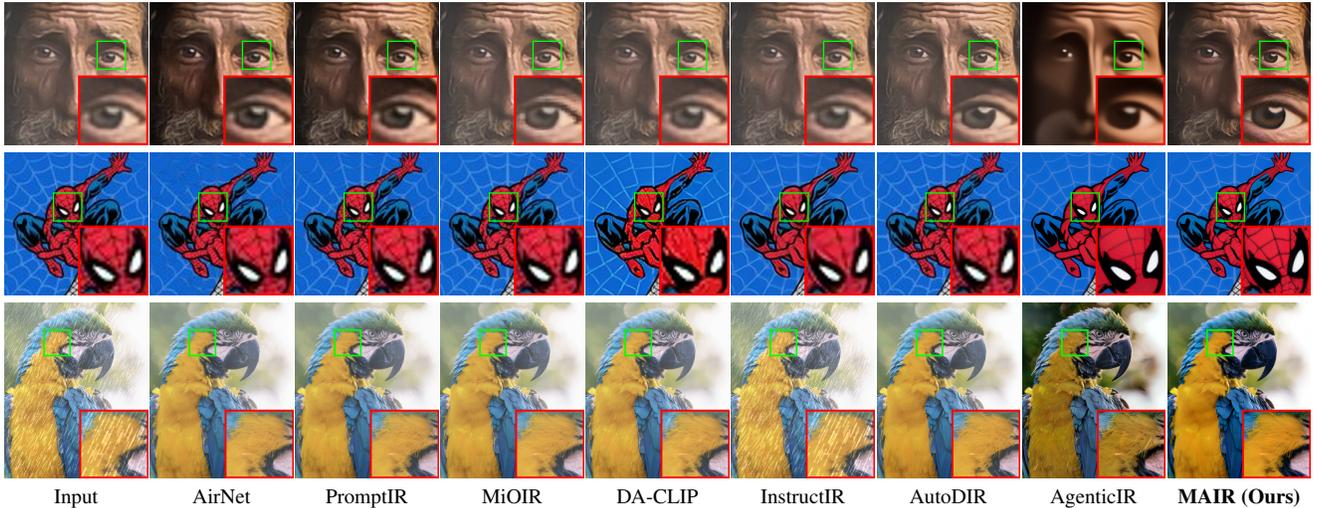


Figure 3. **Qualitative comparison** on three images from real-world paired (top), unpaired (middle) datasets, and Group A [85] (bottom).

Table 2. **Quantitative comparison of different methods on three synthesized sets.** Throughout this paper, the best and second-best results are marked in **bold red** and **underlined blue**, respectively.

Dataset	Method	PSNR	SSIM	LPIPS \downarrow	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	0.3992	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
	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
	MAIR (Ours)	21.02	0.6715	0.2963	0.3330	0.4751	59.19
	Group B	AirNet	19.31	0.6567	0.367	0.2882	0.4274
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
AutoDIR		19.9	0.6643	0.3542	0.2534	0.3986	49.64
AgenticIR		20.55	0.7009	0.3072	0.3204	0.4648	57.57
MAIR (Ours)		20.92	0.7004	0.2788	0.3544	0.5084	60.98
Group C		AirNet	17.95	0.5145	0.5782	0.1854	0.3113
	PromptIR	18.51	0.5166	0.5756	0.1906	0.3104	27.91
	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
	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
	MAIR (Ours)	19.42	0.5544	0.4142	0.2798	0.4239	51.36

age. As shown in Fig. 2 (left), it then processes the LQ image, user instruction, and coarse perception results (in text form) using MLLM GPT-4o [26] to generate a plan that follows our three-stage framework and adheres to the instruction to meet user’s specific needs. The plan is an execution sequence of expert agents that reverse single degradations.

One challenge in the planning of scheduler is that while restoration is constrained by our three-stage framework described in Eq. (4), the optimal ordering of single degradation reversions within each stage is unknown without additional priors. For example, noise and blur can coexist in an image, but their optimal reversion order in Stage 2 is uncertain without further guidance. A straightforward approach would be to rely solely on the internal knowledge of GPT-

Table 3. **Quantitative comparison** on real-world paired dataset.

Method	PSNR	SSIM	LPIPS \downarrow	DISTS \downarrow	NIQE \downarrow	MANIQA	CLIP-IQA	MUSIQ
AirNet	22.24	0.7509	0.3535	0.2256	6.59	0.2708	0.3236	42.86
PromptIR	24.13	0.7724	0.3337	0.2159	6.49	0.2827	0.3301	42.65
MiOIR	23.44	0.7499	0.3241	0.2231	5.61	0.2531	0.3225	44.76
DA-CLIP	22.76	0.7149	0.3694	0.2390	6.46	0.2860	0.3126	45.24
InstructIR	26.01	0.7910	0.3457	0.2268	6.71	0.2882	0.3341	44.25
AutoDIR	20.82	0.6770	0.3352	0.2304	5.49	0.3329	0.3623	55.74
AgenticIR	19.14	0.6574	0.3841	0.2315	5.67	0.3152	0.3779	52.69
MAIR (Ours)	21.67	0.7271	0.3244	0.2171	5.25	0.3199	0.4030	55.21

Table 4. **Quantitative comparison** on real-world unpaired dataset.

Method	NIQE \downarrow	MANIQA	CLIP-IQA	MUSIQ
AirNet	5.68	0.3426	0.5175	51.16
PromptIR	5.89	0.3518	0.5168	51.41
MiOIR	6.19	0.3677	0.5209	52.92
DA-CLIP	6.48	0.3802	0.5301	53.74
InstructIR	7.02	0.3647	0.5258	56.14
AutoDIR	6.32	0.3730	0.5439	53.35
AgenticIR	5.56	0.3773	0.5117	59.13
MAIR (Ours)	5.14	0.3968	0.5308	60.08

Table 5. **Efficiency comparison** of average running time and tool invocations per image on real-world paired and unpaired datasets.

Method	Time (s) \downarrow	Invocations \downarrow
AgenticIR	63.04	5.15
MAIR (Ours)	35.42	1.82

4o. However, this can result in suboptimal plans, as GPT-4o lacks specialized knowledge about intra-stage orderings.

To fully leverage our attempt records, and the powerful understanding and summarization capabilities of GPT-4o, we adopt an experience-driven technique inspired by [85]. Specifically, we reuse the attempt results from our experiment in Tab. 1, incorporating both tool execution sequences and corresponding scores. These records are pre-processed offline by GPT-4o in a separate phase to generate text summaries describing typically effective intra-stage restoration orders as *experience* for planning. Since these attempts contain valuable information about the performance of different plans, the extracted “experience” serves as a valid guidance input, as shown in Fig. 2 (left), helping the scheduler make

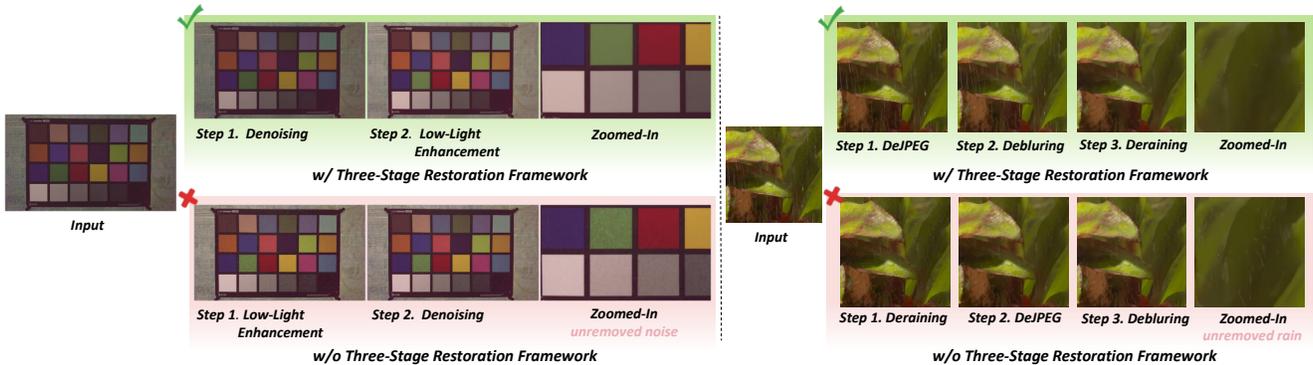


Figure 4. Ablation study of three-stage framework on two images from real-world paired (left) and LHP-/Real-Rain-Val (right) datasets.

Table 6. Ablation study of framework on real-world paired set.

Method	PSNR	SSIM	LPIPS↓	NIQE↓
w/o Three-Stage Framework	21.05	0.7187	0.3256	5.30
w/ Three-Stage Framework (Ours)	21.67	0.7271	0.3244	5.25

more informed decisions when determining the intra-stage restoration order in the inference workflow of our MAIR.

“Experts” for Fine-Grained Degradation Removals.

At the second level, we design experts to recover the image according to the scheduler’s plan in a collaborative manner. As shown in Fig. 2 (right), we assign each expert to handle a single degradation (*e.g.*, rain). Concretely, each expert first uses DepictQA to perceive the degradation and its level in the current image (*e.g.*, high-level additive white Gaussian noise). It then intelligently selects a list of “candidate” tools based on the image, user instruction, and the characteristics of its assigned tools. These tools are applied sequentially on the current image, with each result evaluated by DepictQA to check if the degradation has been successfully reduced below the set threshold (*e.g.*, “low”). If successful, the result is passed to the next agent, and the current agent’s task is complete; otherwise, the expert tries other tools from the list. If all candidate tools fail to produce satisfactory results, the expert compares each outcome pairwise and selects the highest-quality image as its restoration output.

Tool Registry Mechanism. Providing the experts with information about tools is crucial for guiding their selection of restoration candidates. While previous agentic methods achieve tool selection and degradation removal by finetuning perception and planning MLLMs (*e.g.*, LLaVA-Llama3-8B) on pairs of LQ images and tool execution sequences [10], or by directly attempting all tools corresponding to a single degradation [85], they could either require resource-intensive dataset construction and MLLM finetuning, or incur high computational cost due to less intelligence in tool selection. To address this issue, we propose a registry mechanism that equips the experts with specialized knowledge of tools. Specifically, we pre-organize tool information into a set of “registry” forms, each corresponding to a specific tool (*e.g.*, SwinIR [40]). These forms record details about all tools’ functionality (*e.g.*, denoising), applicable scenar-

Table 7. Ablation study of the intelligent candidate tool selection of our “expert” agents on the synthetic dataset Group B.

Method	PSNR	LPIPS↓	MANIQA	CLIP-IQA	MUSIQ
w/o Experts	20.14	0.3427	0.2583	0.3981	50.93
w/ Expert (Ours)	20.92	0.2788	0.3544	0.5084	60.98

ios (*e.g.*, mild levels), efficiency (*e.g.*, fast inference), and other characteristics (*e.g.*, might produce unnecessary details). When an expert agent handles a specific single degradation (*e.g.*, dehazing), the corresponding tool forms are input into GPT-4o as a part of its text input for effective tool selection, eliminating the need for dataset construction and MLLM finetuning, while achieving the flexibility for users to add new tools by simply specifying their corresponding degradation types and filling out the tool registry forms.

Compared to existing single-agent approaches [10, 85], as shown in Fig. 1 and our experiments, this two-level multi-agent design with a tool registry mechanism offers superior effectiveness in complex IR problem-solving. This is due to the specialization and coordination of our “scheduler” and “expert” agents, which decompose the entire problem into manageable sub-tasks, ensuring that both planning and execution are intelligent while preventing any single agent from being overwhelmed. As a result, each agent addresses a focused problem, enhancing both performance and efficiency.

4. Experiment

4.1. Experimental Setting

Implementation Details. We conduct experiments on both synthesized and real-world test image sets. For synthesized datasets, following [85], we employ the pre-trained models of SwinIR [40], FBCNN [27], DiffBIR [41], Restormer [78], X-Restormer [12], DRBNet [50], DehazeFormer [52], RIDCP [66], MPRNet [77], MAIXM [55], and HAT [11] along with traditional operations including gamma correction, constant shift, and histogram equalization as tools for degradation removal. These tools are used to address JPEG artifact removal (Stage 1), denoising, deblurring, and super-resolution (Stage 2), as well as low-light enhancement, deraining, and dehazing (Stage 3). For real-world images, we incorporate RetinexFormer [7], DWGAN [19], and CoTF



Figure 5. **Ablation study of two-level multi-agent system design** regarding the selection of appropriate tools for effective noise removal on an image from Group B (left), and experience-guided formulation of restoration plans on a synthesized LQ image from MiO100 (right).

Table 8. **Ablation study of multi-agent system design** on Group B.

Method	PSNR	SSIM	MANIQA	MUSIQ	Tokens ↓
Single-Agent	20.25	0.6612	0.3170	56.56	2410
Multi-Agent (Ours)	20.92	0.7004	0.3544	60.98	883

[39] into the toolsets of AgenticIR and MAIR for more effective restoration. The DepictQA model of AgenticIR, and GPT-4o, are used for perception, planning, and reflection. All experiments are conducted on two NVIDIA RTX 3090 GPUs. More details on our MAIR’s workflow, tool models, MLLM inputs, experience summarization, and tool registry forms are provided in the [Supplementary Material](#).

Test Datasets. We evaluate MAIR and compare it with other methods using three synthesized test sets: Groups A, B, and C from [85] and two real-world test sets collected by us. The three synthesized test sets contain 1,440 LQ images processed with 16 combinations of mixed 2 or 3 types of degradations applied to images from MiO100 [32]. The two real-world test sets consist of real-world samples suffering from multiple unknown degradations. The first set includes 100 paired LQ-HQ image pairs: 10, 10, 15, 15, 10, 20, and 20 pairs from I-Haze [4], NH-Haze [5], DRealSR [62], RealSR [6], T-OLED [84], SIDD [1], and LHP-Rain [21], which do not overlap with the validation sets in Tab. 1. The second testset contains 100 unpaired real-world LQ images, including 20 hazy images from the internet, 40 images from ImageNet [15], and 40 images from RealSR200 [6].

Compared Methods include six AiO IR models: AirNet [46], PromptIR [48], MiOIR [32], DA-CLIP [42], InstructIR [14], and AutoDIR [29], along with AgenticIR [85].

Evaluation Metrics for assessing the quality of recovered image results include four full-reference IQA metrics: PSNR, SSIM [60], LPIPS [82], and DISTS [16], as well as the four no-reference IQA metrics: NIQE [81], MANIQA [72], CLIP-IQA [58], and MUSIQ [30]. Additionally, FID [24] is also employed to measure the distance between the

Table 9. **Ablation study of “experience”** on real-world paired set.

Method	PSNR	SSIM	LPIPS ↓	FID ↓
w/o Experience	20.99	0.7041	0.3408	120.66
w/ Experience (Ours)	21.67	0.7271	0.3244	115.61

distributions of ground truth and restored images.

4.2. Comparison with State-of-the-Arts

Image Quality Comparison. As reported in Tabs. 2, 3, and 4, MAIR achieves competitive image quality across all five test sets. Specifically, it ranks first or second in Groups A, B, and C, outperforming AgenticIR in all perceptual metrics, including LPIPS, MANIQA, CLIP-IQA, and MUSIQ, while surpassing all the six AiO models. On real-world sets, MAIR delivers competitive results across all metrics, significantly outperforming its baseline agentic approach [85].

Fig. 3 visually demonstrates the superiority of MAIR in restoring three test images. Concretely, it employs a set of tools to effectively remove complex degradations, including haze, noise, low resolution, and compression degradation in the top and middle real-world images, as well as rain and haze in the bottom synthesized image. In contrast, other approaches struggle to reconstruct vivid details around the old man’s eye and the anime character’s face, or fail to remove JPEG artifacts and haze, leading to suboptimal IR outputs. These results comprehensively validate the effectiveness of proposed MAIR in handling diverse degradation scenarios.

Efficiency Comparison. Tab. 5 exhibits that MAIR reduces running time and tool invocations by 44% and 65%, respectively, compared to [85]. This efficiency gain is due to our three-stage framework and multi-agent design, which constrain the search space of plans and enable training-free intelligent tool selection, minimizing unnecessary trials and rollbacks while maintaining competitive performance.

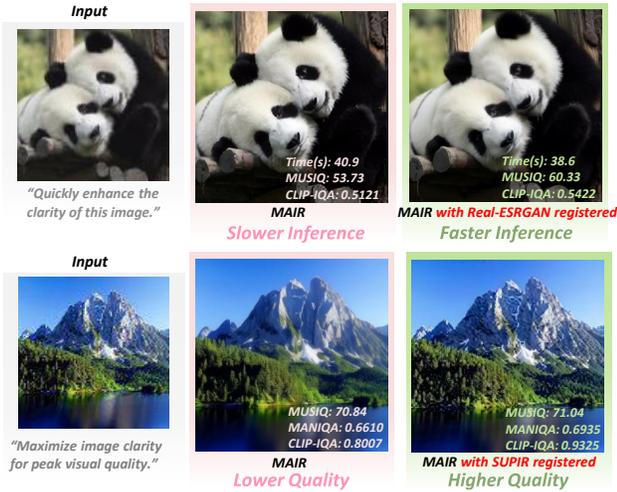


Figure 6. Evaluation of MAIR’s extensibility and flexibility in adding new tools on two images from real-world unpaired dataset.

4.3. Ablation Study

Effect of three-stage framework. Tab. 6 exhibits that removing the framework from scheduler results in consistent performance drops of 0.62dB in PSNR, 0.0084 in SSIM, 0.0012 in LPIPS, and 0.05 in NIQE. Fig. 4 further demonstrates that without our framework, the scheduler often formulates suboptimal execution orders, which are crucial for removing multiple degradations [10, 85]. These findings indicate that our proposed prior and framework are effective.

Effect of the intelligent tool selection of experts. Tab. 7 exhibits that eliminating intelligent selection (*i.e.*, experts’ Step 3 in Fig. 2) and instead using random tool selections for each degradation leads to performance drops of 0.78 dB in PSNR, 0.0639 in LPIPS, 0.0961 in MANIQA, 0.1103 in CLIP-IQA, and 10.05 in MUSIQ, validating the necessity of expert agents selecting appropriate tools using GPT-4o.

Effect of multi-agent system design. Tab. 8 shows that replacing multi-agent design with one single agent that handles all model information and experience—aggregated into the input of GPT-4o to address all the sub-tasks—results in image quality drops of 0.67dB in PSNR, 0.0392 in SSIM, 0.0374 in MANIQA, and 4.42 in MUSIQ. Additionally, it increases GPT-4o’s token consumption by $2.73\times$ due to the aggregation of text descriptions of tool registry forms, experience, *etc.*, leading to higher costs. Fig. 5 further manifests that a single-agent approach often selects inappropriate tools and fails to follow the summarized experience, resulting in suboptimal planning and execution, as evidenced by unremoved noise (left) and agents’ thoughts (right). These results verify the effectiveness of our multi-agent design in collaboratively solving complex IR tasks while maintaining training-free, intelligent tool selection and flexibility.

Effect of experience. Tab. 9 exhibits that the elimination of our summarized experience results in performance drops of 0.68dB in PSNR, 0.023 in SSIM, 0.0164 in LPIPS, and

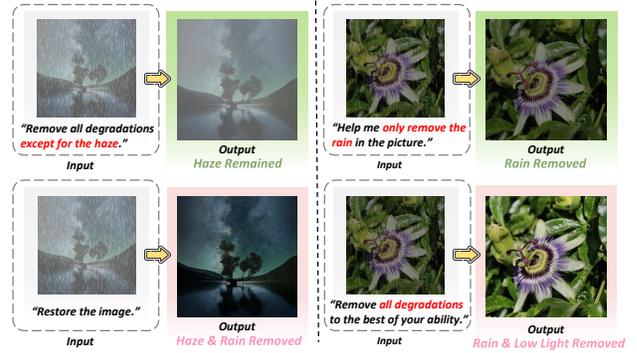


Figure 7. Evaluation of MAIR’s instruction-following capability on two images from the synthetic Groups A (left) and B (right).

5.05 in FID, highlighting its importance for guiding MAIR in real-world scenarios, as also previously validated in [85].

4.4. Evaluation of Extensibility and Flexibility

Fig. 6 shows that MAIR allows users to easily extend its capabilities by adding new tools through our registry mechanism. Specifically, we register two models, Real-ESRGAN [59] (fast inference with moderate quality) and SUPIR [76] (higher quality with lower speed), into super-resolution expert. When users request higher speed or quality, MAIR intelligently calls these models to either reduce running time by 2.3s or enhance details in Stage 2. Fig. 7 demonstrates the instruction-following capabilities of MAIR: when users request retaining certain degradations, such as haze or rain, MAIR is able to understand these preferences and perform personalized restoration accordingly. These results comprehensively verify the extensibility, flexibility, and controllability of MAIR—features that are lacking in RestoreAgent [10] and AgenticIR [85], which require MLLM fine-tuning when adding tools or exhibit less intelligence in selection.

5. Conclusion

This paper introduces MAIR, a novel Multi-Agent system that emulates a team of human specialists to tackle complex IR problems. We model real-world image degradations as a composition of three categories of single degradations and reverse them in the opposite order of their occurrence. Built upon this three-stage framework, we develop a multi-agent system consisting of a “scheduler” for planning and multiple “experts” specialized in counteracting individual degradations using pre-trained IR models (referred to as “tools”). A registry mechanism is further introduced to enable easy integration of tools. Experiments on both synthetic and real-world datasets exhibit that proposed MAIR achieves competitive image quality and higher efficiency than the previous agentic IR method [85], while offering greater flexibility and controllability compared to non-agentic approaches.

Despite being effective, our proposed three-stage framework can not fully cover all real-world degradations, which are often highly complicated and unknown. In addition, al-

though MAIR shows greater flexibility than most AiO models, its inference remains slow, requiring tens of seconds for an image. We leave these limitations for future research.

References

- [1] Abdelrahman Abdelhamed, Stephen Lin, and Michael S Brown. A high-quality denoising dataset for smartphone cameras. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1692–1700, 2018. 4, 7
- [2] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023. 1
- [3] Yuang 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25432–25444, 2024. 1, 2
- [4] Cosmin Ancuti, Codruta O Ancuti, Radu Timofte, and Christophe De Vleeschouwer. I-haze: A dehazing benchmark with real hazy and haze-free indoor images. In *Advanced Concepts for Intelligent Vision Systems: 19th International Conference, ACIVS 2018, Poitiers, France, September 24–27, 2018, Proceedings 19*, pages 620–631. Springer, 2018. 7
- [5] Codruta O Ancuti, Cosmin Ancuti, and Radu Timofte. Nh-haze: An image dehazing benchmark with non-homogeneous hazy and haze-free images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 444–445, 2020. 7
- [6] Jianrui Cai, Hui Zeng, Hongwei Yong, Zisheng Cao, and Lei Zhang. Toward real-world single image super-resolution: A new benchmark and a new model. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3086–3095, 2019. 4, 7
- [7] Yuanhao Cai, Hao Bian, Jing Lin, Haoqian Wang, Radu Timofte, and Yulun Zhang. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12504–12513, 2023. 6
- [8] Bin Chen, Gehui Li, Rongyuan Wu, Xindong Zhang, Jie Chen, Jian Zhang, and Lei Zhang. Adversarial diffusion compression for real-world image super-resolution. *arXiv preprint arXiv:2411.13383*, 2024. 1, 2
- [9] Guangyao Chen, Siwei Dong, Yu Shu, Ge Zhang, Jaward Sesay, Börje F Karlsson, Jie Fu, and Yemin Shi. Autoagents: A framework for automatic agent generation. *arXiv preprint arXiv:2309.17288*, 2023. 3
- [10] 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. *arXiv preprint arXiv:2407.18035*, 2024. 1, 2, 3, 4, 6, 8
- [11] Xiangyu Chen, Xintao Wang, Jiantao Zhou, Yu Qiao, and Chao Dong. Activating more pixels in image super-resolution transformer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 22367–22377, 2023. 6
- [12] Xiangyu Chen, Zheyuan Li, Yuandong Pu, Yihao Liu, Jiantao Zhou, Yu Qiao, and Chao Dong. A comparative study of image restoration networks for general backbone network design. In *European Conference on Computer Vision*, pages 74–91. Springer, 2024. 6
- [13] Wei-Lin Chiang, Zhuohan Li, Ziqing Lin, Ying Sheng, Zhanghao Wu, Hao Zhang, Lianmin Zheng, Siyuan Zhuang, Yonghao Zhuang, Joseph E Gonzalez, et al. Vicuna: An open-source chatbot impressing gpt-4 with 90%* chatgpt quality. See <https://vicuna.lmsys.org> (accessed 14 April 2023), 2(3):6, 2023. 4
- [14] Marcos V Conde, Gregor Geigle, and Radu Timofte. Instructir: High-quality image restoration following human instructions. In *European Conference on Computer Vision*, pages 1–21. Springer, 2024. 1, 2, 7
- [15] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. 7
- [16] Keyan Ding, Kede Ma, Shiqi Wang, and Eero P Simoncelli. Image quality assessment: Unifying structure and texture similarity. *IEEE transactions on pattern analysis and machine intelligence*, 44(5):2567–2581, 2020. 4, 7
- [17] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2015. 2, 3
- [18] Deniz Engin, Anil Genç, and Hazim Kemal Ekenel. Cycle-dehaze: Enhanced cyclegan for single image dehazing. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 825–833, 2018. 2
- [19] Minghan Fu, Huan Liu, Yankun Yu, Jun Chen, and Keyan Wang. Dw-gan: A discrete wavelet transform gan for non-homogeneous dehazing, 2021. 6
- [20] Jinjin Gu, Xianzheng Ma, Xiangtao Kong, Yu Qiao, and Chao Dong. Networks are slacking off: Understanding generalization problem in image deraining. *Advances in Neural Information Processing Systems*, 36, 2024. 2, 3
- [21] Yun Guo, Xueyao Xiao, Yi Chang, Shumin Deng, and Luxin Yan. From sky to the ground: A large-scale benchmark and simple baseline towards real rain removal. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12097–12107, 2023. 2, 4, 7
- [22] Rui Hao, Linmei Hu, Weijian Qi, Qingliu Wu, Yirui Zhang, and Liqiang Nie. Chatllm network: More brains, more intelligence. *arXiv preprint arXiv:2304.12998*, 2023. 2
- [23] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, 33(12):2341–2353, 2010. 2, 3
- [24] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a

- two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017. 7
- [25] Sirui Hong, Xiawu Zheng, Jonathan Chen, Yuheng Cheng, Jinlin Wang, Ceyao Zhang, Zili Wang, Steven Ka Shing Yau, Zijuan Lin, Liyang Zhou, et al. Metagpt: Meta programming for multi-agent collaborative framework. *arXiv preprint arXiv:2308.00352*, 2023. 2
- [26] Aaron Hurst, Adam Lerer, Adam P Goucher, Adam Perelman, Aditya Ramesh, Aidan Clark, AJ Ostrow, Akila Welihinda, Alan Hayes, Alec Radford, et al. Gpt-4o system card. *arXiv preprint arXiv:2410.21276*, 2024. 2, 5
- [27] Jiayi Jiang, Kai Zhang, and Radu Timofte. Towards flexible blind jpeg artifacts removal. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4997–5006, 2021. 6
- [28] Junjun Jiang, Zengyuan Zuo, Gang Wu, Kui Jiang, and Xianming Liu. A survey on all-in-one image restoration: Taxonomy, evaluation and future trends. *arXiv preprint arXiv:2410.15067*, 2024. 1
- [29] 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*, pages 340–359. Springer, 2024. 1, 2, 7
- [30] Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. Musiq: Multi-scale image quality transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5148–5157, 2021. 4, 7
- [31] Sehoon Kim, Suhong Moon, Ryan Tabrizi, Nicholas Lee, Michael W Mahoney, Kurt Keutzer, and Amir Gholami. An llm compiler for parallel function calling. *arXiv preprint arXiv:2312.04511*, 2023. 2
- [32] 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. 1, 2, 3, 7
- [33] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1):492–505, 2018. 3
- [34] Boyun Li, Xiao Liu, Peng Hu, Zhongqin Wu, Jiancheng Lv, and Xi Peng. All-in-one image restoration for unknown corruption. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 17452–17462, 2022. 1, 2
- [35] Chongyi Li, Chunle Guo, Linghao Han, Jun Jiang, Ming-Ming Cheng, Jinwei Gu, and Chen Change Loy. Low-light image and video enhancement using deep learning: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 44(12):9396–9416, 2021. 3
- [36] Guohao Li, Hasan Hammoud, Hani Itani, Dmitrii Khizbullin, and Bernard Ghanem. Camel: Communicative agents for “mind” exploration of large language model society. *Advances in Neural Information Processing Systems*, 36: 51991–52008, 2023. 3
- [37] Gehui Li, Jinyuan Liu, Long Ma, Zhiying Jiang, Xin Fan, and Risheng Liu. Fearless luminance adaptation: A macro-micro-hierarchical transformer for exposure correction. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 7304–7313, 2023. 1
- [38] Wei Li, Qiming Zhang, Jing Zhang, Zhen Huang, Xinmei Tian, and Dacheng Tao. Toward real-world single image deraining: A new benchmark and beyond. *arXiv preprint arXiv:2206.05514*, 2022. 4
- [39] Ziwen Li, Feng Zhang, Meng Cao, Jinpu Zhang, Yuanjie Shao, Yuehuan Wang, and Nong Sang. Real-time exposure correction via collaborative transformations and adaptive sampling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2984–2994, 2024. 7
- [40] Jingyun Liang, Jiezhong Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1833–1844, 2021. 6
- [41] Xinqi Lin, Jingwen He, Ziyan Chen, Zhaoyang Lyu, Bo Dai, Fanghua Yu, Yu Qiao, Wanli Ouyang, and Chao Dong. Diffbir: Toward blind image restoration with generative diffusion prior. In *European Conference on Computer Vision*, pages 430–448. Springer, 2024. 1, 6
- [42] Ziwei Luo, Fredrik K Gustafsson, Zheng Zhao, Jens Sjölund, and Thomas B Schön. Controlling vision-language models for universal image restoration. *arXiv preprint arXiv:2310.01018*, 3(8), 2023. 1, 2, 7
- [43] Michael W Marcellin, Michael J Gormish, Ali Bilgin, and Martin P Boliek. An overview of jpeg-2000. In *Proceedings DCC 2000. Data compression conference*, pages 523–541. IEEE, 2000. 2
- [44] Gary A Mastin. Adaptive filters for digital image noise smoothing: An evaluation. *Computer Vision, Graphics, and Image Processing*, 31(1):103–121, 1985. 2, 3
- [45] Dong Nan, Du-yan Bi, Chang Liu, Shi-ping Ma, and Linyuan He. A bayesian framework for single image dehazing considering noise. *The Scientific World Journal*, 2014(1): 651986, 2014. 2
- [46] Dongwon Park, Byung Hyun Lee, and Se Young Chun. All-in-one image restoration for unknown degradations using adaptive discriminative filters for specific degradations. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5815–5824. IEEE, 2023. 1, 2, 7
- [47] Joon Sung Park, Joseph O’Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th annual acm symposium on user interface software and technology*, pages 1–22, 2023. 1, 2
- [48] Vaishnav Potlapalli, Syed Waqas Zamir, Salman H Khan, and Fahad Shahbaz Khan. Promptir: Prompting for all-in-one image restoration. *Advances in Neural Information Processing Systems*, 36, 2024. 1, 2, 7
- [49] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR, 2021. 2

- [50] Lingyan Ruan, Bin Chen, Jizhou Li, and Miuling Lam. Learning to deblur using light field generated and real defocus images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16304–16313, 2022. 6
- [51] Yongliang Shen, Kaitao Song, Xu Tan, Dongsheng Li, Weiming Lu, and Yueting Zhuang. Hugginggpt: Solving ai tasks with chatgpt and its friends in hugging face. *Advances in Neural Information Processing Systems*, 36, 2024. 1, 2
- [52] Yuda Song, Zhuqing He, Hui Qian, and Xin Du. Vision transformers for single image dehazing. *IEEE Transactions on Image Processing*, 32:1927–1941, 2023. 2, 6
- [53] Jean-Philippe Tarel and Nicolas Hautiere. Fast visibility restoration from a single color or gray level image. In *2009 IEEE 12th international conference on computer vision*, pages 2201–2208. IEEE, 2009. 2
- [54] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*, 2023. 1, 2, 3
- [55] Zhengzhong Tu, Hossein Talebi, Han Zhang, Feng Yang, Peyman Milanfar, Alan Bovik, and Yinxiao Li. Maxim: Multi-axis mlp for image processing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5769–5780, 2022. 6
- [56] Gregory K Wallace. The jpeg still picture compression standard. *Communications of the ACM*, 34(4):30–44, 1991. 2, 3
- [57] Cong Wang, Xiaoying Xing, Yutong Wu, Zhixun Su, and Junyang Chen. Dcsfn: Deep cross-scale fusion network for single image rain removal. In *Proceedings of the 28th ACM international conference on multimedia*, pages 1643–1651, 2020. 2
- [58] Jianyi Wang, Kelvin CK Chan, and Chen Change Loy. Exploring clip for assessing the look and feel of images. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 2555–2563, 2023. 4, 7
- [59] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1905–1914, 2021. 1, 2, 3, 8
- [60] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 4, 7
- [61] Z Wang, S Mao, W Wu, T Ge, F Wei, and H Ji. Unleashing cognitive synergy in large language models: A task-solving agent through multi-persona selfcollaboration. *arXiv 2023. arXiv preprint arXiv:2307.05300*, 2023. 2
- [62] Pengxu Wei, Ziwei Xie, Hannan Lu, Zongyuan Zhan, Qixiang Ye, Wangmeng Zuo, and Liang Lin. Component divide-and-conquer for real-world image super-resolution. In *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VIII 16*, pages 101–117. Springer, 2020. 4, 7
- [63] Chenfei Wu, Shengming Yin, Weizhen Qi, Xiaodong Wang, Zecheng Tang, and Nan Duan. Visual chatgpt: Talking, drawing and editing with visual foundation models. *arXiv preprint arXiv:2303.04671*, 2023. 1, 2
- [64] Qingyun Wu, Gagan Bansal, Jieyu Zhang, Yiran Wu, Shaokun Zhang, Erkang Zhu, Beibin Li, Li Jiang, Xiaoyun Zhang, and Chi Wang. Autogen: Enabling next-gen llm applications via multi-agent conversation framework. *arXiv preprint arXiv:2308.08155*, 2023. 2, 3
- [65] Rongyuan Wu, Tao Yang, Lingchen Sun, Zhengqiang Zhang, Shuai Li, and Lei Zhang. Seesr: Towards semantics-aware real-world image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 25456–25467, 2024. 1, 2
- [66] Rui-Qi Wu, Zheng-Peng Duan, Chun-Le Guo, Zhi Chai, and Chongyi Li. Ridcp: Revitalizing real image dehazing via high-quality codebook priors. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 22282–22291, 2023. 6
- [67] Xiaohu Wu, Ming Liu, Yue Cao, Dongwei Ren, and Wangmeng Zuo. Unpaired learning of deep image denoising. In *European conference on computer vision*, pages 352–368. Springer, 2020. 1, 2
- [68] Zhiheng Xi, Wenxiang Chen, Xin Guo, Wei He, Yiwen Ding, Boyang Hong, Ming Zhang, Junzhe Wang, Senjie Jin, Enyu Zhou, et al. The rise and potential of large language model based agents: A survey. *Science China Information Sciences*, 68(2):121101, 2025. 1
- [69] Bin Xia, Shiyin Wang, Yingfan Tao, Yitong Wang, and Jiaya Jia. Llmga: Multimodal large language model based generation assistant. In *European Conference on Computer Vision*, pages 389–406. Springer, 2024. 1, 2
- [70] Hui Yang, Sifu Yue, and Yunzhong He. Auto-gpt for online decision making: Benchmarks and additional opinions. *arXiv preprint arXiv:2306.02224*, 2023. 2
- [71] Rui Yang, Lin Song, Yanwei Li, Sijie Zhao, Yixiao Ge, Xiu Li, and Ying Shan. Gpt4tools: Teaching large language model to use tools via self-instruction. *Advances in Neural Information Processing Systems*, 36, 2024. 2
- [72] Sidi Yang, Tianhe Wu, Shuwei Shi, Shanshan Lao, Yuan Gong, Mingdeng Cao, Jiahao Wang, and Yujiu Yang. Maniqa: Multi-dimension attention network for no-reference image quality assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1191–1200, 2022. 4, 7
- [73] Shuzhou Yang, Xuanyu Zhang, Yinhuai Wang, Jiwen Yu, Yuhan Wang, and Jian Zhang. Diffle: Diffusion-based domain calibration for weak supervised low-light image enhancement. *International Journal of Computer Vision*, pages 1–20, 2024. 1
- [74] Ziyi Yang, Zaibin Zhang, Zirui Zheng, Yuxian Jiang, Ziyue Gan, Zhiyu Wang, Zijian Ling, Jinsong Chen, Martz Ma, Bowen Dong, et al. Oasis: Open agents social interaction simulations on one million agents. *arXiv preprint arXiv:2411.11581*, 2024. 2
- [75] Zhiyuan You, Jinjin Gu, Zheyuan Li, Xin Cai, Kaiwen Zhu, Chao Dong, and Tianfan Xue. Descriptive image quality

- assessment in the wild. *arXiv preprint arXiv:2405.18842*, 2024. 4
- [76] Fanghua Yu, Jinjin Gu, Zheyuan Li, Jinfan Hu, Xiangtao Kong, Xintao Wang, Jingwen He, Yu Qiao, and Chao Dong. Scaling up to excellence: Practicing model scaling for photo-realistic image restoration in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25669–25680, 2024. 1, 2, 8
- [77] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14821–14831, 2021. 6
- [78] 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 *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5728–5739, 2022. 6
- [79] Kai Zhang, Jingyun Liang, Luc Van Gool, and Radu Timofte. Designing a practical degradation model for deep blind image super-resolution. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4791–4800, 2021. 3
- [80] Kaihao Zhang, Wenqi Ren, Wenhan Luo, Wei-Sheng Lai, Björn Stenger, Ming-Hsuan Yang, and Hongdong Li. Deep image deblurring: A survey. *International Journal of Computer Vision*, 130(9):2103–2130, 2022. 2, 3
- [81] Lin Zhang, Lei Zhang, and Alan C Bovik. A feature-enriched completely blind image quality evaluator. *IEEE Transactions on Image Processing*, 24(8):2579–2591, 2015. 7
- [82] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018. 4, 7
- [83] Yi Zhang, Dasong Li, Ka Lung Law, Xiaogang Wang, Hongwei Qin, and Hongsheng Li. Idr: Self-supervised image denoising via iterative data refinement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2098–2107, 2022. 1, 2
- [84] Yuqian Zhou, David Ren, Neil Emerton, Schoon Lim, and Timothy Large. Image restoration for under-display camera. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9179–9188, 2021. 4, 7
- [85] Kaiwen Zhu, Jinjin Gu, Zhiyuan You, Yu Qiao, and Chao Dong. An intelligent agentic system for complex image restoration problems. In *The Thirteenth International Conference on Learning Representations*, 2025. 1, 2, 3, 4, 5, 6, 7, 8
- [86] Mingchen Zhuge, Haozhe Liu, Francesco Faccio, Dylan R Ashley, Róbert Csordás, Anand Gopalakrishnan, Abdullah Hamdi, Hasan Abed Al Kader Hammoud, Vincent Hermann, Kazuki Irie, et al. Mindstorms in natural language-based societies of mind. *arXiv preprint arXiv:2305.17066*, 2023. 2