Review

Not peer-reviewed version

Language-Driven Image Restoration and Semantic-Aware Quality Assessment: A Survey

Mingyu Liu , Haozhan Shu , Yuning Cui * , Xingcheng Zhou , Hu Cao , Wenqi Ren , Boxin Shi , Alois Knoll

Posted Date: 8 April 2026

doi: 10.20944/preprints202603.2366.v2

Keywords: image restoration; vision-language models; multimodal large language models; image quality assessment

 doi:10.20944/preprints202603.2366.v2

Article

Language-Driven Image Restoration and Semantic-Aware Quality Assessment: A Survey

Mingyu Liu [1], Haozhan Shu [1], Yuning Cui [1,*], Xingcheng Zhou [1], Hu Cao [2], Wenqi Ren [3], Boxin Shi [4] and Alois Knoll [1]

[1] Technical University of Munich, Munich, Germany
[2] Southeast University, Nanjing, China
[3] Sun Yat-sen University, Shenzhen, China
[4] Peking University, Beijing, China
* Correspondence: yuning.cui@in.tum.de

Abstract

Image restoration aims to recover a high-quality image from its degraded counterpart by mitigating distortions introduced during acquisition, transmission, or environmental interaction. Despite the remarkable progress of deep learning–based restoration models, most conventional approaches remain tightly coupled to predefined degradation assumptions and pixel-level supervision, limiting their capability to handle complex and diverse scenarios or user-dependent restoration targets. Recent advances in multimodal large language models (MLLMs) and vision–language models (VLMs) have introduced a new paradigm in which restoration systems incorporate semantic reasoning, language-driven interaction, and cross-modal knowledge. By integrating language models, restoration is extended beyond low-level reconstruction toward degradation interpretation, perceptual alignment, and high-level controllability. In this survey, we provide a systematic review of language-driven image restoration, organized through an interaction-centric taxonomy that characterizes how language models are coupled with restoration pipelines. We analyze representative frameworks from the perspectives of semantic conditioning, perceptual supervision, and execution-level interaction, and discuss how these mechanisms influence restoration objectives and system design. In addition, we review emerging language-driven image quality assessment (IQA) approaches, highlighting their complementary role to conventional fidelity-based metrics. Finally, we identify unresolved challenges and outline potential research directions toward more robust, efficient, and trustworthy restoration techniques.
https://github.com/MingyuLiu1/Language-Driven-IR-and-IQA
CCS Concepts: • General and reference→Surveys and overviews; • Computing methodologies→Computer vision.

Keywords: image restoration; vision-language models; multimodal large language models; image quality assessment

1. Introduction

Image restoration (IR), as a fundamental problem in low-level computer vision, aims to recover high-quality images from degraded counterparts. Over the past decades, numerous IR methods have been extensively studied for a wide range of tasks, including denoising [1–3], deraining [4–6], dehazing [7–9], desnowing [10–12], deblurring [13–15], low-light enhancement (LLIE) [16–18], super-resolution [19–21]. Beyond these tasks, IR techniques have also been widely applied to domain-specific scenarios, such as underwater enhancement [22–24], medical IR [25,26].

IR methodologies have evolved from model-driven approaches based on handcrafted priors to data-driven paradigms powered by deep neural networks, including CNNs [27], Transformers [28], and more recent architectures [29]. In parallel, restoration frameworks have progressed from task-specific designs, where each degradation type is modeled independently [5,9,30], to unified frameworks

 doi:10.20944/preprints202603.2366.v2

such as all-in-one (AiO) restoration, which aim to handle multiple degradation types within a single model [31–34].

Despite these advances, existing methods remain largely constrained by predefined degradation assumptions and are typically optimized with pixel-level supervision, limiting their ability to generalize to complex or user-dependent restoration scenarios. These limitations motivate the exploration of new paradigms beyond conventional frameworks.

Recent breakthroughs in foundation models (FMs), such as vision–language models (VLMs) and multimodal large language models (MLLMs), have opened new opportunities to address these challenges. By encoding rich semantic priors and exhibiting strong reasoning capabilities beyond conventional visual representations, VLMs and MLLMs have been increasingly introduced into IR pipelines [34–37]. Rather than directly performing pixel-level reconstruction, these models are typically employed as auxiliary components to provide high-level information, such as degradation interpretation, semantic guidance, and adaptive control. This emerging paradigm fundamentally reshapes IR from a purely visual mapping problem into a multimodal, semantically informed, and interactive framework.

This paradigm shift introduces a critical challenge in image quality assessment (IQA). Traditional IQA metrics [38–40] are effective for measuring pixel-level fidelity or perceptual similarity. However, they are not designed to assess semantic alignment or instruction consistency, leading to a mismatch between the restoration targets and the evaluation criteria. Recent language-driven IQA methods [41–47] address this gap by leveraging multimodal representations to assess semantic coherence and cross-modal consistency.

Although integrating language for IR and IQA has developed rapidly [34,37,48], a comprehensive survey of this emerging field remains absent, to the best of our knowledge. Existing surveys on IR primarily focus on architectural designs [49] or task-specific learning strategies [4,31,50], while recent reviews on multimodal models [51,52] seldom address low-level vision problems in depth.

In this survey, we provide the first systematic review of language-driven image restoration and language-driven IQA. Figure 1 summarizes the taxonomy of this survey in a hierarchically structured way. We review advances in language-driven restoration frameworks, summarize representative model designs and training strategies, and discuss open challenges and potential future research directions.

The main contributions of this survey are summarized as follows:

- We introduce a unified conceptual framework that interprets language-driven IR as an interaction-centric paradigm, revealing how language models affect restoration behavior beyond architectural modifications, including feature-level, optimization-level, and execution-level coupling.
- We analyze language-driven IQA, highlighting its conceptual distinctions from conventional fidelity metrics and clarifying the challenges of evaluation reliability, calibration stability, and semantic bias.
- We summarize the restoration datasets used in language-driven frameworks and analyze their limitations and emerging requirements from a language-driven perspective, emphasizing the need for semantically enriched, language-aware benchmarks. We also provide comparisons between conventional frameworks and language-driven methods across different settings.
- We investigate open challenges posed by language-integrated restoration systems and outline promising directions for future research that bridge multimodal reasoning, visual perception, and restoration optimization.

 doi:10.20944/preprints202603.2366.v2

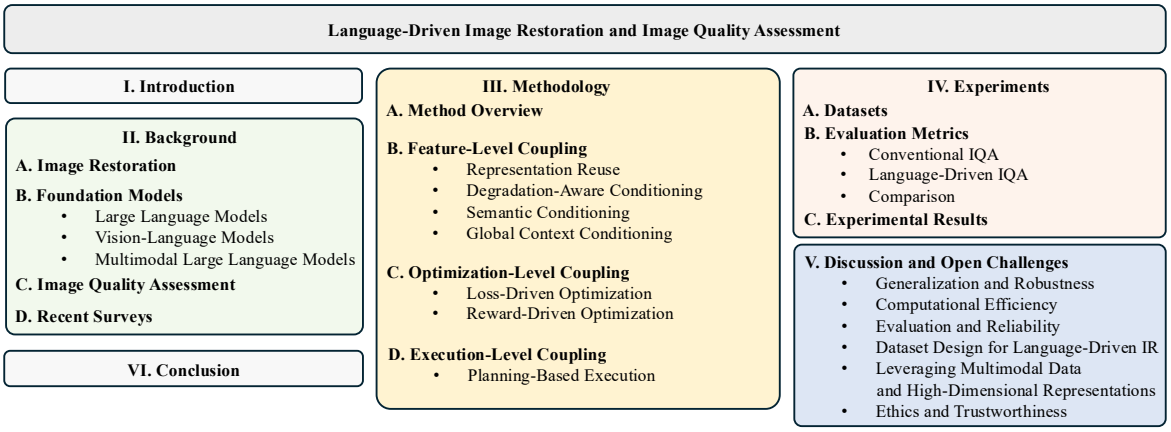

Figure 1. Unified taxonomy and survey structure of language-driven image restoration and image quality assessment.

The rest of the work is organized as follows. Section 2 introduces the necessary preliminaries, including IR fundamentals, multimodal language models, and IQA, providing an overview of the core concepts relevant to this work. Section 3 reviews representative language-driven IR approaches, and the analysis is guided by the proposed interaction-centric taxonomy with a focus on their model architectures and key technical innovations. Section 4 summarizes commonly used datasets for different IR tasks and discusses corresponding evaluation metrics, covering both traditional criteria and recent language-driven assessment methods. Finally, Section 5 analyzes open challenges in current studies and outlines potential solutions and future research directions.

2. Background

In this section, we introduce the fundamental concepts and taxonomy of language-driven image restoration, aiming to provide a structured background for understanding recent language-driven restoration frameworks.

2.1. Image Restoration

Image restoration aims to recover a clean image $\mathcal{I}(x)$ from its degraded observation $\mathcal{D}(x)$ by mitigating various types of distortions:

$$\mathcal{D}(x) = H(\mathcal{I}(x)) + \mathcal{N}, \tag{1}$$

where $H(\cdot)$ denotes the degradation operator and $\mathcal{N}$ represents additive noise.

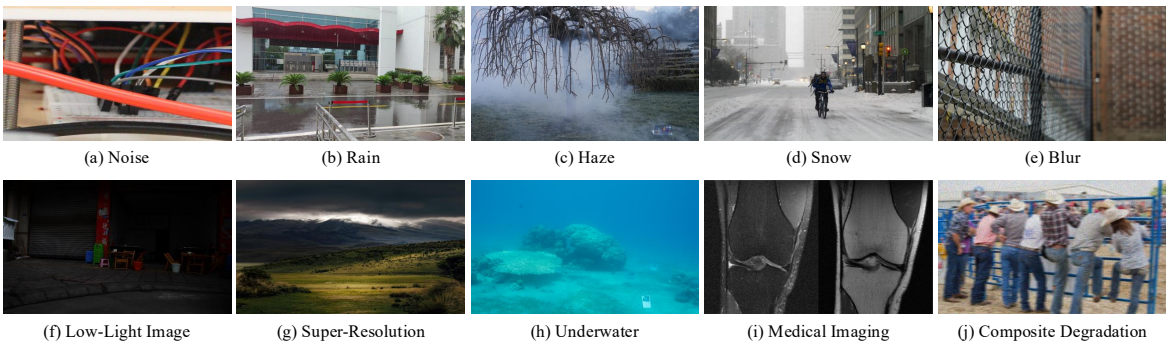

Figure 2. Examples of different degradation types and domains. The images correspond to the following open-source datasets: (a) PolyU [53], (b) LHP [54], (c) NH-HAZE [55], (d) RealSnow10K [56], (e) DPDblur [14], (f) LSRW [57], (g) DIV4K50 [58], (h) Squid [59], (i) FastMRI [60], and (j) MiO100 [61].

Existing IR methods can be broadly categorized into task-specific models and unified frameworks. Task-specific IR methods [8,17] are trained to handle a single degradation type. While such models

 doi:10.20944/preprints202603.2366.v2

often achieve strong performance within their targeted tasks, their specialization inherently restricts generalization to unseen or mixed degradations. This limitation motivates the development of general IR frameworks [62,63], which employ unified architectures capable of addressing multiple restoration tasks. Nevertheless, these approaches typically still rely on task-wise training or fine-tuning procedures. More recently, AiO restoration frameworks [32,33,64,65] have attracted increasing attention. By enabling multiple restoration tasks within a single model without explicit retraining, AiO methods aim to further improve generalization. This is often achieved through degradation-aware representations, prompt-based mechanisms, or multi-branch architectures that explicitly distinguish different degradation types.

Despite these advances, conventional deep learning–based restoration paradigms remain largely constrained by predefined task formulations and limited adaptability to diverse restoration requirements. To address these limitations, recent studies have started to explore the use of language models for IR [34,36,66–68]. Leveraging their strong semantic understanding, reasoning capability, and cross-modal alignment, language-driven approaches enable more flexible and interactive restoration pipelines, such as instruction-guided or context-aware restoration. This emerging paradigm marks a shift from task-centric restoration models toward more general, adaptive, and semantically informed IR frameworks.

2.2. Foundation Models

Recent studies on FMs encompass multiple closely related model families, including large language models (LLMs), VLMs, and MLLMs. These models differ in their input modalities and output objectives, yet collectively serve as the primary computational backbone of language-driven IR methods reviewed in this survey. While capability boundaries between VLMs and MLLMs continue to develop, we adopt a functional distinction based on their dominant usage patterns in restoration frameworks.

Large Language Models. LLMs are FMs trained on large-scale text datasets and operate exclusively on language inputs and outputs. Representative models include GPT [69], LLaMA [70], Qwen [71], and DeepSeek [72]. These models have achieved remarkable progress in natural language understanding, reasoning, and generation, demonstrating strong generalization across a wide range of language-centric tasks. LLMs primarily operate on textual inputs and outputs and do not natively process raw visual data.

Vision-Language Models. In a broad sense, VLMs refer to models that jointly process visual and textual signals for cross-modal understanding. This includes both contrastive representation learning frameworks (e.g., CLIP-like [73,74] and ALIGN-like [75]) and more recent architectures that integrate visual encoders with language modeling components. VLMs primarily enable tasks such as semantic grounding, cross-modal retrieval, and image understanding. In language-driven IR, VLMs often serve as semantic extractors or conditioning providers that bridge visual content and high-level language descriptions.

Multimodal Large Language Models. MLLMs extend LLMs by incorporating modality-specific perception modules, allowing them to process visual and other non-textual inputs (e.g., audio and video) while maintaining a language-centric interface. Representative models include GPT-4o [76], Gemini [77], and Qwen-VL [78,79]. In restoration pipelines, MLLMs are particularly suitable for tasks that require high-level reasoning, such as degradation analysis, restoration planning, and adaptive strategy selection.

In this survey, we emphasize functional roles of VLMs and MLLMs in language-driven IR systems. When referring to methods that leverage language as a high-level interaction signal, regardless of the specific model architecture, we adopt the umbrella term language-driven image restoration.

2.3. Image Quality Assessment

IQA plays a critical role in image processing by enabling the evaluation and optimization of visual content quality [80]. Although subjective assessment based on the human visual system (HVS) is

 doi:10.20944/preprints202603.2366.v2

generally regarded as the most reliable criterion, it is costly and time-consuming, motivating extensive research into objective metrics.

Depending on the availability of reference images, IQA methods are broadly categorized into full-reference (FR-IQA) and no-reference (NR-IQA). FR-IQA methods compare distorted images with their high-quality references using metrics such as PSNR [39], SSIM [38], and MSE, as well as learning-based perceptual metrics [40,81–84]. In contrast, NR-IQA predicts image quality directly from distorted inputs without reference images. Early methods rely on handcrafted natural scene statistics, such as BRISQUE [85], NIQE [86], and PIQE [87], while recent approaches adopt deep learning to model complex perceptual patterns [88–92].

Most recently, language-driven IQA methods [41,42,45–47,93–95] have emerged as a new paradigm. By leveraging multimodal reasoning capabilities and foundational knowledge, these approaches aim to narrow the gap between objective metrics and subjective human assessment. Some methods evaluate image quality using natural languages [95–97], while others directly produce quantitative quality scores [45,94,98].

2.4. Relevant Surveys

A number of surveys have reviewed IR from different perspectives. Many of these works organize the literature according to specific degradation types or application domains [1,4,7,13,16,19,22,31,50]. Along the recent trend of AiO restoration, Jiang *et al.* [31] provided a systematic overview of AiO restoration frameworks. Beyond task-oriented surveys, several studies have examined IR from a model-centric perspective. For example, Su *et al.* [49] reviewed deep learning architectures widely adopted in IR, while Li *et al.* [102] offered an in-depth review of diffusion-based IR methods. Despite these efforts, the rapid emergence of language-driven IR methodologies has not yet been systematically reviewed. In contrast, in this work, we present a comprehensive review of language-driven IR from the following three aspects: 1) recent advances in language-driven IR methods, 2) datasets and language-driven evaluation protocols, and 3) benchmarking and comparative evaluation of language-driven approaches. The comparison between previous surveys and our work is summarized in Table 1.

Table 1. Comparison with representative surveys on image restoration. Tasks: DR (deraining), DH (dehazing), DB (deblurring), DN (denoising), DS (desnowing), LLIE (low-light image enhancement), SR (super-resolution), AiO (all-in-one). Domains: Nat (natural), UW (underwater), Med (medical), HS (hyperspectral). M: multi-task; L: language-guided; A: in-depth analysis; IQA: language-driven image quality assessment.

Survey	Year	Tasks	Domains	M	L	A	IQA	Key contributions
Jiang et al. [1]	2025	DN	Nat	×	×	✓	×	Reviews deep denoising methods
Su et al. [99]	2023	DR	Nat	×	×	×	×	Reviews deraining architectures and benchmarks
Gui et al. [7]	2023	DH	Nat	×	×	×	×	Categorizes dehazing (CNN/GAN/Transformer)
Xiang et al. [13]	2025	DB	Nat	×	×	✓	×	Organizes CNN-based deblurring frameworks
Li et al. [16]	2021	LLIE	Nat	×	×	✓	×	Organizes LLIE by illumination modeling
Zhang et al. [100]	2022	SR	Nat	×	×	✓	×	Taxonomy of SR (architecture/loss/training)
Zhu et al. [101]	2026	DH/DN/Enhancement	UW	✓	✓	✓	×	Reviews underwater enhancement and restoration methods
Wang et al. [50]	2025	DR/DH/DB/DS/LLIE/SR/AiO	Nat (UHD)	✓	×	✓	×	UHD restoration across degradations
Li et al. [102]	2025	DN/DR/DH/DB/LLIE/SR/AiO	Nat	✓	×	✓	×	Diffusion-based restoration across tasks
Jiang et al. [31]	2025	AiO	Nat/UW/Med/HS	✓	×	✓	×	AiO restoration across domains
Ours	2026	DN/DR/DH/DB/DS/LLIE/SR/AiO	Nat/UW/Med/HS	✓	✓	✓	✓	Systematizes language-driven IR and IQA with unified taxonomy

3. Methodology

This section first categorizes language-driven method prototypes and then analyzes each category in detail through representative approaches. Figure 3 summarizes the representative paradigms of existing language-driven IR methods.

3.1. Overview of Language-Driven Restoration and Taxonomy Definition

Recent advances in FMs have significantly influenced the design of IR systems. Rather than solely improving restoration architectures, an increasing number of approaches leverage language-driven semantic priors, cross-modal reasoning, and high-level decision-making to enhance restoration robustness, flexibility, and controllability. As a result, clarifying the roles of language models and how

 doi:10.20944/preprints202603.2366.v2

they interact with restoration pipelines becomes critical for understanding existing frameworks and guiding future research.

- **Feature-Level Coupling:** FM outputs are injected into the forward process to modulate intermediate representations without altering the optimization objective or execution structure. This includes pretrained feature conditioning, degradation-aware conditioning, semantic conditioning, and global context conditioning.
- **Optimization-Level Coupling:** FM outputs define or reshape the optimization objective by introducing differentiable loss terms or scalar reward functions, thereby altering optimization dynamics.
- **Execution-Level Coupling:** FM outputs determine the execution logic of the restoration pipeline by generating high-level plans or control signals, enabling dynamic selection, composition, or scheduling of restoration modules beyond a fixed computational graph.

Existing surveys on IR mainly organize the literature from architecture-driven (e.g., CNN or transformer) [49] or task-driven [4,31] perspectives. From the perspective of model architectures, existing approaches involve different types of foundation models, including VLMs and MLLMs. However, these categorizations become insufficient for language-driven restoration systems, where the primary methodological distinctions arise from interaction mechanisms between language models and restoration networks. Specifically, language-driven restoration systems reshape restoration behavior by modifying information flow structures, supervision signals, control mechanisms, and optimization objectives. Therefore, to provide a structured understanding of this rapidly growing body of work, we adopt an *interaction-centric* taxonomy that categorizes methods according to the functional role of language models within restoration pipelines, while model families are treated as an orthogonal dimension. The interaction-centric taxonomy is described below:

Definition 1: Interaction Interface of Foundation Models in Restoration

We characterize the involvement of FMs in IR through an interaction interface tuple:

$$\mathcal{T} = (\mathcal{I}, \mathcal{O}, \mathcal{G}),$$

where $\mathcal{T}$ represents the interaction interface between the FM and the restoration pipeline.
(1) Input space: $\mathcal{I}$ denotes the modalities consumed by the FM, including degraded images, textual instructions, intermediate features, or auxiliary evaluation signals.
(2) Output space: $\mathcal{O}$ denotes the signals produced by the FM that influence the restoration pipeline, such as semantic embeddings, conditioning signals, differentiable loss terms, restoration plans, or scalar rewards.
(3) Coupling function: $\mathcal{G}$ specifies how $\mathcal{O}$ is integrated into the restoration process, including feature-level conditioning, optimization-level supervision, and execution-level control.

Based on this interface formulation, we distinguish different paradigms according to the dominant interaction type as follows:

In practice, certain tendencies can be observed: VLMs are often used for representation-level conditioning, whereas MLLMs are more frequently involved in higher-level decision-making, such as planning and control. Nevertheless, these associations are not strict, and the same interaction paradigm can be realized with different types of foundation models. To maintain clarity, each method is categorized according to its primary functional influence, although practical systems may exhibit hybrid characteristics across multiple interaction levels.

Figure 4 illustrates the evolution of representative language-driven IR approaches from 2023 to Jan. 2026. In the following subsections, we review each interaction category in detail, summarizing representative methods, core design strategies, as well as their respective strengths and limitations.

 doi:10.20944/preprints202603.2366.v2

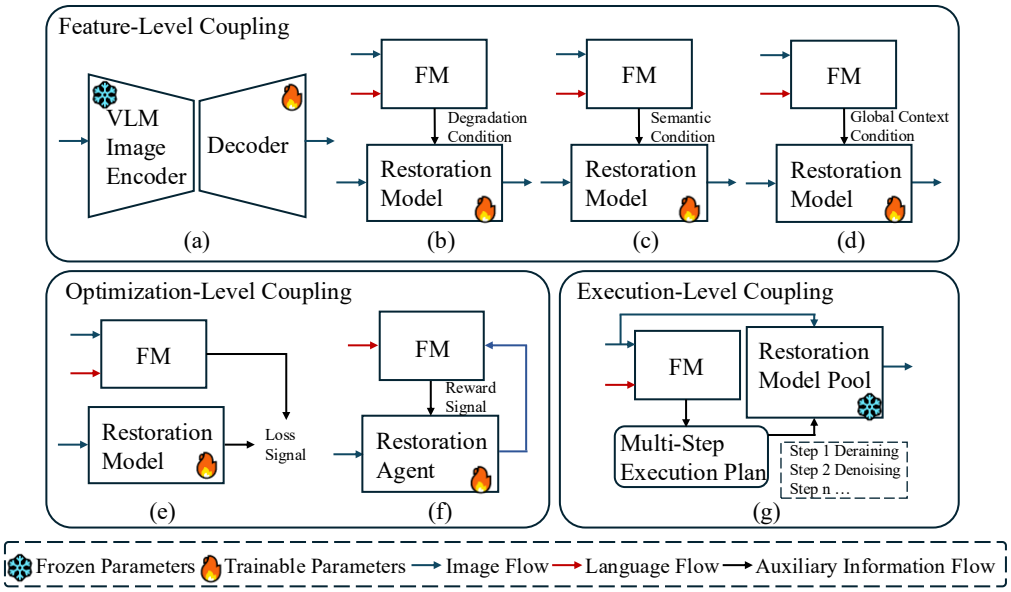

Figure 3. Categories of language-driven image restoration paradigms. We categorize existing language-driven image restoration frameworks into seven representative prototypes. Specifically, feature-level coupling paradigms include: (a) representation reuse, (b) degradation-aware conditioning, (c) semantic conditioning, and (d) global context conditioning. Optimization-level coupling paradigms are categorized into two types: (e) loss-driven optimization and (f) reward-driven optimization. Finally, execution-level coupling paradigms include (g) planning-based execution.

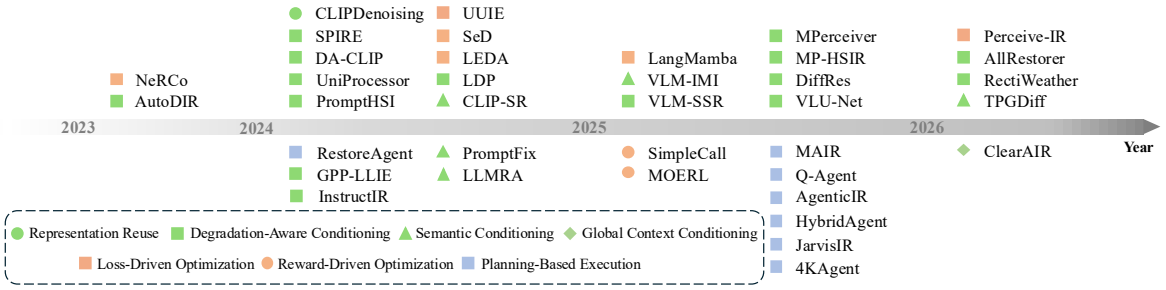

Figure 4. Timeline of representative language-driven IR approaches from 2023 to Jan. 2026. Marker colors denote different interaction paradigms defined in our taxonomy, including feature-level, optimization-level, and execution-level coupling, while marker shapes further distinguish specific interaction types within each category.

3.2. Feature-Level Coupling

Representation Reuse. Representation reuse directly integrates representations from pretrained FMs into restoration backbones, typically by replacing the visual encoding stage. In this paradigm, the pretrained model acts as a fixed feature extractor that provides robust and semantically aligned priors to guide downstream reconstruction. A representative example is CLIP-based feature reuse. During CLIP pretraining, visual representations are optimized to align with text semantics that inherently lack high-frequency details. As a result, these representations tend to overlook noise and are less affected by degradation perturbations, providing more robust guidance for restoration.

For instance, CLIPDenoising [103] incorporates a frozen CLIP [104] visual encoder together with a learnable denoising decoder. By leveraging the distortion-invariant and content-aware properties of CLIP's dense visual features, the method demonstrates promising out-of-distribution generalization under unseen conditions.

Degradation-Aware Conditioning. Degradation-aware conditioning introduces degradation-specific signals, such as type, severity, or mixture, derived from FMs into the restoration pipeline to guide feature modulation. These signals are projected as conditioning inputs to adapt the restoration process to diverse and potentially unseen degradations.

 doi:10.20944/preprints202603.2366.v2

LDP [105] utilizes text-aligned embeddings to describe blur attributes, allowing restoration models to adjust internal feature representations. GPP-LLIE [106] introduces a generative perceptual prior by extracting high-level attributes from LLaVA [44]. In AiO multi-degradation restoration, UniProcessor [107], VLU-Net [67], and AllRestorer [108] exploit textual priors to identify and handle mixed degradations. Furthermore, InstructIR [34] employs natural-language instructions to unify multiple restoration tasks into a single model. DiffRes [109] leverages BLIP-2 [110] to encode degradation information through feature difference instructions. RectiWeather [111] models degradation conditions by estimating weather-related attributes from images and uses these signals to modulate the restoration process. In hyperspectral restoration, MP-HSIR [112] integrates spectral prompts with language–visual prompts, where spectral representations provide low-rank priors and language signals encode degradation characteristics. PromptHSI [113] further represents degradation factors through prompts to support a universal hyperspectral restoration model.

Beyond feed-forward architectures, degradation-aware conditioning has also been widely adopted in diffusion-based frameworks. DA-CLIP [114] explores controllable adaptation of VLM representations for multi-task restoration. Methods such as AutoDIR [115], SPIRE [116], and MPerceiver [117] encode degradation descriptors into language-aligned embeddings for diffusion conditioning. VLM-SSR [118] introduces CLIP-derived features to guide real-world super-resolution.

Semantic Conditioning. Semantic conditioning incorporates high-level semantic signals generated from FMs into the restoration process to guide content-aware feature modulation. Unlike degradation-aware conditioning, which focuses on characterizing degradation attributes, this paradigm emphasizes the extraction and utilization of semantic information, such as structural layouts, textures, and scene descriptions, to improve perceptual fidelity and detail reconstruction.

CLIP-SR [119] couples linguistic guidance with image feature processing to refine structure and textures for super-resolution. VLM-IMI [120] further explores instruction-based guidance by generating textural descriptions from the input images and user prompts. Additionally, TPGDiff [121] extends this idea by incorporating VLM-generated degradation priors into a hierarchical multi-prior formulation, constraining the diffusion trajectory at multiple stages. Semantic conditioning has also been explored in more structured learning frameworks. LLMRA [68] employs an MLLM [122] to perform a high-level description of image content, which is then mapped into a language-aligned embedding space via a pretrained CLIP text encoder to guide restoration. PromptFix [123] utilizes a combination of LLaVA [44] together with a pretrained CLIP visual encoder to extract semantic cues, which are used as auxiliary prompts to guide an instruction-driven diffusion restoration model. While both LLMRA and PromptFix adopt similar language-guided pipelines, they differ in the role of semantic signals. In LLMRA, language-derived representations are integral to the restoration process, whereas in PromptFix, they act as an auxiliary enhancement that complements a generative prior.

Global Context Conditioning. Global context conditioning introduces high-level signals that summarize the overall state of the input and provide holistic guidance for restoration. These signals are derived from aggregated information, such as degradation categories, scene-level descriptions, or perceptual quality assessments. In practice, these global signals are encoded into embeddings and fused into the restoration network, where they act as conditioning features that guide the recovery process.

ClearAIR [66] exemplifies this paradigm by integrating multiple language-driven modules to construct complementary global guidance signals. A VLM-based task identifier [114] encodes degradation categories into prompt-like embeddings, while an MLLM-based IQA module [45] produces perceptual quality scores that are further transformed into conditioning features. Together with semantic guidance modules, these signals are injected into the restoration backbone to refine features.

3.3. Optimization-Level Coupling

Loss-Driven Optimization. Loss-driven optimization incorporates language-derived signals into the training objective by formulating them as auxiliary loss terms. In this paradigm, FMs provide

 doi:10.20944/preprints202603.2366.v2

high-level perceptual or semantic supervision, complementing conventional reconstruction losses and optimizing the restoration process.

In structured medical imaging, LEDA [124] enforces consistency between perceptual representations and semantic tokens through a language-informed codebook, while LangMamba [125] integrates language-aligned features into a Mamba-based backbone for efficient CT denoising. In adversarial settings, methods such as NeRCo [126] and SeD [127] condition discriminators on VLM-derived representations, introducing semantic-aware discriminative signals that complement adversarial objectives. This paradigm is particularly effective in scenarios where paired ground-truth data are limited or unavailable. For instance, in underwater image enhancement [128], a FLIP [129]-pretrained text encoder encodes high-level quality concepts via textual prompts, and the resulting similarity scores are incorporated as perceptual loss terms alongside reconstruction objectives. Furthermore, Perceive-IR [130] enables fine-grained quality control through CLIP-aligned prompt learning and difficulty-adaptive supervision.

Despite these advantages, the supervision provided by FMs is often global, which may limit sensitivity to localized artifacts and make performance dependent on prompt design and model expressiveness. Closely related but conceptually distinct from loss-driven optimization, some recent works employ MLLMs as perceptual evaluators to generate pseudo-label supervision for restoration training, rather than directly incorporating language-derived signals into the optimization objective. For example, SnowMaster [56] leverages MLLM-derived perceptual preference feedback to rank candidate desnowing results and uses the selected outputs as pseudo-labels to support semi-supervised training on real-world snowy images.

Reward-Driven Optimization. Reward-driven optimization treats foundation models as perceptual evaluators that provide feedback signals to guide restoration behavior. Rather than defining explicit reconstruction losses, this paradigm leverages reward signals, scoring mechanisms, or evaluator-based feedback to shape optimization or decision policies, shifting the objective from pixel-level fidelity toward perceptual quality.

MOERL [131] formulates image restoration as a reinforcement learning problem, where a Mixture-of-Experts (MoE) model is optimized using perceptual rewards to adaptively handle complex and mixed degradations without requiring explicit labels. In contrast, SimpleCall [132] employs a lightweight agent that sequentially invokes restoration tools based on MLLM-derived feedback, enabling policy learning in a label-free setting without relying on ground-truth supervision.

3.4. Execution-Level Coupling

Planning-Based Execution. Planning-based execution assigns foundation models the role of high-level controllers that organize the restoration process. Instead of providing auxiliary conditioning signals, FMs structure restoration as a sequential decision procedure, where degradations are identified, tasks are decomposed, and restoration operations are scheduled. In this setting, FMs orchestrate the execution logic, while task-specific restoration networks are responsible for pixel-level reconstruction.

Early works such as RestoreAgent [133] and AgenticIR [37] demonstrate this paradigm by enabling MLLMs to determine restoration tasks and execution sequences. Subsequent studies improve planning stability and efficiency. Q-Agent [134] employs chain-of-thought reasoning [135] to decompose multi-degradation perception and adopts quality-driven greedy planning based on NR-IQA, effectively mitigating unnecessary rollbacks. To enhance scalability and robustness, later works extend this paradigm to multi-agent systems. MAIR [136] introduces a scheduler–expert hierarchy that separates degradation perception from restoration execution, while HybridAgent [137] employs collaborative agents to balance planning accuracy and computational cost. Specialized frameworks such as 4KAgent [58] adapt planning-based execution to ultra-high-definition restoration scenarios. System-level approaches, including Clarity ChatGPT [138] and JarvisIR [36], further emphasize interactive refinement and robustness in real-world settings.

 doi:10.20944/preprints202603.2366.v2

4. Experiments

In this section, we summarize representative datasets widely used in language-driven IR frameworks across different tasks. We then review commonly adopted IQA metrics, covering both classical and recent language-driven evaluation paradigms. Together, these datasets and metrics provide a comprehensive view of current benchmarks and evaluation practices for language-driven IR.

4.1. Datasets

Existing language-driven IR studies commonly adopt subsets of established restoration benchmarks, as summarized in Table 2 and Table 3. These datasets vary substantially in scale, spatial resolution, scene diversity, and data-acquisition protocols, spanning both real-world captures and synthetically generated degradations. While synthetic datasets offer controllability and scalability, they may inadequately capture the ambiguity, complexity, and stochastic characteristics of real-world degradations. This discrepancy becomes particularly critical in language-driven frameworks, where degraded understanding relies on semantic representations that are inherently sensitive to contextual variations.

On the other hand, recent language-driven restoration frameworks increasingly require datasets that explicitly model image-language interactions. However, as shown in Figure 5, most datasets consist primarily of degraded-clean image pairs while rarely providing structured linguistic annotations describing degradation characteristics or perceptual attributes. The absence of structured language annotations may introduce inconsistencies when training or evaluating language-driven restoration models. To bridge this gap, PromptFix [123] constructs a large-scale instruction-following dataset that contains approximately 1.01 million input-goal-instruction triplets across diverse low-level tasks. Furthermore, in safety-critical applications such as autonomous driving, CleanBench [36] defines an instruction sample as a triplet, consisting of a user instruction, a degraded image, and a response. Such datasets reflect a paradigm shift in which restoration is formulated not only as pixel reconstruction but also as a language-conditioned generation problem. Despite these initial explorations, restoration datasets with captions are still underexplored.

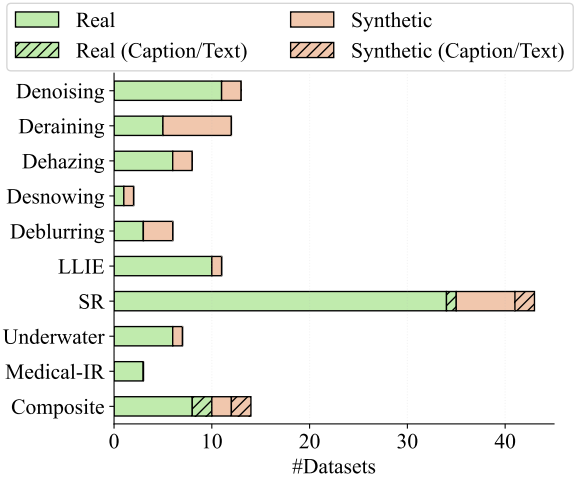

Figure 5. Distribution of datasets across restoration tasks. Bars indicate the number of datasets per task, distinguishing real-world and synthetic datasets. Hatched segments denote datasets accompanied by textual or caption annotations. SR denotes super-resolution.

 doi:10.20944/preprints202603.2366.v2

Table 2. Summary of datasets used in VLM- and MLLM-based image restoration. *r* and *s* denote real and synthetic data, respectively; "–" indicates unavailable train/test splits. LR and HR refer to low- and high-resolution.

Task	Dataset	Year	Type	Domain	Training/Testing	Description
Denoising	Kodak24 [139]	1999	r	Natural	-/24	Clean color images
	McMaster [140]	2011	r	Natural	-/18	18 high-quality color images
	CBSD68 [141]	2001	r	Natural	-/68	68 clean natural images with different noise levels
	Urban100 [142]	2015	r	Natural	-/100	100 high-resolution urban scenes with repetitive structures
	DIV2K [143]	2017	r	Natural	800/100	1000 high-resolution images
	SIDD [144]	2018	r	Natural	-/160	Real-noise image pairs with clean ground truth
	PolyU [53]	2018	r	Natural	-/40	Real-noise paired dataset with 40 scenes
	WED [145]	2016	r&s	Natural	4744/-	Waterloo Exploration Database
	BSD400 [146]	2010	r	Natural	400/-	Training subset from BSD500
	Mayo-2016 [26]	2016	r	Medical	4800/1136	Paired normal-dose and simulated quarter-dose abdominal CT
Deraining	Rain100L [147]	2017	s	Natural	200/100	Images with light rain effect
	Rain100H [147]	2017	s	Natural	1800/100	Images with heavy rain conditions
	Rain800 [148]	2019	s	Natural	700/100	Images with diverse rain patterns
	Rain1400 [149]	2017	s	Natural	12600/1400	14 rain streak types
	Raindrop [150]	2018	r	Natural	1069/58	A paired raindrop dataset captured using dual identical glass setups
	Outdoor-Rain [151]	2019	r&s	Natural	9000/1500	A synthetic outdoor rain dataset with streak and accumulation effects
	RainDS [152]	2021	r&s	Natural	-/5800	Paired deraining dataset organized as a 4-image set
	SSID [153]	2022	r&s	Natural	47600/200	Semi-supervised image deraining sets
	LHP [54]	2023	r	Natural	2100/300	Largest paired real rain dataset with 1920 × 1080 image resolution
Dehazing	FoggyCityscapes [154]	2018	s	Natural	2975/1525	Paired foggy and clear images
	ACDC [155]	2021	r	Natural	1600/2400	Real-world images captured under adverse conditions
	RESIDE [156]	2018	r	Natural	86125/4842	Real and synthetic data across indoor and outdoor scenarios
	NH-HAZE [55]	2020	r	Natural	45/5	A real paired outdoor dehazing set with non-homogeneous haze
	Dense-Haze [157]	2019	r	Natural	45/5	A real paired dehazing dataset for dense, homogeneous haze
Desnowing	RealSnow10K [56]	2025	r	Natural	6406/1047	Real-world snow removal dataset
	Snow100K-L [12]	2018	s	Natural	1872/601	A single-image snow removal benchmark
Deblurring	DPD-blur [14]	2020	r	Natural	350/150	500 real defocus blur image pairs
	DPD-disp [158]	2020	r	Natural	-/350	Reuse the checkpoints trained on the DPD-blur dataset
	DDD-syn [159]	2021	s	Natural	10000/1000	Synthetic deblurring dataset with paired blurry and sharp images
	RDPD [160]	2021	s	Natural	18000/1000	Images captured using a dual-pixel camera
	GoPro [15]	2017	r&s	Natural	2103/1111	Paired images generated from real high-frame-rate GoPro videos
LLIE	LOL-v1 [17]	2018	r	Natural	485/15	Paired low-light and normal-light under controlled conditions
	LSRW [57]	2023	r	Natural	445/50	Paired low-light LR with normal-light HR
	DICM [161]	2013	r	Natural	-/64	Low light images without ground truth for visual comparison
	NPE [162]	2013	r	Natural	-/85	Unpaired low light images
	VV [163]	2018	r	Natural	-/24	24 real-world unpaired low light images
	LOL-v2-real [18]	2021	r	Natural	689/100	Real paired low-light sets
	LOL-v2-syn [18]	2021	s	Natural	900/100	Synthetic paired low-light sets
	MEF [164]	2015	r	Natural	-/17	Multiple images with different exposure levels for the same scene
	SICE [165]	2018	r	Natural	360/229	Multiple reference images of different enhancement levels
	LIME [166]	2016	r	Natural	-/10	10 images without ground truth
Super Resolution	Set5 [167]	2021	r	Natural	-/5	5 real-world natural images
	Set14 [168]	2010	r	Natural	-/14	14 real-world natural images
	Manga109 [169]	2017	r	Natural	-/109	109 real-world manga images
	CelebA [170]	2015	r	Natural	162770/19867	Images with 40 binary attributes
	RealSR [171]	2019	r	Natural	-/35	Real-world low-and high-resolution image pairs
	DrealSR [172]	2020	r	Natural	-/93	93 aligned LR-HR image pairs
	DIV2K-Val [173]	2024	r	Natural	-/100	3K patches from the DIV2K validation set
	RealSRSet [174]	2021	r	Natural	-/20	Contains images captured in practical scenarios
	DIV4K-50 [58]	2024	r	Natural	-/50	256 × 256 distorted images paired with 4096 × 4096 counterparts
	DiffusionDB [175]	2023	s	Natural	-/100	Text-to-image prompt gallery sets
	AID [176]	2017	r	Natural	-/135	Aerial image dataset
	DIOR [177]	2019	r	Natural	-/154	Object detection in optical remote sensing images
	DOTA [178]	2018	r	Natural	-/183	Dataset for object detection in aerial images
	bcSR [179]	2023	r	Medical	-/200	Pathology images patches from breast cancer whole slide images
	US-Case [180]	2025	r	Medical	-/111	Ultrasound cases

Table 3. Summary of datasets used in VLM- and MLLM-based image restoration. *r* and *s* denote real and synthetic data, respectively; "–" indicates unavailable train/test splits. LR and HR refer to low- and high-resolution.

Task	Dataset	Year	Type	Domain	Training/Testing	Description
Underwater	UIEB [23]	2019	r	Underwater	800/90	Underwater image enhancement benchmark
	EUVP [24]	2019	r	Underwater	20000/-	Include both paired and unpaired samples
	RUIE [181]	2020	r	Underwater	-/4230	Real-world underwater image enhancement
Composite	PromptFix [123]	2024	r&s	Natural	101320/-	Paired input–goal–instruction triplets spanning 7 tasks
	MiO100 [61]	2024	r&s	Natural	-/700	Each image is degraded with 7 single degradation types
	AgenticIR [37]	2025	r&s	Natural	-/1440	16 mixed-degradation combinations (2–3 types)
	CleanBench [36]	2025	r&s	Natural	150000/80000	A large-scale, high-quality instruction-response
	MSRS [182]	2022	r	Natural	1163/361	A multi-spectral IR-VIS paired set
	FMB [183]	2023	r	Natural	1220/280	1500 aligned pairs
	CDD-11 [184]	2024	r&s	Natural	13013/2200	1080 × 720 images selected from the RAISE dataset
	TOLED [185]	2021	r	Natural	240/30	A real paired under-display camera restoration set
	AVIRIS [186]	2024	r	HSI	1678/200	Airborne visible/infrared imaging spectrometer
	ARAD [187]	2022	r	HSI	1000/-	A large natural spectral image set

 doi:10.20944/preprints202603.2366.v2

4.2. Evaluation Metrics

Existing IQA methods and evaluation protocols are summarized in Table 4 and Table 5, covering both conventional and language-driven assessment paradigms.

Table 4. Taxonomy of Image Quality Assessment Methods. Language-driven NR-IQA methods are categorized based on the functional role of language, including alignment, reasoning, and scoring paradigms. GT indicates whether references are required.

Category	Sub-category	Representative Methods	GT	Usage
Full-Reference	Non-Learning-Based	PSNR, SSIM [38], FSIM [188], MAE, MSE, RMSE, ERGAS [189]	✓	Pixel-level fidelity or structural consistency
	Learning-Based	LPIPS [40], DISTS [81], CKDN [82], AHIQ [83], TOPIQ-FR [84]	✓	Feature-based perceptual similarity
	Distribution-based	FID [190]	✓	Feature-space distribution alignment
No-Reference	Hand-Crafted	BRISQUE [85], NIQE [86], PIQE [87], LOE [162], PI [191]	×	Blind perceptual quality estimation
	Learning-Based	MUSIQ [88], MANIQA [90], NIMA [89], HyperIQA [91], PAQ2-PIQ [192], DBCNN [193], TOPIQ-NR [84], CNNIQA [92]	×	Learning-based NR-IQA
	Alignment-Based	CLIP-IQA [41], QualiCLIP [43], LIQE [194], SCUIA [195], PromptIQA [196], GRMP-IQA [197], ATTIQA [198], CAP-IQA [199], SFD [200], UniQA [201], RALI [202]	×	Language as representation for perceptual alignment
	Reasoning-Based	DepictQA [95], DepictQA-Wild [203], IQAGPT [204], Co-Instruct [205], Q-Ground [97], SEAGULL [96], AgenticIQA [48]	×	Language-driven quality understanding, explanation, grounding, and decision-making
	Scoring-Based	Q-Align [46], DeQA-Score [45], DogIQA [206], Q-Scorer [98], Compare2Score [94], Q-Insight [207], Q-Ponder [208], Q-Hawkeye [209], LEAF [210]	×	Language-guided quality scoring and calibration
	Resources / Benchmarks	Q-Bench [211], Q-Bench$^+$ [212], Q-Instruct [93]	×	Benchmark datasets and instruction resources for IQA

Table 5. Evaluation protocols for IQA, including human-aligned, task-oriented, and text-based evaluation criteria.

Category	Sub-category	Representative Methods	GT	Usage
Evaluation Protocols	Human-Aligned	PLCC, SRCC, KRCC [213], Weighted Kappa [214], Percent Agreement	×	Correlation with human subjective perception
	Task-Oriented	Precision, Recall, F1, mIoU, Accuracy [95]	✓	Downstream task performance
	Text-Based	BLEU-N [215], ROUGE-L [216], METEOR [217], CIDEr [218]	✓	Textual or semantic fidelity evaluation

Conventional IQA metrics. Image restoration performance is traditionally evaluated using IQA metrics that measure fidelity or perceptual similarity between restored images and references. Full-reference IQA metrics, such as PSNR [39], SSIM [38], FSIM [188], and MAE, quantify pixel-level or structural consistency with ground-truth images. Learning-based metrics (e.g., LPIPS [40], DISTS [81], CKDN [82], and AHIQ [83]) assess perceptual similarity in deep feature spaces and exhibit improved correlation with human judgments. When reference images are unavailable, no-reference IQA metrics are commonly adopted, ranging from hand-crafted statistical measures (e.g., BRISQUE [85], NIQE [86], and PIQE [87]) to learning-based models such as MUSIQ [88] and MANIQA [90]. In addition, distribution-based metrics like FID [190] are often used to evaluate feature distribution alignment between restored and real images, particularly in generative restoration scenarios.

Compared with conventional IQA metrics, recent studies have explored language-driven evaluation paradigms that leverage VLMs and MLLMs. These approaches introduce language as a new modality for modeling perceptual quality, enabling assessment through various mechanisms. Based on the functional role of language, existing methods can be broadly categorized into alignment-based, reasoning-based, and scoring-based IQA.

Alignment-Based IQA. This type of IQA maps images into a language-informed representation space. In this paradigm, textual descriptions, prompts, or attribute embeddings serve as anchors to encode quality semantics, and image quality is inferred through similarity, ranking, or conditional modeling within this shared space.

 doi:10.20944/preprints202603.2366.v2

Early studies, such as CLIP-IQA [41], demonstrate that quality perception can be formulated as a prompt-driven similarity-comparison problem using antonym prompt pairs (e.g., *good* vs. *bad*). Subsequent works extend this principle in various directions. QualiClip [43] proposes a self-supervised, opinion-unaware approach via quality-aware image–text ranking to learn distortion-sensitive representations. PromptIQA [196] further enables requirement-adaptive assessment by incorporating a small set of image–score pairs as prompts. In addition, GRMP-IQA [197] improves data efficiency through meta-prompt learning with gradient regularization, and ATTIQA [198] addresses annotation scarcity through attribute-aware pretraining using pseudo-labels generated by VLMs.

Beyond prompt-based modeling, LIQE [194] establishes a unified vision–language correspondence framework to jointly model semantics, distortions, and quality, while UniQA [201] proposes a multimodal pretraining strategy bridging quality assessment and aesthetic evaluation. SFD [200] further explores semantic feature discrimination as a proxy for quality estimation. In domain-specific settings, SCUIA [195] incorporates semantic-aware contrastive learning for underwater IQA. CAP-IQA [199] integrates text priors with image-specific context for task-aware scoring. Finally, RALI [202] aligns images with quality-aware textual representations distilled from RL-based IQA models, and predicts image quality by similarity matching in the learned text space, thereby preserving generalization without requiring LLM reasoning during inference.

Reasoning-Based IQA. Reasoning-based IQA methods formulate quality assessment as an inference process, where perceptual quality is derived through language-driven analysis, explanation, or decision-making. Instead of directly predicting scores, these methods leverage MLLMs to perform structured reasoning, such as descriptive assessment, comparative judgment, or region-level grounding, to approximate human-like evaluation.

Representative frameworks such as DepictQA [95] and DepictQA-Wild [203] formulate IQA as descriptive and comparative language-based reasoning, enabling human-like assessment. IQAGPT [204] adopts a caption-driven pipeline to jointly produce quality scores and natural-language explanations. Beyond global reasoning, SEAGULL [96] introduces region-aware quality reasoning via SAM-guided feature modeling, while Q-Ground [97] enables fine-grained visual grounding for quality analysis. Co-Instruct [205] further explores open-ended comparative reasoning, allowing MLLMs to generate detailed quality comparisons. Moreover, AgenticIQA [48] advances this paradigm by modeling IQA as a structured decision-making process with planning, execution, and aggregation stages, moving beyond single-pass inference toward agent-based evaluation.

Scoring-Based IQA. Scoring-based IQA methods aim to predict scalar quality scores through direct or learned mappings from images to quality ratings, often enhanced by preference learning, distribution modeling, or distillation strategies.

In particular, Q-Align [46] discretizes subjective scores to a one-hot label to emulate the human judgment process. In contrast, DeQA-Score [45] improves quality prediction by modeling score distributions as soft labels, and Q-Scorer [98] explicitly adapts multimodal representations for direct scalar quality prediction. In addition, Dog-IQA [206] introduces a standard-guided discrete scoring mechanism combined with mix-grained global–local quality aggregation. To address label efficiency, LEAF [210] decouples perceptual knowledge from MOS calibration via distillation.

On the other hand, the comparison-derived scoring framework Compare2Score [94] derives continuous scores from adaptive pairwise comparisons. Q-Insight [207] integrates reinforcement learning with preference modeling, while Q-Ponder [208] introduces explicit joint optimization objectives for both scoring and reasoning. Q-Hawkeye [209] further improves robustness through uncertainty-aware policy optimization.

Resources and Benchmarks. In addition to standalone IQA methods, recent efforts have focused on constructing benchmarks and instruction datasets to evaluate and enhance the perceptual reasoning capabilities of multimodal models. These resources provide standardized evaluation protocols and training signals for language-driven IQA.

 doi:10.20944/preprints202603.2366.v2

For instance, Q-Instruct [93] and Q-Bench [211,212] focus on improving and evaluating low-level perceptual reasoning capabilities of FMs, underscoring the importance of instruction tuning and benchmarking.

Comparison: Conventional vs. Language-Driven Evaluation. Conventional IQA metrics and language-driven evaluation reflect different yet complementary assessment paradigms. Classical full-reference measures (e.g., PSNR, SSIM) quantify pixel fidelity and structural consistency, providing deterministic and reproducible criteria. These metrics remain essential for benchmarking reconstruction accuracy and optimization stability.

By contrast, language-driven evaluation estimates quality through multimodal similarity modeling, preference reasoning, and language-conditioned quality interpretation. Instead of directly measuring pixel fidelity, these approaches estimate quality through mechanisms such as prompt-driven image–text alignment, pairwise comparison, distribution-aware score modeling, and reasoning-guided scoring. Such evaluation strategies are particularly relevant for generative restoration, mixed degradations, and no-reference scenarios. Unlike conventional metrics, language-driven approaches may favor semantic plausibility or contextual consistency.

Importantly, language-driven evaluation should not be interpreted as a direct replacement for conventional IQA metrics, but rather as a complementary perceptual assessment mechanism. Conventional metrics ensure numerical stability and comparability, whereas language-driven evaluators capture complementary perceptual and semantic cues. Recent studies [37,116,132] increasingly adopt hybrid evaluation protocols that combine deterministic fidelity measures with language-driven perceptual assessment to achieve more comprehensive performance evaluation.

However, despite the growing interest in language-driven IQA, a unified evaluation benchmark is still lacking. Existing methods are typically evaluated under task-specific settings, with varying datasets and scoring strategies, making direct quantitative comparison across different works difficult. For instance, recent studies [111,132] demonstrate the effectiveness of language-driven evaluation metrics within their customized experimental setups. However, such results are not directly comparable due to differences in evaluation configurations. Overall, current evidence suggests that language-driven metrics provide complementary insights to conventional measures. Nevertheless, establishing standardized benchmarks for language-driven IQA remains an important open problem.

Evaluation Protocols and Criteria. In addition to direct quality prediction metrics, as shown in Table 5, correlation-based criteria such as PLCC, SRCC, and KRCC [213] are widely adopted to measure the consistency between model-generated assessments and human subjective ratings. These statistics do not evaluate image quality directly but instead quantify the reliability of IQA models with respect to human perception. Additionally, text-based metrics [215,216] are often used to evaluate the semantic fidelity of generated descriptions and reference captions, especially in instruction-following or explanation-based restoration frameworks.

Furthermore, several studies adopt task-oriented evaluation paradigms that assess image quality indirectly through downstream performance [97]. In this setting, restored images are treated as inputs to high-level vision systems, such as detection, segmentation, or recognition models, where improvements in task performance serve as proxies for quality enhancement. The underlying assumption is that degradations often impair feature extraction and perception reliability, and thus, higher downstream accuracy may indicate improved visual quality. Nevertheless, this assumption is not universally valid. Sun et al. [219] have shown that perceptual quality and downstream performance are not strictly causally linked. Despite these limitations, task-driven evaluation remains relevant in application-critical domains, including medical imaging and autonomous driving, where restoration quality is ultimately defined by its functional impact on subsequent perception or decision-making processes.

Metrics for Language-Driven IQA Ability. While the above criteria are commonly used to evaluate IQA outputs, they are still insufficient for systematically assessing the intrinsic IQA capabilities of language-driven models. Leveraging the strong semantic understanding capabilities of

 doi:10.20944/preprints202603.2366.v2

MLLMs and VLMs for IQA has emerged as a promising and rapidly evolving research direction. 2AFC [220] provides an initial attempt by proposing consistency, accuracy, and correlation metrics to analyze judgment robustness and alignment with human opinion scores. Despite this initial effort, the evaluation of language-driven IQA remains largely underexplored, leaving substantial room for further investigation.

4.3. Experimental Results

To better understand the effectiveness of different design choices, we report experimental results of several language-driven all-in-one restoration frameworks under three commonly used experimental settings. Table 6 presents results on three degradations: deraining, denoising ($\sigma = 15, 25, 50$), and dehazing. Meanwhile, Table 7 extends the evaluation to five degradations by additionally including deblurring and LLIE. Table 8 further reports performance on the single LLIE task. All results are obtained from the corresponding original works and reported in terms of PSNR and SSIM.

Table 6. Comparison with state-of-the-art methods on three image restoration tasks. The top rows correspond to non-language-driven approaches, while the bottom rows represent VLM-/MLLM-based methods. Performance is reported in terms of PSNR and SSIM for each dataset.

Method	**Venue**	**Params**	**Deraining**		**Denoising** (BSD68 [141])						**Dehazing**		**Average**	
			Rain100L [147]		$\sigma = 15$		$\sigma = 25$		$\sigma = 50$		SOTS [156]			
AirNet [221]	CVPR'22	9M	34.90	0.967	33.92	0.933	31.26	0.888	28.00	0.797	27.94	0.962	31.20	0.910
IDR [222]	CVPR'23	15M	36.03	0.971	33.89	0.931	31.32	0.884	28.04	0.798	29.87	0.970	31.83	0.911
PromptIR [33]	NeurIPS'23	33M	36.37	0.972	33.98	0.933	31.31	0.888	28.06	0.799	30.58	0.974	32.06	0.913
AdaIR [223]	ICLR'25	29M	38.64	0.983	34.12	0.934	31.45	0.892	28.19	0.802	31.06	0.980	32.69	0.918
DSwinIR [224]	T-PAMI'25	24M	37.73	0.983	34.12	0.933	31.59	0.890	28.31	0.803	31.86	0.980	32.72	0.917
VIVNet [225]	T-PAMI'26	7.42M	38.47	0.983	34.16	0.936	31.50	0.893	28.24	0.806	32.19	0.982	32.91	0.920
InstructIR-3D [34]	ECCV'24	16M	37.98	0.978	34.15	0.933	31.52	0.890	28.30	0.803	30.22	0.959	32.43	0.913
VLU-Net [67]	CVPR'25	35M	38.93	0.984	34.13	0.935	31.48	0.892	28.23	0.804	30.71	0.980	32.70	0.919
Perceive-IR [130]	T-IP'25	42M	38.29	0.980	34.13	0.934	31.53	0.890	28.31	0.804	30.87	0.975	32.63	0.917
ClearAIR [66]	AAAI'26	31M	38.61	0.984	34.18	0.935	31.50	0.891	28.31	0.804	31.08	0.981	32.74	0.919

Table 7. Comparison with state-of-the-art methods on five image restoration tasks. The top rows correspond to non-language-driven approaches, while the bottom rows represent VLM-/MLLM-based methods. Performance is reported in terms of PSNR and SSIM for each dataset.

Method	**Venue**	**Params**	**Dehazing**		**Deraining**		**Denoising**		**Deblurring**		**LLIE**		**Average**	
			SOTS [156]		Rain100L [147]		BSD68$_{\sigma=25}$ [141]		GoPro [15]		LOL [17]			
AirNet [221]	CVPR'22	9M	21.04	0.884	32.98	0.951	30.91	0.882	24.35	0.781	18.18	0.735	25.49	0.847
IDR [222]	CVPR'23	15M	25.24	0.943	35.63	0.965	31.60	0.887	27.87	0.846	21.34	0.826	28.34	0.893
PromptIR [33]	NeurIPS'23	33M	26.54	0.949	36.37	0.970	31.47	0.886	28.71	0.881	22.68	0.832	29.15	0.904
AdaIR [223]	ICLR'25	29M	30.53	0.978	38.02	0.981	31.35	0.888	28.12	0.858	23.00	0.845	30.20	0.910
DSwinIR [224]	T-PAMI'25	24M	30.09	0.975	37.77	0.982	31.34	0.885	29.17	0.879	22.64	0.843	30.20	0.913
VIVNet [225]	T-PAMI'26	7.42M	31.85	0.982	38.67	0.984	31.46	0.892	28.50	0.866	23.03	0.857	30.70	0.916
DA-CLIP [114]	ICLR'24	125M	26.28	0.939	35.91	0.972	25.77	0.653	28.81	0.882	22.57	0.832	29.23	0.898
DiffRes [109]	CVPR'25	45M	27.23	0.958	37.25	0.979	32.07	0.890	29.33	0.883	23.13	0.843	29.78	0.911
InstructIR-5D [34]	ECCV'24	16M	27.10	0.956	36.84	0.973	31.40	0.887	29.40	0.886	23.00	0.836	29.55	0.907
VLU-Net [67]	CVPR'25	35M	30.84	0.980	38.54	0.982	31.43	0.891	27.46	0.840	22.29	0.833	30.11	0.905
Perceive-IR [130]	T-IP'25	42M	28.19	0.964	37.25	0.977	31.44	0.887	29.46	0.886	22.88	0.833	29.84	0.909
ClearAIR [66]	AAAI'26	31M	30.12	0.978	38.20	0.982	31.53	0.888	29.67	0.887	22.83	0.846	30.45	0.916

 doi:10.20944/preprints202603.2366.v2

Table 8. Performance comparison with state-of-the-art approaches on the LOL-v1 [17] dataset.

Method	Venue	PSNR	SSIM
RetinexFormer [30]	ICCV'23	25.16	0.845
LLFormer [226]	AAAI'23	23.65	0.8163
CWNet [227]	ICCV'25	23.60	0.8496
RetinexDiff++ [228]	T-PAMI'25	24.67	0.867
LLMRA [68]	ECCV'24	23.30	0.846
DA-CLIP [114]	ICLR'24	23.40	0.811
DiffRes [109]	CVPR'25	24.55	0.839
Perceive-IR [130]	T-IP'25	23.79	0.841

Under the three-degradation setting (Table 6), language-driven restoration frameworks exhibit comparable overall performance. For instance, VLU-Net [67] achieves the highest deraining PSNR (38.93 dB), whereas ClearAIR [66] exhibits more consistent cross-task behavior, maintaining stable performance across deraining, denoising, and dehazing. A similar trend is observed in the five-degradation setting (Table 7), where ClearAIR achieves the best average performance compared to other language-driven approaches. In contrast, on the single-task LLIE benchmark (Table 8), task-specific non-language-driven methods such as RetinexFormer [30] (25.16 dB) and RetinexDiff++ [228] (24.67 dB) outperform most language-driven models in terms of PSNR.

These observations suggest that, although language-driven frameworks enhance flexibility and generalization in multi-degradation scenarios, their advantages in specialized tasks remain limited. Moreover, gains in pixel-level fidelity (e.g., PSNR) remain limited. This is because language-driven methods focus more on semantic alignment and user intent than on distortion minimization. This leads to a mismatch between current evaluation protocols and the goals of language-driven restoration. Moreover, language-driven IQA metrics are rarely included in existing benchmarks, highlighting the need for evaluation frameworks that jointly consider fidelity, semantic correctness, and user-oriented restoration quality.

5. Discussion and Open Challenges

Although the integration of VLMs and MLLMs has driven the development of IR frameworks, it has introduced new capabilities while simultaneously giving rise to additional challenges. In this section, we analyze key open challenges in language-driven methods, focusing on generalization, computational efficiency, cross-paradigm trade-offs, evaluation reliability, dataset design, high-dimensional representations, and trustworthiness. We further discuss potential research directions that may help address these challenges and inform future developments in language-driven restoration systems.

5.1. Generalization and Robustness

In real-world scenarios, degradations are often complex, mixed, or poorly defined, making generalization and robustness persistent challenges for IR, even when incorporating semantic priors and degradation-aware guidance via VLMs and MLLMs. Specifically, while VLM- or MLLM-derived representations provide high-level semantic context, they do not fully eliminate sensitivity to degradation variations. Second, language-driven frameworks frequently rely on textual descriptions, prompt formulations, or semantic interpretations to characterize degradation types [68,120,123]. Such dependencies may lead to inconsistent restoration behaviors due to linguistic variability and prompt sensitivity. Moreover, language models themselves may exhibit hallucinations and domain biases inherited from large-scale pretraining, which can affect degradation understanding and guidance reliability.

One potential solution involves prompt-invariant modeling, such as template-based or structured prompts, to reduce instability caused by diverse linguistic formulations. Another direction is uncertainty-aware restoration, where degradation cues are accompanied by confidence estimates to im-

 doi:10.20944/preprints202603.2366.v2

prove robustness against hallucination-induced errors. Despite these emerging strategies, developing principled mechanisms for handling degradation ambiguity, prompt variability, and language-model uncertainty remains an important challenge for future research.

5.2. Computational Efficiency

While language-driven IR frameworks introduce enhanced semantic awareness and perceptual reasoning capabilities, computational efficiency remains a central challenge. The computational burden arises from multiple sources. First, pretrained VLMs and MLLMs typically contain a large number of parameters, making semantic feature extraction and cross-modal reasoning inherently expensive. Additional conditioning and reasoning mechanisms further increase memory consumption. Moreover, agentic paradigms often require multiple reasoning iterations, candidate evaluations, or tool invocation loops, leading to high latency and inference cost.

Recent studies have explored several strategies to mitigate these efficiency bottlenecks. For example, policy optimization [132] and dynamic routing mechanisms [134] have been applied to reduce unnecessary reasoning and model invocation steps. Hybrid designs [137] further improve efficiency by restricting expensive language-driven computation to critical stages. Despite these advances, computational efficiency remains a key challenge for practical deployment. Future research may focus on developing compact multimodal models tailored for restoration tasks as well as reducing redundancy in cross-modal representations through mechanisms such as knowledge distillation or token reduction. In addition, resolution-adaptive architectures present a promising direction, enabling language-driven reasoning to operate at coarse semantic scales while preserving high-resolution reconstruction within lightweight visual backbones.

5.3. Cross-Paradigm Trade-offs

Existing language-driven approaches are built upon fundamentally different interaction paradigms, leading to inherent trade-offs in system design. A fundamental distinction lies in how VLMs and MLLMs participate in the restoration process, ranging from static guidance to iterative involvement. In static paradigms, language information is incorporated once through fixed embeddings, prompts, or auxiliary objectives, resulting in stable and computationally efficient pipelines. However, such limited involvement restricts adaptability, as the restoration process cannot revise its behavior during inference. In contrast, interactive paradigms repeatedly engage language models during restoration, enabling dynamic reasoning, planning, and execution. This allows better handling of complex or mixed degradations and supports more flexible restoration strategies. Nevertheless, increased interaction also introduces higher computational cost and the risk of error accumulation across multiple steps, potentially leading to unstable outcomes.

These interaction patterns are further reflected in how language influences restoration, either through implicit guidance (e.g., feature modulation or loss design) or explicit decision-making (e.g., planning, tool invocation, or agent-based control). While implicit mechanisms provide stability and end-to-end optimization, explicit interaction offers greater interpretability and controllability at the cost of increased system complexity.

Overall, rather than converging to a single unified framework, future research may benefit from hybrid strategies that balance static and interactive involvement while combining implicit and explicit forms of language guidance.

5.4. Evaluation Reliability

In contrast to traditional evaluation metrics, language-driven IQA models do not always produce numerically stable or strictly consistent scores. Instead, they often operate in a semantic assessment space, where quality judgments are expressed through preferences, rankings, or linguistic interpretations [95,203]. Recently, several studies [45,98] have explored MLLMs as quantitative quality scorers. While such models demonstrate promising alignment with human perception in many scenarios, their evaluation behavior may exhibit variability and uncertainty. In particular, assessment results can be

 doi:10.20944/preprints202603.2366.v2

sensitive to prompt formulations, where small variations in phrasing or textual context may lead to inconsistent quality predictions. This prompt sensitivity introduces challenges for reproducibility and comparability across evaluation settings. Moreover, score calibration remains a nontrivial challenge, as language models are not explicitly optimized for metric-level numerical stability. This variability becomes particularly critical when language-based IQA outputs are directly used for model selection or benchmark comparisons.

The coexistence of conventional and language-driven IQA metrics introduces new considerations regarding evaluation reliability. Future research may explore standardized prompting protocols, confidence-aware evaluation mechanisms, and hybrid assessment frameworks that combine deterministic metrics with language-driven quality reasoning. Nevertheless, ensuring stable and comparable evaluation remains an open problem.

5.5. Dataset Design for Language-Driven IR

Numerous datasets have been proposed for various restoration tasks. However, several limitations still remain in existing datasets. Current restoration benchmarks rarely incorporate structured linguistic annotations describing degradation characteristics or perceptual attributes. While VLMs and MLLMs benefit from large-scale pretraining and strong semantic understanding, they may still exhibit failure modes in complex or ambiguous scenarios. Consequently, language-driven restoration frameworks often rely on synthetic prompts [112] or automatically generated descriptions [56,68], which can introduce semantic inconsistencies. This highlights the importance of reliable textual supervision for improving model stability and performance. Moreover, most restoration datasets are constructed under predefined degradation models, which may not adequately capture the complexity or compositional nature of real-world degradations. Such limitations may restrict the robustness and generalization capability of language-integrated restoration systems.

Future dataset construction may benefit from multi-level annotations beyond pixel-level ground truth, including degradation semantics, perceptual quality descriptions, and task-oriented linguistic guidance. Furthermore, developing standardized degradation annotation protocols represents a critical research direction for improving cross-modal consistency while reducing annotation ambiguity. Furthermore, incorporating diverse, composed, and open-world degradations may enhance model robustness under realistic conditions.

5.6. Leveraging Multimodal Data and High-dimensional Representations

VLMs and MLLMs enable IR frameworks to incorporate textual priors and cross-modal semantics, opening new opportunities to exploit richer data modalities. However, existing studies mainly focus on RGB images, while other sensing modalities remain largely underexplored, including infrared imagery, event-based data, and depth measurements [229,230]. Integrating heterogeneous sensory inputs may improve the interpretation of degradation, structural reasoning, and contextual consistency. Therefore, extending language-driven restoration frameworks to these modalities presents a promising research direction.

Beyond static images, applying language-driven restoration paradigms to high-dimensional data, such as videos, represents another important direction for future research. Video restoration inherently requires not only accurate frame-level reconstruction but also temporal consistency across frames. In this context, language-driven reasoning mechanisms may provide complementary benefits by scene-level coherence. For example, MLLMs may assist in interpreting dynamic degradation patterns [231] or evaluating video quality to support invoking tools.

5.7. Ethics and Trustworthiness

Ethical considerations and trustworthiness have become increasingly relevant in language-driven IR frameworks. Certain approaches rely on online inference services, which require transmitting visual data for non-local processing [138]. Such designs may raise privacy and data security concerns,

 doi:10.20944/preprints202603.2366.v2

particularly in sensitive scenarios including medical imaging, unmanned aerial vehicle (UAV), and autonomous driving.

Beyond data privacy, the reliability of language-driven components introduces additional challenges. MLLMs are known to exhibit hallucinations, reasoning inconsistencies, and biases inherited from large-scale training data. When integrated into restoration pipelines, these factors may result in incorrect degradation interpretation, unstable guidance signals, or semantically inconsistent restoration outcomes. Unlike conventional restoration errors, failures caused by MLLMs or VLMs may be less predictable and more difficult to attribute.

Future research may investigate privacy-preserving multimodal restoration frameworks and robustness against hallucination-induced errors. In addition, the development of locally deployable multimodal models [72] represents a promising direction for mitigating privacy concerns while preserving semantic reasoning capabilities.

6. Conclusions

This survey presents a unified view of language-driven IR by systematically analyzing how VLMs and MLLMs are integrated into restoration pipelines. Instead of categorizing methods solely by model architecture, we introduced an interaction-centric taxonomy that characterizes the role of language across feature-level, optimization-level, and execution-level paradigms. This perspective reveals that recent advances in IR are not merely driven by architectural modifications, but by fundamentally new ways of incorporating language priors into low-level vision tasks. Building on this taxonomy, we reviewed representative language-driven IR frameworks and highlighted how language information enables more flexible, controllable, and semantically aware restoration. We further summarized emerging language-driven IQA methods, emphasizing their differences from conventional metrics, particularly in capturing perceptual and contextual quality. In addition, we provided a comprehensive overview of commonly used datasets and compared recent methods across multiple restoration tasks.

Despite these advances, language-driven IR still faces several fundamental challenges. These include the trade-off between semantic guidance and pixel-level fidelity, limited generalization under complex and unseen degradations, high computational cost, and the lack of reliable and standardized evaluation protocols. Moreover, the construction of high-quality image–text paired datasets remains underexplored, and the implementation of online MLLMs raises concerns about robustness and trustworthiness in real-world applications. These challenges suggest that language-driven restoration should be viewed not only as an extension of existing pipelines but as a paradigm shift that requires rethinking both system-level model design and evaluation strategies. We further anticipate future research directions: exploring uncertainty-aware restoration mechanisms, more efficient semantic-guided architectures, and scalable annotation strategies for image–text paired datasets. Extending language-driven restoration paradigms to heterogeneous modalities and high-dimensional data also presents promising opportunities.

In summary, as language and vision models continue to develop, their integration is expected to play an increasingly important role in bridging low-level visual restoration with high-level semantic understanding. We hope this survey provides a clear and structured foundation for understanding the evolving design principles of language-driven IR and encourages further investigation into this rapidly developing research direction.s

References

1. Jiang, B.; Li, J.; Lu, Y.; Cai, Q.; Song, H.; Lu, G. Efficient image denoising using deep learning: A brief survey. *Information Fusion* **2025**, p. 103013.
2. Tian, C.; Xu, Y.; Zuo, W. Image denoising using deep CNN with batch renormalization. *Neural Networks* **2020**, *121*, 461–473.
3. 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* **2017**, *26*, 3142–3155.

 doi:10.20944/preprints202603.2366.v2

4. Chen, X.; Pan, J.; Dong, J.; Tang, J. Towards unified deep image deraining: A survey and a new benchmark. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2025**.
5. Chen, X.; Li, H.; Li, M.; Pan, J. Learning a sparse transformer network for effective image deraining. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2023, pp. 5896–5905.
6. Xiao, J.; Fu, X.; Liu, A.; Wu, F.; Zha, Z.J. Image de-raining transformer. *IEEE transactions on pattern analysis and machine intelligence* **2022**, *45*, 12978–12995.
7. Gui, J.; Cong, X.; Cao, Y.; Ren, W.; Zhang, J.; Zhang, J.; Cao, J.; Tao, D. A comprehensive survey and taxonomy on single image dehazing based on deep learning. *ACM Computing Surveys* **2023**, *55*, 1–37.
8. Tsai, F.J.; Peng, Y.T.; Lin, Y.Y.; Lin, C.W. PHATNet: A Physics-guided Haze Transfer Network for Domain-adaptive Real-world Image Dehazing. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision, 2025, pp. 5591–5600.
9. Yu, H.; Huang, J.; Zheng, K.; Zhao, F. High-quality image dehazing with diffusion model. *arXiv preprint arXiv:2308.11949* **2023**.
10. Quan, Y.; Tan, X.; Huang, Y.; Xu, Y.; Ji, H. Image desnowing via deep invertible separation. *IEEE Transactions on Circuits and Systems for Video Technology* **2023**, *33*, 3133–3144.
11. Guo, X.; Wang, X.; Fu, X.; Zha, Z.J. Deep unfolding network for image desnowing with snow shape prior. *IEEE Transactions on Circuits and Systems for Video Technology* **2025**.
12. Liu, Y.F.; Jaw, D.W.; Huang, S.C.; Hwang, J.N. Desnownet: Context-aware deep network for snow removal. *IEEE Transactions on Image Processing* **2018**, *27*, 3064–3073.
13. Xiang, Y.; Zhou, H.; Li, C.; Sun, F.; Li, Z.; Xie, Y. Deep learning in motion deblurring: current status, benchmarks and future prospects. *The Visual Computer* **2025**, *41*, 3801–3827.
14. Abuolaim, A.; Brown, M.S. Defocus deblurring using dual-pixel data. In Proceedings of the European conference on computer vision. Springer, 2020, pp. 111–126.
15. Nah, S.; Hyun Kim, T.; Mu Lee, K. Deep multi-scale convolutional neural network for dynamic scene deblurring. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 3883–3891.
16. Li, C.; Guo, C.; Han, L.; Jiang, J.; Cheng, M.M.; Gu, J.; Loy, C.C. Low-light image and video enhancement using deep learning: A survey. *IEEE transactions on pattern analysis and machine intelligence* **2021**, *44*, 9396–9416.
17. Wei, C.; Wang, W.; Yang, W.; Liu, J. Deep Retinex Decomposition for Low-Light Enhancement. In Proceedings of the BMVC, 2018.
18. Yang, W.; Wang, W.; Huang, H.; Wang, S.; Liu, J. Sparse gradient regularized deep retinex network for robust low-light image enhancement. *IEEE Transactions on Image Processing* **2021**, *30*, 2072–2086.
19. Wang, Z.; Chen, J.; Hoi, S.C. Deep learning for image super-resolution: A survey. *IEEE transactions on pattern analysis and machine intelligence* **2020**, *43*, 3365–3387.
20. Wang, X.; Xie, L.; Dong, C.; Shan, Y. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In Proceedings of the Proceedings of the IEEE/CVF international conference on computer vision, 2021, pp. 1905–1914.
21. Chen, X.; Wang, X.; Zhou, J.; Qiao, Y.; Dong, C. Activating more pixels in image super-resolution transformer. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2023, pp. 22367–22377.
22. Zhang, W.; Dong, L.; Pan, X.; Zou, P.; Qin, L.; Xu, W. A survey of restoration and enhancement for underwater images. *IEEE Access* **2019**, *7*, 182259–182279.
23. Li, C.; Guo, C.; Ren, W.; Cong, R.; Hou, J.; Kwong, S.; Tao, D. An underwater image enhancement benchmark dataset and beyond. *IEEE transactions on image processing* **2019**, *29*, 4376–4389.
24. Islam, M.J.; Xia, Y.; Sattar, J. Fast underwater image enhancement for improved visual perception. *IEEE robotics and automation letters* **2020**, *5*, 3227–3234.
25. Kermany, D.S.; Goldbaum, M.; Cai, W.; Valentim, C.C.; Liang, H.; Baxter, S.L.; McKeown, A.; Yang, G.; Wu, X.; Yan, F.; et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *cell* **2018**, *172*, 1122–1131.
26. McCollough, C.H.; Bartley, A.C.; Carter, R.E.; Chen, B.; Drees, T.A.; Edwards, P.; Holmes III, D.R.; Huang, A.E.; Khan, F.; Leng, S.; et al. Low-dose CT for the detection and classification of metastatic liver lesions: results of the 2016 low dose CT grand challenge. *Medical physics* **2017**, *44*, e339–e352.
27. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems* **2012**, *25*.

 doi:10.20944/preprints202603.2366.v2

28. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Advances in neural information processing systems* **2017**, *30*.
29. Gu, A.; Dao, T. Mamba: Linear-time sequence modeling with selective state spaces. In Proceedings of the First conference on language modeling, 2024.
30. Cai, Y.; Bian, H.; Lin, J.; Wang, H.; Timofte, R.; Zhang, Y. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In Proceedings of the Proceedings of the IEEE/CVF international conference on computer vision, 2023, pp. 12504–12513.
31. Jiang, J.; Zuo, Z.; Wu, G.; Jiang, K.; Liu, X. A survey on all-in-one image restoration: Taxonomy, evaluation and future trends. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2025**.
32. Li, R.; Tan, R.T.; Cheong, L.F. All in one bad weather removal using architectural search. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 3175–3185.
33. Potlapalli, V.; Zamir, S.W.; Khan, S.H.; Shahbaz Khan, F. Promptir: Prompting for all-in-one image restoration. *Advances in Neural Information Processing Systems* **2023**, *36*, 71275–71293.
34. Conde, M.V.; Geigle, G.; Timofte, R. Instructir: High-quality image restoration following human instructions. In Proceedings of the European Conference on Computer Vision. Springer, 2024, pp. 1–21.
35. Guan, C.; Yoshie, O. CLIP-driven rain perception: Adaptive deraining with pattern-aware network routing and mask-guided cross-attention. *arXiv preprint arXiv:2506.01366* **2025**.
36. 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 Proceedings of the Computer Vision and Pattern Recognition Conference, 2025, pp. 22369–22380.
37. Zhu, K.; Gu, J.; You, Z.; Qiao, Y.; Dong, C. An intelligent agentic system for complex image restoration problems. *arXiv preprint arXiv:2410.17809* **2024**.
38. 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* **2004**, *13*, 600–612.
39. Huynh-Thu, Q.; Ghanbari, M. Scope of validity of PSNR in image/video quality assessment. *Electronics letters* **2008**, *44*, 800–801.
40. 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 Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 586–595.
41. Wang, J.; Chan, K.C.; Loy, C.C. Exploring clip for assessing the look and feel of images. In Proceedings of the Proceedings of the AAAI conference on artificial intelligence, 2023, Vol. 37, pp. 2555–2563.
42. Hessel, J.; Holtzman, A.; Forbes, M.; Le Bras, R.; Choi, Y. Clipscore: A reference-free evaluation metric for image captioning. In Proceedings of the Proceedings of the 2021 conference on empirical methods in natural language processing, 2021, pp. 7514–7528.
43. Agnolucci, L.; Galteri, L.; Bertini, M. Quality-aware image-text alignment for opinion-unaware image quality assessment. *arXiv preprint arXiv:2403.11176* **2024**.
44. Liu, H.; Li, C.; Wu, Q.; Lee, Y.J. Visual instruction tuning. *Advances in neural information processing systems* **2023**, *36*, 34892–34916.
45. You, Z.; Cai, X.; Gu, J.; Xue, T.; Dong, C. Teaching large language models to regress accurate image quality scores using score distribution. In Proceedings of the Proceedings of the Computer Vision and Pattern Recognition Conference, 2025, pp. 14483–14494.
46. Wu, H.; Zhang, Z.; Zhang, W.; Chen, C.; Liao, L.; Li, C.; Gao, Y.; Wang, A.; Zhang, E.; Sun, W.; et al. Q-align: Teaching lmms for visual scoring via discrete text-defined levels. *arXiv preprint arXiv:2312.17090* **2023**.
47. Zhang, Z.; Wu, H.; Jia, Z.; Lin, W.; Zhai, G. Teaching lmms for image quality scoring and interpreting. *arXiv preprint arXiv:2503.09197* **2025**.
48. Zhu, H.; Tian, Y.; Ding, K.; Chen, B.; Chen, B.; Wang, S.; Lin, W. Agenticiqa: An agentic framework for adaptive and interpretable image quality assessment. *arXiv preprint arXiv:2509.26006* **2025**.
49. Su, J.; Xu, B.; Yin, H. A survey of deep learning approaches to image restoration. *Neurocomputing* **2022**, *487*, 46–65.
50. Wang, L.; Zhou, W.; Wang, C.; Lam, K.M.; Su, Z.; Pan, J. Deep Learning-Driven Ultra-High-Definition Image Restoration: A Survey. *arXiv preprint arXiv:2505.16161* **2025**.
51. Zhang, J.; Huang, J.; Jin, S.; Lu, S. Vision-language models for vision tasks: A survey. *IEEE transactions on pattern analysis and machine intelligence* **2024**, *46*, 5625–5644.

 doi:10.20944/preprints202603.2366.v2

52. Suhr, A.; Zhou, S.; Zhang, A.; Zhang, I.; Bai, H.; Artzi, Y. A corpus for reasoning about natural language grounded in photographs. In Proceedings of the Proceedings of the 57th annual meeting of the association for computational linguistics, 2019, pp. 6418–6428.
53. Xu, J.; Li, H.; Liang, Z.; Zhang, D.; Zhang, L. Real-world noisy image denoising: A new benchmark. *arXiv preprint arXiv:1804.02603* **2018**.
54. Guo, Y.; Xiao, X.; Chang, Y.; Deng, S.; Yan, L. From sky to the ground: A large-scale benchmark and simple baseline towards real rain removal. In Proceedings of the Proceedings of the IEEE/CVF international conference on computer vision, 2023, pp. 12097–12107.
55. Ancuti, C.O.; Ancuti, C.; Timofte, R. NH-HAZE: An image dehazing benchmark with non-homogeneous hazy and haze-free images. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops, 2020, pp. 444–445.
56. Lai, J.; Chen, S.; Lin, Y.; Ye, T.; Liu, Y.; Fei, S.; Xing, Z.; Wu, H.; Wang, W.; Zhu, L. SnowMaster: Comprehensive Real-world Image Desnowing via MLLM with Multi-Model Feedback Optimization. In Proceedings of the Proceedings of the Computer Vision and Pattern Recognition Conference, 2025, pp. 4302–4312.
57. Hai, J.; Xuan, Z.; Yang, R.; Hao, Y.; Zou, F.; Lin, F.; Han, S. R2rnet: Low-light image enhancement via real-low to real-normal network. *Journal of Visual Communication and Image Representation* **2023**, *90*, 103712.
58. Zuo, Y.; Zheng, Q.; Wu, M.; Jiang, X.; Li, R.; Wang, J.; Zhang, Y.; Mai, G.; Wang, L.V.; Zou, J.; et al. 4kagent: agentic any image to 4k super-resolution. *arXiv preprint arXiv:2507.07105* **2025**.
59. Berman, D.; Levy, D.; Avidan, S.; Treibitz, T. Underwater single image color restoration using haze-lines and a new quantitative dataset. *IEEE transactions on pattern analysis and machine intelligence* **2020**, *43*, 2822–2837.
60. Knoll, F.; Zbontar, J.; Sriram, A.; Muckley, M.J.; Bruno, M.; Defazio, A.; Parente, M.; Geras, K.J.; Katsnelson, J.; Chandarana, H.; et al. fastMRI: A publicly available raw k-space and DICOM dataset of knee images for accelerated MR image reconstruction using machine learning. *Radiology: Artificial Intelligence* **2020**, *2*, e190007.
61. Kong, X.; Dong, C.; Zhang, L. Towards effective multiple-in-one image restoration: A sequential and prompt learning strategy. *arXiv preprint arXiv:2401.03379* **2024**.
62. Wang, Z.; Cun, X.; Bao, J.; Zhou, W.; Liu, J.; Li, H. Uformer: A general u-shaped transformer for image restoration. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022, pp. 17683–17693.
63. 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 Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022, pp. 5728–5739.
64. Liu, M.; Cui, Y.; Liu, X.; Strand, L.; Yin, H.; Knoll, A. Drfir: A dimensionality reduction framework for all-in-one image restoration in spatial and frequency domains. *Expert Systems with Applications* **2025**, p. 128959.
65. Zhang, X.; Zhang, H.; Wang, G.; Zhang, Q.; Zhang, L.; Du, B. UniUIR: Considering Underwater Image Restoration as An All-in-One Learner. *arXiv preprint arXiv:2501.12981* **2025**.
66. Zhang, X.; Zhang, H.; Wang, G.; Zhang, Q.; Zhang, L. ClearAIR: A Human-Visual-Perception-Inspired All-in-One Image Restoration. *arXiv preprint arXiv:2601.02763* **2026**.
67. Zeng, H.; Wang, X.; Chen, Y.; Su, J.; Liu, J. Vision-Language Gradient Descent-driven All-in-One Deep Unfolding Networks. In Proceedings of the Proceedings of the Computer Vision and Pattern Recognition Conference, 2025, pp. 7524–7533.
68. Jin, X.; Shi, Y.; Xia, B.; Yang, W. Llmra: Multi-modal large language model based restoration assistant. *arXiv preprint arXiv:2401.11401* **2024**.
69. Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.D.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. Language models are few-shot learners. *Advances in neural information processing systems* **2020**, *33*, 1877–1901.
70. Grattafiori, A.; Dubey, A.; Jauhri, A.; Pandey, A.; Kadian, A.; Al-Dahle, A.; Letman, A.; Mathur, A.; Schelten, A.; Vaughan, A.; et al. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783* **2024**.
71. Bai, J.; Bai, S.; Chu, Y.; Cui, Z.; Dang, K.; Deng, X.; Fan, Y.; Ge, W.; Han, Y.; Huang, F.; et al. Qwen technical report. *arXiv preprint arXiv:2309.16609* **2023**.
72. Liu, A.; Feng, B.; Xue, B.; Wang, B.; Wu, B.; Lu, C.; Zhao, C.; Deng, C.; Zhang, C.; Ruan, C.; et al. Deepseek-v3 technical report. *arXiv preprint arXiv:2412.19437* **2024**.

 doi:10.20944/preprints202603.2366.v2

73. Luo, H.; Bao, J.; Wu, Y.; He, X.; Li, T. Segclip: Patch aggregation with learnable centers for open-vocabulary semantic segmentation. In Proceedings of the International Conference on Machine Learning. PMLR, 2023, pp. 23033–23044.
74. Yao, L.; Han, J.; Wen, Y.; Liang, X.; Xu, D.; Zhang, W.; Li, Z.; Xu, C.; Xu, H. Detclip: Dictionary-enriched visual-concept paralleled pre-training for open-world detection. *Advances in Neural Information Processing Systems* **2022**, *35*, 9125–9138.
75. Jia, C.; Yang, Y.; Xia, Y.; Chen, Y.T.; Parekh, Z.; Pham, H.; Le, Q.; Sung, Y.H.; Li, Z.; Duerig, T. Scaling up visual and vision-language representation learning with noisy text supervision. In Proceedings of the International conference on machine learning. PMLR, 2021, pp. 4904–4916.
76. Hurst, A.; Lerer, A.; Goucher, A.P.; Perelman, A.; Ramesh, A.; Clark, A.; Ostrow, A.; Welihinda, A.; Hayes, A.; Radford, A.; et al. Gpt-4o system card. *arXiv preprint arXiv:2410.21276* **2024**.
77. Team, G.; Anil, R.; Borgeaud, S.; Alayrac, J.B.; Yu, J.; Soricut, R.; Schalkwyk, J.; Dai, A.M.; Hauth, A.; Millican, K.; et al. Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805* **2023**.
78. Wang, P.; Bai, S.; Tan, S.; Wang, S.; Fan, Z.; Bai, J.; Chen, K.; Liu, X.; Wang, J.; Ge, W.; et al. Qwen2-vl: Enhancing vision-language model's perception of the world at any resolution. *arXiv preprint arXiv:2409.12191* **2024**.
79. Bai, S.; Chen, K.; Liu, X.; Wang, J.; Ge, W.; Song, S.; Dang, K.; Wang, P.; Wang, S.; Tang, J.; et al. Qwen2. 5-vl technical report. *arXiv preprint arXiv:2502.13923* **2025**.
80. Zhai, G.; Min, X. Perceptual image quality assessment: a survey. *Science China Information Sciences* **2020**, *63*, 211301.
81. 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* **2020**, *44*, 2567–2581.
82. Zheng, H.; Yang, H.; Fu, J.; Zha, Z.J.; Luo, J. Learning conditional knowledge distillation for degraded-reference image quality assessment. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 10242–10251.
83. Lao, S.; Gong, Y.; Shi, S.; Yang, S.; Wu, T.; Wang, J.; Xia, W.; Yang, Y. Attentions help cnns see better: Attention-based hybrid image quality assessment network. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022, pp. 1140–1149.
84. Chen, C.; Mo, J.; Hou, J.; Wu, H.; Liao, L.; Sun, W.; Yan, Q.; Lin, W. Topiq: A top-down approach from semantics to distortions for image quality assessment. *IEEE Transactions on Image Processing* **2024**, *33*, 2404–2418.
85. Mittal, A.; Moorthy, A.K.; Bovik, A.C. No-reference image quality assessment in the spatial domain. *IEEE Transactions on image processing* **2012**, *21*, 4695–4708.
86. Mittal, A.; Soundararajan, R.; Bovik, A.C. Making a "completely blind" image quality analyzer. *IEEE Signal processing letters* **2012**, *20*, 209–212.
87. Venkatanath, N.; Praneeth, D.; Sumohana, S.C.; Swarup, S.M.; et al. Blind image quality evaluation using perception based features. In Proceedings of the 2015 twenty first national conference on communications (NCC). IEEE, 2015, pp. 1–6.
88. Ke, J.; Wang, Q.; Wang, Y.; Milanfar, P.; Yang, F. Musiq: Multi-scale image quality transformer. In Proceedings of the Proceedings of the IEEE/CVF international conference on computer vision, 2021, pp. 5148–5157.
89. Talebi, H.; Milanfar, P. NIMA: Neural image assessment. *IEEE transactions on image processing* **2018**, *27*, 3998–4011.
90. 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 Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022, pp. 1191–1200.
91. Su, S.; Yan, Q.; Zhu, Y.; Zhang, C.; Ge, X.; Sun, J.; Zhang, Y. Blindly assess image quality in the wild guided by a self-adaptive hyper network. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 3667–3676.
92. Kang, L.; Ye, P.; Li, Y.; Doermann, D. Convolutional neural networks for no-reference image quality assessment. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 1733–1740.
93. Wu, H.; Zhang, Z.; Zhang, E.; Chen, C.; Liao, L.; Wang, A.; Xu, K.; Li, C.; Hou, J.; Zhai, G.; et al. Q-instruct: Improving low-level visual abilities for multi-modality foundation models. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2024, pp. 25490–25500.

 doi:10.20944/preprints202603.2366.v2

94. Zhu, H.; Wu, H.; Li, Y.; Zhang, Z.; Chen, B.; Zhu, L.; Fang, Y.; Zhai, G.; Lin, W.; Wang, S. Adaptive image quality assessment via teaching large multimodal model to compare. *Advances in Neural Information Processing Systems* **2024**, *37*, 32611–32629.
95. You, Z.; Li, Z.; Gu, J.; Yin, Z.; Xue, T.; Dong, C. Depicting beyond scores: Advancing image quality assessment through multi-modal language models. In Proceedings of the European Conference on Computer Vision. Springer, 2024, pp. 259–276.
96. Chen, Z.; Wang, J.; Wang, W.; Xu, S.; Xiong, H.; Zeng, Y.; Guo, J.; Wang, S.; Yuan, C.; Li, B.; et al. Seagull: No-reference image quality assessment for regions of interest via vision-language instruction tuning. *arXiv preprint arXiv:2411.10161* **2024**.
97. Chen, C.; Yang, S.; Wu, H.; Liao, L.; Zhang, Z.; Wang, A.; Sun, W.; Yan, Q.; Lin, W. Q-ground: Image quality grounding with large multi-modality models. In Proceedings of the Proceedings of the 32nd ACM International Conference on Multimedia, 2024, pp. 486–495.
98. Tang, Z.; Yang, S.; Peng, B.; Wang, Z.; Dong, J. Revisiting MLLM Based Image Quality Assessment: Errors and Remedy. *arXiv preprint arXiv:2511.07812* **2025**.
99. Su, Z.; Zhang, Y.; Shi, J.; Zhang, X.P. A survey of single image rain removal based on deep learning. *ACM Computing Surveys* **2023**, *56*, 1–35.
100. Zhang, K.; Ren, W.; Luo, W.; Lai, W.S.; Stenger, B.; Yang, M.H.; Li, H. Deep image deblurring: A survey. *International Journal of Computer Vision* **2022**, *130*, 2103–2130.
101. Zhu, R.; Sheng, L.; Wu, K.; Boukerche, A.; Long, L.; Yang, Q. Toward Efficient Underwater Visual Perception through Image Enhancement, Compression, and Understanding. *ACM Computing Surveys* **2026**, *58*, 1–46.
102. Li, X.; Ren, Y.; Jin, X.; Lan, C.; Wang, X.; Zeng, W.; Wang, X.; Chen, Z. Diffusion models for image restoration and enhancement: a comprehensive survey. *International Journal of Computer Vision* **2025**, *133*, 8078–8108.
103. Cheng, J.; Liang, D.; Tan, S. Transfer clip for generalizable image denoising. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 25974–25984.
104. Radford, A.; Kim, J.W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; et al. Learning transferable visual models from natural language supervision. In Proceedings of the International conference on machine learning. PmLR, 2021, pp. 8748–8763.
105. Yang, H.; Pan, L.; Yang, Y.; Hartley, R.; Liu, M. Ldp: Language-driven dual-pixel image defocus deblurring network. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 24078–24087.
106. Zhou, H.; Dong, W.; Liu, X.; Zhang, Y.; Zhai, G.; Chen, J. Low-light image enhancement via generative perceptual priors. In Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence, 2025, Vol. 39, pp. 10752–10760.
107. Duan, H.; Min, X.; Wu, S.; Shen, W.; Zhai, G. Uniprocessor: a text-induced unified low-level image processor. In Proceedings of the European Conference on Computer Vision. Springer, 2024, pp. 180–199.
108. Mao, J.; Yang, Y.; Yin, X.; Shao, L.; Tang, H. AllRestorer: All-in-One Transformer for Image Restoration under Composite Degradations. *arXiv preprint arXiv:2411.10708* **2024**.
109. Wang, C.; Fan, H.; Yang, H.; Karimi, S.; Yao, L.; Yang, Y. Adapting Text-to-Image Generation with Feature Difference Instruction for Generic Image Restoration. In Proceedings of the Proceedings of the Computer Vision and Pattern Recognition Conference, 2025, pp. 23539–23550.
110. Li, J.; Li, D.; Savarese, S.; Hoi, S. Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. In Proceedings of the International conference on machine learning. PMLR, 2023, pp. 19730–19742.
111. Dong, W.; Zhou, H.; Ji, T.; Zhao, G.; Zhang, Y.; Zhai, G.; Liu, X.; Chen, J. RectiWeather: Photo-Realistic Adverse Weather Removal via Zero-shot Soft Weather Perception and Rectified Flow **2026**.
112. Wu, Z.; Chen, Y.; Yokoya, N.; He, W. MP-HSIR: A Multi-Prompt Framework for Universal Hyperspectral Image Restoration. *arXiv preprint arXiv:2503.09131* **2025**.
113. Lee, C.M.; Cheng, C.H.; Lin, Y.F.; Cheng, Y.C.; Liao, W.T.; Hsu, C.C.; Yang, F.E.; Wang, Y.C.F. Promptthsi: Universal hyperspectral image restoration framework for composite degradation. *arXiv e-prints* **2024**, pp. arXiv–2411.
114. Luo, Z.; Gustafsson, F.K.; Zhao, Z.; Sjölund, J.; Schön, T.B. Controlling vision-language models for multi-task image restoration. *arXiv preprint arXiv:2310.01018* **2023**.
115. Jiang, Y.; Zhang, Z.; Xue, T.; Gu, J. Autodir: Automatic all-in-one image restoration with latent diffusion. In Proceedings of the European Conference on Computer Vision. Springer, 2024, pp. 340–359.

 doi:10.20944/preprints202603.2366.v2

116. Qi, C.; Tu, Z.; Ye, K.; Delbracio, M.; Milanfar, P.; Chen, Q.; Talebi, H. Spire: Semantic prompt-driven image restoration. In Proceedings of the European Conference on Computer Vision. Springer, 2024, pp. 446–464.
117. 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 Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 25432–25444.
118. Zhang, Z.; Lei, J.; Peng, B.; Zhu, J.; Xu, L.; Huang, Q. Advancing Real-World Stereoscopic Image Super-Resolution via Vision-Language Model. *IEEE Transactions on Image Processing* **2025**.
119. Hu, B.; Liu, H.; Zheng, Z.; Liu, P. CLIP-SR: Collaborative Linguistic and Image Processing for Super-Resolution. *IEEE Transactions on Multimedia* **2025**.
120. Sun, X.; Wang, L.; Wang, C.; Jin, Y.; Lam, K.m.; Su, Z.; Yang, Y.; Pan, J. Adapting Large VLMs with Iterative and Manual Instructions for Generative Low-light Enhancement. *arXiv preprint arXiv:2507.18064* **2025**.
121. Tu, Y.; Yan, Q.; Niu, A.; Tang, J. TPGDiff: Hierarchical Triple-Prior Guided Diffusion for Image Restoration. *arXiv preprint arXiv:2601.20306* **2026**.
122. Hugging Face. Introducing IDEFICS: An Open Reproduction of State-of-the-Art Visual Language Models. https://huggingface.co/blog/idefics, 2023.
123. Yu, Y.; Zeng, Z.; Hua, H.; Fu, J.; Luo, J. Promptfix: You prompt and we fix the photo. *arXiv preprint arXiv:2405.16785* **2024**.
124. Chen, Z.; Chen, T.; Wang, C.; Gao, Q.; Niu, C.; Wang, G.; Shan, H. Low-dose CT denoising with language-engaged dual-space alignment. In Proceedings of the 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 2024, pp. 3088–3091.
125. Chen, Z.; Chen, T.; Wang, C.; Gao, Q.; Xie, H.; Niu, C.; Wang, G.; Shan, H. LangMamba: A Language-driven Mamba Framework for Low-dose CT Denoising with Vision-language Models. *IEEE Transactions on Radiation and Plasma Medical Sciences* **2025**.
126. Yang, S.; Ding, M.; Wu, Y.; Li, Z.; Zhang, J. Implicit neural representation for cooperative low-light image enhancement. In Proceedings of the Proceedings of the IEEE/CVF international conference on computer vision, 2023, pp. 12918–12927.
127. Li, B.; Li, X.; Zhu, H.; Jin, Y.; Feng, R.; Zhang, Z.; Chen, Z. Sed: Semantic-aware discriminator for image super-resolution. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2024, pp. 25784–25795.
128. Song, W.; Liu, C.; Di Mauro, M.; Liotta, A. Unsupervised Underwater Image Enhancement Combining Imaging Restoration and Prompt Learning. In Proceedings of the Chinese Conference on Pattern Recognition and Computer Vision (PRCV). Springer, 2024, pp. 421–434.
129. Li, Y.; Fan, H.; Hu, R.; Feichtenhofer, C.; He, K. Scaling language-image pre-training via masking. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2023, pp. 23390–23400.
130. Zhang, X.; Ma, J.; Wang, G.; Zhang, Q.; Zhang, H.; Zhang, L. Perceive-ir: Learning to perceive degradation better for all-in-one image restoration. *IEEE Transactions on Image Processing* **2025**.
131. Wang, T.; Xia, P.; Li, B.; Jiang, P.T.; Kong, Z.; Zhang, K.; Lu, T.; Luo, W. MOERL: When Mixture-of-Experts Meet Reinforcement Learning for Adverse Weather Image Restoration. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision, 2025, pp. 13673–13683.
132. Lu, J.; Wu, Y.; Zhao, Z.; Wang, H.; Jimenez, F.; Majeedi, A.; Fu, Y. SimpleCall: A Lightweight Image Restoration Agent in Label-Free Environments with MLLM Perceptual Feedback. *arXiv preprint arXiv:2512.18599* **2025**.
133. 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. *Advances in Neural Information Processing Systems* **2024**, *37*, 110643–110666.
134. Zhou, Y.; Cao, J.; Zhang, Z.; Wen, F.; Jiang, Y.; Jia, J.; Liu, X.; Min, X.; Zhai, G. Q-Agent: Quality-Driven Chain-of-Thought Image Restoration Agent through Robust Multimodal Large Language Model. *arXiv preprint arXiv:2504.07148* **2025**.
135. Wei, J.; Wang, X.; Schuurmans, D.; Bosma, M.; Xia, F.; Chi, E.; Le, Q.V.; Zhou, D.; et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems* **2022**, *35*, 24824–24837.
136. Jiang, X.; Li, G.; Chen, B.; Zhang, J. Multi-Agent Image Restoration. *arXiv preprint arXiv:2503.09403* **2025**.
137. Li, B.; Li, X.; Lu, Y.; Chen, Z. Hybrid agents for image restoration. *arXiv preprint arXiv:2503.10120* **2025**.

 doi:10.20944/preprints202603.2366.v2

138. 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**.
139. Franzen, R. Kodak lossless true color image suite, 1999.
140. Zhang, L.; Wu, X.; Buades, A.; Li, X. Color demosaicking by local directional interpolation and nonlocal adaptive thresholding. *Journal of Electronic imaging* **2011**, *20*, 023016–023016.
141. Martin, D.; Fowlkes, C.; Tal, D.; Malik, J. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In Proceedings of the Proceedings eighth IEEE international conference on computer vision. ICCV 2001. IEEE, 2001, Vol. 2, pp. 416–423.
142. Huang, J.B.; Singh, A.; Ahuja, N. Single image super-resolution from transformed self-exemplars. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 5197–5206.
143. Agustsson, E.; Timofte, R. Ntire 2017 challenge on single image super-resolution: Dataset and study. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2017, pp. 126–135.
144. Abdelhamed, A.; Lin, S.; Brown, M.S. A high-quality denoising dataset for smartphone cameras. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 1692–1700.
145. Ma, K.; Duanmu, Z.; Wu, Q.; Wang, Z.; Yong, H.; Li, H.; Zhang, L. Waterloo exploration database: New challenges for image quality assessment models. *IEEE Transactions on Image Processing* **2016**, *26*, 1004–1016.
146. Arbelaez, P.; Maire, M.; Fowlkes, C.; Malik, J. Contour detection and hierarchical image segmentation. *IEEE transactions on pattern analysis and machine intelligence* **2010**, *33*, 898–916.
147. Yang, W.; Tan, R.T.; Feng, J.; Liu, J.; Guo, Z.; Yan, S. Deep joint rain detection and removal from a single image. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1357–1366.
148. Zhang, H.; Sindagi, V.; Patel, V.M. Image de-raining using a conditional generative adversarial network. *IEEE transactions on circuits and systems for video technology* **2019**, *30*, 3943–3956.
149. Fu, X.; Huang, J.; Zeng, D.; Huang, Y.; Ding, X.; Paisley, J. Removing rain from single images via a deep detail network. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 3855–3863.
150. Qian, R.; Tan, R.T.; Yang, W.; Su, J.; Liu, J. Attentive generative adversarial network for raindrop removal from a single image. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 2482–2491.
151. Li, R.; Cheong, L.F.; Tan, R.T. Heavy rain image restoration: Integrating physics model and conditional adversarial learning. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019, pp. 1633–1642.
152. Quan, R.; Yu, X.; Liang, Y.; Yang, Y. Removing raindrops and rain streaks in one go. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021, pp. 9147–9156.
153. Huang, H.; Luo, M.; He, R. Memory uncertainty learning for real-world single image deraining. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2022**, *45*, 3446–3460.
154. Sakaridis, C.; Dai, D.; Van Gool, L. Semantic foggy scene understanding with synthetic data. *International Journal of Computer Vision* **2018**, *126*, 973–992.
155. Sakaridis, C.; Wang, H.; Li, K.; Zurbrügg, R.; Jadon, A.; Abbeloos, W.; Reino, D.O.; Van Gool, L.; Dai, D. ACDC: The adverse conditions dataset with correspondences for robust semantic driving scene perception. *arXiv preprint arXiv:2104.13395* **2021**.
156. Li, B.; Ren, W.; Fu, D.; Tao, D.; Feng, D.; Zeng, W.; Wang, Z. Benchmarking single-image dehazing and beyond. *IEEE transactions on image processing* **2018**, *28*, 492–505.
157. Ancuti, C.O.; Ancuti, C.; Sbert, M.; Timofte, R. Dense-haze: A benchmark for image dehazing with dense-haze and haze-free images. In Proceedings of the 2019 IEEE international conference on image processing (ICIP). IEEE, 2019, pp. 1014–1018.
158. Punnappurath, A.; Abuolaim, A.; Afifi, M.; Brown, M.S. Modeling defocus-disparity in dual-pixel sensors. In Proceedings of the 2020 IEEE International Conference on Computational Photography (ICCP). IEEE, 2020, pp. 1–12.
159. Pan, L.; Chowdhury, S.; Hartley, R.; Liu, M.; Zhang, H.; Li, H. Dual pixel exploration: Simultaneous depth estimation and image restoration. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 4340–4349.

 doi:10.20944/preprints202603.2366.v2

160. Abuolaim, A.; Delbracio, M.; Kelly, D.; Brown, M.S.; Milanfar, P. Learning to reduce defocus blur by realistically modeling dual-pixel data. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 2289–2298.
161. Lee, C.; Lee, C.; Kim, C.S. Contrast enhancement based on layered difference representation of 2D histograms. *IEEE transactions on image processing* **2013**, *22*, 5372–5384.
162. Wang, S.; Zheng, J.; Hu, H.M.; Li, B. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE transactions on image processing* **2013**, *22*, 3538–3548.
163. Vonikakis, V.; Kouskouridas, R.; Gasteratos, A. On the evaluation of illumination compensation algorithms. *Multimedia Tools and Applications* **2018**, *77*, 9211–9231.
164. Ma, K.; Zeng, K.; Wang, Z. Perceptual quality assessment for multi-exposure image fusion. *IEEE Transactions on Image Processing* **2015**, *24*, 3345–3356.
165. Cai, J.; Gu, S.; Zhang, L. Learning a deep single image contrast enhancer from multi-exposure images. *IEEE Transactions on Image Processing* **2018**, *27*, 2049–2062.
166. Guo, X.; Li, Y.; Ling, H. LIME: Low-light image enhancement via illumination map estimation. *IEEE Transactions on image processing* **2016**, *26*, 982–993.
167. Bevilacqua, M.; Roumy, A.; Guillemot, C.; Alberi-Morel, M.L. Low-complexity single-image super-resolution based on nonnegative neighbor embedding **2012**.
168. Zeyde, R.; Elad, M.; Protter, M. On single image scale-up using sparse-representations. In Proceedings of the International conference on curves and surfaces. Springer, 2010, pp. 711–730.
169. Matsui, Y.; Ito, K.; Aramaki, Y.; Fujimoto, A.; Ogawa, T.; Yamasaki, T.; Aizawa, K. Sketch-based manga retrieval using manga109 dataset. *Multimedia tools and applications* **2017**, *76*, 21811–21838.
170. Cheng, D.; Price, B.; Cohen, S.; Brown, M.S. Beyond white: Ground truth colors for color constancy correction. In Proceedings of the Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 298–306.
171. Cai, J.; Zeng, H.; Yong, H.; Cao, Z.; Zhang, L. Toward real-world single image super-resolution: A new benchmark and a new model. In Proceedings of the Proceedings of the IEEE/CVF international conference on computer vision, 2019, pp. 3086–3095.
172. Wei, P.; Xie, Z.; Lu, H.; Zhan, Z.; Ye, Q.; Zuo, W.; Lin, L. Component divide-and-conquer for real-world image super-resolution. In Proceedings of the European conference on computer vision. Springer, 2020, pp. 101–117.
173. Wu, R.; Yang, T.; Sun, L.; Zhang, Z.; Li, S.; Zhang, L. Seesr: Towards semantics-aware real-world image super-resolution. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2024, pp. 25456–25467.
174. Zhang, K.; Liang, J.; Van Gool, L.; Timofte, R. Designing a practical degradation model for deep blind image super-resolution. In Proceedings of the Proceedings of the IEEE/CVF international conference on computer vision, 2021, pp. 4791–4800.
175. Wang, Z.J.; Montoya, E.; Munechika, D.; Yang, H.; Hoover, B.; Chau, D.H. Diffusiondb: A large-scale prompt gallery dataset for text-to-image generative models. In Proceedings of the Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2023, pp. 893–911.
176. Xia, G.S.; Hu, J.; Hu, F.; Shi, B.; Bai, X.; Zhong, Y.; Zhang, L.; Lu, X. AID: A benchmark data set for performance evaluation of aerial scene classification. *IEEE Transactions on Geoscience and Remote Sensing* **2017**, *55*, 3965–3981.
177. Li, K.; Wan, G.; Cheng, G.; Meng, L.; Han, J. Object detection in optical remote sensing images: A survey and a new benchmark. *ISPRS journal of photogrammetry and remote sensing* **2020**, *159*, 296–307.
178. Xia, G.S.; Bai, X.; Ding, J.; Zhu, Z.; Belongie, S.; Luo, J.; Datcu, M.; Pelillo, M.; Zhang, L. DOTA: A large-scale dataset for object detection in aerial images. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 3974–3983.
179. Jia, F.; Tan, L.; Wang, G.; Jia, C.; Chen, Z. A super-resolution network using channel attention retention for pathology images. *PeerJ Computer Science* **2023**, *9*, e1196.
180. FUJIFILM Healthcare Europe.; SonoSkills. US-CASE: Ultrasound Cases Dataset, 2025.
181. Liu, R.; Fan, X.; Zhu, M.; Hou, M.; Luo, Z. Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light. *IEEE transactions on circuits and systems for video technology* **2020**, *30*, 4861–4875.
182. Tang, L.; Yuan, J.; Zhang, H.; Jiang, X.; Ma, J. PIAFusion: A progressive infrared and visible image fusion network based on illumination aware. *Information Fusion* **2022**, *83*, 79–92.

 doi:10.20944/preprints202603.2366.v2

183. Liu, J.; Liu, Z.; Wu, G.; Ma, L.; Liu, R.; Zhong, W.; Luo, Z.; Fan, X. Multi-interactive feature learning and a full-time multi-modality benchmark for image fusion and segmentation. In Proceedings of the Proceedings of the IEEE/CVF international conference on computer vision, 2023, pp. 8115–8124.
184. Guo, Y.; Gao, Y.; Lu, Y.; Zhu, H.; Liu, R.W.; He, S. Onerestore: A universal restoration framework for composite degradation. In Proceedings of the European conference on computer vision. Springer, 2024, pp. 255–272.
185. Zhou, Y.; Ren, D.; Emerton, N.; Lim, S.; Large, T. Image restoration for under-display camera. In Proceedings of the Proceedings of the ieee/cvf conference on computer vision and pattern recognition, 2021, pp. 9179–9188.
186. Lin, C.H.; Hsu, C.C.; Young, S.S.; Hsieh, C.Y.; Tai, S.C. QRCODE: Quasi-residual convex deep network for fusing misaligned hyperspectral and multispectral images. *IEEE Transactions on Geoscience and Remote Sensing* **2024**, *62*, 1–15.
187. Arad, B.; Timofte, R.; Yahel, R.; Morag, N.; Bernat, A.; Cai, Y.; Lin, J.; Lin, Z.; Wang, H.; Zhang, Y.; et al. Ntire 2022 spectral recovery challenge and data set. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 863–881.
188. Zhang, L.; Zhang, L.; Mou, X.; Zhang, D. FSIM: A feature similarity index for image quality assessment. *IEEE transactions on Image Processing* **2011**, *20*, 2378–2386.
189. Du, Q.; Younan, N.H.; King, R.; Shah, V.P. On the performance evaluation of pan-sharpening techniques. *IEEE Geoscience and Remote Sensing Letters* **2007**, *4*, 518–522.
190. Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; Hochreiter, S. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems* **2017**, *30*.
191. Blau, Y.; Mechrez, R.; Timofte, R.; Michaeli, T.; Zelnik-Manor, L. The 2018 PIRM challenge on perceptual image super-resolution. In Proceedings of the Proceedings of the European conference on computer vision (ECCV) workshops, 2018, pp. 0–0.
192. Ying, Z.; Niu, H.; Gupta, P.; Mahajan, D.; Ghadiyaram, D.; Bovik, A. From patches to pictures (PaQ-2-PiQ): Mapping the perceptual space of picture quality. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 3575–3585.
193. Network, A. Blind image quality assessment using a deep bilinear convolutional neural network. *Deep Bilinear Convolutional Neural* **2022**, *5*.
194. Zhang, W.; Zhai, G.; Wei, Y.; Yang, X.; Ma, K. Blind image quality assessment via vision-language correspondence: A multitask learning perspective. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2023, pp. 14071–14081.
195. Zhou, J.; Liu, C.; Jiang, Q.; Fu, X.; Hou, J.; Li, X. Semantic Contrast for Domain-Robust Underwater Image Quality Assessment. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2026**.
196. Chen, Z.; Qin, H.; Wang, J.; Yuan, C.; Li, B.; Hu, W.; Wang, L. Promptiqa: Boosting the performance and generalization for no-reference image quality assessment via prompts. In Proceedings of the European conference on computer vision. Springer, 2024, pp. 247–264.
197. Li, X.; Huang, Z.; Zhang, Y.; Shen, Y.; Li, K.; Zheng, X.; Cao, L.; Ji, R. Few-Shot Image Quality Assessment via Adaptation of Vision-Language Models. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision, 2025, pp. 10442–10452.
198. Kwon, D.; Kim, D.; Ki, S.; Jo, Y.; Lee, H.E.; Kim, S.J. ATTIQA: Generalizable image quality feature extractor using attribute-aware pretraining. In Proceedings of the Proceedings of the Asian Conference on Computer Vision, 2024, pp. 4526–4543.
199. Rifa, K.R.; Zhang, J.; Imran, A. CAP-IQA: Context-Aware Prompt-Guided CT Image Quality Assessment. *arXiv preprint arXiv:2601.01613* **2026**.
200. Dong, G.; Liao, X.; Li, M.; Guo, G.; Ren, C. Exploring semantic feature discrimination for perceptual image super-resolution and opinion-unaware no-reference image quality assessment. In Proceedings of the Proceedings of the Computer Vision and Pattern Recognition Conference, 2025, pp. 28176–28187.
201. Zhou, H.; Tang, L.; Yang, R.; Qin, G.; Zhang, Y.; Li, Y.; Li, X.; Hu, R.; Zhai, G. UniQA: Unified vision-language pre-training for image quality and aesthetic assessment. *arXiv preprint arXiv:2406.01069* **2024**.
202. Zhao, S.; Zhang, X.; Li, W.; Li, J.; Zhang, L.; Xue, T.; Zhang, J. Reasoning as Representation: Rethinking Visual Reinforcement Learning in Image Quality Assessment. *arXiv preprint arXiv:2510.11369* **2025**.
203. You, Z.; Gu, J.; Li, Z.; Cai, X.; Zhu, K.; Dong, C.; Xue, T. Descriptive image quality assessment in the wild. *arXiv preprint arXiv:2405.18842* **2024**.

 doi:10.20944/preprints202603.2366.v2

204. Chen, Z.; Hu, B.; Niu, C.; Chen, T.; Li, Y.; Shan, H.; Wang, G. IQAGPT: computed tomography image quality assessment with vision-language and ChatGPT models. *Visual Computing for Industry, Biomedicine, and Art* **2024**, *7*, 20.
205. Wu, H.; Zhu, H.; Zhang, Z.; Zhang, E.; Chen, C.; Liao, L.; Li, C.; Wang, A.; Sun, W.; Yan, Q.; et al. Towards open-ended visual quality comparison. In Proceedings of the European Conference on Computer Vision. Springer, 2024, pp. 360–377.
206. Liu, K.; Zhang, Z.; Li, W.; Pei, R.; Song, F.; Liu, X.; Kong, L.; Zhang, Y. Dog-IQA: Standard-guided Zero-shot MLLM for Mix-grained Image Quality Assessment. *arXiv preprint arXiv:2410.02505* **2024**.
207. Li, W.; Zhang, X.; Zhao, S.; Zhang, Y.; Li, J.; Zhang, L.; Zhang, J. Q-insight: Understanding image quality via visual reinforcement learning. *arXiv preprint arXiv:2503.22679* **2025**.
208. Cai, Z.; Zhang, J.; Yuan, X.; Jiang, P.T.; Chen, W.; Tang, B.; Yao, L.; Wang, Q.; Chen, J.; Li, B. Q-ponder: A unified training pipeline for reasoning-based visual quality assessment. *arXiv preprint arXiv:2506.05384* **2025**.
209. Xie, W.; Dai, R.; Ding, R.; Liu, K.; Chu, X.; Hou, X.; Wen, J. Q-Hawkeye: Reliable Visual Policy Optimization for Image Quality Assessment. *arXiv preprint arXiv:2601.22920* **2026**.
210. Li, X.; Zhang, Z.; Xu, Z.; Xu, S.; Min, X.; Chen, Y.; Zhai, G. Decoupling Perception and Calibration: Label-Efficient Image Quality Assessment Framework. *arXiv preprint arXiv:2601.20689* **2026**.
211. Wu, H.; Zhang, Z.; Zhang, E.; Chen, C.; Liao, L.; Wang, A.; Li, C.; Sun, W.; Yan, Q.; Zhai, G.; et al. Q-bench: A benchmark for general-purpose foundation models on low-level vision. *arXiv preprint arXiv:2309.14181* **2023**.
212. Zhang, Z.; Wu, H.; Zhang, E.; Zhai, G.; Lin, W. Q-Bench $^+$: A Benchmark for Multi-Modal Foundation Models on Low-Level Vision From Single Images to Pairs. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2024**, *46*, 10404–10418.
213. Kendall, M.G. A new measure of rank correlation. *Biometrika* **1938**, *30*, 81–93.
214. Tinsley, H.E.; Weiss, D.J. Interrater reliability and agreement of subjective judgments. *Journal of Counseling Psychology* **1975**, *22*, 358.
215. Papineni, K.; Roukos, S.; Ward, T.; Zhu, W.J. Bleu: a method for automatic evaluation of machine translation. In Proceedings of the Proceedings of the 40th annual meeting of the Association for Computational Linguistics, 2002, pp. 311–318.
216. Lin, C.Y. Rouge: A package for automatic evaluation of summaries. In Proceedings of the Text summarization branches out, 2004, pp. 74–81.
217. Banerjee, S.; Lavie, A. METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. In Proceedings of the Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization, 2005, pp. 65–72.
218. Vedantam, R.; Lawrence Zitnick, C.; Parikh, D. Cider: Consensus-based image description evaluation. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 4566–4575.
219. Sun, S.; Ren, W.; Wang, T.; Cao, X. Rethinking image restoration for object detection. *Advances in Neural Information Processing Systems* **2022**, *35*, 4461–4474.
220. Zhu, H.; Sui, X.; Chen, B.; Liu, X.; Chen, P.; Fang, Y.; Wang, S. 2AFC prompting of large multimodal models for image quality assessment. *IEEE Transactions on Circuits and Systems for Video Technology* **2024**.
221. Li, B.; Liu, X.; Hu, P.; Wu, Z.; Lv, J.; Peng, X. All-in-one image restoration for unknown corruption. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022, pp. 17452–17462.
222. Zhang, J.; Huang, J.; Yao, M.; Yang, Z.; Yu, H.; Zhou, M.; Zhao, F. Ingredient-oriented multi-degradation learning for image restoration. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2023, pp. 5825–5835.
223. Cui, Y.; Zamir, S.W.; Khan, S.; Knoll, A.; Shah, M.; Khan, F.S. Adair: Adaptive all-in-one image restoration via frequency mining and modulation. In Proceedings of the 13th international conference on learning representations, ICLR 2025. International Conference on Learning Representations, ICLR, 2025, pp. 57335–57356.
224. Wu, G.; Jiang, J.; Jiang, K.; Liu, X.; Nie, L. DSwinIR: Rethinking Window-Based Attention for Image Restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2025**.
225. Cui, Y.; Ren, W.; Shi, B.; Knoll, A. Visual-in-Visual: A Unified and Efficient Baseline for Image Restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2026**.

 doi:10.20944/preprints202603.2366.v2

226. Wang, T.; Zhang, K.; Shen, T.; Luo, W.; Stenger, B.; Lu, T. Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method. In Proceedings of the Proceedings of the AAAI conference on artificial intelligence, 2023, Vol. 37, pp. 2654–2662.
227. Zhang, T.; Liu, P.; Lu, Y.; Cai, M.; Zhang, Z.; Zhang, Z.; Zhou, Q. Cwnet: Causal wavelet network for low-light image enhancement. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision, 2025, pp. 8789–8799.
228. Yi, X.; Xu, H.; Zhang, H.; Tang, L.; Ma, J. Diff-Retinex++: Retinex-driven reinforced diffusion model for low-light image enhancement. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2025**.
229. Deng, X.; Dragotti, P.L. Deep convolutional neural network for multi-modal image restoration and fusion. *IEEE transactions on pattern analysis and machine intelligence* **2020**, *43*, 3333–3348.
230. Wang, Z.; Wu, Y.; Li, D.; Li, G.; Zhu, P.; Zhang, Z.; Jiang, R. LiDAR-assisted image restoration for extreme low-light conditions. *Knowledge-Based Systems* **2025**, *316*, 113382.
231. Janjua, M.K.; Ghasemabadi, A.; Zhang, K.; Salameh, M.; Gao, C.; Niu, D. Grounding Degradations in Natural Language for All-In-One Video Restoration. *arXiv preprint arXiv:2507.14851* **2025**.