

BrightRate: Quality Assessment for User-Generated HDR Videos

Supplementary Material

Anonymous WACV **Algorithms Track** submission

Paper ID 283

001 Supplementary Material Outline

- 002 This supplementary material is organized as follows:
- 003 • **Appendix A:** An overview of UGC and HDR video quality assessment challenges.
 - 004 • **Appendix B.1:** Comprehensive details on video collection, filtering, transcoding, and the bitrate ladder.
 - 005 • **Appendix B.2:** Full description of our AMT study, including instructions, screening procedure, and rejection criteria.
 - 006 • **Appendix B.3:** Analysis of inter-subject consistency and SUREAL-based MOS estimation.
 - 007 • **Appendix C:** Detailed examination of luminance, colorfulness, and spatial-temporal characteristics.
 - 008 • **Appendix D:** Additional technical specifics on resizing, normalization, and training.
 - 009 • **Appendix E:** Extended results, ablation studies, and failure case analyses.

018 A. Background

019 The explosion of UGC on platforms such as YouTube, Facebook, Instagram, and TikTok has transformed video streaming into a ubiquitous, user-driven experience, generating billions of daily views [1, 12, 13]. However, diverse distortion patterns arising from user editing, compression, and platform-specific processing complicate quality assessment [10, 18]. High Dynamic Range (HDR) imaging, supported by mainstream platforms and devices, offers enhanced visual experiences through a broader luminance and color range. Unlike Standard Dynamic Range (SDR) videos, which are limited to 0.1 to 100 cd/m^2 , HDR can represent luminance from 10^{-4} to 10^4 cd/m^2 [8]. HDR10, a widely adopted format, supports 10-bit color depth and Rec. 2020 color gamut (covering 75.8% of the CIE 1931 color space), providing higher peak luminance, improved color accuracy, and more detail in both shadows and highlights, offering a richer, more immersive viewing experience than SDR. The transition to HDR for UGC poses challenges for Video Quality Assessment (VQA) due to in-

creased bit depth, broader luminance range, and complex electro-optical transfer functions (EOTFs) like SMPTE ST 2084 [16]. Traditional SDR-based models fail to capture these HDR-specific features and the variability of distortions from different devices and editing techniques, thereby impeding effective quality prediction. Furthermore, the absence of a publicly available HDR-UGC database has limited the development of HDR-specific VQA models.

046 B. Details of Dataset Construction

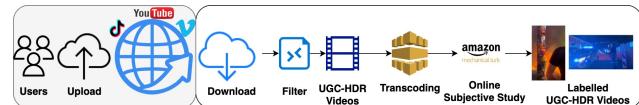


Figure 1. Overview of the dataset preparation approach.

Fig. 1 provides an overview of the entire dataset preparation pipeline. This multi-stage process guarantees that *BrightVQ* reflects authentic HDR-UGC content with diverse distortions

051 B.1. Video data Collection

Table 1. Bitladder used for dataset creation. Each video was encoded at multiple bitrates to simulate real-world streaming conditions¹.

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

HDR-UGC videos were collected from Vimeo under Creative Commons licenses to ensure open-source accessibility. An initial pool of over 10,000 videos was automat-

¹Based on YouTube's streaming guidelines [6] and Apple's HLS authoring specifications [2].



Figure 2. HDR specific challenges, and transcoding (on top of ugc) and ugc challenges in *BrightVQ*.

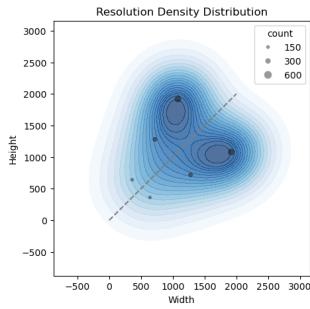


Figure 3. Resolution distribution of *BrightVQ* dataset, maintaining a balanced mix of landscape and portrait videos to study orientation-based perceptual differences.

055 ically filtered using metadata checks for HDR flags, resolution,
056 format consistency, and common categories to remove
057 duplicates and professionally produced content. This was
058 followed by a rigorous manual inspection to verify authen-
059 tic UGC. Given the nature of UGC, the dataset includes an
060 equal mix of landscape and portrait videos. Fig. 4 shows
061 randomly selected frames from *BrightVQ*, illustrating the
062 diversity in video sizes and aspect ratios. This diversity
063 highlights the broad representation of UGC content in terms
064 of resolution, aspect ratios, and distortions.

065 Each selected video was truncated to a maximum of 10
066 seconds using `ffmpeg` [5] and maintained at up to 1080p
067 resolution. To simulate the viewing experience on social
068 media platforms, where videos are often transcoded, we ap-
069 plied a bitrate ladder inspired by industry standards [2, 6]
070 to create the final dataset. Tab. 1 shows the resolution and
071 bitrates used in this bit ladder. The filtered videos were
072 then transcoded following this bitrate ladder to simulate
073 real-world streaming conditions, ensuring a diverse range of
074 compression artifacts. To explore the impact of bitrate se-
075 lection on perceived video quality, Fig. 6 presents the MOS
076 variations across different bitrate ladders, separately ana-

lyzing landscape and portrait videos. The box plot repre-
077 sentation highlights the diversity in perceptual quality rat-
078 ings across different encoding configurations, showing how
079 bitrate and resolution choices affect MOS scores. Fig. 3 il-
080 lustrates the resolution density distribution of videos in the
081 *BrightVQ* dataset. Fig. 5 further visualizes the compres-
082 sion artifacts introduced through this approach. This multi-stage
083 process guarantees that *BrightVQ* reflects authentic HDR-
084 UGC content with diverse distortions.

B.2. Crowdsourced Subjective Study

We employed Amazon Mechanical Turk (AMT) to collect
087 human quality ratings for our HDR-UGC videos, adapting
088 protocols from previous studies [3, 17]. This is the first
089 large-scale HDR-UGC study conducted on AMT, address-
090 ing challenges associated with remote HDR evaluation. To
091 ensure data reliability, we implemented a rigorous multiple-
092 stage filtering process.

The general instruction of this study on AMT is illus-
093 trated in Fig. 7. Initially, subjects were presented with de-
094 tailed instructions and a comprehension quiz (Fig. 8) to con-
095 firm their understanding of the rating process. Only those
096 with HDR-capable displays, verified through automated dy-
097 namic checks for bit depth, codec support, and display res-
098 olution, were allowed to proceed. Before entering the main
099 study, subjects completed a training phase where they rated
100 six HDR videos to familiarize themselves with the interface
101 (Fig. 9). The testing phase followed, in which each partic-
102 ipant rated 94 videos using a 0–100 Likert scale (rating
103 method shown in Fig. 10). To ensure rating consistency, we
104 embedded five golden set videos and five duplicate videos
105 within the test set. Ethical considerations are provided in
106 Fig. 11.

To maintain data integrity, we implemented strict rejec-
107 tion criteria at multiple stages:

- **During Instructions:** Participants with incompatible de-
111 vices were disqualified.
- **During Training:** Continuous HDR and device checks
113 ensured that participants did not switch displays mid-task.
114 Those with incomplete downloads or playback manipula-
115 tions were excluded.
- **During Testing:** Participants were monitored at 25%,
117 50%, and 75% of task completion. Those exhibiting
118 over 50% playback issues or inconsistent ratings on du-
119 plicate/golden set videos (deviations exceeding 20–25%)
120 were removed.

In total, over 200 subjects provided 73,794 ratings (an
122 average of 35 ratings per video).

B.3. Subjective Score Processing

To evaluate inter-subject consistency, we randomly split all
125 MOS ratings into two independent groups and computed the
126 Spearman Rank Correlation Coefficient (SRCC) and Pear-
127

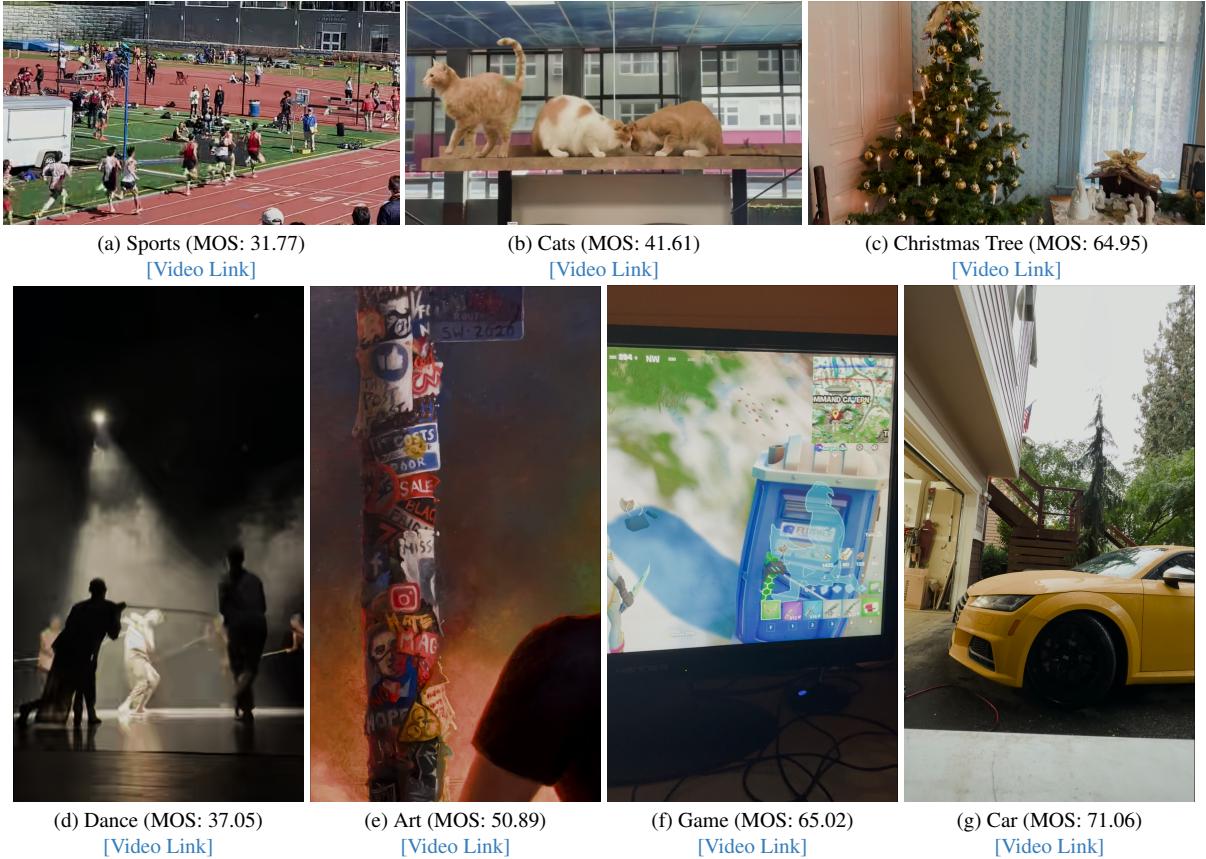


Figure 4. Example frames from *BrightVQ* dataset. Each frame is presented with its category, the MOS for the video and a direct video access link.

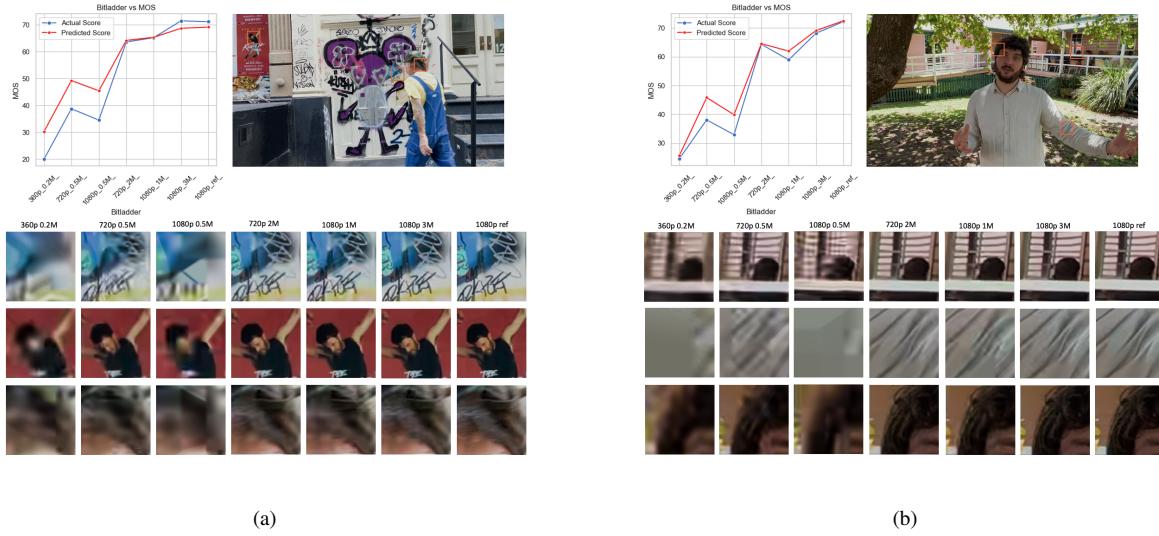


Figure 5. The combination of MOS vs. predicted score plots with visual comparisons of specific image regions to highlight the correlation between distortion and MOS across different bitrates and resolutions.

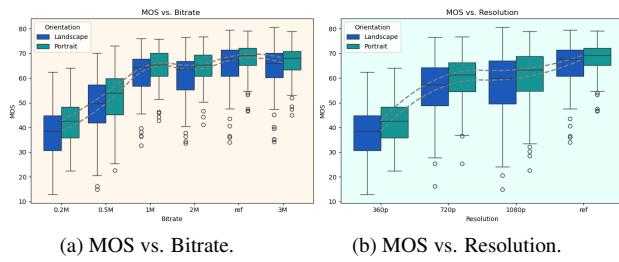


Figure 6. (a) and (b) show the MOS variations across bitrate and resolution respectively for *BrightVQ*.

Subjective Quality Assessment of High Dynamic Range Videos

Please read these instructions carefully. You will take a quiz at the end! You can only Participate, if you use a High Dynamic Range (HDR) capable Display System. PHONES AND TABLETS ARE NOT ALLOWED. We will be publishing this study continuously in several batches. If you find this task interesting, participate as many HITs as you are qualified for. You can skip the instructions and take the quiz [here](#) if you have done this before.

Moving forward you accept with our [terms and conditions](#).

Check out the bottom left corner! If you encounter any error, please click "Help" and follow the steps. If you didn't see or forgot to see the video, please click "I didn't see the video" to load another video.

In this study, you will rate the quality of many videos. Your quality ratings should reflect the **Quality** of the videos, but **NOT** what the video is about. In other words, decide how badly the video is distorted compared to its "ideal appearance", if at all. It is **NOT** important if the videographer did a poor job positioning people or objects in the video scene, or if you don't think the scene is "interesting". In other words, the aesthetics and contents are not important, but the video quality is.

Here are a few example videos along with their quality opinions: **Bad**, **Poor**, **Fair**, **Good**, and **Excellent**.



Figure 7. General instruction of this study on AMT.

QUIZ TIME!

The following quiz is to test your diligence and sincerity. Please choose the appropriate options:

Q1. Where can you find the rating slider?

- On the next page after the video has stopped playing
- Top of the video while it is playing
- Below a video while it is playing

Q2. How do you rate a video using the slider?

- Drag the cursor along the rating scale to the appropriate position
- Enter the rating value in the box below the scale
- Click on the five reference positions shown above the scale

Q3. You are evaluating each video based on its:

- Quality (how good the video looks)
- Content (what is in the video)
- Aesthetics (how good the video scene is framed)

Q4. Which of the following conditions are essential for this study? (Select all that apply)

- Connect your device to a speaker
- Close any other applications running on your device
- Close all other browser tabs
- Switch off the lights
- Set the browser zoom to 100%

Q5. When will the 'survey' appear in this study?

- Immediately after the quiz
- Halfway through the study
- At the end, after all the videos have been rated

Q6. What should you do if you normally wear corrective lens?

- Wear it during the study, as not using it might affect your perception of quality
- Don't wear it since we are measuring "naked" vision
- It does not matter for this study

Submit

Figure 8. Quiz phase on AMT.

TRAINING AND TESTING PHASES:

The study has two phases - a **training phase** and a **testing phase**. The first few videos you see will acquaint you with the rating process and typical video of different qualities. When this training phase is over you can start the testing phase.

Next

Figure 9. Train-test Instruction phase on AMT.

HOW TO RATE A VIDEO:

1. After each video has been played, a rating bar will appear, marked (**scale (0-100)**) from BAD to EXCELLENT. Five pointers - "BAD," "POOR," "GOOD," and "EXCELLENT" are placed at equal intervals on top of the scale to guide you. The rating bar is as shown in the figure below.
2. Each video will play only once, and cannot be paused or replayed. If you did not see a video, you can press the "I didn't see the video" button, note that if you miss too many videos, your HIT will be rejected.
3. Rate videos using the mouse to move your rating to the score (position) you think best represents the quality of the video. **NOTE THAT YOU MAY MOVE THE MARKER ANYWHERE ON THE SLIDER, NOT ONLY AT THE 5 POINTERS BAD-EXCELLENT!**
4. Drag the cursor along the bar. Its final position will be considered as your response when you click **SUBMIT**.
5. For every video we display, marker starts at a point on rating bar.
6. You will not be able to submit your rating and proceed to the next video unless you have moved the cursor. Please do not give random ratings, because we will detect this and remove you from the study.
7. Below the submit button, you will have the option to **report** the video in case you feel the content is "broken", such as a static video, or a still scene, or a *obscene*, or if a video is *misoriented*. The "report" button will only appear AFTER you move the cursor. You can check the corresponding boxes to do so. This is not mandatory and you can proceed to the next video in case there is nothing to report.



Figure 10. Rating instructions on AMT.

Ethics Policy

Thank you again for participating in our Amazon Turk study! One issue we would prefer not to bring up are Turk workers who do not take their task seriously, and instead game or *cheat* by trying to find ways of only appearing to do the task, to get paid without really doing the work. While most Amazon Turk workers are wonderful participants, the number of Turk workers that try to *cheat* has increased.

We therefore tell you that we have sophisticated ways of finding whether a worker is working honestly or not. If a worker does not pass our tests, then their session will end, they will not be paid, and they will not be allowed to participate again, or in future studies!

There are other reasons why we might end your session early, e.g., if we find your set-up cannot download or play videos quickly. In those cases, we will not stop you from future studies, but we will ask you not to try the current study again.

IMPORTANT NOTE: If for some reason the video does not load, please return the HIT and contact us but DO NOT REFRESH the page

Figure 11. Ethics policy on AMT.

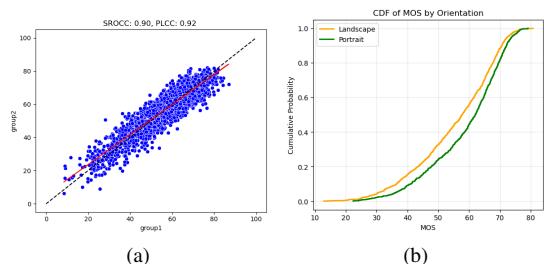


Figure 12. (a). Inter-subject correlation. (b). MOS CDF distribution of all videos in *BrightVQ*.

son Linear Correlation Coefficient (PLCC) between them. As shown in Fig. 12a, the study achieved a median SRCC of 0.90 and a median PLCC of 0.92, demonstrating a high level of agreement between independent participant groups. This strong correlation validates the effectiveness of our data collection methodology, which incorporated device filtering,

128

129

130

131

132

133

134 training phases, and golden set validation to ensure consistent and reliable subjective ratings.
 135

136 To avoid allowing participants with non-HDR displays,
 137 we pre-screen to ensure participants' HDR display capability.
 138 Client-side scripting was used to query display properties
 139 (such as bit-depth, color space, etc.) and verify necessary
 140 codec support in the user's browser (i.e., EDID and
 141 browser capability probing). We collected the HDR display
 142 details for each participant as metadata and in surveys, but
 143 due to privacy reasons, we can not release such sensitive in-
 144 formation. A key goal of BrightVQ was to create a dataset
 145 reflecting real-world UGC HDR viewing experiences, con-
 146 straining participants to a single or very limited range of
 147 "calibrated" HDR displays, as done in traditional lab studies,
 148 would not capture the diversity of consumer HDR de-
 149 vices. Our dataset therefore implicitly includes variations
 150 due to different HDR display capabilities. While this intro-
 151 duces variability, it is precisely this variability that methods
 152 aiming for broad applicability on UGC platforms need to be
 153 robust for.

154 We computed Mean Opinion Scores (MOS) using the
 155 SUREAL method [9], which refines traditional MOS com-
 156 putation by accounting for individual subject bias and rat-
 157 ing inconsistency. Traditional MOS calculations typically
 158 employ a hard rejection approach, where raters failing pre-
 159 defined consistency criteria (e.g., ITU-R BT.500-14 out-
 160 lier detection [7]) are completely excluded from the anal-
 161 ysis. However, this method discards potentially useful data
 162 and does not account for varying levels of rating reliability
 163 among retained subjects. SUREAL takes a soft rejection
 164 approach by modeling each rating probabilistically. Each
 165 rating S_{ij} from subject i for video j is modeled as:

$$166 \quad S_{ij} = \psi_j + \Delta_i + \nu_i X, \quad X \sim \mathcal{N}(0, 1), \quad (1)$$

167 where ψ_j represents the true quality of video j , Δ_i captures
 168 the bias of subject i , and ν_i reflects the rating inconsis-
 169 tency of subject i . The parameters are estimated using Maxi-
 170 mum Likelihood Estimation (MLE), maximizing the like-
 171 lihood. Unlike hard rejection, which entirely removes out-
 172 liers, SUREAL downweights ratings from less reliable sub-
 173 jects. This ensures that MOS values reflect true perceptual
 174 quality while mitigating distortions from inconsistent raters.
 175 By applying SUREAL, we obtained more stable MOS es-
 176 timates, which accurately reflect the perceptual quality of
 177 HDR content across diverse video conditions. The CDF dis-
 178 tribution of all videos in BrightVQ are shown in Fig. 12b.

179 C. Analysis of the HDR content

180 In this section, we provide a detailed analysis of the
 181 dataset's key characteristics, focusing on luminance and
 182 color distribution, spatial-temporal diversity, perceptual
 183 quality trends, and HDR-specific challenges.

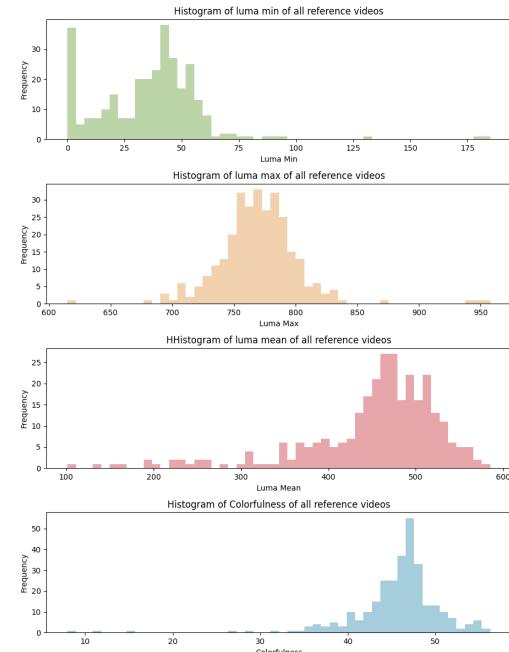


Figure 13. Distribution of luma and colorfulness of the source HDR-UGC videos in *BrightVQ* dataset.

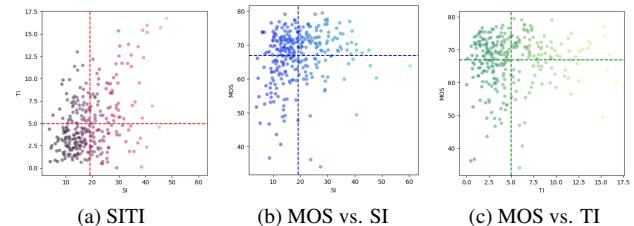


Figure 14. (a) Spatial-Temporal Complexity, (b) MOS vs. Spatial Information (SI), and (c) MOS vs. Temporal Information (TI).

184 Fig. 13 presents the distribution of luma and colorfulness
 185 across the 300 source HDR videos in BrightQA. The first
 186 three histograms illustrate the minimum, maximum, and
 187 mean luma values, highlighting the variation in brightness
 188 levels across different videos. This demonstrates that the
 189 dataset includes both dark and bright HDR scenes, ensur-
 190 ing a wide dynamic range. The fourth histogram shows the
 191 colorfulness distribution, reflecting variations in chromatic
 192 intensity across different videos.

193 To further quantify the diversity in content complex-
 194 ity, Fig. 14 presents an analysis of spatial-temporal com-
 195 plexity, spatial information (SI), and temporal informa-
 196 tion (TI) within the dataset. The scatter plots in Fig. 14 (a)-(c)
 197 demonstrate the variability in SI and TI values, showing a
 198 wide distribution of motion and texture complexity across
 199 the dataset. Higher SI values typically correspond to de-
 200 tailed textures and sharp edges, while higher TI values in-

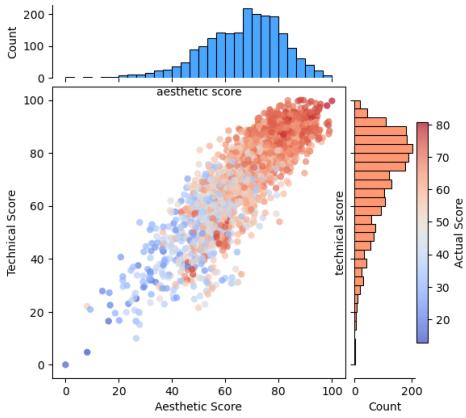


Figure 15. Diversity in aesthetic and technical quality scores in BrightVQ dataset.

201 dicate rapid motion or dynamic scenes. The dataset covers
 202 both high-detail static scenes and fast-moving dynamic content,
 203 ensuring its suitability for evaluating compression and
 204 HDR features across different motion characteristics.

205 Fig. 15 demonstrates the diversity of BrightVQ dataset
 206 in both aesthetic and technical aspects. The scatter plot
 207 shows a wide range of ratings, with each point represent-
 208 ing a video and color-coded by its actual subjective quality
 209 score. The marginal histograms further highlight the
 210 distribution of scores, illustrating the broad variation in per-
 211 ceptual and technical quality across different content. The
 212 BrightVQ dataset presents a diverse range of HDR-UGC
 213 content, covering natural landscapes, indoor scenes, and
 214 various complex lighting conditions, capturing both HDR-
 215 specific and UGC-specific distortions. This diversity en-
 216 sures that BrightVQ provides a comprehensive and realistic
 217 benchmark for evaluating video quality.

218 Figure 16 provides examples of HDR videos subjected
 219 to various spatial resolutions and bitrates, along with their
 220 MOS and luminance histograms. In (a) and (c), the
 221 higher-resolution, higher-bitrate frames retain more de-
 222 tails and exhibit fewer artifacts, whereas lower-resolution,
 223 lower-bitrate versions show noticeable blocking and band-
 224 ing—particularly in the extreme luma regions. Subfig-
 225 ures (b) and (d) illustrate the broader luminance distribu-
 226 tion characteristic of HDR, indicating significant content in
 227 both low and high intensity ranges. Such distributions un-
 228 derscore the importance of HDR-specific processing, since
 229 distortions in extreme luminance regions can disproportio-
 230 nately affect subjective quality.

231 D. More details on Implementation

232 Here we detail the key implementation steps of our Bright-
 233 Rate model. For semantic features, input frames are resized
 234 to 224×224 and passed through the CLIP image encoder

(ViT-B32) [14], yielding high-level semantic representations. UGC features are extracted using CONTRIQUE [11] at two scales: the original frame and a half-resolution version following the original implementation [11]. For HDR features, we convert each frame to YUV, extract the luminance channel $\mathbf{Y}^t \in [0, 1]$, and apply a piecewise expansive non-linearity over a 31×31 window with $\beta = 4$ [4, 15]. We then compute MSCN coefficients on the transformed luminance and model their statistics using GGD/AGGD to obtain HDR features \mathcal{H}^t . Temporal dynamics are captured by computing the absolute differences between consecutive CONTRIQUE [11] features:

$$\Delta \mathcal{U}^t = |\mathcal{U}^t - \mathcal{U}^{t-1}|, \quad (2)$$

which are then globally averaged and concatenated with the spatial features. The final clip-level representation is formed by normalizing and concatenating the UGC, semantic, and HDR features:

$$\mathbf{z} = \text{Norm}(\bar{\mathcal{U}} \oplus \bar{\mathcal{E}} \oplus \bar{\mathcal{H}}). \quad (3)$$

This vector is then fed to a Support Vector Regressor (SVR) to predict the MOS. We train the SVR using 5-fold cross-validation to optimize the regularization parameter and repeat the process over 100 random splits, reporting the median performance. The training loss is given by:

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^N \left(Q_i - \hat{Q}_i \right)^2, \quad (4)$$

where Q_i and \hat{Q}_i denote the ground-truth and predicted MOS, respectively. Only the regressor is trained, while the feature extraction modules remain fixed. These implementation details ensure a robust and efficient extraction of multi-scale UGC, semantic, and HDR features, enabling accurate quality prediction on HDR-UGC videos.

265 E. More Experimental Results

To assess the effectiveness of existing No-Reference Video Quality Assessment models on the BrightVQ dataset, we conducted a comprehensive evaluation of multiple state-of-the-art methods. Fig. 17 expands upon Fig. 9, which presented results for only six models, by providing a more extensive comparison across 13 NR-VQA model. The scatter plots compare predicted scores vs. MOS, with red parametric fitting lines highlighting the correlation trends, while the Pearson correlation coefficient (r) quantifies each model's predictive performance.

Among the evaluated models, some traditional hand-crafted feature-based approaches exhibit limited correlation with MOS, highlighting their challenges in capturing the complexity of HDR-specific distortions in UGC content.

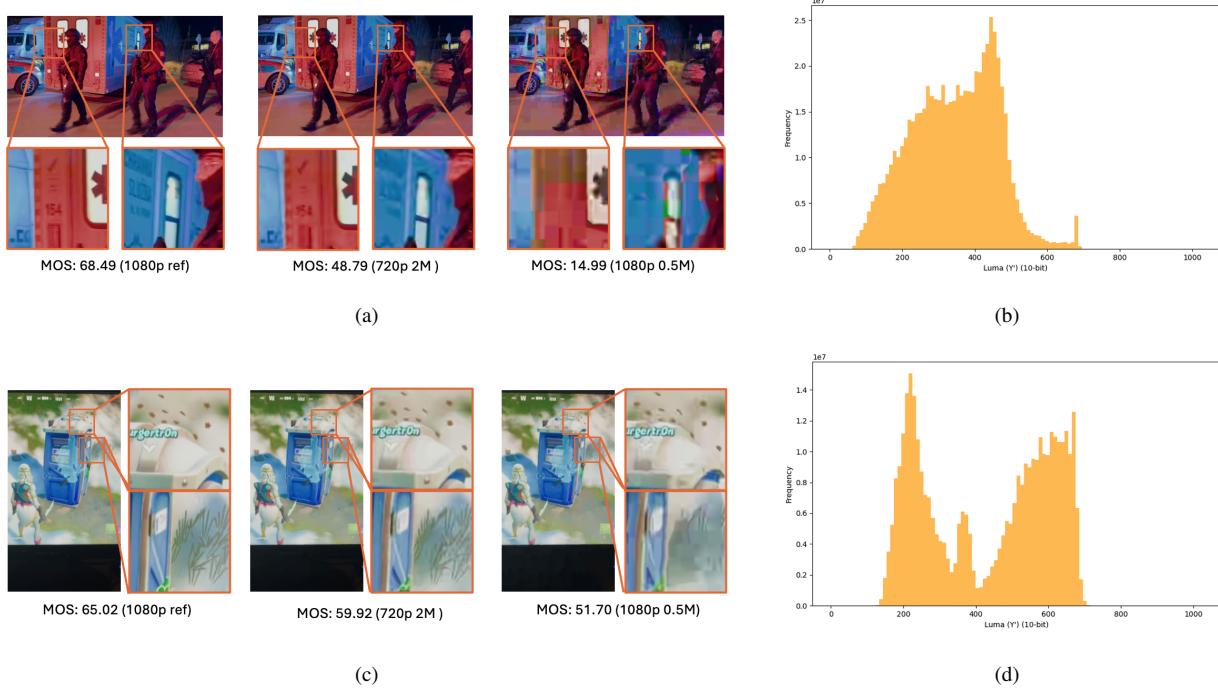


Figure 16. Illustrations of HDR content under different resolutions and bitrates. (a) and (c) show reference frames at 1080p and their lower-resolution, lower-bitrate counterparts, with red boxes highlighting high luma areas with artifacts (e.g., blocking, color banding) become more pronounced. The corresponding MOS values indicate how these distortions affect subjective perception. (b) and (d) present the luminance histograms of the respective frames, revealing a broader distribution for HDR content that spans both low and high luminance ranges. This demonstrates the increased complexity of HDR videos.

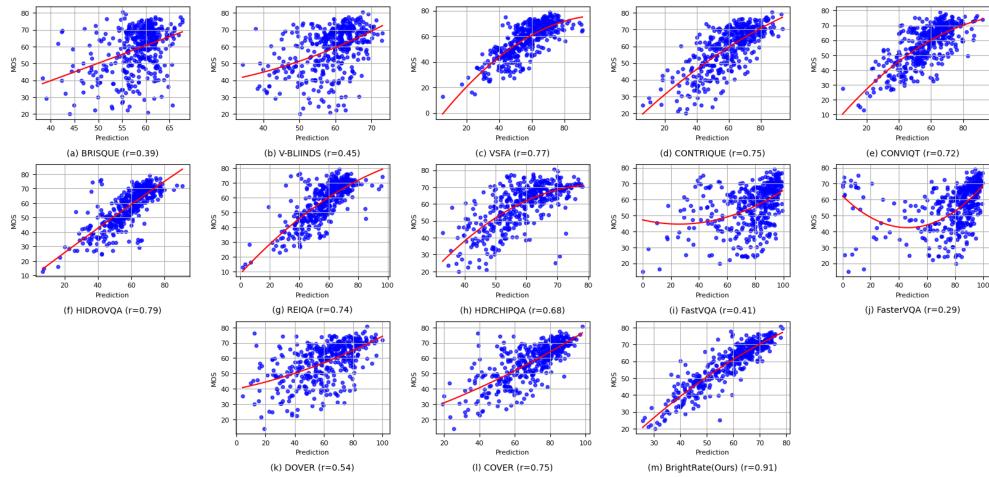
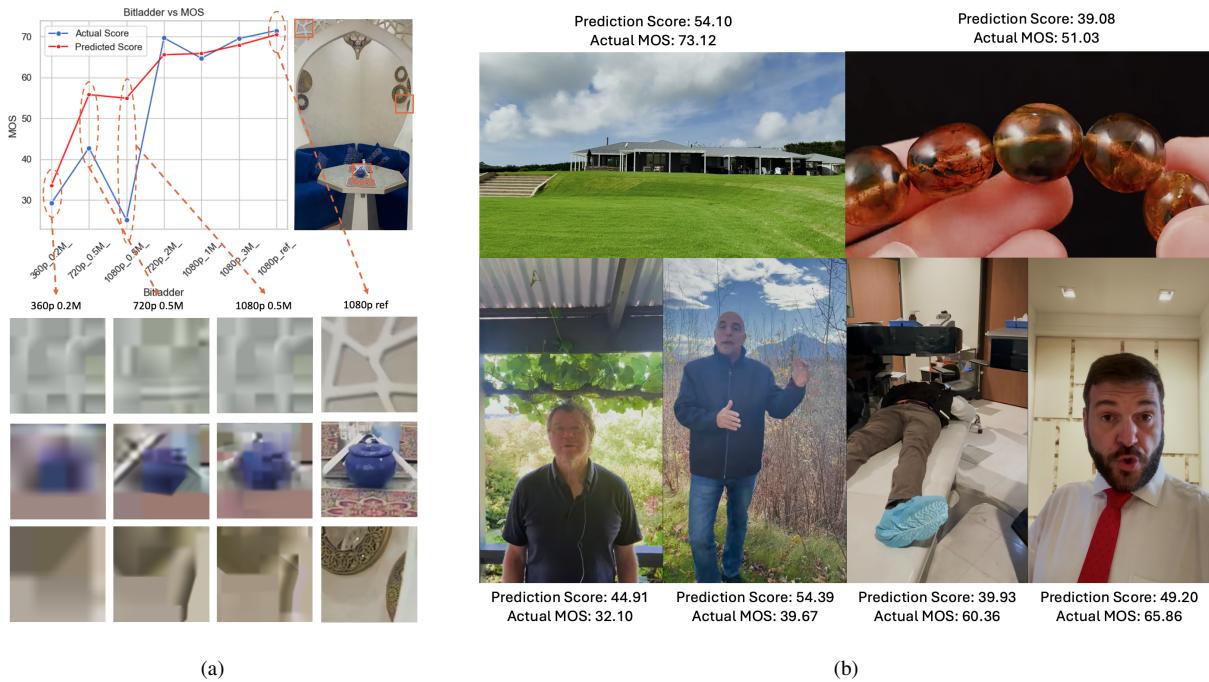


Figure 17. Scatter plots of actual MOS vs. predicted scores for 13 methods evaluated on *BrightVQ*, with parametric fits $l(s)$ in red. A tighter clustering around the diagonal curve indicates a stronger alignment with subjective opinions. Methods yielding narrower scatter demonstrate higher predictive accuracy and consistency, underscoring their ability to capture the underlying perceptual quality cues.

280 Deep learning-based methods show stronger performance,
281 with several achieving a higher degree of correlation by
282 leveraging learned features and spatiotemporal representa-
283 tions. Moreover, HDR-specific VQA models generally out-

perform generic NR-VQA methods, demonstrating the im-
 portance of HDR-aware architectures in perceptual quality
 assessment. Our proposed *BrightRate* model achieves the
 highest correlation ($r = 0.91$), significantly outperforming

284
285
286
287

Figure 18. Failure cases in *BrightRate* predictions.

other approaches. The scatter plot for *BrightRate* shows a strong linear relationship between predicted scores and MOS, indicating its high accuracy and reliability in evaluating HDR video quality.

Fig. 18 presents several failure cases where the predicted video quality scores deviate significantly from the actual MOS. These discrepancies highlight limitations in the model's ability to accurately predict perceptual quality under certain conditions. Fig. 18 (a) illustrates cases where low-resolution, highly compressed videos received higher-than-expected predictions. The close-up patches of compressed video artifacts reveal that blockiness and blurring effects are not always adequately penalized by the model, leading to overestimated quality scores in severely compressed videos. Fig. 18 (b) show video screenshots with complex textures, reflections, or dynamic lighting, where the model struggles to properly assess fine details and HDR characteristics. In videos with human subjects, facial expressions, lighting conditions, or background complexity may lead to misinterpretations of perceptual quality by the model. These failure cases highlight the need for further refinement in *BrightRate*'s HDR-aware feature extraction and compression robustness, ensuring improved alignment with human perception.

References

- [1] 99Firms. Facebook video statistics, 2024. [Online].

- [2] Apple Inc. Hls authoring specification for apple devices, 2024. Accessed: Feb. 2024. 314
- [3] 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. 315
- [4] 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. 316
- [5] FFmpeg Developers. Ffmpeg. <https://ffmpeg.org/>. Accessed: 2025-02-04. 317
- [6] Google Support. Recommended upload encoding settings, 2024. Accessed: Feb. 2024. 318
- [7] International Telecommunication Union. Methodology for the Subjective Assessment of the Quality of Television Pictures. Technical Report BT.500-14, International Telecommunication Union, 2019. 319
- [8] ITU. Bt.2020 : Image parameter values for high dynamic range television for use in production and international programme exchange,. https://www.itu.int/dms_pubrec/itu-r/rec/bt/R-REC-BT.2100-3-202502-I!!PDF-E.pdf. 320
- [9] Zhi Li, Christos G. Bampis, Lucjan Janowski, and Ioannis Katsavounidis. A simple model for subject behavior in subjective experiments. In *Electronic Imaging*, pages 131–1–131–14, 2020. 321

- 344 [10] Yiting Lu, Xin Li, Yajing Pei, Kun Yuan, Qizhi Xie, Yunpeng
345 Qu, Ming Sun, Chao Zhou, and Zhibo Chen. Kvq: Kwai
346 video quality assessment for short-form videos. In *Proceed-*
347 *ings of the IEEE/CVF Conference on Computer Vision and*
348 *Pattern Recognition*, pages 25963–25973, 2024.
- 349 [11] Pavan C. Madhusudana, Neil Birkbeck, Yilin Wang, Balu
350 Adsumilli, and Alan C. Bovik. Image quality assessment
351 using contrastive learning. *IEEE Trans. Image Process.*, 31:
352 4149–4161, 2022.
- 353 [12] Maryam Mohsin. 10 youtube statistics every marketer should
354 know in 2020, 2020. [Online].
- 355 [13] Omnicore. Tiktok by the numbers, 2024. [Online].
- 356 [14] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya
357 Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry,
358 Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen
359 Krueger, and Ilya Sutskever. Learning transferable visual
360 models from natural language supervision. In *ICML*, pages
361 8748–8763. PMLR, 2021.
- 362 [15] Zaixi Shang, Joshua P Ebenezer, Abhinav K Venkatara-
363 manan, Yongjun Wu, Hai Wei, Sriram Sethuraman, and
364 Alan C Bovik. A study of subjective and objective quality
365 assessment of hdr videos. *IEEE Transactions on Image Pro-*
366 *cessing*, 33:42–57, 2023.
- 367 [16] SMPTE. High dynamic range electrooptical transfer func-
368 tion of mastering reference displays. [https://pub.
369 smp-te.org/latest/st2084/st2084-2014.pdf](https://pub.smp-te.org/latest/st2084/st2084-2014.pdf).
- 370 [17] Abhinav K. Venkataraman and Alan C. Bovik. Subjec-
371 tive quality assessment of compressed tone-mapped high dy-
372 namic range videos, 2024.
- 373 [18] Zicheng Zhang, Wei Wu, Wei Sun, Dangyang Tu, Wei Lu,
374 Xiongkuo Min, Ying Chen, and Guangtao Zhai. Md-vqa:
375 Multi-dimensional quality assessment for ugc live videos,
376 2023.