Received 6 February 2023, accepted 13 February 2023, date of publication 15 February 2023, date of current version 23 February 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3245802

RESEARCH ARTICLE

HDR Map Reconstruction From a Single LDR Sky Panoramic Image for Outdoor Illumination Estimation

GYEONGIK SHIN, KYEONGMIN YU, MPABULUNGI MARK, AND HYUNKI HONG
College of Software, Chung-Ang University, Dongjak, Seoul 06973, South Korea

Corresponding author: Hyunki Hong (honghk@cau.ac.kr)

This work was supported in part by the Ministry of Culture, Sports and Tourism and Korea Creative Content Agency, under Project R2021040032; and in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education under Grant NRF-2018R1D1A1B07049932.

ABSTRACT Outdoor low dynamic range (LDR) panoramic images that contain the sun and sky are generally over-saturated because the sun is 10,000 times brighter than the regions that surround it. Because the luminance information in the region that contains the sun in these images is lost, it is difficult to identify the sun's position and generate high dynamic range (HDR) environment maps that can be used to realistically relight virtual objects. Previous methods to reconstruct HDR maps did not consider that the sun covers a small area in an image but contains extremely high luminance values. These methods are therefore insufficient for estimating scene illumination. We propose a multi-faceted approach to reconstructing HDR maps from a single LDR sky panoramic image that considers the sun and sky regions separately. We encode an input image and transfer a multi-dimensional latent representation to two decoders, which reconstruct the luminance information in the sky and sun regions separately. To plausibly model sun illumination, we introduce two networks (Sunpose-net and Sunrad-net) that estimate the position and radiance of the sun. The generated sun radiance map is then merged with the output of the decoder that is responsible for sun regions. We demonstrated that the proposed method more plausibly reconstructs HDR maps than previous methods using the HDR-VDP-2.2 which measures the visual quality of reconstructed HDR maps against ground truth. The accuracy of the overall sun and sky illumination distribution in HDR maps reconstructed using the proposed method was evaluated using histogram distance measures.

INDEX TERMS Sky appearance model, high dynamic range map, deep learning model, sun illumination estimation.

I. INTRODUCTION

In augmented reality (AR), to realistically render an image of a virtual object, it is necessary to simultaneously consider geometric (occlusion based on camera viewpoint) and lighting (effect of light sources on the scene: shadows and shading regions) components. More specifically, the position and luminance of the sun in outdoor environment affects the shadow shape, sharpness, and direction in the rendering image. Though many studies that consider these components have been conducted, there are still several challenges to be addressed regarding lighting estimation for outdoor scenes [1], [2] particularly under various weather conditions. This

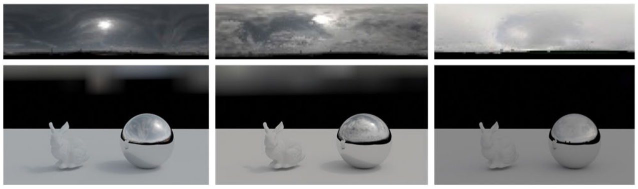

FIGURE 1. HDR maps (top) reconstructed using the proposed method under three weather conditions (left: clear, mid: cloudy, right: overcast); the images rendered using the reconstructed HDR maps (bottom).

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

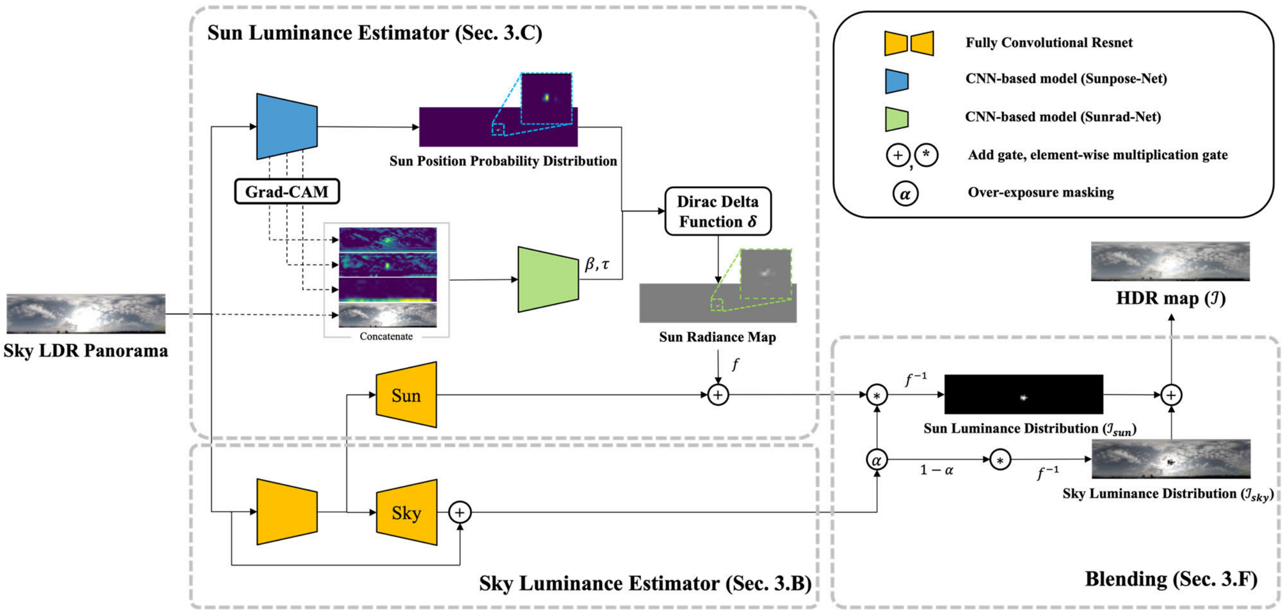
Sun Luminance Estimator (Sec. 3.C)
Fully Convolutional Resnet
CNN-based model (Sunpose-Net)
CNN-based model (Sunrad-Net)
Add gate, element-wise multiplication gate
Over-exposure masking
Sun Position Probability Distribution
Grad-CAM
Dirac Delta Function δ
β, τ
Concatenate
Sky LDR Panorama
Sun Radiance Map
f
HDR map (ℐ)
Sun
f⁻¹
Sun Luminance Distribution (ℐ_sun)
1 − α
f⁻¹
Sky Luminance Distribution (ℐ_sky)
Sky
Sky Luminance Estimator (Sec. 3.B)
Blending (Sec. 3.F)

FIGURE 2. An overview of the proposed method.

is because it is difficult to consider the extremely high luminance of the sun, that causes the surrounding image region to be over-exposed as well as the saturation caused by heavy cloud-cover occluded the sun. The goal of this paper is to facilitate more precise lighting estimation for outdoor environments by taking the presence of the sun and the other regions of the sky panoramic image into account separately. HDR maps were generated using the proposed method and used to render the images shown in Figure 1.

The parameters of models for the appearance of the sky have been estimated, using an analysis of physical phenomena, to simulate the luminance distribution of the sun and sky [3], [4], [5], [6]. This approach enables us to set up the rendering environment for an outdoor scene. Sky appearance models are, however, unable to model complex weather conditions, such as cloudy, partially cloudy, and overcast. Learning-based approaches that retrieve sky parameters from low dynamic range (LDR) images [7], [8], [9], [10], [11] inherit all the limitations of physically-based analytical models.

In rendering engines, real-world lighting information is directly represented using high dynamic range (HDR) images in the form of HDR environment maps, thereby avoiding the challenges associated with estimating sky parameters. However, the input sky panoramic images are generally saturated because the region that contains the sun is 10,000 times brighter than the regions that surround it. Much of the luminance information in the sky regions surrounding the sun is thus lost. Previous methods to reconstruct HDR maps [12], [13], [14], [15], [16], [17], [18], [19], [20] did not consider the presence of the sun, which covers a small area in the image but contains extremely high luminance values, and are therefore insufficient for estimating scene illumination.

In general, HDR maps can be generated by merging multi-exposure (bracketed) LDR images [12], [13], [14]. To create HDR images by merging bracketed images, the scene elements to be synthesized must be static. In order to solve this problem, studies on HDR reconstruction from a single LDR image are being widely conducted [15], [16], [17], [18], [19], [20]. In order to more accurately estimate illumination of an outdoor scene, we reconstructed an HDR map from an LDR image, while separately taking the position and luminance distribution of the sun and sky into account. The light from a distant object like the sun is scattered due to the participating media's transmittance [4], [6]. The light incident (illuminance), $\mathcal{I}$, on a virtual object and observed from a camera is modeled as a combination of the light ($\mathcal{I}_{sun}$) from the sun and the luminance of the sky $\mathcal{I}_{sky}$, as shown in equation (1).

$$\mathcal{I} = \mathcal{I}_{sun} \cdot \tau + I_{sky}, \tag{1}$$

where τ describes the participating media's transmittance. Based on equation (1), we introduce a method to reconstruct the luminance distribution in the sun and sky regions separately and then merge them in an HDR map. $\mathcal{I}_{sun}$ is generated by masking the non-sun regions of the output image of the sun luminance estimator (Sec.III-C). $\mathcal{I}_{sky}$ is generated by masking the sun regions of the output image of the sky luminance estimator (Sec.III-B). $\mathcal{I}_{sun}$ and $\mathcal{I}_{sky}$ applied an inverse tone mapping operation and then blended to generate the final output HDR map $\mathcal{I}$.

Our method encodes the sky information from an input sky panoramic image and transfers the information to two decoders. The two decoders reconstruct the luminance information in the sky and the sun regions separately. To regulate the range of the illumination, the decoder responsible for the sky takes the residuals of the input image and the ground truth

HDR map and combines them with the input LDR image. The decoder for the sun considers the range of luminance distribution in the regions near the sun. Because the sun illumination is lost due to the over-exposure in the input images, two deep learning networks (Sunpose-Net and Sunrad-Net) are used to estimate the sun position and the parameters of the Dirac delta function.

The sun position as well as the estimated parameters (β, τ) are fed into the Dirac delta function to generate a sun radiance map. The output illuminance distributions of these two modules are upscaled and softly blended into a single HDR map. To evaluate our method, HDR images from the Laval HDR sky database are transformed into LDR images using four tone mapping operators (TMO). Figure 2 shows an overview of the proposed method. In this figure, f is a tone mapping operation defined by equation (2)

$$f(x) = \log(1 + vx) / \log(1 + v), \tag{2}$$

where v is the maximum expected luminance value (5000). x is the pixel value (luminance) in the HDR map. f^{-1} represents the upscale (inverse tone mapping) operation.

The simulation results show that our method more accurately reconstructs HDR maps of the sky than previous methods do in terms of HDR-VDP-2.2, peak signal-to-noise ratio (PSNR), structural similarity index measure (SSIM) and histogram distance measures. We compared the rendering results of our method against those of previous methods for outdoor scenes under various weather conditions.

The main contributions of this manuscript are highlighted as follows.

- A multi-faceted HDR reconstruction approach that separately considers the luminance distribution of the sun and that of the sky in outdoor environments.
    - ○ An encoder that generates a multi-dimensional latent representation for the luminance distribution in an input LDR sky panoramic image.
    - ○ Two decoders that consider the heterogeneous luminance properties of the sky and the sun region separately.
- A network that estimates the parameters of the Dirac delta function to retrieve the luminance distribution of the sun which covers a small area in HDR maps but has extremely high luminance values relative to its surroundings.
- A dataset (CAU) containing 505 LDR images captured under multiple exposures and their corresponding 101 HDR images. It is available at https://github.com/ShinYwings/HDR-Map-Reconstruction-From-a-Single-LDR-Sky-Panoramic-Image-For-Outdoor-Illumination-Estimation.

II. RELATED WORK

A. SKY APPEARANCE MODELING

Physically based analytical methods estimate the luminance distribution of the sun and sky using explicit parametric equations [3], [4], [5], [6]. These methods can generate the sky appearance for any given input time, date, or location. The Perez model [3] predicted a sky luminance distribution using five parameters based on global and direct irradiances, building on the 1993 CIE sky model. However, because this model only accounts for luminosity using five parameters, it can only simulate skies under limited weather conditions. The Preetham model [4] improved on the Perez model by adding chromaticity considerations to generate a greater number of sky conditions. The model also proposed a set of linear functions that allow for the coefficients in the Perez model to be retrieved for various levels of turbidity. The Hosek-Wilkie model [5] generated an even more realistic sky appearance for a wider range of turbidity values and low solar altitude (sunset and sunrise) scenes by employing a brute force path tracer to fit the collected reference data. Hosek and Wilkie later used Monte-Carlo simulation and Rayleigh scattering to model solar radiance [6]. Analytical approaches based on physical models are computationally inexpensive and do not require a large dataset. However, they focused on the clear sky and were unable to generate the illumination distribution for the cloudy weather. This is because the limited number of parameters used by analytical models is insufficient for modelling more complex sky conditions, such as partially cloudy and overcast. In addition, the parameters that simulate a real-world sky are retrieved through a time-consuming fine-tuning process.

Learning based methods [7], [8], [9], [10], [11] have focused on estimating the illumination of an outdoor scene from a single input image, rather than generating the sky appearance. Unlike for the analytical methods, manually tuning parameters is not necessary for this process. To account for variations in sun intensity in complex situations like cloudy weather, Matthews [7] added the sun, which inolves an exponential drop-off in the log-intensity domain, to the Preetham model. He also introduced a regression model to consider the dispersion of sunlight that is attenuated by the atmospheric transmission conditions and cloud coverage. Hold-Geoffroy et al. [8] introduced a CNN-based method to reconstruct an outdoor illumination environment map by explicitly predicting the Hosek-Wilkie model's parameters from a single outdoor LDR image of a general scene. This approach inherits the Hosek-Wilkie analytical model's inability to represent complicated sky conditions.

Hold-Geoffroy et al. [10] estimated illumination conditions by implicitly learning the parameters of a low-dimensional, physically-based model. Specifically, he tried to represent sky features as a 64-element latent vector, but this vector was not sufficient to represent the luminance distribution of the sun and various weather conditions. This procedure is performed by mapping the sky appearance in the image to the scene illumination.

To address this problem, Yu et al. [11] developed three auto-encoders that represent sky and sun conditions as two separate vectors for accurate outdoor illumination prediction. Because the encoders were designed to consider the sun and sky intensities separately, the dataset was separated into clear and cloudy weather conditions to train the encoders. Furthermore, the above-mentioned two methods [10] and [11] require user-guided post-processing to compensate for any inaccuracies in the final environment maps.

To account for the variations in light incident on a virtual object caused by changes in spatial properties, [1] and [2] combined the reconstructed environment maps with warped image information according to the geometrical information estimated from intrinsics. Reference [1] considered shadow, depth, normal, and albedo and [2] only considered depth as intrinsic parameters. The HDR enviroment maps are reconstructed using the methods in [20] and [10] for [1] and [2], respectively. Since both methods compressed the environment map into a single vector, they were unable to sufficiently represent the luminance distribution of the sun under various weather conditions.

Zhang and Lalonde [20] proposed a model that learns an inverse tone mapping of LDR sky panoramic images, to HDR sky panoramic images in outdoor scenes. This study employed an encoder–decoder structure to reconstruct HDR images and estimate sun elevation. In general, PSNR, SSIM, and HDR-VDP-2.2 measures are used for HDR image quality assessment [21]. However, in this study, HDR image quality was only evaluated using the mean absolute error between generated and ground truth HDR maps. Another limitation of this approach is that the output resolution is limited to 64 × 128, implying that HDR information in a full resolution LDR background image cannot be extrapolated.

In [22], an HDR environment map, which is represented using spherical harmonics, is estimated from a single LDR monocular spherical panoramic image. This study relied on a global Lambertian assumption of the scene and required its normal map. In addition, only the low-frequency environment map estimates were offered because it regresses up to the third order spherical harmonics coefficients. This implies that the HDR map generated cannot be used to accurately simulate the effects (shadow shape, sharpness, and direction) of the sun in an outdoor scene.

B. HDR RECONSTRUCTION

HDR images can store a wide range of luminance values and are therefore well suited to representing illumination in a real-world sky. It is possible to generate an HDR image from an LDR image using an inverse camera response function (CRF), but the CRF may not always be accessible. In much earlier approaches, HDR images were generated by fusing bracketed exposure LDR images into a single HDR image [12], [13], [14] which may not recover any missing details in under or over-exposed image regions. The need for bracketed (multi-exposed) images limits the use of images from commonly accessible sources like the web, where camera exposure setting information is unknown. Also, ghosting artifacts commonly occurs when bracketed images of dynamic scenes are combined [12].

Many deep learning approaches [15], [16], [17], [18], [19], [23] that perform a direct single LDR to HDR mapping overcome these challenges. Endo et al. [16] developed an encoder-decoder network to infer up/down-exposed (bracketed) images from a single LDR image. The bracketed images were then merged to reconstruct an HDR map using Debevec's method [12]. The number of over/under-exposed images affect much of the HDR reconstruction performance, but an increase in the number of inferred images is computationally expensive in terms of memory and processing. This model is not suited to highly dynamic scenes, such as those containing the sky, because it is trained on an image dataset with a limited exposure range. It does not adequately reconstruct outdoor scenes with the highly dynamic range [16].

More recent approaches employ a direct single LDR-to-HDR conversion with an auto encoder-based model [17], [18], [19], [23]. Eilertsen et al. [17] reconstructed an HDR image from a single exposure by employing an encoder-decoder framework with skip connections to recover details in over-exposed image regions. This approach generalized well to various illumination conditions, however, the use of a fixed inverse CRF such as gamma $g^{-1}(\mathrm{x}) = \mathrm{x}^{\gamma}$ for linearization makes the approach unsuitable for images captured by the camera with dissimilar exposure values.

Liu et al. [18] modeled the HDR-to-LDR image formation pipeline as dynamic range clipping, non-linear mapping from a CRF, and quantization. A couple of U-Nets and CNNs are used to reverse each of these steps while imposing effective physical constraints to facilitate the training of the individual subnetworks. Liu used a separate network to estimate the CRF as a linear combination of PCA basis vectors, as stipulated by the empirical camera models [13]. Yang and Aydın [19] employed two U-Net-based modules to recover low-frequency components and hallucinate the image details clipped due to quantization as well as insufficient dynamic range. However, when the input image is severely clipped, this method tends to suffer from banding effects and color shift artifacts in over-exposed regions. The U-Net structure is generally well suited for image detail reconstruction but completely unable to recover lost details in large over-exposed regions.

C. IMAGE-TO-IMAGE TRANSLATION

Image-to-image translation converts an input image (source domain) to the transferred style image (target domain) while preserving the content representations [24], [25], [26], [27], [28]. A mapping between the source and target domains is learned through training. Isola et al. [24] proposed a method based on conditional generative adversarial network (GAN) [25] to synthesize photos from label maps, reconstruct objects from edge maps, and colorize images. This approach used U-Net for its generator and a Markovian-based method (PatchGAN) for its discriminator. The key limitation of this supervised approach is the need for a source-to-target paired dataset in the training procedure. Unsupervised image-to-image translation approaches overcome this limitation by employing a cycle-consistency constraint [26], [27], [28]

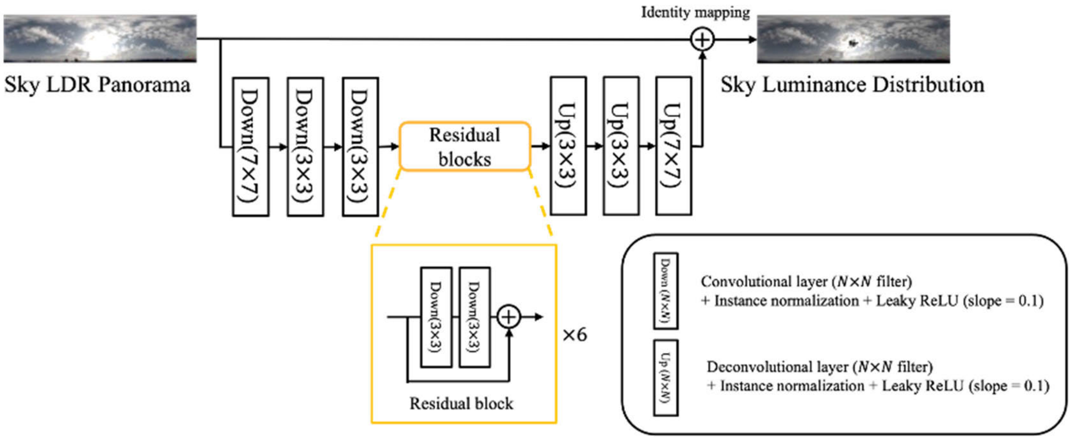
Identity mapping
Sky LDR Panorama
Sky Luminance Distribution
Down(7×7)
Down(3×3)
Down(3×3)
Residual blocks
Up(3×3)
Up(3×3)
Up(7×7)
Down(3×3)
Down(3×3)
×6
Residual block
Down (N×N)
Convolutional layer (N×N filter)
+ Instance normalization + Leaky ReLU (slope = 0.1)
Up (N×N)
Deconvolutional layer (N×N filter)
+ Instance normalization + Leaky ReLU (slope = 0.1)

FIGURE 3. Architecture of the sky luminance estimator.

to learn mapping functions between the source and target domains.

D. CLASSIFICATION ACTIVATION MAP (CAM)

By employing global average pooling (GAP), a CNN trained for classification can be used for object localization without bounding box annotations [29], [30]. This approach identifies exactly which regions of an image are being used for discrimination, but only achieves class-specific feature maps and requires the fully connected layers of the network to be replaced with convolutional layers and GAP [29].

Grad-CAM [30] addressed these issues by employing gradient-based localization to generate saliency maps, showing the location and intensity of the regions of interest in an image for a wide range of CNN-based classification models. Grad-CAM used the gradient information flowing into the last convolutional layer of a CNN to understand the importance of each neuron for a given classification procedure [30]. CAM needs GAP as a last layer, but the Grad-CAM does not require further modification of classification models. In this study, the Grad-CAM is used to identify the regions of the image that are necessary for estimating the position of the sun and its luminance distribution under various weather conditions.

III. METHODOLOGY

A. OVERVIEW

Our goal is to reconstruct an HDR map from a single LDR sky panoramic image. Panoramic images are generally generated using multiple images captured by the camera with a limited field of view. In this procedure, the image pixels from multiple consumer cameras are interpolated, so the pixel values are not explained by a single CRF [31]. Instead, we consider the HDR reconstruction as an image-to-image translation problem for which a GAN-based architecture is well suited. By regarding the HDR map as the output of the generator, an adversarial approach can be used to train both generator and discriminator.

U-Net and encoder-decoder architectures have been widely used in image-to-image translation [24], [25], [26], [27], [28]. The U-Net down-samples input images into a single latent vector, which cannot sufficiently represent the high-frequency components in the luminance distribution of sky images. We employ a modified encoder-decoder architecture that generates a high-dimensional latent representation to sufficiently include the coherence of the sun region and sky as well as the luminance gradation in the sky region. Because the sky image intensities are heavily influenced by the high luminance value of the sun, we employ a single encoder to build a representation of the entire luminance distribution of the sky images. The sky region includes a lot of fine textures, and it has a wider coverage, and much lower luminance distribution than the sun region does. The sun region covers a smaller area but on the log scale, contains luminance values that are five orders of magnitude higher in the sky regions. Therefore, we design and implement two decoders to consider the heterogeneous luminance properties of the sky and the sun region separately. In this paper, the decoder for the sky region is called the sky decoder and that for the sun region is the sun decoder. Separately computed luminance distributions of the sun and the sky are combined to reconstruct an HDR map. A PatchGAN discriminator architecture is adopted because of its ability to model high-frequency components in the image.

B. SKY LUMINANCE ESTIMATOR

The sky luminance estimator (Figure 3) reconstructs the luminance distribution of the sky regions without the over-exposed sun regions of an input sky LDR panoramic image. A latent representation of the input LDR image is generated using an encoder. Then, the latent representation is decoded to generate the sky luminance distribution, which is identity-mapped to the original input sky LDR panoramic image. Additionally, the input LDR image is passed to each of the residual blocks.

Our encoder-decoder architecture is based on the fully convolutional ResNet [32] but uses a modified activation

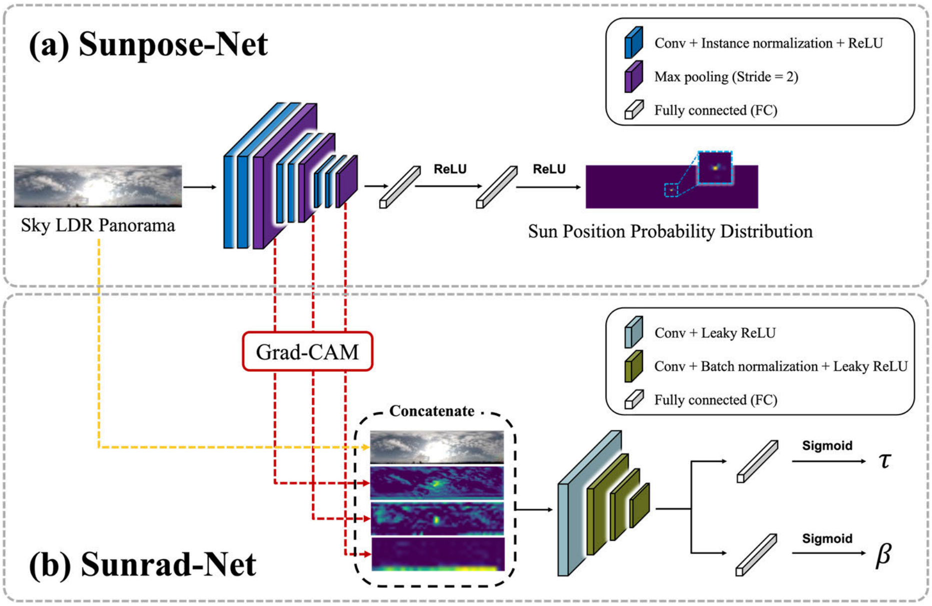

FIGURE 4. Architecture of the (a) Sunrad-Net and (b) Sunpose-Net. For each block in (a), Grad-CAM is used to generate an attention map. The LDR sky panorama is concatenated to each attention map and passed to the Sunrad-Net as input.

function. The encoder down-samples the input image into a multi-dimensional latent vector to represent the overall luminance distribution of the sky panoramic image. The down-sampling layers consist of a 7 × 7 convolutional layer with a stride of 1 and two 3 × 3 convolutional layers with a stride of 2. After each convolutional layer, instance normalization and leaky ReLU activation with a slope of 0.1 are applied. The decoder enables the latent representation to be used to linearly expand the dynamic range in the sky image regions that do not contain the sun. The up-sampling layers (decoder) consist of two 3 × 3 deconvolutional layers with a stride of 2 and a 7 × 7 deconvolutional layer with a stride of 1. After each deconvolutional layer, instance normalization and leaky ReLU activation are applied.

Between the encoder and decoder, there are six residual blocks, each with five operations: a 3 × 3 convolution with a stride of 1, instance normalization, leaky ReLU, a 3 × 3 convolution with a stride of 1, and instance normalization. In each residual block, an input is added to the output of the residual block. By adding the input LDR image intensity to the result of the sky decoder and applying a ReLU operation to clip the negative values, we ensure that the global luminance of the sky region is preserved and avoid sudden variations near over-exposed regions in the training process. The identity mapping is added in the encoder-decoder architecture to learn residual information in the LDR images, as shown in Figure 3. The output of the sky luminance estimator is then softly blended with the sun luminance distribution (Sub-Section III-F). These successive optimizations allow for an even more stable estimation of the illumination distribution of the sky without estimating the CRF.

C. SUN LUMINANCE ESTIMATOR

The sun luminance estimator recovers the luminance distribution of the sun regions, which are highly over-exposed in an LDR image. This module consists of a decoder and two networks: Sunpose-Net estimates the position of the sun in the sky panoramic image, and Sunrad-Net obtains the parameters of the Dirac delta function, which is used to generate the luminance distribution of the sun.

The decoder of the sun luminance estimator uses the multi-dimensional latent representation generated by the encoder to reconstruct the overall luminance distribution for over-exposed regions that contain and surround the sun. The decoder in this module is identical in structure to that in the sky luminance estimator. It is, however, unable to fully recover the illumination distribution of the sun. To additionally consider the extremely high luminance of the sun, we employ the Dirac delta function with three parameters: sharpness (β), transmittance scale (τ) and sun position (x). Here, the sun position is estimated using the Sunpose-Net (in Sub-Section III-D).

The Dirac delta function can generate the sun radiance map for various weather conditions. More specifically, this function can generate various levels of sharpness for the sun regions and maintains the total energy of the radiance map. A map generated using the Dirac delta function spikes as the β value increases and spreads out as it decreases.

In equation (3), τ represents the transmittance value of the sky. The Sunrad-Net estimates these two parameters to simulate the sharpness and the intensity of the sun radiance map under various atmospheric conditions.

$$\delta\left(x_{i,j},\tau,\beta\right)=\frac{\tau}{\beta\sqrt{\pi}}\exp\left(-\frac{\left(1-x_{i,j}\right)^{2}}{\beta}\right), \tag{3}$$

where $x_{i,j}$ is the probability that the sun is present at a given pixel (i, j).

D. SUNPOSE-NET

The Sunpose-Net generates a likelihood map that illustrates the probability that each pixel (i, j) in the input LDR sky panoramic image contains the sun. In order to improve performance in terms of convergence time and accuracy, the Sunpose-Net is pretrained for 1,000 epochs using the Adam optimizer at a learning rate of 1e-04. The Kullback-Leibler (KL) divergence and difference-of-Gaussian (DoG) loss functions that are used in pre-training KL divergence are used to compare predicted and ground truth probability distributions. The DoG loss function, originally designed in the thermal diffusion equation, encourages the network to better represent the gradation (high frequency components) of the luminance distribution of the sun. The DoG loss function is further described in Sub-Section III-G.

The ground truth sun position (zenith angle and azimuth angle) labels are converted to a probability distribution using the von Mises distribution as proposed by Hold-Geoffroy [8]. The numerical sun position is expressed as a probability distribution modeled from the circular normal distribution (von Mises distribution). This representation contains spatial information and achieves better sun position prediction results than that in the numerical representation.

The Sunpose-Net (Figure 4(a)) employs a modified VGG16-Net [32] with three CNN blocks and two affine transform layers. Each CNN block consists of two sub-layers and max pooling. The first CNN block consists of two sub-layers that each have a 7 × 7 convolution filter with a stride of 1. The remaining two CNN blocks consist of two sub-layers that each have a 3 × 3 convolution filter with a stride of 1. Each of the sub-layers is followed by an instance normalization layer and a ReLU activation function. The three CNN blocks output feature maps with 32, 64, and 128 channels, respectively. Each affine layer has a fully connected layer and a ReLU activation function. The dimensions of the final output probability distribution by the Sunpose-Net are identical to the resolution of the input image. The output vector is reshaped to a two-dimensional feature map that represents the probability that each pixel in the input LDR sky panoramic image contains the sun. Instance normalization is used instead of batch normalization to preserve variations in activation maps.

E. SUNRAD-NET

To reconstruct the extremely high luminance in image regions that contain the sun, Sunrad-Net is used to estimate the parameters of the Dirac delta function. The luminance distribution of the sun is generally affected by weather conditions such as cloud coverage. To consider those variations, the Grad-CAM generates an attention map that is used as an input to the Sunrad-Net. Attention maps (Figure 5) are generated by applying Grad-CAM to the outputs of the three blocks in the Sunpose-Net. These attention maps are concatenated to the input LDR image, and then the concatenated maps are passed to the Sunrad-Net as input. As shown in Figure 5, the attention map varies with changes in weather conditions. The first row shows panoramic images of cloud cover varying from clear, as in (a), to heavily overcast skies, as in (d). The remaining rows show the corresponding attention maps generated by applying Grad-CAM at each block in the Sunpose-Net for each input image.

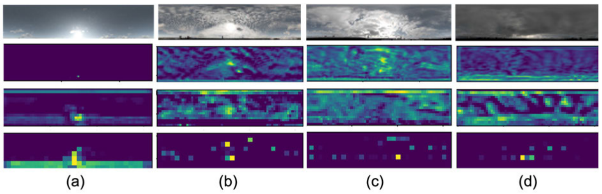

FIGURE 5. Input LDR images and corresponding attention maps generated by applying Grad-CAM at each block in the Sunpose-Net under various weather conditions.

In the Sunrad-Net, a PatchGAN architecture is adopted as shown in Figure 4(b). The Sunrad-Net consists of four sequential 4 × 4 convolutional layers with a stride of 4 and two parallel affine transform layers. As in PatchGAN, batch normalization is excluded from the first convolutional layer and each convolutional layer is followed by batch normalization and leaky ReLU activation with a slope of 0.3. convolutional layers are used to encode the sun radiance distribution of the input image, which is described by the correlated parameters τ and β. The encoded representation is flattened and passed to two separate FC layers that estimate the parameters τ and β. Here, the FC layers have the same number of nodes and apply the sigmoid activation function to generate the final parameters τ and β of the Dirac delta function as a value between 0 and 1.

Because a 0 beta value input to the Dirac delta function results in an infinite sun intensity estimation, we chose to clip the sun radiance at a value of 30,000 in accordance with NASA's tabulated data on spectral radiance [4].

F. BLENDING

The luminance distributions of the sun and the sky are computed separately to take into account the differences in the luminance properties of the sky and the sun regions. These two luminance distributions are upscaled (inverse tone-mapped) into HDR maps using equation (2) and ultimately combined into a single HDR map, which is regarded as the output of our PatchGAN-based generator, similar to the image-to-image translation approaches.

By examining the luminance information retrieved by the sky luminance estimator, we can identify which regions of the sky panoramic image contain the sun. The mask (α) representing the sun region is generated to combine the luminance distribution of the sun with that of the sky. We used a soft mask [17] to blend the illumination distributions generated by the sun and sky illumination estimators into a single HDR map without banding effects.

$$\min\left(1, \max\left(\max_c {}_c \left(D_c - (1-\gamma)/\gamma\right)\right)\right), \qquad (4)$$

where D_c is the output of the sky luminance estimator in channel c and γ is the threshold ($\gamma = 0.12$). The linear ramp for the blending starts at the threshold value and ends at the maximum pixel value. max_c is the maximum value in each channel c.

G. LOSS FUNCTION

Our objective function ($\mathcal{L}_{total}$) is defined as a linear combination of reconstruction, perceptual, DoG, and adversarial losses.

$$\mathcal{L}_{total} = \lambda_1 \mathcal{L}_{rec} + \lambda_2 \mathcal{L}_{perc} + \lambda_3 \mathcal{L}_{adv} + \lambda_4 \mathcal{L}_{DoG}, \qquad (5)$$

Reconstruction loss ($\mathcal{L}_{rec}$), the L1 norm, is used to measure the similarity between the ground truth HDR map and the reconstructed HDR map. Perceptual loss ($\mathcal{L}_{perc}$) computes the perceptual distance between the ground truth and the reconstructed HDR maps, tone-mapped using equation (2) [33]. DoG loss ($\mathcal{L}_{DoG}$) considers the difference between the high-frequency components in the ground truth HDR map and those in the reconstructed HDR map.

Because we consider HDR reconstruction as an image-to-image translation and employ a PatchGAN-based approach, the output HDR map is continuously evaluated using an adversarial loss (L_{adv}) during training. The adversarial loss is used to evaluate how closely the predicted HDR map matches the ground truth HDR map. Here, L_{adv} is computed with the least-squares measure, which stabilizes the training and alleviates the unwanted blurring effects [34]. This ensures that the predicted HDR map is visually more similar to the ground truth HDR map. In our experiment, setting $\lambda_1, \lambda_2, \lambda_3,$ *and* λ_4 to 10, 0.01, 1, and 1000, respectively, resulted in the highest performance.

In our discriminator, 4 × 16 patches in the predicted HDR map are compared with corresponding patches of the ground truth HDR map. The discriminator consists of four convolutional layers with 64, 128, 256, and 512 output channels. Each layer has 4 × 4 convolution (stride of 2), batch normalization, and leaky ReLU activation with a slope of 0.3. Batch normalization is not used in the first convolution layer as in PatchGAN [24]. An additional convolutional layer outputs a probability value describing how correct or fake a given patch is.

IV. EXPERIMENTAL RESULTS

A. TRAINING DETAILS

To evaluate our HDR reconstruction methods, we used TensorFlow 2.4 to conduct experiments on a computer equipped with an Intel® Core™ i9-10920X CPU and an Nvidia RTX 3090 graphics processing unit.

In our experiment, the Laval Sky HDR database [7] and a real image dataset (the CAU dataset) were used in the training and testing processes. The Laval Sky HDR database included 30,000 HDR sky-dome panoramic maps, at a resolution of 1024 × 2048, and their corresponding sun positions (zenith angle and azimuth angle) were provided. For the training, 24,000 HDR maps were used, and the remaining 6,000 images were used in testing. The HDR maps were resized to 32 × 128 for a reasonable trade-off between computational cost and realistic rendering [20]. The sun was horizontally centered in all panoramic images, and the corresponding sun position was converted from real-world angles to pixel positions in the panoramic image. These numerical sun positions were eventually expressed as the probability distributions. Because the focus of this study was the reconstruction of HDR maps from the LDR sky panoramic images, the image regions at a horizontal elevation that was less than 0 degrees were clipped.

In order to evaluate the generalization performance of our method, testing and training data were generated using different TMOs. The test data were generated using four different methods: Drago, Durand, Mantiuk, and Reinhard [35], [36], [37], [38], whereas the training data were generated using the CRF database-based method proposed by Liu et al. [18].

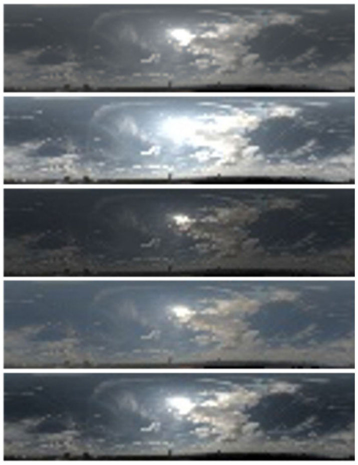

FIGURE 6. HDR input and corresponding Drago, Durand, Mantiuk, Reinhard TMO output LDR images.

Drago employed an adaptive logarithmic mapping technique to compress luminance values, which preserves image details and contrast. Reinhard, inspired by the adaptation processes in the human visual system, proposed intuitive parameters to control intensity, contrast, light, and chromatic adaptation. In Durand's method, the image was decomposed into large-scale variations and edge components using a bilateral filter, to preserve the image details. Mantiuk imposed constraints on the contrast image over a full range of spatial frequencies in order to enhance image contrast without artifacts. In Figure 6, the first image is a ground truth HDR

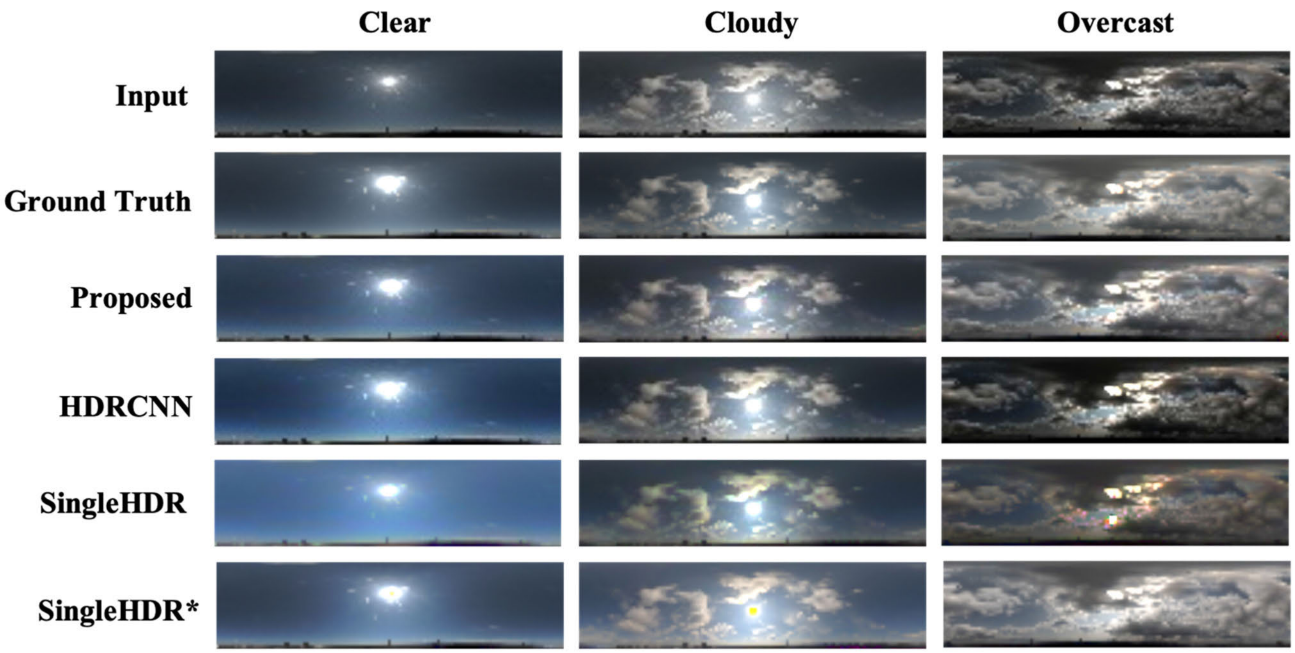

FIGURE 7. Qualitative comparison of HDR maps generated using different HDR reconstruction methods on the Laval dataset.

map, and the susbsequent images are the results found by Drago, Durand, Mantiuk, and Reinhard TMO, respectively. We generated the tone-mapped images using four methods whose parameters were randomly generated 6,000 times. We confirmed that TMOs result in color distributions similar to those in Figure 6.

TABLE 1. Hyperparameters of the compared approaches and proposed method.

Methods	Learning rate	Batch size	Blending parameter (γ)	Inverse Camera Response Function
HDRCNN [17]	5e-5	8	0.05	x^{η} ($\eta = 2$ fixed)
SingleHDR [18]	1e-4	16	0.05	Estimated using CNN [13]
SingleHDR*	1e-4	16	0.12	Estimated using CNN [13]
Proposed	1e-4	32	0.12	Sky luminance estimator (Sec. 3.B)

B. HDR RECONSTRUCTION MAP RESULTS

We compared HDR maps reconstructed using the proposed model with those reconstructed using HDRCNN [17] and SingleHDR [18] in qualitative and quantitative aspects. HDRCNN is a hybrid U-net model that estimates the over-exposure area by applying the inverse camera response function (x^{η}) to the skip connection layers of the U-net. Here, the parameter, η, was set to 2. SingleHDR is also based on the U-net model but estimates a different inverse camera response function for each LDR sky panoramic image by employing a parametric model, whose parameters are estimated using a convolutional neural network [13]. The proposed method uses a sky luminance estimator (Sec. III-B) for estimating inverse camera response functions. In the case of SingleHDR, the results obtained by running inference using the pre-trained weights offered by the author [18] are labeled "SingleHDR", and those attained following training on the Laval dataset are labeled "SingleHDR*". HDRCNN, SingleHDR, SingleHDR*, and the proposed method were trained using batch sizes of 8, 16, 16, and 32, respectively. A learning rate of 1e-4 was used for all methods except HDRCNN where a 5e-5 learning rate was used. In previous methods, a blending parameter was used to combine the linear LDR image with the output of the networks that recover over-exposed and saturated. In our method, the output of the sky luminance estimator is then softly blended with the sun luminance distribution. The blending parameters are shown in Table 1.

Figure 7 shows the HDR maps generated using the proposed method are closer in appearance to the ground truth HDR maps than those generated using HDRCNN [17] and SingleHDR [18] are. The displayed images are the results of tone-mapping using Reinhard's method, which is ideal for qualitative comparison given that it is based on the human visual system [38].

In Table 2, the proposed model was compared with the HDRCNN, SingleHDR, and SingleHDR* using the HDR-VDP-2.2, PSNR, and SSIM evaluation metrics. The best values by each metric are highlighted in bold. The results for these methods were obtained using the source code provided by the authors. The proposed method more accurately reconstructed HDR maps of the outdoor scenes than the two previous methods (HDRCNN and SingleHDR) did.

The method HDR-VDR-2.2, which is based on the visual model for all the luminance conditions, examines visibility and image quality differences between images. For this measure, we paid particular attention to image quality and evaluated it as a mean opinion score (Q score) [39]. A higher Q score implies a better-quality image. The PSNR is the ratio of the maximum possible power of the signal to that of the noise. In our experiment, the PSNR is a pixel-to-pixel comparison between the ground truth HDR map and the predicted HDR map, and it considers the accuracy of the high luminance values [23]. The SSIM estimates the perceived luminance, contrast, and structure quality differences between the ground truth and the predicted images. Here, perceptual uniformity encoding is performed on HDR maps because PSNR and SSIM metrics are originally designed for LDR image comparison [40].

The HDR-VDP-2.2 measure improves as the inverse camera response function is more accurately reconstructed [41]. As described in Section III-B (Sky Luminance Estimator), the proposed sky luminance estimator reconstructs the luminance distribution of the sky without estimating the inverse camera response function. In other words, achieving the highest performance in the HDR-VDP-2.2 measure implies that our GAN-based architecture implicitly reconstructed the most accurate inverse camera response function.

In addition, the histograms show that HDR maps geneated using the proposed method more closely matches the peak luminance values in the sun regions of the ground truth HDR maps than those generated using other methods. Since the peak luminance value is considered in PSNR measure, the proposed method scored highest in two tone-mapping operators: Durand and Mantiuk that preserve a wide range of luminance values. Since the other two tone-mapping operators (Drago and Reinhard) focus on compressing the luminance values to preserve the image details, the HDRCNN achieve slightly higher performance in terms of PSNR measure in Table 2.

HDRCNN often generates overly-bright results and suffers from noise in the under-exposed regions because an aggressive and fixed inverse camera response function is used ($\eta = 2$) [18]. As shown in Table 2, HDR maps reconstructed using HDRCNN were generally worse than those reconstructed using the proposed method. SingleHDR shows better results in terms of SSIM because it explicitly estimates an inverse camera response function for each input image by employing a parametric model, whose parameters are estimated using a convolutional neural network [13]. However, as shown in Figure 7, the small discrepancy in the SSIM can be explained by the fact that quantitative metrics do not always match subjective results and may incorrectly rank the approaches [41].

SingleHDR scores better than other methods on the SSIM metric which examines the image structure properties between two images. This is because HDRCNN and our method simultaneously infer inverse CRFs and over-exposure regions, but SingleHDR uses a canonical method to reconstruct the over-exposed regions after an inverse CRF is inferred. In order to prevent banding effects and to more accurately recover the dynamic range and tone, the proposed method did not estimate a CRF. In some cases, our encoder-decoder architecture results in unexpected noise. Further consideration is required to alleviate this unexpected noise effect.

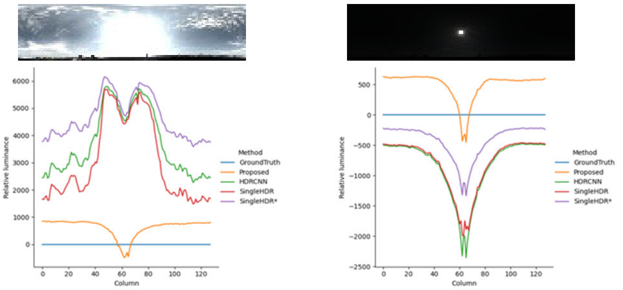

FIGURE 8. Tone mapped LDR images from the Laval dataset under two weather conditions (top); corresponding histograms generated from HDR maps reconstructed using four methods (bottom).

Figure 8 shows tone mapped LDR images from the Laval dataset under cloudy and sunny weather conditions, and the corresponding histograms generated from HDR maps reconstructed using four methods (HDRCNN, SingleHDR, SingleHDR*, and the proposed method). In the histograms, the x axis (bin) represents columns in the HDR map, and the y axis represents the difference between the luminance values of the ground truth (blue line) and reconstructed HDR map. The plots closer to the blue line imply that reconstructed HDR maps have luminance distributions that are more similar to the ground truth HDR map. Figure 8 shows that the HDR maps reconstructed using the proposed method have the luminance distributions most similar to ground truth. This is also confirmed in evaluation using the HDR-VDP-2.2 measure as shown in Table 2.

In order to evaluate how accurately each of the methods reconstructed luminance distribution, histograms of the HDR maps are generated and are directly compared using the earth mover's (EM) and Hellinger (H.) distance measures. The Hellinger distance is used to compute similarity between two probability distributions using a bounded metric [42]. The EM distance is based on the minimal amount of work required to transform one distribution into another. Table 3 shows that the HDR maps generated by our method surpass those generated by previous methods in histogram similarity to the ground truth.

Table 4 shows a quantitative comparison of the HDR maps generated using different combinations of loss functions in our approach on the Laval dataset. The tests were initially conducted with one, two, three, and eventually all loss functions. In Table 5, case A refers to tests conducted using only an encoder-decoder setup. In case B, Sunpose-Net was added prior to inference. In case C, the Sunrad-Net was added to the setup in case B. All four loss terms were used in all cases.

TABLE 2. **Quantitative comparison of HDR maps reconstructed from tone-mapped images (Laval dataset).**

	Drago TMO [35]			Durand TMO [36]			Mantiuk TMO [37]			Reinhard TMO [38]			Mean Value		
	HDR-VDP-2.2	PSNR	SSIM	HDR-VDP-2.2	PSNR	SSIM	HDR-VDP-2.2	PSNR	SSIM	HDR-VDP-2.2	PSNR	SSIM	HDR-VDP-2.2	PSNR	SSIM
HDRCNN [17]	67.17	**28.05**	0.5412	69.03	32.67	0.5904	69.02	32.38	0.5502	66.94	**27.29**	0.4820	68.04	30.10	0.5410
SingleHDR [18]	67.22	27.17	**0.6000**	68.50	32.65	0.6416	68.18	32.42	0.6018	66.23	25.89	0.4903	67.53	29.53	**0.5834**
SingleHDR *	68.08	27.39	0.5152	69.83	33.54	**0.6707**	69.78	33.24	**0.6471**	67.52	26.90	0.4778	68.80	30.27	0.5777
Proposed	**68.51**	27.82	0.5342	**70.55**	**33.98**	0.6302	**70.32**	**33.48**	0.6018	**68.01**	26.85	**0.4915**	**69.35**	**30.53**	0.5644

TABLE 3. **Histogram distance comparison for HDR maps reconstructed from tone-mapped images (Laval dataset).**

	Drago TMO [35]		Durand TMO [36]		Mantiuk TMO [37]		Reinhard TMO [38]		Mean Value	
	EM dist.	H. dist.	EM dist.	H. dist.	EM dist.	H. dist.	EM dist.	H. dist.	EM dist.	H. dist.
HDRCNN [17]	0.1309	10.78	0.04843	4.663	0.05131	5.071	0.08628	7.387	0.07923	6.975
SingleHDR [18]	0.04208	3.711	0.03275	3.193	0.02960	3.050	0.05505	5.146	0.03987	3.775
SingleHDR*	0.06995	5.641	0.03654	3.218	0.03900	3.454	0.04377	3.686	0.04732	4.000
Proposed	**0.02624**	**2.215**	**0.02145**	**1.814**	**0.02306**	**1.979**	**0.03025**	**2.542**	**0.02525**	**2.138**

TABLE 4. **Ablation for the proposed loss functions.**

	Case 1	Case 2			Case 3			Case 4
	$\mathcal{L}_{rec}$	$\mathcal{L}_{rec} + \mathcal{L}_{perc}$	$\mathcal{L}_{rec} + \mathcal{L}_{adv}$	$\mathcal{L}_{rec} + \mathcal{L}_{DoG}$	$\mathcal{L}_{rec} + \mathcal{L}_{perc} + \mathcal{L}_{DoG}$	$\mathcal{L}_{rec} + \mathcal{L}_{adv} + \mathcal{L}_{DoG}$	$\mathcal{L}_{rec} + \mathcal{L}_{adv} + \mathcal{L}_{Perc}$	$\mathcal{L}_{total}$
HDR-VDP-2.2	69.01	67.27	67.22	69.19	69.2852	69.26	**69.57**	69.35
PSNR	30.17	29.44	30.17	30.24	30.37	30.27	30.49	**30.53**
SSIM	0.5670	0.5262	**0.5709**	0.5674	0.5662	0.5602	0.5681	0.5644

Because Mantiuk TMO uses four parameters to ensure that output LDR images match the input HDR scene as closely as possible, an encoder-decoder architecture alone (case A) is sufficient to reconstruct HDR maps.

In our experiment, an input 1024 × 2048 image is reduced to a resolution of 32 × 128 because of the limited computational resources available. In order to examine the effects of reducing the resolution of LDR inputs on HDR reconstruction, we compared the quality of reconstructed HDR images under three input resolutions (32 × 128, 40 × 160, and 48 × 192) in terms of histogram distance measures: Earth mover (EM) and Hellinger (H). Figure 9 and Table 6 show that the proposed method's performance improves as the resolution increases.

The proposed model was additionally evaluated on the real CAU image dataset, which was captured using a CANON EOS 6D Mark II with a CANON EF mount and an 8-15 mm f/4L CANON fisheye USM lens. Figure 10 shows images captured under various exposures and the correspondingsynthesized HDR map. We set the aperture value to f/22 and the ISO value to 100. We then captured each scene with 1/4,000, 1/1,000, 1/250, 1/60, and 1/15 shutter speeds.

TABLE 5. Ablation for the various configurations of the sun luminance estimator.

	Drago TMO [35]			Durand TMO [36]			Mantiuk TMO [37]			Reinhard TMO [38]			CAU dataset		
	HDR-VDP-2.2	PSNR	SSIM	HDR-VDP-2.2	PSNR	SSIM	HDR-VDP-2.2	PSNR	SSIM	HDR-VDP-2.2	PSNR	SSIM	HDR-VDP-2.2	PSNR	SSIM
Case A	**68.67**	27.47	0.5194	**70.55**	33.95	0.6396	**70.34**	**33.56**	0.6210	**68.11**	26.74	0.4910	66.16	13.70	0.1804
Case B	68.50	27.35	0.5167	70.48	33.87	**0.6459**	70.29	33.47	**0.6231**	67.88	26.53	0.4885	66.19	**13.81**	**0.1940**
Case C	68.51	**27.82**	**0.5314**	**70.55**	**33.98**	0.6302	70.32	33.48	0.6018	68.01	**26.85**	**0.4915**	**66.20**	13.68	0.1788

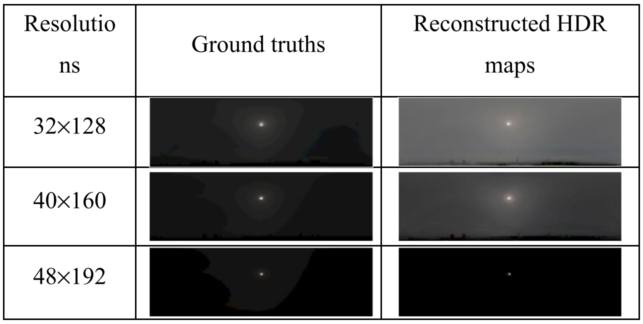

FIGURE 9. Ground truth of three different resolutions and corresponding reconstructed HDR maps.

TABLE 6. Histogram comparison of HDR maps reconstructed for three different resolutions.

Resolutions	EM dist.	H. dist.
32×128	0.026	3.145
40×160	0.017	1.516
48×192	**0.012**	**1.085**

TABLE 7. Quantitative comparison of HDR maps reconstructed from bracketed images (CAU dataset).

	CAU dataset		
	HDR-VDP-2.2	PSNR	SSIM
HDRCNN [17]	64.68	13.56	0.1699
SingleHDR [18]	65.76	13.68	0.1786
SingleHDR*	66.05	**13.80**	**0.1920**
Proposed	**66.20**	13.68	0.1788

Ground-truth HDR images were generated by merging these five LDR images using Photomatix software. The CAU dataset we used in these experiments consists of 505 LDR images and 101 HDR images. The CAU dataset and implemented source code are available at https://github.com/ShinYwings/HDR-Map-Reconstruction-From-a-Single-LDR-Sky-Panoramic-Image-For-Outdoor-Illumination-Estimation.

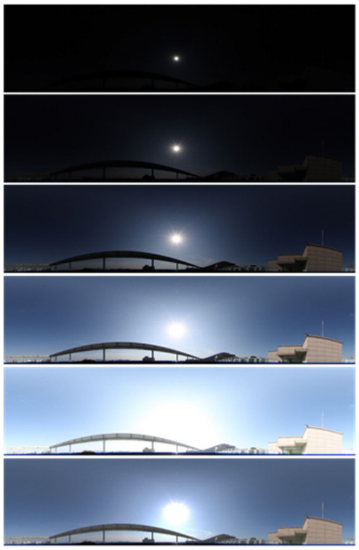

FIGURE 10. Images of a given scene captured with five exposure values, and the corresponding synthesized HDR map (the last row).

TABLE 8. Histogram distance comparison for HDR maps reconstructed from bracketed images (CAU dataset).

	CAU dataset	
	EM dist.	H. dist.
HDRCNN [17]	0.09231	6.542
SingleHDR [18]	0.08323	5.778
SingleHDR*	0.06949	5.010
Proposed	**0.06232**	**4.018**

The HDR maps in the Laval dataset were synthesized from images of 7 different exposure values. However, due to the absence of an appropriate neutral-density (ND) filter, HDR images in the CAU dataset were synthesized from images of just five exposure values. This implies that the CAU dataset images have two missing exposure values (1/8,000 and 1/4,000 with an ND filter), resulting in higher luminance values in synthesized HDR maps. Figure 11 shows the HDR maps generated by the proposed method and the previous methods from the CAU dataset inputs. Tables 7 and 8 show

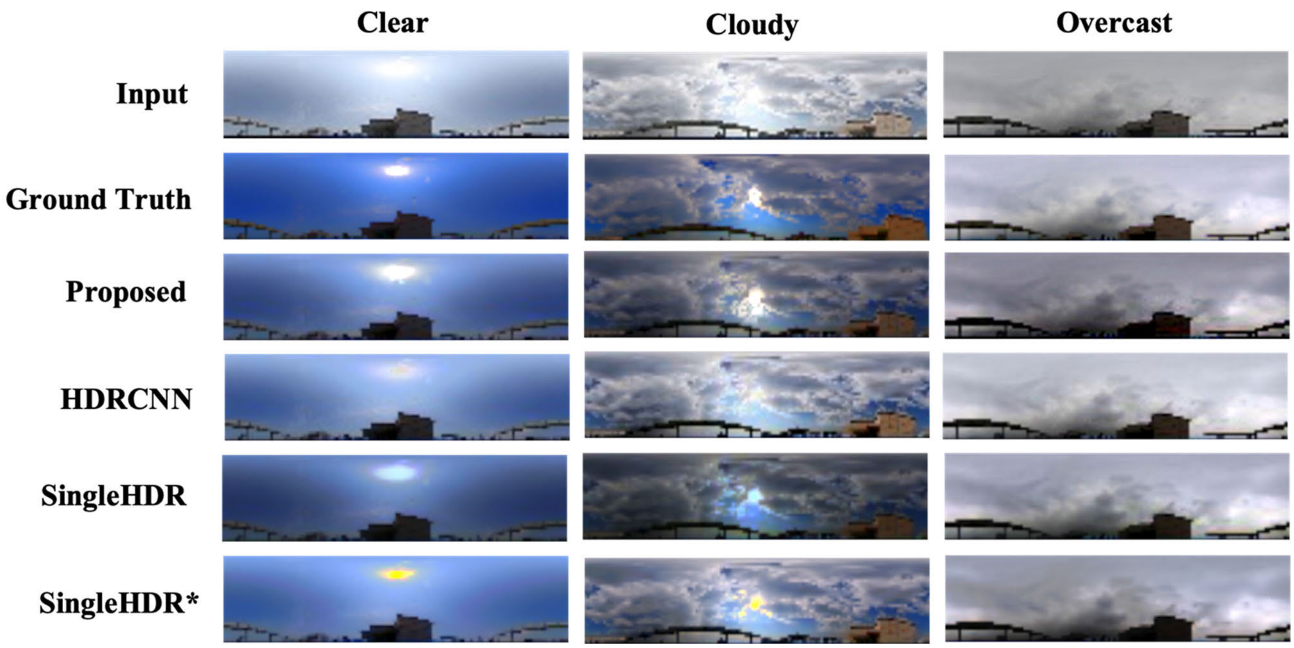

FIGURE 11. Qualitative comparison of HDR maps generated using different HDR reconstruction methods on the CAU dataset.

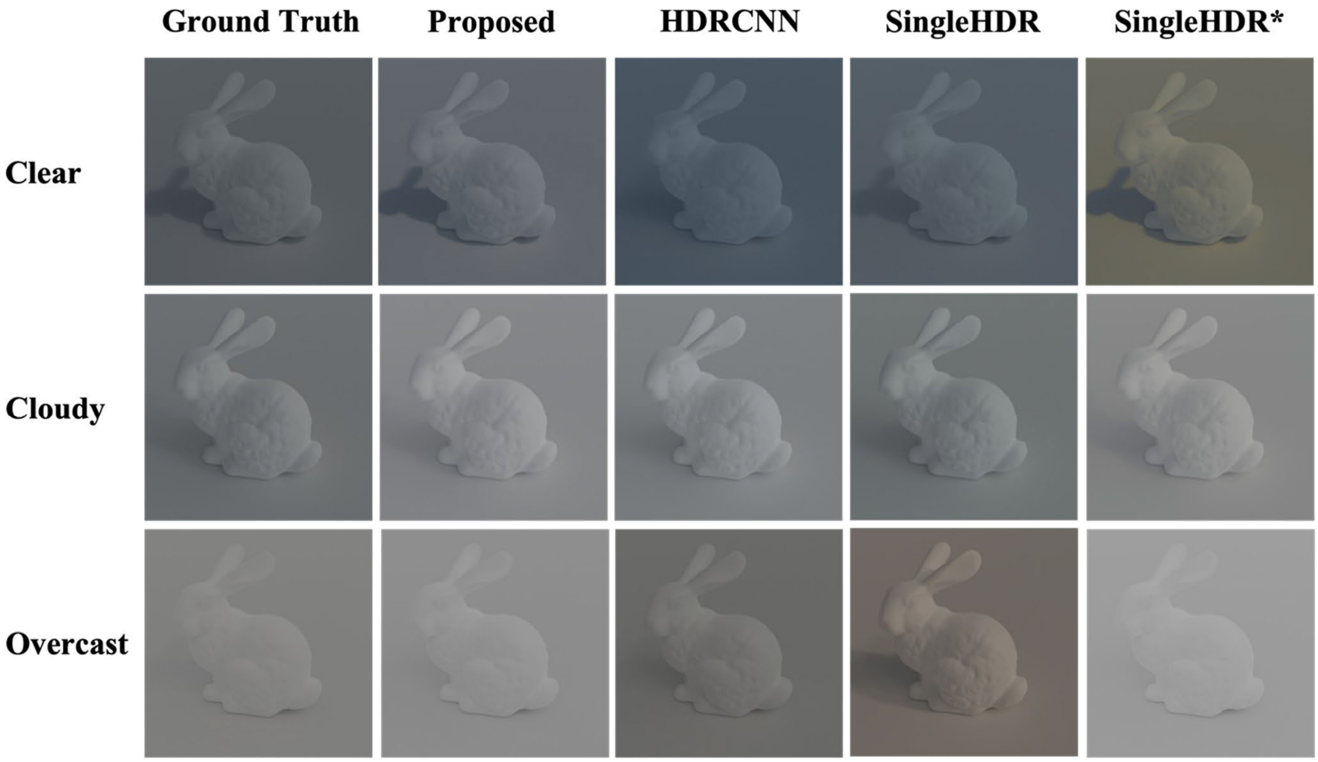

FIGURE 12. Qualitative comparison of images rendered from reconstructed HDR maps on the Laval dataset.

the results of evaluating our proposed method against various HDR reconstruction methods on the CAU dataset using three metrics and histogram distances.

C. RENDERING RESULTS

The rendering engine in Blender 2.93, an open-source 3D computer graphic toolset, was used to generate rendering

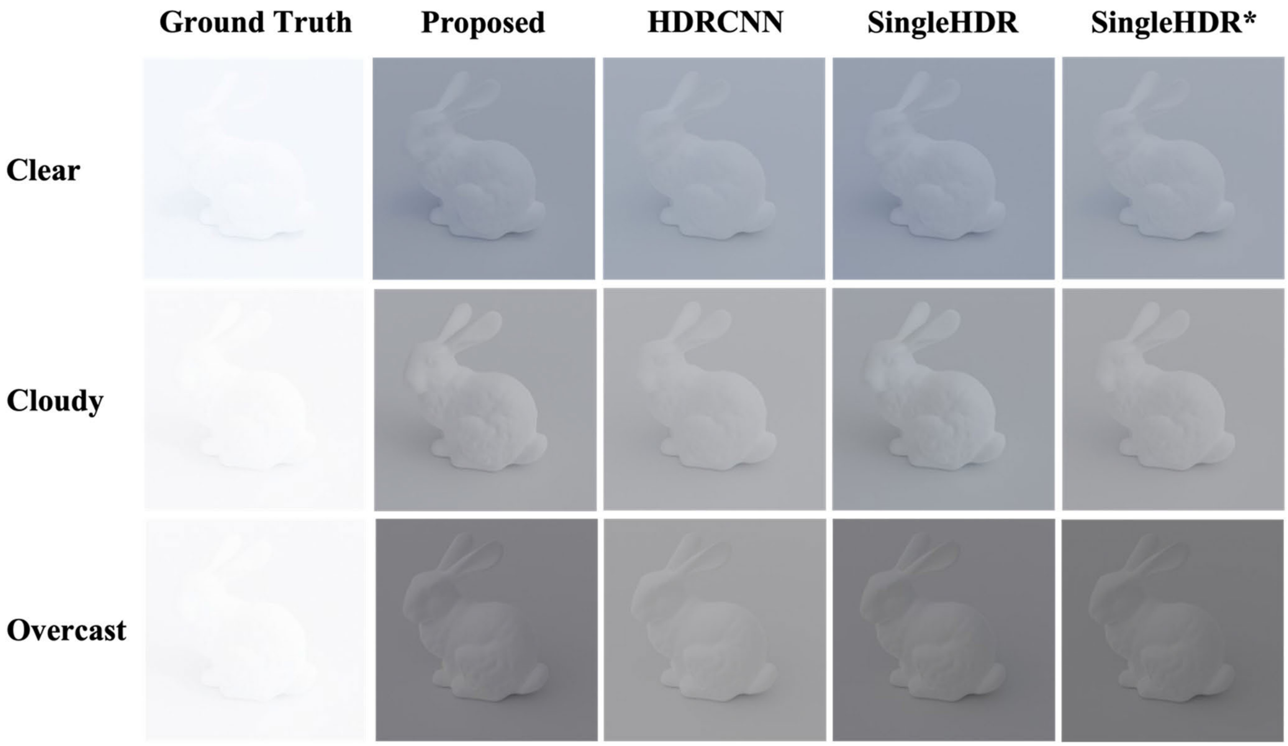

FIGURE 13. Qualitative comparison of images rendered from reconstructed HDR maps on the CAU dataset.

images from the reconstructed HDR maps. In Blender, the reconstructed HDR images from the Laval dataset and CAU dataset were used as environment maps to relight the Stanford bunny models with a diffuse shader. These were ultimately rendered into 256 × 256 images.

Figures 12 and 13 show a qualitative comparison between images rendered using the HDR maps generated by various methods on the Laval and CAU datasets, respectively. Figure 12 shows that the proposed method is robust and produces the most accurate rendering for various weather conditions (clear, cloudy and overcast). Images in the CAU dataset have high luminance values and therefore result in very bright rendered images, as shown in Figure 13.

V. CONCLUSION

In this paper, we introduced a deep learning model for HDR reconstruction from a single LDR sky panoramic image. This model estimated the illumination distribution of sun and sky regions separately and used a blending operator to merge the results into an HDR map. In our deep learning model, an encoder considers coherence of the sun region and sky as well as luminance gradation in the sky region to build a multi-dimensional latent representation of the entire luminance distribution in a sky panoramic image, and the two decoders consider the heterogeneous luminance properties of the sky and the sun region separately. The position and intensity of the sun are estimated from an input LDR sky panoramic image using Sunpose-Net and Sunrad-Net, which are proposed in this study. The performance of the proposed method was evaluated using the Laval Sky HDR database (Laval dataset) and the CAU dataset.

The proposed method has three main limitations. The first is an assumption that the sun is horizontally centered in all input panoramic images. This means that the proposed method only regresses the angle of elevation and not the azimuth. Since the position of the sun in the sky varies through the day, we need to transform input images so that the sun is horizontally centered in the panoramic images. The second is that we did not consider temporal continuity when estimating the sun position and radiance. This implies the HDR maps reconstructed using the proposed method cannot consistently render a virtual object over sequential frames. In this case, unwanted light flickering effects may arise in real-time rendering applications. The third limitation is reduction of the LDR input resolution because of the limited computational resources. In order to represent more precise outdoor illumination, we need to use input images of higher resolution. More specifically, since the sharp variations in the sun image region are low pass filtered in lower resolution images, the luminance distribution of the sun region is diminished. This implies that we may not be able to generate rendering outputs that accurately reflect the effect of bright sun light. The experimental results showed that the proposed method's performance improves as the resolution increases. Therefore, a patch-based training approach is to be explored in order to handle large input panoramic image.

ACKNOWLEDGMENT

The authors would like to thank Jean-Francois Lalonde and his team for creating and sharing the Laval Sky HDR database, which was essential to our study.

REFERENCES

[1] Y. Zhu, Y. Zhang, S. Li, and B. Shi, "Spatially-varying outdoor lighting estimation from intrinsics," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Nashville, TN, USA, Jun. 2021, pp. 12834–12842.

[2] Z. Wang, W. Chen, D. Acuna, J. Kautz, and S. Fidler, "Neural light field estimation for street scenes with differentiable virtual object insertion," in *Proc. Eur. Conf. Comput. Vis.*, Tel Aviv, Israel, 2022, pp. 380–397.

[3] R. Perez, R. Seals, and J. Michalsky, "All-weather model for sky luminance distribution—Preliminary configuration and validation," *Sol. Energy*, vol. 50, no. 3, pp. 235–245, Mar. 1993, doi: 10.1016/0038-092X(93)90017-I.

[4] A. J. Preetham, P. Shirley, and B. Smits, "A practical analytic model for daylight," in *Proc. 26th Annu. Conf. Comput. Graph. Interact. Techn.*, Los Angeles, CA, USA, 1999, pp. 99–100.

[5] L. Hošek and A. Wilkie, "An analytic model for full spectral sky-dome radiance," *ACM Trans. Graph.*, vol. 31, no. 4, pp. 1–9, Jul. 2012, doi: 10.1145/2185520.2185591.

[6] L. Hošek and A. Wilkie, "Adding a solar-radiance function to the Hošek–Wilkie skylight model," *IEEE Comput. Graph. Appl.*, vol. 33, no. 3, pp. 44–52, May/Jun. 2013, doi: 10.1109/MCG.2013.18.

[7] J.-F. Lalonde and I. Matthews, "Lighting estimation in outdoor image collections," in *Proc. IEEE Conf. 3D Vis.*, Tokyo, Japan, Dec. 2014, pp. 131–138.

[8] Y. Hold-Geoffroy, K. Sunkavalli, S. Hadap, E. Gambaretto, and J.-F. Lalonde, "Deep outdoor illumination estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 7312–7321.

[9] J. Zhang, K. Sunkavalli, Y. Hold-Geoffroy, S. Hadap, J. Eisenman, and J.-F. Lalonde, "All-weather deep outdoor lighting estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 10158–10166.

[10] Y. Hold-Geoffroy, A. Athawale, and J.-F. Lalonde, "Deep sky modeling for single image outdoor lighting estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 6927–6935.

[11] P. Yu, J. Guo, F. Huang, C. Zhou, H. Che, X. Ling, and Y. Guo, "Hierarchical disentangled representation learning for outdoor illumination estimation and editing," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Montreal, QC, Canada, Oct. 2021, pp. 15313–15322.

[12] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Proc. ACM SIGGRAPH*, New York, NY, USA, Aug. 2008, pp. 1–10.

[13] M. D. Grossberg and S. K. Nayar, "Modeling the space of camera response functions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 10, pp. 1272–1282, Oct. 2004, doi: 10.1109/TPAMI.2004.88.

[14] J. Stumpfel, A. Jones, A. Wenger, C. Tchou, T. Hawkins, and P. Debevec, "Direct HDR capture of the sun and sky," in *Proc. ACM SIGGRAPH*, New York, NY, USA, 2004, p. 5.

[15] 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*, vol. 37, no. 2, pp. 37–49, May 2018, doi: 10.1111/cgf.13340.

[16] Y. Endo, Y. Kanamori, and J. Mitani, "Deep reverse tone mapping," *ACM Trans. Graph.*, vol. 36, no. 6, pp. 177:1–177:10, Nov. 2017, doi: 10.1145/3130800.3130834.

[17] G. Eilertsen, J. Kronander, G. Denes, R. K. Mantiuk, and J. Unger, "HDR image reconstruction from a single exposure using deep CNNS," *ACM Trans. Graph.*, vol. 36, no. 6, pp. 178:1–178:15, Nov. 2017, doi: 10.1145/3130800.3130816.

[18] Y.-L. Liu, W.-S. Lai, Y.-S. Chen, Y.-L. Kao, M.-H. Yang, Y.-Y. Chuang, and J.-B. Huang, "Single-image HDR reconstruction by learning to reverse the camera pipeline," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 1651–1660.

[19] Y. Zhang and T. O. Aydın, "Deep HDR estimation with generative detail reconstruction," *Comput. Graph. Forum*, vol. 40, no. 2, pp. 179–190, Jun. 2021, doi: 10.1111/cgf.142624.

[20] J. Zhang and J.-F. Lalonde, "Learning high dynamic range from outdoor panoramas," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 4519–4528.

[21] L. Wang and K.-J. Yoon, "Deep learning for HDR imaging: State-of-the-art and future trends," 2021, *arXiv:2110.10394*.

[22] V. Gkitsas, N. Zioulis, F. Alvarez, D. Zarpalas, and P. Daras, "Deep lighting environment map estimation from spherical panoramas," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Seattle, WA, USA, Jun. 2020, pp. 640–641.

[23] X. Chen, Y. Liu, Z. Zhang, Y. Qiao, and C. Dong, "HDRUNet: Single image HDR reconstruction with denoising and dequantization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Nashville, TN, USA, Jun. 2021, pp. 354–363.

[24] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 1125–1134.

[25] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, *arXiv:1411.1784*.

[26] T. Kim, M. Cha, H. Kim, J. K. Lee, and J. Kim, "Learning to discover cross-domain relations with generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, Sydney, NSW, Australia, 2017, pp. 1857–1865.

[27] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 2223–2232.

[28] Z. Yi, H. Zhang, P. Tan, and M. Gong, "DualGAN: Unsupervised dual learning for image-to-image translation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 2849–2857.

[29] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 2921–2929.

[30] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis.*, Venice, Italy, Oct. 2017, pp. 618–626.

[31] H. Li and P. Peers, "CRF-Net: Single image radiometric calibration using CNNs," in *Proc. Eur. Conf. Vis. Media Prod.*, London, U.K., 2017, pp. 1–9.

[32] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[33] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, Amsterdam, The Netherlands, Oct. 2016, pp. 694–711.

[34] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, "Least squares generative adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 2794–2802.

[35] 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, Nov. 2003, doi: 10.1111/1467-8659.00689.

[36] F. Durand and J. Dorsey, "Fast bilateral filtering for the display of high-dynamic-range images," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 257–266, Jul. 2002, doi: 10.1145/566654.566574.

[37] R. Mantiuk, K. Myszkowski, and H.-P. Seidel, "A perceptual framework for contrast processing of high dynamic range images," *ACM Trans. Appl. Perception*, vol. 3, no. 3, pp. 286–308, Jul. 2006, doi: 10.1145/1166087.1166095.

[38] 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. 2005, doi: 10.1109/TVCG.2005.9.

[39] M. Narwaria, R. K. Mantiuk, M. P. Da Silva, and P. Le Callet, "HDR-VDP-2.2: A calibrated method for objective quality prediction of high-dynamic range and standard images," *J. Electron. Imag.*, vol. 24, no. 1, Jan. 2015, Art. no. 010501, doi: 10.1117/1.JEI.24.1.010501.

[40] R. K. Mantiuk and M. Azimi, "PU21: A novel perceptually uniform encoding for adapting existing quality metrics for HDR," in *Proc. Picture Coding Symp. (PCS)*, Bristol, U.K., Jun. 2021, pp. 1–5.

[41] P. Hanji, R. Mantiuk, G. Eilertsen, S. Hajisharif, and J. Unger, "Comparison of single image HDR reconstruction methods—The caveats of quality assessment," in *Proc. ACM SIGGRAPH*, Vancouver, BC, Canada, 2022, pp. 1–8.

[42] M. Afifi, M. A. Brubaker, and M. S. Brown, "HistoGAN: Controlling colors of GAN-generated and real images via color histograms," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Nashville, TN, USA, Jun. 2021, pp. 7941–7950.

GYEONGIK SHIN received the B.S. degree in computer and information communications engineering from Hongik University, Sejong, South Korea, in 2020, and the M.S. degree in computer science and engineering from Chung-Ang University, Seoul, South Korea, in 2022. His research interests include HDR map reconstruction and augmented reality.

KYEONGMIN YU received the B.S. degree in physics and the B.E. degree in computer science and engineering from Chung-Ang University, Seoul, South Korea, in 2022, where she is currently pursuing the M.S. degree in computer science and engineering with Computer Vision and Augmented Reality Laboratory. Her research interests include deep learning and computer vision for sequential image analysis.

MPABULUNGI MARK received the B.S. degree in computer science from Uganda Christian University, Mukono, Uganda, in 2015, and the M.E. degree in IoT electronics engineering from Kangnam University, Yongin, South Korea, in 2020. He is currently pursuing the Ph.D. degree in computer science and engineering with Chung-Ang University, Seoul, South Korea, where he is with the Computer Vision and Augmented Reality Laboratory. His research interests include computer vision and augmented reality with a particular emphasis on scene illumination estimation.

HYUNKI HONG received the B.S., M.S., and Ph.D. degrees in electronic engineering from Chung-Ang University, Seoul, South Korea, in 1993, 1995, and 1998, respectively. He was a Researcher at the Automatic Control Research Center, Seoul National University, until 1999. From 2000 to 2014, he was a Professor at the Department of Imaging Science and Arts, GSAIM, Chung-Ang University, where he has been a Professor with the College of Software, since 2014. He is currently the Dean of the College of Software, Chung-Ang University. From 2002 to 2003, he was a Postdoctoral Researcher at the Department of Computer Science and Engineering, University of Colorado, Denver. His research interests include HDR map reconstruction, camera localization, stereo matching, and augmented reality.