

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

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

<sup>1</sup>Texas A&M University <sup>2</sup>Stanford University <sup>3</sup>Snap Inc. <sup>4</sup>CU Boulder

<sup>5</sup>UT Austin <sup>6</sup>California Institute of Technology <sup>7</sup>Topaz Labs <sup>8</sup>UC Merced

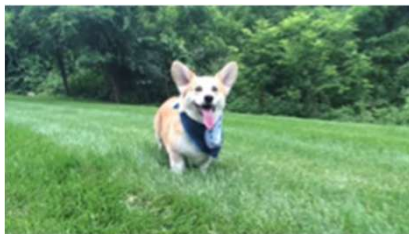
NeurIPS 2025

Presenter: Yu Cao

2026.4.12

## Background: Super Resolution

- SR Aims to Recover High-Resolution Images from Low-Resolution Inputs
- **Ill-Posed Problem**
- Real-World Degradations Amplify the Ill-Posed Nature of Super-Resolution
  - Hard for single model to cover the Vast Solution Space



# Background: SR with Specialist Models

GAN based

Diffusion based

Multi-Agent Methods

## • Real-ESRGAN

- Degradation model

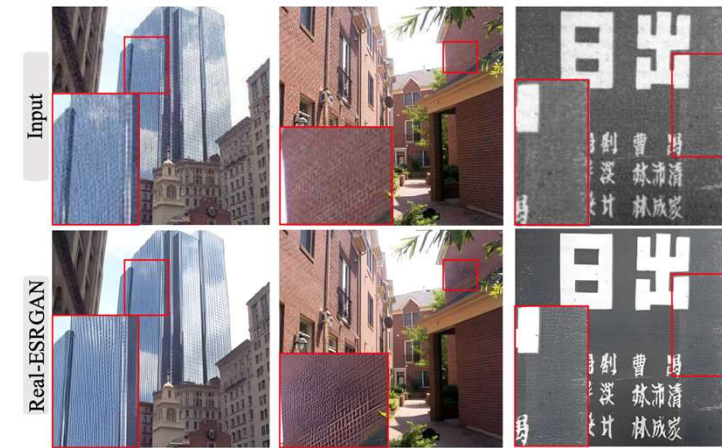
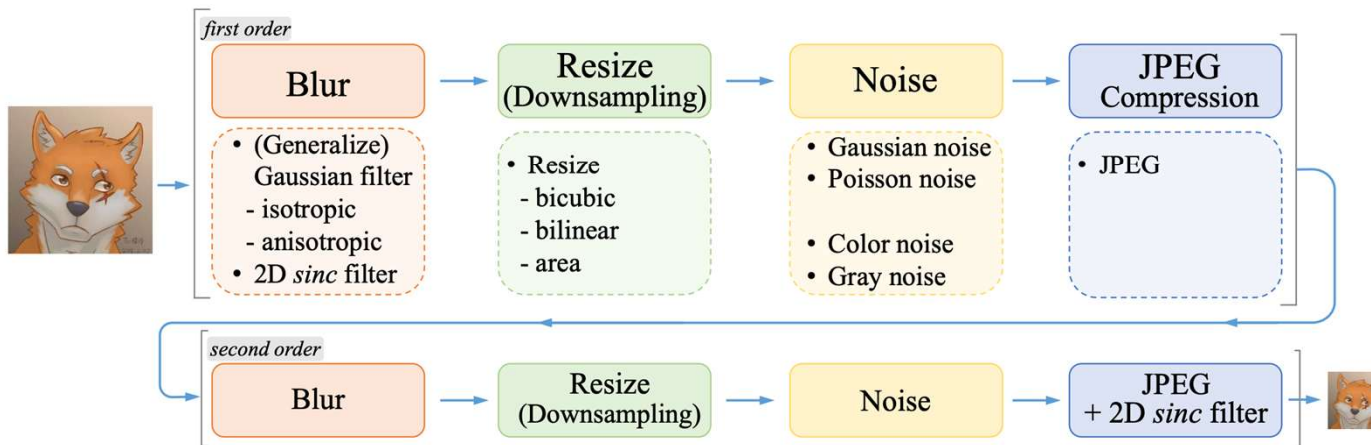


Figure 11: Limitations: 1) twisted lines; 2) unpleasant artifacts caused by GAN training; 3) unknown and out-of-distribution degradations. Zoom in for best view

## Background: SR with Specialist Models

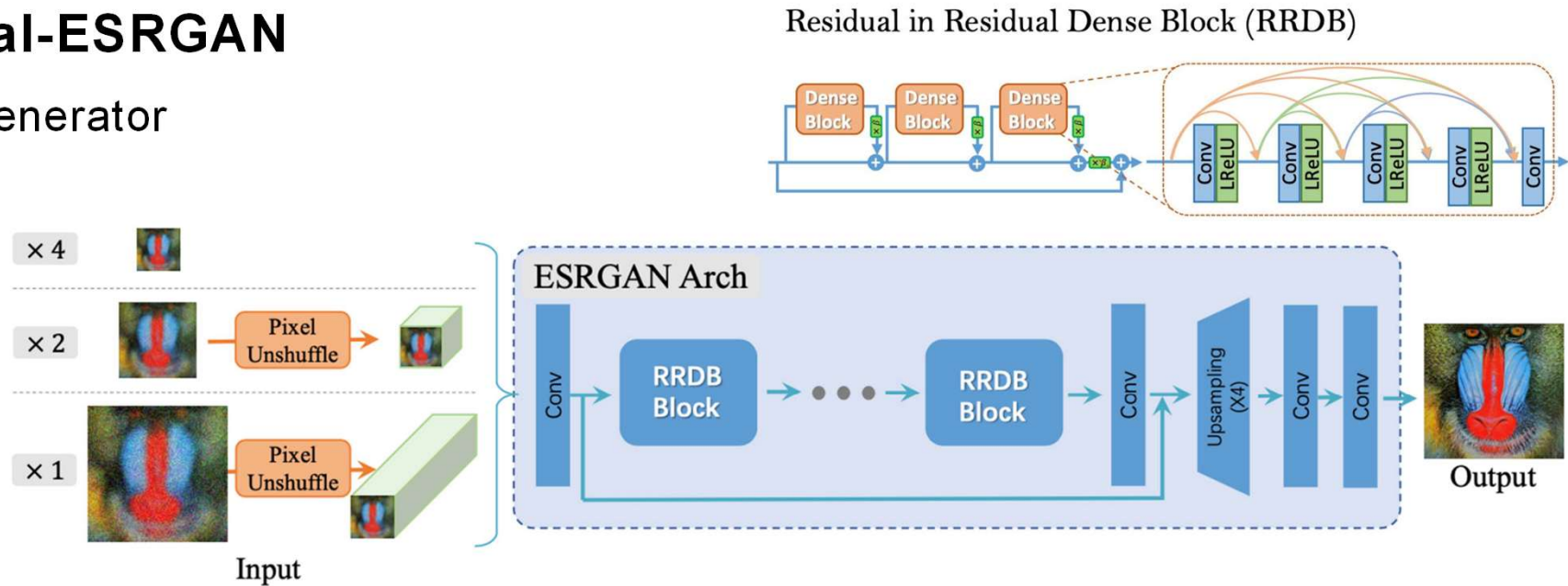
GAN based

Diffusion based

Multi-Agent Methods

- **Real-ESRGAN**

- Generator



- U-Net as Discriminator, with realness values for each pixel as outputs

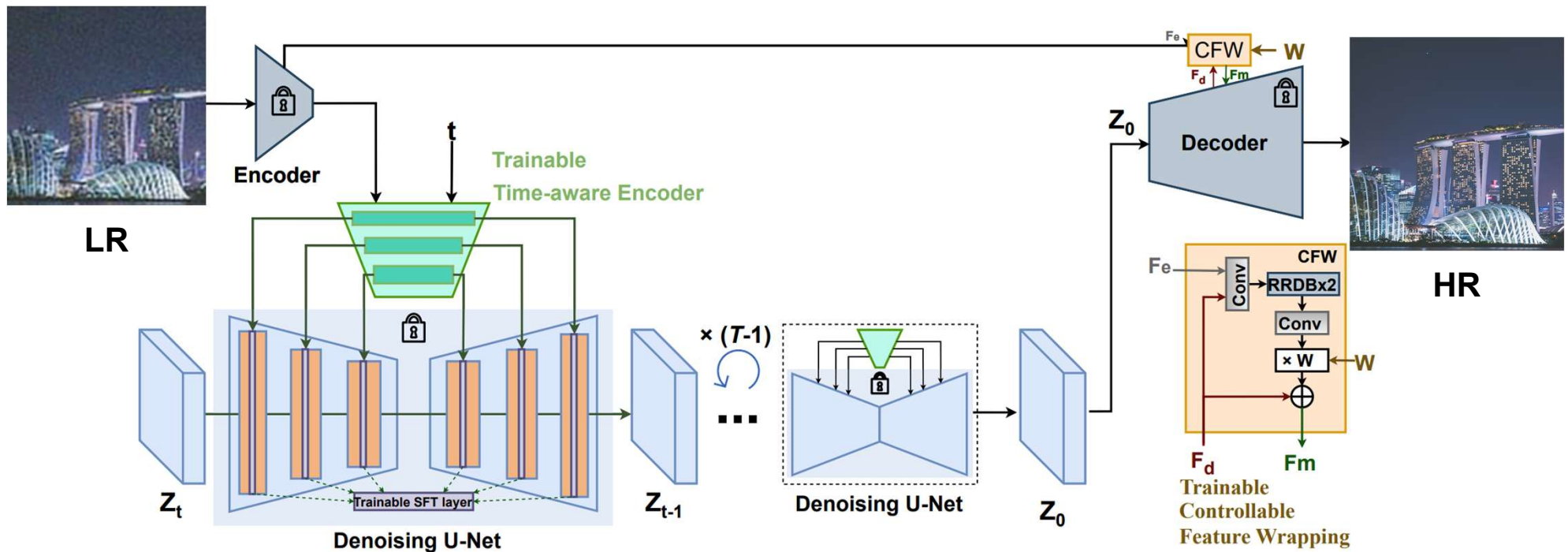
# Background: SR with Specialist Models

GAN based

Diffusion based

Multi-Agent Methods

- **StableSR** Leveraging diffusion prior in SR tasks



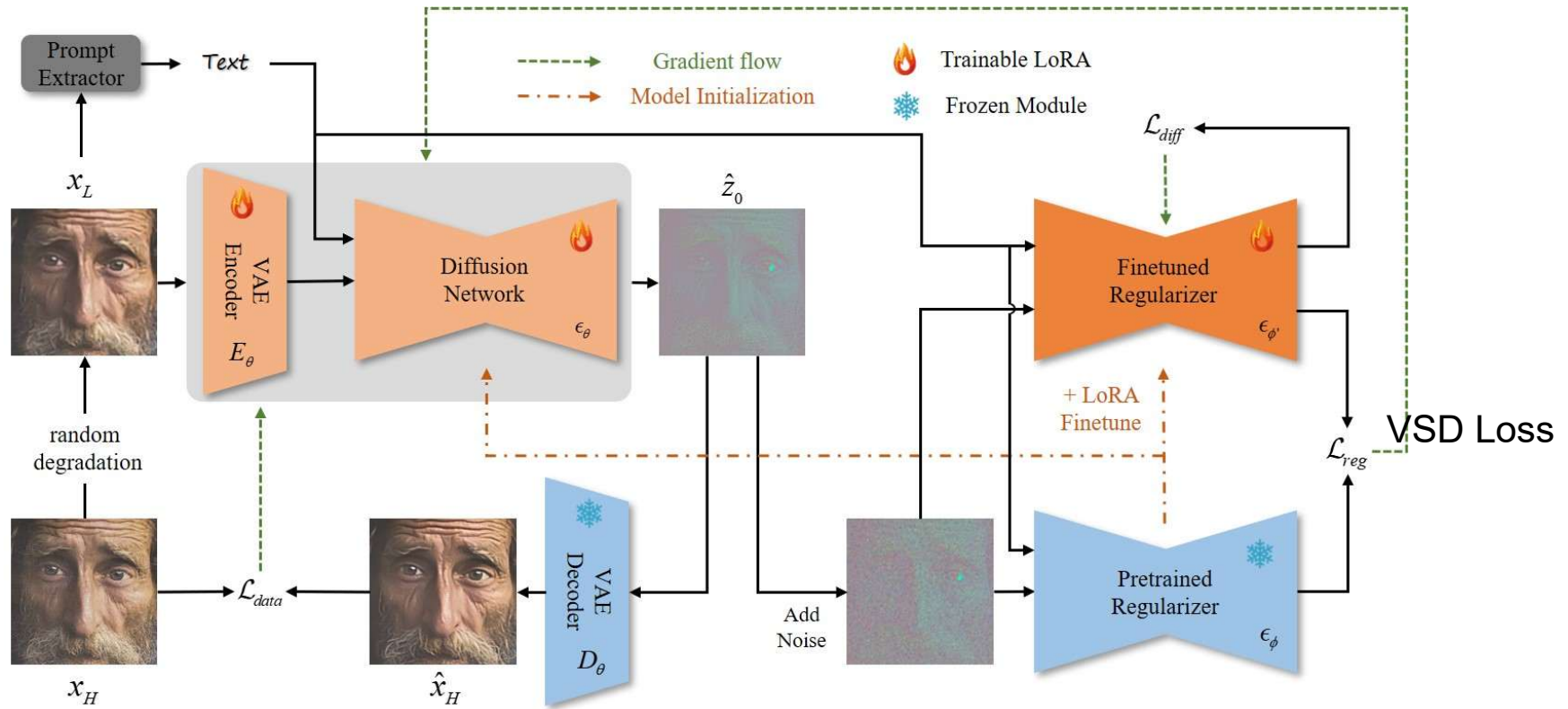
# Background: SR with Specialist Models

GAN based

Diffusion based

Multi-Agent Methods

- OSEDiff**



## Background: SR with Specialist Models

GAN based

Diffusion based

Multi-Agent Methods

- **Problems with Specialist Models**
  - Single model can only cover a small portion of solution space
  - Several generation problems remain poorly addressed
    - E.g. face restoration, text restoration



LR



SR Result



GT

# Background: Image Restoration with Multi-Agent

GAN based

Diffusion based

Multi-Agent Methods

- **RestoreAgent** Propose a framework using multi-agents for IR tasks

Why not all-in-one?

a1. Not truly "all": Models still fail on unseen degradation types



a2. Limited performance: Specialized models outperform generalists



a3. Single Task + All-in-One > All-in-One only



# Background: Image Restoration with Multi-Agent

GAN based

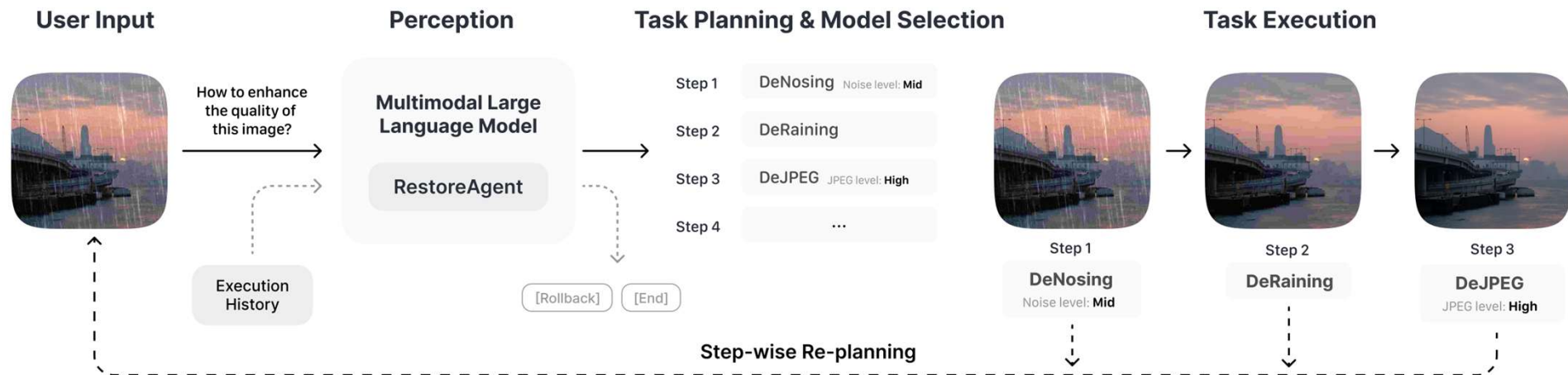
Diffusion based

Multi-Agent Methods

## • RestoreAgent

- Use LoRA to finetune Llava-Llama3-8b as VLM
- RestoreAgent identify degradation type and specify the task execution order

### RestoreAgent Pipeline



# Background: Image Restoration with Multi-Agent

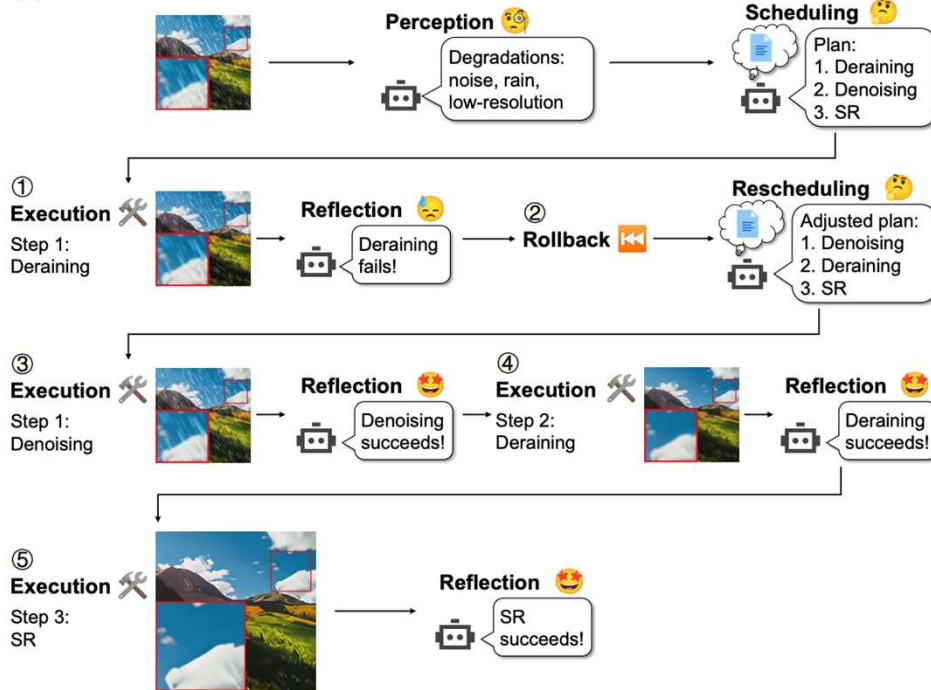
GAN based

Diffusion based

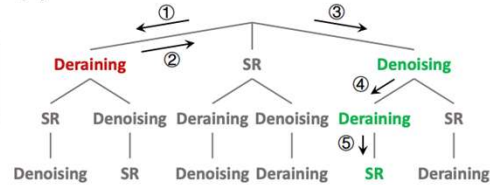
Multi-Agent Methods

- **AgenticIR** Introduce reflection module

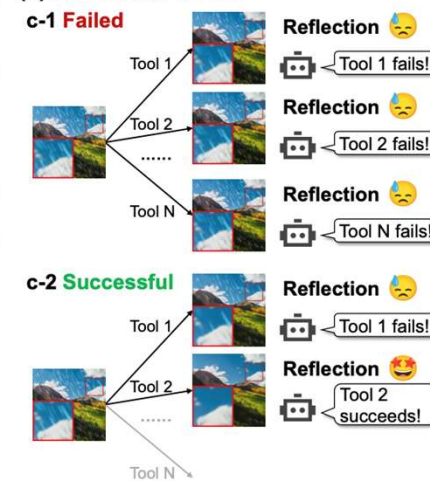
(a) Entire workflow



(b) Workflow in the view of tree search

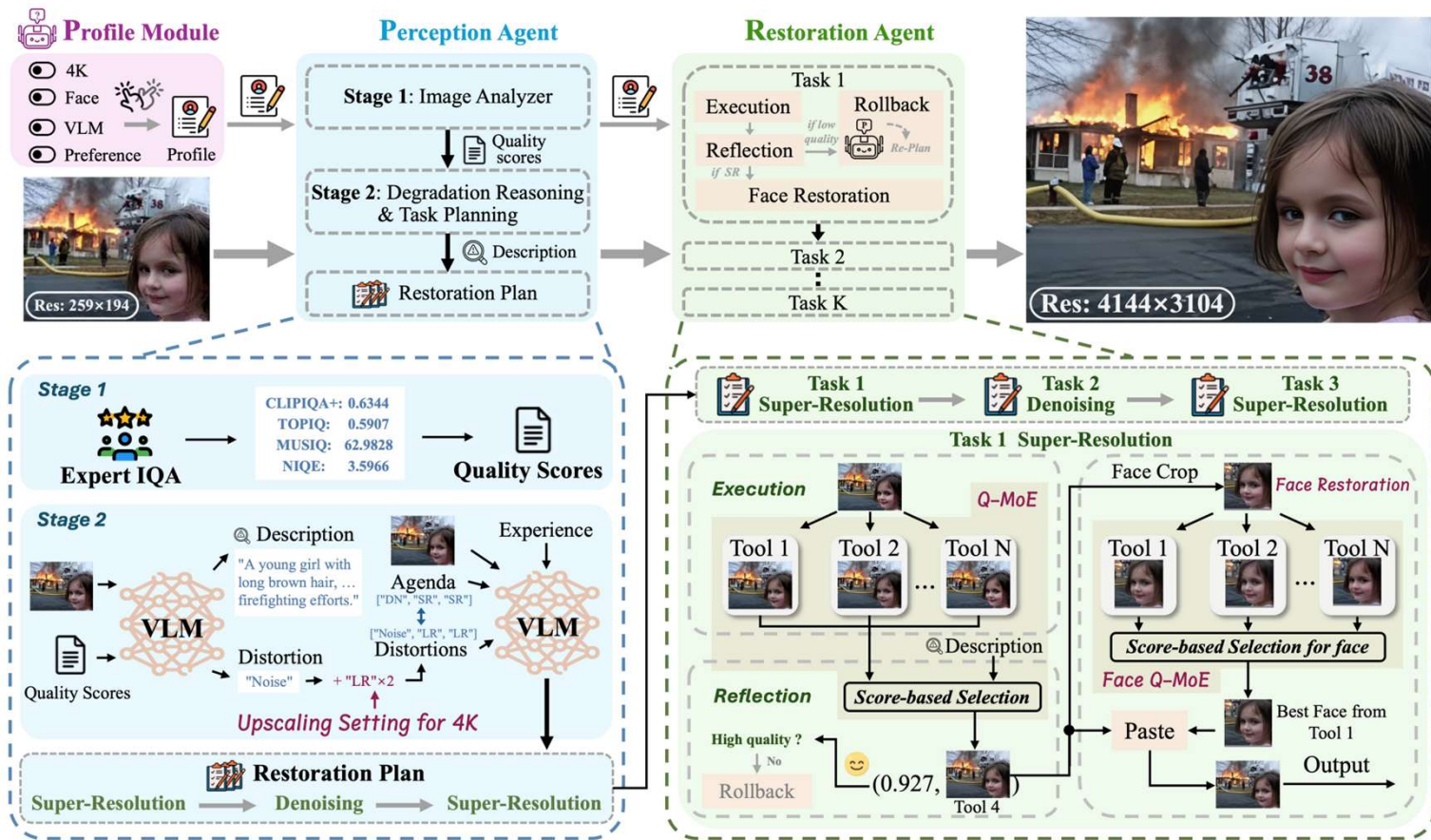


(c) Execution details



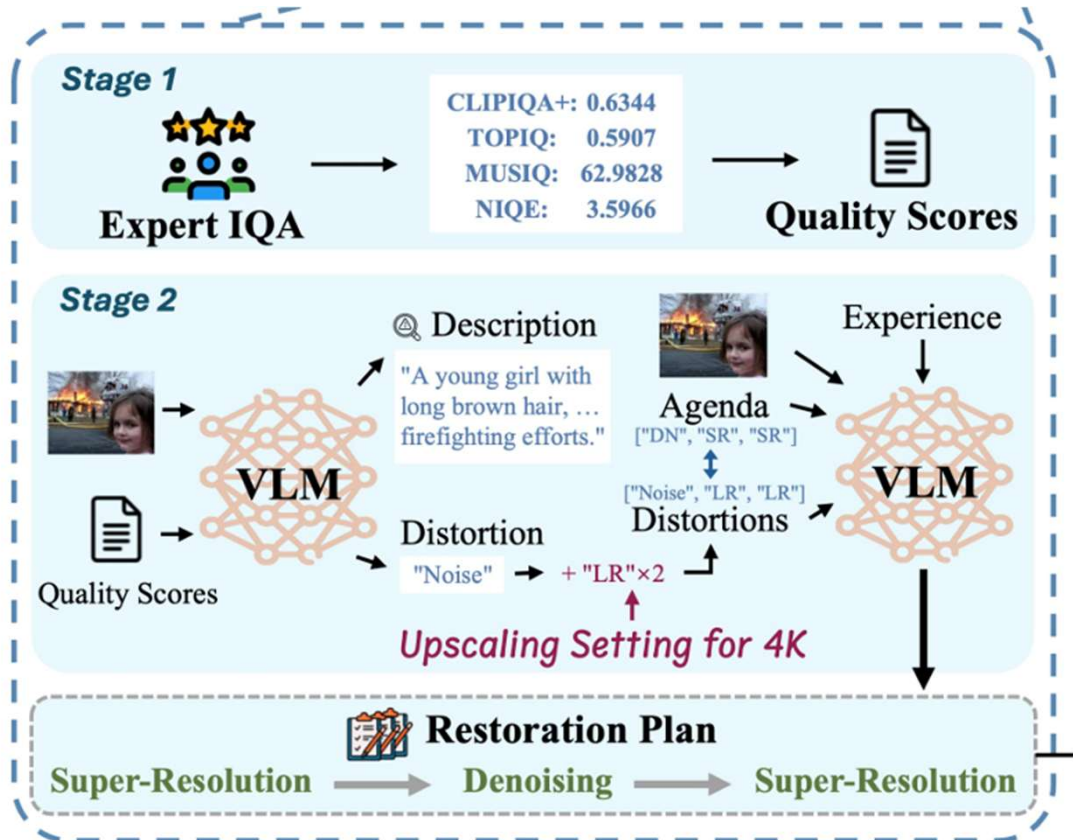
# Methods

- Three main parts: Perception, Restoration, Profile



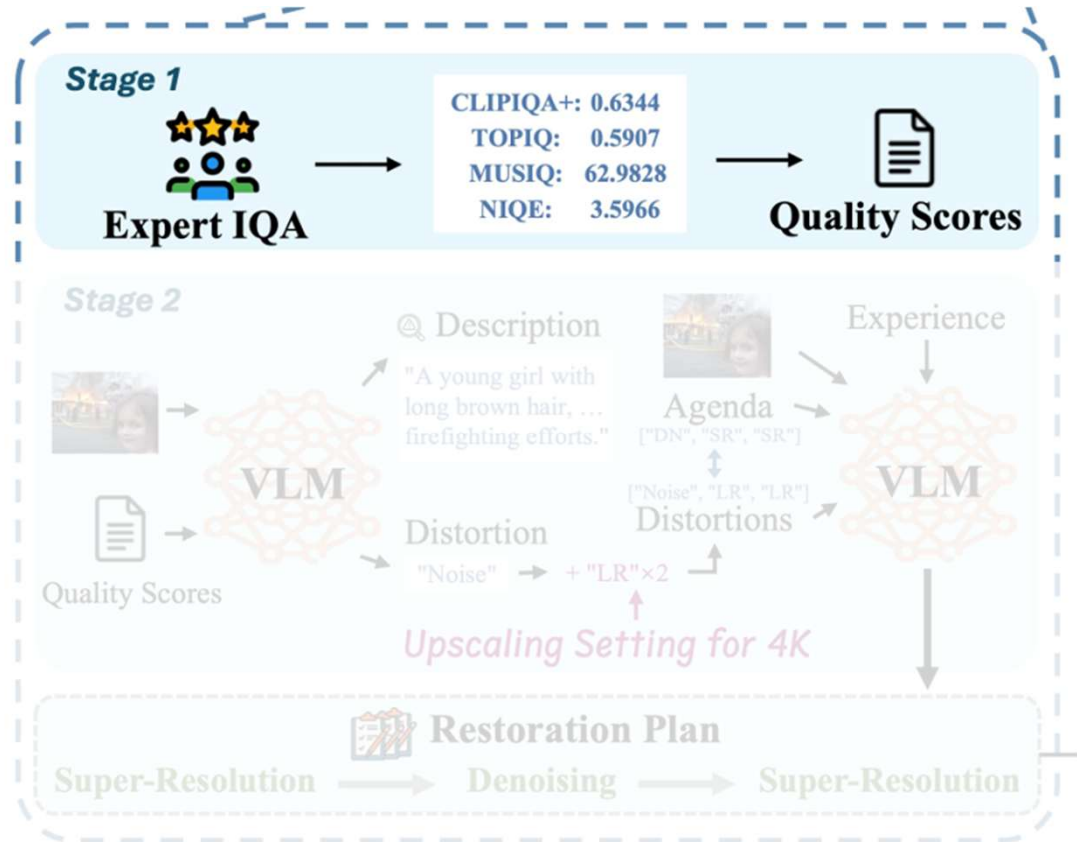
## Methods

- **Perception Agent**
  - **Image Analyser:** Use NR metrics
  - **Degradation Reasoning**
    - Image description  $C_I$
    - Degradation list  $D_I$
    - Initial restoration agenda  $A'_I$
- **Upscaling Factor Configuration**
  - Calculate scale factor and add to  $A'_I$
- **Task Planning**
  - Output initial restoration plan



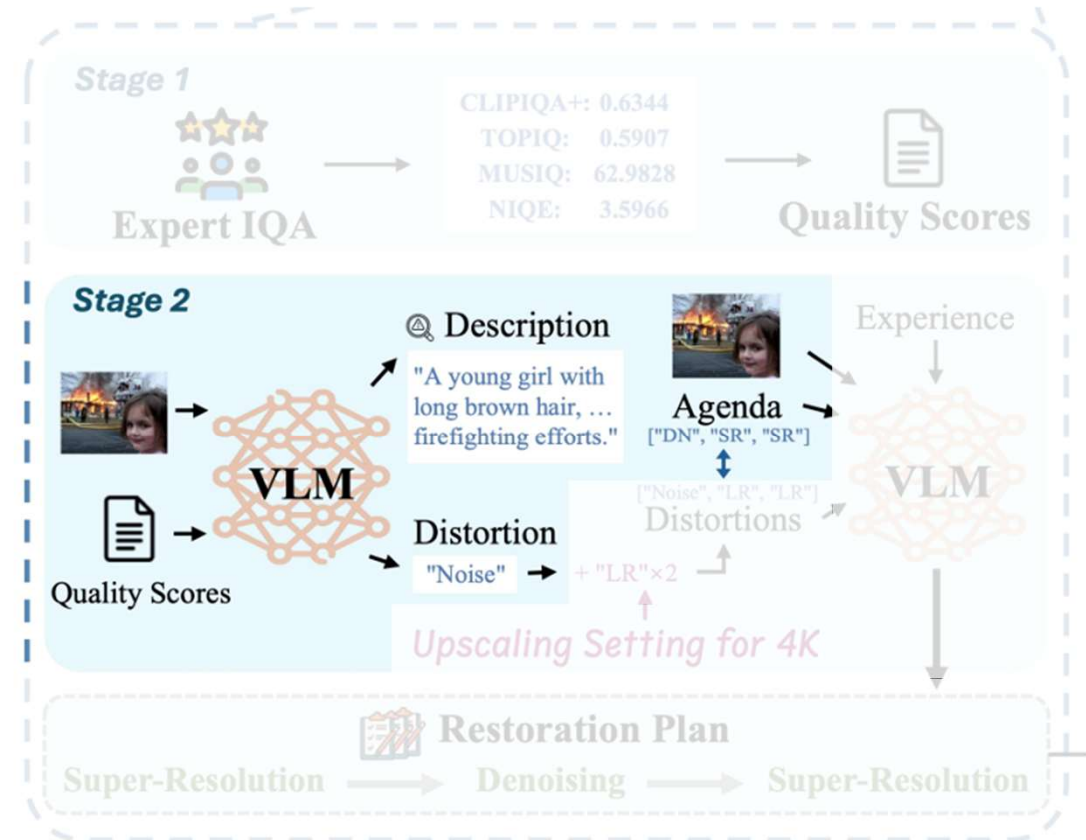
## Methods

- Perception Agent
  - Image Analyser: Use NR metrics



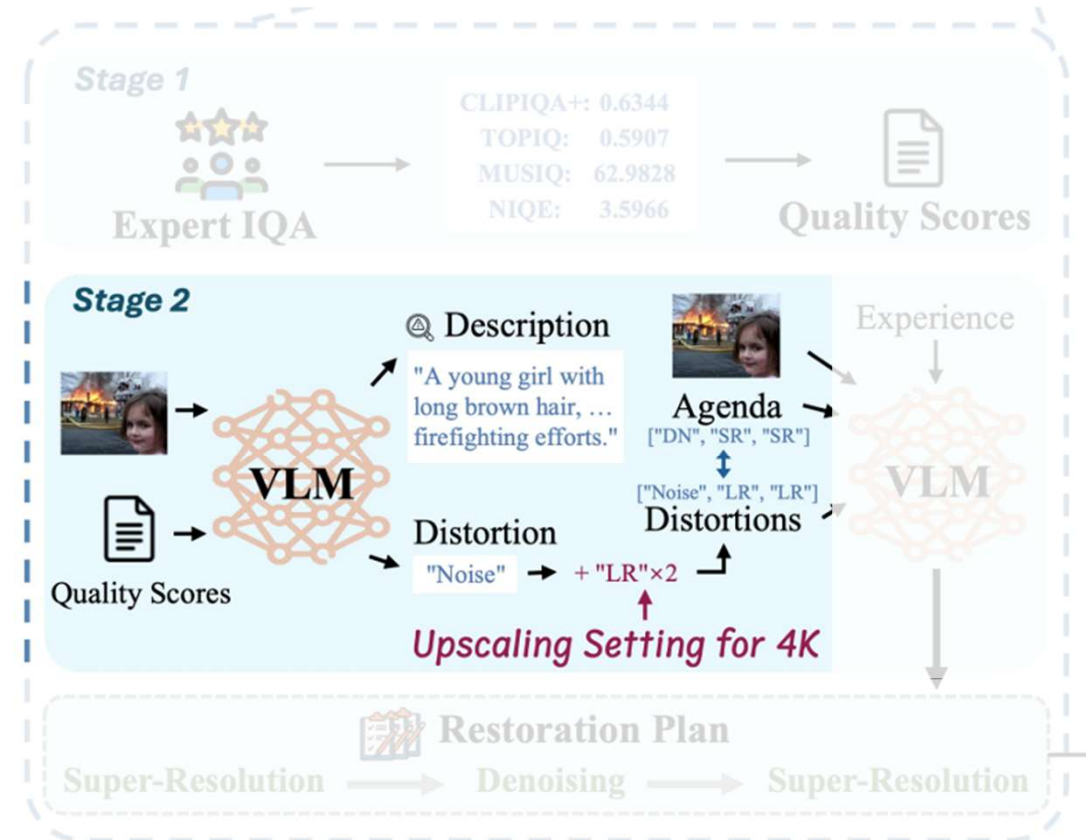
## Methods

- **Perception Agent**
  - **Image Analyser:** Use NR metrics
  - **Degradation Reasoning**
    - Image description  $C_I$
    - Degradation list  $D_I$
    - Initial restoration agenda  $A'_I$



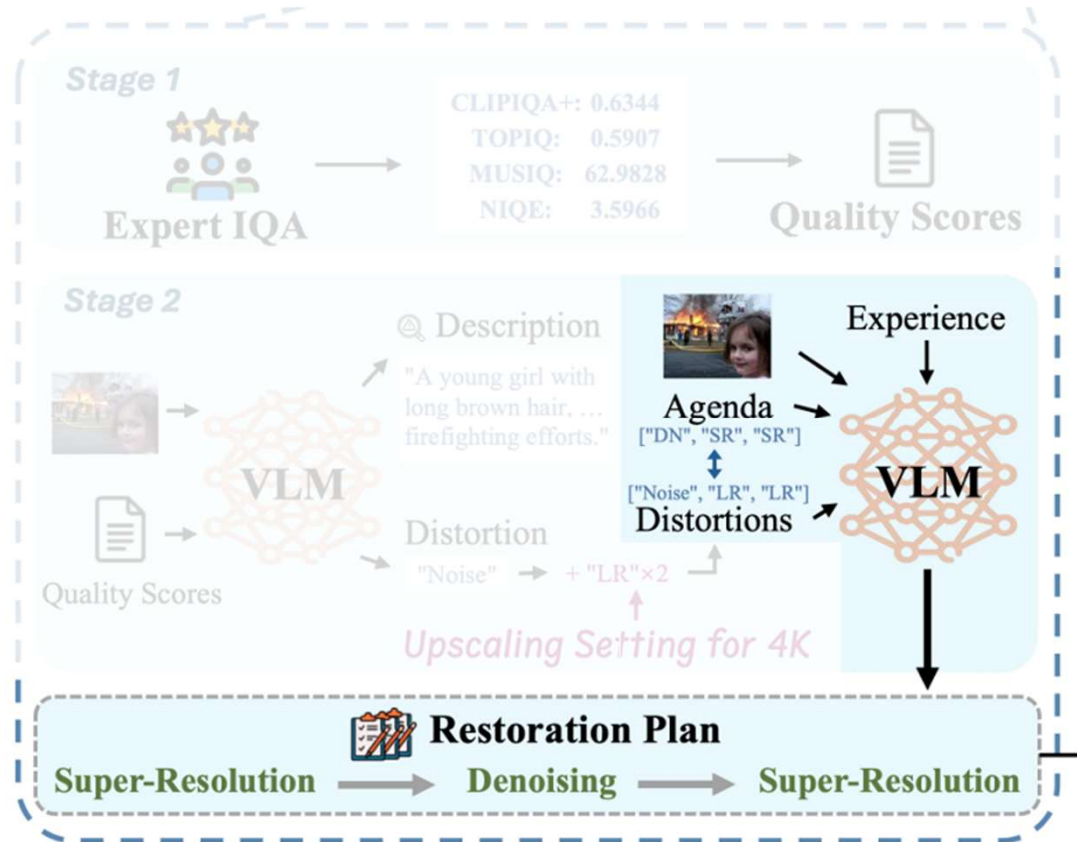
## Methods

- **Perception Agent**
  - **Image Analyser:** Use NR metrics
  - **Degradation Reasoning**
    - Image description  $C_I$
    - Degradation list  $D_I$
    - Initial restoration agenda  $A'_I$
- **Upscaling Factor Configuration**
  - Calculate scale factor and add to  $A'_I$



## Methods

- **Perception Agent**
  - **Image Analyser:** Use NR metrics
  - **Degradation Reasoning**
    - Image description  $C_I$
    - Degradation list  $D_I$
    - Initial restoration agenda  $A'_I$
  - **Upscaling Factor Configuration**
    - Calculate scale factor and add to  $A'_I$
  - **Task Planning**
    - Output initial restoration plan



## Methods

### • Restoration Agent

#### • Execution

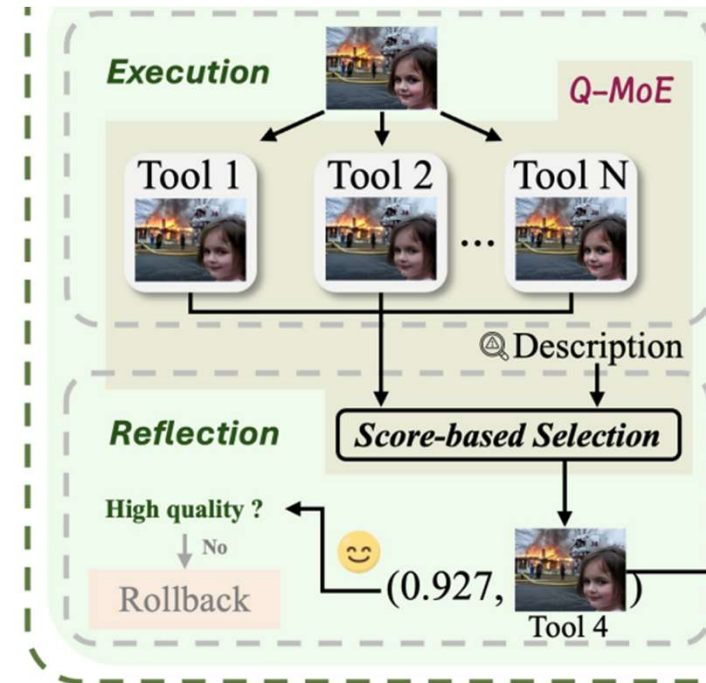
- Each tool (expert) in the toolbox produces a result

#### • Reflection

- HPS v2 (text alignment) + NR metrics
- $Q_s(T_i(I_{k-1})) = H(T_i(I_{k-1}, C_I)) + Q_{nr}(T_i(I_{k-1}))/4,$
- $Q_{nr}(T_i(I_{k-1})) = w_{NIQE} * (1 - Q_{NIQE}/10) + \sum_{j \in \Omega} w_j * Q_j.$
- Choose the result with highest quality

#### • Rollback

- Call perception agent when  $Q_s(I_k) \leq \eta$
- Assign a different restoration task



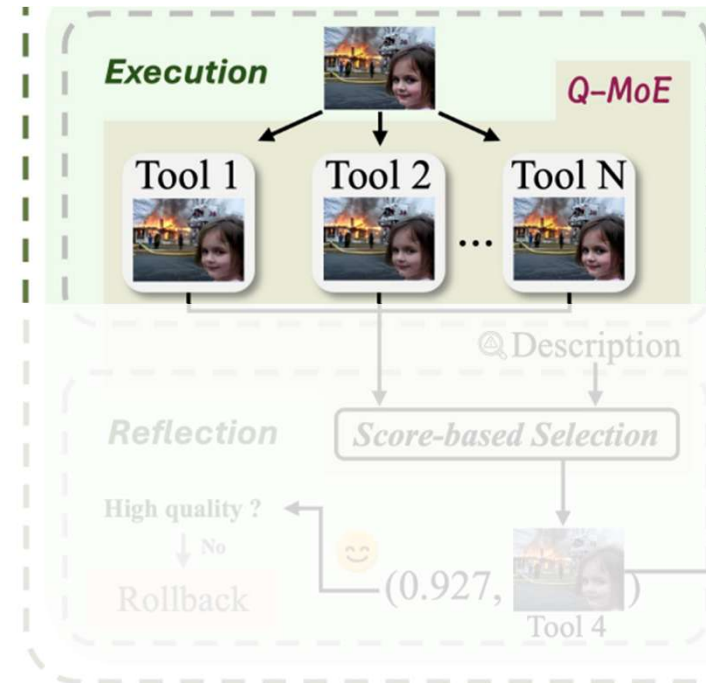
NIQE, MUSIQ, MANIQA, CLIPQA

## Methods

- **Restoration Agent**

- **Execution**

- Each tool (expert) in the toolbox produces a result



NIQE, MUSIQ, MANIQA, CLIPIQA

## Methods

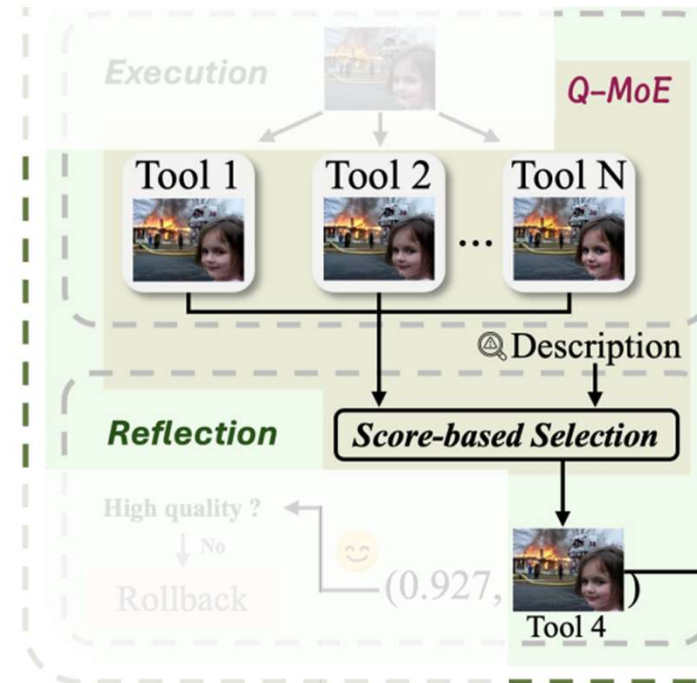
- **Restoration Agent**

- **Execution**

- Each tool (expert) in the toolbox produces a result

- **Reflection**

- HPS v2 (text alignment) + NR metrics
- $Q_s(T_i(I_{k-1})) = H(T_i(I_{k-1}, C_I)) + Q_{nr}(T_i(I_{k-1}))/4,$
- $Q_{nr}(T_i(I_{k-1})) = w_{NIQE} * (1 - Q_{NIQE}/10) + \sum_{j \in \Omega} w_j * Q_j.$
- Choose the result with highest quality



NIQE, MUSIQ, MANIQA, CLIPQA

## Methods

### • Restoration Agent

#### • Execution

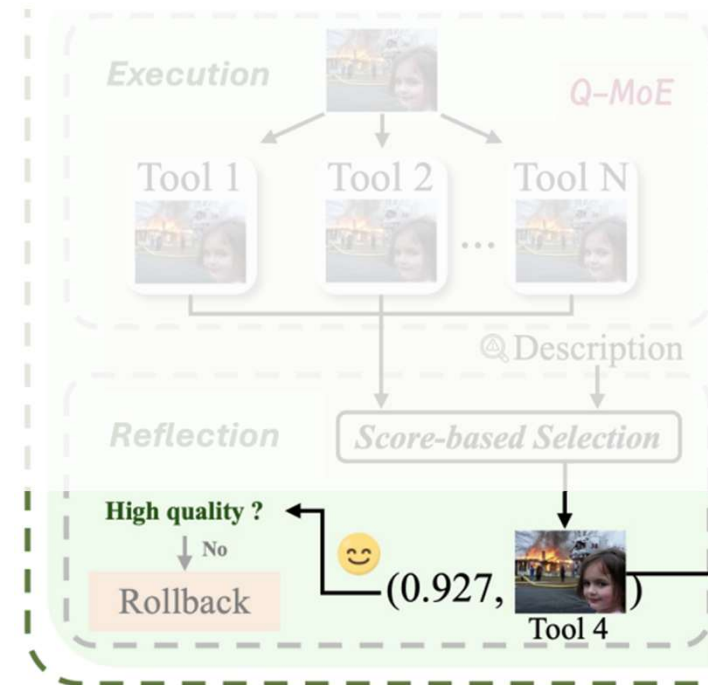
- Each tool (expert) in the toolbox produces a result

#### • Reflection

- HPS v2 (text alignment) + NR metrics
- $Q_s(T_i(I_{k-1})) = H(T_i(I_{k-1}, C_I)) + Q_{nr}(T_i(I_{k-1}))/4,$
- $Q_{nr}(T_i(I_{k-1})) = w_{NIQE} * (1 - Q_{NIQE}/10) + \sum_{j \in \Omega} w_j * Q_j.$
- Choose the result with highest quality

#### • Rollback

- Call perception agent when  $Q_s(I_k) \leq \eta$
- Assign a different restoration task



NIQE, MUSIQ, MANIQA, CLIPQA

## Methods

### • Face Restoration Pipeline

- Only invoked after SR step
- Clip all the faces in the input image
- Use face restoration expert modules
- Q-MoE Decision

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

- IP: Cosine similarity of ArcFace feature
- CF: CLIB-FIQA (Face quality assessment)
- Pick the best result and paste back

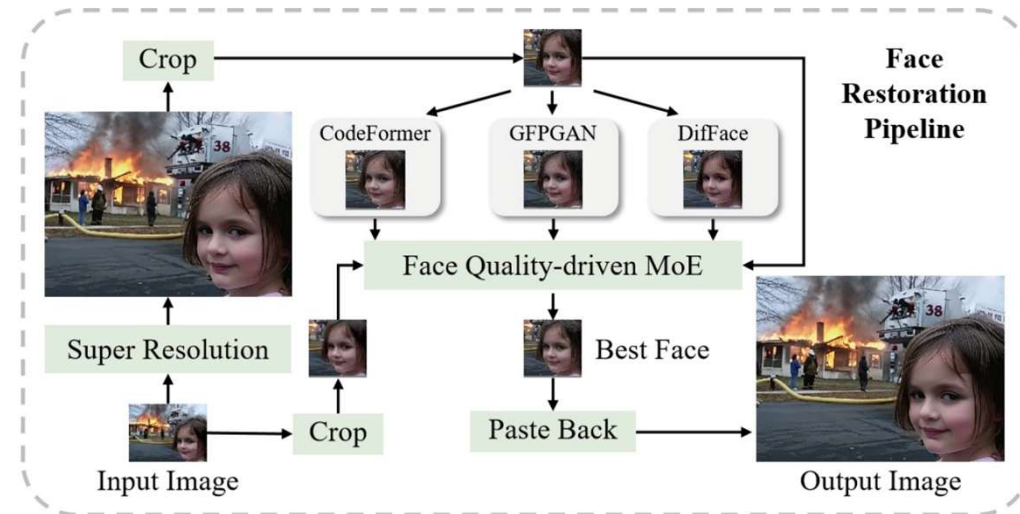


Figure 3: Face restoration pipeline overview.

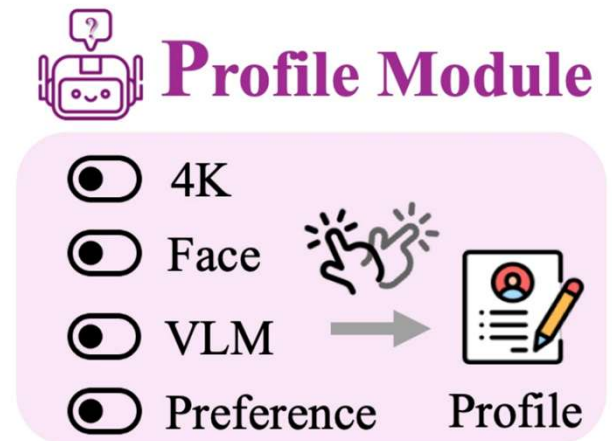
## Methods

- **Profile Module**

- Acts like a system prompt for LLM applications
- 7 Parameters:

- |                     |                       |
|---------------------|-----------------------|
| 1. Perception Agent | 5. Face Restore       |
| 2. Upscale to 4K    | 6. Brightening        |
| 3. Scale Factor     | 7. Restore Preference |
| 4. Restore Option   |                       |

- Each distinct scenario can be addressed by selecting a proper configuration profile
- Adjust a variety of restoration domains



## Model Card

- Model Zoo** Divided into Fidelity-based & Perception-based Methods

<p><b>Brightening</b></p> <p>CLAHE [271]            Constant Shift (C=40)            DiffPlugin [129]            FourierDiff [134]            Gamma Correction (<math>\gamma = 2/3</math>)            MAXIM [183]</p>	<p><b>Defocus Deblurring</b></p> <p>ConvIR [39]            DiffPlugin [129]            DRBNet [165]            IFAN [98]            LaKDNet [164]            Restormer [241]</p>	<p><b>Dehazing</b></p> <p>DehazeFormer [172]            DiffPlugin [129]            MAXIM [183]            RIDCP [213]            X-Restormer [28]</p>	<p><b>Super Resolution</b></p> <p>DiffBIR [121]            DRCT [68]            HAT-L [29]            HAT-GAN [29]            HMA [34]            OSEDiff [211]            PiSA-SR [176]            SwinIR [117]            SwinIR (Real-ISR) [117]            SwinFIR [248]            X-Restormer [28]</p>
<p><b>Denoising</b></p> <p>MAXIM [183]            MPRNet [243]            NAFNet [25]            Restormer [241]            X-Restormer [28]            SwinIR [117]</p>	<p><b>Motion Deblurring</b></p> <p>EVSSM [88]            LaKDNet [164]            MAXIM [183]            MPRNet [243]            NAFNet [25]            Restormer [241]            X-Restormer [28]</p>	<p><b>Deraining</b></p> <p>DiffPlugin [129]            MAXIM [183]            MPRNet [243]            Restormer [241]            X-Restormer [28]</p>	<p><b>JPEG CAR</b></p> <p>FBCNN [77] (QF=5)            FBCNN [77] (QF=90)            FBCNN [77] (BQF)            SwinIR [117] (QF=40)</p>
		<p><b>Face Restoration</b></p> <p>GFPGAN [196]            CodeFormer [266]            DifFace [239]</p>	

# Experiments

## • Overview

- A broad range, including 4k & 16k

Task	Dataset	#Test Images
<b>Classical SR (§5.1)</b>	Set5 [13]	5
	Set14 [244]	14
	B100 [137]	100
	Urban100 [70]	100
	Manga109 [139]	109
<b>Real-World SR (§5.2)</b>	RealSR [17]	93
	DrealSR [209]	100
<b>Multiple-Degradation IR (§5.3)</b>	MiO-Group A [269]	640
	MiO-Group B [269]	400
	MiO-Group C [269]	400
<b>Face Restoration (§5.4)</b>	WebPhoto-Test [196]	407
<b>Large Scale Factor SR (§6.1)</b>	RealSRSet [251]	20

<b>Joint Image Restoration + 4K Upscaling (§6.2)</b>	DIV4K-50 (Ours)	50
<b>AI-Generated Content 4K SR (§7.1)</b>	GenAIBench-4K [101]	100
	DiffusionDB-4K [206]	100
<b>Remote Sensing SR (§8.1)</b>	AID [218]	135
	DIOR [109]	154
	DOTA [217]	183
	WorldStrat [37]	100
<b>Fluorescence Microscopy Image SR (§8.2)</b>	SR-CACO-2 [12]	300
<b>Pathology Image SR (§8.3)</b>	bcSR [76]	200
<b>Medical Image SR (§8.4)</b>	Chest X-ray 2017 [84]	624
	Chest X-ray 14 [195]	880
	US-Case [173]	111
	MMUS1K [152]	100
	DRIVE [174]	20

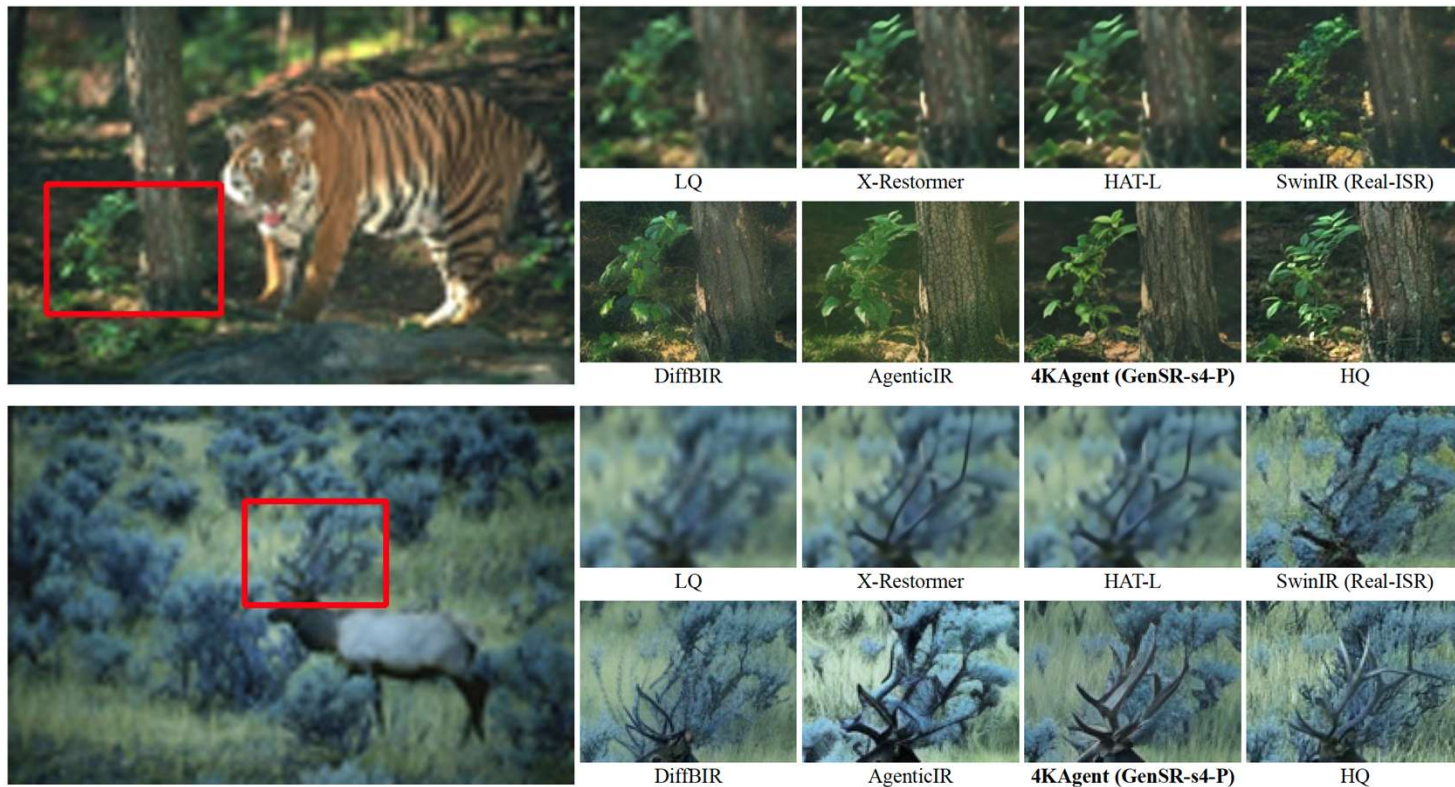
## Experiments

- **Classical Image Super-Resolution (4x)**
  - ExpSR: Only use tools in SR toolbox, GenSR: use all tools

Dataset	Method	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	DISTS $\downarrow$	FID $\downarrow$	NIQE $\downarrow$	CLIPQA $\uparrow$	MUSIQ $\uparrow$	MANIQA $\uparrow$	
Urban100	<i>Fidelity based method</i>										
	SwinIR [117]	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]	<b>28.78</b>	<u>0.8492</u>	0.1623	<b>0.1388</b>	<u>2.92</u>	<u>5.45</u>	<b>0.5271</b>	70.48	0.6778	
	HAT-L [29]	28.58	<u>0.8495</u>	<u>0.1598</u>	0.1411	<b>2.87</b>	5.55	0.5054	70.62	<u>0.6866</u>	
	HMA [34]	<u>28.69</u>	<b>0.8511</b>	<b>0.1583</b>	<i>0.1405</i>	2.93	5.61	<i>0.5084</i>	<b>70.75</b>	<b>0.6893</b>	
	<b>4KAgent (ExpSR-s4-F)</b>	28.59	0.8479	<i>0.1599</i>	<u>0.1399</u>	2.97	<b>5.31</b>	<u>0.5235</u>	<u>70.70</u>	<i>0.6833</i>	
	<i>Perception based method</i>										
	SwinIR (Real-ISR) [117]	<b>23.24</b>	<b>0.7184</b>	<u>0.1908</u>	<u>0.1365</u>	<b>25.36</b>	<b>4.29</b>	0.6169	71.99	0.6578	
	DiffBIR [121]	22.51	0.6397	0.2011	0.1395	26.10	4.79	<b>0.7185</b>	<u>73.10</u>	<i>0.6956</i>	
	OSDiff [211]	21.88	0.6572	0.2185	0.1479	38.13	4.67	0.6593	72.35	0.6822	
	PiSA-SR [176]	22.36	<u>0.6704</u>	<b>0.1823</b>	<b>0.1297</b>	28.51	<u>4.43</u>	<i>0.6814</i>	72.93	<b>0.7020</b>	
	<b>4KAgent (ExpSR-s4-P)</b>	<u>22.56</u>	<i>0.6582</i>	<i>0.1955</i>	<i>0.1378</i>	<u>25.55</u>	4.53	<u>0.7092</u>	<b>73.65</b>	<u>0.6981</u>	
	<i>Agentic System</i>										
	AgenticIR [269]	22.03	<b>0.6615</b>	0.2147	0.1507	<b>31.09</b>	4.65	0.6790	73.10	0.6873	
	<b>4KAgent (GenSR-s4-P)</b>	<b>22.27</b>	0.6545	<b>0.2073</b>	<b>0.1444</b>	32.29	<b>4.43</b>	<b>0.7001</b>	<b>73.57</b>	<b>0.6961</b>	

## Experiments

- **Classical Image Super-Resolution (4x)**
  - Better visual quality & More natural details



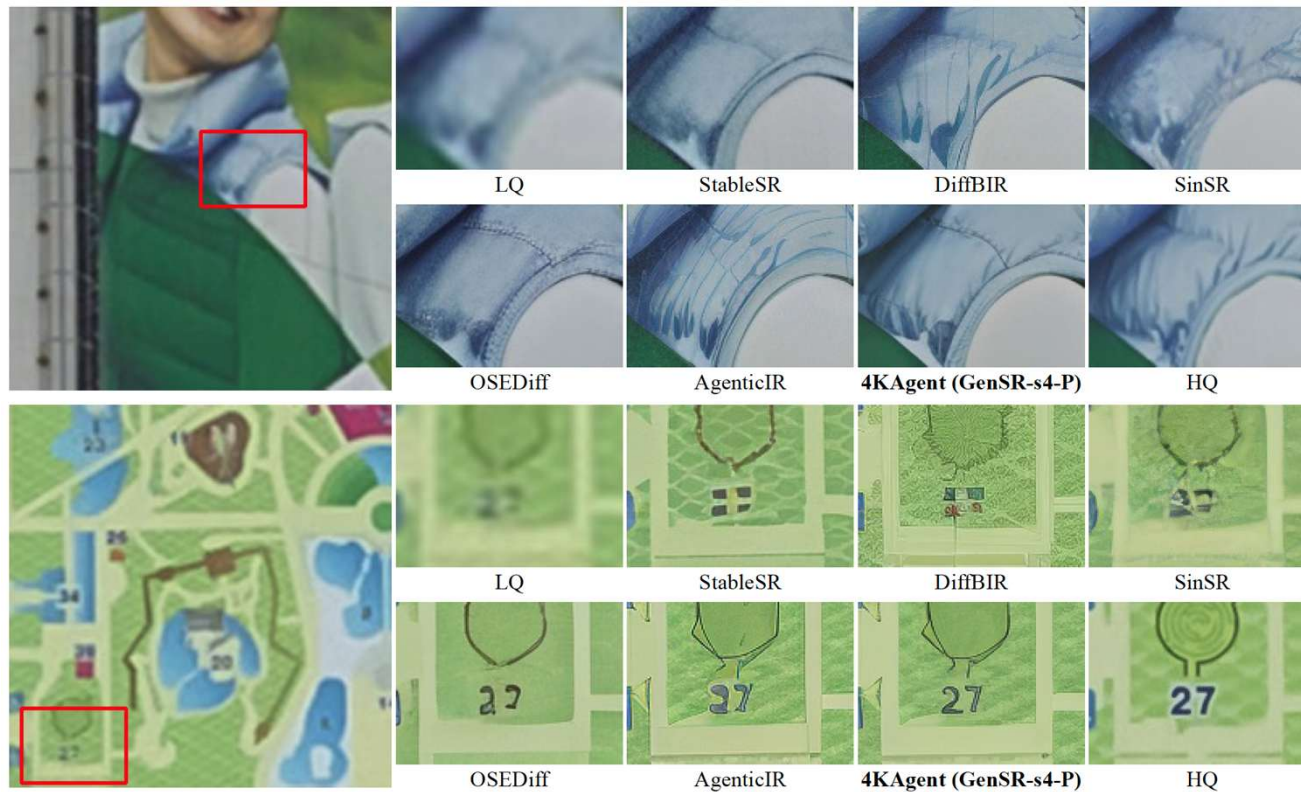
## Experiments

- Real-World Image Super-Resolution (4x)
  - Better results on NR metrics

Dataset	Method	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	DISTS $\downarrow$	FID $\downarrow$	NIQE $\downarrow$	CLIPQA $\uparrow$	MUSIQ $\uparrow$	MANIQA $\uparrow$
RealSR	ResShift [240]	<b>26.31</b>	<u>0.7411</u>	0.3489	0.2498	142.81	7.27	0.5450	58.10	0.5305
	StableSR [192]	24.69	0.7052	0.3091	<i>0.2167</i>	127.20	5.76	0.6195	65.42	0.6211
	DiffBIR [121]	24.88	0.6673	0.3567	0.2290	124.56	5.63	0.6412	64.66	0.6231
	PASD [229]	25.22	0.6809	0.3392	0.2259	<b>123.08</b>	<i>5.18</i>	0.6502	68.74	0.6461
	SeeSR [212]	25.33	0.7273	<i>0.2985</i>	0.2213	125.66	5.38	0.6594	69.37	0.6439
	SinSR [201]	<u>26.30</u>	<i>0.7354</i>	0.3212	0.2346	137.05	6.31	0.6204	60.41	0.5389
	OSDiff [211]	25.15	0.7341	<u>0.2921</u>	<u>0.2128</u>	<u>123.50</u>	5.65	<i>0.6693</i>	69.09	0.6339
	PiSA-SR [176]	25.50	<b>0.7417</b>	<b>0.2672</b>	<b>0.2044</b>	<i>124.09</i>	5.50	<u>0.6702</u>	<i>70.15</i>	<i>0.6560</i>
	AgenticIR [269]	22.45	0.6447	0.3745	0.2503	140.38	5.81	0.6506	65.87	0.6210
	<b>4KAgent (ExpSR-s4-P)</b>	24.60	0.6839	0.3253	0.2292	127.64	<u>5.09</u>	<b>0.7078</b>	<u>70.97</u>	<b>0.6602</b>
<b>4KAgent (GenSR-s4-P)</b>	22.55	0.6557	0.3509	0.2468	134.63	<b>4.78</b>	0.6666	<b>71.77</b>	<u>0.6564</u>	

## Experiments

- **Real-World Image Super-Resolution (4x)**
  - Better reconstruction on characters & texture details



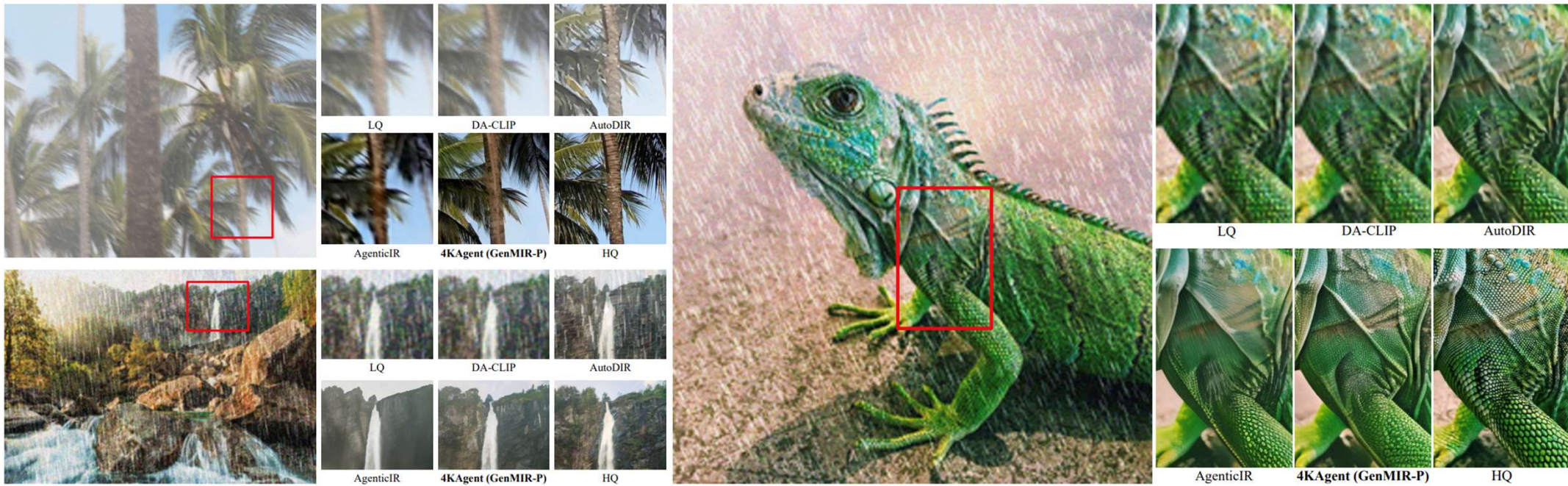
## Experiments

- Multiple-Degradation Image Restoration

Degradations	Method	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	MANIQA $\uparrow$	CLIQQA $\uparrow$	MUSIQ $\uparrow$
Group A	AirNet [105]	19.13	0.6019	0.4283	0.2581	0.3930	42.46
	PromptIR [160]	20.06	0.6088	0.4127	0.2633	0.4013	42.62
	MiOIR [89]	20.84	0.6558	0.3715	0.2451	0.3933	47.82
	DA-CLIP [133]	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 [80]	19.64	0.6286	0.3967	0.2500	0.3767	47.01
	AgenticIR [269]	<u>21.04</u>	<b>0.6818</b>	<i>0.3148</i>	<i>0.3071</i>	<i>0.4474</i>	56.88
	MAIR [79]	<u>21.02</u>	<i>0.6715</i>	<b>0.2963</b>	<u>0.3330</u>	<u>0.4751</u>	<u>59.19</u>
<b>4KAgent (GenMIR-P)</b>	<b>21.48</b>	<u>0.6720</u>	<u>0.3019</u>	<b>0.3748</b>	<b>0.5544</b>	<b>63.19</b>	

# Experiments

- Multiple-Degradation Image Restoration



# Experiments

- Face Restoration

Dataset	Method	NIQE↓	CLIPQA↑	MUSIQ↑	MANIQA↑	CLIB-FIQA↑	DSL-FIQA↑
WebPhoto-Test	GFPGAN [196]	5.12	0.6792	<u>74.21</u>	0.6379	0.6590	<b>0.7732</b>
	CodeFormer [266]	4.58	<u>0.6884</u>	73.87	<u>0.6415</u>	<b>0.6840</b>	0.7435
	DifFace [239]	<u>4.20</u>	0.5831	65.31	0.5891	0.6511	0.6189
	AgenticIR [269]	6.85	0.5731	56.25	0.5465	0.5978	0.5289
	<b>4KAgent (GenSRFR-s4-P)</b>	<b>4.15</b>	<b>0.7077</b>	<b>75.92</b>	<b>0.6576</b>	<u>0.6671</u>	<u>0.7683</u>



LQ



GFPGAN



CodeFormer



DifFace



AgenticIR



**4KAgent (GenSRFR-s4-P)**

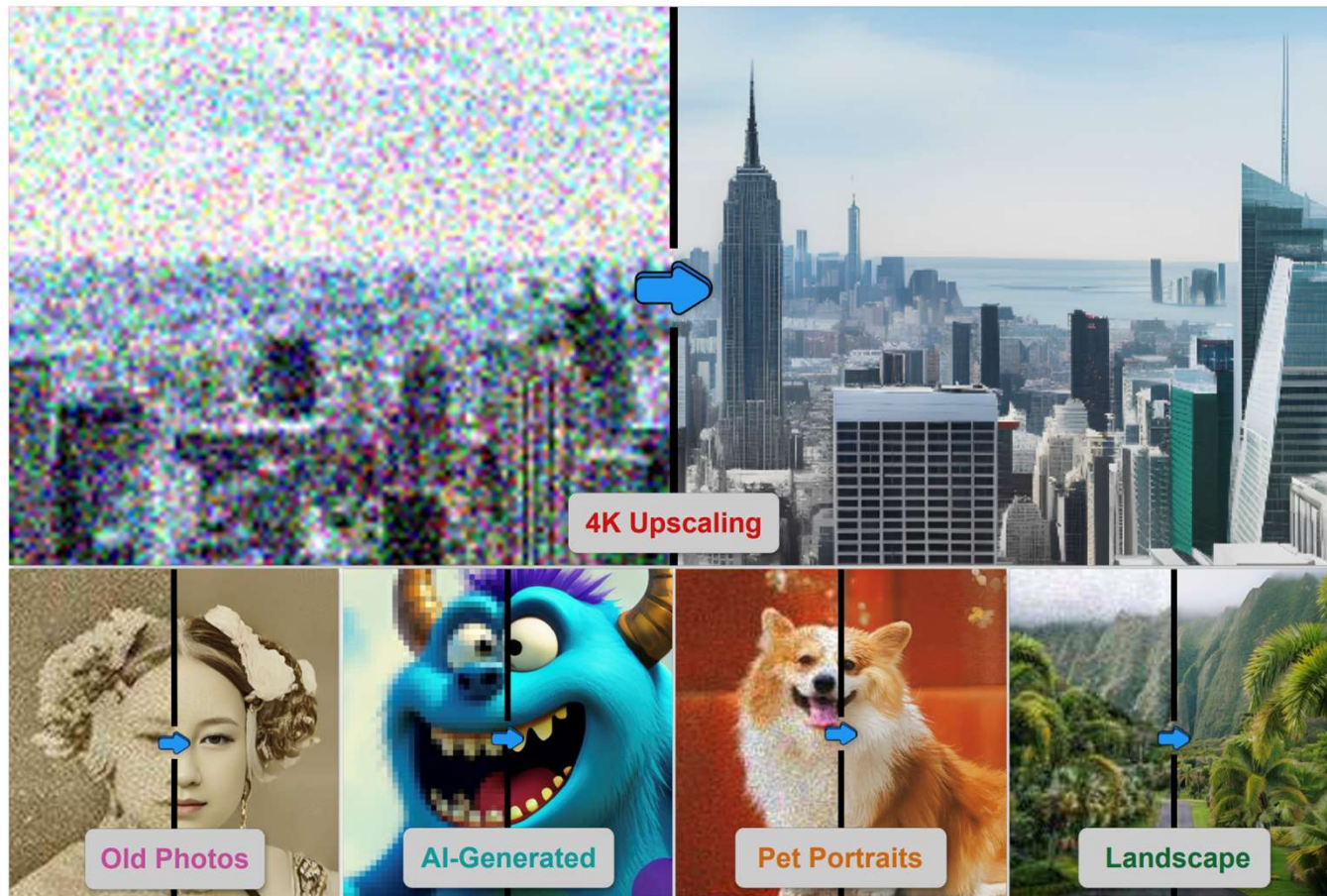
## Experiments

- 16x Super Resolution

Dataset	Method	NIQE↓	CLIPQA↑	MUSIQ↑	MANIQA↑
RealSRSet	HAT-L [29] (4× → 4×)	10.59	0.3885	25.06	0.3060
	DiffBIR [121] (4× → 4×)	3.63	<u>0.7867</u>	44.86	<u>0.6076</u>
	DiffBIR [121] (16×)	<b>2.80</b>	0.7583	47.54	<b>0.6099</b>
	OSEDiff [211] (4× → 4×)	5.40	0.7665	<u>48.42</u>	0.5362
	OSEDiff [211] (16×)	4.66	0.6483	35.33	0.4581
	PiSA-SR [176] (4× → 4×)	5.70	<b>0.7883</b>	48.20	0.5464
	PiSA-SR [176] (16×)	4.88	0.6384	35.90	0.4128
	AgenticIR [269]	4.86	0.6775	44.71	0.5236
	<b>4KAgent (Gen4K-P)</b>	<u>3.53</u>	<u>0.7794</u>	<b>50.84</b>	<u>0.5913</u>

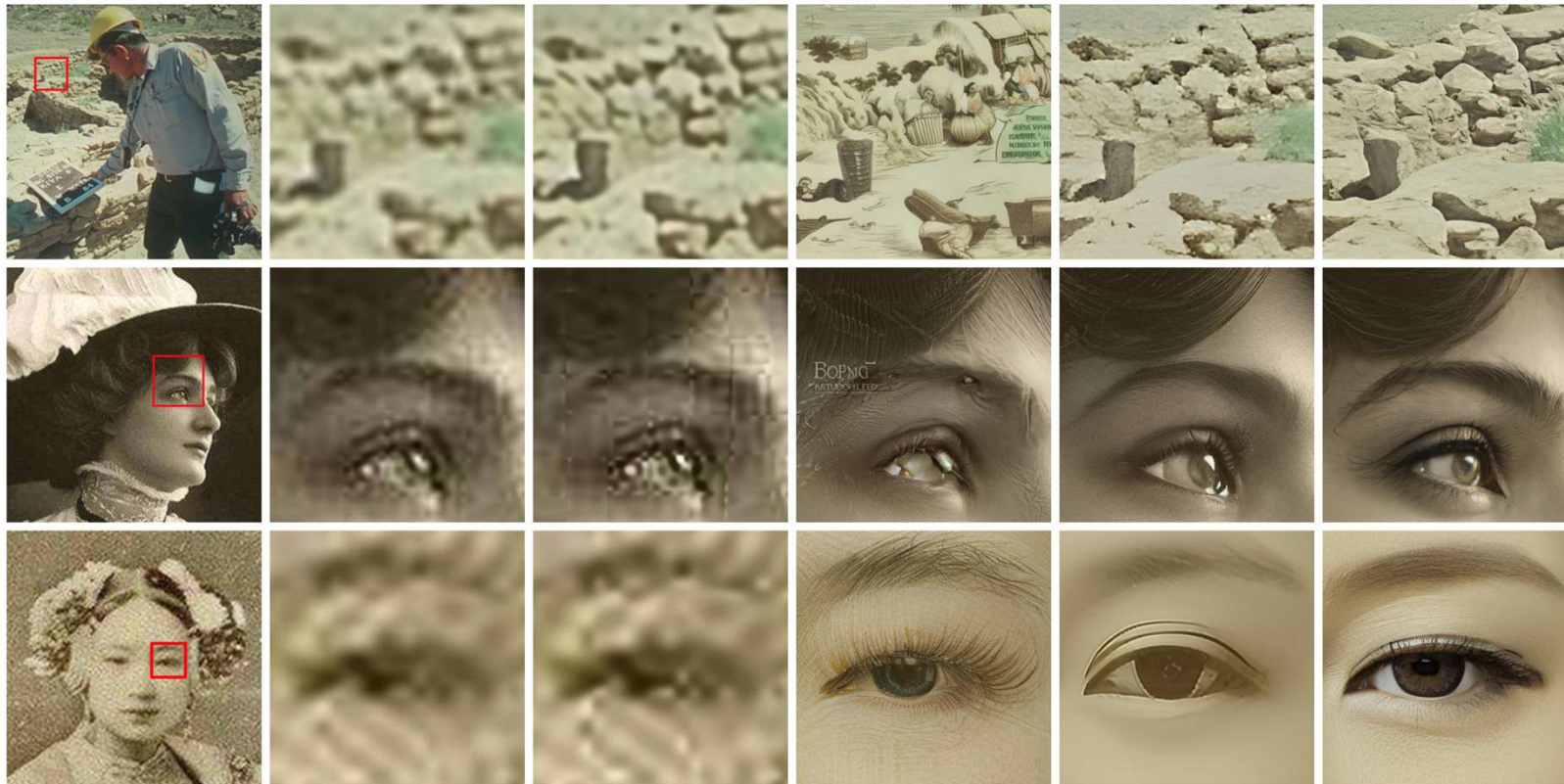
## Experiments

- 16x Super Resolution



# Experiments

- 16x Super Resolution



LQ

HAT-L (4x → 4x)

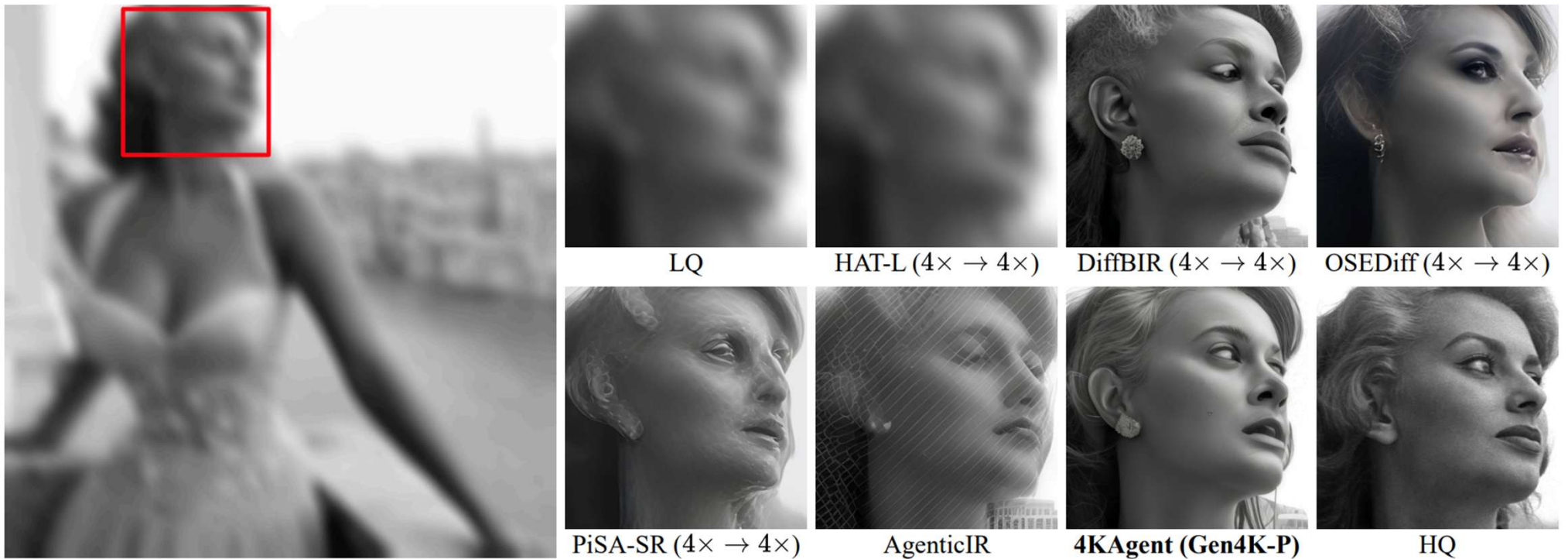
DiffBIR (16x)

AgenticIR

4KAgent (Gen4K-P)

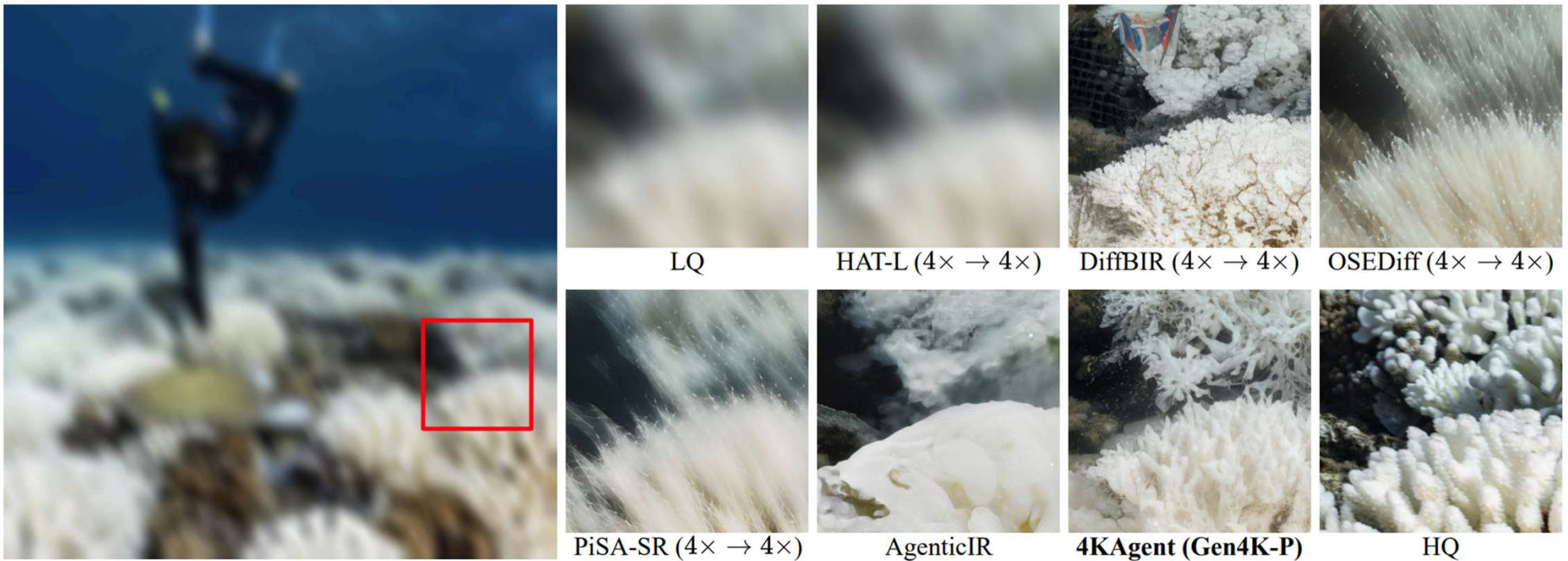
## Experiments

- 16x Super Resolution



## Experiments

- 16x Super Resolution



# Experiments

- AIGC 4k Super Resolution



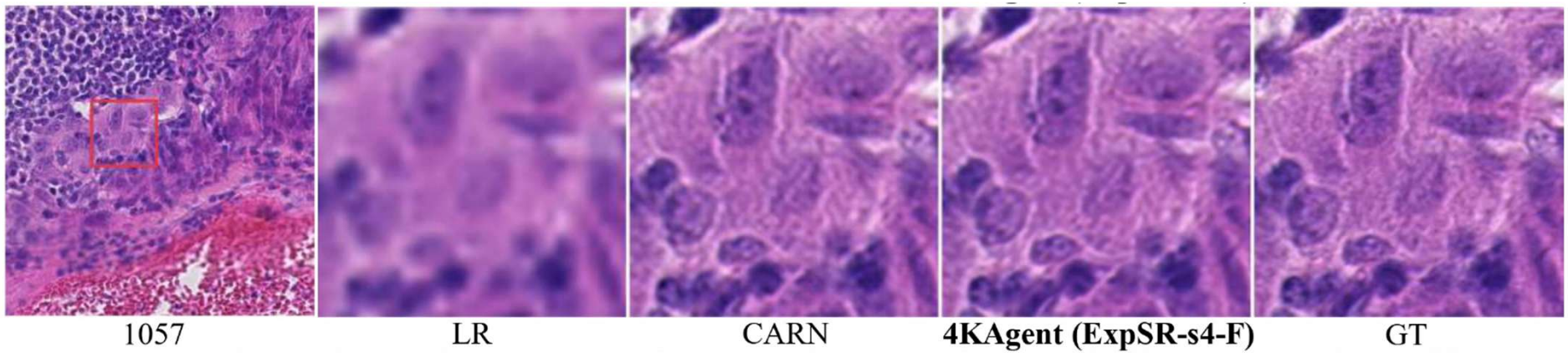
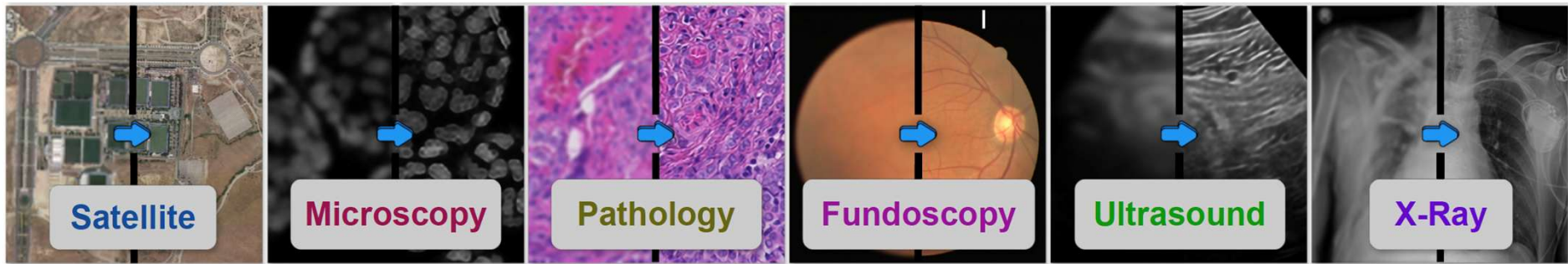
An action photograph of a cat casting fur out of its mouth.

A magician pulling a rabbit out of a hat.



## Experiments

- Restoration on Scientific Image



## Ablation Study

- Ablation of Q-MoE

Dataset	Method	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	MANIQA $\uparrow$	CLIPQA $\uparrow$	MUSIQ $\uparrow$
MiO100 - Group C	4KAgent (DFS)	<b>19.81</b>	<b>0.5785</b>	0.4381	0.3286	0.4854	54.03
	4KAgent ( <b>Q-MoE</b> )	19.77	0.5629	<b>0.4271</b>	<b>0.3545</b>	<b>0.5233</b>	<b>55.56</b>



LQ



4KAgent (DFS)



4KAgent (Q-MoE)



HQ

# Ablation Study

- Ablation of Face Restoration Pipeline

Method	Restore Option	Face Restore	NIQE↓	CLIQQA↑	MUSIQ↑	MANIQA↑	CLIB-FIQA↑	DSL-FIQA↑
4KAgent (ExpSR-s4-P)	super-resolution	False	5.11	<b>0.7119</b>	<u>73.62</u>	<b>0.6601</b>	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	<b>4.15</b>	<u>0.7077</u>	<b>75.92</b>	<u>0.6576</u>	<b>0.6671</b>	<b>0.7683</b>



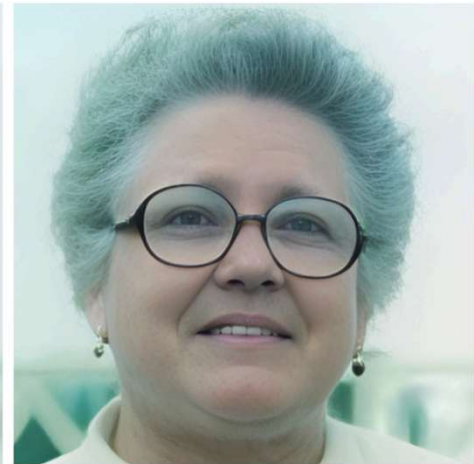
LQ



4KAgent (ExpSR-s4-P)



4KAgent (ExpSRFR-s4-P)



4KAgent (GenSRFR-s4-P)

## Running Time Analysis

- Different configurations lead to various runtime
- Generate a 4k image needs ~1500s (25 min)

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

## Running Time Analysis

- Fast 4K mode (in rebuttal): Disable long-running methods in the toolbox after achieving resolution threshold  $s$

Method	NIQE ↓	CLIQQA ↑	MUSIQ ↑	MANIQA ↑	Avg. Running Time (s)
DiffBIR [1] (16x)	<b>2.65</b>	0.7078	38.59	0.5858	976.2
OSDiff [2] (16x)	8.37	0.5680	25.07	0.4210	106.6
PiSA-SR [3] (16x)	9.30	0.5549	24.51	0.3861	107.8
<b>4KAgent</b> (s=1024)	3.51	0.7310	<b>44.51</b>	0.5689	582.7
<b>4KAgent</b> (s=4096)	3.15	<b>0.7585</b>	44.16	<b>0.5928</b>	1551.8

## Further Discussion

- **Boundary artifacts after copying and pasting face?**
  - Use the toolbox in GFPGAN (CVPR 2021) to extract face and paste back
  - Low occurrence frequency

Method	GFPGAN	CodeFormer	DifFace	4KAgent
Avg. boundary artifacts	9.75 / 100	8.75 / 100	25.75 / 100	7.75 / 100

- Also, the IQA result w & w/o face restoration further validates superior visual quality.

## Further Discussion

- **Winning rate of Q-MoE**
  - Different SR models win from different domains

Task	Benchmark (# number of images)	Leading SR Tools (# number of selection)
Real-World image SR (x4)	RealSR (100)	DiffBIR (55), PiSA-SR (22), OSEDIff (16)
AIGC SR (x4)	GenAIBench (100)	DiffBIR (61), OSEDIff (25), PiSA-SR (14)
Remote Sensing SR (x4)	DOTA (183)	HMA (79), HAT-L (54), PiSA-SR(15)
Fluorescence Microscopic Image SR (x4)	SR-CACO-2 (300)	DRCT (238), HMA (23), HAT-L (17)
Pathology Image SR (x4)	bcSR (200)	X-Restormer (84), HAT-L (64), HMA (30)
Medical Image SR (x4)	Chest X-ray 2017 (624)	HMA (312), DRCT (171), X-Restormer (73)

## Takeaways

- **Multi-agent collaboration is a better solution paradigm**
  - A single expert model struggles to cover the vast solution space of real-world super-resolution
  - Introducing VLM priors and the Q-MoE decision mechanism improves performance
- **Still significant room for improvement in inference efficiency and real-time performance**
  - One-step 16x upscaling has not yet been achieved
  - Edge artifacts of face restoration still need to be improved

**Thanks for listening!**

## Backups

- **Why conducting twice super-resolution for 16x SR?**
  - SOTA SR methods can not achieve 16x resolution or perform limited

Method	NIQE ↓	CLIPQA ↑	MUSIQ ↑	MANIQA ↑
OSDiff [4] (4x → 4x)	4.88	0.7201	39.88	0.5482
OSDiff [4] (16x)	8.37	0.5680	25.07	0.4210
PiSA-SR [5] (4x → 4x)	5.01	0.7141	38.22	0.5364
PiSA-SR [5] (16x)	9.30	0.5549	24.51	0.3861
<b>4KAgent</b> (Direct 16x SR)	8.91	0.4780	22.78	0.3939
<b>4KAgent</b>	<b>3.15</b>	<b>0.7585</b>	<b>44.16</b>	<b>0.5928</b>

## Backups

- **Detailed restoration plan and agenda**

- Building: Denoising@SwinIR( $\sigma=50$ ) → Super-resolution@DiffBIR → Super-resolution@SwinIR (Real-ISR).
- Face: Super-resolution@OSEDiff → Denoising@Restormer → JPEG Compression Artifact Removal@FBCNN (BQF) → Defocus Deblurring@DiffPlugin → Super-resolution@DiffBIR.