HDR-Q: Seeing Beyond 8 Bits

Multimodal LLMs for HDR Video Quality Assessment with HDR-Aware Policy Optimization

Shreshth Saini

Laboratory for Image & Video Engineering (LIVE) · The University of Texas at Austin
Advised by Prof. Alan C. Bovik

CVPR 2026 · December 2025
Saini et al., Seeing Beyond 8bits, CVPR 2026 (arXiv:2603.00938)

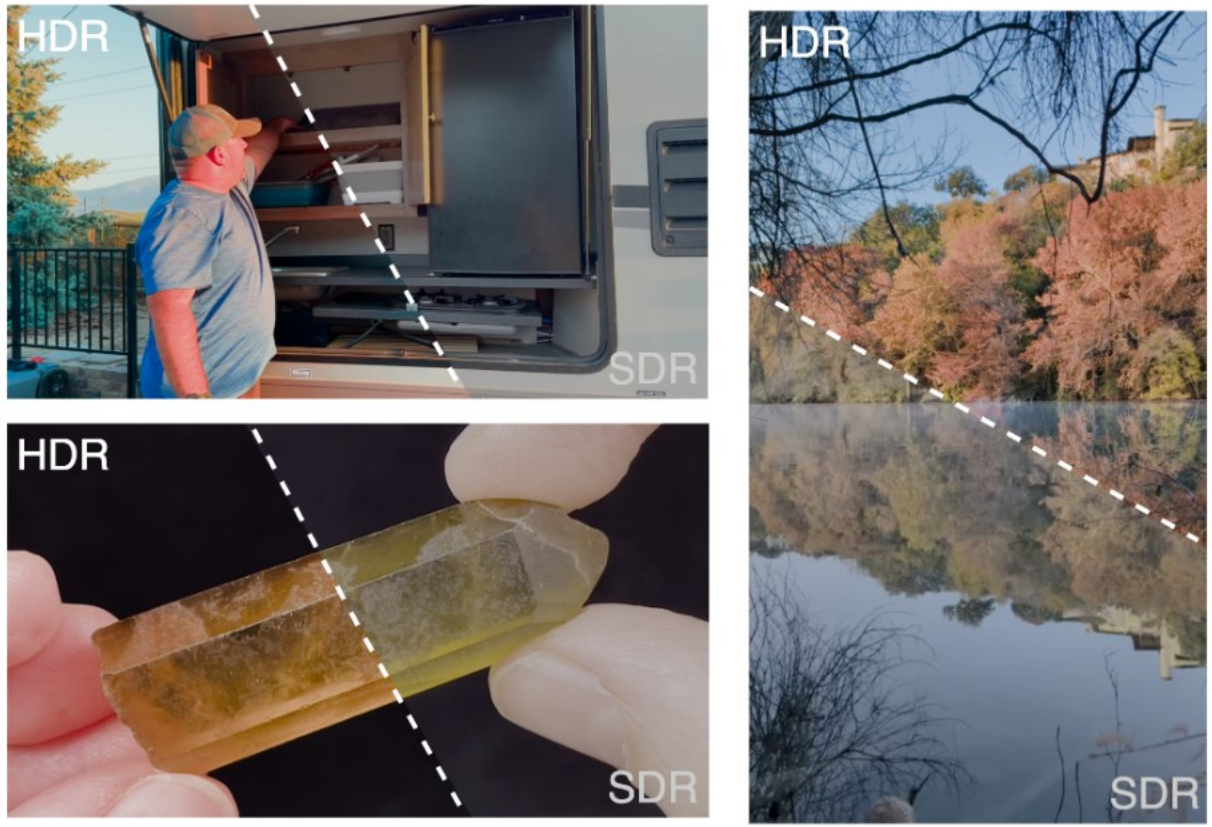
Can AI see and reason about HDR video quality?

Right now the answer is no. Every existing model was built for the 8-bit SDR world.

SDR encoders cannot see what HDR adds

Every modern phone captures 10-bit HDR by default.

- HDR10 is 10-bit, BT.2020 gamut, PQ (perceptual quantizer) transfer.
- Peak of 1000 nits or more, against about 100 for SDR.
- New artifacts such as highlight clipping, near-black banding and color blooming.
- SigLIP and CLIP read 8-bit sRGB, so they are structurally blind to these.



HDR keeps luminance and color detail that SDR collapses.

Where current methods break

Method family	HDR input	Perceptual grounding	Reasoning	Continuous MOS	Interpretable	SROCC
Classical VQA (BRISQUE, VMAF)	✗	✗	✗	✓	✗	0.41
Deep VQA (FastVQA, DOVER)	✗	partial	✗	✓	✗	0.51
HDR-specific (HIDRO-VQA) Saini et al., WACV 2024	✓	✓	✗	✓	✗	0.85
MLLM-VQA (Q-Insight, DeQA)	✗	✗	✓	partial	✓	0.52
HDR-Q (ours)	✓	✓	✓	✓	✓	**0.92**

VQA is video quality assessment, MLLM a multimodal LLM, MOS the mean opinion score, SROCC the Spearman rank correlation with MOS.

Only HDR-Q has HDR-aware perception, a continuous score and readable reasoning at once.

Three obstacles, and the third surprised us

O1. SDR-pretrained encoders

Blind to 10-bit PQ.

O2. A continuous score from tokens

The MOS (mean opinion score) must come out of text.

O3. Modality neglect

GRPO (Group Relative Policy Optimization) ignores HDR tokens.

0.891

SROCC with SDR input only

0.875

standard GRPO, HDR input added

Adding HDR made it worse.

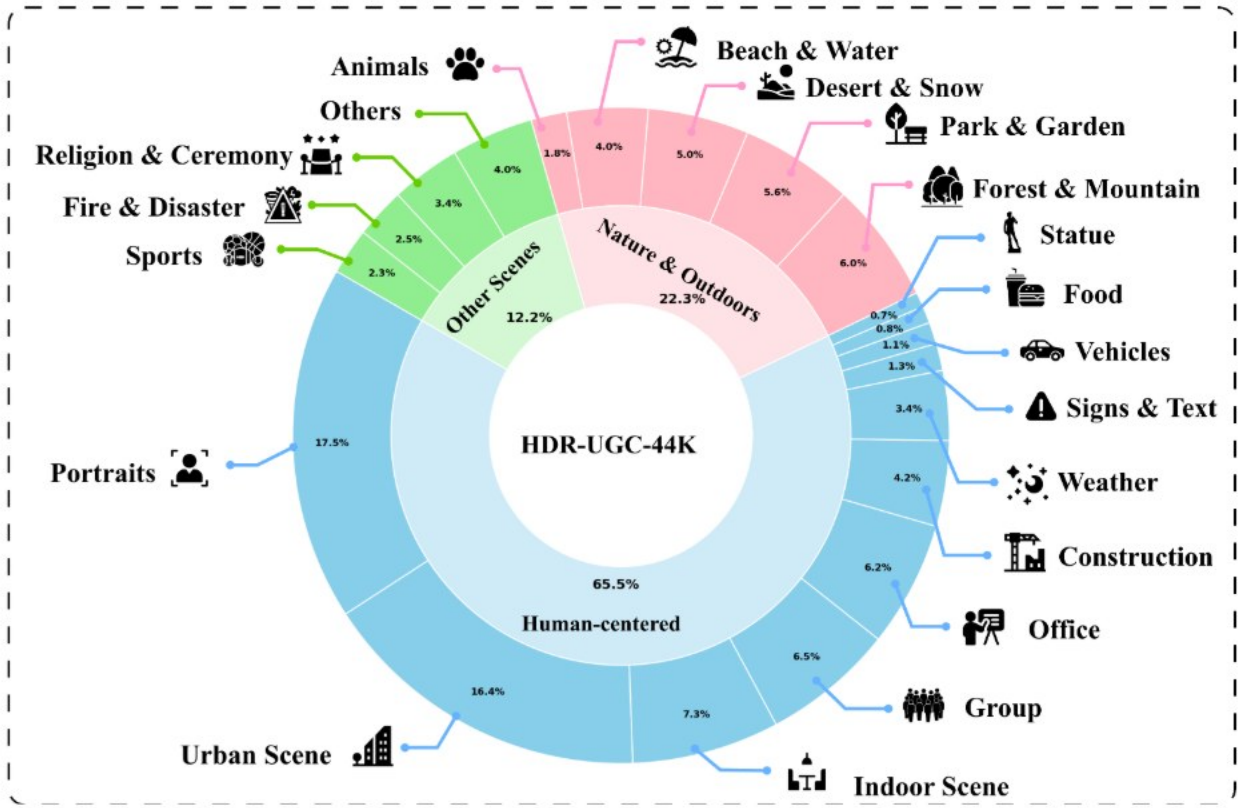
Beyond8Bits, the data HDR-Q trains on Saini et al., CVPR 2026

44,276
HDR-UGC video clips

6,861
unique source videos

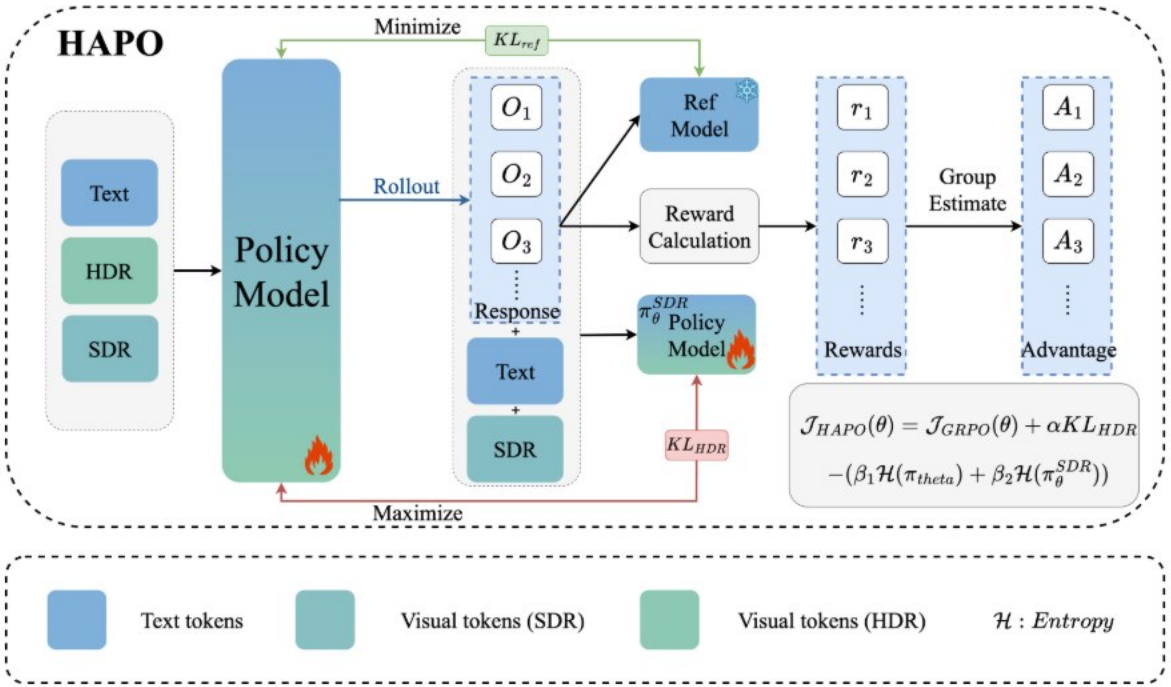
1.5M+
human quality ratings

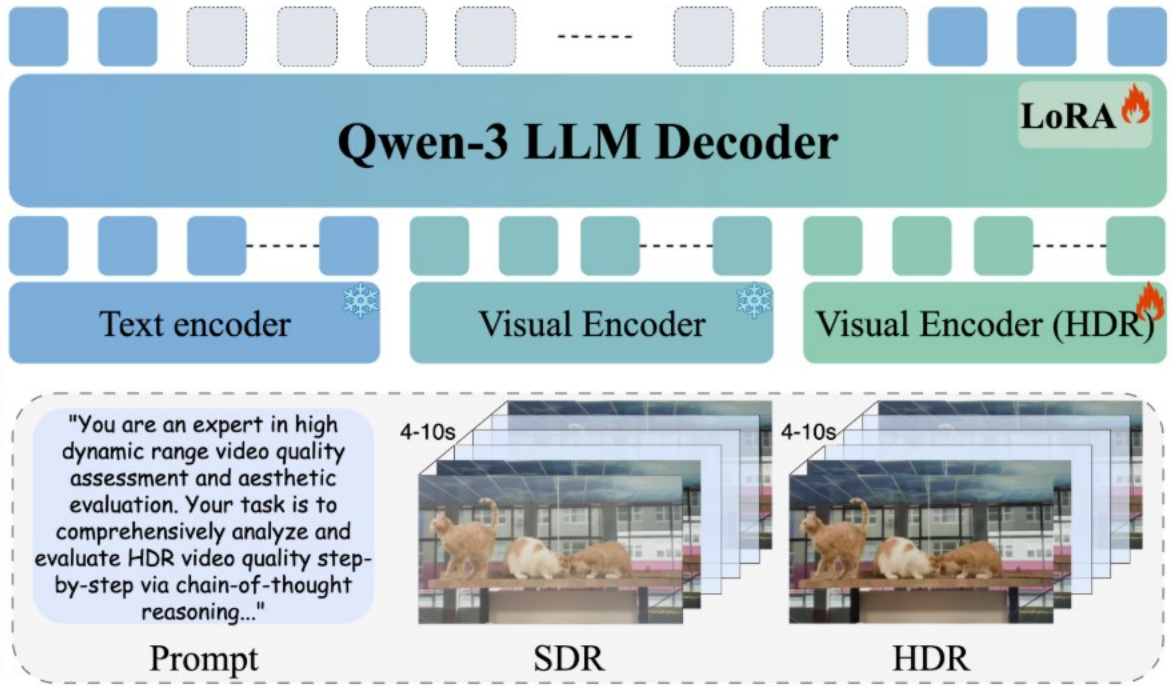
- The first large-scale HDR-UGC (user-generated content) quality dataset.
- About 35 ratings per video, inter-subject SROCC 0.90.
- 10-bit HEVC, PQ, BT.2020, 360p to 1080p, 0.2 to 5 Mbps.
- 2,253 crowdsourced videos plus 4,608 from Vimeo.



Scene categories across the 44K clips. The full figure is on a backup slide.

HDR-Q has one part per obstacle





HDR-aware encoder (O1)

SigLIP-2 on native 10-bit PQ, trained to tell HDR from SDR.

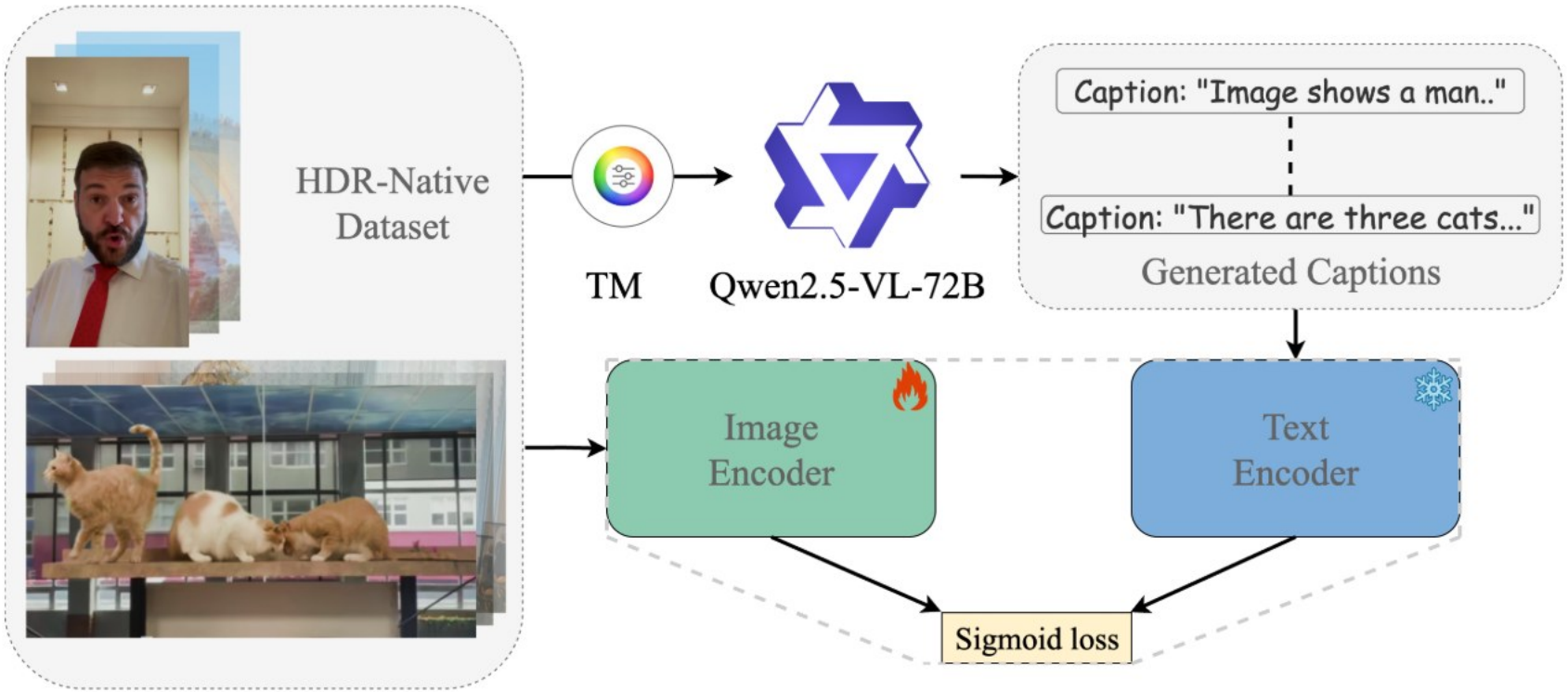
HAPO training (O3)

HDR-Aware Policy Optimization makes the policy use the HDR tokens.

Structured output (O2)

Ovis2.5 (9B), rank-4 LoRA. Reasoning in <think>, the MOS in <answer>.

The encoder separates HDR from SDR



SigLIP-2 finetuned on matched HDR-SDR pairs, with quality captions written by Qwen2.5-VL-72B.

$$L_{\text{ctr}} = \max(0, \delta - D(E(x_{\text{HDR}}), E(c)) + D(E(x_{\text{SDR}}), E(c)))$$

HDR moves toward the caption, SDR moves away.

$$L_{\text{enc}} = L_{\text{Sigmoid}} + \lambda_{\text{ctr}} \cdot L_{\text{ctr}}$$

Input stays native 10-bit PQ.

Adding HDR input made standard GRPO worse

0.8914

SROCC, SDR input only

0.8753

GRPO with HDR added

0.9206

HAPO with HDR added

Why

- GRPO treats every token alike.
- The model can shortcut through the text context.
- SDR features are familiar, HDR ones are foreign.

HAPO rewards a different output for HDR than for SDR.

HAPO pushes HDR and SDR outputs apart

$$K_{\mathrm{HDR}}(\theta) = D_{\mathrm{KL}}(\pi_\theta^{\mathrm{HDR}} \,\|\, \pi_\theta^{\mathrm{SDR}})$$

Ignore HDR and this KL is 0, so we **maximize** it.

$$I_{\theta^*}(\mathrm{output};\, \mathrm{HDR} \mid \mathrm{SDR}) \geq \gamma \cdot K_{\mathrm{HDR}}(\theta^*) - \kappa$$

Mutual information with HDR is bounded below.

HAPO adds three terms to GRPO

$$J_{\text{HAPO}}(\theta) = \mathbf{E}[\Sigma \min(\rho \cdot \tilde{A}, \text{clip}(\ldots) \cdot \tilde{A})] - \beta\, D_{\text{KL}}(\pi^{\text{HDR}} \,\|\, \pi_{\text{ref}}) + \gamma\, K_{\text{HDR}} - H_{\text{dual}}$$

Term	What it does	Setting
$\tilde{A}$	Entropy-weighted advantage (HEW), new	$\lambda_{\text{HEW}} = 0.3$
$\beta\, D_{\text{KL}}$	Reference KL from GRPO, keeps training stable	$\beta = 0.02$
$+\gamma\, K_{\text{HDR}}$	**Contrastive KL, new. Grounds the output in HDR**	$\gamma = 0.5$
$-H_{\text{dual}}$	Dual entropy, new. Prevents collapse	$\eta_1 = 0.01$, $\eta_2 = 0.05$

Training is two RL stages, modality alignment then score calibration, with no supervised fine-tuning (SFT).

0.921 SROCC on Beyond8Bits, and it holds zero-shot

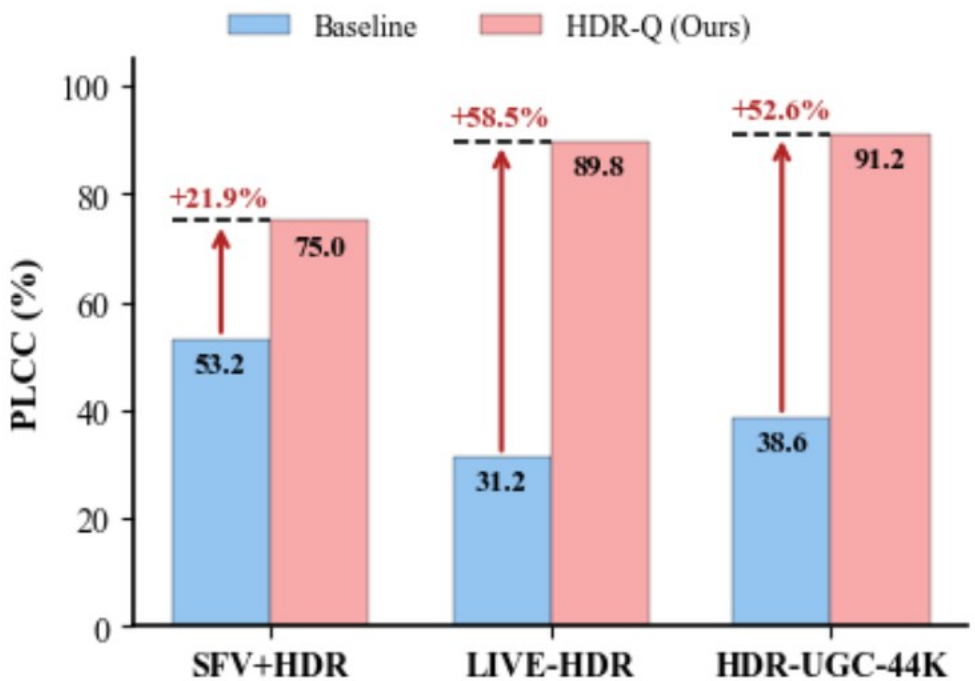
0.921
SROCC on Beyond8Bits

0.908
LIVE-HDR, zero-shot

0.725
SFV+HDR, zero-shot

Method	SROCC ↑	PLCC ↑	RMSE ↓
BRISQUE	0.410	0.469	11.70
CONTRIQUE	0.625	0.605	15.02
CONVIQT	0.799	0.810	8.48
HIDRO-VQA Saini et al., WACV 2024	<u>0.851</u>	<u>0.878</u>	<u>6.09</u>
Q-Insight (best MLLM)	0.517	0.562	20.78
HDR-Q (SDR only)	0.891	0.890	7.42
HDR-Q (full)	**0.921**	**0.912**	**5.16**

Beyond8Bits test set (Saini et al., CVPR 2026). Bold is best, underline is second best.



Zero-shot on LIVE-HDR and SFV+HDR, no fine-tuning on the target datasets.

The baseline invents artifacts, HDR-Q names real ones

Ovis 2.5

<think> The image shows an indoor scene ... Overall, the details are well preserved, the colors look natural. ... However, further checking is necessary. ... The brightness of the image appears uneven: the trees on the left seem clearly underexposed, while the right side appears overexposed due to strong light. The plant colors lean toward a grayish green, suggesting poor color reproduction. The purple flowers also show color blooming, making the hues look unnatural.Therefore, the final score is 85.</think>
<answer>85</answer>

HAPO (ours)

<think>The video shows a lush indoor botanical space... highlights in the mist roll off smoothly without harsh clipping, ... the hues remain stable without notable chroma shifts. Some fine detail is softened in the mistier regions, but this aligns with the natural ambience. Motion through the foliage likely feels gentle, with no obvious signs of jitter or smear. Overall, the scene benefits from balanced luminance, strong color reproduction, and clean gradients, giving the video a polished HDR appearance.</think>
<answer>{"82"}</answer>

Same video for both. HAPO (ours) is HDR-Q. Prompt and frame are on a backup slide.

Both get the same prompt, rate the video 0 to 100 and reason in <think> first.

Ovis2.5 baseline

Reports underexposure, color blooming and unnatural hues that are not in the video.

HDR-Q

Highlights roll off without clipping, hues stay stable, no jitter or smear.

Contrastive KL is the component that matters most

Variant	SROCC	RMSE	Token entropy
GRPO baseline	0.81	10.73	0.20
+ HDR encoder	0.83	8.96	0.24
HAPO without contrastive KL	0.86	7.10	0.29
HAPO without HEW	0.88	6.11	0.27
HAPO without dual entropy	0.91	5.82	0.26
HDR-Q (full)	**0.92**	**5.15**	**0.33**

Each HAPO row removes one component from the full model.

- Without contrastive KL, 0.92 falls to 0.86, the largest drop.
- Without HEW, 0.88. Without dual entropy, 0.91.
- The HDR encoder lifts GRPO from 0.81 to 0.83.

Token entropy rises from 0.20 to 0.33.

What we learned, what is still open

Lessons

- **Modality neglect** is the bottleneck, more than encoder capacity.
- **Collapsing token entropy** signals text shortcuts.
- **Two RL stages** beat SFT then RL.

Open problems

- Only T **= 8 frames**, so slow temporal patterns are missed.
- Little **synthetic or gaming HDR** in training.
- **Too slow** for real time.

Takeaways

1. First MLLM for HDR video quality, 0.92 SROCC against 0.52.
2. HAPO fixes modality neglect, with a mutual-information guarantee.
3. Zero-shot, 0.908 on LIVE-HDR, 0.725 on SFV+HDR.
4. The reasoning names real HDR artifacts.

Related work of mine

- BrightRate, HDR-UGC quality model. WACV 2026 (oral, best student paper)
- CHUG, crowdsourced HDR-UGC dataset. ICIP 2025
- HIDRO-VQA, contrastive HDR features. WACV 2024

Thank you.

TEXAS
The University of Texas at Austin

The University of Texas at Austin
LIVE
Laboratory for Image & Video Engineering

YouTube