

Making Video Quality Assessment Models Robust to Bit Depth

Joshua P. Ebenezer, *Student Member, IEEE*, Zaixi Shang, *Student Member, IEEE*, Yongjun Wu, Hai Wei, Sriram Sethuraman and Alan C. Bovik, *Fellow, IEEE*

Abstract—We introduce a novel feature set, which we call HDRMAX features, that when included into Video Quality Assessment (VQA) algorithms designed for Standard Dynamic Range (SDR) videos, sensitizes them to distortions of High Dynamic Range (HDR) videos that are inadequately accounted for by these algorithms. While these features are not specific to HDR, and also augment the equality prediction performances of VQA models on SDR content, they are especially effective on HDR. HDRMAX features modify powerful priors drawn from Natural Video Statistics (NVS) models by enhancing their measurability where they visually impact the brightest and darkest local portions of videos, thereby capturing distortions that are often poorly accounted for by existing VQA models. As a demonstration of the efficacy of our approach, we show that, while current state-of-the-art VQA models perform poorly on 10-bit HDR databases, their performances are greatly improved by the inclusion of HDRMAX features when tested on HDR and 10-bit distorted videos.

Index Terms—Video Quality Assessment, High Dynamic Range

I. INTRODUCTION

HIGH Dynamic Range (HDR) videos have greater bit depths, wider color gamuts, and different opto-electronic transfer functions than Standard Dynamic Range (SDR) videos. For example, the HDR10 standard requires a bit depth of 10 bits, the BT 2020 color gamut (which covers 75.8% of the perceiveable color gamut), and the Perceptual Quantizer (PQ) OETF, which is designed to represent luminances up to 10,000 nits. SDR videos are typically 8 bit, follow the BT 709 gamut, and use the gamma OETF, which is perceptually inaccurate for luminances above 100 nits. HDR videos also require displays that can radiate the increased ranges of brightnesses and colors, and specially designed hardware, whereas SDR videos were originally designed for Cathode-Ray Tubes, whose capabilities are exceeded by modern displays.

Most state-of-the-art Video Quality Assessment (VQA) algorithms were designed to analyze SDR videos, primarily because the adoption and standardization of HDR is still ongoing. Existing VQA models are thus unable to capture distortions as they affect HDR over increased luminance/color

Corresponding Author: Joshua Ebenezer (joshuaebenezer@utexas.edu). ORCID: 0000-0003-4936-9784.

This work was supported by a grant from Amazon and by grant number 2019844 for the National Science Foundation AI Institute for Foundations of Machine Learning (IFML).

J. P. Ebenezer, Z. Shang, and A. C. Bovik are with The Laboratory for Image and Video Engineering (LIVE) at The University of Texas at Austin, TX 78712 (e-mail: joshuaebenezer@utexas.edu).

Y. Wu, H. Wei, and S. Sethuraman are with Amazon Prime Video, Seattle, WA 98109.

ranges, and they perform poorly on predicting the quality of HDR videos. In a recent study [1], we found that modifying the features of a powerful no-reference VQA model using an expansive nonlinearity significantly enhance its prediction performance on HDR video content. It has occurred to us that applying a similar concept to modify or augment other existing VQA models, including globally deployed algorithms, might also enhance their performances, which could significantly impact the delivery of high-quality HDR content worldwide. We therefore designed a new set of features, that we call HDRMAX, which can be used to augment both No-Reference (NR) VQA and Full-Reference (FR) models. We have found that including HDRMAX features improves robustness to bit depth in every case, dramatically improving their performances on distorted HDR video content. Remarkably, we have found that they also improve VQA model performance on distorted SDR videos although to a lesser degree, eliminating any concerns about their generalizability.

II. NONLINEAR FEATURES FOR VQA

We will focus on processing transformed luminances (luma values) utilized in HDR standards like HDR10. Similar processing can be applied to Chroma values. In order to define HDRMAX features that are sensitized to distortions that affect the quality of HDR, begin by linearly scaling the luma values in each $W \times W$ spatial patch of each video frame to $[-1, 1]$ [1]. It may be observed that this scaling localizes the analysis of videos, and the resulting local dynamic ranges may significantly differ from the global. While this linear scaling to this particular range is not a requirement, it greatly simplifies the following processing steps and our exposition. Then, pass the linearly scaled luma values of each patch through an expansive nonlinearity

$$f(x; \delta) = \begin{cases} \exp(\delta x) - 1 & x > 0 \\ 1 - \exp((- \delta x)) & x < 0 \end{cases}. \quad (1)$$

The function is plotted in Fig.1. This operation, which we refer to as the HDRMAX nonlinearity, expands the brightness ranges at the ends of the local dynamic luma scale while compressing the middle ranges. The rationale for (1) is simple but powerful. The features used in existing VQA algorithms are generally responsive to distortions that are endemic to videos, whether SDR or HDR. However, these responses are generally dominated by distortions of the middle ranges of luma values, because they occupy most of the spatial extents of videos, as well as most of their dynamic ranges. Regions

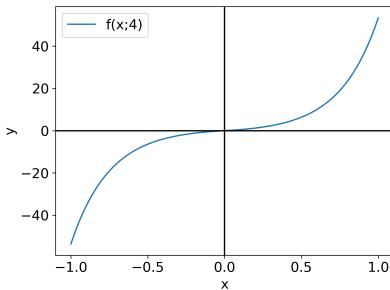


Fig. 1. Plot of $f(x; 4)$ vs x .

at the extremes of the dynamic range, which radiate very high or low (bright whites and dark blacks) luma (or chroma) values, distortions can be quite evident visually, but are often unaccounted for by standard VQA models because of dilution by the dominating responses to regions radiating brightnesses and colors in the middle portion of the dynamic range. The HDRMAX nonlinearity $f(x; \delta)$ counteracts this by isolating the extreme ends of the dynamic range at the expense of the middle range, which is suppressed. In this way, the modified feature responses are able to directly capture perceptual quality information more sensitive to bit-depth. As we shall see, including the HDRMAX nonlinearity to define new features to augment existing VQA models (while not replacing their original features) can produce dramatic improvements on distorted HDR videos, without sacrificing efficacy on SDR videos, with only a modest increment in complexity.

The HDRMAX features are also defined using a second process, which we refer to as +NOISE, which involves applying a small amount of additive white Gaussian noise (AWGN) to the nonlinearly transformed luma values, thereby modeling the random signal fluctuations occurring at different stages along the visual pathway. In addition, on regions that are smooth and lack texture, or are overexposed or underexposed, the addition of random noise provides numerical stability to processes that involve division by local energy or contrast estimates that underlie many VQA models. Denoting a frame processed by the HDRMAX nonlinearity as P , the resultant signal Q (HDRMAX+NOISE) is given by

$$Q[i, j] = P[i, j] + N(0, \sigma^2) \quad (2)$$

where i, j are spatial pixel indices and $N(0, \sigma^2)$ is a zero-mean Gaussian random variable having standard deviation of σ . We fixed $\sigma = 0.001$ in our experiments, although varying it in the range of $[5 \times 10^{-4}, 5 \times 10^{-3}]$ produces little changes in performance.

A. HDRMAX Features for NR VQA

As already described, we modify NVS priors, using HDRMAX features to capture distortions on signal ranges visually expressive of HDR. Given the HDRMAX+NOISE processed luma values, extract quality-aware BRISQUE/NIQE ([2], [3]) features on them as follows. Model the distribution of the Mean-Subtracted Contrast Normalized (MSCN) coefficients [2] of Q in (2) as following a Generalized Gaussian Distribution (GGD). Using the maximum likelihood procedure deployed in BRISQUE/NIQE, find the shape parameter α and

spread parameter σ of the best GGD fit to the empirical distribution $Q[i, j]$. These serve as two features sensitized to video quality at the local extremes of dynamic range.

Likewise, to capture the local correlation structure of $Q[i, j]$ (and the embedded distortions), also compute the products of neighboring pixels along four directions, then model the distributions of these products as following Asymmetric Generalized Gaussian Distributions (AGGD). In this way, an additional 16 features are arrived at (μ, σ_L, σ_R , and γ , each along 4 orientations), computed exactly as they are in [2], [3].

The best fitting GGD and AGGD parameters are then averaged over all the frames of each video at two scales. In addition, the standard deviations of these parameters are computed over each non-overlapping group of five frames, then averaged over the entire video duration, yielding another feature set that is expressive of the temporal variations of quality in the analyzed video. Thus, a total of 72 HDRMAX+NOISE features are found. These are used to supplement learning any given NR VQA model, as exemplified later.

B. HDRMAX Features for FR VQA

We derive a set of nonlinear HDRMAX features for FR VQA using the same nonlinear processing described in (1). Next, we extract the Visual Information Fidelity (VIF) features [4] and the Detail Loss Metric (DLM) [5] features directly on the HDRMAX processed luma values. We do not deploy the +NOISE modification, because VIF (and the related VQA engines ST-RRED [6] and SpEED-QA [7]) already have neural noise models embedded in them. VIF is computed at the same four spatial scales as in VMAF [8], and DLM is also calculated exactly as is done in [8]. These features are extracted from each frame, then averaged over all frames. No additional temporal features are found, since temporal effects are already modeled in the VIF features [4].

III. DATABASES

We used two 10-bit databases and one 8-bit database to study the performance of HDRMAX-enhanced NR and FR VQA models. This allows us to gauge the efficacy of HDRMAX on HDR content, while also ascertaining whether performance is maintained on SDR content.

- The **LIVE HDR database** [9], a new database dedicated to the study of HDR perception, consisting of 310 videos viewed by 66 subjects. The distorted videos were created by compression using the x265 encoder and spatial scaling. All of the videos are compliant with the HDR10 standard.
- The **LIVE ETRI database** [10] consists of 437 videos that had undergone compression, spatial aliasing, and temporal subsampling, along with human subjective scores. The videos in this database are 10 bit SDR, i.e., they are stored with a 10 bit representation but use the BT 709 gamma curve and color gamut.
- The **LIVE Livestream database** [11] contains 315 videos subjected to distortions commonly occurring in

live streamed videos, such as interlacing, judder, compression, aliasing, frame-drops, flicker, and judder. Human scores are also supplied with this dataset. The videos in this database are 8 bit SDR.

IV. RESULTS

For each of the databases, we trained a Support Vector Regressor (SVR) against the given human opinion scores for each video. Each database was divided into train and test sets using a 80:20 content separation. Cross-validation was performed to find the best parameters for the SVR on each database. The Spearman Rank Ordered Correlation Coefficient (SRCC) and Pearson's Linear Correlation Coefficient (PLCC) between the predictions generated by each VQA model against the subjective quality scores were computed, and the median values are reported over 100 random train-test splits. Table I gives the results obtained when HDRMAX+NOISE was included into existing NR VQA algorithms on the two 10 bit databases; the LIVE HDR dataset, and the LIVE ETRI 10 bit SDR database. As may be observed, enhancement with the HDRMAX+NOISE feature set improved performance of every algorithm on every database, often dramatically. VBLIINDS+HDRMAX+NOISE is the best performer on the LIVE HDR dataset, while ChipQA+HDRMAX+NOISE was the best performer on the ETRI dataset.

We also tabulated the results obtained by using the HDRMAX+NOISE feature set to enhance the performances of VQA algorithms on legacy 8 bit SDR video data in Table I. Since HDRMAX+NOISE was designed to improve the quality prediction performance of existing VQA models when applied on HDR, we did not expect any significant improvement in performance. Instead, we were concerned that losses in performance could occur owing to feature dilution. Fortunately, in nearly every instance, the HDRMAX+NOISE-enhanced algorithms performed at least as well as they did without HDRMAX+NOISE augmentation on the LIVE Livestream database, and in most cases performed better. Local contrast plays an important role in SDR video perception, as it does in HDR viewing, which helps explain HDRMAX's strong performance on SDR content as well. These highly satisfactory results strongly suggest that HDRMAX+NOISE can be added to existing VQA algorithms to improve their robustness across bit depths. On average, median SRCCs of NR VQA algorithms improved by 20% on the LIVE HDR database, by 57% on the ETRI 10-bit SDR database, and by 12% on the LIVE Livestream 8-bit SDR database when the HDRMAX+NOISE features were included.

We also modified several leading FR algorithms to include HDRMAX, including peak signal-to-noise ratio (PSNR), SSIM, MS-SSIM [19], VMAF [8], SpEED-QA [7], and ST-RRED [6]. We also included the HDR VQA model HDR-VDP-2 [20] to show that the HDRMAX features are capable of improving the performance VQA models already designed to assess HDR content. Since most FR VQA algorithms directly compute video quality predictions without using machine learning when mapping features to human opinion scores, we modified the compared FR VQA algorithms by breaking them

into their constituent components, using each component as an individual feature. Specifically, we factored SSIM into three features: the three factors representing luminance, contrast and structural similarity features. Likewise, we decomposed MS-SSIM into eleven features, two SSIM features from each of four spatial scales, and three from the coarsest scale. On SpEED-QA, we extracted both the "reduced-reference" version and the "single-number" versions of the spatial and temporal SpEED-QA values to obtain four features. ST-RRED features were obtained from five levels of the Steerable Pyramid used in that algorithm. HDR-VDP-2 features are the pooled quality features over nine spatial scales. For each FR model, the combined feature set was obtained by concatenating the HDRMAX features with the original FR algorithm features. We then trained an SVR to map the features to human quality judgments on each of the 10-bit databases. As maybe observed in Table II, every FR model modified by HDRMAX produced significantly improved quality prediction results on HDR videos, often by rather dramatic amounts. This includes SSIM, MS-SSIM, and VMAF models that are currently globally deployed by streaming and social media platforms. This suggests that HDRMAX can be used to improve the perceptual optimization of HDR video compression, potentially significantly reducing bandwidth consumption at the largest scales. To complete the analysis, we also studied the performances of the same models augmented by HDRMAX, but trained and tested on SDR videos. The results were computed on the LIVE Livestream database and are given in Table II as well. Again, no loss of performance was observed on the SDR video data; rather, there were improvements in all instances. On average, the median SRCCs of FR VQA algorithms improved by 25% on the LIVE HDR database, by 57% on the ETRI 10-bit SDR database, and by 15% on the LIVE Livestream 8-bit SDR database when the HDRMAX features were included.

V. CONCLUSION

The apparent success of our HDRMAX model shows that it is possible to build much stronger predictors of video quality for HDR, by employing it in existing SDR algorithms. Extensive evaluations on HDR and SDR validate these conclusions on both NR and FR VQA models. Although the only currently available subjective HDR VQA databases utilize the HDR10 standard, HDRMAX is likely to be advantageous if applied to HDR standards like HDR10+ and Dolby Vision, which have an even greater capacity to represent local dynamic ranges than HDR10.

REFERENCES

- [1] J. P. Ebenezer, Z. Shang, Y. Wu, H. Wei, S. Sethuraman, and A. C. Bovik, "HDR-ChipQA: No-reference quality assessment on high dynamic range videos," *Submitted to Elsevier Signal Process.: Image Comm.*
- [2] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, 2012.
- [3] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, 2012.

TABLE I

MEDIAN SROCC AND PLCC OF NR MODEL PERFORMANCE AGAINST HUMAN JUDGMENTS ON 10 BIT VQA DATABASES. STANDARD DEVIATIONS ARE SHOWN IN PARENTHESES. THE BEST PERFORMING ALGORITHM IN EACH COLUMN IS BOLD-FACED.

| Dataset | LIVE HDR | | LIVE ETRI (SDR 10 bit) | | Livestream (SDR 8 bit) | |
|--------------------------|-----------------------|-----------------------|------------------------|-----------------------|------------------------|-----------------------|
| | SRCC | PLCC | SRCC | PLCC | SRCC | PLCC |
| HDRMAX | 0.8139(0.0584) | 0.7950(0.0534) | 0.6322(0.1242) | 0.6254(0.1367) | 0.7658(0.0774) | 0.7735(0.0675) |
| HDRMAX+NOISE | 0.7995(0.0643) | 0.7772(0.0941) | 0.6608(0.1042) | 0.6468(0.1076) | 0.7656(0.0854) | 0.7782(0.0722) |
| TLVQM[12] | 0.5781 (0.1014) | 0.5552 (0.0919) | 0.3398(0.1320) | 0.3184(0.1435) | 0.7165(0.0868) | 0.7241(0.0800) |
| TLVQM+HDRMAX | 0.8328(0.0591) | 0.8103(0.0570) | 0.7840(0.0759) | 0.6371(0.1318) | 0.8173(0.0736) | 0.8208(0.0692) |
| TLVQM[12]+HDRMAX+NOISE | 0.8194(0.0526) | 0.7992(0.0534) | 0.6808(0.1265) | 0.6671(0.1255) | 0.8294(0.0745) | 0.8297(0.0687) |
| RAPIQUE[13] | 0.4553 (0.2533) | 0.4864 (0.1171) | 0.1481(0.1419) | -0.0059(0.1633) | 0.3556(0.1637) | 0.3298(0.1545) |
| RAPIQUE[13]+HDRMAX | 0.5458(0.1835) | 0.4363(0.3487) | 0.4221(0.1745) | 0.3671(0.2749) | 0.3793(0.1550) | 0.3456(0.1469) |
| RAPIQUE[13]+HDRMAX+NOISE | 0.6778(0.1583) | 0.6642(0.2645) | 0.4745(0.1619) | 0.4616(0.2971) | 0.5083(0.1543) | 0.4643(0.2099) |
| BRISQUE[2] | 0.7251 (0.0955) | 0.7139 (0.0881) | 0.3891(0.2109) | 0.3796(0.1966) | 0.6564(0.1140) | 0.6840(0.1019) |
| BRISQUE[2]+HDRMAX | 0.8024(0.0743) | 0.7763(0.0576) | 0.6323(0.1230) | 0.6308(0.1320) | 0.7535(0.0898) | 0.7728(0.0758) |
| BRISQUE[2]+HDRMAX+NOISE | 0.8383(0.0737) | 0.8052(0.0816) | 0.6373(0.1136) | 0.6637(0.1348) | 0.7427(0.0771) | 0.7703(0.0664) |
| VIDEVAL[14] | 0.7131(0.1093) | 0.6513(0.1230) | 0.6817(0.1181) | 0.6522(0.1199) | 0.8162(0.0679) | 0.8194(0.0588) |
| VIDEVAL[14]+HDRMAX | 0.8104(0.0682) | 0.7805(0.0672) | 0.6382(0.1318) | 0.6271(0.1368) | 0.8211(0.0657) | 0.8272(0.0602) |
| VIDEVAL[14]+HDRMAX+NOISE | 0.8255(0.0632) | 0.7961(0.0604) | 0.6963(0.1161) | 0.6593(0.1150) | 0.8233(0.0725) | 0.8339(0.0639) |
| VBLINDS[15] | 0.7484(0.1260) | 0.6930(0.1480) | 0.6121(0.1427) | 0.5729(0.1278) | 0.7464(0.0833) | 0.7572(0.0748) |
| VBLINDS[15]+HDRMAX | 0.8463(0.0748) | 0.8163(0.0700) | 0.6784(0.1247) | 0.6858(0.1250) | 0.8181(0.0600) | 0.8300(0.0527) |
| VBLINDS[15]+HDRMAX+NOISE | 0.8492(0.0751) | 0.8190(0.1059) | 0.7172(0.1032) | 0.7132(0.1061) | 0.8166(0.0670) | 0.8228(0.0590) |
| HIGRADE[16] | 0.6985(0.0839) | 0.6548(0.0866) | 0.5018(0.1555) | 0.4675(0.1507) | 0.6801(0.0998) | 0.6913(0.0915) |
| HIGRADE+HDRMAX | 0.8074(0.0766) | 0.7881(0.0740) | 0.7927(0.0581) | 0.6632(0.1251) | 0.7654(0.0886) | 0.7681(0.0773) |
| HIGRADE[16]+HDRMAX+NOISE | 0.8111(0.0690) | 0.7927(0.0674) | 0.6843(0.0963) | 0.6643(0.1108) | 0.7771(0.0894) | 0.7831(0.0773) |
| VSFA[17] | 0.7145(0.1070) | 0.6869(0.1000) | 0.6178(0.2050) | 0.6210(0.2303) | 0.8409(0.0684) | 0.8658(0.0561) |
| VSFA[17]+HDRMAX | 0.7414(0.0960) | 0.7196(0.0889) | 0.6014(0.2160) | 0.5789(0.2571) | 0.8439(0.0656) | 0.8652(0.0526) |
| VSFA[17]+HDRMAX+NOISE | 0.7518(0.0932) | 0.7315(0.0870) | 0.5955(0.2005) | 0.6111(0.2307) | 0.8472(0.0662) | 0.8661(0.0519) |
| ChipQA[18] | 0.7435 (0.0895) | 0.7334 (0.0819) | 0.5934(0.1781) | 0.6033(0.1876) | 0.7980(0.0750) | 0.8066(0.0713) |
| ChipQA[18]+HDRMAX | 0.8117(0.0697) | 0.7766(0.0690) | 0.7025(0.1370) | 0.7069(0.1685) | 0.8049(0.0729) | 0.8188(0.0643) |
| ChipQA[18]+HDRMAX+NOISE | 0.8196(0.0698) | 0.7940(0.0693) | 0.7252(0.1278) | 0.7275(0.1567) | 0.8041(0.0670) | 0.8194(0.0661) |

TABLE II

MEDIAN SROCC, LCC, AND RMSE ON 10 BIT VQA DATABASES OBTAINED USING FR MODELS. STANDARD DEVIATIONS ARE SHOWN IN PARENTHESES. THE BEST PERFORMING ALGORITHM IS BOLD-FACED.

| Dataset | LIVE HDR | | LIVE ETRI (SDR 10 bit) | | Livestream (SDR 8 bit) | |
|------------------|-----------------------|-----------------------|------------------------|-----------------------|------------------------|-----------------------|
| | SRCC | PLCC | SRCC | PLCC | SRCC | PLCC |
| Algorithm | | | | | | |
| HDRMAX | 0.7681(0.0913) | 0.7400(0.0958) | 0.8078(0.1035) | 0.7950(0.0958) | 0.7535(0.0772) | 0.7626(0.0702) |
| PSNR | 0.6242(0.1504) | 0.6357(0.1331) | 0.4941(0.1243) | 0.4289(0.1289) | 0.6063(0.0792) | 0.6238(0.0716) |
| PSNR+HDRMAX | 0.8263(0.0684) | 0.8206(0.0615) | 0.8268(0.1001) | 0.8196(0.0940) | 0.7396(0.0432) | 0.7488(0.0370) |
| SSIM | 0.5208(0.1611) | 0.4898(0.1595) | 0.3568(0.2625) | 0.3358(0.2395) | 0.6539(0.0927) | 0.6584(0.0832) |
| SSIM+HDRMAX | 0.7771(0.0866) | 0.7529(0.0964) | 0.8485(0.0733) | 0.8301(0.0741) | 0.7521(0.0714) | 0.7689(0.0619) |
| MS-SSIM | 0.6007(0.1228) | 0.5810(0.1260) | 0.5234(0.2336) | 0.5319(0.2279) | 0.7306(0.1097) | 0.7377(0.1083) |
| MS-SSIM+HDRMAX | 0.7645(0.0838) | 0.7258(0.0868) | 0.7519(0.1399) | 0.7297(0.1328) | 0.7397(0.0712) | 0.7724(0.0681) |
| ST-RRED | 0.6863(0.0700) | 0.6569(0.0744) | 0.7500(0.0853) | 0.7587(0.0933) | 0.6122(0.0738) | 0.6273(0.0637) |
| ST-RRED+HDRMAX | 0.7896(0.0607) | 0.7595(0.0603) | 0.8628(0.0889) | 0.8535(0.0840) | 0.7685(0.0690) | 0.7902(0.0630) |
| SpEED-QA | 0.611(0.1243) | 0.6196(0.1066) | 0.7031(0.1485) | 0.7179(0.1565) | 0.5561(0.0481) | 0.5891(0.0454) |
| SpEED-QA+HDRMAX | 0.7581(0.0921) | 0.7107(0.0993) | 0.8597(0.0971) | 0.8355(0.0907) | 0.6519(0.0416) | 0.6642(0.0374) |
| VMAF | 0.6753(0.0493) | 0.6086(0.0583) | 0.5617(0.0919) | 0.5069(0.0844) | 0.6424(0.0574) | 0.7050(0.0498) |
| VMAF+HDRMAX | 0.8528(0.0543) | 0.8342(0.0632) | 0.8654(0.1076) | 0.8417(0.0996) | 0.7050(0.0853) | 0.7120(0.0944) |
| HDR-VDP-2 | 0.7041(0.1198) | 0.6722(0.1081) | - | - | - | - |
| HDR-VDP-2+HDRMAX | 0.7431(0.0770) | 0.7208(0.0764) | - | - | - | - |

[4] H. Sheikh and A. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, 2006.

[5] S. Li, F. Zhang, L. Ma, and K. N. Ngan, "Image quality assessment by separately evaluating detail losses and additive impairments," *IEEE Transactions on Multimedia*, vol. 13, no. 5, pp. 935–949, 2011.

[6] R. Soundararajan and A. C. Bovik, "Video quality assessment by reduced reference spatio-temporal entropic differencing," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 4, pp. 684–694, 2012.

[7] C. G. Bampis, P. Gupta, R. Soundararajan, and A. C. Bovik, "SpEED-QA: Spatial efficient entropic differencing for image and video quality," *IEEE Signal Process. Letters*, vol. 24, no. 9, pp. 1333–1337, 2017.

[8] Netflix, "VMAF: The Journey Continues," 2018 (accessed January 13, 2023). [Online]. Available: <https://netfixtechblog.com/vmaf-the-journey-continues-44b51ee9ed12>

[9] Z. Shang, J. P. Ebenezer, A. C. Bovik, Y. Wu, H. Wei, and S. Sethuraman, "Subjective assessment of high dynamic range videos under different ambient conditions," in *IEEE Intl. Conf. Image Process.(ICIP)*, 2022, pp. 786–790.

[10] D. Y. Lee, S. Paul, C. G. Bampis, H. Ko, J. Kim, S. Y. Jeong, B. Homan, and A. C. Bovik, "A subjective and objective study of space-time subsampled video quality," *IEEE Trans. Image Process.*, vol. 31, pp. 934–948, 2021.

[11] Z. Shang, J. P. Ebenezer, Y. Wu, H. Wei, S. Sethuraman, and A. C. Bovik, "Study of the subjective and objective quality of high motion live streaming videos," *IEEE Trans. Image Process.*, vol. 31, pp. 1027–1041, 2022.

[12] J. Korhonen, "Two-level approach for no-reference consumer video quality assessment," *IEEE Trans. Image Process.*, vol. 28, no. 12, pp. 5923–5938, 2019.

[13] Z. Tu, X. Yu, Y. Wang, N. Birkbeck, B. Adsumilli, and A. C. Bovik, "RAPIQUE: Rapid and accurate video quality prediction of user generated content," *IEEE Open J. Signal Process.*, vol. 2, pp. 425–440, 2021.

- [14] Z. Tu, Y. Wang, N. Birkbeck, B. Adsumilli, and A. C. Bovik, “UGC-VQA: Benchmarking blind video quality assessment for user generated content,” *IEEE Trans. Image Process.*, vol. 30, pp. 4449–4464, 2021.
- [15] M. A. Saad, A. C. Bovik, and C. Charrier, “Blind prediction of natural video quality,” *IEEE Trans. Image Process.*, vol. 23, no. 3, pp. 1352–1365, 2014.
- [16] D. Kundu, D. Ghadiyaram, A. C. Bovik, and B. L. Evans, “No-reference quality assessment of tone-mapped HDR pictures,” *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 2957–2971, 2017.
- [17] D. Li, T. Jiang, and M. Jiang, “Quality assessment of in-the-wild videos,” in *Proceedings of the 27th ACM International Conference on Multimedia*, 2019, pp. 2351–2359.
- [18] J. P. Ebenezer, Z. Shang, Y. Wu, H. Wei, S. Sethuraman, and A. C. Bovik, “ChipQA: No-reference video quality prediction via space-time chips,” *IEEE Trans. Image Process.*, vol. 30, pp. 8059–8074, 2021.
- [19] Z. Wang, E. P. Simoncelli, and A. C. Bovik, “Multiscale structural similarity for image quality assessment,” in *The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, vol. 2. Ieee, 2003, pp. 1398–1402.
- [20] R. Mantiuk, K. J. Kim, A. G. Rempel, and W. Heidrich, “HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions,” *ACM Trans. Graphics (TOG)*, vol. 30, no. 4, pp. 1–14, 2011.