Q-Insight: Understanding Image Quality via Visual Reinforcement Learning

Weiqi Li[1], Xuanyu Zhang[1], Shijie Zhao[2♦✉], Yabin Zhang[2], Junlin Li[2], Li Zhang[2], Jian Zhang[1] ✉
[1] School of Electronic and Computer Engineering, Peking University
[2] ByteDance Inc.

Abstract

Image quality assessment (IQA) focuses on the perceptual visual quality of images, playing a crucial role in downstream tasks such as image reconstruction, compression, and generation. The rapid advancement of multi-modal large language models (MLLMs) has significantly broadened the scope of IQA, moving toward comprehensive image quality understanding that incorporates content analysis, degradation perception, and comparison reasoning beyond mere numerical scoring. Previous MLLM-based methods typically either generate numerical scores lacking interpretability or heavily rely on supervised fine-tuning (SFT) using large-scale annotated datasets to provide descriptive assessments, limiting their flexibility and applicability. In this paper, we propose **Q-Insight**, a reinforcement learning-based model built upon group relative policy optimization (GRPO), which demonstrates strong visual reasoning capability for image quality understanding while requiring only a limited amount of rating scores and degradation labels. By jointly optimizing score regression and degradation perception tasks with carefully designed reward functions, our approach effectively exploits their mutual benefits for enhanced performance. Extensive experiments demonstrate that Q-Insight substantially outperforms existing state-of-the-art methods on both score regression and degradation perception tasks, while exhibiting impressive zero-shot generalization and superior comparison reasoning capability. The code and models are available at `https://github.com/bytedance/Q-Insight`.

1 Introduction

Image quality assessment (IQA) is a fundamental task in computer vision, critical for optimizing algorithms, enhancing user experiences, and verifying content authenticity across diverse domains, such as image processing [41, 22, 11, 53] and AI-generated content (AIGC) [37, 8]. Traditional IQA methods rely heavily on hand-crafted metrics, either through reference-based comparisons [50] or statistical measures of natural image properties [64]. However, these approaches typically focus on local image characteristics and fail to comprehensively capture global visual quality, limiting their reliability in complex real-world scenarios. More recently, deep learning-based IQA models [47, 19] have emerged, utilizing neural networks to learn hierarchical image representations. Nevertheless, these methods struggle to face significant challenges in out-of-distribution (OOD) generalization.

With recent advances in multi-modal large language models (MLLMs) [23, 1, 48], researchers have begun to leverage these models' extensive world knowledge and perceptual abilities to enhance IQA performance and broaden its applicability [53, 60, 59, 58, 52]. Existing MLLM-based IQA methods generally fall into two categories: score-based methods, such as Q-Align [53] and DeQA-Score [58], and description-based methods, exemplified by DepictQA [60] and DepictQA-Wild [59]. Score-based methods transform discrete tokens into continuous quality scores, thereby improving

♦Project Lead. ✉: Corresponding authors, zhaoshijie.0526@bytedance.com, zhangjian.sz@pku.edu.cn.
†This work was supported in part by National Natural Science Foundation of China (No. 62372016).

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

 https://doi.org/10.52202/085713-1237

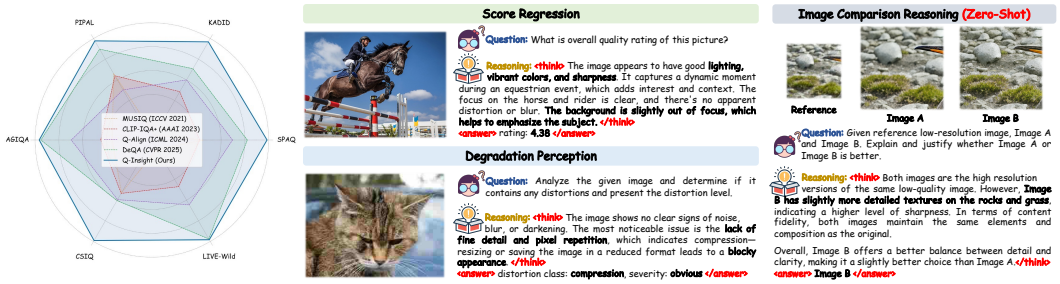


Figure 1: PLCC comparisons between our proposed Q-Insight and existing IQA metrics (left) and three example applications of our Q-Insight (right) are presented. Q-Insight demonstrates significantly improved performance compared to existing methods such as DeQA-Score [58], especially on out-of-domain datasets (e.g., CSIQ [20]). Additionally, Q-Insight effectively supports quality score regression, image degradation perception, and zero-shot image comparison reasoning tasks.

adaptability but typically sacrificing interpretability and neglecting MLLMs' intrinsic reasoning and descriptive capabilities. Meanwhile, **simply regressing a quality score may not be meaningful in certain scenarios**, as image quality scores are subjective, inherently biased, and lack uniform standards across different datasets and content types. For example, when evaluating AIGC-generated data, unusual visual effects and vibrant colors often imply better quality. However, for evaluating super-resolution results, these same features are often considered too painterly, losing the image's authenticity and fidelity. Conversely, description-based methods produce detailed textual explanations of image degradations and comparative assessments, ensuring interpretability yet heavily depending on extensive textual depiction for supervised fine-tuning. Moreover, these models cannot output precise scores, making them unsuitable when used as loss functions or for performing accurate ranking of image quality. Consequently, integrating **numerical scoring** and **descriptive reasoning** into a unified, interpretable MLLM-based IQA framework remains an essential yet unresolved challenge.

In this paper, we move towards a comprehensive understanding of image quality by addressing tasks such as image description, aesthetic and compositional evaluation, degradation perception, and comparative reasoning across images. Rather than teaching the large model "*How to score images*", we aim to inspire it "*How to reason deeply and formulate insightful perspectives on image quality metrics during scoring*". To this end, we resort to Group Relative Policy Optimization (GRPO) [40], a reinforcement learning framework inspired by DeepSeek-R1 [15]. GRPO has recently shown to be highly effective in large language models (LLMs). It uses heuristic reward signals to efficiently guide LLMs in uncovering their intrinsic reasoning capabilities, removing extensive reliance on annotated reasoning chains or additional value models. Recently, researchers have also successfully adapted GRPO to vision-language tasks, including few-shot object detection, reasoning grounding [25], and medical analysis [34]. In the context of image quality understanding, the introduction of GRPO provides at least three distinct advantages: **(1)** no reliance on massive textual training data, **(2)** strong generalization to OOD evaluated images, and **(3)** high diversity in supporting multiple tasks. These benefits align well with our goal of developing a generalized image quality understanding agent.

Specifically, we design **Q-Insight** upon the GRPO framework. In our Q-Insight, we jointly optimize score regression and degradation perception tasks, and carefully design three reward functions: a verifiable score reward for the score regression task, and degradation classification and intensity perception rewards for the degradation perception task. Consequently, Q-Insight effectively exhibits robust reasoning performance using only limited Mean Opinion Scores (MOS) and degradation labels. As shown in Fig. 1, our Q-Insight delivers remarkable performance improvements especially on OOD datasets, while demonstrating comprehensive capabilities across multiple quality assessment and reasoning tasks. For example, it can accurately identify cases where a slightly blurred background, usually regarded as undesirable, effectively helps to emphasize the primary subject of the image. Our empirical investigation reveals that: **(1)** training solely with score labels results in poor perception of image detail degradations (e.g., JPEG compression), while jointly training with the degradation perception task significantly enhances the model's sensitivity to such degradations, and **(2)** score regression and degradation perception tasks are mutually beneficial. Extensive experiments across score regression and degradation perception tasks demonstrate that Q-Insight consistently outperforms existing model-based IQA metrics as well as SFT-driven large language models. Moreover, it exhibits impressive zero-shot generalization on unseen tasks, such as image comparison reasoning, highlighting the robustness and versatility of our method. In summary, our contributions are:

https://doi.org/10.52202/085713-1237

❑ (1) We propose Q-Insight, the first reasoning-style multi-modal large language model specifically designed for comprehensive image quality understanding. Unlike previous methods that depend heavily on detailed textual descriptions for supervised fine-tuning (SFT), our approach achieves superior understanding capability using only limited mean opinion scores or degradation labels.

❑ (2) We introduce a unified framework that jointly optimizes image quality rating and degradation perception, revealing mutual benefits across tasks. Within this framework, we develop three specialized rewards, including verifiable score reward, degradation classification and intensity perception rewards, enabling the GRPO framework to effectively generalize to low-level vision applications.

❑ (3) Extensive experiments across diverse datasets and IQA tasks demonstrate that Q-Insight consistently outperforms existing model-based IQA metrics as well as SFT-driven large language models. Moreover, it exhibits impressive zero-shot generalization on unseen tasks, such as reference-based image comparison reasoning, highlighting the robustness and versatility of our method.

2 Related Work

Score-based IQA methods include full-reference and no-reference approaches. Full-reference methods [50, 42, 65] assess image quality by comparing distorted images with high-quality references using traditional metrics (e.g., SSIM [50]) or advanced deep-learning-based metrics [3, 4, 10, 9, 13, 36] like LPIPS [66]. Non-reference methods evaluate quality without reference images, shifting from traditional handcrafted statistics [27, 28, 29, 30, 31, 38] to deep-learning-derived quality priors [18, 19, 24, 33, 44, 73, 74, 45, 49]. Recent multi-modal large language model (MLLM)-based methods, such as Q-Align [53] and DeQA-Score [58], leverage MLLMs' knowledge and perceptual abilities to produce scores. However, they sacrifice the intrinsic descriptive capabilities of MLLMs.

Description-based IQA methods utilize the foundational knowledge of MLLMs to deliver detailed qualitative assessments and improved interpretability [51, 52, 60, 59, 54, 5, 70, 71, 69]. For instance, Q-Bench [51] and Q-Instruct [52] enhance the low-level perceptual capabilities of MLLMs through specialized datasets and tailored evaluation strategies. Co-Instruct [54] specifically focuses on comparative quality assessments among multiple images. Approaches such as DepictQA [60] and DepictQA-Wild [59] handle both single-image and paired-image evaluations across full-reference and no-reference scenarios. Q-Ground [5] emphasizes a detailed visual quality analysis through visual grounding. However, these methods are highly dependent on extensive textual annotations for supervised fine-tuning, leading to considerable costs in human labor or GPT token consumption.

Reinforcement learning (RL) has emerged as an effective strategy to enhance the reasoning performance of LLMs through feedback-driven refinement [7, 43, 40, 55, 57, 17, 68]. Methods like RLHF [32] and RLAIF [2] employ human or AI-generated feedback to refine model behavior. In vision-language tasks, RL has successfully been employed to align model predictions closely with human preferences and reduce hallucinations [46, 61, 62, 72]. Recently, DeepSeek-R1-Zero [15] introduced group relative policy optimization (GRPO) [40], leveraging rule-based rewards to strengthen reasoning capabilities without supervised fine-tuning. Furthermore, Visual-RFT [25] applied GRPO to visual grounding, and Med-R1 [34] adopted GRPO for medical reasoning tasks. R1-VL [63] extends GRPO through step-wise optimization for multi-modal reasoning. Distinctly, our Q-Insight is the first to integrate RL-based strategies into the foundational visual quality understanding model. It jointly trains on multiple tasks and demonstrates mutually beneficial effects among them.

3 Methodology

3.1 Preliminaries

Group Relative Policy Optimization (GRPO) is an innovative reinforcement learning paradigm that has been widely used in models such as DeepSeek R1-Zero [15], achieving excellent results. Unlike Proximal Policy Optimization (PPO) [39], which requires an explicit critic model to evaluate the performance of the policy model, GRPO [40] directly computes the advantage by comparing a group of responses sampled from the policy model, greatly reducing the computational burden. Specifically, given a query q, GRPO samples N distinct responses $\{o^{(1)}, o^{(2)}, \ldots, o^{(N)}\}$ from the old policy $\pi_{\theta_{\text{old}}}$. Then, the method performs the corresponding actions and receives the respective rewards $\{r^{(1)}, r^{(2)}, ..., r^{(N)}\}$ according to the task-specific rules. By calculating the mean and standard

 https://doi.org/10.52202/085713-1237

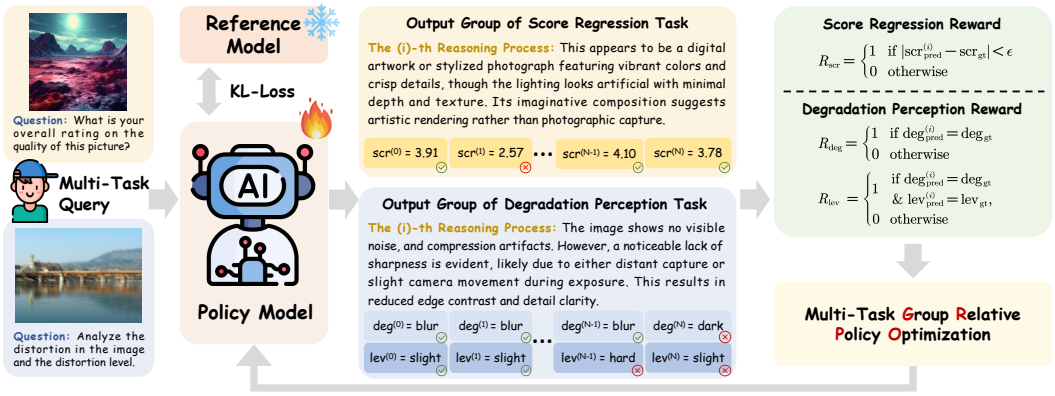


Figure 2: **Overview of the proposed Q-Insight framework.** The policy model receives queries from multiple tasks and generates corresponding groups of responses accompanied by explicit reasoning steps. Task-specific reward functions (R_{scr}, R_{deg}, and R_{lev}) are then applied, and the policy model is subsequently optimized jointly using the multi-task group relative policy optimization algorithm.

deviation of the rewards, the relative advantages of each response can be obtained as follows:

$$\hat{A}^{(i)} = \frac{r^{(i)} - \text{mean}(\{r^{(1)}, r^{(2)} \ldots, r^{(N)}\})}{\text{std}(\{r^{(1)}, r^{(2)} \ldots, r^{(N)}\})}, \tag{1}$$

where $\hat{A}^{(i)}$ represents the normalized relative quality of the i-th answer. Overall, GRPO guides the policy model to prioritize higher-quality answers that receive higher reward values within the group. After obtaining the advantage $\hat{A}^{(i)}$, GRPO calculates the ratio of the probabilities of each response under the new policy $\pi_{\theta_{\text{new}}}$ and the old policy $\pi_{\theta_{\text{old}}}$, denoted as $\rho^{(i)}$. To prevent overly large updates to the model and stabilize training, GRPO restricts the $\rho^{(i)}$ to the range $[1-\delta, 1+\delta]$. To further maintain closeness to the reference distribution π_{ref}, a KL divergence penalty weighted by β is adopted. Finally, the optimization objective of GRPO can be formulated as follows:

$$\mathcal{J}(\theta) = \mathbb{E}_{[q\sim Q, o^{(i)}\sim\pi_{\theta_{\text{old}}}]}\left\{\min\left[\rho^{(i)}\hat{A}^{(i)}, \text{clip}\left(\rho^{(i)}, 1-\delta, 1+\delta\right)\hat{A}^{(i)}\right] - \beta\cdot\mathbb{D}_{\text{KL}}[\pi_{\theta_{\text{new}}}||\pi_{\text{ref}}]\right\} \tag{2}$$

where $\rho^{(i)} = \pi_{\theta_{\text{new}}}(o^{(i)} \mid q) \,/\, \pi_{\theta_{\text{old}}}(o^{(i)} \mid q)$, Q denotes the candidate question set, and $\mathbb{D}_{\text{KL}}$ denotes the KL regularization. π_{ref} is typically a frozen pre-trained MLLM. GRPO effectively integrates consistent policy updates and strong reward signals in a balanced way. To our knowledge, we are the first to apply GRPO to image quality understanding tasks, enabling our model to achieve robust reasoning and generalization performance without heavy reliance on extensive annotated data.

3.2 Overview of Q-Insight

The overall framework of Q-Insight is illustrated in Fig. 2. During training, we jointly optimize two tasks: score regression and degradation perception. Specifically, the multi-modal input for each task comprises an image paired with a task-specific question. Given these inputs, the policy model π_θ generates groups of answers, each accompanied by explicit reasoning steps. Subsequently, each answer is evaluated using its corresponding reward function: R_{scr} for score regression, and R_{deg} and R_{lev} for degradation perception. After computing rewards for each group of answers, the policy model is optimized jointly via the multi-task GRPO algorithm. Additionally, a KL-divergence loss is applied to constrain deviations between the policy model π_θ and the reference model π_{ref}. During inference, the trained Q-Insight generates coherent reasoning processes and outputs precise answers. Further details regarding multi-task GRPO and data construction are provided in Sec. 3.3 and Sec. 3.4.

3.3 Multi-Task Group Relative Policy Optimization

As depicted in Fig. 2, for each input data pair, the policy model π_θ generates a group of N responses, denoted as $\{o^{(i)}\}_{i=1}^{N}$. We then evaluate each of these responses using the proposed reward functions

(R_{scr}, R_{deg}, and R_{lev}) and obtain the overall rewards $\{r^{(i)}\}_{i=1}^{N}$. Proper reward design is crucial, as informative and carefully constructed rewards directly facilitate Q-Insight's ability to effectively learn reasoning and perception patterns, thus ensuring robust performance across multiple tasks. Specifically, Q-Insight employs a general format reward function shared across all tasks, as well as task-specific reward functions tailored to the unique characteristics of each individual task.

Format reward evaluates whether the reasoning steps are properly enclosed within "<think>" and "</think>" tags, and the final answer is correctly enclosed within "<answer>" and "</answer>" tags [15]. Additionally, we require that the content inside "<answer>" tags follow a JSON-like format: beginning with "{", ending with "}", and containing no additional "{" or "}" characters internally. This ensures that Q-Insight can consistently parse results across different tasks. The reward score $r_{\text{fmt}}^{(i)}$ is set to 1 if the i-th response fulfills all the above conditions; otherwise, its reward is 0.

Rewards for score regression task. A standard way to quantify image quality is by using the Mean Opinion Score (MOS). Instead of directly fitting the MLLM predictions to MOS, we use MOS as a general guideline to motivate the model towards deeper reasoning and generating insightful perspectives during the process of evaluating image quality. Inspired by the treatment of mathematical reasoning tasks in DeepSeek-R1 [15], we consider the continuous MOS prediction as either correct or incorrect, thereby adopting a binary reward to avoid extremely large or small reward values. Denoting the predicted score of the i-th response as $\text{scr}_{\text{pred}}^{(i)}$ and the ground-truth score as scr_{gt}, we design the verifiable reward for scoring as follows. The reward value $r_{\text{scr}}^{(i)}$ for the i-th response is determined by:

$$r_{\text{scr}}^{(i)} = 1 \quad \text{if } |\,\text{scr}_{\text{pred}}^{(i)} - \text{scr}_{\text{gt}}\,| < \epsilon, \text{ otherwise } 0, \tag{3}$$

where ϵ is a predefined threshold. In particular, if ϵ is set to 0, the reward simplifies to exact-answer matching. Otherwise, the threshold ϵ allows the model's predicted scores to fluctuate within an acceptable range, rather than strictly requiring exact accuracy. As depicted in Fig. 2, the predicted score receives a reward of 1 if it lies within the threshold ϵ of the ground-truth MOS, and 0 otherwise.

Rewards for degradation perception task. We find that training solely with score labels leads to poor perception of detailed image degradations (e.g., JPEG compression). This may be because generic multimodal models are pre-trained primarily to capture high-level semantic information, causing them to ignore subtle low-level distortions. To address this issue, we jointly train the model with a degradation perception task, leveraging easily obtainable degradation labels, thus enhancing the model's sensitivity to these image degradations. In this task, the model is required to predict both the distortion class and the corresponding distortion level. Since distortion class and level are inherently discrete variables, we similarly design binary rewards for this task. Denoting the predicted distortion class and level of the i-th response as $\text{deg}_{\text{pred}}^{(i)}$ and $\text{lev}_{\text{pred}}^{(i)}$ respectively, we define the degradation classification reward as follows. The reward value $r_{\text{deg}}^{(i)}$ is determined by:

$$r_{\text{deg}}^{(i)} = 1 \quad \text{if } \text{deg}_{\text{pred}}^{(i)} = \text{deg}_{\text{gt}}, \text{ otherwise } 0. \tag{4}$$

This means the predicted distortion class receives a reward of 1 if correct, and 0 otherwise, as illustrated in Fig. 2. Similarly, the intensity perception reward $r_{\text{lev}}^{(i)}$ is determined by:

$$r_{\text{lev}}^{(i)} = 1 \quad \text{if } \text{deg}_{\text{pred}}^{(i)} = \text{deg}_{\text{gt}} \text{ and } \text{lev}_{\text{pred}}^{(i)} = \text{lev}_{\text{gt}}, \text{ otherwise } 0. \tag{5}$$

As depicted in Fig. 2, the predicted distortion level earns a reward of 1 only when both the predicted class and the level exactly match the ground truth; otherwise, it receives 0.

Overall multi-task reward. Finally, the overall reward of i-th response is calculated as:

$$r^{(i)} = r_{\text{fmt}}^{(i)} + \mathbb{1}_{\text{scr}} \cdot r_{\text{scr}}^{(i)} + \mathbb{1}_{\text{deg}} \cdot \left(\alpha_1 \cdot r_{\text{deg}}^{(i)} + \alpha_2 \cdot r_{\text{lev}}^{(i)}\right), \tag{6}$$

where $\mathbb{1}_{\text{scr}}$ equals 1 if the score regression task is selected (and 0 otherwise), and similarly, $\mathbb{1}_{\text{deg}}$ equals 1 if the degradation perception task is selected (and 0 otherwise). Note that the reasoning process illustrated in Fig. 2 emerges naturally from the model's internal capability, without relying on external constraints or additional annotated data. After computing rewards for all generated responses $\{r^{(1)}, r^{(2)}, ..., r^{(N)}\}$, the policy model is updated following Eqs. (1) and (2). With this flexible design, Q-Insight can seamlessly switch between tasks and jointly optimize them during

 https://doi.org/10.52202/085713-1237

Table 1: **PLCC / SRCC comparison on the score regression tasks** between our Q-Insight and other competitive IQA methods. All methods except handcrafted ones are trained on the KonIQ dataset. The best and second-best results of each test setting are highlighted in **bold red** and underlined blue.

Category	Methods	KonIQ	SPAQ	KADID	PIPAL	LiveW	AGIQA	CSIQ	AVG.
Handcrafted	NIQE [29]	0.533	0.679	0.468	0.195	0.493	0.560	0.718	0.521
	(SPL 2012)	/0.530	/0.664	/0.405	/0.161	/0.449	/0.533	/0.628	/0.481
	BRISQUE [28]	0.225	0.490	0.429	0.267	0.361	0.541	0.740	0.436
	(TIP 2012)	/0.226	/0.406	/0.356	/0.232	/0.313	/0.497	/0.556	/0.369
Non-MLLM Deep-learning	NIMA [47]	0.896	0.838	0.532	0.390	0.814	0.715	0.695	0.697
	(TIP 2018)	/0.859	/0.856	/0.535	/0.399	/0.771	/0.654	/0.649	/0.675
	HyperIQA [44]	0.917	0.791	0.506	0.410	0.772	0.702	0.752	0.693
	(CVPR 2020)	/0.906	/0.788	/0.468	/0.403	/0.749	/0.640	/0.717	/0.667
	DBCNN [67]	0.884	0.812	0.497	0.384	0.773	0.730	0.586	0.667
	(TCSVT 2020)	/0.875	/0.806	/0.484	/0.381	/0.755	/0.641	/0.572	/0.645
	MUSIQ [19]	0.924	0.868	0.575	0.431	0.789	0.722	0.771	0.726
	(ICCV 2021)	/0.929	/0.863	/0.556	/0.431	/0.830	/0.630	/0.710	/0.707
	CLIP-IQA+ [49]	0.909	0.866	0.653	0.427	0.832	0.736	0.772	0.742
	(AAAI 2023)	/0.895	/0.864	/0.654	/0.419	/0.805	/0.685	/0.719	/0.720
	ManIQA [56]	0.849	0.768	0.499	0.457	0.849	0.723	0.623	0.681
	(CVPR 2022)	/0.834	/0.758	/0.465	/0.452	/0.832	/0.636	/0.627	/0.658
MLLM-based	C2Score [75]	0.923	0.867	0.500	0.354	0.786	0.777	0.735	0.706
	(NeurIPS 2024)	/0.910	/0.860	/0.453	/0.342	/0.772	/0.671	/0.705	/0.673
	Qwen-SFT [1]	0.889	0.874	0.668	0.473	0.734	**0.813**	0.674	0.732
	(Arxiv 2025)	/0.866	/0.875	/0.663	/0.442	/0.728	/0.739	/0.650	/0.709
	Q-Align [53]	0.941	0.886	0.674	0.403	0.853	0.772	0.671	0.705
	(ICML 2024)	/0.940	/0.887	/0.684	/0.419	/0.860	/0.735	/0.737	/0.752
	DeQA [58]	**0.953**	0.895	0.694	0.472	0.892	0.809	0.787	0.786
	(CVPR 2025)	/**0.941**	/0.896	/0.687	/**0.478**	/**0.879**	/0.729	/0.744	/0.765
	Q-Insight	0.933	**0.907**	**0.742**	**0.486**	**0.893**	0.811	**0.870**	**0.806**
	(Ours)	/0.916	/**0.905**	/**0.736**	/0.474	/0.865	/**0.764**	/**0.824**	/**0.783**

training. During inference, the trained policy model can directly perform image quality understanding without requiring additional fine-tuning. Experimental results presented in Tabs. 3 and 4 and Fig. 4 further demonstrate that jointly addressing the score regression and degradation perception tasks substantially improves performance, highlighting the beneficial interactions between these two tasks.

3.4 Data Construction

We construct multi-modal training data to jointly train Q-Insight on the score regression and degradation perception tasks. The prompts designed for each task are detailed in Tab. A in the appendix. For the score regression task, the input includes a task-specific prompt and the image to be rated, with the Mean Opinion Score (MOS) serving as the guideline to calculate the corresponding reward. In the degradation perception task, the input consists of a prompt and an image characterized by a specific distortion class and severity level. There are five distortion categories: "noise", "blur", "JPEG", "darken", and "null", where "null" indicates no distortion. Each distortion type has five severity levels: "slight", "moderate", "obvious", "serious", and "catastrophic". The distortion class and corresponding severity level constitute the ground-truth labels to calculate the degradation classification and intensity perception rewards. Overall, our carefully-designed GRPO-based framework and multi-task training strategy help Q-Insight achieve robust reasoning and perception capability even from limited annotated labels. More importantly, this flexible approach facilitates effective generalization to various low-level vision applications, as demonstrated by extensive experiments in Sec. 4, clearly highlighting the advantages of our framework in addressing practical vision tasks.

4 Experiments

4.1 Experimental Setup

Datasets and Metrics. For the score regression task, we use diverse IQA datasets across four categories: (a) in-the-wild datasets, including KonIQ [16], SPAQ [11], and LIVE-Wild [12]; (b)

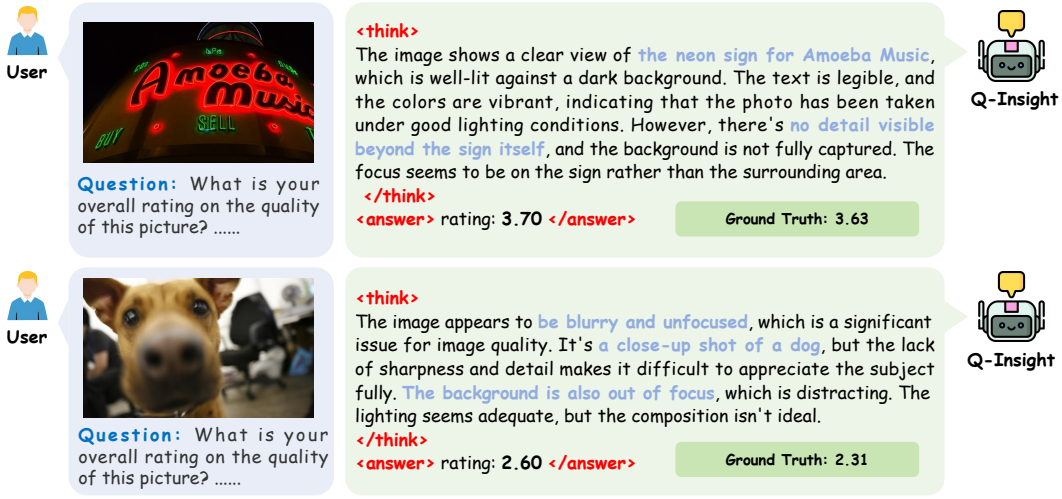


Figure 3: **Score rating and explanation results of our Q-Insight.** Q-Insight is capable of recognizing text, analyzing the lighting and shading conditions of an image, and understanding its composition.

Table 2: **Distortion prediction accuracy (Deg. Acc.) and degradation level accuracy (Lev. Acc.)** comparison between our Q-Insight and AgenticIR [76]. Our method outperforms AgenticIR across all degradations, especially in Noise and JPEG Compression.

Method	Metrics	Noise	Blur	JPEG	Darken	Null	Average
AgenticIR [76] (ICLR 2025)	Deg. Acc.	0.4646	0.8390	0.0135	0.7478	**0.9339**	0.5998
	Lev. Acc.	0.1858	0.3219	0.0000	0.2611	-	0.1922
Q-Insight (Ours)	Deg. Acc.	**1.0000**	**0.9756**	**1.0000**	**0.9027**	0.7603	**0.9277**
	Lev. Acc.	**0.5973**	**0.4438**	**0.5541**	**0.3230**	-	**0.4796**

synthetic distortion datasets, including KADID [22] and CSIQ [20]; (c) model-processed distortions, including PIPAL [14]; and (d) AI-generated images from AGIQA [21]. Following [58], we split KonIQ into training and test sets, with approximately 7000 training images. Mean Opinion Scores (MOS) across these datasets are normalized into the range $[1, 5]$. The remaining datasets are exclusively used to evaluate the model's out-of-distribution (OOD) generalization capability. For degradation perception task, we randomly select 7000 images from DQ-495K [59] that contain a single distortion for training, with an additional 1000 images reserved for testing. We adopt the pearson linear correlation coefficient (PLCC) and spearman rank-order correlation coefficient (SRCC) as metrics to evaluate performance on score regression task, following [19, 53, 58]. For degradation perception, we use the accuracy of distortion class and degradation level as evaluation metrics.

Implementation Details. We adopt Qwen-2.5-VL-7B-Instruct [1] as our base model. In the GRPO algorithm, the generation number N is set to 8, the weight of KL divergence penalty β is set to 1×10^{-3}, while the weights α_1 and α_2 are set to 0.25 and 0.75, respectively. The threshold ϵ is set to 0.35. We employ AdamW [26] as the optimizer, using an initial learning rate of 1×10^{-6} that linearly decays to 1×10^{-9} during training. The model is trained for 10 epochs with a total batch size of 128. Training is completed in approximately one day using 16 NVIDIA A100 GPUs.

4.2 Score Regression

We first evaluate our Q-Insight on the score regression task. We compare Q-Insight with handcrafted methods NIQE [29] and BRISQUE [28]; non-MLLM deep-learning methods including NIMA [47], MUSIQ [19], CLIP-IQA+ [49], and ManIQA [56]; and recent MLLM-based methods such as C2Score [75], Q-Align [53], DeQA-Score [58], and a supervised fine-tuned Qwen [1]. For a fair comparison, all methods (except handcrafted ones) are trained on the KonIQ dataset, and all MLLM-based methods utilize approximately 7B parameters. The comparison results in terms of PLCC and SRCC between Q-Insight and other IQA methods are presented in Tab. 1. Compared with the state-of-the-art method DeQA-Score, our Q-Insight performs slightly worse on the in-domain KonIQ

https://doi.org/10.52202/085713-1237

Table 3: **Ablation study on the score regression task** between multi-task and single-task training. Q-Insight with joint-training significantly outperforms score-only training on PLCC / SRCC metrics.

Method	KonIQ	SPAQ	KADID	PIPAL	LiveW	AGIQA	CSIQ	AVG.
Ours (Score-Only)	0.918 /0.895	0.903 /0.899	0.702 /0.702	0.458 /0.435	0.870 /0.839	**0.816** /**0.766**	0.685 /0.640	0.765 /0.739
Ours (Joint-Training)	**0.933** /**0.916**	**0.907** /**0.905**	**0.742** /**0.736**	**0.486** /**0.474**	**0.893** /**0.865**	0.811 /0.764	**0.870** /**0.824**	**0.806** /**0.783**

Table 4: **Ablation study on the degradation perception task** between multi-task and single-task training. Jointly training with score regression improves the accuracy of degradation perception.

Method	Metrics	Noise	Blur	JPEG	Darken	Null	Average
Ours (Degradation-Only)	Deg. Acc.	0.9867	0.9268	0.9685	0.8805	0.5702	0.8960
	Lev. Acc.	0.4343	0.3951	0.3108	0.2567	-	0.3492
Ours (Joint-Training)	Deg. Acc.	**1.0000**	**0.9756**	**1.0000**	**0.9027**	**0.7603**	**0.9277**
	Lev. Acc.	**0.5973**	**0.4438**	**0.5541**	**0.3230**	-	**0.4796**



Question: What's your overall rating on the quality of this picture? The rating should be a float between 1 and 5, rounded to two decimal places, with 1 representing very poor quality and 5 representing excellent quality.

w/o Joint Training

<think>
The image has good lighting and vibrant colors, which contribute to its overall appeal. However, there is noticeable noise present, which can detract from the image quality. The horse and rider appear slightly blurry, indicating motion or a lower resolution. Despite these issues, the image still maintains a certain level of clarity and detail, especially in the background. Therefore, I would rate the image quality as moderate.
</think>
<answer> rating: 3.50 </answer>

Joint Training

<think>
The image appears to be a digital illustration rather than a photograph, which can affect its quality. It has a somewhat pixelated appearance, suggesting it might not be high resolution. The colors are vibrant but the details are not sharp. The composition is interesting, capturing an action scene well, but the lack of detail and clarity brings down the overall quality.
</think>
<answer> rating: 2.70 </answer>

Figure 4: **Subjective ablation comparison** between joint multi-task training and w/o joint training on the explanation of image scoring. With joint training, our method can better perceive degradation cues in images (such as pixelated appearance), thereby improving the accuracy of quality assessment.

dataset. However, **on out-of-distribution (OOD) datasets, Q-Insight consistently outperforms all baseline methods across nearly all benchmarks**, achieving approximately 0.02 improvements in both PLCC and SRCC. This demonstrates the effectiveness and strong generalization capability of our approach. Fig. 3 illustrates two cases showing the reasoning capability in the score regression task. Specifically, our method goes beyond merely outputting numerical scores and provides detailed, structured reasoning. In the first case (top of Fig. 3), Q-Insight correctly identifies and analyzes textual information displayed on a neon sign, thoroughly examining details such as lighting conditions. In the second case (bottom of Fig. 3), Q-Insight demonstrates its strength in interpreting image composition aspects, such as the arrangement of visual elements and the primary focal point of the image. These examples further illustrate how Q-Insight advances beyond score regression task, offering valuable insights into image quality by examining various perceptual factors from multiple perspectives, ultimately contributing to a comprehensive understanding of image quality.

4.3 Distortion Perception

We further evaluate Q-Insight on the distortion perception task, comparing it with AgenticIR [76], which fine-tunes an MLLM to perform a similar distortion prediction function. The comparative results are presented in Tab. 2. Notably, AgenticIR requires sequential queries for each possible distortion type, whereas Q-Insight identifies distortion types **using only a single query**. Q-Insight consistently outperforms AgenticIR across nearly all distortion categories, resulting in significantly higher average accuracy. However, for the "null" category (no distortion), our performance is slightly

Table 5: **Ablation study on the threshold ϵ for the score regression task.** Q-Insight demonstrates robust and stable performance, indicating it does not require careful tuning of the threshold ϵ.

Method	KonIQ	SPAQ	KADID	PIPAL	LiveW	AGIQA	CSIQ
ϵ = 0.15	0.918/0.900	0.903/0.903	0.749/0.744	0.443/0.441	0.876/0.847	**0.835**/0.764	0.857/0.793
ϵ = 0.25	0.929/0.911	0.902/0.902	**0.750/0.752**	0.472/0.471	0.884/0.855	0.823/0.765	0.889/0.844
ϵ = 0.30	0.932/0.912	0.906/0.903	0.742/0.738	0.471/0.460	0.887/0.856	0.828/**0.773**	0.868/0.822
ϵ = 0.40	0.930/0.912	0.904/0.901	0.721/0.723	0.446/0.428	0.880/0.853	0.819/0.769	0.873/0.831
ϵ = 0.50	0.928/0.906	**0.906**/0.902	**0.750**/0.746	0.475/0.472	0.885/0.854	0.821/0.768	**0.894/0.850**
ϵ = 0.35 **(Ours)**	**0.933/0.916**	**0.907/0.905**	0.742/0.736	**0.486/0.474**	**0.893/0.865**	0.811/0.764	0.870/0.824

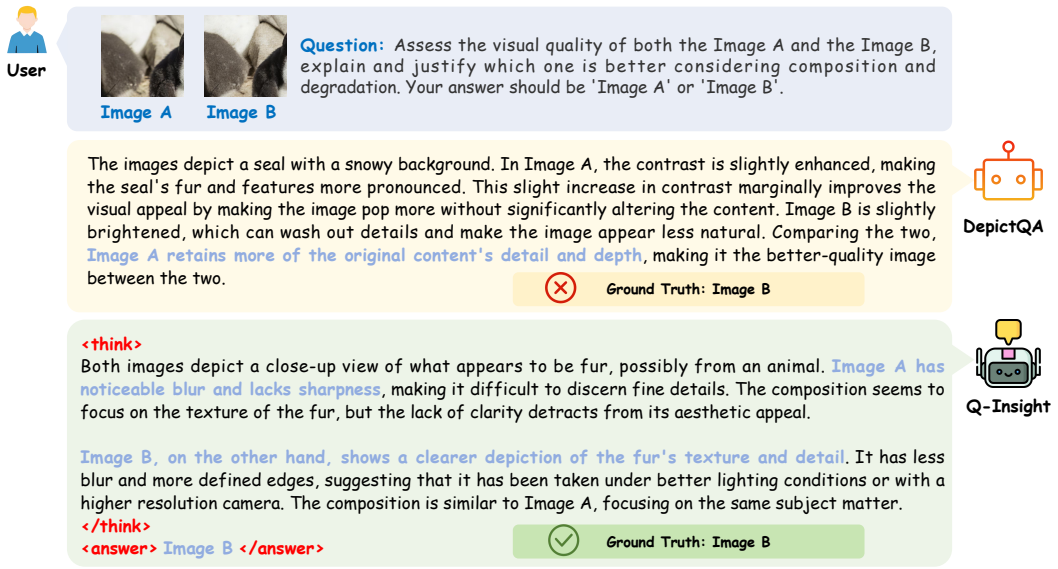


Figure 5: **Image comparison reasoning results** of our Q-Insight and DepictQA [60]. Q-Insight outperforms DepictQA in comprehensive content understanding and accurate degradation perception.

lower than AgenticIR, possibly because the reasoning mechanism of Q-Insight is more sensitive to the presence of potential distortions. Overall, these results highlight Q-Insight's ability to efficiently and accurately perceive image distortions through a concise and unified reasoning framework.

4.4 Ablation Studies

Effect of multi-task training. To validate the effectiveness of multi-task training, we compare our jointly trained Q-Insight model with two single-task variants, each trained independently on a single task. The comparison results are presented in Tabs. 3 and 4. As shown in Tab. 3, the jointly trained Q-Insight significantly outperforms the score-only variant on nearly all datasets, especially on datasets involving synthetic distortions (KADID [22], CSIQ [20]) and those containing model-generated distortions (PIPAL [14]). This demonstrates that incorporating the degradation perception task can effectively enhance performance in the score regression task. Fig. 4 further presents the benefits of multi-task training, showing that Q-Insight can precisely identify detailed degradations such as pixel-level artifacts, thereby improving overall accuracy in quality assessment. Similarly, Tab. 4 indicates that in the degradation perception task, our jointly trained model consistently surpasses the degradation-only variant across all distortion types. This suggests that the score regression task also positively contributes to degradation perception capabilities. These experimental results verify the mutual benefit and effectiveness of the proposed multi-task training strategy. Moreover, our findings clearly show that the visual quality understanding potential of MLLMs can be significantly improved through carefully designed training tasks and learning objectives.

Ablation on the score threshold ϵ. Introducing the threshold allows the model's predictions to vary within an acceptable margin. Tab. 5 reports the ablation results for different choices of ϵ. Q-Insight consistently achieves robust and stable performance across various threshold values, demonstrating that its effectiveness does not depend on careful tuning of ϵ.

https://doi.org/10.52202/085713-1237

Table 6: **Accuracy and PLCC / SRCC results of the reference-based comparison task on the SRbench [6].** Reg-Acc and Gen-Acc represent the accuracy between regression-based and generation-based restoration methods, respectively. Q-Insight outperforms score- and description-based methods.

Category	Method	Reg-Acc	Gen-Acc	Overall-Acc	PLCC	SRCC
Score-Based	PSNR	80.07%	41.70%	34.70%	–	–
	SSIM [50] (TIP 04)	83.00%	45.30%	37.40%	–	–
	LPIPS [66] (CVPR 18)	82.00%	63.90%	65.80%	–	–
	A-FINE [6] (CVPR 25)	83.30%	**78.90%**	82.40%	–	–
Description-Based	DepictQA [60] (ECCV 24)	73.00%	61.64%	62.96%	0.3457	0.3412
	Q-Insight (Zero-Shot)	78.67%	68.64%	75.51%	0.6385	0.6297
	Q-Insight (Trained)	**85.67%**	77.78%	**82.80%**	0.7627	0.7614

4.5 Image Comparison Reasoning

Our Q-Insight effectively generalizes to zero-shot image comparison reasoning tasks in both reference-based and non-reference-based scenarios, as illustrated in Figs. 1 and 5. Specifically, Fig. 1 shows a reference-based comparison scenario, where the reference image is of lower quality, and Images A and B are outputs generated by two different super-resolution methods. Fig. 5 demonstrates Q-Insight's superiority over DepictQA [60], highlighting its enhanced content understanding and precise perception of degradations. These examples illustrate Q-Insight's robust generalization ability, largely enabled by its RL-based framework and multi-task training strategy.

Furthermore, the comparison reasoning performance can be further boosted by training on a small number of labeled comparison pairs using reinforcement learning. Specifically, we randomly sample 5k data pairs from the DiffIQA [6] dataset, where each pair is labeled only with comparison results, without any textual descriptions. Results are shown in Tab. 6. Our Q-Insight consistently surpasses all score-based and description-based methods in terms of overall accuracy and PLCC/SRCC metrics, demonstrating its promising applicability to various image enhancement tasks. Notably, even the zero-shot version of Q-Insight substantially outperforms DepictQA [60], despite the latter relying on large-scale textual datasets. Additionally, A-Fine [6] utilizes more than 200k data pairs collected from four different datasets, combined with a complex three-stage training pipeline, thus requiring over 40 times more data and considerable effort to develop optimal training strategies. In contrast, Q-Insight achieves superior performance through relatively straightforward yet highly effective visual reinforcement learning. Further details and additional results are provided in Sec. B of the Appendix.

5 Conclusion

In this paper, we introduce Q-Insight, a novel GRPO-based model for comprehensive image quality understanding. It jointly optimizes score regression and degradation perception tasks using only a limited amount of labeled data. Unlike traditional methods that rely on extensive textual annotations or purely numerical scoring, our framework combines numerical accuracy with interpretative reasoning, significantly improving the perceptual analysis capabilities of image quality models. Extensive experiments show that Q-Insight consistently outperforms existing state-of-the-art methods across various datasets and tasks, demonstrating impressive zero-shot generalization and superior comparison reasoning capability. Looking ahead, Q-Insight can extend its capabilities to a wide range of tasks, such as image aesthetic assessments, and serve as a powerful discriminative signal to improve image enhancement models. As a unified model for scoring, perception, comparison, and reasoning, Q-Insight can act as a central hub, coordinating image reconstruction tools and providing valuable insights into the enhancement process. This integrated and automated system has the potential to revolutionize image quality understanding and enhancement, providing a unified solution that can transform how image quality is evaluated, improved, and applied across various fields.

Limitations. While achieving promising performance, Q-Insight focuses primarily on natural images. Extending to AI-generated images and videos remains essential and is reserved for future exploration. Besides, using a fixed threshold and discrete distortion levels is not the most elegant solution and may allow for more principled approaches. These issues warrant further exploration.

References

[1] 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.

[2] Yuntao Bai, Saurav Kadavath, Sandipan Kundu, Amanda Askell, Jackson Kernion, Andy Jones, Anna Chen, Anna Goldie, Azalia Mirhoseini, Cameron McKinnon, et al. Constitutional ai: Harmlessness from ai feedback. *arXiv preprint arXiv:2212.08073*, 2022.

[3] Sebastian Bosse, Dominique Maniry, Klaus-Robert Müller, Thomas Wiegand, and Wojciech Samek. Deep neural networks for no-reference and full-reference image quality assessment. *IEEE Transactions on Image Processing (TIP)*, 27(1):206–219, 2017.

[4] Yue Cao, Zhaolin Wan, Dongwei Ren, Zifei Yan, and Wangmeng Zuo. Incorporating semi-supervised and positive-unlabeled learning for boosting full reference image quality assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5851–5861, 2022.

[5] Chaofeng Chen, Sensen Yang, Haoning Wu, Liang Liao, Zicheng Zhang, Annan Wang, Wenxiu Sun, Qiong Yan, and Weisi Lin. Q-ground: Image quality grounding with large multi-modality models. In *Proceedings of the ACM International Conference on Multimedia (ACM MM)*, pages 486–495, 2024.

[6] Du Chen, Tianhe Wu, Kede Ma, and Lei Zhang. Toward generalized image quality assessment: Relaxing the perfect reference quality assumption. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025.

[7] Paul F Christiano, Jan Leike, Tom Brown, Miljan Martic, Shane Legg, and Dario Amodei. Deep reinforcement learning from human preferences. *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)*, 30, 2017.

[8] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)*, 34:8780–8794, 2021.

[9] Keyan Ding, Yi Liu, Xueyi Zou, Shiqi Wang, and Kede Ma. Locally adaptive structure and texture similarity for image quality assessment. In *Proceedings of the ACM International Conference on Multimedia (ACM MM)*, pages 2483–2491, 2021.

[10] 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 (TPAMI)*, 44(5):2567–2581, 2020.

[11] Yuming Fang, Hanwei Zhu, Yan Zeng, Kede Ma, and Zhou Wang. Perceptual quality assessment of smartphone photography. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3677–3686, 2020.

[12] Deepti Ghadiyaram and Alan C Bovik. Live in the wild image quality challenge database. *Online: http://live. ece. utexas. edu/research/ChallengeDB/index. html [Mar, 2017]*, 2015.

[13] Abhijay Ghildyal and Feng Liu. Shift-tolerant perceptual similarity metric. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 91–107. Springer, 2022.

[14] Jinjin GU, Haoming Cai, Haoyu Chen, Xiaoxing Ye, Ren Jimmy S, and Chao Dong. Pipal: a large-scale image quality assessment dataset for perceptual image restoration. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 633–651. Springer, 2020.

[15] Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, et al. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*, 2025.

[16] Vlad Hosu, Hanhe Lin, Tamas Sziranyi, and Dietmar Saupe. Koniq-10k: An ecologically valid database for deep learning of blind image quality assessment. *IEEE Transactions on Image Processing (TIP)*, 29:4041–4056, 2020.

 https://doi.org/10.52202/085713-1237

[17] Binyuan Hui, Jian Yang, Zeyu Cui, Jiaxi Yang, Dayiheng Liu, Lei Zhang, Tianyu Liu, Jiajun Zhang, Bowen Yu, Keming Lu, et al. Qwen2.5-coder technical report. *arXiv preprint arXiv:2409.12186*, 2024.

[18] Le Kang, Peng Ye, Yi Li, and David Doermann. Convolutional neural networks for no-reference image quality assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1733–1740, 2014.

[19] 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 (ICCV)*, pages 5148–5157, 2021.

[20] Eric C Larson and Damon M Chandler. Most apparent distortion: full-reference image quality assessment and the role of strategy. *Journal of Electronic Imaging*, 19(1):011006–011006, 2010.

[21] Chunyi Li, Zicheng Zhang, Haoning Wu, Wei Sun, Xiongkuo Min, Xiaohong Liu, Guangtao Zhai, and Weisi Lin. Agiqa-3k: An open database for ai-generated image quality assessment. *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, 34(8):6833–6846, 2023.

[22] Hanhe Lin, Vlad Hosu, and Dietmar Saupe. Kadid-10k: A large-scale artificially distorted iqa database. In *Proceedings of International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–3. IEEE, 2019.

[23] Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. Visual instruction tuning. *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)*, 36:34892–34916, 2023.

[24] Xialei Liu, Joost Van De Weijer, and Andrew D Bagdanov. Rankiqa: Learning from rankings for no-reference image quality assessment. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 1040–1049, 2017.

[25] Ziyu Liu, Zeyi Sun, Yuhang Zang, Xiaoyi Dong, Yuhang Cao, Haodong Duan, Dahua Lin, and Jiaqi Wang. Visual-rft: Visual reinforcement fine-tuning. *arXiv preprint arXiv:2503.01785*, 2025.

[26] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.

[27] Chao Ma, Chih-Yuan Yang, Xiaokang Yang, and Ming-Hsuan Yang. Learning a no-reference quality metric for single-image super-resolution. *Computer Vision and Image Understanding*, 158:1–16, 2017.

[28] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing (TIP)*, 21(12):4695–4708, 2012.

[29] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a "completely blind" image quality analyzer. *IEEE Signal Processing Letters*, 20(3):209–212, 2012.

[30] Anush Krishna Moorthy and Alan Conrad Bovik. A two-step framework for constructing blind image quality indices. *IEEE Signal Processing Letters*, 17(5):513–516, 2010.

[31] Anush Krishna Moorthy and Alan Conrad Bovik. Blind image quality assessment: From natural scene statistics to perceptual quality. *IEEE Transactions on Image Processing (TIP)*, 20(12):3350–3364, 2011.

[32] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. Training language models to follow instructions with human feedback. *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)*, 35:27730–27744, 2022.

[33] Da Pan, Ping Shi, Ming Hou, Zefeng Ying, Sizhe Fu, and Yuan Zhang. Blind predicting similar quality map for image quality assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6373–6382, 2018.

[34] Jiazhen Pan, Che Liu, Junde Wu, Fenglin Liu, Jiayuan Zhu, Hongwei Bran Li, Chen Chen, Cheng Ouyang, and Daniel Rueckert. Medvlm-r1: Incentivizing medical reasoning capability of vision-language models (vlms) via reinforcement learning. *arXiv preprint arXiv:2502.19634*, 2025.

[35] Nikolay Ponomarenko, Lina Jin, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, Jaakko Astola, Benoit Vozel, Kacem Chehdi, Marco Carli, Federica Battisti, and C.-C. Jay Kuo. Image database tid2013: Peculiarities, results and perspectives. *Signal Processing: Image Communication*, 30:57–77, 2015.

[36] Ekta Prashnani, Hong Cai, Yasamin Mostofi, and Pradeep Sen. Pieapp: Perceptual image-error assessment through pairwise preference. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1808–1817, 2018.

[37] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10684–10695, 2022.

[38] Michele A Saad, Alan C Bovik, and Christophe Charrier. Blind image quality assessment: A natural scene statistics approach in the dct domain. *IEEE Transactions on Image Processing (TIP)*, 21(8):3339–3352, 2012.

[39] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

[40] Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, YK Li, Y Wu, et al. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*, 2024.

[41] H Sheikh. Live image quality assessment database release 2. *http://live. ece. utexas. edu/research/quality*, 2005.

[42] Hamid R Sheikh and Alan C Bovik. Image information and visual quality. *IEEE Transactions on Image Processing (TIP)*, 15(2):430–444, 2006.

[43] David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, et al. Mastering the game of go without human knowledge. *Nature*, 550(7676):354–359, 2017.

[44] Shaolin Su, Qingsen Yan, Yu Zhu, Cheng Zhang, Xin Ge, Jinqiu Sun, and Yanning Zhang. Blindly assess image quality in the wild guided by a self-adaptive hyper network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3667–3676, 2020.

[45] Simeng Sun, Tao Yu, Jiahua Xu, Wei Zhou, and Zhibo Chen. Graphiqa: Learning distortion graph representations for blind image quality assessment. *IEEE Transactions on Multimedia (TMM)*, 25:2912–2925, 2022.

[46] Zhiqing Sun, Sheng Shen, Shengcao Cao, Haotian Liu, Chunyuan Li, Yikang Shen, Chuang Gan, Liang-Yan Gui, Yu-Xiong Wang, Yiming Yang, et al. Aligning large multimodal models with factually augmented rlhf. *arXiv preprint arXiv:2309.14525*, 2023.

[47] Hossein Talebi and Peyman Milanfar. Nima: Neural image assessment. *IEEE Transactions on Image Processing (TIP)*, 27(8):3998–4011, 2018.

[48] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*, 2023.

 https://doi.org/10.52202/085713-1237

[49] 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 (AAAI)*, volume 37, pages 2555–2563, 2023.

[50] 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 (TIP)*, 13(4):600–612, 2004.

[51] Haoning Wu, Zicheng Zhang, Erli Zhang, Chaofeng Chen, Liang Liao, Annan Wang, Chunyi Li, Wenxiu Sun, Qiong Yan, Guangtao Zhai, et al. Q-bench: A benchmark for general-purpose foundation models on low-level vision. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2025.

[52] Haoning Wu, Zicheng Zhang, Erli Zhang, Chaofeng Chen, Liang Liao, Annan Wang, Kaixin Xu, Chunyi Li, Jingwen Hou, Guangtao Zhai, et al. Q-instruct: Improving low-level visual abilities for multi-modality foundation models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 25490–25500, 2024.

[53] Haoning Wu, Zicheng Zhang, Weixia Zhang, Chaofeng Chen, Liang Liao, Chunyi Li, Yixuan Gao, Annan Wang, Erli Zhang, Wenxiu Sun, et al. Q-align: Teaching LMMs for visual scoring via discrete text-defined levels. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2024.

[54] Haoning Wu, Hanwei Zhu, Zicheng Zhang, Erli Zhang, Chaofeng Chen, Liang Liao, Chunyi Li, Annan Wang, Wenxiu Sun, Qiong Yan, et al. Towards open-ended visual quality comparison. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 360–377. Springer, 2024.

[55] An Yang, Beichen Zhang, Binyuan Hui, Bofei Gao, Bowen Yu, Chengpeng Li, Dayiheng Liu, Jianhong Tu, Jingren Zhou, Junyang Lin, et al. Qwen2.5-math technical report: Toward mathematical expert model via self-improvement. *arXiv preprint arXiv:2409.12122*, 2024.

[56] 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 (CVPR)*, pages 1191–1200, 2022.

[57] Huaiyuan Ying, Shuo Zhang, Linyang Li, Zhejian Zhou, Yunfan Shao, Zhaoye Fei, Yichuan Ma, Jiawei Hong, Kuikun Liu, Ziyi Wang, et al. Internlm-math: Open math large language models toward verifiable reasoning. *arXiv preprint arXiv:2402.06332*, 2024.

[58] Zhiyuan You, Xin Cai, Jinjin Gu, Tianfan Xue, and Chao Dong. Teaching large language models to regress accurate image quality scores using score distribution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025.

[59] Zhiyuan You, Jinjin Gu, Zheyuan Li, Xin Cai, Kaiwen Zhu, Chao Dong, and Tianfan Xue. Descriptive image quality assessment in the wild. *arXiv preprint arXiv:2405.18842*, 2024.

[60] 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 *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 259–276. Springer, 2024.

[61] Tianyu Yu, Yuan Yao, Haoye Zhang, Taiwen He, Yifeng Han, Ganqu Cui, Jinyi Hu, Zhiyuan Liu, Hai-Tao Zheng, Maosong Sun, et al. Rlhf-v: Towards trustworthy mllms via behavior alignment from fine-grained correctional human feedback. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 13807–13816, 2024.

[62] Tianyu Yu, Haoye Zhang, Yuan Yao, Yunkai Dang, Da Chen, Xiaoman Lu, Ganqu Cui, Taiwen He, Zhiyuan Liu, Tat-Seng Chua, et al. Rlaif-v: Aligning mllms through open-source ai feedback for super gpt-4v trustworthiness. *arXiv preprint arXiv:2405.17220*, 2024.

[63] Jingyi Zhang, Jiaxing Huang, Huanjin Yao, Shunyu Liu, Xikun Zhang, Shijian Lu, and Dacheng Tao. R1-vl: Learning to reason with multimodal large language models via step-wise group relative policy optimization. *arXiv preprint arXiv:2503.12937*, 2025.

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

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

[66] 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/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 586–595, 2018.

[67] Weixia Zhang, Kede Ma, Jia Yan, Dexiang Deng, and Zhou Wang. Blind image quality assessment using a deep bilinear convolutional neural network. *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, 2020.

[68] Yuxiang Zhang, Shangxi Wu, Yuqi Yang, Jiangming Shu, Jinlin Xiao, Chao Kong, and Jitao Sang. o1-coder: an o1 replication for coding. *arXiv preprint arXiv:2412.00154*, 2024.

[69] Zicheng Zhang, Ziheng Jia, Haoning Wu, Chunyi Li, Zijian Chen, Yingjie Zhou, Wei Sun, Xiaohong Liu, Xiongkuo Min, Weisi Lin, et al. Q-bench-video: Benchmarking the video quality understanding of lmms. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025.

[70] Zicheng Zhang, Tengchuan Kou, Shushi Wang, Chunyi Li, Wei Sun, Wei Wang, Xiaoyu Li, Zongyu Wang, Xuezhi Cao, Xiongkuo Min, et al. Q-eval-100k: Evaluating visual quality and alignment level for text-to-vision content. *arXiv preprint arXiv:2503.02357*, 2025.

[71] Zicheng Zhang, Haoning Wu, Ziheng Jia, Weisi Lin, and Guangtao Zhai. Teaching lmms for image quality scoring and interpreting. *arXiv preprint arXiv:2503.09197*, 2025.

[72] Zhiyuan Zhao, Bin Wang, Linke Ouyang, Xiaoyi Dong, Jiaqi Wang, and Conghui He. Beyond hallucinations: Enhancing lvlms through hallucination-aware direct preference optimization. *arXiv preprint arXiv:2311.16839*, 2023.

[73] Heliang Zheng, Huan Yang, Jianlong Fu, Zheng-Jun Zha, and Jiebo Luo. Learning conditional knowledge distillation for degraded-reference image quality assessment. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 10242–10251, 2021.

[74] Hancheng Zhu, Leida Li, Jinjian Wu, Weisheng Dong, and Guangming Shi. Metaiqa: Deep meta-learning for no-reference image quality assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14143–14152, 2020.

[75] Hanwei Zhu, Haoning Wu, Yixuan Li, Zicheng Zhang, Baoliang Chen, Lingyu Zhu, Yuming Fang, Guangtao Zhai, Weisi Lin, and Shiqi Wang. Adaptive image quality assessment via teaching large multimodal model to compare. In *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)*, 2024.

[76] Kaiwen Zhu, Jinjin Gu, Zhiyuan You, Yu Qiao, and Chao Dong. An intelligent agentic system for complex image restoration problems. *Proceedings of the International Conference on Learning Representations (ICLR)*, 2025.

 https://doi.org/10.52202/085713-1237

A Designed Prompts

The prompts designed for each task are detailed in Tab. A. Specifically, a general system prompt is shared across tasks, which encourages the model to explicitly output its reasoning process and provide structured responses. This general prompt is supplemented by task-specific prompts tailored for score regression and degradation perception, respectively. For the score regression task, the input includes a task-specific prompt and the image to be rated, with the Mean Opinion Score (MOS) serving as the ground-truth. In the degradation perception task, the input consists of a prompt and an image characterized by a specific distortion class and severity level. We define five distortion categories: “noise”, “blur”, “JPEG”, “darken”, and “null”, where “null” indicates no distortion. Each distortion type has five severity levels: “slight”, “moderate”, “obvious”, “serious”, and “catastrophic”. The distortion class and corresponding severity level constitute the ground-truth labels. Additionally, for the comparative reasoning scenario, the inputs include a prompt, two images to be compared, and an optional reference image.

Table A: **Prompts for Different Tasks.** The system prompt is shared across all tasks, while task-specific prompts are additionally designed for each individual task.

System Prompt: A conversation between User and Assistant. The user asks a question, and the Assistant solves it. The assistant first thinks about the reasoning process in the mind and then provides the user with the answer. The reasoning process and answer are enclosed within <think> </think> and <answer> </answer> tags, respectively, i.e., <think> reasoning process here </think><answer> answer here </answer>.
Prompt for Score Regression Task: What is your overall rating on the quality of this picture? The rating should be a float between 1 and 5, rounded to two decimal places, with 1 representing very poor quality and 5 representing excellent quality. Return the final answer in JSON format with the following keys: "rating": The score.
Prompt for Degradation Perception Task: Analyze the given image and determine if it contains any of the following distortions: "noise", "compression", "blur", or "darken". If a distortion is present, classify its severity as "slight", "moderate", "obvious", "serious", or "catastrophic". Return the result in JSON format with the following keys: "distortion_class": The detected distortion (or "null" if none). and "severity": The severity level (or "null" if none).
Prompt for Non-Reference-Based Image Comparison: Given Image A: <image_A> and <Image_B>, assess the visual quality of both the Image A and the Image B, explain and justify which one is better considering composition and degradation. Your answer should be "Image A" or "Image B".
Prompt for Reference-Based Comparison Reasoning: Given a low-quality reference image and two enhanced outputs. Reference Image: <ref_image>, Image A: <image_A> and Image B: <image_B>. Decide which enhanced image is superior or if they are comparable. Evaluate based on: 1) fidelity and consistency with the reference image; 2) overall perceptual quality. Return exactly one of: "Image A", "Image B", or "Similar".

https://doi.org/10.52202/085713-1237

B Experimental Setups and More Results of Comparison Reasoning

B.1 Datasets

For training, we use the DiffIQA [6] dataset, which contains approximately 180k reference–test image pairs generated by applying diffusion-based enhancement methods to reference images of varying quality. Each reference image is paired with multiple test images, and human annotators provide preference labels through triplet-based comparisons, resulting in roughly 180k comparison pairs. Notably, A-Fine [6] aggregates datasets from DiffIQA, TID2013 [35], KADID [22], and PIPAL [14], totaling over 200k comparison pairs for training. In contrast, we randomly sample only 5k pairs from DiffIQA for our training.

For evaluation, we adopt SRIQA-Bench [6], a benchmark specifically designed to evaluate the generalization ability of IQA models in real-world super-resolution (SR) scenarios. This dataset includes 100 low-resolution (LR) reference images, each enhanced by 10 distinct SR methods encompassing both regression-based and generative models. Human raters perform exhaustive pairwise comparisons within each image group, with each comparison annotated by at least 10 annotators to ensure reliability. Since no ground-truth high-resolution reference images are provided, models must evaluate perceptual quality based solely on the degraded LR images. Thus, SRIQA-Bench poses a challenging scenario to rigorously assess the robustness of full-reference IQA models under imperfect reference conditions. we report Reg-Acc, Gen acc, and overall ACC, which denote pairwise ranking accuracy on regression-based SR methods, generation-based SR methods, and all SR outputs, respectively. These metrics measure alignment between model predictions and human judgments under varying super-resolution styles and distortion characteristics. We also report the PLCC and SRCC of Description-based methods.

B.2 Reward Design and Training Details

In the image comparison task, the model is expected to determine which image is superior or if they are comparable. Denote the predicted score of the i-th response as $\text{res}_{\text{pred}}^{(i)}$, we design a comparison reward function R_{comp} as follows. The reward value $r_{\text{comp}}^{(i)}$ for the i-th response is determined by :

$$r_{\text{comp}}^{(i)} = 1 \quad \text{if } \text{res}_{\text{pred}}^{(i)} = \text{res}_{\text{comp}}, \text{ otherwise } 0. \tag{7}$$

Finally, the overall reward of i-th response is calculated as:

$$r^{(i)} = r_{\text{fmt}}^{(i)} + r_{\text{comp}}^{(i)}, \tag{8}$$

where $r_{\text{fmt}}^{(i)}$ is the format reward. For training, we employ AdamW [26] as the optimizer, using an initial learning rate of 1×10^{-6} that linearly decays to 1×10^{-9} during training. The model is trained for 8 epochs with a total batch size of 32. Training is completed in approximately 20 hours using 16 NVIDIA A100 GPUs.

B.3 More Qualitative Results

We provide quantitative results in Tab. 6 in the main paper. Here we present additional qualitative results of reference-based image comparison reasoning in Figs. A, B, C, and D. It can be observed that, despite training on only 5k comparison pairs, Q-Insight surpasses DepictQA [60] by more effectively recognizing subtle image details and distortions, thereby producing more accurate outcomes.

https://doi.org/10.52202/085713-1237

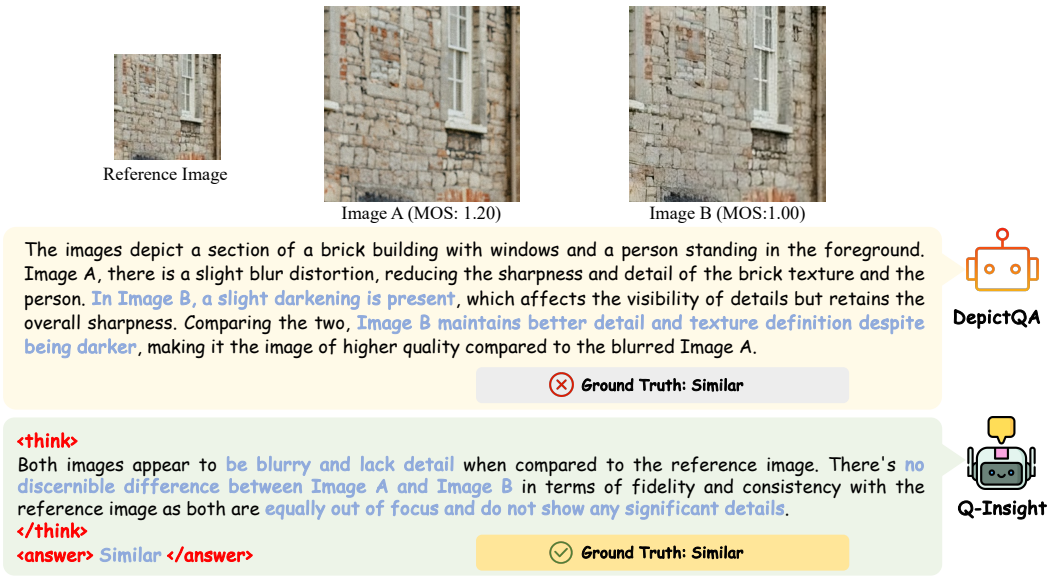


Figure A: **Reference-based image comparison reasoning results** of our Q-Insight and DepictQA [60] on image "online4" from SRBench [6].

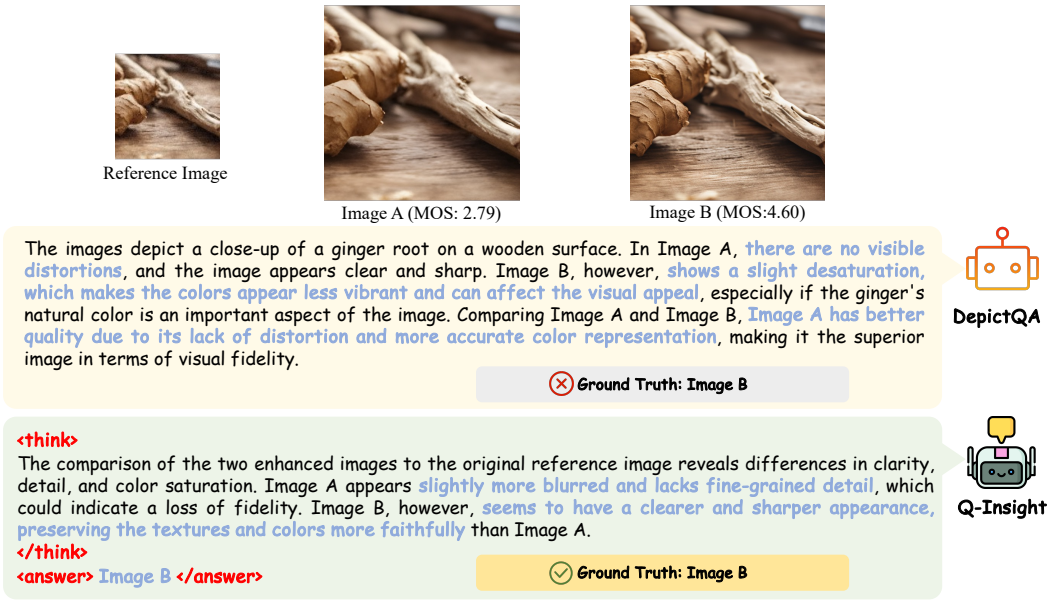


Figure B: **Reference-based image comparison reasoning results** of our Q-Insight and DepictQA [60] on image "online23" from SRBench [6].

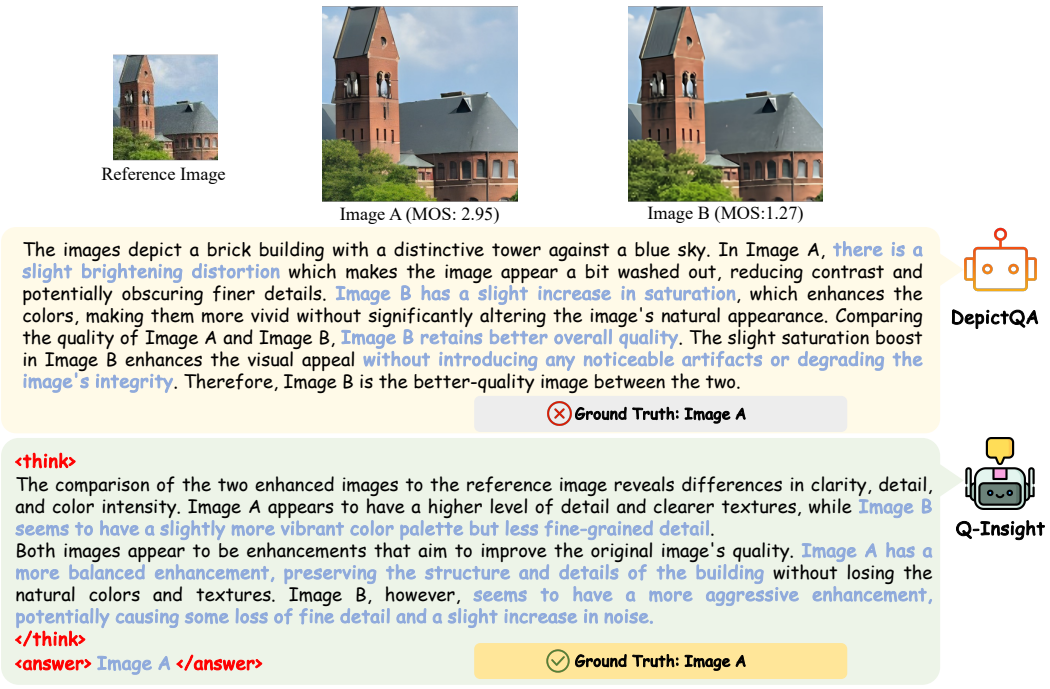


Figure C: **Reference-based image comparison reasoning results** of our Q-Insight and DepictQA [60] on image "realworld33" from SRBench [6].

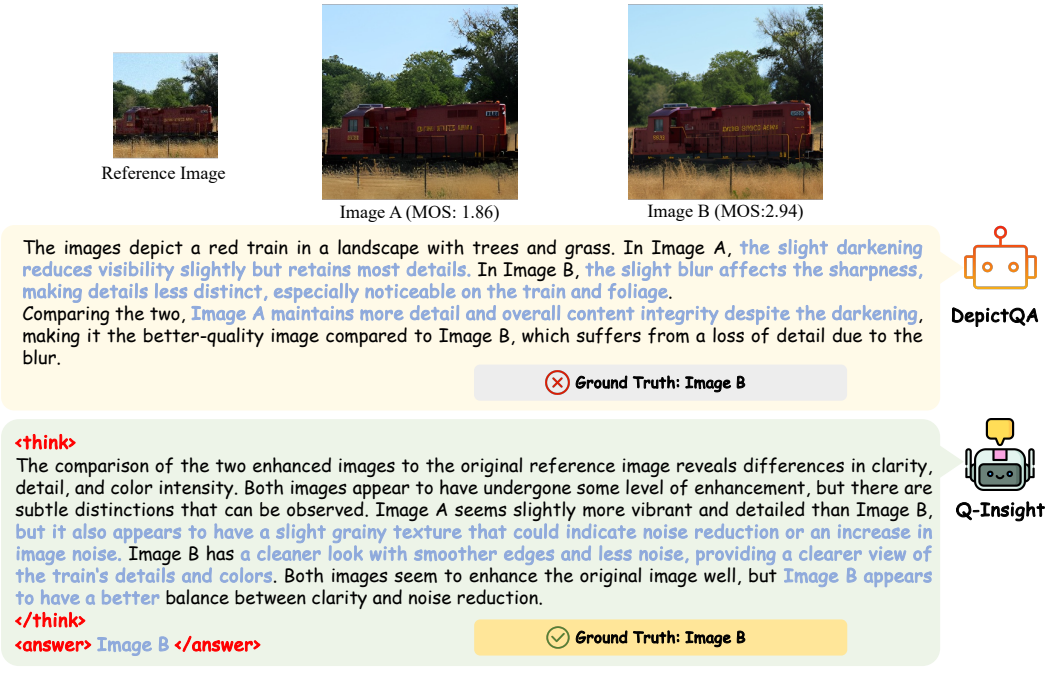


Figure D: **Reference-based image comparison reasoning results** of our Q-Insight and DepictQA [60] on image "srtest55" from SRBench [6].

https://doi.org/10.52202/085713-1237

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: We confirm that the abstract and introduction have clearly stated the claims made in this paper. We have also provided a list of contribution at the end of introduction.

   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: We have discussed the limitations at the end of the paper.

   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: We have provided the data, implementation and training details in Sec. 4.1 of the main paper and Sec. A and Sec. B of the Appendix. We also confirm that the code and pretrained models will be released for reproducible research.

   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**

https://doi.org/10.52202/085713-1237

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: The code and models are available at `https://github.com/bytedance/Q-Insight`.

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: We have provided detailed experimental settings in Sec. 4.1 of the main paper and Sec. A and Sec. B of 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: Error bars are not reported because it would be too computationally expensive.

   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: We have provided sufficient information on the computer resources in Sec. 4.1 of the main paper.

   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 confirm that the research conducted in the paper conform, in every respect, with 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: [NA]

    Justification: This paper focuses solely on technical improvements in image quality assessment and does not have societal impacts.

    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.

https://doi.org/10.52202/085713-1237

- 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: [NA]

    Justification: This paper focuses solely on technical improvements in image quality assessment and does not pose such risks.

    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: All datasets used in the paper are properly cited, and their usage complies with the Apache License 2.0.

    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: [NA]

    Justification: This paper does not release new assets.

    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.

 https://doi.org/10.52202/085713-1237

- 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: [NA]

    Justification: The core method development in this research does not involve LLMs as any important, original, or non-standard components.

    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.