Perceptual Assessment and Optimization of HDR Image Rendering

Peibei Cao[1], Rafał K. Mantiuk[2], and Kede Ma[1,3*]
[1] Department of Computer Science, City University of Hong Kong
[2] Department of Computer Science and Technology, University of Cambridge
[3] Shenzhen Research Institute, City University of Hong Kong
peibeicao2-c@my.cityu.edu.hk, kede.ma@cityu.edu.hk, rkm38@cam.ac.uk
https://github.com/cpb68/HDRQA/

Abstract

High dynamic range (HDR) rendering has the ability to faithfully reproduce the wide luminance ranges in natural scenes, but how to accurately assess the rendering quality is relatively underexplored. Existing quality models are mostly designed for low dynamic range (LDR) images, and do not align well with human perception of HDR image quality. To fill this gap, we propose a family of HDR quality metrics, in which the key step is employing a simple inverse display model to decompose an HDR image into a stack of LDR images with varying exposures. Subsequently, these decomposed images are assessed through well-established LDR quality metrics. Our HDR quality models present three distinct benefits. First, they directly inherit the recent advancements of LDR quality metrics. Second, they do not rely on human perceptual data of HDR image quality for re-calibration. Third, they facilitate the alignment and prioritization of specific luminance ranges for more accurate and detailed quality assessment. Experimental results show that our HDR quality metrics consistently outperform existing models in terms of quality assessment on four HDR image quality datasets and perceptual optimization of HDR novel view synthesis.

1. Introduction

High dynamic range (HDR) images aim to faithfully capture the large luminance variations of natural scenes that low dynamic range (LDR) images are not capable of [16]. In the past few years, numerous HDR imaging and display devices have been developed and commercialized in response to the escalating demand for HDR images in various fields, including photography, gaming, film, and virtual reality. Consequently, HDR image quality assessment (IQA) has become a practically demanding technique for monitoring, ensuring, and optimizing the perceptual quality of HDR images during imaging, compression, communication, and rendering.

At present, HDR quality metrics are largely lacking, which is likely due to the prevailing assumption that most LDR quality models, such as the peak signal-to-noise ratio (PSNR) and the structural similarity (SSIM) index [59], are readily applicable to HDR images. It was not until recently that researchers realized their poor account for human perception of HDR image quality [12, 14]. Mantiuk *et al*. [28] made initial attempts by extending the classic visual difference predictor (VDP) [10] to HDR-VDP, which was subsequently improved from various psychophysical and physiological perspectives [29, 31, 41]. Although the HDR-VDP family embodies many aspects of the early visual system, they contain complex and non-differentiable modules, which may hinder their application scope, especially when adopted as loss functions in perceptual optimization.

Other initiatives have focused on transforming linear luminances into a perceptually more uniform space as a way of improving the applicability of LDR quality metrics. Representative transformations include the logarithmic function [60], the perceptually uniform (PU) encoding curve [1] and its derivative, the PU21 encoding [25], and the perceptual quantizer [37]. The issue with perceptually uniform transformations lies in their tendency to either map luminance values that surpass the LDR image range (*e.g*., PU21 assigns high luminances to values above 255), or to compress the values to a range of $[0, 1]$ (*e.g*., the perceptual quantizer), which alters the image contrast. In the former case, image quality models incapable of handling values beyond the maximum pixel value (255 or 1) will fail to capture distortions in bright regions. In the latter case, the compressed contrast will cause unanticipated changes in the metric predictions.

Inspired by [38], we propose a family of full-reference HDR quality metrics, which rely on a simple inverse display model [26] to transform an (uncalibrated) HDR image to a

*Corresponding author.

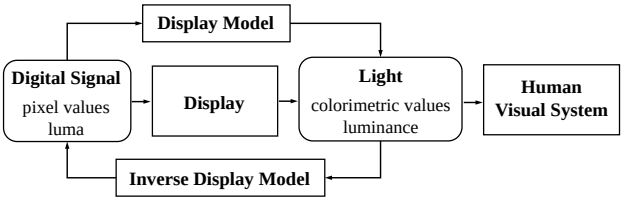


Figure 1. Forward display model simulates the process of converting digital pixel values into physical light in luminances on display. An inverse display model provides the inverse mapping.

stack of LDR images with varying exposures, amenable to LDR-IQA. Our HDR quality metrics offer several key advantages. First, they enjoy the latest developments of and reduce gracefully to LDR quality metrics, with the adoption of a complementary display model (see Fig. 1). Second, they do not need human perceptual data of HDR image quality for re-calibration. Third, they allow for the weighting of specific luminance ranges to highlight their contributions during quality assessment and perceptual optimization. Fourth, they enable the efficient mitigation of possible luminance shifts between the reference and test HDR images for more accurate quality assessment [12, 14]. Experimental results on four human-rated HDR-IQA datasets confirm the superior performance of our metrics, compared to existing models including the HDR-VDP family. We further demonstrate the promise of our HDR quality metrics as the perceptual optimization objectives in HDR novel view synthesis [17]. Importantly, we observe a significant improvement in visual quality for over-exposed regions, which is corroborated by subjective user studies and objective quality estimates.

2. Related Work

In this section, we review two bodies of studies that are related to ours, HDR-IQA and HDR novel view synthesis.

2.1. HDR Quality Metrics

Model-based methods rely on computational models that emulate the physiological responses of neurons in the human visual system, particularly those in the primary visual cortex. HDR-VDP [28] is an excellent example that takes into account aspects of nonlinear photoreceptor response to light, contrast sensitivity, and local adaptation. Similar to VDP [10], HDR-VDP predicts visible difference maps without supplying a numerical quality score. HDR-VDP-2 [29] improves upon HDR-VDP with a revised model of the early visual system. The metric was trained on two LDR-IQA datasets (*i.e.*, LIVE [51] and TID2008 [45]), and was later retrained on two additional HDR-IQA datasets: Narwaria2013 [39] and Narwaria2014 [40], leading to HDR-VDP-2.2 [41]. More recently, HDR-VDP-3 [31] was developed by simulating the impact of aging on the visual system [27], modeling the effect of adaptation to local luminances [57], and re-calibrating the metric on the largest HDR-IQA dataset, UPIQ [34]. Other representative model-based methods include the HDR video quality measure (HDR-VQM) [42] and the normalized Laplacian pyramid distance (NLPD) [20]. Like HDR-VDP, HDR-VQM follows an error visibility paradigm with the PU encoding as the front-end processing, while NLPD incorporates divisive normalization as a form of local gain control [7].

Encoding-based methods transform linear luminances into a perceptually more uniform space for subsequent processing. Xu *et al*. [60] approximated the luminance response curve as a logarithmic function. The PU encoding [1] was derived from the contrast sensitivity function (CSF) in [10], which was optimized to approximate the gamma-encoding in the range from 0.1 to $80\ \mathrm{cd/m^2}$. Similarly, the perceptual quantizer [37] was based on the Barten's CSF [4], and was standardized in the ITU-R Recommendation BT.2100. As an improved version, PU21 encoding [25] relies on a latest CSF [30], which predicts contrast thresholds at luminance levels between 0.0002 and $10,000\ \mathrm{cd/m^2}$. Nevertheless, the PU encoding is designed for the luminance channel only, and is less applicable to chromatic channels.

Our family of HDR quality metrics falls naturally in the category of encoding-based methods. Inspired by Munkberg *et al*. [38], we "encode" an HDR image into a multi-exposure LDR image stack for reliable LDR-IQA.

2.2. Novel View Synthesis

Novel view synthesis, a typical application of image-based rendering, involves generating images from novel viewpoints given a set of input views [52]. The view synthesis can be performed directly in the pixel domain when the input images are densely sampled [13, 21]. It is more common and economic to capture inputs from a wider range of sparse locations, which will be processed through a "proxy" geometry using either a heuristic [6] or learned blending function [15, 48, 49].

Of particular interest in this line of research is NeRF, which represents a scene with a neural radiance field [2, 5, 9, 22, 24, 32, 35, 61]. Mildenhall *et al*. [35] demonstrated that neural implicit representations yield superior results in view synthesis compared to traditional explicit representations such as point clouds, voxels, and octrees. Various aspects of NeRF have been improved, including rendering quality and capability [3], training and rendering efficiency [2, 33], robustness to varying illumination [32] and deformable objects [44], compositionality [43], editability [23], and generalization to novel scenes [8].

Recently, NeRF has been extended to work with HDR image data [17, 36]. Mildenhall *et al*. [36] trained Mip-NeRF [2] using linear noisy RAW images. Huang *et al*. [17] modeled the physical imaging process with two implicit

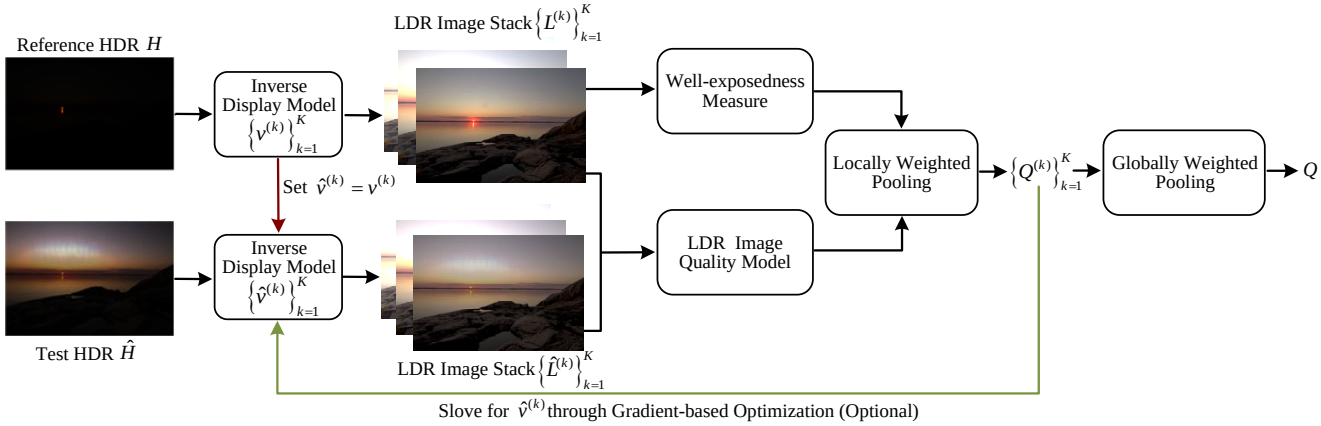


Figure 2. System diagram of the proposed family of HDR-IQA metrics. The default red arrow can be replaced by the optional green arrow, whose goal is to compensate for the possible luminance shifts between the reference and test HDR images, similar to the camera response function correction in [12, 14].

functions: a radiance field and a tone mapper, which are jointly optimized taking multiple LDR images with different exposures as inputs. In this paper, we simplify Huang's method [17] by stripping off the tone mapper and directly optimizing the RAW radiance field guided by the proposed HDR quality metrics.

3. Proposed HDR Quality Metrics

In this section, we propose to transform the problem of HDR-IQA into LDR-IQA, with the help of a simple inverse display model [26]. Fig. 2 shows the system diagram of the proposed family of HDR-IQA metrics.

3.1. Inverse Display Model

A forward display model simulates how the display transforms digital pixel values to physical units of light, while the opposite mapping from physical units to digital values, is referred to as an inverse display model, as illustrated in Fig. 1. Here, we resort to an inverse gain-offset-gamma display model [26]:

$$L^{(k)} = \left(\left[\frac{H \cdot v^{(k)} - b}{1 - b} \right]_0^1 \right)^{\frac{1}{\gamma}}, \quad 1 \le k \le K, \qquad (1)$$

where $v^{(k)}$ is the k-th exposure value, determining the position of the dynamic range window to be mapped to the available luminance range of the display. We assume a fixed display device with the minimum and maximum luminances of $I_{\min} = 1\,\mathrm{cd/m^2}$ and $I_{\max} = 200\,\mathrm{cd/m^2}$, respectively. These are typical specifications of consumer-grade displays of standard dynamic ranges, resulting in the window size $w = \log_2(200/1) = 7.64$ in the logarithmic scale. H denotes the reference HDR image, and $L^{(k)}$ represents the k-th LDR image. b indicates the black-level factor, accounting for the limited contrast of the display due to the light leakage and the ambient light reflections from the display. $[\cdot]_0^1$ denotes the clamping function with the output range $[0, 1]$. $(\cdot)^{1/\gamma}$ represents the gamma correction. We follow the default configurations in [26], and set $b = 1/128$ and $\gamma = 2.2$. Eq. (1) is independently applied to the three color channels.

It is noteworthy that we intentionally avoid employing state-of-the-art tone mapping operators (TMOs) for the HDR-to-LDR conversion. This is because they are essentially dynamic range compressors, leading to the unavoidable loss of information and the emergence of algorithm-dependent artifacts. In contrast, the adopted inverse display model incurs minimal contrast distortions by mapping a portion of the luminance range to that of the LDR display. Moreover, it acts as a local dynamic range magnifier, expanding a specific luminance range for a more detailed examination.

We follow [26] to determine the positions of the sliding windows (*i.e.*, the values of $\{v^{(k)}\}$). Specifically, we select K uniformly spaced overlapping windows such that each eight stops[1] of the luminance range are covered by three windows. This can be done by dividing the eight stops into three equal dynamic ranges and setting the *endpoint* of the k-th window to be

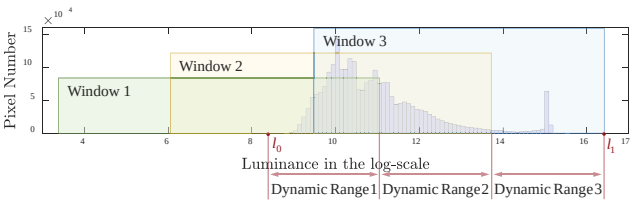
$$l^{(k)} = l_0 + \frac{8}{3}k, \qquad (2)$$

where l_0 represents the minimum log-luminance in the scene. The exposure value $v^{(k)}$ is then computed by

$$v^{(k)} = 2^{-l^{(k)}}. \qquad (3)$$

Fig. 3 (a) shows such an example HDR image with eight stops. Fig. 3 (b)-(d) show the corresponding LDR images, which exhibit different exposures.

[1] When photometric units (*e.g.*, luminances) are plotted on the $\log_2$ axis, each logarithmic unit corresponds to 1 stop.



(a)

(b) (c) (d)

Figure 3. Decomposition of an HDR image with eight stops into three LDR images of different exposures using the inverse display model in Eq. (1). l_0 and l_1 denote the minimum and maximum luminances in the log-scale. **(a)** indicates the positions of the sliding windows by Eq. (2). **(b)-(d)** are the LDR images corresponding to Window 1 to Window 3, respectively.

(a) LDR images

(b) Weighting maps

Figure 4. LDR image stack generated by the inverse display model (in Eq. (1)) and the corresponding local weighting maps (by Eq. (6)) for the "Forest" scene.

3.2. Quality Assessment Model

In the same vein, we may utilize another set of $\{\hat{v}^{(k)}\}_{k=1}^{K}$ to compute an LDR image stack $\{\hat{L}^{(k)}\}_{k=1}^{K}$ from the test HDR image $\hat{H}$. Subsequently, we evaluate the perceptual quality of $\hat{L}^{(k)}$ using $L^{(k)}$ as reference:

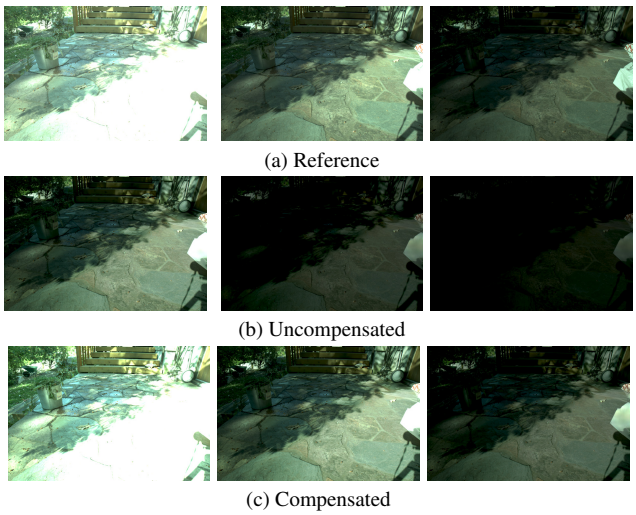
$$Q_i^{(k)} = D_i\left(L^{(k)}, \hat{L}^{(k)}; v^{(k)}, \hat{v}^{(k)}\right), \tag{4}$$

where $D(\cdot,\cdot)$ denotes an LDR quality metric that produces a local quality map, indexed by i. A larger $Q_i^{(k)}$ indicates higher predicted quality at the i-th spatial location and k-th exposure. We pool local quality scores with a local weighting map:

$$Q^{(k)} = \frac{\sum_i W_i^{(k)} Q_i^{(k)}\left(L^{(k)}, \hat{L}^{(k)}; v^{(k)}, \hat{v}^{(k)}\right)}{\sum_i W_i^{(k)}}, \tag{5}$$

where

$$W_i^{(k)} = \begin{cases} 1 & \text{if } 0.1 \le L_i^{(k)} \le 0.9 \\ \epsilon & \text{otherwise,} \end{cases} \tag{6}$$



(a) Reference

(b) Uncompensated

(c) Compensated

Figure 5. Illustration of compensation for the luminance shifts through Eq. (8). **(a)** LDR image stack generated from the reference HDR image. **(b)** LDR image stack generated from the HDR image by MaskHDR [50] *with the same exposure values used in (a)*. **(c)** LDR image stack generated from the HDR image by MaskHDR [50] with optimized exposure values.

is determined by a simple well-exposedness measure to exclude under- and over-exposed regions. ϵ is a small positive constant set to 10^{-5}. In practice, we further normalize the local weightings for the same spatial location across different exposures (*i.e.*, $\sum_k W_i^{(k)} = 1$) to make each spatial location contributes equally in the computation. Fig. 4 shows the local weighting maps corresponding to the LDR image stack of the "Forest" scene. The overall quality score is computed by aggregating global quality estimates across exposures:

$$Q = \sum_{k=1}^{K} G^{(k)} Q^{(k)}\left(L^{(k)}, \hat{L}^{(k)}; v^{(k)}, \hat{v}^{(k)}\right), \tag{7}$$

where $G^{(k)}$ is the k-th global weighting constrained to be non-negative, and $\sum_k G^{(k)} = 1$. It is flexible to put more emphasis on assessing specific luminance ranges by raising the associated $G^{(k)}$ values. Unless otherwise specified, we set $G^{(k)} = 1/K$.

As noticed by Hanji *et al.* [14], the reference and test HDR images may exhibit luminance shifts that will significantly bias quality prediction. To mitigate this issue, we opt to further maximize Q in Eq. (7) with respect to $\{\hat{v}^{(k)}\}_{k=1}^{K}$:

$$Q^{\star} = \max_{\{\hat{v}^{(k)}\}_{k=1}^{K}} Q\left(\{L^{(k)}, \hat{L}^{(k)}; v^{(k)}, \hat{v}^{(k)}\}_{k=1}^{K}\right), \tag{8}$$

which can be decomposed into K one-dimensional optimization problems, and solved efficiently by the golden-section, bisection, or Newton's methods. Fig. 5 provides a

Table 1. Performance comparison in terms of SRCC and PLCC of the proposed HDR quality metrics against 19 existing methods on four HDR-IQA datasets. The left and right numbers separated by "/" indicate the performance on the whole UPIQ dataset and its HDR image subset, respectively. The weightings to compute the average results in the last column are proportion to the numbers of HDR images in respective datasets. The top-2 results are highlighted in bold.

Model	Narwaria2013 [39]		Valenzise2014 [56]		Zerman2017 [62]		UPIQ [34]		Weighted Avg	
	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC
NLPD	0.716	0.747	0.828	0.845	0.752	0.755	0.817/0.814	0.833/0.821	0.785	0.797
HDR-VQM	0.761	0.788	0.865	0.892	0.762	0.774	0.788/0.818	0.817/0.819	0.801	0.812
HDR-VDP-3-Q	0.742	0.770	0.835	0.874	0.700	0.695	0.826/**0.845**	0.871/**0.843**	0.801	0.808
HDR-VDP-3-D	0.723	0.735	0.829	0.854	0.700	0.708	0.801/0.800	0.806/0.784	0.771	0.768
MAE	0.107	0.101	0.225	0.373	0.361	0.355	0.534/0.256	0.570/0.294	0.238	0.269
PU-MAE	0.543	0.620	0.435	0.625	0.553	0.568	0.556/0.620	0.602/0.639	0.579	0.623
PU21-MAE	0.560	0.600	0.470	0.575	0.557	0.555	0.585/0.613	0.625/0.617	0.582	0.601
$Q^{\star}_{\text{MAE}}$	0.624	0.653	0.838	0.868	0.750	0.728	0.602/0.642	0.635/0.646	0.670	0.677
PSNR	0.124	0.139	0.371	0.416	0.465	0.506	0.645/0.299	0.650/0.341	0.293	0.329
PU-PSNR	0.532	0.595	0.529	0.611	0.649	0.677	0.643/0.631	0.651/0.644	0.605	0.636
PU21-PSNR	0.546	0.574	0.588	0.655	0.633	0.662	0.666/0.585	0.665/0.591	0.584	0.603
$Q^{\star}_{\text{PSNR}}$	0.682	0.716	0.766	0.812	0.789	0.774	0.700/0.709	0.701/0.716	0.720	0.733
SSIM	0.126	0.313	0.322	0.502	0.493	0.451	0.677/0.383	0.706/0.475	0.341	0.440
PU-SSIM	0.651	0.690	0.840	0.880	0.754	0.750	0.665/0.736	0.667/0.738	0.729	0.741
PU21-SSIM	0.633	0.679	0.837	0.863	0.757	0.744	0.680/0.674	0.677/0.671	0.691	0.699
$Q^{\star}_{\text{SSIM}}$	0.658	0.664	0.893	0.917	0.814	0.801	0.731/0.750	0.745/0.747	0.752	0.751
LPIPS	0.650	0.695	0.768	0.780	0.684	0.695	0.844/0.829	0.876/0.824	0.765	0.774
PU-LPIPS	0.801	0.823	0.883	0.922	0.779	0.759	0.834/0.832	0.870/0.837	0.822	0.829
PU21-LPIPS	0.815	0.833	0.903	0.921	0.806	0.804	0.779/0.822	0.838/0.828	0.825	0.833
$Q^{\star}_{\text{LPIPS}}$	0.823	0.839	0.905	0.918	0.847	0.837	0.844/0.836	0.880/0.835	**0.840**	**0.843**
DISTS	0.515	0.593	0.794	0.848	0.811	0.846	**0.860**/0.691	**0.882**/0.701	0.680	0.712
PU-DISTS	0.847	0.867	**0.910**	**0.929**	**0.862**	**0.870**	0.805/0.788	0.857/0.804	0.821	0.837
PU21-DISTS	**0.860**	**0.872**	0.907	0.921	0.829	0.831	0.798/0.801	0.854/0.822	0.826	0.842
$Q^{\star}_{\text{DISTS}}$	**0.868**	**0.877**	**0.917**	**0.930**	**0.904**	**0.901**	**0.861/0.853**	**0.881/0.857**	**0.869**	**0.873**

visual comparison of the test LDR image stacks without and with the luminance shift compensation. When adopting Q in Eq. (7) as the loss function for perceptual optimization of HDR image rendering tasks, we can more effectively minimize the luminance shifts in an online fashion by setting $\hat{v}^{(k)} = v^{(k)}$, for $1 \leq k \leq K$.

The proposed HDR quality metric naturally reduces to its base LDR metric ($D(\cdot)$ in Eq. (4)) when assessing LDR images. This is because the LDR images would need to be first transformed from the display-encoded color space (*e.g.*, sRGB) to linear color values through the forward display model:

$$L = (1-b)P^{\gamma} + b \quad \text{and} \quad \hat{L} = (1-b)\hat{P}^{\gamma} + b, \tag{9}$$

where P and $\hat{P}$ are digital pixel values of the reference and test LDR images, respectively. The black-level factor b and gamma parameter γ are the same as in Eq. (1). The maximum luminances of L and $\hat{L}$ are scaled to 200 cd/m^2. The integration of the forward display model with the inverse display model in Eq. (1) results in an identity mapping, thereby leaving the input LDR image intact.

4. Quality Assessment Validation

In this section, we compare our HDR quality metrics with existing model-based and encoding-based methods on four HDR-IQA datasets.

4.1. Experimental Setups

Implementation Details. We adopt five base LDR quality models to implement $D(\cdot,\cdot)$ in Eq. (4): the mean absolute error (MAE), PSNR, SSIM, the learned perceptual image patch similarity (LPIPS) model [63] with VGGNet [53], and the deep image structure and texture similarity (DISTS) metric [11]. To solve the K one-dimensional optimization problems in Eq. (8), we employ the gradient ascent method with an initial learning rate of 10^{-3}, and decay the learning rate by a factor of 5 for every $1,000$ iterations with a maximum of $5,000$ iterations. Early stopping is enabled if the absolute difference of the losses between two consecutive iterations is less than 10^{-3}.

Datasets. Four publicly available HDR-IQA datasets are adopted for benchmarking: Narwaria2013 [39], Valenzise2014 [56], Zerman2017 [62], and UPIQ [34], which contain 140, 50, 100 and $4,159$ images, respectively. The

Table 2. Performance comparison in terms of SRCC and PLCC of the proposed HDR quality metrics without and with the luminance shift compensation.

Method	Weighted average across the four datasets w/o compensation		w/ compensation	
	SRCC	PLCC	SRCC	PLCC
Q_{MAE}	0.525	0.586	0.670	0.677
Q_{PSNR}	0.537	0.546	0.720	0.733
Q_{SSIM}	0.575	0.600	0.752	0.751
Q_{LPIPS}	0.694	0.702	0.840	0.843
Q_{DISTS}	0.708	0.711	0.869	0.873

UPIQ dataset stands out for its collection of 380 HDR and 3,779 LDR images from four sub-datasets [19, 39, 46, 51], whose scores have been carefully re-aligned to a common perceptual scale to ensure consistency.

Competing Metrics. We select nine model-based methods for comparison, including 1) NLPD [20], 2) HDR-VQM [42], 3) the quality score of HDR-VDP-3 [31] (denoted by HDR-VDP-3-Q) and 4) the difference score of HDR-VDP-3[2] (denoted by HDR-VDP-3-D) as four HDR quality metrics, and 5) MAE, 6) PSNR, 7) SSIM [59], 8) LPIPS [63] and 9) DISTS [11] as five LDR quality metrics. We also equip the five LDR quality models with the PU [1] and PU21 encoding, giving rise to 10) PU-MAE, 11) PU-PSNR, 12) PU-SSIM, 13) PU-LPIPS, 14) PU-DISTS, 15) PU21-MAE, 16) PU21-PSNR, 17) PU21-SSIM, 18) PU21-LPIPS, and 19) PU21-DISTS. As suggested in [1, 25, 29], we assume a test HDR display model with a maximum luminance of $1,000$ and $4,000$ cd/m^2 for the PU and PU21 encoding, respectively, which are independently applied to the three color channels. We find empirically that the performance ranking is fairly robust to the selection of the maximum luminance of the display. For the LDR images in UPIQ, we first convert digital pixel values to luminance values via the display model in Eq. (9) before applying HDR quality metrics, and adjust the hyperparameters of the base LDR quality metrics if necessary[3].

Evaluation Criteria. We use two evaluation criteria: Spearman's rank correlation coefficient (SRCC) and Pearson linear correlation coefficient (PLCC). As a standard practice [51, 58], we fit a four-parameter logistic function before computing PLCC.

4.2. Results

Table 1 presents the performance comparison results, where we find that the adopted inverse display model leads to consistent improvements for all base LDR quality models. In particular, the instantiation $Q^{\star}_{\text{DISTS}}$ achieves the best results on all four datasets, even surpassing the re-calibrated HDR-VDP-3-Q on UPIQ. Consistent with previous studies [1, 25], the PU and PU21 encoding can boost the performance of base LDR quality measures, but not as substantial as our metrics. When applied to the LDR images in UPIQ, the PU and PU21 encoding incur noticeable performance degradation. In stark contrast, our metrics maintain reliable LDR-IQA capabilities. Last, there is a clear trend that a better base LDR quality metric generally delivers better performance, affirming our objective of transferring the advancements in LDR-IQA to HDR-IQA.

Table 2 shows the ablation results of the proposed HDR quality metrics without and with the luminance shift compensation (see Eq. (8)). It is evident that our compensated metrics consistently outperform the non-compensated counterparts. This performance gap is expected to be even more pronounced in the presence of large luminance shifts, such as assessing HDR images derived from single image HDR reconstruction methods [12, 14]. Thus, compensating for luminance shifts is recommended as a standard procedure when comparing HDR images.

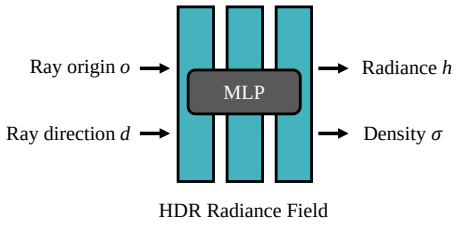


Figure 6. The inputs and outputs of the MLP implicitly model the HDR radiance field. Image adapted from [17].

5. Perceptual Optimization Validation

In this section, we explore the application of the proposed HDR quality metrics for perceptual optimization of HDR novel view synthesis.

5.1. HDR Novel View Synthesis

We select HDR-NeRF [17] as the starting point. The original HDR-NeRF employs a multilayer perceptron (MLP) to implicitly represent the radiance field of an HDR scene, and uses a separate MLP to function as a tone mapper to reconstruct multiple input LDR images of different exposures. Here, we simplify HDR-NeRF by stripping off the tone mapper, and directly reconstruct the HDR scene, guided by the proposed HDR quality metrics (see Fig. 6). We refer to the simplified method as HDR-NeRF†.

Network Design. We employ an eight-layer MLP with 256 channels to implicitly reconstruct the HDR scene radiance. For a given ray $r = o + sd$, where o is the origin, d is the ray

[2]The current version under evaluation is HDR-VDP-3.0.7 with the default parameter setting.

[3]For example, in PU21-SSIM, the two normalizing constants are adjusted to $C_1 = (0.01 \times 4000)^2 = 1,600$ and $C_2 = (0.03 \times 4000)^2 = 14,440$, respectively, as the maximum luminance is $4,000$ cd/m^2.

Table 3. Quantitative comparison of HDR novel view synthesis methods averaged across eight synthetic scenes.

Method	HDR-VDP-3-Q	PSNR	SSIM	Without CRF correction		With CRF correction		$Q^\star$ (Our metric)	
				PU21-PSNR	PU21-SSIM	PU21-PSNR	PU21-SSIM	$Q^\star_{\text{PSNR}}$	$Q^\star_{\text{SSIM}}$
HDR-NeRF	7.023	25.513	0.863	31.388	0.901	38.485	0.929	28.957	0.899
HDR-NeRF+	9.634	27.358	0.929	35.758	0.953	41.350	0.957	32.457	0.937
HDR-NeRF†	**9.863**	**28.754**	**0.933**	**38.202**	**0.967**	**43.483**	**0.973**	**34.539**	**0.968**

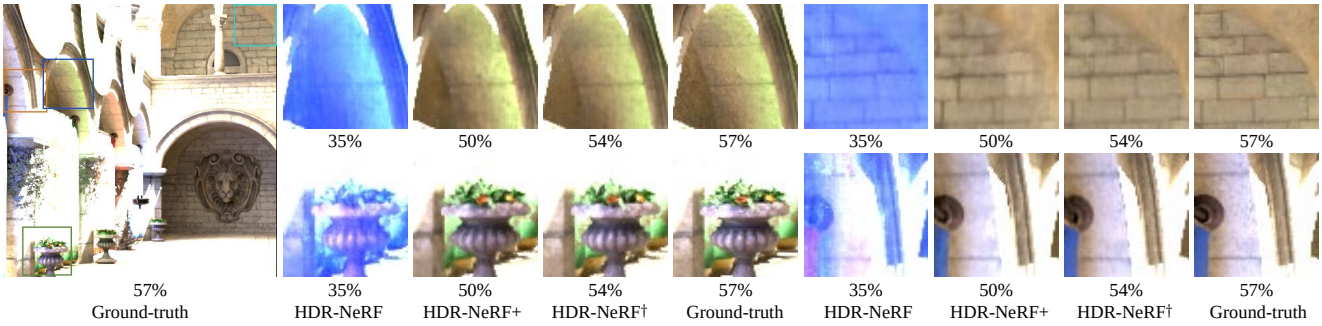


Figure 7. Visual comparison of HDR novel view synthesis methods on the "Sponza" scene. For the reference HDR view as the ground-truth, we set the exposure value v in Eq. (1) to be the 57-th of the full dynamic range. Other percentages are the optimally matched $\hat{v}$ for different synthesis methods by Eq. (8).

direction, and s denotes a position along the ray, the MLP outputs the radiance h and density σ, based on which the luminance value can be computed by

$$\hat{H}(r) = \int_{s_n}^{s_f} T(s)\sigma(r(s))h(r(s))ds, \tag{10}$$

where

$$T(s) = \exp\left(-\int_{s_n}^{s} \sigma(r(v))dv\right). \tag{11}$$

s_n and s_f denote the near and far boundary of the ray, respectively. $T(s)$ denotes the accumulated transmittance along the ray from s_n to s.

Loss Function. For computational convenience, we adopt the proposed HDR quality metric, $Q^\star_{\text{MAE}}$ (rather than $Q^\star_{\text{DISTS}}$), as the loss function to encourage high-fidelity novel view synthesis across all luminance levels.

5.2. Experimental Setups

Model Training and Testing. We employ the dataset in [17], comprising 8 synthetic scenes rendered by Blender[4]. There are 35 HDR views for each scene, and we select 18 views for training, and leave the remaining 17 for testing. The resolution of each view is 400×400.

Training follows the original paper [17]. We employ the positional encoding in [35], and optimize a coarse model and a fine model, where the density predicted by the coarse model is used to bias the sampling of a ray in the fine model. We sample 64 points along each ray in the coarse model and 128 points in the fine model. We employ the Adam optimizer [18] with an initial learning rate 5×10^{-4}, which decays exponentially to 5×10^{-5} with a total of $200,000$ iterations. The batch size of rays is set to $1,024$.

[4]https://www.blender.org/

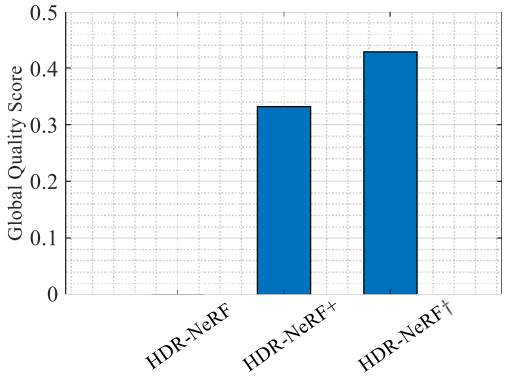


Figure 8. Subjective quality scores across all test views and observers in the 2AFC subjective user study. HDR-NeRF serves as the baseline with the global quality score of zero.

Competing Methods. We compare our method against HDR-NeRF [17] and its variant optimized for HDR views directly, denoted by HDR-NeRF+. When training HDR-NeRF+, the reference and predicted HDR luminance values are tone mapped to LDR values by a simple TMO [47], as suggested in [17].

Evaluation Criteria. We employ several objective quality metrics: 1) HDR-VDP-3-Q [31] 2) PSNR, 3) SSIM [59], 4) PU21-PSNR, 5) PU21-SSIM, 6) PU21-PSNR with camera response function (CRF) correction [14], 7) PU21-SSIM

Table 4. Quantitative comparison of HDR-NeRF† optimized by different loss functions.

Loss	HDR-VDP-3-Q	PSNR	SSIM	Without CRF correction		With CRF correction		$Q^\star$ (Our metric)	
				PU21-PSNR	PU21-SSIM	PU21-PSNR	PU21-SSIM	$Q^\star_{\text{PSNR}}$	$Q^\star_{\text{SSIM}}$
MAE	6.224	17.181	0.454	25.842	0.613	31.577	0.856	22.928	0.849
PU21-MAE	9.717	28.670	0.929	37.293	0.960	42.245	0.970	33.477	0.949
log-MAE	9.720	23.173	0.840	34.006	0.936	41.267	0.961	32.423	0.944
μ-MAE	9.784	25.645	0.894	35.408	0.948	42.104	0.968	33.227	0.948
$Q^\star_{\text{MAE}}$	**9.863**	**28.754**	**0.933**	**38.202**	**0.967**	**43.483**	**0.973**	**34.539**	**0.968**

with CRF correction, 8) the proposed quality metric with PSNR as the base model (*i.e.*, $Q^\star_{\text{PSNR}}$), and 9) the proposed quality metric with SSIM as the base model (*i.e.*, $Q^\star_{\text{SSIM}}$). The CRF correction compensates for the metric sensitivity to the shifts in tone and color [12], and is applied before the PU21 encoding.

5.3. Experimental Results

Quantitative Evaluation. Table 3 lists the average results of rendered novel HDR views of the eight synthetic scenes. The primary observation is that the proposed HDR-NeRF† outperforms HDR-NeRF+ by a clear margin under all evaluation metrics. This demonstrates the superiority of the adopted inverse display model over the simple tone mapper in HDR-NeRF+. Lack of direct supervision, HDR-NeRF performs marginally in synthesizing HDR views, despite its ability to reconstruct a satisfying output LDR image.

Qualitative Evaluation. Fig. 7 visually compares the results on a test view of the "Sponza" scene. The HDR view synthesized by HDR-NeRF suffers from color cast and detail loss. HDR-NeRF+ recovers more details, but not as sharp as those rendered by the proposed HDR-NeRF†.

Subjective User Study. We perform a subjective user study to verify the perceptual advantages of HDR-NeRF†. For each of the eight scenes, we randomly choose five test views, reconstructed by the three competing methods (including HDR-NeRF†). We manually select $\{v^{(k)}\}_{k=1}^{3}$ to zoom in the low, middle, and high luminance range of each reference HDR view, respectively. All LDR images are aligned to the reference LDR images by solving Eq. (8). We adopt the two-alternative forced choice (2AFC) approach to gather human preferences of $\binom{3}{2} \times 8 \times 5 \times 3 = 360$ image pairs from 15 participants. They are given unlimited time to review the images, and are allowed to take a break at any time during subjective testing to mitigate fatigue effects. The global quality scores are aggregated by the maximum likelihood estimation [55]. Fig. 8 shows the results, which verify the perceptual gains of HDR-NeRF† driven by the proposed HDR quality metric.

Ablation study. We evaluate the view synthesis performance of HDR-NeRF† optimized by several different quality metrics as the loss functions: 1) MAE, 2) PU21-MAE, 3) log-encoded MAE (*i.e.*, log-MAE), 4) MAE computed in the LDR domain tone mapped by the μ-law [54] (*i.e.*, μ-MAE), and 5) the proposed $Q^\star_{\text{MAE}}$. Table 4 presents the quantitative comparison results, where HDR-NeRF† optimized by $Q^\star_{\text{MAE}}$ delivers the best results. The encoding-based metrics like PU21-MAE and log-MAE do not necessarily surpass μ-MAE, even though tone mapping would cause detail loss and color distortion.

6. Conclusion and Discussion

We have described a family of HDR quality metrics by augmenting current LDR quality metrics with a simple inverse display model. Our metrics are efficient in inheriting the benefits of advanced LDR quality metrics, flexible to zoom in and align specific luminance ranges for more detailed assessment, and training-free. We have validated our HDR quality metrics in terms of correlation with human perceptual scores on four HDR-IQA datasets and perceptual optimization of HDR novel view synthesis.

Previous studies of HDR image processing are inclined to adopt a global tone mapper for visualizing and comparing the processed results. In contrast, this paper suggests an alternative visualization method of using the inverse display model in Eq. (1). This method allows us to focus on and enhance the visibility of different portions of luminance ranges for a more fine-grained visual comparison (see Fig. 7). Together with the proposed family of HDR quality metrics, we expect more rapid and reliable progress of HDR imaging and rendering in the near future.

As one of the limitations, our metrics do not account for the reduced sensitivity of the visual system at low luminances. That is, the predictions are the same regardless of whether the image is meant to be shown on a dark or bright display. The PU/PU21 encoding and HDR-VDP are designed to model the changes in sensitivity with absolute luminance levels.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 62071407 and the Hong Kong RGC Early Career Scheme (2121382).

References

[1] Tunç O. Aydïn, Rafał K. Mantiuk, and Hans-Peter Seidel. Extending quality metrics to full luminance range images. In *Human Vision and Electronic Imaging XIII*, pages 109–118, 2008. 1, 2, 6

[2] Jonathan T. Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P Srinivasan. Mip-NeRF: A multiscale representation for anti-aliasing neural radiance fields. In *IEEE International Conference on Computer Vision*, pages 5855–5864, 2021. 2

[3] Jonathan T. Barron, Ben Mildenhall, Dor Verbin, Pratul P. Srinivasan, and Peter Hedman. Mip-NeRF 360: Unbounded anti-aliased neural radiance fields. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 5460–5469, 2022. 2

[4] Peter G. J. Barten. Formula for the contrast sensitivity of the human eye. In *Image Quality and System Performance*, pages 231–238, 2003. 2

[5] Mark Boss, Raphael Braun, Varun Jampani, Jonathan T. Barron, Ce Liu, and Hendrik Lensch. NeRD: Neural reflectance decomposition from image collections. In *IEEE International Conference on Computer Vision*, pages 12684–12694, 2021. 2

[6] Chris Buehler, Michael Bosse, Leonard McMillan, Steven Gortler, and Michael Cohen. Unstructured lumigraph rendering. In *Annual Conference on Computer Graphics and Interactive Techniques*, pages 425–432, 2001. 2

[7] Matteo Carandini and David J. Heeger. Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, 13(1):51–62, 2012. 2

[8] Anpei Chen, Zexiang Xu, Fuqiang Zhao, Xiaoshuai Zhang, Fanbo Xiang, Jingyi Yu, and Hao Su. MVSNeRF: Fast generalizable radiance field reconstruction from multi-view stereo. In *IEEE International Conference on Computer Vision*, pages 14124–14133, 2021. 2

[9] Xingyu Chen, Qi Zhang, Xiaoyu Li, Yue Chen, Ying Feng, Xuan Wang, and Jue Wang. Hallucinated neural radiance fields in the wild. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 12943–12952, 2022. 2

[10] Scott J. Daly. Visible differences predictor: An algorithm for the assessment of image fidelity. In *Human Vision, Visual Processing, and Digital Display III*, pages 2–15, 1992. 1, 2

[11] Keyan Ding, Kede Ma, Shiqi Wang, and Eero P. Simoncelli. Image quality assessment: Unifying structure and texture similarity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(5):2567–2581, 2022. 5, 6

[12] Gabriel Eilertsen, Saghi Hajisharif, Param Hanji, Apostolia Tsirikoglou, Rafał K. Mantiuk, and Jonas Unger. How to cheat with metrics in single-image HDR reconstruction. In *IEEE International Conference on Computer Vision Workshops*, pages 3998–4007, 2021. 1, 2, 3, 6, 8

[13] Steven J. Gortler, Radek Grzeszczuk, Richard Szeliski, and Michael F. Cohen. The lumigraph. In *Annual Conference on Computer Graphics and Interactive Techniques*, pages 43–54, 1996. 2

[14] Param Hanji, Rafał K. Mantiuk, Gabriel Eilertsen, Saghi Hajisharif, and Jonas Unger. Comparison of single image HDR reconstruction methods — the caveats of quality assessment. In *Annual Conference on Computer Graphics and Interactive Techniques*, pages 1:1–1:8, 2022. 1, 2, 3, 4, 6, 7

[15] Peter Hedman, Julien Philip, True Price, Jan-Michael Frahm, George Drettakis, and Gabriel Brostow. Deep blending for free-viewpoint image-based rendering. *ACM Transactions on Graphics*, 37(6):257:1–257:15, 2018. 2

[16] Bernd Hoefflinger. *High-Dynamic-Range (HDR) Vision*. Springer, 2007. 1

[17] Xin Huang, Qi Zhang, Ying Feng, Hongdong Li, Xuan Wang, and Qing Wang. HDR-NeRF: High dynamic range neural radiance fields. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 18398–18408, 2022. 2, 3, 6, 7

[18] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 7

[19] Pavel Korshunov, Philippe Hanhart, Thomas Richter, Alessandro Artusi, Rafał K. Mantiuk, and Touradj Ebrahimi. Subjective quality assessment database of HDR images compressed with JPEG XT. In *International Workshop on Quality of Multimedia Experience*, pages 1–6, 2015. 6

[20] Valero Laparra, Alex Berardino, Johannes Ballé, and Eero P. Simoncelli. Perceptually optimized image rendering. *Journal of the Optical Society of America A*, 34(9):1511–1525, 2017. 2, 6

[21] Marc Levoy and Pat Hanrahan. Light field rendering. In *Annual Conference on Computer Graphics and Interactive Techniques*, pages 31–42, 1996. 2

[22] Zhengqi Li, Simon Niklaus, Noah Snavely, and Oliver Wang. Neural scene flow fields for space-time view synthesis of dynamic scenes. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 6498–6508, 2021. 2

[23] Steven Liu, Xiuming Zhang, Zhoutong Zhang, Richard Zhang, Jun-Yan Zhu, and Bryan Russell. Editing conditional radiance fields. In *IEEE International Conference on Computer Vision*, pages 5773–5783, 2021. 2

[24] Li Ma, Xiaoyu Li, Jing Liao, Qi Zhang, Xuan Wang, Jue Wang, and Pedro V. Sander. Deblur-NeRF: Neural radiance fields from blurry images. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 12861–12870, 2022. 2

[25] Rafał K. Mantiuk and Maryam Azimi. PU21: A novel perceptually uniform encoding for adapting existing quality metrics for HDR. In *Picture Coding Symposium*, pages 1–5, 2021. 1, 2, 6

[26] Rafał K. Mantiuk and Wolfgang Heidrich. Visualizing high dynamic range images in a web browser. *Journal of Graphics, GPU, and Game Tools*, 14(1):43–53, 2009. 1, 3

[27] Rafał K. Mantiuk and Giovanni (Gianni) Ramponi. Age-dependent prediction of visible differences in displayed images: Age-dependent prediction of visible differences. *Journal of the Society for Information Display*, 26:1–21, 2018. 2

[28] Rafał K. Mantiuk, Scott J. Daly, Karol Myszkowski, and Hans-Peter Seidel. Predicting visible differences in high dynamic range images: Model and its calibration. In *Human Vision and Electronic Imaging X*, pages 204–214, 2005. 1, 2

[29] Rafał K. Mantiuk, Kil Joong Kim, Allan G. Rempel, and Wolfgang Heidrich. HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. *ACM Transactions on Graphics*, 30(4):40:1–40:14, 2011. 1, 2, 6

[30] Rafał K. Mantiuk, Minjung Kim, Maliha Ashraf, Qiang Xu, Ming R. Luo, Jasna Martinovic, and Sophie Wuerger. Practical color contrast sensitivity functions for luminance levels up to 10000 cd/m^2. In *Color and Imaging Conference*, pages 28:1–28:6, 2020. 2

[31] Rafał K. Mantiuk, Dounia Hammou, and Param Hanji. 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. 1, 2, 6, 7

[32] Ricardo Martin-Brualla, Noha Radwan, Mehdi S. M. Sajjadi, Jonathan T. Barron, Alexey Dosovitskiy, and Daniel Duckworth. NeRF in the wild: Neural radiance fields for unconstrained photo collections. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 7210–7219, 2021. 2

[33] Zhenxing Mi and Dan Xu. Switch-NeRF: Learning scene decomposition with mixture of experts for large-scale neural radiance fields. In *International Conference on Learning Representations*, 2022. 2

[34] Aliaksei Mikhailiuk, María Pérez-Ortiz, Dingcheng Yue, Wilson Suen, and Rafał K. Mantiuk. Consolidated dataset and metrics for high-dynamic-range image quality. *IEEE Transactions on Multimedia*, 24(67):2125–2138, 2021. 2, 5

[35] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021. 2, 7

[36] Ben Mildenhall, Peter Hedman, Ricardo Martin-Brualla, Pratul P. Srinivasan, and Jonathan T. Barron. NeRF in the dark: High dynamic range view synthesis from noisy raw images. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 16190–16199, 2022. 2

[37] Scott Miller, Mahdi Nezamabadi, and Scott J. Daly. Perceptual signal coding for more efficient usage of bit codes. *SMPTE Motion Imaging Journal*, 122(4):52–59, 2013. 1, 2

[38] Jacob Munkberg, Petrik Clarberg, Jon Hasselgren, and Tomas Akenine-Möller. High dynamic range texture compression for graphics hardware. *ACM Transactions on Graphics*, 25(3):698–706, 2006. 1, 2

[39] Manish Narwaria, Matthieu Perreira Da Silva, Patrick Le Callet, and Romuald Pepion. Tone mapping-based high-dynamic-range image compression: Study of optimization criterion and perceptual quality. *Optical Engineering*, 52 (10):102008:1–102008:15, 2013. 2, 5, 6

[40] Manish Narwaria, Matthieu Perreira Da Silva, Patrick Le Callet, and Romuald Pepion. Impact of tone mapping in high dynamic range image compression. In *International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, pages 1–7, 2014. 2

[41] Manish Narwaria, Rafał K. Mantiuk, Mattheiu Perreira Da Silva, and Patrick Le Callet. HDR-VDP-2.2: A calibrated method for objective quality prediction of high-dynamic range and standard images. *Journal of Electronic Imaging*, 24(1):010501:1–010501:10, 2015. 1, 2

[42] 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(4):46–60, 2015. 2, 6

[43] Michael Niemeyer and Andreas Geiger. GIRAFFE: Representing scenes as compositional generative neural feature fields. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 11453–11464, 2021. 2

[44] Keunhong Park, Utkarsh Sinha, Jonathan T. Barron, Sofien Bouaziz, Dan B. Goldman, Steven M. Seitz, and Ricardo Martin-Brualla. Nerfies: Deformable neural radiance fields. In *IEEE International Conference on Computer Vision*, pages 5865–5874, 2021. 2

[45] Nikolay Ponomarenko, Federica Battisti, Karen Egiazarian, Jaakko Astola, and Vladimir Lukin. Metrics performance comparison for color image database. In *International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, pages 1–6, 2009. 2

[46] Nikolay Ponomarenko, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, Lina Jin, Jaakko Astola, Benoit Vozel, Kacem Chehdi, Marco Carli, Federica Battisti, and C.-C. J. Kuo. Image database TID2013: Peculiarities, results and perspectives. *Signal Processing: Image Communication*, 30 (3):57–77, 2015. 6

[47] Erik Reinhard, Michael Stark, Peter Shirley, and James Ferwerda. Photographic tone reproduction for digital images. *ACM Transactions on Graphics*, 21(3):267–276, 2002. 7

[48] Gernot Riegler and Vladlen Koltun. Free view synthesis. In *European Conference on Computer Vision*, pages 623–640, 2020. 2

[49] Gernot Riegler and Vladlen Koltun. Stable view synthesis. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 12211–12220, 2021. 2

[50] Marcel S. Santos, Tsang I. Ren, and Nima K. Kalantari. Single image HDR reconstruction using a CNN with masked features and perceptual loss. *ACM Transactions on Graphics*, 39(4):80:1–80:10, 2020. 4

[51] Hamid R. Sheikh, Muhammad F. Sabir, and Alan C. Bovik. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on Image Processing*, 15(11):3440–3451, 2006. 2, 6

[52] Harry Shum and Sing Bing Kang. Review of image-based rendering techniques. In *Visual Communications and Image Processing*, pages 2–13, 2000. 2

[53] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations*, 2015. 5

[54] Bernard Smith. Instantaneous companding of quantized signals. *The Bell System Technical Journal*, 36(3):653–710, 1957. 8

[55] Kristi Tsukida and Maya R. Gupta. How to analyze paired comparison data. Technical report, Department of Electrical Engineering University of Washington, 2011. 8

[56] Giuseppe Valenzise, Francesca De Simone, Paul Lauga, and Frederic Dufaux. Performance evaluation of objective quality metrics for HDR image compression. In *Applications of Digital Image Processing XXXVII*, pages 78–87, 2014. 5

[57] Peter Vangorp, Karol Myszkowski, Erich W. Graf, and Rafał K. Mantiuk. A model of local adaptation. *ACM Transactions on Graphics*, 34(6):166:1–166:13, 2015. 2

[58] VQEG. Final report from the video quality experts group on the validation of objective models of video quality assessment, 2000. 6

[59] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. 1, 6, 7

[60] Ruifeng Xu, Sumanta N. Pattanaik, and Charles E. Hughes. High-dynamic-range still-image encoding in JPEG 2000. *IEEE Computer Graphics and Applications*, 25(6):57–64, 2005. 1, 2

[61] Alex Yu, Vickie Ye, Matthew Tancik, and Angjoo Kanazawa. PixelNeRF: Neural radiance fields from one or few images. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4578–4587, 2021. 2

[62] Emin Zerman, Giuseppe Valenzise, and Frederic Dufaux. An extensive performance evaluation of full-reference HDR image quality metrics. *Quality and User Experience*, 2(5):1–16, 2017. 5

[63] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 586–595, 2018. 5, 6