An Intelligent Agentic System for Complex Image Restoration Problems

Kaiwen Zhu* Jinjin Gu* Zhiyuan You Yu Qiao Chao Dong

Introduction

- Inspired by human problem-solving, we propose AgenticIR, an agentic system that mimics humans for complex image restoration.
- AgenticIR leverages large language models (LLMs) and vision-language models (VLMs) that interact via text generation to dynamically operate a toolbox of IR models. We fine-tune VLMs for image quality analysis and employ LLMs for reasoning.
- To compensate for LLMs' lack of specific IR knowledge and experience, we introduce a self-exploration method, allowing the LLM to observe and summarize restoration results into referenceable documents.
- Experiments demonstrate AgenticIR's potential in handling complex IR tasks, representing a promising path toward general intelligence in visual processing.

Why Agentic Image Processing?

Limitation of end-to-end paradigm

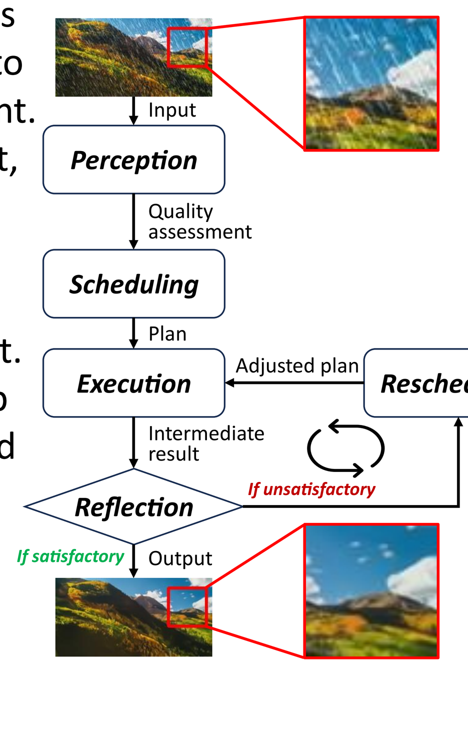
- **Generality-performance dilemma:** Different from image understanding models, image restoration models do not learn "semantics" but overfit to pixel-level transformations. Training on more tasks for generality inevitably compromises performance.
- **Intelligence manifestation:** Human users may anticipate explicit stages like perceiving image content, reasoning about the strategy, and interaction with tools. Such dynamic and adaptive processes can be recognized as intelligence. Monolithic models do not leave room for this.

Promise of agentic paradigm

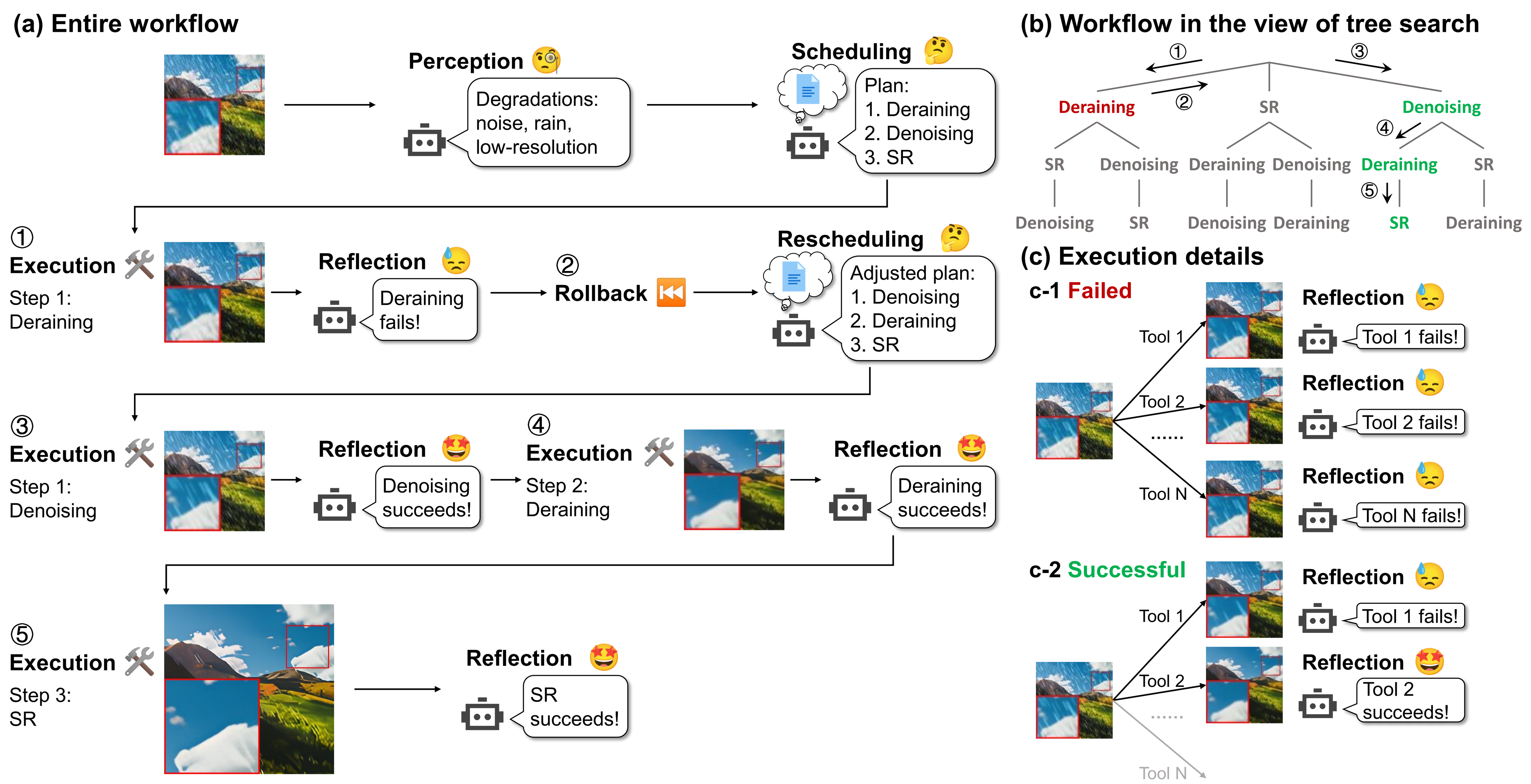
- By invoking various external single-task tools, we can address a wide range of tasks with state-of-the-art performance.
- It is possible to establish architectures that exhibits human-like cognitive characteristics. A proactive agent with this architecture could flexibly process images in a way appreciated by human users, thus demonstrating intelligence in a certain aspect.

Key Agentic Patterns

We abstract how human users use tools to process images. First one needs to analyze the image content. Based on the assessment, a plan is scheduled and then executed. During execution, one would reflect on the tool output. If unsatisfactory, the step should be rolled back and the plan is accordingly adjusted. These stages can be summarized as ***Perception, Scheduling, Execution, Reflection, and Rescheduling.***



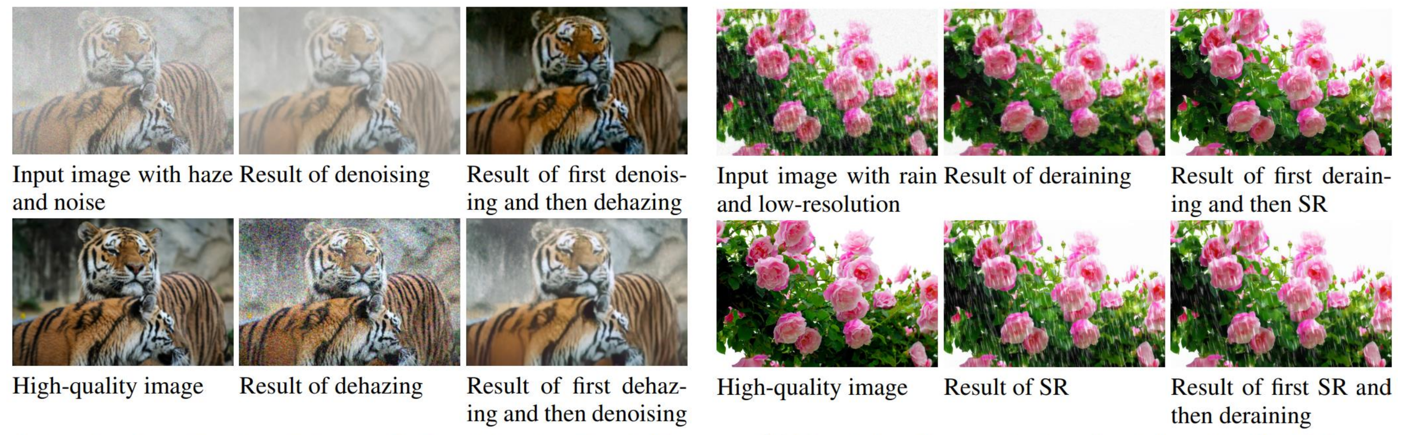
Inference Workflow



Taming LLM and VLM for Low-Level Vision

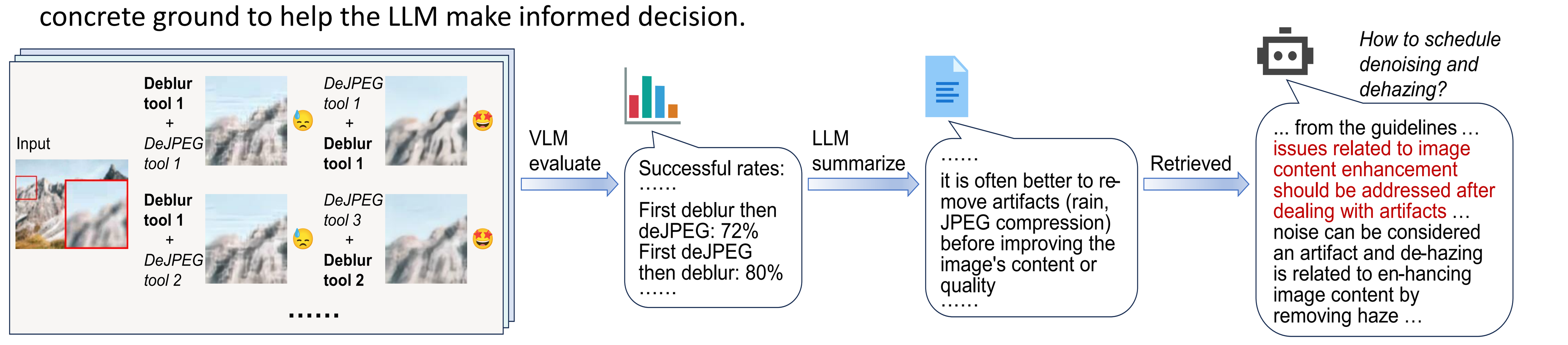
LLM for scheduling image restoration operations

- The order of restoration operations often significantly impacts the result. Different operations may have distinct requirements or side effects on images, forming complex interference.



(a) Dehazing after denoising works but dehazing first fails. (b) Deraining first works but deraining after SR fails.

- LLMs alone suffer from hallucination and thus cannot address the interference to give a reliable plan.
- We propose a self-exploration and experience summarization method, allowing the LLM to observe and summarize restoration results, organizing the experiment knowledge into referenceable document. During inference, the experience is retrieved as a concrete ground to help the LLM make informed decision.



Exploration | Statistics (Experience) | Documents (Knowledge) | Referring to document in inference

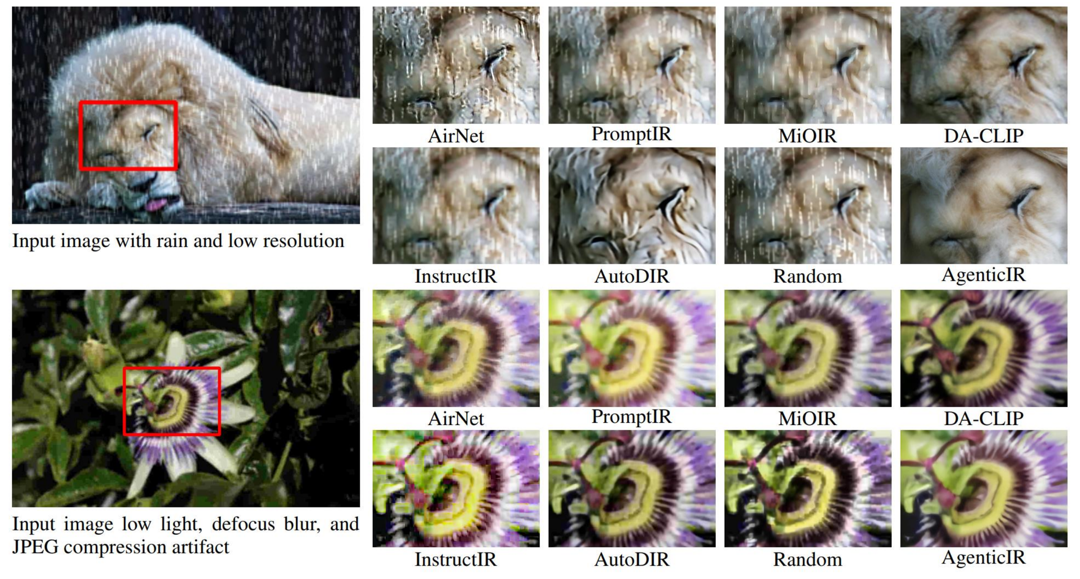
VLM for image quality assessment

- The VLM can describe the image content in natural language to provide necessary information for our agent. Specifically, we need it to analyze the image content and categorize the quality issue that should be addressed.
- DepictQA-Wild is a VLM dedicated to image quality assessment. We fine-tune it by LoRA to further adapt it to the specific requirement of evaluating severities of various degradations.
- The fine-tuning is quite efficient courtesy of the extensive knowledge acquired in pre-training.

Experimental Setup

- Consider 8 types of single degradations
  - Low resolution, noise, motion blur, defocus blur, rain, haze, JPEG compression artifact, low light
- Design 16 combinations of degradations
- Divide the combinations into group A, B, C
  - Group A/B/C contains 8/4/4 combinations
  - Each one in group A and B contain 2 degradations
  - Each one in group C contain 3 degradations
- Apply each combination on 100 images
- Allocate 160 images in group A for self-exploration and experience summarization
- Allocate the remaining 1,440 images for test

Comparison with All-in-one Models



Degradations	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
	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**
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
	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**
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
	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**

Effectiveness of Designs

Self-exploration and experience summarization

Insights distilled from experience
... From these observations, we can infer that generally ... it is often better to remove artifacts (rain, JPEG compression) before improving the image's content or quality ...

Guide: How to schedule ... ?

Deraining, SR (in group A)
According to the collected experience ... it is more effective to address the artifact caused by rain before enhancing the image's resolution ...

Denoising, dehazing (in group B)
... from the guidelines ... issues related to image content enhancement should be addressed after dealing with artifacts ... noise can be considered an artifact and dehazing is related to enhancing image content by removing haze ...

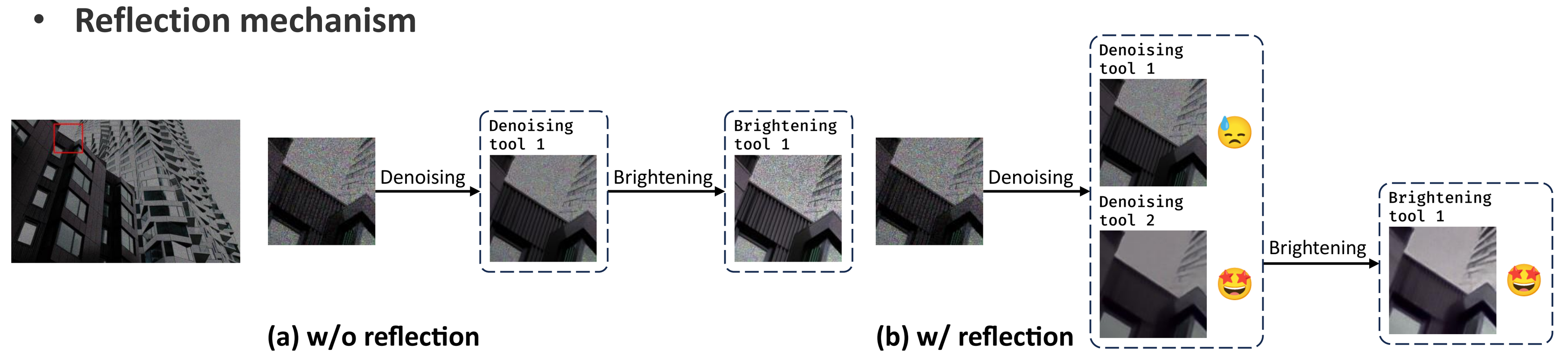
Brightening, defocus deblurring, JPEG artifact removal (in group C)
... more effective to address blurring issues before image content enhancement ... removing artifacts like compression should be done before addressing blurring issues ...

(a) w/o experience (b) w/ experience

Degradations	As planned	PSNR	SSIM	LPIPS↓	MANIQA	CLIP-IQA	MUSIQ
Group A	✓	**21.14**	**0.6836**	**0.2753**	**0.3469**	**0.5091**	**60.77**
	✗	20.79	0.6652	0.3060	0.3385	0.4819	59.85
Group B	✓	**21.14**	**0.7088**	**0.2683**	0.3588	**0.5275**	**61.92**
	✗	20.32	0.6811	0.2976	**0.3623**	0.5257	60.15
Group C	✓	**18.78**	**0.5352**	**0.4239**	**0.3118**	**0.4876**	51.08
	✗	18.49	0.5277	0.4345	0.3058	0.4719	**51.32**

Inference workflow

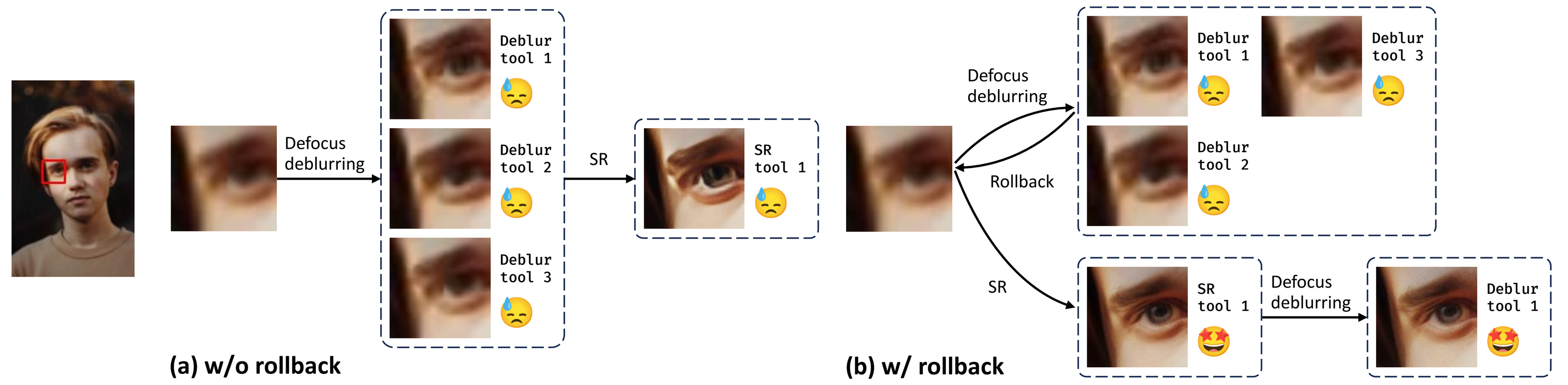
- **Reflection mechanism**



(a) w/o reflection (b) w/ reflection

Degradations	Method Ref.	Method Rb.	PSNR	SSIM	LPIPS↓	MANIQA	CLIP-IQA	MUSIQ
Group A	✓	✗	**21.12**	**0.6809**	**0.3079**	**0.3179**	**0.4617**	**57.52**
	✗	✗	20.47	0.6659	0.3282	0.2906	0.4387	55.56
Group B	✓	✗	**20.74**	**0.6986**	**0.3084**	**0.3126**	**0.4567**	**56.66**
	✗	✗	20.46	0.6798	0.3412	0.2966	0.4359	54.81
Group C	✓	✗	18.85	**0.5510**	**0.4559**	**0.2557**	**0.3771**	**47.38**
	✗	✗	**18.93**	0.5447	0.4764	0.2349	0.3595	43.77

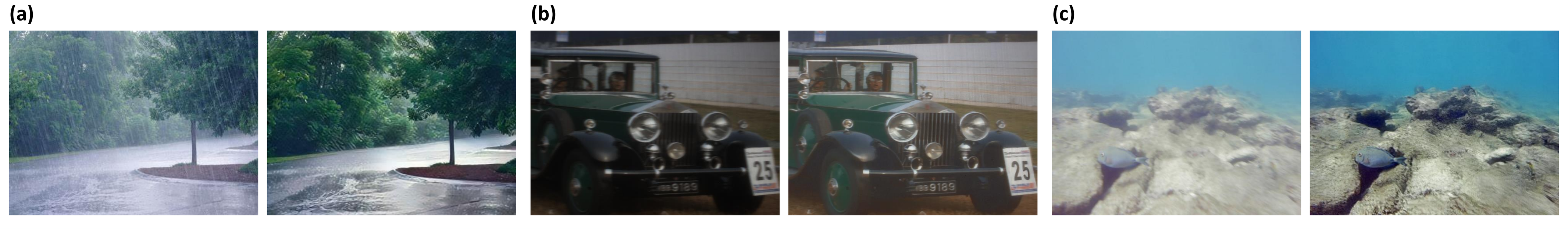
- **Rollback mechanism**



(a) w/o rollback (b) w/ rollback

Degradations	Method Ref.	Method Rb.	PSNR	SSIM	LPIPS↓	MANIQA	CLIP-IQA	MUSIQ
Group A	✓	✓	**20.23**	0.6626	0.3249	**0.3197**	0.4158	**59.87**
	✓	✗	19.77	**0.6725**	**0.3067**	0.3042	**0.4484**	58.70
Group B	✓	✓	**18.76**	**0.6642**	**0.3348**	**0.3251**	0.4525	**57.43**
	✓	✗	18.30	0.6348	0.3591	0.3082	**0.4528**	55.94
Group C	✓	✓	**18.99**	**0.5461**	**0.4604**	**0.2643**	**0.3974**	**49.64**
	✓	✗	18.64	0.5446	0.4634	0.2348	0.3669	46.73

Real-World Application



(a) taken in heavy rain, restored by deraining and dehazing; **(b)** taken by under-display camera, restored by motion deblurring, defocus deblurring, and brightening; **(c)** taken underwater, restored by defocus deblurring, dehazing, and motion deblurring;