BrightRate: Quality Assessment for User-Generated HDR Videos

Anonymous ICCV submission

Paper ID 5421

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

*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 perceptual 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.*

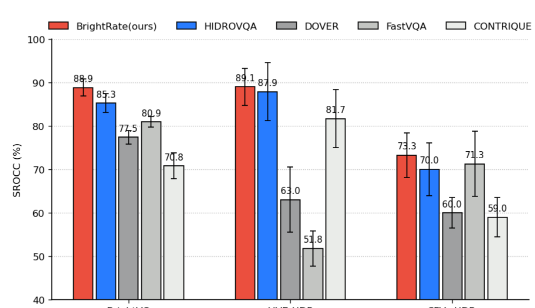


Figure 2. Benchmark performance of *BrightRate*(our) and other leading state-of-the-art (SOTA) VQA models on available HDR VQA datasets.

*Corresponding Author.
†Project Lead.

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

Attribute	Details
Video Specifications	
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
Dataset Statistics	
Reference Videos	300 (150 Landscape, 150 Portrait)
Total Videos	2100
Total Scores	73,794
Avg. Scores/Video	35

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, 28, 31]. However, the diverse distortion patterns from inexpert capture, editing, compressing, and platform-specific processing complicate quality assessment [21, 52]. High Dynamic Range (HDR) imaging further enhances visual experiences with broader luminance and color gamuts as compared to Standard Dynamic Range (SDR) [8, 39]. For example, HDR10 supports 10-bit depth and Rec. 2020 color gamut, delivering superior detail in shadows and highlights [16]. Despite significant advances in VQA [2, 8, 9, 29, 37–39], 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.

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 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 [39].
- 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 [11, 30, 40, 49, 51, 52] capture real-world distortions falling into two categories: In-the-Wild UGC Datasets[14, 40, 43, 47, 49–51], which contain naturally distorted videos but lack control over degradation types, and Simulated UGC Distortion Datasets[11, 19, 30, 52], 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 [32] were small and had restricted availability, while others [6, 34] lacked HDR10 compliance. LIVE-HDR [39] introduced a professionally generated HDR dataset but contains only 31 video contents, limiting its relevance for UGC scenarios. More recently, Wang et al.[44] 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 [7, 17, 25–27, 35] extract powerful distortion-aware statistical and perceptual features but struggle with complex UGC distortions. Deep learning-based models leverage pre-

*Based on YouTube's streaming guidelines [12] and Apple's HLS authoring specifications [4].

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

Dataset	Format	Total Videos (Ref.)	Source	Total Opinions	Orientation	Subjective Study
LIVE-HDR [39]	Rec. 2020, HDR10, PGC	310 (31)	Internet Archive	2,400	Landscape	In-Lab
SFV+HDR [44] (only HDR)	Rec. 2020, HDR10, UGC	300 (300)	YouTube	N/A	Portrait	In-Lab
BrightVQ (Ours)	Rec. 2020, HDR10, UGC	2100 (300)	Vimeo	73,794	Portrait+Landscape	Crowdsourced

trained networks to extract semantic and perceptual features. Among these, for example, VSFA[18] captures temporal variations, FAST/FASTER-VQA[45, 46] uses Transformers, CONTRIQUE[22] applies self-supervised learning, and DOVER[47] integrates aesthetic and technical quality assessment. However, most models, including these, are optimized for SDR and fail to handle HDR-specific distortions. HDR-VQM[29] and HDR-BVQM[2] introduce brightness-aware features but must rely on reference videos or lack HDR-specific adaptations. PU21[24] refines traditional metrics with perceptually uniform encoding but remains content- and display-dependent. HDR-ChipQA[8] extends ChipQA with non-linear luminance transformations, while HIDRO-VQA [37] trains CONTRIQUE [22] on unlabeled HDR videos from YouTube. However, none is able to effectively capture HDR-UGC distortions, limiting their applicability for HDR-UGC quality prediction.

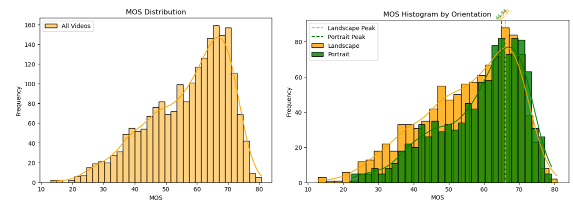
3. Large-Scale Dataset and Human Study

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

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

3.1. Dataset Collection

HDR-UGC videos were sourced from Vimeo under Creative Commons licenses. Over 10,000 videos were automatically filtered by HDR flags, resolution, format, and category, followed by manual verification to ensure authenticity. Videos were truncated to 10 seconds at a maximum resolution of 1080p using `ffmpeg` [10] and transcoded with an industry-standard bitrate ladder [4, 12]. This multi-stage process ensured that *BrightVQ* represents authentic HDR-UGC content with diverse distortions. Please refer to Supplementary Materials for more details.



(a) MOS distribution. (b) Landscape vs. Portrait.

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

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). 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 [15].

To derive robust Mean Opinion Scores (MOS), we employed the SUREAL method [20], which accounts for subject bias and inconsistency. Each rating S_{ij} from subject i for video j is modeled as:

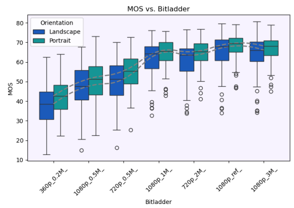
$$S_{ij} = \psi_j + \Delta_i + \nu_i X, \quad X \sim \mathcal{N}(0, 1), \tag{1}$$

where ψ_j represents the true quality of video j, Δ_i captures the bias of subject i, and ν_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. More details are in Supplementary Materials.

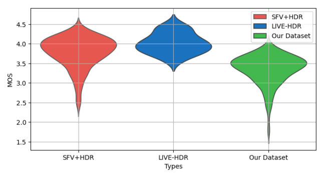
3.3. Analysis of MOS

Fig. 4a depicts the MOS distribution of *BrightVQ*, which is right-shifted, similar to other HDR datasets. This trend suggests that HDR videos, due to their inherently higher luminance and richer color details, often receive higher qual-

ICCV
#5421



(a) MOS vs. Bitadder.



(b) *BrightVQ* vs. LIVE-HDR [39] vs. SFV+HDR [44].

Figure 5. (a) shows the MOS variations across bitladder for *BrightVQ*. (b) compares MOS distributions of *BrightVQ*, LIVE-HDR [39], and SFV+HDR [44], showing that *BrightVQ* has a broader spectrum of MOS with less bias in peak MOS value.

ity 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. 5a 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.

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 [22], semantic cues from a CLIP-based encoder [33, 42], HDR features derived under a piecewise non-linear luminance transform on which distortion-sensitive natural video statistics are computed [8, 26, 35, 41], 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. 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. The self-supervised CONTRIQUE model [22] has demonstrated strong generalization across diverse UGC distortions [37], outperforming other fine-tuned methods [46, 47]. 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 [22] encoder on both the full and a downsampled half-resolution frame versions as:

$$\mathcal{U}^t_{scale} = f_{\text{CONTRIQUE}}(\mathbf{x}^t) \in \mathbb{R}^{d_{UGC}}. \tag{2}$$

where d_{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 [21, 37] and confirmed by our experiments (Sec. 5), CONTRIQUE [22] serves as a robust UGC backbone.

4.2. Semantic Feature Extraction

Perceptual quality depends not only on technical distortions but also on content semantics, which can influence human tolerance to various artifacts [13, 42]. 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, 13, 33, 42]. For each appropriately resized sampled frame $\mathbf{x}^t$, semantic features are extracted as

$$\mathcal{E}^t = f_{\text{CLIP}}(\mathbf{x}^t) \in \mathbb{R}^{d_{\text{SEM}}}. \tag{3}$$

Here, d_{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.

4.3. HDR Feature Extraction

HDR content suffers from distortions in extreme luminance regions that standard SDR-based VQA methods overlook [9, 37]. 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 $\mathbf{Y}^t \in [0, 1]$ is extracted. We then subdivide $\mathbf{Y}^t$ into two intervals (e.g., $[0, 0.5]$ and $[0.5, 1]$) and apply a piecewise expansive non-linearity inspired by [8, 39]. Specifically, the non-linearity is defined as:

$$g(x;\beta) = \begin{cases} e^{\beta x} - 1, & x \geq 0, \\ 1 - e^{-\beta x}, & x < 0, \end{cases} \tag{4}$$

with $\beta = 4$ following [8]. 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 [8]. 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 cap-

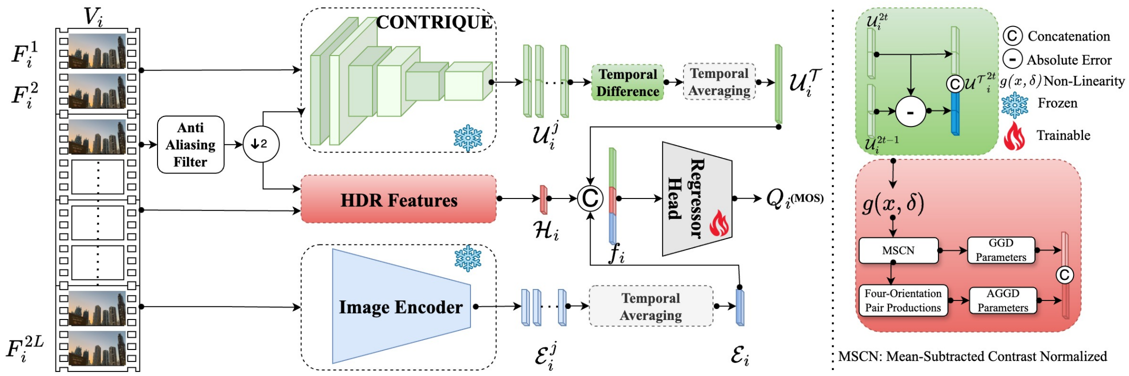


Figure 6. The overall framework of BrightRate for HDR-UGC Video Quality Assessment. BrightRate extracts HDR-specific features, and combines with UGC and Semantic features to give SOTA results on HDR-UGC benchmarks.

ture local image statistics:

$$\mathbf{M}^t(i,j) = \frac{\tilde{\mathbf{Y}}^t(i,j) - \mu(i,j)}{\sigma(i,j) + \epsilon}, \tag{5}$$

where $\mu(i,j)$ and $\sigma(i,j)$ are computed over a 31×31 window and ϵ is a small constant. The MSCN coefficients follow a Generalized Gaussian Distribution (GGD), and their adjacent products are modeled with an Asymmetric GGD (AGGD) [8, 25, 26]. We extract shape and variance parameters from both models and concatenate them across two scales to form the HDR-specific feature vector:

$$\mathcal{H}^t = f_{\text{HDR}}(\tilde{\mathbf{Y}}^t) \in \mathbb{R}^{d_{\text{HDR}}}. \tag{6}$$

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

4.4. Temporal Difference Module

Videos with higher perceptual quality typically exhibit smaller temporal fluctuations, while lower-quality videos show abrupt changes [2, 18, 29]. To capture these dynamics, we compute the absolute difference between consecutive UGC feature vectors:

$$\Delta\mathcal{U}^t = \left|\mathcal{U}^t - \mathcal{U}^{t-1}\right|, \quad t \in \{2, \ldots, T\}, \tag{7}$$

where $\mathcal{U}^t$ denotes the combined UGC feature for frame t (see Sec. 4.1). 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.

4.5. Quality Regression

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 $\overline{\mathbf{z}} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{z}^t$. 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:

$$Q_i = R(\overline{\mathbf{z}}). \tag{8}$$

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(↑)(Std)	PLCC(↑)(Std)	KROCC(↑)(Std)	RMSE(↓)(Std)
NR-IQA	BRISQUE [25]	0.3302 (0.0366)	0.3603 (0.0311)	0.2261 (0.0279)	12.5770 (0.2855)
	HDRMAX [39]	0.6276 (0.0321)	0.6318 (0.0356)	0.4409 (0.0288)	10.2428 (0.4008)
	CONTRIQUE [22]	0.7081 (0.0297)	0.7074 (0.0395)	0.5177 (0.0239)	11.4635 (1.3339)
	REIQA [36]	0.7919 (0.0116)	0.8023 (0.0168)	0.6068 (0.0103)	7.9390 (0.3421)
NR-VQA	VBLIINDS [35]	0.4605 (0.0365)	0.4478 (0.0347)	0.3180 (0.0246)	11.9322 (0.4202)
	CONVIQT [23]	0.7026 (0.0462)	0.7202 (0.0510)	0.5134 (0.0431)	10.5817 (1.4206)
	VSFA [18]	0.7556 (0.0139)	0.7501 (0.0206)	0.5538 (0.0138)	8.8310 (0.1834)
	COVER [13]	0.7609 (0.0201)	0.7917 (0.0252)	0.5597 (0.0181)	7.7352 (0.3104)
	FasterVQA [48]	0.7744 (0.0162)	0.7625 (0.0147)	0.5763 (0.0152)	9.0680 (0.2501)
	DOVER [47]	0.7745 (0.0155)	0.8060 (0.0207)	0.5924 (0.0123)	7.4641 (0.2801)
	FastVQA [45]	0.8094 (0.0121)	0.8530 (0.0156)	0.6445 (0.0106)	7.1336 (0.2402)
NR-HDR-VQA	HDRChipQA [8]	0.6781 (0.0220)	0.6855 (0.0179)	0.4889 (0.0160)	9.5869 (0.3081)
	HIDROVQA [37]	0.8526 (0.0217)	0.8620 (0.0136)	0.6680 (0.0215)	6.5708 (0.3367)
	BrightRate	0.8887(0.0197)	0.8970(0.0171)	0.7059(0.0227)	5.7514(0.4465)

5. Experiment

5.1. Databases

We evaluated *BrightRate* on our newly introduced *BrightVQ* dataset, as well as on SFV+HDR [44] and LIVE-HDR [39]. On For all datasets, we randomly split the videos into 80% training and 20% testing sets based on reference content to ensure that all videos from the same source appeared in the same split [22, 36]. By contrast with UGC-VQA methods such as DOVER [47], KSVQE [21], Fast/Faster-VQA [45, 48], etc. that fine-tune the feature ex-

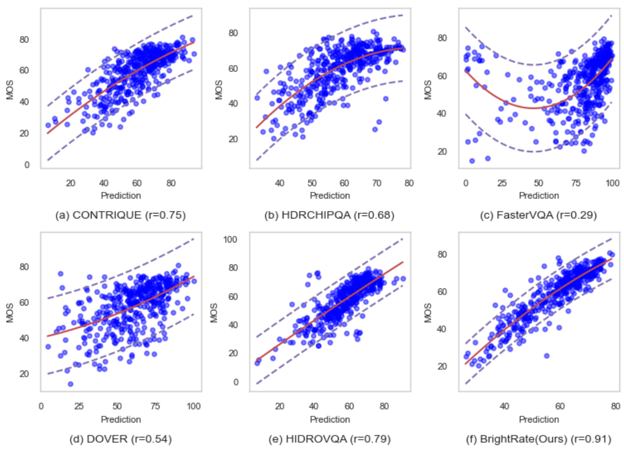
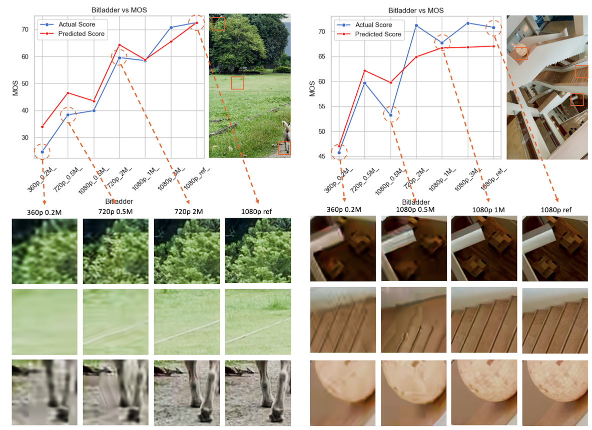


Figure 7. Scatter plots of actual MOS vs. predicted scores for various SOTA models on *BrightVQ*. Red curves show polynomial parametric fits.



(a) (b)

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.

traction 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) [33] for semantic features and the CONTRIQUE model [22] at two scales to extract UGC distribution features. HDR features are extracted by applying an expansive non-linearity over a 31×31 window with an expansion power of 4 [8, 39]. Temporal differences between consecutive CONTRIQUE [22] 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 [21, 22, 36, 37]. More details in Supplementary Material.

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 [37] performed second-best, underscoring its ability to capture HDR-specific distortions. In the NR-VQA/NR-IQA category, although FastVQA [45] performs well among SDR-oriented models, it is outperformed by HDR-specific approaches.

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 [39] and SFV+HDR [44], achieving the highest correlation against MOS. On LIVE-HDR [39], 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 [44], *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.

5.3.3. Cross-dataset Evaluation

We conducted two cross-dataset evaluations: "*BrightRate* dataset→other datasets" and "other datasets→*BrightRate*

Table 4. Performance Comparison on LIVE-HDR [39] and SFV+HDR [44] Datasets.

Method	LIVE-HDR				SFV+HDR			
	SROCC(↑)	PLCC(↑)	KRCC(↑)	RMSE(↓)	SROCC(↑)	PLCC(↑)	KRCC(↑)	RMSE(↓)
BRISQUE [25]	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 [39]	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 [22]	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 [36]	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 [35]	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 [23]	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 [47]	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 [13]	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 [18]	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 [48]	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 [45]	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 [8]	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 [37]	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)
BrightRate	0.8907 (0.0425)	0.8824 (0.0470)	0.7178 (0.0492)	8.3955 (1.9260)	0.7328 (0.0509)	0.7709 (0.0252)	0.5415 (0.0496)	0.2679 (0.0236)

Table 5. Cross Data Validation: Train on *BrightVQ*, Test on LIVE-HDR [39] and SFV+HDR [44].

Method	Test: LIVE-HDR				Test: SFV+HDR			
	SROCC(↑)	PLCC(↑)	RMSE(↓)	KRCC(↑)	SROCC(↑)	PLCC(↑)	RMSE(↓)	KRCC(↑)
BRISQUE [25]	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 [39]	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 [22]	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 [36]	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 [23]	0.6240 (0.1197)	0.6112 (0.1075)	15.0850 (1.2097)	0.4331 (0.0935)	0.4981 (0.0391)	0.5129 (0.0520)	44.0703 (1.3564)	0.4267 (0.0281)
VBLIINDS [35]	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 [8]	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 [18]	0.4597 (0.1622)	0.4349 (0.1609)	17.4869 (2.0345)	0.3342 (0.1329)	0.4581 (0.0849)	0.5404 (0.0899)	51.4192 (1.3247)	0.3179 (0.0602)
HIDROVQA [37]	0.4086 (0.0915)	0.4918 (0.0924)	15.5255 (0.7523)	0.2886 (0.0728)	0.3398 (0.0491)	0.3020 (0.0720)	24.0703 (1.2556)	0.1267 (0.0383)
BrightRate	0.7362 (0.0741)	0.7337 (0.0563)	15.1022 (0.8398)	0.5524 (0.0621)	0.5310 (0.0670)	0.5465 (0.0730)	51.8795 (1.1711)	0.3629 (0.0510)

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 [23] and HIDROVQA [37], 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 [44], indicating its limited diversity in representing HDR distortions. These findings reinforce *BrightVQ*'s value as a robust and comprehensive benchmark for HDR VQA task.

5.4. Ablation Study

To assess the effectiveness of the components in our model, namely UGC Feature Extractor, Semantic Feature Extraction (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.

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 [39] and SFV+HDR [44]. This demonstrates that CLIP strengthens the model's ability to extract meaningful semantic features relevant to video quality assessment.

Effectiveness of Temporal Difference Module: Incorporating the Temporal-Difference Module results in noticeable performance improvements across all datasets. As indicated in Table 7, adding tempporal features increase SROCC (+0.088) and PLCC (+0.075) on *BrightVQ* dataset compared to the baseline, confirming its ability to capture temporal variations in HDR videos. The improvements on LIVE-HDR [39] and SFV+HDR [44] were relatively mod-

Table 6. Cross Data Validation on *BrightVQ* Test Set. Columns under "Train: LIVE-HDR [39]" report metrics when training on LIVE-HDR [39] and testing on *BrightVQ*, while those under "Train: SFV+HDR [44]" report metrics when training on SFV+HDR [44] and testing on *BrightVQ*.

Method	Train: LIVE-HDR [39], Test: *BrightVQ*				Train: SFV+HDR [44], Test: *BrightVQ*			
	SROCC(↑)	PLCC(↑)	RMSE(↓)	KRCC(↑)	SROCC(↑)	PLCC(↑)	RMSE(↓)	KRCC(↑)
BRISQUE [25]	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 [39]	0.1176 (0.0480)	0.0489 (0.0524)	13.7844 (0.3274)	0.0777 (0.0334)	0.2302 (0.0415)	0.2451 (0.0461)	55.0761 (0.6489)	0.1565 (0.0298)
CONTRIQUE [22]	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 [36]	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 [23]	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 [35]	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 [8]	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 [18]	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 [37]	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)
BrightRate	0.6669 (0.0346)	0.7459 (0.0373)	9.4324 (0.7087)	0.4806 (0.0260)	0.5892 (0.0249)	0.5308 (0.0263)	55.0568 (0.6537)	0.4004 (0.0204)

Table 7. Ablation Study I: Effect of Modules on SROCC (↑) and PLCC (↑). Results are reported for *BrightVQ*, LIVE-HDR [39], and SFV+HDR [44] datasets.

Module/s	BrightRate		LIVE-HDR		SFV+HDR	
	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC
Baseline(CONTRIQUE)	0.7081	0.7074	0.7868	0.8016	0.5901	0.5959
+CLIP	0.8431	0.8424	0.8325	0.8230	0.6403	0.6598
+Temporal-Difference	0.7961	0.7821	0.8159	0.8157	0.6161	0.6749
+HDR	0.8485	0.8489	0.8301	0.8129	0.6250	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	BrightRate		LIVE-HDR		SFV+HDR	
	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC
+(CLIP+Temp)	0.8368	0.8578	0.8494	0.8276	0.6318	0.6894
+(HDR+Temp)	0.8389	0.8564	0.8510	0.8319	0.6773	0.6734
+(CLIP+HDR)	0.8463	0.8470	0.8673	0.8301	0.6943	0.7032
BrightRate	0.8887	0.8970	0.8907	0.8824	0.7328	0.7709

est, suggesting that these two datasets may contain fewer temporal artifacts, making motion-aware learning less influential.

Effectiveness of HDR Feature Extraction Module: The HDR-specific feature extraction module enhances the model's ability to detect distortions unique to HDR content. Comparing the baseline with the HDR Feature Extraction module in Table 7, we observe an SROCC increase (+0.140) and PLCC increase of (+0.141) on the *BrightVQ* dataset, emphasizing the module's critical role in HDR quality assessment. The improvements extend to LIVE-HDR [39] (SROCC: 0.8301) and SFV+HDR [44] (SROCC: 0.6250), confirming that HDR-specific feature extraction is essential 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 [39] and SFV+HDR [44]. 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 quality 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

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

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

[3] Sandhini Agarwal, Gretchen Krueger, Jack Clark, Alec Radford, Jong Wook Kim, and Miles Brundage. Evaluating clip: towards characterization of broader capabilities and downstream implications. *arXiv preprint arXiv:2108.02818*, 2021.

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

[5] Maryam Azimi, Amin Banitalebi-Dehkordi, Yuanyuan Dong, Mahsa T. Pourazad, and Panos Nasiopoulos. Evaluating the performance of existing full-reference quality metrics on high dynamic range (hdr) video content, 2018.

[6] Vittorio Baroncini, Federica Mangiatordi, Emiliano Pallotti, and Massimiliano Agostinelli. Visual assessment of hdr video. 2016.

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

[8] Joshua P. Ebenezer, Zaixi Shang, Yongjun Wu, Hai Wei, Sriram Sethuraman, and Alan C. Bovik. Hdr-chipqa: No-reference quality assessment on high dynamic range videos, 2023.

[9] Joshua P Ebenezer, Zaixi Shang, Yixu Chen, Yongjun Wu, Hai Wei, Sriram Sethuraman, and Alan C Bovik. Hdr or sdr? a subjective and objective study of scaled and compressed videos. *IEEE Transactions on Image Processing*, 2024.

[10] FFmpeg Developers. Ffmpeg. https://ffmpeg.org/. Accessed: 2025-02-04.

[11] Deepti Ghadiyaram, Janice Pan, Alan C. Bovik, Anush Krishna Moorthy, Prasanjit Panda, and Kai-Chieh Yang. In-capture mobile video distortions: A study of subjective behavior and objective algorithms. *IEEE Trans. Circuits Syst. Video Technol.*, 28(9):2061–2077, 2018.

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

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

[14] Vlad Hosu, Franz Hahn, Mohsen Jenadeleh, Hanhe Lin, Hui Men, Tamás Szirányi, Shujun Li, and Dietmar Saupe. The konstanz natural video database (konvid-1k). In *QoMEX*, pages 1–6. IEEE, 2017.

[15] International Telecommunication Union. Methodology for the Subjective Assessment of the Quality of Television Pictures. Technical Report BT.500-14, International Telecommunication Union, 2019.

[16] ITU. Bt.2020 : Parameter values for ultra-high definition television systems for production and international programme exchange,. https://glenwing.github.io/docs/ITU-R-BT.2020-1.pdf.

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

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

[19] Yang Li, Shengbin Meng, Xinfeng Zhang, Shiqi Wang, Yue Wang, and Siwei Ma. UGC-VIDEO: perceptual quality assessment of user-generated videos. In *3rd IEEE Conference on Multimedia Information Processing and Retrieval, MIPR 2020, Shenzhen, China, August 6-8, 2020*, pages 35–38. IEEE, 2020.

[20] Zhi Li, Christos G. Bampis, Lucjan Janowski, and Ioannis Katsavounidis. A simple model for subject behavior in subjective experiments. In *Electronic Imaging*, pages 131–1–131–14, 2020.

[21] Yiting Lu, Xin Li, Yajing Pei, Kun Yuan, Qizhi Xie, Yunpeng Qu, Ming Sun, Chao Zhou, and Zhibo Chen. Kvq: Kwai video quality assessment for short-form videos, 2024.

[22] Pavan C. Madhusudana, Neil Birkbeck, Yilin Wang, Balu Adsumilli, and Alan C. Bovik. Image quality assessment using contrastive learning. *IEEE Trans. Image Process.*, 31: 4149–4161, 2022.

[23] Pavan C Madhusudana, Neil Birkbeck, Yilin Wang, Balu Adsumilli, and Alan C Bovik. Conviqt: Contrastive video quality estimator. *IEEE Transactions on Image Processing*, 32:5138–5152, 2023.

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

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

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

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

[28] Maryam Mohsin. 10 youtube statistics every marketer should know in 2020, 2020. [Online].

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

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

[31] Omnicore. Tiktok by the numbers, 2024. [Online].

[32] Xiaofei Pan, Jiaqi Zhang, Shanshe Wang, Shiqi Wang, Yun Zhou, Wenhua Ding, and Yahui Yang. Hdr video quality assessment: Perceptual evaluation of compressed hdr video.

Journal of Visual Communication and Image Representation, 57:76–83, 2018.

[33] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *ICML*, pages 8748–8763. PMLR, 2021.

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

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

[36] Avinab Saha, Sandeep Mishra, and Alan C. Bovik. Re-iqa: Unsupervised learning for image quality assessment in the wild. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023, Vancouver, BC, Canada, June 17-24, 2023*, pages 5846–5855. IEEE, 2023.

[37] Shreshth Saini, Avinab Saha, and Alan C. Bovik. Hidro-vqa: High dynamic range oracle for video quality assessment, 2023.

[38] Zaixi Shang, Yixu Chen, Yongjun Wu, Hai Wei, and Sriram Sethuraman. Subjective and objective video quality assessment of high dynamic range sports content. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 556–564, 2023.

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

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

[41] Zhengzhong Tu, Xiangxu Yu, Yilin Wang, Neil Birkbeck, Balu Adsumilli, and Alan C Bovik. Rapique: Rapid and accurate video quality prediction of user generated content. *IEEE Open Journal of Signal Processing*, 2:425–440, 2021.

[42] Jianyi Wang, Kelvin CK Chan, and Chen Change Loy. Exploring clip for assessing the look and feel of images. In *Proceedings of the AAAI conference on artificial intelligence*, pages 2555–2563, 2023.

[43] Yilin Wang, Sasi Inguva, and Balu Adsumilli. Youtube UGC dataset for video compression research. In *MMSP*, pages 1–5. IEEE, 2019.

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

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

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

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

[48] Haoning Wu, Chaofeng Chen, Liang Liao, Jingwen Hou, Wenxiu Sun, Qiong Yan, Jinwei Gu, and Weisi Lin. Neighbourhood representative sampling for efficient end-to-end video quality assessment. *IEEE Trans. Pattern Anal. Mach. Intell.*, 45(12):15185–15202, 2023.

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

[50] Jiahua Xu, Jing Li, Xingguang Zhou, Wei Zhou, Baichao Wang, and Zhibo Chen. Perceptual quality assessment of internet videos. In *Proceedings of the 29th ACM International Conference on Multimedia*, page 1248–1257, New York, NY, USA, 2021. Association for Computing Machinery.

[51] Zhenqiang Ying, Maniratnam Mandal, Deepti Ghadiyaram, and Alan C. Bovik. Patch-vq: 'patching up' the video quality problem. In *CVPR*, pages 14019–14029. Computer Vision Foundation / IEEE, 2021.

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