Robust-R1: Degradation-Aware Reasoning for Robust Visual Understanding

Jiaqi Tang[1*], Jianmin Chen[2*], Wei Wei[2†], Xiaogang Xu[3], Runtao Liu[1], Xiangyu Wu[4], Qipeng Xie[1], Jiafei Wu[5], Lei Zhang[2], Qifeng Chen[1†]

[1]Hong Kong University of Science and Technology
[2]Northwestern Polytechnical University
[3]Chinese University of Hong Kong
[4]Nanjing University of Science and Technology
[5]University of Hong Kong
cqf@ust.hk, weiweinwpu@nwpu.edu.cn

Abstract

Multimodal Large Language Models struggle to maintain reliable performance under extreme real-world visual degradations, which impede their practical robustness. Existing robust MLLMs predominantly rely on implicit training/adaptation that focuses solely on visual encoder generalization, suffering from limited interpretability and isolated optimization. To overcome these limitations, we propose **Robust-R1**, a novel framework that explicitly models visual degradations through structured reasoning chains. Our approach integrates: (i) supervised fine-tuning for degradation-aware reasoning foundations, (ii) reward-driven alignment for accurately perceiving degradation parameters, and (iii) dynamic reasoning depth scaling adapted to degradation intensity. To facilitate this approach, we introduce a specialized 11K dataset featuring realistic degradations synthesized across four critical real-world visual processing stages, each annotated with structured chains connecting degradation parameters, perceptual influence, pristine semantic reasoning chain, and conclusion. Comprehensive evaluations demonstrate state-of-the-art robustness: **Robust-R1** outperforms all general and robust baselines on the real-world degradation benchmark R-Bench, while maintaining superior anti-degradation performance under multi-intensity adversarial degradations on MMMB, MMStar, and RealWorldQA.

Code — github.com/jqtangust/Robust-R1
Data — huggingface.co/datasets/Jiaqi-hkust/Robust-R1
Model — huggingface.co/Jiaqi-hkust/Robust-R1
Space — huggingface.co/spaces/Jiaqi-hkust/Robust-R1

1 Introduction

Multimodal Large Language Models (MLLMs) have demonstrated remarkable capabilities in visual understanding tasks (Liu et al. 2024; Tang et al. 2024a, 2025; Lu et al. 2024a). However, their performance degrades significantly under real-world visual degradations (e.g., noise, blur, occlusion) (Malik et al. 2025; Schlarmann et al. 2024; Tang et al. 2023, 2024b), compromising reliability in practical applications. Therefore, enhancing robustness against such degradations remains a critical challenge for deploying MLLMs in uncontrolled environments (Long et al. 2025).

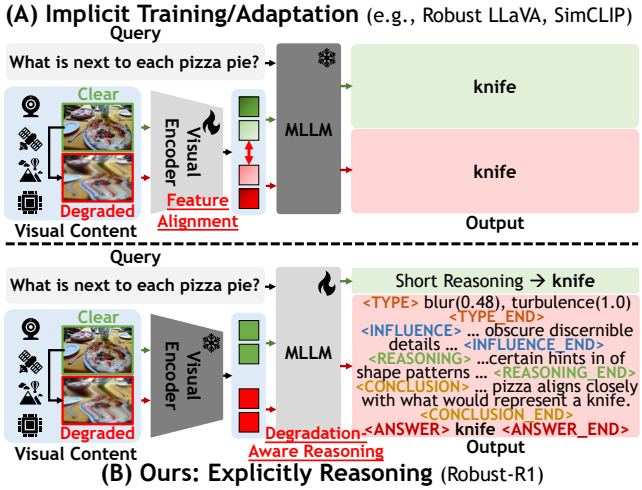

Figure 1: Comparison with other existing robustness enhancement approaches. (A) is based on implicit training/adaptation, which only considers the visual encoder feature alignment. (B) is ours, and we explicitly integrate the degradation-aware reasoning chain into MLLM.

Existing approaches primarily rely on *implicit training/adaptation strategies* to integrate robustness, such as adversarial training (Wang et al. 2024b), robust vision-language alignment (Hossain and Imteaj 2024; Schlarmann et al. 2024; Yuan et al. 2024), or large-scale adversarial pre-training (Malik et al. 2025). These methods focus on fortifying visual encoders against distortions through data-centric optimization. While effective, they suffer from two fundamental limitations (as indicated in Figure 1-A): **(i) Limited**

*These authors contributed equally.
†Corresponding Author: Qifeng Chen; Co-corresponding Author: Wei Wei.

Interpretability: They lack explicit mechanisms to diagnose degradation impacts on original semantic information. **(ii) Isolated Optimization:** They neglect the degradation-propagation relation between the visual encoder and large language model.

To overcome these limitations, we propose **Robust-R1**, a novel framework that explicitly models visual degradations through structured reasoning. Unlike implicit paradigms, **Robust-R1** firstly perceives degradation parameters (type and intensity), then analyzes their semantic impact on visual content, and finally reconstructs distortion-free interpretations to derive robust results. This explicit approach significantly enhances robustness while providing interpretable reasoning traces (as shown in Figure 1-B).

Our implementation comprises three core stages: **First**, we perform Supervised Fine-Tuning (SFT) to equip pre-trained MLLMs with foundational degradation-aware reasoning abilities. **Second**, we design a reward function that aligns model outputs with accurate degradation parameters. **Finally**, we introduce a complementary reward function to dynamically scale the reasoning chain length according to degradation severity, ensuring optimal efficiency.

To support this approach, we construct an 11K dataset from A-OKVQA (Schwenk et al. 2022), comprising 10K training and 1K validation samples. For each sample, we synthesize realistic degradations by simulating four key stages: acquisition $\rightarrow$ transmission $\rightarrow$ environment $\rightarrow$ post-processing with random intensities. We then generate structured reasoning chains that link: (i) degradation parameters ($\mathbf{D}_d$), (ii) their influence (Δ_d), (iii) the pristine semantic reasoning chain ($\mathrm{T}_{\mathbf{X}}$), and (iv) the final conclusion ($\mathbf{Y}_d$). The complexity of these reasoning chains is dynamically scaled with the degradation intensity to balance robustness with computational efficiency.

Comprehensive evaluations demonstrate **Robust-R1**'s superior robustness. On the real-world degradation benchmark R-Bench (Li et al. 2024), **Robust-R1** achieves state-of-the-art (SOTA) performance across all degradation intensities (low, medium, and high), outperforming existing general MLLMs and robust MLLMs. Furthermore, when subjected to adversarial degradation on general visual understanding benchmarks (MMMB (Sun et al. 2025), MMStar (Chen et al. 2024a), and RealWorldQA (xAI 2024)), **Robust-R1** maintains significantly robust performance. It exhibits a markedly smaller performance drop compared to all baselines under multi-level degradation intensities (25%, 50%, and 100%). Our contributions are summarized as:

- We propose **Robust-R1**, a novel approach that explicitly mitigates visual degradations in MLLMs through structured reasoning chains, providing interpretable degradation diagnostics alongside enhanced robustness.
- We construct a dataset of 11K samples featuring realistic degradations synthesized across four critical stages, each annotated with structured reasoning chains for degradation-aware reasoning.
- **Robust-R1** achieves SOTA performance on the real-world robust visual understanding benchmark (R-Bench) and demonstrates superior robustness under adversarial degradation on established general benchmarks (MMMB, MMStar, RealWorldQA), significantly outperforming existing general and robust MLLM baselines.

2 Related Work

Robust Visual Understanding Environmental perturbations pose persistent challenges to multimodal large language models (MLLMs), often significantly degrading their perceptual and reasoning capabilities. As a result, enhancing model robustness has become a critical focus in visual understanding research. Early efforts primarily focused on adversarial training through visual encoder fine-tuning. Approaches like TeCoA (Wang et al. 2024b), SimCLIP (Hossain and Imteaj 2024), and Robust CLIP (Schlarmann et al. 2024) optimized model resilience against localized distortions but faced inherent limitations: reliance on limited adversarial datasets often compromised generalization performance. More recent approaches, such as Robust LLaVA (Malik et al. 2025), have sought to mitigate these issues through large-scale adversarial pre-training. Despite some success, these strategies incur substantial computational and annotation costs, limiting their scalability.

In contrast to these implicit adaptation paradigms, **Robust-R1** introduces a novel degradation-aware reasoning mechanism that explicitly enhances interpretability while improving robustness.

Multimodal Reasoning Multimodal reasoning empowers MLLMs to solve complex tasks by integrating perception, contextual understanding, and logical inference (Wei et al. 2022). Prior work has made considerable progress in domains such as mathematical visual reasoning, where models are required to interpret and reason over problems involving both symbolic notations and visual elements (Wang et al. 2024a; Lu et al. 2024b). Subsequent research has expanded into broader visual reasoning scenarios, exemplified by frameworks like Visual CoT (Shao et al. 2024a) and V* (Wu and Xie 2024), which focus on parsing scene elements and their relational structure.

Robust-R1 builds upon and extends this line of work by harnessing the MLLM's intrinsic reasoning capacity, pioneering its application to explicitly reason about and overcome visual distortions, thereby establishing a new paradigm for robust multimodal understanding.

3 Methodology

Problem Definition Multimodal Large Language Models (MLLMs) frequently exhibit performance degradation when processing visually corrupted inputs in real-world scenarios (Xu et al. 2025b,a), which undermines their interpretation accuracy. This challenge can be represented as Eq. (1),

$$\mathbf{Y}_d = \mathcal{M}_{\text{MLLM}}(\mathbf{X}_d \oplus \mathbf{P}), \tag{1}$$

where $\mathbf{X}_d$ is the degraded visual input, derived as $\mathbf{X}_d = \mathcal{D}(\mathbf{X})$, with $\mathbf{X}$ as the original input and $\mathcal{D}(\cdot)$ representing the degradation function. $\mathbf{P}$ denotes the text prompt. $\mathcal{M}_{\text{MLLM}}(\cdot)$ denotes the original MLLM framework. $\mathbf{Y}_d$ is the generated output under current conditions. $\oplus$ indicates the multimodal

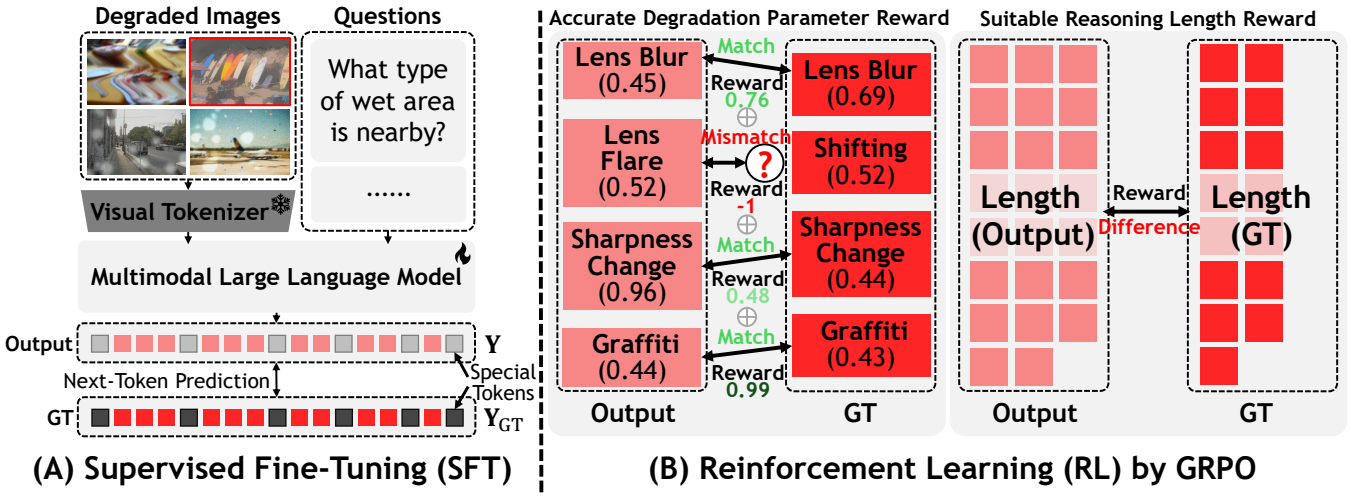

Figure 2: Overview of **Robust-R1**. (A) Supervised Fine-Tuning (SFT): we train the model using reasoning data to equip it with basic degradation-aware reasoning capability; (B) Reinforcement Learning (RL): we propose two reward functions to (i) align precise degradation-aware space while (ii) adaptively scaling to suitable reasoning lengths based on degradation intensity.

combination operator. To tackle this issue, we aim to develop a robust MLLM framework that satisfies:

$$\mathcal{M}_{\text{MLLM}}^{(\text{Robust})}(\mathbf{X}_d \oplus \mathbf{P}) \xrightarrow{\text{approx}} \mathcal{M}_{\text{MLLM}}(\mathbf{X} \oplus \mathbf{P}), \tag{2}$$

where $\mathcal{M}_{\text{MLLM}}^{(\text{Robust})}(\cdot)$ denotes our enhanced model, and the approximation operator $\xrightarrow{\text{approx}}$ signifies the objective of approximating the output under pristine visual conditions.

Overview of Degradation-Aware Reasoning To address the above problem, **Robust-R1** incorporates an explicit degradation-aware reasoning process that perceives degradation parameters (type and intensity), analyzes their impact on visual content, and reconstructs high-fidelity interpretations. This process is formulated as:

$$\begin{aligned} &\mathcal{M}_{\text{MLLM}}^{(\text{Robust})}(\mathbf{X}_d \oplus \mathbf{P}) \Leftrightarrow \\ &\{\mathcal{M}_{\text{p}}(\mathrm{D}_d, \Delta_d \mid \mathbf{X}_d) \rightarrow \mathcal{M}_{\text{r}}(\mathrm{T}_{\mathbf{X}} \mid \mathrm{D}_d, \Delta_d, \mathbf{X}_d, \mathbf{P}) \\ &\rightarrow \mathcal{M}_{\text{MLLM}}(\mathrm{Y}_d \mid (\mathrm{T}_{\mathbf{X}}, \mathrm{D}_d, \Delta_d) \oplus \mathbf{X}_d \oplus \mathbf{P})\}, \end{aligned} \tag{3}$$

where $\mathcal{M}_{\text{p}}(\cdot)$ is degradation parameters perception process, to perceive $\mathrm{D}_d = \{\tau_d^{(i)}, s_d^{(i)}\}_{i=1}^{I}$ (types τ_d and intensities s_d) and their impact Δ_d; $\mathcal{M}_{\text{r}}(\cdot)$ reconstructs the pristine semantic representation $\mathrm{T}_{\mathbf{X}}$ of original $\mathbf{X}$; and original $\mathcal{M}_{\text{MLLM}}(\cdot)$ can generate the robust output $\mathbf{Y}_d$ conditioned on degradation-aware reasoning chain.

Workflow **Firstly**, to integrate degradation-aware reasoning capabilities, We first fine-tune the pretrained vision-language model to establish foundational degradation-aware reasoning capabilities (Section 3.1). **Subsequently**, We employ reinforcement learning with a dedicated reward function to align the model's perception with accurate degradation parameters (D_d) (Section 3.2). **Finally**, we dynamically adjust the reasoning chain length based on degradation intensity to optimize the trade-off between robustness and efficiency (Section 3.3).

3.1 Acquiring Basic Reasoning Ability

Tokenization of Reasoning Chain To enable structured degradation-aware reasoning, we formalize the reasoning chain using special tokens (enclosed in"<" and ">") that segment distinct reasoning phases:

$$\begin{aligned} &\texttt{<TYPE>}\mathrm{D}_d\texttt{<TYPE\_END>}, \\ &\texttt{<INFLUENCE>}\Delta_d\texttt{<INFLUENCE\_END>}, \\ &\texttt{<REASONING>}\mathrm{T}_{\mathbf{X}}\texttt{<REASONING\_END>}, \\ &\texttt{<CONCLUSION>}\mathbf{Y}_d\texttt{<CONCLUSION\_END>}, \\ &\texttt{<ANSWER>}\mathbf{Y}_d^{(\text{answer})}\texttt{<ANSWER\_END>}\ (\text{Optional}), \end{aligned} \tag{4}$$

where $\mathbf{Y}_d^{(\text{answer})}$ denotes the task-specific answer output during benchmark evaluation. This tokenization enforces a sequential reasoning flow to maintain structured output.

Supervised Fine-Tuning (SFT) We optimize model parameters θ through next-token prediction (as shown in Figure 2-A) on the structured reasoning chain:

$$\mathcal{L}_{\text{SFT}} = -\mathbb{E}_{(\mathbf{X}_d, \mathbf{P}, \mathbf{Y}) \sim \mathcal{P}_T} \sum_{n=1}^{N} \log \mathcal{P}_\theta \left(w_n \mid w_{<n}, \mathbf{X}_d, \mathbf{P}\right), \tag{5}$$

where $\mathrm{C} = (w_n, \ldots, w_N) \sim \{\mathrm{D}_d, \Delta_d \rightarrow \mathrm{T}_{\mathbf{X}} \rightarrow \mathbf{Y}_d\}$ represents the output reasoning chain. N denotes the sequence length, $\mathcal{P}_\theta$ is the model's conditional probability distribution, $\mathcal{P}_T$ denotes the distribution of training data. This optimization enables the model to acquire foundational degradation-aware reasoning ability by sequentially generating the structured reasoning chain.

3.2 Aligning Accurate Degradation Parameters

Although SFT equips the MLLM with foundational degradation-aware reasoning ability, it still lacks an accurate perception of degradation parameters (types and intensities). As quantitatively demonstrated in Figure 6-A (w/o

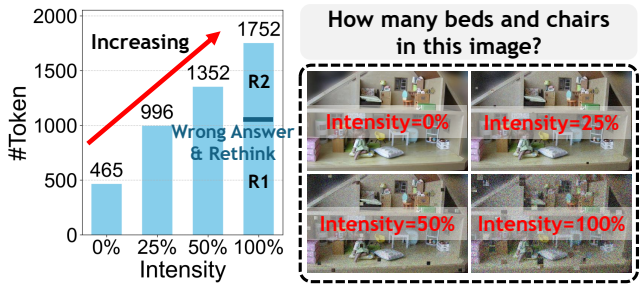

Figure 3: Correlation between degradation intensity and reasoning chain length on Seed-1.5-VL (Guo et al. 2025). Higher degradation intensities require longer chains to maintain accuracy, even multi-step reasoning.

D_d), lacking precise alignment exhibits significant deviation from practical degradation parameters, leading to limited degradation perception ability.

Reward for Accurate Degradation Parameters To achieve high-fidelity alignment, we design a reward function that directly operates in the degradation parameter space (as shown in Figure 2-B (left)). The reward function $r_{\text{deg}}(\mathbf{Y}, \mathbf{Y}_{\text{GT}})$ explicitly evaluates degradation parameter deviation:

$$r_{\text{deg}}(\mathbf{Y}, \mathbf{Y}_{\text{GT}}) = \sum_{i=1}^{I}\sum_{j=1}^{J} \delta(\tau_d^{(i)} = \tau_{\text{GT}}^{(j)}) \cdot \left(1 - \left|s_d^{(i)} - s_{\text{GT}}^{(j)}\right|\right) - \delta(\tau_d^{(i)} \neq \tau_{\text{GT}}^{(j)}), \tag{6}$$

where $\delta(\cdot)$ denotes the Kronecker delta function (web 2025). This formulation specifically: (1) penalizes type mismatches with -1 reward; (2) rewards type matches proportionally to intensity accuracy $(1 - |\Delta s|)$; and (3) aggregates rewards across all instances ($i = 1, \ldots, I$ and $j = 1, \ldots, J$).

3.3 Scaling to Suitable Reasoning Length

Although we achieve accurate D_d alignment, longer reasoning chains may introduce computational redundancy. As identified in (Sui et al. 2025), such "overthinking" reduces inference efficiency without improving output quality.

Observation Through empirical analysis in Figure 3, we observe a strong correlation between degradation intensity and required reasoning length, as:

$$\texttt{len}(\mathbf{Y}) \propto \mathbb{E}\left[\sum_{i=1}^{I} s_d^{(i)}\right], \tag{7}$$

where $\texttt{len}(\mathbf{Y})$ denotes the length of the generated reasoning chain. Higher degradation levels necessitate longer reasoning chains, while simpler degradations only require shorter responses.

Reward for Suitable Reasoning Length To optimize computational efficiency while maintaining robustness, we introduce a length-modulation reward (Figure 2-B (right)):

$$r_{\text{len}}(\mathbf{Y}, \mathbf{Y}_{\text{GT}}) = 1 - \frac{|\texttt{len}(\mathbf{Y}) - \texttt{len}(\mathbf{Y}_{\text{GT}})|}{\texttt{len}(\mathbf{Y}_{\text{GT}})}, \tag{8}$$

where $\texttt{len}(\mathbf{Y}_{\text{GT}})$ is the optimal length from ground truth. This reward equals 1 when lengths match exactly $\texttt{len}(\mathbf{Y}) = \texttt{len}(\mathbf{Y}_{\text{GT}})$, and decreases linearly with relative length discrepancy.

Reinforcement Learning (RL) We integrate these two rewards into a unified optimization framework:

$$\mathcal{R}(\mathbf{Y}, \mathbf{Y}_{\text{GT}}) = r_{\text{deg}}(\mathbf{Y}, \mathbf{Y}_{\text{GT}}) + r_{\text{len}}(\mathbf{Y}, \mathbf{Y}_{\text{GT}}), \tag{9}$$

where $\mathcal{R}(\cdot)$ represents the comprehensive reward function. This composite reward drives Group Relative Preference Optimization (GRPO) (Shao et al. 2024b), and for each input pair $\mathbf{X}_d \oplus \mathbf{P}$, we sample G candidate responses $\{\mathbf{Y}^{(g)}\}_{g=1}^{G}$. The group-relative advantage is computed as:

$$\hat{A}^{(g)} = \frac{\mathcal{R}^{(g)} - \mu_{\mathcal{R}}}{\sigma_{\mathcal{R}}}, \tag{10}$$

where $\mathcal{R}^{(g)} = \mathcal{R}(\mathbf{Y}^{(g)}, \mathbf{Y}_{\text{GT}})$, with:

$$\mu_{\mathcal{R}} = \frac{1}{G}\sum_{g=1}^{G}\mathcal{R}^{(g)}, \ \sigma_{\mathcal{R}} = \sqrt{\frac{1}{G}\sum_{g=1}^{G}(\mathcal{R}^{(g)} - \mu_{\mathcal{R}})^2}, \tag{11}$$

Through GRPO optimization (Shao et al. 2024b), we maximize the expected composite reward:

$$\theta^* = \arg\max_{\theta} \mathbb{E}_{(\mathbf{X}_d, \mathbf{P})\sim\mathcal{P}_T}\left[\mathcal{R}\left(\mathbf{Y}, \mathbf{Y}_{\text{GT}}\right)\right]. \tag{12}$$

This optimization strategy achieves dual objectives: (1) accurate alignment with degradation parameters through r_{deg}, and (2) suitable allocation of computational efficiency through r_{len}. The combined approach ensures robust visual understanding while maintaining efficiency across diverse real-world degradation scenarios.

4 Data Construction

Existing visual understanding datasets (e.g., LLaVA (Liu et al. 2024), R-Bench (Li et al. 2024), A-OKVQA (Schwenk et al. 2022), Conceptual Captions (Sharma et al. 2018)) lack explicit annotations for degradation parameters (D_d), their impacts (Δ_d), and pristine semantic reasoning chains ($\mathrm{T}_{\mathbf{X}}$). This gap hinders training degradation-aware MLLMs. To bridge this gap, we construct a specialized dataset featuring synthetically generated degradations and structured reasoning annotations. Our dataset is built upon a subset of A-OKVQA (Schwenk et al. 2022), comprising 10K samples for training and 1K for validation.

Our whole automated annotation pipeline, illustrated in Figure 4. The procedure consists of the following five steps:

Step (1): Synthesizing Real-World Degradations We construct a comprehensive degradation model $\mathcal{D}(\cdot)$ that simulates degradations introduced across four real-world image processing stages: **1. Acquisition** (Lens Blur, Lens Flare, Motion Blur, Dirty Lens, Saturation), **2. Transmission** (Compression, Block Change, Shifting, Scan Lines), **3. Environment** (Darkness, Atmospheric Turbulence, Noise, Color Diffusion), and **4. Postprocessing** (Sharpness Change, Graffiti, Watermark Damage).

For each pristine image $\mathbf{X}$, we generate a degraded version by:

$$\mathbf{X}_d = \mathcal{D}\left(\mathbf{X}\,;\, \{\tau_d^{(i)}, s_d^{(i)}\}_{i=1}^{I}\right), \tag{13}$$

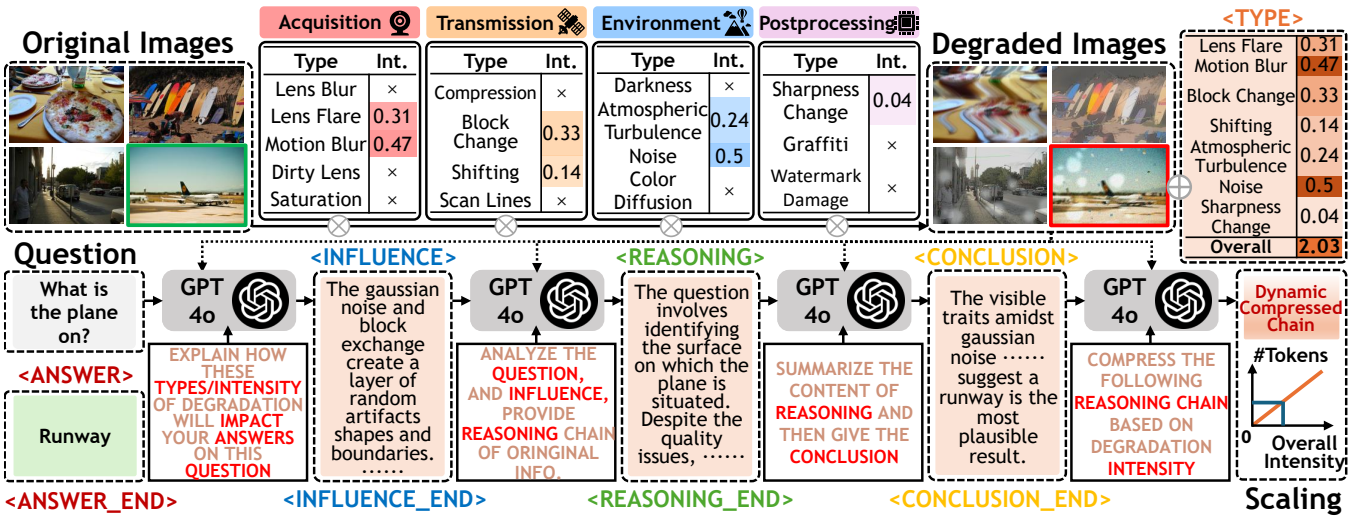

Figure 4: Data generation pipeline. The original images undergo various real-world processing stages, where multiple degradations are randomly added to obtain degraded images and their corresponding degradation <TYPE>s. Based on these and the original question-answering pairs (QAs), the pipeline progressively generates <INFLUENCE>, <REASONING>, and <CONCLUSION>. Finally, the reasoning chain is scaling according to different **intensities** to achieve optimal efficiency.

where the degradation function $\mathcal{D}(\cdot)$ is parameterized by randomly sampled types $\tau_d^{(i)}$ and intensities $s_d^{(i)} \sim \mathcal{U}[0,1]$.

Step (2): Generating Degradation Influence We employ GPT-4o (Hurst et al. 2024) with a fixed prompt template $\Psi_{\text{INFLUENCE}}$ to produce a textual description Δ_d of the degradation's semantic impact:

$$\Delta_d = \mathcal{G}_{\text{GPT-4o}}(\mathbf{X}, \mathbf{X}_d, \mathrm{D}_d, \mathbf{Y}_{\text{GT}}\ ;\ \Psi_{\text{INFLUENCE}}). \quad (14)$$

This narrative establishes a causal link between the visual degradation and its effect on content interpretation, providing the necessary supervision for training the perception module $\mathcal{M}_{\text{p}}(\cdot)$.

Step (3): Generating Pristine Semantic Reasoning Using a distinct prompt template $\Psi_{\text{REASONING}}$, we instruct GPT-4o to infer the original semantic reasoning chain $\mathrm{T}_{\mathbf{X}}$ by compensating for the degradation influence:

$$\mathrm{T}_{\mathbf{X}} = \mathcal{G}_{\text{GPT-4o}}(\mathbf{X}_d, \mathrm{D}_d, \Delta_d, \mathbf{Y}_{\text{GT}}\ ;\ \Psi_{\text{REASONING}}), \quad (15)$$

This step recovers the underlying reasoning process as if performed on the undistorted image, which is crucial for training the reconstruction module $\mathcal{M}\mathrm{r}(\cdot)$.

Step (4): Generating Reasoning Conclusion The final reasoning conclusion $\mathbf{Y}_d$ is generated by conditioning on the pristine semantic reasoning and the ground-truth answer, using a prompt template $\Psi_{\text{CONCLUSION}}$:

$$\mathbf{Y}_d = \mathcal{G}_{\text{GPT-4o}}(\mathrm{T}_{\mathbf{X}}, \mathbf{Y}_{\text{GT}}\ ;\ \Psi_{\text{CONCLUSION}}). \quad (16)$$

Step (5): Scaling Reasoning Chain Length To enable adaptive computational allocation, we dynamically adjust the length of the complete reasoning chain C based on the total degradation intensity:

$$\hat{\mathrm{C}} = \mathcal{G}_{\text{GPT-4o}}\left(\mathrm{C}\ ;\ \Psi_{\text{Len}}\Big(\sum_{i=1}^{I} s_d^{(i)}\Big)\right), \quad (17)$$

where $\hat{\mathrm{C}}$ denotes the scaled reasoning chain, and $\Psi_{\text{Len}}(\cdot)$ is a set of intensity-calibrated prompt templates. This procedure ensures reasoning efficiency and is instrumental for optimizing the length reward r_{len}.

Quality and Robustness The resulting dataset, structured according to the reasoning process defined in Eq. (3), supports both the SFT and the subsequent GRPO optimization of our robust model $\mathcal{M}_{\text{MLLM}}^{(\text{Robust})}(\cdot)$. Besides, the inverse relation between image quality and degradation intensity validates that the distribution of corruptions in our dataset mirrors real-world conditions. The lexical diversity of the reasoning corpus, demonstrates its inherent capacity to model complex logical relationships. This establishes a foundation for achieving robust performance. *More details in the supplementary material.*

5 Experiments

Training Configuration Our model is built upon Qwen2.5-VL-3B (Bai et al. 2025), which employs a redesigned Vision Transformer (ViT) as its vision encoder. We adopt a dual-stage optimization strategy:

- **Supervised Fine-Tuning (SFT)**: 25% training data used to establish basic instruction-following ability.
- **Reinforcement Learning (RL)**: 75% data for align accurate degradation parameters and suitable chain length.

Notably, we freeze both the vision encoder and visual projection layers while performing *full-parameter fine-tuning* on the language model. This design preserves visual feature stability while empowering the MLLM to develop robust degradation-aware reasoning mechanisms.

Baselines We compare against two categories SOTA baselines: (i) **General MLLMs**, including Qwen2.5-VL-3B (Bai

Category	Method	MCQ			VQA			CAP			Overall
		low	mid	high	low	mid	high	low	mid	high	
General MLLM	Qwen2.5-VL-3B (Bai et al. 2025)	0.6411	0.6022	0.5732	0.4872	0.4854	0.4904	0.3778	0.3704	0.3330	0.4845
	Gemma3-4B (Team et al. 2025)	0.5823	0.5776	0.5060	0.4865	0.4630	0.4419	0.4048	0.3746	0.3480	0.4649
	InternVL-4B (Chen et al. 2024b)	0.6235	0.6024	0.5914	0.4982	0.4539	0.5108	0.3667	0.3041	0.2851	0.4706
Robust MLLM	TeCoA (Wang et al. 2024b)	0.4647	0.4223	0.4024	0.4687	0.3994	0.4461	0.2111	0.2195	0.1937	0.3586
	Robust CLIP (Schlarmann et al. 2024)	0.4705	0.4658	0.4024	0.4503	0.4339	0.4743	0.2290	0.2219	0.1983	0.3718
	Robust LLaVA (Malik et al. 2025)	0.3352	0.2608	0.3048	0.2607	0.2212	0.2443	0.0068	0.0065	0.0067	0.1830
Ours	SFT	0.6176	0.6087	0.5610	0.4804	0.4836	0.5012	0.4080	0.3858	0.3518	0.4886
	SFT and RL	0.6529	0.6391	0.6097	0.4914	0.4909	0.4980	0.4068	0.3781	0.3484	0.5017

Table 1: Quantitative performance on R-Bench (Li et al. 2024) on MCQ/VQA/CAP tasks with three degradation strength levels (from low to high). The best/second best results are shown in Red/Blue respectively.

Category	Method	**MMMB** (Sun et al. 2025)				**MMStar** (Chen et al. 2024a)				**RealWorldQA** (xAI 2024)			
		clean	Intensity 25%	50%	100%	clean	Intensity 25%	50%	100%	clean	Intensity 25%	50%	100%
General MLLM	Qwen2.5-VL-3B (Bai et al. 2025)	80.60	79.19	78.68	74.50	54.73	52.90	51.86	48.66	65.22	64.96	63.39	60.65
	Gemma3-4B (Team et al. 2025)	71.01	70.30	70.20	69.14	43.93	43.20	42.60	41.33	55.42	54.77	53.72	52.81
	InternVL-4B (Chen et al. 2024b)	77.97	77.47	76.66	74.59	51.53	50.26	49.60	46.93	57.38	58.16	57.64	54.90
Robust MLLM	TeCoA (Wang et al. 2024b)	57.17	65.71	56.11	51.76	30.46	30.60	30.73	28.06	40.00	39.73	39.47	38.69
	Robust CLIP (Schlarmann et al. 2024)	58.83	58.28	57.97	53.33	33.00	32.26	31.80	29.46	43.26	42.48	42.61	41.43
Ours	SFT	80.85	79.45	78.68	74.94	55.20	53.00	51.86	49.53	68.23	67.58	67.32	63.92
	SFT and RL	81.41	79.49	79.04	75.35	56.86	54.40	53.60	49.53	67.71	66.40	67.05	63.26

Table 2: Quantitative performance for anti-degradation on three visual understanding benchmarks (MMMB (Sun et al. 2025), MMStar (Chen et al. 2024a), and RealWorldQA (xAI 2024)) with three degradation intensity levels (from 25% to 100%). The best/second best results are showed in Red/Blue respectively.

et al. 2025), Gemma3-4B (Team et al. 2025), and InternVL-4B (Chen et al. 2024b); (ii) **Robust MLLMs**, comprising TeCoA (Wang et al. 2024b), Robust CLIP (Schlarmann et al. 2024), and Robust LLaVA (Malik et al. 2025).

Benchmarks We conduct rigorous evaluation across two dimensions: (i) **Real-World Robustness**: Directly assessing robust visual understanding ability on R-Bench (Li et al. 2024); (ii) **Adversarial Robustness**: Evaluation under synthetic degradation attacks by applying multi-type, multi-level real-world degradations to visual content in MMMB (Sun et al. 2025), MMStar (Chen et al. 2024a), and RealWorldQA (xAI 2024). This dual-strategy comprehensively measures both intrinsic degradation comprehension and performance preservation under visual corruption.

5.1 Performance on R-Bench

R-Bench (Li et al. 2024) is a benchmark designed to directly evaluate image understanding capabilities under real-world degradation conditions. It incorporates three distinct task types (Multiple Choice Questions (MCQ), Visual Question Answering (VQA), and Image Captioning (CAP)) with three degradation intensity levels (low, mid, and high) to systematically assess the robustness of visual comprehension.

As shown in Table 1, **`Robust-R1`** (Ours) demonstrates significant improvements in image understanding capabilities following both Supervised Fine-Tuning (SFT) and subsequent Reinforcement Learning optimization (SFT and RL). Experimental results indicate that our model surpasses

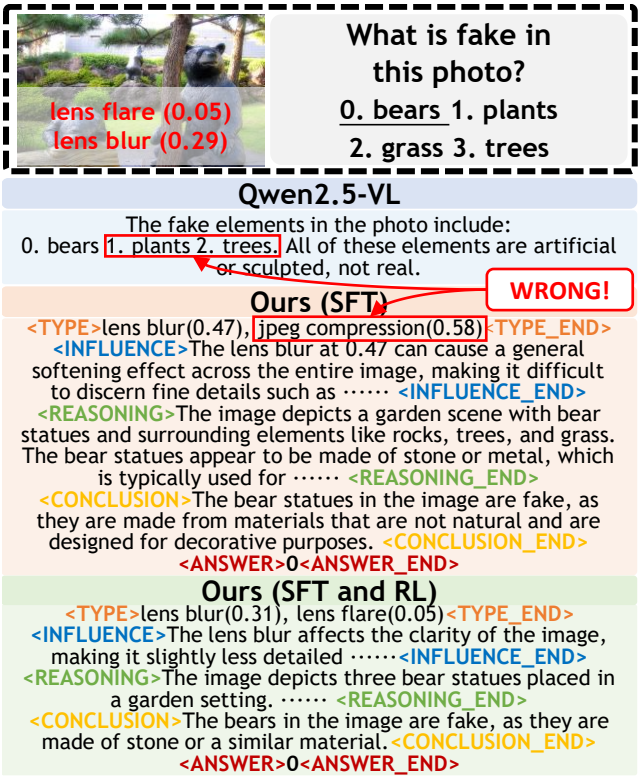

Figure 5: Qualitative evaluation for anti-degradation. Ours (SFT and RL) can provide robust and efficient result.

Method	MCQ			VQA			CAP			Overall
	low	mid	high	low	mid	high	low	mid	high	
Qwen2.5-VL-3B (Bai et al. 2025)	0.6411	0.6022	0.5732	0.4872	0.4854	0.4904	0.3778	0.3704	0.3330	0.4845
Ours (w/o Reasoning)	0.6588	0.5901	0.4756	0.4905	0.4900	0.4862	0.2901	0.2673	0.2758	0.4471
Ours (w/o r_{deg})	0.6647	0.6398	0.5505	0.4912	0.4894	0.5056	0.3684	0.3578	0.3248	0.4880
Ours (w/o r_{len})	0.6647	0.6354	0.5975	0.4904	0.4887	0.4877	0.3656	0.3678	0.3189	0.4907
Ours	0.6529	0.6391	0.6097	0.4914	0.4909	0.4980	0.4068	0.3781	0.3484	0.5017

Table 3: Ablation study on R-Bench (Li et al. 2024) on MCQ/VQA/CAP tasks with three degradation strength levels (from low to high). The best/second best results are showed in Red/Blue respectively.

existing general and robust MLLMs baselines in overall performance on this benchmark.

5.2 Anti-Degradation Performance

To rigorously evaluate our model's robustness against image degradation, we conduct comprehensive experiments on three established visual understanding benchmarks (MMMB (Sun et al. 2025), MMStar (Chen et al. 2024a), and RealWorldQA (xAI 2024)). We introduce random degradations at varying intensity levels (25%, 50%, and 100%) to the original images, creating challenging test conditions that assess the model's anti-degradation capability.

Quantitative Results As demonstrated in Table 2, our model achieves SOTA performance across all degradation levels compared to existing baselines. This evidence confirms our model's exceptional robustness to diverse image degradations under adversarial conditions.

Qualitative Result Figure 5 presents qualitative comparisons of our outputs. Compared to the original baseline, **`Robust-R1`** significantly reduces hallucinations and errors in visual understanding through reasoning. Furthermore, after preference optimization, **`Robust-R1`** achieves a optimal balance between inference efficiency and accurate degradation parameters perception.

5.3 Ablation Study

Reasoning *vs.* Adaptation To validate the effectiveness of explicit reasoning versus implicit adaptation, we conduct an ablation study by removing degradation reasoning chains from our training data, relying solely on fine-tuning for adaptation (Table 3, w/o Reasoning). The experimental results reveal two critical findings: (i) Adaptation provides only marginal performance gains in specific intensity ranges compared to the base model, and fails catastrophically in high-intensity degradation scenarios; (ii) Explicit reasoning demonstrates significantly improved robustness over both the adaptation-only model and the original baseline. These results conclusively demonstrate that explicit reasoning capability is essential for robust visual understanding, enabling systematic analysis and compensation for visual degradations rather than mere adaptation.

Effectiveness of r_{deg} To validate the critical role of the degradation reward r_{deg}, we conduct an ablation study comparing model performance with and without this component.

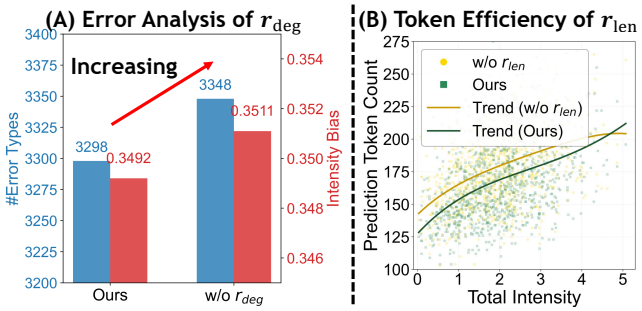

Figure 6: Statistics analysis for (A) r_{deg} and (B) r_{len}.

As shown in Table 3, incorporating r_{deg} substantially improves visual understanding performance on R-Bench compared to the ablated variant. This improvement stems from r_{deg}'s ability to enhance precise alignment with degradation parameters. Furthermore, statistical analysis on our out-of-domain testset (Section 4) in Figure 6-A reveals that r_{deg} significantly reduces two key error types: (i) degradation-type misclassification and (ii) degradation-intensity estimation bias. These results demonstrate that r_{deg} increases model precision in identifying degradation parameters, directly contributing to superior robustness.

Efficiency of r_{len} To evaluate the effectiveness of the length-modulation reward r_{len}, we conduct an ablation study by removing this component. As shown in Figure 6-B, incorporating r_{len} reduces the average reasoning chain length while maintaining performance, demonstrating its ability to improve computational efficiency. Notably, the model adaptively adjusts reasoning depth based on degradation intensity: longer chains are allocated for severe degradation, while simpler cases require less inference. This task-adaptive allocation not only optimizes resource usage but also enhances overall performance, as evidenced by the quantitative improvements in Table 3 (w/o r_{len}).

6 Conclusion

We propose **`Robust-R1`**, a novel paradigm that incorporates explicit degradation reasoning chains to enhance multimodal understanding robustness. We believe this work opens new avenues for building more robust, interpretable, and efficient multimodal systems capable of operating reliably in visually challenging environments.

Acknowledgments

The work described in this paper was supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project Reference Number: AoE/E-601/24-N).

Besides, this work was supported in part by the National Natural Science Foundation of China (No. 62472359, 62372379), in part by the Xi'an's Key Industrial Chain Core Technology Breakthrough Project: AI Core Technology Breakthrough under Grand 24ZDCYJSGG0003. Also, this work was supported by the Key Project of the National Natural Science Foundation of China (No. 62536007), the Zhejiang Province Science Foundation (No. LD24F020002) and the Zhejiang Province's 2025 "Leading Goose + X" Science and Technology Plan (No. 2025C02034).

References

2025. Kronecker delta. https://en.wikipedia.org/wiki/Kronecker_delta.

Bai, S.; Chen, K.; Liu, X.; Wang, J.; Ge, W.; Song, S.; Dang, K.; Wang, P.; Wang, S.; Tang, J.; Zhong, H.; Zhu, Y.; Yang, M.; Li, Z.; Wan, J.; Wang, P.; Ding, W.; Fu, Z.; Xu, Y.; Ye, J.; Zhang, X.; Xie, T.; Cheng, Z.; Zhang, H.; Yang, Z.; Xu, H.; and Lin, J. 2025. Qwen2.5-VL Technical Report. *arXiv*.

Chen, L.; Li, J.; Dong, X.; Zhang, P.; Zang, Y.; Chen, Z.; Duan, H.; Wang, J.; Qiao, Y.; Lin, D.; et al. 2024a. Are We on the Right Way for Evaluating Large Vision-Language Models? In *NeurIPS*.

Chen, Z.; Wu, J.; Wang, W.; Su, W.; Chen, G.; Xing, S.; Zhong, M.; Zhang, Q.; Zhu, X.; Lu, L.; et al. 2024b. Internvl: Scaling up vision foundation models and aligning for generic visual-linguistic tasks. In *CVPR*.

Guo, D.; Wu, F.; Zhu, F.; Leng, F.; Shi, G.; Chen, H.; Fan, H.; Wang, J.; Jiang, J.; Wang, J.; et al. 2025. Seed1. 5-vl technical report. *arXiv*.

Hossain, M. Z.; and Imteaj, A. 2024. Sim-clip: Unsupervised siamese adversarial fine-tuning for robust and semantically-rich vision-language models. *arXiv*.

Hurst, A.; Lerer, A.; Goucher, A. P.; Perelman, A.; Ramesh, A.; Clark, A.; Ostrow, A.; Welihinda, A.; Hayes, A.; Radford, A.; et al. 2024. Gpt-4o system card. *arXiv*.

Li, C.; Zhang, J.; Zhang, Z.; Wu, H.; Tian, Y.; Sun, W.; Lu, G.; Liu, X.; Min, X.; Lin, W.; and Zhai, G. 2024. R-Bench: Are your Large Multimodal Model Robust to Real-world Corruptions? *IEEE JSTSP*.

Liu, H.; Li, C.; Li, Y.; and Lee, Y. J. 2024. Improved Baselines with Visual Instruction Tuning. In *CVPR*.

Long, J.; Xu, Z.; Jiang, T.; Yao, W.; Jia, S.; Ma, C.; and Chen, X. 2025. Robust SAM: On the Adversarial Robustness of Vision Foundation Models. In *AAAI*.

Lu, H.; Niu, X.; Wang, J.; Wang, Y.; Hu, Q.; Tang, J.; Zhang, Y.; Yuan, K.; Huang, B.; Yu, Z.; et al. 2024a. Gpt as psychologist? preliminary evaluations for gpt-4v on visual affective computing. In *CVPR*.

Lu, P.; Bansal, H.; Xia, T.; Liu, J.; Li, C.; Hajishirzi, H.; Cheng, H.; Chang, K.-W.; Galley, M.; and Gao, J. 2024b. Mathvista: Evaluating mathematical reasoning of foundation models in visual contexts. In *ICLR*.

Malik, H. S.; Shamshad, F.; Naseer, M.; Nandakumar, K.; Khan, F.; and Khan, S. 2025. Robust-llava: On the effectiveness of large-scale robust image encoders for multi-modal large language models. In *ICCVW*.

Schlarmann, C.; Singh, N. D.; Croce, F.; and Hein, M. 2024. Robust clip: Unsupervised adversarial fine-tuning of vision embeddings for robust large vision-language models. In *ICML*.

Schwenk, D.; Khandelwal, A.; Clark, C.; Marino, K.; and Mottaghi, R. 2022. A-OKVQA: A Benchmark for Visual Question Answering using World Knowledge. In *ECCV*.

Shao, H.; Qian, S.; Xiao, H.; Song, G.; Zong, Z.; Wang, L.; Liu, Y.; and Li, H. 2024a. Visual cot: Advancing multimodal language models with a comprehensive dataset and benchmark for chain-of-thought reasoning. *NeurIPS*.

Shao, Z.; Wang, P.; Zhu, Q.; Xu, R.; Song, J.; Bi, X.; Zhang, H.; Zhang, M.; Li, Y.; Wu, Y.; et al. 2024b. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv*.

Sharma, P.; Ding, N.; Goodman, S.; and Soricut, R. 2018. Conceptual Captions: A Cleaned, Hypernymed, Image Alt-text Dataset For Automatic Image Captioning. In *ACL*.

Sui, Y.; Chuang, Y.-N.; Wang, G.; Zhang, J.; Zhang, T.; Yuan, J.; Liu, H.; Wen, A.; Zhong, S.; Chen, H.; and Hu, X. 2025. Stop Overthinking: A Survey on Efficient Reasoning for Large Language Models. *TMLR*.

Sun, H.-L.; Zhou, D.-W.; Li, Y.; Lu, S.; Yi, C.; Chen, Q.-G.; Xu, Z.; Luo, W.; Zhang, K.; Zhan, D.-C.; and Ye, H.-J. 2025. Parrot: Multilingual Visual Instruction Tuning. *arxiv*.

Tang, J.; Lu, H.; Wu, R.; Xu, X.; Ma, K.; Fang, C.; Guo, B.; Lu, J.; Chen, Q.; and Chen, Y.-C. 2024a. HAWK: Learning to Understand Open-World Video Anomalies. In *NeurIPS*.

Tang, J.; Wu, R.; Xu, X.; Hu, S.; and Chen, Y.-C. 2024b. Learning to Remove Wrinkled Transparent Film with Polarized Prior. In *CVPR*.

Tang, J.; Xia, Y.; Wu, Y.-F.; Hu, Y.; Chen, Y.; Chen, Q.-G.; Xu, X.; Wu, X.; Lu, H.; Ma, Y.; Lu, S.; and Chen, Q. 2025. LPO: Towards Accurate GUI Agent Interaction via Location Preference Optimization. *arxiv*.

Tang, J.; Xu, X.; Hu, S.; and Chen, Y.-C. 2023. High Dynamic Range Image Reconstruction via Deep Explicit Polynomial Curve Estimation. In *ECAI*.

Team, G.; Kamath, A.; Ferret, J.; Pathak, S.; Vieillard, N.; Merhej, R.; Perrin, S.; Matejovicova, T.; Ramé, A.; Rivière, M.; et al. 2025. Gemma 3 technical report. *arXiv*.

Wang, K.; Pan, J.; Shi, W.; Lu, Z.; Ren, H.; Zhou, A.; Zhan, M.; and Li, H. 2024a. Measuring multimodal mathematical reasoning with math-vision dataset. *NeurIPS*.

Wang, S.; Zhang, J.; Yuan, Z.; and Shan, S. 2024b. Pre-trained model guided fine-tuning for zero-shot adversarial robustness. In *CVPR*.

Wei, J.; Wang, X.; Schuurmans, D.; Bosma, M.; Xia, F.; Chi, E.; Le, Q. V.; Zhou, D.; et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *NeurIPS*.

Wu, P.; and Xie, S. 2024. V?: Guided visual search as a core mechanism in multimodal llms. In *CVPR*.

xAI. 2024. Grok-1.5 Vision Preview.

Xu, X.; Wu, J.; Yan, Q.; Cui, J.; Hong, R.; and Yu, B. 2025a. Learnable Feature Patches and Vectors for Boosting Low-light Image Enhancement without External Knowledge. In *CVPR*.

Xu, X.; Zhou, K.; Hu, T.; Wu, J.; Wang, R.; Peng, H.; and Yu, B. 2025b. Low-Light Video Enhancement via Spatial-Temporal Consistent Decomposition. In *IJCAI*.

Yuan, F.; Qin, C.; Xu, X.; and Li, P. 2024. Helpd: Mitigating hallucination of lvlms by hierarchical feedback learning with vision-enhanced penalty decoding. In *EMNLP*.