Seeing Beyond 8bits: Subjective and Objective Quality Assessment of HDR-UGC Videos

Supplementary Material

This supplementary material expands on the technical, experimental, and ethical components of HDR-Q that could not be included in the main paper due to space constraints. It provides additional dataset details, methodological clarification, and extended empirical analysis to fully support the claims made in the primary manuscript.

- **Related Work:** Extended discussion of HDR-VQA and MLLM-based perceptual quality literature.
- **Dataset Details:** Comprehensive documentation of collection pipeline, AMT protocol, QC procedures, HDR verification, reliability analyses, and MOS behaviors.
- **Dataset Analysis:** Content diversity, SI–TI characteristics, MOS distributions, bitrate–resolution trends, and demographic insights.
- **Method Details:** Additional exposition of HAPO, reward design, two-stage RL training, and HDR-aware encoder formulation.
- **Qualitative Examples:** Additional visualizations of reasoning output of our model.
- **Ethical Considerations:** Discussion of privacy, dataset usage constraints, bias, and responsible deployment.

A. Related Work

A.1. HDR-VQA: Datasets and Models

Subjective video quality datasets such as CVD2014 [23], LIVE-VQA [30], LIVE-VQC [33], LSVQ [45], MD-VQA [50], and Maxwell [40] have driven progress in SDR VQA by supporting the development of handcrafted models [9, 13, 19–21, 28] and deep learning architectures [12, 14, 17, 37–39]. These datasets however operate entirely in SDR, with limited luminance ranges and no notion of PQ-transfer or wide color gamut content, making them insufficient for algorithms intended to handle HDR-specific perceptual phenomena. To address HDR scenarios, several early HDR VQA datasets were introduced [3–5, 24, 27]. Many of these collections involve small content diversity, outdated HDR standards, or limited accessibility, which restricts their modern usability. LIVE-HDR [31] provides a more contemporary benchmark with 310 annotated clips generated under controlled distortions. SFV+HDR [35] expands toward short-form user content with 2k clips, yet only 300 are annotated due to high labeling cost. The field therefore lacks sufficiently large, diverse HDR datasets capturing practical user-generated conditions, tone-mapping inconsistencies, sensor noise, or extreme dynamic-range variation. Parallel to dataset development, several HDR VQA algorithms have been proposed. Full-reference metrics such as HDR-VQM [22], HDR-BVQM [1], and PU21 [18] leverage perceptually uniform luminance transforms or brightness-adaptive operators. These methods however assume pristine references and are not suited to the unpredictable variability of UGC. Blind HDR models emerged more recently. HDR-ChipQA [10] extends ChipQA with nonlinear luminance modeling, while HIDRO-VQA [29] adapts CONTRIQUE [17] through large-scale pretraining on unlabeled YouTube HDR videos. Although these approaches incorporate HDR-specific priors, they do not robustly handle HDR-UGC distortions such as clipped highlights, crushed shadows, quantization banding, or gamut shifts that dominate modern consumer-captured HDR. This motivates the construction of larger HDR-UGC datasets and more principled modeling frameworks capable of reasoning about HDR-specific perceptual cues.

A.2. MLLM-Based Perceptual Quality Assessment

Multimodal large language models have increasingly been applied to perceptual quality assessment. Q-Bench [41] established that general-purpose MLLMs remain far from human perception, with large inconsistencies in both ranking and scoring tasks. Instruction tuning approaches such as Q-Instruct [43] and Q-Bench's instruction-aligned variants improved distortion awareness by coupling low-level quality cues with textual descriptions. DepictQA [47] and DepictQA-Wild [46] moved toward more naturalistic distortion explanations by prompting models to produce descriptive rationales before scoring. Several studies have proposed more structured formulations. Compare2Score [53] demonstrated that pairwise comparison signals are easier for MLLMs to model and can be meta-learned to form continuous scores. Reinforcement learning-to-rank was explored in VisualQuality-R1 [44], which improved alignment with human preferences. Discrete quality level calibration methods such as Q-Align [42] improved robustness but sacrificed continuous regression fidelity. DeQA-Score [48] incorporated multi-dataset training with soft labels to better capture MOS distributions. Q-Insight [15] used reinforcement learning to improve joint score prediction and degradation perception.

Video-focused extensions include Q-Bench-Video [51], which provided the first unified benchmark for evaluating MLLMs on spatiotemporal distortions, and MVQA-68K [25], which introduced large-scale multi-attribute labels and textual rationales for training video-aware reason-

ing models. These works collectively point toward the feasibility of MLLMs for perceptual quality assessment, yet they remain fundamentally SDR-based. None explicitly address HDR, which involves different perceptual sensitivities, tone-mapping behaviors, and signal statistics. Moreover, existing approaches often rely on supervised or instruction tuning, whereas reinforcement learning for perceptual tasks is still in early stages.

Our work extends this direction by introducing an HDR-aware vision encoder and a reinforcement learning framework explicitly optimized for HDR grounding, filling the gap between HDR-specific perception research and modern multimodal reasoning frameworks.

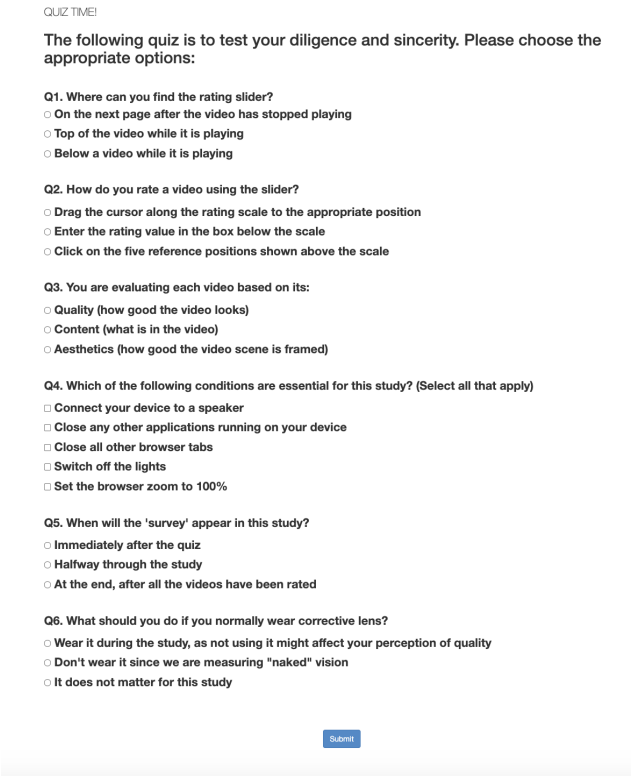


Figure 1. Quiz phase on AMT.

B. Dataset Details

B.1. Data Collection Sources

As mentioned in the main paper, source videos were collected from:

- **Crowdsourcing Campaign:** An open call was made for users to submit HDR videos captured on their personal devices (various recent models of iPhones, Samsung Galaxy, Google Pixel, etc.). Submissions required user consent for research use. This yielded diverse, authentic UGC footage.
- **Vimeo:** Videos licensed under Creative Commons were identified using Vimeo's search filters for HDR content. Manual screening ensured the videos were genuinely UGC (or representative of high-quality UGC) rather than professional productions. Categories included travel vlogs, personal events, amateur sports recordings, nature footage, etc.

Some example frames from our HDR video dataset are shown in Figure 2.

B.2. Video Filtering and Processing

- **Initial Filtering:** Automated checks removed videos with non-HDR flags, incompatible codecs, very low resolutions, or durations outside a reasonable range (e.g., $<$ 4s or $>$ 60s). Duplicate detection was performed. Manual screening removed clearly PGC content, static videos, and content violating ethical guidelines (privacy, safety).
- **Trimming:** Videos were trimmed to a maximum of 10 seconds, typically selecting a segment with representative motion and content complexity.
- **Bitrate Ladder Transcoding:** Using FFmpeg, source videos (considered 'Reference' quality) were transcoded to various resolution/bitrate combinations as specified in Table 1. Encoding used HEVC Main 10 profile, with PQ transfer function and Rec.2020 color primaries, matching typical HDR10 standards. Constant Quality Rate Factor (CRF)was used to target the specified bitrates.

Table 1. Bitrate ladder used for dataset creation, simulating real-world streaming conditions [2, 11].

Resolution	Bitrates (Mbps)
360p	0.2, 0.5
720p	0.5, 1.0, 2.0
1080p	1.0, 2.0, 3.0, 5.0
1080p (Source)	Reference

B.3. Crowdsourced Subjective Study

We conducted a large-scale crowdsourced study on Amazon Mechanical Turk (AMT) to collect human quality judgments for HDR user-generated content (HDR-UGC), adapting best practices from prior work on crowdsourced video quality assessment [7, 16, 34, 45]. To address the unique challenges of remote HDR evaluation, we implemented strict device qualification (10bit HDR Display, HDR playback capable browser, stable internet, etc.), persistent HDR capability checks, multi-stage quality control, and robust score aggregation.

B.4. Platform and Participants

Platform. AMT was used as the crowdsourcing platform.

Geography and compensation. We primarily recruited

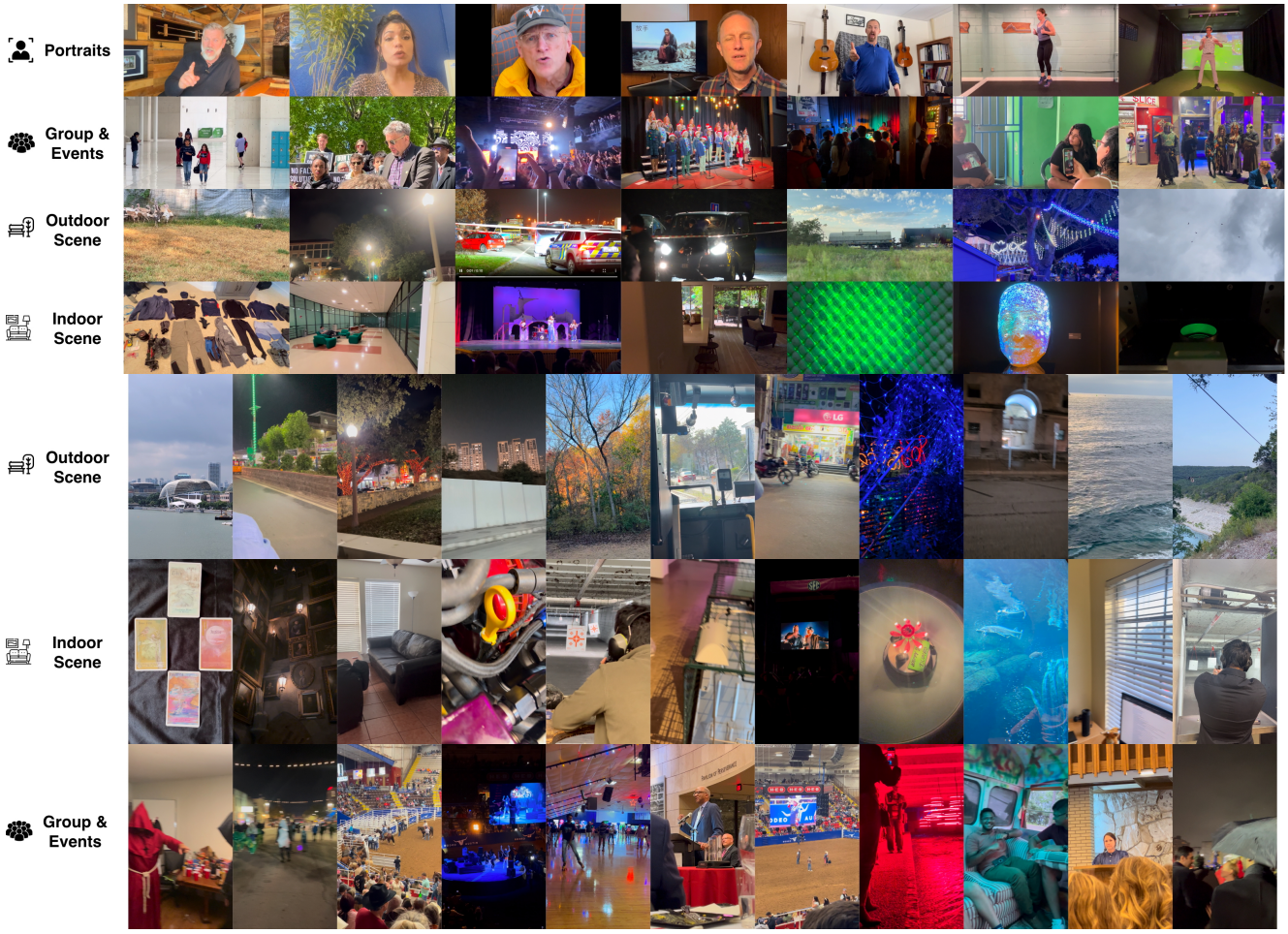


Figure 2. Overview of our video dataset illustrated through sampled frames.

workers from regions with higher HDR device penetration (e.g., North America and Europe). Compensation was set above typical local minimum wage for the estimated task duration.

Eligibility and qualification. Workers had to pass a multi-part qualification: (i) verification of HDR10-capable display, (ii) stable internet connection, (iii) English instruction comprehension, and (iv) a short training/quiz on the rating task (examples and multiple-choice questions), see Fig. 1. Only workers who passed all checks were admitted to the main study.

B.5. Stimuli and HIT Design

The main instructions of the study shown in Fig. 3. Each Human Intelligence Task (HIT) presented a batch of videos in a custom web interface capable of rendering HDR content leveraging browser support. Playback controls were limited to play/pause and replay; seeking was disabled to ensure that the entire clip was viewed prior to rating. Subjects used a likert-scale (0–100) to provide quality scores (rating instructions shown in Fig. 4).

B.6. HDR Capability Verification

Ensuring true HDR playback is critical for data validity. We implemented:

- **Pre-screening:** Client-side scripts probed display and browser capabilities indicative of HDR10 playback (e.g., bit depth, EOTF/codec support, and resolution proxies).
- **Persistent checks:** The same probes were re-run periodically during the HIT (e.g., at section boundaries) to detect display or window changes mid-task.

Workers whose devices failed initial or persistent checks were disqualified; ratings collected during failed sessions were discarded.

Training phase. Before the main test, subjects rated six HDR videos to familiarize themselves with the interface and the intended use of the likert-scale scale (Fig. 5). Feedback and brief reminders reinforced proper use of the range.

Testing phase. Each participant then rated 94 videos. To monitor consistency and calibration, we embedded five

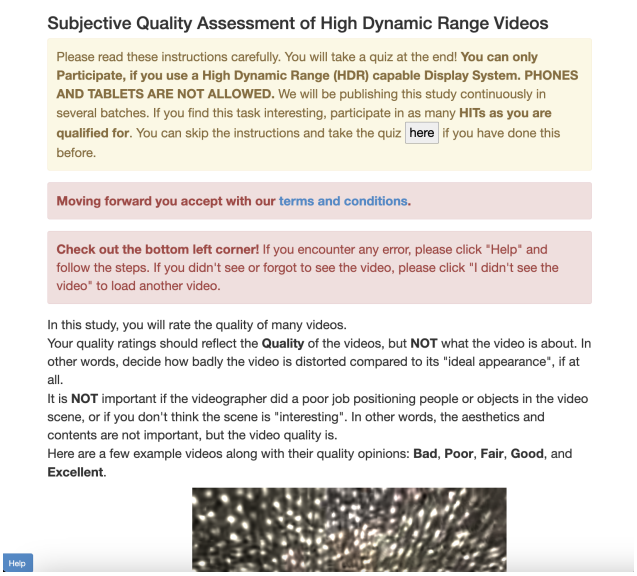


Figure 3. General instruction of this study on AMT

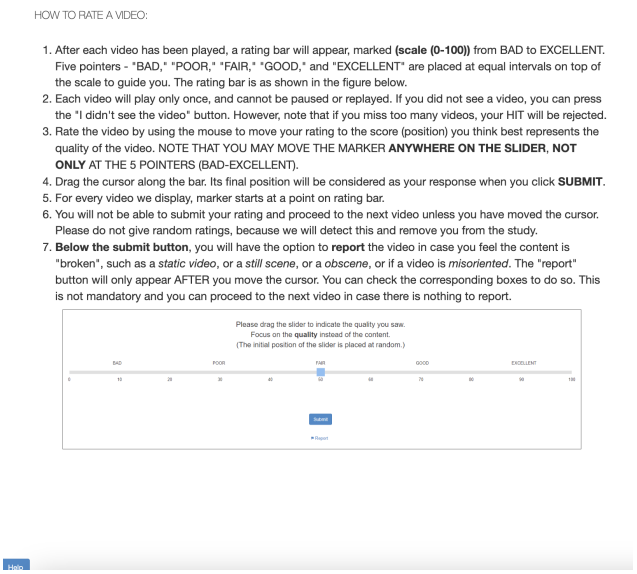


Figure 4. Rating instructions on AMT.

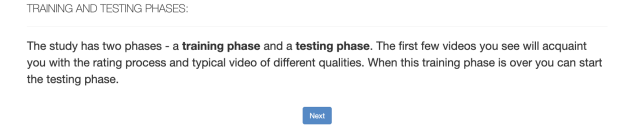


Figure 5. Train-test Instruction phase on AMT.

golden-set videos (with MOS obtained from lab/pilot studies) and five repeat (duplicate) videos within the test set.

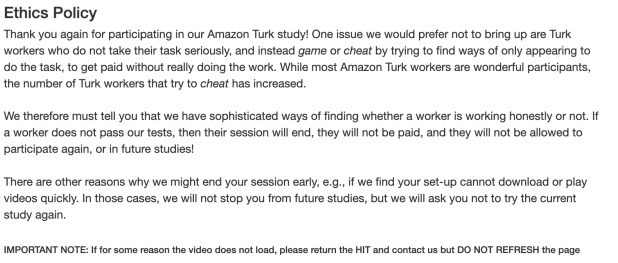


Figure 6. Ethics policy on AMT.

Across the full dataset, each video received on average $\sim$35 ratings.

B.7. Quality Control (QC)

We combined a priori design constraints with a posteriori checks:

- **Golden-set agreement:** Ratings deviating by more than two standard deviations from pilot MOS were flagged.
- **Repeat consistency:** For duplicated videos, absolute differences > 20 points indicated inconsistency.
- **Timing anomalies:** Exceptionally fast or slow completion times (relative to clip length and page dwell) were flagged.
- **Playback integrity:** We monitored download stalls, abnormal replay patterns, and other playback issues. Progress checkpoints at $\sim$25%, 50%, and 75% of the HIT facilitated mid-task intervention.

Rejection policy. Participants exhibiting multiple QC failures (e.g., repeated golden-set deviations, excessive duplicate inconsistencies, or $>$50% problematic playbacks) were disqualified and their ratings removed. Device-incompatible or non-HDR sessions were also rejected.

B.8. Ethics and Privacy

All workers provided informed consent within the AMT interface shown in Fig. 6. No personally identifiable information was stored beyond the minimum necessary for task operation and payment; HDR capability logs were retained only as anonymous technical flags for QC. The study design adhered to crowdsourcing ethics.

B.9. Reliability Analysis

We assessed internal consistency and subject reliability using:

- **Intra-subject repeatability:** We calculated the correlation between each subject's ratings on duplicate pairs and absolute repeat error.
- **Inter-subject Correlation:** We randomly Split all MOS ratings into two independent groups and computed the Spearman Rank Correlation Coefficient (SRCC) and Pearson Linear Correlation Coefficient (PLCC) between

them. The study achieved a median SRCC of 0.90 and a median PLCC of 0.92, shown in Fig. 8

B.10. Summary of Scale and Outcomes

In total, we processed ∼44K encoded sequences, each receiving on average ∼35 crowd ratings. After BT.500-style screening and SUREAL aggregation, we obtained robust MOS for HDR videos with associated confidence intervals. This constitutes, to our knowledge, the first crowdsourced large-scale HDR-UGC subjective study, designed to capture real-world device diversity while ensuring statistical reliability.

C. Dataset Analysis

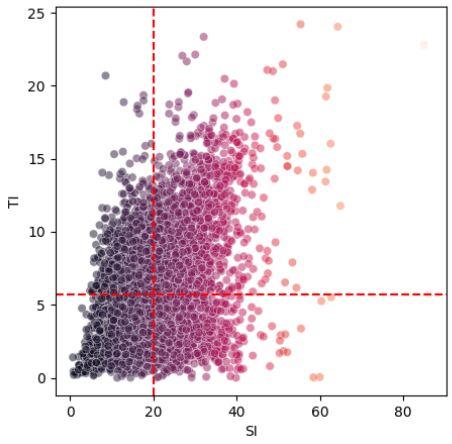
C.1. Content Analysis

SI-TI: To better characterize the diversity of content complexity, Fig. 7 provides an analysis of spatial-temporal complexity along with spatial information (SI) and temporal information (TI) in the dataset. The scatter plots in Fig. 7 (a)–(c) highlight the variation in SI and TI values, revealing a broad range of texture and motion complexity. Higher SI values are associated with scenes containing rich textures and sharp edges, whereas higher TI values reflect sequences with fast motion or dynamic activity. The dataset encompasses both highly detailed static scenes and rapidly changing dynamic content, making it well-suited for assessing compression performance and HDR characteristics under diverse motion conditions.

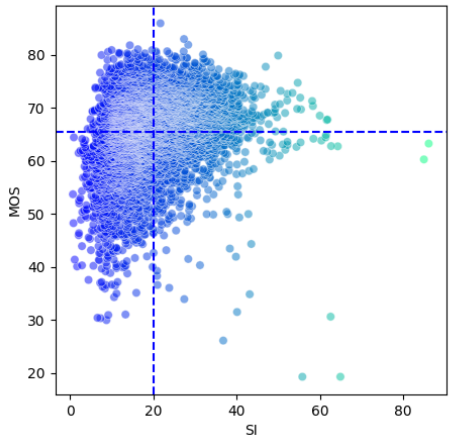
C.2. MOS Analysis

MOS Distribution: Fig. 9 presents the distribution and CDF of mean scores across all videos. Fig. 10(a) shows the MOS distributions across video orientations, where landscape videos achieve higher scores on average than portrait videos.Fig. 10(b) illustrates the score distributions for Vimeo and Crowd. The orange histogram and curve represent Vimeo, while the purple histogram and curve represent Crowd, with probability density estimates shown as smooth lines.

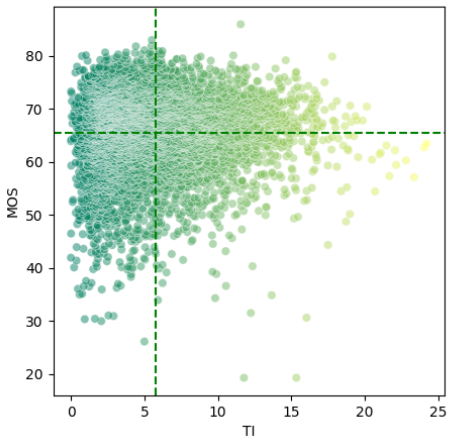
MOS vs. Bitrate ladders: Fig. 11 presents the relationship between MOS and encoding parameters, namely bitrate and resolution, for both landscape and portrait videos. As shown in Fig. 11(a) MOS consistently increases with higher bitrates, though the improvement plateaus at higher levels, with reference videos achieving the best quality. Similarly, Fig. 11(b) demonstrates that higher resolutions lead to higher MOS, with 1080p and reference videos rated the highest. Across both analyses, landscape videos generally obtain slightly higher MOS than portrait videos, particularly at lower bitrates and resolutions.



((a)) SI vs. TI.



((b)) MOS vs. SI.



((c)) MOS vs. TI.

Figure 7. (a) Spatial-Temporal Complexity, (b) MOS vs. Spatial Information (SI), and (c) MOS vs. Temporal Information (TI).

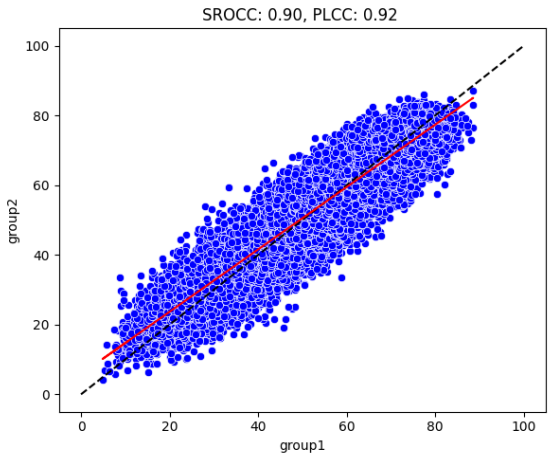
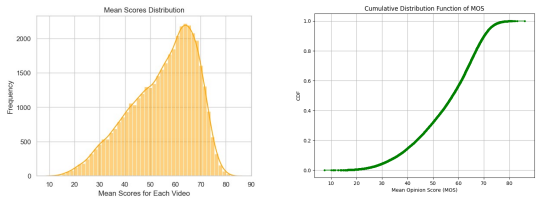


Figure 8. Inter-subject correlation.



((a)) MOS distribution. ((b)) MOS CDF distribution.

Figure 9. (a) MOS distribution of all videos in UGC-HDR-44K. (b) MOS distributions of Vimeo and Crowd videos in UGC-HDR-44K.

HDR UGC VQA: System Prompt Layout	
Role	**Expert in HDR video quality assessment, HDR perceptual evaluation. Evaluates full-video quality holistically.**
Inputs	`[Video Frame], [User Query / Context].`
Constraints	MOS must be a single integer (0–100).
Output Tags	`<think>` Continuous reasoning `</think>` then `<answer>mos</answer>`.

Table 2. **Prompt Structure** for HDR UGC VQA under the updated system prompt.

C.3. Dataset Limitations

Limited motion features and distortion: we observed that collected HDR videos didn't have extereme motions in them as observed manually and confirmed from fig. 7.

D. Method Details

D.1. HDR-Aware Policy Optimization (HAPO)

While GRPO [32] has proven effective for text-only reasoning tasks [49], it offers no guarantee that the policy leverages perception cues effectively [36], rather than shortcutting through generic textual signals. This makes our task even more challenging which require us to capture HDR cues from input signals. To address this gap, we propose HDR-Aware Policy Optimization (HAPO), a reinforcement learning framework that explicitly enforces HDR grounding, stabilizes entropy, and improves credit assignment for reasoning-heavy HDR-UGC VQA.

HDR–SDR Contrastive KL. A key challenge in multimodal RL is modality neglect [52], where the model ignores modality-specific features if they are not explicitly rewarded [26, 52]. In HDR-UGC VQA, this manifests as policies producing valid scores and rationales without exploiting HDR cues. Let $\pi_\theta^{\mathrm{HDR}}(\cdot) = \pi_\theta(\cdot \mid \text{text}, v, v^{SDR})$ denote the policy conditioned on HDR, SDR, and text,$\pi_\theta^{\mathrm{SDR}}(\cdot) = \pi_\theta(\cdot \mid \text{text}, v^{\mathrm{SDR}})$, the policy deprived of HDR video. In case of $\pi_\theta^{\mathrm{SDR}}$, we only give SDR video and text as input, and mask the HDR visual tokens in language decoder input space. Define:

$$D_{\mathrm{KL}}\big(\pi_\theta^{\mathrm{HDR}} \,\|\, \pi_\theta^{\mathrm{SDR}}\big) = \mathbb{E}_{\{o_i\}\sim\pi_{\theta_{\mathrm{old}}}(\cdot|\text{text},v,v^{\mathrm{SDR}})} \left[\frac{1}{K}\sum_{i=1}^{K}\frac{1}{|o_i|}\sum_{t=1}^{|o_i|}\left(\frac{\pi_\theta^{\mathrm{HDR}}(o_{i,t})}{\pi_\theta^{\mathrm{SDR}}(o_{i,t})} - \log\frac{\pi_\theta^{\mathrm{HDR}}(o_{i,t})}{\pi_\theta^{\mathrm{SDR}}(o_{i,t})} - 1\right)\right]. \tag{1}$$

Maximizing $\mathcal{K}_{\mathrm{HDR}}$ ensures that removing HDR tokens significantly perturbs the decoding distribution, thereby incentivizing the model to exploit HDR-specific information rather than collapsing into SDR-only reasoning.

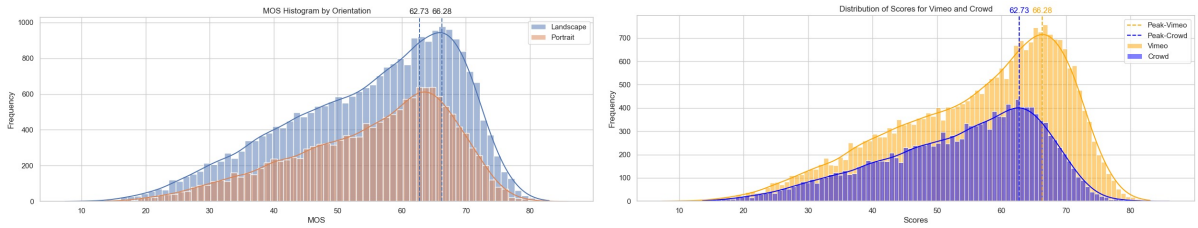
D.2. Rewards

Format Reward (R_{fmt}). This reward encourages the model to generate reasoning outputs structured with the designated special token, i.e. using `<think>`..Reasoning..`</think>` and `<answer>`..Final Answer..`</answer>` tags, and it should be correctly follow a JSON format.

$$R_{\mathrm{fmt}} = \begin{cases} 1, & \text{if ANSWER adheres to the required format,} \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

This prevents degenerate completions and guarantees that the policy remains aligned to the expected response template.

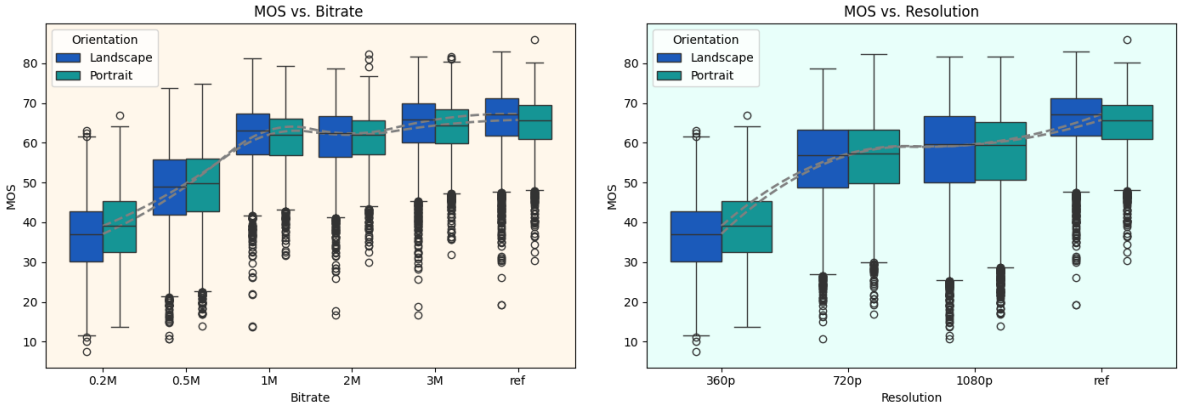
Score Reward (R_{sc}). Majority of existing reward formulations for image and video quality assessment use binary



((a)) MOS Distribution for Portrait and Landscape.

((b)) MOS Distribution for Vimeo and Crowd.

Figure 10. (a) MOS distribution of all videos in *UGC-HDR-44K*. (b) MOS distributions of Vimeo and Crowd videos in *UGC-HDR-44K*.



((a)) MOS vs. Bitrate.

((b)) MOS vs. Resolution.

Figure 11. (a) The MOS variations across bitrate. (2) The MOS variations across resolution.

rewards [15, 44], assigning the same credit to all predictions within a threshold, ignoring their relative proximity to the ground truth. While some other uses linear L1-based rewards [48], which improve upon binary reward but treat large and small errors with equal slope, offering little guidance once the model predictions are already close to accurate. All these methods provide only weak or coarse training signals, which leads to diminishing learning signals precisely in the regime where fine-grained calibration is most critical. Given perceptual quality assessment is inherently a regression problem, a more fitting reward would be that adjusts weight based on closeness to ground truth. Inspired from this, We introduce a gaussian weighted regression reward. The reward grows sharply as predictions approach the ground-truth MOS and gradually saturates near it, providing high-resolution feedback in the fine-grained regime while avoiding sensitivity to distant outliers. Formally, for predicted score $\hat{s}_i$ and ground truth s_*:

$$R_{\text{sc}}(\hat{s}_i, s_*) = \alpha \cdot \exp\left(-\frac{(\hat{s}_i - s_*)^2}{2\sigma^2}\right), \tag{3}$$

where σ controls the tolerance to deviations and $\alpha \in (0, 1]$ scales the overall magnitude. This formulation naturally emphasizes precision close to the target while still allowing coarse signals for distant errors, leading to more accurate and stable MOS predictions.

Self-Reward (R_{self}). Exploits within-group consensus. Given a group $\{o_i\}_{i=1}^{K}$, the majority answer $o^{\star}_{\text{maj}}$ is identified:

$$o^{\star}_{\text{maj}} = \arg \max_{o \in \{o_1, \ldots, o_K\}} \sum_{i=1}^{K} \mathbb{I}[o_i = o], \tag{4}$$

and each response is rewarded as

$$R_{\text{self}}(o_i) = \mathbb{I}[o_i = o^{\star}_{\text{maj}}]. \tag{5}$$

This stabilizes learning when external verifiers are noisy and prevents vanishing-advantage issues. Finally, each completion $o_i = \langle \hat{r}, \hat{s} \rangle$ receives a total reward:

$$\mathcal{R}_i = w_{\text{fmt}} R_{\text{fmt}} + w_{\text{sc}} R_{\text{sc}} + w_{\text{self}} R_{\text{self}}, \tag{6}$$

Together, these rewards jointly enforce structural validity, fine-grained MOS accuracy, HDR-aware attribute calibration, interpretable reasoning, and stable consensus-driven optimization.

D.3. Two-Stage Training Pipeline

Our training follows a two-stage RL-based paradigm [6, 8], both optimized with the same objective but serving distinct purposes.

Figure 12. Visual Reasoning Sample.

Stage 1: Modality Alignment. Since HDR tokens and their projection layer are initialized from scratch, we first align the HDR-aware vision encoder, projection module, and the LLM decoder. Instead of conventional supervised fine-tuning (SFT), we employ the same HAPO-based training used in Stage 2. This ensures that from the outset the policy learns to integrate HDR tokens into the language input space while producing structured reasoning outputs. Such modality alignment stages are common in multimodal RL [36, 49], where early grounding improves subsequent optimization.

Stage 2: Full RFT. After alignment, we continue training on the full HDR-UGC corpus using the complete HAPO objective. This balances training across samples of varying distortion severity, improving MOS prediction, and HDR-aware reasoning quality. Unlike the conventional SFT followed by RL pipeline [6, 26], both our stages use RL training, guaranteeing HDR grounding throughout.

HDR-Q provides a unified framework for HDR-UGC quality assessment. First, the HDR-aware vision encoder pro-

duces tokens that are sensitive to extreme contrast, peak highlights, near-black detail, and wide color gamut properties unique to HDR content. Second, maximizing the HDR–SDR contrastive KL divergence forces the language decoding distribution to shift when HDR cues are removed, thereby increasing the conditional mutual information between outputs and HDR inputs. Third, policy entropy regularization suppresses trivial entropy inflation and stabilizes optimization. Fourth, high-entropy weighting (HEW) allocates stronger learning signals to explorative tokens, which are most critical for accurate artifact identification and quality calibration, addressing the credit assignment problem in reasoning. Finally, self-rewarding consolidates group consensus, reinforcing consistent rationales and calibrated MOS predictions across sampled responses, leading to stable and interpretable HDR-aware quality assessment.

D.4. Ethical Considerations

- **Dataset Collection:** Videos collected through crowdsourcing involved explicit user consent for research purposes. Videos sourced from Vimeo were filtered for appropriate Creative Commons licenses. Efforts were made to filter personally identifiable information, though UGC content inherently carries privacy risks. The dataset will be released under a license restricting non-research use.
- **Subjective Study:** Participants were informed about the study's purpose and duration. Compensation was designed to be fair, exceeding typical platform rates. Data collected was anonymized. The study protocol was reviewed for ethical considerations regarding participant effort and potential exposure to diverse, unmoderated UGC.
- **Model Bias:** The dataset, despite efforts for diversity, may reflect biases present in the source platforms or participant demographics. The trained MLLM may inherit or amplify these biases. Potential biases related to perceived demographics, content types, or specific artifacts should be acknowledged.
- **Potential Misuse:** VQA models could potentially be used for unintended purposes, such as automated censorship or unfair content moderation. The focus of this work is perceptual quality assessment for improving user experience and system optimization.

References

[1] Naima Aamir, Junaid Mir, Imran Fareed Nizami, Furqan Shaukat, and Muhammad Majid. Hdr-bvqm: High dynamic range blind video quality model. *Multimedia Tools and Applications*, 80:27701 – 27715, 2021. 1

[2] Apple Inc. Hls authoring specification for apple devices, 2024. Accessed: Feb. 2024. 2

[3] Shahrukh Athar, Thilan Costa, Kai Zeng, and Zhou Wang. Perceptual quality assessment of UHD-HDR-WCG videos. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 1740–1744. IEEE, 2019. 1

[4] Maryam Azimi et al. PU21: A novel perceptually uniform encoding for adapting existing quality metrics for HDR. In *2021 Picture Coding Symposium (PCS)*, pages 1–5. IEEE, 2021.

[5] V Baroncini, K Andersson, AK Ramasubramonian, and G Sullivan. Verification test report for HDR/WCG video coding using HEVC main 10 profile. In *Proc. JCTVC-X1018 24th JCT-VC Meeting*, pages 293–303, 2016. 1

[6] Hardy Chen, Haoqin Tu, Fali Wang, Hui Liu, Xianfeng Tang, Xinya Du, Yuyin Zhou, and Cihang Xie. Sft or rl? an early investigation into training r1-like reasoning large vision-language models. *arXiv preprint arXiv:2504.11468*, 2025. 7, 8

[7] Yu-Chih Chen, Avinab Saha, Alexandre Chapiro, Christian Häne, Jean-Charles Bazin, Bo Qiu, Stefano Zanetti, Ioannis Katsavounidis, and Alan C. Bovik. Subjective and objective quality assessment of rendered human avatar videos in virtual reality. *IEEE Transactions on Image Processing*, 33: 5740–5754, 2024. 2

[8] Wei Dai, Peilin Chen, Chanakya Ekbote, and Paul Pu Liang. Qoq-med: Building multimodal clinical foundation models with domain-aware grpo training. *arXiv preprint arXiv:2506.00711*, 2025. 7

[9] Joshua Peter Ebenezer, Zaixi Shang, Yongjun Wu, Hai Wei, Sriram Sethuraman, and Alan C Bovik. Chipqa: No-reference video quality prediction via space-time chips. *IEEE Transactions on Image Processing*, 30:8059–8074, 2021. 1

[10] Joshua P Ebenezer, Zaixi Shang, Yongjun Wu, Hai Wei, Sriram Sethuraman, and Alan C Bovik. Hdr-chipqa: No-reference quality assessment on high dynamic range videos. *Signal Processing: Image Communication*, 129:117191, 2024. 1

[11] Google Support. Recommended upload encoding settings, 2024. Accessed: Feb. 2024. 2

[12] Chenlong He, Qi Zheng, Ruoxi Zhu, Xiaoyang Zeng, Yibo Fan, and Zhengzhong Tu. Cover: A comprehensive video quality evaluator. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 5799–5809, 2024. 1

[13] Jari Korhonen. Two-level approach for no-reference consumer video quality assessment. *IEEE Trans. Image Process.*, 28(12):5923–5938, 2019. 1

[14] Dingquan Li, Tingting Jiang, and Ming Jiang. Quality assessment of in-the-wild videos. In *ACM Multimedia*, pages 2351–2359. ACM, 2019. 1

[15] Weiqi Li, Xuanyu Zhang, Shijie Zhao, Yabin Zhang, Junlin Li, Li Zhang, and Jian Zhang. Q-insight: Understanding image quality via visual reinforcement learning. *arXiv preprint arXiv:2503.22679*, 2025. 1, 7

[16] Yiting Lu, Xin Li, Yajing Pei, Kun Yuan, Qizhi Xie, Yunpeng Qu, Ming Sun, Chao Zhou, and Zhibo Chen. Kvq: Kwai video quality assessment for short-form videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25963–25973, 2024. 2

[17] Pavan C. Madhusudana, Neil Birkbeck, Yilin Wang, Balu Adsumilli, and Alan C. Bovik. Image quality assessment

using contrastive learning. *IEEE Trans. Image Process.*, 31: 4149–4161, 2022. 1

[18] RK Mantiuk and M Azimi. Pu21: A novel perceptually uniform encoding for adapting existing quality metrics for hdr. 2021. 1

[19] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*, 21 (12):4695–4708, 2012. 1

[20] Anish Mittal, Rajiv Soundararajan, and Alan C. Bovik. Making a "completely blind" image quality analyzer. *IEEE Signal Processing Letters*, 20(3):209–212, 2013.

[21] Anish Mittal, Michele A. Saad, and Alan C. Bovik. A completely blind video integrity oracle. *IEEE Trans. Image Process.*, 25(1):289–300, 2016. 1

[22] Manish Narwaria, Matthieu Perreira Da Silva, and Patrick Le Callet. Hdr-vqm: An objective quality measure for high dynamic range video. *Signal Processing: Image Communication*, 35:46–60, 2015. 1

[23] Mikko Nuutinen, Toni Virtanen, Mikko Vaahteranoksa, Tero Vuori, Pirkko Oittinen, and Jukka Häkkinen. CVD2014 - A database for evaluating no-reference video quality assessment algorithms. *IEEE Trans. Image Process.*, 25(7):3073–3086, 2016. 1

[24] Xiaofei Pan, Jiaqi Zhang, Shanshe Wang, Shiqi Wang, Yun Zhou, Wenhua Ding, and Yahui Yang. Hdr video quality assessment: Perceptual evaluation of compressed hdr video. *Journal of Visual Communication and Image Representation*, 57:76–83, 2018. 1

[25] Yanyun Pu, Kehan Li, Zeyi Huang, Zhijie Zhong, and Kaixiang Yang. Mvqa-68k: A multi-dimensional and causally-annotated dataset with quality interpretability for video assessment. *arXiv preprint arXiv:2509.11589*, 2025. 1

[26] Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. Direct preference optimization: Your language model is secretly a reward model. *Advances in neural information processing systems*, 36:53728–53741, 2023. 6, 8

[27] Martin Rerabek, Philippe Hanhart, Pavel Korshunov, and Touradj Ebrahimi. Subjective and objective evaluation of hdr video compression. In *9th International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM)*, 2015. 1

[28] Michele A. Saad, Alan C. Bovik, and Christophe Charrier. Blind prediction of natural video quality. *IEEE Trans. Image Process.*, 23(3):1352–1365, 2014. 1

[29] Shreshth Saini, Avinab Saha, and Alan C Bovik. Hidro-vqa: High dynamic range oracle for video quality assessment. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 469–479, 2024. 1

[30] Kalpana Seshadrinathan, Rajiv Soundararajan, Alan Conrad Bovik, and Lawrence K. Cormack. Study of subjective and objective quality assessment of video. *IEEE Transactions on Image Processing*, 19(6):1427–1441, 2010. 1

[31] Zaixi Shang, Joshua P Ebenezer, Abhinau K Venkataramanan, Yongjun Wu, Hai Wei, Sriram Sethuraman, and Alan C Bovik. A study of subjective and objective quality assessment of hdr videos. *IEEE Transactions on Image Processing*, 33:42–57, 2023. 1

[32] Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, YK Li, Yang Wu, et al. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*, 2024. 6

[33] Zeina Sinno and Alan Conrad Bovik. Large-scale study of perceptual video quality. *IEEE Trans. Image Process.*, 28(2): 612–627, 2019. 1

[34] Abhinau K. Venkataramanan and Alan C. Bovik. Subjective quality assessment of compressed tone-mapped high dynamic range videos, 2024. 2

[35] Yilin Wang, Joong Gon Yim, Neil Birkbeck, and Balu Adsumilli. Youtube sfv+ hdr quality dataset. In *2024 IEEE International Conference on Image Processing (ICIP)*, pages 96–102. IEEE, 2024. 1

[36] Zhenhailong Wang, Xuehang Guo, Sofia Stoica, Haiyang Xu, Hongru Wang, Hyeonjeong Ha, Xiusi Chen, Yangyi Chen, Ming Yan, Fei Huang, et al. Perception-aware policy optimization for multimodal reasoning. *arXiv preprint arXiv:2507.06448*, 2025. 6, 8

[37] Haoning Wu, Chaofeng Chen, Jingwen Hou, Liang Liao, Annan Wang, Wenxiu Sun, Qiong Yan, and Weisi Lin. Fast-vqa: Efficient end-to-end video quality assessment with fragment sampling. In *Computer Vision – ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part VI*, page 538–554, Berlin, Heidelberg, 2022. Springer-Verlag. 1

[38] Haoning Wu, Chaofeng Chen, Liang Liao, Jingwen Hou, Wenxiu Sun, Qiong Yan, Jinwei Gu, and Weisi Lin. Neighbourhood representative sampling for efficient end-to-end video quality assessment, 2022.

[39] Haoning Wu, Liang Liao, Chaofeng Chen, Jingwen Hou, Annan Wang, Wenxiu Sun, Qiong Yan, and Weisi Lin. Disentangling aesthetic and technical effects for video quality assessment of user generated content. *CoRR*, abs/2211.04894, 2022. 1

[40] Haoning Wu, Erli Zhang, Liang Liao, Chaofeng Chen, Jingwen Hou, Annan Wang, Wenxiu Sun, Qiong Yan, and Weisi Lin. Towards explainable in-the-wild video quality assessment: A database and a language-prompted approach. In *Proceedings of the 31st ACM International Conference on Multimedia, MM 2023, Ottawa, ON, Canada, 29 October 2023- 3 November 2023*, pages 1045–1054. ACM, 2023. 1

[41] Haoning Wu, Zicheng Zhang, Erli Zhang, Chaofeng Chen, Liang Liao, Annan Wang, Chunyi Li, Wenxiu Sun, Qiong Yan, Guangtao Zhai, et al. Q-bench: A benchmark for general-purpose foundation models on low-level vision. *arXiv preprint arXiv:2309.14181*, 2023. 1

[42] Haoning Wu, Zicheng Zhang, Weixia Zhang, Chaofeng Chen, Liang Liao, Chunyi Li, Yixuan Gao, Annan Wang, Erli Zhang, Wenxiu Sun, et al. Q-align: Teaching lmms for visual scoring via discrete text-defined levels. *arXiv preprint arXiv:2312.17090*, 2023. 1

[43] Haoning Wu, Zicheng Zhang, Erli Zhang, Chaofeng Chen, Liang Liao, Annan Wang, Kaixin Xu, Chunyi Li, Jingwen

Hou, Guangtao Zhai, et al. Q-instruct: Improving low-level visual abilities for multi-modality foundation models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 25490–25500, 2024. 1

[44] Tianhe Wu, Jian Zou, Jie Liang, Lei Zhang, and Kede Ma. Visualquality-r1: Reasoning-induced image quality assessment via reinforcement learning to rank. *arXiv preprint arXiv:2505.14460*, 2025. 1, 7

[45] Zhenqiang Ying, Maniratnam Mandal, Deepti Ghadiyaram, and Alan Bovik. Patch-vq:'patching up'the video quality problem. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14019–14029, 2021. 1, 2

[46] Zhiyuan You, Jinjin Gu, Zheyuan Li, Xin Cai, Kaiwen Zhu, Chao Dong, and Tianfan Xue. Descriptive image quality assessment in the wild. *arXiv preprint arXiv:2405.18842*, 2024. 1

[47] Zhiyuan You, Zheyuan Li, Jinjin Gu, Zhenfei Yin, Tianfan Xue, and Chao Dong. Depicting beyond scores: Advancing image quality assessment through multi-modal language models. In *European Conference on Computer Vision*, pages 259–276. Springer, 2024. 1

[48] Zhiyuan You, Xin Cai, Jinjin Gu, Tianfan Xue, and Chao Dong. Teaching large language models to regress accurate image quality scores using score distribution. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 14483–14494, 2025. 1, 7

[49] Qiying Yu, Zheng Zhang, Ruofei Zhu, Yufeng Yuan, Xiaochen Zuo, Yu Yue, Weinan Dai, Tiantian Fan, Gaohong Liu, Lingjun Liu, et al. Dapo: An open-source llm reinforcement learning system at scale. *arXiv preprint arXiv:2503.14476*, 2025. 6, 8

[50] Zicheng Zhang, Wei Wu, Wei Sun, Dangyang Tu, Wei Lu, Xiongkuo Min, Ying Chen, and Guangtao Zhai. Md-vqa: Multi-dimensional quality assessment for ugc live videos, 2023. 1

[51] Zicheng Zhang, Ziheng Jia, Haoning Wu, Chunyi Li, Zijian Chen, Yingjie Zhou, Wei Sun, Xiaohong Liu, Xiongkuo Min, Weisi Lin, et al. Q-bench-video: Benchmark the video quality understanding of lmms. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 3229–3239, 2025. 1

[52] Xu Zheng, Chenfei Liao, Yuqian Fu, Kaiyu Lei, Yuanhuiyi Lyu, Lutao Jiang, Bin Ren, Jialei Chen, Jiawen Wang, Chengxin Li, et al. Mllms are deeply affected by modality bias. *arXiv preprint arXiv:2505.18657*, 2025. 6

[53] Hanwei Zhu, Haoning Wu, Yixuan Li, Zicheng Zhang, Baoliang Chen, Lingyu Zhu, Yuming Fang, Guangtao Zhai, Weisi Lin, and Shiqi Wang. Adaptive image quality assessment via teaching large multimodal model to compare. *Advances in Neural Information Processing Systems*, 37: 32611–32629, 2024. 1