

Review

Not peer-reviewed version

Agentic Image Restoration Links Closed-Loop Controllers with Protocol-Aware Evaluation

Yuezhe Yang [†], Chengru Li [†], Zhuodong Chai [†], [Xingbo Dong](#) ^{*}, [Zhe Jin](#)

Posted Date: 15 September 2026

Keywords: agentic image restoration; AI agents; multimodal large language models; reinforcement learning; memory-augmented agents; multi-agent systems



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC, OpenAlex.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Review

Agentic Image Restoration Links Closed-Loop Controllers with Protocol-Aware Evaluation

Yuezhe Yang [†], Chengru Li [†], Zhuodong Chai [†], Xingbo Dong ^{*} and Zhe Jin

Anhui Provincial International Joint Research Center for Advanced Technology in Medical Imaging, Anhui University, Hefei, China

^{*} Correspondence: xingbo.dong@ahu.edu.cn

[†] These authors contributed equally to this work.

Abstract

Image restoration is shifting from fixed inference pipelines towards systems that diagnose degradation, choose tools, assess intermediate results, and revise their actions. However, the term “agentic” currently spans architectures with different control mechanisms and evaluation assumptions. We review representative studies published from 2018 to 2026 and offer a controller-centred synthesis of agentic image restoration (AIR). Our perspective separates four functional components, perception, decision, action, and reflection or memory, from five controller families: reinforcement learning, prompt conditioning, multimodal large language model tool use, memory-augmented planning, and multi-agent coordination. Evidence from photography, autonomous driving, medical reconstruction, remote sensing, and microscopy shows that controller comparisons are often confounded by incompatible degradations, tool pools, feedback signals, and budgets. We therefore pair the taxonomy with protocol-aware reporting priorities for output quality, downstream utility, execution cost, failure recovery, and provenance. This framework clarifies the boundary between conditioned restoration and autonomous control, identifies unresolved evaluation gaps, and provides a practical map for developing reliable and auditable AIR systems.

Keywords: agentic image restoration; AI agents; multimodal large language models; reinforcement learning; memory-augmented agents; multi-agent systems

1. Introduction

Image restoration supports photography, autonomous perception, medical imaging, remote sensing, and scientific measurement. It estimates a latent image $\mathbf{x}$ from a degraded observation

$$\mathbf{y} = \mathcal{D}(\mathbf{x}; \boldsymbol{\eta}) + \mathbf{n}, \tag{1}$$

where $\mathcal{D}$ represents acquisition or environmental degradation, $\boldsymbol{\eta}$ its parameters, and $\mathbf{n}$ measurement noise. We use restoration broadly to include degradation removal, reconstruction, super-resolution, and quality enhancement.

Classical priors and modern neural networks have substantially improved restoration quality [1–7]. Most systems nevertheless execute a predetermined inference procedure. Prompt-conditioned models such as PromptIR and InstructIR adapt internal features, but their standard implementations do not evaluate external tool sequences or revise failed plans [8,9]. This distinction matters under composite or changing degradations, where the correct operation and its order may be unknown.

Agentic image restoration (AIR) adds runtime control to the restoration pipeline. A controller can diagnose degradation, select and sequence tools, inspect intermediate results, and revise its actions or memory [10–12]. These capabilities appear in different combinations, and the term “agentic” is consequently applied to systems with substantially different control structures.

Figure 1 distinguishes fixed mapping, prompt conditioning, and closed-loop restoration. The defining feature of AIR is not model size or iterative computation alone. It is an observable decision process that connects perception, action, feedback, and optional memory.

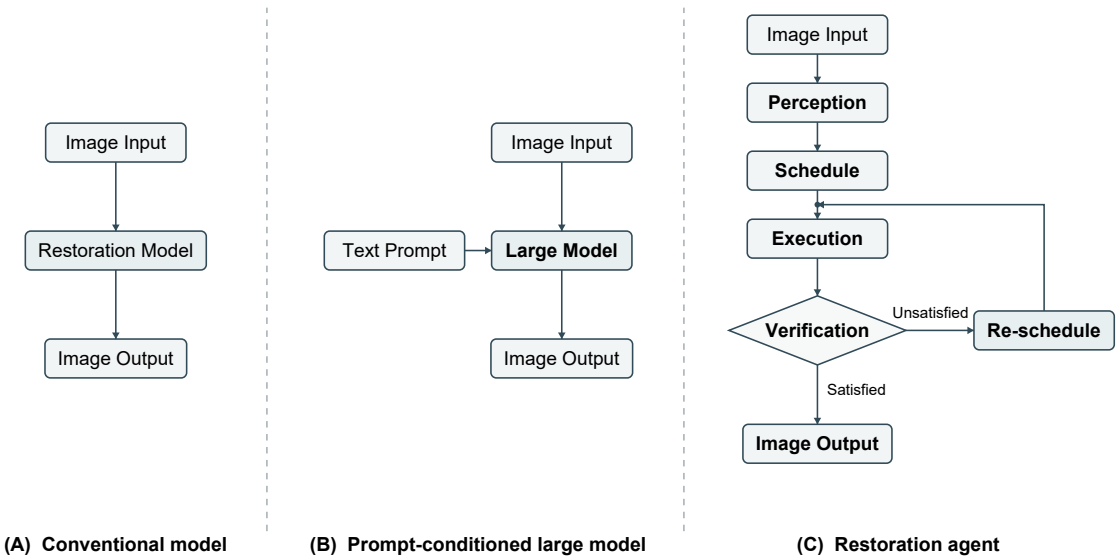


Figure 1. Three control formulations for image restoration. Conventional models apply a fixed mapping. Prompt-conditioned models adapt a shared backbone to an instruction or learned condition. Restoration agents observe image conditions, schedule and execute operations, verify quality, and revise unsatisfactory plans.

This narrative review asks three questions. What makes a restoration system agentic at inference time? Which controller architectures have been reported? How should their outputs and decision processes be evaluated? We organise the evidence into five controller families: cybernetic reinforcement learning (P1), prompt-conditioned restoration (P2), multimodal large language model (MLLM) reasoning and tool use (P3), memory-augmented agents (P4), and multi-agent systems (P5). An orthogonal four-component anatomy describes perception, decision, action, and reflection or memory.

The review covers primary AIR studies published from January 2018 to August 2026. It includes natural photography, 4K super-resolution [13,14], autonomous driving [15], medical reconstruction [16–19], microscopy [20], and cultural heritage [21,22]. We synthesise these studies by controller mechanism rather than chronology. We then identify protocol differences that prevent pooled rankings and propose evaluation priorities for reliable, auditable AIR.

2. Review Methodology and Unified Taxonomy

2.1. Review Methodology

We searched PubMed, IEEE Xplore, ACM Digital Library, SpringerLink, ScienceDirect, Wiley, IOP, CVF Open Access, OpenReview, NeurIPS proceedings, arXiv, and Research Square. Queries combined agent-related terms with restoration, reconstruction, denoising, super-resolution, enhancement, correction, or sampling. Modality-specific terms covered medical imaging, microscopy, remote sensing, photography, and cultural heritage. Citation chaining, exact-title searches, DOI checks, and official project repositories extended and verified the candidate set. The search covered 1 January 2018 to 9 August 2026, with metadata rechecked on 16 August 2026.

Studies were eligible when a runtime controller affected restoration, reconstruction, enhancement, acquisition, or correction through an identifiable decision interface, action space, feedback process, or memory mechanism. Prompt-conditioned systems were retained as a boundary category when their condition explicitly controlled restoration behaviour. We excluded papers in which “agent” denoted only an attention representation, or reinforcement learning controlled only architecture

development. Duplicate preprints and published articles were merged, with the version of record preferred. Withdrawn or retracted records were excluded.

We extracted publication metadata, domain, controller, perception input, action space, feedback, memory, and public resources. Numerical results were recorded with their dataset, degradation, comparator, metric, and protocol, then checked against the primary source. The reference database contains 89 records. The technical evidence set comprises 49 studies, including 45 controller or boundary studies, two medical extensions, and two agent-evaluation suites. Supplementary Tables S1–S7 provide detailed evidence matrices and source-traceable numerical examples.

We coded each system by perception, decision, action, and reflection or memory. Controller classes follow the locus and mechanism of runtime control. Boundary labels identify hybrid designs or iterative learned controllers without an external MLLM tool loop. Because datasets, degradation mixtures, tools, stopping rules, and metric implementations differ, quantitative values are interpreted within their source protocols rather than pooled.

2.2. Four-Component Anatomy

AIR extends feed-forward restoration into a feedback process (Figure 2). Perception translates images and acquisition information into an actionable observation. A decision component selects operations from an available tool space. Reflection evaluates intermediate or final outputs and can stop, retry, or roll back. Memory may retain useful outcomes for later decisions.

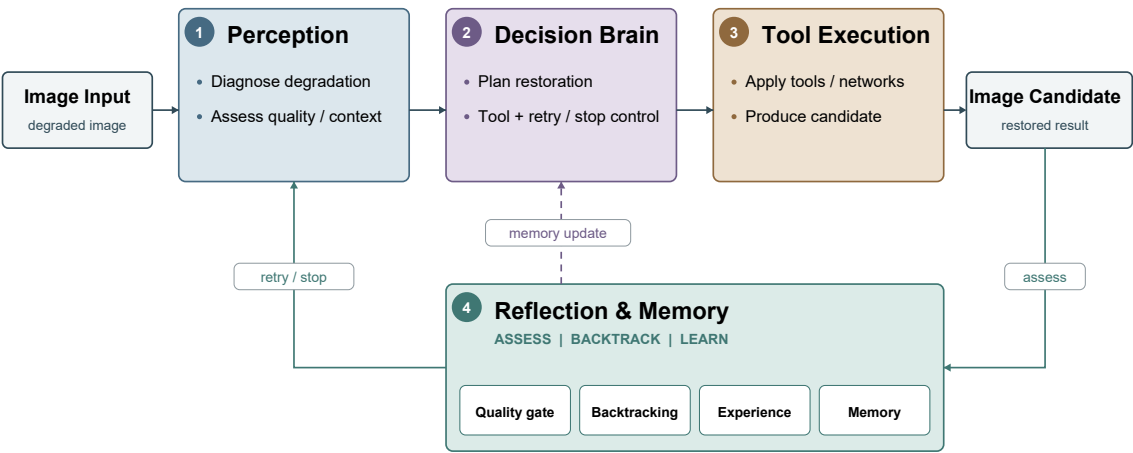


Figure 2. Four-component anatomy of AIR. Perception converts a degraded input into an actionable observation. The controller selects and executes tools. Reflection uses quality feedback to stop, retry, or roll back, while memory can retain useful experience.

Perception.

Systems use learned degradation embeddings, MLLM diagnoses, acquisition states, or image-quality measurements. DA-CLIP learns content and degradation representations [23]. RestoreAgent and JarvisIR generate semantic diagnoses for planning [11,15]. RAR instead uses learned quality assessment inside an iterative latent process [14].

Decision.

The controller maps the current observation and history to an action. Implementations range from reinforcement-learning policies [24] to MLLM planners, memory-augmented planning [12], and distributed coordination [25,26].

Action.

Available operations may include classical filters, specialised neural models, continuous controls, or sandboxed code. Tool strength and diversity affect both flexibility and planning difficulty.

Comparisons must therefore disclose exact checkpoints, parameters, permitted order, and execution budgets.

Reflection and memory.

Feedback may be absent, episode-local, or persistent. RestoreAgent uses execution history to revise assignments, whereas RAR performs learned assessment and restoration in latent space [11,14]. AgenticIR and SEAR retrieve prior trajectories [12,27]. Causal-AgentIR updates structured relations among degradations, actions, and outcomes [28].

2.3. Five Controller Families

Figure 3 situates representative systems from 2018 to 2026. The families overlap historically and describe dominant runtime control, not a strict maturity ladder.

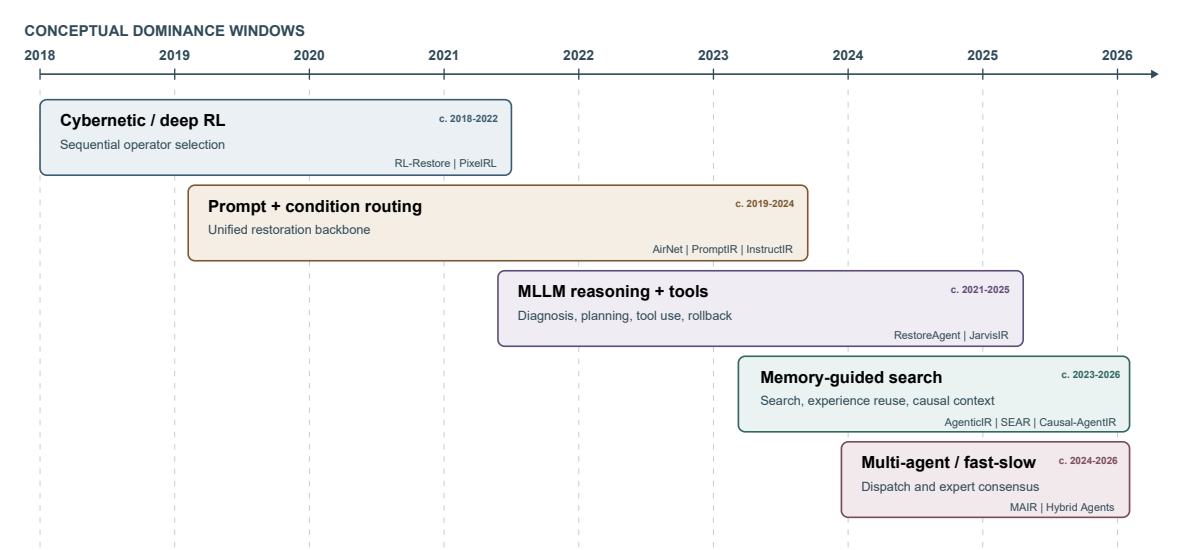


Figure 3. Representative AIR developments from 2018 to 2026. The five families are organised by dominant controller: reinforcement learning, prompt conditioning, MLLM tool orchestration, memory-augmented planning, and multi-agent coordination.

P1: cybernetic reinforcement learning.

A policy optimised through reinforcement learning selects tools, parameters, spatial routes, or acquisition steps. Rewards can combine restoration quality, physical fidelity, or computational cost.

P2: prompt-conditioned restoration.

A learned degradation representation, visual prompt, or instruction modulates a shared backbone. Most systems perform one conditioned inference without external tool verification, retry, or persistent memory.

P3: MLLM reasoning and tool use.

An MLLM diagnoses degradation, decomposes the task, and invokes specialised restoration tools. Some systems inspect intermediate results and replan, while others retain a fixed execution path.

P4: memory-augmented restoration.

The controller retrieves successful trajectories or reasons over structured memory. Memory changes later decisions across inputs and distinguishes these systems from episode-local execution history.

P5: multi-agent restoration.

Control is distributed among specialised roles, such as a scheduler, restoration experts, and a feedback agent. Coordination and communication become additional sources of cost and failure.

Hybrid attributes do not automatically change the controller class. Generated-code execution, for example, can expand the action space of a P3, P4, or P5 system. We assign each system by its primary decision mechanism and treat benchmarks as evaluation resources outside P1–P5.

Tables 1–3 compare representative systems across controller, perception, action, feedback, memory, domain, and publication status.

Table 1. Representative P1 and P2 systems. P denotes the dominant controller paradigm; B marks a boundary case.

Work	Venue/ year	P	Perception	Action	Feedback/ memory	Domain
RL-Restore [24]	CVPR 2018	P1	CNN features	Twelve operators	PSNR reward and stop action	Natural
Distort-Recover [31]	CVPR 2018	P1	Global state	Colour actions	Unpaired reward	Retouching
PixelRL [29]	AAAI 2019	P1	Pixel context	Pixel actions	Multi-step reward	Natural
Path-Restore [30]	TPAMI 2022	P1	Patch difficulty	Routing paths	Efficiency reward	Natural
CT-DRL [32]	TMI 2018	P1	Iteration state	Reconstruction parameters	Quality reward	CT
MRI-RL [34]	Neurocomputing 2024	P1	Dual <i>k</i> -space states	Sampling actions	SSIM reward	MRI
AirNet [36]	CVPR 2022	P2	Degradation embedding	Feature modulation	Open loop	Natural
PromptIR [8]	NeurIPS 2023	P2	Learned prompts	Prompted Transformer	Open loop	Natural
InstructIR [9]	ECCV 2024	P2	Instruction	Conditioned network	Open loop	Natural
DA-CLIP [23]	ICLR 2024	P2	CLIP embeddings	Feature control	Open loop	Natural
AutoDIR [39]	ECCV 2024	P2/B	BIQA or text	Latent diffusion	Iterative routing	Natural
OneRestore [40]	ECCV 2024	P2	Scene and degradation	Composite restorer	Open loop	Natural

Table 2. Representative P3 systems and learned boundary cases.

Work	Venue/ year	P	Perception	Action	Feedback/ memory	Domain
Clarity [10]	Preprint 2023	P3	ChatGPT and CLIP	Classical and neural tools	User feedback	Natural
RestoreAgent [11]	NeurIPS 2024	P3	MLLM diagnosis	Model registry	Replan and rollback	Natural
JarvisIR [15]	CVPR 2025	P3	Driving VLM	Weather pipeline	Perception feedback	Driving
4KAgent [13]	NeurIPS 2025	P3	VLM and IQA	Mixture of experts	Recursive reflection	Super-resolution
RAR [14]	CVPR 2026	P2/B	Latent assessor	Latent restoration	Restore–assess–repeat	Natural
AgentMRI [16]	JiIM 2026	P3	Medical VLM	Correction paths	Consensus gate	MRI
EpiAgent [21]	CVPR 2026	P3	Iterative observation	Restoration tools	Reevaluation	Heritage
SIMBA [20]	Preprint 2026	P3	Language description	Tool calls	Verification	Microscopy
FAPE-IR [51]	CVPR 2026	P3	Frequency planner	LoRA experts	Conditioned execution	Natural
Restore-R1 [52]	CVPR Findings 2026	P1/B	Visual encoder	RL tool selector	Quality reward	Natural

Table 3. Representative P4 and P5 systems.

Work	Venue/ year	P	Perception	Action	Feedback/ memory	Domain
AgentIR [12]	ICLR 2025	P4	Multimodal diagnosis	Tool registry	Rollback and experience	Natural
SEAR [27]	Preprint 2026	P4	Episodic fingerprint	Dual-process planner	P-MCTS and memory	Natural
Causal-AgentIR [28]	Preprint 2026	P4	Quality and history	Causal graph	Persistent updates	Natural
MAIR [25]	IJCV 2026	P5	Scheduler	Seven experts	Candidate comparison	Natural
Hybrid Agents [26]	CVPR 2026	P5	Fast, slow, feedback	Conditional restoration	Feedback selection	Natural
MAPGR [22]	npj Heritage Sci. 2026	P5	Specialised roles	Residual diffusion	Evidence and abstention	Heritage

3. Cybernetic Reinforcement Learning

3.1. Foundational MDP Formulation of Cybernetic RL Agents

Early reinforcement-learning (RL) restoration systems introduced explicit states, actions, rewards, and termination decisions. They formulate image recovery as a Markov Decision Process (MDP) governed by the tuple $(\mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R}, \gamma)$ instead of a static mapping $f_\theta(\mathbf{y}) = \hat{\mathbf{x}}$. A policy π_ϕ observes the degraded image state $\mathbf{s}_t \in \mathcal{S}$, selects an action $a_t \in \mathcal{A}$, and receives the updated image $\mathbf{x}_{t+1} = \mathcal{T}(\mathbf{x}_t, a_t)$. Systems with a learned stop action terminate when the policy predicts that further processing is unnecessary.

The main design variables are the structure and spatial granularity of the action space $\mathcal{A}$. In a discrete MDP, the agent selects an index from a predefined catalogue of lightweight operators. In a continuous MDP, the action specifies numerical parameters such as colour-curve control points, regularisation weights $\lambda(x)$ in iterative CT reconstruction, or smoothing coefficients (σ_s, σ_r) in bilateral filtering. Spatial scope also varies. Some policies apply global actions, while PixelRL operates at pixel level and Path-Restore routes image patches through branches of different depth.

Policy learning is driven by a stepwise quality reward $r_t = \mathcal{R}(\mathbf{s}_t, a_t, \mathbf{s}_{t+1})$, which can quantify incremental gains in restoration metrics or physical fidelity. Some systems also include computational cost or acquisition constraints. This closed-loop evaluation allows the policy to adapt its trajectory to the observed state and establishes the control formulation used by the systems discussed in Section 3.2.

3.2. Representative Systems and Methodological Evolution

We compare representative RL systems through four control designs. These are global discrete tool selection, spatially adaptive routing, continuous parameter tuning, and dual-state active acquisition.

RL-Restore [24] treats restoration as a discrete multi-step decision problem. A Q-learning agent observes global image features, selects from a fixed toolbox of lightweight CNNs for blur, noise, and compression, and may emit a stop action. The reward is based on stepwise PSNR gain. Its action selection operates at image level and therefore applies uniform control within each image.

PixelRL [29] assigns an agent to each pixel in a fully convolutional A3C architecture. Reward Map Convolution incorporates neighbouring future states into the reward calculation to coordinate local decisions. Path-Restore [30] uses a Pathfinder module trained with a difficulty-regulated reward. It routes easier patches through shallow paths and harder patches through deeper residual paths, trading restoration quality against computation.

Discrete action spaces can still support fine-grained sequential adjustment. Distort-and-Recover [31] uses discrete global colour-editing actions and learns from synthetically distorted high-quality images, avoiding paired input and retouched-target training data. Medical inverse problems instead expose reconstruction or filtering parameters as the action interface. Shen *et al.* [32] use a parameter-tuning policy for iterative CT reconstruction. Patwari *et al.* [33] apply reinforcement learning

to joint bilateral filtering parameters under an image-quality reward. These systems retain an explicit processing interface while adapting parameters to the observed reconstruction state.

Online reconstruction can make active sampling expensive when a deep reconstructor is evaluated at every decision. DUAL [34] separates a hidden DNN state used to compute training rewards from a visible state used by the online phase-selection policy. The source reports 9 ms decision time, compared with 1363 ms and 1333 ms for two reference methods, corresponding to an approximately 150-fold speed-up in that experimental setting.

PaAgent [35] combines an RL tool-selection policy with subjective and objective feedback. Its name refers to a portrait-aware bank of restoration tools. The reported action space contains 14 restoration tools and covers 14 degradation scenarios. Group Relative Policy Optimisation trains the policy using MLLM semantic assessment together with no-reference image-quality signals. The MLLM therefore contributes reward information while the learned policy remains the action selector.

4. Prompt-Modulated and Condition-Routing Systems

P2 systems broaden restoration coverage by conditioning a shared network on a learned degradation representation, a prompt, or a natural-language instruction. They expose an explicit control signal that adapts the restoration network. In most cases, this signal modulates the backbone within one inference procedure, without a verified sequence of external tool calls.

4.1. Implicit and Learned Degradation Prompts

AirNet [36] learns a degradation representation contrastively and uses it to support all-in-one restoration. PromptIR [8] introduces learned prompts into a Transformer restoration backbone, allowing degradation-dependent feature modulation. MPerceiver [37] extends multi-task restoration with a multi-modal prompt-perception design. These models may adapt computation inside the network through model-internal routing and conditioning.

4.2. Language and Vision–Language Conditioning

InstructIR [9] conditions restoration on human instructions, making the requested operation explicit. DA-CLIP [23] learns degradation-aware representations in the representation space of CLIP [38] for restoration. AutoDIR [39] uses BIQA to identify a predominant degradation and can apply its structural-corrected latent-diffusion restorer iteratively, while OneRestore [40] represents scenes and degradations for composite restoration. Their conditioning mechanisms differ and are therefore described separately.

Language-driven adverse-weather restoration (LDR) [41] uses a pre-trained vision–language model to generate description-based priors for degradation occurrence, type, and severity, then sparsely selects restoration experts in a mixture-of-experts network. SPIRE [42] uses content prompts and language-specified restoration strength to control a generative restoration model. UniRes [43] addresses complex degradations by combining specialised models during diffusion sampling and exposing a fidelity–quality trade-off. These systems use model-internal routing or conditioning mechanisms.

SUPIR [44] and DreamClear [45] use generative priors for real-world restoration and can use semantic conditions. Their perceptual-quality objectives differ from full-reference PSNR-oriented restoration. Numerical comparison therefore requires a common dataset and metric implementation.

4.3. P2–P3 Boundary

The classification criterion is the locus of control. When a prompt is encoded once and the restoration backbone produces an output without explicit observe–act–verify decisions, the system is P2. AutoDIR is treated as P2/B because its closed-set BIQA module can repeatedly select a dominant degradation within a fixed routing space. If an MLLM diagnoses the image, invokes external tools, examines results, and can revise the plan, it is P3 or above. RAR [14] is an iterative learned boundary case based on internal latent control.

Supplementary Table S1 provides the mechanism-level comparison of representative prompt-conditioned systems. Pooled PSNR is not reported because the sources use different tasks and protocols.

P2 establishes that a single backbone can be controlled by an interpretable or learned condition. Runtime tool orchestration, auditable retry, and persistent experience define the transition to higher-level agentic controllers.

5. Single-MLLM Reasoning and Tool-Orchestrating Agents

Multimodal large language models (MLLMs) enable restoration systems to express a degradation diagnosis, select specialised tools, and interpret intermediate outputs. We use P3 for systems in which one language or vision–language controller is the principal runtime decision maker. Prompt-conditioned single-pass networks remain P2. Learned iterative models such as RAR [14] form a boundary group because their assessment and restoration occur end-to-end in latent space.

5.1. Explicit Diagnosis and Task-Coupled Perception

RestoreAgent [11] made degradation identification and restoration-model assignment explicit, selecting an ordered sequence of degradation types and corresponding models under a scoring function. JarvisIR [15] extends this idea to autonomous driving, where restoration is evaluated through downstream perception utility. On its CleanBench-Real evaluation, the paper reports a 50% average improvement across the perception metrics relative to using the degraded input.

4KAgent [13] combines a profiling module, a VLM-and-IQA perception agent, and a restoration agent that uses quality-driven mixture-of-experts selection in a recursive execution–reflection process. Domain-specific systems adapt the observation interface to acquisition physics or scientific semantics: AgentMRI [16] targets accelerated MRI, EpiAgent [21] cultural-inscription restoration, and SIMBA [20] natural-language control of single-molecule imaging workflows. The Research Square preprint describing FMIRAgent [46] reports an MLLM-based microscopy assistant that identifies degradation, selects or adapts restoration algorithms, explains its choices, and refines decisions using user or simulated feedback. RIR-Agent [47] applies MLLM perception, LLM planning, and a predefined specialist toolbox for composite remote-sensing degradation. Its executor supports adaptation and rollback, and a separate manual mode provides user-customised workflows with real-time MLLM feedback. Their inputs, feedback signals, and success criteria differ, so quantitative results are reported within the setting defined by each source.

5.2. Planning and Tool Selection

RestoreAgent expresses tool selection as

$$\sigma^* = \arg \max_{\sigma \in \mathcal{S}(\mathcal{D}, \mathcal{M})} S(I, \sigma), \quad (2)$$

where $\mathcal{S}(\mathcal{D}, \mathcal{M})$ is the set of valid degradation–model sequences and S is the chosen scoring function [11]. Subsequent systems modify the planning space or the coupling between planning and execution.

- Q-Agent [48] combines robust multimodal degradation perception with quality-driven greedy restoration.
- DiTTo [49] models degradation order and uses a simulator to reduce the cost of constructing order-aware action trajectories.
- OPERA [50] jointly optimises the restoration plan and the specialised executors, addressing the mismatch between an independently trained planner and tool pool.
- FAPE-IR [51] combines a frozen MLLM planner with a frequency-aware LoRA mixture-of-experts diffusion executor.

- Restore-R1 [52] is an adjacent RL boundary case. Its lightweight actor–critic policy is trained with DeQA-Score improvement using PPO-style optimisation, with GRPO evaluated as an alternative, as detailed in Section 3.

Creative enhancement has related but distinct objectives. PhotoArtAgent [53] uses language-model artist agents for iterative photo retouching. PhotoAgent [54] plans long-horizon aesthetic edits by tree search with memory and visual feedback and introduces UGC-Edit. RetouchIQ [55] controls Lightroom for instruction-based retouching and trains MLLM agents with a generalist reward. These systems optimise instruction adherence and aesthetic preference, requiring evaluation separate from full-reference restoration fidelity.

5.3. Verification, Iteration, and Boundary Cases

P3 feedback ranges from execution-history-based retry in RestoreAgent to downstream perception feedback in JarvisIR and task-specific review in EpiAgent. The methods use different assessors, stopping rules, and revision procedures. An iterative architecture is also distinct from an MLLM tool controller. RAR [14] is a fully trainable restore–assess–repeat framework operating in latent space and is included as a learned boundary case.

Supplementary Table S2 compares the general restoration and editing controllers, including their controller, execution, and feedback mechanisms.

Supplementary Table S3 provides the corresponding source-traceable comparison for application-specific restoration controllers.

The main synthesis is architectural. Explicit diagnosis can make tool choice interpretable. Task-coupled feedback can align restoration with downstream utility. Joint planner–executor training can improve coordination. The literature uses different tools, test mixtures, stopping rules, and quality metrics. Quantitative conclusions are therefore reported within each source-defined evaluation setting.

6. Memory-Augmented and Search-Guided Evolving Agents

P4 systems retain information beyond the current action and use it to guide later restoration decisions. The retained state may be an execution trace, a reusable experience document, a search tree, or a causal memory. The following discussion separates these mechanisms because the sources implement different search and memory procedures.

6.1. Experience and Episodic Memory

AgenticIR [12] divides its workflow into perception, scheduling, execution, reflection, and rescheduling. It constructs referenceable documentation through self-exploration and can retrieve that experience for later inputs. The tool pool covers eight degradation types, with three to six specialised models per type. The paper describes rollback and rescheduling as depth-first tree search, a procedure distinct from Monte Carlo Tree Search and UCT-style rollouts.

SEAR [27] combines an Intuitive Executor with a Deliberate Planner. It uses episodic state fingerprints for memory retrieval and P-MCTS for planning. In the paper’s Group B ablation, the full system averages 8.15 tool calls compared with 16.75 without memory.

6.2. Search-Guided Decision Making

Search is useful when restoration order is non-commutative and a locally attractive operation can harm later recovery. The sources implement three distinct forms.

- AgenticIR performs rollback-guided depth-first exploration and records successful experience [12].
- SEAR uses Pruning-Aware Monte Carlo Tree Search (P-MCTS) in its deliberate pathway and can accept retrieved shortcuts in its intuitive pathway [27].
- PhotoAgent applies tree search to long-horizon aesthetic editing, a different objective from degradation-removal tool routing [54].

Each algorithm is described using the search formulation reported by its source.

6.3. Causal Memory

Causal-AgentIR [28] uses multiple roles to update and reason over a causal memory. Its documented memory operations include add, update, merge, reinforce, ignore, and discard. This structure is intended to retain relations among degradations, restoration actions, and outcomes in addition to retrieving similar past trajectories. The paper reports a 35.55-dB average PSNR for its own all-in-one evaluation protocol. This number uses a different setting from the MiO100-derived group protocols used by AgenticIR and SEAR.

Supplementary Table S4 compares the verified search and memory mechanisms while retaining each source-defined quantitative protocol.

The evidence supports two conclusions. First, reusable experience reduces redundant planning in the SEAR ablation setting. Second, causal memory broadens the representation of prior experience beyond nearest-neighbour retrieval. Common data, tools, budgets, and evaluation code are still needed to measure cold-start behaviour and cross-system quality.

7. Multi-Agent and Heterogeneous Collaborative Systems

P5 distributes control across agents with distinct roles. The sources describe scheduler and specialist designs, fast and slow routing, and domain-specific coordination.

7.1. Scheduler and Specialist Agents

MAIR [25] contains a scheduler and seven expert agents. They are DeJPEG, DeBlur, DeNoise, SuperRes, DeRain, DeHaze, and LightEnh. The scheduler diagnoses the input and activates relevant experts in sequence. Within an expert, candidate restoration models can be tried, reflected on, and compared pairwise when necessary. Thus MAIR has one scheduler and seven listed experts. Candidate evaluation occurs within the expert workflow.

The paper reports 35.42 seconds and 1.82 model invocations for MAIR, compared with 63.04 seconds and 5.15 invocations for AgenticIR in its efficiency comparison. These figures are tied to the authors' environment and test protocol. Its paired real-world evaluation reports 21.67 dB PSNR.

7.2. Fast–Slow Collaboration

Hybrid Agents [26] uses three roles, FastAgent, SlowAgent, and FeedbackAgent. FastAgent is a lightweight language model (Llama-3.2-1B-Instruct) used through in-context learning. FastAgent is distinct from a fixed look-up table. FastAgent handles direct restoration instructions, whereas SlowAgent reasons about vague requests and FeedbackAgent evaluates the result. For direct prompts, the reported fast route uses approximately 12% of SlowAgent's runtime. Table 1 of the source reports task-specific average agent inference times of 0.08–0.13 seconds for this route and 0.75–1.05 seconds when it is disabled. A reported 72.9% FastAgent success rate concerns the denoising condition.

7.3. Domain-Specific Coordination

MAPGR [22] applies multi-agent prompt guidance with residual diffusion to mural restoration. Its prompts, damage representation, and evaluation are specific to cultural-heritage restoration.

CLEAR [56] is an adjacent dual-process remote-sensing system that decides whether restoration is needed before maritime object detection. The verified paper reports 86.92% mAP50, an improvement of 7.74 percentage points over its stated baseline, and an overall F1 improvement of 9.15 percentage points. Its 11.8 FPS practical estimate assumes a 5% trigger rate. These performance and efficiency quantities come from distinct source-defined settings.

Supplementary Table S5 compares the roles, coordination mechanisms, and source-reported examples of representative multi-agent systems.

Evaluation of multi-agent systems requires reporting the number and size of controllers, expert-model executions, communication rounds, and the downstream metric. In particular, MAIR's

scheduler–expert design, Hybrid Agents’ request-routing design, and CLEAR’s conditional preprocessing solve different control problems and require separate quality and latency interpretations.

8. Benchmarks, Evaluation Protocols, and System Metrics

AIR requires evaluation of both the restored image and the process that produced it. Output assessment uses fidelity, perceptual, scientific, or downstream-task metrics. Process assessment measures tool calls, planning steps, latency, resource use, and failure recovery. Current mixed-degradation protocols differ in test images, corruption order, severity, and tool pools. Results must therefore remain tied to their source protocols.

8.1. Evaluation Suites and Protocol Families

ImagingBench covers 20 computational-imaging tasks in Expert, Planner, and Forward settings [57]. Imaging-101 covers 57 expert-verified tasks and evaluates planning, unit-test validity, and end-to-end reconstruction [58]. These suites complement output-only benchmarks by testing whether a plan is executable and physically meaningful. They are evaluation resources rather than P1–P5 controllers.

AIR studies also use natural, medical, remote-sensing, and specialised datasets [59]. Supplementary Table S6 maps principal datasets to their task and protocol families. Supplementary Table S7 reports checked numerical examples without combining incompatible settings.

8.2. Multi-dimensional Evaluation

Full-reference fidelity commonly uses PSNR and SSIM [60]. Perceptual assessment can use LPIPS or DISTS [61,62]. Real-world no-reference metrics include BRISQUE, MUSIQ, MANIQA, and CLIP-IQA [63–66]. Reports should state metric direction, implementation, aggregation, and reference availability.

Restoration may instead be judged by downstream detection, segmentation, recognition, or scientific measurement. Such comparisons require a fixed downstream model, test split, and degraded-input baseline. Visual improvement alone cannot establish task benefit.

Agent evaluation should report model and tool invocations, planning and retry counts, latency, peak memory, token or API cost, failures, and the fraction of worsened inputs. Code-generating agents also require compilation, runtime, and physical-validity checks. Repetitions and uncertainty are essential when controller decisions are stochastic.

Supplementary Tables S8 and S9 consolidate the minimum metric families and disclosure fields. Cross-system comparison is credible only under a shared harness with fixed inputs, tools, feedback, stopping rules, and budgets.

9. Open Challenges and Strategic Future Roadmaps

Agentic image restoration remains constrained by distribution shift, acquisition physics, reasoning cost, and the risk that a generative prior introduces unsupported detail. These limitations define four closely related research directions.

Many tool-using systems expose restoration models as external functions. OPERA addresses the coordination gap between a planner and independently trained executors [50]. A physics-aware controller could additionally test candidate outputs against a known image-formation model [67]. Forward-model residuals and uncertainty estimates would make physical inconsistency measurable. Clinical or scientific validity still requires domain-specific evaluation.

Acquisition and restoration can also be treated as a joint control problem. JarvisIR links restoration to downstream driving perception [15], while MRI policies select measurements under reconstruction feedback [34]. Future systems could choose exposure, gain, sampling, or re-acquisition actions under a fixed dose, time, energy, or motion budget. Evaluation would require a matched passive acquisition baseline under the same budget.

Memory-based systems offer a basis for controlled stress testing. Causal-AgentIR updates a structured memory through add, update, merge, reinforce, ignore, and discard operations [28]. Adversarial or curriculum-generated degradations could extend this process by searching for controller failures. The resulting synthetic corruptions would need validation against real sensor statistics and evaluation on held-out real degradations.

Auditable restoration requires a record of the source asset, controller, tool versions, parameters, ordered actions, intermediate outputs, and uncertainty. C2PA Technical Specification 2.4 provides a current mechanism for signed content credentials and provenance assertions [68]. Such provenance establishes lineage and reproducibility, while domain experts remain responsible for judging whether the restored content is scientifically or clinically valid.

10. Conclusions

This review presents agentic image restoration as a transition from fixed restoration mappings to closed-loop systems that perceive degradation, coordinate restoration actions, evaluate intermediate outcomes, and refine subsequent decisions. By integrating four functional components with five controller families, the proposed framework clarifies how reinforcement learning, prompt conditioning, multimodal reasoning, memory, and multi-agent collaboration contribute to adaptive restoration. Together, this synthesis offers a common conceptual and evaluative foundation for developing more adaptive, transparent, and reliable restoration systems across imaging domains.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on [Preprints.org](https://www.preprints.org).

Author Contributions: X.D. is the designated Lead Author for this article. He provided conceptual leadership by defining the scope and guiding questions of the review, developing the controller-centred analytical framework, and integrating evidence across controller families and application domains. He also contributed to methodology, literature investigation, evidence curation, formal analysis, visualization and original-draft writing, and takes responsibility for the review's overall intellectual coherence and final content. Y.Y. contributed to conceptualization, methodology, literature investigation, evidence curation, formal analysis, visualization and original-draft writing. He contributed to the synthesis and revision of the manuscript under the research supervision of X.D. and Z.J. C.L. contributed to conceptualization, methodology, literature investigation, evidence curation, formal analysis, visualization and original-draft writing. He contributed to the synthesis, review and editing of the manuscript under the research supervision of X.D. and Z.J. Z.C. contributed to conceptualization, methodology, literature investigation, evidence curation, formal analysis, visualization and original-draft writing. He contributed to the synthesis, review and editing of the manuscript under the research supervision of X.D. and Z.J. Z.J. provided senior supervision and critically reviewed and edited the manuscript for intellectual content. Y.Y., C.L., Z.C. and X.D. jointly carried out the evidence curation and formal analysis underlying the review's synthesis. Y.Y., C.L. and Z.C. contributed equally to this work. All authors reviewed and approved the final manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (Grant No. 62306003).

Data Availability Statement: No new data were collected or generated for this narrative review. Evidence matrices and source-traceable tables are provided in the Supplementary Information. Supporting materials are also available in the [GitHub repository](#).

Conflicts of Interest: X.D. is a Guest Editor of the related Scientific Reports Collection and has not been involved in editorial handling or peer review of this manuscript. Y.Y., C.L., Z.C. and Z.J. declare no competing interests.

Acknowledgments: As Lead Author, X.D. defined the review's scope and guiding questions, established the controller-centred framework through which the literature was analysed, and integrated evidence spanning reinforcement learning, prompt-conditioned, multimodal-language-model, memory-augmented and multi-agent controllers into a coherent critical synthesis. He is accountable for the intellectual integrity of the review and for ensuring that the contributions and limitations are represented accurately. No new code was developed. Links to publicly available code for the reviewed studies are provided in the [GitHub repository](#). Generative AI was used for language polishing.

References

1. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena* **60**(1–4), 259–268 (1992) [https://doi.org/10.1016/0167-2789\(92\)90242-F](https://doi.org/10.1016/0167-2789(92)90242-F)
2. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Transactions on Image Processing* **16**(8), 2080–2095 (2007) <https://doi.org/10.1109/TIP.2007.901238>
3. Dong, C., Loy, C.C., He, K., Tang, X.: Learning a deep convolutional network for image super-resolution. In: *European Conference on Computer Vision*, pp. 184–199 (2014). https://doi.org/10.1007/978-3-319-10593-2_13
4. Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L.: Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE Transactions on Image Processing* **26**(7), 3142–3155 (2017) <https://doi.org/10.1109/TIP.2017.2662206>
5. Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., Timofte, R.: SwinIR: Image restoration using Swin Transformer. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pp. 1833–1844 (2021). <https://doi.org/10.1109/ICCVW54120.2021.00210>
6. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.-H.: Restormer: Efficient transformer for high-resolution image restoration. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5728–5739 (2022). <https://doi.org/10.1109/CVPR52688.2022.00564>
7. Chen, L., Chu, X., Zhang, X., Sun, J.: Simple baselines for image restoration. In: *European Conference on Computer Vision*, pp. 17–33 (2022). https://doi.org/10.1007/978-3-031-20071-7_2
8. Potlapalli, V., Zamir, S.W., Khan, S., Khan, F.S.: PromptIR: Prompting for all-in-one image restoration. In: *Advances in Neural Information Processing Systems*, vol. 36, pp. 71275–71293 (2023). <https://doi.org/10.52202/075280-3121> . https://proceedings.neurips.cc/paper_files/paper/2023/hash/e187897ed7780a579a0d76fd4a35d107-Abstract-Conference.html
9. Conde, M.V., Geigle, G., Timofte, R.: InstructIR: High-quality image restoration following human instructions. In: *European Conference on Computer Vision*, pp. 1–21 (2024). https://doi.org/10.1007/978-3-031-72764-1_1
10. Wei, Y., Zhang, Z., Ren, J., Xu, X., Hong, R., Yang, Y., Yan, S., Wang, M.: Clarity ChatGPT: An interactive and adaptive processing system for image restoration and enhancement. *arXiv preprint arXiv:2311.11695* (2023) <https://doi.org/10.48550/arXiv.2311.11695> [arXiv:2311.11695](https://arxiv.org/abs/2311.11695) [cs.CV]
11. Chen, H., Li, W., Gu, J., Ren, J., Chen, S., Ye, T., Pei, R., Zhou, K., Song, F., Zhu, L.: RestoreAgent: Autonomous image restoration agent via multimodal large language models. In: *Advances in Neural Information Processing Systems*, vol. 37, pp. 110643–110666 (2024). <https://doi.org/10.52202/079017-3512>
12. Zhu, K., Gu, J., You, Z., Qiao, Y., Dong, C.: An intelligent agentic system for complex image restoration problems. In: *International Conference on Learning Representations* (2025). https://proceedings.iclr.cc/paper_files/paper/2025/hash/921ac785fa9edc73cacaf2664f43d234-Abstract-Conference.html
13. Zuo, Y., Zheng, Q., Wu, M., Jiang, X., Li, R., Wang, J., Zhang, Y., Mai, G., Wang, L.V., Zou, J.Y., Wang, X., Yang, M.-H., Tu, Z.: 4KAgent: Agentic any image to 4K super-resolution. In: *Advances in Neural Information Processing Systems*, vol. 38, pp. 182016–182085 (2025). <https://doi.org/10.52202/085713-5476> . https://proceedings.neurips.cc/paper_files/paper/2025/hash/f0075fe4e59652cf43148dcfab8d3c93-Abstract-Conference.html
14. Chen, I.-H., Hadji, I., Sánchez, E., Bulat, A., Kuo, S.-Y., Timofte, R., Tzimiropoulos, G., Martinez, B.: Restore, assess, repeat: A unified framework for iterative image restoration. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15506–15515 (2026). https://openaccess.thecvf.com/content/CVPR2026/html/Chen_Restore_Assess_Repeat_A_Unified_Framework_for_Iterative_Image_Restoration_CVPR_2026_paper.html
15. Lin, Y., Lin, Z., Chen, H., Pan, P., Li, C., Chen, S., Wen, K., Jin, Y., Li, W., Ding, X.: JarvisIR: Elevating autonomous driving perception with intelligent image restoration. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 22369–22380 (2025). https://openaccess.thecvf.com/content/CVPR2025/html/Lin_JarvisIR_Elevating_Autonomous_Driving_Perception_with_Intelligent_Image_Restoration_CVPR_2025_paper.html
16. Saju, G.A., Akhi, M., Chang, Y.: Agentmri: A vision language model-powered ai system for self-regulating mri reconstruction with multiple degradations. *Journal of Imaging Informatics in Medicine* **39**(2), 1422–1440 (2026) <https://doi.org/10.1007/s10278-025-01617-0> . Published online in 2025

17. Xie, X., Zhao, W., Nan, M., *et al.*: Prompt-agent-driven integration of foundation model priors for low-count pet reconstruction. *IEEE Transactions on Medical Imaging* **44**(10), 4073–4086 (2025) <https://doi.org/10.1109/TMI.2025.3527155>
18. Maier, A., Kachelrieß, L., Bayer, S., Huang, Y., Xia, Y., Simpson, A., Zaiss, M.: Agentic Autoresearch for CT Reconstruction (2026). <https://doi.org/10.48550/arXiv.2607.22824> . <https://arxiv.org/abs/2607.22824>
19. Yang, Y., Bi, L., Yang, B., Wang, Y., He, Y., Peng, Y., Jin, Z., Dong, X., Kim, J.: Representation paradigms in AI-based 3d radiological image reconstruction: A literature review. *Research Square* (2026) <https://doi.org/10.21203/rs.3.rs-10102575/v1>
20. Mao, H., Mauny, H., Venkataraman, O.K., Laplante, C., Xu, D., Zhang, Y.: SIMBA: An agentic AI platform for single-molecule multi-dimensional imaging. *bioRxiv* (2026) <https://doi.org/10.64898/2026.04.16.719005> . Preprint
21. Zhu, S., Chen, A., Nie, N., Fang, P., Zhang, M.-L., Xue, H.: EpiAgent: An agent-centric system for ancient inscription restoration. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 39304–39313 (2026). https://openaccess.thecvf.com/content/CVPR2026/html/Zhu_EpiAgent_An_Agent-Centric_System_for_Ancient_Inscription_Restoration_CVPR_2026_paper.html
22. Ren, N., Huang, K.: MAPGR: Multi-agent prompt-guided residual diffusion for ancient mural restoration. *npj Heritage Science* **14**, 319 (2026) <https://doi.org/10.1038/s40494-026-02607-3>
23. Luo, Z., Gustafsson, F.K., Zhao, Z., Sjölund, J., Schön, T.B.: Controlling vision-language models for multi-task image restoration. In: *International Conference on Learning Representations* (2024). <https://openreview.net/forum?id=t3vnnLeajU>
24. Yu, K., Dong, C., Lin, L., Loy, C.C.: Crafting a toolchain for image restoration by deep reinforcement learning. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2443–2452 (2018). https://openaccess.thecvf.com/content_cvpr_2018/html/Yu_Crafting_a_Toolchain_CVPR_2018_paper.html
25. Jiang, X., Li, G., Chen, B., Zhang, J.: Multi-agent image restoration. *International Journal of Computer Vision* **134**(5), 205 (2026) <https://doi.org/10.1007/s11263-026-02792-5>
26. Li, B., Li, X., Lu, Y., Chen, Z.: Hybrid agents for image restoration. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 22636–22647 (2026). https://openaccess.thecvf.com/content/CVPR2026/html/Li_Hybrid_Agents_for_Image_Restoration_CVPR_2026_paper.html
27. Cui, S., Ji, F., Sun, G., Guo, Y., Tang, X., Li, J., Xu, F.: Self-evolving agentic image restoration via deliberate planning and intuitive execution. *arXiv preprint arXiv:2606.28971* (2026) <https://doi.org/10.48550/arXiv.2606.28971> [cs.CV]
28. Gao, H., Chen, Y., Ma, L.: Causal-AgentIR: Self-evolving causal memory for adaptive image restoration agents. *arXiv preprint arXiv:2607.21125* (2026) <https://doi.org/10.48550/arXiv.2607.21125> [cs.CV]
29. Furuta, R., Inoue, N., Yamasaki, T.: Fully convolutional network with multi-step reinforcement learning for image processing. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 3598–3605 (2019). <https://doi.org/10.1609/aaai.v33i01.33013598> . <https://ojs.aaai.org/index.php/AAAI/article/view/4240>
30. Yu, K., Wang, X., Dong, C., Tang, X., Loy, C.C.: Path-restore: Learning network path selection for image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **44**(10), 7078–7092 (2022) <https://doi.org/10.1109/TPAMI.2021.3096255> . Published online in 2021
31. Park, J., Lee, J.-Y., Yoo, D., Kweon, I.S.: Distort-and-recover: Color enhancement using deep reinforcement learning. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5928–5936 (2018). https://openaccess.thecvf.com/content_cvpr_2018/html/Park_Distort-and-Recover_Color_Enhancement_CVPR_2018_paper.html
32. Shen, C., Gonzalez, Y., Chen, L., Jiang, S.B., Jia, X.: Intelligent parameter tuning in optimization-based iterative ct reconstruction via deep reinforcement learning. *IEEE Transactions on Medical Imaging* **37**(6), 1430–1439 (2018) <https://doi.org/10.1109/TMI.2018.2823679>
33. Patwari, M., Gutjahr, R., Raupach, R., Maier, A.: Low dose ct denoising via joint bilateral filtering and intelligent parameter optimization. In: *Proceedings of the 6th International Conference on Image Formation in X-Ray Computed Tomography*, pp. 174–177 (2020). <https://doi.org/10.48550/arXiv.2007.04768> . <https://arxiv.org/abs/2007.04768>

34. Liu, Y., Pang, Y., Sun, X., Hou, Y., Yang, Z., Wang, Z.: Dual states based reinforcement learning for fast mr scan and image reconstruction. *Neurocomputing* **568**, 127067 (2024) <https://doi.org/10.1016/j.neucom.2023.127067>
35. Wang, Y., Yan, Q., Zhou, J., Dai, D., Dong, W.: Paagent: Portrait-aware image restoration agent via subjective-objective reinforcement learning. *arXiv preprint arXiv:2603.17055* (2026) <https://doi.org/10.48550/arXiv.2603.17055> [cs.CV]
36. Li, B., Liu, X., Hu, P., Wu, Z., Lv, J., Peng, X.: All-in-one image restoration for unknown corruption. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17452–17462 (2022). https://openaccess.thecvf.com/content/CVPR2022/html/Li_All-in-One_Image_Restoration_for_Unknown_Corruption_CVPR_2022_paper.html
37. Ai, Y., Huang, H., Zhou, X., Wang, J., He, R.: 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*, pp. 25432–25444 (2024). https://openaccess.thecvf.com/content/CVPR2024/html/Ai_Multimodal_Prompt_Perceiver_Empower_Adaptiveness_Generalizability_and_Fidelity_for_All-in-One_CVPR_2024_paper.html
38. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., Sutskever, I.: Learning transferable visual models from natural language supervision. In: *Proceedings of the International Conference on Machine Learning. Proceedings of Machine Learning Research*, vol. 139, pp. 8748–8763. PMLR, Virtual Event (2021). <https://proceedings.mlr.press/v139/radford21a.html>
39. Jiang, Y., Zhang, Z., Xue, T., Gu, J.: AutoDIR: Automatic all-in-one image restoration with latent diffusion. In: *European Conference on Computer Vision*, pp. 340–359 (2024). https://doi.org/10.1007/978-3-031-73661-2_19
40. Guo, Y., Gao, Y., Lu, Y., Zhu, H., Liu, R.W., He, S.: OneRestore: A universal restoration framework for composite degradation. In: *European Conference on Computer Vision*, pp. 255–272 (2024). https://doi.org/10.1007/978-3-031-72655-2_15
41. Yang, H., Pan, L., Yang, Y., Liang, W.: Language-driven all-in-one adverse weather removal. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 24902–24912 (2024). https://openaccess.thecvf.com/content/CVPR2024/html/Yang_Language-driven_All-in-one_Adverse_Weather_Removal_CVPR_2024_paper.html
42. Qi, C., Tu, Z., Ye, K., Delbracio, M., Milanfar, P., Chen, Q., Talebi, H.: SPIRE: Semantic prompt-driven image restoration. In: *European Conference on Computer Vision*, pp. 446–464 (2024). https://doi.org/10.1007/978-3-031-73661-2_25
43. Zhou, M., Ye, K., Delbracio, M., Milanfar, P., Patel, V.M., Talebi, H.: UniRes: Universal image restoration for complex degradations. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 13237–13247 (2025). https://openaccess.thecvf.com/content/ICCV2025/html/Zhou_UniRes_Universal_Image_Restoration_for_Complex_Degradations_ICCV_2025_paper.html
44. Yu, F., Gu, J., Li, Z., Hu, J., Kong, X., Wang, X., He, J., Qiao, Y., Dong, C.: 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*, pp. 25669–25680 (2024). https://openaccess.thecvf.com/content/CVPR2024/html/Yu_Scaling_Up_to_Excellence_Practicing_Model_Scaling_for_Photo-Realistic_Image_CVPR_2024_paper.html
45. Ai, Y., Zhou, X., Huang, H., Han, X., Chen, Z., You, Q., Yang, H.: Dreamclear: High-capacity real-world image restoration with privacy-safe dataset curation. In: *Advances in Neural Information Processing Systems*, vol. 37, pp. 55443–55469 (2024). <https://doi.org/10.52202/079017-1761> . https://proceedings.neurips.cc/paper_files/paper/2024/hash/6452474601429509f3035dc81c233226-Abstract-Conference.html
46. He, R., Tan, W., Ma, C., Zhang, Z., Yan, B.: Self-explained thinking agent for autonomous microscopy restoration. *Research Square* (2025) <https://doi.org/10.21203/rs.3.rs-7116422/v1> . Preprint, version 1; not peer reviewed
47. Liu, J., Li, T., Liang, L., Fu, G., Wang, G., Rui, Q., Tang, X., Zhu, S., Yang, Y.: RIR-Agent: An interactive framework for effective and adaptive restoration of remote sensing imagery. *Expert Systems with Applications* **324**, 132495 (2026) <https://doi.org/10.1016/j.eswa.2026.132495>
48. Zhou, Y., Cao, J., Wen, F., Zhang, Z., Zhou, Y., Shi, Y., Liu, X., Timofte, R., Van Gool, L., Zhai, G.: Q-Agent: Quality-driven chain-of-thought image restoration agent through robust multimodal large language

- model. arXiv preprint arXiv:2504.07148 (2025) <https://doi.org/10.48550/arXiv.2504.07148> arXiv:2504.07148 [eess.IV]
49. Choi, S., Oh, J.: DiTTo: Scalable order-aware all-in-one image restoration agent. arXiv preprint arXiv:2605.30915 (2026) <https://doi.org/10.48550/arXiv.2605.30915> arXiv:2605.30915 [cs.CV]
 50. Zhu, F., Xie, S., Zeng, Y., Liu, M., Zuo, W.: OPERA: An agent for image restoration with end-to-end joint planning–execution optimization. arXiv preprint arXiv:2605.22104 (2026) <https://doi.org/10.48550/arXiv.2605.22104> arXiv:2605.22104 [cs.CV]
 51. Liu, J., Xu, S., Yang, Q., Wang, Y., Chen, X., Ji, Z.: FAPE-IR: Frequency-aware planning and execution framework for all-in-one image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 15561–15573 (2026). https://openaccess.thecvf.com/content/CVPR2026/html/Liu_FAPE-IR_Frequency-Aware_Planning_and_Execution_Framework_for_All-in-One_Image_Restoration_CVPR_2026_paper.html
 52. Lu, J., Wu, Y., Zhao, Z., Wang, H., Jimenez, F., Majeedi, A., Fu, Y.: Restore-R1: Efficient image restoration agents via reinforcement learning with multimodal LLM perceptual feedback. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Findings, pp. 8629–8639 (2026). https://openaccess.thecvf.com/content/CVPR2026F/html/Lu_Restore-R1_Efficient_Image_Restoration_Agents_via_Reinforcement_Learning_with_Multimodal_CVPRF_2026_paper.html
 53. Chen, H., Tao, K., Wang, Y., Wang, X., Zhu, L., Gu, J.: PhotoArtAgent: Intelligent photo retouching with language model-based artist agents. arXiv preprint arXiv:2505.23130 (2025) <https://doi.org/10.48550/arXiv.2505.23130> arXiv:2505.23130 [cs.CV]
 54. Yao, M., You, Z., Tam, K.-M., Wang, M., Xue, T.: PhotoAgent: Exploratory visual aesthetic planning with large vision models. arXiv preprint arXiv:2602.22809 (2026) <https://doi.org/10.48550/arXiv.2602.22809> arXiv:2602.22809 [cs.CV]. ICML 2026 oral presentation; arXiv version 3
 55. Wu, Q., Shi, J., Jenni, S., Kafle, K., Wang, T., Chang, S., Zhao, H.: RetouchIQ: MLLM agents for instruction-based image retouching with generalist reward. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 12279–12288 (2026). https://openaccess.thecvf.com/content/CVPR2026/html/Wu_RetouchIQ_MLLM_Agents_for_Instruction-Based_Image_Retouching_with_Generalist_Reward_CVPR_2026_paper.html
 56. Li, M., Zhao, X., Wan, Y.: CLEAR: A cognitive LLM-empowered adaptive restoration framework for robust ship detection in complex maritime scenarios. Remote Sensing 18(8), 1142 (2026) <https://doi.org/10.3390/rs18081142>
 57. Chung, E., Zheng, C., Tan, J., Li, J., Zhang, H., Chen, H.: Does AI Understand Imaging? A Systematic Benchmark of Agentic AI for Computational Imaging Tasks (2026). <https://doi.org/10.48550/arXiv.2607.07189> . <https://arxiv.org/abs/2607.07189>
 58. Chen, S., Ying, J., Jia, Y., Gu, Y., Ye, E., Bai, W., Zeng, Z., Ren, S., Gao, B., Li, Y., Zhang, T., Sun, H.: Imaging-101: Benchmarking LLM Coding Agents on Scientific Computational Imaging (2026). <https://doi.org/10.48550/arXiv.2607.10789> . <https://arxiv.org/abs/2607.10789>
 59. Yang, Y., Chen, Y., Dong, X., Zhang, J., Long, C., Jin, Z., Dai, Y.: An annotated heterogeneous ultrasound database. Scientific Data 12(1), 148 (2025) <https://doi.org/10.1038/s41597-025-04464-4>
 60. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing 13(4), 600–612 (2004) <https://doi.org/10.1109/TIP.2003.819861>
 61. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 586–595 (2018). <https://doi.org/10.1109/CVPR.2018.00068>
 62. Ding, K., Ma, K., Wang, S., Simoncelli, E.P.: Image quality assessment: Unifying structure and texture similarity. IEEE Transactions on Pattern Analysis and Machine Intelligence 44(5), 2567–2581 (2022) <https://doi.org/10.1109/TPAMI.2020.3045810> . Published online in 2020
 63. Mittal, A., Moorthy, A.K., Bovik, A.C.: No-reference image quality assessment in the spatial domain. IEEE Transactions on Image Processing 21(12), 4695–4708 (2012) <https://doi.org/10.1109/TIP.2012.2214050>
 64. Ke, J., Wang, Q., Wang, Y., Milanfar, P., Yang, F.: MUSIQ: Multi-scale image quality transformer. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 5148–5157 (2021). https://openaccess.thecvf.com/content/ICCV2021/html/Ke_MUSIQ_Multi-Scale_Image_Quality_Transformer_ICCV_2021_paper.html

65. Yang, S., Wu, T., Shi, S., Lao, S., Gong, Y., Cao, M., Wang, J., Yang, Y.: MANIQA: Multi-dimension attention network for no-reference image quality assessment. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 1191–1200 (2022). https://openaccess.thecvf.com/content/CVPR2022W/NTIRE/html/Yang_MANIQA_Multi-Dimension_Attention_Network_for_No-Reference_Image_Quality_Assessment_CVPRW_2022_paper.html
66. Wang, J., Chan, K.C.K., Loy, C.C.: Exploring CLIP for assessing the look and feel of images. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 37, pp. 2555–2563 (2023). <https://doi.org/10.1609/aaai.v37i2.25353>
67. Yang, Y., Ruan, Q., Cai, W., Dong, Y., Yang, D., Dong, X., Jin, Z., Dai, Y.: UltraGS: Real-time physically-decoupled gaussian splatting for ultrasound novel view synthesis. In: IEEE International Conference on Multimedia & Expo (2026). <https://arxiv.org/abs/2511.07743>
68. Coalition for Content Provenance and Authenticity: C2PA Technical Specification, Version 2.4. Accessed 14 August 2026 (2026). https://spec.c2pa.org/specifications/specifications/2.4/specs/C2PA_Specification.html

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.