

BrightRate: Quality Assessment for User-Generated HDR Videos Supplementary Material

Anonymous ICCV submission

Paper ID 5421

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
 - 005 • **Appendix B.1:** Comprehensive details on video collection, filtering, transcoding, and the bitrate ladder.
 - 006
 - 007 • **Appendix B.2:** Full description of our AMT study, including instructions, screening procedure, and rejection criteria.
 - 008
 - 009
 - 010 • **Appendix B.3:** Analysis of inter-subject consistency and SUREAL-based MOS estimation.
 - 011
 - 012 • **Appendix C:** Detailed examination of luminance, colorfulness, and spatial-temporal characteristics.
 - 013
 - 014 • **Appendix D:** Additional technical specifics on resizing, normalization, and training.
 - 015
 - 016 • **Appendix E:** Extended results, ablation studies, and failure case analyses.
 - 017

018 A. Background

019 The explosion of UGC on platforms such as YouTube,
020 Facebook, Instagram, and TikTok has transformed video
021 streaming into a ubiquitous, user-driven experience, gen-
022 erating billions of daily views [1, 12, 13]. However, di-
023 verse distortion patterns arising from user editing, compres-
024 sion, and platform-specific processing complicate quality
025 assessment [10, 18]. High Dynamic Range (HDR) imag-
026 ing, supported by mainstream platforms and devices, offers
027 enhanced visual experiences through a broader luminance
028 and color range. Unlike Standard Dynamic Range (SDR)
029 videos, which are limited to 0.1 to 100 cd/m^2 , HDR can
030 represent luminance from 10^{-4} to 10^4 cd/m^2 [8]. HDR10,
031 a widely adopted format, supports 10-bit color depth and
032 Rec. 2020 color gamut (covering 75.8% of the CIE 1931
033 color space), providing higher peak luminance, improved
034 color accuracy, and more detail in both shadows and high-
035 lights, offering a richer, more immersive viewing experi-
036 ence than SDR. The transition to HDR for UGC poses
037 challenges for Video Quality Assessment (VQA) due to in-
038 creased bit depth, broader luminance range, and complex

039 electro-optical transfer functions (EOTFs) like SMPTE ST
040 2084 [16]. Traditional SDR-based models fail to capture
041 these HDR-specific features and the variability of distortions
042 from different devices and editing techniques, thereby
043 impeding effective quality prediction. Furthermore, the ab-
044 sence of a publicly available HDR-UGC database has limited
045 the development of HDR-specific VQA models.

046 B. Details of Dataset Construction

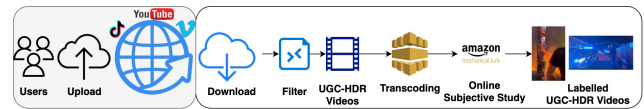


Figure 1. Overview of the dataset preparation approach.

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

051 B.1. Video data Collection



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

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

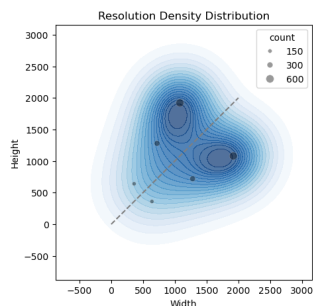


Figure 3. Resolution distribution of *BrightVQ* dataset, maintaining a balanced mix of landscape and portrait videos to study orientation-based perceptual differences.

052 HDR-UGC videos were collected from Vimeo under
 053 Creative Commons licenses to ensure open-source access-
 054 ibility. An initial pool of over 10,000 videos was automat-
 055 ically filtered using metadata checks for HDR flags, resolu-
 056 tion, 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-
 077 lyzing landscape and portrait videos. The box plot repre-
 078 sentation highlights the diversity in perceptual quality rat-

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

ings across different encoding configurations, showing how
 bitrate and resolution choices affect MOS scores. Fig. 3 il-
 lustrates the resolution density distribution of videos in the
BrightVQ dataset. Fig. 5 further visualizes the compression
 artifacts introduced through this approach. This multi-stage
 process guarantees that *BrightVQ* reflects authentic HDR-
 UGC content with diverse distortions.

B.2. Crowdsourced Subjective Study

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

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

To maintain data integrity, we implemented strict rejection
 criteria at multiple stages:

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

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

B.3. Subjective Score Processing

To evaluate inter-subject consistency, we randomly split all
 MOS ratings into two independent groups and computed the
 Spearman Rank Correlation Coefficient (SRCC) and Pear-
 son Linear Correlation Coefficient (PLCC) between them.
 As shown in Fig. 12a, the study achieved a median SRCC of

079
080
081
082
083
084
085
086
087
088
089
090
091
092
093
094
095
096
097
098
099
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129

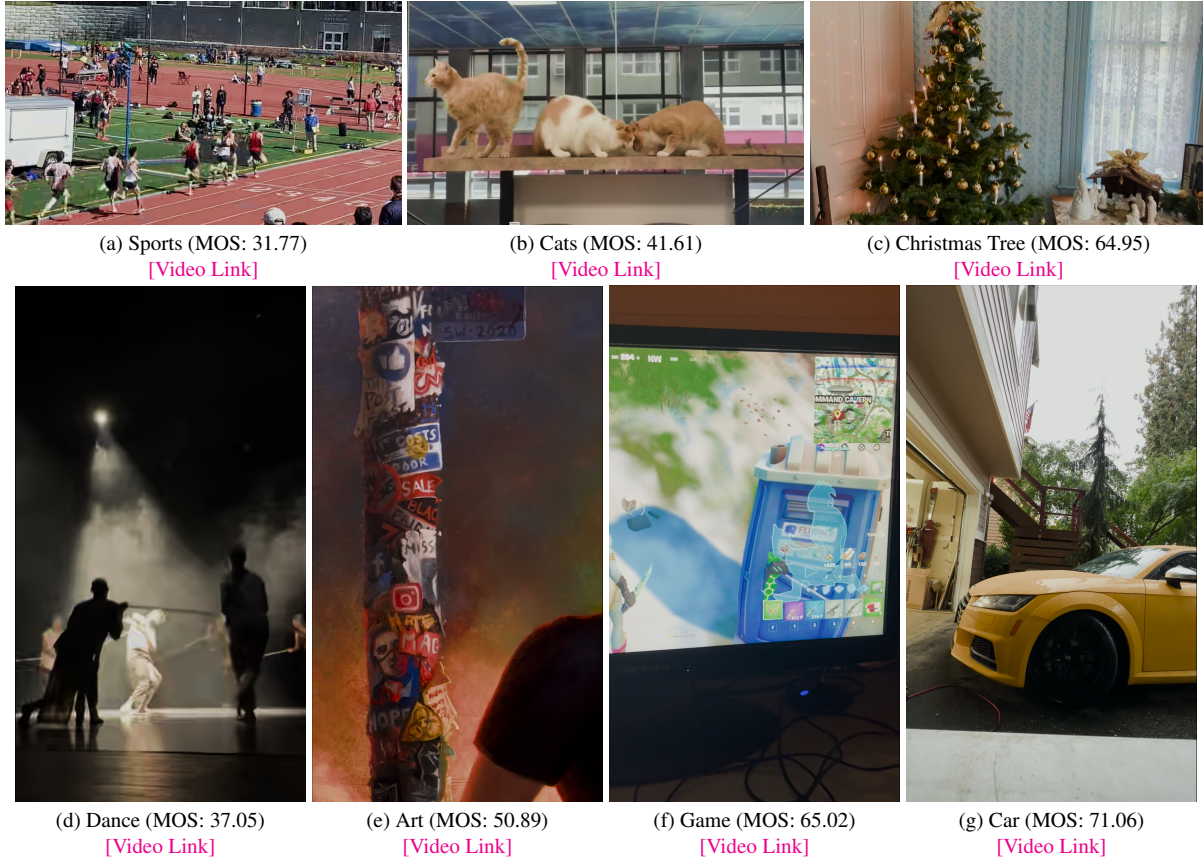


Figure 4. Example frames from *BrightVQ* dataset. Each frame is presented with its category, the MOS for the video and a direct video access link.

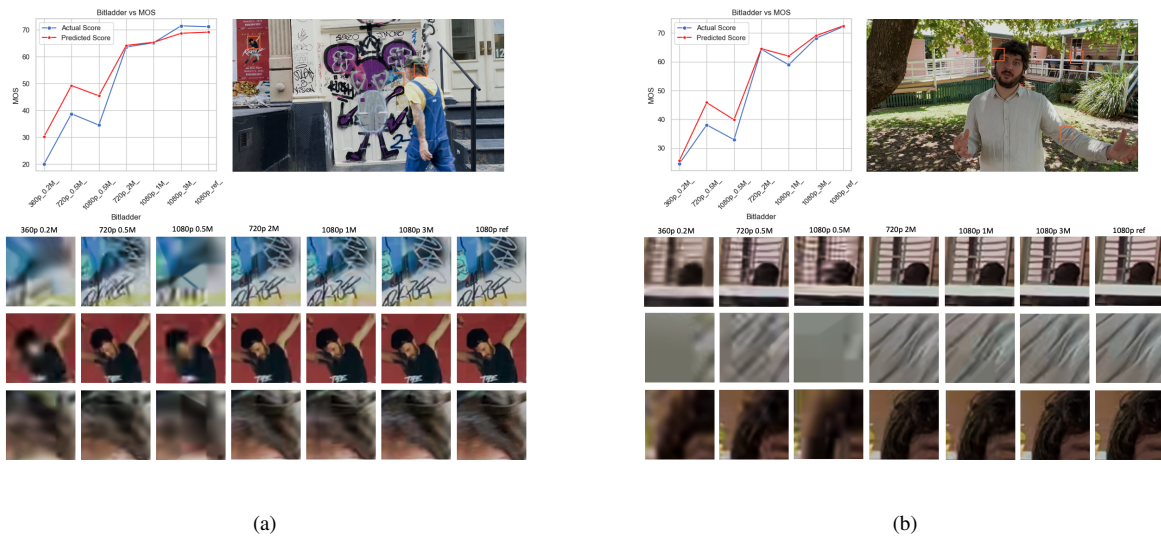


Figure 5. The combination of MOS vs. predicted score plots with visual comparisons of specific image regions to highlight the correlation between distortion and MOS across different bitrates and resolutions.

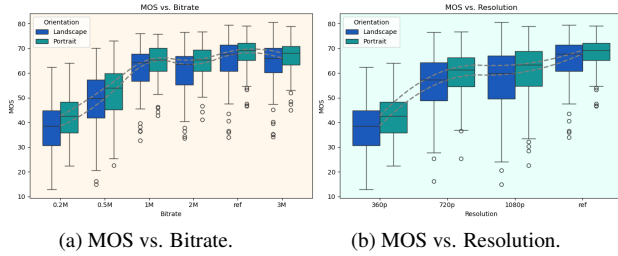


Figure 6. (a) and (b) show the MOS variations across bitrate and resolution respectively for *BrightVQ*.

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 in 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," "FAIR," "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 passed or replayed. If you did not see a video, you can press the "I didn't see the video" button. However, note that if you miss too many videos, your HIT will be rejected.
3. Rate the video by 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 must 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.

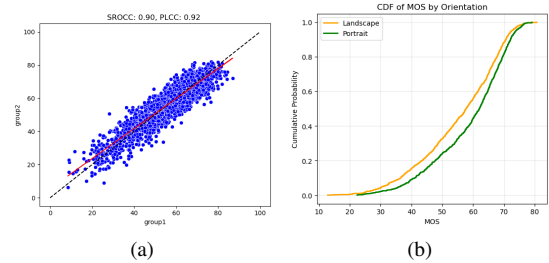


Figure 12. (a). Inter-subject correlation. (b). MOS CDF distribution of all videos in *BrightVQ*.

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

130
131
132
133
134
135

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

148
$$S_{ij} = \psi_j + \Delta_i + \nu_i X, \quad X \sim \mathcal{N}(0, 1), \quad (1)$$

149 where ψ_j represents the true quality of video j , Δ_i captures
 150 the bias of subject i , and ν_i reflects the rating inconsis-
 151 tency of subject i . The parameters are estimated using Maxi-
 152 mum Likelihood Estimation (MLE), maximizing the like-
 153 lihood. Unlike hard rejection, which entirely removes out-
 154 liers, SUREAL downweights ratings from less reliable sub-
 155 jects. This ensures that MOS values reflect true perceptual
 156 quality while mitigating distortions from inconsistent raters.
 157 By applying SUREAL, we obtained more stable MOS esti-
 158 mates, which accurately reflect the perceptual quality of
 159 HDR content across diverse video conditions. The CDF dis-
 160 tribution of all videos in *BrightVQ* are shown in Fig. 12b.

161 **C. Analysis of the HDR content**

162 In this section, we provide a detailed analysis of the
 163 dataset’s key characteristics, focusing on luminance and
 164 color distribution, spatial-temporal diversity, perceptual
 165 quality trends, and HDR-specific challenges.

166 Fig. 13 presents the distribution of luma and colorfulness
 167 across the 300 source HDR videos in *BrightQA*. The first
 168 three histograms illustrate the minimum, maximum, and
 169 mean luma values, highlighting the variation in bright-
 170 ness levels across different videos. This demonstrates that
 171 the dataset includes both dark and bright HDR scenes, ensur-
 172 ing a wide dynamic range. The fourth histogram shows the
 173 colorfulness distribution, reflecting variations in chromatic
 174 intensity across different videos.

175 To further quantify the diversity in content complex-
 176 ity, Fig. 14 presents an analysis of spatial-temporal com-
 177 plexity, spatial information (SI), and temporal information
 178 (TI) within the dataset. The scatter plots in Fig. 14 (a)-(c)
 179 demonstrate the variability in SI and TI values, showing a
 180 wide distribution of motion and texture complexity across
 181 the dataset. Higher SI values typically correspond to de-
 182 tailed textures and sharp edges, while higher TI values in-
 183 dicate rapid motion or dynamic scenes. The dataset covers
 184 both high-detail static scenes and fast-moving dynamic con-
 185 tent, ensuring its suitability for evaluating compression and
 186 HDR features across different motion characteristics.

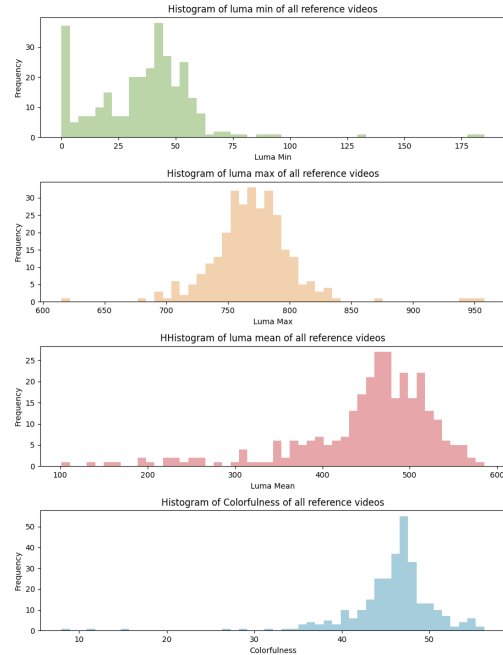


Figure 13. Distribution of luma and colorfulness of the source HDR-UGC videos in *BrightVQ* dataset.

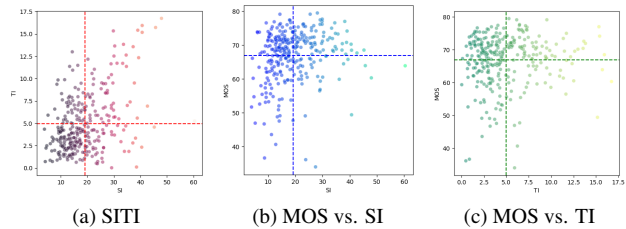


Figure 14. (a) Spatial-Temporal Complexity, (b) MOS vs. Spatial Information (SI), and (c) MOS vs. Temporal Information (TI).

187 Fig. 15 demonstrates the diversity of *BrightVQ* dataset
 188 in both aesthetic and technical aspects. The scatter plot
 189 shows a wide range of ratings, with each point represent-
 190 ing a video and color-coded by its actual subjective quality
 191 score. The marginal histograms further highlight the dis-
 192 tribution of scores, illustrating the broad variation in per-
 193 ceptual and technical quality across different content. The
 194 *BrightVQ* dataset presents a diverse range of HDR-UGC
 195 content, covering natural landscapes, indoor scenes, and
 196 various complex lighting conditions, capturing both HDR-
 197 specific and UGC-specific distortions. This diversity ensur-
 198 es that *BrightVQ* provides a comprehensive and realistic
 199 benchmark for evaluating video quality.

200 Figure 16 provides examples of HDR videos subjected
 201 to various spatial resolutions and bitrates, along with their
 202 MOS and luminance histograms. In (a) and (c), the
 203 higher-resolution, higher-bitrate frames retain more de-

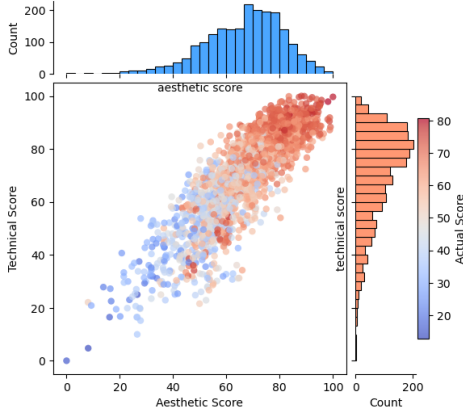


Figure 15. Diversity in aesthetic and technical quality scores in *BrightVQ* dataset.

tails and exhibit fewer artifacts, whereas lower-resolution, lower-bitrate versions show noticeable blocking and banding—particularly in the extreme luma regions. Subfigures (b) and (d) illustrate the broader luminance distribution characteristic of HDR, indicating significant content in both low and high intensity ranges. Such distributions underscore the importance of HDR-specific processing, since distortions in extreme luminance regions can disproportionately affect subjective quality.

D. More details on Implementation

Here we detail the key implementation steps of our *BrightRate* model. For semantic features, input frames are resized 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 $Y^t \in [0, 1]$, 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:

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

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.

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. Deep learning-based methods show stronger performance, with several achieving a higher degree of correlation by leveraging learned features and spatiotemporal representations. Moreover, HDR-specific VQA models generally outperform generic NR-VQA methods, demonstrating the importance of HDR-aware architectures in perceptual quality assessment. Our proposed *BrightRate* model achieves the highest correlation ($r = 0.91$), significantly outperforming 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,

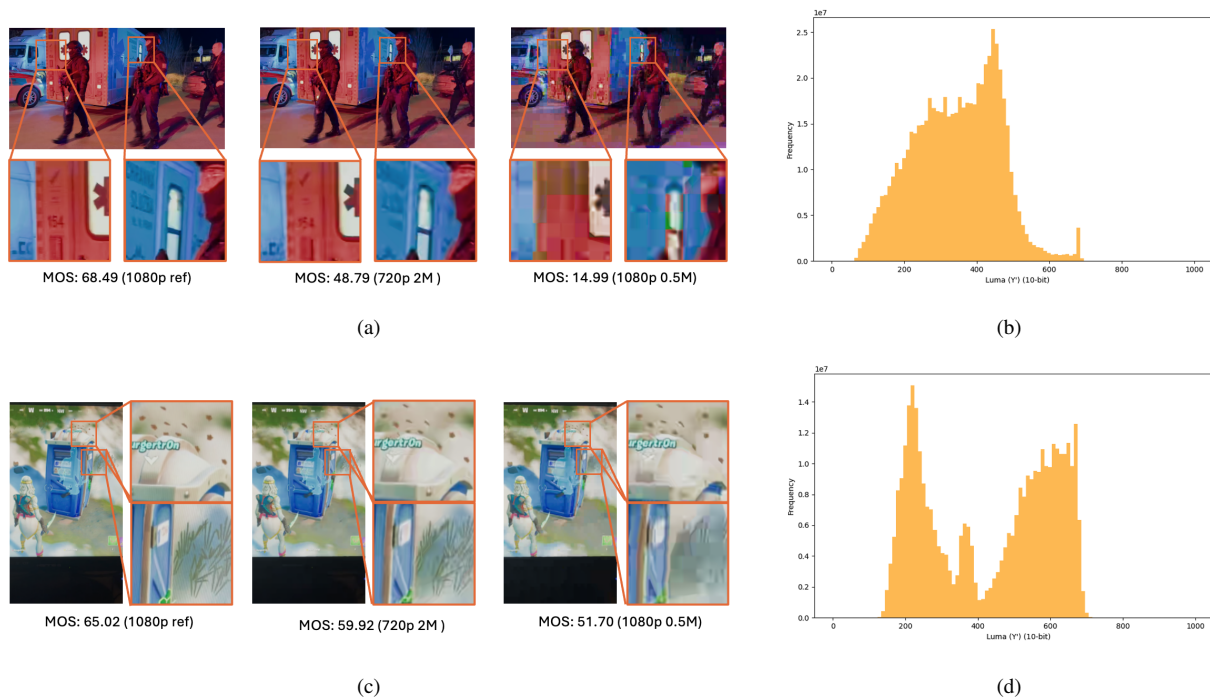


Figure 16. Illustrations of HDR content under different resolutions and bitrates. (a) and (c) show reference frames at 1080p and their lower-resolution, lower-bitrate counterparts, with red boxes highlighting high luma areas with artifacts (e.g., blocking, color banding) become more pronounced. The corresponding 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.

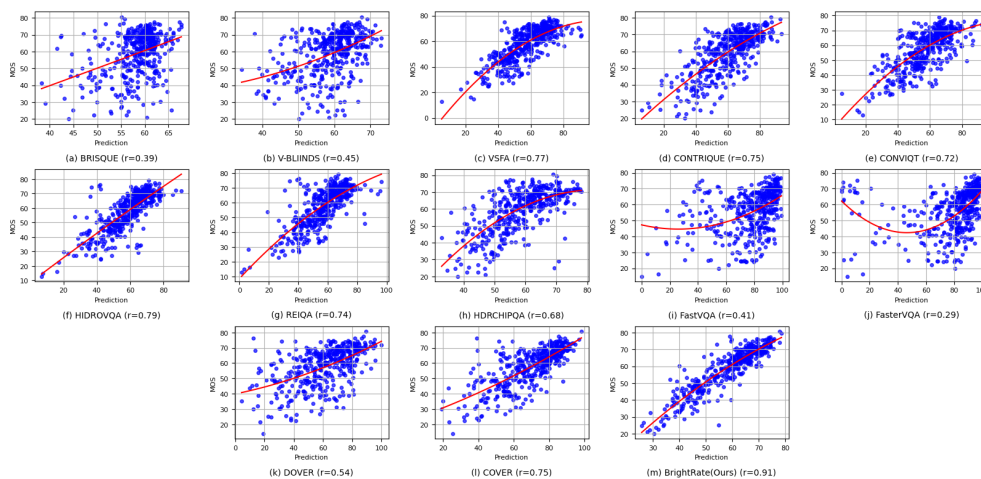


Figure 17. Scatter plots of actual MOS vs. predicted scores for 13 methods evaluated on *BrightVQ*, with parametric fits $l(s)$ in red. A tighter clustering around the diagonal curve indicates a stronger alignment with subjective opinions. Methods yielding narrower scatter demonstrate higher predictive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

283
284
285
286

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 re-

287
288
289
290

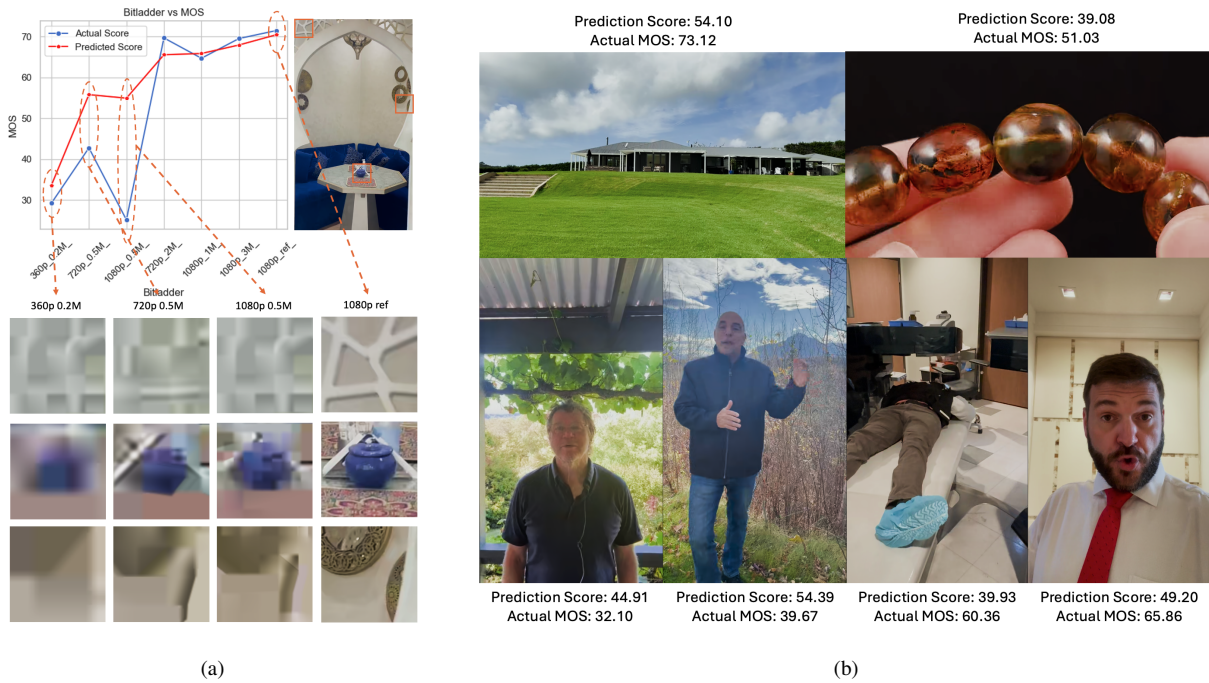


Figure 18. Failure cases in *BrightRate* predictions.

291 finement in *BrightRate*'s HDR-aware feature extraction and
 292 compression robustness, ensuring improved alignment with
 293 human perception.

294 **References**

295 [1] 99Firms. Facebook video statistics, 2024. [Online].
 296 [2] Apple Inc. Hls authoring specification for apple devices,
 297 2024. Accessed: Feb. 2024.
 298 [3] Yu-Chih Chen, Avinab Saha, Alexandre Chapiro, Christian
 299 Häne, Jean-Charles Bazin, Bo Qiu, Stefano Zanetti, Ioannis
 300 Katsavounidis, and Alan C. Bovik. Subjective and objec-
 301 tive quality assessment of rendered human avatar videos in
 302 virtual reality. *IEEE Transactions on Image Processing*, 33:
 303 5740–5754, 2024.
 304 [4] Joshua P. Ebenezer, Zaixi Shang, Yongjun Wu, Hai Wei,
 305 Sriram Sethuraman, and Alan C. Bovik. Hdr-chipqa: No-
 306 reference quality assessment on high dynamic range videos,
 307 2023.
 308 [5] Ffmpeg Developers. Ffmpeg. <https://ffmpeg.org/>.
 309 Accessed: 2025-02-04.
 310 [6] Google Support. Recommended upload encoding settings,
 311 2024. Accessed: Feb. 2024.
 312 [7] International Telecommunication Union. Methodology for
 313 the Subjective Assessment of the Quality of Television Pic-
 314 tures. Technical Report BT.500-14, International Telecom-
 315 munication Union, 2019.
 316 [8] ITU. Bt.2020 : Parameter values for ultra-high defini-
 317 tion television systems for production and international pro-

gramme exchange,. <https://glenwing.github.io/docs/ITU-R-BT.2020-1.pdf>. 318
 319
 [9] Zhi Li, Christos G. Bampis, Lucjan Janowski, and Ioannis 320
 Katsavounidis. A simple model for subject behavior in sub- 321
 jective experiments. In *Electronic Imaging*, pages 131–1– 322
 131–14, 2020. 323
 [10] Yiting Lu, Xin Li, Yajing Pei, Kun Yuan, Qizhi Xie, Yunpeng 324
 Qu, Ming Sun, Chao Zhou, and Zhibo Chen. Kvg: Kwai 325
 video quality assessment for short-form videos, 2024. 326
 [11] Pavan C. Madhusudana, Neil Birkbeck, Yilin Wang, Balu 327
 Adsumilli, and Alan C. Bovik. Image quality assessment 328
 using contrastive learning. *IEEE Trans. Image Process.*, 31:
 329 4149–4161, 2022. 330
 [12] Maryam Mohsin. 10 youtube statistics every marketer should 331
 know in 2020, 2020. [Online]. 332
 [13] Omnicore. Tiktok by the numbers, 2024. [Online]. 333
 [14] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya 334
 Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, 335
 Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen 336
 Krueger, and Ilya Sutskever. Learning transferable visual 337
 models from natural language supervision. In *ICML*, pages 338
 8748–8763. PMLR, 2021. 339
 [15] Zaixi Shang, Joshua P Ebenezer, Abhinav K Venkatara- 340
 manan, Yongjun Wu, Hai Wei, Sriram Sethuraman, and 341
 Alan C Bovik. A study of subjective and objective quality 342
 assessment of hdr videos. *IEEE Transactions on Image Pro- 343
 cessing*, 33:42–57, 2023. 344
 [16] SMPTE. High dynamic range electrooptical transfer func- 345
 tion of mastering reference displays. <https://pub.smpete.org/latest/st2084/st2084-2014.pdf>. 346
 347

- 348 [17] Abhinav K. Venkataramanan and Alan C. Bovik. Subjec-
349 tive quality assessment of compressed tone-mapped high dy-
350 namic range videos, 2024.
- 351 [18] Zicheng Zhang, Wei Wu, Wei Sun, Dangyang Tu, Wei Lu,
352 Xionghuo Min, Ying Chen, and Guangtao Zhai. Md-vqa:
353 Multi-dimensional quality assessment for ugc live videos,
354 2023.