Hybrid Loss for Learning Single-Image-based HDR Reconstruction

Kenta Moriwaki[1]   Ryota Yoshihashi[1]   Rei Kawakami[1]
Shaodi You[2,3]   Takeshi Naemura[1]

[1]The University of Tokyo   [2]Data61-CSIRO   [3]Australian National University

{moriwaki,yoshi,rei,naemura}@nae-lab.org, Shaodi.You@data61.csiro.au

Low-dynamic-range image (input)

High-dynamic-range image reconstructed by our method

Figure 1: The top row shows multiple exposure levels of an LDR image, the original exposure level of which is the third one from left. Given the original LDR image as an input, our method reconstructs an HDR image, as shown in the bottom row with the multiple exposure levels. No structures in the over-exposed region (green) are visible in the LDR image and the under-exposed region (red) is grossly quantized. Our method can inpaint these lost structures plausibly and recover intensity gradients in the under-exposed region.

Abstract

This paper tackles high-dynamic-range (HDR) image reconstruction given only a single low-dynamic-range (LDR) image as input. While the existing methods focus on minimizing the mean-squared-error (MSE) between the target and reconstructed images, we minimize a hybrid loss that consists of perceptual and adversarial losses in addition to HDR-reconstruction loss. The reconstruction loss instead of MSE is more suitable for HDR since it puts more weight on both over- and under- exposed areas. It makes the reconstruction faithful to the input. Perceptual loss enables the networks to utilize knowledge about objects and image structure for recovering the intensity gradients of saturated and grossly quantized areas. Adversarial loss helps to select the most plausible appearance from multiple solutions. The hybrid loss that combines all the three losses is calculated in logarithmic space of image intensity so that the outputs retain a large dynamic range and meanwhile the learning becomes tractable. Comparative experiments conducted with other state-of-the-art methods demonstrated

that our method produces a leap in image quality.

1. Introduction

High-dynamic-range (HDR) imaging is capable of expressing a wide range of light intensities. It can avoid over- and under- exposures, and can express image brightnesses beyond the quantization resolution of the sensor. It enhances the viewing experience when images are shown on HDR displays. Moreover, it is used in image-based rendering for accurate simulation of environmental lighting, and in artistic image editing, owing to its rich representation capability. For an overview of the technique, see [32, 1, 26]. To obtain an HDR image, we currently need either an expensive HDR camera or multiple shots from a low-dynamic-range (LDR) camera with different exposures [4]. However, if an HDR image could be reconstructed from a single LDR image, billions of photographs shot in LDR could be utilized for HDR applications.

To enable single-image HDR reconstruction, conven-

tional inverse-tone-mapping (iTM) methods focus on saturated areas and extrapolate the light intensity from the surrounding regions on the basis of local heuristics. As the heuristics are not universal, they often fail to recover over- and under- exposed areas or introduce unnatural artifacts. More recently, deep-learning-based methods have been applied to single-image-based HDR reconstruction [6, 5, 27, 22, 39]. However, the images produced by the existing methods still suffer from artifacts and insufficient contrast. This is partly because they only focus on minimizing the mean-squared error (MSE) between the reconstructed and target images.

Single-image-based HDR reconstruction is a highly ill-posed problem that prevents MSE-based learning from producing satisfactory images. An ideal HDR reconstruction must have the following properties: 1) Inpainting of over-exposed areas: clipped values due to saturation must be extrapolated for higher intensity, but since the information is lost from those regions, how to extrapolate them is often rather ambiguous. 2) Recovering the intensity gradients of under-exposed areas: due to quantization, gradations are lost in very dark regions, which causes unnatural artifacts. 3) Visual fidelity: after inpainting and restoration, the reconstructed HDR image must appear natural to the human eye. Blur or artifacts harm the visual impression even if they are negligible when measured by MSE. 4) Semantic consistency: inpainted or restored HDR images must be plausible with the context of the scene and its objects. For example, in a sunset image, saturated regions in the sky may have to be re-colorized in reddish orange. Given such four properties, multiple solutions may exist, and as MSE-based solutions tend to be the average of a number of possible solutions, they may often be blurred.

This hybrid nature of HDR reconstruction makes it difficult to design one good loss function. To overcome the problem, we introduce a *hybrid loss* that consists of HDR-reconstruction loss, adversarial loss [10], and perceptual loss [8, 15, 21]. HDR-reconstruction loss is our novel loss that minimizes per-pixel difference between images, and is useful for image reconstruction. This loss is designed to put more weight on over- and under- exposed areas so that it achieves the properties of 1) and 2). Adversarial loss is utilized for better image quality, addressing the property of 3). In contrast to the other losses which are fixed during optimization, the loss from the framework of generative adversarial networks (GAN) [10] can be formulated as a two-player game where the generator and the discriminator are updated competitively and whose solution is the Nash equilibrium between the players. A GAN can select a solution from among the set of possible ones [9]. Therefore, it may find a better solution than those only relying on MSE minimization. Perceptual loss is a high-level similarity between images, which can be calculated by using the output of the high-level layer in pre-trained convolutional neural networks (CNN) that encode broad knowledge of object appearances. The loss causes the reconstructed image to be more natural and plausible, addressing the property of 4).

All of the losses are essential to guide the optimization of our network, and multiple losses can also avoid artifacts that may be produced by a single loss measures. To ensure the output values that have a high dynamic range and still keep the learning tractable, all losses are calculated in logarithmic space of image intensity. This is found to be very effective. We also introduce a new evaluation protocol for HDR image reconstruction that considers the intensity gradient recovery in under-exposed regions as equally as that in over-exposed regions. The experiments show that the proposed method makes a leap in the resulted image quality compared to the state-of-the-art methods.

The contributions of the paper are summarized as follows. First, we propose a hybrid loss consisting of reconstruction loss, adversarial loss, and perceptual loss. The reconstruction loss is devised so that the intensity gradients in saturated and dark regions can be reconstructed. All losses are defined in logarithmic space of image intensity, which is crucial to make the learning tractable. To our knowledge, this is the first study that introduces a perceptual loss for HDR reconstruction. Second, the results of the proposed method are a leap in quality in comparison with the state of the art. Third, we introduced a new evaluation protocol that takes the under-exposed regions into account as well as the over-exposed regions. Finally, we collected a new HDR dataset that are publicly available on the web, and will release the URL list of the images. The code and trained model will be published upon acceptance of this paper.

2. Related work

Single-image HDR reconstruction Conventionally, HDR reconstruction has been performed by non-learning-based brightness enhancement through filtering or light-source detection. For example, bilateral filters applied to x-y-range three-dimensional grids work as brightness enhancement functions [20, 19]. However, non-learning-based approaches cannot estimate physically accurate amounts of light due to the lack of knowledge about real HDR images; thus, the quality of the estimated HDR images is limited.

A few studies have used deep learning in HDR reconstruction from a single LDR image. Such methods can be categorized into multi-step and single-step methods. The multi-step methods generate bracketed images with multiple exposures and then merge them. The single-step methods generate an HDR image directly in one step.

An example of a multi-step methods is Deep Reverse Tone Mapping (DrTMO) [6], which generates multiple images with different exposures using an encoder-decoder network [13, 37]. To train the network, LDR images are simu-

lated using various camera curves [12] from an HDR image dataset and input. ChainHDRI [22] and Recursive-HDRI [23] are similar to DrTMO [6], the difference being that they recurrently generate higher or lower exposure images from images generated in the previous time steps. However, such recurrent methods need multiple forward computations in one HDR generation; in contrast, ours can generate HDR images in one forward pass.

With the growing popularity of end-to-end learning, single-step networks that directly estimate the desired HDR images may be preferable to multi-step methods. HDR-CNN [5] and Deep Reciprocating HDR [39] share the same encoder-decoder structure that directly generates an HDR image from an LDR image. While the architecture itself is similar to UNet for segmentation [34], they train networks to recover from over-/under-exposures of moderate extent that are artificially added to the training LDR images. ExpandNet [27] has a three-branch architecture designed for single-step HDR image generation, and the branches are for global, semi-local, and local feature extraction. In contrast, we show that the simple encoder-decoder architecture performs well with our augmented loss functions.

GANs The essential difficulty with single-image HDR reconstruction is in the restoration of over- or under-exposed regions, where structures in the original scenes are totally lost or heavily corrupted. Even the deep-learning methods discussed above suffer from imperfect restoration and unnatural artifacts. For this reason, GANs [10], which have successfully restored and inpainted natural appearing images [30, 38], are considered promising. If the reconstructive error (for example, the MSE between the outputs and the training images) is the only loss function, the restored images are easily blurred. A GAN can mitigate such artifacts and recover more detailed texture.

GANs have already been used for HDR image generation, by Lee *et al*. [23] and Ning *et al*. [29]. By introducing GAN, the restoration quality is further improved than the simple encoder-decoder networks. We found that GAN combined with reconstructive error still generates blur or unnatural artifacts. In this paper, by further introducing perceptual loss and reconstruction loss optimized for HDR, the image quality can be improved.

Deep image processing Apart from HDR reconstruction, we can see wider variety of deep-learning methods for image processing within LDR images, which are still useful as references. For example, convolutional GANs well-performed in superresolution [21], denoising [3], or inpaiting [30]. Other than GANs, there are some promising approaches such as multiscale [38, 24], perceptual losses [15], attention [31], or reinforcement learning [40, 7]. While their insights are useful also for our task, such methods for LDR images are not directly applicable to HDR images.
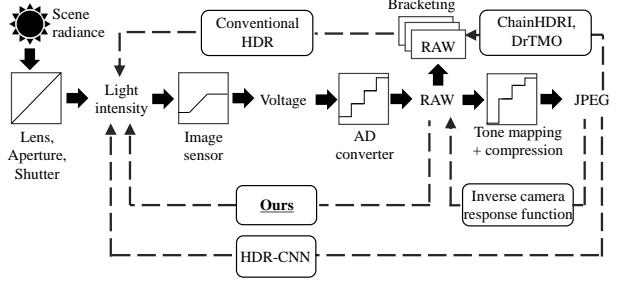

Figure 2. The pipelines of image formation in cameras and HDR reconstruction methods. Ours estimates HDR light intensity from linearized (raw) images, while other deep-learning-based methods use images after tone mapping and compression. Our method can accurately estimate the light intensity using measurement-based linearization.

3. Method

3.1. Problem statement

Single-image-based HDR reconstruction can be defined as a task to estimate physical light intensity from single LDR images. Since the imaging pipeline in cameras is a lossy process, estimating physical light intensity from RAW or JPEG images is an ill-posed problem. The imaging pipeline [17], as shown in Fig 2, consists of the followings: First, a lens gathers light rays and forms an image on a sensor. The amount of light that reaches the sensor is controlled by an aperture and shutter speed, which decides the exposure value of the image. The sensor outputs voltages corresponding to the amount of light, but too large voltages are cropped due to saturation. The voltages from the sensor are digitized by an AD converter, and in this part the small voltage values are quantized, leading to the lost tones in under-exposed regions. The images after AD conversion are called RAW images, and they are further tone-mapped and compressed into JPEG images. From such images, single-image-based HDR reconstruction methods need to estimate the original light intensity.

Given the pipeline of image formation, there is a degree of freedom in from which stage a method reconstruct HDR images. The most conventional way to reconstruct HDR images from LDR images is exposure bracketing [4], which is to capture a single scene by multiple LDR images with various exposure values and merge them later into an HDR image. In single-image-based HDR reconstruction, most of the learning-based HDR-reconstruction methods use JPEG images after tone mapping [5, 6]. While this is useful for applying to daily JPEG images, it may make the reconstruction more difficult. The tone mapping makes the nonlinearity between light intensity and pixel values larger. In addition, the mapping functions differ by cameras, which increases the uncertainty of reconstruction. In contrast, we estimate HDR images from raw images, which preserve lin-
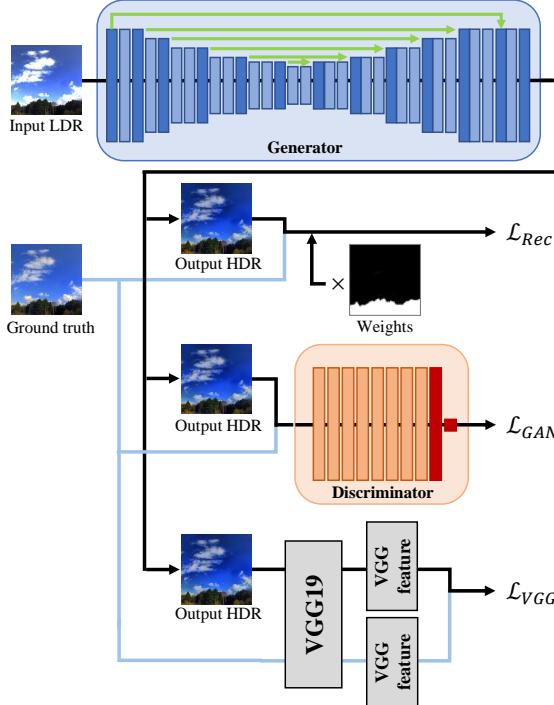
Figure 3. The overview of our method. Our generator is an encoder-decoder network with skip connections. In addition to the adversarial loss $\mathcal{L}_{GAN}$ from the discriminator, we also incorporate HDR-reconstruction loss $\mathcal{L}_{Rec}$ and perceptual loss $\mathcal{L}_{VGG}$ to improve image quality.

ear relationship to light intensity within non-saturating regions. This does not reduce the applicability of our method, since raw images can be easily recovered from JPEG images when the camera response function is known, and even when it is unknown, a number of methods are available to estimate the inverse camera response function from images [11, 36].

3.2. The hybrid loss

The outline of the method is shown in Fig. 3. The generator is an encoder-decoder network with skip connections. The input for the generator is a color LDR image with 8 bits per channel, and the output is an image of the float data type with 32 bits per channel.[1] The detail on how to synthesize an input LDR image is described in Sec. 4.

To train the generator, we use a hybrid loss, combining HDR-reconstruction loss $\mathcal{L}_{Rec}$, adversarial loss $\mathcal{L}_{GAN}$, and perceptual loss $\mathcal{L}_{VGG}$. The hybrid loss of the generator $\mathcal{L}_G$ can be written as follows:

$$\mathcal{L}_G = \mathcal{L}_{Rec} + \alpha \mathcal{L}_{GAN} + \beta \mathcal{L}_{VGG}, \quad (1)$$

where α and β are weights to balance the losses.

---

[1]Note that our method outputs much more bits compared to that of the HDR image format.

HDR-reconstruction loss Our $\mathcal{L}_{Rec}$, the reconstruction loss, penalizes the per-pixel ℓ^2 distance of intensities between the reconstructed and the ground-truth HDR images. A problem in defining $\mathcal{L}_{Rec}$ w.r.t HDR images is the wide range of values; naive loss functions, such as the mean-squared error, may depend too much on the high-luminance regions, and errors in the lower range will be negligible. To avoid this, we define the loss in the logarithmic domain of pixel intensity. We also introduce weights to put more attention on over- and under- exposed regions. Thus, the $\mathcal{L}_{Rec}$ can be expressed as follows:

$$\mathcal{L}_{Rec}(\hat{y}, y) = \frac{1}{N} \sum_{i=1}^{N} w_i |\log(\hat{y}_i) - \log(y_i)|^2, \quad (2)$$

where $\hat{y}$ is the reconstructed HDR image by the generator, y is the original HDR image in the training set (ground truth), w_i corresponds to the pixel-wise weights, i is for each pixel, and N is the number of pixels. For simplicity, Eq. 2 shows the HDR-reconstruction loss in the per-image form, but the loss is averaged over the mini batch during training.

Since over-exposed regions can naturally have large intensity difference, we introduce weights for emphasizing under-exposed regions. Thus, w_i is defined as

$$w_i = 1 + \gamma \max\left(0, 1 - \frac{1}{\tau} \min_c(x_{i,c})\right), \quad (3)$$

where $x_{i,c}$ is the LDR image normalized to $[0, 1]$, and c is for color channel. τ is the threshold, and γ is a weight to enhance the loss of under-exposed regions. In this paper, the threshold τ is set to 0.05.

Adversarial loss The adversarial loss $\mathcal{L}_{GAN}$, as in a GAN, is introduced so that the generator can deceive the discriminator. The loss is useful for making the generated images close to the distribution of the original dataset. It is expressed as follows:

$$\mathcal{L}_{GAN} = \sum_{j=1}^{M} -\log D(\log(\hat{y}_j)), \quad (4)$$

where $D(\hat{y})$ is the probability of classifying whether $\hat{y}$ is real or fake, and M is the number of training images in the mini batch. Here, we also use logarithm of $\hat{y}$ for calculating the loss, to keep the high dynamic range of the images and still make the learning tractable. The loss is smaller when $\hat{y}$ is closer to the original input, such that the discriminator D considers it is real.

The discriminator D is composed of eight convolutional layers and two fully connected layers. In the training, the input for the discriminator is either of a pair, one of which is $\hat{y}$ and the other is y. The loss function $\mathcal{L}_D$ for training the

discriminator can be written as follows [10]:

$$\mathcal{L}_D = -\frac{1}{M} \sum_{j=1}^{M} \left(\log\big(1 - D\big(\log(\hat{y}_j)\big)\big) + \log\big(D(\log(y_j))\big) \right).$$

(5)

The logarithm of $\hat{y}$ and y is introduced for the same reason as in Eq. 4. The two losses $\mathcal{L}_G$ and $\mathcal{L}_D$ are used for updating the weights of the generator and the discriminator, respectively.

Perceptual loss The perceptual loss enhances perceptual similarity for human eyes and mitigate artifacts in the output image by utilizing pre-trained image-classification networks. We adopt VGG19 [35] for this purpose following [2, 21, 15]. VGG19 is pre-trained in ILSVRC2012, which is an LDR-image-classification dataset, and it is not directly applicable to HDR images due to the difference of domains. However, we found that the perceptual loss is still useful by applying logarithmic transformation on input HDR images. Specifically, we denote the output of the l-th pooling layer in VGG19 as ϕ_l, and $\mathcal{L}_{VGG}$ is defined as follows:

$$\mathcal{L}_{VGG}(\hat{y}, y) = \frac{1}{N_l} \sum_{k=1}^{N_l} |\phi_l(\log(\hat{y}))_k - \phi_l(\log(y))_k|^2, \quad (6)$$

where N_l represents the total number of pixels in the feature space, and $\phi_l(\cdot)_k$ represents the feature vector at the pixel k. In this paper, we used $l = 5$ in the same manner as in [21]. The perceptual loss is also averaged over the mini batch during training.

4. Experiments

We conducted experiments on publicly available HDR image sets to compare the reconstructed HDR quality of our method and existing ones. Furthermore, we show extensive visualization of HDR-reconstruction results and analysis.

Datasets We used a part of the dataset used in HDR-CNN [5] and images crawled from the web newly by us. The motivation is that approximately half of the training images used in [5] are private data of the author which are not publicly available. Due to the lack of public large-scale HDR image set sufficient to train deep networks, prior studies partially used private HDR image sets [5, 22, 23], which may be a problem in reproduction. In contrast, all of our training data is publicly available on the web. Our dataset consists of images of indoor and outdoor scenes. We used 999 HDR images and 61 HDR videos for training. We will release the URL list of the images, although distribution of the original images is not allowed due to the copyright.

To train an HDR reconstruction network, we need a collection of pairs, one of which is an HDR image as ground
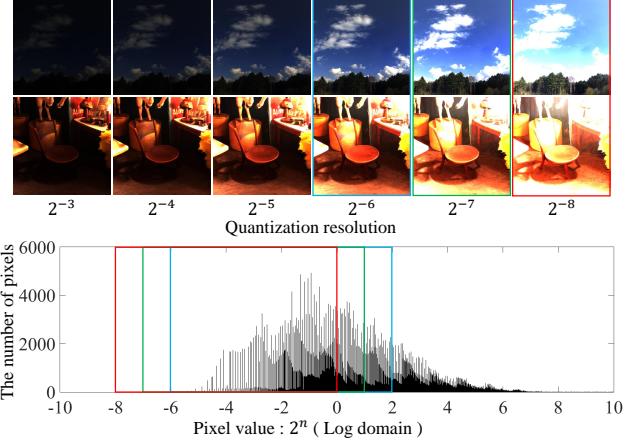

Figure 4. The visualization of decomposed LDR images from an HDR image. Each exposure level is referred by its quantization resolution. The lower figure shows the entire histogram of HDR pixel values and LDRs capture part of it. The red, green, and blue rectangles correspond to the same exposure level in the top row.

truth and the other is an LDR image as the input. Thus, we automatically generated LDR images from the HDR dataset. First, the HDR images were cropped at random positions and resized to 256×256. Then, each HDR image was normalized by its median of the luminance value. Finally, we decomposed the HDR image into multiple LDR images with various quantization resolutions, each of which had the range of 8-bit for each channel. Figure 4 shows the visualization of the range of decomposed LDR images. Specifically, we decomposed an HDR image into six exposure levels, and we refer to each of the levels by using their quantization resolution. Among the six decomposed images, images in the five ranges corresponding to the quantization resolution of 2^{-8} to 2^{-4} are randomly used as inputs in the training. The total number of LDR images for training is 127,831.

Training We followed the GAN training framework [10], where a generator and a discriminator were updated alternatively. We used our hybrid loss (Eq. 1) to update our generator, and the discriminator loss (Eq. 5) to update our discriminator during training. The loss minimization was performed with the ADAM optimizer [18]. For ADAM's parameter, the initial learning rate was set to 2.0×10^{-5}, the batch size was 16, and the total number of epochs was 60. For the parameters in the hybrid loss, α and β in Eq. 1 were set as follows: $\alpha = 1.0 \times 10^{-3}, \beta = 5.0 \times 10^{-4}$. α was set to align the order of magnitude of L_{Rec} and L_{GAN}, and β was set so that L_{VGG} is approximately 10 times larger than the other loss functions. γ was set to 5.0, so that the under-exposed region was enhanced approximately 5 times.

We utilized the weights of the trained model of HDR-CNN [5] as the initial values of the generator. In HDR-CNN, the model was pre-trained with Places database [42],

Table 1. Comparison of the ground truth and HDR images by the proposed and other state-of-the-art methods. The input images have brightness coressponding to 2^{-6}. (See Fig.4 for the corresponding range.)

| | Reinhard's TMO | | | | Kim and Kautz's TMO | | | | VDP quality score | |
| | PSNR(dB) | | SSIM | | PSNR(dB) | | SSIM | | | |
	m	σ	m	σ	m	σ	m	σ	m	σ
Proposed	**31.53**	5.60	**0.948**	0.028	**29.71**	2.74	**0.931**	0.034	**51.30**	4.79
HDR-CNN [5]	18.33	1.27	0.791	0.063	21.68	2.28	0.846	0.061	51.09	4.57
DrTMO [6]	23.70	7.26	0.846	0.165	21.97	7.72	0.819	0.184	43.59	3.70
ExpandNet [27]	18.60	3.28	0.729	0.129	17.35	2.43	0.721	0.093	46.92	6.16
Huo *et al.* [14]	14.97	0.90	0.665	0.072	15.82	1.62	0.716	0.069	39.77	3.87
KOEO [19]	16.75	1.71	0.706	0.066	16.96	2.32	0.737	0.069	39.00	3.10
RecursiveHDRI [23]	26.71	2.78	/	/	22.31	3.20	/	/	48.85	4.91

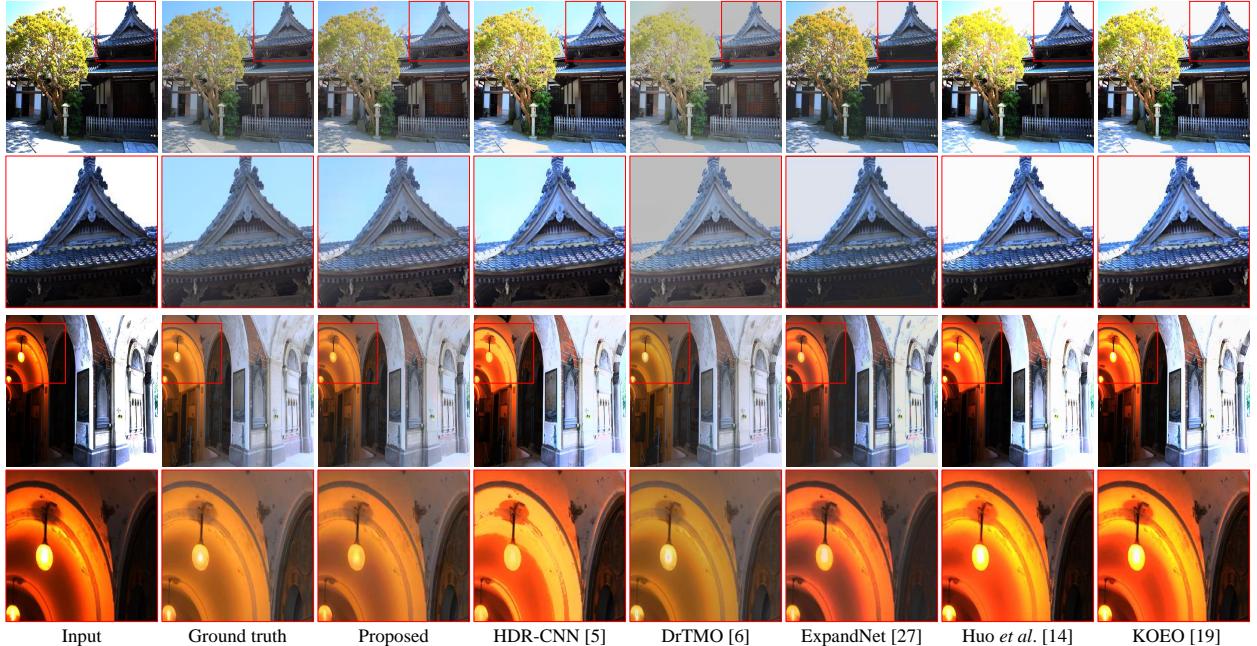| Input | Ground truth | Proposed | HDR-CNN [5] | DrTMO [6] | ExpandNet [27] | Huo *et al.* [14] | KOEO [19] |

Figure 5. Comparison between the ground truth and HDR images reconstructed by the proposed and other methods. HDR images are tone-mapped using the method of Reinhard *et al.* [33]

Table 2. PSNR(dB) compared in the LDR image stacks. (See Fig. 4 for the corresponding range.)

Method	2^{-3}	2^{-4}	2^{-5}	2^{-6}	2^{-7}	2^{-8}	Mean
Proposed	**34.57**	**30.66**	**27.67**	**27.82**	**32.44**	**32.16**	**30.88**
HDR-CNN [5]	27.66	23.20	19.44	17.67	18.90	16.39	20.54
DrTMO [6]	28.56	24.05	20.26	18.97	23.19	26.59	23.60
ExpandNet [27]	25.44	20.89	17.22	16.15	19.95	18.84	19.74
Huo *et al.* [14]	19.09	14.13	11.35	12.74	16.48	14.50	14.71
KOEO [19]	17.13	12.26	11.26	13.50	16.97	15.21	14.38

Table 3. SSIM compared in the LDR image stack.

Method	2^{-3}	2^{-4}	2^{-5}	2^{-6}	2^{-7}	2^{-8}	Mean
Proposed	**0.937**	**0.931**	**0.931**	**0.951**	**0.978**	**0.964**	**0.948**
HDR-CNN [5]	0.833	0.784	0.752	0.753	0.782	0.793	0.782
DrTMO [6]	0.859	0.828	0.812	0.843	0.906	0.953	0.866
ExpandNet [27]	0.786	0.738	0.718	0.756	0.841	0.877	0.786
Huo *et al.* [14]	0.597	0.534	0.516	0.581	0.681	0.753	0.610
KOEO [19]	0.567	0.515	0.534	0.616	0.706	0.762	0.616

and then trained using the collected HDR-image dataset. We fine-tuned the model further with our dataset. The network architecture of the discriminator is based on [21], which is composed of 10 layers, 8 of which are convolutional layers and 2 are fully connected layers. It is trained from scratch.

Comparisons with state-of-the-art methods First, we show the quantitative comparisons to existing single-image-based HDR reconstruction methods. Following the latest work [23], we use HDREye dataset [28] for the test set. We report PSNR and SSIM between the ground truth and each HDR image inferred by the methods, both of which were tone-mapped by the tone-mapping operator (TMO) of Reinhard *et al.* [33] and Kim and Kautz [16]. Also, we report the metric of HDR-VDP-2 [25], which is based on the human visual system to evaluate the estimated HDR images. The parameters used for HDR-VDP-2 are exactly
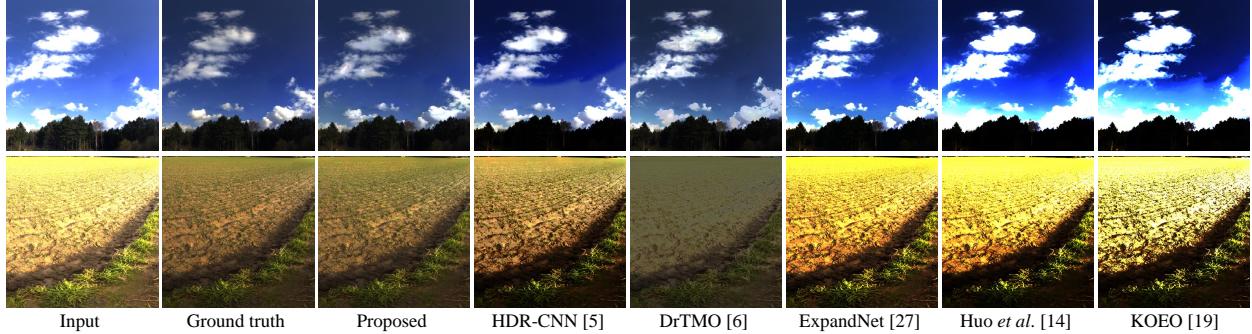
Figure 6. Comparison of the ground truth LDR and reconstructed LDR images. The image corresponding to the range of 2^{-7} was used for the input, and the estimated HDR image was decomposed into six exposure levels of LDR images and evaluated. The figure shows the ground truth and the results of each method corresponding to the range of 2^{-6} in Fig. 4.

the same as in [23]: a 24-inch display with a resolution of $1,900 \times 1,200$, and a view distance of 0.5 m.

Table 1 and Fig. 5 show the evaluation results. We compared ours with the iTM methods of Huo *et al.* [14] and Kovaleski and Oliveira expansion operator (KOEO) [19], in addition to DrTMO [6], HDR-CNN [5] and Expand-Net [27], which are the state-of-the-art methods using deep learning. Table 1 is divided into three blocks. The top block shares the same training dataset of ours. The middle block is the results of using the models that have been released to public, though the training dataset may be partially different from ours (though we believe there should be a lot of overlap). The bottom block is the results reported in Recursive-HDRI [23] using exactly the same test set and evaluation protocol, though we do not have access to the network or the dataset. DrTMO [6] is supposed to use the processed image with camera response function as an input, but we report the results inputting the linear LDR image (the same as others), since the scores are slightly better with the linear inputs.

As Table 1 shows, the proposed method is superior to the other methods in all the metrics. As shown in Fig. 5, even if qualitatively evaluated, we can see that our results are closer to the ground truth images. As shown in the images in the top two rows, the saturated regions of the sky area were successfully recovered by the proposed method. As shown in the images in the bottom two rows, the texture inside the light bulb is successfully recovered.

Range-wise evaluation In the evaluation metrics in Table 1, the errors in the bright regions will be dominant and under-exposed regions will have negligible effect. To visualize the errors at each exposure levels equally, in this paper, we introduce a new evaluation protocol; namely, the errors are evaluated in the decomposed LDR images, ranging from 2^{-3} to 2^{-8}. For this evaluation, our dataset was split into training and testing sets. We used 900 images for training. For testing, we excluded images that may be used for training in HDR-CNN [5], and used the remainder which contain

33 images. The input LDR images are in the range of 2^{-7}.

Tables 2 and 3, and Fig. 6 show the results evaluated with the decomposed LDR image sets. Tables 2 and 3 include two blocks as in Table 1, where the difference between them is the training datasets. PSNR and SSIM shown in Tables 2 and 3 respectively show that the proposed method is superior to other methods in all the ranges. As shown in Fig.6, iTMs fail to recover over-exposed areas, and lead to unnatural boundaries in the near-saturation areas. The proposed method can recover the contrast accurately, and the results are visually closer to the ground truth image. More results can be found in the supplementary material.

Ablation study We compare our full hybrid loss with its ablations. We use the HDREye dataset and the range-wise evaluation protocol for this study. The input images are either in the range of 2^{-5} or 2^{-6}. Table 4 and Fig. 7 show the results of the combinations of the three losses, $\mathcal{L}_{Rec}$, $\mathcal{L}_{GAN}$, and $\mathcal{L}_{VGG}$, used for the generator. Although in Table 4 the PSNR and SSIM are slightly degraded by introducing $\mathcal{L}_{GAN}$ or $\mathcal{L}_{VGG}$, as shown in Fig. 7, result images such as those using $\mathcal{L}_{GAN}+\mathcal{L}_{VGG}$ are visually more plausible, while results of $\mathcal{L}_{Rec}$ are smoothed and texture-less. By using the hybrid loss, the reconstructed image is visually plausible and also faithful to the input image. In addition, each of the networks, i.e., the generator, the discriminator, and the VGG19, has unique artifacts because of aliasing of using the spatial re-sampling. By combining all of them, such patterns can be reduced effectively.

Restoration of under-exposed areas Most existing methods focus on the saturated region for intensity recovery. We show that our method is effective for restoring the dark region as shown in Fig. 8. Information in the under-exposed region is grossly quantized as can be seen in the image with the adjusted gain. The boundaries of the staircase and the texture of the walls are restored with the weights in Eq. 3.

| Input | Ground truth | Proposed | $\mathcal{L}_{Rec}$ | $\mathcal{L}_{Rec} + \mathcal{L}_{GAN}$ | $\mathcal{L}_{Rec} + \mathcal{L}_{VGG}$ | $\mathcal{L}_{GAN} + \mathcal{L}_{VGG}$ |

Figure 7. Comparison of results using different combinations of the loss functions, $\mathcal{L}_{Rec}$, $\mathcal{L}_{GAN}$, and $\mathcal{L}_{VGG}$. The results of the proposed method are visually the closest to the ground truth, and have less artifacts.
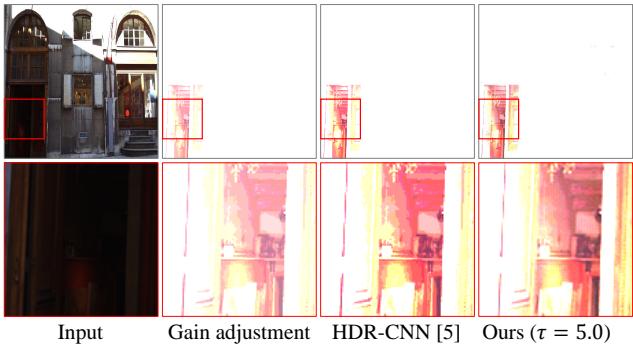

| Input | Gain adjustment | HDR-CNN [5] | Ours ($\tau = 5.0$) |

Figure 8. Zoomed views of under-exposed region. The intensity gradients are recovered successfully in those grossly quantized areas with our method. (Best viewed in color with zoom in.)

Table 4. Ablation study of the loss functions. In terms of PSNR and SSIM, reconstruction loss $\mathcal{L}_{Rec}$ is the most effective. However, $\mathcal{L}_{GAN}$ and $\mathcal{L}_{VGG}$ produce visually plausible results as shown in Fig. 7.

					Ours
$\mathcal{L}_{Rec}$	✓	✓	✓		✓
$\mathcal{L}_{GAN}$		✓		✓	✓
$\mathcal{L}_{VGG}$			✓	✓	✓
PSNR	**31.37**	30.83	30.69	29.54	30.75
SSIM	**0.957**	0.951	0.952	0.945	0.953

5. Conclusion

We presented a method to reconstruct HDR images directly from a single LDR image by designing a hybrid loss incorporating HDR reconstruction loss, adversarial loss, and perceptual loss, which are all calculated using logarith-

mic space of image intensity. The method produces superior results compared with existing methods, and successfully recovers both over- and under- exposed regions.

The limitation is that when saturated areas are too large, the generator struggles to inpaint the regions. Deepening the network, collecting larger datasets, and conducting user study are our future work.

6. Acknowledgement

References

[1] F. Banterle, A. Artusi, K. Debattista, and A. Chalmers. *Advanced High Dynamic Range Imaging: Theory and Practice (2nd Edition)*. AK Peters (CRC Press), Natick, MA, USA, July 2017. 1

[2] J. Bruna, P. Sprechmann, and Y. LeCun. Super-resolution with deep convolutional sufficient statistics. In *International Conference on Learning Representations (ICLR)*, 2015. 5

[3] J. Chen, J. Chen, H. Chao, and M. Yang. Image blind denoising with generative adversarial network based noise modeling. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 3

[4] P. E. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. In G. S. Owen, T. Whitted, and B. Mones-Hattal, editors, *SIGGRAPH*, pages 369–378. ACM, 1997. 1, 3

[5] G. Eilertsen, J. Kronander, G. Denes, R. Mantiuk, and J. Unger. Hdr image reconstruction from a single exposure using deep cnns. *ACM Transactions on Graphics (TOG)*, 36(6), 2017. 2, 3, 5, 6, 7, 10

[6] Y. Endo, Y. Kanamori, and J. Mitani. Deep reverse tone mapping. *ACM Transactions on Graphics (Proc. of SIGGRAPH ASIA 2017)*, 36(6), Nov. 2017. 2, 3, 6, 7, 10

[7] R. Furuta, N. Inoue, , and T. Yamasaki. Fully convolutional network with multi-step reinforcement learning for image processing. In *AAAI Conference on Artificial Intelligence (AAAI)*, 2019. 3

[8] L. A. Gatys, A. S. Ecker, and M. Bethge. Texture synthesis using convolutional neural networks. In *NIPS*, 2015. 2

[9] I. J. Goodfellow. NIPS 2016 tutorial: Generative adversarial networks. *CoRR*, abs/1701.00160, 2017. 2

[10] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'14, pages 2672–2680, Cambridge, MA, USA, 2014. MIT Press. 2, 3, 5

[11] M. D. Grossberg and S. K. Nayar. Determining the camera response from images: What is knowable? *IEEE*

Transactions on Pattern Analysis & Machine Intelligence, (11):1455–1467, 2003. 4

[12] M. D. Grossberg and S. K. Nayar. What is the space of camera response functions? In *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003.Proceedings.*, volume 2, pages 602–612, 2003. 3

[13] G. E. Hinton and R. R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, July 2006. 2

[14] Y. Huo, F. Yang, L. Dong, and V. Brost. Physiological inverse tone mapping based on retina response. *The Visual Computer*, 30(5):507–517, 2014. 6, 7, 10

[15] J. Johnson, A. Alahi, and L. Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European Conference on Computer Vision*, 2016. 2, 3, 5

[16] M. H. Kim and J. Kautz. Consistent tone reproduction. In *Proceedings of the Tenth IASTED International Conference on Computer Graphics and Imaging*, CGIM '08, pages 152–159, Anaheim, CA, USA, 2008. ACTA Press. 6

[17] S. J. Kim, H. T. Lin, Z. Lu, S. Süsstrunk, S. Lin, and M. S. Brown. A new in-camera imaging model for color computer vision and its application. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(12):2289–2302, 2012. 3

[18] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*, 2015. 5

[19] R. Kovaleski and M. M. Oliveira. High-quality reverse tone mapping for a wide range of exposures. In *Graphics, Patterns and Images (SIBGRAPI), 2014 27th SIBGRAPI Conference on*, pages 49–56, Aug 2014. 2, 6, 7, 10

[20] R. P. Kovaleski and M. M. Oliveira. High-quality brightness enhancement functions for real-time reverse tone mapping. *The Visual Computer*, 25(5-7):539–547, 2009. 2

[21] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. P. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi. Photo-realistic single image super-resolution using a generative adversarial network. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 105–114, 2017. 2, 3, 5, 6

[22] 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*, 6:49913–49924, 2018. 2, 3, 5

[23] S. Lee, G. Hwan An, and S.-J. Kang. Deep recursive hdri: Inverse tone mapping using generative adversarial networks. In *The European Conference on Computer Vision (ECCV)*, September 2018. 3, 5, 6, 7

[24] K. Lu, S. You, and N. Barnes. Deep texture and structure aware filtering network for image smoothing. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 217–233, 2018. 3

[25] R. Mantiuk, K. J. Kim, A. G. Rempel, and W. Heidrich. Hdr-vdp-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. *ACM Trans. Graph.*, 30(4):40:1–40:14, July 2011. 6

[26] R. K. Mantiuk, K. Myszkowski, and H.-P. Seidel. High dynamic range imaging, 2015. 1

[27] D. Marnerides, T. Bashford-Rogers, J. Hatchett, and K. Debattista. Expandnet: A deep convolutional neural network for high dynamic range expansion from low dynamic range content. *Comput. Graph. Forum*, 37(2):37–49, 2018. 2, 3, 6, 7, 10

[28] H. Nemoto, P. Korshunov, P. Hanhart, and T. Ebrahimi. Visual attention in ldr and hdr images. 2015. 6

[29] S. Ning, H. Xu, L. Song, R. Xie, and W. Zhang. Learning an inverse tone mapping network with a generative adversarial regularizer. 04 2018. 3

[30] D. Pathak, P. Krähenbühl, J. Donahue, T. Darrell, and A. Efros. Context encoders: Feature learning by inpainting. In *CVPR*, 2016. 3

[31] R. Qian, R. T. Tan, W. Yang, J. Su, and J. Liu. Attentive generative adversarial network for raindrop removal from a single image. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. 3

[32] 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. 1

[33] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda. Photographic tone reproduction for digital images. *ACM Trans. Graph.*, 21(3):267–276, July 2002. 6

[34] O. Ronneberger, P.Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 9351 of *LNCS*, pages 234–241. Springer, 2015. 3

[35] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *In International Conference on Learning Recognition(ICLR)*, 2015. 5

[36] J. Takamatsu, Y. Matsushita, and K. Ikeuchi. Estimating camera response functions using probabilistic intensity similarity. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8. IEEE, 2008. 4

[37] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol. Extracting and composing robust features with denoising autoencoders. pages 1096–1103, 2008. 2

[38] C. Yang, X. Lu, Z. Lin, E. Shechtman, O. Wang, and H. Li. High-resolution image inpainting using multi-scale neural patch synthesis. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017. 3

[39] X. Yang, K. Xu, Y. Song, Q. Zhang, X. Wei, and R. W. Lau. Image correction via deep reciprocating hdr transformation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. 2, 3

[40] K. Yu, C. Dong, L. Lin, and C. C. Loy. Crafting a toolchain for image restoration by deep reinforcement learning. 3

[41] G. Zaal. HDRI Haven. https://hdrihaven.com/hdris/. [Online; accessed 23-Nov-2018]. 10

[42] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva. Learning deep features for scene recognition using places database. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *NIPS*, pages 487–495. Curran Associates, Inc., 2014. 5

7. List of the HDR Dataset

Table 5 shows the list of URLs that provide the datasets used in our paper for training and testing. Copyrights of the images are owned by the photographers and the authors of the websites. HDRIhaven and HDReye are only used for testing in our experiments. Please note that the two links to EMPA and Pouli_hdrldr are no longer available as of submission, and for those data, we recommend readers contacting the copyright holders or us directly.

8. Additional Experimental Results

8.1. Range-wise evaluation

Figures 9– 13 are additional visualizations for the range-wise evaluation. We again compared ours with the state-of-the-art methods using deep learning, namely DrTMO [6], HDR-CNN [5] and ExpandNet [27], and the iTM methods of Huo *et al*. [14] and Kovaleski and Oliveira expansion operator (KOEO) [19]. As the figures show, the proposed method can reconstruct very bright light intensities that are clearly visible in the range [2] of 2^{-3}, and the recovered intensities are closer to the ground truths in most cases compared to the other methods.

8.2. Qualitative evaluation of the proposed method

We further show qualitative results using images in HDRIhaven [41], which includes HDR images taken with a wide range of exposures (around 20 EVs). Figures 14– 17 show the inputs, the ground truths, and the reconstructed images by the proposed method. The results show that our method can reconstruct the light intensities well, in a very high dynamic range and with complex textures.

---

[2]2^{-3} stands for the quantization resolution as explained in Fig.4 of the main text.

Table 5. URL list of datasets

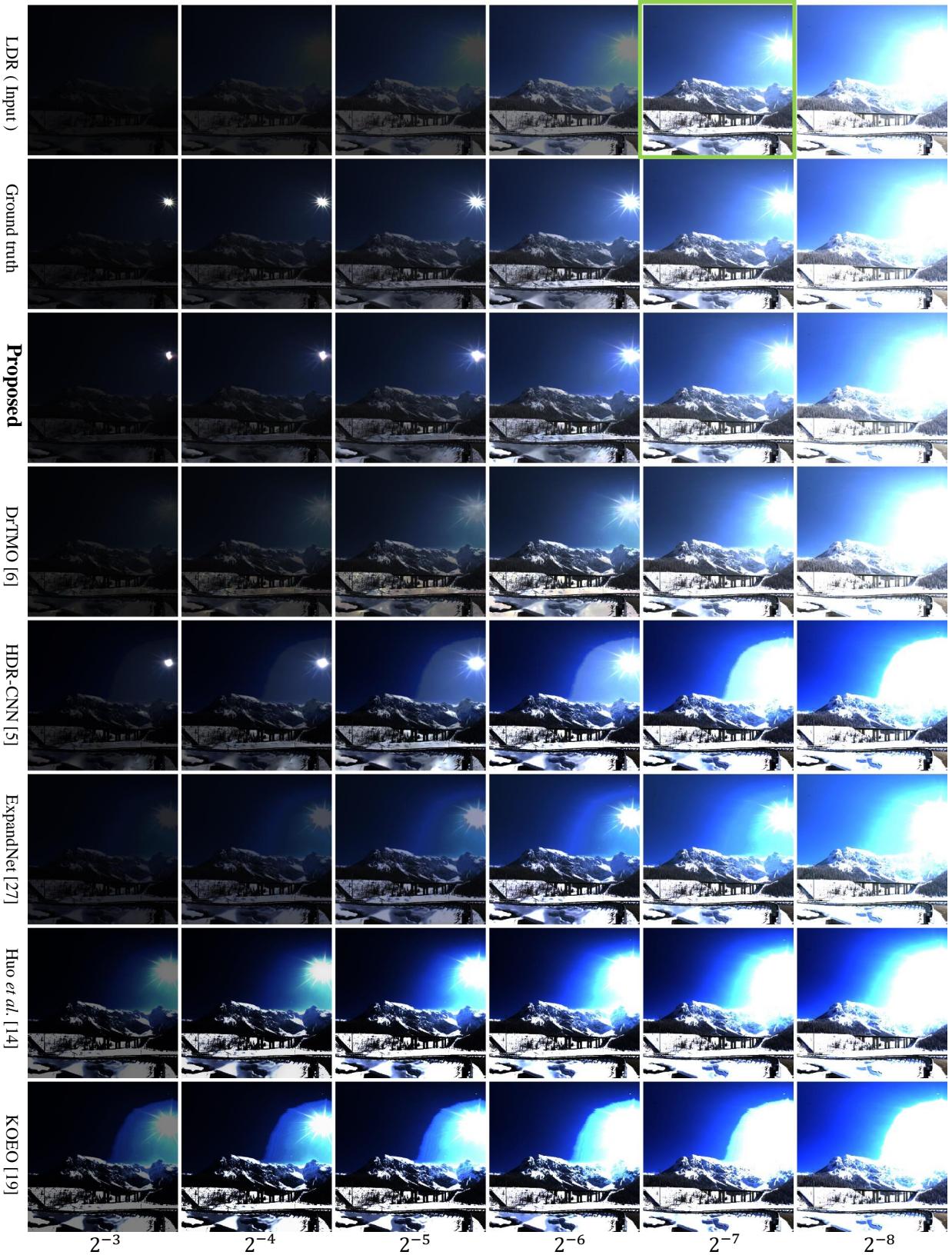|	Name	URL	Size
Image	HDRIhaven	https://hdrihaven.com/hdris/	248
	Funt	http://www.cs.sfu.ca/~colour/data/funt_hdr/#DATA	107
	Fairchild	http://rit-mcsl.org/fairchild//HDRPS/HDRthumbs.html	104
	Stanford	http://scarlet.stanford.edu/~brian/hdr/hdr.html	88
	HDRMAPS	http://hdrmaps.com/freebies	71
	HDR-dataset	http://www.hdrlabs.com/sibl/archive.html	64
	HDReye	https://mmspg.epfl.ch/hdr-eye	46
	Ward	http://www.anyhere.com/gward/hdrenc/pages/originals.html	33
	Freeskies	https://joost3d.com/hdris/	23
	Noemotion	http://noemotionhdrs.net/hdrother.html	21
	BOCO	http://bocostudio.com/boco-pano/	15
	Dutch_360	https://www.dutch360hdr.com/shop/product-category/free-360-hdri/	14
	Openfootage	http://www.openfootage.net/category/high-dynamic-range-panorama/hdris-with-a-much-higher-dynamic-range/	14
	HDRI_hub	https://www.hdri-hub.com/hdrishop/freesamples/freehdri/item/323-hdr-city-road-night-lights-free	11
	Viz people	https://www.viz-people.com/portfolio/free-hdri-maps/	10
	pfstools	http://pfstools.sourceforge.net/hdr_gallery.html	9
	Giantcow	http://giantcowfilms.com/2015/11/23/hdr-morning-sun-winter/	4
	HDRishop	https://www.hdrishop.com/collections/free-hdris/products/free-bathroom-hdri	3
	Dylan sisson	http://www.dylansisson.com/project/panoramas/	2
	EMPA	http://www.empamedia.ethz.ch/hdrdatabase/index.php	33
	Pouli_hdrldr	Statistical Regularities in Low and High Dynamic Range Images by Pouli et al. [2010]	327
Video	Stuttgart	https://hdr-2014.hdm-stuttgart.de/	33
	DML-HDR	http://dml.ece.ubc.ca/data/DML-HDR/	10
	LiU HDRV	http://hdrv.org/Resources.php	10
	Boltard	https://people.irisa.fr/Ronan.Boitard/	7
	MPI	http://resources.mpi-inf.mpg.de/hdr/video/	1

Figure 9. The top row shows multiple exposure levels of an LDR image (the original LDR is highlighted with a green rectangle), where each column corresponds to the levels from 2^{-8} to 2^{-3}. The second row shows the ground truth, and the rest are the HDR images reconstructed by each method, given the LDR image as an input. The light intensity of the sun is partly recovered by the proposed method.
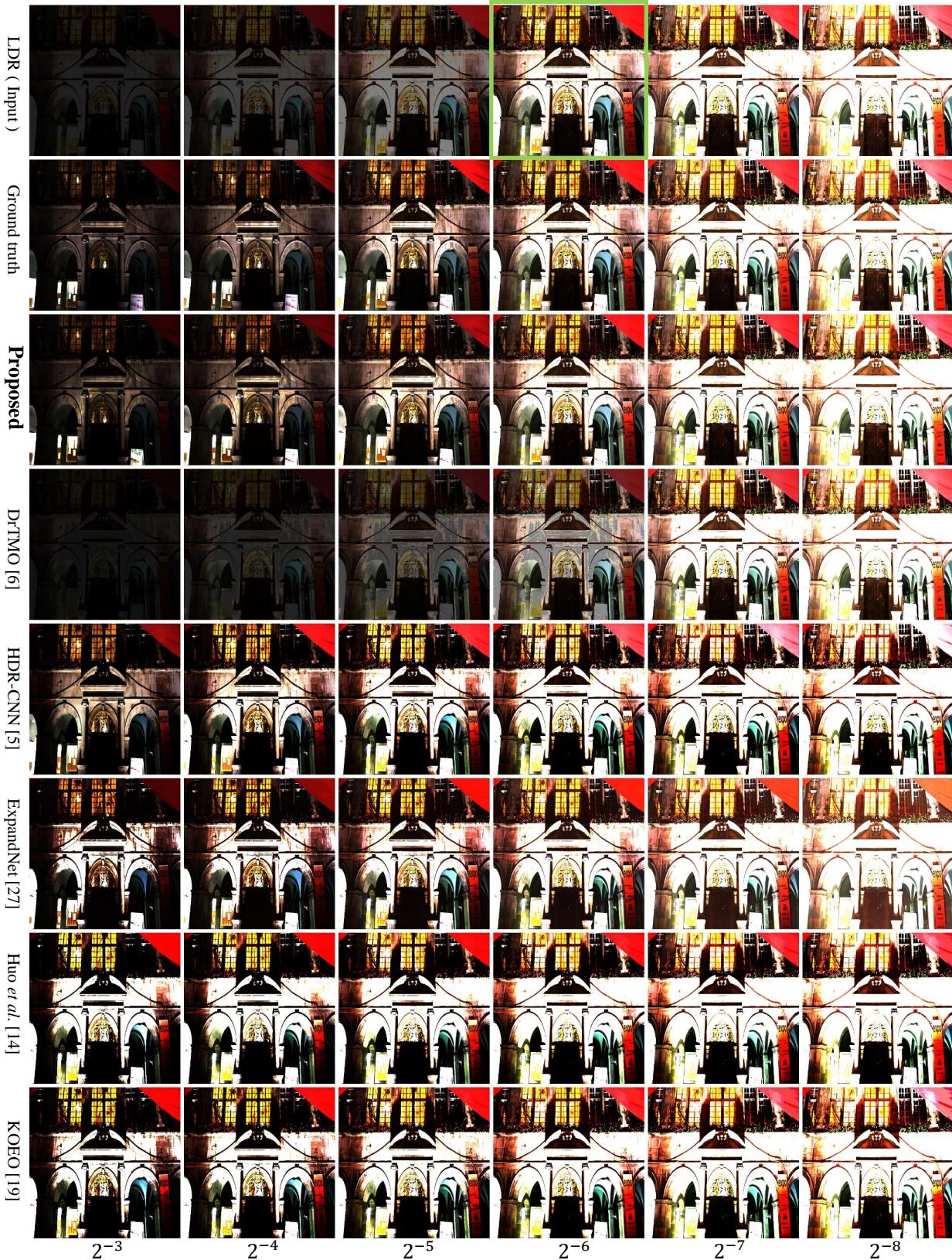
Figure 10. An example of night views of classic buildings. The proposed method can recover the indoor lighting and the color of the building.

Figure 11. An example of sunset scenes. The proposed method successfully recovers the texture of the clouds, and the luminance of the sun.

Figure 12. An example of sunny outdoor scenes. The proposed method can recover the sky, the clouds, and the color of the walls.

Figure 13. An example of indoor scenes with windows. The proposed method recovers the intensities of the sky, the trees, the person at the back, and some color of the road.

Figure 14. 'Colosseum' from the HDRIhaven dataset. The top row shows multiple exposure levels of an LDR image (the original LDR is highlighted with a green rectangle), where each column corresponds to the levels from 2^{-5} to 2^2. The second row shows the ground truth, and the third row shows the HDR images reconstructed by the proposed method. The images in the range of 2^0 are highlighted with a blue rectangle and the fourth row shows the zoomed images of them. The regions highlighted with red rectangles are further zoomed and shown in the bottom row. The proposed method recovers the light intensities that is still observable in the range 64 times higher than the range of the original image.
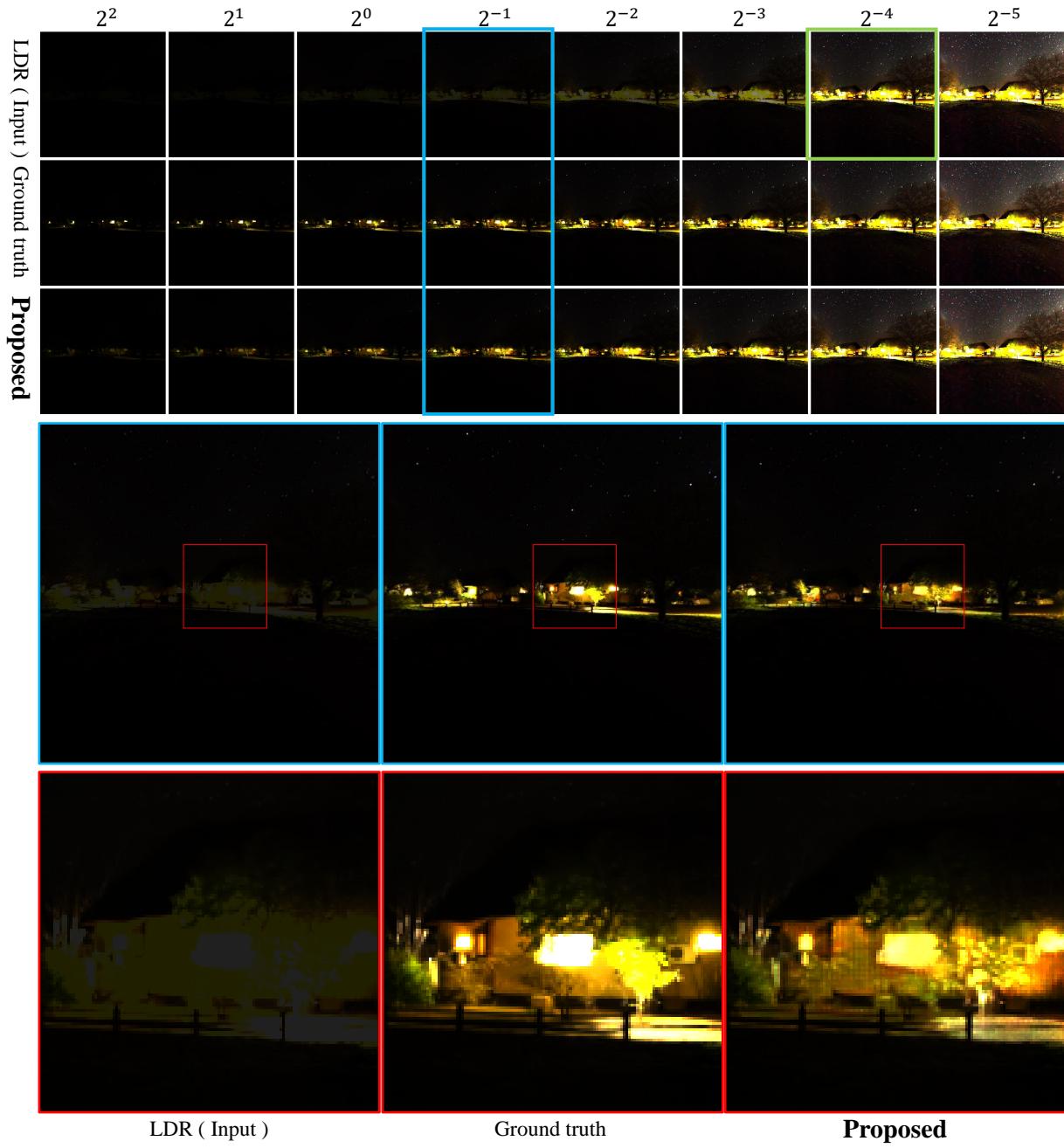
Figure 15. 'Satara night.' The proposed method recovers the intensities of not only the light sources but also non-illuminating objects such as trees and houses around the light sources.
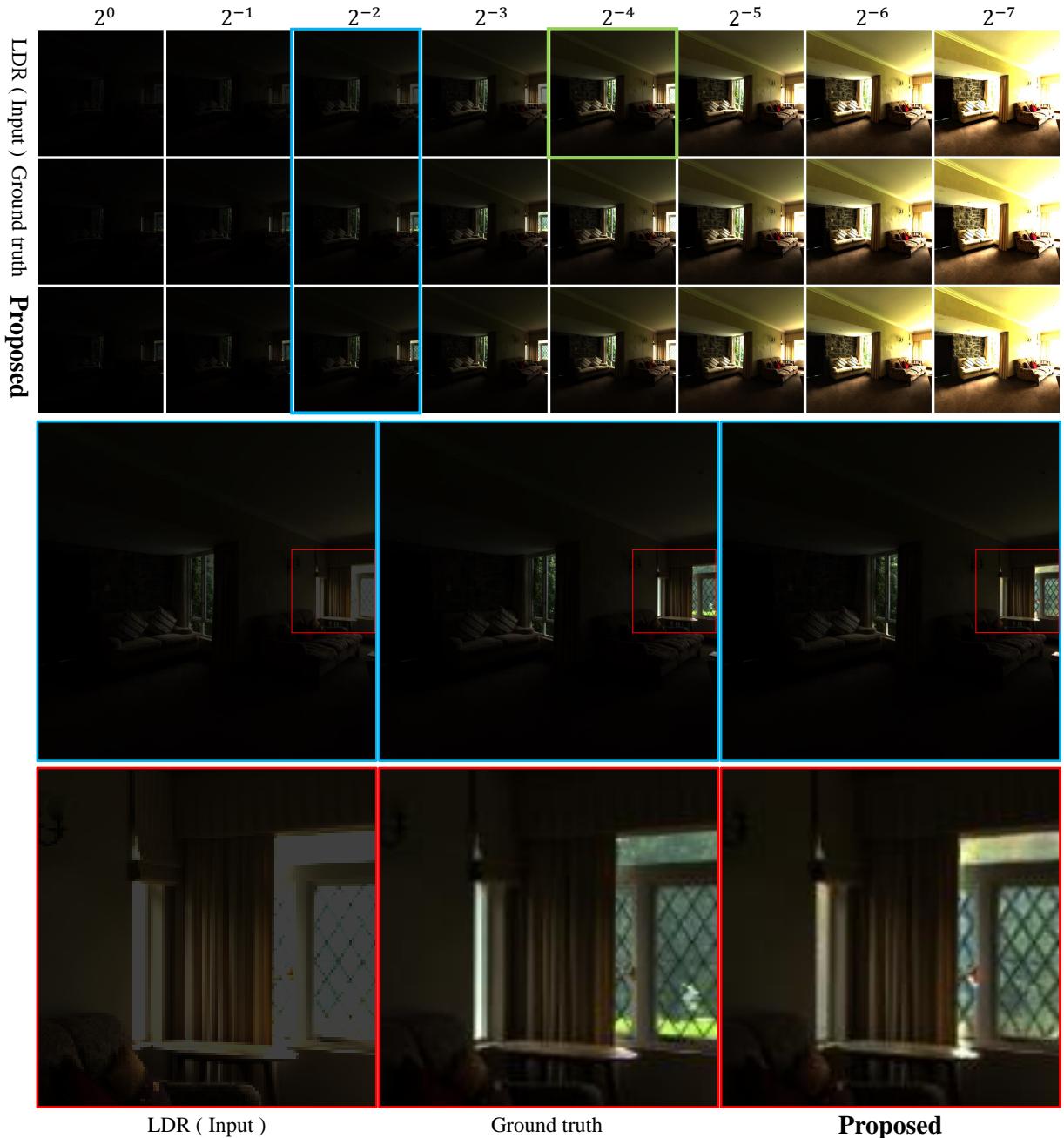
Figure 16. 'Lythwood lounge.' The proposed method inpainted the trees outside the window, which are totally lost in the LDR image.
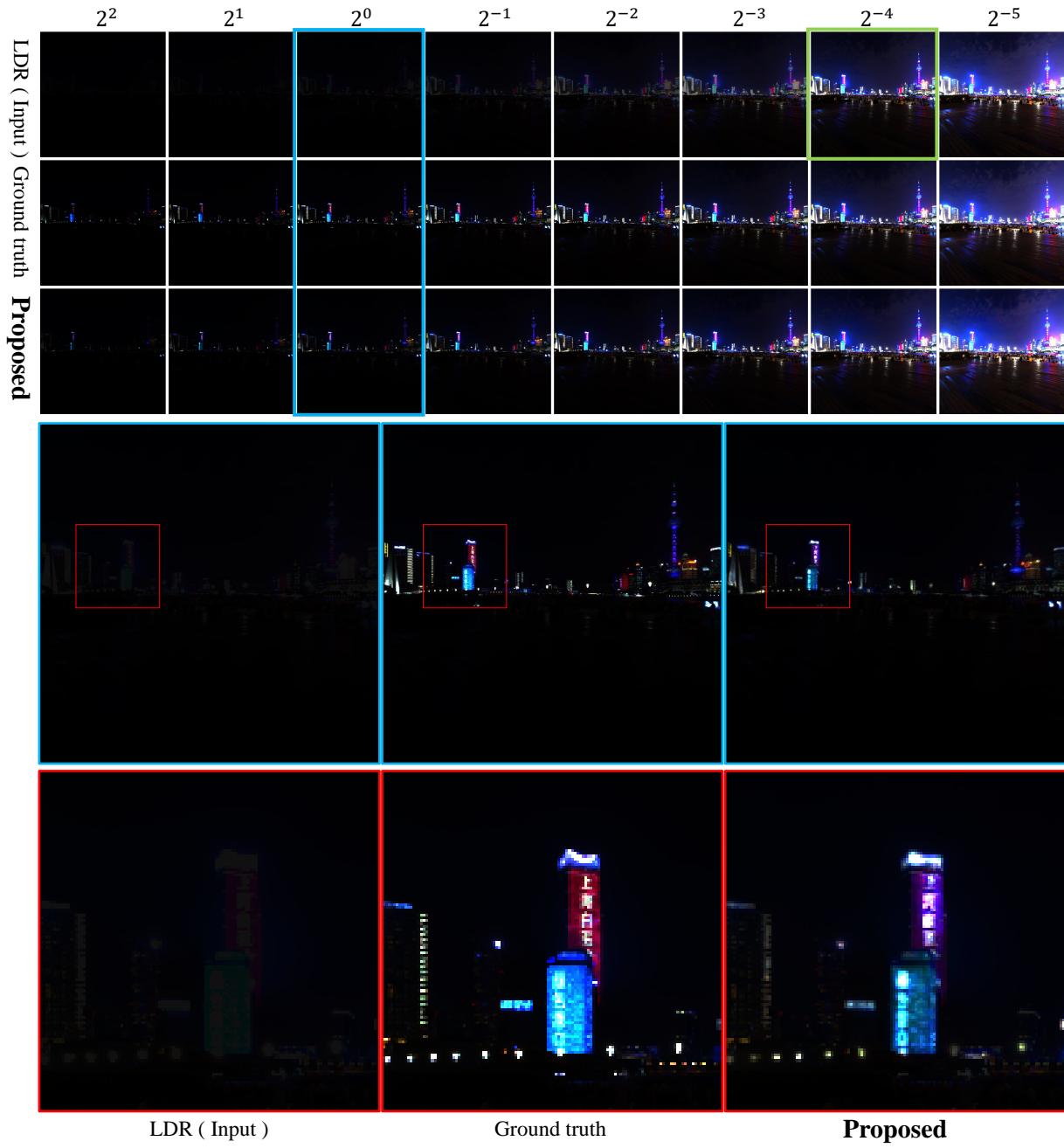
Figure 17. 'Shanghai bund.' The proposed method recovers the buildings' red and blue light-ups. The hue of the red color is slightly shifted, but visually the restoration is natural.