Zoom-IQA: Image Quality Assessment with Reliable Region-Aware Reasoning

Guoqiang Liang[1], Jianyi Wang[1], Zhonghua Wu[2], Shangchen Zhou[1*], and Chen Change Loy[1*]

[1] S-Lab, Nanyang Technological University, Singapore
[2] SenseTime Research
Project Page: https://ethanliang99.github.io/ZOOMIQA-Projectpage/

Abstract. Image Quality Assessment (IQA) is a long-standing problem in computer vision. Previous methods typically focus on predicting numerical scores without explanation or provide low-level descriptions lacking precise scores. Recent reasoning-based vision language models (VLMs) have shown strong potential for IQA, enabling joint generation of quality descriptions and scores. However, we notice that existing VLM-based IQA methods tend to exhibit unreliable reasoning due to their limited capability of integrating visual and textual cues. In this work, we introduce Zoom-IQA, a VLM-based IQA model to explicitly emulate key cognitive behaviors: uncertainty awareness, region reasoning, and iterative refinement. Specifically, we present a two-stage training pipeline: 1) supervised fine-tuning (SFT) on our Grounded-Rationale-IQA (GR-IQA) dataset to teach the model to ground its assessments in key regions; and 2) reinforcement learning (RL) for dynamic policy exploration, primarily stabilized by our KL-Coverage regularizer to prevent reasoning and scoring diversity collapse, and supported by a Progressive Re-sampling Strategy to mitigate annotation bias. Extensive experiments show that Zoom-IQA achieves improved robustness, explainability, and generalization. The application to downstream tasks, such as image restoration, further demonstrates the effectiveness of Zoom-IQA.

Keywords: Image Quality Assessment · Multimodal Large Language Models · Reinforcement Learning

1 Introduction

Image Quality Assessment (IQA) is a fundamental task in computer vision, aiming to evaluate the perceptual quality of images in alignment with human perception. Its importance has grown rapidly as IQA models increasingly serve as critical perceptual reward signals for improving modern algorithms. Specifically, IQA serves as a component in frameworks like Reinforcement Learning (RL) from Human Feedback (RLHF) [15,18,45,61] to align outputs with human preferences. Similarly, in image restoration, IQA scores are used as differentiable

* Corresponding author

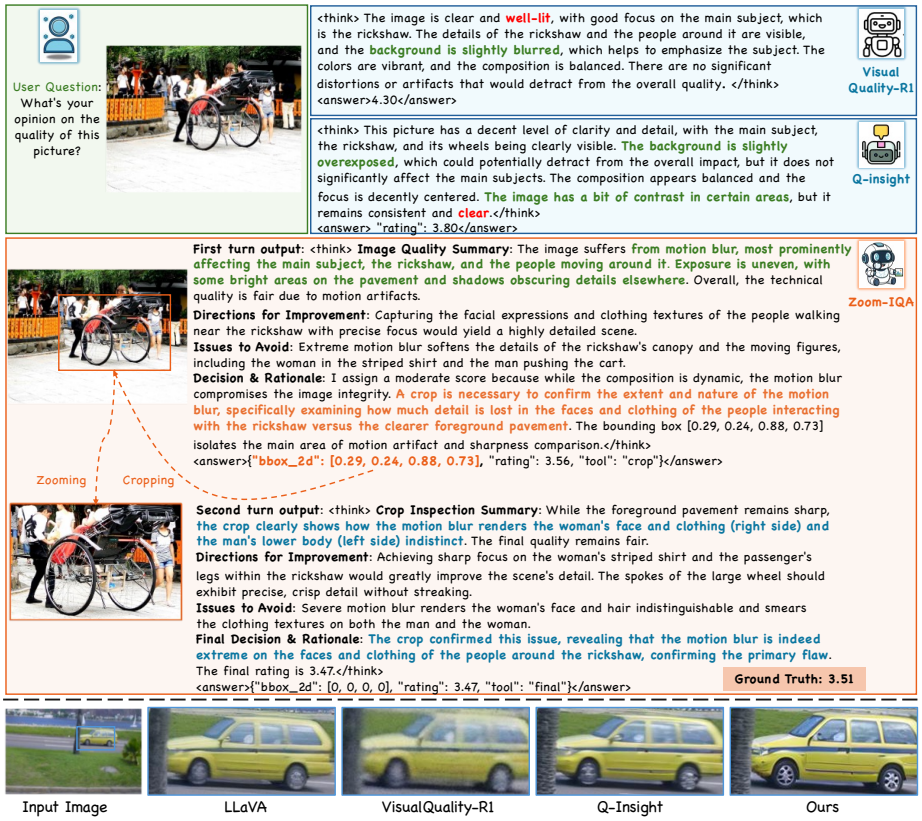

Fig. 1: (Upper) Current IQA methods are **non-interactive**, leading to inferior assessments. They either spot only partial flaws (e.g., **slightly overexposed** or **slightly blurred**) or make factually incorrect claims (**clear** and **well-lit**), resulting in erroneous judgments. Our Zoom-IQA first hypothesizes flaws (green text), then grounds them by cropping (orange text), and finally verifies the degradation (blue text). The predicted score rankings for this image are top 37.1% for Q-Insight, top 35.1% for VisualQuality-R1, top 34.9% for ours, and top 31.9% for the ground truth. **(Lower)** Our model's reasoning outputs also benefit downstream tasks, such as text-guided image restoration with SUPIR [72]. Our prompt enables a far superior restoration compared to those guided by other IQA methods or SUPIR's default VLM, LLaVA-1.5-13b [31].

rewards [53] or in Direct Preference Optimization (DPO) [4,64] to guide models toward perceptually superior results.

The emergence of vision language models (VLMs) like CLIP [43] opened a promising direction for IQA [57], further advanced by large-scale VLMs [2,31,56] that leverage broad knowledge to better align with human perception. Existing VLM-based IQA methods fall into two categories: (1) *Score-based methods* (*e.g.*, Q-Align [63] and DeQA-Score [68]), which emphasize accurate scores but lack the ability to provide textual descriptions; and (2) *Description-based methods*

(*e.g.*, DepictQA series [69, 70]), which offer detailed explanations but rely on SFT data with descriptions generated from ground-truth synthetic distortions. To bridge this gap, recent methods like Q-Insight [29] and VisualQuality-R1 [65] introduce RL to unify quality scoring and textual reasoning using only score labels as rewards.

Despite impressive capabilities, these VLM-based IQA methods remain *non-interactive*. They generate responses in a single pass without any mechanism for iterative visual refinement or correction. As highlighted in complex visual tasks (*e.g.*, object detection [35, 48] and VQA [37, 71]), the lack of intermediate visual grounding may lead to unreliable responses and reasoning, especially under complex scenarios. Similarly, the nature of IQA requires subjects to interpret complex images by "zooming in" on key regions. The importance of this behavior is highlighted by the DiffIQA dataset [8], which provided annotators with a zoom-in feature to inspect details. Such a region-aware interaction plays a helpful role in understanding the quality of an image, but is absent in existing VLM-based IQA models, confining their reasoning to the text domain and limiting effective use of visual information, as shown in Fig. 1.

Enabling a VLM-based IQA method to dynamically "crop and zoom" for iterative region-aware assessment faces two core challenges. (1) *Region-aware Learning.* The model must learn where to focus and how to transform regions (*e.g.*, crop, zoom) based on its own partial textual deliberations, similar to the grounding in VQA [47, 78]. However, grounding IQA is more challenging than VQA. In VQA, grounding is often explicit, *i.e.*, an answer can be tied to discrete, localizable semantic objects, for which extensive annotations are available [47, 78]. In contrast, an IQA score is a holistic judgment aggregated from numerous, complex factors. The core difficulty is the lack of supervision specifying which regions a human prioritized to arrive at their final score. While recent IQA grounding datasets [6, 9] provide static distortion masks, these methods fail to reveal the criticality of those regions to the overall human assessment or capture the dynamic reasoning path that led to the final judgment. (2) *Self-Guided Reasoning Policy.* The model must learn a dynamic policy on when to trigger detailed visual inspection rather than relying on random exploration or preset rules for zooming. This requires an unsupervised iterative cognitive process, *i.e.*, performing a holistic assessment, identifying its own uncertainty about a specific region, and only then deciding to "zoom in" for refinement.

To bridge such gaps, we make **two primary contributions**. **First**, we introduce Grounded-Rationale-IQA (GR-IQA), a fine-grained dataset curated to facilitate the development of interleaved text-image Chain-of-Thought (CoT) reasoning. Directly harnessing advanced VLMs such as Gemini [12] to label IQA data with reasoning and scores can suffer from misalignment between visual inputs and reasoning outputs due to hallucination [27, 33]. Our GR-IQA is designed to avoid such hallucination by providing rationales that are verifiably grounded in visual regions. Specifically, our curation pipeline introduces two key modules: (1) *Visual Reliance Filtering (VRF)*, which enforces grounding by measuring the generative output shift (with vs. without the image); and

(2) *Hint-Augmented Consistency Filtering (HACF)*, which performs sample-level checks to filter hallucination-like descriptions. **Second**, we propose Zoom-IQA (Zoomable Region Reasoning for Reliable Image Quality Assessment), a novel framework designed to enhance the reasoning reliability of IQA with region awareness. Zoom-IQA is trained in two stages: it first learns formatted grounding (how to "zoom") via supervised fine-tuning (SFT) on our GR-IQA dataset, and then learns a dynamic policy (when to "zoom") via reinforcement learning (RL). Such a learned policy allows Zoom-IQA to operate iteratively, moving beyond "single-pass" methods to identify uncertainty and refine its assessment, achieving truly interactive visual reasoning. To stabilize the training process, we further propose the KL-Coverage regularizer, designed to prevent a collapse in reasoning path diversity, which often leads to a severe "mode collapse". A Progressive Re-sampling Strategy is also developed to mitigate bias from imbalanced annotations.

Our Zoom-IQA is evaluated across diverse datasets and IQA tasks, demonstrating superior performance over both conventional IQA metrics and recent SFT-driven large language models. Moreover, Zoom-IQA exhibits impressive zero-shot generalization, such as effectively guiding image restoration models at test time, which highlights the robustness and real-world applicability of its region reasoning.

2 Related Work

Image Quality Assessment. Previous IQA works are broadly divided into full-reference (FR) and no-reference (NR) approaches, based on the availability of a pristine reference image. As our work does not require a reference, we focus on the more challenging NR-IQA task. Conventional NR-IQA methods [36,38–41] relied on hand-crafted, degradation-aware features to predict the final quality score. Subsequent deep learning-based models [3, 11, 24, 25, 34, 42, 51, 52, 79] replaced this pipeline, directly predicting quality scores using end-to-end trainable neural networks. *Nonetheless, these models often suffer from significant performance degradation on out-of-distribution (OOD) data, limiting their practical usage and generalizability.*

Vision Language Models in Image Quality Assessment. Vision language models (VLMs) [2, 31, 43, 56] have been extensively studied in image quality assessment (IQA), leveraging their powerful cross-modal understanding and strong generalization capabilities. These works typically focus on one of two objectives: providing numerical quality scores [32, 57, 63, 68, 80] or generating visual quality descriptions [9, 62, 69, 70]. Specifically, CLIP-IQA proposes to harness CLIP [43] for image quality assessment from multiple aspects. DOG-IQA [32] attempts to mimic the human evaluation process (*e.g.*, zooming in to evaluate specific areas). However, these models lack the necessary reasoning capabilities and cannot dynamically decide which regions to inspect. A common way to alleviate such limitations is via pre-processing, *i.e.*, using a pre-trained segmentation model to crop sub-images and then computing the final score as a weighted average

of these crops. Recent works like Q-Insight [29], VisualQuality-R1 [65], and Q-Ponder [5] propose to employ reinforcement learning (RL) to leverage the reasoning capabilities of VLMs, enabling image quality rating as well as textual justifications. Building upon this paradigm, subsequent advancements have expanded RL-based reasoning to video quality assessment [76] and employed contrastive learning to directly align images with the generalizable text representations learned via RL [77]. However, the reasoning chains generated by these methods remain purely textual. They lack dynamic interaction with the image (*e.g.*, cropping or zooming) to evaluate specific regions—a process crucial to human assessment. *Consequently, visual evidence is insufficiently explored, and the models fail to ground their textual reasoning in verifiable image regions, limiting both the reliability and interpretability of their outputs.*

Vision Language Models with Multimodal Reasoning. Recent advances in enhancing the reasoning capabilities of VLMs have significantly improved their performance on challenging tasks, such as mathematical problem solving [20, 37, 73], general VQA [37, 71], and object detection [35, 48, 71]. However, these models typically generate reasoning chains composed solely of natural language. This text-only reasoning can be opaque and often lacks sufficient grounding in the visual input's fine-grained details. To address this, recent works in general VQA [21, 50, 75, 78] propose to integrate evidence regions into the reasoning process. By equipping VLMs with capabilities like iterative zoom-in and region-of-interest selection, these methods demonstrated boosted performance. Improved interpretability is further gained via visual-linguistic interactive reasoning. However, these VQA methods generally rely on heavily annotated evidence regions (*e.g.*, bounding boxes) for training. *Such fine-grained regional labels and corresponding reasoning trajectories are expensive and lacking in the IQA domain.*

3 Methodology

Our method, Zoom-IQA, is trained via a two-stage pipeline (illustrated in Fig. 2): **1)** Supervised Fine-Tuning for Grounded Quality Rationale Learning: We first leverage our GR-IQA dataset to teach the VLM the foundational "how-to" skills: grounding textual rationales in visual regions and executing the "zoom" action (Sec. 3.1). **2)** Reinforcement Learning for Self-Guided Exploration: We then use the RL-based training process (Group Relative Policy Optimization) to learn a dynamic policy that decides when to deploy these skills for iterative refinement (Sec. 3.2).

3.1 Grounded Quality Rationale Learning

Recalling the challenges from the Introduction, grounding IQA is uniquely challenging. While VQA supervision can link answers to semantic regions, IQA supervision is often limited to static distortion masks. This static approach fails to capture the dynamic reasoning path of how or why specific regions influence the final holistic assessment.

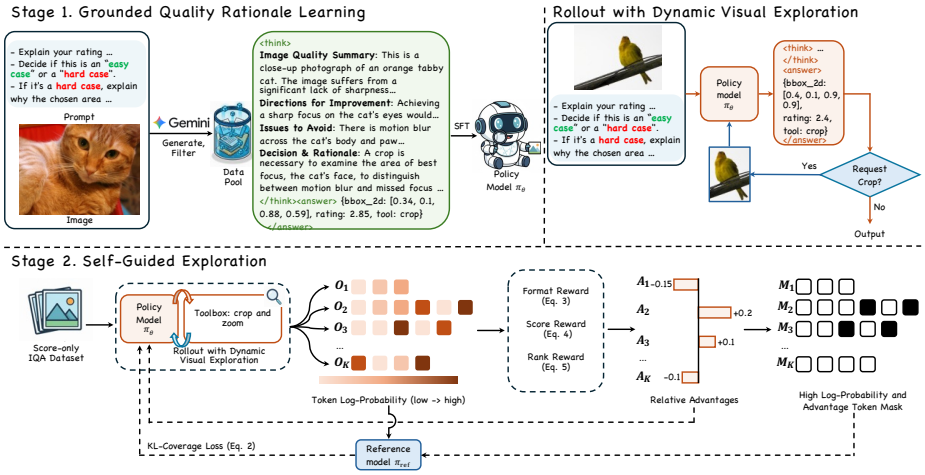

Fig. 2: An overview of our two-stage framework. Stage **(1)**, **Grounded Quality Rationale Learning** (Sec. 3.1), first uses SFT to teach the model how to correctly execute the crop action. Stage **(2)**, **Self-Guided Exploration** (Sec. 3.2), then uses RL to let the model learn what to crop, allowing it to discover regions that lead to a deeper understanding of image quality.

Grounded-Rationale-IQA (GR-IQA) Dataset. To bridge this gap, we curate the Grounded-Rationale-IQA (GR-IQA) dataset with approximately 7,000 reasoning trajectories. This curation is performed through a novel pipeline (Fig. 3) that features our two key modules: **1)** Visual Reliance Filtering (VRF); **2)** Hint-Augmented Consistency Filtering (HACF).

Data Generation. We prompt the closed-source VLM, Gemini-2.5-pro [12], on KonIQ dataset images using a structured prompt. This compels the VLM to generate a two-part response: a textual rationale (within `<think>`) and a JSON action (within `<answer>`). The textual rationale is strictly constrained to a four-part format: (1) a holistic *Image Quality Summary*; (2) *Directions for Improvement*; (3) *Issues to Avoid*; and (4) a *Decision & Rationale* where the model must decide if the image is an "easy case" (requiring a "final" tool) or a "hard case" (requiring a "crop" tool) and justify its choice. The `<answer>` block then contains the chosen "tool" ("final" or "crop"), the rating, and a conditional "bbox". This structured process forces the VLM to link its score to **regional evidence** and to **perform self-assessment on its own uncertainty** (*i.e.*, "zoom" or not). This raw output forms the data for our GR-IQA dataset, which is then passed to our filtering modules (VRF and HACF).

Visual Reliance Filtering (VRF). Despite strong language priors, VLMs can over-rely on textual co-occurrence and generate responses that are weakly grounded in the image [27, 33]. Detecting such hallucinations is non-trivial in our setting: (i) for closed-source VLMs, token-level log-probabilities are often unavailable, making probability-based contrastive analysis inapplicable; and (ii) online consistency checks that inject image perturbations are ill-suited for IQA,

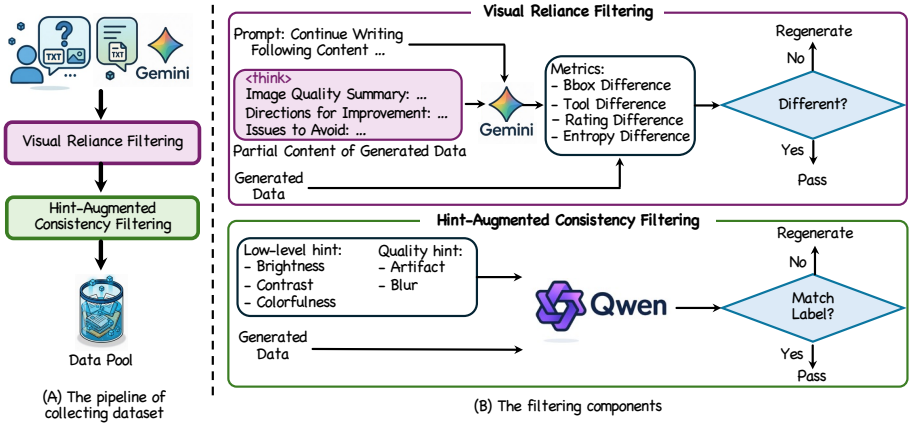

Fig. 3: The GR-IQA dataset curation pipeline. It uses **(1)** Visual Reliance Filtering (VRF) to ensure visual grounding by measuring the output shift with vs. without the image, and **(2)** Hint-Augmented Consistency Filtering (HACF) to perform sample-level, hint-based checks for unfaithful text.

since added distortions can directly alter the target degradation to be assessed. We therefore propose *Visual Reliance Filtering* (VRF), which removes samples with low *visual* reliance without modifying the input image. Specifically, we contrast the VLM outputs under two inference conditions: **(1)** conditioned on both the image I and a partial rationale R_A, and **(2)** conditioned only on R_A (without I). We measure their discrepancy using task-relevant signals (e.g., predicted quality score difference, localization overlap, and output uncertainty); samples with overly small discrepancies are discarded, indicating that the response can be reproduced from text alone.

Hint-Augmented Consistency Filtering (HACF). While VRF encourages visual dependence for the final `<answer>`, the intermediate rationale R (the `<think>` block) may still contain statements unsupported by the visual evidence. We thus employ a strong LLM rater, Qwen-2.5-32b [66], denoted as $\text{LLM}_{\text{Rater}}$, to perform a sample-level assessment of the rationale's image-consistency. To aid this judgment, $\text{LLM}_{\text{Rater}}$ is provided with (R, I) together with *dataset-provided* low-level, non-semantic hints H (e.g., brightness, sharpness, and color statistics). The rater outputs a binary accept/reject decision: $D_{\text{HACF}} = \text{LLM}_{\text{Rater}}(R, I, H)$. We only retain samples with $D_{\text{HACF}} =$ "Pass", ensuring that the training data exhibits image-consistent reasoning. *Additional examples of filtered-out samples are provided in the supplementary material.*

Grounded Rationale Fine-Tuning. Following the curation of our high-fidelity GR-IQA dataset, we proceed to the SFT stage. We fine-tune the VLM, parameterized by θ, to auto-regressively generate the complete ground-truth response $S_{\text{structured}} = (R, A)$, given the image I and the prompt T_{CoT}. This is achieved by minimizing the standard cross-entropy (CE) loss $\mathcal{L}_{\text{SFT}}$: $\mathcal{L}_{\text{SFT}}(\theta) =$

$-\sum_{t=1}^{|S|} \log p_\theta(S_t \mid S_{<t}, I, T_{\text{CoT}})$ where S_t is the t-th token in the ground-truth sequence $S_{\text{structured}}$, and $|S|$ is the total length of the sequence.

3.2 Self-Guided Exploration

KL-Coverage Regularizer. Recent studies on applying Reinforcement Learning (RL) to large language models (LLMs) [10,13] highlight a critical challenge: policy entropy often drops sharply at the onset of training, declining monotonically to near zero. This "entropy collapse" severely limits the model's ability to explore, leading to performance plateaus. In the context of IQA, this issue is particularly detrimental. It leads to a collapse in the diversity of both reasoning paths and predicted rating scores. For instance, existing RL-based IQA methods, such as VisualQuality-R1 [65], suffer from "score collapse." On the KonIQ [19] test set, this method's output unique score ratio is merely 2.04%, in stark contrast to the 71.34% of the ground-truth Mean Opinion Scores (MOS) distribution (when rounded to two decimal places).

To address this problem, we propose the KL-Coverage regularizer. This approach is inspired by the use of KL penalties to constrain policy updates [46] and the recent finding that high covariance between action log-probabilities and logit changes leads to policy entropy collapse [13]. Our regularizer is thus designed to specifically suppress numerical tokens that exhibit this high covariance.

Given a batch of N rollout tokens, let $\pi_\theta(y_i|y_{<i})$ denote the policy's probability for token y_i given its prefix $y_{<i}$, and let $A(y_i)$ be its associated advantage. We first compute the batch-level mean log-probability $\overline{\log \pi}$ and mean advantage $\overline{A}$: $\overline{\log \pi} = \frac{1}{N}\sum_{j=1}^{N} \log \pi_\theta(y_j|y_{<j}), \quad \overline{A} = \frac{1}{N}\sum_{j=1}^{N} A(y_j)$. We then define a token-wise covariance score $Cov(y_i)$ as the centered cross-product:

$$Cov(y_i) = \left(\log \pi_\theta(y_i|y_{<i}) - \overline{\log \pi}\right)\left(A(y_i) - \overline{A}\right). \tag{1}$$

Crucially, our regularizer mainly targets the numerical tokens responsible for the final score. We first define a candidate set $\mathcal{N}_{ans}$ comprising all numerical tokens within the `<answer>...</answer>` tags. We then rank the tokens in this candidate set $\mathcal{N}_{ans}$ by their $Cov(y_i)$ scores. We define a binary mask M_i for tokens $y_i \in \mathcal{N}_{ans}$, where $M_i = 1$ if the token is in the top-p proportion (e.g., $p = 0.02$), and $M_i = 0$ otherwise. Here, p is a hyperparameter defining the fraction of these candidate tokens to be regularized.

Finally, we impose the KL penalty only on these selected tokens (where $M_i = 1$). The KL-Coverage loss, $\mathcal{L}_{KLC}$, is computed as the mean KL divergence between the old policy $\pi_{\theta_{\text{old}}}$ and the current policy π_θ, averaged only over these masked-in tokens:

$$\mathcal{L}_{KLC} = \frac{\sum_{y_i \in \mathcal{N}_{ans}} M_i \cdot D_{KL}\left(\pi_{\theta_{\text{old}}}(y_i|y_{<i}) \| \pi_\theta(y_i|y_{<i})\right)}{\sum_{y_i \in \mathcal{N}_{ans}} M_i}. \tag{2}$$

Progressive Re-sampling Strategy. Our training data suffers from a long-tailed score distribution, leading to poor performance on scarce score intervals

(*e.g.*, very high or low quality). To mitigate this data bias, we adopt a multi-stage re-sampling strategy. The model is first trained on the original data distribution, and in subsequent stages, we progressively increase the sampling frequency of these under-represented score intervals. This allows the model to first learn the general distribution and then fine-tune on rarer data, improving its robustness across the entire score range.

Format Reward. This reward verifies that the model's output strictly adheres to our required structured reasoning format. Specifically, the reasoning process, enclosed in `<think>...</think>` tags, must explicitly articulate key components such as `"Directions for Improvement"` and `"Issues to Avoid"`. Furthermore, the final decision must be in a structured `<answer>...</answer>` tag, containing elements like `bbox_2d` and `rating`. If any of these formats are incorrect, the format score is 0. Only when all formats are correct can the model achieve the format score of 1.0 as defined:

$$R_{\text{format}}(O) = \begin{cases} 1.0 & \text{if } O \text{ satisfies all format requirements} \\ 0 & \text{otherwise} \end{cases}. \tag{3}$$

Score Reward. This reward encourages the model to predict a quality rating r_{pred} that is close to the ground-truth score r_{gt}. We define this as a continuous Gaussian reward based on their difference:

$$R_{\text{score}} = \exp\left(-\frac{(r_{\text{pred}} - r_{\text{gt}})^2}{2\sigma^2}\right), \tag{4}$$

where σ is a hyperparameter controlling the sensitivity of the reward.

Rank Reward. To ensure the model learns relative quality ordering, we define a rank reward $R_{\text{rank}}(x_i)$ based on pairwise comparisons within a batch, inspired by the Thurstone model [55]. It is computed as:

$$R_{\text{rank}}(x_i) = \frac{1}{B-1}\sum_{j\neq i}\left(\sqrt{\hat{p}_{ij}p^*_{ij}} + \sqrt{(1-\hat{p}_{ij})(1-p^*_{ij})}\right), \tag{5}$$

where $p^*_{ij} = p(x_i, x_j)$ is the ground-truth preference derived from MOS, indicating whether $\text{MOS}(x_i) > \text{MOS}(x_j)$, and $\hat{p}_{ij}$ is the model's predicted preference probability, calculated using the Thurstone model: $\hat{p}_{ij} = \Phi\left(\frac{\mu_i - \mu_j}{\sqrt{v_i + v_j}}\right)$. Here, μ and v represent the estimated mean and variance of the model's rating distribution for an input, and Φ is the standard normal CDF.

Total Reward. The total reward R_{total} for a trajectory comprises several components designed to guide the behavior of Zoom-IQA, formulated as:

$$R_{\text{total}} = R_{\text{format}} + \alpha_{score}R_{\text{score}} + \alpha_{rank}R_{\text{rank}}, \tag{6}$$

where α_{score} and α_{rank} are coefficients that balance the importance of score prediction and rank consistency. In our experiments, we set $\alpha_{score} = 1$ and $\alpha_{rank} = 2$.

Fig. 4: Qualitative comparison of Zoom-IQA with competing methods (Q-Insight [29], VisualQuality-R1 [65]). We highlight **correct** descriptions, **incorrect** descriptions, and the **uncertainty-aware** reasoning unique to our model. Specifically, the predicted score rankings for this image are top 42.5% for Q-Insight, top 44.7% for VisualQuality-R1, top 50.4% for ours, and top 54.3% for the ground truth.

4 Experiments

4.1 Experimental Settings

Implementation Details. We initialize Qwen2.5-VL-7b [2] as our base model during the first cold start stage, in which training is performed with a batch size of 2, 4 gradient accumulation steps, a learning rate of 2.5×10^{-6}, and a warm-up ratio of 0.3. For GRPO, we train the finetuned model after the cold start stage with a batch size of 1, 2 gradient accumulation steps, a learning rate of 1×10^{-6}, and a KL penalty coefficient of $\beta = 0.04$. The number of generated responses N is set to 8.

Datasets and Metrics. For the first cold start stage, we applied SFT with our collected high-quality CoT datasets using cross-entropy loss. For the score regression task, we conduct training and evaluation on seven IQA datasets grouped

Table 1: PLCC / SRCC comparison on the score regression tasks between our method and other competitive IQA methods. All methods except handcrafted ones are trained on the **KonIQ dataset**.

Metric	Category	Methods	KonIQ	SPAQ	KADID	PIPAL	LiveW	AGIQA	CSIQ
PLCC	Handcrafted	NIQE [39]	0.533	0.679	0.468	0.195	0.493	0.560	0.718
		BRISQUE [38]	0.225	0.490	0.429	0.267	0.361	0.541	0.740
	Non-VLM	NIMA [54]	0.896	0.838	0.532	0.390	0.814	0.715	0.695
		HyperIQA [51]	0.917	0.791	0.506	0.410	0.772	0.702	0.752
		DBCNN [74]	0.884	0.812	0.497	0.384	0.773	0.730	0.586
		MUSIQ [25]	0.924	0.868	0.575	0.431	0.789	0.722	0.771
		ManIQA [67]	0.849	0.768	0.499	0.457	0.849	0.723	0.623
	VLM (w/o & w/ reasoning)	CLIP-IQA+ [57]	0.909	0.866	0.653	0.427	0.832	0.736	0.772
		C2Score [80]	0.923	0.867	0.500	0.354	0.786	0.777	0.735
		Q-Align [63]	0.941	0.886	0.674	0.403	0.853	0.772	0.671
		DeQA-Score [68]	0.953	0.895	0.694	0.472	0.892	0.809	0.787
		Q-Insight [29]	0.918	0.903	0.702	0.458	0.870	0.816	0.685
		VisualQuality-R1 [65]	0.910	0.889	0.703	0.451	0.856	0.817	0.768
		Zoom-IQA (Ours)	0.938	0.902	0.701	0.468	0.887	0.816	0.797
SRCC	Handcrafted	NIQE [39]	0.530	0.664	0.405	0.161	0.449	0.533	0.628
		BRISQUE [38]	0.226	0.406	0.356	0.232	0.313	0.497	0.556
	Non-VLM	NIMA [54]	0.859	0.856	0.535	0.399	0.771	0.654	0.649
		HyperIQA [51]	0.906	0.788	0.468	0.403	0.749	0.640	0.717
		DBCNN [74]	0.875	0.806	0.484	0.381	0.755	0.641	0.572
		MUSIQ [25]	0.929	0.863	0.556	0.431	0.830	0.630	0.710
		ManIQA [67]	0.834	0.758	0.465	0.452	0.832	0.636	0.627
	VLM (w/o & w/ reasoning)	CLIP-IQA+ [57]	0.895	0.864	0.654	0.419	0.805	0.685	0.719
		C2Score [80]	0.910	0.860	0.453	0.342	0.772	0.671	0.705
		Q-Align [63]	0.940	0.887	0.684	0.419	0.860	0.735	0.737
		DeQA-Score [68]	0.941	0.896	0.687	0.478	0.879	0.729	0.744
		Q-Insight [29]	0.895	0.903	0.702	0.435	0.839	0.766	0.640
		VisualQuality-R1 [65]	0.896	0.892	0.712	0.441	0.827	0.760	0.707
		Zoom-IQA (Ours)	0.922	0.900	0.700	0.465	0.870	0.765	0.754

into three categories: (1) In-the-wild datasets, including KonIQ [19], SPAQ [16], and LIVE-Wild [17]; (2) Synthetic distortion datasets, including KADID [30], PIPAL [23], and CSIQ [26]; (3) AI-generated image datasets, including AGIQA [28]. We adopt the Pearson linear correlation coefficient (PLCC) and Spearman rank-order correlation coefficient (SRCC) as metrics to evaluate performance on the score regression task, following previous works [29, 68].

4.2 Comparison and Evaluation

Image Quality Score Regression. We compare our method with SOTA IQA methods in three different categories: **(I)** handcrafted, including NIQE [39] and BRISQUE [38]; **(II)** deep learning-based, NIMA [54], HyperIQA [51], DBCNN [74],

Table 2: Quantitative results for image quality description. We evaluate four metrics: Accuracy (Acc.), Reasonableness (Reason.), Completeness (Compl.), and Confidence (Conf.) using closed-source VLM evaluators: Gemini-2.5-Flash and GPT-5-mini.

Dataset	Method	Gemini-2.5-flash				GPT-5-mini			
		Acc. ↑	Reason. ↑	Compl. ↑	Conf. ↑	Acc. ↑	Reason. ↑	Compl. ↑	Conf. ↑
KonIQ	DepictQA [70]	5.40	5.49	5.51	7.96	4.54	5.09	4.41	7.80
	VisualQuality-R1 [65]	7.29	7.60	7.29	7.57	6.10	6.05	6.02	6.79
	Q-Insight [29]	7.17	7.44	7.08	6.93	5.32	5.74	5.29	6.15
	Zoom-IQA (Ours)	8.72	8.80	8.30	8.60	6.93	6.93	6.61	7.98
SPAQ	DepictQA [70]	6.04	6.39	6.14	7.86	4.80	5.08	4.46	6.87
	VisualQuality-R1 [65]	8.32	8.35	7.70	7.55	6.61	6.67	5.51	6.82
	Q-Insight [29]	7.84	8.02	7.51	6.98	6.22	6.35	5.54	5.90
	Zoom-IQA (Ours)	8.63	8.69	8.47	8.63	6.97	7.27	6.79	7.99

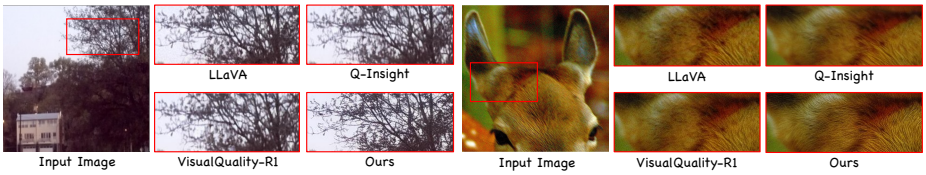

Fig. 5: Qualitative evaluation of reasoning quality on the image restoration task.

MUSIQ [25], and ManIQA [67]; **(III)** VLM-based models, CLIP-IQA+ [57], C2Score [80], Q-Align [63], DeQA-Score [68], VisualQuality-R1 [65], and Q-Insight [29]. Since VisualQuality-R1 [65] did not report a KonIQ-only trained model, we retrained it with its official training code. As shown in Table 1, our approach achieves comparable performance to existing baselines across various synthetic and real-world benchmarks. When comparing with state-of-the-art IQA methods [29, 65] w/ reasoning capability, our Zoom-IQA presents consistently superior performance across almost all the benchmarks. Furthermore, a qualitative comparison demonstrates the superiority of our region-aware reasoning over competing methods (Fig. 4).

Image Quality Reasoning. To validate the effectiveness and accuracy of our reasoning chains, we follow common practices [7, 22] to employ a VLM-as-judge evaluation methodology on the KonIQ and SPAQ datasets. We utilize two powerful, closed-source VLMs (Gemini-2.5-Flash [12] and GPT-5-mini [49]) as evaluators, which are tasked to score the generated descriptions on a 1-to-9 scale across four key criteria: Accuracy, Reasonableness, Completeness, and Confidence. To ground the assessments and ensure objectivity, the VLMs are prompted with the image, the generated reasoning chain, and corresponding low-level image indicators (e.g., brightness, sharpness) for cross-referencing. The detailed definitions of each metric and the full prompt structure are provided in the Appendix. As shown in Table 2, our method (Zoom-IQA) consistently and significantly outperforms all baselines [29, 65, 70] across both datasets and under the scrutiny of both VLM evaluators, indicating the superiority of our reasoning reliability.

Reasoning-guided Restoration. High-quality visual reasoning provides reliable guidance for downstream tasks such as generative image restoration [58–60,

Table 3: Quantitative results of image restoration using different reasoning guidance.

Guidance	DIV2K (SUPIR)			RealLQ250 (DreamClear)			
	PSNR↑	MANIQA↑	CLIPIQA↑	NIQE↓	MANIQA↑	MUSIQ↑	CLIPIQA↑
LLaVA-13b	21.57	0.4974	0.6184	3.922	**0.4369**	66.55	0.6827
VisualQuality-R1	22.38	0.4898	0.5927	3.968	0.4273	65.55	0.6846
Q-Insight	**22.44**	0.4861	0.5754	3.965	0.4297	65.96	0.6820
Zoom-IQA	21.31	**0.5455**	**0.6773**	**3.914**	0.4365	**66.97**	**0.6946**

72]. To validate the efficacy and transferability of Zoom-IQA's reasoning capability, we leverage its outputs to guide two state-of-the-art text-guided restoration frameworks: SUPIR [72] and DreamClear [1]. Both frameworks follow a cascaded paradigm where initial results from a lightweight restoration network are conditioned by VLM-generated guidance. For the SUPIR framework, we replace its default prompt generator (LLaVA-1.5-13b [31]) with the textual reasoning outputs derived from Q-Insight [29], VisualQuality-R1 [65], and Zoom-IQA. As illustrated in Fig. 5, Zoom-IQA generates highly specific and context-aware guidance, enabling SUPIR to accurately restore fine-grained, detailed textures that are otherwise overlooked or poorly reconstructed by competing methods. Concurrently, we extend our evaluation to DreamClear [1]. This choice is motivated by its T5 text encoder [44], which supports a longer context window than the CLIP encoder [43] used in SUPIR, thereby accommodating the detailed reasoning prompts generated by Zoom-IQA. Under this framework, we compare the default captions from LLaVA-1.6-13b [31] against the reasoning content from VisualQuality-R1, Q-Insight, and Zoom-IQA's reasoning trajectory.

To quantitatively evaluate the quality of the reasoning guidance across these different configurations, we conduct comprehensive experiments using SUPIR on DIV2K (synthetic) and DreamClear on RealLQ250 (real-world), as summarized in Tab. 3. Zoom-IQA attains the best CLIPIQA on both setups, the best MANIQA under SUPIR, and the best NIQE and MUSIQ under DreamClear. Moreover, supplementary DreamClear comparisons reveal superior recovery of high-frequency details.

4.3 Ablation Study

We conduct a comprehensive ablation study to evaluate the contribution of each proposed component. The results, measured by PLCC and SRCC across seven commonly used benchmarks, are presented in Table 4. Our analysis begins with the SFT model (Row 1), which serves as the baseline.

Impact of Reward Signals. The comparison between Row 2 (Rank Reward only) and Row 3 (Score Reward only) indicates the effectiveness of Rank Reward for synthetic benchmarks (e.g., KADID: 0.709 vs. 0.677; CSIQ: 0.772 vs. 0.715). Besides, the Score Reward benefits real-world benchmarks (e.g., KonIQ: 0.928 vs. 0.906; SPAQ: 0.896 vs. 0.892). Such a comparison shows that both rewards contribute to the assessment of image quality.

Table 4: Ablation studies on each component with PLCC / SRCC metrics. Models are trained on KonIQ.

Metric	SFT	Score Reward	Rank Reward	KL-Coverage Regularizer Loss	Prog. Training	KonIQ	SPAQ	KADID	PIPAL	LIVE-Wild	AGIQA	CSIQ
PLCC	✓					0.836	0.849	0.632	0.431	0.806	0.766	0.688
	✓		✓	✓		0.906	0.892	0.709	0.450	0.854	0.784	0.772
	✓	✓		✓		0.928	0.896	0.677	0.379	0.875	0.807	0.715
	✓	✓	✓			0.908	0.888	0.683	0.455	0.874	0.806	0.759
	✓	✓	✓	✓		0.932	0.898	0.665	0.458	0.881	0.810	0.791
	✓	✓	✓	✓	✓	0.938	0.902	0.701	0.468	0.887	0.816	0.797
SRCC	✓					0.806	0.839	0.621	0.426	0.762	0.697	0.645
	✓		✓	✓		0.887	0.884	0.699	0.447	0.827	0.739	0.710
	✓	✓		✓		0.915	0.892	0.664	0.392	0.865	0.745	0.696
	✓	✓	✓			0.890	0.882	0.669	0.446	0.847	0.738	0.718
	✓	✓	✓	✓		0.918	0.895	0.652	0.455	0.865	0.758	0.749
	✓	✓	✓	✓	✓	0.922	0.900	0.700	0.465	0.870	0.765	0.754

Table 5: Ablation studies on cropping strategies and training paradigms.

Setting	KonIQ	KADID	LIVE-Wild	CSIQ
Central Crop	0.898/0.876	0.672/0.661	0.871/0.842	0.743/0.677
Single-Stage SFT	0.834/0.812	0.614/0.613	0.801/0.775	0.672/0.637
Ours	0.938/0.922	0.701/0.700	0.887/0.870	0.797/0.754

Effect of KL-Coverage Regularizer Loss. We evaluate the impact of our KL-Coverage loss by comparing Row 4 (without KL) to Row 5 (with KL). Introducing such a regularizer brings consistent performance gains across the majority of datasets, such as on KonIQ (0.908 to 0.932), SPAQ (0.888 to 0.898), and CSIQ (0.759 to 0.791), indicating its effectiveness in preventing mode collapse and encouraging diverse reasoning.

Effect of Progressive Training. Comparing Row 5 (full model without) with Row 6 (full model with), progressive training provides a consistent performance lift across all datasets. Such an improvement stems from the strategy's ability to effectively handle imbalanced quality data, *i.e.*, conducting a superior evaluation on images at the extremes of the quality spectrum.

4.4 Discussion

Necessity of the Learned Zooming Policy. Our learned policy consistently outperforms a simple heuristic that always zooms into the image center ("Center Crop"), while keeping the rest of the pipeline unchanged (Tab. 5; e.g., +0.040 PLCC on KonIQ). This clear margin suggests that quality-relevant distortions are not necessarily centered, and that RL-based active zooming effectively helps locate degradation-critical regions.

Saliency Grounding of Zoomed Regions. To evaluate whether the generated crops are semantically meaningful, we introduce a Saliency Grounding Benchmark using QAGNet [14]. We assess performance using two metrics: Saliency

Table 6: Saliency Grounding Benchmark on KonIQ and SPAQ.

Method	KonIQ			SPAQ		
	Area	$F_{0.5}\uparrow$	SDL↑	Area	$F_{0.5}\uparrow$	SDL↑
Qwen2.5-VL-7b	0.65	0.19	1.04	0.66	0.06	1.01
Zoom-IQA	**0.36**	**0.70**	**1.90**	**0.37**	**0.78**	**2.13**

Density Lift (SDL)—measuring the relative saliency concentration within the crop compared to the whole image—and Tight-Coverage ($F_{0.5}$), which jointly rewards high saliency coverage and compact crop sizes. As shown in Tab. 6, the baseline Qwen2.5-VL-7b outputs excessively large crops (Area $\sim 65\%$) that far overstep the actual salient regions due to its limited grounding capabilities. In contrast, Zoom-IQA precisely localizes necessary and meaningful regions, achieving much tighter crops (Area $\sim 36\%$) alongside substantially higher SDL and $F_{0.5}$ (e.g., 0.78 vs. 0.06 on SPAQ). This demonstrates that our zoom-in mechanism is highly semantics-aware rather than blindly cropping.

Superiority of RL over Pure Supervised Fine-Tuning. Fine-tuning the backbone VLM on GR-IQA using a "Single-Stage SFT" baseline (without the multi-stage zoom pipeline or RL) consistently underperforms our full method (Tab. 5; e.g., +0.110 SRCC on KonIQ). These results indicate that while GR-IQA provides useful supervision for learning basic zooming/grounding behaviors, supervised imitation alone is insufficient to induce an effective region-selection strategy. In contrast, the RL objective enables self-guided exploration and improves the learned policy's ability to identify quality-critical regions, leading to better generalization across datasets.

5 Conclusion

In this work, we proposed Zoom-IQA, a novel IQA framework that uses an iterative process of reasoning and zooming to focus on quality-relevant regions and generate accurate chain-of-thought reasoning. To encourage reliable reasoning, we first build the fine-grained Grounded-Rationale-IQA (GR-IQA) dataset. We further present key training strategies, including a two-stage scheme, reward designs, a KL-Coverage regularizer, and progressive resampling. We verify the effectiveness of our designs via extensive experiments from multiple aspects, including score prediction, reasoning examination, and the downstream application, together with a thorough ablation study. We believe our work could motivate future development in various domains, such as designing IQA data pipelines with automated data labeling, enhancing the reasoning reliability of IQA, and building more robust, interactive perceptual models.

6 Acknowledgment

This research is supported by cash and in-kind funding from NTU S-Lab and industry partner(s). It is also supported by Singapore MOE AcRF Tier 2 (MOE-T2EP20224-0003).

References

1. Ai, Y., Zhou, X., Huang, H., Han, X., Chen, Z., You, Q., Yang, H.: DreamClear: High-capacity real-world image restoration with privacy-safe dataset curation. In: NeurIPS (2024)
2. Bai, S., Chen, K., Liu, X., Wang, J., Ge, W., Song, S., Dang, K., Wang, P., Wang, S., et al.: Qwen2.5-VL technical report. arXiv preprint arXiv:2502.13923 (2025)
3. Bosse, S., Maniry, D., Müller, K.R., Wiegand, T., Samek, W.: Deep neural networks for no-reference and full-reference image quality assessment. IEEE TIP (2017)
4. Cai, M., Li, S., Li, W., Huang, X., Chen, H., Hu, J., Wang, Y.: DSPO: Direct semantic preference optimization for real-world image super-resolution. arXiv preprint arXiv:2504.15176 (2025)
5. Cai, Z., Zhang, J., Yuan, X., Jiang, P.T., Chen, W., Tang, B., Yao, L., Wang, Q., Chen, J., Li, B.: Q-Ponder: A unified training pipeline for reasoning-based visual quality assessment. arXiv preprint arXiv:2506.05384 (2025)
6. Chen, C., Yang, S., Wu, H., Liao, L., Zhang, Z., Wang, A., Sun, W., Yan, Q., Lin, W.: Q-Ground: Image quality grounding with large multi-modality models. In: ACM MM (2024)
7. Chen, D., Chen, R., Zhang, S., Wang, Y., Liu, Y., Zhou, H., Zhang, Q., Wan, Y., Zhou, P., Sun, L.: MLLM-as-a-judge: Assessing multimodal LLM-as-a-judge with vision-language benchmark. In: ICML (2024)
8. Chen, D., Wu, T., Ma, K., Zhang, L.: Toward generalized image quality assessment: Relaxing the perfect reference quality assumption. In: CVPR (2025)
9. Chen, Z., Zhang, X., Li, W., Pei, R., Song, F., Min, X., Liu, X., Yuan, X., Guo, Y., Zhang, Y.: Grounding-IQA: Multimodal language grounding model for image quality assessment. arXiv preprint arXiv:2411.17237 (2024)
10. Cheng, D., Huang, S., Zhu, X., Dai, B., Zhao, W.X., Zhang, Z., Wei, F.: Reasoning with exploration: An entropy perspective. arXiv preprint arXiv:2506.14758 (2025)
11. Cheon, M., Yoon, S.J., Kang, B., Lee, J.: Perceptual image quality assessment with transformers. In: CVPR (2021)
12. Comanici, G., Bieber, E., Schaekermann, M., Pasupat, I., Sachdeva, N., Dhillon, I., Blistein, M., Ram, O., Zhang, D., Rosen, E., et al.: Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities. arXiv preprint arXiv:2507.06261 (2025)
13. Cui, G., Zhang, Y., Chen, J., Yuan, L., Wang, Z., Zuo, Y., Li, H., Fan, Y., Chen, H., Chen, W., et al.: The entropy mechanism of reinforcement learning for reasoning language models. arXiv preprint arXiv:2505.22617 (2025)
14. Deng, B., Song, S., French, A.P., Schluppeck, D., Pound, M.P.: Advancing saliency ranking with human fixations: Dataset models and benchmarks. In: CVPR (2024)
15. Dhariwal, P., Nichol, A.: Diffusion models beat GANs on image synthesis. In: NeurIPS (2021)
16. Fang, Y., Zhu, H., Zeng, Y., Ma, K., Wang, Z.: Perceptual quality assessment of smartphone photography. In: CVPR (2020)
17. Ghadiyaram, D., Bovik, A.C.: Live in the wild image quality challenge database. Online: http://live.ece.utexas.edu/research/ChallengeDB/index.html [Mar, 2017] (2015)
18. He, X., Jiang, D., Zhang, G., Ku, M., Soni, A., Siu, S., Chen, H., Chandra, A., Jiang, Z., Arulraj, A., et al.: VideoScore: Building automatic metrics to simulate fine-grained human feedback for video generation. EMNLP (2024)

19. Hosu, V., Lin, H., Sziranyi, T., Saupe, D.: KonIQ-10k: An ecologically valid database for deep learning of blind image quality assessment. IEEE TIP (2020)
20. Huang, W., Jia, B., Zhai, Z., Cao, S., Ye, Z., Zhao, F., Xu, Z., Hu, Y., Lin, S.: Vision-R1: Incentivizing reasoning capability in multimodal large language models. arXiv preprint arXiv:2503.06749 (2025)
21. Jiang, C., Heng, Y., Ye, W., Yang, H., Xu, H., Yan, M., Zhang, J., Huang, F., Zhang, S.: VLM-R^3: Region recognition, reasoning, and refinement for enhanced multimodal chain-of-thought. arXiv preprint arXiv:2505.16192 (2025)
22. Jiang, D., Zhang, R., Guo, Z., Li, Y., Qi, Y., Chen, X., Wang, L., Jin, J., Guo, C., Yan, S., et al.: MME-CoT: Benchmarking chain-of-thought in large multimodal models for reasoning quality, robustness, and efficiency. arXiv preprint arXiv:2502.09621 (2025)
23. Jinjin, G., Haoming, C., Haoyu, C., Xiaoxing, Y., Ren, J.S., Chao, D.: PIPAL: a large-scale image quality assessment dataset for perceptual image restoration. In: ECCV (2020)
24. Kang, L., Ye, P., Li, Y., Doermann, D.: Convolutional neural networks for no-reference image quality assessment. In: CVPR (2014)
25. Ke, J., Wang, Q., Wang, Y., Milanfar, P., Yang, F.: MUSIQ: Multi-scale image quality transformer. In: ICCV (2021)
26. Larson, E.C., Chandler, D.M.: Most apparent distortion: full-reference image quality assessment and the role of strategy. Journal of Electronic Imaging (2010)
27. Leng, S., Zhang, H., Chen, G., Li, X., Lu, S., Miao, C., Bing, L.: Mitigating object hallucinations in large vision-language models through visual contrastive decoding. In: CVPR (2024)
28. Li, C., Zhang, Z., Wu, H., Sun, W., Min, X., Liu, X., Zhai, G., Lin, W.: AGIQA-3k: An open database for AI-generated image quality assessment. IEEE TCSVT (2023)
29. Li, W., Zhang, X., Zhao, S., Zhang, Y., Li, J., Zhang, L., Zhang, J.: Q-Insight: Understanding image quality via visual reinforcement learning. In: NeurIPS (2025)
30. Lin, H., Hosu, V., Saupe, D.: KADID-10k: A large-scale artificially distorted IQA database. In: 2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX) (2019)
31. Liu, H., Li, C., Wu, Q., Lee, Y.J.: Visual instruction tuning. In: NeurIPS (2023)
32. Liu, K., Zhang, Z., Li, W., Pei, R., Song, F., Liu, X., Kong, L., Zhang, Y.: DOG-IQA: Standard-guided zero-shot MLLM for mix-grained image quality assessment. arXiv preprint arXiv:2410.02505 (2024)
33. Liu, S., Zheng, K., Chen, W.: Paying more attention to image: A training-free method for alleviating hallucination in LVLMs. In: ECCV (2024)
34. Liu, X., Van De Weijer, J., Bagdanov, A.D.: RankIQA: Learning from rankings for no-reference image quality assessment. In: ICCV (2017)
35. Liu, Z., Sun, Z., Zang, Y., Dong, X., Cao, Y., Duan, H., Lin, D., Wang, J.: Visual-RFT: Visual reinforcement fine-tuning. arXiv preprint arXiv:2503.01785 (2025)
36. Ma, C., Yang, C.Y., Yang, X., Yang, M.H.: Learning a no-reference quality metric for single-image super-resolution. CVIU (2017)
37. Meng, F., Du, L., Liu, Z., Zhou, Z., Lu, Q., Fu, D., Han, T., Shi, B., Wang, W., He, J., et al.: MM-Eureka: Exploring the frontiers of multimodal reasoning with rule-based reinforcement learning. arXiv preprint arXiv:2503.07365 (2025)
38. Mittal, A., Moorthy, A.K., Bovik, A.C.: No-reference image quality assessment in the spatial domain. IEEE TIP (2012)
39. Mittal, A., Soundararajan, R., Bovik, A.C.: Making a "completely blind" image quality analyzer. IEEE Signal processing letters (2012)

40. Moorthy, A.K., Bovik, A.C.: A two-step framework for constructing blind image quality indices. IEEE Signal Processing Letters (2010)
41. Moorthy, A.K., Bovik, A.C.: Blind image quality assessment: From natural scene statistics to perceptual quality. IEEE TIP (2011)
42. Pan, D., Shi, P., Hou, M., Ying, Z., Fu, S., Zhang, Y.: Blind predicting similar quality map for image quality assessment. In: CVPR (2018)
43. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al.: Learning transferable visual models from natural language supervision. In: ICML (2021)
44. Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., Liu, P.J.: Exploring the limits of transfer learning with a unified text-to-text transformer. JMLR (2020)
45. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: CVPR (2022)
46. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O.: Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347 (2017)
47. Shao, H., Qian, S., Xiao, H., Song, G., Zong, Z., Wang, L., Liu, Y., Li, H.: Visual CoT: Advancing multi-modal language models with a comprehensive dataset and benchmark for chain-of-thought reasoning. NeurIPS (2024)
48. Shen, H., Liu, P., Li, J., Fang, C., Ma, Y., Liao, J., Shen, Q., Zhang, Z., Zhao, K., Zhang, Q., et al.: VLM-R1: A stable and generalizable R1-style large vision-language model. arXiv preprint arXiv:2504.07615 (2025)
49. Singh, A., Fry, A., Perelman, A., Tart, A., Ganesh, A., El-Kishky, A., McLaughlin, A., Low, A., Ostrow, A., Ananthram, A., et al.: OpenAI GPT-5 system card. arXiv preprint arXiv:2601.03267 (2025)
50. Su, A., Wang, H., Ren, W., Lin, F., Chen, W.: Pixel reasoner: Incentivizing pixel-space reasoning with curiosity-driven reinforcement learning. arXiv preprint arXiv:2505.15966 (2025)
51. Su, S., Yan, Q., Zhu, Y., Zhang, C., Ge, X., Sun, J., Zhang, Y.: Blindly assess image quality in the wild guided by a self-adaptive hyper network. In: CVPR (2020)
52. Sun, S., Yu, T., Xu, J., Zhou, W., Chen, Z.: GraphIQA: Learning distortion graph representations for blind image quality assessment. IEEE TMM (2022)
53. Sun, X., Lin, Q., Gao, Y., Zhong, Y., Feng, C., Li, D., Zhao, Z., Hu, J., Ma, L.: RFSR: Improving ISR diffusion models via reward feedback learning. arXiv preprint arXiv:2412.03268 (2024)
54. Talebi, H., Milanfar, P.: NIMA: Neural image assessment. IEEE TIP (2018)
55. Thurstone, L.L.: A law of comparative judgment. In: Scaling, pp. 81–92. Routledge (2017)
56. Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., et al.: LLaMA: Open and efficient foundation language models. arXiv preprint arXiv:2302.13971 (2023)
57. Wang, J., Chan, K.C., Loy, C.C.: Exploring CLIP for assessing the look and feel of images. In: AAAI (2023)
58. Wang, J., Lin, S., Lin, Z., Ren, Y., Wei, M., Yue, Z., Zhou, S., Chen, H., Zhao, Y., Yang, C., Xiao, X., Loy, C.C., Jiang, L.: SeedVR2: One-step video restoration via diffusion adversarial post-training. In: ICLR (2026)
59. Wang, J., Lin, Z., Wei, M., Zhao, Y., Yang, C., Loy, C.C., Jiang, L.: SeedVR: Seeding infinity in diffusion transformer towards generic video restoration. In: CVPR (2025)
60. Wang, J., Yue, Z., Zhou, S., Chan, K.C., Loy, C.C.: Exploiting diffusion prior for real-world image super-resolution. IJCV (2024)

61. Wang, Y., Li, Z., Zang, Y., Wang, C., Lu, Q., Jin, C., Wang, J.: Unified multimodal chain-of-thought reward model through reinforcement fine-tuning. NeurIPS (2025)
62. Wu, H., Zhang, Z., Zhang, E., Chen, C., Liao, L., Wang, A., Xu, K., Li, C., Hou, J., Zhai, G., et al.: Q-Instruct: Improving low-level visual abilities for multi-modality foundation models. In: CVPR (2024)
63. Wu, H., Zhang, Z., Zhang, W., Chen, C., Liao, L., Li, C., Gao, Y., Wang, A., Zhang, E., Sun, W., et al.: Q-Align: Teaching LMMs for visual scoring via discrete text-defined levels. In: ICML (2024)
64. Wu, R., Sun, L., Zhang, Z., Wang, S., Wu, T., Yi, Q., Li, S., Zhang, L.: DP^{2}O-SR: Direct perceptual preference optimization for real-world image super-resolution. NeurIPS (2025)
65. Wu, T., Zou, J., Liang, J., Zhang, L., Ma, K.: VisualQuality-R1: Reasoning-induced image quality assessment via reinforcement learning to rank. In: NeurIPS (2025)
66. Yang, A., Yang, B., Zhang, B., Hui, B., Zheng, B., Yu, B., Li, C., Liu, D., Huang, F., Wei, H., Lin, H., Yang, J., Tu, J., Zhang, J., Yang, J., et al.: Qwen2.5 technical report. arXiv preprint arXiv:2412.15115 (2024)
67. Yang, S., Wu, T., Shi, S., Lao, S., Gong, Y., Cao, M., Wang, J., Yang, Y.: ManIQA: Multi-dimension attention network for no-reference image quality assessment. In: CVPR (2022)
68. You, Z., Cai, X., Gu, J., Xue, T., Dong, C.: Teaching large language models to regress accurate image quality scores using score distribution. In: CVPR (2025)
69. You, Z., Gu, J., Li, Z., Cai, X., Zhu, K., Dong, C., Xue, T.: Descriptive image quality assessment in the wild. arXiv preprint arXiv:2405.18842 (2024)
70. You, Z., Li, Z., Gu, J., Yin, Z., Xue, T., Dong, C.: Depicting beyond scores: Advancing image quality assessment through multi-modal language models. In: ECCV (2024)
71. Yu, E., Lin, K., Zhao, L., Yin, J., Wei, Y., Peng, Y., Wei, H., Sun, J., Han, C., Ge, Z., et al.: Perception-R1: Pioneering perception policy with reinforcement learning. arXiv preprint arXiv:2504.07954 (2025)
72. Yu, F., Gu, J., Li, Z., Hu, J., Kong, X., Wang, X., He, J., Qiao, Y., Dong, C.: Scaling up to excellence: Practicing model scaling for photo-realistic image restoration in the wild. In: CVPR (2024)
73. Zhang, J., Huang, J., Yao, H., Liu, S., Zhang, X., Lu, S., Tao, D.: R1-VL: Learning to reason with multimodal large language models via step-wise group relative policy optimization. arXiv preprint arXiv:2503.12937 (2025)
74. Zhang, W., Ma, K., Yan, J., Deng, D., Wang, Z.: Blind image quality assessment using a deep bilinear convolutional neural network. IEEE TCSVT (2018)
75. Zhang, X., Gao, Z., Zhang, B., Li, P., Zhang, X., Liu, Y., Yuan, T., Wu, Y., Jia, Y., Zhu, S.C., et al.: Chain-of-Focus: Adaptive visual search and zooming for multimodal reasoning via RL. arXiv preprint arXiv:2505.15436 (2025)
76. Zhang, X., Li, W., Zhao, S., Li, J., Zhang, L., Zhang, J.: VQ-Insight: Teaching VLMs for AI-generated video quality understanding via progressive visual reinforcement learning. In: AAAI (2026)
77. Zhao, S., Zhang, X., Li, W., Li, J., Zhang, L., Xue, T., Zhang, J.: Reasoning as representation: Rethinking visual reinforcement learning in image quality assessment. arXiv preprint arXiv:2510.11369 (2025)
78. Zheng, Z., Yang, M., Hong, J., Zhao, C., Xu, G., Yang, L., Shen, C., Yu, X.: DeepEyes: Incentivizing "thinking with images" via reinforcement learning. arXiv preprint arXiv:2505.14362 (2025)
79. Zhu, H., Li, L., Wu, J., Dong, W., Shi, G.: MetaIQA: Deep meta-learning for no-reference image quality assessment. In: CVPR (2020)

80. Zhu, H., Wu, H., Li, Y., Zhang, Z., Chen, B., Zhu, L., Fang, Y., Zhai, G., Lin, W., Wang, S.: Adaptive image quality assessment via teaching large multimodal model to compare. In: NeurIPS (2024)