4KAgent: Agentic Any Image to 4K Super-Resolution

Yushen Zuo[1], **Qi Zheng**[1†], **Mingyang Wu**[1†], **Xinrui Jiang**[2†], **Renjie Li**[1], **Jian Wang**[3], **Yide Zhang**[4], **Gengchen Mai**[5], **Lihong V. Wang**[6], **James Zou**[2], **Xiaoyu Wang**[7], **Ming-Hsuan Yang**[8], **Zhengzhong Tu**[1⋆]

[1]Texas A&M University [2]Stanford University [3]Snap Inc. [4]CU Boulder
[5]UT Austin [6]California Institute of Technology [7]Topaz Labs [8]UC Merced

[⋆]Corresponding Author: tzz@tamu.edu. [†]Equal contributions.

Project Website: 4kagent.github.io

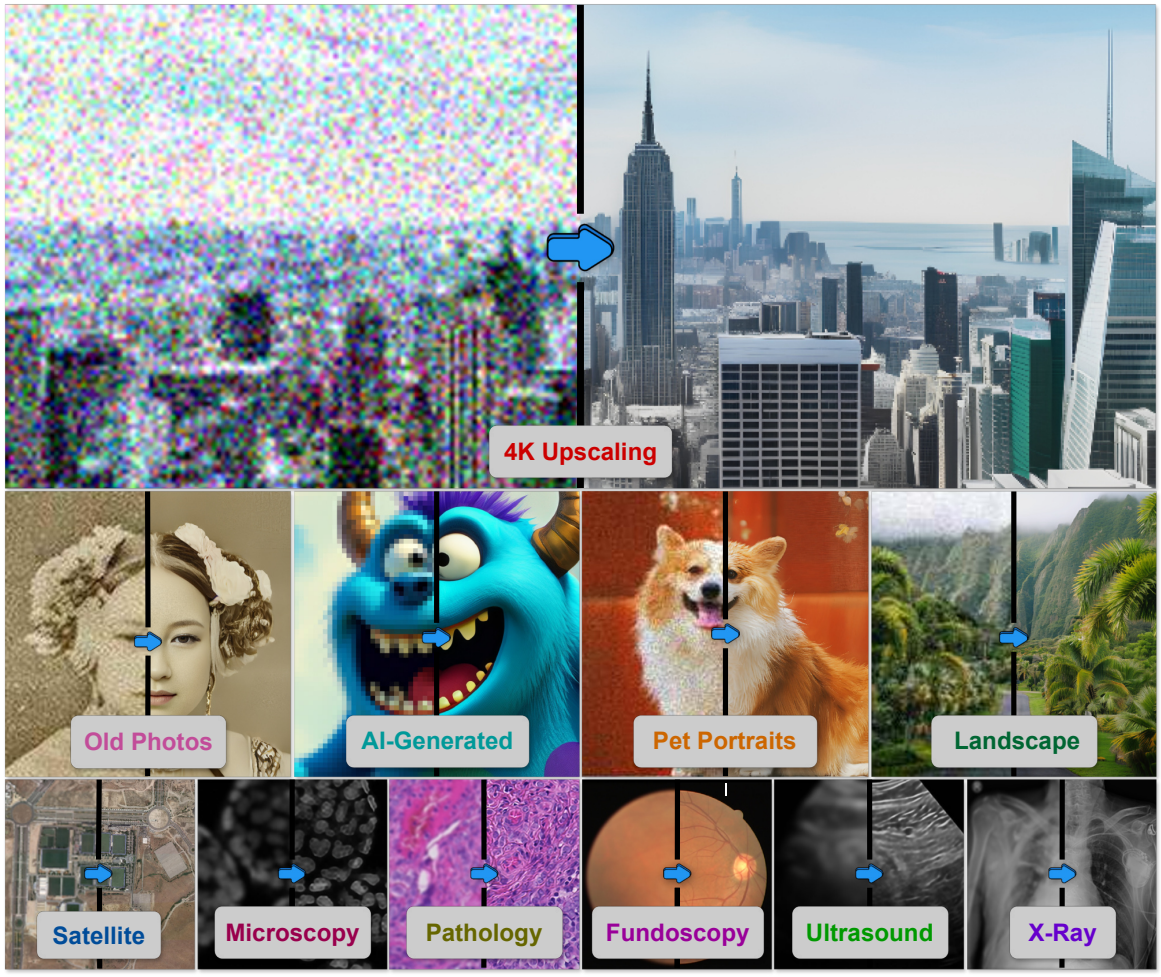


Figure 1: We present 4KAgent, an agentic image super-resolution generalist designed to universally upscale **any image to 4K**, regardless of input *type*, *degradation level*, or *domain*. That is, **4KAgent** effectively restores diverse imagery, spanning from natural scenes, severely degraded captures (*e.g.*, old photos), human/pet portraits, AI-generated content (AIGC), as well as *specialized scientific imaging domains*, such as remote sensing, fluorescence microscopy, pathology, and various medical modalities like X-ray, ultrasound, and funduscopy—all **without the need** for any re-training or domain-specific adaptation.

39th Conference on Neural Information Processing Systems (NeurIPS 2025).

Abstract

We present **4KAgent**, a unified agentic super-resolution generalist system designed to universally upscale any image to 4K resolution (and even higher, if applied iteratively). Our system can transform images from extremely low resolutions with severe degradations, for example, highly distorted inputs at 256×256, into crystal-clear, photorealistic 4K outputs. 4KAgent comprises three core components: **(1)** *Profiling*, a module that customizes the 4KAgent pipeline based on bespoke use cases; **(2)** A *Perception Agent*, which leverages vision-language models alongside image quality assessment experts to analyze the input image and make a tailored restoration plan; and **(3)** A *Restoration Agent*, which executes the plan, following a recursive execution-reflection paradigm, guided by a quality-driven mixture-of-experts policy to select the optimal output for each step. Additionally, 4KAgent embeds a specialized face restoration pipeline, significantly enhancing facial details in portrait and selfie photos. We rigorously evaluate our 4KAgent across **11** distinct task categories encompassing a total of **26** diverse benchmarks, setting new state-of-the-art on a broad spectrum of imaging domains. Our evaluations cover natural images, portrait photos, AI-generated content, satellite imagery, fluorescence microscopy, and medical imaging like fundoscopy, ultrasound, and X-ray, demonstrating superior performance in terms of both perceptual (*e.g.*, NIQE, MUSIQ) and fidelity (*e.g.*, PSNR) metrics. By establishing a novel agentic paradigm for low-level vision tasks, we aim to catalyze broader interest and innovation within vision-centric autonomous agents across diverse research communities. We release all the code, models, and results at: https://4kagent.github.io.

1 Introduction

Image super-resolution (SR) is a fundamental task in computer vision that aims to reconstruct high-resolution (HR) images from their low-resolution (LR) counterparts [232, 45, 46, 98, 207, 267, 202, 201, 168, 195]. It serves as a bedrock for various low-level vision tasks [257, 119, 246, 186], including deblurring [182, 33], dehazing [63, 106], deraining [165, 79], and low-light enhancement [211, 59]. Beyond its classical role in computational photography and imaging, SR techniques significantly influence numerous domains, such as biomedical imaging [55, 169], remote sensing [171, 64, 92], surveillance [259], and embodied artificial intelligence applications [60, 170, 75].

Traditional SR methods [45, 202] typically assume known synthetic degradation during training, which limits their generalization to real-world captures that suffer from complex, heterogeneous, and unpredictable degradations [201]. Recent research has increasingly shifted to a more practical real-world super-resolution (RealSR) task [21, 216], which attempts to explicitly address diverse and unknown degradations in naturally captured photographs and videos. RealSR requires models not only to handle multiple combined degradations effectively but also to exhibit strong adaptability and generalization across varied scenarios [238, 187]. Many effective solutions have been proposed to solve the RealSR problem, via simulating complex real-world degradations [256, 201], leveraging the powerful generative prior of pre-trained diffusion models [81, 216, 217, 179, 144], enabling robust restoration under unknown conditions. Inspired by the advanced planning and reasoning capabilities of large language models (LLMs) [213, 71, 237, 48], agentic restoration frameworks [22, 274] have emerged as an advanced tool that can adaptively handle multiple degradations through sequential planning and dynamic restoration strategies.

Despite their successes in certain scenarios, existing performant generative approaches [216, 179] can only handle limited degradation ranges, *e.g.*, up to $4\times$ upscaling, failing to recover extremely low-quality images with highly complex and diverse degradations in the wild. Moreover, SR specialist models are known to generalize poorly to out-of-distribution domains [23], let alone when applied to a different scaling factor. This is mainly due to heavy reliance on supervised learning on synthetic image pairs that cannot fully capture the complex real-world image degradations, not to mention other domains, ranging from AI-generated imagery, scientific computing, to biomedical images. Last but not least, practically, users often demand highly specific workflows, *e.g.*, either denoising, upscaling, or prioritizing high fidelity over perceptual quality, and hence a one-size-fits-all system that can flexibly adapt to satisfy diverse user needs and application scenarios is in pressing need.

To fill this gap, we present **4KAgent**, the first-of-its-kind agentic framework for generic, flexible, and interpretable super-resolution of **any image to 4K**. As illustrated in Fig. 1, 4KAgent is capable of upscaling any low-resolution image (*e.g.*, 0.065 megapixels) to 4K×4K, (*i.e.*, 16.7 megapixels) via a 16× upscaling factor[1] (§3.4). It also sets new state-of-the-art (SoTA) on classical image super-resolution (§3.1), real-world image super-resolution (§3.2), face restoration (Appendix), and multiple-degradation image restoration (§3.3) benchmarks, in terms of perceptual quality. We also show that 4KAgent generalizes across widespread applications in low-level tasks, such as joint restoration & 4K upscaling (§3.5), and AI-generated content 4K upscaling (§3.6). Lastly, thanks to the mixture-of-experts and profile design, 4KAgent demonstrates broader impacts on interdisciplinary areas such as scientific super-resolution (§3.7), including ❶ Satellite image super-resolution, ❷ fluorescence microscopy super-resolution, and ❸ medical image super-resolution.

Our contributions are as follows:

- **[Framework]** We present **4KAgent**, the first AI agent framework for universal any-image-to-4K upscaling, capable of handling **all image categories**, ranging from classical and realistic degraded and extremely low-quality images to AI-generated imagery and scientific imaging domains such as remote sensing, microscopy, and biomedical imaging.
- **[System Design]** We design a multi-agent system in 4KAgent. The **Perception Agent** employs large vision-language models (VLMs) to analyze the content and distortion in the image and provide the restoration plan for the restoration agent to execute. The **Restoration Agent**, which sets up an execution–reflection–rollback procedure for recursive restoration and upscaling.
- **[Q-MoE & Face Restoration pipeline]** In each restoration step of the restoration plan, we propose a Quality-Driven Mixture-of-Expert (**Q-MoE**) policy in execution and reflection to select the optimal image. We further develop a face restoration pipeline to enhance faces in images.
- **[Profile Module]** To expand the applicability of 4KAgent, we propose a **Profile Module** to bring the availability to customize the system for different restoration tasks. 4KAgent can adapt to different restoration tasks without extra training.
- **[DIV4K-50 Dataset]** To evaluate 4K super-resolution performances, we build the **DIV4K-50** dataset as a challenging test set to upscale a low-quality (LQ) image in 256×256 resolution with multiple degradations to a high-quality (HQ) 4K image in 4096×4096 resolution.
- **[Experiments]** Extensive experimental results demonstrate the superiority of 4KAgent as a **generalist 4K upscaling agent**: 4KAgent sets new state-of-the-art on a variety of real-world image super-resolution benchmarks, multiple-degradation restoration benchmarks, face restoration, 4K upscaling task, and various scientific imaging tasks, including satellite image super-resolution, fluorescence microscopic imaging, X-ray radiography, and biomedical imaging super-resolution.

2 Method

2.1 System Overview

We introduce **4KAgent**, a multi-agent framework designed to upscale arbitrary images to 4K resolution. Fig. 2 illustrates the overall workflow of our proposed 4KAgent, which decomposes the restoration pipeline into a collection of specialized agents. The **Perception Agent** analyzes degradations (noise, blur, *etc.*), extracts semantic and structural cues, and schedules a restoration plan containing a sequence of operators (denoising, deblurring, super-resolution, *etc.*). The **Restoration Agent** follows the restoration plan using our proposed Quality-Driven Mixture-of-Experts (Q-MoE) to select the best output from multiple restoration tools. The rollback mechanism is activated if the quality of the restored image falls below a threshold. Additionally, a dedicated **Face Restoration Pipeline** further enhances facial regions by triggering expert face restoration models. A user-configurable **Profile Module** allows users to customize the system (*e.g.*, prioritize fidelity or perceptual quality), enabling robust, high-quality 4K SR across diverse content and degradation types.

2.2 Perception Agent

The **Perception Agent** is designed as a four-stage analytical module that bridges low-level image quality assessment with high-level reasoning. Its core function is to extract a robust and holistic

[1] 4KAgent can actually achieve large-scale super-resolution (*e.g.*, 32×, 64×, ...) if applied recursively [168].

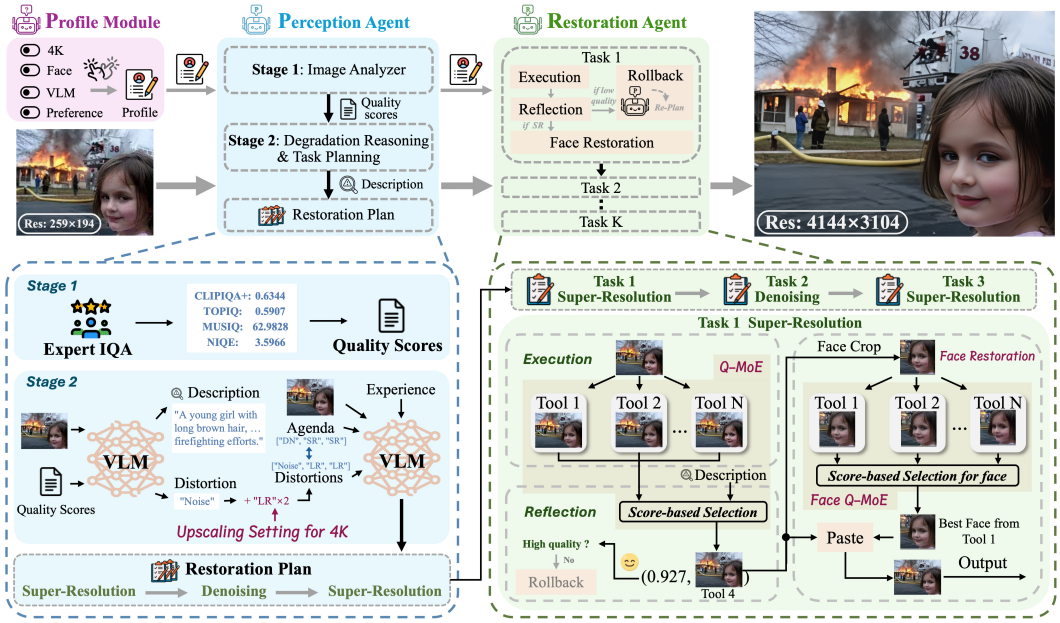


Figure 2: **4KAgent system overview.**

understanding of the input image in terms of both semantic content and low-level degradations, and to create a restoration plan that guides the subsequent restoration process.

Image Analyzer. Perception agent invokes a suite of expert Image Quality Assessment (IQA) tools that evaluate the input image I across multiple quality dimensions $Q_I = (Q_1, Q_2, \dots)$. Specifically, we adopt CLIPIQA [194], TOPIQ [20], MUSIQ [83], and NIQE [260] as the IQA metrics. These metrics represent perceptual quality from diverse aspects (due to their different model designs and training data), which are employed as *context* for the next step of degradation reasoning.

Degradation Reasoning. Perception agent leverages a VLM M_R to reason over the obtained IQA metrics. Specifically, by incorporating the input image I, IQA metrics Q_I, the VLM M_R predicts the degradation list D_I from the input image, which corresponds to an initial restoration agenda A'_I. Meanwhile, M_D also analyzes the content in the image and outputs the corresponding image descriptions C_I (*i.e.*, captioning). The whole process can be expressed as $C_I, D_I, A'_I = M_R(I, Q_I)$.

Upscaling Factor Configuration. 4KAgent is able to automatically determine and apply an appropriate super-resolution scale to reach 4K. Given an input image I with height H_I and width W_I, the scale factor s is calculated by $s = \min\Big(\{\,\hat{s} \in \{2, 4, 8, 16\} \mid \max(H_I, W_I) \cdot \hat{s} \geq 4000\} \cup \{16\}\Big)$. After obtaining the initialized agenda A'_I from M_D, 4KAgent calculates the scale factor s and appends super-resolution task(s) into A'_I to obtain the final agenda A_I. Under this setting, 4KAgent is able to upscale any image (resolution larger than 250×250) to 4K resolution in a single process.

Task Planning. After obtaining the degradation list D_I present in the input image and the restoration agenda A_I, the perception agent employs an LLM / VLM M_P to provide the restoration plan. Specifically, by incorporating image descriptions C_I, degradation list D_I, restoration experience E, and input image itself I (available when using VLM as M_P), M_P outputs an initial restoration plan $P_I = M_P(C_I, D_I, A_I, E, I)$, which contains a sequence of restoration tasks.

2.3 Restoration Agent

Building upon the task plan P_I provided by the Perception Agent, the Restoration Agent executes an iterative process, each stage of which tightly couples restoration and evaluation using an execution–reflection–rollback triplet. Within this agent, we propose a Quality-Driven Mixture-of-Experts (**Q-MoE**) policy, both in execution and reflection, to select the optimal image for each restoration step. We also employ a rollback mechanism to adjust the restoration plan if necessary.

Execution. Guided by the task plan P_I, this agent executes the restoration step by step. In each restoration step, the input image goes through all tools in the toolbox, which contains a number of

advanced restoration models (detailed in the Appendix) corresponding to each individual restoration task. In 4KAgent, we curate **9** different restoration tasks that are useful to enhance image quality: **Brightening**, **Defocus Deblurring**, **Motion Deblurring**, **Dehazing**, **Denoising**, **Deraining**, **JPEG Compression Artifact Removal**, **Super Resolution**, and **Face Restoration**. Specifically, for step k in the restoration plan, it produces multiple restoration results $\{T_i(I_{k-1}), i = 1 \sim N\}$ (T_i is i-th tool in the toolbox, N is the number of tools in the toolbox) based on the input image I_{k-1}.

Reflection. After obtaining restoration results $\{T_i(I_{k-1}), i = 1 \sim N\}$, restoration agent selects the best image based on their quality. To evaluate the quality of image $T_i(I_{k-1})$, we compute the image quality score by combining the preference model HPSv2 [219] and no-reference IQA metrics. Specifically, we use HPSv2 to assess the human preference of the resulting image $T_i(I_{k-1})$ based on the image content description C_I. For no-reference IQA metrics, we employ NIQE [148], MANIQA [233], MUSIQ [83], and CLIPIQA [194] to calculate a weighted sum as its no-reference quality score: $Q_s(T_i(I_{k-1})) = \mathrm{H}(T_i(I_{k-1}), C_I) + Q_{nr}(T_i(I_{k-1}))/4$, where $Q_{nr}(T_i(I_{k-1})) = w_{\mathrm{NIQE}} \cdot (1 - Q_{\mathrm{NIQE}}/10) + \sum_{j\in\Omega} w_j \cdot Q_j$, $\Omega = \{\mathrm{MUSIQ}, \mathrm{MANIQA}, \mathrm{CLIPIQA}\}$, H indicates the HPSv2 evaluation. After obtaining the quality score of each result image, the final result of this restoration step is obtained by the highest quality score: $I_k = T_{i^*}(I_{k-1}), i^* = \arg\max_{i\in\{1,\dots,N\}} Q_s(T_i(I_{k-1}))$. The combination of execution and reflection can be viewed as a Mixture-of-Experts (MoE) system, which we refer to as a Quality-Driven Mixture-of-Experts (**Q-MoE**): the input image is processed through each expert (execution), and the Reflection function selects the optimal image among all.

Rollback. Following previous AI Agent systems [160, 274, 268, 113, 69], we also design a rollback mechanism in the 4KAgent system. Specifically, if the quality score of I_k is lower than a threshold η, *i.e.*, $Q_s(I_k) \leq \eta$, the restoration step is seen as a failure step, and 4KAgent generates a failure message S_I. Then the system rolls back to the input image I_{k-1} and assigns a different restoration task in this step. (Details are provided in the Appendix)

2.4 Face Restoration Pipeline

Human face regions are often the most visually sensitive and semantically important components in an image. However, conventional super-resolution methods struggle to maintain identity consistency, natural skin textures, and perceptual quality when restoring faces, especially in heavily degraded portraits. To address this, 4KAgent incorporates a dedicated **Face Restoration Pipeline**, which is selectively triggered within the restoration workflow. The Face Restoration Pipeline is embedded as a submodule in 4KAgent and is only invoked *after the super-resolution restoration step*, ensuring that face quality refinement is seamlessly integrated into the iterative restoration loop.

The overall framework of the Face Restoration Pipeline in 4KAgent is shown in Fig. 3. First, 4KAgent detects and crops faces in the input image $\{F_I^l, l = 1 \sim L\}$ (L is the number of faces in the image I). Then, if **super-resolution** is in the restoration plan and the resulting image I_k of super-resolution step does not trigger the rollback mechanism, 4KAgent will detect and crop faces in the resulting image $\{F_{I_k}^l, l = 1 \sim L'\}$ (L' is the number of faces in the image I_k). If $L = L'$, then for each face in I_k, different advanced face restoration methods are applied, yielding restored faces $\{T_i^f(F_{I_k}^l), i = 0 \sim N^f\}$. Here, T_i^f is a face restoration tool in the toolbox, T_0^f is an identical function, and N^f is the number of face restoration tools in the toolbox. Likewise, we also apply the Q-MoE policy here: 4KAgent selects the best face based on the quality score Q_s^f. The quality score Q_s^f not only considers face quality, but also identity preservation:

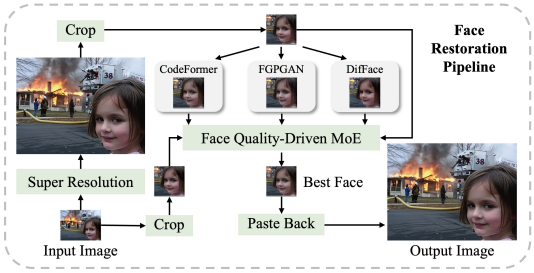


Figure 3: Face Restoration Pipeline overview.

$$Q_s^f(T_i^f(F_{I_k}^l)) = w_{\mathrm{IP}} \cdot \mathrm{IP}(T_i^f(F_{I_k}^l), F_I^l) + w_{\mathrm{IQA}} \cdot (Q_{nr}(T_i^f(F_{I_k}^l))/4 + Q_{\mathrm{CF}}(T_i^f(F_{I_k}^l))), \quad (1)$$

where $l = 1 \sim L$. IP calculates the cosine similarity of face features, extracted using ArcFace [42]. CF indicates CLIB-FIQA [156], which is an advanced face IQA metric. 4KAgent combines the no-reference quality score used in the reflection stage and the CLIB-FIQA score to assess the face quality. After obtaining quality score Q_s^f, 4KAgent selects the best face F_{out}^l: $F_{out}^l = T_{i^*}^f(F_{I_k}^l), i^* =$

$\arg\max_{i \in \{0,\ldots,N^f\}} Q_s^f(T_i^f(F_{I_k}^l))$. 4KAgent will paste F_{out}^l back to the original image I_k, then proceeding to the next step.

2.5 Profile Module

To enhance the flexibility and applicability of our 4KAgent system, we develop the **Profile Module**, enabling dynamic customization for diverse image restoration scenarios, according to user requirements. Specifically, the Profile Module serves a role analogous to customization module in LLM applications, allowing fine-grained control through the following seven configuration parameters:

1. **Perception Agent**: Specifies the choice of LLM / VLM employed by the Perception Agent. [`Default: Llama-vision`]
2. **Upscale to 4K**: Determines whether to upscale to 4K resolution. [`Default: True`]
3. **Scale Factor**: Explicitly defines the upscale factor for the entire pipeline. (`Default: 4`, `Options: [2,4,8,16]`). This parameter overrides "**Upscale to 4K**" when specified.
4. **Restore Option**: Explicitly sets the restoration task(s) to be applied. If set to `None`, restoration task(s) are determined automatically by the Perception Agent. (`Default: None`)
5. **Face Restore**: Toggles activation of the dedicated face restore pipeline. (`Default: True`)
6. **Brightening**: Controls the activation of image brightening, which may cause color shifts in restored images. Provided as [Optional] to maintain image color fidelity. (`Default: False`)
7. **Restore Preference**: Defines whether to prioritize higher perceptual quality or higher fidelity in image restoration. (`Options: [Perception, Fidelity]`, `Default: Perception`). Here we respect the perception-distortion tradeoff [14, 270], deeming models that optimize for distortion metrics (*e.g.*, PSNR, SSIM [208]) as `Fidelity` models while methods trained for perceptual quality (*e.g.*, NIQE [148], MUSIQ [83]) as `Perception` models.

The Profile Module offers exceptional configurability, enabling seamless adaptation to a wide range of restoration tasks without requiring model retraining or domain-specific fine-tuning. To the best of our knowledge, 4KAgent is a first-of-its-kind framework that enjoys unprecedented robustness and generalizability: each distinct restoration scenario can be addressed by simply selecting an appropriate configuration profile, thanks to which 4KAgent consistently achieves excellent performance across a variety of challenging restoration domains. Comprehensive details on predefined profiles in 4KAgent and their naming conventions are further elaborated in the Appendix.

3 Experiments

We evaluate 4KAgent on a wide range of **11** image super-resolution tasks, including classical image SR (4×) (§3.1), real-world image SR (4×) (§3.2), multiple-degradation image restoration (§3.3), face restoration (4×) (Appendix), large scale factor SR (16×) (§3.4), joint restoration with 4K upscaling (§3.5), and AI-generated content 4K SR (§3.6), remote sensing image SR (Appendix), fluorescence microscopy image SR (Appendix), and medical image SR (Appendix). In total, we evaluate 4KAgent on **26** benchmarks. The summary of datasets used in experiments is shown in the Appendix. We also present the parameter setting, the details of the toolbox, and prompts used in the LLM / VLM component of 4KAgent in the Appendix.

3.1 Classical Image Super-Resolution

In this section, we follow the classical SR experiment setting [266, 119] and evaluate 4KAgent on classical SR benchmarks, including Set5 [13], Set14 [249], B100 [140], Urban100 [70], and Manga109 [142]. In addition to PSNR and SSIM [208], we evaluate images with LPIPS [262], FID [66], NIQE [148], and MUSIQ [83] for more comprehensive evaluation.

Due to the high flexibility of 4KAgent, which is governed by configurable pro-

Table 1: Quantitative comparison on classical image SR benchmark (B100). The best and second-best results are marked in **bold** and underline.

Method	PSNR↑	SSIM↑	LPIPS↓	FID↓	NIQE↓	MUSIQ↑
SwinIR [119]	27.92	0.7489	0.3548	94.57	6.27	57.71
X-Restormer [28]	27.99	0.7508	0.3521	90.52	6.21	57.91
HAT-L [29]	<u>28.08</u>	**0.7547**	0.3440	89.52	6.20	58.71
DiffBIR [123]	24.99	0.6156	0.2719	84.99	<u>3.92</u>	68.23
OSEDiff [216]	24.35	0.6495	<u>0.2408</u>	**73.23**	4.08	68.54
AgenticIR [274]	22.51	0.5853	0.3078	102.92	4.08	68.36
4KAgent (ExpSR-s4-F)	**28.09**	<u>0.7540</u>	0.3453	88.89	6.02	59.12
4KAgent (ExpSR-s4-P)	24.64	0.6294	**0.2387**	<u>73.64</u>	**3.86**	<u>69.42</u>
4KAgent (GenSR-s4-P)	23.64	0.6246	0.2572	78.80	3.93	**69.44**

files, we use three different profiles to customize the 4KAgent in this experiment: **ExpSR-s4-F**, **ExpSR-s4-P**, and **GenSR-s4-P**. For comparison, we use state-of-the-art fidelity-based methods (e.g., SwinIR [119], X-Restormer [28], HAT [29]) and perception-based methods (e.g., DiffBIR [123], OSEDiff [216]). We also include AgenticIR [274] for agentic system-level comparison. We present experimental results on the B100 dataset in Tab. 1. Detailed experimental results and visual comparisons are shown in the Appendix. Based on quantitative comparisons, 4KAgent outperforms AgenticIR across nearly all evaluation metrics on classical SR benchmarks. Moreover, 4KAgent can easily output images with high perceptual quality or high fidelity by setting different profiles.

3.2 Real-World Image Super-Resolution

In this section, we follow previous Real-ISR methods and evaluate 4KAgent on real-world image super-resolution datasets (RealSR [17], DrealSR [214]). We use two different profiles to customize 4KAgent in this experiment: **ExpSR-s4-P** and **GenSR-s4-P**. We compare 4KAgent with state-of-the-art methods, including ResShift [245], StableSR [195], DiffBIR [123], PASD [234], SeeSR [217], SinSR [205], and OSEDiff [216]. We also employ AgenticIR in this experiment for agentic system comparison. We adopt the same evaluation metrics as in the classical SR experiments. Experimental results on the RealSR dataset are shown in Tab. 2. 4KAgent outperforms AgenticIR in every metric regardless of profile setting. Moreover, 4KAgent sets a new state-of-the-art performance on no-reference perceptual metrics (*e.g.*, NIQE, MUSIQ). Detailed experimental results on both datasets and visual comparisons are shown in the Appendix.

Table 2: Quantitative comparison on Real-World image SR benchmark (RealSR). The best and second-best results are marked in **bold** and underline.

Method	PSNR↑	SSIM↑	LPIPS↓	FID↓	NIQE↓	MUSIQ↑
ResShift [245]	**26.31**	**0.7411**	0.3489	142.81	7.27	58.10
StableSR [195]	24.69	0.7052	0.3091	127.20	5.76	65.42
DiffBIR [123]	24.88	0.6673	0.3567	124.56	5.63	64.66
PASD [234]	25.22	0.6809	0.3392	**123.08**	5.18	68.74
SeeSR [217]	25.33	0.7273	0.2985	125.66	5.38	69.37
SinSR [205]	26.30	0.7354	0.3212	137.05	6.31	60.41
OSEDiff [216]	25.15	0.7341	**0.2921**	123.50	5.65	69.09
AgenticIR [274]	22.45	0.6447	0.3745	140.38	5.81	65.87
4KAgent (ExpSR-s4-P)	24.60	0.6839	0.3253	127.64	5.09	70.97
4KAgent (GenSR-s4-P)	22.55	0.6557	0.3509	134.63	**4.78**	**71.77**

3.3 Multiple-Degradataion Image Restoration

In this section, we follow the setting of AgenticIR, using the Group A, B, and C test sets, which contains 1,440 LQ images processed with 16 combinations of degradations applied to images from the MiO100 dataset [91]. In this experiment, we configure 4KAgent with **GenMIR-P** profile. We compare 4KAgent with several all-in-one models: AirNet [107], PromptIR [162], MiOIR [90], DA-CLIP [136], InstructIR [36], AutoDIR [81], and agentic systems: AgenticIR [274] and MAIR [80]. Experimental results are shown in Tab. 3 and Fig. 4.

Table 3: Quantitative comparison of multiple-degradataion image restoration tasks on the **Group C** subset. The best and second-best results are marked in **bold** and underline.

Method	PSNR↑	SSIM↑	LPIPS↓	MANIQA↑	CLIPIQA↑	MUSIQ↑
AirNet [107]	17.95	0.5145	0.5782	0.1854	0.3113	30.12
PromptIR [162]	18.51	0.5166	0.5756	0.1906	0.3104	29.71
MiOIR [90]	15.63	0.4896	0.5376	0.1717	0.2891	37.95
DA-CLIP [136]	18.53	0.5320	0.5335	0.1916	0.3476	33.87
InstructIR [36]	17.09	0.5135	0.5582	0.1732	0.2537	33.69
AutoDIR [81]	18.61	0.5443	0.5019	0.2045	0.2939	37.86
AgenticIR [274]	18.82	0.5474	0.4493	0.2698	0.3948	48.68
MAIR [80]	19.42	0.5544	**0.4142**	0.2798	0.4239	51.36
4KAgent (GenMIR-P)	**19.77**	**0.5629**	0.4271	**0.3545**	**0.5233**	**55.56**

Figure 4: Visual comparisons on the MiO100 dataset. (Please zoom in to see details.)

We present experimental results on Group C here. Experimental results for all groups are shown in the Appendix. As summarized in Tab. 3, 4KAgent achieves state-of-the-art performance on all metrics excluding LPIPS. Fig. 4 shows that 4KAgent generates images with finer-grained details, which are

more consistent with the HQ images (*e.g.*, the hand and bottle on the left and the skin of the lizard on the right). These results demonstrate the superiority of 4KAgent under complex degradations.

3.4 Large Scale Factor (16×) Image Super-Resolution

In this section, we target a challenging setting, 16× real-world image upscaling. For the dataset used in this experiment, we select RealSRSet [256], a real-world dataset consisting of 20 low-quality images for large-scale super-resolution experiment. Specifically, we configure 4KAgent with the **Gen4K-P** profile for this experiment. Based on the resolution of images in the dataset, 4KAgent will upscale each image with a scale factor of 16.

We compare 4KAgent against HAT-L [29], DiffBIR [123] under different settings: (1) 4×→ 4×: first upscale the low-quality image by a scale factor of 4, then upscale the upscaled images to obtain the 16× upscaled image. (2) 16×: directly upscale the low-quality image by a scale factor of 16. We also extend AgenticIR for large scale factor (16×) image super-resolution for agentic system comparison. Visual comparisons of different methods are shown in Fig. 5. 4KAgent generates fine-grained and realistic details that are more visually pleasant than the comparing methods (*e.g.*, the rock and grass textures). Quantitative results and more visual comparisons are shown in the Appendix.

Figure 5: Visual comparisons on the RealSRSet dataset. (Please zoom in to see details)

3.5 Joint Restoration & 4K Upscaling

In this section, we bring 4KAgent to the most challenging setting: Joint multiple-degradation image restoration and 4K upscaling. As there are no previous methods and datasets targeted at this setting, we construct a new evaluation dataset, **DIV4K-50**, based on the Aesthetic-4K dataset [254] to rigorously test end-to-end restoration and ultra-high-scale SR. Specifically, we select 50 images from the Aesthetic-4K dataset based on its content, then center-crop each image to 4096×4096 as the high-quality (HQ) ground truth image. We downsample HQ images to 256×256 and randomly apply combinations of defocus blur, motion blur, additive Gaussian noise, and JPEG compression to generate the corresponding low-quality (LQ) images. In this experiment, we also configure 4KAgent with the **Gen4K-P** profile. We compare 4KAgent with more methods in this experiment (OSEDiff [216], PiSA-SR [179]) with more upscaling settings.

As shown in Fig. 6, 4KAgent reconstructs finer and more natural details than the comparing methods (*e.g.*, the facial features). Quantitative results and more visual comparisons are shown in the Appendix.

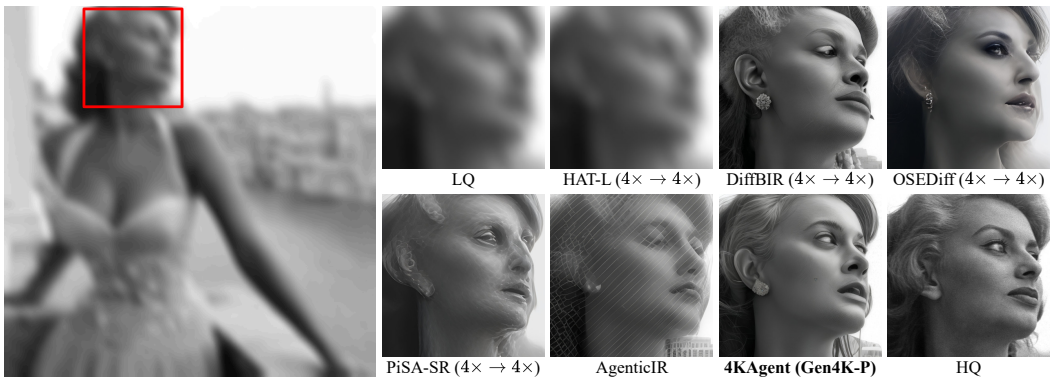


Figure 6: Visual comparisons on the DIV4K-50 dataset (Please zoom in to see details).

3.6 AI-Generated Content (AIGC) 4K Super-Resolution

In this section, we investigate super-resolution in AIGC scenarios by comparing direct 4K generation with 1K followed by upscaling using 4KAgent. As no prior method and dataset targets this setting, we sample 200 prompts from two standard AIGC benchmarks [210, 103] and generate both 1K and native 4K images using several generative models [95, 49, 24, 229, 72, 254], which we denote as GenAIBench-4K and DiffusionDB-4K. We use the **ExpSR-s4-P** profile in 4KAgent in this experiment. To better capture local degradations in 4K images, we introduce MUSIQ-P, a patch-based metric that averages MUSIQ scores over non-overlapping 512×512 regions.

Table 4: Quantitative comparison of AIGC $4\times$ SR in GenAIBench-4K [103]. Bold denotes top performers; underlines indicate second and third. MUSIQ-P* is a patch-applied MUSIQ metric for evaluating 4K images.

Model	NIQE↓	MANIQA↑	MUSIQ-P*↑	CLIPIQA↑
SANA-1K [229]	4.18	0.4814	**66.30**	<u>0.7147</u>
+ 4KAgent	3.03	0.4735	57.97	0.7050
GPT4o [72]	5.69	0.4997	<u>64.43</u>	0.6607
+ 4KAgent	3.56	0.4976	58.28	0.7016
FLUX.1-dev [95]	6.18	<u>0.5018</u>	61.02	0.6768
+ 4KAgent	<u>2.98</u>	<u>0.5034</u>	58.19	<u>0.7078</u>
PixArt-Σ [24]	4.12	0.4415	63.74	0.6960
+ 4KAgent	**2.76**	0.4699	56.71	0.7077
SD3-Medium [49]	5.03	0.4767	<u>64.68</u>	0.6922
+ 4KAgent	<u>2.99</u>	**0.5155**	60.22	**0.7169**

Fig. 7 presents models enhanced by 4KAgent produce finer-grained details compared to native 4K generation methods [254, 229]. As shown in Tab. 4, 4KAgent consistently boosts both perceptual quality and semantic alignment on GenAIBench-4K, surpassing their original 1K generation baselines. In addition, we observe that 4KAgent demonstrates stronger alignment with human preferences compared to native 4K generation methods, as evidenced by higher PickScore [88]. Additional comparisons with native 4K methods and results on DiffusionDB-4K are presented in the Appendix.

Figure 7: Qualitative comparison between native 4K image generation and 1K image generation methods with 4KAgent, using identical prompts.

3.7 Scientific Images

In this section, we extend the evaluation of 4KAgent across a broad spectrum of cross-domain scientific image super-resolution applications, where high spatial fidelity is crucial but often limited by sensor cost and physical constraints [184, 185, 188]. The explored imaging domains and corresponding benchmark datasets are as follows: For remote sensing image super-resolution, we evaluate on four benchmark datasets covering varied land-use patterns and sensing characteristics, including AID [223], DIOR [111], DOTA [222], and WorldStrat [37]. For fluorescence microscopy super-resolution, we compare the performance of 4KAgent with a set of representative deep learning-based single-image SR (SISR) methods on SR-CACO-2 dataset [12]. For biomedical super-resolution, datasets from 4 distinct modalities are considered: bcSR [77] in pathology microscopy, Chest X-ray 2017 [85] and Chest X-ray 14 [199] in X-ray, US-Case [175] and MMUS1K [154] in ultrasound, and DRIVE retinal vessel dataset [176] in fundoscopy.

Visual comparisons are shown in Fig. 8, showcasing six imaging modalities side by side: remote sensing, fundoscopy, confocal fluorescence microscopy, pathology, X-ray, and ultrasound from left to

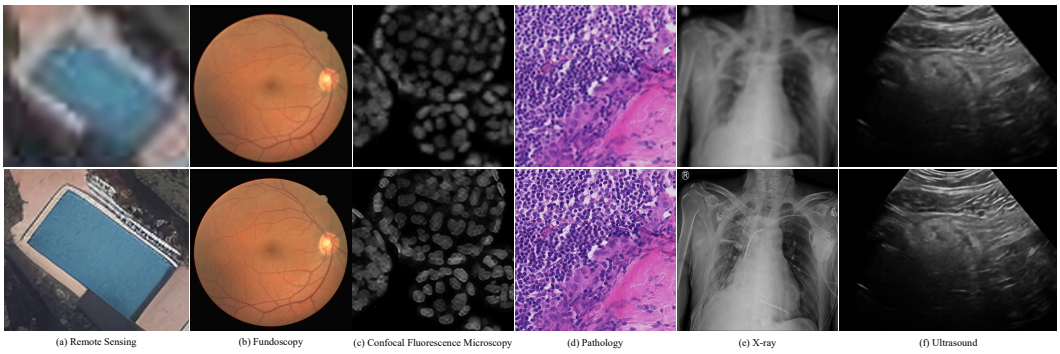


Figure 8: 4KAgent in processing scientific images. (Low-quality input vs. 4KAgent result)

right. The top and bottom rows display the LQ inputs and 4KAgent outputs, respectively. Detailed quantitative and qualitative results are presented in the Appendix.

4 Related Work

We briefly review related works in three key areas. In **image super-resolution**, early CNN-based methods such as SRCNN [45] laid the foundation, followed by architectural innovations such as residual and dense connections [87, 121, 267], attention mechanisms [19, 41, 266], and generative models including GAN-based [98, 201, 256, 21, 118, 119, 105] and diffusion-based methods [195, 234, 123, 217, 168, 205, 216]. In **image restoration**, progress has been driven by residual learning [255, 131, 257], attention modules [242, 25, 186, 57, 247], transformers [189, 275, 246, 206], as well as GAN-based [58, 11, 56, 73, 147, 158] and diffusion-based models [123, 221, 81, 204, 82, 50], with recent efforts addressing multi-degradation scenarios via unified models [107, 138, 159, 236, 252]. In **LLM agents**, foundational reasoning frameworks like CoT [212], ReAct [237], and CoALA [178] have inspired domain-specific systems such as MMC-TAgent [93], VideoAgent [198], and MedCoT [129]. Notably, RestoreAgent [22], AgenticIR [274], MAIR [80], HybridAgent [104], and Q-Agent [272] demonstrate how LLMs can orchestrate multi-step visual restoration workflows via perception, planning, and quality-aware decision-making. For a full discussion, please refer to the Appendix.

5 Concluding Remarks

In this paper, we introduce 4KAgent, a versatile agentic image super-resolution generalist model designed to universally upscale images of diverse types and degradation levels to 4K resolution. By leveraging advanced multi-expert integration, adaptive decision-making, and specialized tools for perception and fidelity optimization, 4KAgent significantly enhances restoration quality across various challenging domains, including severely degraded images, natural scenes, portraits, AI-generated content, and specialized scientific modalities such as remote sensing, microscopy, and medical imaging. Extensive evaluations on both standard benchmarks and specialized datasets demonstrate that 4KAgent consistently outperforms existing state-of-the-art approaches, especially in complex scenarios where conventional super-resolution methods fall short. This robust performance, achieved without domain-specific retraining, highlights the generalizability and practical utility of 4KAgent for generic deployment in both consumer, commercial, and scientific applications.

Future Work Looking ahead, we identify several promising directions to further enhance the capabilities and applicability of 4KAgent. First, we plan to improve system efficiency by developing more accurate distortion–perception modeling and refining the execution–reflection–rollback mechanism for faster and higher-quality image restoration. Second, we will strengthen the safety and robustness of 4KAgent to mitigate risks such as privacy leakage and harmful content generation. Finally, we aim to continuously expand the 4KAgent toolbox by incorporating additional domain-specific restoration methods and designing targeted profiles to better support diverse user needs and application scenarios.

Acknowledgements Gengchen Mai is supported by the NSF under Grant No. 2521631.

References

[1] NVIDIA DLSS 4 Supreme Speed. Superior Visuals. Powered by AI. https://www.nvidia.com/en-us/geforce/technologies/dlss/. 58

[2] Video quality in public safety (VQiPS) workshop report. https://www.nist.gov/system/files/documents/2016/10/06/final_vqipsworkshopreport_092912.pdf. 59

[3] Ramzi Abiantun, Felix Juefei-Xu, Utsav Prabhu, and Marios Savvides. Ssr2: Sparse signal recovery for single-image super-resolution on faces with extreme low resolutions. *Pattern Recognition*, 90:308–324, 2019. 59

[4] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023. 27

[5] Waqar Ahmad, Hazrat Ali, Zubair Shah, and Shoaib Azmat. A new generative adversarial network for medical images super resolution. *Scientific Reports*, 12(1):9533, 2022. 54, 55

[6] Meta AI. Llama 3.2-vision. https://huggingface.co/meta-llama/Llama-3.2-11B-Vision-Instruct, 2024. 27

[7] Simone Angarano, Francesco Salvetti, Mauro Martini, and Marcello Chiaberge. Generative adversarial super-resolution at the edge with knowledge distillation. *Engineering Applications of Artificial Intelligence*, 123:106407, 2023. 60

[8] Vegard Antun, Francesco Renna, Clarice Poon, Ben Adcock, and Anders C Hansen. On instabilities of deep learning in image reconstruction and the potential costs of ai. *Proceedings of the National Academy of Sciences*, 117(48):30088–30095, 2020. 62

[9] Aditya Arora, Zhengzhong Tu, Yufei Wang, Ruizheng Bai, Jian Wang, and Sizhuo Ma. Guidesr: Rethinking guidance for one-step high-fidelity diffusion-based super-resolution. *arXiv preprint arXiv:2505.00687*, 2025. 62

[10] Shuai Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, Sibo Song, Kai Dang, Peng Wang, Shijie Wang, Jun Tang, et al. Qwen2. 5-vl technical report. *arXiv preprint arXiv:2502.13923*, 2025. 27

[11] David Bau, Hendrik Strobelt, William Peebles, Jonas Wulff, Bolei Zhou, Jun-Yan Zhu, and Antonio Torralba. Semantic photo manipulation with a generative image prior. *arXiv preprint arXiv:2005.07727*, 2020. 10, 62

[12] Soufiane Belharbi, Mara Whitford, Phuong Hoang, Shakeeb Murtaza, Luke McCaffrey, and Eric Granger. Sr-caco-2: A dataset for confocal fluorescence microscopy image super-resolution. *Advances in Neural Information Processing Systems*, 37:59948–59983, 2024. 9, 30, 49

[13] Marco Bevilacqua, Aline Roumy, Christine Guillemot, and Marie Line Alberi-Morel. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. 2012. 6, 30

[14] Yochai Blau and Tomer Michaeli. The perception-distortion tradeoff. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6228–6237, 2018. 6

[15] BytePlus. Business growth through superior technology. https://www.byteplus.com, 2025. 61

[16] BytePlus. Unleashing the power of super resolution: A game-changer for visual content. https://www.byteplus.com/en/topic/96403, 2025. 61

[17] Jianrui Cai, Hui Zeng, Hongwei Yong, Zisheng Cao, and Lei Zhang. Toward real-world single image super-resolution: A new benchmark and a new model. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3086–3095, 2019. 7, 30

[18] Jung-Woo Chang, Keon-Woo Kang, and Suk-Ju Kang. An energy-efficient fpga-based deconvolutional neural networks accelerator for single image super-resolution. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(1):281–295, 2018. 61

[19] Chaofeng Chen, Dihong Gong, Hao Wang, Zhifeng Li, and Kwan-Yee K Wong. Learning spatial attention for face super-resolution. *IEEE Transactions on Image Processing*, 30:1219–1231, 2020. 10, 62

[20] Chaofeng Chen, Jiadi Mo, Jingwen Hou, Haoning Wu, Liang Liao, Wenxiu Sun, Qiong Yan, and Weisi Lin. Topiq: A top-down approach from semantics to distortions for image quality assessment. *IEEE Transactions on Image Processing*, 2024. 4

[21] Chaofeng Chen, Xinyu Shi, Yipeng Qin, Xiaoming Li, Xiaoguang Han, Tao Yang, and Shihui Guo. Real-world blind super-resolution via feature matching with implicit high-resolution priors. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 1329–1338, 2022. 2, 10, 62

[22] Haoyu Chen, Wenbo Li, Jinjin Gu, Jingjing Ren, Sixiang Chen, Tian Ye, Renjing Pei, Kaiwen Zhou, Fenglong Song, and Lei Zhu. Restoreagent: Autonomous image restoration agent via multimodal large language models. *arXiv preprint arXiv:2407.18035*, 2024. 2, 10, 63

[23] Haoyu Chen, Wenbo Li, Jinjin Gu, Jingjing Ren, Haoze Sun, Xueyi Zou, Zhensong Zhang, Youliang Yan, and Lei Zhu. Low-res leads the way: Improving generalization for super-resolution by self-supervised learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25857–25867, 2024. 2

[24] Junsong Chen, Chongjian Ge, Enze Xie, Yue Wu, Lewei Yao, Xiaozhe Ren, Zhongdao Wang, Ping Luo, Huchuan Lu, and Zhenguo Li. Pixart-σ: Weak-to-strong training of diffusion transformer for 4k text-to-image generation. In *European Conference on Computer Vision*, pages 74–91. Springer, 2024. 9, 40

[25] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *European conference on computer vision*, pages 17–33. Springer, 2022. 10, 28, 62

[26] Shengjie Chen, Shuo Chen, Zhenhua Guo, and Yushen Zuo. Low-resolution palmprint image denoising by generative adversarial networks. *Neurocomputing*, 358:275–284, 2019. 62

[27] Wei-Ting Chen, Gurunandan Krishnan, Qiang Gao, Sy-Yen Kuo, Sizhou Ma, and Jian Wang. Dsl-fiqa: Assessing facial image quality via dual-set degradation learning and landmark-guided transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2931–2941, 2024. 36

[28] Xiangyu Chen, Zheyuan Li, Yuandong Pu, Yihao Liu, Jiantao Zhou, Yu Qiao, and Chao Dong. A comparative study of image restoration networks for general backbone network design. In *European Conference on Computer Vision*, pages 74–91. Springer, 2024. 6, 7, 28, 31, 32

[29] Xiangyu Chen, Xintao Wang, Wenlong Zhang, Xiangtao Kong, Yu Qiao, Jiantao Zhou, and Chao Dong. Hat: Hybrid attention transformer for image restoration. *arXiv preprint arXiv:2309.05239*, 2023. 6, 7, 8, 28, 31, 32, 37, 39, 43, 44, 45

[30] Xiangyu Chen, Xintao Wang, Jiantao Zhou, Yu Qiao, and Chao Dong. Activating more pixels in image super-resolution transformer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 22367–22377, 2023. 55

[31] Yuhua Chen, Yibin Xie, Zhengwei Zhou, Feng Shi, Anthony G Christodoulou, and Debiao Li. Brain mri super resolution using 3d deep densely connected neural networks. In *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*, pages 739–742. IEEE, 2018. 60

[32] Zhen Chen, Xiaoqing Guo, Chen Yang, Bulat Ibragimov, and Yixuan Yuan. Joint spatial-wavelet dual-stream network for super-resolution. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part V 23*, pages 184–193. Springer, 2020. 52

[33] Sung-Jin Cho, Seo-Won Ji, Jun-Pyo Hong, Seung-Won Jung, and Sung-Jea Ko. Rethinking coarse-to-fine approach in single image deblurring. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4641–4650, 2021. 2

[34] Shu-Chuan Chu, Zhi-Chao Dou, Jeng-Shyang Pan, Shaowei Weng, and Junbao Li. Hmanet: Hybrid multi-axis aggregation network for image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6257–6266, 2024. 28, 30, 31, 32

[35] Joseph Paul Cohen, Margaux Luck, and Sina Honari. Distribution matching losses can hallucinate features in medical image translation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part I*, pages 529–536. Springer, 2018. 62

[36] Marcos V Conde, Gregor Geigle, and Radu Timofte. Instructir: High-quality image restoration following human instructions. In *European Conference on Computer Vision*, pages 1–21. Springer, 2024. 7, 35

[37] Julien Cornebise, Ivan Oršolić, and Freddie Kalaitzis. Open high-resolution satellite imagery: The worldstrat dataset–with application to super-resolution. *Advances in Neural Information Processing Systems*, 35:25979–25991, 2022. 9, 30, 43

[38] Kate Crawford. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press, 2021. 62

[39] Yuning Cui, Wenqi Ren, Xiaochun Cao, and Alois Knoll. Revitalizing convolutional network for image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024. 28

[40] Ryan Dahl, Mohammad Norouzi, and Jonathon Shlens. Pixel recursive super resolution. In *Proceedings of the IEEE international conference on computer vision*, pages 5439–5448, 2017. 62

[41] Tao Dai, Jianrui Cai, Yongbing Zhang, Shu-Tao Xia, and Lei Zhang. Second-order attention network for single image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11065–11074, 2019. 10, 62

[42] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4690–4699, 2019. 5

[43] Keyan Ding, Kede Ma, Shiqi Wang, and Eero P Simoncelli. Image quality assessment: Unifying structure and texture similarity. *IEEE transactions on pattern analysis and machine intelligence*, 44(5):2567–2581, 2020. 30

[44] Chao Dong, Yubin Deng, Chen Change Loy, and Xiaoou Tang. Compression artifacts reduction by a deep convolutional network. In *Proceedings of the IEEE international conference on computer vision*, pages 576–584, 2015. 62

[45] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV 13*, pages 184–199. Springer, 2014. 2, 10, 50, 52, 62

[46] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2015. 2

[47] Chao Dong, Chen Change Loy, and Xiaoou Tang. Accelerating the super-resolution convolutional neural network. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*, pages 391–407. Springer, 2016. 54

[48] Zane Durante, Qiuyuan Huang, Naoki Wake, Ran Gong, Jae Sung Park, Bidipta Sarkar, Rohan Taori, Yusuke Noda, Demetri Terzopoulos, Yejin Choi, et al. Agent ai: Surveying the horizons of multimodal interaction. *arXiv preprint arXiv:2401.03568*, 2024. 2

[49] Patrick Esser, Sumith Kulal, Andreas Blattmann, Rahim Entezari, Jonas Müller, Harry Saini, Yam Levi, Dominik Lorenz, Axel Sauer, Frederic Boesel, et al. Scaling rectified flow transformers for high-resolution image synthesis. In *Forty-first international conference on machine learning*, 2024. 9, 40

[50] Ben Fei, Zhaoyang Lyu, Liang Pan, Junzhe Zhang, Weidong Yang, Tianyue Luo, Bo Zhang, and Bo Dai. Generative diffusion prior for unified image restoration and enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9935–9946, 2023. 10, 62

[51] Chun-Mei Feng, Yunlu Yan, Huazhu Fu, Li Chen, and Yong Xu. Task transformer network for joint mri reconstruction and super-resolution. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VI 24*, pages 307–317. Springer, 2021. 60

[52] João Gama, Indrė Žliobaitė, Albert Bifet, Mykola Pechenizkiy, and Abdelhamid Bouchachia. A survey on concept drift adaptation. *ACM computing surveys (CSUR)*, 46(4):1–37, 2014. 61

[53] Shangqi Gao and Xiahai Zhuang. Multi-scale deep neural networks for real image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2019. 62

[54] Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé Iii, and Kate Crawford. Datasheets for datasets. *Communications of the ACM*, 64(12):86–92, 2021. 61

[55] Hayit Greenspan. Super-resolution in medical imaging. *The computer journal*, 52(1):43–63, 2009. 2

[56] Jinjin Gu, Yujun Shen, and Bolei Zhou. Image processing using multi-code gan prior. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3012–3021, 2020. 10, 62

[57] Shuhang Gu, Yawei Li, Luc Van Gool, and Radu Timofte. Self-guided network for fast image denoising. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2511–2520, 2019. 10, 62

[58] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C Courville. Improved training of wasserstein gans. *Advances in neural information processing systems*, 30, 2017. 10, 62

[59] Xiaojie Guo, Yu Li, and Haibin Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on image processing*, 26(2):982–993, 2016. 2

[60] Muhammad Haris, Greg Shakhnarovich, and Norimichi Ukita. Task-driven super resolution: Object detection in low-resolution images. In *Neural Information Processing: 28th International Conference, ICONIP 2021, Sanur, Bali, Indonesia, December 8–12, 2021, Proceedings, Part V 28*, pages 387–395. Springer, 2021. 2

[61] Muhammad Haris, Gregory Shakhnarovich, and Norimichi Ukita. Deep back-projection networks for super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1664–1673, 2018. 50

[62] Woodrow Hartzog. *Privacy's blueprint: The battle to control the design of new technologies*. Harvard University Press, 2018. 62

[63] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, 33(12):2341–2353, 2010. 2

[64] Yutong He, Dingjie Wang, Nicholas Lai, William Zhang, Chenlin Meng, Marshall Burke, David Lobell, and Stefano Ermon. Spatial-temporal super-resolution of satellite imagery via conditional pixel synthesis. *Advances in Neural Information Processing Systems*, 34:27903–27915, 2021. 2, 42

[65] Zhe He, Yide Zhang, Xin Tong, Lei Li, and Lihong V Wang. Quantum microscopy of cells at the heisenberg limit. *Nature Communications*, 14(1):2441, 2023. 60

[66] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017. 6, 30

[67] Shane M Hickey, Ben Ung, Christie Bader, Robert Brooks, Joanna Lazniewska, Ian RD Johnson, Alexandra Sorvina, Jessica Logan, Carmela Martini, Courtney R Moore, et al. Fluorescence microscopyan outline of hardware, biological handling, and fluorophore considerations. *Cells*, 11(1):35, 2021. 49

[68] Chih-Chung Hsu, Chia-Ming Lee, and Yi-Shiuan Chou. Drct: Saving image super-resolution away from information bottleneck. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6133–6142, 2024. 28, 30, 31, 32

[69] Minda Hu, Tianqing Fang, Jianshu Zhang, Junyu Ma, Zhisong Zhang, Jingyan Zhou, Hongming Zhang, Haitao Mi, Dong Yu, and Irwin King. Webcot: Enhancing web agent reasoning by reconstructing chain-of-thought in reflection, branching, and rollback. *arXiv preprint arXiv:2505.20013*, 2025. 5

[70] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5197–5206, 2015. 6, 30

[71] Xu Huang, Weiwen Liu, Xiaolong Chen, Xingmei Wang, Hao Wang, Defu Lian, Yasheng Wang, Ruiming Tang, and Enhong Chen. Understanding the planning of llm agents: A survey. *arXiv preprint arXiv:2402.02716*, 2024. 2

[72] Aaron Hurst, Adam Lerer, Adam P Goucher, Adam Perelman, Aditya Ramesh, Aidan Clark, AJ Ostrow, Akila Welihinda, Alan Hayes, Alec Radford, et al. Gpt-4o system card. *arXiv preprint arXiv:2410.21276*, 2024. 9, 40

[73] Shady Abu Hussein, Tom Tirer, and Raja Giryes. Image-adaptive gan based reconstruction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 3121–3129, 2020. 10, 62

[74] Maxar Intelligence. Supercharging maxar intelligence's imagery basemaps with worldview legion satellite imagery. https://blog.maxar.com/earth-intelligence/2025/supercharging-maxar-intelligences-imagery-basemaps-with-worldview-legion-satellite-imagery, 2025. 61

[75] Md Jahidul Islam, Sadman Sakib Enan, Peigen Luo, and Junaed Sattar. Underwater image super-resolution using deep residual multipliers. In *2020 IEEE international conference on robotics and automation (ICRA)*, pages 900–906. IEEE, 2020. 2

[76] Md Jahidul Islam, Peigen Luo, and Junaed Sattar. Simultaneous enhancement and super-resolution of underwater imagery for improved visual perception. In *16th Robotics: Science and Systems, RSS 2020*. MIT Press Journals, 2020. 60

[77] Feng Jia, Lei Tan, Guang Wang, Cheng Jia, and Zhi Chen. A super-resolution network using channel attention retention for pathology images. *PeerJ Computer Science*, 9:e1196, 2023. Published 2023 Jan 17. 9, 30, 51, 52

[78] Jiaxi Jiang, Kai Zhang, and Radu Timofte. Towards flexible blind jpeg artifacts removal. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4997–5006, 2021. 28

[79] Kui Jiang, Zhongyuan Wang, Peng Yi, Chen Chen, Baojin Huang, Yimin Luo, Jiayi Ma, and Junjun Jiang. Multi-scale progressive fusion network for single image deraining. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8346–8355, 2020. 2

[80] Xu Jiang, Gehui Li, Bin Chen, and Jian Zhang. Multi-agent image restoration. *arXiv preprint arXiv:2503.09403*, 2025. 7, 10, 35, 63

[81] Yitong Jiang, Zhaoyang Zhang, Tianfan Xue, and Jinwei Gu. Autodir: Automatic all-in-one image restoration with latent diffusion. In *European Conference on Computer Vision*, pages 340–359. Springer, 2024. 2, 7, 10, 35, 62

[82] Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. *Advances in Neural Information Processing Systems*, 35:23593–23606, 2022. 10, 62

[83] Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. Musiq: Multi-scale image quality transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5148–5157, 2021. 4, 5, 6, 30

[84] Christopher J Kelly, Alan Karthikesalingam, Mustafa Suleyman, Greg Corrado, and Dominic King. Key challenges for delivering clinical impact with artificial intelligence. *BMC medicine*, 17:1–9, 2019. 62

[85] Daniel S Kermany, Michael Goldbaum, Wenjia Cai, Carolina CS Valentim, Huiying Liang, Sally L Baxter, Alex McKeown, Ge Yang, Xiaokang Wu, Fangbing Yan, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *cell*, 172(5):1122–1131, 2018. 9, 30, 54

[86] Chanwoo Kim, Soham U. Gadgil, Alex J. DeGrave, Jesutofunmi A. Omiye, Zhuo Ran Cai, Roxana Daneshjou, and Su-In Lee. Transparent medical image AI via an image–text foundation model grounded in medical literature. *Nature Medicine*, 30:1154–1165, 2024. 54

[87] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1646–1654, 2016. 10, 50, 62

[88] Yuval Kirstain, Adam Polyak, Uriel Singer, Shahbuland Matiana, Joe Penna, and Omer Levy. Pick-a-pic: An open dataset of user preferences for text-to-image generation. *Advances in Neural Information Processing Systems*, 36:36652–36663, 2023. 9, 41

[89] Lingshun Kong, Jiangxin Dong, Ming-Hsuan Yang, and Jinshan Pan. Efficient visual state space model for image deblurring. *arXiv preprint arXiv:2405.14343*, 2024. 28

[90] Xiangtao Kong, Chao Dong, and Lei Zhang. Towards effective multiple-in-one image restoration: A sequential and prompt learning strategy. *arXiv preprint arXiv:2401.03379*, 2024. 7, 35

[91] Xiangtao Kong, Jinjin Gu, Yihao Liu, Wenlong Zhang, Xiangyu Chen, Yu Qiao, and Chao Dong. A preliminary exploration towards general image restoration. *arXiv preprint arXiv:2408.15143*, 2024. 7, 34

[92] Pawel Kowaleczko, Tomasz Tarasiewicz, Maciej Ziaja, Daniel Kostrzewa, Jakub Nalepa, Przemyslaw Rokita, and Michal Kawulok. A real-world benchmark for sentinel-2 multi-image super-resolution. *Scientific Data*, 10(1):644, 2023. 2, 42

[93] Somnath Kumar, Yash Gadhia, Tanuja Ganu, and Akshay Nambi. Mmctagent: Multi-modal critical thinking agent framework for complex visual reasoning, 2024. 10

[94] Somnath Kumar, Yash Gadhia, Tanuja Ganu, and Akshay Nambi. Mmctagent: Multi-modal critical thinking agent framework for complex visual reasoning. *arXiv preprint arXiv:2405.18358*, 2024. 63

[95] Black Forest Labs. Flux. https://github.com/black-forest-labs/flux, 2024. 9, 40

[96] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. Deep laplacian pyramid networks for fast and accurate super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 624–632, 2017. 54

[97] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. Fast and accurate image super-resolution with deep laplacian pyramid networks. *IEEE transactions on pattern analysis and machine intelligence*, 41(11):2599–2613, 2018. 50

[98] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017. 2, 10, 52, 54, 55, 62

[99] Juhyoung Lee, Jinsu Lee, and Hoi-Jun Yoo. Srnpu: An energy-efficient cnn-based super-resolution processor with tile-based selective super-resolution in mobile devices. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 10(3):320–334, 2020. 61

[100] Junyong Lee, Hyeongseok Son, Jaesung Rim, Sunghyun Cho, and Seungyong Lee. Iterative filter adaptive network for single image defocus deblurring. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2034–2042, 2021. 28

[101] Royson Lee, Stylianos I Venieris, Lukasz Dudziak, Sourav Bhattacharya, and Nicholas D Lane. Mobisr: Efficient on-device super-resolution through heterogeneous mobile processors. In *The 25th annual international conference on mobile computing and networking*, pages 1–16, 2019. 61

[102] Sen Lei, Zhenwei Shi, and Wenjing Mo. Transformer-based multistage enhancement for remote sensing image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–11, 2022. 43, 44, 45

[103] Baiqi Li, Zhiqiu Lin, Deepak Pathak, Jiayao Li, Yixin Fei, Kewen Wu, Tiffany Ling, Xide Xia, Pengchuan Zhang, Graham Neubig, et al. Genai-bench: Evaluating and improving compositional text-to-visual generation. *arXiv preprint arXiv:2406.13743*, 2024. 9, 30, 40, 41

[104] Bingchen Li, Xin Li, Yiting Lu, and Zhibo Chen. Hybrid agents for image restoration. *arXiv preprint arXiv:2503.10120*, 2025. 10, 63

[105] Bingchen Li, Xin Li, Hanxin Zhu, Yeying Jin, Ruoyu Feng, Zhizheng Zhang, and Zhibo Chen. Sed: Semantic-aware discriminator for image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 25784–25795, 2024. 10, 62

[106] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1):492–505, 2018. 2

[107] Boyun Li, Xiao Liu, Peng Hu, Zhongqin Wu, Jiancheng Lv, and Xi Peng. All-in-one image restoration for unknown corruption. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 17452–17462, 2022. 7, 10, 35, 63

[108] Gen Li, Jie Ji, Minghai Qin, Wei Niu, Bin Ren, Fatemeh Afghah, Linke Guo, and Xiaolong Ma. Towards high-quality and efficient video super-resolution via spatial-temporal data overfitting. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10259–10269. IEEE, 2023. 58

[109] Jinlong Li, Baolu Li, Zhengzhong Tu, Xinyu Liu, Qing Guo, Felix Juefei-Xu, Runsheng Xu, and Hongkai Yu. Light the night: A multi-condition diffusion framework for unpaired low-light enhancement in autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15205–15215, 2024. 59

[110] Juncheng Li, Faming Fang, Kangfu Mei, and Guixu Zhang. Multi-scale residual network for image super-resolution. In *Proceedings of the European conference on computer vision (ECCV)*, pages 517–532, 2018. 62

[111] Ke Li, Gang Wan, Gong Cheng, Liqiu Meng, and Junwei Han. Object detection in optical remote sensing images: A survey and a new benchmark. *ISPRS journal of photogrammetry and remote sensing*, 159:296–307, 2020. 9, 30, 43

[112] Xin Li, Kun Yuan, Bingchen Li, Fengbin Guan, Yizhen Shao, Zihao Yu, Xijun Wang, Yiting Lu, Wei Luo, Suhang Yao, et al. Ntire 2025 challenge on short-form ugc video quality assessment and enhancement: Methods and results. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 1092–1103, 2025. 58

[113] Xingzuo Li, Kehai Chen, Yunfei Long, Xuefeng Bai, Yong Xu, and Min Zhang. Generator-assistant stepwise rollback framework for large language model agent. *arXiv preprint arXiv:2503.02519*, 2025. 5

[114] Yawei Li, Yuchen Fan, Xiaoyu Xiang, Denis Demandolx, Rakesh Ranjan, Radu Timofte, and Luc Van Gool. Efficient and explicit modelling of image hierarchies for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18278–18289, 2023. 50

[115] Yiming Li, Shuo Shao, Yu He, Junfeng Guo, Tianwei Zhang, Zhan Qin, Pin-Yu Chen, Michael Backes, Philip Torr, Dacheng Tao, and Kui Ren. Rethinking data protection in the (generative) artificial intelligence era. *arXiv preprint arXiv:2507.03034*, 2025. 62

[116] Yinxiao Li, Pengchong Jin, Feng Yang, Ce Liu, Ming-Hsuan Yang, and Peyman Milanfar. Comisr: Compression-informed video super-resolution. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 2543–2552, 2021. 61

[117] Zhen Li, Jinglei Yang, Zheng Liu, Xiaomin Yang, Gwanggil Jeon, and Wei Wu. Feedback network for image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3867–3876, 2019. 50

[118] Jie Liang, Hui Zeng, and Lei Zhang. Efficient and degradation-adaptive network for real-world image super-resolution. In *European Conference on Computer Vision*, pages 574–591. Springer, 2022. 10, 62

[119] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1833–1844, 2021. 2, 6, 7, 10, 28, 31, 32, 43, 44, 45, 50, 55, 62

[120] Lighthouse Guild. High tech for low vision: How technology is changing the world for people with vision loss. https://lighthouseguild.org/news/high-tech-for-low-vision-how-technology-is-changing-the-world-for-people-with-vision-loss/, 2025. 61

[121] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 136–144, 2017. 10, 52, 55, 62

[122] Tianwei Lin, Wenqiao Zhang, Sijing Li, Yuqian Yuan, Binhe Yu, Haoyuan Li, Wanggui He, Hao Jiang, Mengze Li, Xiaohui Song, et al. Healthgpt: A medical large vision-language model for unifying comprehension and generation via heterogeneous knowledge adaptation. *arXiv preprint arXiv:2502.09838*, 2025. 54

[123] Xinqi Lin, Jingwen He, Ziyan Chen, Zhaoyang Lyu, Bo Dai, Fanghua Yu, Yu Qiao, Wanli Ouyang, and Chao Dong. Diffbir: Toward blind image restoration with generative diffusion prior. In *European Conference on Computer Vision*, pages 430–448. Springer, 2024. 6, 7, 8, 10, 28, 31, 32, 33, 37, 39, 43, 44, 45, 62

[124] Yunlong Lin, Zixu Lin, Haoyu Chen, Panwang Pan, Chenxin Li, Sixiang Chen, Kairun Wen, Yeying Jin, Wenbo Li, and Xinghao Ding. Jarvisir: Elevating autonomous driving perception with intelligent image restoration. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 22369–22380, 2025. 63

[125] Yunlong Lin, Zixu Lin, Kunjie Lin, Jinbin Bai, Panwang Pan, Chenxin Li, Haoyu Chen, Zhongdao Wang, Xinghao Ding, Wenbo Li, et al. Jarvisart: Liberating human artistic creativity via an intelligent photo retouching agent. *arXiv preprint arXiv:2506.17612*, 2025. 63

[126] G. Litjens, P. Bandi, B. Ehteshami Bejnordi, O. Geessink, M. Balkenhol, P. Bult, A. Halilovic, M. Hermsen, R. van de Loo, R. Vogels, Q.F. Manson, N. Stathonikos, A. Baidoshvili, P. van Diest, C. Wauters, M. van Dijk, and J. van der Laak. 1399 H&E-stained sentinel lymph node sections of breast cancer patients: the CAMELYON dataset. *GigaScience*, 7(6):giy065, June 2018. 51

[127] Jiaming Liu, Zihao Liu, Xuan Huang, Ruoxi Zhu, Qi Zheng, Zhijian Hao, Tao Liu, Jun Tao, and Yibo Fan. Auto-isp: An efficient real-time automatic hyperparameter optimization framework for isp hardware system. In *Proceedings of the 61st ACM/IEEE Design Automation Conference*, pages 1–6, 2024. 60

[128] Jiaming Liu, Qi Zheng, Zihao Liu, Yilian Zhong, Peiye Liu, Tao Liu, Shusong Xu, Yanheng Lu, Sicheng Li, Dimin Niu, et al. Frequency-biased synergistic design for image compression and compensation. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 12820–12829, 2025. 62

[129] Jiaxiang Liu, Yuan Wang, Jiawei Du, Joey Tianyi Zhou, and Zuozhu Liu. Medcot: Medical chain of thought via hierarchical expert, 2024. 10

[130] Jiaxiang Liu, Yuan Wang, Jiawei Du, Joey Tianyi Zhou, and Zuozhu Liu. Medcot: Medical chain of thought via hierarchical expert. *arXiv preprint arXiv:2412.13736*, 2024. 63

[131] Xing Liu, Masanori Suganuma, Zhun Sun, and Takayuki Okatani. Dual residual networks leveraging the potential of paired operations for image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7007–7016, 2019. 10, 62

[132] Yuhao Liu, Zhanghan Ke, Fang Liu, Nanxuan Zhao, and Rynson WH Lau. Diff-plugin: Revitalizing details for diffusion-based low-level tasks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4197–4208, 2024. 28

[133] Jie Lu, Anjin Liu, Fan Dong, Feng Gu, Joao Gama, and Guangquan Zhang. Learning under concept drift: A review. *IEEE transactions on knowledge and data engineering*, 31(12):2346–2363, 2018. 61

[134] Zhisheng Lu, Juncheng Li, Hong Liu, Chaoyan Huang, Linlin Zhang, and Tieyong Zeng. Transformer for single image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 457–466, 2022. 55

[135] Luma. Dream-machine. https://lumalabs.ai/dream-machine, 2024. 59

[136] Ziwei Luo, Fredrik K Gustafsson, Zheng Zhao, Jens Sjölund, and Thomas B Schön. Controlling vision-language models for multi-task image restoration. *arXiv preprint arXiv:2310.01018*, 2023. 7, 35

[137] Xiaoqian Lv, Shengping Zhang, Chenyang Wang, Yichen Zheng, Bineng Zhong, Chongyi Li, and Liqiang Nie. Fourier priors-guided diffusion for zero-shot joint low-light enhancement and deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25378–25388, 2024. 28

[138] Jiaqi Ma, Tianheng Cheng, Guoli Wang, Qian Zhang, Xinggang Wang, and Lefei Zhang. Prores: Exploring degradation-aware visual prompt for universal image restoration. *arXiv preprint arXiv:2306.13653*, 2023. 10, 63

[139] Varun Mannam, Yide Zhang, Xiaotong Yuan, and Scott Howard. Deep learning-based super-resolution fluorescence microscopy on small datasets. In *Single Molecule Spectroscopy and Superresolution Imaging XIV*, volume 11650, pages 60–68. SPIE, 2021. 49

[140] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings eighth IEEE international conference on computer vision. ICCV 2001*, volume 2, pages 416–423. IEEE, 2001. 6, 30

[141] Daniel Enrique Martinez, Waiman Meinhold, John Oshinski, Ai-Ping Hu, and Jun Ueda. Super resolution for improved positioning of an mri-guided spinal cellular injection robot. *Journal of Medical Robotics Research*, 6(01n02):2140002, 2021. 60

[142] Yusuke Matsui, Kota Ito, Yuji Aramaki, Azuma Fujimoto, Toru Ogawa, Toshihiko Yamasaki, and Kiyoharu Aizawa. Sketch-based manga retrieval using manga109 dataset. *Multimedia tools and applications*, 76:21811–21838, 2017. 6, 30

[143] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. A survey on bias and fairness in machine learning. *ACM computing surveys (CSUR)*, 54(6):1–35, 2021. 61

[144] Kangfu Mei, Hossein Talebi, Mojtaba Ardakani, Vishal M Patel, Peyman Milanfar, and Mauricio Delbracio. The power of context: How multimodality improves image super-resolution. *arXiv preprint arXiv:2503.14503*, 2025. 2

[145] Yiqun Mei, Yuchen Fan, and Yuqian Zhou. Image super-resolution with non-local sparse attention. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3517–3526, 2021. 50

[146] Jacob Menick and Nal Kalchbrenner. Generating high fidelity images with subscale pixel networks and multidimensional upscaling. *arXiv preprint arXiv:1812.01608*, 2018. 62

[147] Sachit Menon, Alexandru Damian, Shijia Hu, Nikhil Ravi, and Cynthia Rudin. Pulse: Self-supervised photo upsampling via latent space exploration of generative models. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition*, pages 2437–2445, 2020. 10, 62

[148] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a completely blind image quality analyzer. *IEEE Signal processing letters*, 20(3):209–212, 2012. 5, 6

[149] Yogendra Rao Musunuri, Oh-Seol Kwon, and Sun-Yuan Kung. Srodnet: Object detection network based on super resolution for autonomous vehicles. *Remote Sensing*, 14(24):6270, 2022. 60, 61

[150] Saeel Sandeep Nachane, Ojas Gramopadhye, Prateek Chanda, Ganesh Ramakrishnan, Kshitij Sharad Jadhav, Yatin Nandwani, Dinesh Raghu, and Sachindra Joshi. Few shot chain-of-thought driven reasoning to prompt llms for open ended medical question answering. *arXiv preprint arXiv:2403.04890*, 2024. 63

[151] Babak Naderi, Ross Cutler, Juhee Cho, Nabakumar Khongbantabam, and Dejan Ivkovic. Icme 2025 grand challenge on video super-resolution for video conferencing. *arXiv preprint arXiv:2506.12269*, 2025. 58

[152] Valfride Nascimento, Rayson Laroca, Jorge de A Lambert, William Robson Schwartz, and David Menotti. Super-resolution of license plate images using attention modules and sub-pixel convolution layers. *Computers & Graphics*, 113:69–76, 2023. 59

[153] Ngoc Long Nguyen, Jérémy Anger, Axel Davy, Pablo Arias, and Gabriele Facciolo. Self-supervised multi-image super-resolution for push-frame satellite images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1121–1131, 2021. 60, 61

[154] Zhangkai Ni, Runyu Xiao, Wenhan Yang, Hanli Wang, Zhihua Wang, Lihua Xiang, and Liping Sun. M2trans: Multi-modal regularized coarse-to-fine transformer for ultrasound image super-resolution. *IEEE Journal of Biomedical and Health Informatics*, pages 1–12, 2024. 9, 30, 54, 55

[155] OpenAI, Tim Brooks, Bill Peebles, Connor Holmes, Will DePue, Yufei Guo, Li Jing, David Schnurr, Joe Taylor, Troy Luhman, Eric Luhman, Clarence Ng, Ricky Wang, and Aditya Ramesh. Video generation models as world simulators. https://openai.com/research/video-generation-models-as-world-simulators, 2024. 59

[156] Fu-Zhao Ou, Chongyi Li, Shiqi Wang, and Sam Kwong. Clib-fiqa: face image quality assessment with confidence calibration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1694–1704, 2024. 5, 36

[157] Suraj Pai, Ibrahim Hadzic, Dennis Bontempi, Keno Bressem, Benjamin H Kann, Andriy Fedorov, Raymond H Mak, and Hugo JWL Aerts. Vision foundation models for computed tomography. *arXiv preprint arXiv:2501.09001*, 2025. 54

[158] Xingang Pan, Xiaohang Zhan, Bo Dai, Dahua Lin, Chen Change Loy, and Ping Luo. Exploiting deep generative prior for versatile image restoration and manipulation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11):7474–7489, 2021. 10, 62

[159] Dongwon Park, Byung Hyun Lee, and Se Young Chun. All-in-one image restoration for unknown degradations using adaptive discriminative filters for specific degradations. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5815–5824. IEEE, 2023. 10, 63

[160] Shishir G Patil, Tianjun Zhang, Vivian Fang, Roy Huang, Aaron Hao, Martin Casado, Joseph E Gonzalez, Raluca Ada Popa, Ion Stoica, et al. Goex: Perspectives and designs towards a runtime for autonomous llm applications. *arXiv preprint arXiv:2404.06921*, 2024. 5

[161] Leonardo Peroni and Sergey Gorinsky. An end-to-end pipeline perspective on video streaming in best-effort networks: a survey and tutorial. *ACM Computing Surveys*, 2024. 61

[162] Vaishnav Potlapalli, Syed Waqas Zamir, Salman H Khan, and Fahad Shahbaz Khan. Promptir: Prompting for all-in-one image restoration. *Advances in Neural Information Processing Systems*, 36:71275–71293, 2023. 7, 35

[163] Darren Pouliot, Rasim Latifovic, Jon Pasher, and Jason Duffe. Landsat super-resolution enhancement using convolution neural networks and sentinel-2 for training. *Remote Sensing*, 10(3):394, 2018. 61

[164] Chang Qiao, Di Li, Yuting Guo, Chong Liu, Tao Jiang, Qionghai Dai, and Dong Li. Evaluation and development of deep neural networks for image super-resolution in optical microscopy. *Nature methods*, 18(2):194–202, 2021. 50

[165] Dongwei Ren, Wangmeng Zuo, Qinghua Hu, Pengfei Zhu, and Deyu Meng. Progressive image deraining networks: A better and simpler baseline. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3937–3946, 2019. 2

[166] Lingyan Ruan, Mojtaba Bemana, Hans-peter Seidel, Karol Myszkowski, and Bin Chen. Revisiting image deblurring with an efficient convnet. *arXiv preprint arXiv:2302.02234*, 2023. 28

[167] Lingyan Ruan, Bin Chen, Jizhou Li, and Miuling Lam. Learning to deblur using light field generated and real defocus images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16304–16313, 2022. 28

[168] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement. *IEEE transactions on pattern analysis and machine intelligence*, 45(4):4713–4726, 2022. 2, 3, 10, 62

[169] Lothar Schermelleh, Alexia Ferrand, Thomas Huser, Christian Eggeling, Markus Sauer, Oliver Biehlmaier, and Gregor PC Drummen. Super-resolution microscopy demystified. *Nature cell biology*, 21(1):72–84, 2019. 2

[170] Tixiao Shan, Jinkun Wang, Fanfei Chen, Paul Szenher, and Brendan Englot. Simulation-based lidar super-resolution for ground vehicles. *Robotics and Autonomous Systems*, 134:103647, 2020. 2

[171] Jacob Shermeyer and Adam Van Etten. The effects of super-resolution on object detection performance in satellite imagery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019. 2, 42, 60, 61

[172] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. Real-time single image and video super-resolution using an efficient subpixel convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1874–1883, 2016. 54

[173] Dehua Song, Yunhe Wang, Hanting Chen, Chang Xu, Chunjing Xu, and DaCheng Tao. Addersr: Towards energy efficient image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 15648–15657, 2021. 61

[174] Yuda Song, Zhuqing He, Hui Qian, and Xin Du. Vision transformers for single image dehazing. *IEEE Transactions on Image Processing*, 32:1927–1941, 2023. 28

[175] FUJIFILM Healthcare Europe & SonoSkills. US-CASE: Ultrasound Cases Dataset. http://www.ultrasoundcases.info/Cases-Home.aspx, 2025. 9, 30, 54

[176] J. J. Staal, M. D. Abràmoff, M. Niemeijer, M. A. Viergever, and B. van Ginneken. Ridge-based vessel segmentation in color images of the retina. *IEEE Transactions on Medical Imaging*, 23(4):501–509, 2004. 9, 30, 54

[177] Theodore Sumers, Shunyu Yao, Karthik Narasimhan, and Thomas Griffiths. Cognitive architectures for language agents. *Transactions on Machine Learning Research*, 2023. 63

[178] Theodore R. Sumers, Shunyu Yao, Karthik Narasimhan, and Thomas L. Griffiths. Cognitive architectures for language agents, 2024. 10

[179] Lingchen Sun, Rongyuan Wu, Zhiyuan Ma, Shuaizheng Liu, Qiaosi Yi, and Lei Zhang. Pixel-level and semantic-level adjustable super-resolution: A dual-lora approach. *arXiv preprint arXiv:2412.03017*, 2024. 2, 8, 28, 31, 32, 33, 37, 39, 43, 44, 45, 62

[180] Ying Tai, Jian Yang, and Xiaoming Liu. Image super-resolution via deep recursive residual network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3147–3155, 2017. 50

[181] Ying Tai, Jian Yang, Xiaoming Liu, and Chunyan Xu. Memnet: A persistent memory network for image restoration. In *Proceedings of the IEEE international conference on computer vision*, pages 4539–4547, 2017. 50

[182] Xin Tao, Hongyun Gao, Xiaoyong Shen, Jue Wang, and Jiaya Jia. Scale-recurrent network for deep image deblurring. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8174–8182, 2018. 2

[183] Muhammed Telçeken, Devrim Akgun, Sezgin Kacar, and Bunyamin Bingol. A new approach for super resolution object detection using an image slicing algorithm and the segment anything model. *Sensors*, 24(14):4526, 2024. 60

[184] Jennifer A Thorley, Jeremy Pike, and Joshua Z Rappoport. Super-resolution microscopy: a comparison of commercially available options. In *Fluorescence microscopy*, pages 199–212. Elsevier, 2014. 9, 49

[185] Kalina L Tosheva, Yue Yuan, Pedro Matos Pereira, Siân Culley, and Ricardo Henriques. Between life and death: strategies to reduce phototoxicity in super-resolution microscopy. *Journal of Physics D: Applied Physics*, 53(16):163001, 2020. 9, 49

[186] Zhengzhong Tu, Hossein Talebi, Han Zhang, Feng Yang, Peyman Milanfar, Alan Bovik, and Yinxiao Li. Maxim: Multi-axis mlp for image processing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5769–5780, 2022. 2, 10, 28, 62

[187] Zhengzhong Tu, Yilin Wang, Neil Birkbeck, Balu Adsumilli, and Alan C Bovik. Ugc-vqa: Benchmarking blind video quality assessment for user generated content. *IEEE Transactions on Image Processing*, 30:4449–4464, 2021. 2

[188] Sabina Umirzakova, Shabir Ahmad, Latif U Khan, and Taegkeun Whangbo. Medical image super-resolution for smart healthcare applications: A comprehensive survey. *Information Fusion*, 103:102075, 2024. 9, 53

[189] Jeya Maria Jose Valanarasu, Rajeev Yasarla, and Vishal M Patel. Transweather: Transformer-based restoration of images degraded by adverse weather conditions. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2353–2363, 2022. 10, 62, 63

[190] Aaron Van den Oord, Nal Kalchbrenner, Lasse Espeholt, Oriol Vinyals, Alex Graves, et al. Conditional image generation with pixelcnn decoders. *Advances in neural information processing systems*, 29, 2016. 62

[191] Veo-Team, Agrim Gupta, Ali Razavi, Andeep Toor, Ankush Gupta, Dumitru Erhan, Eleni Shaw, Eric Lau, Frank Belletti, Gabe Barth-Maron, Gregory Shaw, Hakan Erdogan, Hakim Sidahmed, Henna Nandwani, Hernan Moraldo, Hyunjik Kim, Irina Blok, Jeff Donahue, José Lezama, Kory Mathewson, Kurtis David, Matthieu Kim Lorrain, Marc van Zee, Medhini Narasimhan, Miaosen Wang, Mohammad Babaeizadeh, Nelly Papalampidi, Nick Pezzotti, Nilpa Jha, Parker Barnes, Pieter-Jan Kindermans, Rachel Hornung, Ruben Villegas, Ryan Poplin, Salah Zaiem, Sander Dieleman, Sayna Ebrahimi, Scott Wisdom, Serena Zhang, Shlomi Fruchter, Signe Nørly, Weizhe Hua, Xinchen Yan, Yuqing Du, and Yutian Chen. Veo 2. https://deepmind.google/technologies/veo/veo-2, 2024. 59

[192] W3C. Research - low vision accessibility task force. https://www.w3.org/WAI/GL/low-vision-a11y-tf/wiki/Research, 2025. 61

[193] Hang Wang, Xuanhong Chen, Bingbing Ni, Yutian Liu, and Jinfan Liu. Omni aggregation networks for lightweight image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22378–22387, 2023. 50

[194] Jianyi Wang, Kelvin CK Chan, and Chen Change Loy. Exploring clip for assessing the look and feel of images. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pages 2555–2563, 2023. 4, 5, 30

[195] Jianyi Wang, Zongsheng Yue, Shangchen Zhou, Kelvin CK Chan, and Chen Change Loy. Exploiting diffusion prior for real-world image super-resolution. *International Journal of Computer Vision*, 132(12):5929–5949, 2024. 2, 7, 10, 33, 62

[196] Jiarui Wang, Binglu Wang, Xiaoxu Wang, Yongqiang Zhao, and Teng Long. Hybrid attention-based u-shaped network for remote sensing image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–15, 2023. 43, 44, 45

[197] Xiaohan Wang, Yuhui Zhang, Orr Zohar, and Serena Yeung-Levy. Videoagent: Long-form video understanding with large language model as agent. In *European Conference on Computer Vision*, pages 58–76. Springer, 2024. 63

[198] Xiaohan Wang, Yuhui Zhang, Orr Zohar, and Serena Yeung-Levy. Videoagent: Long-form video understanding with large language model as agent. In Aleš Leonardis, Elisa Ricci, Stefan Roth, Olga Russakovsky, Torsten Sattler, and Gül Varol, editors, *Computer Vision – ECCV 2024*, pages 58–76, Cham, 2025. Springer Nature Switzerland. 10

[199] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and Ronald M Summers. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2097–2106, 2017. 9, 30, 54

[200] Xintao Wang, Yu Li, Honglun Zhang, and Ying Shan. Towards real-world blind face restoration with generative facial prior. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9168–9178, 2021. 28, 30, 36

[201] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1905–1914, 2021. 2, 10, 62

[202] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision (ECCV) workshops*, 2018. 2

[203] Yifan Wang, Federico Perazzi, Brian McWilliams, Alexander Sorkine-Hornung, Olga Sorkine-Hornung, and Christopher Schroers. A fully progressive approach to single-image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 864–873, 2018. 50

[204] Yinhuai Wang, Jiwen Yu, and Jian Zhang. Zero-shot image restoration using denoising diffusion null-space model. *arXiv preprint arXiv:2212.00490*, 2022. 10, 62

[205] Yufei Wang, Wenhan Yang, Xinyuan Chen, Yaohui Wang, Lanqing Guo, Lap-Pui Chau, Ziwei Liu, Yu Qiao, Alex C Kot, and Bihan Wen. Sinsr: diffusion-based image super-resolution in a single step. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 25796–25805, 2024. 7, 10, 33, 62

[206] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 17683–17693, 2022. 10, 62

[207] Zhihao Wang, Jian Chen, and Steven CH Hoi. Deep learning for image super-resolution: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 43(10):3365–3387, 2020. 2

[208] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 6, 30

[209] Zhou Wang, Eero P Simoncelli, and Alan C Bovik. Multiscale structural similarity for image quality assessment. In *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, volume 2, pages 1398–1402. Ieee, 2003. 54

[210] Zijie J Wang, Evan Montoya, David Munechika, Haoyang Yang, Benjamin Hoover, and Duen Horng Chau. Diffusiondb: A large-scale prompt gallery dataset for text-to-image generative models. *arXiv preprint arXiv:2210.14896*, 2022. 9, 30, 40, 41

[211] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018. 2

[212] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, brian ichter, Fei Xia, Ed Chi, Quoc V Le, and Denny Zhou. Chain-of-thought prompting elicits reasoning in large language models. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 24824–24837. Curran Associates, Inc., 2022. 10

[213] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837, 2022. 2, 63

[214] Pengxu Wei, Ziwei Xie, Hannan Lu, Zongyuan Zhan, Qixiang Ye, Wangmeng Zuo, and Liang Lin. Component divide-and-conquer for real-world image super-resolution. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VIII 16*, pages 101–117. Springer, 2020. 7, 30

[215] Martin Weigert, Uwe Schmidt, Tobias Boothe, Andreas Müller, Alexandr Dibrov, Akanksha Jain, Benjamin Wilhelm, Deborah Schmidt, Coleman Broaddus, Siân Culley, et al. Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nature methods*, 15(12):1090–1097, 2018. 60, 61

[216] Rongyuan Wu, Lingchen Sun, Zhiyuan Ma, and Lei Zhang. One-step effective diffusion network for real-world image super-resolution. *Advances in Neural Information Processing Systems*, 37:92529–92553, 2024. 2, 6, 7, 8, 10, 28, 31, 32, 33, 37, 39, 43, 44, 45, 62

[217] Rongyuan Wu, Tao Yang, Lingchen Sun, Zhengqiang Zhang, Shuai Li, and Lei Zhang. Seesr: Towards semantics-aware real-world image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 25456–25467, 2024. 2, 7, 10, 33, 62

[218] Rui-Qi Wu, Zheng-Peng Duan, Chun-Le Guo, Zhi Chai, and Chongyi Li. Ridcp: Revitalizing real image dehazing via high-quality codebook priors. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 22282–22291, 2023. 28

[219] Xiaoshi Wu, Yiming Hao, Keqiang Sun, Yixiong Chen, Feng Zhu, Rui Zhao, and Hongsheng Li. Human preference score v2: A solid benchmark for evaluating human preferences of text-to-image synthesis. *arXiv preprint arXiv:2306.09341*, 2023. 5

[220] Bin Xia, Yucheng Hang, Yapeng Tian, Wenming Yang, Qingmin Liao, and Jie Zhou. Efficient non-local contrastive attention for image super-resolution. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pages 2759–2767, 2022. 50

[221] Bin Xia, Yulun Zhang, Shiyin Wang, Yitong Wang, Xinglong Wu, Yapeng Tian, Wenming Yang, and Luc Van Gool. Diffir: Efficient diffusion model for image restoration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 13095–13105, 2023. 10, 62

[222] Gui-Song Xia, Xiang Bai, Jian Ding, Zhen Zhu, Serge Belongie, Jiebo Luo, Mihai Datcu, Marcello Pelillo, and Liangpei Zhang. Dota: A large-scale dataset for object detection in aerial images. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3974–3983, 2018. 9, 30, 43

[223] Gui-Song Xia, Jingwen Hu, Fan Hu, Baoguang Shi, Xiang Bai, Yanfei Zhong, Liangpei Zhang, and Xiaoqiang Lu. Aid: A benchmark data set for performance evaluation of aerial scene classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(7):3965–3981, 2017. 9, 30, 42

[224] Peng Xia, Jinglu Wang, Yibo Peng, Kaide Zeng, Xian Wu, Xiangru Tang, Hongtu Zhu, Yun Li, Shujie Liu, Yan Lu, et al. Mmedagent-rl: Optimizing multi-agent collaboration for multimodal medical reasoning. *arXiv preprint arXiv:2506.00555*, 2025. 63

[225] Jun Xiao, Xinyang Jiang, Ningxin Zheng, Huan Yang, Yifan Yang, Yuqing Yang, Dongsheng Li, and Kin-Man Lam. Online video super-resolution with convolutional kernel bypass grafts. *IEEE Transactions on Multimedia*, 25:8972–8987, 2023. 58

[226] Jun Xiao, Tianshan Liu, Rui Zhao, and Kin-Man Lam. Balanced distortion and perception in single-image super-resolution based on optimal transport in wavelet domain. *Neurocomputing*, 464:408–420, 2021. 62

[227] Yi Xiao, Qiangqiang Yuan, Kui Jiang, Yuzeng Chen, Qiang Zhang, and Chia-Wen Lin. Frequency-assisted mamba for remote sensing image super-resolution. *IEEE Transactions on Multimedia*, 2024. 43

[228] Yi Xiao, Qiangqiang Yuan, Kui Jiang, Jiang He, Xianyu Jin, and Liangpei Zhang. Ediffsr: An efficient diffusion probabilistic model for remote sensing image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–14, 2023. 43

[229] Enze Xie, Junsong Chen, Junyu Chen, Han Cai, Haotian Tang, Yujun Lin, Zhekai Zhang, Muyang Li, Ligeng Zhu, Yao Lu, et al. Sana: Efficient high-resolution image synthesis with linear diffusion transformers. *arXiv preprint arXiv:2410.10629*, 2024. 9, 40

[230] Liming Xu, Xianhua Zeng, Zhiwei Huang, Weisheng Li, and He Zhang. Low-dose chest x-ray image super-resolution using generative adversarial nets with spectral normalization. *Biomedical Signal Processing and Control*, 55:101600, 2020. 54, 55

[231] Feng Yang, Yue-Min Zhu, Jian-Hua Luo, Marc Robini, Jie Liu, and Pierre Croisille. A comparative study of different level interpolations for improving spatial resolution in diffusion tensor imaging. *IEEE Journal of Biomedical and Health Informatics*, 18(4):1317–1327, 2014. 54

[232] Jianchao Yang, John Wright, Thomas S Huang, and Yi Ma. Image super-resolution via sparse representation. *IEEE transactions on image processing*, 19(11):2861–2873, 2010. 2

[233] Sidi Yang, Tianhe Wu, Shuwei Shi, Shanshan Lao, Yuan Gong, Mingdeng Cao, Jiahao Wang, and Yujiu Yang. Maniqa: Multi-dimension attention network for no-reference image quality assessment. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1191–1200, 2022. 5, 30

[234] Tao Yang, Rongyuan Wu, Peiran Ren, Xuansong Xie, and Lei Zhang. Pixel-aware stable diffusion for realistic image super-resolution and personalized stylization. In *European Conference on Computer Vision*, pages 74–91. Springer, 2024. 7, 10, 33, 62

[235] Tianjie Yang, Yaoru Luo, Wei Ji, and Ge Yang. Advancing biological super-resolution microscopy through deep learning: a brief review. *Biophysics Reports*, 7(4):253, 2021. 49

[236] Mingde Yao, Ruikang Xu, Yuanshen Guan, Jie Huang, and Zhiwei Xiong. Neural degradation representation learning for all-in-one image restoration. *IEEE Transactions on Image Processing*, 2024. 10, 63

[237] Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. React: Synergizing reasoning and acting in language models. In *International Conference on Learning Representations (ICLR)*, 2023. 2, 10, 63

[238] Zhenqiang Ying, Haoran Niu, Praful Gupta, Dhruv Mahajan, Deepti Ghadiyaram, and Alan Bovik. From patches to pictures (paq-2-piq): Mapping the perceptual space of picture quality. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3575–3585, 2020. 2

[239] Jinsu Yoo, Taehoon Kim, Sihaeng Lee, Seung Hwan Kim, Honglak Lee, and Tae Hyun Kim. Enriched cnn-transformer feature aggregation networks for super-resolution. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 4956–4965, 2023. 50

[240] Chenyu You, Guang Li, Yi Zhang, Xiaoliu Zhang, Hongming Shan, Mengzhou Li, Shenghong Ju, Zhen Zhao, Zhuiyang Zhang, Wenxiang Cong, et al. Ct super-resolution gan constrained by the identical, residual, and cycle learning ensemble (gan-circle). *IEEE transactions on medical imaging*, 39(1):188–203, 2019. 54

[241] Zhiyuan You, Zheyuan Li, Jinjin Gu, Zhenfei Yin, Tianfan Xue, and Chao Dong. Depicting beyond scores: Advancing image quality assessment through multi-modal language models. In *European Conference on Computer Vision*, pages 259–276. Springer, 2024. 27

[242] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. Free-form image inpainting with gated convolution. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4471–4480, 2019. 10, 62

[243] Miao Yu, Zhenghua Xu, and Thomas Lukasiewicz. A general survey on medical image super-resolution via deep learning. *Computers in Biology and Medicine*, 193:110345, Jul 2025. 53

[244] Zongsheng Yue and Chen Change Loy. Difface: Blind face restoration with diffused error contraction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024. 28, 36

[245] Zongsheng Yue, Jianyi Wang, and Chen Change Loy. Resshift: Efficient diffusion model for image super-resolution by residual shifting. *Advances in Neural Information Processing Systems*, 36:13294–13307, 2023. 7, 33

[246] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5728–5739, 2022. 2, 10, 28, 62

[247] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Learning enriched features for real image restoration and enhancement. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXV 16*, pages 492–511. Springer, 2020. 10, 62

[248] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14821–14831, 2021. 28

[249] Roman Zeyde, Michael Elad, and Matan Protter. On single image scale-up using sparse-representations. In *International conference on curves and surfaces*, pages 711–730. Springer, 2010. 6, 30

[250] Zheng Zhan, Yifan Gong, Pu Zhao, Geng Yuan, Wei Niu, Yushu Wu, Tianyun Zhang, Malith Jayaweera, David Kaeli, Bin Ren, et al. Achieving on-mobile real-time super-resolution with neural architecture and pruning search. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4821–4831, 2021. 58

[251] Aoyang Zhang, Qing Li, Ying Chen, Xiaoteng Ma, Longhao Zou, Yong Jiang, Zhimin Xu, and Gabriel-Miro Muntean. Video super-resolution and cachingan edge-assisted adaptive video streaming solution. *IEEE Transactions on Broadcasting*, 67(4):799–812, 2021. 61

[252] Cheng Zhang, Yu Zhu, Qingsen Yan, Jinqiu Sun, and Yanning Zhang. All-in-one multi-degradation image restoration network via hierarchical degradation representation. In *Proceedings of the 31st ACM international conference on multimedia*, pages 2285–2293, 2023. 10, 63

[253] Dafeng Zhang, Feiyu Huang, Shizhuo Liu, Xiaobing Wang, and Zhezhu Jin. Swinfir: Revisiting the swinir with fast fourier convolution and improved training for image super-resolution. *arXiv preprint arXiv:2208.11247*, 2022. 28, 30

[254] Jinjin Zhang, Qiuyu Huang, Junjie Liu, Xiefan Guo, and Di Huang. Diffusion-4k: Ultra-high-resolution image synthesis with latent diffusion models. *arXiv preprint arXiv:2503.18352*, 2025. 8, 9, 38, 40

[255] Kai Zhang, Yawei Li, Wangmeng Zuo, Lei Zhang, Luc Van Gool, and Radu Timofte. Plug-and-play image restoration with deep denoiser prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10):6360–6376, 2021. 10, 62

[256] Kai Zhang, Jingyun Liang, Luc Van Gool, and Radu Timofte. Designing a practical degradation model for deep blind image super-resolution. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4791–4800, 2021. 2, 8, 10, 30, 62

[257] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE transactions on image processing*, 26(7):3142–3155, 2017. 2, 10, 62

[258] Kaibing Zhang, Dacheng Tao, Xinbo Gao, Xuelong Li, and Jie Li. Coarse-to-fine learning for single-image super-resolution. *IEEE transactions on neural networks and learning systems*, 28(5):1109–1122, 2016. 54

[259] Liangpei Zhang, Hongyan Zhang, Huanfeng Shen, and Pingxiang Li. A super-resolution reconstruction algorithm for surveillance images. *Signal Processing*, 90(3):848–859, 2010. 2

[260] Lin Zhang, Lei Zhang, and Alan C Bovik. A feature-enriched completely blind image quality evaluator. *IEEE Transactions on Image Processing*, 24(8):2579–2591, 2015. 4, 30

[261] Lin Zhang, Lei Zhang, Xuanqin Mou, and David Zhang. Fsim: A feature similarity index for image quality assessment. *IEEE transactions on Image Processing*, 20(8):2378–2386, 2011. 54

[262] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018. 6, 30

[263] Xindong Zhang, Hui Zeng, Shi Guo, and Lei Zhang. Efficient long-range attention network for image super-resolution. In *European conference on computer vision*, pages 649–667. Springer, 2022. 55

[264] Xindong Zhang, Hui Zeng, and Lei Zhang. Edge-oriented convolution block for real-time super resolution on mobile devices. In *Proceedings of the 29th ACM international conference on multimedia*, pages 4034–4043, 2021. 61

[265] Yide Zhang, Zhe He, Xin Tong, David C Garrett, Rui Cao, and Lihong V Wang. Quantum imaging of biological organisms through spatial and polarization entanglement. *Science Advances*, 10(10):eadk1495, 2024. 60

[266] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European conference on computer vision (ECCV)*, pages 286–301, 2018. 6, 10, 52, 62

[267] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2472–2481, 2018. 2, 10, 52, 62

[268] Zhisong Zhang, Tianqing Fang, Kaixin Ma, Wenhao Yu, Hongming Zhang, Haitao Mi, and Dong Yu. Enhancing web agents with explicit rollback mechanisms. *arXiv preprint arXiv:2504.11788*, 2025. 5

[269] Xiaole Zhao, Yulun Zhang, Tao Zhang, and Xueming Zou. Channel splitting network for single mr image super-resolution. *IEEE transactions on image processing*, 28(11):5649–5662, 2019. 60

[270] Qi Zheng, Yibo Fan, Leilei Huang, Tianyu Zhu, Jiaming Liu, Zhijian Hao, Shuo Xing, Chia-Ju Chen, Xiongkuo Min, Alan C Bovik, et al. Video quality assessment: A comprehensive survey. *arXiv preprint arXiv:2412.04508*, 2024. 6

[271] Shangchen Zhou, Kelvin Chan, Chongyi Li, and Chen Change Loy. Towards robust blind face restoration with codebook lookup transformer. *Advances in Neural Information Processing Systems*, 35:30599–30611, 2022. 28, 36

[272] Yingjie Zhou, Jiezhang Cao, Zicheng Zhang, Farong Wen, Yanwei Jiang, Jun Jia, Xiaohong Liu, Xiongkuo Min, and Guangtao Zhai. Q-agent: Quality-driven chain-of-thought image restoration agent through robust multimodal large language model. *arXiv preprint arXiv:2504.07148*, 2025. 10, 63

[273] Yiyang Zhou, Yangfan He, Yaofeng Su, Siwei Han, Joel Jang, Gedas Bertasius, Mohit Bansal, and Huaxiu Yao. Reagent-v: A reward-driven multi-agent framework for video understanding. *arXiv preprint arXiv:2506.01300*, 2025. 63

[274] Kaiwen Zhu, Jinjin Gu, Zhiyuan You, Yu Qiao, and Chao Dong. An intelligent agentic system for complex image restoration problems. *arXiv preprint arXiv:2410.17809*, 2024. 2, 5, 6, 7, 10, 29, 30, 31, 32, 33, 35, 36, 37, 39, 44, 45, 57, 63

[275] Ruoxi Zhu, Zhengzhong Tu, Jiaming Liu, Alan C Bovik, and Yibo Fan. Mwformer: Multi-weather image restoration using degradation-aware transformers. *IEEE Transactions on Image Processing*, 2024. 10, 62

[276] Karel J Zuiderveld et al. Contrast limited adaptive histogram equalization. *Graphics gems*, 4(1):474–485, 1994. 28

A Model Card

A.1 Profile

4KAgent is highly flexible through configurable profiles. Users can easily customize 4KAgent by selecting a pre-defined profile or creating a new one. We provide a set of representative pre-defined profiles in Tab. 5, which cover most use cases and include all the profiles used in our experiments. This design enables easy and intuitive customization for unseen scenarios.

Table 5: Pre-defined Profiles in 4KAgent.

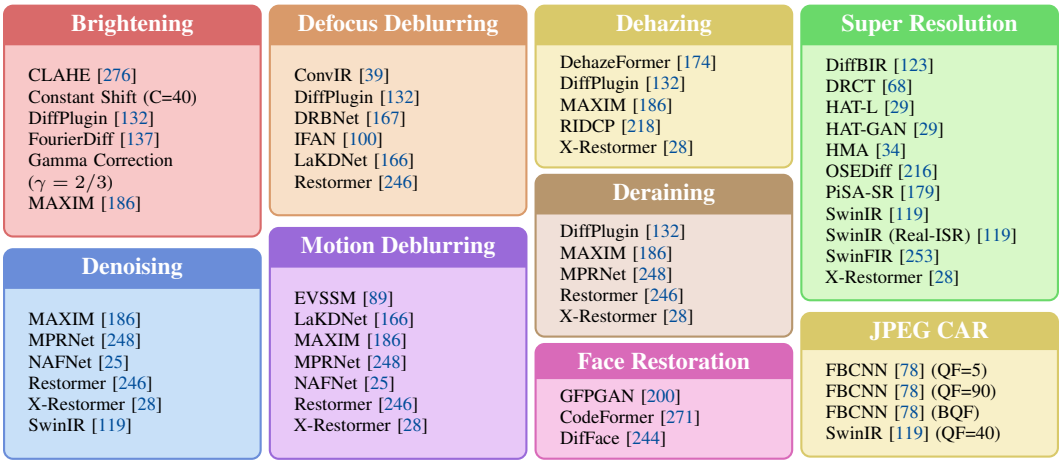
Profile Nickname	Perception Agent	Upscale to 4K	Scale Factor	Restore Option	Face Restore	Brightening	Restore Preference
Gen4K-P	DepictQA [241]	True	None	None	True	False	Perception
Gen4K-F	DepictQA [241]	True	None	None	True	False	Fidelity
Aer4K-P	Llama-3.2-Vision [6]	True	None	None	False	False	Perception
Aer4K-F	Llama-3.2-Vision [6]	True	None	None	False	False	Fidelity
AerSR-s4-P	Llama-3.2-Vision [6]	False	4	None	False	False	Perception
AerSR-s4-F	Llama-3.2-Vision [6]	False	4	None	False	False	Fidelity
ExpSR-s4-P	Llama-3.2-Vision [6]	False	4	super-resolution	False	False	Perception
ExpSR-s4-F	Llama-3.2-Vision [6]	False	4	super-resolution	False	False	Fidelity
ExpSR-s2-F	Llama-3.2-Vision [6]	False	2	super-resolution	False	False	Fidelity
ExpSR-s8-F	Llama-3.2-Vision [6]	False	8	super-resolution	False	False	Fidelity
GenSR-s4-P	DepictQA [241]	False	4	None	False	False	Perception
GenMIR-P	DepictQA [241]	False	4	None	False	True	Perception
ExpSRFR-s4-P	Llama-3.2-Vision [6]	False	4	super-resolution	True	False	Perception
GenSRFR-s4-P	DepictQA [241]	False	4	None	True	False	Perception

Profile naming convention: We combine restoration type, restoration task, and restoration preference to construct the profile name. For example, **Gen** indicates a General image, **Aer** indicates Aerial image, **4K** indicates "Upscale to 4K" on, and **P** indicates to restore the image with high Perceptual quality. **Exp** corresponds to Explicit setting, indicating that the profile has explicitly set the restoration task (*e.g.*, **SR**, which indicates Super-Resolution). **MIR** indicates Multiple Image Restoration. **FR** indicates Face Restoration. **s4** indicates to upscale the image by a scale factor of 4.

4KAgent supports various VLMs and LLMs within the Perception Agent, enabling effective analysis of image content and degradation. Specifically, users can select from DepictQA [241] or Llama-3.2-Vision (11B) [6], while the system can be readily extended to incorporate more recent VLMs, *e.g.*, Qwen2.5-VL [10]. For the VLMs or LLMs to schedule the restoration plan, users can choose from GPT-4 [4], or Llama-3.2-Vision. This is configured by the **Perception Agent** in the profile module. For example, when it is set to **Llama-3.2-Vision**, the Llama-3.2-Vision model serves as the core engine to perceive image content and degradation, and then schedules the restoration plan P_I. As DepictQA is fine-tuned for image quality assessment (IQA), when it is set as the VLM in the perception agent, 4KAgent employs Llama-3.2-Vision to obtain the image description C_I and GPT-4 [4] to schedule the restoration plan.

A.2 Model Zoo

4KAgent constructs a restoration toolbox comprising nine distinct image restoration tasks: ❶ Brightening, ❷ Defocus Deblurring, ❸ Motion Deblurring, ❹ Dehazing, ❺ Denoising, ❻ Deraining, ❼ JPEG Compression Artifact Removal (JPEG CAR), ❽ Super Resolution, and ❾ Face Restoration. For each of these tasks, we integrate advanced state-of-the-art methods into our comprehensive restoration toolbox. Detailed correspondences between restoration tasks and their representative methods are presented below, where 'QF' denotes the JPEG Quality Factor and 'BQF' indicates methods that are blind to the Quality Factor in the JPEG CAR task.



As previously mentioned in Appendix A.1, users can customize 4KAgent by adjusting the **Restore Preference** setting, which prioritizes either perceptual quality or fidelity. We achieve this by categorizing the methods in the restoration toolbox into perception-oriented and fidelity-oriented groups. For example, the Super-Resolution tools in the toolbox are split into:

1. **Fidelity-based:** HAT-L [29], X-Restormer [28], SwinFIR [253], HMA [34], DRCT [68]
2. **Perception-based:** DiffBIR [123], HAT-GAN [29], OSEDiff [216], PiSA-SR [179], SwinIR (Real-ISR) [119]

Accordingly, when **Restore Preference** is set to **Perception**, 4KAgent restricts its execution to perception-based methods, thereby aligning the restoration process with the user's preference.

We further develop a **Fast4K** mode for 4KAgent. Specifically, when the size of the input image at the current step of the restoration plan exceeds a predefined threshold s_t, 4KAgent automatically filters out methods with long inference times from the toolbox, such as DiffBIR (a 50-step diffusion-based method) in the super-resolution toolbox. Users can adjust s_t to control the running time of 4KAgent. To comprehensively evaluate the performance of 4KAgent, Fast4K is disabled in all experiments.

A.3 Inference Details

A.3.1 Rollback

In this section, we present the detailed workflow of the rollback mechanism in 4KAgent. If the quality score of I_k after step k in the initial restoration plan P_I is lower than a threshold η, *i.e.*, $Q_s(I_k) \leq \eta$, this step is regarded as a failure step and 4KAgent generates a failure message S_I. Then 4KAgent invokes the perception agent to adjust the subsequent plan based on the degradation list D_I, the remaining restoration tasks A_I^R of the restoration agenda A_I, restoration experience E, and failure message S_I: $P_I^{adj} = M_P(D_I, A_I^R, E, S_I)$. After that, 4KAgent assigns an alternative restoration task for the current step. If all subsequent restoration tasks assigned to this step lead to rollback, 4KAgent adopts a fallback strategy and reverts to the original plan to execute subsequent tasks.

A.3.2 Implementation Details

Computational Resources. As a multi-agent system, 4KAgent supports multi-GPU deployment. Specifically, different agents (Perception Agent, Restoration Agent) are assigned to separate GPUs to conserve memory. Most of our experiments are conducted using two NVIDIA RTX 4090 GPUs.

Hyper-parameters. Hyperparameters in 4KAgent mainly reside in the Restoration Agent, including the weights used to compute the quality scores Q_s and Q_s^f in execution, as well as the quality threshold η for the rollback mechanism.. Specifically, we set $w_{\text{NIQE}} = 1.0, w_{\text{MUSIQ}} = 0.01, w_{\text{MANIQA}} = 1.0, w_{\text{CLIPIQA}} = 1.0$ for Q_s, $w_{\text{IP}} = 0.001, w_{\text{IQA}} = 1.0$ for Q_s^f, and $\eta = 0.5$ for rollback.

A.3.3 Prompts

In 4KAgent, we enable the VLM / LLM to perceive image degradations and formulate a restoration plan via customized system prompts. In this section, we present the details of these prompts.

When the **Perception Agent** in the profile module selects **DepictQA**, we use the same prompt as in AgenticIR [274] for DepictQA to assess the image degradations, and GPT-4 to generate the restoration plan. When **Llama-3.2-Vision** is selected in the **Perception Agent**, we design tailored prompts for degradation reasoning and restoration planning, as shown below, where **{·}** represents slots to fill according to the context, and the content inside comes from external input. For the restoration experience E in 4KAgent, we employ the restoration experience from AgenticIR.

Prompt for Llama-Vision in Degradation Reasoning

```
You are an expert tasked with image quality assessment (IQA) and
well-versed in popular IQA metrics, including CLIPIQA+, TOPIQ_NR,
MUSIQ, and NIQE. Note that for NIQE, a lower score indicates
better image quality, whereas for the other metrics, higher scores
generally reflect better quality.  Here's an image to restore,
along with its corresponding quality scores evaluated using the
aforementioned IQA metrics.  First, please describe the content
and style of the input image, the description must not contain
its image quality.  Second, please assess the image based on both
the metric scores and your prior visual knowledge, with respect to
the following two degradations:  noise, motion blur, defocus blur,
haze, rain, jpeg compression artifact.  Images may suffer from one
or more of these degradations.  **Do not output any explanations
or comments.** **Strictly return only a JSON object** containing
degradation types and image content/style description.  The keys in
the JSON object should be:  'degradations' and 'image_description'.
Information about the input image:  IQA metrics:  {iqa_result}.
({iqa_result} corresponds to Q_I.)
```

Prompt for Llama-Vision in Planning (Rollback)

```
You are an expert in image restoration.  Given an image of low
quality, your task is to guide the user to utilize various tools
to enhance its quality.  The input image requires a list of
restoration tasks.  Your goal is to make a plan (the order of
the tasks) based on the task list.  The final output should be
formatted as a JSON object containing the restoration plan (the
correct order of the tasks).  The key in the JSON object should be:
'plan'.  Information about the input image:
Its description is:  {image_description} (C_I),
It suffer from degradations {degradations} (D_I),
The list of restoration tasks:  {tasks} (A_I / A_I^R),
For your information, based on past trials, we have the following
experience in making a restoration plan:  {experience} (E).  Based
on this experience, please give the correct order of the tasks in
the restoration plan.  The restoration plan must be a permutation
of {tasks} in the order you determine.  (Besides, in attempts just
now, we found the result is unsatisfactory if {failed_tries} (S_I)
is conducted first.  Remember not to arrange {failed_tries} in the
first place.)  **Do not output any explanations or comments.**
**Strictly return only a JSON object** containing plan.  The keys
in the JSON object should be:  'plan'.
```

B Experiment Overview

We evaluate 4KAgent on a variety of complex image restoration and super-resolution (SR) tasks to demonstrate its profile-driven flexibility under different restoration requirements, validate its generalization across multiple image domains, and quantify the contribution of each core component via ablation studies. Specifically, we evaluate 4KAgent across a wide range of **11** image SR tasks on **26** benchmarks. The summary of datasets used in experiments is shown in Tab. 6, which can be classified as natural degraded images (§C,§D), AI-generated images (§E), and scientific images (§F).

Then, we perform an ablation study on **Q-MoE** policy and **Face Restoration Pipeline** in 4KAgent, together with a running time analysis (§G).

First, we evaluate 4KAgent on natural image restoration and super-resolution tasks under standard settings, including classical image super-resolution (4×) (§C.1), real-world image super-resolution (4×) (§C.2), multiple-degradation image restoration (§C.3), and face restoration (4×) (§C.4). Then, we consider more challenging scenarios, such as large scale factor super-resolution (16×) (§D.1) and joint restoration with 4K upscaling (§D.2). Finally, we extend 4KAgent to diverse domains by evaluating it on AIGC images (§E.1) and scientific imagery (§F), including remote sensing (§F.1), microscopy (§F.2), pathology (§F.3), and medical images (§F.4). To comprehensively evaluate the performance of 4KAgent, we disable the Fast4K mode in all experiments.

C Experiment Part I: 4× Natural Image Super-Resolution

C.1 Classical Image Super-Resolution

Settings In this section, we provide detailed experimental results on Set5 [13], Set14 [249], B100 [140], Urban100 [70], and Manga109 [142] datasets. For a more comprehensive comparison, we include recent methods, including HMA [34], DRCT [68], and SwinFIR [253], as they are contained in the toolbox. In addition to metrics used in the main paper (PSNR, SSIM [208], LPIPS [262], FID [66], NIQE [260], MUSIQ [83]), we employ DISTS [43], CLIPIQA [194], and MANIQA-pipal [233] for evaluation.

Specifically, PSNR and SSIM are computed on the Y channel in the YCbCr color space and are used to measure the fidelity of images. LPIPS and DISTS are computed in the RGB space and evaluate perceptual quality with reference images. FID is used to evaluate the distance of distributions between the ground truth and the restored images. NIQE, CLIP-IQA, MUSIQ, and MANIQA-pipal are used to evaluate the perceptual quality of images without reference images.

Table 6: Dataset summary in 4KAgent experiments.

Task	Dataset	#Test Images
Classical SR (§C.1)	Set5 [13]	5
	Set14 [249]	14
	B100 [140]	100
	Urban100 [70]	100
	Manga109 [142]	109
Real-World SR (§C.2)	RealSR [17]	100
	DrealSR [214]	93
Multiple-Degradation IR (§C.3)	MiO-Group A [274]	640
	MiO-Group B [274]	400
	MiO-Group C [274]	400
Face Restoration (§C.4)	WebPhoto-Test [200]	407
Large Scale Factor SR (§D.1)	RealSRSet [256]	20
Joint Image Restoration + 4K Upscaling (§D.2)	DIV4K-50 (Ours)	50
AI-Generated Content 4K SR (§E.1)	GenAIBench-4K [103]	100
	DiffusionDB-4K [210]	100
Remote Sensing SR (§F.1)	AID [223]	51
	DIOR [111]	154
	DOTA [222]	183
	WorldStrat [37]	85
Fluorescence Microscopy Image SR (§F.2)	SR-CACO-2 [12]	300
Pathology Image SR (§F.3)	bcSR [77]	200
Medical Image SR (§F.4)	Chest X-ray 2017 [85]	624
	Chest X-ray 14 [199]	880
	US-Case [175]	111
	MMUS1K [154]	100
	DRIVE [176]	20

Quantitative Comparison It should be noted that, once the user sets the **Restore Option** to **super-resolution** in the profile, the 4KAgent system can be seen as a quality-driven Mixture-of-Expert system for image super-resolution. In this mode, the system sequentially invokes every super-resolution tool in its toolbox based on the **Restore Option** setting in the profile, then selects the best result based on the quality score Q_s. Accordingly, we group 4KAgent with **ExpSR-s4-F** and **ExpSR-s4-P** profile into ***Fidelity-based method*** and ***Perception-based method***.

Experimental results are shown in Tabs. 7 and 8. For the commonly used fidelity metrics PSNR and SSIM in the classical image SR task, 4KAgent with **ExpSR-s4-F** profile shows competitive performance compared to state-of-the-art fidelity-based methods, ranking among the top three on Set5, B100, Urban100, and Manga109 datasets. For perception-based methods, we focus more on perceptual metrics such as NIQE, CLIPIQA, MUSIQ, and MANIQA. By simply switching the profile from **ExpSR-s4-F** to **ExpSR-s4-P**, 4KAgent achieves strong performance among state-of-

Table 7: Quantitative comparison on classical image super-resolution benchmarks (Set5, Set14, B100). The top three results for each metric are marked in **bold**, underline, and *italic*. For Agentic systems, we only **bold** the best performance.

Dataset	Method	PSNR↑	SSIM↑	LPIPS↓	DISTS↓	FID↓	NIQE↓	CLIPIQA↑	MUSIQ↑	MANIQA↑
Set5	***Fidelity-based method***									
	SwinIR [119]	32.92	0.9044	0.1669	0.1567	57.37	7.24	0.6179	59.98	0.6095
	X-Restormer [28]	33.15	0.9057	0.1636	0.1564	60.24	*7.07*	*0.6368*	60.09	0.6169
	DRCT [68]	33.26	0.9067	0.1616	*0.1526*	**52.25**	6.94	**0.6406**	*60.21*	0.6100
	HAT-L [29]	*33.29*	0.9082	**0.1582**	0.1542	56.95	7.11	0.6389	**60.44**	0.6212
	HMA [34]	**33.39**	**0.9089**	0.1587	**0.1535**	54.61	7.11	0.6338	60.39	**0.6241**
	4KAgent (ExpSR-s4-F)	33.34	*0.9081*	*0.1589*	0.1549	*56.62*	**6.90**	0.6294	60.02	*0.6177*
	Perception-based method									
	SwinIR (Real-ISR) [119]	**28.48**	**0.8446**	0.1632	0.1590	63.58	7.46	0.7072	62.43	0.6153
	DiffBIR [123]	26.41	0.7510	0.2059	0.1888	72.79	6.06	**0.8405**	**70.23**	*0.6767*
	OSEDiff [216]	26.21	*0.8063*	0.1583	*0.1647*	*67.50*	**5.78**	0.7973	68.76	0.6698
	PiSA-SR [179]	27.56	0.8189	**0.1318**	**0.1516**	**62.94**	*5.87*	*0.8086*	*69.87*	**0.6904**
	4KAgent (ExpSR-s4-P)	*26.88*	0.7899	*0.1591*	0.1657	70.63	5.79	0.8245	69.93	0.6808
	Agentic System									
	AgenticIR [274]	23.68	0.6711	0.2737	0.2190	124.96	**6.59**	**0.7750**	**71.88**	**0.7079**
	4KAgent (GenSR-s4-P)	**26.25**	**0.7672**	**0.1785**	**0.1836**	**89.02**	6.72	0.7396	70.39	0.6811
Set14	***Fidelity-based method***									
	SwinIR [119]	29.09	0.7950	0.2671	0.1574	70.49	6.19	0.5252	63.10	0.5891
	X-Restormer [28]	29.16	0.7963	0.2659	0.1557	69.86	6.22	0.5332	62.91	0.5925
	DRCT [68]	**29.57**	*0.8009*	0.2617	*0.1524*	*67.84*	6.09	**0.5362**	*63.12*	0.5932
	HAT-L [29]	*29.46*	0.8014	**0.2565**	0.1516	**66.61**	*6.11*	0.5267	63.23	0.5986
	HMA [34]	29.51	**0.8019**	0.2567	**0.1510**	69.41	6.25	0.5278	63.00	**0.6012**
	4KAgent (ExpSR-s4-F)	29.43	0.7989	*0.2593*	0.1528	67.83	**5.95**	*0.5315*	**63.45**	*0.5970*
	Perception-based method									
	SwinIR (Real-ISR) [119]	**25.91**	**0.7187**	*0.2244*	*0.1508*	96.19	4.45	0.6506	66.82	0.6054
	DiffBIR [123]	24.73	0.6349	0.2338	0.1545	*100.51*	*4.34*	*0.7553*	**72.97**	*0.6869*
	OSEDiff [216]	24.30	*0.6663*	0.2389	0.1524	101.03	4.61	0.7264	70.02	0.6674
	PiSA-SR [179]	*24.76*	0.6716	**0.1993**	**0.1343**	**89.91**	4.16	0.7643	*71.81*	**0.7015**
	4KAgent (ExpSR-s4-P)	24.76	0.6471	0.2158	0.1467	101.99	**4.01**	**0.7740**	72.54	0.6956
	Agentic System									
	AgenticIR [274]	21.98	0.6064	0.2807	0.1812	129.29	4.58	0.7449	72.48	0.6804
	4KAgent (GenSR-s4-P)	**23.40**	**0.6340**	**0.2484**	**0.1749**	**125.29**	**4.29**	**0.7604**	**73.64**	**0.7061**
B100	***Fidelity-based method***									
	SwinIR [119]	27.92	0.7489	0.3548	0.2005	94.57	6.27	0.5373	57.71	0.5860
	X-Restormer [28]	27.99	0.7508	0.3521	0.1972	90.52	6.21	0.5427	57.91	0.5935
	DRCT [68]	28.10	0.7535	0.3480	**0.1947**	**87.76**	6.06	*0.5499*	*58.78*	0.5895
	HAT-L [29]	28.08	0.7547	**0.3440**	*0.1952*	89.52	6.20	0.5477	58.71	*0.5991*
	HMA [34]	**28.12**	**0.7559**	0.3442	0.1953	88.46	*6.17*	**0.5534**	59.11	**0.6043**
	4KAgent (ExpSR-s4-F)	*28.09*	*0.7540*	*0.3453*	0.1950	*88.89*	**6.02**	0.5516	**59.12**	0.5994
	Perception-based method									
	SwinIR (Real-ISR) [119]	**25.42**	**0.6711**	0.2500	0.1699	92.65	*4.00*	0.6322	62.78	0.6085
	DiffBIR [123]	*24.99*	0.6156	0.2719	0.1666	84.99	3.92	0.7483	68.23	*0.6750*
	OSEDiff [216]	24.35	*0.6495*	*0.2408*	*0.1634*	73.23	4.08	*0.7422*	68.54	0.6725
	PiSA-SR [179]	25.00	0.6520	**0.2111**	**0.1471**	**61.82**	4.04	0.7384	*68.47*	0.6829
	4KAgent (ExpSR-s4-P)	24.64	0.6294	0.2387	0.1606	*73.64*	**3.86**	**0.7546**	**69.42**	**0.6851**
	Agentic System									
	AgenticIR [274]	22.51	0.5853	0.3078	0.1907	102.92	4.08	**0.7474**	68.36	0.6752
	4KAgent (GenSR-s4-P)	**23.64**	**0.6246**	**0.2572**	**0.1702**	**78.80**	**3.93**	0.7354	**69.44**	**0.6844**

the-art perception-based methods, ranking among the top two across most metrics on all classical SR benchmarks. For comparison across agentic systems, 4KAgent outperforms AgenticIR in most metrics on classical SR benchmarks, especially on the Set14 and B100 datasets.

Qualitative Comparison For visual comparison, we select two leading fidelity-based methods (X-Restormer, HAT-L), two perception-based methods (SwinIR (Real-ISR), DiffBIR), as well as one agentic system (AgenticIR) as baselines. For 4KAgent, we present result images under **GenSR-s4-P** profile for a comprehensive comparison. Visual comparisons in Fig. 9 reveal that fidelity-based

Table 8: Quantitative comparison on classical image super-resolution benchmarks (Urban100 and Manga109). The top three results for each metric are marked in **bold**, underline, and *italic*. For Agentic systems, we only **bold** the best performance.

Dataset	Method	PSNR↑	SSIM↑	LPIPS↓	DISTS↓	FID↓	NIQE↓	CLIPIQA↑	MUSIQ↑	MANIQA↑
Urban100	***Fidelity-based method***									
	SwinIR [119]	27.45	0.8254	0.1840	0.1533	3.58	*5.50*	0.5003	70.00	0.6693
	X-Restormer [28]	27.64	0.8288	0.1805	0.1504	3.65	5.61	0.4953	70.00	0.6746
	DRCT [68]	**28.78**	*0.8492*	0.1623	**0.1388**	2.92	5.45	**0.5271**	70.48	0.6778
	HAT-L [29]	28.58	0.8495	0.1598	0.1411	**2.87**	5.55	0.5054	*70.62*	0.6866
	HMA [34]	28.69	**0.8511**	**0.1583**	*0.1405*	*2.93*	5.61	*0.5084*	**70.75**	**0.6893**
	4KAgent (ExpSR-s4-F)	*28.59*	0.8479	*0.1599*	0.1399	2.97	**5.31**	0.5235	70.70	*0.6833*
	Perception-based method									
	SwinIR (Real-ISR) [119]	**23.24**	**0.7184**	0.1908	0.1365	**25.36**	**4.29**	0.6169	71.99	0.6578
	DiffBIR [123]	*22.51*	0.6397	0.2011	0.1395	*26.10*	4.79	**0.7185**	73.10	*0.6956*
	OSEDiff [216]	21.88	0.6572	0.2185	0.1479	38.13	4.67	0.6593	72.35	0.6822
	PiSA-SR [179]	22.36	0.6704	**0.1823**	**0.1297**	28.51	4.43	*0.6814*	*72.93*	**0.7020**
	4KAgent (ExpSR-s4-P)	22.56	*0.6582*	*0.1955*	*0.1378*	25.55	*4.53*	0.7092	**73.65**	0.6981
	Agentic System									
	AgenticIR [274]	22.03	**0.6615**	0.2147	0.1507	**31.09**	4.65	0.6790	73.10	0.6873
	4KAgent (GenSR-s4-P)	**22.27**	0.6545	**0.2073**	**0.1444**	32.29	**4.43**	**0.7001**	**73.57**	**0.6961**
Manga109	***Fidelity-based method***									
	SwinIR [119]	32.05	0.9260	0.0926	0.0761	1.88	5.32	**0.6385**	**70.32**	0.6117
	X-Restormer [28]	32.40	0.9279	0.0909	0.0748	1.88	5.48	0.6325	70.05	0.6123
	DRCT [68]	32.84	0.9307	0.0889	0.0685	1.49	5.08	0.6362	69.77	0.6087
	HAT-L [29]	33.08	0.9334	0.0845	*0.0684*	*1.48*	5.26	0.6160	69.76	0.6145
	HMA [34]	**33.20**	**0.9344**	**0.0835**	**0.0682**	**1.47**	*5.24*	0.6208	69.92	**0.6196**
	4KAgent (ExpSR-s4-F)	*32.87*	*0.9316*	*0.0860*	0.0683	1.48	**4.95**	*0.6329*	*69.99*	*0.6125*
	Perception-based method									
	SwinIR (Real-ISR) [119]	**26.29**	**0.8553**	**0.1367**	**0.0948**	**24.59**	**4.30**	0.6316	70.28	0.5868
	DiffBIR [123]	23.57	0.7297	0.1923	0.1275	30.11	4.55	**0.7804**	*74.51*	0.6787
	OSEDiff [216]	23.74	*0.7980*	*0.1703*	*0.1181*	41.54	4.78	0.6874	72.51	0.6538
	PiSA-SR [179]	24.02	0.8119	0.1450	0.1161	34.11	*4.35*	*0.7277*	74.76	*0.6779*
	4KAgent (ExpSR-s4-P)	*23.76*	0.7615	0.1776	0.1231	*33.36*	4.32	0.7678	**75.08**	**0.6801**
	Agentic System									
	AgenticIR [274]	**23.70**	0.7550	0.1862	**0.1246**	**34.01**	4.38	0.7450	73.98	0.6597
	4KAgent (GenSR-s4-P)	23.12	**0.7556**	**0.1834**	0.1264	34.58	**4.23**	**0.7652**	**75.02**	**0.6797**

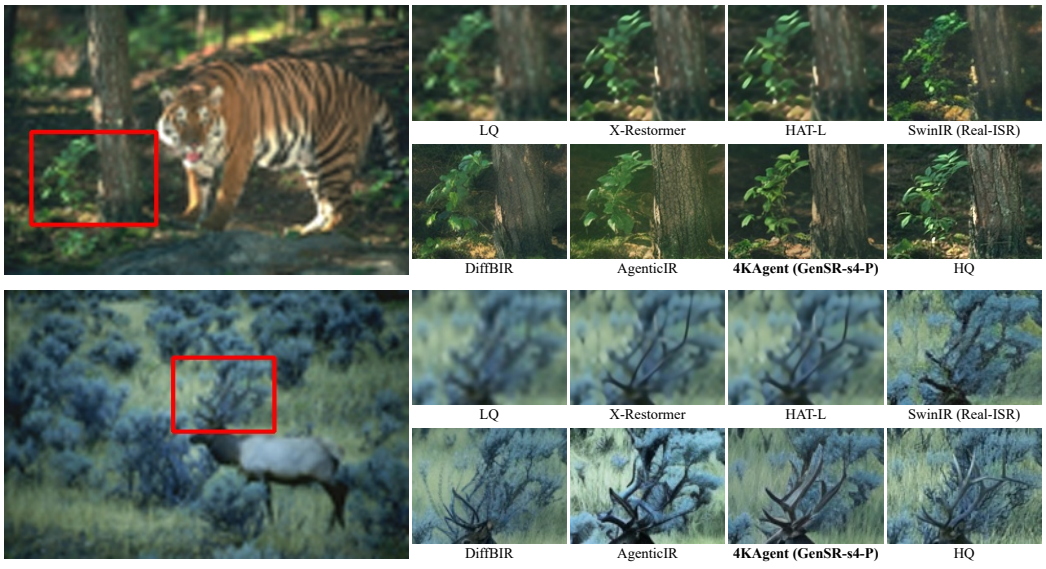


Figure 9: Visual comparisons on the classical image SR task (please zoom in to see details).

methods tend to produce overly smooth or blurred details (*e.g.*, HAT-L), even when trained under real-world image SR settings (*e.g.*, SwinIR (Real-ISR)), which is visually unpleasant. The diffusion-based

method DiffBIR generates rich but unrealistic details. AgenticIR performs well in detail generation but still lacks realism and exhibits noticeable color shifts. 4KAgent delivers both richer and more accurate details than these methods. For instance, it faithfully reproduces the fine stripes on tree bark in the top row and the intricate structure of antlers in the bottom row.

Discussions In the context of classical image super-resolution (SR), fidelity-based methods prioritize reconstruction accuracy, measured by PSNR and SSIM, resulting in outputs that often appear overly smooth or blurred. In contrast, perception-based methods optimize for high perceptual quality, reflected in metrics like NIQE, CLIPIQA, MUSIQ, and MANIQA, though often at the expense of fidelity. For example, diffusion-based approaches (*e.g.*, DiffBIR) may hallucinate rich but unrealistic textures. AgenticIR, while capable of generating sharper details, sometimes introduces color shifts or artifacts that undermine visual plausibility. 4KAgent offers configurable flexibility through its profile system, allowing it to operate either as a fidelity-based system (**ExpSR-s4-F**) or as a perception-based system (**ExpSR-s4-P**). Quantitatively, 4KAgent delivers competitive PSNR and SSIM scores under the fidelity-based profile, and achieves leading performance in perceptual metrics (NIQE, CLIPIQA, MUSIQ, MANIQA) under the perception-based profile. Qualitatively, 4KAgent consistently produces images with rich, realistic details. The flexibility of 4KAgent allows it to strike a superior balance: it can be easily tuned for maximum visual fidelity or for maximum perceptual appeal without extra training or adaptation, which avoids the common drawbacks of existing SR systems.

C.2 Real-World Image Super-Resolution

Settings In this section, we provide a detailed analysis of 4KAgent in the real-world image super-resolution task by presenting detailed experiment results on both the RealSR and DRealSR datasets, as well as visual comparisons.

Quantitative Comparison Experiment results on real-world image super-resolution datasets are shown in Tab. 9. For this task, we focus more on perceptual metrics, such as NIQE, CLIPIQA, MUSIQ, and MANIQA. Real-world image SR methods have achieved promising results on these metrics. AgenticIR, which contains the DiffBIR in its toolbox, outperforms DiffBIR in most perceptual metrics, indicating that agentic systems have better potential in solving the real-world SR problem. 4KAgent goes a step further and outperforms AgenticIR in most metrics, achieving better perceptual quality with better fidelity (*e.g.*, PSNR, SSIM), regardless of profile setting. In addition, 4KAgent sets a new state-of-the-art performance on perceptual metrics (*e.g.*, NIQE, MUSIQ).

Table 9: Quantitative comparison on real-world image super-resolution benchmarks (RealSR and DrealSR). The top three results for each metric are marked in **bold**, underline, and *italic*.

Dataset	Method	PSNR↑	SSIM↑	LPIPS↓	DISTS↓	FID↓	NIQE↓	CLIPIQA↑	MUSIQ↑	MANIQA↑
RealSR	ResShift [245]	**26.31**	0.7411	0.3489	0.2498	142.81	7.27	0.5450	58.10	0.5305
	StableSR [195]	24.69	0.7052	0.3091	*0.2167*	127.20	5.76	0.6195	65.42	0.6211
	DiffBIR [123]	24.88	0.6673	0.3567	0.2290	124.56	5.63	0.6412	64.66	0.6231
	PASD [234]	25.22	0.6809	0.3392	0.2259	**123.08**	*5.18*	0.6502	68.74	0.6461
	SeeSR [217]	25.33	0.7273	*0.2985*	0.2213	125.66	5.38	0.6594	69.37	0.6439
	SinSR [205]	26.30	*0.7354*	0.3212	0.2346	137.05	6.31	0.6204	60.41	0.5389
	OSEDiff [216]	25.15	0.7341	0.2921	0.2128	123.50	5.65	*0.6693*	69.09	0.6339
	PiSA-SR [179]	*25.50*	**0.7417**	**0.2672**	**0.2044**	*124.09*	5.50	0.6702	*70.15*	*0.6560*
	AgenticIR [274]	22.45	0.6447	0.3745	0.2503	140.38	5.81	0.6506	65.87	0.6210
	4KAgent (ExpSR-s4-P)	24.60	0.6839	0.3253	0.2292	127.64	5.09	**0.7078**	70.97	**0.6602**
	4KAgent (GenSR-s4-P)	22.55	0.6557	0.3509	0.2468	134.63	**4.78**	0.6666	**71.77**	0.6564
DrealSR	ResShift [245]	**28.45**	0.7632	0.4073	0.2700	175.92	8.28	0.5259	49.86	0.4573
	StableSR [195]	28.04	0.7460	0.3354	*0.2287*	*147.03*	6.51	0.6171	58.50	0.5602
	DiffBIR [123]	26.84	0.6660	0.4446	0.2706	167.38	6.02	0.6292	60.68	0.5902
	PASD [234]	27.48	0.7051	0.3854	0.2535	157.36	*5.57*	0.6714	64.55	0.6130
	SeeSR [217]	28.26	*0.7698*	*0.3197*	0.2306	149.86	6.52	0.6672	64.84	0.6026
	SinSR [205]	28.41	0.7495	0.3741	0.2488	177.05	7.02	0.6367	55.34	0.4898
	OSEDiff [216]	27.92	**0.7835**	0.2968	**0.2165**	135.29	6.49	0.6963	64.65	0.5899
	PiSA-SR [179]	*28.31*	0.7804	**0.2960**	0.2169	**130.61**	6.20	*0.6970*	*66.11*	*0.6156*
	AgenticIR [274]	23.06	0.6145	0.4775	0.2973	182.02	6.11	0.6542	63.59	0.5927
	4KAgent (ExpSR-s4-P)	26.00	0.6535	0.4257	0.2717	170.19	5.51	**0.7167**	67.72	**0.6397**
	4KAgent (GenSR-s4-P)	23.11	0.6126	0.4579	0.2866	178.36	**4.65**	0.7092	**69.30**	0.6219

Qualitative Comparison For visual comparison, we select four representative real-world image super-resolution methods (StableSR, DiffBIR, SinSR, OSEDiff) as well as one agentic system (AgenticIR) as baselines. The visual results are presented in Fig. 10. While these methods are able to recover rich details from the LQ image, their results often lack realism and visual fidelity. For example, in the top row, OSEDiff reconstructs clothing that appears more like jackets, whereas the HQ reference image shows down jackets. 4KAgent produces sharper and more realistic details, such as the texture of the down jacket in the top row and the clarity of the number '27' in the bottom row.

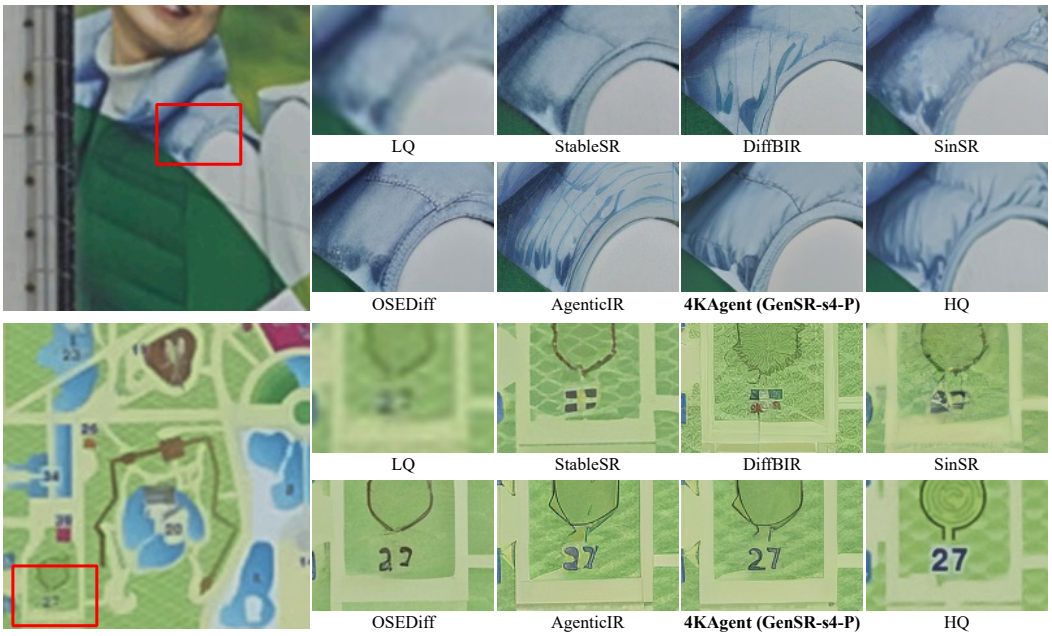


Figure 10: Visual comparisons on the real-world image SR task (please zoom in to see details).

Discussions The real-world image super-resolution task is substantially more challenging than the classical image super-resolution task, as it involves complex and unknown degradations beyond synthetic downsampling processes, which can also be seen from the comparison of the LQ image and HQ image in the dataset. Under this challenging setting, agentic systems prove their advantage by analyzing the distortion and restoring the image properly. 4KAgent further proves its superiority by consistently outperforming AgenticIR in most quantitative metrics. In particular, 4KAgent sets a new state-of-the-art performance for no-reference perceptual metrics, demonstrating that its design effectively elevates perceived realism. Qualitatively, these gains translate into visibly sharper and more believable details. By dynamically leveraging multiple SR experts and selecting the optimal result, 4KAgent shows its superiority in the challenging real-world image super-resolution task.

C.3 Multiple-Degradation Image Restoration

Settings In this section, we present detailed experimental results on Group A, B, and C test sets from the MiO100 dataset [91] with more visual comparisons.

Quantitative Comparison Experimental results are shown in Tab. 10. In the multiple-degradation image restoration (IR) task, agentic systems once again prove their superiority, outperforming all-in-one methods in all metrics. Among the agentic systems, 4KAgent performs the best, achieving a new state-of-the-art performance on PSNR, MANIQA, CLIPIQA, and MUSIQ. Specifically, for no-reference perceptual metrics (MANIQA, CLIPIQA, MUSIQ), 4KAgent outperforms all compared methods by a noticeable margin (*e.g.*, 4.2 lead in MUSIQ on Group C). For SSIM and LPIPS metrics, 4KAgent remains competitive, ranking within the top two on Group A and Group C subsets.

Qualitative Comparison For visual comparison, we select two leading all-in-one methods (DA-CLIP, AutoDIR) as well as an agentic system (AgenticIR) as baselines. Visual comparisons are shown

Table 10: Quantitative comparison of multiple-degradation image restoration tasks on three subsets (Group A, B, and C) from the MiO100 dataset. The top three results for each metric are marked in **bold**, underline, and *italic*.

Degradations	Method	PSNR↑	SSIM↑	LPIPS↓	MANIQA↑	CLIPIQA↑	MUSIQ↑
Group A	AirNet [107]	19.13	0.6019	0.4283	0.2581	0.3930	42.46
	PromptIR [162]	20.06	0.6088	0.4127	0.2633	0.4013	42.62
	MiOIR [90]	20.84	0.6558	0.3715	0.2451	0.3933	47.82
	DA-CLIP [136]	19.58	0.6032	0.4266	0.2418	0.4139	42.51
	InstructIR [36]	18.03	0.5751	0.4429	0.2660	0.3528	45.77
	AutoDIR [81]	19.64	0.6286	0.3967	0.2500	0.3767	47.01
	AgenticIR [274]	21.04	**0.6818**	*0.3148*	*0.3071*	*0.4474*	*56.88*
	MAIR [80]	*21.02*	*0.6715*	**0.2963**	0.3330	0.4751	59.19
	4KAgent (GenMIR-P)	**21.48**	0.6720	0.3019	**0.3748**	**0.5544**	**63.19**
Group B	AirNet [107]	19.31	0.6567	0.3670	0.2882	0.4274	47.88
	PromptIR [162]	20.47	0.6704	0.3370	0.2893	0.4289	48.10
	MiOIR [90]	*20.56*	*0.6905*	0.3243	0.2638	0.4330	51.87
	DA-CLIP [136]	18.56	0.5946	0.4405	0.2435	0.4154	43.70
	InstructIR [36]	18.34	0.6235	0.4072	0.3022	0.3790	50.94
	AutoDIR [81]	19.90	0.6643	0.3542	0.2534	0.3986	49.64
	AgenticIR [274]	20.55	**0.7009**	*0.3072*	*0.3204*	*0.4648*	*57.57*
	MAIR [80]	20.92	0.7004	**0.2788**	0.3544	0.5084	60.98
	4KAgent (GenMIR-P)	**20.95**	0.6727	0.3017	**0.3734**	**0.5505**	**62.69**
Group C	AirNet [107]	17.95	0.5145	0.5782	0.1854	0.3113	30.12
	PromptIR [162]	18.51	0.5166	0.5756	0.1906	0.3104	29.71
	MiOIR [90]	15.63	0.4896	0.5376	0.1717	0.2891	37.95
	DA-CLIP [136]	18.53	0.5320	0.5335	0.1916	0.3476	33.87
	InstructIR [36]	17.09	0.5135	0.5582	0.1732	0.2537	33.69
	AutoDIR [81]	18.61	0.5443	0.5019	0.2045	0.2939	37.86
	AgenticIR [274]	*18.82*	*0.5474*	*0.4493*	*0.2698*	*0.3948*	*48.68*
	MAIR [80]	19.42	0.5544	**0.4142**	0.2798	0.4239	51.36
	4KAgent (GenMIR-P)	**19.77**	**0.5629**	0.4271	**0.3545**	**0.5233**	**55.56**

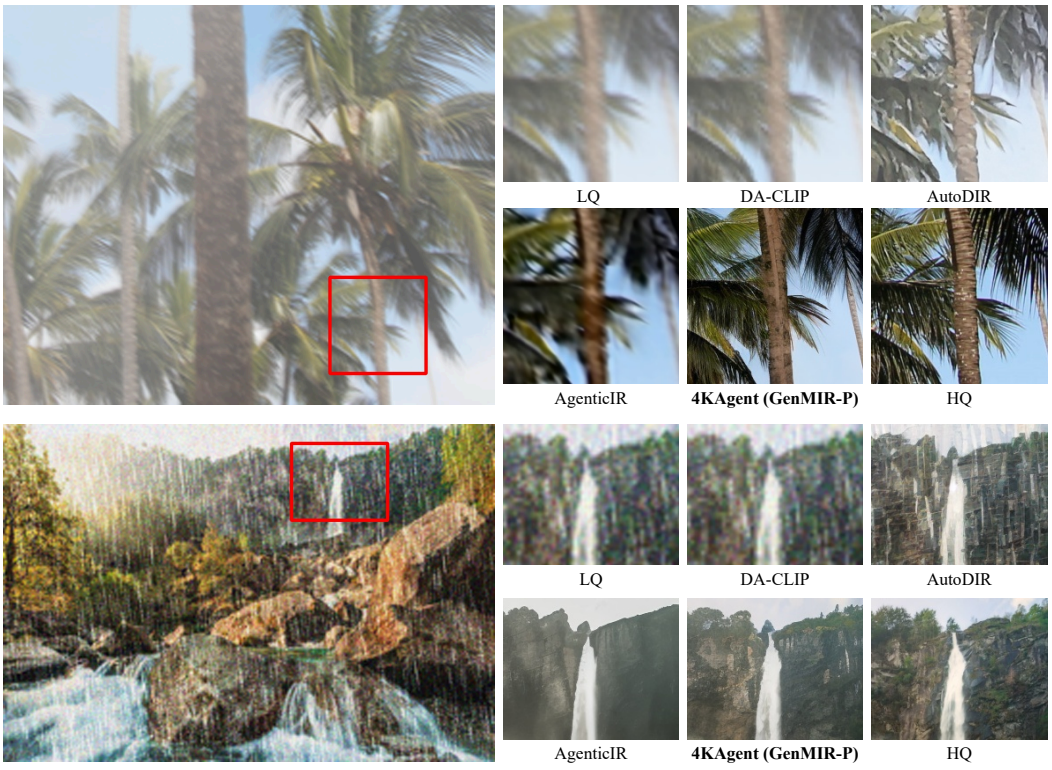


Figure 11: Additional visual comparisons on the MiO100 dataset (please zoom in to see details).

in Fig. 11. All-in-one methods exhibit limited performance in this setting, especially when restoring complex distortions, such as raindrops. AgenticIR achieves promising results, demonstrating the potential of agentic systems in dealing with complex distortion tasks. 4KAgent goes a step further by generating images with finer-grained details and better consistency with the high-quality (HQ) reference image. For instance, the natural stripes on the tree trunks and the fine leaf textures in the top row, as well as the intricate waterfall ripples and mountain contours in the bottom row.

Discussions Based on the experimental results, agentic systems have shown their superiority in the multiple-degradation image restoration tasks, where each low-quality (LQ) image is affected by $2 \sim 3$ types of distortions. Under this challenging setting, 4KAgent exhibits a clear advantage in handling complex, multi-degraded inputs, outperforming both conventional all-in-one methods and previous agentic systems. Quantitatively, 4KAgent achieves state-of-the-art results across multiple metrics, including PSNR, MANIQA, CLIPIQA, and MUSIQ, highlighting its strong capability in enhancing both fidelity and perceptual quality. Qualitatively, this metric superiority translates into more faithful restoration of fine-grained patterns and textures, even under severe and heterogeneous distortions.

C.4 Face Restoration

Settings In this section, we evaluate 4KAgent on the real-world face restoration benchmark, the WebPhoto-test [200] dataset, which contains 407 low-quality face images collected from the Internet. As the face restoration pipeline is a module subsequent to the super-resolution task in 4KAgent, we first downsample the images by a factor of 4 to generate low-quality (LQ) images. In this experiment, we configure 4KAgent with the **GenSRFR-s4-P** profile.

We compare 4KAgent with state-of-the-art face restoration methods, including CodeFormer [271], GFPGAN [200], and DifFace [244], as well as an agentic system AgenticIR [274]. For face restoration methods, we set the scaling factor to 4. As there are no high-quality (HQ) references, we evaluate performance with four no-reference perceptual metrics (NIQE, CLIPIQA, MUSIQ, and MANIQA-pipal) and two advanced face-specific IQA metrics (CLIB-FIQA [156] and DSL-FIQA [27]).

Quantitative Comparison Experimental results are shown in Tab. 11. AgenticIR performs inferior to prior face restoration methods in terms of no-reference perceptual metrics and face IQA metrics. 4KAgent outperforms AgenticIR on every metric by a clear margin (*e.g.*, 19.67 lead on MUSIQ). Moreover, 4KAgent achieves the best performance on no-reference perceptual metrics and delivers competitive results on face IQA metrics, ranking second on both CLIB-FIQA and DSL-FIQA.

Table 11: Quantitative comparison on face restoration benchmark (WebPhoto-Test). The top three results for each metric are marked in **bold**, underline, and *italic*.

Dataset	Method	NIQE↓	CLIPIQA↑	MUSIQ↑	MANIQA↑	CLIB-FIQA↑	DSL-FIQA↑
WebPhoto-Test	GFPGAN [200]	5.12	*0.6792*	74.21	*0.6379*	*0.6590*	**0.7732**
	CodeFormer [271]	*4.58*	0.6884	*73.87*	0.6415	**0.6840**	*0.7435*
	DifFace [244]	4.20	0.5831	65.31	0.5891	0.6511	0.6189
	AgenticIR [274]	6.85	0.5731	56.25	0.5465	0.5978	0.5289
	4KAgent (GenSRFR-s4-P)	**4.15**	**0.7077**	**75.92**	**0.6576**	0.6671	0.7683

Qualitative Comparison Visual comparisons are shown in Fig. 12. Compared with other methods, 4KAgent demonstrates a clear advantage in restoring realistic facial details, such as fine hair strands and natural skin textures. Moreover, it achieves superior restoration performance in non-facial regions, such as the wall and leaves in the first row and the logo on the hat in the second row. By consistently delivering high-quality restoration in both facial and non-facial areas, 4KAgent produces more visually pleasing and perceptually balanced results overall.

Discussions In the face restoration scenario, 4KAgent effectively addresses both facial and contextual degradations. Quantitatively, 4KAgent achieves the best performance on general no-reference perceptual metrics and delivers competitive scores on face IQA metrics, demonstrating the superiority of its system design and face Q-MoE policy. Qualitatively, this translates into more natural and richly detailed facial features, such as individual hair strands and realistic skin texture, while also enhancing background elements, producing more visually pleasing outputs. Among agentic systems, AgenticIR

Figure 12: Visual comparisons on the face restoration task (please zoom in to see details).

applies a uniform processing pipeline without a dedicated face restoration module, which limits its performance in this scenario. Benefiting from the face restoration pipeline and profile design, 4KAgent can be tailored as a face restoration expert, achieving superior results. We further present how these two designs enhance the face restoration ability of 4KAgent in the ablation study.

D Experiment Part II: 16× Natural Image Super-Resolution

Experiment Part I (§C) demonstrates the flexibility and superiority of 4KAgent on general image restoration and super-resolution tasks. In this section, we evaluate 4KAgent on more challenging restoration tasks. First, we assess its performance on the 16× real-world image super-resolution task. Next, we evaluate 4KAgent on our proposed **DIV4K-50** dataset.

D.1 Large Scale Factor (16×) Image Super-Resolution

Settings In the main paper, we present visual comparisons between 4KAgent and state-of-the-art image super-resolution methods (*e.g.*, HAT-L, DiffBIR) and agentic systems (*e.g.*, AgenticIR) on the RealSRSet dataset under a large-scale factor (16×) upscaling setting. Here, we extend our experiment with detailed quantitative evaluations with more methods and additional qualitative examples. As there are no corresponding high-quality (HQ) reference images in the dataset, we evaluate the result images on four no-reference perceptual metrics: NIQE, CLIPIQA, MUSIQ, and MANIQA-pipal.

Quantitative Comparison Experimental results are shown in Tab. 12. As the largest scale factor of the pre-trained model in HAT-L is 4, we apply the 4×→ 4× setting for HAT-L for 16× upscaling. Fidelity-based method struggles to deliver satisfactory performance on perceptual metrics under this setting. Recent perception-based real-world image super-resolution methods perform well on these metrics, even with the 16× setting. For example, DiffBIR with 16× setting achieves the best NIQE and MANIQA scores. Among agentic systems, 4KAgent outperforms AgenticIR on every metric by a clear margin (*e.g.*, 6.13 lead on MUSIQ). In addition, 4KAgent achieves the best performance on MUSIQ and the second-best performance on NIQE. For CLIPIQA and MANIQA metrics, 4KAgent also delivers competitive performance, ranking among the top three across all methods.

Table 12: Quantitative comparison on RealSRSet dataset under 16× upscaling. The top three results for each metric are marked in **bold**, underline, and *italic*.

Dataset	Method	NIQE↓	CLIPIQA↑	MUSIQ↑	MANIQA↑
RealSRSet	HAT-L [29] (4×→ 4×)	10.59	0.3885	25.06	0.3060
	DiffBIR [123] (4×→ 4×)	*3.63*	0.7867	44.86	0.6076
	DiffBIR [123] (16×)	**2.80**	0.7583	47.54	**0.6099**
	OSEDiff [216] (4×→ 4×)	5.40	0.7665	48.42	0.5362
	OSEDiff [216] (16×)	4.66	0.6483	35.33	0.4581
	PiSA-SR [179] (4×→ 4×)	5.70	**0.7883**	*48.20*	0.5464
	PiSA-SR [179] (16×)	4.88	0.6384	35.90	0.4128
	AgenticIR [274]	4.86	0.6775	44.71	0.5236
	4KAgent (Gen4K-P)	3.53	*0.7794*	**50.84**	*0.5913*

Figure 13: Additional visual comparisons on the RealSRSet dataset (please zoom in to see details).

Qualitative Comparison For visual comparison, we select three representative methods to benchmark against 4KAgent: (1) HAT-L (4×→ 4×): As a representative fidelity-based method, we investigate its performance under a large-scale upscaling setting. (2) DiffBIR (16×): As shown in Tab. 12, DiffBIR with 16× setting achieves the best performance on NIQE and MANIQA. Therefore, we include it to assess its visual quality. (3) AgenticIR: Selected for agentic system comparison. Visual comparisons are shown in Fig. 13.

HAT-L (4×→ 4×) shows limited enhancement over the low-quality input, leading to notably blurred textures. DiffBIR (16×) produces visually rich but often unrealistic hallucinations, and in some cases even alters the semantic content of the scene (*e.g.*, the first row), which is visually unappealing. AgenticIR generates visually plausible results but lacks sufficient sharpness and fine-grained details. 4KAgent generates fine-grained and realistic details: the rock and grass textures in the first row, and the hair strands, eyebrow patterns, and naturally expressive eyes in the second and third rows are all much more faithfully restored with finer-grained details.

Discussions In the challenging 16× upscaling scenario, 4KAgent delivers competitive quantitative results alongside fine-grained and realistic qualitative results, compared to other methods. In addition, traditional fidelity-oriented methods such as HAT-L struggle to recover fine details and instead produce overly smoothed and blurred results. This highlights the limitation of fidelity-driven pipelines under extreme magnification levels. Therefore, for such high-scale upscaling tasks, it is essential to configure 4KAgent with a perception-oriented profile (*e.g.*, setting **Restore Option** to **Perception** in the profile module) to better prioritize realistic texture synthesis. As image resolution approaches 4K and beyond, existing no-reference perceptual metrics may become misaligned with human judgment of visual quality. This discrepancy underscores the need for developing new no-reference perceptual metrics specifically designed for ultra-high-resolution images.

D.2 Joint Restoration & 4K Upscaling

Settings In this section, we bring 4KAgent to the most challenging setting: Joint multiple image restoration and 4K upscaling. As there are no previous methods and datasets targeted at this setting, we propose a new evaluation dataset, **DIV4K-50**, constructed from the Aesthetic-4K dataset [254] to rigorously test end-to-end restoration and ultra-high-scale SR. In this experiment, we evaluate 4KAgent on **DIV4K-50** and we configure 4KAgent with the **Gen4K-P** profile. Comparing methods and experimental settings are the same as in Appendix D.1.

Quantitative Comparison Quantitative comparisons are shown in Tab. 13. Similar to the experiment results in Appendix D.1, real-world image super-resolution methods perform competitively on perceptual metrics under this challenging setting. For example, DiffBIR achieves the best score on NIQE and CLIPIQA metrics. For agentic systems, 4KAgent outperforms AgenticIR on every metric. Additionally, 4KAgent achieves the best performance on MUSIQ and MANIQA metrics, and the second-best performance on NIQE and CLIPIQA metrics.

Table 13: Quantitative comparison on DIV4K-50 dataset. The top three results for each metric are marked in **bold**, underline, and *italic*.

Dataset	Method	NIQE↓	CLIPIQA↑	MUSIQ↑	MANIQA↑
DIV4K-50	HAT-L [29] (4× → 4×)	11.86	0.4699	22.82	0.3270
	DiffBIR [123] (4× → 4×)	*3.36*	**0.7588**	37.17	0.5916
	DiffBIR [123] (16×)	**2.65**	0.7078	38.59	*0.5858*
	OSEDiff [216] (4× → 4×)	4.88	*0.7201*	39.88	0.5482
	OSEDiff [216] (16×)	8.37	0.5680	25.07	0.4210
	PiSA-SR [179] (4× → 4×)	5.01	0.7141	38.22	0.5364
	PiSA-SR [179] (16×)	9.30	0.5549	24.51	0.3861
	AgenticIR [274]	5.13	0.5614	*39.55*	0.4814
	4KAgent (Gen4K-P)	3.15	0.7585	**44.16**	**0.5928**

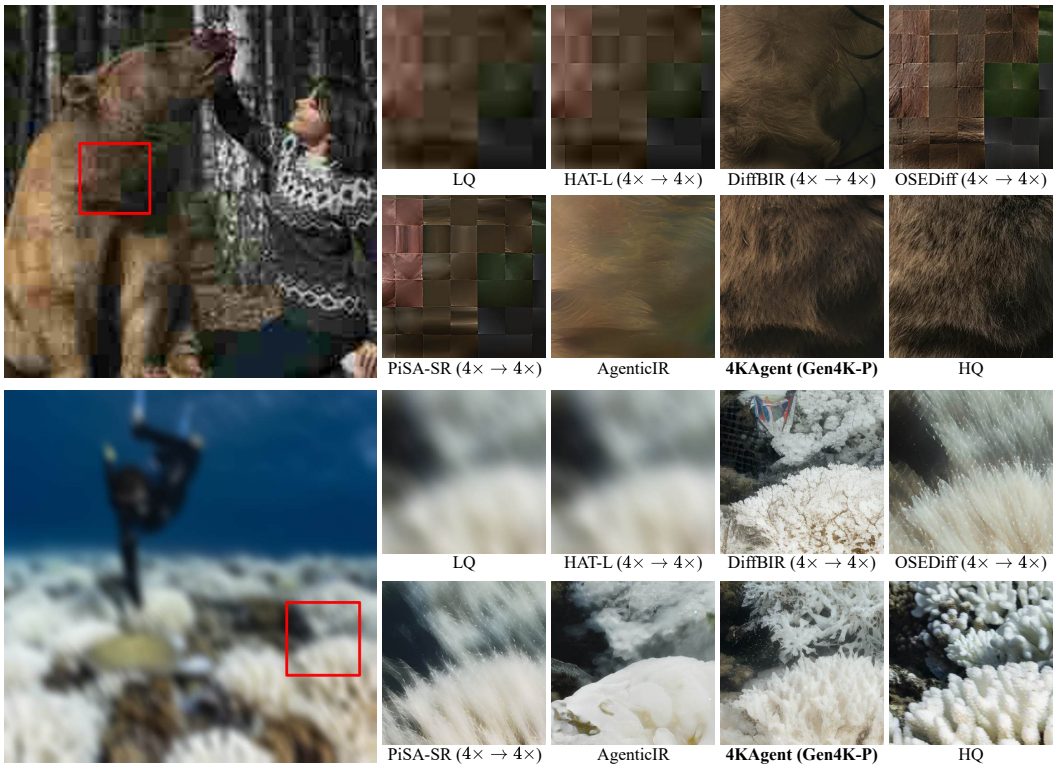


Figure 14: Additional visual comparisons on the DIV4K-50 dataset (please zoom in to see details).

Qualitative Comparison As shown in Appendix D.1, directly upscaling images with the 16× setting often produces visually rich but unrealistic artifacts. To ensure a fair and meaningful qualitative comparison in this experiment, we select previous methods with the 4×→ 4× setting, along with AgenticIR, as baselines. Qualitative comparisons are shown in Fig. 14. Real-world image super-resolution methods generally recover more details than fidelity-based method. However, their outputs still exhibit noticeable distortions. For example, the generated patches from OSEDiff and PiSA-SR in the middle row retain visible JPEG compression artifacts, which degrade the overall visual quality. While DiffBIR achieves the most favorable visual results among these methods, its outputs still suffer from either blurring or unrealistic artifacts. AgenticIR performs competitively but tends to produce insufficiently sharp details. 4KAgent consistently reconstructs finer and more natural details. Figure 14 presents additional visual comparisons: even in this challenging setting, 4KAgent faithfully reconstructs finer, more natural details, such as the bears fur in the top row and the intricate mountain textures in the bottom row, highlighting the superiority of our method.

Discussions The DIV4K-50 benchmark presents an extremely challenging setting, where fully recovering low-quality 256×256 inputs to match the 4096×4096 ground-truth images is virtually

unattainable. Therefore, our focus shifts towards generating richly detailed and visually authentic textures rather than exact pixel-wise reconstruction. Existing real-world image super-resolution methods struggle to fully correct the compounded degradations present in challenging scenarios. While these methods achieve competitive scores on perceptual metrics, their outputs tend to suffer from unrealistic hallucinated textures. Prior agentic systems, while effective at handling multiple degradations, show limitations in maintaining sufficient sharpness when upscaling to 4K resolutions. 4KAgent demonstrates its capability to simultaneously address multiple degradations and extreme scaling factors, effectively reconstructing natural, fine-grained details with high visual fidelity.

E Experiment Part III: AI-Generated Content (AIGC) 4K Super-Resolution

Text-to-visual models have ushered in a new era of high-quality image synthesis in AI-generated content. While existing models exhibit impressive capabilities in interpreting and following complex user instructions, their limitation to relatively low-resolution outputs (*e.g.*, 1024×1024) poses significant challenges for applications requiring ultra-high visual fidelity, such as digital content creation and cinematic production. Scaling diffusion models for high-resolution generation entails computational overhead and access to large-scale high-resolution training data. As a practical alternative, pre-trained diffusion models can be repurposed for ultra-high-resolution image generation.

E.1 AI-Generated Content 4K Super-Resolution Experiment

In this section, we present comprehensive experimental results comparing the effectiveness of end-to-end generation of ultra-high-resolution images with upscaling 1K images to 4K using our 4KAgent. The comparsions are conducted on two curated datasets: GenAIBench-4K and DiffusionDB-4K.

Settings We curated a set of 200 prompts sampled from two widely-used AIGC benchmarks [210, 103]. For each prompt, 1K-resolution images were generated using several representative text-to-image models, including Flux.1-dev [95], Stable Diffusion 3 (SD3) [49], PixArt-Σ [24], SANA [229], and GPT-4o [72]. In parallel, we employed native 4K-capable models such as SANA [229] and Diffusion-4K [254] to directly synthesize 4K-resolution outputs. Due to more stringent safety protocols, GPT-4o yielded only 39 valid 1K-resolution images from DiffusionDB prompts.

We use the **ExpSR-s4-P** profile in 4KAgent here. To assess perceptual quality, we employ no-reference perceptual metrics. However, we observe that these metrics, particularly MUSIQ, are not tailored for evaluating ultra-high-resolution images, likely due to their inability to capture fine-grained details through a multi-scale architecture. To mitigate this limitation, we introduce **MUSIQ-P**, a patch-applied variant that computes MUSIQ scores over non-overlapping 512×512 patches and averages them, thereby improving sensitivity to localized artifacts in ultra-high-resolution content.

Table 14: Comprehensive quantitative comparison of AIGC 4× Super-Resolution. The top three performances of each metric are marked in **bold**, underline, *italic* respectively. MUSIQ-P* indicates a patch-applied variant of the MUSIQ metric for evaluating ultra-high-resolution (4K) images.

Dataset	**GenAIBench-4K [103]**				**DiffusionDB-4K [210]**			
Model	**NIQE↓**	**CLIPIQA↑**	**MUSIQ-P*↑**	**MANIQA↑**	**NIQE↓**	**CLIPIQA↑**	**MUSIQ-P*↑**	**MANIQA↑**
SANA-4K [229]	4.02	0.6172	47.93	0.3673	3.74	0.6005	48.66	0.3425
Diffusion-4K [254]	6.38	0.5049	35.07	0.3535	6.55	0.5056	35.87	0.3404
SANA-1K [229]	4.18	0.7147	**66.30**	0.4814	3.80	0.6910	67.99	0.5104
+ 4KAgent	3.03	0.7050	57.97	0.4735	3.04	0.7082	60.48	0.4715
GPT4o [72]	5.69	0.6607	*64.43*	0.4997	5.13	0.6275	62.53	0.4398
+ 4KAgent	3.56	0.7016	58.28	0.4976	3.40	0.6867	56.67	0.4711
FLUX.1-dev [95]	6.18	0.6768	61.02	*0.5018*	5.33	**0.7509**	**69.69**	**0.5835**
+ 4KAgent	*2.98*	*0.7078*	58.19	0.5034	*3.04*	0.7440	60.88	*0.5056*
PixArt-Σ [24]	4.12	0.6960	63.74	0.4415	3.66	0.6892	*66.54*	0.4386
+ 4KAgent	**2.76**	0.7077	56.71	0.4699	**2.88**	*0.7092*	58.85	0.4659
SD3-Medium [49]	5.03	0.6922	64.68	0.4767	4.38	0.6667	65.99	0.4413
+ 4KAgent	2.99	**0.7169**	60.22	**0.5155**	2.99	0.7066	59.35	0.4747

Quantitative Comparison. Tab. 14 presents the quantitative results across three strategies: (1) native 4K generation, (2) 1K-resolution generation, and (3) 1K-resolution images upscaled by 4KAgent. On GenAIBench-4K, the SANA-1K + 4KAgent pipeline achieves a NIQE score of 3.03 and a CLIPIQA of 0.7050, significantly outperforming SANA-4K (NIQE 4.02, CLIPIQA 0.6172). Similarly, PixArt-Σ + 4KAgent obtains the best NIQE score (2.76) among all methods, while SD3-Medium + 4KAgent achieves the best CLIPIQA (0.7169) and MANIQA (0.5155) scores. On DiffusionDB-4K, several models such as SANA-1K, GPT-4o, PixArt-Σ, and SD3-Medium, when upscaled with 4KAgent, achieve significantly lower NIQE and higher CLIPIQA scores. Although MUSIQ-P scores for the upscaled images show a slight decrease relative to their 1K counterparts, the difference remains marginal, suggesting limited perceptual degradation during upscaling.

To further assess semantic and aesthetic fidelity, we report PickScore [88] in Tab. 15, which quantitatively captures diversity and human-aligned visual quality. On GenAIBench-4K, models enhanced with 4KAgent outperform their native 4K counterparts. On DiffusionDB-4K, the performance gap is smaller, which may be attributed to the dataset's richer and more descriptive prompt content.

Table 15: Comparison of PickScore-Based [88] Quantitative Evaluation Between Native 4K and 4KAgent-Upscaled 1K Models.

Dataset	Avg. Prompt Length	SANA-4K	SANA-1K + 4KAgent	Diffusion-4K	Flux.1-dev + 4KAgent
GenAIBench-4K [103]	12.13	0.4482	**0.5518**	0.2389	**0.7611**
DiffusionDB-4K [210]	25.29	0.4893	**0.5107**	0.2406	**0.7594**

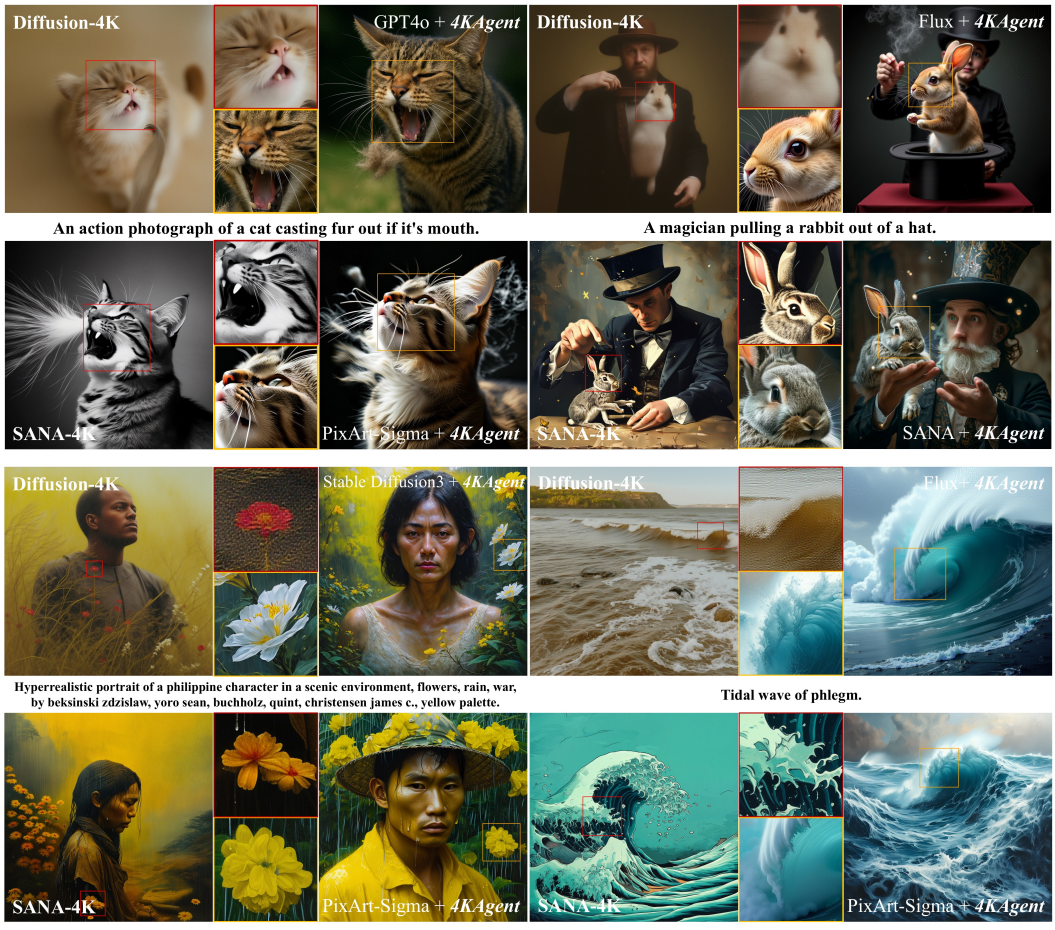


Figure 15: Visual comparison between native 4K image generation and 1K image generation methods with 4KAgent, using identical prompts. 4K images from Diffusion-4K and SANA-4K are displayed on the left, while the corresponding outputs enhanced by 4KAgent are shown on the right.

Qualitative Comparison Fig. 15 shows qualitative results of applying 4KAgent to various base models under identical prompts. Across different models, 4KAgent consistently enhances visual fidelity and preserves fine-grained details. As shown in Fig. 16, images upscaled from SANA-1K using 4KAgent exhibit richer textures and stronger aesthetic alignment than those generated natively at 4K resolution by SANA-4K.

SANA-4K **SANA-1K + 4KAgent**

Figure 16: Visual comparison of aesthetic preference alignment between SANA-4K and SANA-1K+4KAgent using identical prompts sampled from GenAIBench-4K. SANA-1K+4KAgent yields superior aesthetic alignment and richer high-resolution details, highlighted in the zoomed-in patches.

Discussions The application of our 4KAgent in AIGC scenarios leads to substantial improvements in image quality when upscaling 1K-resolution images to 4K. **First**, when applied to 1K-resolution inputs, 4KAgent consistently achieves notable gains across multiple quantitative benchmarks, enabling more detailed and accurate reconstructions in the resulting 4K outputs. As traditional metrics are not specifically tuned for ultra-high-resolution images, we adopted the adaptive MUSIQ-P metric to enable a perceptually-focused evaluation. The results indicate that 4K-upscaled images achieve perceptual quality scores comparable to their original 1K counterparts. **Second**, 4KAgent demonstrates strong capability in synthesizing high-fidelity visual details and intricate textures. However, as with many traditional super-resolution methods, this training-free framework occasionally introduces unintended bokeh-like artifacts, particularly in blurred background regions. Given 4KAgent's modular and scalable design, we believe integrating task-specific profile configurations and perceptual alignment strategies could further reduce artifacts and improve robustness in diverse AIGC applications.

F Experiment Part IV: Scientific Imagery Super-Resolution

F.1 Remote Sensing Image Super-Resolution

High-resolution satellite imagery is foundational for a wide spectrum of remote sensing tasks, including urban planning, environmental monitoring, and disaster response [64, 171]. However, due to cost, bandwidth, and sensing constraints, acquiring such high-resolution imagery globally at very high frequency remains very expensive or even impractical. Recent advances in deep learning-based super-resolution have provided a promising alternative by reconstructing high-fidelity imagery from lower-resolution observations [92]. In this section, we evaluate 4KAgent against state-of-the-art baselines on a diverse set of real-world satellite image super-resolution datasets.

Settings We evaluate our models on four benchmark datasets covering varied land-use patterns and sensing characteristics:

- AID [223] is a large-scale dataset constructed to benchmark aerial scene classification methods. It includes over 10,000 high-resolution aerial images across 30 scene categories, such as airports, industrial areas, and farmlands. Each image has a resolution of 600×600 with a spatial resolution

of 0.5-8 m/pixel. Images exhibit high intra-class diversity and low inter-class variation, making them well-suited for evaluating generalization in SR tasks.

- DIOR [111] is a comprehensive object detection benchmark in the remote sensing domain, containing 23,463 images and 192,472 annotated object instances across 20 categories. The resolution of images is 800×800, and the spatial resolutions range from 0.5m to 30m. These images exhibit high diversity in resolution, imaging conditions, and object scale.
- DOTA [222] consists of 2,806 ultra-high-resolution aerial images collected from various sensors. It features over 188,000 labeled object instances with arbitrary orientations. Each image is in resolution about 4000×4000. The combination of large-scale, fine-grained annotations and high inter-scene variability makes DOTA particularly valuable for evaluating perceptual fidelity.
- WorldStrat [37] is a unique dataset designed for real-world satellite image super-resolution tasks, with globally stratified land use coverage across 10,000 km^2. It pairs 1054×1054 pixel high-resolution (SPOT 6/7, 1.5 m/pixel) and temporally matched low-resolution (Sentinel-2, 10 m/pixel) imagery for thousands of regions worldwide. Importantly, unlike synthetic degradation benchmarks, WorldStrat contains real cross-sensor-captured low-resolution (LR) and high-resolution (HR) image pairs, introducing natural misalignment and color mismatches due to different sensor characteristics.

From each dataset, we select 100-200 representative scenes. Following [227, 228], for AID, DIOR, and DOTA, HR images are downsampled using bicubic interpolation to generate corresponding LR inputs. For WorldStrat, we adopt the datasets official pre-processing pipeline, selecting LR images that are temporally closest to each HR acquisition. Notably, due to the different sensors used for LR and HR captures, RGB content may exhibit significant variation, posing a realistic challenge for super-resolution. To test the generalization ability of our models, we evaluate on a spectrum of resolution scales: 1) 4× SR (128→512) on DIOR and DOTA datasets; 2) 4× SR (160→640) on AID and WorldStrat datasets; 3) 4× SR (512→2048) for high-res DOTA scenes; 4) 16× SR (e.g., 256→4096) for DOTA scenes.

We evaluate 4KAgent with the **AerSR-s4-F** profile and **AerSR-s4-P** profile in 4× super resolution for **Fidelity** and **Perception** preference, respectively. Then we evaluate 4KAgent with the **Aer4K-F** profile and **Aer4K-P** profile in 16× super resolution for **Fidelity** and **Perception** preference, respectively. We benchmark 4KAgent against the following categories of SR models: 1) Expert aerial SR models: HAUNet [196], TransENet [102]; 2) Fidelity-based SR models: HAT-L [29], PiSA-SR-PSNR [179], and SwinIR [119]; 3) Perception-based SR Models: DiffBIR [123], OSEDiff [216], HAT-GAN [29], PiSA-SR [179], and SwinIR (Real-ISR) [119]. For 'PiSA-SR-PSNR', we set the pixel guidance factor $\lambda_{pix} = 1.0$ and semantic guidance factor $\lambda_{sem} = 0$ for PiSA-SR [179] in inference. Additionally, we include AgenticIR for agentic system comparison. Together, these diverse datasets and models enable a comprehensive evaluation of 4KAgent in terms of both pixel fidelity and perceptual quality, across synthetic and real-world degradation settings.

Quantitative Comparison We report the comparison results on AID (4x SR, 160→640), DIOR (4× SR, 128→512), DOTA (4× SR, 128→512), WorldStrat (4× SR, 160→640), DOTA (4× SR, 512→2048), and DOTA (16× 4K SR) in Tabs. 16 to 21, respectively. Across all six benchmark settings, 4KAgent with **Fidelity** preference consistently demonstrates superior performance in terms of **pixel-level reconstruction**. It ranks within the top three PSNR in four out of six tasks, and ranks within the top three in SSIM across all synthetic scenarios. This confirms its ability to preserve structural details across scales and domains. Importantly, 4KAgent also consistently outperforms AgenticIR by a large margin across all fidelity metrics and tasks, highlighting its effectiveness.

In terms of perceptual quality, 4KAgent with **Perception** preference achieves top performance in **perceptual quality** assessment metrics across multiple datasets. Notably, on the WorldStrat, 4KAgent (AerSR-s4-P) ranks first in MUSIQ, MANIQA, and CLIPIQA. These results indicate that 4KAgent not only produces photorealistic outputs but also maintains robustness in real-world, sensor-misaligned scenarios. Compared to AgenticIR, 4KAgent with **Perception** preference shows clear gains across all perceptual IQA metrics, reaffirming the value of 4KAgent in balancing realism and structure across diverse and challenging remote sensing settings.

Qualitative Comparison Figs. 17 to 21 present a comprehensive visual comparison of all evaluated models across the tested datasets of 4× AID, 4× DIOR, 4× DOTA, 4× WorldStrat, and 16× DOTA. **Firstly**, 4KAgent with the perception preference consistently delivers superior perceptual quality on

Table 16: 4× performance comparison of evaluated models on the AID dataset (160→640). The top three results for each metric are marked in **bold**, underline, and *italic*.

DataSet	Model	PSNR↑	SSIM↑	LPIPS↓	DISTS↓	FID↓	NIQE↓	MUSIQ↑	MANIQA↑	CLIPIQA↑
Fidelity-based SR	SwinIR [119]	*28.4887*	**0.7422**	0.4355	0.2473	186.1843	7.4295	50.1222	0.3108	0.2563
	HAT-L [29]	27.1630	0.6835	0.4635	0.2330	126.9911	7.3586	36.4299	0.4149	0.4111
	PiSA-SR-PSNR [179]	27.9079	0.7251	0.4273	0.4273	144.3109	7.2371	41.6239	0.4024	0.1831
Perception-based SR	SwinIR (Real-ISR) [119]	26.5090	0.6700	**0.3344**	**0.1928**	129.3879	*3.8690*	60.6544	0.5641	0.5205
	HAT-GAN [29]	25.7860	0.6643	*0.3522*	*0.2137*	140.2364	4.8258	55.4862	0.5618	0.3556
	DiffBIR [123]	24.8343	0.5554	0.4466	0.2374	130.6386	4.8871	*65.9636*	*0.6342*	**0.7302**
	OSEDiff [216]	25.2220	0.6164	0.3497	0.3497	**91.5957**	3.6661	63.9855	0.6219	0.6074
	PiSA-SR [179]	24.5971	0.5903	0.3541	0.3541	115.8287	**3.4859**	66.0433	**0.6555**	0.6346
Expert Aerial SR	HAUNet [196]	28.5136	0.7146	0.4327	0.2083	*122.1656*	7.2497	35.5021	0.4162	0.1706
	TransENet [102]	28.0317	0.6983	0.4179	0.2109	125.9495	6.7700	35.1140	0.3776	0.1162
Agentic System	AgenticIR [274]	21.3431	0.5147	0.4600	0.2539	149.7191	4.5325	67.0257	0.6283	*0.6693*
	4KAgent (AerSR-s4-F)	**28.5481**	*0.7157*	0.4436	0.2263	127.4411	7.5916	37.5713	0.3774	0.4322
	4KAgent (AerSR-s4-P)	24.4212	0.5354	0.4696	0.2566	139.7775	4.8915	**68.1858**	0.6439	0.6897

Table 17: 4× performance comparison of evaluated models on the DIOR dataset (128→512). The top three results for each metric are marked in **bold**, underline, and *italic*.

Type	Model	PSNR↑	SSIM↑	LPIPS↓	DISTS↓	FID↓	NIQE↓	MUSIQ↑	MANIQA↑	CLIPIQA↑
Fidelity-based SR	SwinIR [119]	**27.8751**	**0.7257**	0.4474	0.2488	223.4653	7.1247	51.6481	0.3122	0.2441
	HAT-L [29]	*27.7355*	0.6962	0.4586	0.2195	134.4649	7.1142	37.1784	0.4151	0.4194
	PiSA-SR-PSNR [179]	27.4176	0.7118	0.4378	0.2202	167.7604	6.9455	42.5087	0.4095	0.1881
Perception-based SR	SwinIR (Real-ISR) [119]	26.4708	0.6698	**0.3391**	0.1983	144.3900	*3.8921*	60.6319	0.5552	0.5091
	HAT-GAN [29]	26.8015	0.6848	0.3398	*0.2073*	149.8121	4.8459	55.8046	0.5592	0.3392
	DiffBIR [123]	24.9254	0.5742	0.4201	0.2317	146.2642	4.9198	*66.4572*	*0.6315*	**0.7078**
	OSEDiff [216]	25.0470	0.6207	*0.3506*	**0.1875**	**127.7888**	3.6641	65.3934	0.6245	0.5976
	PiSA-SR [179]	24.4078	0.5932	0.3534	0.3534	*129.2724*	**3.5111**	67.6365	**0.6571**	0.6229
Expert Aerial SR	HAUNet [196]	27.8221	0.6992	0.4527	0.2100	128.8770	6.9586	35.5885	0.4089	0.1572
	TransENet [102]	27.3002	0.6824	0.4391	0.2113	129.4893	6.4986	34.7713	0.3750	0.0984
Agentic System	AgenticIR [274]	22.4811	0.5654	0.4668	0.2388	169.2341	4.7446	63.1399	0.5938	*0.6252*
	4KAgent (AerSR-s4-F)	27.6761	*0.7062*	0.4309	0.2250	146.5618	7.2555	37.5543	0.3811	0.4368
	4KAgent (AerSR-s4-P)	24.4893	0.5795	0.4374	0.2471	160.6006	4.6522	**68.0117**	0.6358	0.6456

Table 18: 4× performance comparison of evaluated models on the DOTA dataset (128→512). The top three results for each metric are marked in **bold**, underline, and *italic*.

Type	Model	PSNR↑	SSIM↑	LPIPS↓	DISTS↓	FID↓	NIQE↓	MUSIQ↑	MANIQA↑	CLIPIQA↑
Fidelity-based SR	HAT-L [29]	**33.0720**	**0.8656**	**0.2448**	0.1471	58.0105	6.6527	51.7547	0.5858	0.3725
	PiSA-SR-PSNR [179]	28.9623	0.7999	0.3415	0.2093	133.7664	7.7350	47.4847	0.4878	0.3094
	SwinIR [119]	30.5969	0.8254	0.3275	0.2215	143.2866	7.5391	54.3111	0.4526	0.2700
Perception-based SR	DiffBIR [123]	25.8326	0.6489	0.3724	0.2340	115.1440	6.0906	64.9539	*0.6535*	*0.6772*
	OSEDiff [216]	26.3616	0.7156	0.3324	0.2133	126.4670	*5.4257*	64.1220	0.6278	0.6736
	HAT-GAN [29]	28.6557	0.7869	0.2751	0.1818	115.1743	5.7245	57.0159	0.5929	0.3691
	PiSA-SR [179]	25.8447	0.6921	0.3220	0.2081	112.1042	4.9062	66.3901	0.6676	**0.6855**
	SwinIR (Real-ISR) [119]	28.9000	0.7883	*0.2657*	0.1769	110.4552	**4.7712**	59.7489	0.5886	0.4519
Expert Aerial SR	HAUNet [196]	32.8286	0.8627	0.2480	**0.1428**	**57.3008**	6.5917	50.7492	0.5711	0.3824
	TransENet [102]	30.7214	0.8176	0.2883	*0.1553*	*68.5120*	6.2878	42.8957	0.4856	0.3441
Agentic System	AgenticIR [274]	19.9655	0.5973	0.4227	0.2620	137.2777	6.3126	*65.5596*	0.6375	0.6198
	4KAgent (AerSR-s4-F)	*31.3589*	*0.8478*	0.2853	0.1776	88.0366	7.0808	50.6815	0.5515	0.3799
	4KAgent (AerSR-s4-P)	24.9224	0.6427	0.3884	0.2555	131.0346	6.1609	**67.0355**	**0.6701**	0.6800

low-resolution SR datasets, as demonstrated in Figs. 17 to 20. In contrast, 4KAgent with the fidelity preference excels on high-resolution 4K SR datasets, producing the most faithful reconstructions in Fig. 21. **Secondly**, 4KAgent exhibits a clear advantage in reconstructing fine structures such as lines and patterns, as evident in Figs. 18 and 19. **Finally**, in the challenging cross-sensor super-resolution scenario of WorldStrat, where LR and HR images originate from different sensors. 4KAgent with

Table 19: 4× performance comparison of evaluated models on the WorldStrat dataset (160→640). The top three results for each metric are marked in **bold**, underline, and *italic*.

Type	Model	PSNR↑	SSIM↑	LPIPS↓	DISTS↓	FID↓	NIQE↓	MUSIQ↑	MANIQA↑	CLIPIQA↑
Fidelity-based SR	HAT-L [29]	21.2238	0.6480	0.3468	0.2108	145.3798	9.0026	30.6655	0.2844	0.2339
	PiSA-SR-PSNR [179]	24.4312	**0.7271**	0.3312	0.2283	142.8870	10.5327	29.6462	0.2994	0.2509
	SwinIR [119]	18.0937	0.6136	0.3451	0.2279	180.4954	8.2607	27.1954	0.2104	0.2355
Perception-based SR	DiffBIR [123]	20.6485	0.5150	0.6781	0.3712	227.6764	7.5514	*53.5666*	0.5475	0.6075
	OSEDiff [216]	25.9716	0.6316	0.4460	0.2562	176.2589	8.6342	46.5092	0.4988	0.5096
	HAT-GAN [29]	27.0796	0.7241	0.3199	0.1978	*137.4441*	9.8474	30.3741	0.3587	0.2623
	PiSA-SR [179]	23.9304	0.6179	0.4581	0.2748	170.0426	**7.2214**	48.6414	0.5152	0.5010
	SwinIR (Real-ISR) [119]	**27.9062**	0.7120	0.3473	*0.2074*	149.4428	10.1237	32.9484	0.3497	0.3060
Expert Aerial SR	HAUNet [196]	*26.1895*	*0.7143*	**0.3141**	**0.1976**	**128.2747**	10.6318	28.4401	0.3129	0.2701
	TransENet [102]	24.4879	0.6943	*0.3270*	0.2106	133.6959	*7.7765*	27.8965	0.3152	0.2246
Agentic System	AgenticIR [274]	19.5883	0.5188	0.6716	0.3686	224.8042	8.7079	54.0649	*0.5166*	*0.5402*
	4KAgent (AerSR-s4-F)	22.3529	0.6470	0.3702	0.2324	166.2731	9.5364	34.3698	0.3011	0.2875
	4KAgent (AerSR-s4-P)	20.1510	0.5379	0.6363	0.3664	223.4866	8.5528	**56.8421**	**0.5547**	**0.6236**

Table 20: 4× performance comparison of evaluated models on the DOTA dataset (512→2048). The top three results for each metric are marked in **bold**, underline, and *italic*.

Type	Model	PSNR↑	SSIM↑	LPIPS↓	DISTS↓	FID↓	NIQE↓	MUSIQ↑	MANIQA↑	CLIPIQA↑
Fidelity-based SR	HAT-L [29]	**38.4856**	**0.9101**	**0.1956**	0.1007	0.2722	6.8076	39.5562	0.5006	0.3408
	PiSA-SR-PSNR [179]	31.5336	0.8519	0.3183	0.1797	40.7879	7.0981	38.6277	0.4226	0.2444
	SwinIR [119]	33.7463	0.8759	0.2990	0.1809	15.6964	6.7591	43.3724	0.4200	0.2847
Perception-based SR	DiffBIR [123]	26.3032	0.6490	0.4035	0.1953	76.3522	*4.0360*	**57.2133**	**0.6265**	**0.7306**
	OSEDiff [216]	28.5768	0.7567	0.3632	0.2127	70.4222	4.1286	51.4468	0.5871	0.6907
	HAT-GAN [29]	31.2735	0.8408	0.2720	0.1527	47.0606	4.6791	47.6011	0.5508	0.3551
	PiSA-SR [179]	27.6765	0.7303	0.3499	0.2175	54.5844	**3.8036**	*51.7604*	*0.6135*	0.6353
	SwinIR (Real-ISR) [119]	31.6933	0.8423	0.2564	0.1453	37.3714	4.1055	48.3376	0.5561	0.4038
Expert Aerial SR	HAUNet [196]	38.2237	0.9075	0.2002	**0.0974**	*0.2984*	6.6907	38.8776	0.4926	0.3471
	TransENet [102]	35.9776	0.8824	0.2431	*0.1267*	**0.2137**	6.3886	34.8775	0.4345	0.2942
Agentic System	AgenticIR [274]	21.4719	0.7284	0.4157	0.2167	77.7286	4.7902	49.5913	0.5496	0.4853
	4KAgent (AerSR-s4-F)	*36.7655*	*0.9017*	*0.2343*	0.1283	11.0017	7.0361	38.7286	0.4738	0.3493
	4KAgent (AerSR-s4-P)	28.4281	0.7513	0.3440	0.2181	41.1425	3.9267	52.1735	0.6264	*0.6608*

Table 21: 16× performance comparison of evaluated models on the DOTA dataset (4K resolution). The top three results for each metric are marked in **bold**, underline, and *italic*.

Type	Model	PSNR↑	SSIM↑	LPIPS↓	DISTS↓	FID↓	NIQE↓	MUSIQ↑	MANIQA↑	CLIPIQA↑
Fidelity-based SR	HAT-L [29]	**23.9586**	**0.6362**	0.6471	0.3219	**82.7644**	9.0807	31.5394	0.2565	0.3062
	PiSA-SR-PSNR [179]	22.6265	0.5994	0.7279	0.3368	110.5112	9.2583	24.0154	0.2066	0.1766
	SwinIR [119]	22.9425	0.6095	0.6860	0.3815	162.1128	9.7613	37.0268	0.2792	0.2814
Perception-based SR	DiffBIR [123]	21.4093	0.4612	0.5595	**0.2214**	114.8595	**3.4046**	**57.5771**	0.5030	**0.7588**
	OSEDiff [216]	22.0602	0.5544	0.5450	0.2647	107.3622	4.1667	52.5278	*0.4430*	0.7287
	HAT-GAN [29]	21.7525	0.5901	0.5590	0.2668	139.2047	5.4465	45.6411	0.2791	0.3448
	PiSA-SR [179]	22.1022	0.5761	0.5552	0.2517	100.3336	4.2723	48.0151	0.3194	0.5972
	SwinIR (Real-ISR) [119]	21.6770	0.5731	**0.5431**	*0.2431*	129.7745	*3.7377*	50.5413	0.3033	0.4885
Expert Aerial SR	HAUNet [196]	23.6649	0.6268	0.6922	0.3304	86.2487	9.0018	26.2489	0.2207	0.2567
	TransENet [102]	22.9690	0.5992	0.7449	0.3531	*97.5895*	7.6931	21.3092	0.1903	0.1765
Agentic System	AgenticIR [274]	17.8736	0.4675	0.5928	0.2451	135.6437	3.8950	*54.3685*	0.4301	0.6551
	4KAgent (Aer4K-F)	*23.4348*	*0.6255*	0.6520	0.3312	105.6710	9.0064	33.6645	0.2725	0.3314
	4KAgent (Aer4K-P)	21.9826	0.5515	*0.5525*	0.2415	112.2518	3.7230	55.7730	**0.5175**	*0.7159*

the perception preference still maintains promising visual performance, as shown in Fig. 20. This demonstrates the robustness of 4KAgent across both resolution scales and sensor domains.

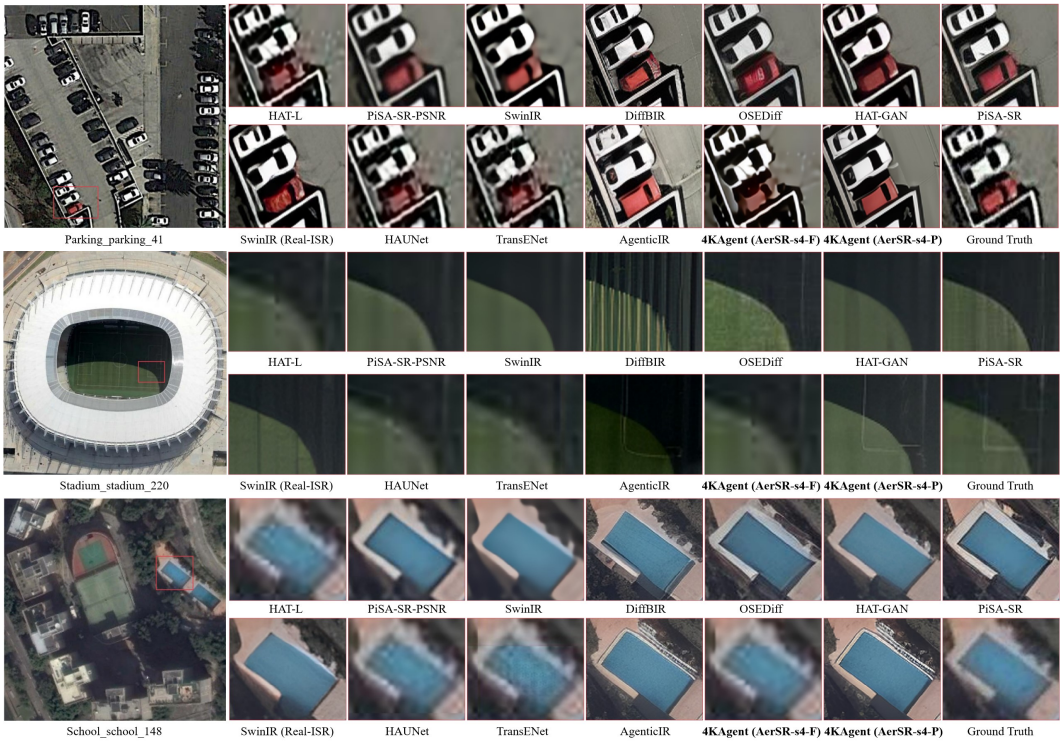


Figure 17: Visual comparison on the AID dataset (160→640).

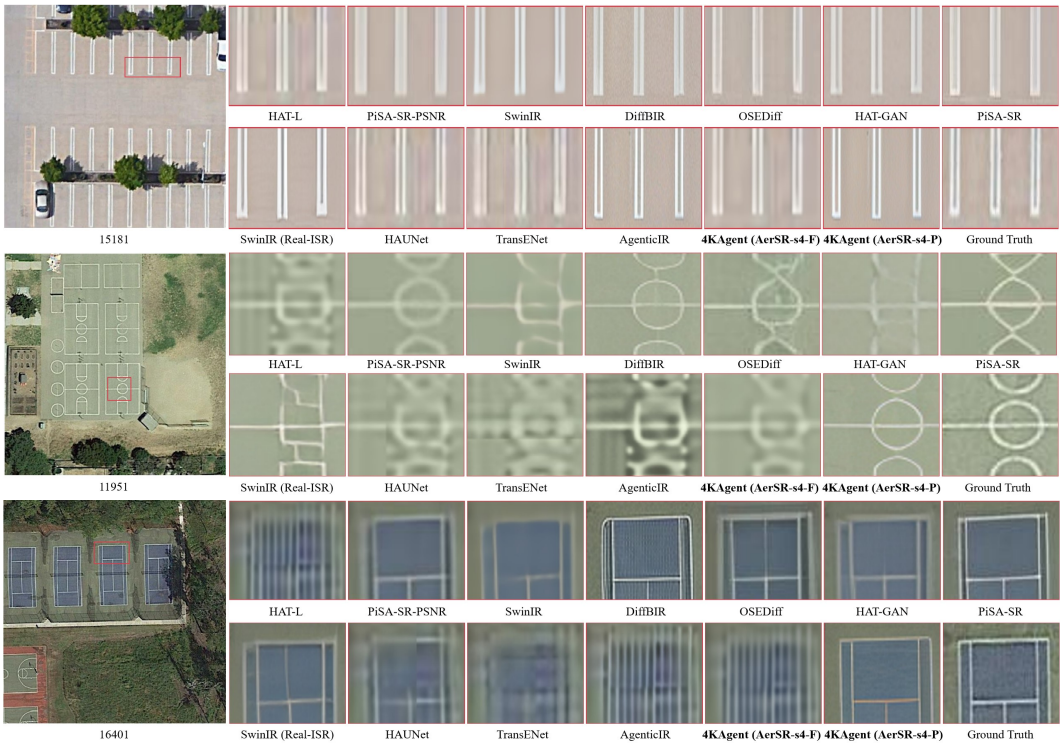


Figure 18: Visual comparison on the DIOR dataset (128→512).

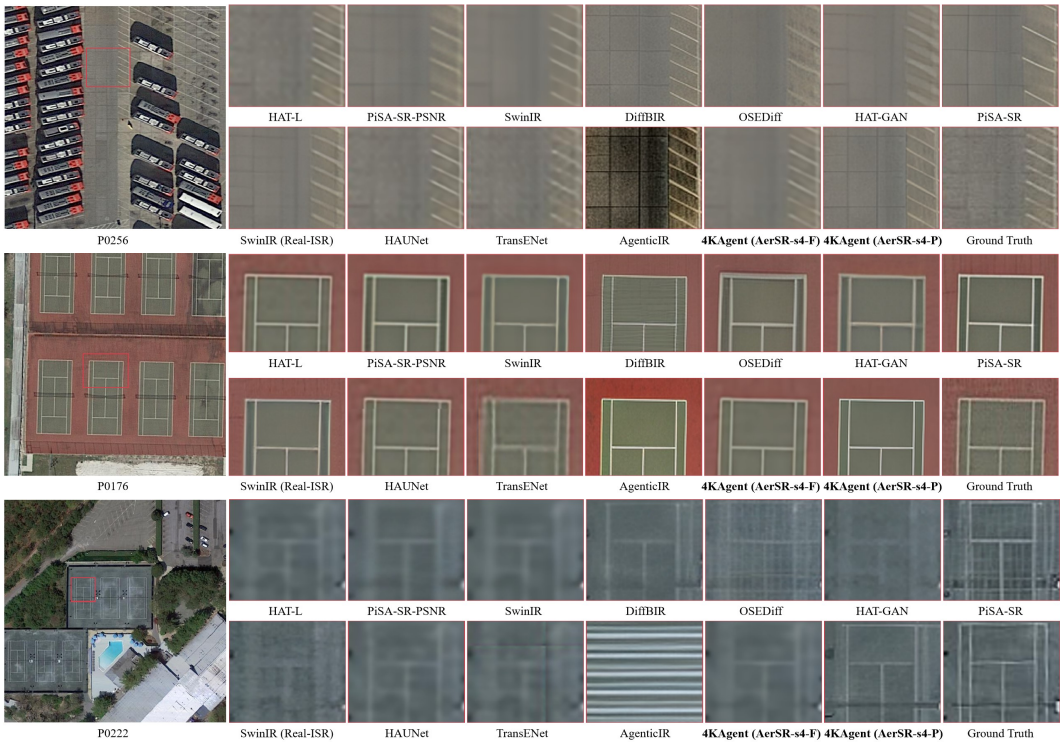


Figure 19: Visual comparison on DOTA dataset (128→512).

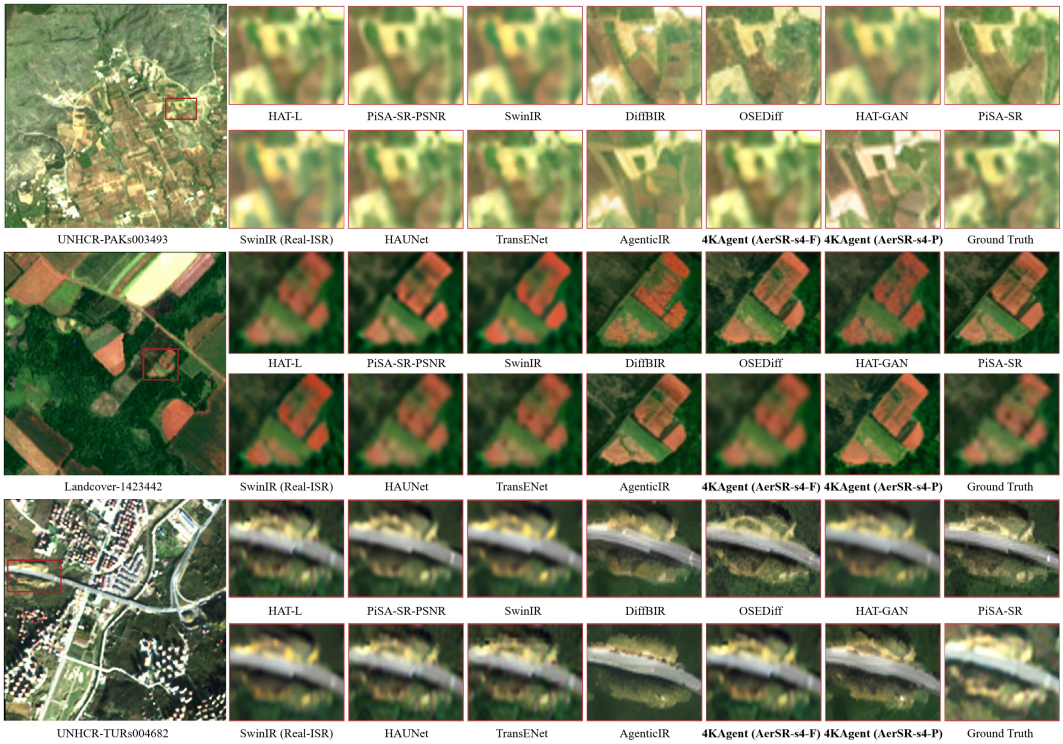


Figure 20: Visual comparison on the WorldStrat dataset (160→640).

Figure 21: Visual comparison on the DOTA dataset (4K upscaling).

Discussions These results across fidelity and perception metrics, combined with qualitative visual comparisons, provide several key insights into the advantages of 4KAgent. **First**, the consistent top-tier performance of 4KAgent with fidelity-based profile across a wide range of datasets and scaling factors suggests that the agent's analytical pipeline and adaptive control provide more precise reconstruction than traditional feedforward models. Unlike conventional SR networks that are typically optimized for either low-level fidelity or high-level realism, 4KAgents architecture decouples these objectives through specialized profiles, allowing it to excel in both domains without compromise. **Second**, the perceptual strength of 4KAgent is reflected not only in numerical scores but also in its sharper textures, reduced artifacts, and more semantically coherent outputs in qualitative results, demonstrating the value of integrating agentic reasoning with perceptual priors, especially in real-world datasets like WorldStrat. **Finally**, the margin by which both 4KAgent variants outperform AgenticIR, demonstrating the superiority of 4KAgent in terms of agentic system. Our contributions in the design of specialized profiles, adaptive modulation, and perceptual alignment mechanisms are crucial to bridging the gap between task generality and SR specialization. Together, these findings indicate that agentic architectures, when properly aligned with SR objectives, enable scalable, generalizable, and controllable super-resolution across diverse remote sensing domains.

F.2 Fluorescence Microscopic Image Super-Resolution

Confocal fluorescence microscopy is one of the most accessible and widely used techniques for studying cellular and subcellular structures [67, 235]. It builds a sharp image by using either a single pinhole to scan point-by-point or an array of pinholes on a spinning disk to scan multiple points simultaneously to reject out-of-focus light, offering molecular specificity and 3D sectioning capabilities. However, it is constrained by diffraction-limited resolution down to 200 nm under visible light [184]. Meanwhile, high-intensity illumination required for improved resolution leads to photobleaching and phototoxicity, limiting live-cell imaging duration and data throughput [185]. Deep learning-based single-image super-resolution (SISR) methods have shown great promise in recovering high-frequency details from lower-resolution inputs in biological microscopy, overcoming some limitations of hardware-based SR techniques [139], despite the scarcity of large, publicly available fluorescence microscopy datasets. To extend the evaluation of our 4KAgent on this scientific application of different modalities, we conducted experiments on a representative dataset against baselines from major SISR families.

Settings We evaluate 4KAgent on *SR-CACO-2* benchmark dataset [12], which contains 2,200 unique images of the Caco-2 human epithelial cell line, labeled with three distinct fluorescent markers: Survivin (CELL0), E-cadherin / Tubulin (CELL1), and Histone H2B (CELL2), at $\times 2$ (256→512), $\times 4$ (128→512), and $\times 8$ (64→512) scales. To generate high-resolution images, each tile was scanned with a 1024×1024 pixel resolution, and 8 scans were captured and then averaged together to reduce noise. Meanwhile, low-resolution images were captured directly by the microscope at three different scales without averaging. The full dataset contains 9,937 patches for each cell extracted from scanning confocal volumes, with tiles 9, 10, 14, 20 used as the test set. In our experiments, we randomly sampled 100 patches from each marker category in the test set at three super-resolution scales.

We evaluate 4KAgent with the **ExpSR-s2-F** profile, **ExpSR-s4-F** profile, and **ExpSR-s8-F** profile, considering the demands and requirements of the microscopy image super-resolution task. We benchmark 4KAgent against 15 representative SISR models, broadly spanning pre-upsampling, post-upsampling, iterative up-and-down sampling and progressive upsampling SR methods. Each model has been trained on the SR-CACO-2 training set before deployment. Encompassing a wide spectrum of upsampling strategies, this rigorous benchmark ensures a comprehensive comparison between our 4KAgent and other specialized and general-purpose SR methods, assessing 4KAgent's blind inference performance on the novel microscopy data domain.

Quantitative Comparison All quantitative results are in Tab. 22. PSNR, SSIM, and NRMSE were selected as criteria. We noticed that the background of a cell constituted a significant proportion of an image, as was also reported in the original SR-CACO-2 benchmark. Because including the non-informative dark background in evaluation can lead to inflated and biased performance metrics, we adopted the masking strategy described in [12] to define our Regions-of-Interest (ROIs) and calculated performance metrics based only on these areas. Across different scales and cell types, 4KAgent with **Fidelity** preference consistently achieves top performance in **pixel-level reconstruction** in ROI. The superior result on PSNR, SSIM, and NRMSE confirms 4KAgent's effectiveness in reconstructing fine fluorescence-labeled structures and low-level pixel fidelity, under various downsampling conditions. Furthermore, when compared with ENLCN, one of the most competitive methods, our 4KAgent with fidelity mode consistently exhibits a clear advantage in all numerical metrics, underscoring its ability to handle the blind super-resolution task for real-world microscopy data.

Qualitative Comparison Representative qualitative results for the highly challenging $8\times$ super-resolution task are shown in Fig. 22. At such a high magnification, where information loss is severe, the ability to reconstruct distinct biological structures for each of the three cellular markers becomes a critical test for any SISR method.

Our visual analysis reveals clear performance differences. For the inherently dim and sparse CELL0 Survivin marker, 4KAgents reconstruction is markedly clearer and closer to the ground truth. It successfully restores the faint midbody structure with higher fidelity than top-performing baselines like ENLCN and ACT, which struggle to resolve this signal from the background. This superior performance is also evident for CELL1, where 4KAgent delineates the membrane and cytoskeletal framework with sharp, continuous lines. In contrast, the outputs from most other methods appear

Table 22: Performance comparison of evaluated models on the selected sr-caco-2 test set **on ROI only, *i.e.*, cells**. The top three results for each metric are marked in **bold**, underline, and *italic*.

SISR Methods	Scale	PSNR↑				NRMSE↓				SSIM↑			
		CELL0	CELL1	CELL2	Mean	CELL0	CELL1	CELL2	Mean	CELL0	CELL1	CELL2	Mean
Bicubic	x2	34.93	32.61	30.24	32.59	0.0887	0.0681	0.0661	0.0743	0.7899	0.7826	0.7038	0.7588
	x4	35.08	31.99	30.43	32.50	0.0793	0.0690	0.0641	0.0708	0.8411	0.7998	0.7718	0.8042
	x8	32.01	28.77	26.27	29.02	0.1311	0.1071	0.1240	0.1207	0.7280	0.6677	0.6808	0.6922
Pre-upsampling SR													
SRCNN [45]	x2	37.04	34.47	33.03	34.85	0.0610	0.0516	0.0471	0.0532	0.8733	0.8566	0.8283	0.8527
	x4	35.39	32.73	31.48	33.20	0.0704	0.0610	0.0563	0.0626	0.8707	0.8193	0.8104	0.8335
	x8	32.52	29.16	26.53	29.41	0.1074	0.0924	0.1143	0.1047	0.8117	0.7219	0.7207	0.7514
VDSR [87]	x2	37.50	34.29	33.00	34.93	0.0602	0.0554	0.0479	0.0545	0.8921	0.8608	0.8385	0.8638
	x4	36.18	32.52	31.44	33.38	0.0663	0.0638	0.0571	0.0624	0.8777	0.8218	0.8180	0.8392
	x8	32.03	28.80	26.42	29.08	0.1307	0.1062	0.1237	0.1202	0.7291	0.6712	0.6877	0.6960
DRRN [180]	x2	37.35	34.23	33.08	34.89	0.0609	0.0555	0.0475	0.0546	0.8917	0.8597	0.8386	*0.8634*
	x4	36.00	32.50	31.43	33.31	0.0678	0.0637	0.0570	0.0628	0.8772	0.8216	0.8167	0.8385
	x8	31.92	28.31	26.39	28.88	0.1310	0.1096	0.1230	0.1212	0.7290	0.6583	0.6860	0.6911
MemNet [181]	x2	35.69	33.40	30.81	33.30	0.0776	0.0557	0.0607	0.0647	0.8295	0.8280	0.7759	0.8111
	x4	34.61	32.48	30.26	32.45	0.0808	0.0610	0.0654	0.0690	0.8465	0.8067	0.7651	0.8061
	x8	32.00	28.76	26.52	29.09	0.1272	0.0993	0.1183	0.1149	0.7528	0.6972	0.7102	0.7201
Post-upsampling SR													
NLSN [145]	x2	37.57	34.31	33.14	35.01	0.0588	0.0527	0.0465	0.0527	0.8911	0.8563	0.8375	0.8616
	x4	36.39	32.75	31.68	33.61	0.0630	0.0600	0.0548	0.0593	0.8754	0.8179	0.8131	0.8355
	x8	32.56	29.13	26.30	29.33	0.1147	0.0939	0.1205	0.1097	0.7909	0.7092	0.6989	0.7330
DFCAN [164]	x2	37.21	34.20	32.74	34.72	0.0614	0.0561	0.0493	0.0556	0.8899	0.8603	0.8375	0.8626
	x4	35.92	32.49	31.29	33.23	0.0684	0.0653	0.0582	0.0640	0.8770	0.8223	0.8174	*0.8389*
	x8	31.25	28.15	25.45	28.28	0.1344	0.1079	0.1276	0.1233	0.7447	0.6749	0.6841	0.7012
SwinIR [119]	x2	24.55	34.48	33.08	30.71	0.2349	0.0527	0.0473	0.1116	0.3785	0.8626	0.8385	0.6932
	x4	35.93	32.66	31.57	33.39	0.0673	0.0618	0.0559	0.0617	0.8772	0.8198	0.8161	0.8377
	x8	31.34	28.43	25.86	28.54	0.1314	0.1035	0.1230	0.1193	0.7516	0.6838	0.6923	0.7092
ENLCN [220]	x2	37.59	34.41	33.15	35.05	0.0574	0.0518	0.0462	0.0518	0.8876	0.8569	0.8340	0.8595
	x4	36.30	32.74	31.63	33.56	0.0638	0.0606	0.0553	*0.0599*	0.8766	0.8196	0.8148	0.8370
	x8	32.69	29.28	26.31	29.43	0.1108	0.0921	0.1205	0.1078	0.7998	0.7109	0.6984	0.7364
GRL [114]	x2	31.28	34.54	32.81	32.88	0.1088	0.0522	0.0492	0.0701	0.8043	0.8625	0.8337	0.8335
	x4	35.76	32.81	31.48	33.35	0.0678	0.0581	0.0565	0.0608	0.8774	0.8133	0.8144	0.8350
	x8	28.14	28.93	26.22	27.76	0.1555	0.0953	0.1104	0.1204	0.7296	0.7110	0.7197	0.7201
ACT [239]	x2	37.24	34.66	33.14	*35.01*	0.0619	0.0496	0.0459	0.0525	0.8890	0.8604	0.8288	0.8594
	x4	36.17	32.76	31.56	33.50	0.0652	0.0590	0.0554	*0.0599*	0.8761	0.8134	0.8065	0.8320
	x8	32.74	29.13	26.39	*29.42*	0.1064	0.0915	0.1152	0.1044	0.8083	0.7128	0.7063	0.7425
Omni-SR [193]	x2	37.35	34.19	33.02	34.85	0.0597	0.0548	0.0475	0.0540	0.8896	0.8562	0.8370	0.8609
	x4	35.86	32.53	31.49	33.29	0.0680	0.0635	0.0563	0.0626	0.8737	0.8165	0.8117	0.8340
	x8	30.44	28.21	25.32	27.99	0.1418	0.1075	0.1265	0.1253	0.7231	0.6673	0.6808	0.6904
Iterative up-and-down sampling SR													
DBPN [61]	x2	37.44	34.54	33.02	35.00	0.0588	0.0512	0.0470	*0.0523*	0.8872	0.8601	0.8339	0.8604
	x4	36.22	32.85	31.64	*33.57*	0.0638	0.0591	0.0548	0.0593	0.8745	0.8216	0.8091	0.8351
	x8	32.39	28.89	26.36	29.21	0.1103	0.0946	0.1157	*0.1069*	0.8060	0.7084	0.7149	*0.7431*
SRFBN [117]	x2	36.02	33.49	31.38	33.63	0.0767	0.0611	0.0576	0.0651	0.8319	0.8205	0.7592	0.8038
	x4	35.66	32.49	31.05	33.07	0.0729	0.0636	0.0592	0.0653	0.8589	0.8131	0.7921	0.8214
	x8	32.33	29.05	26.62	29.33	0.1243	0.1019	0.1181	0.1148	0.7553	0.6869	0.7081	0.7168
Progressive upsampling SR													
ProSR [203]	x2	36.92	34.66	32.80	34.79	0.0621	0.0494	0.0485	0.0533	0.8879	0.8577	0.8385	0.8614
	x4	36.11	32.71	31.61	33.48	0.0656	0.0614	0.0556	0.0609	0.8771	0.8217	0.8164	0.8384
	x8	32.13	29.43	26.36	29.31	0.1268	0.0909	0.1224	0.1134	0.7504	0.7200	0.6919	0.7208
MS-LapSRN [97]	x2	32.73	32.49	28.34	31.19	0.1014	0.0593	0.0805	0.0804	0.7957	0.8177	0.7735	0.7956
	x4	30.91	31.36	30.69	30.99	0.1118	0.0672	0.0611	0.0801	0.8124	0.7820	0.7858	0.7934
	x8	30.67	27.64	24.68	27.66	0.1206	0.1056	0.1305	0.1189	0.7829	0.6902	0.6649	0.7127
Agentic System													
4KAgent (ExpSR-sN-F) (N ∈ [2, 4, 8])	x2	39.92	36.95	33.93	**36.94**	0.0508	0.0337	0.0426	**0.0424**	0.9321	0.9105	0.8745	**0.9057**
	x4	41.25	36.86	35.07	**37.73**	0.0389	0.0318	0.0366	**0.0358**	0.9555	0.9314	0.9089	**0.9319**
	x8	38.93	33.66	31.99	**34.86**	0.0532	0.0483	0.0602	**0.0539**	0.9378	0.9033	0.8929	**0.9113**

noticeably blurry, failing to preserve the cells essential structural integrity. In the case of the bright nuclear marker CELL2, the diffuse nature of the chromatin structure means even the ground truth image itself lacks hard, well-defined edges. In this difficult context, 4KAgent reconstructs a complex, high-frequency textural pattern that is visually competitive with the other methods. While the intricate nature of the target makes absolute fidelity hard to judge, our method effectively generates a detailed result on par with other models, even under the zero-shot blind inference setting.

Discussions Our experiments show that 4KAgent delivers leading performance on the challenging SR-CACO-2 dataset, with quantitative metrics and qualitative results surpassing those of the evaluated specialist models. The superior performance of 4KAgent underscores its strong applicability to fluorescence microscopy SISR. First, it showcases strong zero-shot generalization, achieving highly competitive super-resolution performance on microscopy data, and could be further strengthened by adapting more domain-specific tools. Second, 4KAgent exhibits impressive cross-domain transferability, successfully adapting methods originally optimized for natural scenes to the distinct characteristics of fluorescence microscopy images. Third, the agent-based architecture enables the flexible and

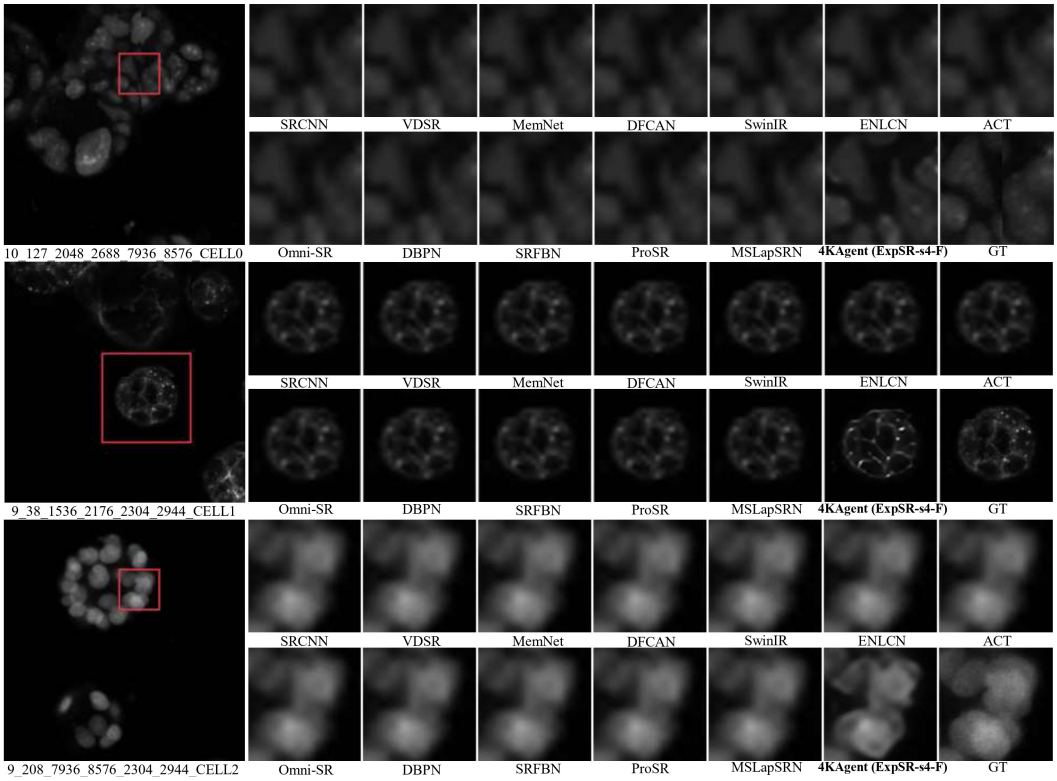


Figure 22: Visualization of fluorescence microscopy image SR on SR-CACO-2 dataset (64→512).

modular integration of existing models without requiring expensive retraining or model modification. Beyond immediate applications to super-resolution, the modularity and domain-agnostic nature of 4KAgent also suggest its broad potential for other real-world biomedical imaging domains where data scarcity or retraining costs are limiting factors.

F.3 Pathology Image Super-Resolution

Pathology images, particularly whole-slide images (WSIs) and their extracted patches, play a critical role in digital diagnostics and disease detection. Typically, glass slides containing tissue sections stained with hematoxylin and eosin are digitized using high-speed scanners at resolutions approaching $\sim$0.25 μm per pixel, resulting in gigapixel-scale images characterized by distinct color profiles and high-frequency textures unique to cellular structures. However, the substantial costs and data storage requirements associated with ultra-high-resolution scanning have led many workflows to rely on computational upscaling from lower-resolution acquisitions. This task is challenging, as pathology images possess specialized characteristics that pose significant difficulties for conventional single-image super-resolution (SISR) methods originally optimized for natural scenes. To address this challenge, we evaluate our 4KAgent on the bcSR dataset [77], comparing it against several established techniques to assess its effectiveness in this specialized domain.

Settings Our evaluation is conducted on the bcSR benchmark dataset curated for pathology image super-resolution. The bcSR dataset was derived from the larger CAMELYON [126] dataset, which contains WSIs of H&E-stained breast cancer sentinel lymph node sections. To create bcSR, the authors first sampled representative 1024×1024 patches from the original WSIs. Subsequently, a filtering process was applied to remove patches with large blank areas and to select for images with high color channel variance, ensuring the dataset was rich in informative and challenging tissue structures. The final bcSR dataset consists of 1,200 unique images, which were split into a 1,000-image training set and a 200-image test set.

Following the standard protocol established by the bcSR benchmark, the high-resolution ground truth images were downsampled using bicubic interpolation to generate the low-resolution inputs. We evaluated 4KAgent using the **ExpSR-s4-F** and **ExpSR-s8-F** profiles for the 4× and 8× tasks, respectively, prioritizing pixel fidelity which is critical for preserving fine diagnostic details. Performance was measured using the PSNR and SSIM metrics.

Quantitative Comparison Quantitative results for the pathology image super-resolution task are summarized in Tab. 23. Across both 4× and 8× upsampling tasks, our 4KAgent achieves the highest SSIM score among all evaluated methods. While CARN, a model specifically designed for this pathology dataset, attains a marginally higher Peak PSNR, 4KAgent's superior SSIM is more indicative of its ability to accurately preserve the complex tissue morphology and textures that are essential for pathological diagnosis, which is more critical for the reliability of features used for clinical assessment. This demonstrates the effectiveness of 4KAgent in recovering diagnostically relevant details from downsampled pathology patches, with performance comparable or superior to other specialized, fully-trained models.

Table 23: Quantitative comparison on Pathology dataset. The top three results for each metric are marked in **bold**, underline, and *italic*.

Method	4×		8×	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑
Bicubic	27.019	0.6659	22.475	0.2776
SRCNN [45]	27.475	0.7329	22.489	0.3624
SRGAN [98]	28.606	0.7719	23.729	0.5580
EDSR [121]	29.830	0.8058	24.366	0.5715
RDN [267]	*29.913*	0.8074	24.392	0.5711
RCAN [266]	29.916	*0.8085*	*24.404*	0.5749
SWD-Net [32]	29.853	0.8000	24.465	*0.5755*
CARN [77]	**29.964**	0.8408	**24.479**	0.5763
4KAgent (ExpSR-sN-F) (N ∈ [4, 8])	29.746	**0.8602**	24.300	**0.5826**

Qualitative Comparison Fig. 23 presents a qualitative comparison for 4× super-resolution on representative patches from the bcSR test set. While both methods significantly improve upon the heavily blurred low-resolution inputs, a closer inspection of the ROIs reveals that our training-free 4KAgent consistently produces results that are on par with, and often superior to, the fully-trained, domain-specific CARN model. This superior performance is particularly evident in challenging cases. In image 1010, 4KAgent successfully delineates individual cell boundaries and restores the heterogeneous texture of the tissue architecture. In contrast, CARN's output suffers from a loss of sharpness and definition, resulting in noisier and blurrier cell regions and poorly defined tissue architecture, while also introducing a slight color deviation. Similarly, in image 1175, 4KAgent accurately reconstructs the intricate internal structures, preserving the sharp outlines of the nuclei and cytoplasm. CARN's output, conversely, suffers from a loss of sharpness and inaccurate detail, while also exhibiting subtle grid-like artifacts. Across all examples, 4KAgent consistently generates nuclei with sharper boundaries and more distinct internal textures, along with clearer cell membranes, demonstrating a higher fidelity to the ground truth.

These visual improvements also correlate with higher SSIM scores of 4KAgent, confirming its enhanced ability to preserve the structural integrity of the tissue. The accurate recovery of such fine-grained morphological details is critical for potential downstream clinical applications. High-fidelity reconstructions like those from 4KAgent can enable more reliable automated analysis, such as precise nuclei segmentation for cell counting, classification of cellular atypia, and grading of cancerous tissue, thereby highlighting the potential value of our approach in digital pathology workflows.

Discussions The combined quantitative and qualitative results underscore the significant potential of 4KAgent for pathology image super-resolution. Although not leading in PSNR, 4KAgent's superior SSIM scores demonstrate a more accurate reconstruction of high-frequency textures and tissue morphology, which are paramount for pathological interpretation. Furthermore, because 4KAgent is not trained on a specific pathology dataset, it is less susceptible to overfitting to the characteristics of a single data source. This provides a significant advantage when performing SR

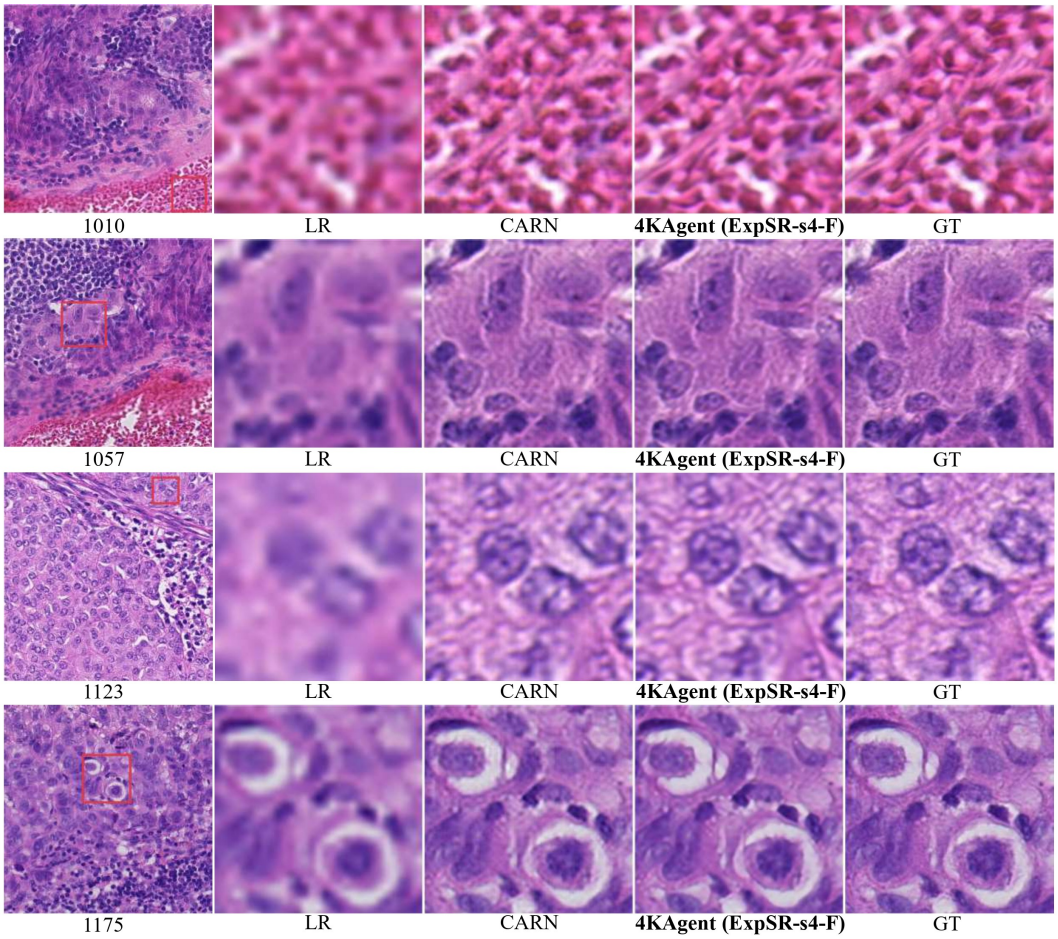


Figure 23: Visual comparison of pathology image super-resolution on bcSR dataset (256→1024).

on real-world pathology images acquired by different scanners and staining protocols. Additionally, 4KAgent's agentic framework allows for flexible expansion of its tool profile to better adapt to pathology imaging modalities. Its leading performance on the bcSR dataset validates the potential of this agentic approach as a robust and generalizable solution for biomedical imaging.

The ability to generate reconstructions with high structural fidelity has direct implications for critical downstream applications in computational pathology. By restoring sharper nuclei, clearer cell membranes, and more intelligible tissue architecture, 4KAgent provides a more reliable input for automated analysis pipelines. This can enhance the accuracy of tasks such as nuclei segmentation, cell counting, and the classification of cancerous tissue, ultimately making it a more robust and practically useful tool for AI-assisted biomedical diagnostics.

F.4 Medical Image Super-Resolution: X-Ray, Ultrasound, and Fundoscopy

In this chapter, we shift our focus to the super-resolution of clinical diagnostic imaging modalities, where the primary goal is to enhance anatomical and pathological details for improved diagnostic accuracy while minimizing patient burden, such as reducing exposure to ionizing radiation in X-ray imaging [188]. Although often grouped together, these modalities can be fundamentally diverse, operating on different physical principles: from X-rays utilizing ionizing radiation to ultrasound relying on acoustic waves. This diversity gives rise to unique image characteristics and modality-specific challenges, such as maintaining pathological invariance in chest X-rays or avoiding the generation of pseudo-structures in ultrasound images.

The prevailing approach in medical image SR has been the development of highly specialized models, each trained on specific datasets for a single modality [243]. The rise of foundation models

has led to more powerful specialist systems, such as those tailored for a single modality like CT or dermatology [157, 86], though a few pioneering works have begun to explore more universal solutions [122]. A significant drawback of such a specialized paradigm is poor generalization across different datasets and modalities, which creates a major bottleneck for practical clinical deployment and motivates our evaluation of 4KAgent's performance on these challenges. To this end, this chapter evaluates 4KAgent's performance across several distinct and clinically important modalities.

Settings We evaluate 4KAgent with the **ExpSR-s4-F** profile across three medical imaging modalities with distinct imaging principles: X-ray, ultrasound, and fundoscopy, benchmarking against their respective baselines. Benchmark datasets of each imaging modality are summarized as follows:

- **X-ray**. Chest X-ray 2017 [85] and Chest X-ray 14 [199]. Specifically, Chest X-ray 2017 is a dataset of 5,856 pediatric images from Guangzhou Women and Childrens Medical Centre, split into 5,232 images for training and 624 images for testing. Chest X-ray 14 contains 112,120 frontal-view X-rays of 30,805 patients with 14 disease labels mined from radiology reports. Among them, 880 images additionally contain expert-annotated bounding boxes. Following the settings in [230], we evaluate on the Chest X-ray 2017 test set and the 880 annotated data in Chest X-ray 14.

- **Ultrasound Image**. US-Case [175] and MMUS1K [154]. The US-Case collection comprises over 7,000 sonographic images spanning organs such as the liver, heart, and mediastinum. Adopting the selection protocol from [154], we reused the subset of 111 images in the test set for benchmarking, excluding 11 scans whose small field of view limited their diagnostic value. MMUS1K features 1,023 anonymized multi-organ ultrasound scans, including bladder, gallbladder, thyroid, kidney, etc., sourced from Shanghai Tenth Peoples Hospital. All images meet a minimum resolution of 448×600 px and were cleansed of watermarks and blurring artifacts via LabelImg. The test set with label numbers from 0801 to 0900 was used for evaluation.

- **Fundoscopy Image**. DRIVE [176] consists of 40 color fundus images from a diabetic retinopathy screening program in the Netherlands collected by a Canon CR5 non-mydriatic 3CCD camera. The dataset is equally divided into 20 for training and 20 for testing. As all images are in 584×565 resolution, following the setup in [5], the original high-resolution (HR) images were resized to 512×512 and before being further utilized to generate LR pairs.

To consist with baseline methods for each dataset, the LR input images were generated by downsampling HR images via bicubic interpolation. For X-ray images, we use SSIM, FSIM [261], and MSIM [209] as metrics, while PSNR and SSIM are used for Ultrasound and Fundoscopy images.

Table 24: Quantitative comparison on X-ray datasets. The top three results for each metric are marked in **bold**, underline, and *italic*.

Method	Chest X-ray 2017			Chest X-ray 14		
	SSIM↑	FSIM↑	MSIM↑	SSIM↑	FSIM↑	MSIM↑
Nearest-Neighbor	0.637	0.672	0.668	0.701	0.724	0.713
Interpolation [231]	0.615	0.663	0.644	0.687	0.698	0.681
CTF [258]	0.889	0.933	*0.954*	0.917	0.955	0.943
ESPCN [172]	0.756	0.825	0.804	0.795	0.822	0.815
FSRCNN [47]	*0.897*	0.943	0.953	0.917	0.959	*0.953*
LapSRN [96]	0.893	0.942	*0.954*	0.915	0.956	0.949
SRGAN [98]	0.821	0.896	0.868	0.844	0.903	0.897
GAN-CIRCLE [240]	*0.897*	*0.947*	0.923	*0.919*	*0.969*	0.945
SNSRGAN [230]	0.911	0.981	0.983	0.925	0.995	0.986
4KAgent (ExpSR-s4-F)	**0.933**	**0.996**	**0.987**	**0.960**	**0.999**	**0.993**

Quantitative Comparison X-ray Quantitative results are summarized in Tab. 24. Ultrasound Quantitative results are summarized in Tab. 25. Fundoscopy Quantitative results are summarized in Tab. 26, which collectively demonstrate 4KAgent's consistently superior performance across all

three distinct medical imaging modalities. On the X-ray datasets, 4KAgent with Fidelity profile surpasses the specialized SNSRGAN [230] model across all structure-focused metrics. For Ultrasound imaging, it also achieves a significant performance leap, boosting the PSNR on the MMUS1K dataset by nearly 3 dB over the previous state-of-the-art, M2Trans [154]. Similarly, on the DRIVE Fundoscopy dataset, 4KAgent again sets a new performance benchmark, improving the PSNR from 37.72 to 41.52 and the SSIM from 0.91 to 0.95. This consistent outperformance across modalities, from the need for pathological invariance in X-rays to the clarity of fine vessels in fundoscopy, highlights the effectiveness and robustness of 4KAgent for diverse medical SR tasks.

Qualitative Comparison Representative qualitative results for X-Ray SR, ultrasound SR, and fundoscopy SR are shown in Figs. 24 to 26. The visual outcomes generally align with our quantitative findings and suggest the potential benefits of 4KAgent for clinical imaging applications.

For X-ray super-resolution, where maintaining pathological invariance is important, 4KAgent produces reconstructions with improved delineation of lung parenchyma and clearer visibility of rib cage contours. It appears to achieve this clarity while reducing some of the over-smoothing artifacts occasionally seen in other SR methods, thus helping to preserve diagnostic details that are crucial for identifying pulmonary abnormalities with greater confidence.

In the ultrasound comparisons, 4KAgent shows a notable advantage. On the US-CASE example, it restores clearer tissue boundaries and more internal detail compared to the blurrier reconstruction from the M2Trans baseline. Similarly, for the MMUS1K image, 4KAgent appears to reduce speckle noise while enhancing anatomical definition, whereas the baseline result is affected by some noise and artifacts. In both cases, 4KAgent generates echogenic patterns that more closely resemble the ground truth, improving overall image fidelity.

Table 25: Quantitative comparison on Ultrasound dataset. The top three results for each metric are marked in **bold**, underline, and *italic*.

Method	US-CASE		MMUS1K	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑
Bicubic	28.90	0.7892	28.24	0.7817
EDSR [121]	30.82	*0.8497*	30.04	*0.8326*
SwinIR [119]	28.50	0.7834	27.66	0.7758
ELAN [263]	*31.02*	0.8464	*30.40*	0.8309
ESRT [134]	30.84	0.8374	30.25	0.8235
HAT [30]	28.72	0.7812	28.08	0.7582
M2Trans [154]	31.32	0.8516	30.68	0.8392
4KAgent (ExpSR-s4-F)	**33.27**	**0.8895**	**33.58**	**0.8678**

Table 26: Quantitative comparison on Fundoscopy dataset. The top three results for each metric are marked in **bold**, underline, and *italic*.

Dataset	Method	PSNR↑	SSIM↑
DRIVE	Bicubic	25.20	0.86
	SRGAN [98]	*34.22*	*0.88*
	Ahmad *et al.* [5]	37.72	0.91
	4KAgent (ExpSR-s4-F)	**41.52**	**0.95**

The fundoscopy results demonstrate 4KAgent's effectiveness in restoring details from degraded inputs. Compared to the LR image, 4KAgent's reconstruction of the retinal vascular network shows a clear improvement. The method produces sharper and more continuous vessels, resolving many of the fine micro-vessels and bifurcation points that are obscured in the LR version. The resulting image more closely resembles the HR ground truth, suggesting its potential to aid in retinopathy screening from lower-resolution captures without sacrificing critical diagnostic details.

Discussions From both quantitative and qualitative perspectives, our evaluation suggests that 4KAgent is a capable system for cross-domain super-resolution across diverse clinical imaging modalities, showing competitive performance on X-ray, ultrasound, and fundoscopy datasets. This result is notable, as the prevailing approach often involves developing specialized models for each modality, a paradigm that can be limited by poor generalization across different scanners and protocols. By not relying on domain-specific training, 4KAgents agentic framework offers a flexible alternative, adaptively deploying its tools to address the unique challenges of each image, from enhancing the clarity of lung markings in chest radiographs to defining subtle echogenic interfaces in ultrasound and resolving fine vascular networks in fundoscopy.

The ability to generate reconstructions with improved structural details may have implications for downstream applications. For example, improved sharpness in retinal vessels could aid in retinopathy screening; clearer ultrasound images help with tissue boundary delineation for segmentation; and more detailed X-rays could enhance the visibility of subtle pulmonary abnormalities. Ultimately, the

Figure 24: Visual comparison of X-Ray image SR on Chest X-ray 2017 and Chest X-ray 14 dataset.

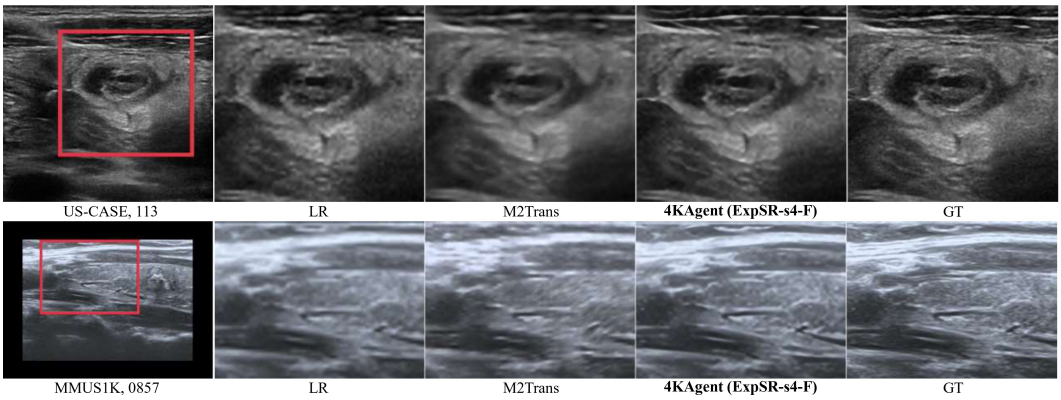


Figure 25: Visual comparison of Ultrasound image SR on US-CASE and MMUS1K dataset.

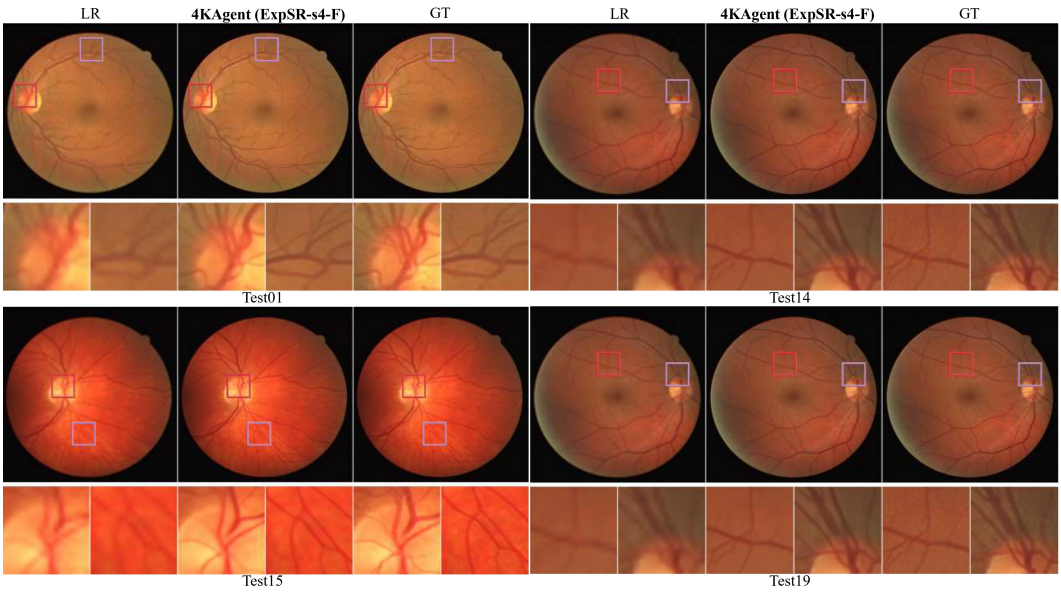


Figure 26: Visual comparison of fundoscopy image SR on DRIVE dataset (128→512).

performance of our agentic approach indicates its significant potential for robust deployment across more real-world clinical workflows and imaging modalities, driven by its adaptability and inherent extensibility for incorporating more specialized medical profiles.

G Ablation Studies and Running Time Analysis

In this section, we conduct ablation studies on core components in 4KAgent: (1) Q-MoE policy and (2) Face Restoration Pipeline. Then, we perform a running time analysis of 4KAgent.

Q-MoE policy To assess the contribution of our Q-MoE mechanism during execution and reflection, we perform an ablation study in which Q-MoE is replaced by the DFS strategy from AgenticIR [274], denoting this variant as **4KAgent (DFS)**. Experiments are conducted on the MiO-100 Group C dataset under the multiple-degradation image restoration setting.

As shown in Tab. 27, integrating Q-MoE leads to substantial improvements in perceptual quality. Specifically, metrics such as LPIPS, MANIQA, CLIPIQA, and MUSIQ exhibit significant gains, with minimal impact on fidelity metrics like PSNR and SSIM. Furthermore, the visual comparisons presented in Fig. 27 provide additional evidence, showing that 4KAgent equipped with Q-MoE generates noticeably sharper and more realistic details compared to the DFS-based variant.

Table 27: Ablation study on Q-MoE policy. The better result is marked in **bold**.

Dataset	Method	PSNR↑	SSIM↑	LPIPS↓	MANIQA↑	CLIPIQA↑	MUSIQ↑
MiO100 - Group C	4KAgent (DFS)	**19.81**	**0.5785**	0.4381	0.3286	0.4854	54.03
	4KAgent (**Q-MoE**)	19.77	0.5629	**0.4271**	**0.3545**	**0.5233**	**55.56**

Figure 27: Visual comparisons for ablation study on Q-MoE.

Face restoration pipeline To evaluate the impact of our Face Restoration Pipeline, we conduct an ablation study on the WebPhoto-test dataset using three profiles: **ExpSR-s4-P**, **ExpSRFR-s4-P**, and **GenSRFR-s4-P**. Experimental results and the difference among these profiles are shown in Tab. 28.

Enabling the face restoration module (i.e., switching profile from **ExpSR-s4-P** to **ExpSRFR-s4-P** and **GenSRFR-s4-P**) yields higher face IQA scores (CLIB-FIQA and DSL-FIQA). Moreover, when setting the 'Restore Option' to 'None' rather than 'super-resolution', we observe further improvements across both generic image IQA metrics (NIQE and MUSIQ) and face IQA metrics.

Table 28: Ablation study on face restoration pipeline on WebPhoto-Test dataset. The best and second-best results are marked in **bold** and underline respectively.

Method	Restore Option	Face Restore	NIQE↓	CLIPIQA↑	MUSIQ↑	MANIQA↑	CLIB-FIQA↑	DSL-FIQA↑
4KAgent (ExpSR-s4-P)	super-resolution	False	5.11	**0.7119**	<u>73.62</u>	**0.6601**	0.6415	0.7194
4KAgent (ExpSRFR-s4-P)	super-resolution	True	<u>4.53</u>	0.6600	72.89	0.6405	<u>0.6602</u>	<u>0.7237</u>
4KAgent (GenSRFR-s4-P)	None	True	**4.15**	<u>0.7077</u>	**75.92**	<u>0.6576</u>	**0.6671**	**0.7683**

Visual comparisons are shown in Fig. 28. 4KAgent with **GenSRFR-s4-P** profile produces the finest facial details (hair details, harmony of facial area and background area). This trend indicates that WebPhoto-test images suffer from complex, mixed degradations, and 4KAgent benefits from integrating multiple restoration tasks with Q-MoE driven selection to achieve superior visual quality.

Running Time Analysis The inference time of 4KAgent varies depending on the selected profile, the quality of the input image, and the length of the restoration plan. In this section, we analyze the inference time of 4KAgent using NVIDIA RTX 4090 GPUs. Specifically, we report the fastest and slowest cases observed in our experiments. The fastest case involves super-resolving images

LQ | 4KAgent (ExpSR-s4-P) | 4KAgent (ExpSRFR-s4-P) | **4KAgent (GenSRFR-s4-P)**

Figure 28: Visual comparisons for ablation study on face restoration pipeline.

($\times 4$) from the B100 dataset using the **ExpSR-s4-F** profile. The slowest case corresponds to jointly restoring and upscaling low-quality images from the DIV4K-50 dataset to 4K resolution under the **Gen4K-P** profile. The inference times for these two cases are summarized in Tab. 29.

Table 29: Inference time of 4KAgent (fastest and slowest cases on our experiments).

Profile Nickname	Task	Resolution	Benchmark	Length of Plan	Inference Time (s)
ExpSR-s4-F	Super-resolution ($4\times$)	$120 \times 80 \rightarrow 480 \times 320$	B100	1.0 ± 0.0	50.96 ± 2.01
Gen4K-P	Joint restoration + 4K Upscaling	$256 \times 256 \rightarrow 4096 \times 4096$	DIV4K-50	3.4 ± 0.6	1551.76 ± 230.73

As 4KAgent currently executes its tools sequentially, there is substantial potential for acceleration, for example, by running independent restoration tools in parallel at each step.

H Applications and Broader Impacts

H.1 Applications

High-resolution On-Demand Media Streaming 4KAgent offers significant potential for enabling network operators, such as YouTube, Netflix, Instagram, Amazon Prime Video, TikTok, Kwai, Snap, Twitch, to name a few, to deliver 4K-quality video services from much lower-bitrate streams. For example, edge-based SR can upscale a 1K stream to 4K at the user's device [250], allowing providers to store and transmit mostly lower-resolution content (*e.g.*, 1K) which can then be upscaled to 4K quality on the end-user's device using edge-based processing. This approach dramatically cuts storage and bandwidth costs (and even energy use) compared to naïvely streaming native 4K [108]. Technologies like NVIDIA's Deep Learning Super Sampling (DLSS) [1] demonstrate the feasibility and usability of real-time super-resolution on GPU chips. Integrating such real-time upscaling into adaptive streaming protocols could also improve user experience by minimizing disruptive quality shifts often associated with variable network conditions, ensuring viewers consistently receive high-resolution playback on capable displays.

Video Conferencing and Telepresence Network bandwidth constraints and limitations inherent in typical webcams or smartphone cameras often necessitate transmitting video streams at resolutions lower than 4K. Implementing SR algorithms, such as 4KAgent, on the receiver's end can effectively upscale these lower-resolution feeds. This process restores fine-grained details in facial features or gestures that might otherwise be lost, thereby enhancing the perceived visual quality and potentially aiding communication cues like lip-reading or the interpretation of subtle expressions [151, 225, 112]. Consequently, even devices with modest camera capabilities can deliver an experience approximating 4K quality to the viewer, without requiring increased upload bandwidth from the sender. This democratization of high-resolution video conferencing can improve remote collaboration, making it more accessible and effective for users constrained by network limitations or hardware capabilities.

Surveillance and Security Image SR technologies like 4KAgent (with fidelity-based profile) offer significant value in enhancing footage from law enforcement operations, particularly from body-worn cameras and dashcams. These devices often capture video at resolutions like 720p or 1080p with

wide fields of view, resulting in low-detail imagery, especially in challenging conditions such as low light [109]. Faces or license plates captured at a distance may span only a few dozen pixels, far below recommended thresholds for identification (*e.g.*, $\sim$90$\times$60 pixels per face for courtroom evidence [2]). The quality is often further compromised by heavy compression and sensor limitations, introducing noise and motion blur. Modern SR approaches, particularly "blind" methods that model complex real-world degradations, can effectively mitigate these issues and restore detail in practical bodycam footage. By enhancing critical regions (faces, license plates) in police videos, SR can improve both human and automated identification, while preserving the veracity required for judicial use.

Similarly, public surveillance systems, including city-wide CCTV networks, border security cameras, and transit hub monitoring, face comparable challenges related to resolution and image quality. Fixed cameras covering wide areas often render persons or objects of interest with very low pixel counts, with quality impacted by distance, illumination, camera motion, and aggressive compression techniques employed to manage bandwidth and storage [152]. SR provides a means to enhance detail retroactively without costly hardware upgrades. Field studies have also reported the effectiveness of SR. For example, a National Institute of Justice study [3] showed that multi-resolution SR could reconstruct identifiable features from extremely low-resolution facial images comparable to those from real-world security cameras. Overall, SR can act as a force multiplier for legacy surveillance infrastructure, enhancing situational awareness and forensic capabilities. However, the enhanced capability for identification also raises potential privacy concerns, which will be discussed in Appendix H.3.

Gaming and Entertainment SR techniques are extensively utilized in the entertainment sector to enhance visual quality while sustaining high frame rates, particularly in demanding gaming, VR, and AR applications. A prominent example is NVIDIA's DLSS, a suite of AI-powered neural rendering techniques that upscale lower-resolution frames to higher target resolutions, as high as 4K. DLSS can significantly improve performance, often more than doubling GPU throughput and leading to substantial frame rate increasesfor instance, one report indicated a boost of up to approximately 360% (*e.g.*, from 8 to 36.8 fps on an RTX 2060 at 4K). Successive iterations like DLSS 3 with Frame Generation and DLSS 3.5 with Ray Reconstruction have introduced further advancements by using AI to generate additional frames or improve ray-traced effects.

VR, AR, and XR This need for efficient, high-quality rendering extends critically to the domain of spatial intelligence and computing, as seen in advanced devices like the Apple Vision Pro, which aims to deliver experiences with more pixels than a 4K TV per eye. While such platforms boast high native display resolutions, SR techniques could play a crucial role in rendering complex mixed-reality scenes or high-fidelity passthrough video efficiently, maintaining visual clarity without overwhelming the processing capabilities. Similarly, as smart glasses like the Ray-Ban Meta Wayfarer evolve and potentially incorporate more advanced display capabilities for augmented reality overlays, SR will be key to delivering crisp digital information without excessive battery drain. Broader XR initiatives, such as Google's development of Android XR, also stand to benefit from robust SR solutions to enable a diverse ecosystem of devices to achieve compelling visual experiences. For all these platforms, from gaming consoles to sophisticated XR headsets and smart glasses, the ability of SR systems like 4KAgent to adaptively enhance visual quality from various inputs will be paramount in balancing immersive, high-resolution experiences with practical performance and power constraints.

AI-Generated Content (AIGC) Production Industry Photographers, digital artists, and filmmakers increasingly leverage SR tools to enlarge, restore, and enhance the quality of both conventional (*e.g.*, old photos, archival digital footage) and even AI-generated images and video footage. We have demonstrated in Appendix E.1 that 4KAgent is capable of synthesizing high-fideltiy details in generated media, coinciding with a recent trend that generates a high-resolution advertisement for KFC[2], leveraging outputs from generative video models such as Google's Veo [191], Luma AI's Dream Machine [135], and OpenAI's Sora [155], further enhanced using Topaz Labs Video Upscaler to achieve higher resolutions (*e.g.*, 4K) and professional quality suitable for broader use. SR techniques are crucial for bridging this gap towards generating ultra-high-resolution content, enabling creators to enhance these AI-generated visuals. For instance, images generated for concept art, marketing materials, or virtual environments can be significantly improved in detail and clarity through SR, making them suitable for 4K displays or large-format printing. Similarly, generative

[2] https://x.com/Wesley_Kibande/status/1908091178723029193

video content, which might be created at lower resolutions to manage computational costs, can be upscaled using specialized tools like Topaz Video AI to achieve crisper, higher-resolution results (*e.g.*, 4K) ready for distribution or integration into larger productions. State-of-the-art SR methods, including GAN-based approaches, can synthesize photorealistic details, effectively transforming AIGC outputs into polished, professional-grade assets. The ability of robust and adaptive SR solutions like 4KAgent to handle the diverse and sometimes unpredictable nature of AIGC makes them particularly valuable for ensuring that AI-driven creative endeavors can meet high-quality benchmarks.

Scientific Imaging High-resolution imagery is crucial across numerous scientific disciplines, particularly when native sensor capabilities are constrained. In remote sensing, deep super-resolution (SR) methods significantly enhance spatial details of satellite imagery, facilitating accurate land-use classification and environmental monitoring [153, 171]. For instance, self-supervised SR techniques trained on sequences of satellite images yield sharper and less noisy results compared to raw captures, substantially improving downstream geospatial analysis [153]. Microscopy and biomedical imaging similarly benefit from SR, particularly through novel quantum imaging techniques. Recent advancements by Zhang et al. and He et al. leverage quantum entanglement to achieve unprecedented imaging resolution, demonstrating quantum microscopy at the Heisenberg limit and significantly enhancing cellular and sub-cellular visualization [265, 65]. Additionally, computational SR methods applied to microscopy, like content-aware restoration for fluorescence images [215], complement these quantum techniques by computationally reconstructing detailed 3D biological structures from limited optical inputs. Thus, versatile and advanced SR frameworks such as 4KAgent, coupled with emerging quantum imaging methods, can revolutionize scientific research by providing richer, more precise imagery across multiple imaging modalities.

Medical Image Applications In medical imaging, SR facilitates detailed diagnostics by transforming low-dose or rapidly acquired imaging scans into high-fidelity medical images. Techniques employing deep learning-based SR on modalities such as Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) have shown promise in generating accurate, high-resolution images from suboptimal inputs, thus reducing patient radiation exposure and acquisition time without sacrificing diagnostic quality [31, 269]. For instance, generative adversarial networks (GANs) and transformer-based SR approaches like TransMRI demonstrate substantial improvements in enhancing anatomical details critical for diagnostic accuracy [51]. Consequently, methods like 4KAgent, which provide universal super-resolution capabilities, can significantly impact clinical diagnostics by offering highly detailed and diagnostically reliable imagery from resource-efficient imaging procedures.

Embodied AI and Robotics Embodied AI systems, including robotics platforms, leverage SR to enhance visual perception, critical for tasks such as navigation, object manipulation, and human-robot interaction. Robotic visual systems frequently face limitations in sensor resolution and onboard processing capacity, challenges that SR methods can effectively address. Recent studies indicate that integrating SR into robotic vision pipelines notably improves object detection and localization, particularly for distant or small-scale objects critical in dynamic environments [141]. Furthermore, real-time lightweight SR models tailored for robotic platforms have been developed, improving perception accuracy and enabling robots to perform complex tasks efficiently, such as precise grasping and navigation through cluttered or visually challenging scenarios [7, 76]. Consequently, robust SR algorithms substantially advance robotic autonomy and operational effectiveness.

Autonomous Vehicles, Drones, and Intelligent Transportation Systems (ITS) Autonomous vehicles, aerial drones, and intelligent transportation systems increasingly depend on high-quality visual inputs for safe and efficient operation. SR technologies significantly boost the capability of onboard cameras to discern critical details from lower-resolution captures under real-world constraints. For instance, SR-enhanced imagery improves the detection accuracy of pedestrian, vehicle, and traffic sign recognition systems, crucial for autonomous navigation and safety-critical decision-making [183, 149, 127]. Research demonstrates that training object detection models with super-resolved images significantly enhances their effectiveness in challenging scenarios, including poor visibility or long-range object detection from drones or vehicle-mounted cameras [149]. Furthermore, ITS can employ SR-enhanced camera networks for robust and precise traffic monitoring and anomaly detection, improving urban mobility management and public safety.

GIScience and GeoAI Geographic Information Science (GIScience) heavily utilizes spatially explicit, high-resolution data for mapping, analysis, and policy-making. Deep SR has recently emerged as a powerful method for enhancing geospatial imagery resolution, substantially improving the interpretability and accuracy of satellite-derived GIS datasets. For example, SR methods significantly improve the extraction accuracy of buildings, roads, vegetation, and other geographic features from medium- to low-resolution imagery, providing a cost-effective alternative to deploying expensive high-resolution satellite sensors [171, 163]. Additionally, high-quality SR outputs are vital for applications such as precision agriculture and disaster response planning, demonstrating the broad utility of universal SR frameworks like 4KAgent within GIScience research and applications.

H.2 Broader Impacts

Economic Impacts. SR technologies, exemplified by 4KAgent, drive significant economic advantages by enhancing operational efficiency, creating new markets, and supporting environmental sustainability. By reducing bandwidth, storage, and infrastructure costs associated with high-resolution content delivery, SR solutions enable economical, high-quality media dissemination over constrained networks, benefiting digital platforms and smaller businesses [116, 251]. Companies such as BytePlus and Maxar Intelligence have successfully leveraged SR technologies to open new markets in healthcare diagnostics, geospatial intelligence, and media restoration [16, 74]. Additionally, by minimizing data storage and transmission demands, SR contributes meaningfully to the environmental goals of reduced energy consumption and lower carbon emissions [101].

Accessibility. SR substantially promotes digital equity by enabling high-quality visual content access for users in regions with limited bandwidth or resource constraints without necessitating advanced infrastructure or expensive devices [161]. Particularly in education and healthcare, SR supports remote learning and telemedicine by delivering clearer instructional and diagnostic imagery, significantly benefiting underserved communities [15]. Moreover, SR-integrated assistive technologies offer improved accessibility for individuals with visual impairments by enhancing image clarity and text readability, facilitating greater inclusion and interaction within the digital sphere [120, 192].

Vertical Impacts to Industry. Demand for real-time, high-quality SR capabilities stimulates technological progress across various industries, including media, robotics, autonomous systems, and scientific visualization. SR advancement drives innovation in specialized neural hardware, edge computing solutions, and embedded AI, spurring the development of powerful and efficient Neural Processing Units (NPUs) in consumer devices [18, 99]. Additionally, SR techniques significantly enhance autonomous vehicle, drone, and robotic platform capabilities by improving object detection accuracy, scene understanding, and decision-making reliability, particularly in challenging operational environments [149, 171]. These impacts extend to scientific instrumentation, such as microscopy and geospatial imaging, where SR enables unprecedented detail and precision [153, 215].

H.3 Limitations and Potential Negative Societal Impacts

Efficiency and Computational Cost. SR methods, particularly deep learning-based and agentic frameworks like 4KAgent, often require substantial computational resources for training and inference. High-resolution SR models usually rely on resource-intensive GPU or TPU clusters, imposing significant energy consumption [173]. Even optimized inference can become computationally burdensome at the edge, potentially limiting deployment on low-powered or mobile devices unless significant model compression and acceleration techniques are applied [99, 264]. Balancing performance and efficiency remains a critical open challenge, particularly for real-time applications in resource-constrained environments.

Bias, Fairness, and Model Drift. Data-driven SR models inherit biases from their training datasets, potentially leading to uneven quality across different image categories, demographics, or scenarios [54, 143]. Such biases might systematically disadvantage specific groups, for instance, by inadequately resolving images related to underrepresented populations or environments. Model drift over timewhere the model gradually becomes less accurate due to changing real-world distributionsalso poses a serious issue for practical deployment, requiring continuous monitoring and recalibration to ensure fairness and reliability [52, 133].

Ethical Issues and Privacy. Enhanced imaging capabilities enabled by SR, particularly in surveillance contexts, can amplify privacy risks. The ability to recover detailed features such as faces or license plates from previously anonymized or low-resolution imagery might lead to unauthorized or unethical identification of individuals [62]. This capability necessitates clear regulatory guidelines and ethical oversight to avoid misuse [115]. Concerns are especially pronounced in contexts of law enforcement, border surveillance, and public monitoring, where SR technologies must be carefully governed to prevent potential violations of civil liberties and personal privacy [38].

Failure Modes in High-stake Settings. The adoption of SR techniques in high-stakes environments, including medical diagnostics, autonomous vehicles, and security monitoring, introduces risks associated with model hallucinations or misleading image reconstructions [35]. SR models, particularly generative approaches, may produce plausible yet incorrect details absent from the original low-resolution inputs, potentially leading to erroneous interpretations or decisions [8]. In clinical settings, for example, SR-generated artifacts could result in incorrect diagnoses or overlooked medical conditions, underscoring the importance of rigorous validation and transparency regarding model uncertainty and reliability [84].

I Related Work

I.1 Image Super-Resolution

Deep learning has significantly advanced the field of single-image Super-Resolution (SR). The seminal work, SRCNN [45], introduced a convolutional net for SR, with a primary focus on optimizing the Mean Squared Error between the super-resolved and high-resolution images. Following this, numerous studies have enhanced reconstruction accuracy by improving network architectures, including residual and dense connections [87, 121, 267], attention mechanisms [19, 41, 266], and multi-scale networks [53, 110]. While these methods perform well in modeling the posterior distribution of the training data, they inevitably suffer from the issue of overly smooth visual results [98, 147, 226]. In recent years, significant efforts have been made to develop generative model-based SISR techniques that produce more visually appealing results. These include autoregressive models [40, 146, 190], GAN-based models [98, 201, 256, 21, 118, 105], and diffusion-based models [195, 234, 123, 217, 168, 205, 216]. SRGAN [98], as a pioneering GAN-based SR model, assumes image degradation through bicubic downsampling and generates photo-realistic images. BSRGAN [256] and Real-ESRGAN [201] achieve promising real-world SR results by using randomly shuffled degradation and higher-order degradation. SwinIR [119] replaces the CNN-based generator network with visual transformers, leading to more stable training and more realistic textures. Additionally, SeD [105] introduces a semantic-aware discriminator to capture fine-grained distributions by incorporating image semantics as a condition. Recent diffusion-based models have focused on fine-tuning the Stable Diffusion model for reconstructing high-quality images, using low-quality images as control signals. Notably, StableSR [195] fine-tunes a time-aware encoder and employs feature warping to balance fidelity and perceptual quality. SeeSR [217] introduces degradation-robust, tag-style text prompts to enhance the semantic awareness of the Real-ISR model. Furthermore, recent studies on diffusion-based models, such as SinSR [205], OSEDiff [216], PiSA-SR [179], and GuideSR [9] achieve one-step image super-resolution.

I.2 Image Restoration

Recent advances in deep learning have led to remarkable progress in blind image restoration tasks, including denoising, deblurring, deraining, dehazing, and removal of JPEG compression artifacts. Early works such as ARCNN [44] demonstrated the potential of compact convolutional neural networks, particularly in the context of image denoising. Since then, a broad range of sophisticated network architectures and training strategies have been developed to further enhance restoration performance. These include the use of residual blocks [255, 131, 257], attention mechanisms [242, 25, 186, 57, 247], and Transformer-based designs [189, 275, 246, 206]; as well as generative paradigms such as GANs [58, 26, 11, 56, 73, 147, 158, 128] and diffusion models [123, 221, 81, 204, 82, 50]. Notably, general-purpose restoration models like Uformer [206], MAXIM [186], Restormer [246], and NAFNet [25] have demonstrated strong performance across diverse restoration tasks, often trained independently for each specific degradation type. However, such single-degradation methods often struggle in real-world scenarios where multiple types of degradations co-exist. This limitation

has sparked growing interest in the emerging field of All-in-One image restoration, which aims to build unified models capable of handling a wide range of degradations with a single network [107, 138, 159, 236, 252, 189]. For instance, AirNet [107] introduces a degradation classifier trained via contrastive learning to guide restoration, while ADMS [159] employs a multi-type degradation classifier to dynamically select Adaptive Discriminant filters, enabling degradation-specific parameter modulation within the restoration network.

I.3 LLM Agents

Advancements in LLM-based frameworks have enabled more structured reasoning and agent designs, particularly for complex multimodal tasks. Initial efforts emphasized improving reasoning capabilities through refined prompting strategies and modular architecture. Chain-of-Thought (CoT) prompting [213] introduced stepwise reasoning, facilitating decomposition and interpretability across diverse tasks. ReAct [237] combined reasoning with tool interaction by interleaving thought traces and external actions, supporting more adaptive behavior. Extending this direction, CoALA [177] formalized components such as memory, reasoning, and control within a cognitive architecture, offering a modular design space for building general-purpose language agents. These developments established a basis for domain-specific agent systems with integrated reasoning pipelines. Building on these foundations, application-driven LLM agents have been developed to incorporate tool use and dynamic decision-making within specialized domains. In vision tasks, MMCTAgent [94], VideoAgent [197], and ReAgent-V [273] implement planning and evaluation pipelines for image and video analysis, incorporating external modules for retrieval and verification. In the medical domain, agents such as MedCoT [130], CLINICR [150], and MMedAgent-RL [224] employ hierarchical reasoning frameworks to address clinical questions, integrating structured logic and domain-specific knowledge to enhance interpretability and decision quality.

Similarly, LLM-based agents have also emerged as a promising paradigm for tackling complex image restoration tasks involving multiple degradations. RestoreAgent [22] pioneered the use of MLLMs for autonomous task identification, model selection, and execution planning. AgenticIR [274] introduced a five-stage human-inspired workflowPerception, Scheduling, Execution, Reflection, and Reschedulingaugmented with self-exploration to build IR-specific experience. MAIR [80] advanced this by employing a multi-agent system guided by real-world degradation priors, improving both efficiency and scalability. HybridAgent [104] proposed a hybrid interaction scheme with fast and slow agents, along with a mixed-distortion removal strategy to mitigate error propagation. Q-Agent [272] further introduces a quality-driven chain-of-thought framework, leveraging no-reference IQA metrics to guide greedy restoration without costly rollbacks. These works demonstrate the growing potential of combining general-purpose language intelligence with visual tools for robust, adaptive image restoration. More recently, JarvisIR [124] and JarvisArt [125] leveraged intelligent agent workflows to perform task-oriented image restoration and creative photo retouching.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction accurately reflect the papers contributions and scope.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: This paper discusses the limitations of the work performed by the authors in the Appendix.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [NA]

Justification: This paper does not include theoretical results.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: Information needed to reproduce the main experimental results of the paper is provided in the Appendix.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: Code, checkpoints, and datasets are available at https://github.com/taco-group/4KAgent.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so No is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: This paper provides these details in the Appendix.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [No]

Justification: Following common practice in the image restoration literature, we do not report error bars in this paper because of the heavy computation overheads.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).

- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: This paper provides sufficient information on the computer resources, as detailed in the Appendix, to reproduce the experiments.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics https://neurips.cc/public/EthicsGuidelines?

   Answer: [Yes]

   Justification: We follow the NeurIPS Code of Ethics.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: We discuss the societal impacts of our proposed method in the Appendix.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.

- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [Yes]

    Justification: We will require the users to adhere to usage guidelines for our released dataset.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: We properly cite the original assets in the paper.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.

- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: We release a new dataset for image restoration and 4K upscaling alongside documentation.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: This paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: This paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.

- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: [Yes]

    Justification: This paper describes the usage of LLM components in the proposed method in the paper.

    Guidelines:

    - The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy (https://neurips.cc/Conferences/2025/LLM) for what should or should not be described.