VQAThinker: Exploring Generalizable and Explainable Video Quality Assessment via Reinforcement Learning

Linhan Cao[1*], Wei Sun[2*♡], Weixia Zhang[1], Xiangyang Zhu[3], Jun Jia[1], Kaiwei Zhang[1], Dandan Zhu[2], Guangtao Zhai[1], Xiongkuo Min[1†]

[1]Shanghai Jiao Tong University
[2]East China Normal University
[3]Shanghai Artificial Intelligence Laboratory

Abstract

Video quality assessment (VQA) aims to objectively quantify perceptual quality degradation in alignment with human visual perception. Despite recent advances, existing VQA models still suffer from two critical limitations: *poor generalization to out-of-distribution (OOD) videos* and *limited explainability*, which restrict their applicability in real-world scenarios. To address these challenges, we propose **VQAThinker**, a reasoning-based VQA framework that leverages large multimodal models (LMMs) with reinforcement learning to jointly model video quality understanding and scoring, emulating human perceptual decision-making. Specifically, we adopt group relative policy optimization (GRPO), a rule-guided reinforcement learning algorithm that enables reasoning over video quality under score-level supervision, and introduce three VQA-specific rewards: (1) a **bell-shaped regression reward** that increases rapidly as the prediction error decreases and becomes progressively less sensitive near the ground truth; (2) a **pairwise ranking reward** that guides the model to correctly determine the relative quality between video pairs; and (3) a **temporal consistency reward** that encourages the model to prefer temporally coherent videos over their perturbed counterparts. Extensive experiments demonstrate that VQAThinker achieves state-of-the-art performance on both in-domain and OOD VQA benchmarks, showing strong generalization for video quality scoring. Furthermore, evaluations on video quality understanding tasks validate its superiority in distortion attribution and quality description compared to existing explainable VQA models and LMMs. These findings demonstrate that reinforcement learning offers an effective pathway toward building generalizable and explainable VQA models solely with score-level supervision.

Code — https://github.com/clh124/VQAThinker

Introduction

Video quality assessment (VQA) (Min et al. 2024) plays a fundamental role in video processing systems, serving as the primary criterion for evaluating perceptual quality across stages such as acquisition, compression, and enhancement. Most established VQA models (*e.g.,* VMAF (Li et al. 2018)) rely on full-reference (FR) settings, where a pristine reference video is available for comparison. In many real-world scenarios, particularly those involving user-generated content (UGC), reference videos are unavailable, making FR methods inapplicable. Therefore, no-reference (NR) VQA, which assesses quality without relying on reference videos, offers broader applicability and is the focus of this work [1].

NR-VQA has witnessed significant advancements, progressing from early handcrafted feature-based methods, such as natural scene statistics (NSS) (Mittal, Moorthy, and Bovik 2012) to deep neural network (DNN) architectures, including convolutional neural networks (CNNs) (Li, Jiang, and Jiang 2019), Transformers (Wu et al. 2022), and more recently, large multimodal models (LMMs) (Ge et al. 2024). However, most existing methods rely on supervised fine-tuning (SFT) using specific video content or distortion datasets (Ying et al. 2021; Li et al. 2019), resulting in models that act as specialists for narrow scenarios but exhibit poor generalization to unseen video domains. To enhance generalization, recent studies have explored strategies, such as multi-dataset joint training (Li, Jiang, and Jiang 2019), unsupervised and self-supervised learning (Cao et al. 2025), etc., aiming to improve training diversity while minimizing labeling costs. Nevertheless, these approaches often incur significantly higher training costs due to the increased data scale and model complexity.

In addition, interpretability has become a crucial capability for practical NR-VQA applications. While conventional VQA models typically output singl- or multi-dimensional quality scores, real-world scenarios often require diagnostic assessments that reveal the underlying causes of quality degradation, such as compression artifacts, temporal instability, etc. Such interpretability not only enhances the reliability of VQA predictions but also enables closed-loop video processing by guiding targeted restoration strategies based on identified distortions. Recently, the rise of LMMs has advanced the development of explainable VQA, where models are typically fine-tuned on human-annotated quality instruction datasets, such as VQA^2 (Jia et al. 2024) and OmniVQA-Chat-400K (Jia et al. 2025). However, these datasets are typically bootstrapped from LMM-generated descriptions

*Equal contribution.
♡Project lead.
†Corresponding authors.

[1]In this paper, we use VQA to refer to NR-VQA for simplicity.

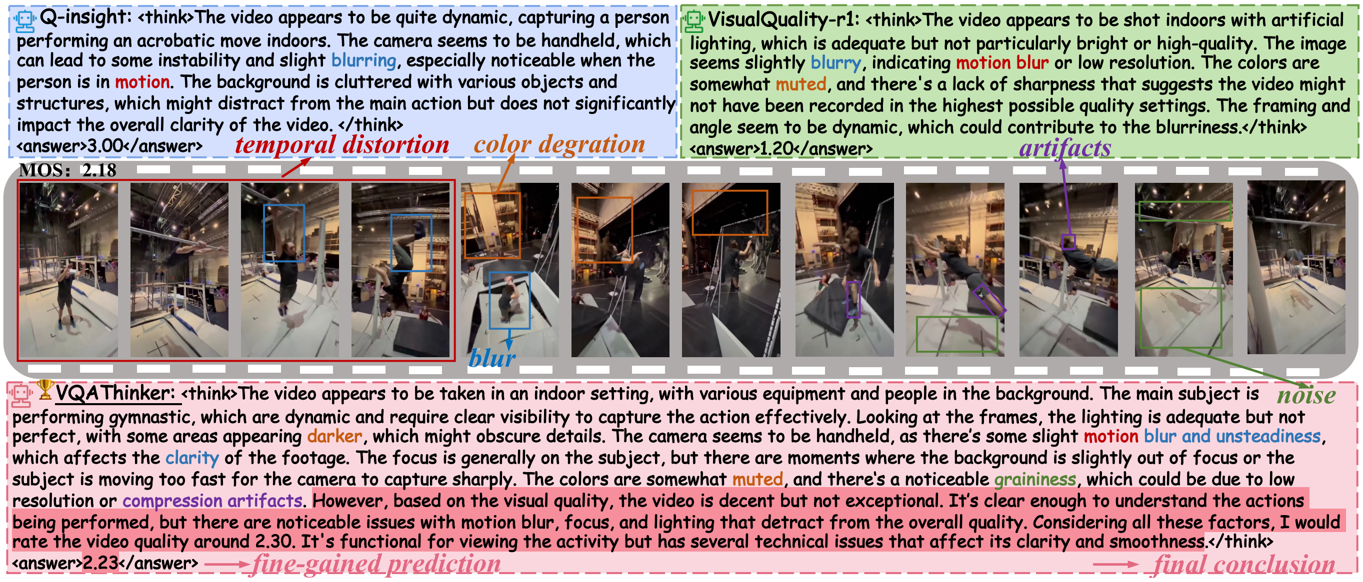

Figure 1: Comparison of Q-Insight, VisualQuality-R1, and our VQAThinker in video quality understanding and scoring. All three models are trained on the LSVQ dataset. Compared to Q-Insight and VisualQuality-R1, VQAThinker generates a more comprehensive reasoning trace that covers major distortion types, thereby producing more accurate video quality scores.

and later refined by human annotators, making the final annotations susceptible to the biases and limitations of the source models. Furthermore, they often lack associated quality scores, hindering the joint learning of quality understanding and scoring, a process believed to mirror human perceptual decision-making (Mazurek et al. 2003).

Recent advances in reinforcement learning (RL), particularly using group relative policy optimization (GRPO) (Shao et al. 2024), have demonstrated strong potential in enhancing the reasoning capabilities of LMMs across various tasks (Li et al. 2025b) while without the need for explicit value function. In the context of quality assessment, this learning paradigm is inherently aligned with the human perceptual process, which involves first reasoning about potential sources of quality degradation in a video and then making a quality judgment accordingly. Although GRPO has been adapted for image quality assessment (IQA) (Li et al. 2025a; Wu et al. 2025) and has shown improved generalizability, existing reward designs in such studies often result in coarse-grained quality predictions and lack temporal distortion modeling capabilities.

To address above challenges, we propose **VQAThinker**, a novel reasoning-driven VQA framework that leverages LMMs with reinforcement learning to jointly model video quality understanding and scoring. Specifically, we adopt GRPO to encourage the model to generate a reasoning process for video quality assessment prior to producing the final quality score, and optimize this process using score-level supervision. Under this framework, we design three VQA-specific reward functions: (1) a **bell-shaped regression reward** that enables fine-grained quality prediction by rapidly increasing rewards as the prediction error decreases and becoming progressively less sensitive near the ground truth, overcoming the limitations of constant or linearly scaled rewards in existing GRPO-based methods such as Q-Insight (Li et al. 2025a) and VQ-Insight (Zhang et al. 2025a); (2) a **pairwise ranking reward** that encourages relative quality consistency across video pairs by explicitly modeling inter-video comparison with fidelity loss (Tsai et al. 2007); and (3) a **temporal consistency reward** that captures temporal distortions across video frames by comparing the model's predictions on raw and temporally perturbed videos.

We train VQAThinker on the LSVQ dataset (Ying et al. 2021) with score-level supervision, and experimental results show that VQAThinker not only achieves state-of-the-art performance on UGC VQA benchmarks but also outperforms existing methods on out-of-distribution (OOD) VQA benchmarks that feature diverse content and distortion types. In addition, we validate its quality understanding ability on the video distortion attribution and video quality description tasks, and the results also demonstrate that VQAThinker delivers competitive performance despite being trained without any video quality instruction data.

Our main contributions are summarized as follows:

- We propose VQAThinker, a reasoning-based VQA method that combines LMMs with reinforcement learning to jointly model quality understanding and scoring.
- We design a GRPO-based training scheme with three VQA-specific rewards to enable fine-grained score regression, quality order preservation, and temporal distortion awareness under score-level supervision.
- Extensive experiments show that VQAThinker achieves state-of-the-art performance on both in-domain and OOD benchmarks, and delivers strong interpretability without using any instruction-tuned datasets.

Related Work

Generalizable Video Quality Assessment

Developing generalizable VQA models remains a long-standing objective. Early studies primarily utilize hand-crafted features based on perceptual quality priors, such as NSS (Mittal, Moorthy, and Bovik 2012), motion vectors (Konrad and Dubois 1992), optical flow (Beauchemin and Barron 1995), etc. While effective on synthetic distortions, these methods demonstrated limited efficacy on videos containing complex real-world distortions. Subsequently, DNN-based methods have been widely adopted for VQA, which typically comprise two main modules: a feature extraction module and a regression module. The feature extraction module often employs off-the-shelf networks, such as CNNs (Li, Jiang, and Jiang 2019), Transformers (Wu et al. 2022), and LMMs (Wu et al. 2023d), which are either pre-trained on large-scale datasets (*e.g.*, IQA datasets (Li et al. 2022)) or fine-tuned directly on the target VQA datasets (Sun et al. 2022; Wu et al. 2022) to capture the spatial and temporal quality representation. The regression module then maps the extracted features to quality scores, using either MLPs that aggregate frame-level predictions (Sun et al. 2024) or sequence modeling methods, such as GRUs (Li, Jiang, and Jiang 2019), InceptionTime (Ying et al. 2021), or Transformers (Wu et al. 2023a), to capture temporal dependencies across frames. DNN-based methods significantly improve the performance on in-the-wild distortions, but they still suffer from limited generalization to unseen video content and distortion types (Cao et al. 2025).

To improve generalization, recent studies have explored various training strategies beyond standard supervised learning. These include mixed-dataset training with scale alignment (Li, Jiang, and Jiang 2019), contrastive learning with proxy tasks (Chen et al. 2021; Madhusudana et al. 2023), distortion-aware reconstruction via encoder-decoder architectures (Xie et al. 2024), and learning-to-rank frameworks using pseudo-labeled video pairs (Cao et al. 2025). Although these methods improve generalization to some extent by increasing the diversity of training data, they typically incur substantial costs in constructing large-scale datasets and demand significantly higher training overhead. In this paper, we demonstrate that VQAThinker enhances generalization without increasing the number of training samples, highlighting its efficiency.

Explainable Video Quality Assessment

Explainable VQA aims to identify the underlying distortions that impact perceptual quality. Some studies focus on developing multi-dimensional VQA models (Wu et al. 2023c; Duan et al. 2025) that produce quality scores across several dimensions, enabling diagnostic insights into the specific distortions affecting video quality. Recently, with the rapid advancement of LMMs, several works have proposed instruction-based datasets for image and video quality understanding, such as Q-Bench (Wu et al. 2023d), Q-Bench-Video (Zhang et al. 2025a), Q-Instruct (Wu et al. 2024), VQA^2 (Jia et al. 2024), and OmniVQA-Chat (Jia et al. 2025), etc. These methods fine-tune LMMs on such datasets to generate descriptive language explanations of visual quality. While effective in producing interpretive outputs, these approaches require large-scale annotated datasets with detailed quality descriptions and often lack a direct connection to quantitative quality scoring.

DeepSeek-R1 (Guo et al. 2025) has shown remarkable effectiveness in promoting reasoning and generalization capabilities of LLMs through reinforcement learning. Q-Insight (Li et al. 2025a) leverages GRPO with a threshold-based binary reward for score prediction and a classification-based binary rewards for distortion type and severity level, to enable reasoning-driven IQA. VisualQuality-R1 (Wu et al. 2025) formulates IQA as a ranking problem and introduces a rank-based reinforcement learning framework with a continuous fidelity loss to model the relative quality between image pairs. As a concurrent work, VQ-Insight (Zhang et al. 2025a) extends Q-Insight by adopting a cold-start reinforcement learning strategy that first warms up LMMs with image-level scoring, then incorporates temporal modeling and multi-task rewards across three training stages to enhance VQA. In contrast to VQ-Insight, our method eliminates the need for a cold-start procedure and multi-stage training, and introduces the dedicated VQA-specific rewards to achieve superior performance and efficiency.

Method

Model Architecture

As illustrated in Figure 2, we utilize an off-the-shelf LMM as the backbone of VQAThinker, which takes a video $\boldsymbol{v}$ and a text prompt p as input, and produces a quality-aware response q that consists of a quality reasoning trace z and a scalar quality score s, delimited by <think> </think> and <answer> </answer> tags, respectively. To enable temporal-aware quality modeling, we additionally incorporate a frozen motion feature extractor to extract local temporal dynamics, along with a motion projector that maps the extracted temporal features into the language space. The detailed model architecture and video preprocessing procedure are provided in the supplementary material. We formalize this inference process as:

$$q = \mathcal{Q}_\theta(p, \boldsymbol{v}, \mathcal{M}(\boldsymbol{v})), \tag{1}$$

where $\mathcal{Q}$ denotes the LMM augmented with the motion projector parameterized by fine-tuned weights θ and $\mathcal{M}$ represents the frozen motion feature extractor.

Reinforcement Learning Strategy

We adopt GRPO to train VQAThinker, which eliminates the need for an explicit value function by leveraging relative comparisons among grouped responses. During training, for a batch of videos $\mathcal{V} = \{\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_N\}$, where N denotes the batch size, GRPO generates K quality-aware response for $\boldsymbol{v}_i$, denoted as $q(\boldsymbol{v}_i) = [q_1(\boldsymbol{v}_i), q_2(\boldsymbol{v}_i), ..., q_K(\boldsymbol{v}_i)]$ using the previous policy $\pi_{\theta_{\text{old}}}$, parameterized by the weights from the previous training epoch θ_{old}. These responses are then evaluated by **VQA-specific rewards** to obtain reward values $r(\boldsymbol{v}_i) = [r_1(\boldsymbol{v}_i), r_2(\boldsymbol{v}_i), ..., r_K(\boldsymbol{v}_i)]$.

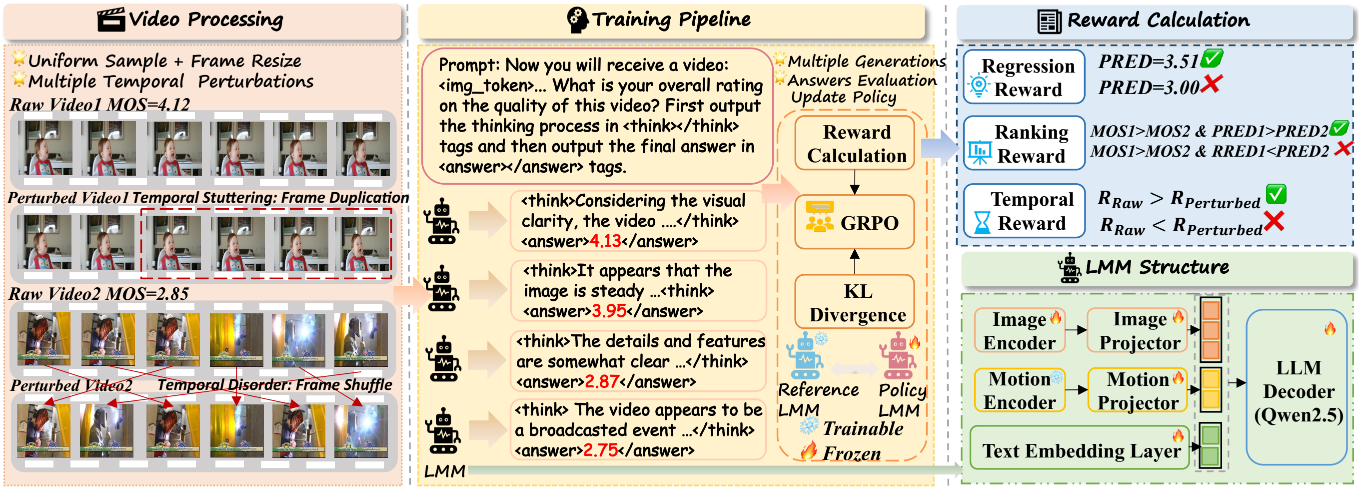

Figure 2: Overall framework of VQAThinker. The architecture consists of a LMM equipped with a frozen motion encoder and a motion projector. During training, the raw video is degraded using a frame perturbation operator to generate a perturbed version. Regression and ranking rewards are computed for both the raw and perturbed videos, while the temporal reward is obtained by comparing the reward differences between them. The regression, ranking, and temporal rewards derived from the raw videos are then used to optimize VQAThinker via the GRPO algorithm.

To evaluate the relative merit of the k-th response $r_k(\boldsymbol{v}_i)$ within the group of responses for video $\boldsymbol{v}_i$, GRPO computes its standardized advantage score $a_k(\boldsymbol{v}_i)$ by normalizing its reward over the K responses:

$$a_k(\boldsymbol{v}_i) = \frac{r_k(\boldsymbol{v}_i) - \mu(r(\boldsymbol{v}_i))}{\sigma(r(\boldsymbol{v}_i))}, \tag{2}$$

where $\mu(\cdot)$ and $\sigma(\cdot)$ denote the mean and standard deviation operators respectively.

The optimization objective of GRPO is defined as:

$$\mathcal{J}(\theta) = \mathbb{E}_{\boldsymbol{v}_i \sim \mathcal{V},\, q_k \sim \pi_{\theta_{\text{old}}}(\cdot|\boldsymbol{v}_i)} \Big[\min\big(\rho_i a_k(\boldsymbol{v}_i),\ \bar{\rho}_i a_k(\boldsymbol{v}_i)\big) - \beta \cdot \mathbb{D}_{\text{KL}}(\pi_\theta(q_k|\boldsymbol{v}_i) \,\|\, \pi_{\text{ref}}(\pi_\theta(q_k|\boldsymbol{v}_i))\Big], \tag{3}$$

where $\rho_i = \frac{\pi_\theta(q_k|\boldsymbol{v}_i)}{\pi_{\theta_{\text{old}}}(q_k|\boldsymbol{v}_i)}$ is the importance sampling ratio between the current policy π_θ and the previous policy $\pi_{\theta_{\text{old}}}$, adjusting the magnitude of the advantage-weighted reward. $\bar{\rho}_i = \text{clip}(\rho_i,\ 1-\varepsilon_s,\ 1+\varepsilon_s)$ clips the ratio within a small range controlled by $\varepsilon_s > 0$ to ensure stable policy updates. β is a coefficient that balances the KL divergence regularization term, which penalizes deviation from a reference policy π_{ref}, typically instantiated by the model parameters before reinforcement learning.

VQA-specific Rewards

The reward function provides training signals that guide the policy toward desirable behaviors. To this end, we design a set of VQA-specific reward functions that enable the model to learn video quality understanding and quality scoring. For the response q_k^i (a simplified notation for $q_k(\boldsymbol{v}_i)$), its rewards include:

Format Reward. To ensure that the model response q_k^i conforms to the expected format, where the reasoning trace is enclosed within <think> tags and the predicted score is enclosed within <answer> tags, we introduce a binary format reward inspired by DeepSeek-R1, defined as:

$$\mathcal{R}_{\text{fmt}}(q_k^i) = \mathbb{I}\left[\text{PatternMatch}(q_k^i)\right], \tag{4}$$

where $\mathbb{I}[\cdot]$ denotes the indicator function that equals 1 when the condition holds and 0 otherwise, and PatternMatch(q_k^i) returns `True` if the output q_k^i satisfies the required format constraints.

Bell-shaped Regression Reward. VQA is naturally formulated as a regression task. However, existing regression-based rewards struggle to provide fine-grained training signals. Some methods use constant binary rewards (Li et al. 2025a), assigning fixed values when predictions fall within a threshold, without capturing their proximity to the ground truth. Others adopt L1-norm-based rewards (Zhang et al. 2025a), where the reward scales linearly with prediction error at a constant rate. In practice, learning becomes increasingly difficult as the predicted score approaches the ground truth. Therefore, the reward function should provide stronger and more rapidly changing signals in this fine-grained regime to better guide the model toward high-precision predictions.

To address this, we propose a bell-shaped regression reward based on a Gaussian function, which increases rapidly as the prediction error decreases and becomes progressively less sensitive near the ground truth, thereby facilitating fine-grained score optimization:

$$\mathcal{R}_{\text{reg}}(q_k^i) = \alpha \cdot \exp\left(-\frac{(s_k^i - g^i)^2}{2\sigma^2}\right), \tag{5}$$

where s_k^i is the predicted quality score extracted from the response q_k^i, g^i is the ground-truth quality score of $\boldsymbol{v}_i$, σ

In-domain Datasets			LSVQ$_{\text{test}}$		LSVQ$_{\text{1080p}}$		KoNViD-1k		LIVE-VQC		YouTube-UGC		Overall	
# of videos			7,182		3,573		1,200		585		1,020		—	
Methods	**Publication**	**Training data**	**SRCC**	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**
Unsupervised Methods														
NIQE	SPL'13	None	0.442	0.332	0.489	0.459	0.541	0.553	0.596	0.628	0.278	0.290	0.457	0.395
VIIDEO	TIP'15		0.080	0.080	0.009	0.019	0.299	0.300	0.033	0.215	0.058	0.154	0.077	0.095
STEM	TIP'21		0.206	0.243	0.434	0.381	0.619	0.627	0.594	0.629	0.284	0.318	0.325	0.336
CLIP-IQA	AAAI'23		0.438	0.413	0.553	0.505	0.696	0.651	0.704	0.683	0.391	0.394	0.499	0.469
Supervised Methods														
SimpleVQA	MM'22	LSVQ (28K)	0.864	0.861	0.756	0.801	0.861	0.860	0.762	0.799	0.808	0.808	0.827	0.838
FAST-VQA	ECCV'22	LSVQ (28K)	0.880	0.880	0.781	0.813	0.859	0.854	**0.826**	0.845	0.730	0.747	0.838	0.849
DOVER	ICCV'23	LSVQ (28K)	0.878	0.866	0.782	0.813	0.874	0.869	0.817	0.840	0.771	0.781	0.842	0.845
MinimalisticVQA	TPAMI'24	LSVQ (28K)	0.885	0.882	0.792	0.828	0.862	0.859	0.775	0.821	0.826	0.821	0.849	0.859
Q-Align	ICML'24	fused (287K)	**0.886**	**0.884**	0.761	0.822	0.876	0.878	0.783	0.819	0.834	0.846	0.844	0.861
VQA^2-Scorer	MM'25	VQA^2-ID (157K)	0.878	0.872	0.794	0.821	**0.881**	0.880	0.785	0.830	0.811	0.823	0.847	0.854
RL-based Methods														
Q-Insight	NeurIPS'25	LSVQ (28K)	0.644	0.639	0.601	0.648	0.751	0.753	0.624	0.708	0.560	0.591	0.635	0.651
VisualQuality-R1	NeurIPS'25	LSVQ (28K)	0.795	0.796	0.716	0.744	0.784	0.792	0.732	0.781	0.717	0.730	0.765	0.776
VQ-Insight	arXiv'25	LSVQ & KonIQ (38K)	0.875	0.876	0.786	0.823	0.875	**0.884**	0.790	0.835	NA	NA	NA	NA
VQAThinker	—	LSVQ (28K)	0.883	0.880	**0.798**	**0.834**	**0.881**	**0.884**	0.808	**0.847**	**0.860**	**0.863**	**0.855**	**0.866**

Out-of-distribution Datasets			LIVE-YT-Gaming		CGVDS		LIVE-YT-HFR		Waterloo-IVC-4K		VDPVE		Overall	
# of videos			600		357		480		1,200		839		—	
Methods	**Publication**	**Training data**	**SRCC**	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**
Unsupervised Methods														
NIQE	SPL'13	None	0.240	0.247	0.473	0.496	0.354	0.413	0.048	0.002	0.415	0.335	0.256	0.232
VIIDEO	TIP'15		0.077	-0.199	0.157	-0.257	0.276	0.244	0.114	0.078	0.298	0.254	0.179	0.061
STEM	TIP'21		0.103	0.111	0.498	0.492	0.288	0.317	0.184	0.097	0.389	0.313	0.266	0.223
CLIP-IQA	AAAI'23		0.358	0.384	0.556	0.565	0.362	0.355	0.133	0.213	0.415	0.446	0.269	0.281
Supervised Methods														
SimpleVQA	MM'22	LSVQ (28K)	0.657	0.728	0.773	0.809	0.416	0.502	0.379	0.425	0.643	0.647	0.536	0.581
FAST-VQA	ECCV'22	LSVQ (28K)	0.631	0.677	0.725	0.747	0.326	0.415	0.327	0.363	0.611	0.620	0.489	0.526
DOVER	ICCV'23	LSVQ (28K)	0.647	0.728	0.694	0.747	0.360	0.465	0.368	0.418	0.627	0.631	0.511	0.563
MinimalisticVQA	TPAMI'24	LSVQ (28K)	0.686	0.746	0.797	0.816	0.301	0.388	0.459	0.502	0.639	0.641	0.555	0.594
Q-Align	ICML'24	fused (287K)	0.611	0.681	0.756	0.798	0.329	0.342	0.414	0.497	0.639	0.649	0.526	0.575
VQA^2-Scorer	MM'25	VQA^2-ID (157K)	0.613	0.698	0.656	0.741	0.332	0.413	0.415	0.474	0.684	0.692	0.527	0.584
RL-based Methods														
Q-Insight	NeurIPS'25	LSVQ (28K)	0.310	0.326	0.372	0.384	0.256	0.268	0.218	0.206	0.547	0.564	0.334	0.340
VisualQuality-R1	NeurIPS'25	LSVQ (28K)	0.472	0.548	0.493	0.574	0.340	0.347	0.227	0.298	0.622	0.637	0.408	0.458
VQ-Insight	arXiv'25	LSVQ & KonIQ (38K)	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
VQAThinker	—	LSVQ (28K)	**0.767**	**0.806**	**0.856**	**0.845**	**0.528**	**0.610**	**0.573**	**0.624**	**0.706**	**0.716**	**0.661**	**0.698**

Table 1: Performance comparison of our model against other competitive methods. The best results are highlight in **boldface**, the second best is underlined. NA in the table indicates unavailable results. “Overall” represents the weighted average results based on the number of videos in each dataset.

controls the sensitivity of the reward to deviations from the ground truth, and $\alpha \in (0, 1]$ is a scaling factor that modulates the overall magnitude of the reward.

Pairwise Ranking Reward. VQA can also be formulated as a ranking problem, where the objective is to learn the relative quality ordering between video pairs based on their predicted scores. This formulation is typically optimized using ranking-based loss functions, such as fidelity loss (Tsai et al. 2007) or approximate Spearman's Rank Correlation Coefficient (SRCC) loss (Li et al. 2022), which measure the consistency between predicted and ground-truth rankings.

Inspired by VisualQuality-R1 (Wu et al. 2025), we adopt a pairwise ranking reward that assigns higher rewards to video pairs whose predicted ranking aligns with the ground-truth rank order. Specifically, for s_k^i, the predicted score for the k-th response q_k^i of video $\boldsymbol{v}_i$, we compare it against the average prediction $\mu(s_j)$ of another video $\boldsymbol{v}_j$ to evaluate whether the predicted ranking aligns with the ground-truth rank order between g_i and g_j. To quantify the likelihood that s_k^i ranks higher than $\mu(s_j)$, we compute comparative probability $p_k(\boldsymbol{v}_i, \boldsymbol{v}_j)$ using the cumulative distribution function of the standard normal distribution:

$$p_k(\boldsymbol{v}_i, \boldsymbol{v}_j) = \Phi\left(\frac{s_i^j - \mu(s_j)}{\sqrt{\sigma(s_i)^2 + \sigma(s_j)^2 + \epsilon}}\right),$$

where $\sigma(s_i)^2$ and $\sigma(s_j)^2$ denote the variance of K predicted scores for $\boldsymbol{v}_i$ and $\boldsymbol{v}_j$, respectively, and ϵ is a small positive constant for numerical stability, and the subsequent occurrences ϵ serve the same purpose.

The pairwise ranking reward is computed based on the fidelity loss (Tsai et al. 2007), and is defined as:

$$\mathcal{R}_{\text{rank}}(q_k^i) = \sqrt{p_k(\boldsymbol{v}_i, \boldsymbol{v}_j) \cdot \mathbb{I}[g_i > g_j] + \epsilon} \quad (6)$$

$$+ \sqrt{(1 - p_k(\boldsymbol{v}_i, \boldsymbol{v}_j)) \cdot \mathbb{I}[g_i < g_k] + \epsilon}, \quad (7)$$

where $\mathbb{I}[\cdot]$ is the indicator function.

Temporal Consistency Reward. We have introduced a motion feature extractor to capture local temporal dynamics. However, this module is limited in handling long-range temporal distortions that may occur across the sampled frames.

To address this issue, we propose a temporal consistency reward that explicitly encourages the model to be sensitive to broader temporal degradations. Specifically, for video $\boldsymbol{v}_i$, we construct a temporally perturbed counterpart $\boldsymbol{v}_{i,\text{temp}}$ by applying various frame-level distortions, such as frame repetition and frame dropping, to simulate real-world temporal artifacts like frame lagging and frame loss (see supplementary material for details). For the responses q_i and $q_{i,\text{temp}}$, we compute both the regression reward $\mathcal{R}_{\text{reg}}(\cdot)$ and the ranking reward $\mathcal{R}_{\text{rank}}(\cdot)$. The temporal consistency reward is then defined to encourage model behavior that yields a higher regression or ranking reward for $\boldsymbol{v}_i$ than for $\boldsymbol{v}_{i,\text{temp}}$:

$$\mathcal{R}_{\text{temp}}^{(t)}(q_k^i) = \begin{cases} \delta, & \text{if } \mu(r_i^{(t)}) \geq \mu(r_{i,\text{temp}}^{(t)}) \text{ and } \mu(r_i^{(t)}) > \tau \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

where $t \in \{\text{reg}, \text{rank}\}$ indicates the reward type, $\mu(\cdot)$ denotes the average reward across all responses, δ is a fixed bonus, and τ is a confidence threshold to ensure that temporal consistency rewards are only granted when the model is sufficiently accurate. The final temporal consistency reward [2] is the sum of the two sub-rewards:

$$\mathcal{R}_{\text{temp}}(q_k^i) = \mathcal{R}_{\text{temp}}^{(\text{reg})}(q_k^i) + \mathcal{R}_{\text{temp}}^{(\text{rank})}(q_k^i). \tag{9}$$

The final reward is computed as the sum of the three components:

$$r_k(\boldsymbol{v}_i) = \mathcal{R}_{\text{fmt}}(q_k^i) + \mathcal{R}_{\text{reg}}(q_k^i) + \mathcal{R}_{\text{rank}}(q_k^i) + \mathcal{R}_{\text{temp}}(q_k^i). \tag{10}$$

Experiments

Experimental Setups

Benchmark Datasets. VQAThinker is trained on the training set of LSVQ (Ying et al. 2021) containing $28,056$ UGC videos. We evaluate VQAThinker from two perspectives: **video quality scoring** and **video quality understanding**.

For **video quality scoring**, we evaluate performance on ten VQA benchmarks: LSVQ Test (Ying et al. 2021), LSVQ 1080p (Ying et al. 2021), KoNViD-1k (Hosu et al. 2017), LIVE-VQC (Sinno and Bovik 2018), YouTube-UGC (Wang, Inguva, and Adsumilli 2019), LIVE-YT-Gaming (Yu et al. 2022), CGVDS (Saha et al. 2023), LIVE-YT-HFR (Madhusudana et al. 2021), Waterloo-IVC-4K (Li et al. 2019), and VDPVE (Gao et al. 2023). The first five datasets are considered in-domain, as they primarily consist of UGC videos with in-the-wild distortions. The remaining five datasets are treated as OOD due to their distinct content types (*e.g.,* gaming videos) or distortion patterns (*e.g.,* compression and enhancement artifacts). For **video quality understanding**, we adopt two evaluation benchmarks: FineVD (Duan et al. 2025) and Q-Bench-Video (Zhang et al. 2025b). FineVD is used to assess the model's ability to identify quality-related distortion attributes, while Q-Bench-Video measures the accuracy of video quality descriptions.

[2]Note that the rewards from $\boldsymbol{v}_{i,\text{temp}}$ are used only for comparison, and are not included in the reinforcement learning objective.

Model / Attribute	Color	Noise	Artifact	Blur	Temporal	Overall
Open-source LMMs						
VideoChat2	18.25	18.80	24.26	31.15	15.19	21.53
VideoLLaMA2	57.87	63.98	51.87	61.91	**86.42**	64.41
Video-ChatGPT	28.35	27.66	44.49	62.89	18.60	36.40
MiniCPM-V	70.47	65.85	<u>52.66</u>	46.26	71.95	61.44
LLaVA-NeXT	26.87	27.26	43.90	64.76	13.09	35.18
Video-LLaVA	26.48	27.26	43.90	64.67	13.09	35.08
Qwen2.5-VL	68.49	48.74	47.51	52.08	56.25	54.61
InternVL3	64.63	62.18	52.49	52.98	68.01	60.06
Explainable VQAs						
FineVQ	<u>73.52</u>	**72.74**	51.87	64.76	65.06	<u>65.59</u>
Q-Insight	63.99	52.52	48.21	64.58	68.01	59.46
VisualQuality-R1	65.27	55.46	48.17	<u>65.77</u>	63.24	59.58
VQAThinker	**82.64**	<u>70.17</u>	**55.15**	**70.83**	<u>83.46</u>	**72.45**

Table 2: Performance comparison of explainable VQA methods and LMMs on the FineVD dataset for the video distortion attribution task. All values are in % (percentage).

Competing Methods We compare VQAThinker with three categories of methods: 1) unsupervised methods: NIQE (Mittal, Soundararajan, and Bovik 2012), VIIDEO (Mittal, Saad, and Bovik 2015), STEM (Kancharla and Channappayya 2021), and CLIP-IQA (Wang, Chan, and Loy 2023); 2) supervised methods: SimpleVQA (Sun et al. 2022), FAST-VQA (Wu et al. 2022), DOVER (Wu et al. 2023b), MinimalisticVQA (IX) (Sun et al. 2024), Q-Align (Wu et al. 2023d) and VQA^2-Scorer (Jia et al. 2024); 3) reinforcement learning methods: Q-Insight (Li et al. 2025a), VisualQuality-R1 (Wu et al. 2025), and VQ-Insight (Zhang et al. 2025a). To ensure fairness, we retrain the IQA-based methods such as Q-Insight and VisualQuality-R1 on the LSVQ training set, using the same sampled frames as VQAThinker to adapt them for the VQA task.[3]

Implementation Details. We utilize InternVL3-8B (Zhu et al. 2025) as the backbone LMM and SlowFast (Feichtenhofer et al. 2019) as the motion feature extractor, both initialized with their publicly released weights. During training, the motion feature extractor is kept frozen, while all other components are fine-tuned. In the GRPO trainer, the number of response K is set to 4, and the weight of the KL penalty β is set to 0.04. The model is trained for three epochs on 8 NVIDIA A800 GPUs with a total batch size $N = 64$ and a learning rate of 1×10^{-6}. We set $\sigma = 0.5$ and $\alpha = 0.8$ in Eq. (5), and $\delta = 0.3$ and $\tau = 0.5$ in Eq. (8).

Performance Analysis

Video Quality Scoring. The SRCC and PLCC results across ten VQA benchmarks are summarized in Table 1. For in-domain evaluation, VQAThinker achieves state-of-the-art performance on nearly all benchmarks and outperforms all competing methods in terms of overall accuracy. Compared to VQ-Insight, a recent RL-based VQA method, VQAThinker demonstrates consistently superior performance on the available benchmarks, despite adopting a more streamlined training process with fewer training samples. For OOD evaluation, VQAThinker exhibits a substantial advantage

[3]Since VQ-Insight is not open-sourced, we report its performance based on the official results.

Model	Global↑	Referring↑	Overall↑
Open-source LMMs			
VideoChat2	38.91	40.47	39.68
VILA1.5	46.55	48.69	47.61
InternVL-Chat	49.43	54.05	51.72
LLaVA-OneVision	49.41	49.35	49.38
LLaVA-Next-Video	47.01	52.23	49.59
mPLUG-Owl3	51.71	48.78	50.26
Video-LLaVA	43.66	47.73	45.67
Qwen2.5-VL	55.71	53.75	54.66
InternVL3	50.87	50.75	50.80
Proprietary LMMs			
Gemini 1.5 Pro	51.98	**60.80**	56.35
GPT-4o	57.98	54.39	56.17
GPT-4o mini	50.03	49.88	49.95
Explainable VQAs			
VQA^2-Assistant	**61.59**	<u>57.36</u>	**59.32**
Q-Insight	55.71	53.45	54.50
VisualQuality-R1	57.09	54.35	55.63
VQAThinker	<u>58.82</u>	56.46	<u>57.56</u>

Table 3: Performance comparison of explainable VQA methods and LMMs on the Q-Bench-Video (dev) dataset for the video quality description task. All values are in %.

over existing methods, regardless of content domain (*e.g.,* gaming videos in LIVE-YT-Gaming and CGVDS) or distortion type (*e.g.,* frame rate in LIVE-YT-HFR, compression in Waterloo-IVC-4K, and enhancement in VDPVE). Overall, it achieves a 19.1% relative improvement in average SRCC on OOD datasets, without relying on additional training samples to increase data diversity. These results demonstrate the effectiveness and strong generalization capability of our approach.

Video Quality Understanding. The results on FineVD and Q-Bench-Video are presented in Table 2 and Table 3 respectively. *Notably, VQAThinker is not explicitly trained on any video quality instruction datasets; instead, we directly evaluate its reasoning traces as a measure of the model's quality understanding ability.* Despite this, VQAThinker achieves strong performance on both benchmarks. On FineVD, which assesses distortion attribution capabilities, VQAThinker significantly outperforms open-source LMMs and RL-based quality assessment models. Remarkably, it even surpasses FineVQ, a model specifically fine-tuned on the FineVD training set. Similarly, on Q-Bench-Video, which evaluates the overall quality description ability, VQAThinker outperforms all open-source and proprietary LMMs as well as RL-based baselines, and is only 1.76% below VQA^2-Assistant, a model trained on 157K instruction–answer pairs. These results demonstrate that the proposed reinforcement learning framework, equipped with VQA-specific rewards, effectively enables the model to reason about potential video distortions that affect perceptual quality.

Ablation Study

Motion Feature Extractor. We incorporate a frozen motion feature extractor to capture local temporal dynamics. To assess its standalone effect, we ablate it under the setting without the temporal consistency reward. As shown in Table 4, introducing the motion feature extractor consistently improves both in-domain and OOD performance, validating its effectiveness in modeling local temporal distortions.

Reward			Motion	Reasoning	In-domain		OOD	
Rank	Reg.	Temp.			SRCC	PLCC	SRCC	PLCC
✔			✔	✔	0.780	0.785	0.506	0.541
	✔		✔	✔	0.823	0.834	0.620	0.652
✔	✔		✔	✔	0.849	0.860	0.638	0.677
✔	✔			✔	0.840	0.853	0.614	0.650
✔	✔	✔	✔		0.844	0.848	0.627	0.660
✔	✔	✔	✔	✔	0.855	0.866	0.662	0.698

Table 4: Performance of ablation studies. "Rank", "Reg.", and "Temp." denote the pairwise ranking reward, bell-shaped regression reward, and temporal consistency reward, respectively. "Motion" refers to whether the motion feature extractor are included, and "Reasoning" indicates whether the reasoning output is required. The in-domain and OOD results are averaged across the corresponding test sets.

VQA-specific Rewards. We first evaluate the individual impact of the bell-shaped regression reward and the pairwise ranking reward. From Table 4, the bell-shaped regression reward yields significantly better performance than the pairwise ranking reward, demonstrating its effectiveness in promoting fine-grained quality assessment. When the two rewards are combined, performance further improves, indicating that the pairwise ranking reward complements the regression objective by enhancing the model's ability to discern the relative quality order between video pairs. Adding the temporal consistency reward, which is designed to account for long-term temporal distortions, leads to additional performance gains, validating its contribution to improving temporal sensitivity in VQA.

Reasoning. We incorporate the format reward to guide the model to generate a reasoning trace prior to producing the final quality score, which better aligns with the human perceptual decision-making process. To assess its impact, we remove the <think> tags from the format reward, disabling the reasoning generation. As shown in Table 4, we observe a consistent performance drop across both in-domain and OOD benchmarks, with more pronounced degradation on the OOD datasets. It indicates that the explicit reasoning procedure is important for enhancing the model's video quality scoring capability.

Conclusion

In this work, we propose VQAThinker, a reasoning-based VQA framework that leverages GRPO to achieve generalizable and explainable VQA. Specifically, we introduce three VQA-specific rewards: a bell-shaped regression reward for fine-grained score prediction, a pairwise ranking reward for relative quality discrimination, and a temporal consistency reward for modeling temporal degradations. Extensive experiments across ten VQA benchmarks and two quality description datasets demonstrate that VQAThinker achieves state-of-the-art performance in both in-domain and OOD settings, while also providing strong interpretability through explicit reasoning traces.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grant 62522116, Grant 62271312, Grant 62132006, and Grant 62301316, and in part by STCSM under Grant 22DZ2229005.

References

Beauchemin, S. S.; and Barron, J. L. 1995. The computation of optical flow. *ACM Computing Surveys*, 27(3): 433–466.

Cao, L.; Sun, W.; Zhu, X.; Zhang, K.; Jia, J.; Peng, Y.; Zhu, D.; Zhai, G.; and Min, X. 2025. Generalizable video quality assessment via weak-to-strong learning. *arXiv preprint arXiv:2505.03631*.

Chen, P.; Li, L.; Wu, J.; Dong, W.; and Shi, G. 2021. Contrastive self-supervised pre-training for video quality assessment. *IEEE Transactions on Image Processing*, 31: 458–471.

Duan, H.; Hu, Q.; Wang, J.; Yang, L.; Xu, Z.; Liu, L.; Min, X.; Cai, C.; Ye, T.; Zhang, X.; et al. 2025. Finevq: Fine-grained user generated content video quality assessment. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 3206–3217.

Feichtenhofer, C.; Fan, H.; Malik, J.; and He, K. 2019. Slowfast networks for video recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 6202–6211.

Gao, Y.; Cao, Y.; Kou, T.; Sun, W.; Dong, Y.; Liu, X.; Min, X.; and Zhai, G. 2023. Vdpve: Vqa dataset for perceptual video enhancement. In *IEEE Conference on Computer Vision and Pattern Recognition*, 1474–1483.

Ge, Q.; Sun, W.; Zhang, Y.; Li, Y.; Ji, Z.; Sun, F.; Jui, S.; Min, X.; and Zhai, G. 2024. LMM-VQA: Advancing video quality assessment with large multimodal models. *arXiv preprint arXiv:2408.14008*.

Guo, D.; Yang, D.; Zhang, H.; Song, J.; Zhang, R.; Xu, R.; Zhu, Q.; Ma, S.; Wang, P.; Bi, X.; et al. 2025. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*.

Hosu, V.; Hahn, F.; Jenadeleh, M.; Lin, H.; Men, H.; Szirányi, T.; Li, S.; and Saupe, D. 2017. The konstanz natural video database (KoNViD-1k). In *International Conference on Quality of Multimedia Experience*, 1–6.

Jia, Z.; Zhang, Z.; Qian, J.; Wu, H.; Sun, W.; Li, C.; Liu, X.; Lin, W.; Zhai, G.; and Min, X. 2024. VQA ˆ2: Visual question answering for video quality assessment. *arXiv preprint arXiv:2411.03795*.

Jia, Z.; Zhang, Z.; Zhang, Z.; Liang, Y.; Zhu, X.; Li, C.; Han, J.; Wu, H.; Wang, B.; Zhang, H.; et al. 2025. Scaling-up perceptual video quality assessment. *arXiv preprint arXiv:2505.22543*.

Kancharla, P.; and Channappayya, S. S. 2021. Completely blind quality assessment of user generated video content. *IEEE Transactions on Image Processing*, 31: 263–274.

Konrad, J.; and Dubois, E. 1992. Bayesian estimation of motion vector fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(09): 910–927.

Li, B.; Zhang, W.; Tian, M.; Zhai, G.; and Wang, X. 2022. Blindly assess quality of in-the-wild videos via quality-aware pre-training and motion perception. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(9): 5944–5958.

Li, D.; Jiang, T.; and Jiang, M. 2019. Quality assessment of in-the-wild videos. In *Proceedings of the ACM international Conference on Multimedia*, 2351–2359.

Li, W.; Zhang, X.; Zhao, S.; Zhang, Y.; Li, J.; Zhang, L.; and Zhang, J. 2025a. Q-insight: Understanding image quality via visual reinforcement learning. *arXiv preprint arXiv:2503.22679*.

Li, Y.; Liu, Z.; Li, Z.; Zhang, X.; Xu, Z.; Chen, X.; Shi, H.; Jiang, S.; Wang, X.; Wang, J.; et al. 2025b. Perception, reason, think, and plan: A survey on large multimodal reasoning models. *arXiv preprint arXiv:2505.04921*.

Li, Z.; Bampis, C.; Novak, J.; Aaron, A.; Swanson, K.; Moorthy, A.; and Cock, J. 2018. VMAF: The journey continues. *Netflix Technology Blog*, 25.

Li, Z.; Duanmu, Z.; Liu, W.; and Wang, Z. 2019. AVC, HEVC, VP9, AVS2 OR AV1?—A comparative study of state-of-the-art video encoders on 4K videos. In *International Conference on Image Analysis and Recognition*, 162–173. Springer.

Madhusudana, P. C.; Birkbeck, N.; Wang, Y.; Adsumilli, B.; and Bovik, A. C. 2023. Conviqt: Contrastive video quality estimator. *IEEE Transactions on Image Processing*, 32: 5138–5152.

Madhusudana, P. C.; Yu, X.; Birkbeck, N.; Wang, Y.; Adsumilli, B.; and Bovik, A. C. 2021. Subjective and objective quality assessment of high frame rate videos. *IEEE Access*, 9: 108069–108082.

Mazurek, M. E.; Roitman, J. D.; Ditterich, J.; and Shadlen, M. N. 2003. A role for neural integrators in perceptual decision making. *Cerebral Cortex*, 13(11): 1257–1269.

Min, X.; Duan, H.; Sun, W.; Zhu, Y.; and Zhai, G. 2024. Perceptual video quality assessment: A survey. *Science China Information Sciences*, 67(11): 211301.

Mittal, A.; Moorthy, A. K.; and Bovik, A. C. 2012. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*, 21(12): 4695–4708.

Mittal, A.; Saad, M. A.; and Bovik, A. C. 2015. A completely blind video integrity oracle. *IEEE Transactions on Image Processing*, 25(1): 289–300.

Mittal, A.; Soundararajan, R.; and Bovik, A. C. 2012. Making a "completely blind" image quality analyzer. *IEEE Signal Processing Letters*, 20(3): 209–212.

Saha, A.; Chen, Y.-C.; Davis, C.; Qiu, B.; Wang, X.; Gowda, R.; Katsavounidis, I.; and Bovik, A. C. 2023. Study of subjective and objective quality assessment of mobile cloud gaming videos. *IEEE Transactions on Image Processing*, 32: 3295–3310.

Shao, Z.; Wang, P.; Zhu, Q.; Xu, R.; Song, J.; Bi, X.; Zhang, H.; Zhang, M.; Li, Y.; Wu, Y.; et al. 2024. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*.

Sinno, Z.; and Bovik, A. C. 2018. Large-scale study of perceptual video quality. *IEEE Transactions on Image Processing*, 28(2): 612–627.

Sun, W.; Min, X.; Lu, W.; and Zhai, G. 2022. A deep learning based no-reference quality assessment model for ugc videos. In *Proceedings of the ACM International Conference on Multimedia*, 856–865.

Sun, W.; Wen, W.; Min, X.; Lan, L.; Zhai, G.; and Ma, K. 2024. Analysis of video quality datasets via design of minimalistic video quality models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Tsai, M.-F.; Liu, T.-Y.; Qin, T.; Chen, H.-H.; and Ma, W.-Y. 2007. Frank: a ranking method with fidelity loss. In *Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 383–390.

Wang, J.; Chan, K. C.; and Loy, C. C. 2023. Exploring clip for assessing the look and feel of images. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 2555–2563.

Wang, Y.; Inguva, S.; and Adsumilli, B. 2019. YouTube UGC dataset for video compression research. In *International Workshop on Multimedia Signal Processing*, 1–5. IEEE.

Wu, H.; Chen, C.; Hou, J.; Liao, L.; Wang, A.; Sun, W.; Yan, Q.; and Lin, W. 2022. Fast-vqa: Efficient end-to-end video quality assessment with fragment sampling. In *European Conference on Computer Vision*, 538–554. Springer.

Wu, H.; Chen, C.; Liao, L.; Hou, J.; Sun, W.; Yan, Q.; and Lin, W. 2023a. Discovqa: Temporal distortion-content transformers for video quality assessment. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(9): 4840–4854.

Wu, H.; Zhang, E.; Liao, L.; Chen, C.; Hou, J.; Wang, A.; Sun, W.; Yan, Q.; and Lin, W. 2023b. Exploring video quality assessment on user generated contents from aesthetic and technical perspectives. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 20144–20154.

Wu, H.; Zhang, E.; Liao, L.; Chen, C.; Hou, J.; Wang, A.; Sun, W.; Yan, Q.; and Lin, W. 2023c. Towards explainable in-the-wild video quality assessment: a database and a language-prompted approach. In *Proceedings of the ACM International Conference on Multimedia*, 1045–1054.

Wu, H.; Zhang, Z.; Zhang, E.; Chen, C.; Liao, L.; Wang, A.; Xu, K.; Li, C.; Hou, J.; Zhai, G.; et al. 2024. Q-instruct: Improving low-level visual abilities for multi-modality foundation models. In *IEEE Conference on Computer Vision and Pattern Recognition*, 25490–25500.

Wu, H.; Zhang, Z.; Zhang, W.; Chen, C.; Liao, L.; Li, C.; Gao, Y.; Wang, A.; Zhang, E.; Sun, W.; et al. 2023d. Q-align: Teaching lmms for visual scoring via discrete text-defined levels. *arXiv preprint arXiv:2312.17090*.

Wu, T.; Zou, J.; Liang, J.; Zhang, L.; and Ma, K. 2025. VisualQuality-R1: Reasoning-Induced image quality sssessment via reinforcement learning to rank. *arXiv preprint arXiv:2505.14460*.

Xie, Q.; Yuan, K.; Qu, Y.; Wu, M.; Sun, M.; Zhou, C.; and Zhu, J. 2024. QPT-V2: Masked image modeling advances visual scoring. In *Proceedings of the ACM International Conference on Multimedia*, 2709–2718.

Ying, Z.; Mandal, M.; Ghadiyaram, D.; and Bovik, A. 2021. Patch-vq:'patching up'the video quality problem. In *IEEE Conference on Computer Vision and Pattern Recognition*, 14019–14029.

Yu, X.; Tu, Z.; Ying, Z.; Bovik, A. C.; Birkbeck, N.; Wang, Y.; and Adsumilli, B. 2022. Subjective quality assessment of user-generated content gaming videos. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 74–83.

Zhang, X.; Li, W.; Zhao, S.; Li, J.; Zhang, L.; and Zhang, J. 2025a. VQ-Insight: Teaching VLMs for AI-Generated video quality understanding via progressive visual reinforcement learning. *arXiv preprint arXiv:2506.18564*.

Zhang, Z.; Jia, Z.; Wu, H.; Li, C.; Chen, Z.; Zhou, Y.; Sun, W.; Liu, X.; Min, X.; Lin, W.; et al. 2025b. Q-Bench-Video: Benchmark the video quality understanding of LMMs. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 3229–3239.

Zhu, J.; Wang, W.; Chen, Z.; Liu, Z.; Ye, S.; Gu, L.; Tian, H.; Duan, Y.; Su, W.; Shao, J.; et al. 2025. Internvl3: Exploring advanced training and test-time recipes for open-source multimodal models. *arXiv preprint arXiv:2504.10479*.