Adapting Quality Metrics to Tone Mapping
Supplementary Material

KENNETH CHEN*, New York University, USA
DONGYEON KIM, University of Cambridge, United Kingdom
YUTA ASANO, Reality Labs Research, Meta, USA
ALEXANDRE CHAPIRO, Reality Labs Research, Meta, USA
QI SUN, New York University, USA
RAFAŁ K. MANTIUK, University of Cambridge, United Kingdom

This document provides supplementary material for the paper 'Adapting Quality Metrics to Tone Mapping'. The content includes additional details referenced in the main paper.

ACM Reference Format:
Kenneth Chen, Dongyeon Kim, Yuta Asano, Alexandre Chapiro, Qi Sun, and Rafał K. Mantiuk. 2026. Adapting Quality Metrics to Tone Mapping *Supplementary Material*. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers (SIGGRAPH Conference Papers '26), July 19–23, 2026, Los Angeles, CA, USA.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3799902.3811107

Contents

*All data access, collection, and experiments were performed by New York University and University of Cambridge.

Authors' Contact Information: Kenneth Chen, kennychen@nyu.edu, New York University, Brooklyn, New York, USA; Dongyeon Kim, dongyeon.kim93@gmail.com, University of Cambridge, Cambridge, United Kingdom; Yuta Asano, yasano@meta.com, Reality Labs Research, Meta, Redmond, USA; Alexandre Chapiro, alex@chapiro.net, Reality Labs Research, Meta, Sunnyvale, USA; Qi Sun, qisun@nyu.edu, New York University, Brooklyn, USA; Rafał K. Mantiuk, mantiuk@gmail.com, University of Cambridge, Cambridge, United Kingdom.

SIGGRAPH Conference Papers '26, Los Angeles, CA, USA

ACM ISBN 979-8-4007-2554-8/2026/07
https://doi.org/10.1145/3799902.3811107

S1 Additional Background and Related Work

High dynamic range (HDR) content has become increasingly prevalent. HDR content is typically encoded in a source container with a very large dynamic range, such as ITU-R Rec. 2100 [ITU-R 2025]. This range typically cannot be reproduced on a target display. Instead, *tone mapping*, an image processing operation that maps HDR content into a range suitable for the target displays' dynamic range, is employed. Consequently, ensuring the perceptual accuracy of tone-mapped content in relation to the HDR reference is key to maintaining high display quality.

S1.1 What is tone mapping?

A diversity of approaches to tone mapping have been proposed, including heuristics based on photographic techniques [Reinhard et al. 2002], perceptually-inspired methods [Pattanaik et al. 1998; Tariq et al. 2023], algorithms that adapt to display parameters [Chen et al. 2023, 2025; Mantiuk et al. 2008], learning-based techniques [Rana et al. 2020; Vinker et al. 2021], and more. Tone mapping operators have also been developed specifically for HDR video [Eilertsen et al. 2015, 2017] or stereo displays [Zhong et al. 2019]. While these distortion types are manifold, in this work we argue that the problem of accounting for the absolute luminance difference between HDR reference and tone-mapped test is the main concern when predicting quality.

S2 Metrics

We show, in Table S1, an extended version of the table of quality metrics in the main manuscript, which lists all quality metrics evaluated in this paper.

S3 Scenes

Here, we describe the scenes used in our tone mapping quality assessment study in more detail:

Anime. This is a frame from Sol Levante, which is an animated film. The frame includes a bright character and local highlights in the surroundings. The image comes from the Netflix open source content.

Man. This is a close-up of a man's face from the Meridian film. The image contains highlights in the eyes, and details in the man's skin and hair. The image comes from the Netflix open source content.

Table S1. *Summary of quality metrics.* We summarize the quality metrics studied in this work. This table includes the strategy employed by the metric and whether the input types to the metric are photometric.

Metrics	e	s	s	n	f	v	p	t	Ref.	Test
PSNR (RGB)	✔	✘	✘	✘	✘	✘	✘	✘	✘	✘
PSNR (Y)	✔	✘	✘	✘	✘	✘	✘	✘	✔	✔
CIE ΔE 2000 [CIE 2018]	✔	✘	✘	✘	✘	✘	✘	✘	✔	✔
IC_tC_p [Pytlarz and Atkins 2023]	✔	✘	✘	✘	✘	✘	✘	✘	✔	✔
HyAB [Abasi et al. 2020]	✔	✘	✘	✘	✘	✘	✘	✘	✔	✔
Spatial ΔE 2000	✔	✔	✘	✘	✘	✘	✘	✘	✔	✔
sCIELab [Zhang and Wandell 1997]	✔	✔	✘	✘	✘	✘	✘	✘	✘	✘
ΔE^S_{ITP} [Choudhury et al. 2021]	✔	✔	✘	✘	✘	✘	✘	✘	✔	✔
DSS [Balanov et al. 2015]	✘	✔	✘	✘	✘	✘	✘	✘	✘	✘
MDSI [Ziaei Nafchi et al. 2016]	✘	✔	✘	✘	✘	✘	✘	✘	✘	✘
GMSD [Xue et al. 2013]	✘	✔	✘	✘	✘	✘	✘	✘	✘	✘
MS-GMSD [Zhang et al. 2017]	✘	✔	✘	✘	✘	✘	✘	✘	✘	✘
SSIM [Wang et al. 2004]	✘	✔	✘	✘	✘	✘	✘	✘	✘	✘
MS-SSIM [Wang et al. 2003]	✘	✔	✘	✘	✘	✘	✘	✘	✘	✘
IW-SSIM [Wang and Li 2011]	✘	✔	✘	✘	✘	✘	✘	✘	✘	✘
FSIM [Zhang et al. 2011]	✘	✔	✘	✘	✘	✘	✘	✘	✘	✘
FSIMc [Zhang et al. 2011]	✘	✔	✘	✘	✘	✘	✘	✘	✘	✘
NLPD [Laparra et al. 2017]	✘	✔	✘	✘	✘	✘	✘	✘	✘	✘
HaarPSI [Reisenhofer et al. 2018]	✘	✔	✘	✘	✘	✘	✘	✘	✘	✘
WaDIQaM [Bosse et al. 2017]	✘	✔	✘	✘	✘	✘	✘	✘	✘	✘
MS-SWD [He et al. 2024]	✘	✔	✔	✘	✘	✘	✘	✘	✘	✘
VIF [Sheikh and Bovik 2006]	✔	✘	✔	✘	✘	✘	✘	✘	✘	✘
VSI [Zhang et al. 2014]	✔	✘	✔	✘	✘	✘	✘	✘	✘	✘
BRISQUE [Mittal et al. 2012a]	✘	✘	✔	✔	✘	✘	✘	✘	–	✘
NIQE [Mittal et al. 2012b]	✘	✘	✔	✔	✘	✘	✘	✘	–	✘
PIQE [Venkatanath et al. 2015]	✘	✔	✔	✔	✘	✘	✘	✘	–	✘
MANIQA [Yang et al. 2022]	✘	✘	✘	✔	✔	✘	✘	✘	–	✘
TOPIQ-NR [Chen et al. 2024]	✘	✘	✘	✔	✔	✘	✘	✘	–	✘
MUSIQ [Ke et al. 2021]	✘	✘	✘	✔	✔	✘	✘	✘	–	✘
CKDN [Zheng et al. 2021]	✘	✘	✘	✘	✔	✘	✘	✘	✘	✘
CLIP-IQA [Wang et al. 2023]	✘	✘	✘	✘	✔	✘	✘	✘	✘	✘
DISTS [Ding et al. 2022]	✘	✘	✘	✘	✔	✘	✘	✘	✘	✘
IQT [Cheon et al. 2021]	✘	✘	✘	✘	✔	✘	✘	✘	✘	✘
MILO [Çoğalan et al. 2025]	✘	✘	✘	✘	✔	✘	✘	✘	✘	✘
LPIPS (AlexNet) [Zhang et al. 2018]	✘	✘	✘	✘	✔	✘	✘	✘	✘	✘
LPIPS (VGG) [Zhang et al. 2018]	✘	✘	✘	✘	✔	✘	✘	✘	✘	✘
PieAPP [Prashnani et al. 2018]	✘	✘	✘	✘	✔	✘	✘	✘	✘	✘
ST-LPIPS (AlexNet)	✘	✘	✘	✘	✔	✘	✘	✘	✘	✘
TOPIQ [Chen et al. 2024]	✘	✘	✘	✘	✔	✘	✘	✘	✘	✘
AHIQ [Lao et al. 2022]	✘	✘	✘	✘	✔	✘	✘	✘	✘	✘
STRRED [Soundararajan and Bovik 2012]	✘	✘	✔	✘	✘	✔	✘	✘	✘	✘
VMAF [Li et al. 2018]	✘	✔	✘	✘	✘	✔	✘	✘	✘	✘
CGVQM [Jindal et al. 2025]	✘	✘	✘	✘	✔	✔	✘	✘	✘	✘
FLIP [Andersson et al. 2020]	✔	✘	✘	✘	✔	✘	✔	✘	✘	✘
HDR-FLIP [Andersson et al. 2021]	✔	✘	✘	✘	✔	✘	✔	✘	✔	✔
HDR-VDP-3 [Mantiuk et al. 2023]	✘	✘	✘	✘	✘	✘	✔	✘	✔	✔
ColorVideoVDP [Mantiuk et al. 2024]	✘	✘	✘	✘	✘	✔	✔	✘	✔	✔
FFTMI [Krasula et al. 2020]	✘	✔	✘	✘	✘	✘	✘	✔	✔	✘
FSITM [Ziaei Nafchi et al. 2015]	✘	✔	✘	✘	✘	✘	✘	✔	✔	✘
FSITM-TMQI [Ziaei Nafchi et al. 2015]	✘	✔	✘	✘	✘	✘	✘	✔	✔	✘
TMQI [Yeganeh and Wang 2013]	✘	✔	✘	✘	✘	✘	✘	✔	✔	✘
CIVDM [Aydin et al. 2008]	✘	✘	✘	✘	✘	✘	✔	✔	✔	✔
ColorVideoVDP-tm	✘	✘	✘	✘	✘	✔	✔	✔	✔	✔

Woman. This includes a shot of a showgirl in front of a mirror with bright bulbs. Her dress contains bright, reflective sequins. The image comes from Froehlich et al. [2014].

Lounge. A photograph of a common room with a bright window and highlights reflecting off a glossy chess board. This image was taken by the authors.

Toys. This is a photograph of small plush toys with high-frequency details under a bright lamp. It includes a cube with specular highlights. This image was taken by the authors.

Type. This is a photograph of a typewriter with text on a paper illuminated by a bright lamp. This image is from Fairchild [2007].

Bloom. This is a photograph of a detailed flower bloom in the foreground with a bright sky. This image is from Fairchild [2007].

Water. This is a photograph of a waterfall with bright sky region and detail in the foreground. This image is from Fairchild [2007].

Glade. This is a photograph of a grassy glade with a bright sky. This image was taken by the authors.

Rushmore. This is a photograph of Mt. Rushmore with bright sky and dark shadows. This image is from Fairchild [2007].

Snow. This is a photograph of bright snow and dark trees with highlights in the snow and a bright region behind the trees. This image is from Fairchild [2007].

College. This is a photograph of a university building with intricate details, in front of a bright sky. This image was taken by the authors.

S4 Tone Mapper

Here, we show in additional detail our tone mapping operator as a flow diagram in Figure S1. Color coding matches the respective inputs/outputs of each function. The steps of our tone mapper, in detail, are as follows:

(1) *Exposure adjustment* maps the HDR image to an SDR range,

$$\mathbf{I}'_{(x,y)} = \log_2\left(\mathbf{I}_{(x,y)}\right) - \texttt{percentile}(\mathbf{I}, p), \tag{S1}$$

where p is the p^{th} percentile photometric value. In this work, we used the 95^{th} percentile.

(2) *Tone encoding* via cross-channel maximum,

$$\mathbf{T}_{(x,y)} = \max(\mathbf{c}_R, \mathbf{c}_G, \mathbf{c}_B),\ \mathbf{c} = \mathbf{I}'_{(x,y)} \tag{S2}$$

where $\mathbf{T}$ is what we refer to as *tone*. In prior works, tone can be encoded as relative luminance. Chen et al. [2025] showed that encoding tone as the max of photometric values avoids color clipping in bright pixels.

(3) *Smooth clipping* of values greater than 1 after the exposure adjustment step,

$$\mathbf{T}'_{(x,y)} = \mathcal{G}\left(\mathbf{T}_{(x,y)}, C, t_s, t_e\right), \tag{S3}$$

where $\mathcal{G}(\cdot)$ is a spline that takes as input the exposure-adjusted tone and applies a smooth roll-off to highlights, with parameters t_s and t_e representing the start and end of the roll-off, respectively, and C representing the slope of the curve (contrast).

(4) *Contrast compression* (optionally) of component images with the bilateral filtering technique of Durand and Dorsey [2002],

$$\mathbf{T}_i = \mathcal{B}(\mathbf{T}', \sigma_s, \sigma_r, n), \tag{S4}$$

where $\mathbf{T}_i$ is the illumination layer of tone, and σ_s, σ_r, n are parameters controlling the effect of the bilateral filter. In our implementation, we used $\sigma_s = 0.02 \cdot \max(H, W)$ and $\sigma_r = 2$, where H, W are image dimensions. The reflectance layer is computed by separating the illumination layer from tone,

$$\mathbf{T}_r = \mathbf{T}' - \mathbf{T}_i. \tag{S5}$$

The contrast of each component is compressed by applying a power function (multiplication in the logarithmic domain),

$$\mathbf{T}' = \gamma_i \cdot \mathbf{T}_i + \gamma_r \cdot \mathbf{T}_r. \tag{S6}$$

Increasing γ_i is equivalent to boosting overall contrast, while γ_r sharpens the image (improves contrast of details).

(5) *Color reconstruction* is done with the technique of Mantiuk et al. [2009],

$$\mathbf{c}' = s \cdot \left(\mathbf{T}'_{(x,y)} - \mathbf{T}_{(x,y)}\right) + \mathbf{c}, \tag{S7}$$

which adjusts colors so that original color ratios are preserved, with s controlling the extent of this adjustment. Modulating s is equivalent to modifying the saturation of the image. The image is then brought back to the linear domain,

$$\mathbf{c}' = 2^{\mathbf{c}'}. \tag{S8}$$

S5 Metric Optimization Study Parameters

Here, we describe the details of the metric optimization study. A non-uniform patter search algorithm was run on the tone mapping parameters. We used the MATLAB `patternsearch` algorithm with `"nups"` setting for optimization. As discussed in the main body, we optimize three parameters: the exposure, contrast, and smooth clipping range ($\mathbf{k} = \left[p, C, t_e - t_s\right]$). The initial parameters are $\mathbf{k} = \left[95, 0.8, 1\right]$, and the upper and lower bounds were $\left[87.5, 0.6, 0.1\right]$ and $\left[95, 1, 5\right]$, respectively. These bounds were chosen so that no extreme parameter values were selected during the optimization.

S6 Additional Evaluation Results

We showed the per-scene and per-transfer function results individually, with the linear encoding in Figure S2, PQ in Figure S3, PU21 in Figure S4, μ-law in Figure S5, and PU21-PU21 in Figure S6. Note that psychophysics-based metrics like ColorVideoVDP-tm, ColorVideoVDP, and HDR-VDP-3 perform significantly better on average for all encodings except PU21-PU21 because they do not require a specific display encoding to work.

References

Saeedeh Abasi, Mohammad Amani Tehran, and Mark D. Fairchild. 2020. Distance metrics for very large color differences. *Color Research & Application* 45, 2 (2020), 208–223. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/col.22451 doi:10.1002/col.22451

Pontus Andersson, Jim Nilsson, Tomas Akenine-Möller, Magnus Oskarsson, Kalle Åström, and Mark D. Fairchild. 2020. FLIP: A Difference Evaluator for Alternating Images. *Proc. ACM Comput. Graph. Interact. Tech.* 3, 2, Article 15 (Aug. 2020), 23 pages. doi:10.1145/3406183

Pontus Andersson, Jim Nilsson, Peter Shirley, and Tomas Akenine-Möller. 2021. Visualizing errors in rendered high dynamic range images. In *Eurographics-Short Papers*. Eurographics-European Association for Computer Graphics, 25–28.

Tunç Ozan Aydin, Rafał Mantiuk, Karol Myszkowski, and Hans-Peter Seidel. 2008. Dynamic range independent image quality assessment. *ACM Trans. Graph.* 27, 3 (Aug. 2008), 1–10. doi:10.1145/1360612.1360668

Amnon Balanov, Arik Schwartz, Yair Moshe, and Nimrod Peleg. 2015. Image quality assessment based on DCT subband similarity. In *2015 IEEE international conference on image processing (ICIP)*. IEEE, 2105–2109.

Sebastian Bosse, Dominique Maniry, Klaus-Robert Müller, Thomas Wiegand, and Wojciech Samek. 2017. Deep neural networks for no-reference and full-reference image quality assessment. *IEEE Transactions on image processing* 27, 1 (2017), 206–219.

Bin Chen, Akshay Jindal, Michal Piovarči, Chao Wang, Hans-Peter Seidel, Piotr Didyk, Karol Myszkowski, Ana Serrano, and Rafał K. Mantiuk. 2023. The effect of display capabilities on the gloss consistency between real and virtual objects. In *SIGGRAPH Asia 2023 Conference Papers* (Sydney, NSW, Australia) *(SA '23)*. Association for Computing Machinery, New York, NY, USA, Article 90, 11 pages. doi:10.1145/3610548.3618226

Chaofeng Chen, Jiadi Mo, Jingwen Hou, Haoning Wu, Liang Liao, Wenxiu Sun, Qiong Yan, and Weisi Lin. 2024. TOPIQ: A Top-Down Approach From Semantics to Distortions for Image Quality Assessment. *IEEE Transactions on Image Processing* 33 (2024), 2404–2418. doi:10.1109/TIP.2024.3378466

Kenneth Chen, Nathan Matsuda, Jon McElvain, Yang Zhao, Thomas Wan, Qi Sun, and Alexandre Chapiro. 2025. What is HDR? Perceptual Impact of Luminance and Contrast in Immersive Displays. In *Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers (SIGGRAPH Conference Papers '25)*. Association for Computing Machinery, New York, NY, USA, Article 40, 11 pages. doi:10.1145/3721238.3730629

Manri Cheon, Sung-Jun Yoon, Byungyeon Kang, and Junwoo Lee. 2021. Perceptual Image Quality Assessment With Transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. 433–442.

Anustup Choudhury, Robert Wanat, Jaclyn Pytlarz, and Scott Daly. 2021. Image quality evaluation for high dynamic range and wide color gamut applications using visual spatial processing of color differences. *Color Research & Application* 46, 1 (2021), 46–64. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/col.22588 doi:10.1002/col.22588

CIE. 2018. CIE 015: 2018 Colorimetry. (2018).

Ugur Çoğalan, Mojtaba Bemana, Karol Myszkowski, Hans-Peter Seidel, and Colin Groth. 2025. MILO: A Lightweight Perceptual Quality Metric for Image and Latent-Space Optimization. *ACM Transactions on Graphics (TOG)* 44, 6 (2025). doi:10.1145/3763340

Keyan Ding, Kede Ma, Shiqi Wang, and Eero P. Simoncelli. 2022. Image Quality Assessment: Unifying Structure and Texture Similarity. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44, 5 (2022), 2567–2581. doi:10.1109/TPAMI.2020.3045810

Frédo Durand and Julie Dorsey. 2002. Fast bilateral filtering for the display of high-dynamic-range images. In *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques* (San Antonio, Texas) *(SIGGRAPH '02)*. Association for Computing Machinery, New York, NY, USA, 257–266. doi:10.1145/566570.566574

Gabriel Eilertsen, Rafał K. Mantiuk, and Jonas Unger. 2015. Real-time noise-aware tone mapping. *ACM Trans. Graph.* 34, 6, Article 198 (Nov. 2015), 15 pages. doi:10.1145/2816795.2818092

G. Eilertsen, R. K. Mantiuk, and J. Unger. 2017. A comparative review of tone-mapping algorithms for high dynamic range video. *Computer Graphics Forum* 36, 2 (2017), 565–592. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.13148 doi:10.1111/cgf.13148

Mark D. Fairchild. 2007. The HDR Photographic Survey. *Color and Imaging Conference* 15, 1 (2007), 233–233. doi:10.2352/CIC.2007.15.1.art00044

Jan Froehlich, Stefan Grandinetti, Bernd Eberhardt, Simon Walter, Andreas Schilling, and Harald Brendel. 2014. Creating cinematic wide gamut HDR-video for the evaluation of tone mapping operators and HDR-displays. In *Digital Photography X*, Nitin Sampat, Radka Tezaur, Sebastiano Battiato, and Boyd A. Fowler (Eds.), Vol. 9023. International Society for Optics and Photonics, SPIE, 90230X. doi:10.1117/12.2040003

Jiaqi He, Zhihua Wang, Leon Wang, Tsein-I Liu, Yuming Fang, Qilin Sun, and Kede Ma. 2024. Multiscale Sliced Wasserstein Distances as Perceptual Color Difference Measures. In *European Conference on Computer Vision*. 1–18. http://arxiv.org/abs/2407.10181

ITU-R. 2025. *ITU-R BT.2100: Image Parameter Values for High Dynamic Range Television for Use in Production and International Programme Exchange*. ITU-R Recommendation BT.2100-3. International Telecommunication Union, Radiocommunication Sector. https://www.itu.int/rec/R-REC-BT.2100/

Akshay Jindal, Nabil Sadaka, Manu Mathew Thomas, Anton Sochenov, and Anton Kaplanyan. 2025. CGVQM+D: Computer Graphics Video Quality Metric and Dataset. *Computer Graphics Forum* (2025). doi:10.1111/cgf.70221

Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. 2021. MUSIQ: Multi-Scale Image Quality Transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 5148–5157.

Lukáš Krasula, Karel Fliegel, and Patrick Le Callet. 2020. FFTMI: Features Fusion for Natural Tone-Mapped Images Quality Evaluation. *IEEE Transactions on Multimedia* 22, 8 (2020), 2038–2047. doi:10.1109/TMM.2019.2952256

Shanshan Lao, Yuan Gong, Shuwei Shi, Sidi Yang, Tianhe Wu, Jiahao Wang, Weihao Xia, and Yujiu Yang. 2022. Attentions Help CNNs See Better: Attention-based Hybrid Image Quality Assessment Network. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 1139–1148. doi:10.1109/CVPRW56347.2022.00123

Valero Laparra, Alexander Berardino, Johannes Ballé, and Eero P. Simoncelli. 2017. Perceptually optimized image rendering. *J. Opt. Soc. Am. A* 34, 9 (Sep 2017), 1511–1525. doi:10.1364/JOSAA.34.001511

Zhi Li, Christos Bampis, Julie Novak, Anne Aaron, Kyle Swanson, Anush Moorthy, and JD Cock. 2018. VMAF: The journey continues. *Netflix Technology Blog* 25, 1 (2018).

Rafał Mantiuk, Scott Daly, and Louis Kerofsky. 2008. Display adaptive tone mapping. In *ACM SIGGRAPH 2008 Papers* (Los Angeles, California) *(SIGGRAPH '08)*. Association

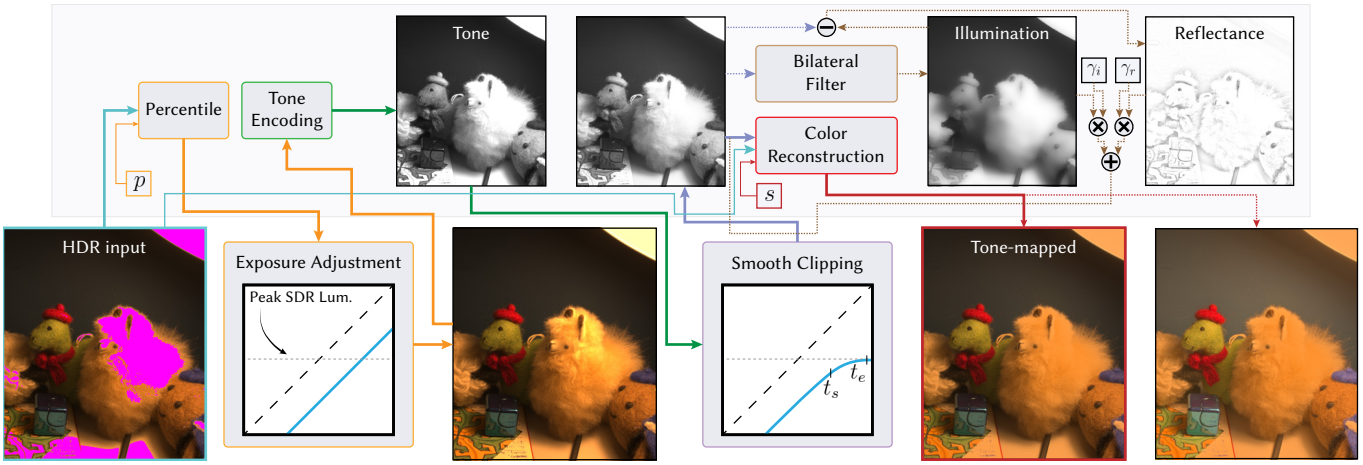

Fig. S1. *Tone mapper schematic diagram.* The generic tone mapping operator is described in detail here as a flow diagram. First, an HDR image is exposure-adjusted by its p-th percentile photometric value. Tone is then extracted, and then smooth clipped. The clipped tone is then optionally decomposed, using a bilateral filter, into illumination and reflectance components. The contrast of these components are then adjusted, and combined to form the modified tone component. Finally, color is reconstructed with saturation adjustment.

for Computing Machinery, New York, NY, USA, Article 68, 10 pages. doi:10.1145/1399504.1360667

Rafał Mantiuk, Radosław Mantiuk, Anna Tomaszewska, and Wolfgang Heidrich. 2009. Color correction for tone mapping. *Computer Graphics Forum* 28, 2 (2009), 193–202. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-8659.2009.01358.x doi:10.1111/j.1467-8659.2009.01358.x

Rafal K Mantiuk, Dounia Hammou, and Param Hanji. 2023. HDR-VDP-3: A multi-metric for predicting image differences, quality and contrast distortions in high dynamic range and regular content. *arXiv preprint arXiv:2304.13625* (2023).

Rafal K. Mantiuk, Param Hanji, Maliha Ashraf, Yuta Asano, and Alexandre Chapiro. 2024. ColorVideoVDP: A visual difference predictor for image, video and display distortions. *ACM Trans. Graph.* 43, 4, Article 129 (July 2024), 20 pages. doi:10.1145/3658144

Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. 2012a. No-Reference Image Quality Assessment in the Spatial Domain. *IEEE Transactions on Image Processing* 21, 12 (2012), 4695–4708. doi:10.1109/TIP.2012.2214050

Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. 2012b. Making a "completely blind" image quality analyzer. *IEEE Signal processing letters* 20, 3 (2012), 209–212.

Sumanta N. Pattanaik, James A. Ferwerda, Mark D. Fairchild, and Donald P. Greenberg. 1998. A multiscale model of adaptation and spatial vision for realistic image display. In *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '98)*. Association for Computing Machinery, New York, NY, USA, 287–298. doi:10.1145/280814.280922

Ekta Prashnani, Hong Cai, Yasamin Mostofi, and Pradeep Sen. 2018. PieAPP: Perceptual Image-Error Assessment Through Pairwise Preference. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Jaclyn Pytlarz and Robin Atkins. 2023. Perceptual Color Volume Measuring the Distinguishable Colors of HDR and WCG Displays. (2023).

Aakanksha Rana, Praveer Singh, Giuseppe Valenzise, Frederic Dufaux, Nikos Komodakis, and Aljosa Smolic. 2020. Deep Tone Mapping Operator for High Dynamic Range Images. *IEEE Transactions on Image Processing* 29 (2020), 1285–1298. doi:10.1109/TIP.2019.2936649

Erik Reinhard, Michael Stark, Peter Shirley, and James Ferwerda. 2002. Photographic tone reproduction for digital images. *ACM Trans. Graph.* 21, 3 (July 2002), 267–276. doi:10.1145/566654.566575

Rafael Reisenhofer, Sebastian Bosse, Gitta Kutyniok, and Thomas Wiegand. 2018. A Haar wavelet-based perceptual similarity index for image quality assessment. *Signal Processing: Image Communication* 61 (2018), 33–43. doi:10.1016/j.image.2017.11.001

Hamid R Sheikh and Alan C Bovik. 2006. Image information and visual quality. *IEEE Transactions on image processing* 15, 2 (2006), 430–444.

Rajiv Soundararajan and Alan C Bovik. 2012. Video quality assessment by reduced reference spatio-temporal entropic differencing. *IEEE Transactions on Circuits and Systems for Video Technology* 23, 4 (2012), 684–694.

Taimoor Tariq, Nathan Matsuda, Eric Penner, Jerry Jia, Douglas Lanman, Ajit Ninan, and Alexandre Chapiro. 2023. Perceptually Adaptive Real-Time Tone Mapping. In *SIGGRAPH Asia 2023 Conference Papers* (Sydney, NSW, Australia) *(SA '23)*. Association for Computing Machinery, New York, NY, USA, Article 36, 10 pages. doi:10.1145/3610548.3618222

Narasimhan Venkatanath, D Praneeth, S Channappayya Sumohana, S Medasani Swarup, et al. 2015. Blind image quality evaluation using perception based features. In *2015 twenty first national conference on communications (NCC)*. IEEE, 1–6.

Yael Vinker, Inbar Huberman-Spiegelglas, and Raanan Fattal. 2021. Unpaired Learning for High Dynamic Range Image Tone Mapping. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 14657–14666.

Jianyi Wang, Kelvin C.K. Chan, and Chen Change Loy. 2023. Exploring CLIP for Assessing the Look and Feel of Images. *Proceedings of the AAAI Conference on Artificial Intelligence* 37, 2 (Jun. 2023), 2555–2563. doi:10.1609/aaai.v37i2.25353

Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing* 13, 4 (2004), 600–612. doi:10.1109/TIP.2003.819861

Zhou Wang and Qiang Li. 2011. Information Content Weighting for Perceptual Image Quality Assessment. *IEEE Transactions on Image Processing* 20, 5 (2011), 1185–1198. doi:10.1109/TIP.2010.2092435

Zhou Wang, Eero P Simoncelli, and Alan C Bovik. 2003. Multiscale structural similarity for image quality assessment. In *The thrity-seventh asilomar conference on signals, systems & computers, 2003*, Vol. 2. Ieee, 1398–1402.

Wufeng Xue, Lei Zhang, Xuanqin Mou, and Alan C Bovik. 2013. Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. *IEEE transactions on image processing* 23, 2 (2013), 684–695.

Sidi Yang, Tianhe Wu, Shuwei Shi, Shanshan Lao, Yuan Gong, Mingdeng Cao, Jiahao Wang, and Yujiu Yang. 2022. MANIQA: Multi-dimension Attention Network for

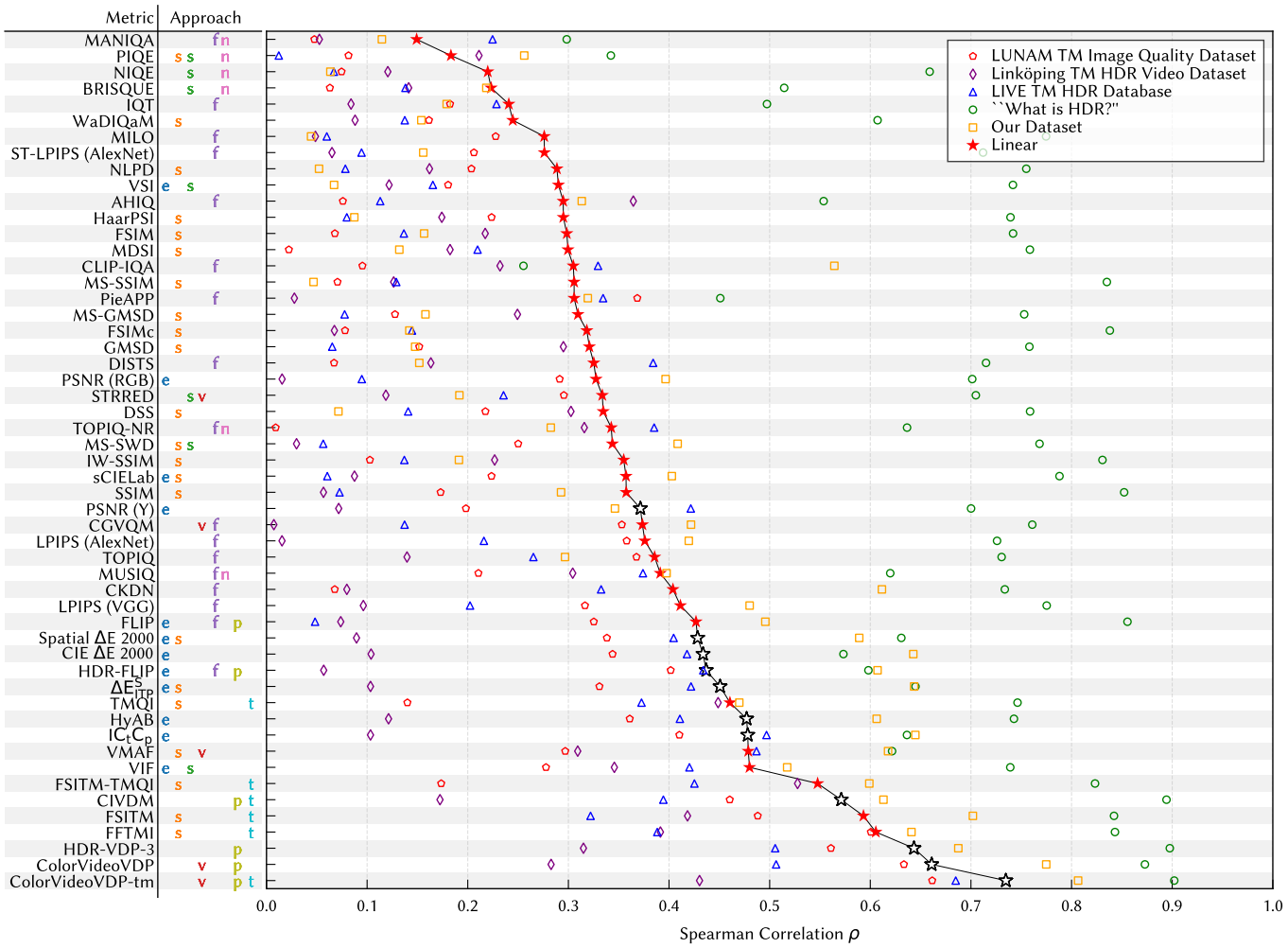

Fig. S2. *Metric evaluation of linear encoding.* Here, we plot the Spearman correlation coefficients for the linear encoding. Individual datasets are plotted as unfilled markers. Metric approaches include error-based (e), structural (s), statistical (s), no-reference (n), feature-based (f), video (v, psychophysical (p), and tone mapping (t) metrics.

No-Reference Image Quality Assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 1191–1200.

Hojatollah Yeganeh and Zhou Wang. 2013. Objective Quality Assessment of Tone-Mapped Images. *IEEE Transactions on Image Processing* 22, 2 (2013), 657–667. doi:10.1109/TIP.2012.2221725

Bo Zhang, Pedro V Sander, and Amine Bermak. 2017. Gradient magnitude similarity deviation on multiple scales for color image quality assessment. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 1253–1257.

Lin Zhang, Ying Shen, and Hongyu Li. 2014. VSI: A visual saliency-induced index for perceptual image quality assessment. *IEEE Transactions on Image processing* 23, 10 (2014), 4270–4281.

Lin Zhang, Lei Zhang, Xuanqin Mou, and David Zhang. 2011. FSIM: A feature similarity index for image quality assessment. *IEEE transactions on Image Processing* 20, 8 (2011), 2378–2386.

Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. 2018. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

X. Zhang and B. A. Wandell. 1997. A spatial extension of CIELAB for digital color-image reproduction. *Journal of the Society for Information Display* 5, 1 (1997), 61–63. arXiv:https://sid.onlinelibrary.wiley.com/doi/pdf/10.1889/1.1985127 doi:10.1889/1.1985127

Heliang Zheng, Huan Yang, Jianlong Fu, Zheng-Jun Zha, and Jiebo Luo. 2021. Learning Conditional Knowledge Distillation for Degraded-Reference Image Quality Assessment. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 10242–10251.

Fangcheng Zhong, George Alex Koulieris, George Drettakis, Martin S. Banks, Mathieu Chambe, Frédo Durand, and Rafał K. Mantiuk. 2019. DiCE: dichoptic contrast enhancement for VR and stereo displays. *ACM Trans. Graph.* 38, 6, Article 211 (Nov. 2019), 13 pages. doi:10.1145/3355089.3356552

Hossein Ziaei Nafchi, Atena Shahkolaei, Reza Farrahi Moghaddam, and Mohamed Cheriet. 2015. FSITM: A Feature Similarity Index For Tone-Mapped Images. *IEEE Signal Processing Letters* 22, 8 (2015), 1026–1029. doi:10.1109/LSP.2014.2381458

Hossein Ziaei Nafchi, Atena Shahkolaei, Rachid Hedjam, and Mohamed Cheriet. 2016. Mean Deviation Similarity Index: Efficient and Reliable Full-Reference Image Quality Evaluator. *IEEE Access* 4 (2016), 5579–5590. doi:10.1109/ACCESS.2016.2604042

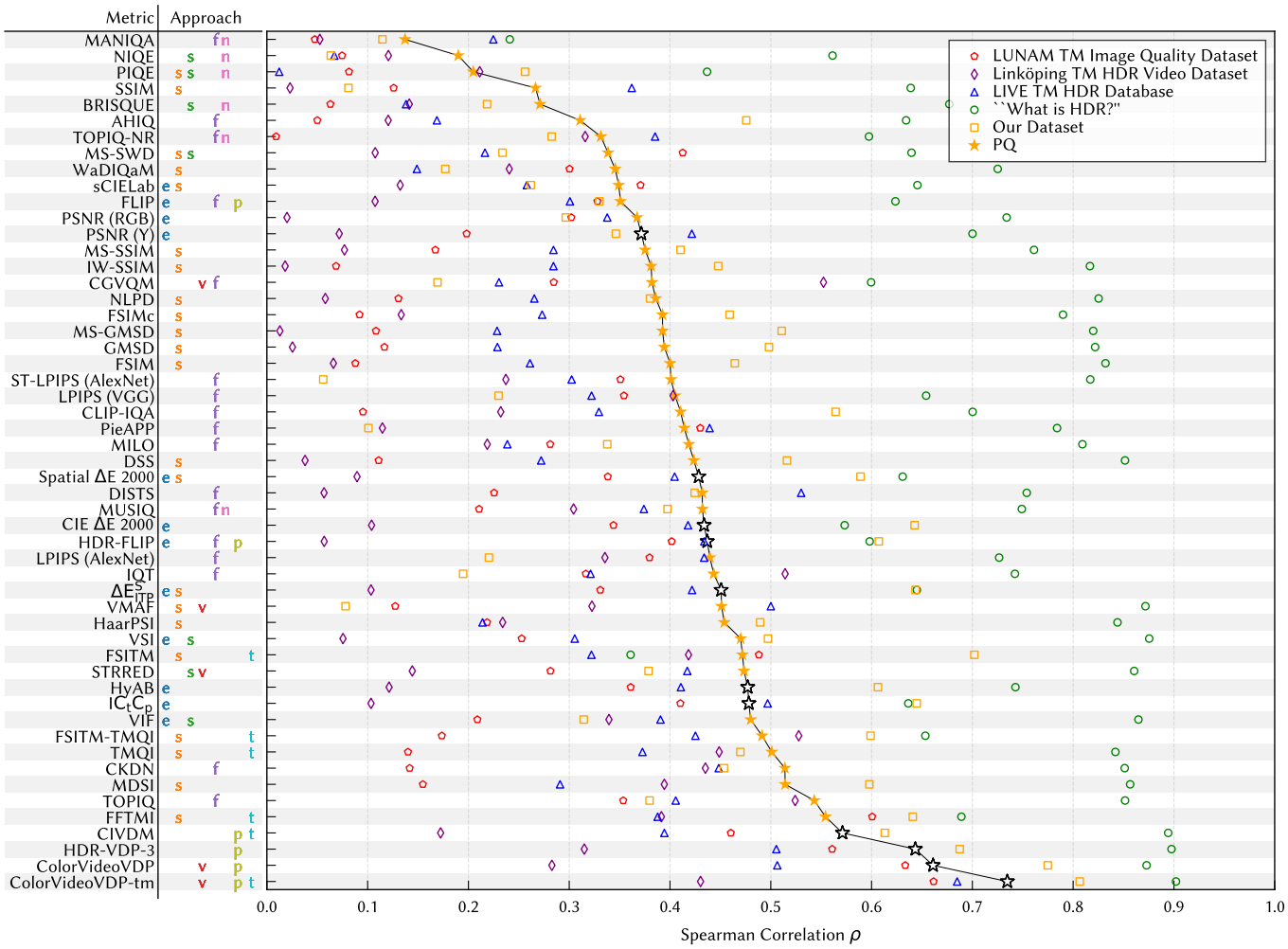

Fig. S3. *Metric evaluation of PQ encoding.* Here, we plot the Spearman correlation coefficients for the PQ encoding. Individual datasets are plotted as unfilled markers. Metric approaches include error-based (e), structural (s), statistical (s), no-reference (n), feature-based (f), video (v, psychophysical (p), and tone mapping (t) metrics.

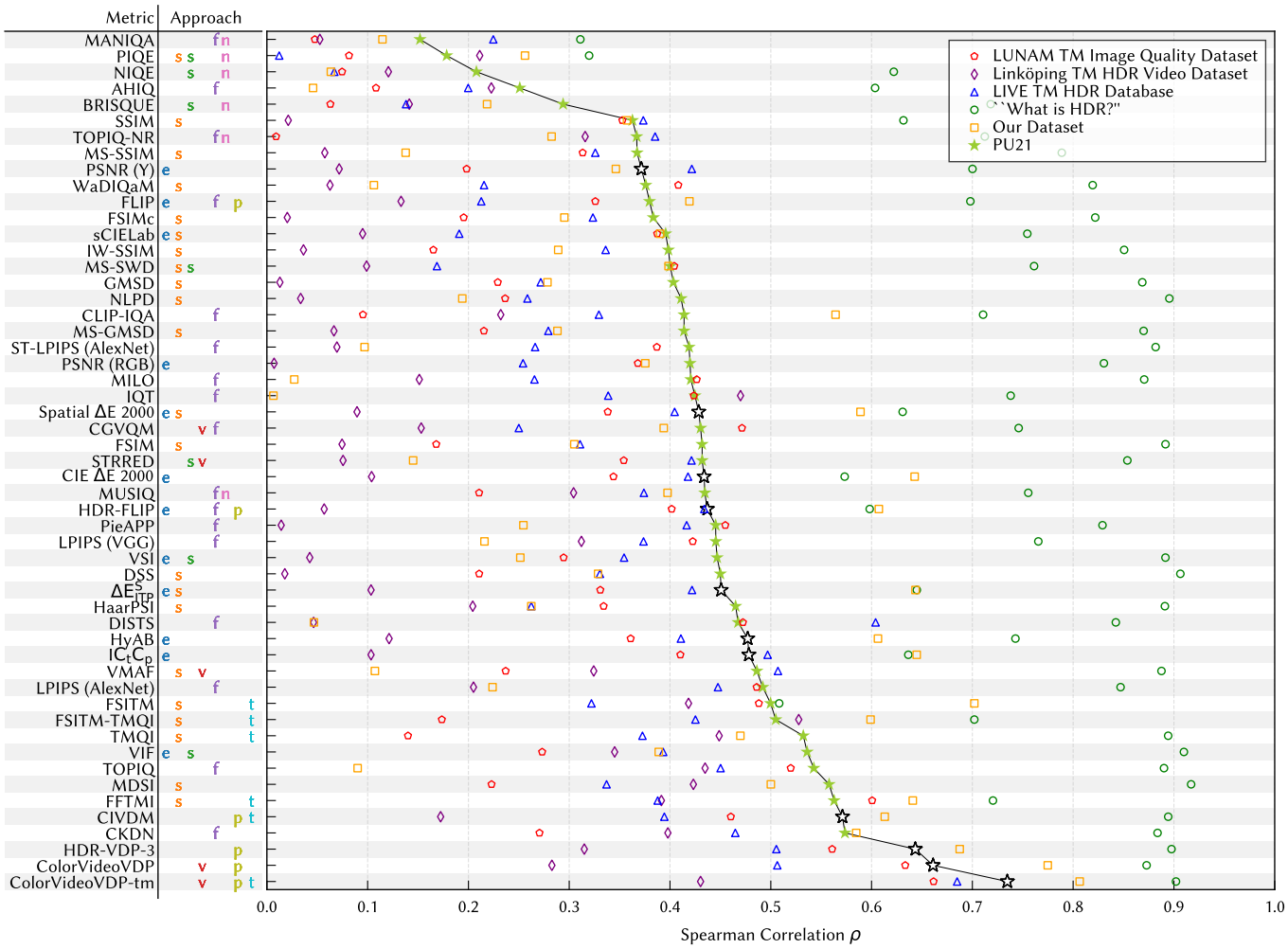

Fig. S4. *Metric evaluation of PU21 encoding.* Here, we plot the Spearman correlation coefficients for the PU21 encoding. Individual datasets are plotted as unfilled markers. Metric approaches include error-based (e), structural (s), statistical (s), no-reference (n), feature-based (f), video (v, psychophysical (p), and tone mapping (t) metrics.

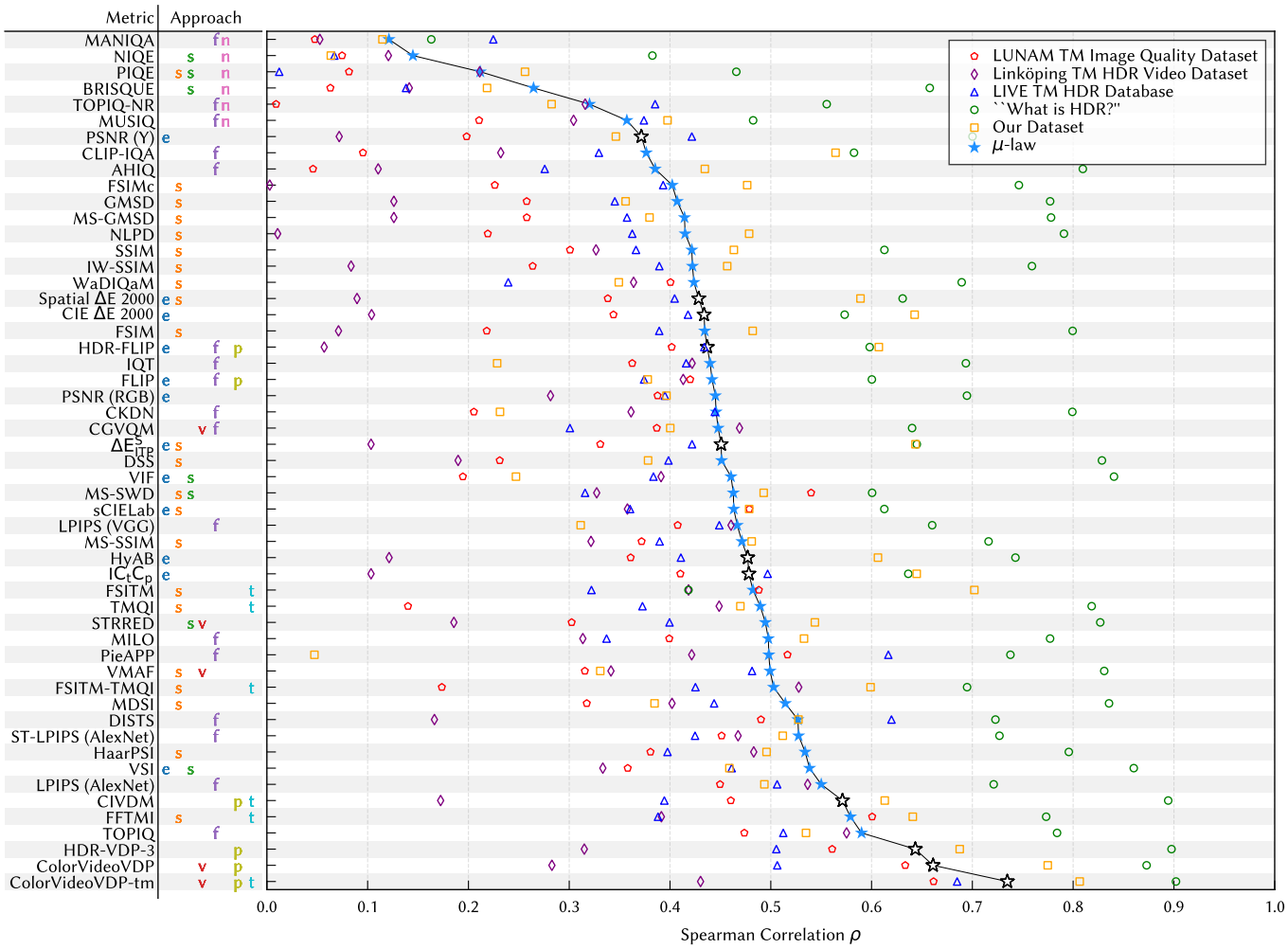

Fig. S5. *Metric evaluation of μ-law encoding.* Here, we plot the Spearman correlation coefficients for the μ-law encoding. Individual datasets are plotted as unfilled markers. Metric approaches include error-based (e), structural (s), statistical (s), no-reference (n), feature-based (f), video (v, psychophysical (p), and tone mapping (t) metrics.

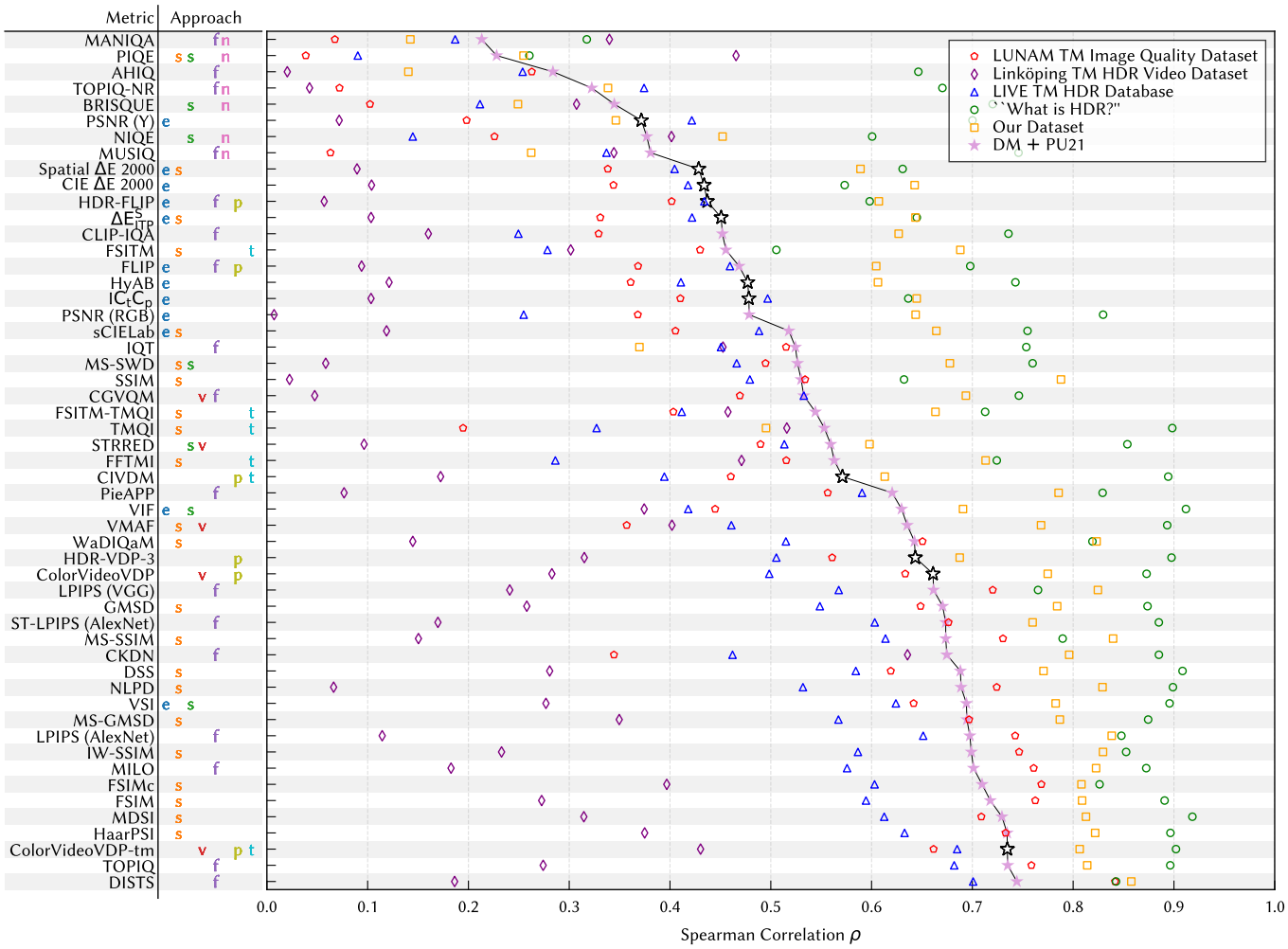

Fig. S6. *Metric evaluation of DM + PU21 encoding.* Here, we plot the Spearman correlation coefficients for the DM + PU21 encoding. Individual datasets are plotted as unfilled markers. Metric approaches include error-based (e), structural (s), statistical (s), no-reference (n), feature-based (f), video (v, psychophysical (p), and tone mapping (t) metrics.