Perceptual Optimization of a Biologically-Inspired Tone Mapping Operator

Peibei Cao, Chenyang Le, Yuming Fang, Senior Member, IEEE, and Kede Ma, Member, IEEE

Abstract—With the increasing popularity and accessibility of high dynamic range (HDR) photography, tone mapping operators (TMOs) for dynamic range compression and medium presentation are practically demanding. In this paper, we develop a two-stage neural network-based HDR image TMO that is biologically-inspired, computationally efficient, and perceptually optimized. In Stage one, motivated by the physiology of the early stages of the human visual system (HVS), we first decompose an HDR image into a normalized Laplacian pyramid. We then use two lightweight deep neural networks (DNNs) that take this normalized representation as input and estimate the Laplacian pyramid of the corresponding LDR image. We optimize the tone mapping network by minimizing the normalized Laplacian pyramid distance (NLPD), a perceptual metric calibrated against human judgments of tone-mapped image quality. In Stage two, we generate a pseudo-multi-exposure image stack with different color saturation and detail visibility by inputting an HDR image “calibrated” with different maximum luminances to the learned tone mapping network. We then train another lightweight DNN to fuse the LDR image stack into a desired LDR image by maximizing a variant of MEF-SSIM, another perceptually calibrated metric for image fusion. By doing so, the proposed TMO is fully automatic to tone map uncalibrated HDR images. Across an independent set of HDR images, we find that our method produces images with consistently better visual quality, and is among the fastest local TMOs.

Index Terms—High dynamic range imaging, tone mapping, image fusion, Laplacian pyramid, perceptual optimization.

1 INTRODUCTION

With the steady improvements in photography technologies, current image sensors (often powered by computational imaging methods [1]) are able to capture pictures with a high dynamic range up to eight orders of magnitude, closely approximating the sensitivity of human vision in the photopic regime [2]. However, existing monitors, projectors, and print-outs, including the best ones, are limited to a lower dynamic range than that can be captured by current sensors [3], thus inadequate to reproduce the full spectrum of luminance values presented in natural scenes. When rendering high dynamic range (HDR) images on low dynamic range (LDR) display devices, tone mapping operators (TMOs) are a prerequisite for dynamic range compression, preserving visual features that are important to describe the original scenes and perceptually noticeable to the human eye. An example is given in Fig. 1.

The naive way of HDR image tone mapping is to linearly rescale the luminances of the HDR image to the range that the display can reproduce. However, images produced this way are often under- or over-exposed, due to the existence of local regions with high luminances (see Fig. 1(a)). In the past decades, extensive efforts have been dedicated to developing TMOs for non-linear dynamic range compression with faithful tone reproduction and detail preservation. These can be broadly categorized into global and local methods. Global TMOs perform the same computation to all pixels (i.e., translation-invariant), which are more computationally efficient at the cost of contrast decrease and detail loss [4], [8]–[12]. Local TMOs [5], [13]–[17], on the other hand, aim to preserve and enhance local contrast often within a two-layer decomposition framework [13]. Although these methods can produce images with better visual quality, it remains difficult to balance global and local contrast, and prevent edge-related artifacts. Most TMOs rely on pre-defined computational graphs with few justifications of perceptual optimality of such computation. In addition, manual hyper-parameter adjustment (e.g., setting maximum luminances for uncalibrated HDR images) is often needed to produce reasonable results, which are, however, no better than conventional photographs on challenging HDR scenes.

Recently, deep neural networks (DNNs) began to show their potential in HDR image tone mapping [18]–[21]. However, unlike traditional image processing tasks such as Gaussian image denoising and image compression, there is no easy-to-obtain ground-truth images available for supervised training in the LDR domain. One popular strategy is to choose the best tone-mapped image from a set of candidate ones produced by existing TMOs with the help of objective quality metrics [20], [22] or subjective experiments [23]. Although with the effort of creating a “super-method”, the resulting TMO may be biased by the common failures of base TMOs. Another approach is to ask photography experts to manually compress the dynamic range of HDR images [7], [18], [21], which is prohibitively slow and suffers from subjective biases. To alleviate this, semi-supervised [7] and adversarial learning [6] techniques have been explored for tone mapping, which turn out to be less accurate and robust (see Fig. 1).

Tumblin and Rushmeier [24] pioneered perceptual optimization of HDR image tone mapping in a cross-dynamic-range setting. They first advocated selecting a TMO capable of producing tone-mapped images that perceptually match the appearance of
In this paper, we describe a two-stage DNN-based TMO for rendering HDR images with the following three desirable design principles.

- It is biologically-inspired. We explicitly model how the early stages of the human visual system (HVS) respond to different light levels by decomposing the input HDR image into a normalized Laplacian pyramid, a multiscale non-linear representation derived from Laplacian pyramid. This allows us to artificially manipulate the maximum luminance of the original scene, giving rise to different detail visibility. As will be clear shortly, we will take advantage of this property to generate a pseudo-multi-exposure image stack for final LDR image fusion.

- It is computationally efficient. Instead of iteratively optimizing over the space of all feasible tone-mapped images, we train two feed-forward DNNs (collectively referred to as the tone mapping network) in Stage one: one accepts all bandpass channels and the highpass channel, while the other processes the lowpass channel of the normalized Laplacian pyramid of an (arbitrarily calibrated) HDR image. Together, they predict the Laplacian pyramid of the corresponding LDR image. By varying the maximum luminance of the original scene, we generate a pseudo-multi-exposure image stack, consisting of LDR images of the same content but with different detail visibility using the learned tone mapping network.

- It is perceptually optimized in a cross-dynamic-range setting. Unlike most TMOs, we train the tone mapping network by optimizing a perceptual metric - the normalized Laplacian pyramid distance (NLPD) between the input HDR scenes and the estimated LDR images in Stage one. We then train the fusion network sequentially in Stage two by maximizing a variant of another perceptual metric - MEF-SSIM for multi-exposure image fusion (MEF). Both NLPD and MEF-SSIM have been subject-calibrated on databases of human perceptual scores and proven effective in optimizing image rendering algorithms and runs in real-time on GPUs.

Extensive experiments demonstrate that the proposed method performs consistently better than state-of-the-art TMOs both qualitatively (via formal subjective experiments) and quantitatively (in terms of objective metrics, TMQI and NLPD). Meanwhile, it is fully automatic to work with uncalibrated HDR images with unknown maximum luminances. Moreover, it is among the fastest local TMOs, and runs in real-time on GPUs.

In this section, we provide a brief review of existing TMOs, with emphasis on DNN-based ones. As our method involves fusing a pseudo-multi-exposure image stack, we will also introduce MEF, an alternative approach to HDR rendering.

### 2 Related Work

In this section, we provide a brief review of existing TMOs, with emphasis on DNN-based ones. As our method involves fusing a pseudo-multi-exposure image stack, we will also introduce MEF, an alternative approach to HDR rendering.

#### 2.1 Existing TMOs

##### 2.1.1 Conventional TMOs

TMOs can be classified into several categories under different sets of criteria, among which we adopt the taxonomy of global and local operators. Global TMOs are dependent on a family of parametric functions, which are specified by some global image statistics and applied to all pixels in an HDR image. These include histogram equalization, homography, gamma mapping, logarithmic function, and sigmoid non-linearity. Glocal methods remain the fastest TMOs as each pixel in S undergoes the same simple transformation. They preserve global contrast well but may lose some local details. Local TMOs are a set of sophisticated methods, which preserve relative contrast between neighboring pixels (e.g., in the form of local gradients) that the human eye is more sensitive.
A common design principle is the layer decomposition originated from the retinex theory [33]. Among many variants [34], [35], the two-layer decomposition by Durand and Dorsey [13] was the most widely accepted, in which tone compression is applied to the base layer, while detail reproduction or enhancement is applied to the detail layer. Many subsequent methods [5], [14], [17], [32] have been proposed based on this design principle differing mainly in how the two-layer image decomposition is performed in a more effective and perceptually plausible way. On many HDR scenes, local TMOs lead to excellent improvements in local contrast preservation. However, this often comes at the cost of increased computational complexity and manual hyper-parameter tuning [13]. Besides, global contrast may be compromised, and local artifacts such as halo-like glows may appear, resulting in unnatural and unrealistic tone-mapped images.

The design of the above-mentioned TMOs is mostly based on empirical rules, with little validity of perceptual optimality of such rules. Perceptual optimization of tone mapping in a cross-dynamic-range setting has been investigated by Tomblin and Rushmeier [24] and later by the authors in [36], [37], who employed simple parametric functions with limited expressiveness. More generally, HDR image tone mapping can be formulated as a constrained optimization problem [25], [26]:

$$I^* = \arg \min_I \ell(S, I), \quad \text{s.t.} \quad I \in \mathcal{C},$$

where $S$ denotes a calibrated HDR image, and $\mathcal{C}$ is the set of feasible tone-mapped images given physical constraints. $\ell(\cdot, \cdot)$ denotes an objective metric that is capable of measuring the perceptual distance between two images of different dynamic ranges, $I^*$ is the optimal tone-mapped image under the criterion $\ell(S, \cdot)$. Note that traditional objective quality metrics such as the mean squared error (MSE) and the structural similarity (SSIM) index [38] are not suitable here because they assume that the two images being compared have the same dynamic range (see Sec. 4.1.4 and Fig. 11). Common choices for $\ell(\cdot, \cdot)$ include TMQI [22] and NLPD [26]. Due to the non-convexity of TMQI and NLPD and the high-dimensionality of the constrained optimization problem, gradient-based iterative solvers were originally proposed, which are computationally prohibitive.

2.1.2 DNN-based TMOs

The primary effort of many DNN-based TMOs [20], [23] is to create a number of ground-truth LDR images for paired training in the LDR domain. Montulet and Briassouli [18] applied a list of existing TMOs to each HDR image, and selected the best tone-mapped one in terms of TMQI as the ground-truth. Many subsequent works have followed this path [20], [21], except for Yang et al. [23] who resorted to formal subjective experiments for best image selection. Panetta et al. [19] trained the tone mapping network over a combination of low-light datasets, which contain the ground-truth normal-light images. Despite the effort, the created ground-truths may be biased by the adopted objective metrics or human annotators. For instance, although TMQI performs well in quality assessment of tone-mapped images, it has its own “blind spots” especially when used as a perceptual optimization metric (see Sec. 4.1.4 and Fig. 11). As a result, different combinations of loss functions have been proposed to encourage the creation of better-quality images in a rather empirical way. Candidate losses for combination include mean absolute error (MAE), MSE, gradient profile loss [19], and VGG content loss [18], [20], [23]. Zhang et al. [7] proposed a semi-supervised learning scheme, where they employed the adversarial loss and the cycle-consistency loss to match the distribution of high-quality LDR images. Vinker et al. [6] achieved tone mapping with a deep generative adversarial network, where the structural similarity is enforced by patch-wise Pearson correlation. Instead of working in the LDR domain with difficult-to-obtain ground-truths, we perform perceptual optimization of HDR image tone mapping in a cross-dynamic-range setting, where we treat the available HDR image containing richer information of the captured natural scene as the ground-truth. This is made possible by cross-dynamic-range quality metrics such as NLPD [26].

2.2 MEF Methods

MEF refers to a class of techniques that fuse a sequence of LDR images with different exposures into a single high-quality LDR image with a better overall appearance [1]. The prevailing scheme for MEF follows a weighted summation framework, where each exposure image is associated with a weight map of the same size. Burt and Adelson proposed the Laplacian pyramid in 1983 [27].
which has a profound impact on MEF \cite{1, 39}. To reproduce or enhance the local details, various edge-preserving filters, including bilateral filter\cite{40} and guided filter\cite{41} have been used for weight map computation. For quality assessment, Ma et al.\cite{28} developed one of the first quality metrics - MEF-SSIM, and successfully applied it to perceptual optimization of MEF methods in the space of raw pixels\cite{42} and DNN parameters\cite{29}, respectively. At Stage two, our method generates a sequence of LDR images, which can be treated as a pseudo-multi-exposure image stack because they correspond to the same HDR scene with different (simulated) maximum luminances. Similar techniques for pseudo-multi-exposure generation have been widely practiced in the related field of inverse tone mapping\cite{43}.

3 Proposed Method

In this section, we describe the proposed two-stage TMO that is biologically-inspired, computationally efficient, and perceptually optimized. Fig.\cite{2} shows the schematic diagrams. At Stage one, after preprocessing, we decompose the color-space-transformed calibrated HDR image into a normalized Laplacian pyramid, and input it to the tone mapping network, consisting of two DNNs for Laplacian pyramid estimation. At Stage two, we use the trained tone mapping network to produce a pseudo-multi-exposure image stack, corresponding to the same HDR image calibrated with different maximum luminances. We then train a fusion network for weight map estimation. The final high-quality LDR image is computed by a weighted fusion.

3.1 Stage One: Tone Mapping Network

3.1.1 Preprocessing

It is pivotal for biologically-inspired TMOs, including the proposed one, to work with calibrated HDR images, meaning that all pixels record the true luminance values (in the unit of candela per square meter, cd/m$^2$). This is because the responses of the HVS to different light levels are highly non-linear\cite{43}. HDR image calibration (also known as the photometric calibration) allows TMOs to make correct distinctions between bright and dim scenes. Otherwise, a day-lit HDR image in arbitrary units may be tone-mapped to a night scene with loss of structural details. In practice, the majority of HDR images circulated on the Internet are acquired without calibration, in which the recorded measurements $R$ are highly non-linear\cite{43}. HDR image calibration (also known as the photometric calibration) allows TMOs to make correct distinctions between bright and dim scenes. Otherwise, a day-lit HDR image in arbitrary units may be tone-mapped to a night scene with loss of structural details. In practice, the majority of HDR images circulated on the Internet are acquired without calibration, in which the recorded measurements $R$ are linearly proportional to the true luminances $S$ with an unknown scaling factor. To apply HVS-based TMOs to an uncalibrated HDR image, educated guesses about the minimum and maximum luminances of the original scene\cite{26}, denoted by $S_{\text{min}}$ and $S_{\text{max}}$, respectively, need to be made. Nevertheless, this is by itself a very challenging computer vision task. One significant advantage of the proposed TMO is that during training the tone mapping network, the HDR scenes can be calibrated with arbitrary (and even inaccurate) minimum and maximum luminances as a form of data augmentation. After specifying $S_{\text{min}}$ and $S_{\text{max}}$, we first convert the HDR image from RGB to HVS color space\cite{45}, and then linearly rescale the luminance measurements:

$$R = \frac{R - R_{\text{min}}}{R_{\text{max}} - R_{\text{min}}} \in [0, 1],$$

$$S = (S_{\text{max}} - S_{\text{min}}) \cdot R + S_{\text{min}}.$$  

We then decompose the “calibrated” luminance channel into the normalized Laplacian pyramid\cite{26}.

TABLE 1

| Layer | 1 | 2 | 3 | 4 | 5 | 6 |
|-------|---|---|---|---|---|---|
| Convolution | 3 | 3 | 3 | 3 | 3 | 3 |
| Dilatation | 1 | 2 | 4 | 8 | 1 | 1 |
| Width | 32 | 32 | 32 | 32 | 32 | 1 |
| Bias | $\times$ | $\times$ | $\times$ | $\times$ | $\times$ | $\times$ |
| Adaptive Normalization | $\checkmark$ | $\checkmark$ | $\checkmark$ | $\checkmark$ | $\checkmark$ | $\checkmark$ |
| LReLU Non-linearity | $\checkmark$ | $\checkmark$ | $\checkmark$ | $\checkmark$ | $\checkmark$ | $\checkmark$ |

TABLE 2

| Layer | 1 | 2 | 3 | 4 |
|-------|---|---|---|---|
| Convolution | 3 | 3 | 3 | 3 |
| Dilatation | 1 | 2 | 4 | 4 |
| Width | 24 | 24 | 24 | 1 |
| Bias | $\times$ | $\times$ | $\times$ | $\times$ |
| Adaptive Normalization | $\checkmark$ | $\checkmark$ | $\checkmark$ | $\times$ |
| LReLU Non-linearity | $\checkmark$ | $\checkmark$ | $\checkmark$ | $\times$ |

3.1.2 Network Architecture

The core of our tone mapping network are two DNNs to estimate the Laplacian pyramid of the LDR image from the normalized Laplacian pyramid of the corresponding HDR image. One DNN is shared to process all bandpass channels and the highpass channel, while the other is reserved for the lowpass channel. From a number of alternative networks, we employ the context aggregation network (CAN)\cite{46, 47} as our default architecture, which has been used to approximate and accelerate a wide range of image processing applications, including $\ell_0$ smoothing, style transfer, and pencil drawing. It allows receptive field expansion without compromising spatial resolution, which effectively aggregates global context information. The two CANs share the same architecture with six convolution layers, whose output has the same resolution as the input. The details are specified in Table\cite{1} which are manually optimized to be highly lightweight. Convolutions except the last one, are followed by the adaptive normalization (AN):

$$\text{AN}(Z) = \lambda_1 Z + \lambda_2 \text{BN}(Z),$$

where $\lambda_1$ and $\lambda_2 \in \mathbb{R}$ are two learnable parameters, and $Z$ denotes intermediate representation. The weight sharing across all bandpass channels and the highpass channel allows our method to process a normalized Laplacian pyramid of arbitrary levels. We employ the leaky rectified linear unit (LReLU) as the non-linear activation function:

$$\text{LReLU}(Z) = \max(\lambda_3 Z, 0),$$

where the parameter $0 \leq \lambda_3 < 1$ is made fixed during training. The lowpass channel is compressed by the other CAN with the same architecture. The output LDR image, constrained to have a luminance range of $[5, 300]$ cd/m$^2$, is reconstructed by collapsing the estimated Laplacian pyramid from the tone mapping network. In other words, we assume a fixed display device with the minimum and maximum luminances of $I_{\text{min}} = 5$ and $I_{\text{max}} = 300$, respectively, which is typical specifications of consumer-grade standard dynamic range displays.
A worth-mentioning difference of our tone-mapping network compared to the original CAN [47] is that all bias terms, including those in adaptive normalization are removed. As proved in [48], a bias-free DNN with piece-wise linear activation function (e.g., LReLU) is locally scale-invariant. That is, if the input is rescaled by a constant value, the output will be rescaled by the same amount:

$$g(\alpha Z_j) = \alpha g(Z_j), \quad (6)$$

where $j$ indexes the coefficients of the intermediate representation $Z$. Empirically, scale-invariance renders the tone-mapping network more robust to maximum luminance variations during training and testing (see Fig. 3).

### 3.1.3 NLPD as the Perceptual Metric

The NLPD metric, proposed in [26] and adopted as the objective function for our tone mapping network, is inspired by the physiology of the early visual system. Specifically, the luminances of the calibrated HDR image $S$ are firstly preprocessed by a power function, approximating the transformation of light to the response of retinal photoreceptors [26]:

$$S^{(1)} = S^\gamma, \quad (7)$$

After that, $S^{(1)}$ is partitioned recursively into frequency subbands via luminance subtraction, which mimics the center-surround receptive fields in the retina and the lateral geniculate nucleus [26]:

$$X^{(i+1)} = DLX^{(i)}, \quad i \in \{1, \ldots, M - 1 \}, \quad (8)$$

$$Z^{(i)} = X^{(i)} - LX^{(i+1)}, \quad (9)$$

$$Z^{(M)} = X^{(M)}, \quad (10)$$

where $D$ and $U$ represent linear down-/up-sampling operations, respectively, and $L$ denotes the lowpass filter, which is inherited from the Laplacian pyramid [27]. $M$ is the number of pyramid levels. The normalized Laplacian pyramid can be computed by dividing each coefficient with a weighted summation of neighboring coefficients (plus a constant) within each subband:

$$Y^{(i)} = Z^{(i)} \odot (P|Z^{(i)}| + C_0), \quad (11)$$

where $\odot$ represents the Hadamard division, and $P$ is a convolution filter optimized to eliminate the statistical redundancies [26]. $C_0$ is a small positive constant to avoid potential division by zero. Benefiting from such division normalization, we transform images with different dynamic ranges to a similar operating range. The normalized Laplacian pyramid representations of the HDR and tone-mapped images can be expressed as

$$f(S) = \left\{Y^{(i)}\right\}_{i=1}^{M} \quad \text{and} \quad f(I) = \left\{\tilde{Y}^{(i)}\right\}_{i=1}^{M}, \quad (12)$$

based on which we compute the NLPD metric:

$$\ell(S, I) = \left[ \frac{1}{M} \sum_{i=1}^{M} \left( \frac{1}{n^{(i)}} \sum_{j=1}^{n^{(i)}} |Y^{(i)}_{j} - \tilde{Y}^{(i)}_{j}|^{\alpha} \right)^{\frac{1}{\beta}} \right]^{\frac{1}{\gamma}}, \quad (13)$$

where $n^{(i)}$ denotes the number of coefficients in the $i$-th subband. The two exponents $\alpha$ and $\beta$ are computed for each frequency subband and for all subbands, respectively, which are optimized to match the human perception of image quality using a subject-rated image quality database [49].

### 3.2 Stage Two: Fusion Network

#### 3.2.1 Generation of Pseudo-Multi-Exposure Image Stack

The forward propagation of the tone mapping network and the computation of the NLPD metric in Stage one require the exact specification of the minimum and maximum luminances for uncalibrated HDR scenes. While most HVS-based TMOs are fairly robust to the minimum luminance $S_{\text{min}}$, making its setting straightforward, it is not the case for the maximum luminance $S_{\text{max}}$, which involves extensive human expertise and is thus time-consuming. Empirically, a higher estimated $S_{\text{max}}$ means that more simulated light is cast into the original scene [26], leading to better visibility of local structures, especially in dark regions (see Fig. 4 (a) and (c)). As there is no free lunch in computational photography, the measurement noise is also likely to be amplified, when $S_{\text{max}}$ is set extremely high (e.g., $S_{\text{max}} = 10^7$). Here, instead of manually picking a scene-dependent $S_{\text{max}}$ as done previously [26], [30], we make our method fully automatic by means of pseudo-MEF. Specifically, we first linearly sample $K$ maximum luminances from the range of $[10^3, 10^7]$ cd/m$^2$ in logarithmic scale, and calibrate the HDR image with each of the $K$ values. We then feed the calibrated images to the trained tone mapping network to create a sequence of $K$ candidate LDR images, which we call the pseudo-multi-exposure image stack, and will be fused to produce the final output image. We use the prefix “pseudo” because the image stack does not contain under- and over-exposure distortions, but instead may suffer from color saturation and noise artifacts. We will take advantage of these distortion characteristics to make a slight modification of the MEF-SSIM metric [28] such that the optimized fusion network is noise-aware.

1. Throughout this paper, we set $S_{\text{min}} = I_{\text{min}} = 5$ cd/m$^2$.
Fig. 4. Pseudo-multi-exposure image stacks produced by the tone mapping network and the corresponding weight maps, together with the final images by the fusion network. A brighter pixel in the weight map indicates that the corresponding LDR image pixel contributes more to the final image. It is clear that images with higher maximum luminances are tone-mapped with richer structural details but also severer degree of noise, while images with lower maximum luminances are more color-saturated and less detailed. The fusion network optimized by a variant of MEF-SSIM is noise-aware, generating reasonable weight maps that show strong preference to clean, high-contrast, well-exposed, and well-saturated patches. As a result, the output images combine the best perceptual aspects of the LDR image stacks.

3.2.2 Network Architecture

As shown in the right panel of Fig. 2, our fusion network is also implemented by a CAN that predicts the weight maps \{W^{(k)}\} with the same resolution of the input pseudo-multiple-exposure image stack \{I^{(k)}\}. The network specification is given in Table 2, which is shallower than the tone mapping network. The parameters are shared by all pseudo-exposure images, allowing an arbitrary-length stack to be handled. The last layer predicts the weight maps, which are used to compute the final output image by a weighted summation:

\[ F = \sum_{k=1}^{K} W^{(k)} \odot I^{(k)}, \]

where \( \odot \) denotes the Hadamard product.

3.2.3 MEF-SSIM Variant as the Perceptual Metric

The MEF-SSIM metric proposed in [28] provides an accurate quality characterization of multi-exposure fused images. It first decomposes an image patch \( x^{(k)} \) into three components - mean intensity, signal contrast and signal structure:

\[
x^{(k)} = \|x^{(k)} - \mu_k\| \cdot \frac{x^{(k)} - \mu_k}{\|x^{(k)} - \mu_k\|} + \mu_k
\]

\[
= \|\tilde{x}^{(k)}\| \cdot \frac{\tilde{x}^{(k)}}{\|\tilde{x}^{(k)}\|} + \mu_k
\]

\[
= c_k \cdot s_k + \hat{l}_k,
\]

where \( \|\cdot\| \) denotes the \( \ell_2 \)-norm. \( l_k = \mu_k \), \( c_k = \|\tilde{x}^{(k)}\| \), and \( s_k = \frac{\tilde{x}^{(k)}}{\|\tilde{x}^{(k)}\|} \) represent the mean intensity, the signal contrast, and the signal structure, respectively. MEF-SSIM computes the intensity of the desired patch by

\[
\hat{i} = \frac{\sum_{k=1}^{K} w_l(g_k, \hat{l}_k) \mu_k}{\sum_{k=1}^{K} w_l(g_k, l_k)},
\]

where \( w_l(\cdot) \) is specified by a two-dimensional Gaussian to measure the well-exposedness:

\[
w_l(g_k, l_k) = \exp \left(-\frac{(g_k - \tau)^2}{2\sigma^2} - \frac{(l_k - \tau)^2}{2\sigma^2} \right).
\]
\( \sigma_g \) and \( \sigma_l \) are photometric spreads, and \( \tau = 0.5 \) stands for the mid-intensity value. The desired contrast is the highest one across all exposures:

\[
\hat{c} = \max_{1 \leq k \leq K} c_k. \tag{18}
\]

The desired structure is calculated by a weighted summation followed by \( \ell_2 \)-normalization:

\[
\hat{s} = \frac{\tilde{s}}{\| \tilde{s} \|}, \quad \text{where} \quad \tilde{s} = \sum_{k=1}^{K} w_k(\hat{x}^{(k)}) s_k / \sum_{k=1}^{K} w_k(\hat{x}^{(k)}). \tag{19}
\]

In the original MEF-SSIM for perceptual optimization \([29], [42]\), \( w_k(\cdot) \) is a Kronecker delta function that identifies the structure vector corresponding to the maximum contrast (i.e., \( \hat{c} \)). This is primarily motivated to avoid selecting under- or over-exposed patches with essentially no structures. However, such choice is not wise for the pseudo-multi-exposure image stack created in Sec. 3.2.1, where the structure vector with the maximum contrast is highly likely to contain (amplified) measurement noise (see Fig. 4(a)). Here, we propose a simple yet effective remedy for MEF-SSIM to make it noise-aware: change the delta function \( w_k(\cdot) \) to select the structure vector with the median signal contrast instead. As shown in Fig. 4 the fusion network optimized by the MEF-SSIM variant is able to generate weight maps that show strong preference to clean, high-contrast, well-exposed, and well-saturated patches.

The remaining construction of MEE-SSIM is left intact, where we first compute the desired image patch by fusing the three components:

\[
\hat{x} = \hat{c} \cdot \hat{s} + \hat{I}, \tag{20}
\]

and make SSIM-like local quality measurements:

\[
S \left( \left\{ \hat{x}^{(k)} \right\}, f \right) = \frac{(2 \mu_{\hat{x}} \mu_{f} + C_1)(2 \sigma_{\hat{x}f} + C_2)}{\left( \mu_{\hat{x}}^2 + \mu_{f}^2 + C_1 \right) \left( \sigma_{\hat{x}}^2 + \sigma_{f}^2 + C_2 \right)}, \tag{21}
\]

where \( \mu_{\hat{x}} \) and \( \mu_{f} \) represent the mean intensities of the desired patch \( \hat{x} \) and a given fused patch \( f \), respectively, and \( \sigma_{\hat{x}} \), \( \sigma_{f} \), and \( \sigma_{\hat{x}f} \) indicate the local variances of \( \hat{x} \) and \( f \), and their covariance, respectively. \( C_1 \) and \( C_2 \) are two small positive constants for numerical stability. The local quality measurements are averaged to produce an overall quality estimate for the fused image.

### 3.3 Color Reproduction

As discussed before, both the tone mapping network and the fusion network work with the luminance channel due primarily to the two perceptual metrics NLPD \([26]\) and MEF-SSIM \([28]\) accept grayscale images only. To recover the color appearance of the final output image, we adopt the method in \([35], [45]\) that is also widely adopted by other recent methods \([5], [30]\):

\[
F^{(c)} = \left( \frac{S^{(c)}}{S} \right)^\rho F, \tag{22}
\]

where \( c \in \{ R, G, B \} \) indexes the RGB channels, and \( \rho \) controls the color saturation. \( S \) and \( F \) represent the luminance channel before and after tone mapping, respectively.

### 3.4 Model Training and Testing

We collect a dataset of 435 HDR images mainly from \([10], [50]–[55]\), among which 395 are utilized for training while 40 for testing. We choose to train the entire method sequentially. We first optimize the tone mapping network by minimizing the NLPD metric with the default setting \([26]\). Specifically, the non-linearity parameter \( \gamma \) in Eq. (7) is set to \( 1/\gamma \). For bandpass and highpass channels, the convolutional filter \( P \) is set to \( [0.05, 0.25, 0.4, 0.25, 0.05] \), and the constant \( C_0 \) in Eq. (11) is set to 0.17. While for the lowpass channel, the filter \( P \) is set to the identity matrix and \( C_0 \) is set to 4.86. The two optimized exponents \( \alpha \) and \( \beta \) in Eq. (13) are set to 2.0 and 0.6, respectively. We set the negative slope \( \lambda_3 = 0.2 \) in LReLU. During the training of the tone mapping network, each HDR image is arbitrarily calibrated by maximum luminances sampled randomly from \( \{10^3, 10^4, 10^5, 10^6, 10^7\} \) cd/m\(^2\), and each calibrated image is decomposed into a normalized Laplacian pyramid with five levels.

Adam \([56]\) is employed as the stochastic optimizer with an initial learning rate of \( 10^{-3} \) and a mini-batch size of 4. We decay the learning rate every 1,000 epochs by a factor of 10, and we train the tone mapping network for a total of 2,000 epochs. Random cropping and flipping have been employed to augment the training data. After Stage one optimization, we use the trained tone mapping network to create the pseudo-multi-exposure image stack for each HDR scene by setting the number \( K = 5 \). That is, we sample the same five discrete values \( \{10^3, 10^4, 10^5, 10^6, 10^7\} \) cd/m\(^2\) for stack generation. It is noteworthy that the fusion network is designed to accept an LDR image stack of arbitrary length and resolution. The two Gaussian spread parameters in Eq. (17) of MEF-SSIM are inherited from previous publications \([29], [42]\): \( \sigma_g = 0.2 \) and \( \sigma_l = 0.5 \).

The training of the fusion network is nearly identical to that of the tone mapping network, except that we allocate the LDR image stack along the batch dimension as an efficient implementation of parameter sharing. This makes the mini-batch size to be one.

During testing, we resize each HDR image so that the short side has a size of 512, and calibrate it with five maximum luminance values \( \{10^3, 10^4, 10^5, 10^6, 10^7\} \). We feed the calibrated HDR images of the same content to the tone mapping network to generate the pseudo-multi-exposure image stack, which will be subsequently fed into the fusion network to produce the final high-quality LDR image (with color reproduction).

### 4 Experiments

In this section, we compare the proposed method with classical and recent DNN-based TMOs in terms of subjective quality, objective quality (by TMQI \([22]\) and NLPD \([26]\)), and computational time. Moreover, we carry out a subjective experiment to verify the perceptual gains achieved by our method. In addition, we conduct a series of ablation experiments to identify the core components of the proposed method.

We select 15 TMOs for comparison, including Drago03 \([4]\), Reinhard05 \([11]\), Kim08 \([12]\), WLS \([13]\), LLF \([14]\), Bruce14 \([15]\), GR \([17]\), NLPD-Opt \([26]\), Khan18 \([58]\), Liang18 \([5]\), Zhang20 \([57]\), Rana20 \([20]\), Zhang21 \([7]\), Vinker21 \([6]\), and Yang21 \([23]\). Rana20, Zhang21, Vinker21, and Yang21 are DNN-based TMOs, while the others are conventional operators, among which Drago03 Reinhard05 and Kim08 are global operators, and the rest are local operators.
Drago03 relies on an adaptive logarithmic mapping, while Reinhard05 uses a practical S-shaped curve. Kim08 improves upon Drago03 by refining the logarithmic curve from the perspective of photosensitive material characteristics. WLS casts HDR tone mapping into a weighted least squares problem, and LLF involves ingenious manipulation of Laplacian pyramid coefficients. Bruce14 achieves tone mapping via MEF. GR and Liang18 are based on the two-layer decomposition in the gradient domain. Zhang20 is a retina-inspired TMO (by modeling retina horizontal and bipolar cells). NLPD-Opt compresses the HDR image by directly minimizing the NLPD metric in the image space. Thus, given sufficient iterations, NLPD-Opt is regarded as the lower bound for all TMOs in terms of NLPD. All DNN-based methods except Vinker21 require paired images for supervision. Zhang21 acquires the ground-truth LDR images from expert manipulation, while Rana20 and Yang21 select the best LDR image produced by a list of existing TMOs objectively (via TMQI) and subjectively (via human inspection), respectively. Zhang21 introduces a semi-supervised strategy to further leverage the real-world LDR image distribution as a form of regularization. Like our method, Vinker21 does not need the ground-truth LDR images for training, which is, however, achieved by a rather empirical combination of several loss functions. All algorithms are implemented either by Banterle et al. [59] in their great MATLAB Toolbox or by the respective authors. We test them with the default settings.

4.1 Main Results

4.1.1 Qualitative Comparison

Fig. 1 compares the tone mapping results of linear scaling, Drago03 [4], Liang18 [5], Vinker21 [6], Zhang21 [7], and our method on an “Outdoor Table” HDR scene. The linear scaling creates an over-exposed image in most regions. Drago03 generates a relatively dark appearance with reduced global contrast. The results produced by Zhang21 and Vinker21 look pale with over-saturation appearances; Liang18 performs better than the two methods with enhanced local details. In contrast, our method significantly outperforms the competing ones in terms of color and detail reproduction, giving rise to a more perceptually appealing overall appearance.

Fig. 5 compares the tone mapping results of Kim08 [12], WLS [14], GR [17], Zhang20 [57], Rana20 [20], Zhang21 [7], Vinker21 [6], and our method on a “Forest” HDR scene. The global TMO Kim08 [12] contains noticeable under-exposed areas. Local TMOs like WLS [14] and Zhang20 [17] focus on local detail enhancement, and ultimately lead to edge-related artifacts. Such over-enhancement is less pronounced for GR. The results by DNN-based methods Rana20 [20] and Zhang21 [7] are slightly over-saturated with fewer details. Vinker21 [6] generates a natural-looking LDR image similar to that of the proposed method, despite that the former is weaker at reproducing warm colors.

Fig. 6 compares the tone mapping results of LLF [15], NLPD-Opt [26], Khan18 [58], Zhang20 [57], Zhang21 [7], Vinker21 [6], Yang21 [23], and our method on a “Classroom” HDR scene. Our method gives a more faithful color reproduction for the chairs and the carpet. Moreover, it does an excellent job in recovering the fine structures of the wood floor and the soft shadow. On the contrary, most competing methods suffer from the problems of reduced global contrast, color cast, and detail loss. Of particular interest, NLPD-Opt [26] tends to overshoot the details and is slightly noisy with a manually optimized $S_{\text{max}} = 10^6 \text{ cd/m}^2$ for perceptual
quality (not physical plausibility). Comparison with NLPD-Opt provides strong evidence that our method automatically combines the best perceptual aspects of the LDR stack for noise-aware tone mapping.

Fig. 7 compares the tone mapping results of Drago03, Reinhard05, Kim08, Khan18, NLPD-Opt ($S_{\text{max}} = 10^5 \text{ cd/m}^2$), Zhang20, Yang21, and our method on a “Night Street” HDR scene. Global TMOs - Drago03, Reinhard05, and Kim08 lose many details in the regions of the dark sky and the bright light sources. Local TMOs - Khan18, NLPD-Opt, and Zhang20 perform better in detail preservation. Nevertheless, the colors reproduced by Khan18 and Zhang20 are unnatural. The result by Yang21 looks over-exposed and over-saturated. In contrast, our method gives a more vivid color appearance with balanced local contrast and details.

4.1.2 Quantitative Comparison

To evaluate the performance of the competing TMOs quantitatively, we adopt two objective metrics: TMQI and NLPD. TMQI is specifically designed for cross-dynamic-range image...
outperform global operators in terms of TMQI. This is not surprising because TMQI is biased towards comparing local structure similarity, which is the design focus of local TMOs. Such result discrepancy is less pronounced in terms of NLPD. Second, DNN-based methods are not necessarily better than the conventional methods. This is also reasonable as it is generally difficult (and conceptually impossible) to specify ground-truth LDR images, otherwise the problem of HDR image tone mapping is readily solvable. As a consequence, with a limited number of paired data (and a set of unpaired data) for supervised (and semi-supervised) training, the learned DNNs may be weak at generalizing to unseen challenging HDR scenes, producing unexpected visually annoying appearances. Third, as expected, NLPD-Opt achieves the best performance in terms of NLPD, followed by the proposed TMO and LLF. Fourth, it is interesting to see that the proposed method achieves the best performance measured by TMQI, which provides strong justifications for the perceptual optimality of our method.

We test the computation time of our method with the 15 TMOs on a computer with a 4.4GHz CPU, a 64G RAM, and an NVIDIA GTX 3080Ti GPU. All conventional methods are based on the MATLAB implementations [59], while the DNN-based methods are implemented using PyTorch (or Tensorflow for Yang21). It can be observed from Table 3 that our CPU-only method runs among the fastest local TMOs thanks to the manually optimized lightweight network architectures. Moreover, when equipped with the GPU, the running time of our method is shortened to 0.0177 second, which facilitates real-time applications.

### 4.1.3 Subjective Experiment

In order to further validate that NLPD and MEF-SSIM optimization indeed result in perceptual gains of our method, we carry out a subjective experiment in a normal indoor office. In particular, we select 20 HDR images of diverse content variations and quality evaluation. It combines structural fidelity (denoted by SF) and statistical naturalness (denoted by SN) measurements to assess a tone-mapped image with reference to the corresponding HDR image. NLPD can also be applied to the cross-dynamic-range scenario because of the divisive normalization step, which serves as a form of local gain control. A larger TMQI or a smaller NLPD value indicates better predicted quality. Table 3 shows the ten-fold cross-validation results, from which we have several interesting observations. First, local operators generally

| TMO     | TMQI↑ | SF↑   | SN↑   | NLPD↓ | Time  |
|---------|-------|-------|-------|-------|-------|
| Drago03 | 0.8394| 0.8839| 0.2520| 0.2404| 0.28  |
| Reinhard05 | 0.8692| 0.8308| 0.4923| 0.2474| 0.27  |
| Kim08   | 0.8385| 0.8899| 0.2412| 0.2262| 0.66  |
| WLS     | 0.7997| 0.8117| 0.1966| 0.2303| 3.35  |
| LLF     | 0.9187| 0.9178| 0.6200| 0.2076| 365.43|
| Bruce14 | 0.8277| 0.7304| 0.4188| 0.2972| 10.23 |
| GR      | 0.8738| 0.8394| 0.4888| 0.2943| 12.98 |
| NLPD-Opt| 0.8782| 0.8809| 0.4551| 0.1776| 172.36|
| Khan18  | 0.8739| 0.8807| 0.4000| 0.2479| 3.21  |
| Liang18 | 0.8955| 0.8458| 0.6142| 0.2167| 1.98  |
| Zhang20 | 0.8916| 0.8301| 0.6061| 0.2456| 3.26  |
| Rana20  | 0.8678| 0.8818| 0.4101| 0.2484| 0.12  |
| Zhang21 | 0.8462| 0.7964| 0.4496| 0.2450| —     |
| Vinker21| 0.8857| 0.8612| 0.5320| 0.2143| 0.10  |
| Yang21  | 0.8940| 0.8458| 0.6005| 0.2259| 0.33  |
| Ours    | 0.9202| 0.8817| 0.6780| 0.1913| 1.06  |

Fig. 8. Comparison of our method with Reinhard05 [11], WLS [14], LLF [15], Bruce14 [16], Liang18 [5], Vinker21 [6], and Yang21 [23] on a “Windowsill” HDR scene.
Fig. 9. Subjective rankings of the competing TMOs in our experiment.

(a) One-level  (b) Two-level

(c) Three-level  (d) Four-level

(e) Five-level  (f) Six-level

Fig. 10. Tone mapping results of the “Workshop” HDR scene with different input pyramid levels.

Table 4: Ten-fold cross-validation results with different input pyramid levels.

| Pyramid Level | TMQI↑ | SF↑ | SN↑ | NLPD↓ | Time |
|---------------|-------|-----|-----|-------|------|
| One           | 0.8832| 0.7765| 0.6412| 0.2082| 0.0138|
| Two           | 0.8962| 0.8205| 0.6396| 0.2005| 0.0140|
| Three         | 0.9113| 0.8565| 0.6671| 0.1946| 0.0166|
| Four          | 0.9202| 0.8817| 0.6780| 0.1913| 0.0174|
| Five          | 0.9224| 0.8885| 0.6799| 0.1905| 0.0177|
| Six           | 0.9238| 0.8919| 0.6827| 0.1897| 0.0186|

worst to the bottom right and the best visual result to the top left.

We treat the raw ranking data directly as quality scores, with 1 to 16 denoting the worst to the best quality. The performance of each TMO in terms of the mean opinion score (MOS) across all subjects and all scenes is illustrated in Fig. 9. Our method performs the best, even outperforming NLPD-Opt that optimizes for the same objective in the image space. We believe this arises because NLPD-Opt sometimes overfits NLPD during the single-example optimization, and creates an over-enhanced (and even noisy) appearance similar to GR [17]. In Stage two, our MEF-SSIM-optimized method is able to rectify the over-enhancement problem. Nevertheless, NLPD-Opt ranks second in our subjective experiment, verifying the suitability of NLPD as an objective quality measure for benchmarking existing TMOs and guiding the design of more perceptual TMOs. Moreover, some DNN-based methods achieve low rankings, and are far behind some conventional methods. We treat this as caveats of the current ad-hoc combination of loss functions as the training objective without verifying its perceptual relevance.

4.1.4 Ablation Experiments

We conduct a series of ablation experiments to single out the contributions of the biologically-inspired design (i.e., normalized Laplace decomposition) and the perceptual optimization (i.e., NLPD for the tone mapping network and MEF-SSIM for the fusion network).

We first analyze the effect of the input pyramid level on final visual quality. Note that one level corresponds to directly feeding the raw HDR image into a single network for tone mapping. As shown in Fig. 10, more levels lead to improved detail reproduction at the cost of increased computational complexity, which is also evidenced by the quantitative results in Table 4. The default five-level pyramid keeps a good balance between visual quality and computational speed.

We then disable the fusion network, and switch NLPD to three other objective functions: mean absolute error (MAE), SSIM [38], and TMQI [22], while fixing the tone mapping network architecture. Fig. 11 shows the optimization results, which are optimal under their respective objectives. As can be seen, the NLPD-optimized image better preserves structures outside the window with few artifacts. Qualitatively, we find these results consistent across a wide range of HDR scenes.

In our experiments, we set the luminance range of \( S_{\text{max}} \) to be \([10^3, 10^7] \text{ cd}/\text{m}^2\) for photometric calibration, which covers the maximum luminances of most challenging HDR scenes. Here, we compare the tone mapping results calibrated by three luminance ranges, and tone-map them using the competing TMOs. We invite 20 participants, including 14 males and 6 females aged between 18 and 30. All subjects have general knowledge of image processing but are blind to the detailed purpose of this study. We display each set of 16 tone-mapped images (with the same visual content) simultaneously to the subjects on two calibrated monitors, and the subjects are free to zoom in to any portion of the images for more careful comparison. We ask the subjects to position (i.e., rank) the images according to their perceived quality, placing the

4. For example, a luminance of \( 10^7 \text{ cd}/\text{m}^2 \) corresponds to the filament of a clear incandescent lamp. See https://en.wikipedia.org/wiki/Orders_of_magnitude_luminance for more information.
We have introduced a biologically-inspired TMO based on nine pseudo-multi-exposure images, respectively. The maximum luminances used for calibration are all sampled uniformly from different maximum luminance ranges - $[10^3, 10^5]$, $[10^2, 10^6]$, and $[10^0, 10^2]$ cd/m$^2$. The results are shown in Fig. 12, where we find that the “Arched Roof” scene with a higher dynamic range benefits from a higher $S_{\text{max}}$, and the tone-mapped image is well-saturated and more detailed. For the “Outer Wall” with a lower dynamic range, the tone mapping result is relatively insensitive to the setting of $S_{\text{max}}$. To make our method more widely applicable, it is preferred to work with a wider maximum luminance range, and let the fusion network decide which pseudo-exposures to rely on (see Fig. 4).

We last probe the robustness of the fusion network by varying the length of the pseudo-multi-exposure image stack. Specifically, we generate four image stacks consisting of three, five, seven, and nine pseudo-multi-exposure images, respectively. The maximum luminances used for calibration are all sampled uniformly from $[10^3, 10^7]$ cd/m$^2$ in the logarithmic scale. The results are shown in Fig. 13 where we observe that although TMQI and NLPD improve slightly with the length of the image stack, such improvements are barely noticeable by the human eye.

5 Conclusion and Discussion

We have introduced a biologically-inspired TMO based on lightweight tone mapping and fusion networks, optimized sequentially for two perceptual metrics, NLPD and MEF-SSIM. The tone mapping network is trained to generate the pseudo-multi-exposure image stack by varying the maximum luminance of the input HDR image. The fusion network is responsible for fusing the image stack into a final high-quality image that is noise-free, high-contrast, well-exposed, and well-saturated. Without using the ground-truth LDR images for supervised training, our optimized method matches and exceeds the state-of-the-art across a variety of HDR natural scenes. The perceptual advantages of our method are further verified by another perceptual quality metric - TMQI and in a form subjective experiment.

Our biologically-inspired TMO gives us an opportunity to artificially manipulate the light source in the scene by linearly rescaling the maximum luminance $S_{\text{max}}$. We take advantage of this to fully automate the proposed TMO. Alternatively, it is worth exploring our tone mapping network in low-light and normal-light image enhancement (see Fig. 14). Meanwhile, in our experiments, we assume a fixed display constraint with a minimum luminance of $I_{\text{min}} = 5$ and a maximum luminance of $I_{\text{max}} = 300$, while the luminance ranges for displays on the market vary. Therefore, in the future, we will take steps to incorporate various display constraints into the proposed perceptual optimization framework.

ACKNOWLEDGMENT

The authors would like to dedicate this humble piece of work to Prof. Ted Adelson for his 70th birthday. Ted was the great mind behind the Laplacian pyramid, which was proposed nearly 40 years ago, and is the core building block of our current work.

REFERENCES

[1] T. Mertens, J. Kautz, and F. Van Reeth, “Exposure fusion: A simple and practical alternative to high dynamic range photography,” Computer Graphics Forum, vol. 28, no. 1, pp. 161–171, 2009.
[2] B. Hoefflinger, High-Dynamic-Range (HDR) Vision. Springer, 2007.
[3] E. Reinhard, W. Heidrich, P. Debevec, S. Pattanaik, G. Ward, and K. Myszkowski, High Dynamic Range Imaging: Acquisition, Display, and Image-based Lighting. Morgan Kaufmann, 2010.
[4] F. Drago, K. Myszkowski, T. Annen, and N. Chiba, "Adaptive logarithmic mapping for displaying high contrast scenes," Computer Graphics Forum, vol. 22, no. 3, pp. 419–426, 2003.
[5] Z. Liang, J. Xu, D. Zhang, Z. Cao, and L. Zhang, “A hybrid $\ell_1-\ell_0$ layer decomposition model for tone mapping,” in IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 4758–4766.
[6] Y. Vinker, I. Huberman-Spiegelglas, and R. Fattal, “Unpaired learning for high dynamic range image tone mapping,” in IEEE International Conference on Computer Vision, 2021, pp. 14 657–14 666.
[7] N. Zhang, Y. Zhao, C. Wang, and R. Wang, “A real-time semi-supervised deep tone mapping network,” IEEE Transactions on Multimedia, vol. 24, no. 217, pp. 2815–2827, 2022.

5. https://sites.google.com/view/ted-fest-2022/
Fig. 13. Tone mapping results of (a)-(d) the “Leafy Plant” and the (e)-(h) the “Outdoor Corridor” HDR scenes with pseudo-multi-exposure image stacks of different lengths. (a): TMQI = 0.9649, NLPQ = 0.1633. (b): TMQI = 0.9655, NLPQ = 0.1630. (c): TMQI = 0.9659, NLPQ = 0.1629. (d): TMQI = 0.9660, NLPQ = 0.1628. (e): TMQI = 0.9114, NLPQ = 0.1666. (f): TMQI = 0.9126, NLPQ = 0.1664. (g): TMQI = 0.9133, NLPQ = 0.1661. (h): TMQI = 0.9236, NLPQ = 0.1660.

Fig. 14. Visual examples of low-light image enhancement. (a)/(c) Low-light images. (b)/(d) Enhanced images corresponding to (a)/(c) by the proposed TMO.

8. G. J. Ward, “The RADIANCE lighting simulation and rendering system,” in Annual Conference on Computer Graphics and Interactive Techniques, 1994, pp. 459–472.
9. G. W. Larson, H. Rushmeier, and C. Piatko, “A visibility matching tone reproduction operator for high dynamic range scenes,” IEEE Transactions on Visualization and Computer Graphics, vol. 3, no. 4, pp. 291–306, 1997.
10. J. Tumblin and H. Rushmeier, “Tone reproduction for realistic images,” IEEE Computer Graphics and Applications, vol. 13, no. 6, pp. 42–48, 1993.
11. E. Reinhard and K. Devlin, “Dynamic range reduction inspired by photoreceptor physiology,” IEEE Transactions on Visualization and Computer Graphics, vol. 11, no. 1, pp. 13–24, 2005.
12. M. H. Kim and J. Kautz, “Consistent tone reproduction,” in International Conference on Computer Graphics and Imaging, 2008, pp. 152–159.
13. F. Durand and J. Dorsey, “Fast bilateral filtering for the display of high-dynamic-range images,” in Annual Conference on Computer Graphics and Interactive Techniques, 2002, pp. 257–266.
14. Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski, “Edge-preserving decompositions for multi-scale tone and detail manipulation,” in Annual Conference on Computer Graphics and Interactive Techniques, 2008, pp. 67:1–67:10.
15. S. Paris, S. W. Hasinoff, and J. Kautz, “Local Laplacian filters: Edge-aware image processing with a Laplacian pyramid,” in Annual Conference on Computer Graphics and Interactive Techniques, 2011, pp. 68:1–68:12.
16. N. D. Bruce, “ExpoBlend: Information preserving exposure blending based on normalized log-domain entropy,” Computers & Graphics, vol. 39, no. 4, pp. 12–23, 2014.
17. T. Shibata, M. Tanaka, and M. Okutomi, “Gradient-domain image reconstruction framework with intensity-range and base-structure constraints,” in IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2745–2753.
18. R. Montulet and A. Briassouli, “Deep learning for robust end-to-end tone mapping,” in British Machine Vision Conference, 2019, pp. 1–12.
19. K. Panetta, L. Kezebou, V. Vludare, S. Agaian, and Z. Xia, “TMO-Net: A parameter-free tone mapping operator using generative adversarial network, and performance benchmarking on large scale HDR dataset,” IEEE Access, vol. 9, no. 3227, pp. 39500–39517, 2021.
20. A. Rana, P. Singh, G. Valenzise, F. Dufaux, N. Komodakis, and A. Smolic, “Deep tone mapping operator for high dynamic range images,” IEEE Transactions on Image Processing, vol. 29, no. 98, pp. 1285–1298, 2020.
21. N. Zhang, C. Wang, Y. Zhao, and R. Wang, “Deep tone mapping network in HSV color space,” in IEEE Visual Communications and Image Processing, 2019, pp. 1–4.
22. H. Yeganeh and Z. Wang, “Objective quality assessment of tone-mapped images,” IEEE Transactions on Image Processing, vol. 22, no. 2, pp. 657–667, 2013.
23. J. Yang, Z. Liu, M. Lin, S. Yanushkevich, and O. Yadid-Pecht, “Deep reformulated Laplacian tone mapping,” arXiv preprint arXiv:2102.00348, 2021.
24. J. Tumblin and H. Rushmeier, “Tone reproduction for realistic images,” IEEE Computer Graphics and Applications, vol. 13, no. 6, pp. 42–48, 1993.
25. K. Ma, H. Yeganeh, K. Zeng, and Z. Wang, “High dynamic range image compression by optimizing tone mapped image quality index,” IEEE Transactions on Image Processing, vol. 24, no. 10, pp. 3086–3097, 2015.
26. V. Laparra, A. Berardino, J. Ballé, and E. P. Simoncelli, “Perceptually optimized image rendering,” Journal of the Optical Society of America A, vol. 34, no. 9, pp. 1511–1525, 2017.
27. P. Burt and E. Adelson, “The Laplacian pyramid as a compact image code,” IEEE Transactions on Communications, vol. 31, no. 4, pp. 532–540, 1983.
Kede Ma (S'13–M'18) received the B.E. degree from the University of Science and Technology of China, Hefei, China, in 2012, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2014 and 2017, respectively. He was a Research Associate with the Howard Hughes Medical Institute and New York University, New York, NY, USA, in 2018. He is currently an Assistant Professor with the Department of Computer Science, City University of Hong Kong. His research interests include perceptual image processing, computational vision, and computational photography.