IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 10, 2021, accepted February 27, 2021, date of publication March 8, 2021, date of current version March 16, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3064295

TMO-Net: A Parameter-Free Tone Mapping Operator Using Generative Adversarial Network, and Performance Benchmarking on Large Scale HDR Dataset

KAREN PANETTA[1], (Fellow, IEEE), LANDRY KEZEBOU[1], (Graduate Student Member, IEEE), VICTOR OLUDARE[1], (Graduate Student Member, IEEE), SOS AGAIAN[2], (Fellow, IEEE), AND ZEHUA XIA[1]

[1]Department of Electrical and Computer Engineering, Tufts University, Medford, MA 02155, USA
[2]Department of Computer Science, City University New York (CUNY), New York, NY 100161, USA

Corresponding author: Landry Kezebou (landry.kezebou@tufts.edu)

ABSTRACT Currently published tone mapping operators (TMO) are often evaluated on a very limited test set of high dynamic range (HDR) images. Thus, the resulting performance index is highly subject to extensive hyperparameter tuning, and many TMOs exhibit sub-optimal performance when tested on a broader spectrum of HDR images. This indicates that there are deficiencies in the generalizable applicability of these techniques. Finally, it is a challenge developing parameter-free tone mapping operators using data-hungry advanced deep learning methods due to the paucity of large scale HDR datasets. In this paper, these issues are addressed through the following contributions: a) a large scale HDR image benchmark dataset (LVZ-HDR dataset) with multiple variations in sceneries and lighting conditions is created to enable performance evaluation of TMOs across a diverse conditions and scenes that will also contribute to facilitate the development of more robust TMOs using state-of-the-art deep learning methods; b) a deep learning-based tone mapping operator (TMO-Net) is presented, which offers an efficient and parameter-free method capable of generalizing effectively across a wider spectrum of HDR content; c) finally, a comparative analysis, and performance benchmarking of 19 state-of-the-art TMOs on the new LVZ-HDR dataset are presented. Standard metrics including the Tone Mapping Quality Index (TMQI), Feature Similarity Index for Tone Mapped images (FSITM), and Natural Image Quality Evaluator (NIQE) are used to qualitatively evaluate the performance index of the benchmarked TMOs. Experimental results demonstrate that the proposed TMO-Net qualitatively and quantitatively outperforms current state-of-the-art TMOs.

INDEX TERMS Deep learning, GAN-based tone mapping, TMO benchmarking, tone mapping, TMO-Net HDR dataset, parameter-free tone mapping.

I. INTRODUCTION

Despite 20 years of research and dozens of proposed algorithms, tone-mapping for high dynamic range images remains a difficult problem. The basic challenge is straightforward: the real world contains many scenes with enormous dynamic range, but most image formats store brightness using only 8 bits and most displays are likewise limited. Even displays with extended dynamic range continue to fall short of reproducing the range of brightness visible in the world. Thus, it is nearly always necessary to reduce the overall contrast of an HDR image so that the image can be transmitted and displayed on low dynamic range (LDR) displays.

A plethora of tone mapping operators have been proposed to compress the dynamic range of HDR images while maintaining a pleasing and natural appearance. Each method strives to provide improvements over previous methods to achieve high-quality rendering [1]–[6]. However, due to the lack of availability of a standardized dataset, these methods are often only tested on a handful of images and do not consider a wide variety of scenes and illuminations [7].

Overcoming the challenge of hyperparameter tuning to achieve good results is often understated. Moreover, it is difficult to assess the informational context of hyperparameters

The associate editor coordinating the review of this manuscript and approving it for publication was Gulistan Raja.

when they are assigned non-descriptive names such as 'alpha' and 'beta', with no clear guidance or analysis on how to intuitively choose their optimal values. As a result, tone mapping operators are severely challenged when presented with other HDR images for which these of the hyperparameters are not known.

It is common to find that some operators work well on daytime imagery, for example, but much less effectively on nighttime imagery or vice versa. The process of manually navigating the hyperparameter space to produce the most pleasing result is inefficient and impedes the wider adoption of TMOs. Due to these challenges, some popular commercial applications [8] have abandoned tone mapping as a viable solution. Instead, multiple exposures are used to create an HDR image and from this, two LDR images are synthesized followed by exposure fusion [9], [10]. There remains a need for a more robust and generalizable tone mapping operator free of hyperparameters.

More recent tone mapping operators have used a deep neural network approach (particularly generative adversarial networks) to address some of the limitations described above [7], [11]. However, the lack of a standard large-scale HDR dataset makes it difficult to develop and compare deep-learning algorithms for tone mapping. This lack of a standard dataset also makes it harder to quantitatively benchmark traditional TMOs and evaluate progress in the field.

This work addresses these with the following contributions:

1) First, a large scale HDR image benchmark dataset (LVZ-HDR dataset) with high variability in sceneries and lighting conditions is created to promote performance evaluation of TMOs, and to facilitate the development of more robust and more generalizable TMOs using state-of-the-art deep learning methods.
2) Second, a deep learning-based tone mapping operator (TMO-Net) is proposed. This operator is efficient and parameter-free and performs well across a wide range of HDR scenes.
3) Third, a comparative analysis and performance benchmarking of 19 state-of-the-art TMOs on the proposed large scale LVZ-HDR dataset is performed using TMQI, FSITM, and other relevant evaluation metrics.
4) Finally, a subjective study to validate the qualitative performance and the merits of the proposed TMO-Net is conducted.

The rest of this paper is organized as follows: Section 2 takes a closer look at various categories of traditional TMOs, as well as the most recent Neural Network approach. Section 3 details the network architecture, objective functions, and working principles of the proposed TMO-Net. Section 4 details how the LVZ-HDR dataset was created, and also describes chosen evaluation metrics. Section 5 evaluates and analyzes the performance of selected tone mapping algorithms on the proposed dataset, and presents benchmark results. Finally, Section 6 summarizes findings, presents limitations and suggests directions for future work.

II. RELATED WORK

Given the continued relevance of tone mapping in the image processing pipeline, numerous tone mapping operators have been proposed. Traditional TMOs can be categorized into global [2], [12]–[16], local [13], [15], [17], segmentation [1], [18], and gradient-based operators [4], [12], [19]–[21], depending on how these operators are being applied to the HDR image.

Global operators seek to preserve the global contrast in an image, and as such, the operator is applied to all pixels at once. Tumblin *et al.* [12] proposed a human visual system inspired operator for displaying high contrast images. By compressing HDR images, this TMO can preserve global brightness but it requires calibrated luminance values to produce good results. Another example of a global operator is the time-dependent visual adaptation proposed by Pattanaik *et al.* [14]. The operator accounts for appearance changes in animations or interactive real-time simulations and tries to match a user's visual responses to those the user would experience in the real-world scene. The tone mapping operator is based on global histogram adjustment which reproduces the high dynamic range of images, and then a bilateral weighting scheme is used to remove boundary and halo artifacts often introduced after segmentation. Drago *et al.* [13] proposed using adaptive logarithmic mapping for more robust rendering, while Durand and Dorsey [17] and Schlick [2] proposed histogram adjustment-based tone mapping, and quantization-based techniques respectively, for better global contrast enhancement in the output tone mapped images. Unfortunately, the output of the tone mapped image can lose its local details in the tone mapping process.

Local operators [13], [15]–[17], on the other hand, in contrast to global operators, apply the mapping operator to each pixel with respect to its neighboring pixels. Factoring the local statistics of the neighboring pixels around each pixel of interest guides the operator to reproduce both the local and global contrast of the original scene. Reinhard *et al.* [16] proposed a photographic tone reproduction using local statistics of neighborhood pixels which simulate the burning (adding of more light) and dodging (withholding light from a portion of the print) effect applied by photographers. First, a scaling correlative to camera exposure is applied, then automatic dodging-and-burning is applied to accomplish dynamic range compression when needed. Automatic dodging-and-burning acts as a local operator to improve results quality and avoid halo artifacts. Kim *et al.* [15] proposed a local operator based on the retinex model by introducing a k factor decision for enhancing the appearance and naturalness of HDR images and eliminate the burden of setting the parameters. k is a parameter for the retinex algorithm and is chosen based on the dynamic range in images. Ledda *et al.* [18] used a local model of eye adaptation to ensure proper local and global consistency.

Segmentation-based operators [1], [18], [20] divide the image into different uniform regions and apply a global

operator to each region before merging them. This helps in minimizing modifications to the original gamut. Yee and Pattanaik [20] proposed using image segmentation which involves dividing the image into several regions using a histogram-based technique and grouping neighboring pixels using a flood-fill approach. Krawcyzk *et al.* [21] also proposed using lightness perception in tone reproduction; and Lischinski *et al.* [1] proposed using interactive local adjustment of tonal values.

Frequency/gradient-based operators [4], [12], [14], [19]–[21], try to separate the lower and higher frequency components of an image with the goal of preserving the high-frequency components which the human visual system is more sensitive to and applying the operator to the lower frequencies. Tumblin and Turk [4] proposed a low curvature image simplifier (LCIS) filter inspired by anisotropic diffusion using partial differential equations. This filter preserves local contrast without losing the fine details and texture in the original image. Fattal *et al.* [22] manipulate the gradient field of the luminance image by suppressing the magnitudes of large gradients and keeping small gradients to preserve fine details. By solving a Poisson equation on the modified gradient field, a low dynamic range image can be produced.

The major drawback in the above-mentioned traditional tone mapping operators is the high parameter sensitivity, which prevents these methods from generalizing well on broader HDR imaging contents. The most recent methods have favored leveraging the tremendous success of deep neural networks, to develop parameter-free operators to address the challenges of parameter sensitivity and the generalization issues of traditional tone mapping operators [7], [11].

Particularly, the revolutionary concept of generative adversarial networks (GANs) [23] has led to immense success on image-to-image translation tasks such as image super-resolution [24], [25], image inpainting [26]–[29], image style-transfer [30]–[33], and pixel-level image segmentation [34], [35].

Inspired by these concepts, recent approaches have explored generative models for translating images from the high dynamic range to its counterpart low dynamic range or vice versa, by automatically learning a mapping function.

Particularly, Rana *et al.* [7] proposed DeepTMO network by exploring single-scale and multi-scale variants of both the generator and discriminator network of the pix2pixHD [32] architecture. Although DeepTMO is parameter-free and achieves better results than most traditional methods, the tone-mapped images often exhibit poor contrast as well as blurry patches for certain scenes. We believe these artifacts could be a result of a) the structure of the generator architecture model, and, b) the limited scale of data on which the model is trained. Lee *et al.* [36] also proposed using a deep recursive GAN model for performing the inverse mapping from LDR to HDR domain.

The proposed TMO-Net addresses the shortcomings of DeepTMO by introducing an attention-guided generator architecture for generating visually pleasing, tone-mapped images. Gradient profile loss is also introduced in the generator loss function to improve the visual perception of the tone mapped images. Section 3 provides a more detailed explanation of the model architecture and losses.

Performance benchmarking and comparative analysis on the large scale LVZ-HDR is conducted to properly evaluate the generalization abilities and visual output of the proposed TMO-Net in contrast with traditional and recent neural net TMOs.

III. PROPOSED APPROACH

The thrust of the approach is to view the HDR tone mapping as a low-light image enhancement problem. By considering a 16-bit HDR image as a poorly underexposed image with very low noise. Leveraging on the tremendous success of deep convolutional neural networks, we can design an adequate neural network with the proper objective function, and train it to learn a complex dynamic mathematical mapping function to produce well-exposed images with global and local contrast consistency, from underexposed image inputs. Such a trained model can serve as a dynamic tone mapping operator.

The proposed TMO-Net uses a generative adversarial network approach that tries to automatically learn a mapping function between images in two distinct domains.

The input is an 8-bit RGB image created by applying a gamma of 2.2 to the HDR image and normalizing the result to utilize the complete 0-255 range. Since HDR images typically have a small number of very bright pixels, this results in an image that appears very underexposed when displayed on a screen. TMO-Net attempts to map this image into a well-exposed image, which is also saved with 8 bits and standard sRGB gamma encoding.

The learned complex mapping function is parameter-free and can generalize well and consistently. TMO-Net also seeks to address the challenges of parameter-free and GAN-based model DeepTMO which exhibits blurry patches, global contrast inconsistency, halo artifacts, and color smearing, mainly due to the nature of its generator network architecture and objective function. Unlike DeepTMO which adopts the multiscale generator and multiscale discriminator network architecture of the Pix2PixHD model, TMO-Net only consists of a single generator network G, and a single discriminator network D pitched against each other. Fig. 4 shows the model architecture of the proposed TMO-Net. The discriminator D is based on PatchGAN, a Markovian discriminator model proposed by Isola *et al.* [37]. By introducing an attention module and a cascaded-in-cascaded block in our generator architecture, and incorporating a gradient profile loss into the objective function, TMO-Net is able to overcome the above-mentioned limitations of DeepTMO.

A. NETWORK ARCHITECTURE

TMO-Net is inspired by pix2pixHD [33] from where we derived the base generator architecture model; PatchGAN [37] whose discriminator network we adopted; FFA-Net [38] which inspired the attention guided module

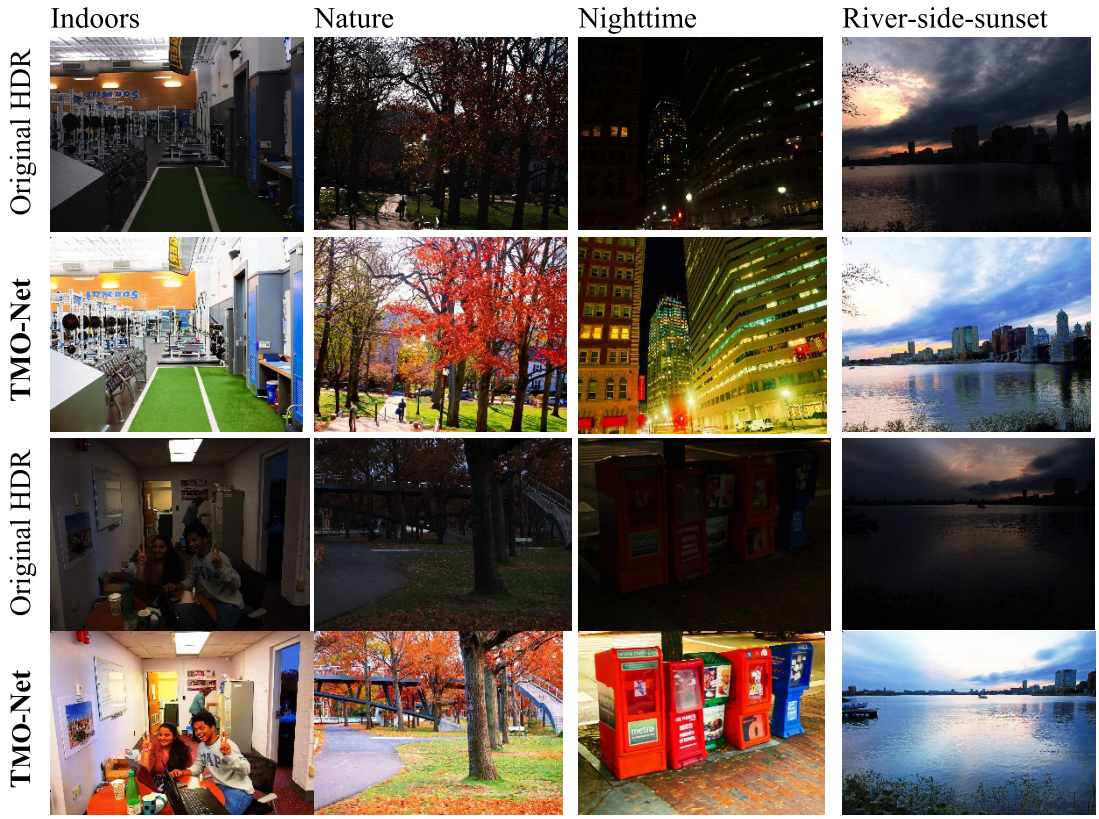


FIGURE 1. Sample tone mapping results of the proposed parameter-free TMO-Net. Images in the first and third rows are original sample HDR images from each major scene category of the proposed LVZ-HDR dataset. Images in the second and fourth rows represent the corresponding tone mapped images using the proposed TMO-Net method. The column titles represent scene categories in the LVZ-HDR Dataset.

in our generator architecture; and Cascaded Residual Dense Net [39], [40] where we derived the intuition for residual-in-residual dense block used in our generator architecture. As shown in Fig. 4, TMO-Net from end-to-end is a U-Net [41] like architecture with a set of downsampling blocks serving as the encoder network, followed by a series of cascaded dense blocks, a set of Upsampling blocks, and finally a fusion layer. The input size to the network is set to 256×256. Each cascaded dense block (CB) is constituted by a series of residual blocks with skip connections and concatenation layers. The output of the last cascaded dense block is simultaneously fed to the decoder network (Upsampling blocks) and an attention guided module, whose output is fused in the final layer to generate the tone mapped output image. The attention module is decomposed into channel-wise attention and pixel-wise attention blocks. Its purpose is to assist the generator in regulating light sources, to combat color smearing while ensuring global and local contrast and color consistency in the generated images. The guided attention module also helps eliminate the problem of under-exposure and over-exposure during image synthesis.

Downsampling and Upsampling blocks mirror each other but grow in inverse directions. The generator network consists of 3 down-sampling blocks and 3 up-sampling blocks, each consisting of a 2D convolutional layers in combination with an instance normalization layer, and a rectified linear unit (ReLU) activation layer.

Each cascaded dense block (CB) consists of 2 residual blocks and 2 output convolution layers. Residual blocks, in turn, are also made up of 2 convolutional layers with filter size of 3×3, and each followed by ReLU activation layer, with skip connections. The cascaded dense block aim at ensuring that the generator mapping function learns the finest textural and structural details in the image, and then transmits them to the up-sampling blocks, while also allowing for proper transmission of more global features learned in earlier layers. Bilinear sampling is used in the decoder network in combination with instance normalization and ReLU activation to ensure the elimination of potential checkerboard artifacts [31].

The discriminator adopted is a PatchGAN network, and has been largely adopted in recent literature as it is able to discriminate based on NxN patches rather than the entire image. Such a discriminator network architecture properly discerns the local textural and structural information in images and forces the discriminator to learn how to generate more globally and locally consistent images. It is made up of 5 convolutional layers which is used to transform a $6 \times 256 \times 256$ input

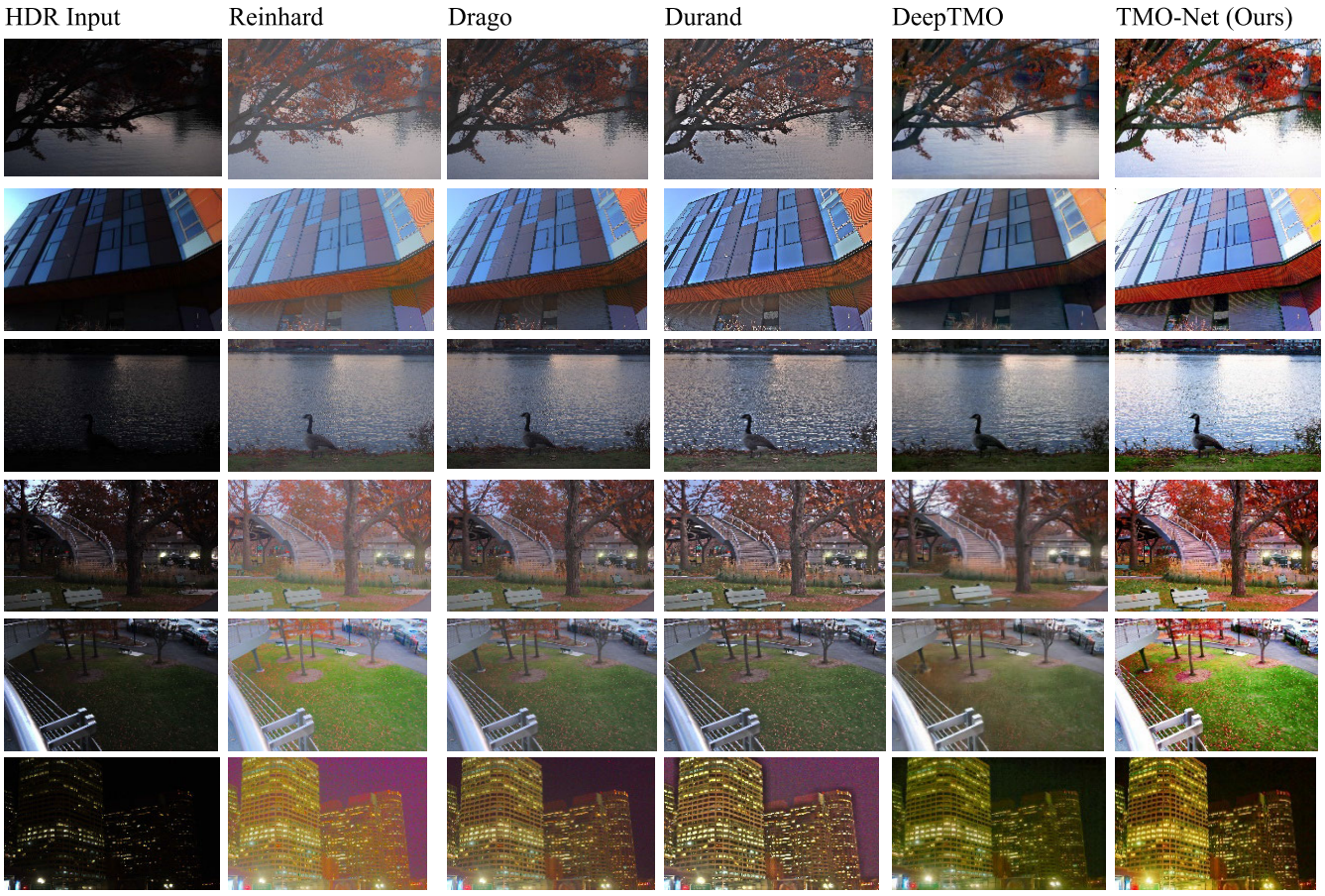


FIGURE 2. Sample tone mapped image outputs of some of the best traditional TMOs, along with parameter-free DeepTMO and the proposed parameter-free TMO-Net. The original images in the first column are images randomly selected from the proposed LVZ-HDR dataset and show great variability in scene types and visual content. It is evident that TMO-Net rendition yields much superior visual quality (aesthetic and naturalness) than all traditional methods, and even better than the deep learning and parameter-free counterpart (Deep-TMO). Input HDR images shown in the first column are obtained by applying a gamma map of 2.2 and scaling it to 8-bit range.

(including the real and generated image pair) to a $1 \times 35 \times 35$ output that represents the average validity responses of the discriminator.

B. TMO-NET GENERATOR OBJECTIVE FUNCTION

The objective function of the proposed TMO-Net is derived as the cumulation of four independent losses, namely: the adversarial loss, the perceptual loss, the feature matching loss and the gradient profile loss; each addressing specific issues and collectively guiding efficiency of the learned mapping function.

a) The Adversarial loss is formulated as a minmax game between the generator and discriminator pitched against each other, where the generator tries to minimize the loss while the discriminator tries to maximize it. The generator G tries to learn a forward mapping function $G : X \rightarrow Y$ to translate images in the high dynamic range domain X to the low dynamic range domain Y, and must successfully fool the discriminator which learns to differentiate between generated LDR images and original LDR images. L_{GAN} loss is defined by (1) [37]:

$$L_{GAN}(G, D) = \mathbb{E}_X[logD(X)] + \mathbb{E}_X[log(1 - D(G(\hat{X}))] \quad (1)$$

where X *and* $\hat{X}$ denote the input HDR image and the corresponding LDR ground truth, while $G(\hat{X})$ represent the output of the generator G.

b) The perceptual loss has been found to help guide the generator to properly evaluate the perceptual feature difference between the generated images and images in the target domain [42]. As such, incorporating the VGG loss into our generator's objective function will help the generator during training to evaluate the high-level perceptual and semantic difference between synthesized LDR images and corresponding ground truth LDR images, using the pixel-wise distance between both images. For optimal performance and effective impact on the training, pretrained features of the VGG19 network are reused. The perceptual loss is defined as [42]:

$$L_{VGG} = \sum_{i=1}^{N} \frac{1}{M_i} \left[\left\| F^{(i)}(Y) - F^{(i)}(G(X)) \right\|_1 \right] \quad (2)$$

$F^{(i)}$ denotes the i^{th} layer of VGG19 network, and M_i denotes the i^{th} element of that layer.

FIGURE 3. Visual comparison between the proposed parameter-free TMO-Net and DeepTMO counterpart. A closer look at both deep learning-based methods shows the strength of TMO-Net model in terms of rendering contrast, sharpness, colorfulness, aesthetic, and most importantly, naturalness in tone mapped images.

c) The Feature matching loss is essential for stabilizing training while directing the generator to produce both natural and reasonable statistical information at multiple scales. This loss is incorporated with respect to the discriminator which helps match intermediate feature maps between the real and synthesized images. The feature matching loss is defined as [43]:

$$L_{FM}(G, D) = \mathbb{E}_{X,Y} \sum_{i=1}^{T} \frac{1}{N_i} \left[\left\| D^{(i)}(X) - D^{(i)}(G(\hat{X})) \right\|_1 \right] \quad (3)$$

With T as the total number of layers, N_i the number of elements in each layer, and $D^{(i)}$ the i^{th}-layer feature extractor of the discriminator.

d) Gradient profile loss is further introduced to help measure the difference between the edge information between synthesized LDR image and ground truth LDR image [44], and is given as:

$$L_{GPL}(X, \hat{Y}) = \sum_{c} \left(\frac{1}{H} trace \left(\nabla G \left(\hat{Y} \right)_c \cdot \nabla X_c^{\tau} \right) + \frac{1}{W} trace(\nabla G(\hat{Y})_c^{\tau} \cdot \nabla X_c) \right) \quad (4)$$

where $(\cdot)^{\tau}$ represents transpose, H and W are the height and width of the image, X is the target LDR image and $\hat{Y}$ is the synthesized LDR image.

The overall objective function of TMO-Net is therefore defined by the following equation:

$$L_{TMO\text{-}Net} = L_{GAN} + \lambda_1 L_{FM} + \lambda_2 L_{VGG} + \lambda_3 L_{GPL} \quad (5)$$

where the distinctive weights $\lambda_1 = 10.0, \lambda_2 = 1.0$, and $\lambda_3 = 0.8$ are used to control the relative importance or contribution of individual losses to the overall objective. Adam optimizer with an initial learning rate of 0.0002 is used to optimize both the generator and discriminator.

The downside of deep neural nets approach to solving a problem is that they are data-hungry methods and require thousands of image samples and equivalent ground truth. This means that to properly train a GAN-based model to perform automatic tone mapping operation on input images as we are trying to accomplish in this paper, there is a need to gather a dataset with thousands of HDR images, and the corresponding ground truth tone mapped images. Such data normalized paired HDR/LDR dataset is currently unavailable. To mitigate this challenge, the proposed TMO-Net is intuitively trained end-to-end on low-light images dataset. The insight here is that such low-light datasets contain under-exposed images whose characteristics model the distribution of HDR images considered under-exposed when viewed on LDR devices. The advantage here is that there are several such datasets that contain thousands of images with corresponding well-exposed pairs, and a huge variety of exposure, scenes

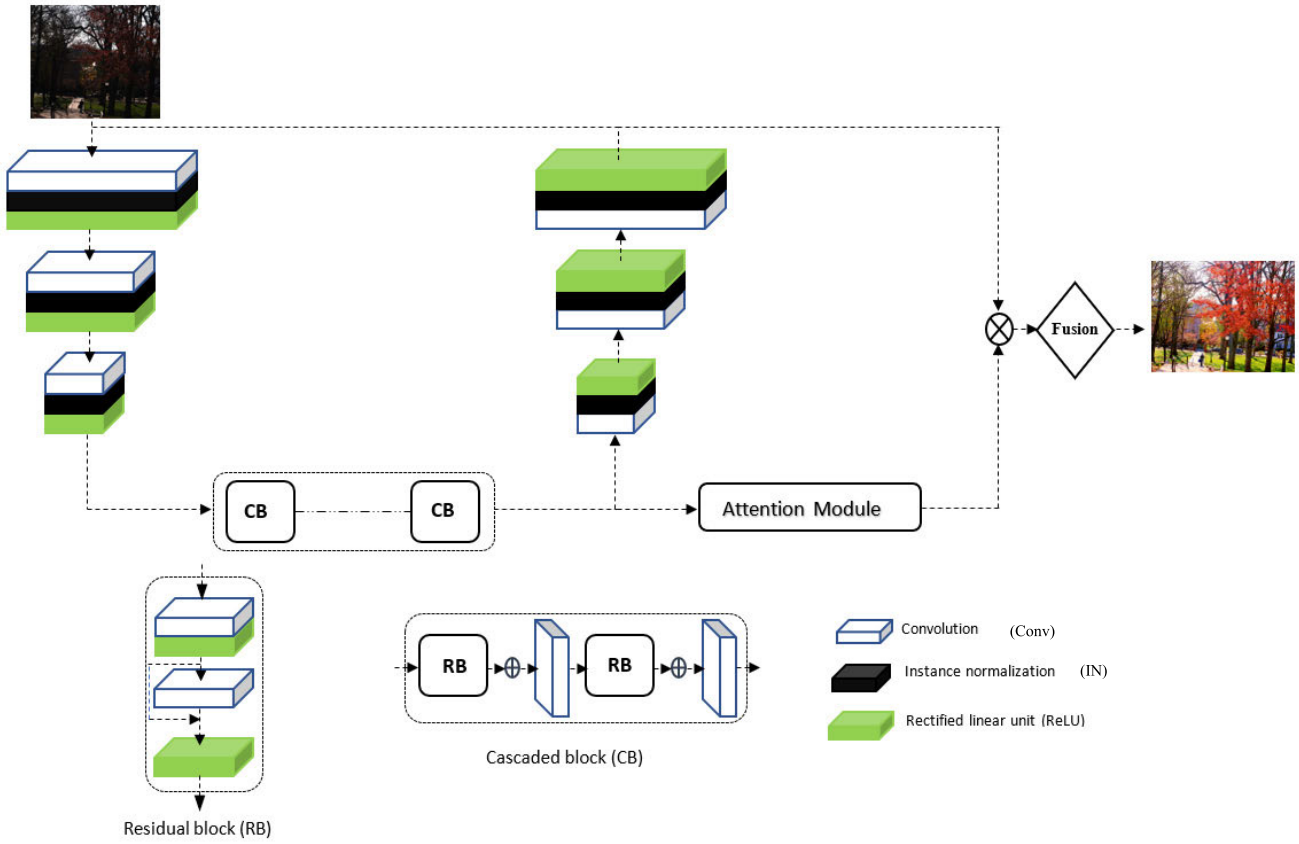


FIGURE 4. TMO-Net generator architecture. Conv denotes a 2D convolution, IN denotes Instance normalization and ReLU denotes a Rectified Linear Unit activation function. The Attention block is same as the Upsampling block with the tanh activation replaced with a sigmoid. The fusion operation sums the products of the multiplication operation to create an enhanced image.

and illuminations. Essentially, we combine the Low-Light dataset[1] which contains up to 2274 lowlight images and corresponding 2274 enhanced image pairs; 789 images from the Dark Face dataset [45]; 485 images from the LOL dataset; and 1000 images from [46], respectively. This makes our training set very diverse.

The formulation of our final training objective resulted from extensive computational experimentations and hyperparameter search. We tried several combinations of components in our loss function over numerous alterations of the proposed network architecture and found that the final objective function presented yielded the most desirable outputs.

Because of the apparent complexity of the overall training objective function, an ablation study has been conducted to validate the importance of individual components. Fig. 5 shows sample image input and corresponding output from the same network trained with variations of the proposed objective function. Local patches are extracted from the images and zoomed in for further comparative analysis. We zoom in on three sample patches at corresponding locations on each of the output images which are color coded. It can be seen on patches (d), (e),(f), that removing the SPL loss component results in output lacking sharpness and global contrast consistency. Removing the VGG loss further exacerbates this issue as blurry patches become more apparent and unnatural. These artifacts can be visualized in image patches (g), (h), (i). The importance of the feature matching loss can be better visualized by comparing image patches (j), (k), and (l), to image patches (a), (b) and (c). Feature matching loss helps preserve intrinsic and finer details from the original scene during reconstruction, hence yielding a much better toning. Comparing patches (a), (b), and (c) from the output image of TMO-Net to corresponding patches from output of the same network with varied loss function components, show relevance of each component in the final training objective.

[1]https://sites.google.com/site/vonikakis/datasets

IV. LVZ-HDR BENCHMARK DATASET AND EVALUATION METRICS

A. LVZ-HDR BENCHMARK DATASET

A new dataset called the LVZ-HDR dataset was created. Each HDR image in the proposed dataset is formed by fusing 5 LDR images of the same scene taken at multiple exposure brackets, −2EV, −1EV, 0EV, +1EV, +2EV, using Photoshop HDR tools. The software also offers handy functions to help eliminate imaging fusion problems such as

FIGURE 5. Ablation study showing the relevance of individual components of the overall objective function. Images patched (a)-(l) are extracted at corresponding locations from each of the output with varied combination of loss function. Image patches (a), (b),and (c) are extracted from the output with all four loss components. Similarly, patches (d), (e), (f) are taken from output of the same network trained without the spl loss in the objective. Patches are color coded to facilitate visualization and comparision.

misalignment, noise and ghosting. The images were taken using an Olympus E-P1 camera. The LVZ-HDR benchmark dataset consists of 456 High Dynamic Range images taken from different states, cities, time of the day and most importantly covering a wide range of sceneries. Images in the LVZ dataset are grouped into four broad scene categories: Indoors (71 images), Nature (173 images), Nighttime (80 images), and River-side-Sunset (133) images. The dataset is available on Kaggle for download here,[2] with free license for educational and research purposes shall be provided. Original multi-exposure LDR images used to create each HDR image can also be made available upon request. Fig. 6 shows diversified sample HDR images from the proposed large scale LVZ-HDR dataset.

The proposed dataset strives to address the lack of large scale HDR test set for properly evaluating the merits and performance of tone mapping operators. Published research works on tone mapping have so far limited their evaluation to a small number of images, with limited sceneries and visual content. Results presented in conventional tone mapping publications are often selective of best output images or resulting from excessive hyperparameter tuning. Hence, it is difficult to ascertain the generalizability of an operator across a wider range of sceneries. We benchmark the performance of 19 tone mapping operators on the LVZ-HDR dataset, including 17 traditional state-of-the-art TMOs, and 2 deep learning-based and parameter-free TMOs, among which the proposed TMO-Net, and DeepTMO counterpart.

B. EVALUATION METRICS

Two widely accepted evaluation metrics, TMQI [47] and FSITM [48], have been proposed for quantitatively evaluating the performance of tone mapping operators. These metrics are used to benchmark the performance of 19 tone mapping operators including the proposed TMO-Net.

The TMQI algorithm uses a combination of multiscale structural similarity index, and intensity statistics to measure three major scores, including: a) the structural fidelity score (S-score), b) the statistical naturalness score (N-score) and finally c) the TMQI score (Q-Score).

The structural fidelity measure between a local HDR and LDR patches x and y respectively is defined as:

$$S_{local}(x,y) = \frac{2\sigma_x'\sigma_y' + C_1}{\sigma_x'^2 + \sigma_y'^2 + C_1} \cdot \frac{\sigma_{xy} + C_2}{\sigma_x\sigma_y + C_2} \tag{6}$$

where $\boldsymbol{\sigma_x}$ and $\boldsymbol{\sigma_y}$ represent local standard deviations, $\boldsymbol{\sigma_{xy}}$ represents the local cross-correlation between any corresponding x and y patches, $\boldsymbol{C_1}$ ***and*** $\boldsymbol{C_2}$ represent stabilizing constants. Combining the structural fidelity of local patches at different scales yields the overall structural fidelity as follows:

$$S = \prod_{l}^{L} S_l^{\beta_l} \begin{cases} L = number\ of\ scales \\ \beta_l = weight\ assigned\ to\ a\ scale \\ l = l^{th} scale \end{cases} \tag{7}$$

[2]https://www.kaggle.com/landrykezebou/lvzhdr-tone-mapping-benchmark-dataset-tmonet

Similarly, the statistical naturalness score is calculated as:

$$N = \frac{1}{K} P_m P_d \begin{cases} P_m = gaussian\ probability\ density\ function \\ P_d = Beta\ probability\ density\ functions \\ K = \max(P_m, P_d) normalization\ factor \end{cases} \tag{8}$$

Finally, the TMQI Q score is given as the linear combination of the S and N scores as follows:

$$\boldsymbol{Q = aS^{\alpha} + (1-a)N^{\beta}} \times \begin{cases} \boldsymbol{0 \le a \le 1} \\ \boldsymbol{\alpha, \beta = S\ and\ N\ reps.\ sensitivities} \\ \boldsymbol{S, N = S\ and\ N\ scores\ reps.} \end{cases} \tag{9}$$

Note that all S, N and Q scores are all in the range of 0 to 1, and the higher the score, the better the visual image quality. Further details and proof of the mathematical equations in (6), (7), (8) and (9) can be found in the original TMQI paper [17]. The MATLAB code for TMQI can be found here[3]

The Feature Similarity Index for Tone Mapped images (FSITM) metric on the other hand, leverages on the local phase information embedded in images to compute the Q-score by comparing the locally weighted mean phase angle maps between the original HDR and the output LDR tone mapped images [48]. The MATLAB code available for FSITM can be found here.[4]

V. EXPERIMENTAL RESULTS AND BENCHMARKING

In this section, the performance of the selected state-of-the-art tone mapping operators on the proposed LVZ-HDR dataset is discussed and their visual output compared. Performance evaluation on individual scene categories is also implemented and the TMOs are ranked according to the overall benchmark results.

The performance of 19 TMOs chosen across all categories of tone mapping discussed in Section 2 including global, local, segmentation-based, gradient/frequency-based, and deep learning-based operators are evaluated.

The benchmarked TMOs include: Drago TMO [13], Durand TMO [17], Exponential TMO,[5] Kim Kautz Consistent TMO [15], Krawczyk TMO [21], Lischinski TMO [1], Logarithmic TMO, Mertens TMO [9], Normalize TMO[7], Pattanaik TMO [14], Raman TMO [6], Reinhard Devlin TMO [3], Reinhard TMO [16], Schlick TMO [2], Tumblin TMO [4], Ward Global TMO, and WardHistAdj TMO, DeepTMO, and TMO-Net. The MATLAB code for all the traditional TMOs is available from the HDR toolbox GitHub repository.[6]

Fig. 7 shows the output of each selected TMOs on a few test images, from different categories in the LVZ-HDR benchmark dataset. The results for all 456 images in the LVZ-HDR

[3]https://ece.uwaterloo.ca/~z70wang/research/tmqi/

[4]https://www.mathworks.com/matlabcentral/fileexchange/5981

[5]Francesco Banterle, "Hdr toolbox: https://github.com/banterle/hdrtoolbox/blob/master/sourcecode," GitHub, 2010-2015

[6]https://github.com/banterle/HDR_Toolbox

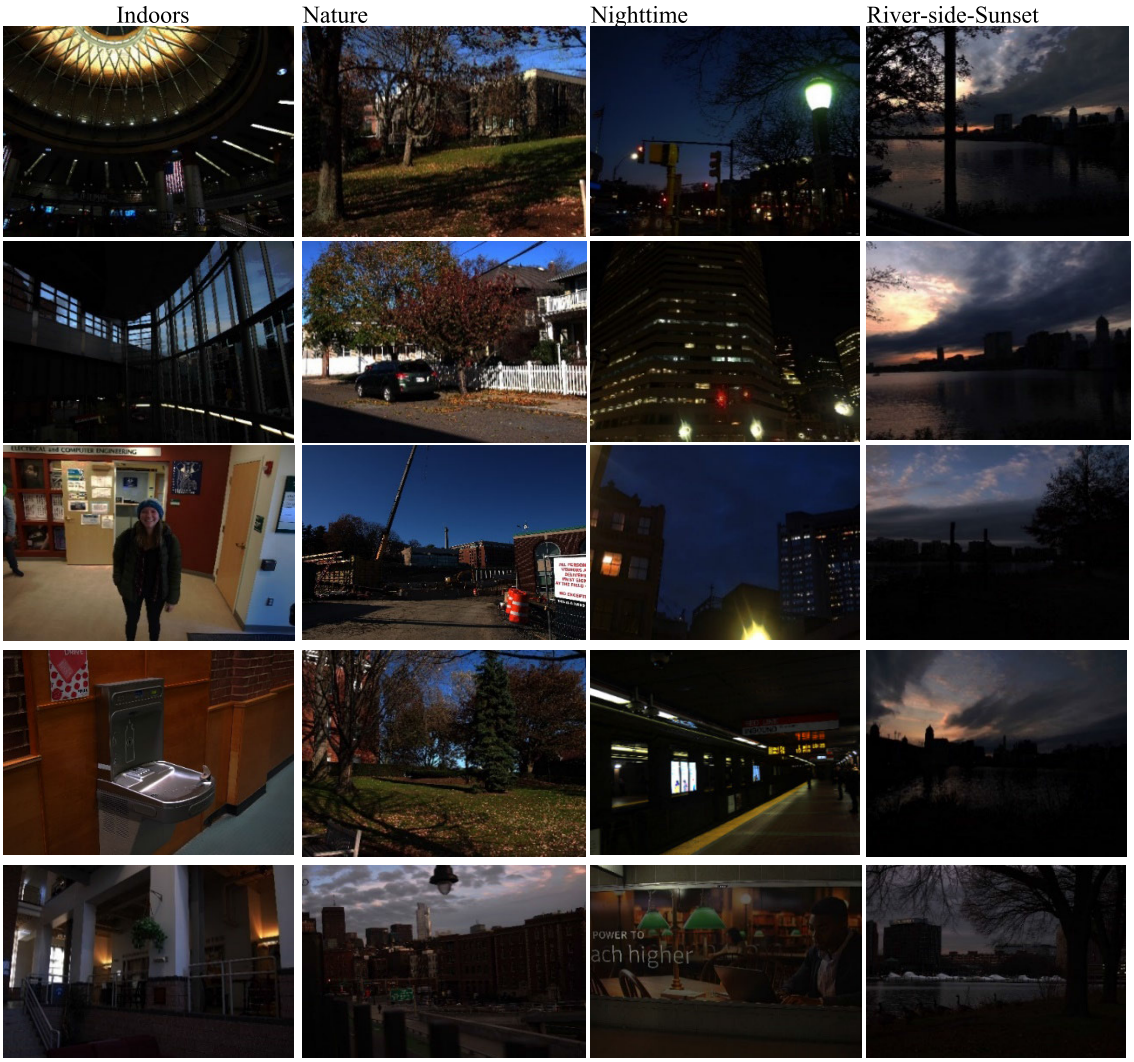


FIGURE 6. Random sample HDR images drawn from the proposed large scale LVZ-HDR benchmark dataset. Scene contents such as people, roofs, grass fields, buildings, sky, river shores, subways, stairwells and sunset can be apparent in the sample HDR images shown.

benchmark dataset for all 19 TMOs are available on Kaggle for visualization and download[7].

Fig. 7 shows that the proposed TMO-Net qualitatively outperforms all other methods. TMO-Net generates more visually appealing and more natural-looking images with high definition color rendering, high contrast, and sharpness.

TMO-Net not only eliminates the parameter sensitivity of traditional methods, but also addresses key issues such as color smearing, low contrast, blurriness, halo artifacts, color misrepresentation and lack of sharpness, some of which are apparent even in DeepTMO, the parameter-free and deep learning counterpart. These qualitative results are further substantiated quantitatively using dedicated evaluation metrics such as TMQI, FSTIM, NIQE and UIQM [49], which are presented in Table 1.

[7]https://www.kaggle.com/landrykezebou/lvzhdr-tone-mapping-benchmark-dataset-tmonet

A. BENCHMARKING WITH TMQI

As discussed earlier, the TMQI measures 3 useful scores: a) the S-score which measure the structural fidelity index, b) the N-score which measure the natural realness of the image, and c) the Q-score which is the linear combination of the S and N scores and represents the TMQI index. TMQI is also the most used evaluation metric for tone mapping algorithms. The performance of the operators are ranked based on the average TMQI score on the LVZ-HDR dataset.

The benchmark plot in Fig. 8, shows that the proposed TMO-Net ranks and the best performing tone mapping operator based on the TMQI average scores. The ranking indicates that the top 2 algorithms are parameter-free operators TMO-Net and DeepTMO, which confirms just how well these 2 deep learning-based operators can automatically adapt to varying scenes and a wider range of HDR content. It is equally important to observe the difference between the

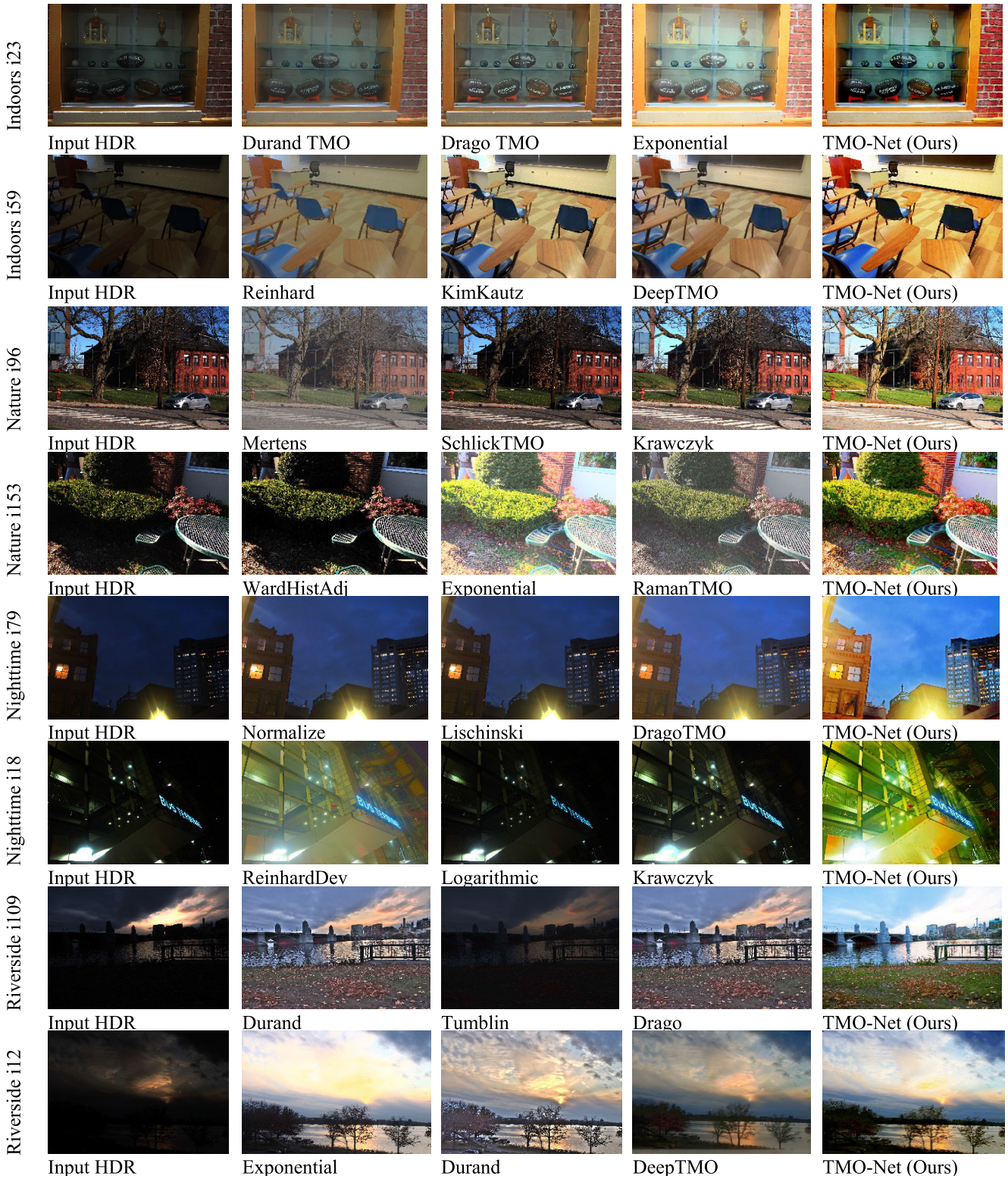


FIGURE 7. Comparison between the proposed TMO-Net and existing state-of-the-art TMOs. Images in the first column are original HDR images from different scene categories of the LVZ-HDR benchmark dataset. Images in the last columns are corresponding output from the proposed parameter-free tone mapping algorithm TMO-Net. Images in intermediary columns are output from other TMOs.

average naturalness scores and the corresponding TMQI-Q scores. The corresponding Q score and Naturalness scores for the proposed TMO-Net are very high and very close to each other in terms of values, which indicates that TMO-Net is able to maintain high tone mapping consistency across all scenes.

TABLE 1. Quantitative results of existing TMOs in comparison with the proposed TMO-Net, using state of the art evaluation metrics. All reported scores are the average scores on all 456 HDR images and across all scene categories in the proposed LVZ-HDR benchmark dataset. TMQI-Q represents the Q-score or overall TMQI score while TMQI-N represents the corresponding naturalness score of the TMQI metric. UIQM-Colorfulness represents the colorfulness component of the UIQM[8] non-reference image quality metric; NIQE[9] is the Naturalness Image Quality Evaluator. Note that the smaller the NIQE, the better, but the higher the other scores, the better.

Tone Mapping Operator	TMQI-Q score (↑)	TMQI-N score (↑)	FSTIM (↑)	NIQE (↓)	UIQM Overall (↑)	UIQM-colorfulness (↑)	UIQM-sharpness (↑)	UIQM-Contrast (↑)
TMO-Net (ours)	**0.875**	**0.84**	0.84	**3.23**	**1.48**	**7.83**	**7.09**	0.97
DeepTMO	0.833	0.66	0.57	3.61	1.25	5.01	3.68	1.03
Durand	0.820	0.53	0.91	3.72	1.13	5.75	3.60	0.89
KimKautz	0.820	0.30	0.88	4.34	1.12	5.10	5.10	3.91
Drago	0.787	0.21	0.89	4.64	0.95	4.99	2.96	0.75
ReinhardDevlin	0.783	0.24	0.86	4.69	0.79	4.17	2.30	0.63
Krawczyk	0.782	0.24	0.88	4.22	1.07	4.84	3.72	0.82
Lischinski	0.779	0.13	0.92	4.43	1.12	4.18	3.85	0.86
WardHistAdj	0.778	0.10	0.91	3.99	1.32	4.09	4.95	1.00
Raman	0.777	0.23	0.90	4.69	0.77	2.42	2.10	0.64
Exponential	0.769	0.13	0.83	3.80	0.97	6.82	2.86	0.77
Mertens	0.767	0.18	0.89	4.68	0.72	2.10	1.80	0.618
Reinhard	0.762	0.10	0.92	4.72	1.06	4.17	3.55	0.83
Normalize	0.758	0.13	0.93	4.86	1.14	3.75	4.17	0.87
Schlick	0.682	0.02	0.97	5.50	1.17	2.67	4.63	0.88
Logarithmic	0.677	0.008	0.92	5.67	1.10	2.18	4.10	0.84
Tumblin	0.641	0.004	0.89	6.45	0.90	1.71	3.06	0.72
WardGlobal	0.607	0.004	0.90	6.97	1.16	1.26	4.59	0.88
Patternaik	0.353	0.003	0.75	14.08	0.77	0.03	5.65	1.44

Note that for adequate comparison, DeepTMO model was also retrained on the same training set as TMO-Net, and for the same number of epochs.

This also demonstrates that TMO-Net generates not only the most visually appealing images but also the most natural-looking image rendering with very little to no artifacts.

Fig. 9 also shows that contrary to traditional tone mapping operators which perform better on some scene categories than others, TMO-Net maintains a consistent performance across the different categories of the benchmark dataset, which means that the proposed model generalizes well to a wider range of HDR visual content and scenes.

B. TMO-NET VS DeepTMO

In this section, the performance of TMO-Net is compared to another deep-learning-based, parameter-free tone mapping operator, DeepTMO.

Fig. 10 compares sample images output from the TMO-Net and DeepTMO operators. Zoom-in extracts show several artifacts presents in the DeepTMO rendering. TMO-Net renders images with higher definition, consistent global and local color and contrast, and better sharpness. DeepTMO, on the other hand, exhibits local blurriness, local color smearing, and lacks sharpness.

C. USER SURVEY

To further assess and validate the qualitative performance of the proposed TMO-Net in contrast with existing state-of-the-art TMOs, an online user survey was carried out. To make the survey most efficient and less crowded, 12 HDR images presented in this paper were selected, 3 from each scene category. The corresponding tone mapping outputs for the top 7 operators based on the TMQI ranking in Fig. 8 were chosen for the survey. As such, the survey is presented to the user in 12 groups, each group showing a separate input HDR image, and the corresponding tone mapping output for each of the 7 TMOs chosen. We asked the users to select for each group, 2 images based on what they feel is: i) the most visually appealing image, ii) the most enhanced output image

[8]UIQM code available for download on request at: https://www.karenpanetta.com/download

[9]NIQE code available at: https://github.com/dsoellinger/blind_image_quality_toolbox

Rank	Operators	TMQI_Q
1	**TMO-Net (ours)**	**0.875**
2	DeepTMO	0.833
3	Durand	0.82
4	KimKautz	0.82
5	Drago	0.787
6	ReinhardDevlin	0.783
7	Krawczyk	0.782
8	Lischinski	0.779
9	WardHistAdj	0.778
10	Raman	0.777
11	Exponential	0.769
12	Merteens	0.767
13	Reinhard	0.762
14	Normalize	0.758
15	Schilick	0.682
16	Logarithmic	0.677
17	Tumblin	0.641
18	WardGlobal	0.607
19	Patternaik	0.353

FIGURE 8. Performance benchmarking on the LVZ-HDR dataset using TMQI evaluation metrics. The table on the right ranks the selected tone mapping operators from the best performing to the least. Scores used are the overall TMQI score averages for each TMO on all 456 HDR images of the LVZ-HDR benchmark dataset. The blue curve represents the TMQI-Q or overall TMQI average score, while TMQI-N measures the corresponding average naturalness scores.

FIGURE 9. Performance comparison of selected operators on scene categories of the LVZ-HDR benchmark dataset. plot (a) shows the average naturalness score while plot (b) shows the average TMQI-Q score per scene category of the dataset. It can be observed that TMO-Net exhibits better performance index across all scene categories.

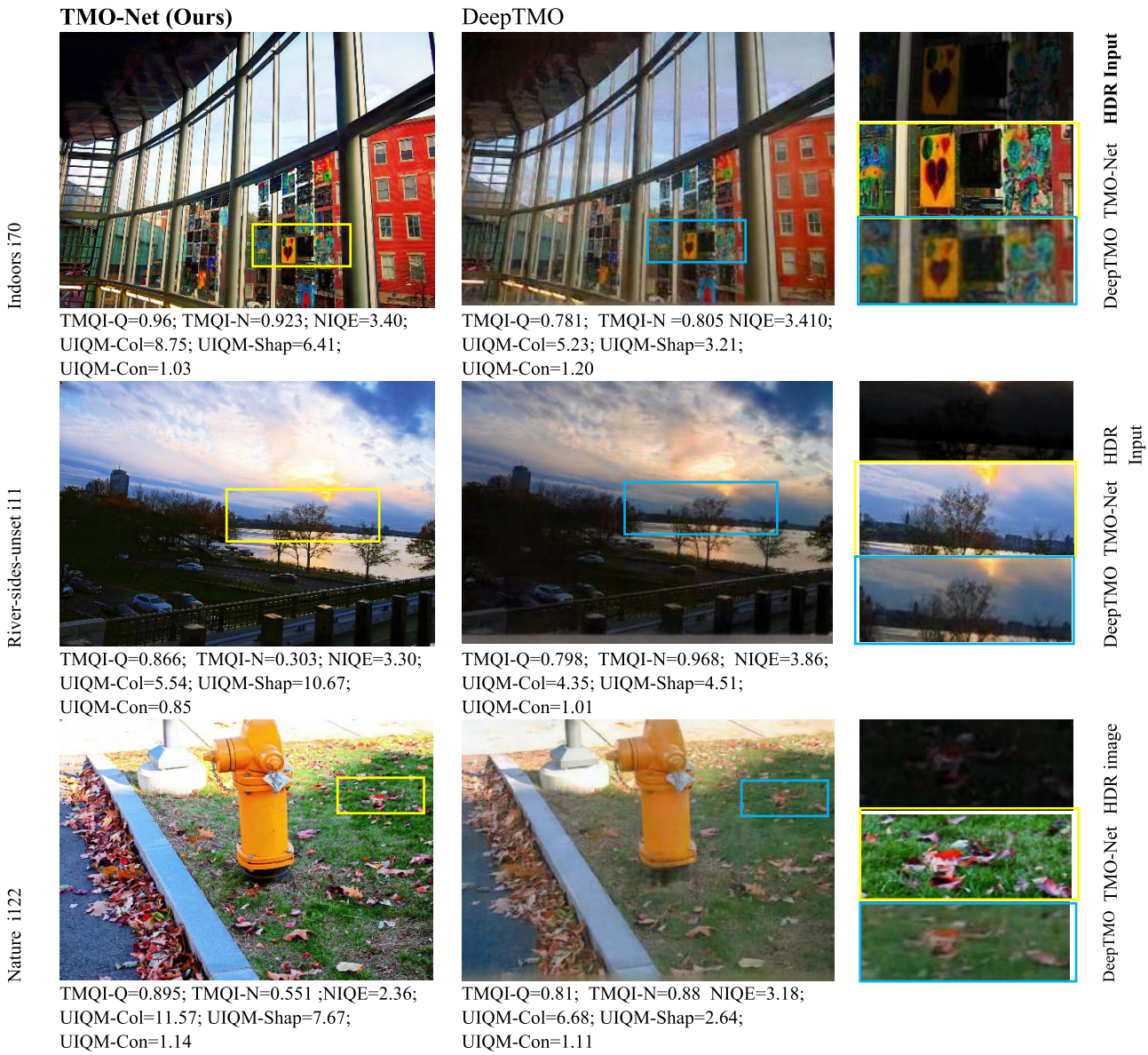


FIGURE 10. A closer look and low-level comparison between TMO-Net and DeepTMO. Images in the first column are output results using the proposed TMO-Net while images in the second column are outputs of the DeepTMO. The third column highlights zoom- extracts of corresponding images. TMQI-Q is TMQI score; TMQI-N is the TMWI naturalness score; UIQM-Col, UIQM-Sharp and UIQM-Con represent the colorfulness, sharpness and contrast scores of the UIQM metric.

based on input HDR image, and iii) the most natural looking image among the 7 TMO output presented.

A total of 91 people took the survey,[10] the vast majority of whom were college students with no image processing background, and the rest being image processing experts or research students.

Survey results presented in Table 2 shows that for every one of the 12 test images presented to the 91 users, at least 61 users on average selected TMO-Net output image as the best tone mapping operator based on the selection criteria described above. This substantiates the claim that TMO-Net qualitatively outperforms existing TMOs. The survey equally supports the benchmarking results presented in Fig. 8 based on the quantitative performance of various TMOs. A tentative ranking of TMOs based on average survey index score also reveals TMO-Net as the best performing operator. Note that Durand and DeepTMO operators rank as second and third respectively on the survey ranking as opposed to third and second on the TMQI ranking. However, it is safe to say that the survey ranking is still within margin of error of the TMQI benchmarking presented in Fig. 8, given that average TMQI scores of DeepTMO, Durand and KimKautz are very close to each other. As such, outputs of these 3 operators are very similar and equally likely (33% chance) to be chosen

[10]https://www.wenjuan.com/s/VnmMja/

TABLE 2. Survey results. For each input HDR image (3/category X 4 categories = 12 images), users are presented a group of 8 images at once, including the original HDR image and the tone mapping output of 7 TMOs. Users are then asked to choose 2 images based on 3 criteria: i) the most visually appealing, ii) the most enhanced image output based on input HDR image, and iii) the most natural looking output image. There was a total of 91 users, each voting for 2 images per group (12 groups of 8 images available). Therefore, a total of 2,184 (91 × 2 × 12 = 2,184) votes were cast. Since an image can only be chosen once in a group of 8, each image can be voted a maximum total of (91 × 1 × 12)= 1092.

Image Category	HDR image	Durand	**TMO-Net**	DeepTMO	Exponential	Drago	KimKautz	Krawczyk
Indoors	indoors_i23	30	**74**	15	40	7	13	3
	indoors_i34	39	**53**	32	29	17	8	4
	indoors_i59	12	**54**	39	28	2	39	8
Nature	nature_i68	19	**69**	35	28	15	5	11
	nature_i153	48	**57**	7	26	5	2	37
	nature_i174	48	**48**	16	15	29	10	16
Nighttime	nighttime_i18	35	**72**	26	16	11	19	3
	nightime_i24	32	**62**	12	13	12	28	23
	nightime_i65	45	**73**	19	15	11	11	8
River-side-sunset	riverside_i12	26	**52**	25	30	13	21	15
	riverside_i62	24	**62**	24	15	4	27	26
	riverside_i85	18	**66**	39	22	10	17	10
Total		378	**742**	288	277	136	200	164
Ranking		2	**1**	3	4	7	5	6

as second best image for each group after choosing TMO-Net as the best output.

VI. CONCLUSION

In this paper, a large scale HDR dataset (LVZ-HDR) is introduced to serve as a standard benchmark dataset for evaluating the performance of both current and future tone mapping operators. Furthermore, an efficient deep learning-based and parameter-free tone mapping operator, TMO-Net is proposed to effectively synthesize color-rich and visually pleasing low dynamic range images from the input high dynamic range images. TMO-Net can generalize well and produce consistent results on a wider range of HDR visual content. Qualitative results broached in section 5 reveals that the proposed TMO-Net renders images that are more visually appealing, more natural-looking, with consistent global and local colorfulness and contrast, and higher sharpness. A closer comparison with DeepTMO shows that TMO-Net outperforms DeepTMO, both qualitatively and quantitatively.

Experimental results also show that the proposed TMO-Net quantitatively outperforms all current state-of-the-art tone mapping operators across all scene categories. TMO-Net addresses the shortcomings of existing tone mapping operators such as blurry patches, color smearing, halo artifact, inadequate sharpness and improper color rendering. To further substantiate and validate such performance ranking, the performance of the selected algorithms are re-evaluated using other widely used blind evaluation metrics such as NIQE, and the colorfulness, and sharpness components of the UIQM.

VII. LIMITATIONS

Although TMO-Net addresses most of the challenges and limitations of DeepTMO and other traditional tone mapping operators, there are several areas for improvement. The most significant is that TMO-Net does not natively use HDR images, and therefore some image quality is lost due to the 8-bit quantization of the input, particularly in darker areas.

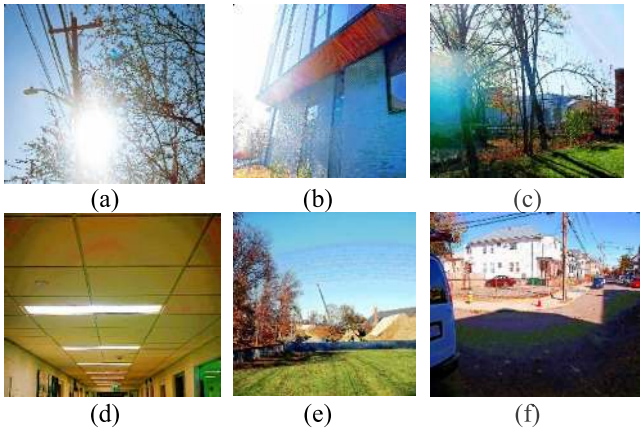


FIGURE 11. Images showing failure cases of the TMO-Net. Images (a),(b) and (c) show rare cases where the TMO-Net's attention module over-enhances specific regions of the image and result in such colors dominating the image. Images in (d), (e), and (f) show sample images where artifacts such as misalignment and ghost present in the original hdr images are made worst in the output tone mapped image.

While the generator is capable of partially compensating for this, better results could be achieved by using the full bit depth of the HDR input.

The attention module in the generator architecture model sometimes over-enhances specific regions in the images and this could result in some colors dominating the rest of the image. Furthermore, in few cases where the original HDR images contain artifacts such as misalignment, and ghosting, which were not properly addressed during scene exposure fusion, TMO-Net seems to make such artifacts worse in the output tone mapped image. Some of these failure modes are shown in Fig. 11.

One way of mitigating these issues in to adequately penalize the attention guided module in the generator architecture. It is equally important to ensure proper alignment and deghosting in the output hdr during fusion exposure.

Another important limitation in the benchmarking is that we use the TMQI scores. TMQI metric does not follow the

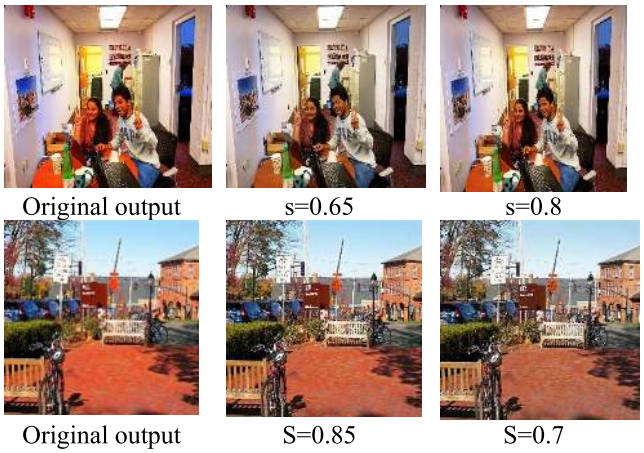


FIGURE 12. ***TMO-Net results before and after color correction*** **[2]. Image in the first column are the original tone mapped output from the TMO-Net model, while images in second and third columns represent corresponding output after applying color correction algorithm. The parameter s controls the color saturation.**

human visual system and as such exhibit some limitations. TMQI sometimes assigns lower score to images of higher visual quality, which suggest there is a strong need to develop a more robust and human visual system inspired tone mapping evaluation metric. Because of such limitations of the TMQI and FSTIM metrics, we also validated the merits of TMO-Net by using other evaluation metrics including NIQE and the human visual system inspired UIQM. These limitations will constitute the basis for our future work and new research directions.

Color over-emphasis/saturation is also one of the limitations observed with the proposed TMO-Net model. Although TMO-Net was trained to generate high quality tone mapped images which are perceptually more pleasing and more natural looking to the finest details, it can be seen in some of the output images that some colors appear saturated. One easy way to fix this would be to apply a color correction algorithm such as the one proposed by Mantiuk *et al.* [50] and suggested by [7], or even a much older color correction algorithm proposed by Schlick [2], to the output images. The color correction proposed by Mantiuk defined as $C_{out} = (\left(\frac{C_{in}}{L_{in}} - 1\right).s + 1)L_{out}$ where s controls color saturation.

Fig. 12 shows sample image output before and after color correction at multiple color saturation control values.

However, for fair and unbiased comparative analysis, we only used the raw output of the proposed model without further color correction. One room for improving the TMO-Net will be to incorporate this color adjustment to the objective function, but this could result in new artifacts in the output images.

APPENDIX

A few hdr images have become very popular in the decades-long history of tone mapping research as they are commonly used for evaluation in traditional tone mapping publications. Although we propose a new large scale hdr tone mapping benchmark dataset that we hope will become the new landmark for properly evaluating the performance of tone mapping operators, it is still important to also showcase the performance of our proposed parameter-free deep learning-based tone mapping operator- TMO-Net, on some of these popular images.

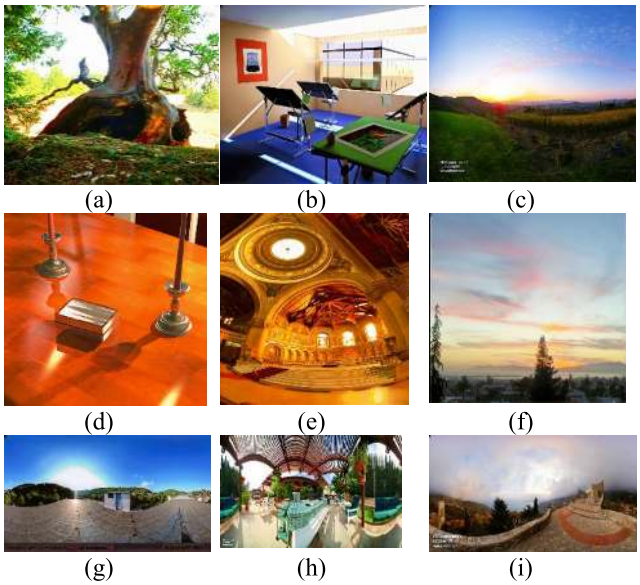


FIGURE 13. TMO-Net result on some of the most popular hdr images in hdr tone mapping research. (a) Tree, (b) rend05, (c) SpheronNapaValley, (d) rend06, (e) memorial, (f) vinesunset, (g) Spheron3, (h) SpheronPriceWestern, (i) SpheronNice. Original hdr images can be found at: http://www.anyhere.com/gward/hdrenc/pages/originals.html Image courtesy: ILM(a); Charles Ehrlich (b,d); Paul Debevec (e); Spheron (c,gh,i).

Fig. 13 shows tone mapped results of TMO-Net on some of these most popular hdr images.

We can observe from the results shown in Fig. 13 that TMO-Net also achieves state of the art performance on these well know evaluation images.

ACKNOWLEDGMENT

The authors would like to thank the valuable contributions of Dr. Steven Bell in achieving the overall goals of this research endeavor. They also wish to sincerely acknowledge and thank the editors and reviewers for their insightful comments and constructive criticism which have helped in significantly improving the overall quality of this manuscript.

REFERENCES

[1] D. Lischinski, Z. Farbman, M. Uyttendaele, and R. Szeliski, "Interactive local adjustment of tonal values," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 646–653, Jul. 2006, doi: 10.1145/1141911.1141936.

[2] C. Schlick, "Quantization techniques for visualization of high dynamic range pictures," in *Photorealistic Rendering Techniques*. Berlin, Germany: Springer, 1995, pp. 7–20.

[3] E. Reinhard and K. Devlin, "Dynamic range reduction inspired by photoreceptor physiology," *IEEE Trans. Vis. Comput. Graphics*, vol. 11, no. 1, pp. 13–24, Jan./Feb. 2005. Accessed: Apr. 10, 2020.

[4] J. Tumblin and G. Turk, "LCIS: A boundary hierarchy for detail-preserving contrast reduction," in *Proc. 26th Annu. Conf. Comput. Graph. Interact. Techn. (SIGGRAPH)*, 1999, pp. 83–90, doi: 10.1145/311535.311544.

[5] G. W. Larson, H. Rushmeier, and C. Piatko, "A visibility matching tone reproduction operator for high dynamic range scenes," *IEEE Trans. Vis. Comput. Graphics*, vol. 3, no. 4, pp. 291–306, Oct. 1997, doi: 10.1109/2945.646233.

[6] S. Raman and S. Chaudhuri, "Bilateral filter based compositing for variable exposure photography," in *Proc. Eurographics*, 2009, pp. 1–4. Accessed: Apr. 10, 2020.

[7] A. Rana, P. Singh, G. Valenzise, F. Dufaux, N. Komodakis, and A. Smolic, "Deep tone mapping operator for high dynamic range images," *IEEE Trans. Image Process.*, vol. 29, pp. 1285–1298, 2020, doi: 10.1109/tip.2019.2936649.

[8] *Google AI Blog: Introducing the HDR+Burst Photography Dataset*. Accessed: Jun. 12, 2020. [Online]. Available: https://ai.googleblog.com/2018/02/introducing-hdr-burst-photography.html

[9] T. Mertens, J. Kautz, and F. Van Reeth, "Exposure fusion: A simple and practical alternative to high dynamic range photography," *Comput. Graph. Forum*, vol. 28, no. 1, pp. 161–171, Mar. 2009, doi: 10.1111/j.1467-8659.2008.01171.x.

[10] T. Jinno and M. Okuda, "Multiple exposure fusion for high dynamic range image acquisition," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 358–365, Jan. 2012, doi: 10.1109/TIP.2011.2160953.

[11] R. Montulet and A. Briassouli, "Deep learning for robust end-to-end tone mapping," in *Proc. BMVC*, 2019, p. 194. Accessed: Apr. 24, 2020.

[12] J. Tumblin, J. K. Hodgins, and B. K. Guenter, "Two methods for display of high contrast images," *ACM Trans. Graph.*, vol. 18, no. 1, pp. 56–94, Jan. 1999, doi: 10.1145/300776.300783.

[13] F. Drago, K. Myszkowski, T. Annen, and N. Chiba, "Adaptive logarithmic mapping for displaying high contrast scenes," *Comput. Graph. Forum*, vol. 22, no. 3, pp. 419–426, Sep. 2003, doi: 10.1111/1467-8659.00689.

[14] S. N. Pattanaik, J. Tumblin, H. Yee, and D. P. Greenberg, "Time-dependent visual adaptation for fast realistic image display," in *Proc. 27th Annu. Conf. Comput. Graph. Interact. Techn. (SIGGRAPH)*, 2000, pp. 47–54, doi: 10.1145/344779.344810.

[15] K. Kim, J. Bae, and J. Kim, "Natural HDR image tone mapping based on retinex," *IEEE Trans. Consum. Electron.*, vol. 57, no. 4, pp. 1807–1814, Nov. 2011, doi: 10.1109/TCE.2011.6131157.

[16] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," in *Proc. 29th Annu. Conf. Comput. Graph. Interact. Techn. (SIGGRAPH)*, 2002, pp. 267–276, doi: 10.1145/566570.566575.

[17] F. Durand and J. Dorsey, "Fast bilateral filtering for the display of high-dynamic-range images," in *Proc. 29th Annu. Conf. Comput. Graph. Interact. Techn. (SIGGRAPH)*, 2002, pp. 257–266, doi: 10.1145/566570.566574.

[18] P. Ledda, L. P. Santos, and A. Chalmers, "A local model of eye adaptation for high dynamic range images," in *Proc. 3rd Int. Conf. Comput. Graph., Virtual Reality, Visualisation Interact. Afr. (AFRIGRAPH)*, 2004, pp. 151–160, doi: 10.1145/1029949.1029978.

[19] F. Banterle, A. Artusi, K. Debattista, and A. Chalmers, *Advanced High Dynamic Range Imaging*, 2nd ed. Boca Raton, FL, USA: CRC Press, Sep. 2017, p. 328, doi: 10.1201/9781315119526.

[20] Y. H. Yee and S. Pattanaik, "Segmentation and adaptive assimilation for detail-preserving display of high-dynamic range images," *Vis. Comput.*, vol. 19, nos. 7–8, pp. 457–466, Dec. 2003, doi: 10.1007/s00371-003-0211-5.

[21] G. Krawczyk, K. Myszkowski, and H.-P. Seidel, "Lightness perception in tone reproduction for high dynamic range images," *Comput. Graph. Forum*, vol. 24, no. 3, pp. 635–645, Sep. 2005, doi: 10.1111/j.1467-8659.2005.00888.x.

[22] R. Fattal, D. Lischinski, and M. Werman, "Gradient domain high dynamic range compression," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 249–256, Jul. 2002, doi: 10.1145/566654.566573.

[23] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680. Accessed: Mar. 1, 2020. [Online]. Available: http://www.github.com/goodfeli/adversarial.

[24] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," 2017, *arXiv:1609.04802*. Accessed: May 3, 2020. [Online]. Available: https://arxiv.org/abs/1609.04802

[25] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. C. Loy, "ESRGAN: Enhanced super-resolution generative adversarial networks," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 1–16. Accessed: May 3, 2020. [Online]. Available: https://github.com/xinntao/ESRGAN

[26] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang. (2018). *Generative Image Inpainting With Contextual Attention*. Accessed: Apr. 26, 2020. [Online]. Available: https://github.com/

[27] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. Huang. (2019). *Free-Form Image Inpainting With Gated Convolution*. Accessed: Apr. 26, 2020. [Online]. Available: https://github.com/JiahuiYu/

[28] R. A. Yeh, C. Chen, T. Yian Lim, A. G. Schwing, M. Hasegawa-Johnson, and M. N. Do, "Semantic image inpainting with deep generative models," 2017, *arXiv:1607.07539*. Accessed: Apr. 26, 2020. [Online]. Available: https://arxiv.org/abs/1607.07539

[29] L. Yuan, C. Ruan, H. Hu, and D. Chen, "Image inpainting based on patch-GANs," *IEEE Access*, vol. 7, pp. 46411–46421, 2019, doi: 10.1109/ACCESS.2019.2909553.

[30] L. Kezebou, V. Oludare, K. Panetta, and S. S. Agaian, "TR-GAN: Thermal to RGB face synthesis with generative adversarial network for cross-modal face recognition," *Proc. SPIE*, vol. 11399, p. 23, Apr. 2020, doi: 10.1117/12.2558166.

[31] V. Oludare, L. Kezebou, K. Panetta, and S. S. Agaian, "Attention-guided cascaded networks for improved face detection and landmark localization under low-light conditions," *Proc. SPIE*, vol. 11399, p. 13, Apr. 2020, doi: 10.1117/12.2558397.

[32] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional GANs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8798–8807, doi: 10.1109/CVPR.2018.00917.

[33] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," Nov. 2016, *arXiv:1611.07004*. Accessed: Nov. 9, 2019. [Online]. Available: http://arxiv.org/abs/1611.07004

[34] N. Souly, C. Spampinato, and M. Shah, "Semi supervised semantic segmentation using generative adversarial network," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 5688–5696. Accessed: May 3, 2020.

[35] Y. Xue, T. Xu, H. Zhang, L. R. Long, and X. Huang, "SegAN: Adversarial network with multi-scale l1 loss for medical image segmentation," *Neuroinformatics*, vol. 16, nos. 3–4, pp. 383–392, Oct. 2018, doi: 10.1007/s12021-018-9377-x.

[36] S. Lee, G. H. An, and S.-J. Kang, "Deep chain HDRI: Reconstructing a high dynamic range image from a single low dynamic range image," *IEEE Access*, vol. 6, pp. 49913–49924, Sep. 2018, doi: 10.1109/ACCESS.2018.2868246.

[37] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. (2017). *Image-to-Image Translation With Conditional Adversarial Networks*. Accessed: Mar. 21, 2020. [Online]. Available: https://github.com/phillipi/pix2pix

[38] X. Qin, Z. Wang, Y. Bai, X. Xie, and H. Jia, "FFA-Net: Feature fusion attention network for single image dehazing," Nov. 2019, *arXiv:1911.07559*. Accessed: Mar. 16, 2020. [Online]. Available: http://arxiv.org/abs/1911.07559

[39] J. Pang, W. Sun, J. S. Ren, C. Yang, and Q. Yan, "Cascade residual learning: A two-stage convolutional neural network for stereo matching," 2017, *arXiv:1708.09204*. Accessed: May 3, 2020. [Online]. Available: https://arxiv.org/abs/1708.09204

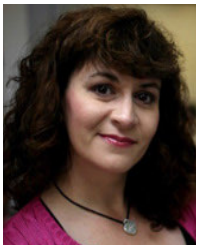
[40] N. Ahn, B. Kang, and K.-A. Sohn, "Fast, accurate, and lightweight super-resolution with cascading residual network," 2018, *arXiv:1803.08664*. Accessed: May 3, 2020. [Online]. Available: https://arxiv.org/abs/1803.08664

[41] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 9351, 2015, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.

[42] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 9906, 2016, pp. 694–711, doi: 10.1007/978-3-319-46475-6_43.

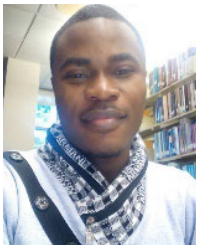
[43] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional GANs," 2018, *arXiv:1711.11585*. Accessed: Mar. 11, 2020. [Online]. Available: https://arxiv.org/abs/1711.11585

[44] M. S. Sarfraz, C. Seibold, H. Khalid, and R. Stiefelhagen, "Content and colour distillation for learning image translations with the spatial profile loss," Aug. 2019, *arXiv:1908.00274*. Accessed: Mar. 18, 2020. [Online]. Available: http://arxiv.org/abs/1908.00274
[45] Y. Yuan, W. Yang, W. Ren, J. Liu, W. J. Scheirer, and Z. Wang, "UG^{2+} track 2: A collective benchmark effort for evaluating and advancing image understanding in poor visibility environments," Apr. 2019, *arXiv:1904.04474*. Accessed: May 3, 2020. [Online]. Available: http://arxiv.org/abs/1904.04474
[46] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, Aug. 2018, pp. 1–12. Accessed: May 3, 2020. [Online]. Available: http://arxiv.org/abs/1808.04560
[47] H. Yeganeh and Z. Wang, "Objective quality assessment of tone-mapped images," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 657–667, Feb. 2013, doi: 10.1109/TIP.2012.2221725.
[48] H. Z. Nafchi, A. Shahkolaei, R. F. Moghaddam, and M. Cheriet, "FSITM: A feature similarity index for tone-mapped images," *IEEE Signal Process. Lett.*, vol. 22, no. 8, pp. 1026–1029, Aug. 2015, doi: 10.1109/LSP.2014.2381458.
[49] K. Panetta, C. Gao, and S. Agaian, "Human-visual-system-inspired underwater image quality measures," *IEEE J. Ocean. Eng.*, vol. 41, no. 3, pp. 541–551, Jul. 2016, doi: 10.1109/JOE.2015.2469915.
[50] R. Mantiuk, R. Mantiuk, A. Tomaszewska, and W. Heidrich, "Color correction for tone mapping," *Comput. Graph. Forum*, vol. 28, no. 2, pp. 193–202, Apr. 2009, doi: 10.1111/j.1467-8659.2009.01358.x.



KAREN PANETTA (Fellow, IEEE) received the B.S. degree in computer engineering from Boston University, Boston, MA, USA, in 1985, and the M.S. and Ph.D. degrees in electrical engineering from Northeastern University, Boston, in 1987 and 1994, respectively. She is currently the Dean of Graduate Engineering Education and a Professor with the Department of Electrical and Computer Engineering, Tufts University, Medford, MA, and the Director of the Dr. Panetta's Vision and Sensing System Laboratory. She is the 2019 IEEE-HKN (Eta Kappa Nu) President and the Vice-President of the IEEE Systems, Man and Cybernetics Society. Her research interests include developing efficient algorithms for simulation, modeling, signal, and image processing for biomedical and security applications

LANDRY KEZEBOU (Graduate Student Member, IEEE) received the Bachelor of Engineering degree in electrical engineering from Ahmadu Bello University, Zaria, Nigeria, in 2014, and the M.S. degree in electrical and computer engineering from Tufts University, Medford, MA, USA, in 2019, where he is currently pursuing the Ph.D. degree with the ECE Department. His current research interests include latest computer vision and deep learning technologies for highway intelligent transportation systems, object detection, object tracking, image enhancement, and generative adversarial networks.



VICTOR OLUDARE (Graduate Student Member, IEEE) received the B.S. degree in electrical engineering from the Federal University of Technology Minna, Nigeria, in 2014, and the M.S. degree from Tufts University, Medford, MA, USA, in 2018, where he is currently pursuing the Ph.D. degree in electrical engineering. His research interests include the use of computer vision techniques and deep learning architectures for conservation purposes, object detection, recognition and tracking, and image enhancement.

SOS AGAIAN (Fellow, IEEE) received the M.S. degree in mathematics and mechanics from Yerevan University, Yerevan, Armenia, in 1968, the Ph.D. degree in math and physics from the Steklov Institute of Mathematics, Russian Academy of Sciences, Moskva, Russia, in 1975, and the Doctor of Engineering Sciences degree from the Institute of the Control System, Russian Academy of Sciences, in 1985. He is currently a Distinguished Professor with the City University of New York, New York, NY, USA. His research interests include computational vision and machine learning, multimodal data fusion, signal/image processing modeling, multimodal biometric and digital forensics, three-dimensional imaging sensors, information processing and security, and biomedical and health informatics. He is a Fellow of SPIE, the Society for Imaging Science and Technology, and the American Association for the Advancement of Science.

ZEHUA XIA received the B.S. degree in electrical engineering, China. He is currently pursuing the master's degree with Tufts University. His current research interest includes image and signal processing.