

# BrightRate: Quality Assessment for User-Generated HDR Videos

Anonymous WACV **Algorithms Track** submission

Paper ID 283

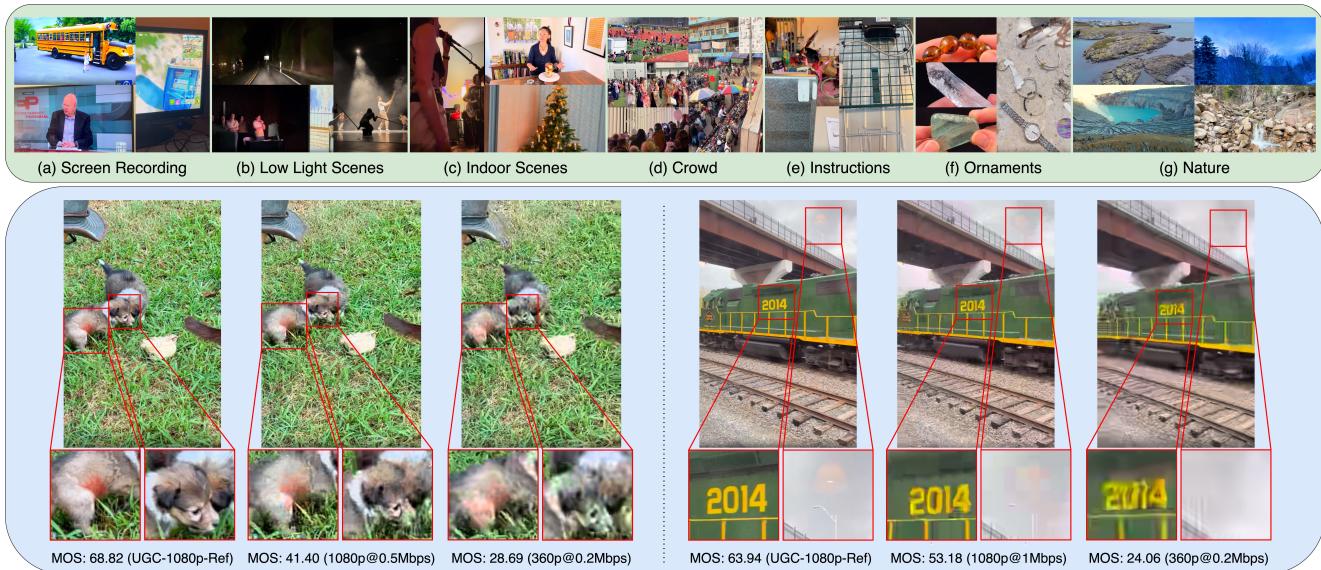


Figure 1. Upper green-bordered box illustrates a variety of content categories and challenging scenes in *BrightVQA*. The images in the lower blue-bounded box show the impact of compression on UGC video quality. Some heavily distorted regions are highlighted in red.

## Abstract

**001** High Dynamic Range (HDR) videos offer superior luminance and color fidelity as compared to Standard Dynamic Range (SDR) content. The rapid growth of User-Generated Content (UGC) on platforms such as YouTube, Instagram, and TikTok has brought a significant increase in the volumes of streamed and shared UGC videos. This newer category of videos brings new challenges to the development of effective No-Reference (NR) video quality assessment (VQA) models specialized to HDR UGC, because of the extreme variety and severities of distortions, arising from diverse capture, editing, and processing outcomes. Towards addressing this issue, we introduce **BrightVQ**, a sizeable new psychometric data resource. It is the first large-scale subjective video quality database dedicated to the quality modelling of HDR UGC videos. BrightVQ comprises 2,100 videos, on which we collected 73,794 percep-

tual quality ratings. Using this dataset, we also developed **BrightRate**, a novel video quality prediction model designed to capture both UGC-specific distortions coexisting with HDR-specific artifacts. Extensive experimental results demonstrate that BrightRate achieves state-of-the-art performance across HDR databases. Project page: <https://brightvqa.github.io/BrightVQ/>

017  
018  
019  
020  
021  
022  
023

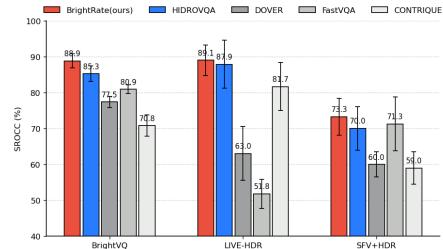


Figure 2. Benchmark performance of *BrightRate*(our) and other leading state-of-the-art (SOTA) VQA models on available HDR VQA datasets.

\*Corresponding Author.

†Project Lead.

024

## 1. Introduction

The explosion of User-Generated Content (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, 29, 32]. However, the diverse distortion patterns from in-expert capture, editing, compressing, and platform-specific processing complicate quality assessment [22, 55]. High Dynamic Range (HDR) imaging further enhances visual experiences with broader luminance and color gamuts as compared to Standard Dynamic Range (SDR) [10, 40]. For example, HDR10 supports 10-bit depth and Rec. 2020 color gamut, delivering superior detail in shadows and highlights [17]. Despite significant advances in VQA [2, 9, 10, 30, 38–40], current SDR-based models fail to capture HDR-specific features and in particular, tremendously diverse HDR-UGC distortions, impeding the development of effective quality prediction models.

Table 1. Overview of the *BrightVQ* Dataset, summarizing video specifications (format, resolution, duration), the encoding bitrate ladder,\* with extensive subjective quality annotations.

Attribute	Details
<b>Video Specifications</b>	
Format	Rec. 2020, 10-bit, PQ
Resolutions	1920×1080, 1080×1920, 1280×720, 720×1280, 640×360, 360×640
Bitrates(Mbps)	0.2, 0.5, 1.0, 2.0, 3.0, Reference
Duration(Sec.)	4 – 10
<b>Dataset Statistics</b>	
Reference Videos	300 (150 Landscape, 150 Portrait)
Total Videos	2100
Total Scores	73,794
Avg. Scores/Video	35

Towards aiding progress on this increasingly important problem, we introduce *BrightVQ*, the first large-scale, open-source video quality database designed for the quality analysis of UGC in HDR format. This dataset includes 2,100 videos derived from 300 diverse HDR-UGC source videos, and spans a wide range of content types, from action sequences to vlogs and natural landscapes (see Table 1). *BrightVQ* captures coincident HDR-specific and UGC-specific distortions, reflecting the complexities of real-world HDR-UGC VQA. We also conducted the first large-scale crowdsourced subjective study for HDR-UGC, collecting 73,794 subjective ratings from participants using HDR-capable displays. This rich collection of diverse contents and human annotations establishes *BrightVQ* as a

\*Based on YouTube’s streaming guidelines [13] and Apple’s HLS authoring specifications [4].

powerful resource for advancing HDR-UGC VQA.

Additionally, we created *BrightRate*, a novel model for HDR-UGC quality assessment. *BrightRate* employs multiple branches to capture UGC-specific distortions, semantic cues, and HDR-specific artifacts, especially in extreme luminance regions. This hybrid approach yields state-of-the-art prediction accuracy and interpretability, outperforming existing VQA methods. Our experiments on *BrightVQ* and other HDR datasets (Fig. 2) validate the effectiveness of *BrightRate* on handling both UGC and HDR-specific distortions. The contributions of this paper are summarized below:

- We introduce *BrightVQ*, the first large-scale HDR-UGC video quality database, that is-ten times larger than previous public HDR datasets [40].
- We conducted the first large-scale crowdsourced subjective study on HDR-UGC videos, collecting 73,794 ratings from over 200 participants.
- We created *BrightRate*, a novel HDR-UGC video quality prediction model that fuses UGC, HDR-specific, semantic, and temporal features to achieve state-of-the-art prediction performance.
- We conducted extensive experiments on *BrightVQ* and other HDR datasets to study the effectiveness and broad applicability of *BrightRate* against other SOTA models.

## 2. Related Work

**HDR-UGC VQA Databases.** Various UGC VQA databases [12, 31, 41, 52, 54, 55] capture real-world distortions falling into two categories: In-the-Wild UGC Datasets[15, 41, 46, 50, 52–54], which contain naturally distorted videos but lack control over degradation types, and Simulated UGC Distortion Datasets[12, 20, 31, 55], which model compression and transmission artifacts in controlled settings. However, HDR VQA databases remain limited in scale, accessibility, and compliance with modern standards. Early datasets like DML-HDR [5] and Compressed-HDR [33] were small and had restricted availability, while others [6, 35] lacked HDR10 compliance. LIVE-HDR [40] introduced a professionally generated HDR dataset but contains only 31 video contents, limiting its relevance for UGC scenarios. More recently, Wang et al.[47] created a short-form HDR dataset with 2,000 videos, but only 300 include subjective scores. As HDR adoption in UGC grows, a large-scale VQA database is needed to effectively capture real-world distortions and quality variations. Table 2 compares existing HDR datasets.

**HDR-UGC VQA Methods.** Modern VQA models may be broadly categorized into handcrafted feature-based and deep learning-based methods. Handcrafted approaches [8, 18, 26–28, 36] extract powerful distortion-aware statistical and perceptual features but struggle with complex UGC distortions. Deep learning-based models leverage pre-

Table 2. Comparison of *BrightVQ* with existing HDR VQA datasets.

Dataset	Format	Total Videos (Ref.)	Source	Total Opinions	Orientation	Subjective Study
LIVE-HDR [40]	Rec. 2020, HDR10, PGC	310 (31)	Internet Archive	2,400	Landscape	In-Lab
SFV+HDR [47] (only HDR)	Rec. 2020, HDR10, UGC	300 (300)	YouTube	N/A	Portrait	In-Lab
<b>BrightVQ (Ours)</b>	Rec. 2020, HDR10, UGC	2100 (300)	Vimeo	73,794	Portrait+Landscape	Crowdsourced

108 trained networks to extract semantic and perceptual features. Among these, for example, VSFA[19] captures temporal variations, FAST/FASTER-VQA[48, 49] uses Transformers, CONTRIQUE[23] applies self-supervised learning, and DOVER[50] integrates aesthetic and technical quality assessment. However, most models, including these, 109 are optimized for SDR and fail to handle HDR-specific distortions. 110 HDR-VQM[30] and HDR-BVQM[2] introduce 111 brightness-aware features but must rely on reference videos or lack HDR-specific adaptations. 112 PU21[25] refines traditional metrics with perceptually uniform encoding but 113 remains content- and display-dependent. 114 HDR-ChipQA[10] extends ChipQA with non-linear luminance transformations, while HIDRO-VQA [38] trains CONTRIQUE [23] 115 on unlabeled HDR videos from YouTube. However, none is 116 able to effectively capture HDR-UGC distortions, limiting 117 their applicability for HDR-UGC quality prediction. 118

### 119 3. Large-Scale Dataset and Human Study

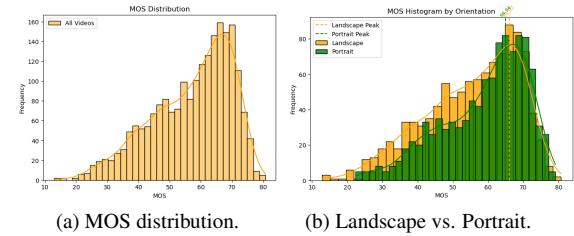


Figure 3. Overview of crowdsourced online subjective study on Amazon Mechanical Turk (AMT).

126 In this section, we discuss the newly proposed HDR- 127 UGC VQA dataset-*BrightVQ*. *BrightVQ* comprises 2,100 128 videos generated from 300 diverse HDR-UGC source clips 129 that span a wide range of real-world contents—including 130 indoor and outdoor scenes, food, vlogs, and natural land- 131 scapes (see Fig. 1). Table 1 provides an overview of key 132 video specifications and the encoding bitrate ladder used to 133 simulate realistic streaming conditions.

#### 134 3.1. Dataset Collection

135 HDR-UGC videos were sourced from Vimeo under Creative 136 Commons licenses. Over 10,000 videos were automatically 137 filtered by HDR flags, resolution, format, and category, 138 followed by manual verification to ensure authenticity. 139 Videos were truncated to 10 seconds at a maximum res- 140 olution of 1080p using `ffmpeg` [11] and transcoded with 141 an industry-standard bitrate ladder [4, 13]. This multi-stage 142 process ensured that *BrightVQ* represents authentic HDR-

Figure 4. (a) MOS distribution of all videos in *BrightVQ*. (b) MOS distributions of landscape and portrait orientation videos in *BrightVQ*.

UGC content with diverse distortions. Please refer to [Supplementary Materials Sec. B. 1](#) for more details.

#### 3.2. Subjective Quality Study

To obtain reliable human subjective quality annotations, we conducted the first large-scale crowdsourced HDR-UGC study on AMT (Fig. 3). Over 200 participants with HDR-capable devices provided 73,794 ratings (averaging 35 ratings per video). The study included a comprehension quiz, a training phase with six HDR videos, and a testing phase where each subject rated 94 videos (with 5 golden set videos and 5 repeated videos). During the data collection, rigorous device checks, playback monitoring, and golden set validation ensured that unreliable raters were excluded, with subject rejections following the ITU-R BT.500-14 standard [16].

To derive robust Mean Opinion Scores (MOS), we employed the SUREAL method [21], which accounts for subject bias and inconsistency. Each rating  $S_{ij}$  from subject  $i$  for video  $j$  is modeled as:

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

where  $\psi_j$  represents the true quality of video  $j$ ,  $\Delta_i$  captures the bias of subject  $i$ , and  $\nu_i$  reflects the rating inconsistency of subject  $i$ . Parameters are estimated using Maximum Likelihood Estimation (MLE), resulting in MOS values that are robust to outliers and unreliable ratings. Full description of our AMT study is in [Supplementary Materials Sec. B. 2](#).

#### 3.3. Analysis of MOS

Fig. 4a depicts the MOS distribution of *BrightVQ*, which is right-shifted, similar to other HDR datasets. This trend

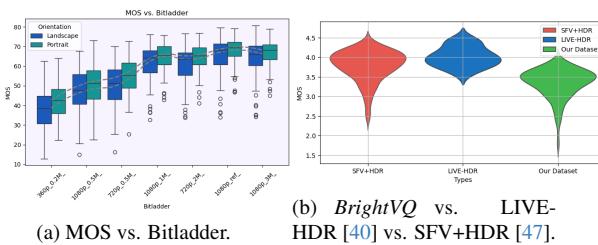


Figure 5. (a) shows the MOS variations across bitladder for *BrightVQ*. (b) compares MOS distributions of *BrightVQ*, LIVE-HDR [40], and SFV+HDR [47], showing that *BrightVQ* has a broader spectrum of MOS with less bias in peak MOS value.

suggests that HDR videos, due to their inherently higher luminance and richer color details, often receive higher quality ratings. Fig. 4b compares MOS distributions for landscape and portrait videos, showing significant overlap that indicates orientation has minimal impact on perceived quality. Furthermore, Fig. 14a demonstrates that bitrate strongly influences MOS, with lower bitrates resulting in greater variability. A comparative analysis in Fig. 5b reveals that *BrightVQ* covers a wider range of MOS values and exhibits less bias toward high scores than to existing HDR databases, underscoring its ability to capture severe distortions often absent from professional HDR collections. We get a inter-subject correlation of 0.92, more detailed analysis for the MOS scores is in [Supplementary Materials Sec. B.3 and Sec. C](#).

## 4. Proposed Method

Our proposed *BrightRate* model (Fig. 6) is a novel no-reference VQA framework designed for HDR-UGC videos. It combines UGC-specific features from the pretrained CONTRIQUE [23], semantic cues from a CLIP-based encoder [34, 45], HDR features derived under a piecewise non-linear luminance transform on which distortion-sensitive natural video statistics are computed [10, 27, 36, 43], and temporal differences, which are then regressed to MOS. Extensive experiments on our *BrightVQ* and other HDR benchmarks demonstrate state-of-the-art performance.

### 4.1. HDR Feature Extraction

HDR content suffers from distortions in extreme luminance regions that standard SDR-based VQA methods overlook [9, 38]. To address this limitation, we propose a two-step HDR feature extraction module that combines an expansive non-linearity with natural scene statistics (NSS) modeling. First, each frame  $\mathbf{x}^t$  is converted to YUV and its normalized luminance channel  $\mathbf{Y}^t \in [-1, 1]$  is extracted. We then subdivide  $\mathbf{Y}^t$  into two intervals (e.g.,  $[-1, 0]$  and  $[0, 1]$ ) and apply a piecewise expansive non-linearity in-

spired by [10, 40]. Specifically, the non-linearity is defined as:

$$g(x; \beta) = \begin{cases} e^{\beta x} - 1, & x \geq 0, \\ 1 - e^{-\beta x}, & x < 0, \end{cases} \quad (2)$$

with  $\beta = 4$  following [10]. This transformation stretches the extreme ends of the luminance scale while compressing mid-range values, thereby amplifying distortions in highlights and shadows that would otherwise be masked. The expansive non-linearity is applied within a sliding window of size  $w \times w$ , where we choose  $w = 31$  following [10]. The output is an enhanced luminance channel  $\tilde{\mathbf{Y}}^t = g(\mathbf{Y}^t; 4)$  that more clearly reveals HDR-specific artifacts in very dark or very bright regions. Next, we compute Mean-Subtracted Contrast Normalized (MSCN) coefficients from  $\tilde{\mathbf{Y}}^t$  to capture local image statistics:

$$\mathbf{M}^t(i, j) = \frac{\tilde{\mathbf{Y}}^t(i, j) - \mu(i, j)}{\sigma(i, j) + \epsilon}, \quad (3)$$

where  $\mu(i, j)$  and  $\sigma(i, j)$  are computed over a  $31 \times 31$  window and  $\epsilon$  is a small constant. The MSCN coefficients follow a Generalized Gaussian Distribution (GGD), and their adjacent products are modeled with an Asymmetric GGD (AGGD) [10, 26, 27]. We extract shape and variance parameters from both models and concatenate them across two scales to form the HDR-specific feature vector:

$$\mathcal{H}^t = f_{\text{HDR}}(\tilde{\mathbf{Y}}^t) \in \mathbb{R}^{d_{\text{HDR}}}. \quad (4)$$

This two-step approach—expanding extreme luminance details and extracting NSS-based features—effectively highlights HDR-specific distortions critical for accurate quality assessment.

### 4.2. UGC Feature Extraction

UGC typically exhibits a wide range of distortions—including noise, over/under exposure, camera shake, blur, and compression artifacts—stemming from the variability of user skills, capture devices, and post-processing techniques. For our UGC backbone we use CONTRIQUE features [23]. Let  $\mathbf{x}^t \in \mathbb{R}^{H \times W \times 3}$  denote the  $t$ -th frame of a HDR-UGC video. We extract multi-scale features by running the CONTRIQUE [23] encoder on both the full and a downsampled half-resolution frame versions as:

$$\mathcal{U}_{\text{scale}}^t = f_{\text{CONTRIQUE}}(\mathbf{x}^t) \in \mathbb{R}^{d_{\text{UGC}}}. \quad (5)$$

where  $d_{\text{UGC}}$  represents the dimensionality of the extracted feature space. The final UGC feature map is denoted  $\mathcal{U}^t$ , which is a concatenation of both the full and half scale UGC features. As demonstrated in prior work [22, 38] and confirmed by our experiments (Sec. 5), CONTRIQUE [23] serves as a robust UGC backbone.

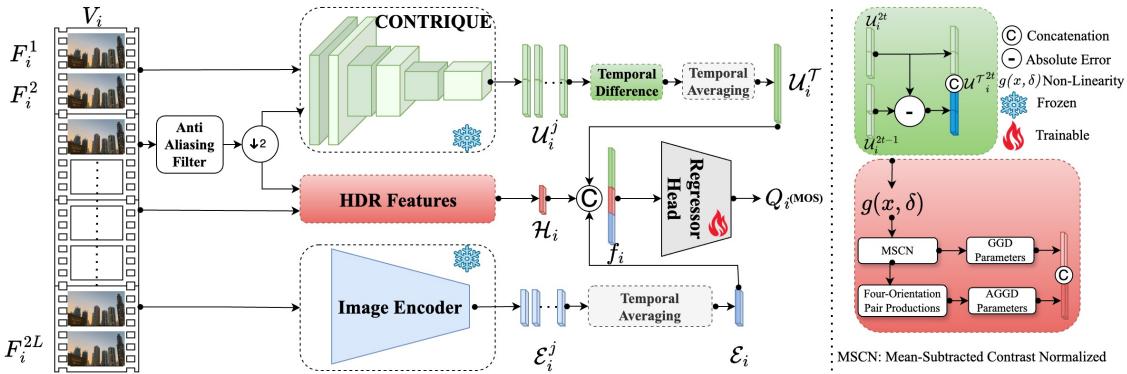


Figure 6. The overall framework of BrightRate for HDR-UGC Video Quality Assessment. BrightRate extracts HDR-specific features, and combines with UGC and Semantic features to give SOTA results on HDR-UGC benchmarks.

255

### 4.3. Semantic Feature Extraction

256

Perceptual quality depends not only on technical distortions but also on content semantics, which can influence human tolerance to various artifacts [14, 45]. For instance, compression artifacts may be more perceptible on homogeneous, flat regions than on richly textured areas. To improve content understanding in our HDR-UGC VQA framework, we employ the CLIP Image Encoder [3, 14, 34, 45]. For each appropriately resized sampled frame  $\mathbf{x}^t$ , semantic features are extracted as

265

$$\mathcal{E}^t = f_{\text{CLIP}}(\mathbf{x}^t) \in \mathbb{R}^{d_{\text{SEM}}}.$$

266

Here,  $d_{\text{SEM}}$  denotes the semantic feature dimension. Leveraging CLIP's fine-grained semantics from millions of image-text pairs, we capture high-level contextual cues that affect perceptual quality. By fusing these with UGC-specific distortion features, we form a holistic representation that enhances sensitivity to both content and technical distortions.

273

### 4.4. Temporal Difference Module

274

Videos with higher perceptual quality typically exhibit smaller temporal fluctuations, while lower-quality videos show abrupt changes [2, 19, 30]. To capture these dynamics, we compute the absolute difference between consecutive UGC feature vectors:

279

$$\Delta \mathcal{U}^t = |\mathcal{U}^t - \mathcal{U}^{t-1}|, \quad t \in \{2, \dots, T\}, \quad (7)$$

280

where  $\mathcal{U}^t$  denotes the combined UGC feature for frame  $t$  (see Sec. 4.2). We then concatenate these temporal differences with the original features, and normalize the result, yielding an enriched representation that captures both static distortions and their temporal fluctuations.

Table 3. Comparison of SOTA IQA and VQA methods on the *BrightVQ* dataset, with median (standard deviations) values reported. Best and second-best results are highlighted in red and blue, respectively, while our proposed *BrightRate* is shaded in gray.

	Method	SROCC( $\uparrow$ )(Std)	PLCC( $\uparrow$ )(Std)	KROCC( $\uparrow$ )(Std)	RMSE( $\downarrow$ )(Std)
NR-IQA	BRISQUE [26]	0.3302 (0.0366)	0.3603 (0.0311)	0.2261 (0.0279)	12.5770 (0.2855)
	HDRMAX [40]	0.6276 (0.0321)	0.6318 (0.0356)	0.4409 (0.0288)	10.2428 (0.4008)
	CONTRIQUE [23]	0.7081 (0.0297)	0.7074 (0.0395)	0.5177 (0.0239)	11.4635 (1.3339)
	REIQA [37]	0.7919 (0.0116)	0.8023 (0.0168)	0.6068 (0.0103)	7.9390 (0.3421)
NR-VQA	VBLIINDS [36]	0.4605 (0.0365)	0.4478 (0.0347)	0.3180 (0.0246)	11.9322 (0.4202)
	CONVIQT [24]	0.7026 (0.0462)	0.7202 (0.0510)	0.5134 (0.0431)	10.5817 (1.4206)
	VSFA [19]	0.7556 (0.0139)	0.7501 (0.0206)	0.5538 (0.0138)	8.8310 (0.1834)
	COVER [14]	0.7609 (0.0201)	0.7917 (0.0252)	0.5597 (0.0181)	7.7352 (0.3104)
	FasterVQA [51]	0.7744 (0.0162)	0.7625 (0.0147)	0.5763 (0.0152)	9.0680 (0.2501)
	DOVER [50]	0.7745 (0.0155)	0.8060 (0.0207)	0.5924 (0.0123)	7.4641 (0.2801)
	FastVQA [48]	0.8094 (0.0121)	0.8530 (0.0156)	0.6445 (0.0106)	7.1336 (0.2402)
NR-HDR-VQA	HDRChipQA [10]	0.6781 (0.0220)	0.6855 (0.0179)	0.4889 (0.0160)	9.5869 (0.3081)
	HIDROVQA [38]	<b>0.8526 (0.0217)</b>	<b>0.8620 (0.0136)</b>	<b>0.6680 (0.0215)</b>	<b>6.5708 (0.3367)</b>
<b>BrightRate</b>					
0.8887(0.0197) 0.8970(0.0171) 0.7059(0.0227) 5.7514(0.4465)					

### 4.5. Quality Regression

285

At each frame  $t$ , we concatenate the four feature types (UGC, temporal difference, semantic, and HDR) into a feature vector  $\mathbf{z}^t$ , with normalization to ensure balanced magnitudes. Averaging over  $T$  frames yields the clip descriptor  $\bar{\mathbf{z}} = \frac{1}{T} \sum_{t=1}^T \mathbf{z}^t$ . A Support Vector Regressor (SVR), known for its stable training and strong generalization ability, is employed as the regressor  $R(\cdot)$  to predict the MOS:

$$Q_i = R(\bar{\mathbf{z}}). \quad (8)$$

## 5. Experiment

295

### 5.1. Databases

296

We evaluated *BrightRate* on our newly introduced *BrightVQ* dataset, as well as on SFV+HDR [47] and LIVE-HDR [40]. For all datasets, we randomly split the videos into 80% training and 20% testing sets based on refer-

297

298

299

300

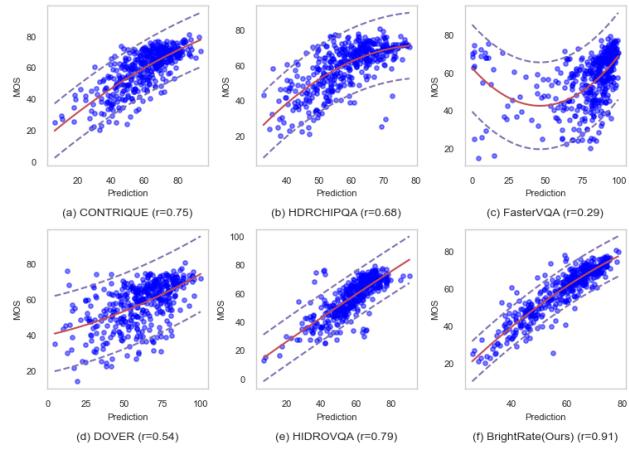


Figure 7. Scatter plots of actual MOS vs. predicted scores for various SOTA models on *BrightVQ*. Red curves show polynomial parametric fits.

ence content to ensure that all videos from the same source appeared in the same split [23, 37]. By contrast with UGC-VQA methods such as DOVER [50], KSVQE [22], Fast/Faster-VQA [48, 51], etc. that fine-tune the feature extraction backbones, we train only a lightweight regressor, preserving the generalization capabilities of the pre-trained modules.

## 5.2. Implementation Details

We use the CLIP image encoder (ViT-B32) [34] for semantic features and the CONTRIQUE model [23] at two scales to extract UGC distribution features. HDR features are extracted by applying an expansive non-linearity over a  $31 \times 31$  window with an expansion power of 4 [10, 40]. Temporal differences between consecutive CONTRIQUE [23] features are computed and averaged. The resulting normalized, concatenated clip-level descriptor is then fed into an SVR, optimized via 5-fold cross validation and evaluated as the median over 100 splits using PLCC, SROCC, RMSE, and KRCC [22, 23, 37, 38]. More implementation details in Supplementary Material Sec. D.

## 5.3. Experiment Results

### 5.3.1. Evaluation on BrightVQ Dataset

Table 3 shows that *BrightRate* consistently outperforms state-of-the-art methods on the *BrightVQ* dataset by an average of  $\approx 3\%$  across metrics, achieving the highest SROCC of 0.8887, PLCC of 0.8970, and KROCC of 0.7059, while maintaining the lowest RMSE of 5.7514. Notably, among existing NR-HDR-VQA methods, HIDROVQA [38] performed second-best, underscoring its ability to capture HDR-specific distortions. In the NR-VQA/NR-IQA category, although FastVQA [48] performs well among SDR-

oriented models, it is outperformed by HDR-specific approaches.

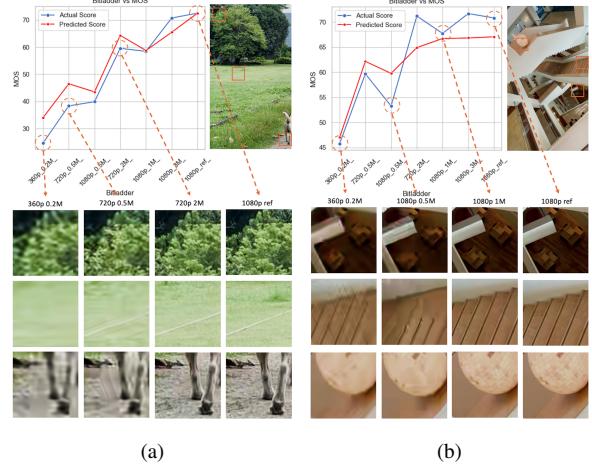


Figure 8. The combination of MOS vs. predicted score plots with visual comparisons of specific image regions to highlight the correlation between distortion and MOS across different bitrates and resolutions.

Fig. 7 compares predicted scores to actual MOS across several state-of-the-art methods on the *BrightVQ* dataset. Compared to other methods, *BrightRate* demonstrates a narrower distribution, indicating stronger alignment with subjective opinions. Fig. 8 illustrates MOS vs. predicted scores across different bitrates and resolutions on the *BrightVQ* dataset, highlighting the model’s ability to capture UGC and compression distortions. While the overall correlation is strong, deviations occur at lower bitrates where the model tends to overestimate quality. Visual comparisons further demonstrate how blurring, blocking, and texture loss degrade perceptual quality, especially in highly compressed videos. These results confirm *BrightVQ* dataset as a challenging benchmark for HDR-UGC VQA tasks.

### 5.3.2. Evaluation on existing HDR Datasets

Table 4 shows that *BrightRate* outperforms all existing models on both LIVE-HDR [40] and SFV+HDR [47], achieving the highest correlation against MOS. On LIVE-HDR [40], it improves SROCC and PLCC by approximately 1.3% and 1.5%, respectively, over the second-best model, demonstrating its effectiveness at capturing HDR-specific distortions. Similarly, on SFV+HDR [47], *BrightRate* outperforms by 2.6% in SROCC and 1.0% in PLCC, further confirming its robustness across different HDR datasets. Compared to SDR-oriented models, *BrightRate* achieves significantly higher correlations and reduces RMSE by a large margin, indicating its superior ability to handle both UGC and HDR content. These results validate the effectiveness of *BrightRate* in predicting HDR perceptual quality across diverse content and compression settings.

332  
333

334  
335  
336  
337

338  
339  
340  
341  
342  
343  
344  
345  
346  
347

348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363

Table 4. Performance Comparison on LIVE-HDR [40] and SFV+HDR [47] Datasets.

Method	LIVE-HDR				SFV+HDR			
	SROCC(↑)	PLCC(↑)	KRCC(↑)	RMSE(↓)	SROCC(↑)	PLCC(↑)	KRCC(↑)	RMSE(↓)
BRISQUE [26]	0.7251 (0.0955)	0.7139 (0.0881)	0.3424 (0.0579)	12.6404 (2.1651)	0.4664 (0.0846)	0.4186 (0.0628)	0.3165 (0.0646)	0.3811 (0.0321)
HDRMAX [40]	0.6308 (0.1214)	0.5088 (0.0911)	0.4509 (0.0962)	15.4146 (5.0564)	0.5371 (0.0654)	0.5463 (0.0660)	0.3821 (0.0529)	0.3495 (0.0170)
CONTRIQUE [23]	0.8170 (0.0672)	0.7875 (0.0705)	0.5876 (0.0420)	11.2514 (2.0548)	0.5901 (0.0450)	0.5959 (0.0455)	0.4204 (0.0330)	0.3368 (0.0264)
REIQA [37]	0.7196 (0.1634)	0.6883 (0.1191)	0.5197 (0.1208)	15.1653 (1.6896)	0.5822 (0.0669)	0.5998 (0.0367)	0.4145 (0.0499)	0.3072 (0.0275)
VBLIINDS [36]	0.7483 (0.1446)	0.7193 (0.1141)	0.2541 (0.1233)	12.7794 (2.3715)	0.3335 (0.1133)	0.2713 (0.1254)	0.2300 (0.0802)	0.3988 (0.0527)
CONVIQT [24]	0.7922 (0.0855)	0.8001 (0.0837)	0.6041 (0.0842)	11.9681 (1.9134)	0.5736 (0.0408)	0.6017 (0.0324)	0.4170 (0.0328)	0.3412 (0.0237)
DOVER [50]	0.6303 (0.0750)	0.6832 (0.0870)	0.4692 (0.0950)	17.0005 (2.0130)	0.6001 (0.0354)	0.6154 (0.1570)	0.4270 (0.0910)	0.5750 (0.0721)
COVER [14]	0.5022 (0.0848)	0.5013 (0.1508)	0.3731 (0.1447)	21.3297 (1.8020)	0.6613 (0.0557)	0.7048 (0.1103)	0.4705 (0.1802)	0.6831 (0.0577)
VSFA [19]	0.7127 (0.1079)	0.6918 (0.1114)	0.5760 (0.1469)	13.0511 (2.4003)	0.6449 (0.0704)	0.7233 (0.0449)	0.4783 (0.0646)	0.2911 (0.0347)
FasterVQA [51]	0.3385 (0.0505)	0.4114 (0.0850)	0.2282 (0.0443)	22.1425 (1.8504)	0.6948 (0.0905)	0.6889 (0.0755)	0.5089 (0.0390)	0.3081 (0.0225)
FastVQA [48]	0.5182 (0.0410)	0.5727 (0.0547)	0.3822 (0.0411)	18.8379 (1.3507)	0.7130 (0.0747)	0.7295 (0.0297)	0.5193 (0.0357)	0.7467 (0.0208)
HDRChipQA [10]	0.8250 (0.0589)	0.8344 (0.0562)	0.4501 (0.0500)	9.8038 (1.7334)	0.6296 (0.0734)	0.6508 (0.0316)	0.4440 (0.0475)	0.3271 (0.0231)
HIDROVQA [38]	0.8793 (0.0672)	0.8678 (0.0643)	0.6919 (0.0508)	8.8743 (1.7538)	0.7003 (0.0606)	0.7320 (0.0514)	0.5156 (0.0541)	0.2735 (0.0250)
<b>BrightRate</b>	<b>0.8907 (0.0425)</b>	<b>0.8824 (0.0470)</b>	<b>0.7178 (0.0492)</b>	<b>8.3955 (1.9260)</b>	<b>0.7328 (0.0509)</b>	<b>0.7709 (0.0252)</b>	<b>0.5415 (0.0496)</b>	<b>0.2679 (0.0236)</b>

Table 5. Cross Data Validation: Train on *BrightVQ*, Test on LIVE-HDR [40] and SFV+HDR [47].

Method	Test: LIVE-HDR				Test: SFV+HDR			
	SROCC(↑)	PLCC(↑)	RMSE(↓)	KRCC(↑)	SROCC(↑)	PLCC(↑)	RMSE(↓)	KRCC(↑)
BRISQUE [26]	0.4201 (0.1371)	0.4267 (0.1092)	16.6469 (1.4088)	0.2882 (0.0989)	0.2078 (0.1270)	0.1485 (0.1448)	54.0184 (0.9749)	0.1466 (0.0895)
HDRMAX [40]	0.1788 (0.0856)	0.2235 (0.0893)	17.6386 (1.1955)	0.1263 (0.0584)	0.4335 (0.1052)	0.4512 (0.1036)	54.3283 (0.8694)	0.3000 (0.0740)
CONTRIQUE [23]	0.5528 (0.0648)	0.5809 (0.0683)	15.2477 (1.0483)	0.3901 (0.0544)	0.4798 (0.0491)	0.5020 (0.0720)	54.0703 (1.2556)	0.3267 (0.0383)
REIQA [37]	0.4255 (0.1413)	0.4919 (0.0936)	15.7472 (1.0587)	0.2911 (0.1002)	0.4573 (0.0624)	0.4349 (0.0493)	52.5308 (0.9482)	0.3119 (0.0454)
CONVIQT [24]	<b>0.6240 (0.1197)</b>	<b>0.6112 (0.1075)</b>	<b>15.0850 (1.2097)</b>	<b>0.4331 (0.0935)</b>	<b>0.4981 (0.0391)</b>	0.5129 (0.0520)	<b>44.0703 (1.3564)</b>	<b>0.4267 (0.0281)</b>
VBLIINDS [36]	0.1524 (0.0823)	0.1520 (0.1416)	24.5003 (3.1376)	0.1194 (0.0594)	0.2949 (0.1330)	0.3871 (0.1441)	56.3203 (0.9249)	0.2072 (0.0928)
HDRChipQA [10]	0.3240 (0.1127)	0.3460 (0.1049)	17.4732 (1.8048)	0.2472 (0.0859)	0.2334 (0.1225)	0.1923 (0.1309)	56.7852 (1.4586)	0.1631 (0.0841)
VSFA [19]	0.4597 (0.1622)	0.4349 (0.1609)	17.4869 (2.0345)	0.3342 (0.1329)	0.4581 (0.0849)	<b>0.5404 (0.0899)</b>	51.4192 (1.3247)	0.3179 (0.0602)
HIDROVQA [38]	0.4086 (0.0915)	0.4918 (0.0924)	15.5255 (0.7523)	0.2886 (0.0728)	0.3398 (0.0491)	0.3020 (0.0720)	<b>24.0703 (1.2556)</b>	0.1267 (0.0383)
<b>BrightRate</b>	<b>0.7362 (0.0741)</b>	<b>0.7337 (0.0563)</b>	<b>15.1022 (0.8398)</b>	<b>0.5524 (0.0621)</b>	<b>0.5310 (0.0670)</b>	<b>0.5465 (0.0730)</b>	51.8795 (1.1711)	<b>0.3629 (0.0510)</b>

364

### 5.3.3. Cross-dataset Evaluation

365

We conducted two cross-dataset evaluations: “*BrightRate* dataset→other datasets” and “other datasets→*BrightRate* dataset” in Table 5 and Table 6. The cross-dataset evaluation results highlight *BrightVQ*’s strong generalization ability, as models trained on it perform well across different HDR datasets. While some models, such as CONVIQT [24] and HIDROVQA [38], achieve competitive results in certain metrics, *BrightRate*-trained models consistently demonstrate higher correlations against MOS and lower RMSE in most cases. Moreover, models trained on other datasets struggled to generalize effectively to *BrightVQ*, especially those trained on SFV+HDR [47], indicating its limited diversity in representing HDR distortions. These findings reinforce *BrightVQ*’s value as a robust and comprehensive benchmark for HDR VQA task.

380

### 5.4. Ablation Study

381

To assess the effectiveness of the components in our model, namely UGC Feature Extractor, Semantic Feature Extrac-

tion (CLIP), Temporal Difference Module (Temp), and HDR Feature Extraction (HDR)- we conducted an ablation study, with results present in Table 7 and 8. The baseline model is trained without these components, while our full proposed model integrates all these three components. The findings indicate that each module enhances performance, with the best results achieved when all three are combined.

383

384

385

386

387

388

389

**Effectiveness of CLIP Model:** Comparing the baseline to the model incorporating CLIP shown in Table 7, we observe significant improvement in SROCC (+0.135) and PLCC (+0.135) on the *BrightVQ* dataset, along with consistent gains across LIVE-HDR [40] and SFV+HDR [47]. This demonstrates that CLIP strengthens the model’s ability to extract meaningful semantic features relevant to video quality assessment.

390

391

392

393

394

395

396

397

**Effectiveness of Temporal Difference Module:** Incorporating the Temporal-Difference Module results in noticeable performance improvements across all datasets. As indicated in Table 7, adding temporal features increase SROCC (+0.088) and PLCC (+0.075) on *BrightVQ* dataset

398

399

400

401

402

Table 6. Cross Data Validation on *BrightVQ* Test Set. Columns under “Train: LIVE-HDR [40]” report metrics when training on LIVE-HDR [40] and testing on *BrightVQ*, while those under “Train: SFV+HDR [47]” report metrics when training on SFV+HDR [47] and testing on *BrightVQ*.

Method	Train: LIVE-HDR [40], Test: <i>BrightVQ</i>				Train: SFV+HDR [47], Test: <i>BrightVQ</i>			
	SROCC(↑)	PLCC(↑)	RMSE(↓)	KRCC(↑)	SROCC(↑)	PLCC(↑)	RMSE(↓)	KRCC(↑)
BRISQUE [26]	0.1411 (0.0778)	0.1420 (0.1052)	15.6298 (1.3612)	0.0971 (0.0515)	0.1388 (0.0820)	0.1486 (0.1023)	55.0998 (0.6600)	0.0890 (0.0554)
HDRMAX [40]	0.1176 (0.0480)	0.0489 (0.0524)	13.7844 (0.3274)	0.0777 (0.0344)	0.2302 (0.0415)	0.2451 (0.0461)	55.0761 (0.6489)	0.1565 (0.0298)
CONTRIQUE [23]	0.6392 (0.0196)	0.7135 (0.0204)	22.5554 (0.8008)	0.4588 (0.0154)	0.5538 (0.0241)	0.5248 (0.0258)	55.0629 (0.6549)	0.3780 (0.0194)
REIQA [37]	0.6056 (0.0232)	0.6119 (0.0168)	10.4203 (0.2005)	0.4196 (0.0173)	0.3995 (0.0378)	0.3554 (0.0461)	55.0996 (0.6411)	0.2850 (0.0276)
CONVIQT [24]	0.6563 (0.0650)	0.6858 (0.0642)	10.5680 (1.1620)	0.4744 (0.0492)	0.5294 (0.0392)	0.5387 (0.0394)	55.0753 (0.6525)	0.3637 (0.0302)
VBLIINDS [36]	0.1036 (0.0620)	0.0541 (0.0613)	13.5020 (0.2067)	0.0679 (0.0424)	0.2093 (0.0572)	0.1731 (0.0694)	55.0335 (0.2067)	0.1462 (0.0391)
HDRChipQA [10]	0.3817 (0.0503)	0.3811 (0.0703)	13.4357 (0.5963)	0.2652 (0.0353)	0.0523 (0.0687)	0.0382 (0.0606)	55.0512 (0.6565)	0.0334 (0.0460)
VSFA [19]	0.5770 (0.0577)	0.6066 (0.0550)	10.5367 (0.4795)	0.4104 (0.0471)	0.3551 (0.0448)	0.3361 (0.0494)	55.1327 (0.6452)	0.2425 (0.0339)
HIDROVQA [38]	0.6931 (0.0456)	0.7015 (0.0435)	12.9803 (0.8618)	0.4918 (0.0346)	0.5261 (0.0426)	0.5041 (0.0423)	55.1434 (0.6609)	0.3597 (0.0307)
<b>BrightRate</b>	<b>0.6669 (0.0346)</b>	<b>0.7459 (0.0373)</b>	<b>9.4324 (0.7087)</b>	<b>0.4806 (0.0260)</b>	<b>0.5892 (0.0249)</b>	<b>0.5308 (0.0263)</b>	<b>55.0568 (0.6537)</b>	<b>0.4004 (0.0204)</b>

Table 7. Ablation Study I: Effect of Modules on SROCC (↑) and PLCC (↑). Results are reported for *BrightVQ*, LIVE-HDR [40], and SFV+HDR [47] datasets.

Module/s	<b>BrightRate</b>		<b>LIVE-HDR</b>		<b>SFV+HDR</b>	
	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC
Baseline(CONTRIQUE)	0.7081	0.7074	0.7868	0.8016	0.5901	0.5959
+CLIP	<b>0.8431</b>	<b>0.8424</b>	<b>0.8325</b>	<b>0.8230</b>	<b>0.6403</b>	<b>0.6598</b>
+Temporal-Difference	0.7961	0.7821	0.8159	<b>0.8157</b>	0.6161	<b>0.6749</b>
+HDR	<b>0.8485</b>	<b>0.8489</b>	<b>0.8301</b>	0.8129	<b>0.6250</b>	0.6408

Table 8. Ablation Study II: Effect of Combinations of Modules on SROCC (↑) and PLCC (↑). Note: “Temp” here represents Temporal-Difference Module.

Module/s	<b>BrightRate</b>		<b>LIVE-HDR</b>		<b>SFV+HDR</b>	
	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC
+(CLIP+Temp)	0.8368	<b>0.8578</b>	0.8494	0.8276	0.6318	0.6894
+(HDR+Temp)	0.8389	0.8564	0.8510	<b>0.8319</b>	0.6773	0.6734
+(CLIP+HDR)	<b>0.8463</b>	0.8470	<b>0.8673</b>	0.8301	<b>0.6943</b>	<b>0.7032</b>
<b>BrightRate</b>	<b>0.8887</b>	<b>0.8970</b>	<b>0.8907</b>	<b>0.8824</b>	<b>0.7328</b>	<b>0.7709</b>

403 compared to the baseline, confirming its ability to capture  
404 temporal variations in HDR videos. The improvements on  
405 LIVE-HDR [40] and SFV+HDR [47] were relatively mod-  
406 est, suggesting that these two datasets may contain fewer  
407 temporal artifacts, making motion-aware learning less in-  
408 fluential.

#### Effectiveness of HDR Feature Extraction Module:

409 The HDR-specific feature extraction module enhances the  
410 model’s ability to detect distortions unique to HDR content.  
411 Comparing the baseline with the HDR Feature Extraction  
412 module in Table 7, we observe an SROCC increase (+0.140)  
413 and PLCC increase of (+0.141) on the *BrightVQ* dataset,  
414 emphasizing the module’s critical role in HDR quality as-  
415 sessment. The improvements extend to LIVE-HDR [40]  
416 (SROCC: 0.8301) and SFV+HDR [47] (SROCC: 0.6250),  
417 confirming that HDR-specific feature extraction is essential  
418

for accurate VQA performance.

**Effectiveness of Combining Components:** The combination results shown in Table 8 indicate that different combinations of CLIP, Temp, and HDR lead to varying degrees of improvement, highlighting the complementary roles of these components. CLIP+HDR achieves the highest performance in SROCC among two-component combinations across all datasets, demonstrating the strong synergy between semantic understanding and HDR-specific feature learning in assessing HDR video quality. Despite its relatively weaker impact compared to CLIP and HDR, Temp enhances overall performance when included in the overall model, particularly on LIVE-HDR [40] and SFV+HDR [47]. This confirms that while Temp alone is not the primary driver of performance, it refines and stabilizes predictions in dynamic scenes, making it a valuable addition in a comprehensive HDR video quality assessment framework. The best performance is achieved when all components—UGC, CLIP, Temp, and HDR—are combined, as this allows the model to leverage semantic understanding, HDR-aware distortion modeling, and temporal consistency with UGC features simultaneously.

## 6. Conclusion

In this paper, we introduce ***BrightVQ***, the first large-scale HDR-UGC video database, and ***BrightRate***, a novel no-reference VQA model for HDR-UGC content. *BrightVQ*, comprising 2,100 videos and 73,794 subjective ratings, offers a comprehensive benchmark for real-world HDR quality assessment. BrightRate fuses UGC distortion, semantic, HDR-specific (via expansive non-linearity), and temporal features to robustly predict quality scores. Extensive experiments on BrightVQ and other HDR datasets demonstrate its state-of-the-art performance. Our dataset and model are publicly available, providing a valuable resource for future research in HDR-UGC VQA.

**References**

- 454
- 455 [1] 99Firms. Facebook video statistics, 2024. [Online].  
456 [2] Naima Aamir, Junaid Mir, Imran Fareed Nizami, Furqan  
457 Shaukat, and Muhammad Majid. Hdr-bvqm: High dynamic  
458 range blind video quality model. *Multimedia Tools and Applications*, 80:27701 – 27715, 2021.  
459 [3] Sandhini Agarwal, Gretchen Krueger, Jack Clark, Alec Rad-  
460 ford, Jong Wook Kim, and Miles Brundage. Evaluating  
461 clip: towards characterization of broader capabilities and  
462 downstream implications. *arXiv preprint arXiv:2108.02818*,  
463 2021.  
464 [4] Apple Inc. Hls authoring specification for apple devices,  
465 2024. Accessed: Feb. 2024.  
466 [5] Maryam Azimi, Amin Banitalebi-Dehkordi, Yuanyuan  
467 Dong, Mahsa T. Pourazad, and Panos Nasiopoulos. Evaluat-  
468 ing the performance of existing full-reference quality metrics  
469 on high dynamic range (hdr) video content, 2018.  
470 [6] Vittorio Baroncini, Federica Mangiatordi, Emiliano Pallotti,  
471 and Massimiliano Agostinelli. Visual assessment of hdr  
472 video. 2016.  
473 [7] Yu-Chih Chen, Avinab Saha, Alexandre Chapiro, Christian  
474 Häne, Jean-Charles Bazin, Bo Qiu, Stefano Zanetti, Ioannis  
475 Katsavounidis, and Alan C. Bovik. Subjective and objec-  
476 tive quality assessment of rendered human avatar videos in  
477 virtual reality. *IEEE Transactions on Image Processing*, 33:  
478 5740–5754, 2024.  
479 [8] Joshua Peter Ebenezer, Zaixi Shang, Yongjun Wu, Hai  
480 Wei, Sriram Sethuraman, and Alan C Bovik. Chipqa:  
481 No-reference video quality prediction via space-time chips.  
482 *IEEE Transactions on Image Processing*, 30:8059–8074,  
483 2021.  
484 [9] Joshua P Ebenezer, Zaixi Shang, Yixu Chen, Yongjun Wu,  
485 Hai Wei, Sriram Sethuraman, and Alan C Bovik. Hdr or sdr?  
486 a subjective and objective study of scaled and compressed  
487 videos. *IEEE Transactions on Image Processing*, 2024.  
488 [10] Joshua P Ebenezer, Zaixi Shang, Yongjun Wu, Hai Wei,  
489 Sriram Sethuraman, and Alan C Bovik. Hdr-chipqa: No-  
490 reference quality assessment on high dynamic range videos.  
491 *Signal Processing: Image Communication*, 129:117191,  
492 2024.  
493 [11] FFmpeg Developers. Ffmpeg. <https://ffmpeg.org/>.  
494 Accessed: 2025-02-04.  
495 [12] Deepti Ghadiyaram, Janice Pan, Alan C. Bovik, Anush Kr-  
496 ishna Moorthy, Prasanjit Panda, and Kai-Chieh Yang. In-  
497 capture mobile video distortions: A study of subjective be-  
498 havior and objective algorithms. *IEEE Trans. Circuits Syst.*  
499 *Video Technol.*, 28(9):2061–2077, 2018.  
500 [13] Google Support. Recommended upload encoding settings,  
501 2024. Accessed: Feb. 2024.  
502 [14] Chenlong He, Qi Zheng, Ruoxi Zhu, Xiaoyang Zeng, Yibo  
503 Fan, and Zhengzhong Tu. Cover: A comprehensive video  
504 quality evaluator. In *2024 IEEE/CVF Conference on Com-*  
505 *puter Vision and Pattern Recognition Workshops (CVPRW)*,  
506 pages 5799–5809, 2024.  
507 [15] Vlad Hosu, Franz Hahn, Mohsen Jenadeleh, Hanhe Lin, Hui  
508 Men, Tamás Szirányi, Shujun Li, and Dietmar Saupe. The  
509 konstanz natural video database (konvid-1k). In *QoMEX*,  
510 pages 1–6. IEEE, 2017.  
511 [16] International Telecommunication Union. Methodology for  
512 the Subjective Assessment of the Quality of Television Pic-  
513 tures. Technical Report BT.500-14, International Telecom-  
514 munication Union, 2019.  
515 [17] ITU. Bt.2020 : Image parameter values for high dy-  
516 namic range television for use in production and interna-  
517 tional programme exchange., [https://www.itu.int/dms\\_pubrec/itu-r/rec/bt/R-REC-BT.2100-3-202502-I!!PDF-E.pdf](https://www.itu.int/dms_pubrec/itu-r/rec/bt/R-REC-BT.2100-3-202502-I!!PDF-E.pdf).  
518 [18] Jari Korhonen. Two-level approach for no-reference con-  
519 sumer video quality assessment. *IEEE Trans. Image Pro-  
520 cess.*, 28(12):5923–5938, 2019.  
521 [19] Dingquan Li, Tingting Jiang, and Ming Jiang. Quality as-  
522 sessment of in-the-wild videos. In *ACM Multimedia*, pages  
523 2351–2359. ACM, 2019.  
524 [20] Yang Li, Shengbin Meng, Xinfeng Zhang, Shiqi Wang, Yue  
525 Wang, and Siwei Ma. UGC-VIDEO: perceptual quality as-  
526 sessment of user-generated videos. In *3rd IEEE Confer-  
527 ence on Multimedia Information Processing and Retrieval,*  
528 *MIPR 2020, Shenzhen, China, August 6–8, 2020*, pages 35–  
529 38. IEEE, 2020.  
530 [21] Zhi Li, Christos G. Bampis, Lucjan Janowski, and Ioannis  
531 Katsavounidis. A simple model for subject behavior in sub-  
532 jective experiments. In *Electronic Imaging*, pages 131–1–  
533 131–14, 2020.  
534 [22] Yiting Lu, Xin Li, Yajing Pei, Kun Yuan, Qizhi Xie, Yunpeng  
535 Qu, Ming Sun, Chao Zhou, and Zhibo Chen. Kvq: Kwai  
536 video quality assessment for short-form videos. In *Pro-  
537 ceedings of the IEEE/CVF Conference on Computer Vision and*  
538 *Pattern Recognition*, pages 25963–25973, 2024.  
539 [23] Pavan C. Madhusudana, Neil Birkbeck, Yilin Wang, Balu  
540 Adsumilli, and Alan C. Bovik. Image quality assessment  
541 using contrastive learning. *IEEE Trans. Image Process.*, 31:  
542 4149–4161, 2022.  
543 [24] Pavan C Madhusudana, Neil Birkbeck, Yilin Wang, Balu  
544 Adsumilli, and Alan C Bovik. Convigt: Contrastive video  
545 quality estimator. *IEEE Transactions on Image Processing*,  
546 32:5138–5152, 2023.  
547 [25] RK Mantiuk and M Azimi. Pu21: A novel perceptually uni-  
548 form encoding for adapting existing quality metrics for hdr.  
549 2021.  
550 [26] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad  
551 Bovik. No-reference image quality assessment in the spa-  
552 tial domain. *IEEE Transactions on Image Processing*, 21  
553 (12):4695–4708, 2012.  
554 [27] Anish Mittal, Rajiv Soundararajan, and Alan C. Bovik. Mak-  
555 ing a “completely blind” image quality analyzer. *IEEE Sig-*  
556 *nal Processing Letters*, 20(3):209–212, 2013.  
557 [28] Anish Mittal, Michele A. Saad, and Alan C. Bovik. A com-  
558 pletely blind video integrity oracle. *IEEE Trans. Image Pro-  
559 cess.*, 25(1):289–300, 2016.  
560 [29] Maryam Mohsin. 10 youtube statistics every marketer should  
561 know in 2020, 2020. [Online].  
562 [30] Manish Narwaria, Matthieu Perreira Da Silva, and Patrick  
563 Le Callet. Hdr-vqm: An objective quality measure for high  
564

- 567 dynamic range video. *Signal Processing: Image Communi- 623  
568 cation*, 35:46–60, 2015. 624  
569 [31] Mikko Nuutinen, Toni Virtanen, Mikko Vaahteranoksa, Tero 625  
570 Vuori, Pirkko Oittinen, and Jukka Häkkinen. CVD2014 - 626  
571 A database for evaluating no-reference video quality assess- 627  
572 ment algorithms. *IEEE Trans. Image Process.*, 25(7):3073– 628  
573 3086, 2016. 629  
574 [32] Omnicore. Tiktok by the numbers, 2024. [Online]. 630  
575 [33] Xiaofei Pan, Jiaqi Zhang, Shanshe Wang, Shiqi Wang, Yun 631  
576 Zhou, Wenhua Ding, and Yahui Yang. Hdr video quality 632  
577 assessment: Perceptual evaluation of compressed hdr video. 633  
578 *Journal of Visual Communication and Image Representation*, 634  
579 57:76–83, 2018. 635  
580 [34] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya 636  
581 Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, 637  
582 Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen 638  
583 Krueger, and Ilya Sutskever. Learning transferable visual 639  
584 models from natural language supervision. In *ICML*, pages 640  
585 8748–8763. PMLR, 2021. 641  
586 [35] Martin Rerabek, Philippe Hanhart, Pavel Korshunov, and 642  
587 Touradj Ebrahimi. Subjective and objective evaluation of 643  
588 hdr video compression. In *9th International Workshop on Video 644  
589 Processing and Quality Metrics for Consumer Electronics 645  
590 (VPQM)*, 2015. 646  
591 [36] Michele A. Saad, Alan C. Bovik, and Christophe Charrier. 647  
592 Blind prediction of natural video quality. *IEEE Trans. Image 648  
593 Process.*, 23(3):1352–1365, 2014. 649  
594 [37] Avinab Saha, Sandeep Mishra, and Alan C. Bovik. Re-iqa: 650  
595 Unsupervised learning for image quality assessment in 651  
596 the wild. In *IEEE/CVF Conference on Computer Vision and 652  
597 Pattern Recognition, CVPR 2023, Vancouver, BC, Canada, June 653  
598 17-24, 2023*, pages 5846–5855. IEEE, 2023. 654  
599 [38] Shreshth Saini, Avinab Saha, and Alan C Bovik. Hidro-vqa: 655  
600 High dynamic range oracle for video quality assessment. In 656  
601 *Proceedings of the IEEE/CVF Winter Conference on Applica- 657  
602 tions of Computer Vision*, pages 469–479, 2024. 658  
603 [39] Zaixi Shang, Yixu Chen, Yongjun Wu, Hai Wei, and Sri- 659  
604 ram Sethuraman. Subjective and objective video quality 660  
605 assessment of high dynamic range sports content. In *Proceed- 661  
606 ings of the IEEE/CVF Winter Conference on Applications of 662  
607 Computer Vision*, pages 556–564, 2023. 663  
608 [40] Zaixi Shang, Joshua P Ebenezer, Abhinau K Venkataraman, 664  
609 Yongjun Wu, Hai Wei, Sriram Sethuraman, and Alan C Bovik. 665  
610 A study of subjective and objective quality 666  
611 assessment of hdr videos. *IEEE Transactions on Image 667  
612 Processing*, 33:42–57, 2023. 668  
613 [41] Zeina Sinno and Alan Conrad Bovik. Large-scale study of 669  
614 perceptual video quality. *IEEE Trans. Image Process.*, 28(2): 670  
615 612–627, 2019. 671  
616 [42] SMPTE. High dynamic range electrooptical transfer function 672  
617 of mastering reference displays. [https://pub.  
618 smpste.org/latest/st2084/st2084-2014.pdf](https://pub.smpste.org/latest/st2084/st2084-2014.pdf). 673  
619 [43] Zhengzhong Tu, Xiangyu Yu, Yilin Wang, Neil Birkbeck, 674  
620 Balu Adsumilli, and Alan C Bovik. Rapique: Rapid and 675  
621 accurate video quality prediction of user generated content. 676  
622 *IEEE Open Journal of Signal Processing*, 2:425–440, 2021. 677  
623

679 

## Supplementary Material Outline

680 This supplementary material is organized as follows:  
 681 • **Appendix A:** An overview of UGC and HDR video qual-  
 682 ity assessment challenges.  
 683 • **Appendix B.1:** Comprehensive details on video collec-  
 684 tion, filtering, transcoding, and the bitrate ladder.  
 685 • **Appendix B.2:** Full description of our AMT study, in-  
 686 cluding instructions, screening procedure, and rejection  
 687 criteria.  
 688 • **Appendix B.3:** Analysis of inter-subject consistency and  
 689 SUREAL-based MOS estimation.  
 690 • **Appendix C:** Detailed examination of luminance, color-  
 691 fullness, and spatial-temporal characteristics.  
 692 • **Appendix D:** Additional technical specifics on resizing,  
 693 normalization, and training.  
 694 • **Appendix E:** Extended results, ablation studies, and fail-  
 695 ure case analyses.

696 

## A. Background

697 The explosion of UGC on platforms such as YouTube,  
 698 Facebook, Instagram, and TikTok has transformed video  
 699 streaming into a ubiquitous, user-driven experience, gener-  
 700 ating billions of daily views [1, 29, 32]. However, di-  
 701 verse distortion patterns arising from user editing, compres-  
 702 sion, and platform-specific processing complicate quality  
 703 assessment [22, 55]. High Dynamic Range (HDR) imag-  
 704 ing, supported by mainstream platforms and devices, of-  
 705 fers enhanced visual experiences through a broader lumi-  
 706 nance and color range. Unlike Standard Dynamic Range  
 707 (SDR) videos, which are limited to 0.1 to 100  $cd/m^2$ , HDR  
 708 can represent luminance from  $10^{-4}$  to  $10^4 cd/m^2$  [17].  
 709 HDR10, a widely adopted format, supports 10-bit color  
 710 depth and Rec. 2020 color gamut (covering 75.8% of the  
 711 CIE 1931 color space), providing higher peak luminance,  
 712 improved color accuracy, and more detail in both shadows  
 713 and highlights, offering a richer, more immersive viewing  
 714 experience than SDR. The transition to HDR for UGC poses  
 715 challenges for Video Quality Assessment (VQA) due to in-  
 716 creased bit depth, broader luminance range, and complex  
 717 electro-optical transfer functions (EOTFs) like SMPTE ST  
 718 2084 [42]. Traditional SDR-based models fail to capture  
 719 these HDR-specific features and the variability of distor-  
 720 tions from different devices and editing techniques, thereby  
 721 impeding effective quality prediction. Furthermore, the ab-  
 722 sence of a publicly available HDR-UGC database has lim-  
 723 ited the development of HDR-specific VQA models.

724 

## B. Details of Dataset Construction

725 Fig. 9 provides an overview of the entire dataset preparation  
 726 pipeline. This multi-stage process guarantees that *BrightVQ*  
 727 reflects authentic HDR-UGC content with diverse distor-  
 728 tions

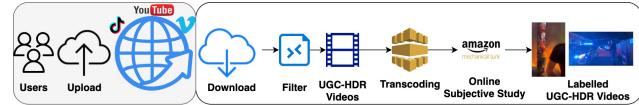


Figure 9. Overview of the dataset preparation approach.

729 

### B.1. Video data Collection



Figure 10. HDR specific challenges, and transcoding (on top of ugc) and ugc challenges in *BrightVQ*.

Table 9. Bitladder used for dataset creation. Each video was encoded at multiple bitrates to simulate real-world streaming conditions<sup>†</sup>.

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

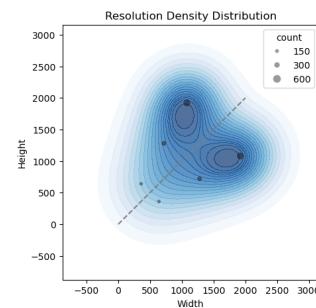


Figure 11. Resolution distribution of *BrightVQ* dataset, main-  
 taining a balanced mix of landscape and portrait videos to study  
 orientation-based perceptual differences.

<sup>†</sup>Based on YouTube’s streaming guidelines [13] and Apple’s HLS au-  
 thoring specifications [4].

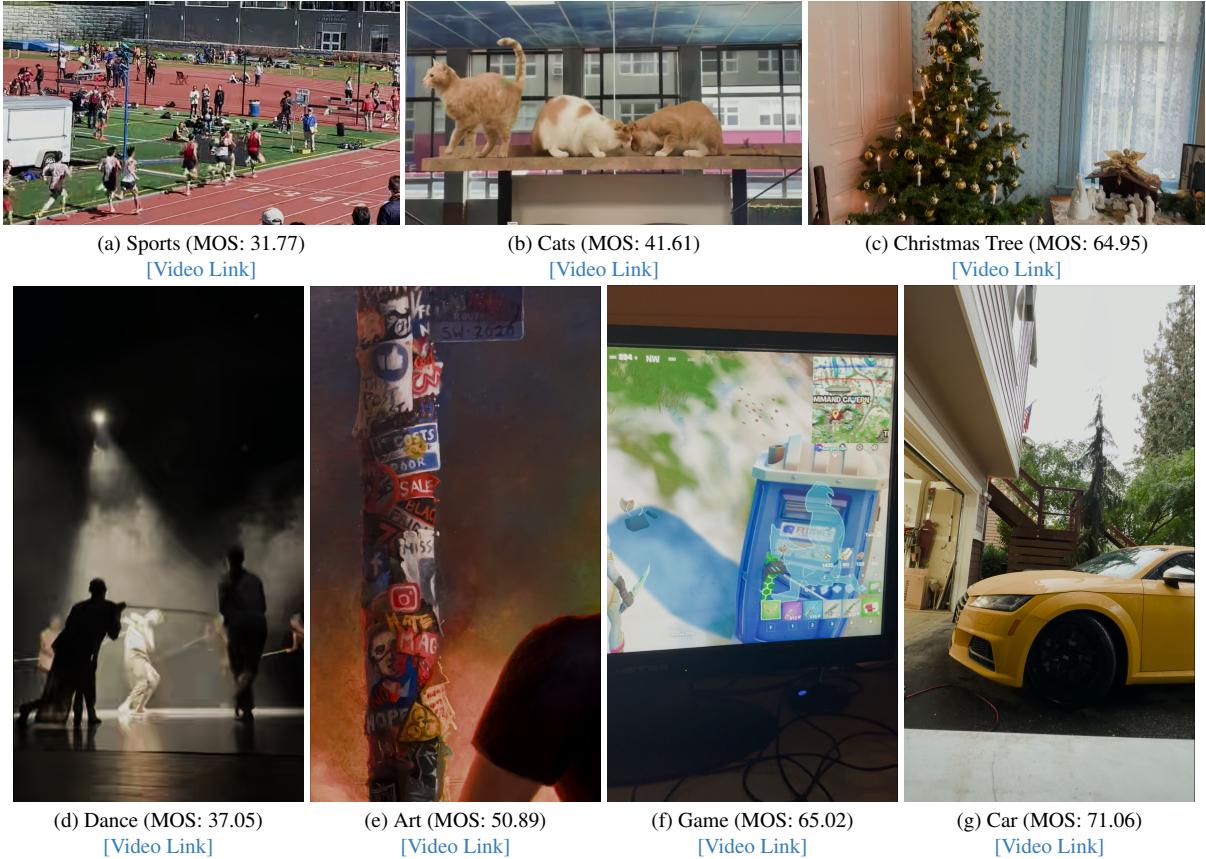


Figure 12. Example frames from *BrightVQ* dataset. Each frame is presented with its category, the MOS for the video and a direct video access link.

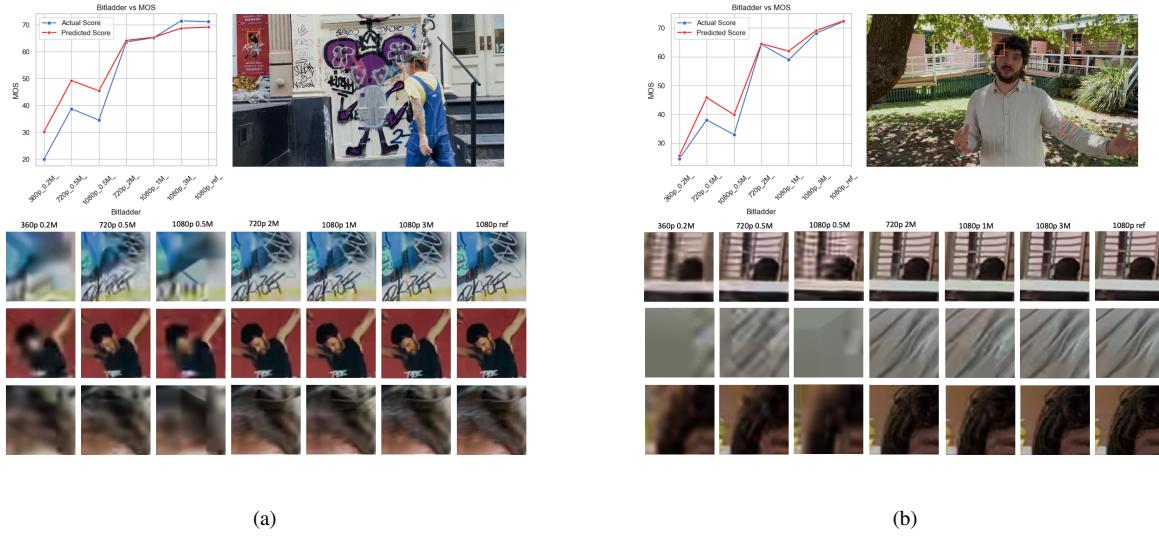


Figure 13. The combination of MOS vs. predicted score plots with visual comparisons of specific image regions to highlight the correlation between distortion and MOS across different bitrates and resolutions.

730 HDR-UGC videos were collected from Vimeo under  
 731 Creative Commons licenses to ensure open-source accessibility.  
 732 An initial pool of over 10,000 videos was automatically filtered using metadata checks for HDR flags, resolution,  
 733 format consistency, and common categories to remove  
 734 duplicates and professionally produced content. This was  
 735 followed by a rigorous manual inspection to verify authentic  
 736 UGC. Given the nature of UGC, the dataset includes an  
 737 equal mix of landscape and portrait videos. Fig. 12 shows  
 738 randomly selected frames from *BrightVQ*, illustrating the  
 739 diversity in video sizes and aspect ratios. This diversity  
 740 highlights the broad representation of UGC content in terms  
 741 of resolution, aspect ratios, and distortions.

742 Each selected video was truncated to a maximum of 10  
 743 seconds using `ffmpeg` [11] and maintained at up to 1080p  
 744 resolution. To simulate the viewing experience on social  
 745 media platforms, where videos are often transcoded, we ap-  
 746 plied a bitrate ladder inspired by industry standards [4, 13]  
 747 to create the final dataset. Tab. 9 shows the resolution and  
 748 bitrates used in this bit ladder. The filtered videos were  
 749 then transcoded following this bitrate ladder to simulate  
 750 real-world streaming conditions, ensuring a diverse range  
 751 of compression artifacts. To explore the impact of bitrate  
 752 selection on perceived video quality, Fig. 14 presents the  
 753 MOS variations across different bitrate ladders, separately  
 754 analyzing landscape and portrait videos. The box plot rep-  
 755 resentation highlights the diversity in perceptual quality rat-  
 756 ings across different encoding configurations, showing how  
 757 bitrate and resolution choices affect MOS scores. Fig. 11  
 758 illustrates the resolution density distribution of videos in  
 759 the *BrightVQ* dataset. Fig. 13 further visualizes the com-  
 760 pression artifacts introduced through this approach. This  
 761 multi-stage process guarantees that *BrightVQ* reflects au-  
 762 thentic HDR-UGC content with diverse distortions.

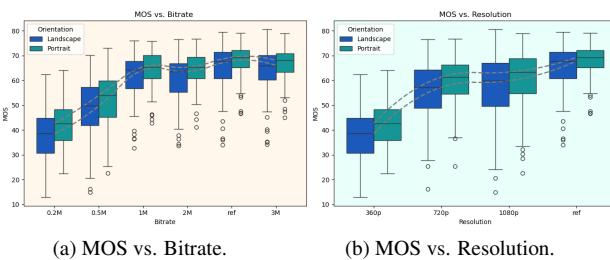


Figure 14. (a) and (b) show the MOS variations across bitrate and resolution respectively for *BrightVQ*.

## 764 B.2. Crowdsourced Subjective Study

765 We employed Amazon Mechanical Turk (AMT) to collect  
 766 human quality ratings for our HDR-UGC videos, adapting  
 767 protocols from previous studies [7, 44]. This is the first  
 768 large-scale HDR-UGC study conducted on AMT, address-  
 769 ing challenges associated with remote HDR evaluation. To

770 ensure data reliability, we implemented a rigorous multiple-  
 771 stage filtering process.

772 The general instruction of this study on AMT is illus-  
 773 trated in Fig. 15. Initially, subjects were presented with de-  
 774 tailed instructions and a comprehension quiz (Fig. 16) to  
 775 confirm their understanding of the rating process. Only  
 776 those with HDR-capable displays, verified through auto-  
 777 mated dynamic checks for bit depth, codec support, and  
 778 display resolution, were allowed to proceed. Before en-  
 779 tering the main study, subjects completed a training phase  
 780 where they rated six HDR videos to familiarize themselves  
 781 with the interface (Fig. 17). The testing phase followed, in  
 782 which each participant rated 94 videos using a 0–100 Lik-  
 783 er scale (rating method shown in Fig. 18). To ensure rating  
 784 consistency, we embedded five golden set videos and five  
 785 duplicate videos within the test set. Ethical considera-  
 786 tions are provided in Fig. 19.

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

- **During Instructions:** Participants with incompatible devices were disqualified.
- **During Training:** Continuous HDR and device checks ensured that participants did not switch displays mid-task. Those with incomplete downloads or playback manipulations 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 duplicate/golden set videos (deviations exceeding 20–25%) were removed.

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

### 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 15. 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 16. 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 17. 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 paused 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 them 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 obscure, 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.

**Next**

Figure 18. Rating instructions on AMT.

802

**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 Pearson Linear Correlation Coefficient (PLCC) between them.

## 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 19. Ethics policy on AMT.

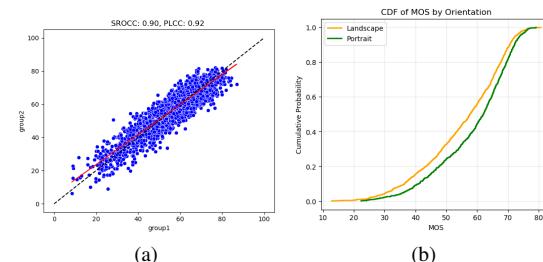


Figure 20. (a). Inter-subject correlation. (b). MOS CDF distribution of all videos in BrightVQ.

As shown in Fig. 20a, 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, training phases, and golden set validation to ensure consistent and reliable subjective ratings.

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

We computed Mean Opinion Scores (MOS) using the SUREAL method [21], which refines traditional MOS computation by accounting for individual subject bias and rating inconsistency. Traditional MOS calculations typically employ a hard rejection approach, where raters failing predefined consistency criteria (e.g., ITU-R BT.500-14 outlier

807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837

838 detection [16]) are completely excluded from the analysis.  
 839 However, this method discards potentially useful data  
 840 and does not account for varying levels of rating reliability  
 841 among retained subjects. SUREAL takes a soft rejection  
 842 approach by modeling each rating probabilistically. Each  
 843 rating  $S_{ij}$  from subject  $i$  for video  $j$  is modeled as:

$$844 \quad S_{ij} = \psi_j + \Delta_i + \nu_i X, \quad X \sim \mathcal{N}(0, 1), \quad (9)$$

845 where  $\psi_j$  represents the true quality of video  $j$ ,  $\Delta_i$  captures  
 846 the bias of subject  $i$ , and  $\nu_i$  reflects the rating inconsistency  
 847 of subject  $i$ . The parameters are estimated using Maxi-  
 848 mum Likelihood Estimation (MLE), maximizing the like-  
 849 lihood. Unlike hard rejection, which entirely removes out-  
 850 liers, SUREAL downweights ratings from less reliable sub-  
 851 jects. This ensures that MOS values reflect true perceptual  
 852 quality while mitigating distortions from inconsistent raters.  
 853 By applying SUREAL, we obtained more stable MOS es-  
 854 timates, which accurately reflect the perceptual quality of  
 855 HDR content across diverse video conditions. The CDF dis-  
 856 tribution of all videos in *BrightVQ* are shown in Fig. 20b.

### 857 C. Analysis of the HDR content

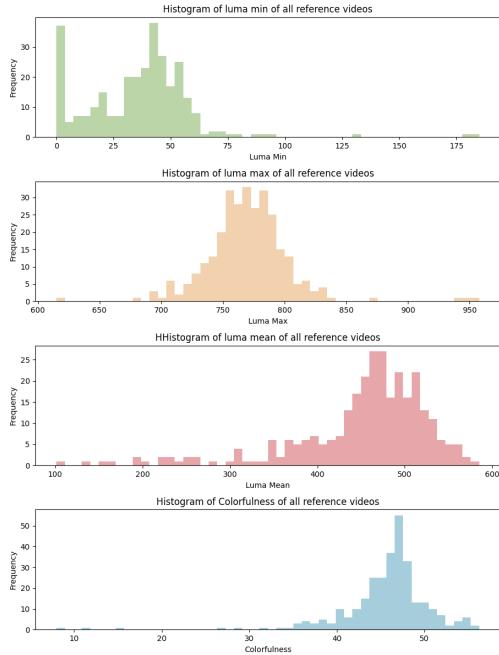


Figure 21. Distribution of luma and colorfulness of the source HDR-UGC videos in *BrightVQ* dataset.

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

862 Fig. 21 presents the distribution of luma and colorfulness  
 863 across the 300 source HDR videos in BrightQA. The first

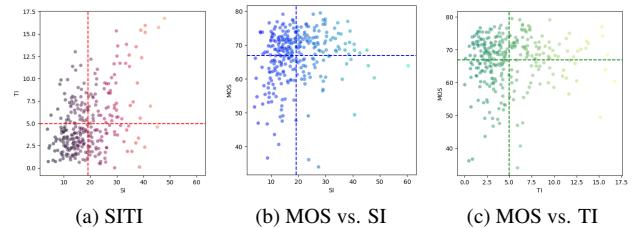


Figure 22. (a) Spatial-Temporal Complexity, (b) MOS vs. Spatial Information (SI), and (c) MOS vs. Temporal Information (TI).

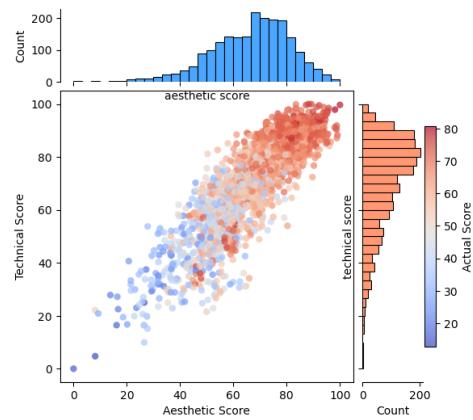


Figure 23. Diversity in aesthetic and technical quality scores in *BrightVQ* dataset.

864 three histograms illustrate the minimum, maximum, and  
 865 mean luma values, highlighting the variation in brightness  
 866 levels across different videos. This demonstrates that the  
 867 dataset includes both dark and bright HDR scenes, ensur-  
 868 ing a wide dynamic range. The fourth histogram shows the  
 869 colorfulness distribution, reflecting variations in chromatic  
 870 intensity across different videos.

871 To further quantify the diversity in content complex-  
 872 ity, Fig. 22 presents an analysis of spatial-temporal com-  
 873 plexity, spatial information (SI), and temporal informa-  
 874 tion (TI) within the dataset. The scatter plots in Fig. 22 (a)-(c)  
 875 demonstrate the variability in SI and TI values, showing a  
 876 wide distribution of motion and texture complexity across  
 877 the dataset. Higher SI values typically correspond to de-  
 878 tailed textures and sharp edges, while higher TI values in-  
 879 dicate rapid motion or dynamic scenes. The dataset covers  
 880 both high-detail static scenes and fast-moving dynamic con-  
 881 tent, ensuring its suitability for evaluating compression and  
 882 HDR features across different motion characteristics.

883 Fig. 23 demonstrates the diversity of BrightVQ dataset  
 884 in both aesthetic and technical aspects. The scatter plot  
 885 shows a wide range of ratings, with each point represent-  
 886 ing a video and color-coded by its actual subjective quality  
 887 score. The marginal histograms further highlight the dis-  
 888 tribution of scores, illustrating the broad variation in per-

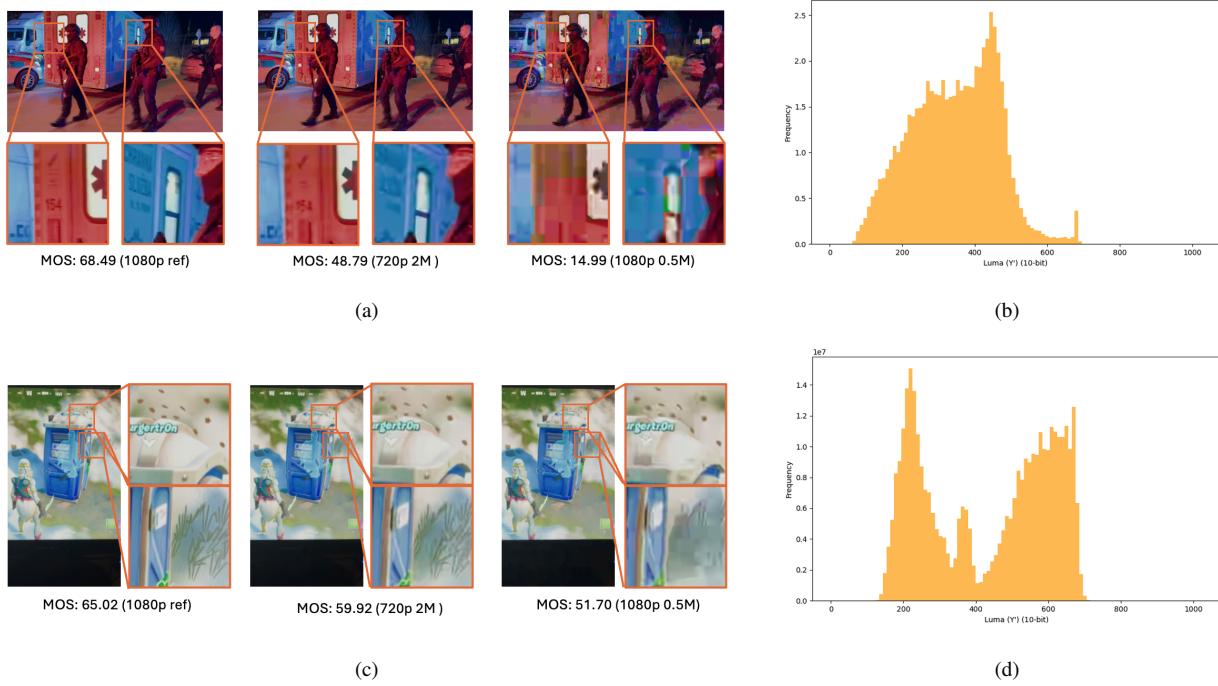


Figure 24. Illustrations of HDR content under different resolutions and bitrates. (a) and (c) show reference frames at 1080p and their lower-resolution, lower-bitrate counterparts, with red boxes highlighting high luma areas with artifacts (e.g., blocking, color banding) become more pronounced. The corresponding 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.

ceptual and technical quality across different content. The BrightVQ dataset presents a diverse range of HDR-UGC content, covering natural landscapes, indoor scenes, and various complex lighting conditions, capturing both HDR-specific and UGC-specific distortions. This diversity ensures that BrightVQ provides a comprehensive and realistic benchmark for evaluating video quality.

Figure 24 provides examples of HDR videos subjected to various spatial resolutions and bitrates, along with their MOS and luminance histograms. In (a) and (c), the higher-resolution, higher-bitrate frames retain more details 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  $224 \times 224$  and passed through the CLIP image encoder (ViT-B32) [34], yielding high-level semantic representations. UGC features are extracted using CONTRIQUE [23] at two scales: the original frame and a half-resolution version following the original implementation [23]. 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  $31 \times 31$  window with  $\beta = 4$  [10, 40]. 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 [23] features:

$$\Delta \mathcal{U}^t = |\mathcal{U}^t - \mathcal{U}^{t-1}|, \quad (10)$$

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:

$$\mathbf{z} = \text{Norm}(\overline{\mathcal{U}} \oplus \overline{\mathcal{E}} \oplus \overline{\mathcal{H}}). \quad (11)$$

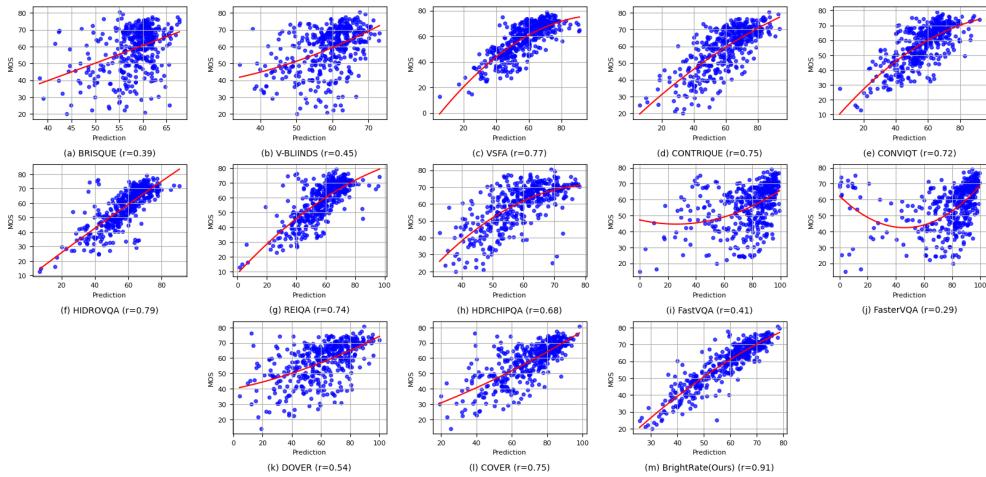


Figure 25. Scatter plots of actual MOS vs. predicted scores for 13 methods evaluated on *BrightVQ*, with parametric fits  $l(s)$  in red. A tighter clustering around the diagonal curve indicates a stronger alignment with subjective opinions. Methods yielding narrower scatter demonstrate higher predictive accuracy and consistency, underscoring their ability to capture the underlying perceptual quality cues.

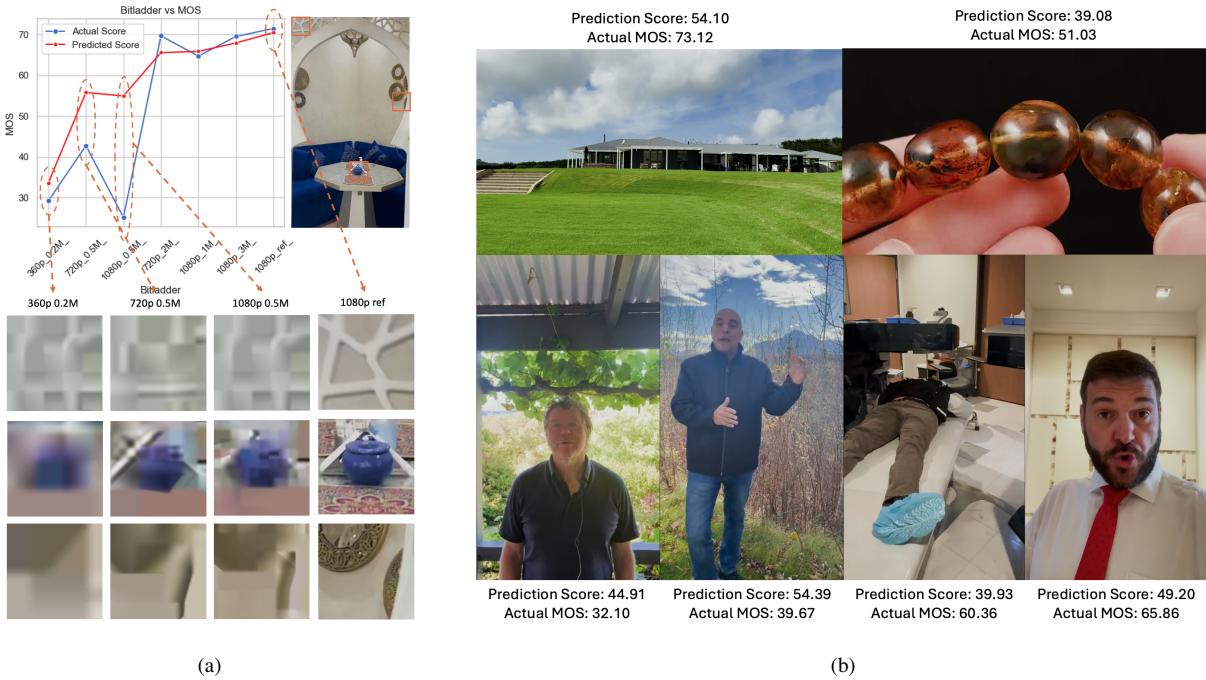


Figure 26. Failure cases in *BrightRate* predictions.

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

$$936 \quad \mathcal{L} = \frac{1}{N} \sum_{i=1}^N \left( Q_i - \hat{Q}_i \right)^2, \quad (12)$$

937 where  $Q_i$  and  $\hat{Q}_i$  denote the ground-truth and predicted  
 938 MOS, respectively. Only the regressor is trained, while  
 939 the feature extraction modules remain fixed. These imple-  
 940 mentation details ensure a robust and efficient extraction of  
 941 multi-scale UGC, semantic, and HDR features, enabling ac-  
 942 curate quality prediction on HDR-UGC videos.

**943 E. More Experimental Results**

944 To assess the effectiveness of existing No-Reference Video  
945 Quality Assessment models on the *BrightVQ* dataset, we  
946 conducted a comprehensive evaluation of multiple state-of-  
947 the-art methods. Fig. 25 expands upon Fig. 9, which pre-  
948 sented results for only six models, by providing a more ex-  
949 tensive comparison across 13 NR-VQA model. The scatter  
950 plots compare predicted scores vs. MOS, with red paramet-  
951 ric fitting lines highlighting the correlation trends, while the  
952 Pearson correlation coefficient ( $r$ ) quantifies each model's  
953 predictive performance.

954 Among the evaluated models, some traditional hand-  
955 crafted feature-based approaches exhibit limited correlation  
956 with MOS, highlighting their challenges in capturing the  
957 complexity of HDR-specific distortions in UGC content.  
958 Deep learning-based methods show stronger performance,  
959 with several achieving a higher degree of correlation by  
960 leveraging learned features and spatiotemporal representa-  
961 tions. Moreover, HDR-specific VQA models generally out-  
962 perform generic NR-VQA methods, demonstrating the im-  
963 portance of HDR-aware architectures in perceptual quality  
964 assessment. Our proposed *BrightRate* model achieves the  
965 highest correlation ( $r = 0.91$ ), significantly outperforming  
966 other approaches. The scatter plot for *BrightRate* shows  
967 a strong linear relationship between predicted scores and  
968 MOS, indicating its high accuracy and reliability in eval-  
969 uating HDR video quality.

970 Fig. 26 presents several failure cases where the pre-  
971 dicted video quality scores deviate significantly from the ac-  
972 tual MOS. These discrepancies highlight limitations in the  
973 model's ability to accurately predict perceptual quality un-  
974 der certain conditions. Fig. 26 (a) illustrates cases where  
975 low-resolution, highly compressed videos received higher-  
976 than-expected predictions. The close-up patches of com-  
977 pressed video artifacts reveal that blockiness and blurring  
978 effects are not always adequately penalized by the model,  
979 leading to overestimated quality scores in severely com-  
980 pressed videos. Fig. 26 (b) show video screenshots with  
981 complex textures, reflections, or dynamic lighting, where  
982 the model struggles to properly assess fine details and HDR  
983 characteristics. In videos with human subjects, facial ex-  
984 pressions, lighting conditions, or background complexity  
985 may lead to misinterpretations of perceptual quality by the  
986 model. These failure cases highlight the need for further re-  
987 finement in *BrightRate*'s HDR-aware feature extraction and  
988 compression robustness, ensuring improved alignment with  
989 human perception.