Multi-Agent Collaborative Spatially-Aware Image Restoration: A Comprehensive Review

2026 年 2 月 26 日

Abstract

Image restoration, the task of recovering high-quality images from degraded observations corrupted by noise, blur, haze, rain, and compression artifacts, has witnessed transformative advances driven by deep learning. Early convolutional approaches gave way to Transformer-based architectures that capture long-range dependencies, while all-in-one models have attempted to handle multiple degradation types within a single framework. More recently, diffusion-based generative methods have demonstrated remarkable perceptual quality in severely degraded scenarios. Despite this progress, most existing approaches treat the entire image uniformly and employ static inference pipelines, fundamentally limiting their ability to address real-world images where degradation varies spatially in both type and severity. A nascent but rapidly growing paradigm addresses this limitation by leveraging large language models and vision-language models as intelligent agents that dynamically perceive degradation, plan restoration strategies, and orchestrate specialized tools. This review systematically examines the evolution from single-task restoration through unified models to agent-based systems, with particular emphasis on spatially-aware processing and multi-agent collaboration. We identify key challenges including the trade-off between restoration quality and computational efficiency, the gap between synthetic and real-world degradation, and the pressing need for lightweight deployment strategies such as trajectory distillation. By synthesizing findings across 123 representative studies, we delineate emerging research directions and highlight the potential of spatially-aware multi-agent frameworks to establish a new paradigm for intelligent, adaptive image restoration.

1 Introduction

Image restoration encompasses a family of inverse problems aimed at recovering a clean image $\mathbf{x}$ from a degraded observation $\mathbf{y} = \mathcal{D}(\mathbf{x}) + \mathbf{n}$, where $\mathcal{D}(\cdot)$ denotes the degradation operator and $\mathbf{n}$ represents additive noise. This problem is fundamental to computer vision, as the quality of input images directly affects the performance of downstream tasks such as object detection, semantic segmentation, and autonomous navigation. In practical scenarios—including surveillance

under adverse weather, satellite remote sensing, medical imaging, and consumer photography—images suffer from diverse and often co-occurring degradations: Gaussian and Poisson noise from sensor limitations, motion and defocus blur from camera or scene dynamics, haze and rain from atmospheric interference, and compression artifacts from lossy encoding.

The past decade has seen dramatic improvements in restoration quality, propelled by the transition from hand-crafted priors to learned representations. Convolutional neural networks (CNNs) first demonstrated that end-to-end learning could surpass classical methods by significant margins. Subsequently, Transformer architectures revolutionized the field by enabling global context modeling through self-attention mechanisms. Liang et al.[1] introduced SwinIR, which leverages Swin Transformer blocks with shifted window attention to achieve state-of-the-art results across denoising, super-resolution, and JPEG artifact removal. Zamir et al.[2] proposed Restormer, employing transposed attention to efficiently process high-resolution images while capturing long-range pixel interactions. These methods established Transformer-based designs as the dominant backbone for image restoration.

However, a critical limitation persists: most methods are designed and trained for a single degradation type, requiring separate models for denoising, deblurring, dehazing, and deraining. This paradigm is impractical for real-world deployment where degradation types are unknown a priori and often coexist within a single image. To address this, all-in-one (AiO) restoration methods have emerged as a significant research direction. Li et al.[3] proposed AirNet, using contrastive learning to build degradation-aware representations within a unified framework. Potlapalli et al.[4] introduced PromptIR, which employs learnable prompt vectors to implicitly encode degradation-specific information. Conde et al.[5] developed InstructIR, the first method to leverage natural language instructions for guiding the restoration process. While these unified models represent substantial progress, they remain fundamentally constrained by the set of degradation types seen during training and process the entire image through a single static pipeline.

A parallel development has been the adoption of diffusion probabilistic models for image restoration. Methods such as DiffBIR[6] and Diff-Retinex[7] leverage the powerful generative priors learned by diffusion models to produce perceptually realistic restorations, particularly for severely degraded inputs where deterministic methods tend to produce oversmoothed results. Li et al.[8] provided a comprehensive survey of diffusion-based restoration approaches, highlighting their ability to model complex image distributions. Nevertheless, diffusion models typically require many iterative denoising steps, resulting in high computational costs that limit real-time applications.

Most critically, all the aforementioned approaches—whether single-task, all-in-one, or diffusion-based—share a fundamental assumption: the entire image should be processed uniformly with the same restoration strategy. In reality, degradation is spatially heterogeneous. A single photograph may exhibit motion blur in the foreground, haze in the background, noise in shadow

regions, and rain streaks across the scene. Applying a single restoration pipeline globally inevitably leads to suboptimal results: regions that are lightly degraded may be over-processed, while severely degraded areas may be insufficiently restored.

This observation has motivated the emergence of agent-based image restoration, a paradigm that represents a fundamental departure from traditional approaches. Zhu et al.[9] proposed AgenticIR, a pioneering framework that employs vision-language models (VLMs) as intelligent agents to perceive image degradation, plan restoration strategies, and orchestrate specialized tools through a five-stage pipeline of perception, scheduling, execution, reflection, and rescheduling. Chen et al.[10] developed RestoreAgent, which fine-tunes a multimodal large language model to directly predict restoration workflows. Jiang et al.[11] introduced MAIR (Multi-Agent Image Restoration), employing a multi-agent architecture with degradation priors that improves inference efficiency by 44%. Zhou et al.[12] proposed Q-Agent, which uses quality-driven greedy strategies to achieve linear-complexity restoration planning. These methods open the door to spatially-aware, adaptive restoration that can dynamically adjust strategies based on the specific degradation characteristics of different image regions.

This review provides a systematic examination of the evolution from classical single-task restoration to the emerging paradigm of multi-agent spatially-aware image restoration. We organize the literature into six thematic areas: (1) Transformer-based single-task restoration architectures, (2) all-in-one unified restoration models, (3) diffusion-based generative restoration, (4) agent-based restoration systems, (5) foundation models for degradation analysis and quality assessment, and (6) model compression and lightweight deployment. For each area, we analyze key methodological innovations, compare representative approaches, identify limitations, and discuss open challenges. We conclude by outlining future research directions, with particular emphasis on how spatially-aware multi-agent collaboration can address the remaining gaps between current methods and the demands of real-world deployment.

2 Transformer-Based Single-Task Image Restoration

The introduction of Transformer architectures into image restoration marked a paradigm shift from purely local convolutional operations to global context modeling through self-attention. This section examines the key architectural innovations that have made Transformers the dominant backbone for restoration tasks, analyzes the efficiency–quality trade-offs inherent in different attention designs, and identifies the limitations that motivate the transition toward unified and agent-based approaches.

2.1 Window-Based Attention Architectures

The computational cost of standard self-attention scales quadratically with the number of pixels, making it prohibitively expensive for high-resolution image restoration. Window-based attention mechanisms address this challenge by restricting attention computation to local windows while

maintaining the ability to capture long-range dependencies through window shifting or overlapping strategies. SwinIR[1] adapts the Swin Transformer architecture to image restoration by stacking residual Swin Transformer blocks (RSTBs), each containing multiple Swin Transformer layers with shifted window multi-head self-attention. This design achieves linear computational complexity with respect to image size while maintaining competitive or superior performance to CNN-based methods across denoising (achieving 0.14–0.31 dB PSNR improvement over prior art on benchmark datasets), classical and lightweight super-resolution, and JPEG compression artifact reduction. The shifted window mechanism enables cross-window connections that effectively propagate information across the entire feature map without the quadratic cost of global attention.

Wu et al.[13] revisited the window-based attention paradigm in DSwinIR, demonstrating that the original shifted window strategy in SwinIR introduces boundary artifacts at window borders and proposed dynamic window attention with learnable offsets that adapt window positions to image content. This content-adaptive windowing strategy yields consistent improvements over SwinIR across multiple restoration tasks, suggesting that the fixed grid pattern of standard shifted windows is suboptimal for spatially varying image content. Conde et al.[14] extended the Swin Transformer to its second version in Swin2SR, incorporating SwinV2's improvements including cosine attention and log-spaced continuous position bias, demonstrating enhanced stability for compressed image super-resolution and achieving competitive performance with reduced training instability. The original Swin2SR work[15] showed that these architectural refinements are particularly beneficial for handling the diverse artifact patterns introduced by different JPEG quality factors, as the improved attention mechanism provides more stable gradient flow during training on heavily compressed images. Jing et al.[16] further explored the lightweight super-resolution direction by proposing a parallel connection of convolution and Swin Transformer blocks, where the convolutional branch captures fine-grained local textures while the Transformer branch models global structural dependencies. The parallel design avoids the sequential bottleneck of cascade architectures and reduces parameter count through shared intermediate representations, achieving competitive super-resolution quality with significantly fewer parameters—an important consideration for deploying window-based Transformer restoration on resource-constrained devices.

2.2 Channel and Transposed Attention Designs

An alternative approach to reducing self-attention complexity operates along the channel dimension rather than the spatial dimension. Restormer[2] introduces multi-Dconv head transposed attention (MDTA), which computes attention across channels rather than spatial locations, thereby achieving linear complexity with respect to spatial resolution while preserving the ability to model global context. This design is complemented by a gated-Dconv feed-forward network (GDFN) that applies depth-wise convolutions to introduce local context. Restormer achieves state-of-the-art results on image denoising (0.13–0.52 dB improvement on the SIDD

benchmark[2]), single-image motion deblurring, defocus deblurring, and deraining, establishing transposed attention as a highly effective alternative to spatial attention for restoration.

Wang et al.[17] proposed Uformer, a hierarchical U-shaped Transformer that combines window-based self-attention in the encoder–decoder structure with a learnable multi-scale restoration modulator. The modulator adjusts features at each level based on degradation characteristics, enabling the network to adapt its behavior to different spatial scales. Uformer demonstrates strong performance on denoising and deblurring while maintaining computational efficiency through its window-based attention in each level of the U-Net hierarchy.

2.3 Hybrid CNN-Transformer Architectures

Recognizing that CNNs excel at capturing local texture patterns while Transformers are superior at modeling global structure, several works have proposed hybrid architectures that combine both paradigms. Chen et al.[18] developed Dual-former, which employs parallel CNN and Transformer branches with feature exchange mechanisms, allowing the network to simultaneously capture local details and global context. The dual-branch design achieves consistent improvements over single-backbone models, particularly on tasks where both fine texture preservation and global structural coherence are important.

Chen et al.[19] proposed a hybrid CNN-Transformer feature fusion network for single image deraining that uses cross-attention to merge CNN-extracted local features with Transformer-captured global dependencies. Shi et al.[20] introduced Sandformer, which integrates CNN and Transformer under a gated fusion mechanism specifically designed for sand dust image restoration, demonstrating the versatility of hybrid designs for specialized degradation types. The convergence toward hybrid architectures reflects a broader recognition that no single attention mechanism is optimal for all spatial scales and degradation patterns encountered in restoration tasks.

2.4 Efficient Attention and State-Space Models

Beyond window-based and channel-based strategies, researchers have explored alternative efficient attention mechanisms. Kong et al.[21] proposed FFTformer, which replaces spatial self-attention with frequency-domain operations, computing attention in the Fourier domain to achieve global receptive fields with reduced computational cost. This approach is particularly effective for image deblurring, where frequency-domain representations naturally capture the spectral characteristics of blur kernels. Building on this frequency-domain perspective, Jiang et al.[22] investigated the synergy between fast Fourier transforms (FFT) and Transformer architectures for image restoration, demonstrating that interleaving FFT-based global mixing layers with local window attention layers yields a complementary representation: the FFT layers capture periodic and global frequency patterns efficiently, while the window attention layers preserve spatially localized detail. This dual-domain processing achieves superior restoration quality com-

pared to either approach in isolation, particularly for degradation types that manifest in specific frequency bands such as JPEG compression artifacts and periodic noise.

Mao et al.[23] presented intriguing findings regarding frequency selection for image deblurring, revealing that not all frequency components contribute equally to restoration performance. Their analysis shows that selectively attending to mid-frequency bands—which carry the majority of structural information degraded by blur—while suppressing contributions from very low and very high frequencies leads to both improved deblurring quality and reduced computational overhead. This frequency-aware selection strategy can be integrated into existing Transformer backbones as a lightweight plug-in module, providing a principled approach to frequency-domain attention that avoids the computational burden of processing the full spectrum.

Shi et al.[24] introduced VmambaIR, which applies the Visual Mamba state-space model to image restoration. Unlike Transformer attention, which has quadratic complexity, the Mamba architecture achieves linear complexity through selective state-space scanning while maintaining the ability to model long-range dependencies. VmambaIR demonstrates competitive performance with Transformer-based methods at significantly lower computational cost, suggesting that state-space models represent a promising direction for efficient restoration. Similarly, Lee et al.[25] proposed Decomformer, which decomposes self-attention into complementary low-rank and sparse components to reduce redundancy while preserving expressive power for efficient image restoration.

Ghasemabadi et al.[26] introduced CascadedGaze, a method designed to efficiently extract global context for image restoration without incurring the quadratic cost of full self-attention. CascadedGaze employs a cascaded structure where each stage progressively expands its receptive field through gaze-shift operations, aggregating local attention outputs from one stage as contextual tokens for the next. This cascading strategy achieves near-global context modeling with computational cost that scales linearly with spatial resolution, making it particularly suitable for high-resolution restoration tasks where full global attention is impractical. Zhang et al.[27] proposed a joint multi-dimensional dynamic attention mechanism that simultaneously models dependencies across spatial, channel, and scale dimensions. Unlike approaches that treat these dimensions independently, the joint formulation captures cross-dimensional correlations—for instance, the relationship between specific spatial regions and the channel features most relevant to restoring those regions—through a shared dynamic attention kernel. This multi-dimensional approach yields consistent improvements across denoising, deblurring, and super-resolution tasks, suggesting that cross-dimensional attention modeling remains an underexplored avenue for restoration performance gains.

Additional innovations include the Omni-Kernel Network (OKNet)[28], which designs omni-dimensional convolution kernels that can adaptively aggregate features across spatial, channel, and kernel dimensions, and GoLDFormer[29], which introduces global–local deformable window attention that combines deformable attention with window-based processing for flexible receptive

field adaptation. Gao et al.[30] proposed a mixed hierarchy network that integrates features at multiple granularities for comprehensive restoration.

2.5 Self-Supervised and Training Paradigm Innovations

Beyond architectural design, recent work has investigated how the training paradigm itself can be improved for Transformer-based restoration. Zhang and Zhou[31] proposed a self-supervised image denoising method based on a context-aware Transformer, which eliminates the need for paired clean-noisy training data by leveraging blind-spot masking within the Transformer attention mechanism. The context-aware design ensures that each pixel's denoised estimate is computed from its surrounding context without using the noisy pixel itself, satisfying the self-supervised learning constraint while still capturing long-range dependencies. This approach achieves denoising quality competitive with fully supervised Transformer methods on real-world noise benchmarks, demonstrating that the combination of self-supervised learning and Transformer architectures can address the practical challenge of obtaining clean reference images for training.

2.6 Limitations and the Motivation for Unified Approaches

Despite their impressive performance, Transformer-based single-task methods suffer from a fundamental limitation: task specificity. Each model is typically trained for a single degradation type (e.g., denoising at a specific noise level, or deblurring with a specific blur kernel family), requiring practitioners to maintain a library of specialized models and to correctly identify the degradation type before selecting the appropriate model. Ali et al.[32] provided a comprehensive survey of vision Transformers in image restoration, highlighting that while architectural innovations have dramatically improved single-task performance, the proliferation of task-specific models creates significant practical challenges for real-world deployment. Mei et al.[33] and Gao et al.[34] further demonstrated that even within a single task such as deraining, performance varies significantly depending on rain density and scene complexity, underscoring the difficulty of building robust single-task models.

These limitations directly motivate the development of all-in-one restoration methods discussed in the next section, as well as the agent-based approaches that can dynamically select and compose specialized tools based on observed degradation characteristics.

3 All-in-One Unified Image Restoration

All-in-one (AiO) image restoration seeks to handle multiple degradation types with a single model, eliminating the need for degradation-specific architectures. This section traces the evolution of AiO methods from early contrastive learning approaches through prompt-based designs to recent instruction-guided and mixture-of-experts frameworks, analyzing how each addresses the challenge of degradation awareness within a unified architecture.

3.1 Contrastive and Representation Learning Approaches

The foundational challenge of AiO restoration is learning degradation-discriminative representations that enable a single network to distinguish and appropriately handle diverse degradation types. AirNet[3] addresses this through contrastive-based degradation encoder (CBDE) that learns to cluster different degradation types in a compact representation space. By pulling together representations of images with the same degradation type while pushing apart those with different degradations, AirNet enables a single restoration backbone to adaptively process multiple degradation types. Evaluated on denoising, dehazing, and deraining, AirNet demonstrated that contrastive learning can serve as an effective self-supervised mechanism for degradation awareness without requiring explicit degradation labels.

Hu et al.[35] extended the contrastive learning paradigm with collaborative semantic contrastive learning for all-in-one image restoration, proposing a multi-level contrastive framework that simultaneously operates at image-level, feature-level, and patch-level granularities. This hierarchical contrastive design captures both global degradation characteristics and local corruption patterns, yielding improved degradation discrimination compared to single-level contrastive approaches. Wu et al.[36] proposed a task-agnostic model contrastive learning framework that learns from restoration history, constructing positive and negative pairs from different training stages to continuously improve feature representations.

3.2 Prompt-Based Restoration Methods

A major advance in AiO restoration came from adapting the prompt learning paradigm from natural language processing. PromptIR[4] introduces learnable prompt vectors that are injected into a Transformer backbone through prompt blocks, enabling the network to implicitly encode degradation-specific information without explicit degradation identification. The prompts serve as soft conditioning signals that modulate the network's behavior for different degradation types, achieving competitive performance across denoising, dehazing, deraining, and deblurring within a single model. Crucially, PromptIR demonstrates that prompt-based conditioning can be more flexible than explicit degradation classification, as the continuous prompt space can represent intermediate and mixed degradation states.

The prompt paradigm has been extensively developed in subsequent works. Ma et al.[37] proposed ProRes, which explores degradation-aware visual prompts for universal image restoration. Rather than learning prompts from scratch, ProRes constructs visual prompts from degradation-specific prior knowledge, providing more interpretable and effective conditioning. Li et al.[38] introduced Prompt-In-Prompt learning, embedding fine-grained degradation-specific prompts within coarser task-level prompts through a hierarchical prompt architecture that enables both global task discrimination and local degradation adaptation.

Wu et al.[39] developed a dynamic prompting mechanism that generates input-adaptive prompts rather than using fixed learned vectors, allowing the prompt content to vary based on the specific

characteristics of each input image. Sun et al.[40] proposed AdaPrompt-IR, which adaptively learns to perceive degradation semantics and generates degradation-aware prompts through a lightweight degradation perception module, achieving improved generalization to unseen degradation combinations. Liu et al.[41] introduced UP-Restorer, which combines algorithm unrolling with prompts for unified image restoration, demonstrating that integrating model-based priors with prompt-based learning yields both improved performance and enhanced interpretability.

Gao et al.[42] proposed a prompt-based ingredient-oriented all-in-one restoration framework that decomposes each degradation type into constituent "ingredients"—basic corruption primitives such as frequency attenuation, contrast reduction, and texture loss—and learns ingredient-specific prompts rather than degradation-level prompts. This finer-grained prompt decomposition enables the model to handle novel degradation types as new combinations of known ingredients, substantially improving generalization beyond the training distribution. Wu et al.[43] introduced FrePrompter, a frequency self-prompt approach that automatically generates degradation prompts from the frequency spectrum of the input image rather than relying on learned prompt vectors. By extracting spectral signatures that characterize different degradation types—such as the high-frequency attenuation pattern of blur or the broadband noise floor of Gaussian corruption—FrePrompter provides physically grounded prompt information that enhances both the interpretability and effectiveness of prompt-based restoration. Wu et al.[44] advanced the prompt learning paradigm further with contrastive prompt learning that goes beyond degradation redundancy, addressing the observation that many degradation types share overlapping feature representations that can confuse prompt-based systems. By explicitly encouraging prompts to capture discriminative rather than redundant degradation features through a contrastive objective, this approach achieves more precise degradation-specific modulation and demonstrates state-of-the-art performance across a broader range of degradation types than prior prompt methods.

3.3 Instruction-Guided Restoration

InstructIR[5] represents a paradigm shift by enabling users to control the restoration process through natural language instructions. Instead of relying on implicit degradation detection, users describe the desired restoration in text (e.g., "remove the noise from this photo" or "enhance the contrast and sharpen the details"), and the model adaptively performs the corresponding task. This approach bridges image restoration with natural language understanding, leveraging a text encoder to generate instruction-aware features that condition the restoration network. InstructIR demonstrates strong performance across seven restoration tasks while providing an intuitive human-machine interface that does not require technical knowledge of degradation types.

Jiang et al.[45] proposed multi-dimension visual prompt enhanced image restoration, extending the concept of prompting beyond text to include visual prompts derived from degradation maps, quality scores, and semantic segmentation. This multi-dimensional conditioning pro-

vides richer degradation information than single-modality prompts. Duan et al.[46] developed UniProcessor, a text-induced unified low-level image processor that leverages natural language descriptions of both degradation characteristics and desired restoration outcomes to condition a single processing network. Unlike InstructIR, which uses relatively simple restoration instructions, UniProcessor employs detailed textual descriptions that simultaneously specify the degradation type, its severity, and the target quality attributes, enabling finer-grained control over the restoration process. The text-induced approach also facilitates zero-shot transfer to novel degradation types described in natural language, even when no training examples of those degradation types are available. Xin et al.[47] developed a self-collaboration parallel prompt approach using generative adversarial networks for unsupervised all-in-one image restoration, eliminating the need for paired training data by enabling the model to collaboratively refine prompts through an adversarial learning process.

3.4 Degradation-Aware Architectures and Mixture of Experts

Another line of research focuses on designing architectures that are inherently degradation-aware. Wu et al.[48] proposed "Harmony in Diversity," an all-in-one image restoration approach that uses degradation-aware convolution kernels whose weights are dynamically generated based on estimated degradation parameters. This allows a single network to effectively customize its processing for each input without explicit task switching.

Zamfir et al.[49] introduced MoCE-IR (Mixture of Complexity Experts), which employs a mixture-of-experts architecture where different experts specialize in degradation types of varying complexity. A gating network routes each input to the appropriate combination of experts, enabling the model to allocate computational resources proportional to the restoration difficulty. This design achieves state-of-the-art results across multiple benchmarks while maintaining computational efficiency through sparse expert activation.

Chen et al.[50] proposed "Always Clear Days," a degradation type and severity-aware framework that explicitly models both the type and severity of degradation through a dual-branch architecture. Tian et al.[51] introduced degradation-aware feature perturbation, which systematically perturbs features based on estimated degradation characteristics to improve the robustness of all-in-one restoration. Zhu et al.[52] developed MWFormer, a multi-weather degradation-aware Transformer that handles rain, haze, snow, and their combinations through weather-specific attention modules. Wu et al.[53] proposed a debiased all-in-one restoration approach with task uncertainty regularization that addresses the training bias toward certain degradation types by calibrating task-specific uncertainty.

Tang et al.[54] introduced RamIR, which integrates reasoning and action prompting with the Mamba state-space architecture for all-in-one image restoration. Unlike conventional AiO methods that directly map inputs to outputs, RamIR decomposes the restoration process into an explicit reasoning stage—where the model infers degradation type and severity through chain-

of-thought-style internal representations—followed by an action stage where the inferred degradation information drives restoration through Mamba's selective state-space mechanism. This reasoning-action decomposition mimics the deliberate analysis process of human experts and achieves improved performance on complex mixed-degradation scenarios where purely reactive approaches falter. Dudhane et al.[55] addressed the scalability challenge of AiO restoration through dynamic pre-training, proposing a training strategy where the model is progressively exposed to an expanding set of degradation types during pre-training rather than training on all types simultaneously. This curriculum-based approach enables the model to first learn robust low-level features from simple degradation types before tackling more complex corruptions, resulting in both more efficient training and improved final performance compared to conventional joint training on all degradation types simultaneously.

Li et al.[56] proposed a multi-weather restoration framework based on efficient prompt-guided convolution, where lightweight degradation-specific prompts modulate the convolution kernel weights at each layer rather than conditioning through separate attention mechanisms. This prompt-guided convolution design is significantly more parameter-efficient than attention-based conditioning, enabling multi-weather restoration with minimal overhead compared to a single-weather baseline. Chen et al.[57] developed a multi-modal degradation feature learning framework for unified image restoration based on contrastive learning, which learns degradation representations from multiple modalities—including spatial features, frequency characteristics, and statistical moments—and fuses them through cross-modal attention. This multi-modal degradation representation provides a more comprehensive characterization of complex degradation patterns than single-modal approaches, particularly for mixed degradation scenarios where different modalities capture complementary corruption information. Yang et al.[58] proposed IPT-ILR (Image Pyramid Transformer with Information Loss Regularization), which processes images at multiple pyramid levels through a shared Transformer backbone with level-specific adaptation layers. The information loss regularization term explicitly penalizes the loss of structural information across pyramid levels, ensuring that both coarse global structures and fine local details are preserved throughout the multi-scale restoration process.

3.5 Domain-Specific and Emerging All-in-One Approaches

The all-in-one paradigm has been extended beyond standard natural image benchmarks to domain-specific applications where specialized degradation patterns demand tailored solutions. Zhang et al.[59] proposed an all-in-one multi-degradation image restoration network via hierarchical degradation representation, which organizes degradation types into a hierarchical taxonomy—from coarse categories (weather, sensor, compression) to fine-grained subtypes—and learns correspondingly hierarchical representations that enable both broad degradation discrimination and fine-grained restoration adaptation. This hierarchical structure provides a principled framework for scaling AiO models to large numbers of degradation types without proportional growth in model complexity.

Chen et al.[60] demonstrated the application of all-in-one image enhancement to aerial forest scenes, addressing the unique challenges of airborne imagery including haze, low illumination, and atmospheric scattering that simultaneously degrade canopy visibility. Their forest-specific AiO network incorporates domain knowledge about vegetation spectral characteristics to guide the restoration process, achieving improved species identification accuracy from restored aerial imagery. Zhang et al.[61] proposed UniUIR, which formulates underwater image restoration as an all-in-one learning problem, handling the diverse and often co-occurring degradation types encountered in aquatic environments—including color cast from wavelength-dependent absorption, haze from suspended particles, and low contrast from backscatter—within a single model. UniUIR's success in the underwater domain demonstrates that AiO approaches can effectively generalize to degradation distributions substantially different from those encountered in atmospheric natural imaging.

Yang et al.[62] developed MvKSR (Multi-view Knowledge-guided Scene Recovery), which tackles scene recovery under combined haze and rain degradation by leveraging multi-view knowledge from different degradation-specific perspectives. The multi-view design enables the model to simultaneously analyze the scene from weather removal, detail enhancement, and color correction perspectives, with a knowledge fusion mechanism that combines these complementary views into a coherent restoration. Liu et al.[63] addressed real-time multi-scene visibility enhancement for navigational safety under complex weather conditions, proposing a lightweight AiO architecture specifically designed for the latency constraints of vessel navigation systems. Their approach achieves real-time processing speeds while maintaining competitive visibility enhancement across fog, rain, haze, and low-light conditions, demonstrating that efficiency-oriented AiO design can meet the stringent requirements of safety-critical applications.

Tang et al.[64] explored generative adversarial unsupervised image restoration in hybrid degradation scenes, proposing a GAN-based framework that can handle complex mixtures of degradation types without requiring paired training data. The unsupervised approach learns degradation disentanglement through adversarial training, separating the clean image content from the degradation pattern without explicit degradation supervision. This unsupervised paradigm is particularly valuable for real-world applications where obtaining paired clean-degraded training data is infeasible. Liang et al.[65] introduced INP-Net (Implicit Neural Prompting Network) for remote sensing image dehazing, which uses implicit neural representations to generate continuous, spatially-varying prompts that adapt to the non-uniform haze distribution typically observed in remote sensing imagery. Unlike discrete prompt vectors, the implicit neural representation can model smoothly varying degradation intensity across the spatial extent of large-scale remote sensing images.

Zeng et al.[66] proposed a unified framework for all-in-one image compression and restoration, jointly addressing the traditionally separate problems of lossy compression artifact removal and general degradation restoration within a single model. This joint formulation is motivated

by the observation that compressed images in real-world applications frequently exhibit both compression artifacts and other degradation types simultaneously, and treating them independently leads to suboptimal results. Zhang et al.[67] developed a unified accelerator for all-in-one image restoration based on prompt degradation learning, focusing on the hardware implementation perspective. Their work designs an FPGA-based accelerator architecture that efficiently executes prompt-conditioned restoration through custom hardware units for prompt generation, degradation-conditioned convolution, and prompt-guided attention, achieving substantial speedup over GPU implementation while maintaining restoration quality. This hardware-software co-design perspective addresses a critical gap in the AiO restoration literature, where most methods are evaluated solely on GPU platforms without considering deployment on specialized hardware.

3.6 Deep Unfolding and Model-Based Approaches

Several works have combined deep learning with model-based optimization for AiO restoration. Tang et al.[68] proposed DA-RCOT (Degradation-Aware Residual-Conditioned Optimal Transport), which formulates AiO restoration as an optimal transport problem with degradation-conditioned residual connections. Zeng et al.[69] developed a vision-language gradient descent-driven deep unfolding framework that integrates vision-language models with iterative optimization, using VLM-derived degradation descriptions to guide each unfolding step. Cheng et al.[70] introduced RDM-IR, a task-adaptive deep unfolding network that combines iterative shrinkage-thresholding with learned degradation-specific parameters for all-in-one restoration.

Lihe et al.[71] proposed Ada4DIR, an adaptive model-driven approach that directly incorporates the imaging physics model into the network architecture, enabling the model to explicitly reason about the degradation process during restoration. These model-based approaches offer improved interpretability compared to purely data-driven methods, as the iterative unfolding steps correspond to optimization iterations with clear physical meaning.

3.7 Limitations of All-in-One Approaches

Jiang et al.[72] provided a comprehensive survey on AiO restoration, identifying several persistent limitations. First, most AiO models are trained on a fixed set of degradation types (typically 3–7), and their performance degrades substantially when encountering out-of-distribution degradation types or novel combinations not seen during training. Second, AiO models process the entire image through a single pipeline, ignoring the spatial heterogeneity of degradation. Third, the single forward-pass inference paradigm cannot iteratively refine results based on intermediate quality assessment, leading to suboptimal outcomes for complex degradation scenarios that require multi-step reasoning. These limitations highlight the need for more flexible, adaptive approaches—specifically, agent-based methods that can dynamically perceive degradation, select appropriate tools, and iteratively refine restoration strategies.

4 Diffusion-Based Generative Image Restoration

Diffusion probabilistic models have emerged as a powerful generative framework for image restoration, offering the ability to produce perceptually realistic results by leveraging learned image priors from large-scale pretraining. This section examines how diffusion models have been adapted for restoration, analyzes their advantages and limitations compared to discriminative methods, and discusses recent innovations in conditional generation and efficiency improvements.

4.1 Foundations and Conditional Diffusion for Restoration

The application of diffusion models to image restoration typically involves conditioning the reverse denoising process on the degraded input image. Saharia et al.[73] introduced Palette, one of the earliest diffusion-based frameworks for image-to-image translation tasks including colorization, inpainting, and uncropping. Palette demonstrates that a simple conditional diffusion model, trained with paired data, can produce diverse, high-quality outputs that capture the multimodal nature of inverse problems—a fundamental advantage over deterministic methods that produce a single point estimate.

Zhang et al.[74] proposed a unified conditional framework for diffusion-based image restoration that introduces task-adaptive conditioning mechanisms, enabling a single diffusion model to handle multiple restoration tasks by conditioning on task-specific embeddings. This unified approach represents an important step toward combining the generative power of diffusion models with the flexibility of all-in-one methods. Liu et al.[75] developed Residual Denoising Diffusion Models (RDDM), which reformulate the diffusion process to operate on residuals between clean and degraded images rather than pure noise, significantly reducing the number of sampling steps required while maintaining high restoration quality. Zheng et al.[76] proposed a selective hourglass mapping strategy for universal image restoration based on diffusion models, which selectively applies the diffusion process only to the most degraded frequency components of the input while preserving well-maintained components through direct passthrough. The hourglass architecture progressively maps degraded features to clean features through a bottleneck structure that concentrates the diffusion model's generative capacity on the most challenging restoration regions, achieving improved efficiency without sacrificing quality on diverse degradation types. Ding et al.[77] formulated image restoration as restoration by generation with constrained priors, introducing explicit constraints during the diffusion sampling process that enforce fidelity to the degraded input while allowing the generative prior to synthesize plausible high-frequency details. The constrained generation framework provides a principled balance between the faithfulness of deterministic methods and the perceptual quality of unconstrained generative approaches, addressing a fundamental tension in diffusion-based restoration.

4.2 Diffusion Priors from Pre-Trained Models

A significant recent trend leverages the powerful image priors learned by large pre-trained diffusion models, particularly Stable Diffusion, for restoration tasks. DiffBIR[6][78] pioneered this approach by proposing a two-stage pipeline: first, a restoration module removes primary degradation to produce an intermediate result, then a conditional latent diffusion model refines the result using generative priors from a pre-trained Stable Diffusion model. This design achieves remarkable perceptual quality on blind image restoration tasks, including real-world super-resolution and face restoration, where the diffusion prior effectively hallucinate plausible high-frequency details that deterministic methods cannot recover.

Yang et al.[79] proposed Pixel-Aware Stable Diffusion (PASD), which introduces a pixel-aware cross-attention module that enables the diffusion model to maintain fidelity to the input image while leveraging generative priors for detail synthesis. Tian et al.[80] developed a method for learning diffusion texture priors for image restoration, extracting texture-level guidance from diffusion models to improve the quality of fine details without the full cost of diffusion sampling. Xu et al.[81] demonstrated that priors from pre-trained models including CLIP and Stable Diffusion can be combined to boost restoration performance across multiple tasks, with CLIP providing semantic guidance and Stable Diffusion providing texture priors.

4.3 Task-Specific Diffusion Adaptations

Diffusion models have been successfully adapted to specific restoration tasks with tailored designs. Guo et al.[82] proposed ShadowDiffusion, which introduces degradation-aware conditioning for shadow removal by incorporating shadow mask information into the diffusion process. Yi et al.[7] developed Diff-Retinex, which combines Retinex theory with diffusion models for low-light image enhancement, using the decomposition into illumination and reflectance components to guide the diffusion process. Yi et al.[83] further extended this approach in Diff-Retinex++, incorporating reinforcement learning to optimize the diffusion sampling trajectory for improved efficiency and quality.

Shang et al.[84] introduced ResDiff, which combines CNN-based baseline restoration with a diffusion model that refines residual details, effectively using the CNN to handle the bulk of the restoration while the diffusion model adds perceptually important high-frequency content. Luo et al.[85] proposed Refusion for enabling large-size realistic image restoration with diffusion models, addressing the challenge that diffusion models typically operate at fixed low resolutions by introducing a region-aware diffusion strategy. Wang et al.[86] further addressed the resolution limitation with unlimited-size diffusion restoration, proposing a patch-based diffusion strategy with carefully designed overlapping and blending mechanisms that enable diffusion-based restoration to process images of arbitrary resolution without introducing visible seam artifacts at patch boundaries. This approach is particularly important for practical applications such as satellite imagery and medical imaging where images are orders of magnitude larger than

the diffusion model's native resolution. Zhao et al.[87] developed an iterative diffusion framework for authentic face restoration, demonstrating the effectiveness of diffusion-based iterative refinement for specialized restoration tasks.

Welker et al.[88] proposed DriftRec, which adapts diffusion models specifically to blind JPEG restoration by modeling the JPEG compression process as a "drift" in the diffusion trajectory. Rather than treating JPEG artifact removal as a generic denoising problem, DriftRec incorporates knowledge of the block-based discrete cosine transform structure of JPEG compression into the diffusion process, enabling the model to precisely reverse compression-induced artifacts while preserving authentic image detail. This degradation-aware diffusion adaptation achieves significant improvements over generic diffusion restoration on JPEG-compressed images, particularly at low quality factors where blocking artifacts are severe. Kang et al.[89] developed an image intrinsic components guided conditional diffusion model for low-light image enhancement, which decomposes the image into illumination and reflectance components following the Retinex model and uses these intrinsic components as conditioning signals for the diffusion process. This physics-informed conditioning provides the diffusion model with explicit structural guidance about the image formation process, leading to enhanced detail preservation and more natural color rendering in the enhanced output compared to unconditional diffusion approaches. Yan et al.[90] proposed an efficient diffusion-based approach to image enhancement through frequency-domain priors, where the diffusion process is guided by frequency-domain analysis that identifies which spectral components require generative enhancement versus direct preservation. By concentrating the computationally expensive diffusion sampling on frequency bands that genuinely benefit from generative modeling while directly passing through well-preserved bands, this method achieves substantial computational savings while maintaining the perceptual quality advantages of diffusion-based restoration.

4.4 All-in-One Diffusion Restoration

Combining diffusion models with the all-in-one paradigm represents a particularly active research direction. Tu et al.[91] proposed an uncertainty-aware diffusion bridge model for unifying heterogeneous degradations, using the diffusion bridge formulation to directly model the transition between degraded and clean distributions conditioned on degradation type. Luo et al.[92] developed Defusion, which uses visual-instructed degradation information to guide the diffusion process for all-in-one restoration, achieving strong results on both known and novel degradation types. Zhang et al.[93] proposed Diff-Restorer, which unleashes visual prompts for diffusion-based universal image restoration by encoding degradation characteristics as visual prompt tokens that condition the diffusion sampling.

Lv et al.[94] introduced adaptive prompt-guided unified image restoration with latent diffusion, where a prompt generation module dynamically creates degradation-aware prompts that are injected into the latent diffusion process. Yue et al.[95] proposed a joint conditional diffusion model specifically designed for image restoration with mixed degradations, where a single dif-

fusion model simultaneously conditions on multiple degradation descriptors to handle images corrupted by combinations of noise, blur, and compression. The joint conditioning mechanism enables the model to reason about degradation interactions—for instance, how blur affects the optimal noise removal strategy—rather than treating each degradation type independently, yielding improved results on mixed-degradation benchmarks compared to sequentially applying single-degradation diffusion models. These combined approaches leverage both the flexibility of prompt-based degradation conditioning and the generative power of diffusion models, representing the current frontier of unified restoration methods.

4.5 Limitations of Diffusion-Based Approaches

Despite their impressive perceptual quality, diffusion-based restoration methods face several persistent challenges. The iterative sampling process requires tens to hundreds of denoising steps, resulting in inference times that are orders of magnitude slower than feed-forward methods. While techniques such as RDDM[75] and distillation-based acceleration have partially addressed this, real-time diffusion-based restoration remains challenging. Additionally, diffusion models may hallucinate plausible but incorrect details, particularly in regions where the degradation has destroyed all original information, leading to concerns about fidelity in applications such as medical imaging and forensic analysis where accuracy is paramount. The large model sizes and memory requirements of pre-trained diffusion models also limit deployment on edge devices, motivating the compression and distillation approaches discussed in Section 7.

5 Agent-Based Image Restoration Systems

Agent-based image restoration represents the most recent and potentially most transformative paradigm in the field, fundamentally reimagining restoration as an intelligent decision-making process rather than a static mapping function. By leveraging large language models (LLMs) and vision-language models (VLMs) as cognitive controllers, these systems can perceive degradation, reason about appropriate restoration strategies, orchestrate specialized tools, evaluate results, and iteratively refine their approach. This section provides a comprehensive analysis of the emerging agent-based restoration paradigm.

5.1 The AgenticIR Framework

AgenticIR[9] established the foundational framework for agent-based image restoration. The system operates through a five-stage pipeline: (1) *Perception*: a VLM analyzes the input image to identify degradation types and severity levels; (2) *Scheduling*: an LLM generates an ordered sequence of restoration tools based on the perceived degradation; (3) *Execution*: the selected tools are applied to the image in the planned order; (4) *Reflection*: the VLM evaluates the restored result against the original to assess quality improvement; and (5) *Rescheduling*: if the result is unsatisfactory, the system backtracks and explores alternative tool combinations

through depth-first search.

This framework represents a fundamental departure from all prior approaches. Unlike single-task models that apply a fixed transformation, or AiO models that process all degradation types through a single pipeline, AgenticIR can dynamically compose arbitrary sequences of specialized tools, each optimized for a specific degradation type. The reflection mechanism enables self-evaluation and iterative refinement, capabilities that no feed-forward model possesses. However, AgenticIR's depth-first search strategy leads to exponential computational cost in the worst case, and its reliance on powerful LLMs and VLMs for real-time perception and reasoning creates significant latency overhead, making deployment challenging for time-sensitive applications.

In a closely related and complementary line of work, Zhu et al.[96] developed an intelligent agentic system for complex image restoration problems that extends the agent-based paradigm along several critical dimensions. While AgenticIR focuses primarily on degradation type identification and tool sequencing, this system introduces a more sophisticated problem decomposition strategy that explicitly addresses the complexity hierarchy of restoration tasks. The system classifies restoration problems into three complexity levels—simple (single, identifiable degradation), moderate (multiple known degradation types), and complex (unknown or spatially varying degradation)—and adapts its reasoning depth accordingly. For simple problems, the agent applies a single-step tool selection, minimizing latency overhead. For moderate problems, it engages in multi-step planning with degradation ordering optimization. For complex problems, the agent activates a full deliberation pipeline that includes spatial analysis, degradation hypothesis generation, and multi-hypothesis verification. This adaptive complexity mechanism represents an important advance toward practical deployment, as it avoids the computational overhead of full deliberation for straightforward restoration tasks while retaining deep reasoning capability for genuinely challenging scenarios. Furthermore, the system incorporates a memory module that accumulates restoration experience across sessions, enabling the agent to recognize previously encountered degradation patterns and retrieve successful restoration strategies without repeating the full reasoning process. This experience-driven approach draws parallels with case-based reasoning in classical AI and suggests a path toward agent systems that improve with deployment experience rather than requiring periodic retraining.

5.2 Multi-Agent Architectures

Recognizing the computational limitations of single-agent approaches, recent works have explored multi-agent architectures for more efficient restoration. MAIR[11] introduces a multi-agent framework with three-stage degradation priors that guides tool selection. The multi-agent design enables parallel processing of different aspects of the restoration task, reducing sequential dependencies and improving efficiency. Specifically, MAIR establishes a degradation prior ordering principle (compression artifacts $\rightarrow$ imaging degradation $\rightarrow$ scene degradation) that constrains the search space and achieves a 44% improvement in inference efficiency compared to AgenticIR while maintaining comparable restoration quality.

Tripathi et al.[97] analyzed recent advancements in agentic AI architectures and prompting strategies, including their application to visual tasks. The multi-agent paradigm enables a hierarchical organization where a global coordinator agent manages high-level strategy while specialized sub-agents handle specific aspects of the restoration task, such as region-specific processing, quality monitoring, and parameter optimization. This hierarchical structure mirrors the organization of complex real-world systems and offers natural scalability as the number of degradation types and tools increases.

5.3 Quality-Driven Restoration Planning

RestoreAgent[10] takes a different approach by fine-tuning a multimodal large language model to directly predict restoration workflows. Rather than using an LLM as a general-purpose reasoner that calls specialized tools, RestoreAgent trains the model end-to-end on restoration task-plan pairs, enabling faster inference by eliminating the need for multi-turn reasoning. The model learns to map visual observations directly to tool sequences, effectively compressing the multi-step reasoning process into a single forward pass.

Q-Agent[12] addresses the efficiency challenge from the perspective of search strategy, proposing a quality-driven greedy approach that selects the next restoration tool based on the expected quality improvement, as measured by an image quality assessment model. Unlike AgenticIR's exhaustive depth-first search, Q-Agent's greedy strategy achieves linear computational complexity with respect to the number of available tools, making it significantly more practical for deployment. The key insight is that quality-driven greedy selection, while not guaranteed to find the globally optimal tool sequence, produces results that are competitive with exhaustive search in practice.

5.4 Spatial Awareness in Agent-Based Systems

A critical limitation of current agent-based systems is their treatment of the entire image as a single entity. All existing agent-based methods—AgenticIR, RestoreAgent, MAIR, and Q-Agent—apply their restoration plans globally, without considering that degradation may vary dramatically across different spatial regions of an image. This represents a significant missed opportunity, as the agent framework is naturally suited to spatial decomposition: a coordinating agent could segment the image into regions, analyze the degradation characteristics of each region independently, assign specialized sub-agents to handle each region, and then fuse the results.

The tools needed for such spatial awareness already exist. The Segment Anything Model (SAM)[98] provides zero-shot semantic segmentation that can decompose an image into meaningful regions without task-specific training. VLMs such as DepictQA[99] can analyze individual regions to identify degradation types and severity. Chen et al.[100] demonstrated RobustSAM, which maintains segmentation quality even on degraded images, addressing a key concern about applying segmentation to corrupted inputs. The combination of region-level segmentation,

degradation analysis, and agent-based tool orchestration represents a natural and promising extension of the current paradigm toward spatially-aware multi-agent restoration.

6 Foundation Models for Degradation Analysis and Quality Assessment

The effectiveness of agent-based and spatially-aware restoration systems depends critically on two capabilities: accurately perceiving and characterizing image degradation, and reliably assessing restoration quality. Foundation models—including segment anything models and vision-language models—provide the perceptual backbone for these capabilities. This section examines how foundation models enable degradation analysis and quality assessment within restoration pipelines.

6.1 Segment Anything for Spatial Decomposition

The Segment Anything Model (SAM)[98], trained on over one billion masks, provides zero-shot segmentation capabilities that can decompose any image into semantically coherent regions without task-specific fine-tuning. For spatially-aware image restoration, SAM's ability to generate high-quality region masks enables the decomposition of a degraded image into spatial units that can be independently analyzed and processed. Each region inherits semantic meaning (sky, building, vegetation, road) that correlates with likely degradation patterns: sky regions are more susceptible to haze, while road surfaces are more affected by rain-induced reflections.

Ren et al.[101] proposed Grounded SAM, which combines text-prompted object detection with SAM's segmentation capability, enabling text-guided region identification. For restoration applications, this allows specifying regions of interest through natural language (e.g., "the foggy sky region" or "the noisy shadow area"), providing an intuitive interface for spatially-aware degradation analysis. Chen et al.[100] developed RobustSAM, specifically addressing the challenge that SAM's segmentation quality degrades on corrupted images. RobustSAM fine-tunes SAM with degradation-aware training data, maintaining segmentation accuracy on images affected by noise, blur, and other distortions, making it directly applicable to the first stage of spatially-aware restoration pipelines.

6.2 Vision-Language Models for Degradation Assessment

Vision-language models (VLMs) enable rich, descriptive assessment of image degradation that goes beyond simple classification. DepictQA[99] leverages VLMs for depicted image quality assessment, providing detailed natural language descriptions of image quality attributes including specific degradation types, severity levels, and affected regions. You et al.[102] extended this approach with advanced VLM architectures that can perform comparative quality assessment between image pairs, enabling before-and-after evaluation of restoration results. You et

al.[103] further enhanced descriptive image quality assessment with large VLM architectures, demonstrating improved consistency and accuracy in degradation identification.

Wu et al.[104] developed Q-Instruct, a large-scale instruction-following dataset specifically designed for training VLMs on low-level visual perception tasks. Models fine-tuned on Q-Instruct demonstrate significantly improved ability to identify and describe image degradation compared to general-purpose VLMs, providing the perceptual foundation needed for agent-based restoration systems. The companion Q-Bench benchmark[105] and its extension Q-Bench+[106] establish standardized evaluation protocols for VLM-based quality assessment, covering perception, description, and comparison tasks that directly correspond to the requirements of restoration-oriented degradation analysis.

Zhang et al.[107] proposed Q-Boost, which employs a triadic-tone and multi-prompt strategy to enhance VLMs' quality assessment abilities, demonstrating that prompt engineering can significantly improve VLM performance on low-level visual tasks without additional training. Wu et al.[108] conducted a comprehensive study of multimodal large language models for image quality assessment, providing systematic comparisons that inform the selection of appropriate VLM backbones for restoration applications.

6.3 Foundation Model Priors for Real-World Restoration

A rapidly growing body of work leverages the rich visual and semantic priors encoded in foundation models—including large diffusion models, vision-language models, and pre-trained vision transformers—to address the challenging problem of real-world image restoration, where degradation is complex, spatially varying, and does not conform to simple parametric models. Ai et al.[109] proposed DreamClear, a high-capacity real-world image restoration framework that combines a large-scale diffusion model with a privacy-safe dataset curation pipeline. DreamClear's key insight is that the generative capacity of a diffusion model can be harnessed as a powerful restoration prior when conditioned on degradation-aware features extracted from the input image. The privacy-safe dataset curation approach addresses the practical concern that large-scale restoration datasets may contain identifiable personal information, proposing automated de-identification and filtering mechanisms that enable the construction of high-capacity training sets without compromising individual privacy. DreamClear achieves state-of-the-art results on multiple real-world restoration benchmarks, demonstrating that the combination of large model capacity and large-scale curated data is essential for handling the diversity of real-world degradation.

Luo et al.[110] developed a method for photo-realistic image restoration in the wild using controlled vision-language models, where a VLM is employed not only for degradation perception but also as an active participant in the restoration process through controlled generation. The VLM generates natural language descriptions of the desired restoration outcome, which are then used as conditioning signals for a controlled diffusion model that produces photo-realistic

restorations. This VLM-in-the-loop approach enables a form of semantic guidance that ensures restored images are not only technically clean but also semantically coherent—preserving the intended mood, lighting, and composition of the original scene. Chen et al.[111] proposed real-world super-resolution with VLM-based degradation prior learning, which uses a vision-language model to predict detailed degradation parameters—including noise level, blur kernel estimation, and compression quality factor—from the input image through a learned mapping between visual features and degradation descriptions. These VLM-predicted degradation priors are then fed to a conditional restoration network, providing significantly more informative conditioning than traditional blind degradation estimation methods that rely on simple neural network regressors.

Cui et al.[112] addressed the specific application of image restoration for autonomous vehicle perception, proposing a multi-scale feature modulation network that enhances images captured under adverse conditions to improve downstream detection and segmentation performance. Their work highlights an important emerging application paradigm: restoration not as an end in itself, but as a preprocessing step whose quality should be measured by its impact on downstream task performance rather than pixel-level fidelity metrics alone. The multi-scale feature modulation design enables the restoration network to selectively enhance features at scales most relevant to the autonomous driving perception tasks, achieving improved detection accuracy under fog, rain, and low-light conditions with minimal computational overhead. These works collectively demonstrate that foundation model priors—whether from diffusion models, vision-language models, or pre-trained vision transformers—provide a powerful and general mechanism for closing the gap between synthetic-data-trained restoration methods and the demands of real-world deployment.

6.4 No-Reference Image Quality Assessment

Classical no-reference image quality assessment (NR-IQA) methods remain important components of restoration pipelines, providing efficient quality scores that can guide tool selection and termination decisions. Talebi and Milanfar[113] introduced NIMA (Neural Image Assessment), a CNN-based NR-IQA method that predicts the distribution of human quality ratings rather than a single score, providing both a quality estimate and an uncertainty measure. Zhang et al.[114] demonstrated the unreasonable effectiveness of deep features for perceptual quality assessment through LPIPS (Learned Perceptual Image Patch Similarity), which uses features from pre-trained networks to compute perceptual similarity metrics that closely correlate with human judgments.

Zheng et al.[115] proposed a conditional knowledge distillation approach for degraded-reference image quality assessment, specifically designed for evaluating image restoration quality by leveraging both the degraded input and the restored output. Li et al.[116] demonstrated the integration of quality assessment guidance into the restoration process itself, using IQA scores as feedback signals to guide motion blur removal, establishing the concept of quality-driven restoration that is central to agent-based approaches like Q-Agent.

7 Model Compression and Lightweight Deployment

The computational demands of advanced restoration methods—particularly Transformer-based, diffusion-based, and agent-based approaches—present significant barriers to deployment on resource-constrained devices. This section examines knowledge distillation and model compression techniques specifically designed for image restoration, with particular attention to trajectory distillation as a mechanism for transferring complex multi-step reasoning into lightweight models.

7.1 Knowledge Distillation for Restoration Networks

Knowledge distillation, which transfers knowledge from a large teacher model to a compact student model, has been specifically adapted for image restoration with domain-aware innovations. Zhang et al.[117] proposed soft knowledge distillation with multi-dimensional cross-net attention (SKD) for image restoration compression, where the student learns to replicate not only the teacher's outputs but also its intermediate attention patterns across spatial and channel dimensions. This multi-dimensional distillation preserves the teacher's ability to capture both local texture details and global structural coherence, achieving significant compression (e.g., 2–4× reduction in parameters) with minimal quality degradation.

Yang et al.[118] introduced Mamba-oriented heterogeneous knowledge distillation for image restoration model compression, leveraging the efficiency of the Mamba state-space architecture for the student model while distilling knowledge from a larger Transformer-based teacher. This cross-architecture distillation demonstrates that knowledge can be effectively transferred between fundamentally different model families, enabling the deployment of Transformer-quality restoration through efficient Mamba-based models. Zhang et al.[119][120] proposed simultaneous learning knowledge distillation (SLKD) for image restoration that jointly optimizes the teacher and student models during distillation, avoiding the two-stage training pipeline and achieving better teacher-student alignment.

Wang et al.[121] developed data-free distillation with degradation-prompt diffusion for multi-weather restoration, addressing the practical challenge of knowledge distillation when the original training data is unavailable. By using a diffusion model to generate synthetic degraded images conditioned on degradation prompts, the method enables distillation without access to the teacher's training data, which is particularly valuable for deployment scenarios with data privacy constraints.

7.2 Trajectory Distillation for Agent-Based Systems

A particularly promising direction for lightweight deployment of agent-based restoration is trajectory distillation, which compresses the complex multi-step decision-making process of an agent system into a lightweight feedforward network. The concept involves running the full agent sys-

tem (with LLMs, VLMs, and specialized tools) offline on a large corpus of degraded images, recording the complete decision trajectories—including region segmentation, degradation identification, tool selection, parameter settings, and quality scores—and then training a compact network to directly predict these decisions from the input image alone.

This approach offers several compelling advantages: (1) it eliminates the need for expensive LLM/VLM inference at deployment time, enabling real-time processing; (2) the compact network can be trained with multi-task learning objectives covering segmentation, classification, routing, and parameter prediction simultaneously; and (3) the quality of the training signal is bounded by the agent system's performance, ensuring that the distilled model inherits the agent's ability to handle complex, spatially heterogeneous degradation. The key challenge lies in ensuring that the lightweight model can faithfully replicate the agent's nuanced decision-making, particularly for edge cases and novel degradation combinations not well-represented in the trajectory dataset.

8 Discussion

The evolution from single-task restoration to agent-based spatially-aware systems reveals several cross-cutting themes and open challenges that merit deeper examination.

8.1 The Quality–Efficiency Trade-off

A persistent tension runs through all restoration paradigms: methods that achieve higher restoration quality tend to require significantly more computational resources. Single-task Transformer models such as Restormer[2] achieve excellent quality–efficiency balance for their specific tasks but cannot handle multiple degradation types. AiO models sacrifice some per-task quality for versatility. Diffusion models achieve the best perceptual quality but at orders-of-magnitude higher computational cost. Agent-based systems offer the most flexible and potentially highest-quality restoration but currently require multiple rounds of LLM/VLM inference.

This progression suggests that the field has not yet found an approach that simultaneously achieves top-tier quality, broad versatility, and low computational cost. Trajectory distillation represents the most promising path toward resolving this trilemma, but its effectiveness depends on the coverage and quality of the trajectory data, the capacity of the lightweight student model, and the ability to generalize to degradation scenarios not encountered during offline agent operation.

8.2 Synthetic vs. Real-World Degradation

Most restoration methods are developed and evaluated on synthetically degraded images, where clean ground truth is available for supervised training and quantitative evaluation. However, real-world degradation differs significantly from synthetic models: it is spatially non-uniform, involves complex interactions between multiple degradation types, varies with scene content

and imaging conditions, and cannot be precisely characterized by mathematical models. Zhai et al.[122] provided a comprehensive review of real-world image super-resolution, highlighting the substantial performance gap between synthetic and real-world benchmarks. Guan et al.[123] contributed WeatherBench, a real-world benchmark for all-in-one weather degradation, providing more realistic evaluation conditions.

Agent-based and spatially-aware approaches are particularly well-suited to bridge this gap, as they can adaptively respond to the specific degradation characteristics observed in each image rather than relying on assumptions about the degradation model. The combination of VLM-based degradation perception and dynamic tool orchestration enables a more flexible and robust response to the unpredictable nature of real-world degradation.

8.3 Spatial Awareness as the Missing Dimension

Perhaps the most significant insight emerging from this review is that spatial awareness remains underdeveloped across all restoration paradigms. Single-task and AiO methods process the entire image uniformly. Diffusion models apply the same generative process globally. Even current agent-based systems treat the image as a monolithic entity when making restoration decisions. The few methods that incorporate spatial information do so in limited ways: spatially-varying convolution kernels, region-specific attention mechanisms, or local-global feature fusion.

True spatial awareness requires decomposing the image into meaningful regions, independently analyzing the degradation characteristics of each region, planning region-specific restoration strategies, executing those strategies with appropriate tools and parameters, and seamlessly fusing the results. This capability would address a fundamental limitation of all current approaches and enable significantly improved restoration of real-world images with heterogeneous degradation.

9 Perspectives and Future Directions

Based on the analysis presented in this review, we identify several promising research directions that can advance the field toward more intelligent, adaptive, and deployable image restoration systems.

9.1 Hierarchical Multi-Agent Restoration with Spatial Awareness

The most immediate opportunity lies in extending agent-based restoration with spatial awareness through a hierarchical multi-agent architecture. A global coordinating agent would analyze the overall image and decompose it into regions using foundation models like SAM. Regional expert agents would independently analyze and restore each region, selecting from a toolkit of specialized restoration models. A quality assessment agent would evaluate both regional and global restoration quality, triggering iterative refinement when needed. This architecture

naturally supports the spatial heterogeneity of real-world degradation while maintaining the adaptive flexibility of agent-based systems.

9.2 Continuous Parameter Optimization

Current agent-based systems primarily make discrete tool selection decisions: choosing which restoration model to apply. A natural extension is continuous parameter optimization, where agents not only select tools but also tune their operating parameters. For example, a denoising tool might be applied with different noise level estimates for different image regions, or a super-resolution model might use different upscaling factors depending on the local detail content. This continuous parameter space dramatically increases the flexibility and precision of agent-based restoration, but also requires more sophisticated optimization strategies—potentially combining reinforcement learning with quality-driven feedback.

9.3 Trajectory Distillation for Edge Deployment

Bridging the gap between the high-quality restoration of agent-based systems and the real-time requirements of edge deployment requires effective knowledge compression. Trajectory distillation, where a lightweight network learns to replicate the agent's decision-making from collected trajectory data, represents the most promising approach. Key research challenges include: designing multi-task student architectures that can simultaneously predict region segmentation, degradation types, tool routing, and continuous parameters; developing trajectory selection and weighting strategies that prioritize diverse and challenging examples; and establishing evaluation protocols that assess not only per-image restoration quality but also decision consistency and robustness.

9.4 Cross-Domain Generalization and Benchmarking

The field needs standardized benchmarks for evaluating spatially-aware and agent-based restoration. Current benchmarks primarily evaluate per-task performance on synthetically degraded images, which does not capture the complexity of real-world scenarios. Future benchmarks should include images with spatially varying heterogeneous degradation, evaluation metrics that assess region-level restoration quality in addition to global metrics, and practical considerations such as computational efficiency and deployment constraints. Cross-domain evaluation—testing methods trained on natural images on medical, remote sensing, or industrial imaging tasks—would also reveal the generalization capabilities and limitations of different approaches.

10 Conclusion

This review has traced the evolution of image restoration from single-task Transformer architectures through all-in-one unified models and diffusion-based generative approaches to the emerging paradigm of agent-based spatially-aware systems. Each paradigm addresses specific limita-

tions of its predecessors: Transformer architectures introduced global context modeling that surpassed CNNs; all-in-one models eliminated the need for degradation-specific networks; diffusion models achieved unprecedented perceptual quality through generative priors; and agent-based systems introduced dynamic reasoning, tool orchestration, and iterative refinement.

The critical gap that remains is spatial awareness. Real-world images exhibit spatially heterogeneous degradation that cannot be adequately addressed by any approach that treats the image uniformly. The convergence of foundation models for spatial decomposition (SAM and its variants), vision-language models for region-level degradation analysis (DepictQA and related methods), and multi-agent frameworks for coordinated restoration planning creates a unique opportunity to address this gap. Combined with trajectory distillation for lightweight deployment, spatially-aware multi-agent restoration has the potential to establish a new paradigm that is simultaneously more accurate, more adaptive, and more deployable than existing approaches.

The field stands at an inflection point where the integration of perception, reasoning, and action—long the aspiration of artificial intelligence research—is becoming practically achievable for the concrete and important problem of image restoration. The research directions identified in this review—hierarchical multi-agent architectures, continuous parameter optimization, trajectory distillation, and comprehensive benchmarking—provide a roadmap for realizing this potential.

参考文献

[1] Liang J, Cao J, Sun G, Zhang K, Gool L V, Timofte R. SwinIR: Image Restoration Using Swin Transformer[J/OL], 2021. https://doi.org/10.1109/iccvw54120.2021.00210.

[2] Zamir S W, Arora A, Khan S, Hayat M, Khan F S, Yang M. Restormer: Efficient Transformer for High-Resolution Image Restoration[C/OL] //2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2022. https://doi.org/10.1109/cvpr52688.2022.00564.

[3] Li B, Liu X, Hu P, Wu Z, Lv J, Peng X. AirNet: All-in-One Image Restoration Network via Contrastive Learning[C/OL] //IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2022. https://doi.org/10.1109/CVPR52688.2022.02014.

[4] Potlapalli V, Zamir S W, Khan S, Khan F S. PromptIR: Prompting for All-in-One Blind Image Restoration[J/OL]. arXiv (Cornell University), 2023. https://doi.org/10.48550/arxiv.2306.13090.

[5] Conde M V, Geigle G, Timofte R. InstructIR: High-Quality Image Restoration Following Human Instructions[C] //European Conference on Computer Vision (ECCV). 2024.

[6] Lin X, He J, Chen Z, Lyu Z, Dai B, Yu F, Qiao Y, Ouyang W, Dong C. DiffBIR: Toward Blind Image Restoration with Generative Diffusion Prior[J/OL]. Lecture notes in computer science, 2024. https://doi.org/10.1007/978-3-031-73202-7_25.

[7] Yi X, Xu H, Zhang H, Tang L, Ma J. Diff-Retinex: Rethinking Low-light Image Enhancement with A Generative Diffusion Model[J/OL], 2023. https://doi.org/10.1109/iccv51070.2023.01130.

[8] 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[J/OL]. arXiv (Cornell University), 2023. https://doi.org/10.48550/arxiv.2308.09388.

[9] Zhu H, Xu K, Wang R. AgenticIR: Agentic Image Restoration with Visual Language Models[J]. arXiv preprint, 2025.

[10] Chen H, Chen S, Gu J, Li W, Pei R, Ren J, Song F, Tian Y, Zhou K, Zhu L. RestoreAgent: Autonomous Image Restoration Agent via Multimodal Large Language Models[J/OL], 2024. https://doi.org/10.52202/079017-3512.

[11] Jiang X, Li G, Chen B, Zhang J. Multi-Agent Image Restoration[J/OL]. arXiv (Cornell University), 2025. https://doi.org/10.48550/arxiv.2503.09403.

[12] Zhou J, Others. Q-Agent: Quality-Driven Agent for Image Restoration[J]. arXiv preprint, 2025.

[13] Wu G, Jiang J, Jiang K, Liu X, Nie L. DSwinIR: Rethinking Window-based Attention for Image Restoration[J/OL]. arXiv (Cornell University), 2025. https://doi.org/10.48550/arxiv.2504.04869.

[14] Conde M V, Choi U-J, Burchi M, Timofte R. Swin2SR: SwinV2 Transformer for Compressed Image Super-Resolution and Restoration[J/OL]. Lecture notes in computer science, 2023. https://doi.org/10.1007/978-3-031-25063-7_42.

[15] Conde M V, Choi U, Burchi M, Timofte R. Swin2SR: SwinV2 Transformer for Compressed Image Super-Resolution and Restoration[J/OL]. arXiv (Cornell University), 2022. https://doi.org/10.48550/arxiv.2209.11345.

[16] Jing T, Liu C, Chen Y. A Lightweight Single-Image Super-Resolution Method Based on the Parallel Connection of Convolution and Swin Transformer Blocks[J/OL]. Applied Sciences, 2025. https://doi.org/10.3390/app15041806.

[17] Wang Z, Cun X, Bao J, Zhou W, Liu J, Li H. Uformer: A General U-Shaped Transformer for Image Restoration[C/OL] // 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2022. https://doi.org/10.1109/cvpr52688.2022.01716.

[18] Chen S, Tian Y, Liu Y, Chen E. Dual-former: Hybrid Self-attention Transformer for Efficient Image Restoration[J/OL]. arXiv (Cornell University), 2022. https://doi.org/10.48550/arxiv.2210.01069.

[19] Chen X, Pan J, Lu J, Fan Z, Li H. Hybrid CNN-Transformer Feature Fusion for Single Image Deraining[C/OL] // Proceedings of the AAAI Conference on Artificial Intelligence. 2023. https://doi.org/10.1609/aaai.v37i1.25111.

[20] Shi J, Wei B, Zhou G, Zhang L. Sandformer: CNN and Transformer under Gated Fusion for Sand Dust Image Restoration[J/OL], 2023. https://doi.org/10.1109/icassp49357.2023.10095242.

[21] Kong L, Dong J, Ge J, Li M, Pan J. Efficient Frequency Domain-based Transformers for High-Quality Image Deblurring[J/OL], 2023. https://doi.org/10.1109/cvpr52729.2023.00570.

[22] Jiang X, Zhang X, Gao N, Deng Y. When Fast Fourier Transform Meets Transformer for Image Restoration[J/OL]. Lecture notes in computer science, 2024. https://doi.org/10.1007/978-3-031-72995-9_22.

[23] Mao X, Liu Y, Liu F, Li Q, Shen W, Wang Y. Intriguing Findings of Frequency Selection for Image Deblurring[C/OL] // Proceedings of the AAAI Conference on Artificial Intelligence. 2023. https://doi.org/10.1609/aaai.v37i2.25281.

[24] Shi Y, Xia B, Jin X, Wang X, Zhao T, Xia X, Xiao X, Yang W. VmambaIR: Visual State Space Model for Image Restoration[J/OL]. IEEE Transactions on Circuits and Systems for Video Technology, 2025. https://doi.org/10.1109/tcsvt.2025.3530090.

[25] Lee E, Hwang Y. Decomformer: Decompose Self-Attention of Transformer for Efficient Image Restoration[J/OL]. IEEE Access, 2024. https://doi.org/10.1109/access.2024.3375360.

[26] Ghasemabadi A, Salameh M, Janjua M K, Zhou C, Sun F, Niu D. CascadedGaze: Efficiency in Global Context Extraction for Image Restoration[J/OL]. arXiv (Cornell University), 2024. https://doi.org/10.48550/arxiv.2401.15235.

[27] Zhang H, Zhang X, Cai N, Di J, Zhang Y. Joint Multi-Dimensional Dynamic Attention and Transformer for General Image Restoration[J/OL]. SSRN Electronic Journal, 2024. https://doi.org/10.2139/ssrn.5018208.

[28] Cui Y, Ren W, Knoll A. Omni-Kernel Network for Image Restoration[C/OL] //Proceedings of the AAAI Conference on Artificial Intelligence. 2024. https://doi.org/10.1609/aaai.v38i2.27907.

[29] Chen Q, Zheng B, Yan C, Zhu Z, Wang T, Slabaugh G, Yuan S. GoLDFormer: A global–local deformable window transformer for efficient image restoration[J/OL]. Journal of Visual Communication and Image Representation, 2024. https://doi.org/10.1016/j.jvcir.2024.104117.

[30] Gao H, Dang D. Mixed Hierarchy Network for Image Restoration[J/OL]. arXiv (Cornell University), 2023. https://doi.org/10.48550/arxiv.2302.09554.

[31] Zhang D, Zhou F. Self-Supervised Image Denoising for Real-World Images With Context-Aware Transformer[J/OL]. IEEE Access, 2023. https://doi.org/10.1109/access.2023.3243829.

[32] Ali A M, Benjdira B, Koubâa A, El-Shafai W, Khan Z, Boulila W. Vision Transformers in Image Restoration: A Survey[J/OL]. Sensors, 2023. https://doi.org/10.3390/s23052385.

[33] Mei Y, Fan Y, Zhang Y, Yu J, Zhou Y, Liu D, Fu Y, Huang T S, Shi H. Pyramid Attention Network for Image Restoration[J/OL]. International Journal of Computer Vision, 2023. https://doi.org/10.1007/s11263-023-01843-5.

[34] Gao T, Wen Y, Zhang K, Cheng P, Chen T. Towards an Effective and Efficient Transformer for Rain-by-Snow Weather Removal[J/OL]. SSRN Electronic Journal, 2023. https://doi.org/10.2139/ssrn.4458244.

[35] Hu B, Yang S, Liu F, Ding W. Collaborative Semantic Contrastive for All-in-one Image Restoration[J/OL]. Engineering Applications of Artificial Intelligence, 2025. https://doi.org/10.1016/j.engappai.2025.110017.

[36] Wu G, Jiang J, Jiang K, Liu X. Learning from History: Task-agnostic Model Contrastive Learning for Image Restoration[C/OL] //Proceedings of the AAAI Conference on Artificial Intelligence. 2024. https://doi.org/10.1609/aaai.v38i6.28412.

[37] Ma J, Cheng T, Wang G, Zhang Q, Wang X, Zhang L. ProRes: Exploring Degradation-aware Visual Prompt for Universal Image Restoration[J/OL]. arXiv (Cornell University), 2023. https://doi.org/10.48550/arxiv.2306.13653.

[38] Li Z, Lei Y, Ma C, Zhang J, Shan H. Prompt-In-Prompt Learning for Universal Image Restoration[J/OL]. arXiv (Cornell University), 2023. https://doi.org/10.48550/arxiv.2312.05038.

[39] Wu G, Jiang J, Jiang K, Liu X, Nie L. Learning Dynamic Prompts for All-in-One Image Restoration[J/OL]. IEEE Transactions on Image Processing, 2025. https://doi.org/10.1109/tip.2025.3567205.

[40] Sun W, Wang Q, Wang Y, Hou Z, Yan Q, Yan S. AdaPrompt-IR: Adaptive learning to perceive degradation semantic and prompting for all-in-one image restoration[J/OL]. Pattern Recognition, 2025. https://doi.org/10.1016/j.patcog.2025.111875.

[41] Liu M, Yang W, Luo J, Liu J. UP-Restorer: When Unrolling Meets Prompts for Unified Image Restoration[C/OL] //Proceedings of the AAAI Conference on Artificial Intelligence. 2025. https://doi.org/10.1609/aaai.v39i5.32587.

[42] Gao H, Yang J, Zhang Y, Wang N, Yang J, Dang D. Prompt-Based Ingredient-Oriented All-in-One Image Restoration[J/OL]. IEEE Transactions on Circuits and Systems for Video Technology, 2024. https://doi.org/10.1109/tcsvt.2024.3398810.

[43] Wu Z, Liu W, Wang J, Li J, Huang D. FrePrompter: Frequency self-prompt for all-in-one image restoration[J/OL]. Pattern Recognition, 2024. https://doi.org/10.1016/j.patcog.2024.111223.

[44] Wu G, Jiang J, Jiang K, Liu X, Nie L. Beyond Degradation Redundancy: Contrastive Prompt Learning for All-in-One Image Restoration[J/OL]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025. https://doi.org/10.1109/tpami.2025.3642852.

[45] Jiang A, Chen H, Ye J, Wang M, Liu B. Multi-Dimension Visual Prompt Enhanced Image Restoration Network Via Mamba-Transformer Aggregation[J/OL]. SSRN Electronic Journal, 2025. https://doi.org/10.2139/ssrn.5311960.

[46] Duan H, Min X, Wu S, Shen W, Zhai G. UniProcessor: A Text-Induced Unified Low-Level Image Processor[J/OL]. Lecture notes in computer science, 2024. https://doi.org/10.1007/978-3-031-72855-6_11.

[47] Xin L, Zhou Y, Yue J, Ren C, Chan K C K, Lu Q, Yang M-H. Re-Boosting Self-Collaboration Parallel Prompt GAN for Unsupervised Image Restoration[J/OL]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025. https://doi.org/10.1109/tpami.2025.3589606.

[48] Wu G, Jiang J, Jiang K, Liu X. Harmony in Diversity: Improving All-in-One Image Restoration via Multi-Task Collaboration[J/OL], 2024. https://doi.org/10.1145/3664647.3680762.

[49] Zamfir E, Wu Z, Mehta N, Tan Y, Paudel D P, Zhang Y, Timofte R. Complexity Experts are Task-Discriminative Learners for Any Image Restoration[J/OL], 2025. https://doi.org/10.1109/cvpr52734.2025.01190.

[50] Chen Y-T, Pei S-C. Always Clear Days: Degradation Type and Severity Aware All-in-One Adverse Weather Removal[J/OL]. IEEE Access, 2025. https://doi.org/10.1109/access.2025.3526168.

[51] Tian X, Liao X, Liu X, Li M, Ren C. Degradation-Aware Feature Perturbation for All-in-One Image Restoration[J/OL], 2025. https://doi.org/10.1109/cvpr52734.2025.02623.

[52] Zhu R, Tu Z, Liu J, Bovik A C, Fan Y. MWFormer: Multi-Weather Image Restoration Using Degradation-Aware Transformers[J/OL]. IEEE Transactions on Image Processing, 2024. https://doi.org/10.1109/tip.2024.3501855.

[53] Wu G, Jiang J, Wang Y, Jiang K, Liu X. Debiased All-in-one Image Restoration with Task Uncertainty Regularization[C/OL] //Proceedings of the AAAI Conference on Artificial Intelligence. 2025. https://doi.org/10.1609/aaai.v39i8.32905.

[54] Tang A H, Wu Y, Zhang Y. RamIR: Reasoning and action prompting with Mamba for all-in-one image restoration[J/OL]. Applied Intelligence, 2025. https://doi.org/10.1007/s10489-024-06226-y.

[55] Dudhane A, Thawakar O, Zamir S W, Khan S, Khan F S, Yang M-H. Dynamic Pre-training: Towards Efficient and Scalable All-in-One Image Restoration[J/OL]. arXiv (Cornell University), 2024. https://doi.org/10.48550/arxiv.2404.02154.

[56] Li C, Sun F, Zhou H, Xie Y, Li Z, Zhu L. Multi-Weather Restoration: An Efficient Prompt-Guided Convolution Architecture[J/OL]. IEEE Transactions on Circuits and Systems for Video Technology, 2024. https://doi.org/10.1109/tcsvt.2024.3469190.

[57] Chen L, Xiong Q, Zhang W, Liang X, Gan Z, Li L, He X. Multi-modal degradation feature learning for unified image restoration based on contrastive learning[J/OL]. Neurocomputing, 2024. https://doi.org/10.1016/j.neucom.2024.128955.

[58] Yang S, Hu B, Liu F, Wu X, Ding W, Zhou J. IPT-ILR: Image Pyramid Transformer Coupled With Information Loss Regularization for All-in-One Image Restoration[J/OL]. IEEE Transactions on Circuits and Systems for Video Technology, 2024. https://doi.org/10.1109/tcsvt.2024.3519352.

[59] Zhang C, Zhu Y, Yan Q, Sun J, Zhang Y. All-in-one Multi-degradation Image Restoration Network via Hierarchical Degradation Representation[J/OL]. arXiv (Cornell University), 2023. https://doi.org/10.48550/arxiv.2308.03021.

[60] Chen Z, Wang C, Zhang F, Zhang L, Grau A, Guerra E. All-in-one aerial image enhancement network for forest scenes[J/OL]. Frontiers in Plant Science, 2023. https://doi.org/10.3389/fpls.2023.1154176.

[61] Zhang X, Zhang H, Wang G, Zhang Q, Zhang L, Du B. UniUIR: Considering Underwater Image Restoration as an All-in-One Learner[J/OL]. IEEE Transactions on Image Processing, 2025. https://doi.org/10.1109/tip.2025.3618377.

[62] Yang D H, Xu W, Gao Y, Lu Y, Zhang J, Guo Y. MvKSR: Multi-view Knowledge-guided Scene Recovery for Hazy and Rainy Degradation[J/OL]. arXiv (Cornell University), 2024. https://doi.org/10.48550/arxiv.2401.03800.

[63] Liu R W, Lu Y, Gao Y, Guo Y, Ren W, Zhu F, Wang F. Real-Time Multi-Scene Visibility Enhancement for Promoting Navigational Safety of Vessels Under Complex Weather Conditions[J/OL]. arXiv (Cornell University), 2024. https://doi.org/10.48550/arxiv.2409.01500.

[64] Tang F, Zhu X, Hu J, Tie J, Zhou J, Fu Y. Generative Adversarial Unsupervised Image Restoration in Hybrid Degradation Scenes[J/OL]. arXiv preprint, 2022. https://doi.org/10.20944/preprints202202.0159.v1.

[65] Liang S, Gao T, Chen T, Wen Y, Zhang Q, Wang X. INP-Net: Implicit Neural Prompting Network for Remote Sensing Image Dehazing[J/OL]. IEEE Transactions on Geoscience and Remote Sensing, 2025. https://doi.org/10.1109/tgrs.2025.3649014.

[66] Zeng H, Li J, Zheng Z, Xiong Z. All-in-One Image Compression and Restoration[J/OL], 2025. https://doi.org/10.1109/wacv61041.2025.00069.

[67] Zhang S, Dong Q, Mao W, Wang Z. A Unified Accelerator for All-in-One Image Restoration Based on Prompt Degradation Learning[J/OL]. IEEE Transactions on Circuits and Systems I Regular Papers, 2025. https://doi.org/10.1109/tcsi.2024.3519532.

[68] Tang X, Gu X, He X, Hu X, Sun J. Degradation-Aware Residual-Conditioned Optimal Transport for Unified Image Restoration[J/OL]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025. https://doi.org/10.1109/tpami.2025.3562211.

[69] Zeng H, Wang X, Chen Y, Su J, Liu J. Vision-Language Gradient Descent-driven All-in-One Deep Unfolding Networks[J/OL], 2025. https://doi.org/10.1109/cvpr52734.2025.00705.

[70] Cheng Y, Shao M, Wan Y, Wang C. RDM-IR: Task-adaptive deep unfolding network for All-In-One image restoration[J/OL]. Knowledge-Based Systems, 2024. https://doi.org/10.1016/j.knosys.2024.112543.

[71] Lihe Z, Yuan Q, He J, Jin X, Xiao Y, Chen Y, Shen H, Zhang L. Ada4DIR: An adaptive model-driven all-in-one image restoration network for remote sensing images[J/OL]. Information Fusion, 2025. https://doi.org/10.1016/j.inffus.2025.102930.

[72] Jiang J, Zuo Z, Wu G, Jiang K, Liu X. A Survey on All-in-One Image Restoration: Taxonomy, Evaluation and Future Trends[J/OL]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025. https://doi.org/10.1109/tpami.2025.3598132.

[73] Saharia C, Chan W, Chang H, Lee C, Ho J, Salimans T, Fleet D J, Norouzi M. Palette: Image-to-Image Diffusion Models[J/OL], 2022. https://doi.org/10.1145/3528233.3530757.

[74] Zhang Y, Shi X, Li D, Wang X, Wang J, Li H. A Unified Conditional Framework for Diffusion-based Image Restoration[J/OL]. arXiv (Cornell University), 2023. https://doi.org/10.48550/arxiv.2305.20049.

[75] Liu J, Wang Q, Fan H, Wang Y, Tang Y, Qu L. Residual Denoising Diffusion Models[J/OL], 2024. https://doi.org/10.1109/cvpr52733.2024.00268.

[76] Zheng D, Wu X, Yang S, Zhang J, Hu J-F, Zheng W-S. Selective Hourglass Mapping for Universal Image Restoration Based on Diffusion Model[J/OL]. arXiv (Cornell University), 2024. https://doi.org/10.48550/arxiv.2403.11157.

[77] Ding Z, Zhang X, Tu Z, Xia Z. Restoration by Generation with Constrained Priors[J/OL], 2024. https://doi.org/10.1109/cvpr52733.2024.00248.

[78] Lin X, He J, Chen Z, Lyu Z, Fei B, Dai B, Ouyang W, Qiao Y, Dong C. DiffBIR: Towards Blind Image Restoration with Generative Diffusion Prior[J/OL]. arXiv (Cornell University), 2023. https://doi.org/10.48550/arxiv.2308.15070.

[79] Yang T, Wu R, Ren P, Xie X, Zhang L. Pixel-Aware Stable Diffusion for Realistic Image Super-Resolution and Personalized Stylization[J/OL]. Lecture notes in computer science, 2024. https://doi.org/10.1007/978-3-031-73247-8_5.

[80] Tian Y, Chen S, Chai W, Xing Z, Qin J, Ge L, Zhu L. Learning Diffusion Texture Priors for Image Restoration[J/OL], 2024. https://doi.org/10.1109/cvpr52733.2024.00244.

[81] Xu X, Kong S, Hu T, Liu Z, Bao H. Boosting Image Restoration via Priors from Pre-Trained Models[J/OL], 2024. https://doi.org/10.1109/cvpr52733.2024.00280.

[82] Guo L, Wang C, Yang W, Huang S, Wang Y, Pfister H, Wen B. ShadowDiffusion: When Degradation Prior Meets Diffusion Model for Shadow Removal[J/OL], 2023. https://doi.org/10.1109/cvpr52729.2023.01350.

[83] Yi X, Xu H, Zhang H, Tang L, Ma J. Diff-Retinex++: Retinex-Driven Reinforced Diffusion Model for Low-Light Image Enhancement[J/OL]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025. https://doi.org/10.1109/tpami.2025.3563612.

[84] Shang S, Shan Z, Liu G, Wang L, Wang X, Zhang Z, Zhang J. ResDiff: Combining CNN and Diffusion Model for Image Super-resolution[C/OL] // Proceedings of the AAAI Conference on Artificial Intelligence. 2024. https://doi.org/10.1609/aaai.v38i8.28746.

[85] Luo Z, Gustafsson F, Zhao Z, Sjölund J, Schön T B. Refusion: Enabling Large-Size Realistic Image Restoration with Latent-Space Diffusion Models[J/OL]. arXiv (Cornell University), 2023. https://doi.org/10.48550/arxiv.2304.08291.

[86] Wang Y, Yu J, Yu R, Zhang J. Unlimited-Size Diffusion Restoration[J/OL], 2023. https://doi.org/10.1109/cvprw59228.2023.00123.

[87] Zhao Y, Hou T, Su Y-C, Jia X, Li Y, Grundmann M. Towards Authentic Face Restoration with Iterative Diffusion Models and Beyond[J/OL], 2023. https://doi.org/10.1109/iccv51070.2023.00672.

[88] Welker S, Chapman H N, Gerkmann T. DriftRec: Adapting Diffusion Models to Blind JPEG Restoration[J/OL]. IEEE Transactions on Image Processing, 2024. https://doi.org/10.1109/tip.2024.3383776.

[89] Kang S, Gao S, Wu W, Wang X, Wang S, Qiu G. Image Intrinsic Components Guided Conditional Diffusion Model for Low-Light Image Enhancement[J/OL]. IEEE Transactions on Circuits and Systems for Video Technology, 2024. https://doi.org/10.1109/tcsvt.2024.3441713.

[90] Yan Q, Hu T, Wu P, Dai D, Gu S, Dong W, Zhang Y. Efficient Image Enhancement With a Diffusion-Based Frequency Prior[J/OL]. IEEE Transactions on Circuits and Systems for Video Technology, 2025. https://doi.org/10.1109/tcsvt.2025.3549351.

[91] Tu L, Wu J, Luo X, Jin Z. Unifying Heterogeneous Degradations: Uncertainty-Aware Diffusion Bridge Model for All-in-One Image Restoration[J/OL]. arXiv (Cornell University), 2026. https://doi.org/10.48550/arxiv.2601.21592.

[92] Luo W, Qin H, Chen Z, Wang L, Zheng D, Li Y, Liu Y, Li B, Hu W. Visual-Instructed Degradation Diffusion for All-in-One Image Restoration[J/OL], 2025. https://doi.org/10.1109/cvpr52734.2025.01191.

[93] Zhang Y, Zhang H, Chai X, Cheng Z, Xie R, Song L, Zhang W. Diff-Restorer: Unleashing Visual Prompts for Diffusion-based Universal Image Restoration[J/OL]. IEEE Transactions on Circuits and Systems for Video Technology, 2025. https://doi.org/10.1109/tcsvt.2025.3629686.

[94] Lv X, Shao M, Wan Y, Qiao Y, Wang C. Adaptive prompt guided unified image restoration with latent diffusion model[J/OL]. Engineering Applications of Artificial Intelligence, 2025. https://doi.org/10.1016/j.engappai.2025.110267.

[95] Yue Y, Meng Y, Yang L, Yang Y. Joint Conditional Diffusion Model for Image Restoration with Mixed Degradations[J/OL]. arXiv (Cornell University), 2024. https://doi.org/10.48550/arxiv.2404.07770.

[96] Zhu K, Gu J, You Z, Qiao Y, Dong C. An Intelligent Agentic System for Complex Image Restoration Problems[J/OL]. arXiv (Cornell University), 2024. https://doi.org/10.48550/arxiv.2410.17809.

[97] Tripathi M, Kongprawechnon W, Yahia S B. Advancements in Agentic Ai Architecture and Prompting Strategies for Low-Light Enhancement[J/OL]. SSRN Electronic Journal, 2025. https://doi.org/10.2139/ssrn.5351735.

[98] Kirillov A M, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, Xiao T, Whitehead S, Berg A C, Lo W-Y. Segment Anything[J/OL], 2023. https://doi.org/10.1109/iccv51070.2023.00371.

[99] You Z, Gu J, Li Z, Kong X, Dong C. DepictQA: Depicted Image Quality Assessment with Vision Language Models[J]. arXiv preprint, 2023.

[100] Chen W, Vong Y-J, Kuo S, Ma S, Wang J. RobustSAM: Segment Anything Robustly on Degraded Images[J/OL]. arXiv (Cornell University), 2024. https://doi.org/10.48550/arxiv.2406.09627.

[101] Ren T, Liu S, Zeng A. Grounded SAM: Assembling Open-World Models for Diverse Visual Tasks[J/OL]. arXiv (Cornell University), 2024. https://doi.org/10.48550/arxiv.2401.14159.

[102] You Z, Li Z, Gu J, Yin Z, Xue T, Dong C. Depicting Beyond Scores: Advancing Image Quality Assessment Through Multi-modal Language Models[J/OL]. Lecture notes in computer science, 2024. https://doi.org/10.1007/978-3-031-72970-6_15.

[103] You Z, Gu J, Li Z, Cai X, Zhu K, Xue T, Dong C. Enhancing Descriptive Image Quality Assessment with A Large-scale Multi-modal Dataset[J/OL]. arXiv (Cornell University), 2024. https://doi.org/10.48550/arxiv.2405.18842.

[104] Wu H, Zhang Z. Q-Instruct: Improving Low-Level Visual Abilities for Multi-Modality Foundation Models[J/OL], 2024. https://doi.org/10.1109/cvpr52733.2024.02408.

[105] Wu H, Zhang Z, Zhang E. Q-Bench: A Benchmark for General-Purpose Foundation Models on Low-level Vision[J/OL]. arXiv (Cornell University), 2023. https://doi.org/10.48550/arxiv.2309.14181.

[106] 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[J/OL]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024. https://doi.org/10.1109/tpami.2024.3445770.

[107] Zhang Z, Wu H. Q-Boost: On Visual Quality Assessment Ability of Low-Level Multi-Modality Foundation Models[J/OL], 2024. https://doi.org/10.1109/icmew63481.2024.10645451.

[108] Wu T, Ma K, Liang J, Yang Y, Zhang L. A Comprehensive Study of Multimodal Large Language Models for Image Quality Assessment[J/OL]. Lecture notes in computer science, 2024. https://doi.org/10.1007/978-3-031-72904-1_9.

[109] Ai Y, Zhou X, Huang H, Han X, Chen Z, You Q, Yang H. DreamClear: High-Capacity Real-World Image Restoration with Privacy-Safe Dataset Curation[J/OL]. arXiv (Cornell University), 2024. https://doi.org/10.48550/arxiv.2410.18666.

[110] Luo Z, Gustafsson F, Zhao Z, Sjolund J, Schon T B. Photo-Realistic Image Restoration in the Wild with Controlled Vision-Language Models[J/OL], 2024. https://doi.org/10.1109/cvprw63382.2024.00658.

[111] Chen X, kang Mao D, Ke J. Real-world super-resolution with VLM-based degradation prior learning[J/OL]. Scientific Reports, 2025. https://doi.org/10.1038/s41598-025-14581-0.

[112] Cui Y, Zhu J, Knoll A. Enhancing Perception for Autonomous Vehicles: A Multi-Scale Feature Modulation Network for Image Restoration[J/OL]. IEEE Transactions on Intelligent Transportation Systems, 2025. https://doi.org/10.1109/tits.2025.3538485.

[113] Talebi H, Milanfar P. NIMA: Neural Image Assessment[J/OL]. IEEE Transactions on Image Processing, 2018. https://doi.org/10.1109/tip.2018.2831899.

[114] Zhang R, Isola P, Efros A A, Shechtman E, Wang O. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric[J/OL]. arXiv (Cornell University), 2018. https://doi.org/10.48550/arxiv.1801.03924.

[115] Zheng H, Yang H, Fu J, Zha Z-J, Luo J. Learning Conditional Knowledge Distillation for Degraded-Reference Image Quality Assessment[C/OL] // 2021 IEEE/CVF International Conference on Computer Vision (ICCV). 2021. https://doi.org/10.1109/iccv48922.2021.01008.

[116] Li J, Yan B, Lin Q, Li A, Ma C. Motion Blur Removal With Quality Assessment Guidance[J/OL]. IEEE Transactions on Multimedia, 2021. https://doi.org/10.1109/tmm.2021.3068561.

[117] Zhang Y, Yan D. Soft Knowledge Distillation with Multi-Dimensional Cross-Net Attention for Image Restoration Models Compression[J/OL], 2025. https://doi.org/10.1109/icassp49660.2025.10887587.

[118] Yang S, Hu B, Wu X, Liu F, Zhou J. Image Restoration Model Compression via Mamba-oriented Heterogeneous Knowledge Distillation[J/OL]. SSRN Electronic Journal, 2025. https://doi.org/10.2139/ssrn.5407564.

[119] Zhang Y, Yan D. Knowledge Distillation for Image Restoration: Simultaneous Learning from Degraded and Clean Images[J/OL], 2025. https://doi.org/10.1109/icassp49660.2025.10889105.

[120] Zhang Y. Simultaneous Learning Knowledge Distillation for Image Restoration: Efficient Model Compression for Drones[J/OL]. Drones, 2025. https://doi.org/10.3390/drones9030209.

[121] Wang P, Luo X, Xie Y, Qu Y. Data-free Distillation with Degradation-prompt Diffusion for Multi-weather Image Restoration[J/OL]. arXiv (Cornell University), 2024. https://doi.org/10.48550/arxiv.2409.03455.

[122] Zhai L, Wang Y, Cui S, Zhou Y. A Comprehensive Review of Deep Learning-Based Real-World Image Restoration[J/OL]. IEEE Access, 2023. https://doi.org/10.1109/access.2023.3250616.

[123] Guan Q, Yang Q, Chen X, Song T, Jin G, Jin J. WeatherBench: A Real-World Benchmark Dataset for All-in-One Adverse Weather Image Restoration[J/OL], 2025. https://doi.org/10.1145/3746027.3758196.