

Received August 13, 2018, accepted September 1, 2018, date of publication September 17, 2018, date of current version October 12, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2870295

# Inverse Tone Mapping Operator Using Sequential Deep Neural Networks Based on the Human Visual System

HANBYOL JANG, KIHUN BANG<sup>ID</sup>, JINSEONG JANG, AND DOSIK HWANG<sup>ID</sup>

School of Electrical and Electronic Engineering, Yonsei University College of Engineering, Seoul 03722, South Korea

Corresponding author: Dosik Hwang (dosik.hwang@yonsei.ac.kr)

This work was supported by Samsung Electronics Co., Ltd.

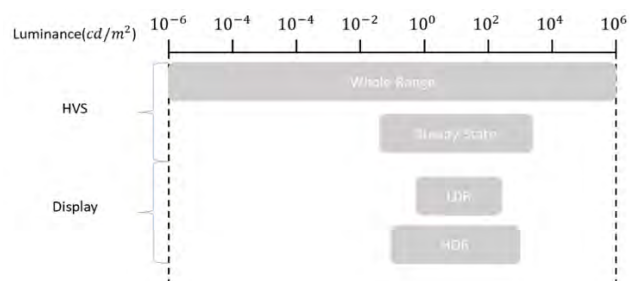
**ABSTRACT** Conventional digital displays show images with much less dynamic range than that of human visual perception. High-dynamic range (HDR) displays are being developed for viewing images with higher dynamic range than that of conventional low-dynamic range (LDR) images. However, to view existing LDR images on HDR displays, an inverse tone mapping operator (ITMO), which is a process to extend the dynamic range of LDR images, is required. In this paper, we propose an adaptive ITMO by effectively learning the differences between LDR and HDR images using a sequential learning process: dynamic range learning followed by color difference learning. Our proposed method enables visualization of colors similar to real-world colors better than conventional ITMOs by learning color differences based on the human visual system properties. For the objective comparison, seven different evaluation metrics optimized for HDR image evaluation were used. Our method resulted in 10%–25% improved HDR-visual difference predictor-2.2 values over those of other ITMOs. Other metrics also demonstrated the superior performance of our method. For the subjective comparison, eight human observers evaluated the estimated HDR images in terms of color appearance and overall preferences. Our method received an average 4.7 out of 5 score, whereas other ITMOs received below 3.5 scores in the evaluation of color appearance. The objective and subjective evaluations' results showed that our proposed method outperformed the conventional ITMOs in estimating the dynamic range and color appearance of ground-truth HDR images.

**INDEX TERMS** Color difference, deep neural network, high dynamic range, inverse tone mapping operator, human visual system.

## I. INTRODUCTION

Humans have the following two types of photoreceptor cells in the retina: cone cells that can function under more intense light ranges from  $10^{-2}$  to  $10^8 \text{ cd/m}^2$  and rod cells that can function under less intense light ranges from  $10^{-6}$  to  $10 \text{ cd/m}^2$  [1]. The total amount of light that humans can perceive ranges from  $10^{-6}$  to  $10^8 \text{ cd/m}^2$ , and the whole dynamic range is as wide as  $10^{14}$ . However, humans cannot perceive that entire range of light at once. The human visual system (HVS) works in a way in which rhodopsin of the rod cell enables the eyes to adapt to light, so light can be seen only in one portion of the broad spectrum. The range of light seen in this adaptive visual system is called simultaneous dynamic range or steady-state dynamic range, and the simultaneous dynamic range is  $\leq 10^{3.7}$  [2].

However, displays, such as TVs, monitors, and mobile phones emit as much as  $10^{-1}$  to  $10^3 \text{ cd/m}^2$ , and the contrast



**FIGURE 1.** The overall dynamic range of the HVS. The LDR display cannot show a real-world scene because the LDR display cannot emit light less than the HVS can accommodate [2].

ratio is  $<10^3$ , which is far below the human-perceivable dynamic range and contrast ratio. Fig. 1 compares the dynamic range of HVS and the display. In recent years, the need for dynamic range improvement has led to the

development of a wide dynamic range display, which is termed a high dynamic range (HDR) display. An HDR display not only extends the dynamic range, but also enhances the contrast ratio of the panel for a more realistic expression. HDR displays require higher than 10-bit panels, whereas traditional low dynamic range (LDR) displays have 8-bit panels. There are two standards of HDR technologies: HDR10 and Dolby vision. HDR10 supports  $\leq 10$ -bit precision with  $1,000 \text{ cd/m}^2$  peak luminance, and Dolby Vision supports  $\leq 12$ -bit precision with  $4,000 \text{ cd/m}^2$  peak luminance [3], [4]. The major display manufacturers, such as LG, Philips, and Samsung, follow these standards for HDR technologies.

As HDR displays are being developed, the number of bits used to represent a single pixel in the image are typically increased. Conventional LDR images have 8 bits for a single pixel per R, G, and B channel. This bit depth is not effective for display of these LDR images on a 10-bit HDR display. Therefore, HDR images with a high bit and contrast ratio that can be properly seen on HDR displays are required.

Most cameras today have an HDR mode, and there are two ways to obtain HDR images. The first method used in most HDR-enabled cameras is to make HDR images by taking multiple LDR images of varying exposure values (EV) over a short period and synthesizing them [5], [6]. This has the advantage of being able to obtain HDR images even with a simple LDR camera. However, because the shooting time to obtain multiple images with various exposure values is long, temporal resolution is a problem while obtaining an HDR video. In addition, if there is camera shake or subject movement in the shooting process, the HDR image will be blurred, and ghost artifacts will appear [7]. Some solutions for reducing the ghost artifacts have been proposed, but they cannot be completely removed in cases of severe motion [8], [9]. The second method is to acquire the HDR image in a single shot by using an advanced HDR camera. However, advanced HDR cameras are too expensive to be used for general purposes.

Furthermore, there are a tremendous number of LDR images that have already been shot with only a single exposure value. The aforementioned methods to obtain HDR images cannot be applied to these LDR contents.

To overcome the limitations of acquiring HDR images, making an HDR image from a single LDR image has been studied actively, and this method is called inverse tone mapping. HDR images made by using an inverse tone mapping operator (ITMO) have high usability because there is no limitation in shooting images in a single shot or combining multiple LDR images. In addition, ITMO enables the conversion of LDR images already shot into HDR images. Over the past decade, several ITMO techniques have been proposed. However, the performances of existing ITMOs are not satisfactory, especially when estimating ground-truth HDR images. Conventional ITMOs work for only a few types of images, and they are mostly focused only on widening the dynamic range; therefore, they often fail to estimate the

ground-truth HDR images and cannot handle changes in chromatic information.

Over the past few years, deep convolutional neural networks have shown great performance in the field of computer vision. Deep learning techniques have overcome many problems in classification by using single-class labels [10]–[18]. Image segmentation by labeling the class for each pixel of the image outperforms the conventional methods [10]–[12]. Similarly, deep learning techniques have achieved good results in regression problems such as super resolution [13]–[15], artifact reduction [16], and image translation [17], [18]. In this paper, we designed a deep learning architecture that was optimized for ITMO and demonstrated that a dynamic range can be effectively improved by luminance learning and that the color appearance can also be well estimated by subsequent chrominance learning.

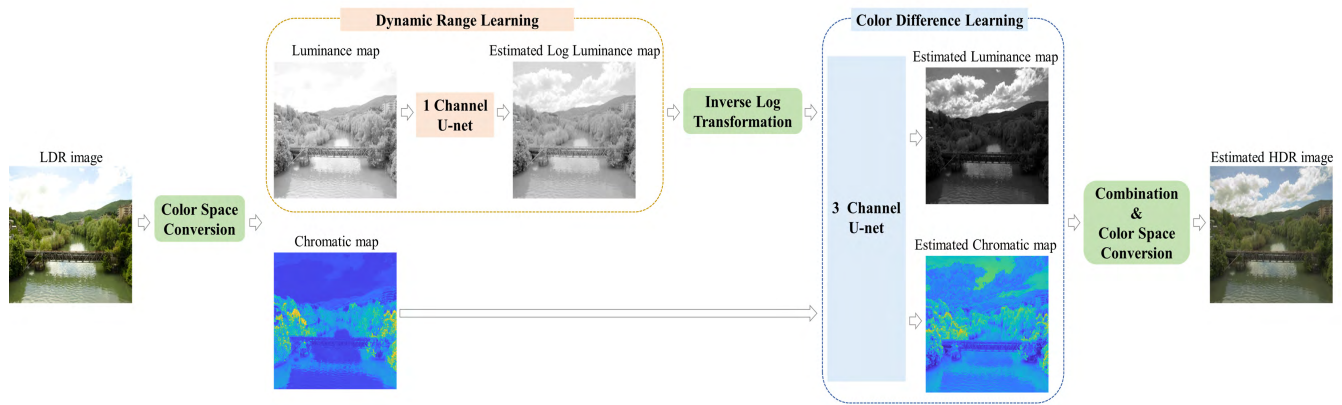
In the remaining part of this paper, Section 2 introduces previous related studies. Section 3 details the proposed algorithm. Section 4 describes the data and shows HDR images generated by the proposed and conventional ITMOs, and Section 5 presents subjective and objective comparisons of ITMOs. Finally, Section 6 summarizes the main findings of this paper.

## II. RELATED WORKS

### A. INVERSE TONE MAPPING OPERATORS

Various ITMO techniques mostly based on mathematical models have been developed previously. Akyüz *et al.* [19] proposed a simple linear expansion method without considering image characteristics. Masia *et al.* [20] proposed a gamma-curve based global mapping function that can handle global luminance variation, and selective tone mapping methods for local areas with different characteristics. However, a limitation is that the contrast of dark areas is not well improved when the expansion is focused on the over-exposed areas and vice versa.

Several methods have also been proposed to handle saturated bright regions because the difference between HDR and LDR images is most noticeable in bright regions. Wang *et al.* [21] interpolated over-exposed and under-exposed areas selectively to reconstruct the details of high-light and texture. Didyk *et al.* [22] classified bright regions into three types: diffuse surface, reflection, and light source. They applied different expansion functions for each type of bright region. Banterle *et al.* [23] predicted ground-truth images through the inverse process of the Reinhard tone mapping operator [24]. The *Expand Map*, which extracts the light source in images by applying a median-cut algorithm [25], can reconstruct lost luminance in saturated areas and attenuate quantization artifacts. Rempel *et al.* [26] used the brightness enhancement function (BEF) instead of an *Expand Map* for videos. However, these methods have a limitation in which the user must manually select several parameters. Some studies have proposed fully-automated ITMOs. Kovalski and Oliveira [27] used a bilateral



**FIGURE 2.** Overall framework of our proposed method comprising two main parts. The first part, *Dynamic Range Learning*, estimates HDR luminance from LDR luminance. The second part, *Color Difference Learning*, estimates HDR images by using the estimated HDR luminance and chromatic channels of LDR.

filtering-based high-quality edge-aware BEF that works automatically in real time. Eilertsen *et al.* [28] proposed a method for predicting clipped bright regions through a convolutional neural network (CNN), which works adaptively and automatically. However, they still focused on processing bright areas, and therefore showed effective improvement of the dynamic range of bright areas; however, the contrast in dark areas is insufficient. Actually, the HVS is more sensitive to bright areas than to dark areas [19], but contrast in dark areas is also very important to express real-world scenes. Therefore, we propose an adaptive ITMO using deep neural networks that improves the dynamic range in bright and dark areas and is applicable to various images.

### B. COLOR APPEARANCE FOR HIGH DYNAMIC RANGE

Most ITMOs apply their designed luminance mapping algorithms equally to each red (R), green (G), and blue (B) channel, equally. However, processing the same operation on each RGB channel does not produce good color representation of ground-truth HDR images because the RGB color space does not consider the color perception of the HVS. Therefore, another adequate operation considering the interaction between the luminance and the chromatic information is required to express the HDR image like a real scene.

Human visual perception has different sensitivities depending on the color and has high resolution in the green component whose wavelength is in the range 540–570nm [29], [30]. Because the RGB color space cannot consider human sensitivity to color, the CIEXYZ color space that considers the human color perception system has been developed [31], [32]. Furthermore, the CIEL\*a\*b\* color space has been developed for the purpose of independently representing luminance and chromatic information [33]; the L\* channel indicates luminance, and the a\*, and b\* channels indicate chromatic information.

In this paper, we propose an effective learning method in the CIEL\*a\*b\* color space rather than the RGB color space

to represent the color realistically by using a loss function designed for HVS.

### III. HDR GENERATION MODEL

Fig. 2 shows a diagram of the overall process of the proposed method. The procedure is as follows: 1) extract the luminance and chromatic information from the LDR image, 2) estimate the luminance of the HDR from the luminance of the LDR via the trained *Dynamic Range Learning* model, 3) estimate the chromatic information of the HDR by using the estimated HDR luminance obtained in step 2 and the chromatic information of the LDR via the trained *Color Difference Learning* model, and 4) recombine the estimated luminance and chromatic information.

#### A. COLOR SPACE CONVERSION

The primary difference between LDR and HDR images is the luminance. Therefore, the color space that expresses luminance and chromatic information independently is suitable for analyzing the characteristics of LDR and HDR. However, in RGB color space, luminance and chromatic information are mixed and not independent of each other in all R, G, and B channels. In contrast, they are independently expressed in the CIEL\*a\*b\* color space. Thus, we converted images into the CIEL\*a\*b\* color space for efficient analysis of the characteristics of LDR and HDR and their differences. RGB values can be converted into a CIEXYZ color space [31] that considers color sensitivity through equation (1). Then, CIEXYZ values can be converted into the CIEL\*a\*b\* color space through equation (2).  $X_n$ ,  $Y_n$ ,  $Z_n$  are the CIEXYZ tristimulus values of the reference white point. We set the CIE illuminant E ( $X_n = Y_n = Z_n = 1$ ) as the ideal white point [34], [35]

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.4124 & 0.3576 & 0.1805 \\ 0.2126 & 0.7152 & 0.0722 \\ 0.0193 & 0.1192 & 0.9505 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

$$L^* = 116f\left(\frac{Y}{Y_n}\right) - 16$$

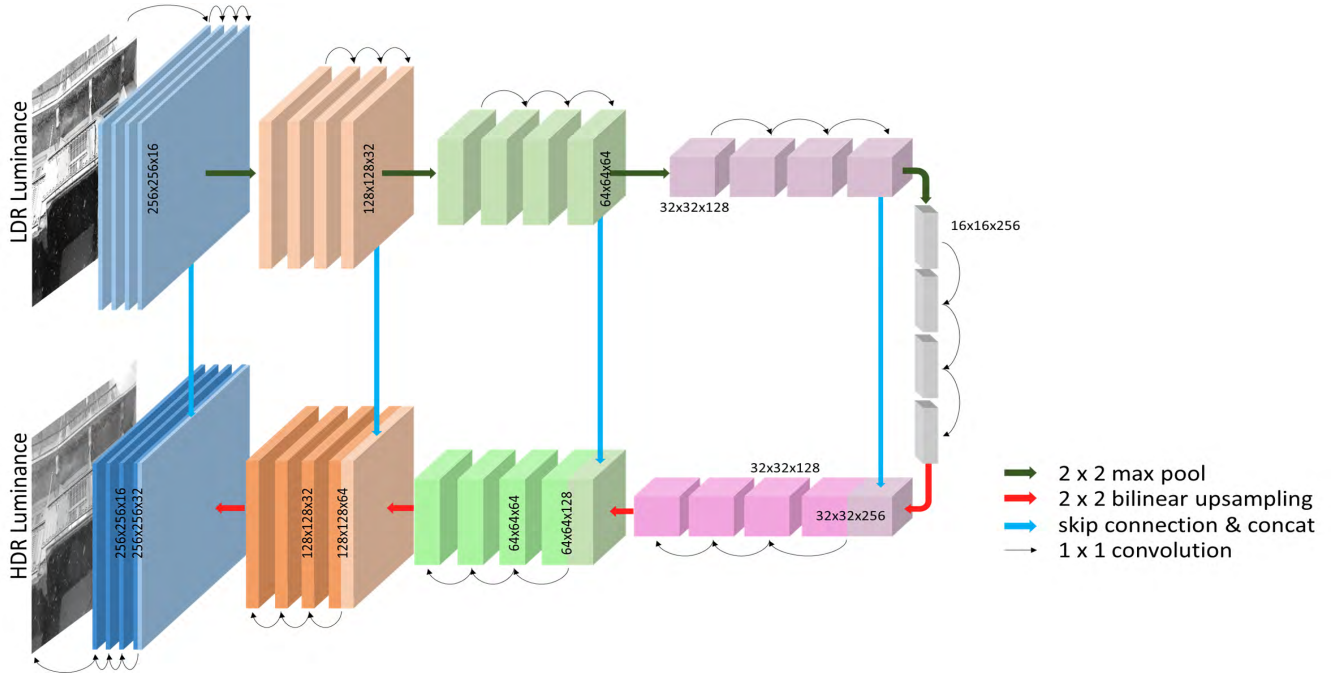


FIGURE 3. Architecture of Dynamic Range Learning Network.

$$\begin{aligned}
 a^* &= 500 \left( f \left( \frac{X}{X_n} \right) - f \left( \frac{Y}{Y_n} \right) \right) \\
 b^* &= 200 \left( f \left( \frac{Y}{Y_n} \right) - f \left( \frac{Z}{Z_n} \right) \right) \\
 f(t) &= \begin{cases} \sqrt[3]{t} & \text{if } t > \delta^3 \\ \frac{t}{3\delta^2} + \frac{4}{29} & \text{otherwise,} \end{cases} \quad \text{where } \delta = \frac{6}{29} \quad (2)
 \end{aligned}$$

### B. LOG TRANSFORMATION

Since LDR images use 8 bits to express the levels of intensity for a single pixel and HDR images use  $\geq 10$  bits, there are big differences in the luminance distributions over the entire images between the LDR and HDR pairs. Even when the luminance values are expressed between 0 and 100 in CIEL\*a\*b\* color space, the distribution of luminance values is quite different between them. This big difference makes it difficult for the artificial neural networks to learn the relationship between LDR and HDR pairs. In our preliminary studies, our neural networks failed to learn this relationship. Therefore, we applied log transformation on HDR luminance as a pre-processing step to reduce the numerical difference in luminance distribution between LDR and HDR pairs. After log transformation, we adjusted the converted luminance range to be from 0 to 100 to match that of the LDR images. We found that this transformation helped our neural networks to effectively learn the luminance relationship between LDR and HDR pairs.

### C. DYNAMIC RANGE LEARNING

We designed the learning architecture to estimate the luminance of HDR from the luminance of LDR, as shown

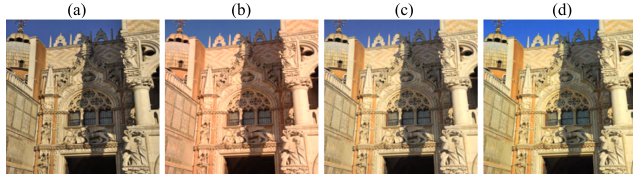
in Fig. 3. Our model is based on a U-net architecture;  $2 \times 2$  max pooling layers and upsampling layers enabled for extracting features at multiscale, and our proposed model is designed to learn not only simple low-level features but also high-level features [11].

For learning-based on local information, a  $3 \times 3$  convolution layer is used in general. However, our experiment shows that the  $3 \times 3$  convolution layer generates halo artifacts and blur (the results of our experiments are presented in Section 4). Halo artifacts occurred in most operators for tone reproduction [36]–[39]. Halo artifacts are easily observed in edges with a large intensity gradient. Halo artifacts occur when filters are used without considering edges so that the intensity of the bright region affects the dark region and vice versa [40]. Thus, we designed a model using  $1 \times 1$  layers instead of  $3 \times 3$  convolution layers.

The  $1 \times 1$  layer helps in obtaining a pixel-wise global mapping function. However, a learning model that is stacked only on  $1 \times 1$  layers cannot design an adaptive global mapping function considering various characteristics of images, such as the maximum and minimum intensities of luminance. Therefore, we used the whole-size LDR–HDR pairs instead of using the patches cropped from the whole images to learn the characteristics of the images. The pooling and upsampling layer can help to acquire an adaptive global mapping function by learning features on multiscale.

Autoencoder models, such as U-net referenced in this paper, generally use deconvolution layers (sometimes called transposed convolution) for upsampling feature maps. It has been reported that the deconvolution layer overlaps pixels in the upsampled image if the kernel size is not divided by the stride and the image generated through multiple





**FIGURE 4.** Examples of incomplete color representation in HDR estimation. (a) and (b) are the LDR and its ground-truth HDR image, respectively. The estimated HDR images were produced by using the ground-truth HDR luminance with (c): its application to RGB channels, and (d): use of the LDR chrominance. Even though the luminance is the same as that of the ground-truth HDR image, the color appearance is different. (all HDR images were tone mapped by using Reinhard's et al).

deconvolution layers causes checkerboard artifacts of various scales [41]. Therefore, we used a bilinear upsampling method instead of deconvolution layers to prevent checkerboard artifacts so that the overlapping does not occur during the upsampling process. Regarding the loss function in the learning process, we used  $L_2$ -norm of  $L_{HDR}$  and  $L_{LDR}$  as described in equation (3).

$$Loss_{L^*} = \sqrt{\sum (L^*_{HDR} - L^*_{LDR})^2} \quad (3)$$

#### D. COLOR DIFFERENCE LEARNING

Most ITMO methods focus on predicting only the luminance information and represent HDR images by computing the predicted luminance mapping function for each color channel. However, because the RGB color space does not consider the HVS, even if the luminance information is accurately predicted, processing in the RGB channels does not result in proper HDR color representation of the ground-truth HDR images. Fig. 4(c) shows the HDR image produced by multiplying the luminance ratio of the ground-truth HDR and the LDR images by the RGB color channels of the LDR image based on equation (4) [23]. This HDR estimate does not properly render the color information of the ground-truth HDR image.

$$\begin{bmatrix} R_{HDR} \\ G_{HDR} \\ B_{HDR} \end{bmatrix} = \frac{L^*_{HDR}}{L^*_{LDR}} \begin{bmatrix} R_{LDR} \\ G_{LDR} \\ B_{LDR} \end{bmatrix} \quad (4)$$

Similarly, a simple processing in the CIEL\*a\*b\*, even if the HDR luminance is perfectly predicted, it cannot display the proper color of the ground-truth HDR image if the chromatic channels are simply taken from the LDR image. Fig. 4(d) shows the HDR image reconstructed by replacing  $L^*$  of LDR with  $L^*$  of HDR, and the chromatic channels are retained from LDR. There is a noticeable color difference between this HDR estimate and the ground-truth HDR image. Therefore, it is necessary to predict the chromatic information as well as the luminance information of HDR; thus, we designed a *Color Difference Learning* method.

*Color Difference Learning* is performed after *Dynamic Range Learning*. The main difference between LDR and HDR images is luminance; therefore, reducing the luminance

difference by performing *Dynamic Range Learning* first is very efficient for the subsequent *Color Difference Learning*. The learning architecture for *Color Difference Learning* shown in Fig. 5 is similar to the *Dynamic Range Learning* architecture. The estimated HDR luminance via *Dynamic Range Learning* and the chromatic channels of LDR are used as input to this network.

In *Color Difference Learning*, we used a loss function that is based on the HVS characteristics.  $\Delta E_{RGB}$  in equation (5), defined in Euclidean distance in the RGB color space, has a very low correlation with human color perception because the human eye has different sensitivities to the R, G, and B channels, and changes in the intensity of each component produce different sensations [42]. However,  $\Delta E^*_{ab}$  in equation (6), defined as the root mean square of  $L^*$ ,  $a^*$ , and  $b^*$ , has a high correlation with the color perception of the human retina [43]. Therefore, to make the color appear more similar to the ground-truth HDR in a human sense, we used  $\Delta E^*_{ab}$  as a loss function instead of  $\Delta E_{RGB}$ .

$$\Delta E_{RGB} = \sqrt{\Delta R^2 + \Delta G^2 + \Delta B^2} \quad (5)$$

$$\Delta E^*_{ab} = \sqrt{\Delta L^{*2} + \Delta a^{*2} + \Delta b^{*2}} \quad (6)$$

## IV. EXPERIMENT

### A. DATASET

A large number of various datasets are needed to develop an adaptive ITMO via deep learning. We used a raw image dataset (RAISE) comprised of 8,156 high resolution raw images with various scenes [44]. All images were shot by using three different cameras with average exposure values of 12.5, 13.9, and 14.0 EVs, which had various color temperatures ranging from 2,000 K to 8,000 K and scenes classified as “Outdoor,” “Indoor,” “Landscape,” “Nature,” “People,” “Objects,” and “Buildings.” The images with camera specifications are available from a website (<http://mmlab.science.unitn.it/RAISE/>).

We generated a set of LDR–HDR pair images from a single raw image. The raw images were 12- to 14-bit images taken from digital cameras without any compression and contained all of the original intensity values [45]. From a single raw image, five LDR images were generated with different exposure values (−2, −1, 0, 1, 2) by using a camera response function [46]. An HDR image was then produced by using the HDR Pro algorithm developed by Debevec and Malik [5] (built in Adobe Photoshop CC 2018) to synthesize these five LDR images. The LDR image with an exposure of 0 was paired with the HDR image to be an LDR–HDR pair image set for our experiments. Examples of LDR–HDR images from the RAISE dataset are shown in Fig. 6. All HDR images were tone mapped by using Reinhard's TMO for displaying on a typical 8-bit LDR panel display [24] for readers of this paper.

In addition to the RAISE dataset, totally different types of datasets were also used to verify our learning-based ITMO. A total of 50 LDR–HDR pair images were obtained from Fairchild [47] and EPFL's HDR-Eye [48] datasets. The HDR

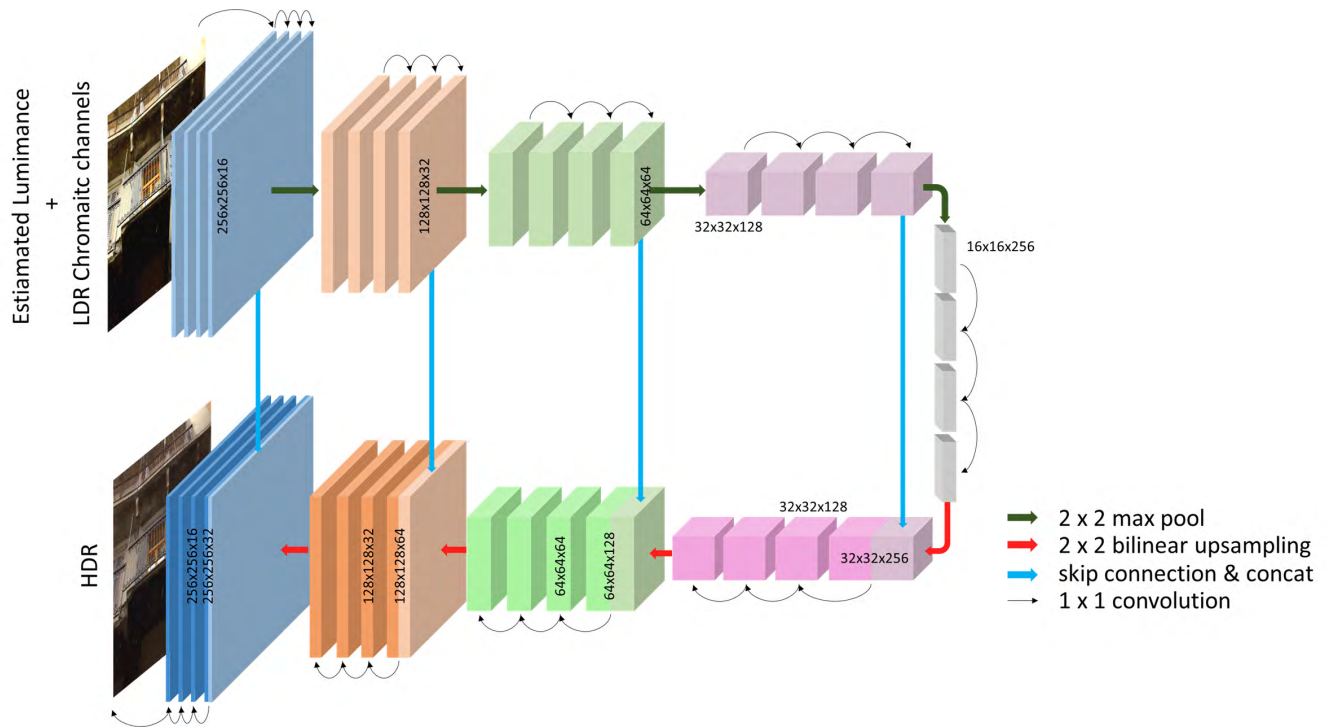


FIGURE 5. Architecture of color difference learning network.

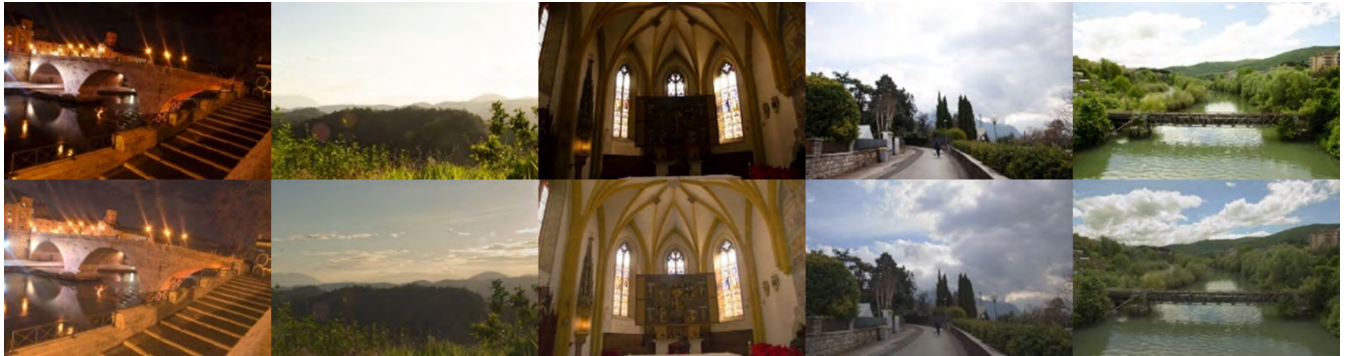


FIGURE 6. LDR-HDR image pairs in the RAISE dataset. Images on the top row are the LDR images, and the corresponding ground-truth HDR images are shown on the bottom row. HDR images were tone mapped by using Reinhard's TMO for a typical LDR panel display.

images in these datasets were produced from several LDR images that were physically and individually taken with different exposures in camera settings [49]. HDR images in the Fairchild dataset were produced by synthesizing nine LDR images with different exposures taken by a Nikon D2x camera. HDR images in the HDR-Eye dataset were produced by synthesizing nine LDR images with different exposures taken by several cameras, including a Sony DSC-RX100 II, Sony NEX-5N, and Sony alpha 6000. Our ITMO trained with the RAISE datasets was tested on these different types of LDR-HDR pairs for verification.

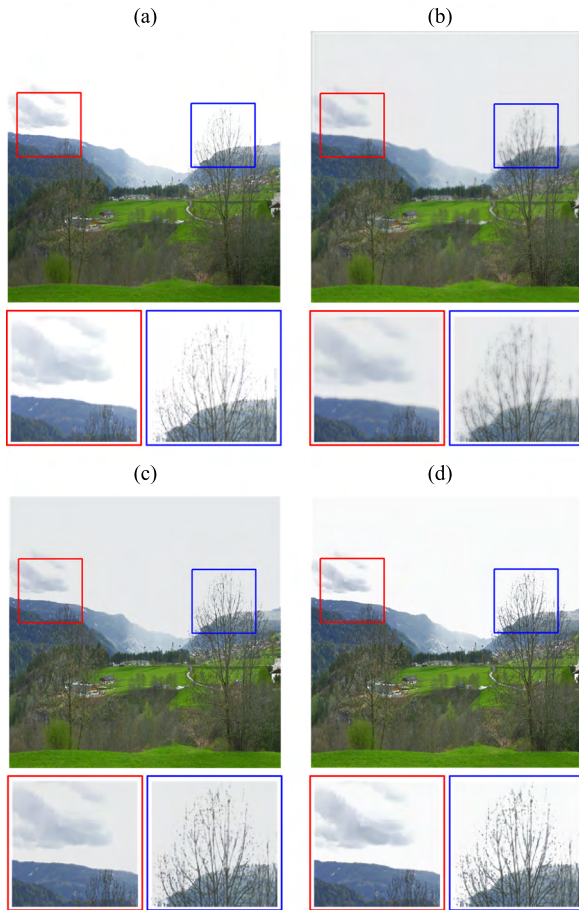
The total 8,156 pairs of the RAISE images were divided into 8,000 and 156 images. The 156 images were used as a test set, and the remaining 8,000 images were used for training

and validation sets. Furthermore, the 8,000 images were divided into 5-folds for extensive cross-validation: 4 subsets were used for training, and the remaining subset was used for validation. For all combinations of training and validation sets, a total of 5 models were trained and checked for overfitting.

### B. GLOBAL ITMO AND HALO ARTIFACT

Fig. 7 compared the estimated HDR images by using three different models: The first model was simply stacked with 20 convolution layers without pooling layers, and each layer comprised 64 convolution kernels with a kernel size of  $3 \times 3$ . The second model differed from the first model by using a kernel size of only  $1 \times 1$ . The third model was the proposed





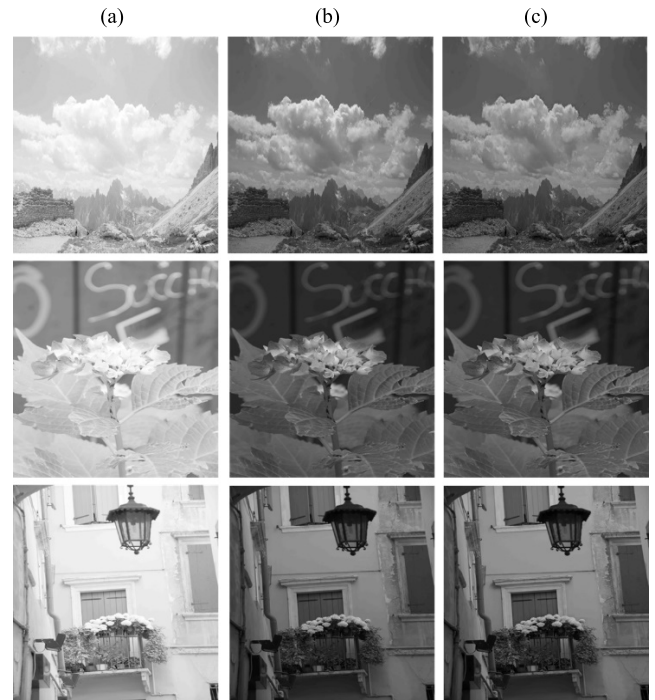
**FIGURE 7.** Comparison of the estimated HDR images with various training models. (a) shows the ground-truth HDR image and (b), (c), (d) are reconstructed HDR images using chromatic channels from the LDR image and estimated luminance from various trained models. Each trained model comprises (b) 20 convolution layers with kernel size  $3 \times 3$ , (c) 20 convolution layers with kernel size  $1 \times 1$ , and (d) the proposed model shown in Fig. 3, respectively. Halo artifacts and blur artifacts occur in (b). The proposed model can estimate luminance more accurately than that of the other models.

model comprising  $1 \times 1$  kernels,  $2 \times 2$  max pooling, and bilinear upsampling layers. All estimated HDR images were reconstructed from the estimated HDR luminance by using these models and the same chromatic channels of the LDR image.

The HDR images estimated by using  $3 \times 3$  kernels generated blur artifacts and halo artifacts. Blur artifacts were observed in the branches in the blue box and in the clouds in the red box. Halo artifacts can be easily detected in an edge with a large intensity gradient, such as the edge between the mountain and the sky in the red box of Fig. 7(b). The estimated HDR images using  $1 \times 1$  kernels do not generate blur artifacts and halo artifacts. However, both models that use only convolution layers cannot consider the characteristics of the whole image, and the global mapping function from the two trained models cannot estimate the luminance accurately. The estimated luminance using only  $3 \times 3$  kernels and  $1 \times 1$  kernels are different from the luminance of the



**FIGURE 8.** Comparison of HDR images generated using the deconvolution layers and the bilinear upsampling layers. Top-row images are reconstructed using deconvolution layers. The deconvolution layer causes the checkerboard artifacts and they can be easily detected in the sky. Bottom-row images are reconstructed by using bilinear upsampling layers, and the checkerboard artifacts are reduced.



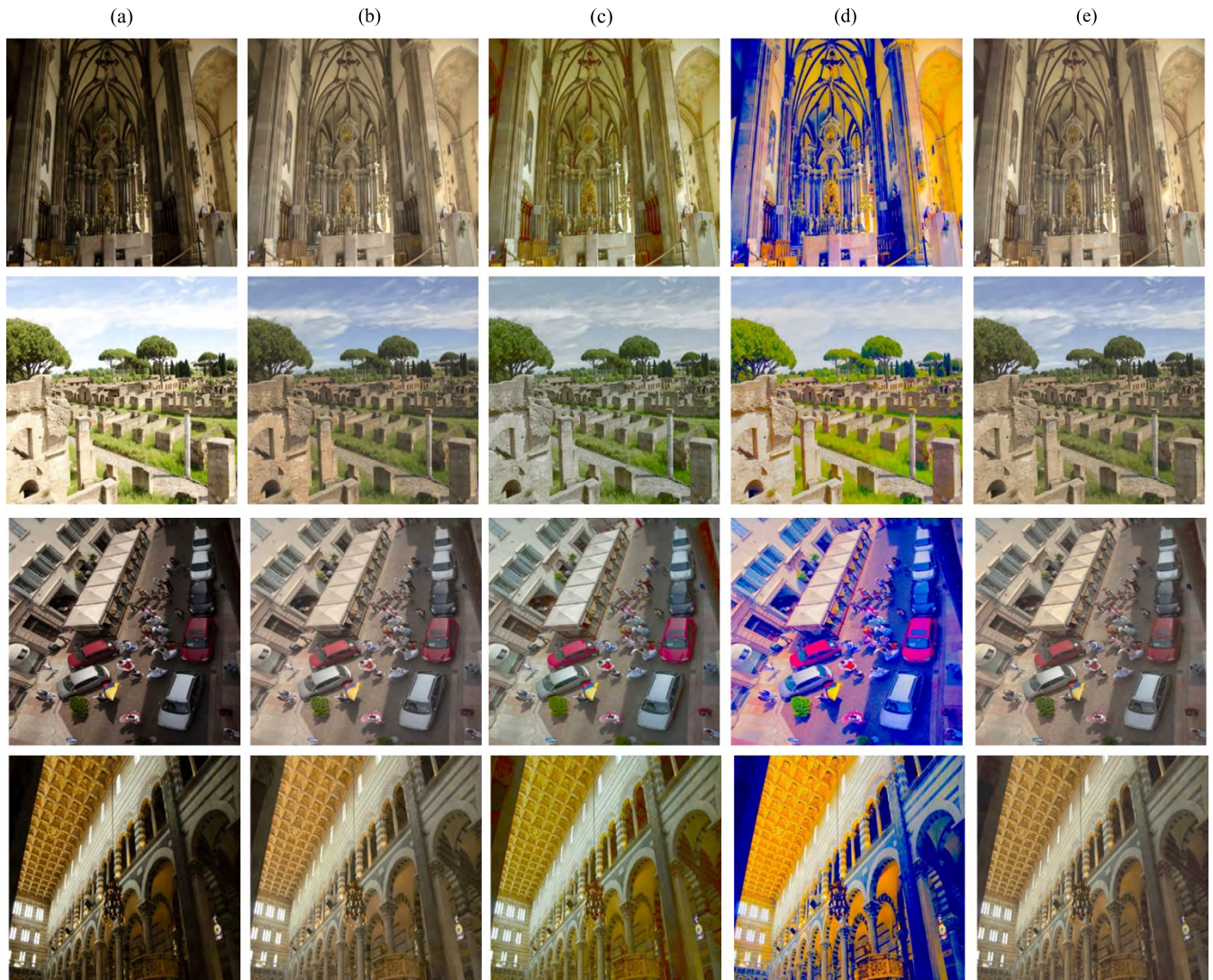
**FIGURE 9.** Estimated luminance from *Dynamic Range Learning*. *Dynamic Range learning* can recover saturated luminance similar to that of ground-truth HDR. (a) LDR. (b) Ground-truth HDR. (c) Estimated luminance.

ground-truth HDR, such as that of the sky in the blue boxes of Fig. 7(b) and Fig. 7(c). The proposed model can consider the characteristics of the whole image. Therefore, the estimated HDR images using the proposed model can accurately estimate the luminance, similar to that of the ground-truth HDR without blur artifacts and halo artifacts shown in Fig. 7(d).

### C. CHECKERBOARD ARTIFACT

The high-level features extracted through *Dynamic Range Learning* are represented in the image through multiple upsampling operations. Fig. 8 compares the estimated luminance of HDR for two models by using deconvolution and bilinear upsampling layers.





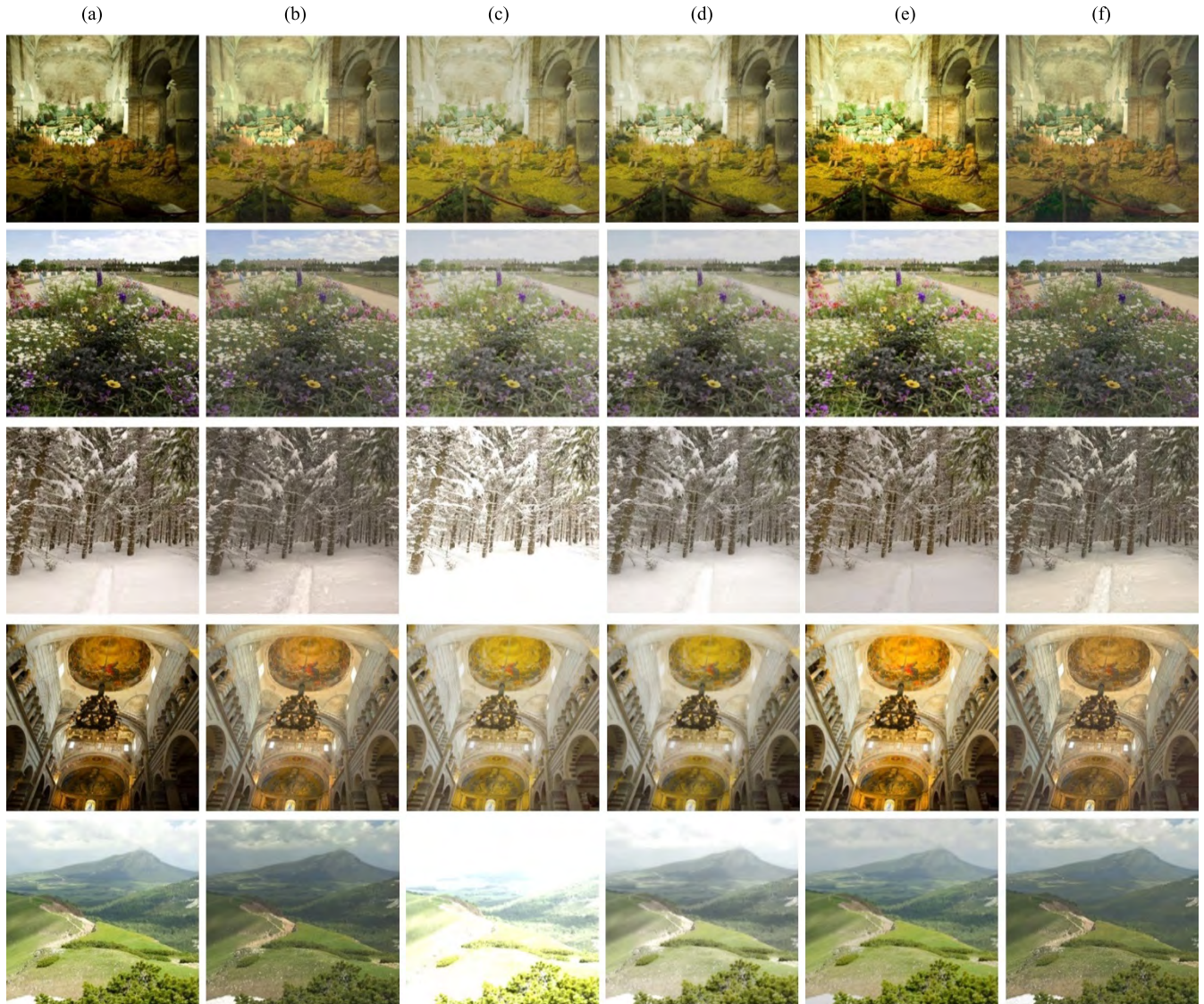
**FIGURE 10.** The results of *Color Difference Learning*. (a) and (b) are the LDR and ground-truth HDR images, respectively. (c) is reconstructed from the luminance estimated from *Dynamic Range Learning*. Chromatic information is the same as that of the LDR images. (d) and (e) are estimated images from RGB channel learning and *Color Difference Learning*, respectively. The transferred red colors are over-expressed, and the transferred green colors, such as those of leaves, are unnatural when using  $\Delta E_{RGB}$  as a loss function. However, the estimated chromatic information that is learned by using  $\Delta E_{ab}^*$  as a loss function is very similar to that of the ground-truth HDR images.

Some artifactual checkerboard patterns are detected when using deconvolution layers, as shown in the first-row images of Fig. 8. The high-level features, which are in low resolution, have to be represented in high resolution through the upsampling process in the proposed training model. Because the deconvolution operation can overlap high-level features in the upsampling process, the multiple deconvolution layers cause checkerboard artifacts in various scales. The checkerboard artifacts can be easily detected in flat areas better than in complex areas because the differences in luminance intensities in repetition are very small. However, the bilinear interpolation operations can upsample high-level features without overlapping; therefore the checkerboard artifacts are reduced, as shown in the second-row images of Fig. 8.

#### D. DYNAMIC RANGE LEARNING

Fig. 9 shows the HDR luminance images predicted by *Dynamic Range Learning*. The left, center, and right images are the LDR, ground-truth HDR, and estimated HDR luminance images, respectively. There are significant differences in luminance between the LDR and the ground-truth HDR luminance images. The detailed contrasts within the clouds (top row), flowers and leaves (middle row), and textures on the wall (bottom row) are not easily perceived in the LDR luminance images (a), whereas the HDR luminance images clearly visualize those detailed contrasts. Comparison of the ground-truth HDR luminance images (b) with the estimated HDR images (c) demonstrates that our *Dynamic Range Learning* networks more accurately estimated the ground-truth HDR luminance from the single LDR luminance (a).





**FIGURE 11.** Comparison with other existing ITMOs (Akyuz, Huo, Eilertsen ITMOs). In terms of both color appearance and dynamic range, compared with the other ITMOs, the proposed method produced HDR images that are the most similar to the ground-truth HDR images. (a) LDR. (b) Ground-truth HDR. (c) Akyuz. (d) Huo. (e) Eilertsen. (f) Proposed.

### E. COLOR DIFFERENCE LEARNING

Fig. 10 shows the final resulting images of *Color Difference Learning* after *Dynamic Range Learning*. Fig. 10(a) and (b) are the LDR and ground-truth images, respectively. Fig. 10(c) shows the images that combine the estimated luminance by *Dynamic Range Learning* with the chromatic information of the LDR image. Fig. 10(d) and (e) are the HDR images estimated by using the deep learning techniques with  $\Delta E_{RGB}$  and  $\Delta E_{ab}^*$  as loss functions, respectively.

Colors cannot be distinguished in very bright or dark regions in LDR images. In contrast, all of the estimated HDR images can be distinguished by their colors. However, the color is still not sufficiently similar to the ground-truth HDR. The HDR images reconstructed by combining the estimated HDR luminance and chromatic information of the

LDR images can only attain the HDR luminance and cannot properly represent the color of the ground-truth HDR images. The HDR images estimated by using  $\Delta E_{RGB}$  as a loss function also show different color representations from those in the ground-truth HDR images. Colors became excessively red or green. In particular, human perception has a high sensitivity in the green component [30]; therefore, the trained model using  $\Delta E_{RGB}$  cannot express green colors well. In contrast, the HDR images estimated by using the proposed method show natural color representations and contrasts similar to those of the ground-truth HDR images.

### F. COMPARISON WITH CONVENTIONAL ITMO

Fig. 11 shows the HDR images estimated by using the proposed and conventional ITMO methods; Akyüz [19],





**FIGURE 12.** The results from other types of datasets not used for training. Although these LDR–HDR images were produced differently and provided by other sources (Fairchild and HDR-Eye datasets), HDR images were more accurately estimated from the corresponding single LDR images. (a) LDR. (b) Ground-truth HDR. (c) Proposed.

Eilertsen *et al.* [28], and Huo *et al.* [50] ITMOs are used for comparison. As mentioned in Section 2, some limitations of the conventional ITMOs can be observed. The estimated luminance from Huo and Eilertsen ITMOs could not sufficiently improve the dynamic range of under-exposed regions. However, the proposed method did improve the dynamic range in those under-exposed areas. Akyuz ITMO improved the visibility of the under-exposed area, but the estimated images are quite different from the ground-truth HDR images in terms of the global contrast because the Akyuz ITMO extends the dynamic range linearly. In terms of color appearance, only the Eilertsen ITMO tries to change the chromatic information via a learning process using R, G, and B channels. However, the inferred HDR images deviate from the ground-truth HDR images in terms of luminance contrast and color appearance. The proposed ITMO method resulted in high quality HDR images that are most comparable to the ground-truth HDR images.

## G. RESULTS FROM OTHER TYPES OF DATASETS NOT USED FOR TRAINING

For the training of our neural networks, the LDR–HDR pairs obtained from the RAISE dataset were used as the input and label images. Because both the LDR and HDR images used in the training stage were produced from the corresponding single raw file as explained earlier, we performed additional tests of our trained model on totally different types of datasets: the Fairchild [47] and HDR-Eye [48] datasets. Fig. 12(a) and (b) show the LDR images and ground-truth HDR images provided by the Fairchild/ HDR-Eye datasets, respectively. It is noted that nine different LDR images were physically taken with nine different exposure values and combined to produce the corresponding HDR image [49]. Fig. 12(c) shows the estimated HDR images obtained by our trained model using the single LDR images in (a). Comparison between (b) and (c) demonstrates that our trained model successfully estimated the ground-truth HDR images from the single LDR images even though these LDR–HDR datasets were totally different from the datasets used in the training.

## V. EVALUATION RESULTS

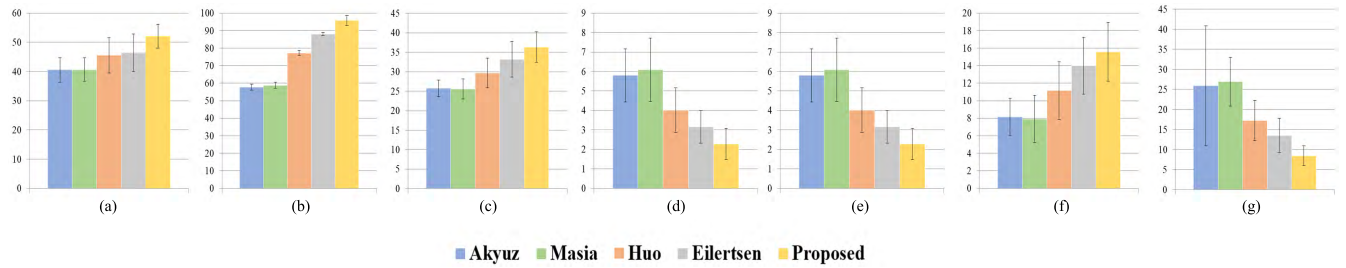
### A. OBJECTIVE EVALUATION

The performance of the proposed method was evaluated by using objective image quality evaluation metrics to compare it with other ITMO technologies. Akyuz, Masia [20], Huo, and Eilertsen ITMOs were included for comparison.

Because HDR images have a wider dynamic range than that of LDRs, general evaluation methods, such as the mean square error (MSE) or peak signal to noise ratio (PSNR), are not appropriate; the correlation with human subjective evaluation is slightly lower [51]. Therefore, we used the following HDR-specific image quality metrics: the high dynamic range visual difference predictor-2.2 (HDR-VDP-2.2) [52], perceptually uniform peak signal to noise ratio (PU\_PSNR), perceptually uniform structural similarity index measure (PU\_SSIM), perceptually uniform root mean square error (PU\_RMSE) [53], root mean squared logarithmic error (RMSLE) and multiple-exposure peak signal to noise ratio (mPSNR) [54], which have high correlations with human observed image characteristics. The deltaE1976 color difference metric is additionally used to evaluate the color appearance, because the other metrics only evaluate the dynamic range [51].

Fig. 13 and Table 1 show the overall comparison of several ITMO methods performed by using these metrics. The metric values are averaged values from 206 test images. (156 RAISE test images, 50 Fairchild & HDR-Eye test images)

Our proposed method performed significantly better than other ITMOs in all metrics considering the dynamic range. Regarding color appearance, the smaller values of deltaE1976 indicate that the color representation is closer to that of the reference. The performance of our proposed



**FIGURE 13.** Comparison with conventional ITMOs evaluated by using the following objective HDR metrics: (a) HDR-VDP-2.2, (b) PU-SSIM-Y, (c) PU-PSNR-Y, (d) PU-RMSE-Y, (e) RMSLE (f) mPSNR, and (g) deltaE1976. The proposed method shows superior performance in all metrics.

**TABLE 1.** Quantitative performance evaluation of several ITMOs using HDR-specific evaluation metrics.

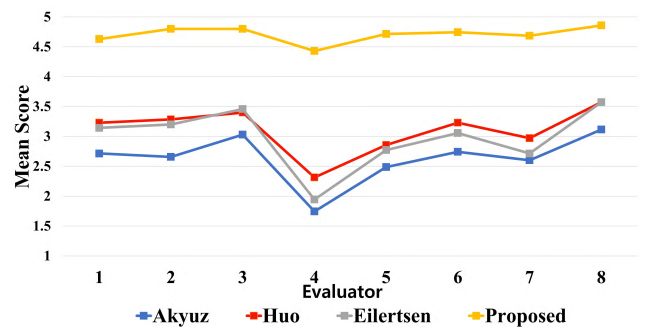
	Akyuz	Masia	Huo	Eilertsen	Proposed
HDR_VDP_2.2	41.501	42.091	45.967	47.187	52.067
PU_SSIM_Y	0.528	0.591	0.721	0.873	0.952
PU_PSNR_Y	25.695	26.335	29.161	32.933	36.524
PU_RMSE_Y	30.474	29.017	21.502	14.26	9.350
RMSLE	5.2555	5.0875	3.9964	2.9174	2.2775
mPSNR	8.361	8.743	11.153	14.07	15.577
deltaE1976	25.602	24.721	18.073	13.358	8.480

method in deltaE1976 was also substantially superior to that of other ITMOs.

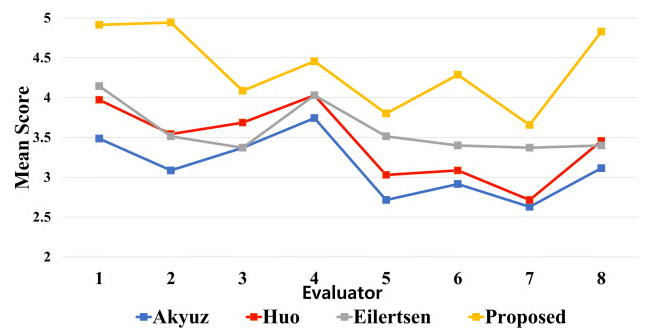
## B. SUBJECTIVE COMPARISON

We performed two subjective assessments. The first was an evaluation of the similarity of the proposed method images with the ground-truth HDR images in terms of color appearance. The overall evaluation method followed the recommendation of the Double-Stimulus Continuous-Quality Scale method by Rec. ITU-R BT.500-10 [55]. We provided the test images together with the ground-truth images to evaluators who were carefully trained to score the color similarity between two images from 1 to 5. A score of 5 indicated that the images were very similar, and a score of 1 indicated that the images were very different from each other in terms of color naturalness and similarity. Evaluations were performed three times in three different sessions. For displaying the HDR images, a 49-inch Samsung UN49KS8000F HDR display was used with a maximum displayable luminance of  $1,000 \text{ cd/m}^2$ . The backlight of the evaluation room was removed and the distance between the evaluators and the HDR TV was 2 m. Eight human evaluators with normal visual acuity and color vision participated. Among 206 test images, 40 images were randomly selected, and their HDR estimates by Akyuz, Huo, Eilertsen, and the proposed ITMOs were evaluated. In total, 160 images ( $40 \times 4$ ) were scored

by individual evaluators. Test images for evaluation were resized to a  $1024 \times 1024$  resolution for a side-by-side comparison with the ground-truth HDR images. Eight evaluators scored the test images three times in three different sessions. Fig. 14 shows the mean scores of individual evaluators for two sessions (the first session was excluded because it was for practice). Our proposed method demonstrated a significantly better performance than that of all other methods.



**FIGURE 14.** Subjective evaluation result of color appearance for each ITMO. The scores from eight evaluators show that the proposed method HDR images are more similar to the ground-truth HDR images than that of other ITMO methods in color appearance.



**FIGURE 15.** Subjective evaluation results of how similar the estimated HDR images are to a real-world scene for each ITMO. The scores from eight evaluators show that the proposed method HDR images are more realistic than those of the other ITMO method images.

Another subjective assessment was performed to evaluate how realistic the estimated HDR images were. Only HDR estimates were provided without reference images and scored within the range of 1 to 5 considering the naturalness of images and the contrasts for bright and dark areas, with



5 indicating the best and 1 indicating the worst. The resolution of all test images in this assessment was set to  $2048 \times 2048$ . Eight evaluators scored the test images three times in three different sessions. Fig. 15 shows the mean scores of the individual evaluators for two sessions (the first session was excluded because it was for practice). The high scores for the proposed method in this assessment suggest that the HDR images estimated by using the proposed method look more realistic than the images estimated by other ITMO methods on an HDR display.

## VI. DISCUSSION AND CONCLUSION

Various ITMOs have been proposed to generate HDR images. Most ITMOs are based on mathematical models to extend the dynamic range of LDR images. Very recently, a few papers were published that described the use of CNN learning techniques for ITMOs. All of these methods show good performance but still have some limitations in displaying realistic HDR images. In this paper, we designed an effective learning model based on a loss function considering the HVS and trained the model by using various LDR – HDR pair datasets to overcome the limitations. The experimental results showed that our trained model could estimate HDR images from a single LDR image that were similar to a real-world scene. Both objective and subjective evaluations have demonstrated performance of our model superior to that of the conventional ITMOs in terms of both dynamic range and color appearance.

However, there are some limitations. Images made by using our method show a slight difference in bright areas relative to those in ground-truth HDR images. This difference might have been caused by using deltaE1976 in designing the loss function for training. It has been reported that deltaE1976 is not as perceptually uniform as intended in saturated regions [56]. Training with deltaE2000 [57], which complements perceptual uniformity, is expected to show improved color appearance, particularly in saturated areas.

The LDR – HDR pair images used in the training of our model were produced from the corresponding single raw files in the RAISE dataset. Nevertheless, our trained model successfully estimated the ground-truth HDR images in different types of datasets provided by other sources (Fairchild and HDR-Eye datasets). The generation of LDR images, exposure values, number of LDR images used to produce HDR images, and types of cameras were all different from those of the LDR – HDR pairs of the RAISE datasets used for the training. This test result on different types of image sets demonstrated that our trained model was robust in estimating HDR images from various types of single LDR images.

In conclusion, a robust inverse tone mapping operator using cascaded neural networks was designed and trained by using a sequential learning process to effectively learn the differences between LDR and HDR images. The luminance learning process first reduced the significant gap between LDR and HDR luminance, and then the chrominance learning process effectively learned the color differences based on the properties of the HVS and finally inferred the

HDR color appearance. Both objective and subjective evaluations demonstrated the robust performance of our trained model for various types of LDR – HDR image pairs.

## APPENDIX

### HIGH DYNAMIC RANGE IMAGE QUALITY ASSESSMENT

#### 1. HDR-VDP-2.2

HDR-VDP-2.2 shows the highest correlation with the subjective quality evaluation among objective quality assessment metrics [51]. The quality  $Q$  of HDR images was evaluated per equation (7),

$$Q_{HDR-VDP} = \frac{1}{F \times Q} \sum_{f=1}^F \sum_{o=1}^O w_f \log\left(\frac{1}{I} \sum_{i=1}^I D_p^2[f, o](i) + \delta\right) \quad (7)$$

where  $f$  and  $o$  are the spatial frequency from 1 to  $F$  and orientation from 1 to  $O$  respectively,  $D_p$  is the difference in normalized noise between the  $f^{\text{th}}$  and  $o^{\text{th}}$  of the steerable pyramid for the reference and test images,  $i$  is the index of the pixel,  $I$  is the total number of pixels,  $\delta$  is small constants that is set  $10^{-5}$  to avoid singularities when  $D$  is close to 0, and  $w$  is pooling weights per band and is set by constrained optimization to avoid overfitting.

2. Perceptually Uniform (PU) Encoding-based Metrics  
Luminance values are not perceptually uniform for humans. Many quality metrics uses gamma correction and a log function to transform the input image to achieve perceptual uniformity. Therefore, we also used a log function to evaluate image quality more perceptually. The root mean square logarithmic error (RMSLE) metric is computed per equation (8).

$$Q_{RMSLE} = \sqrt{\frac{1}{M \times N} \sum_i^M \sum_j^N [\log G(x, y) - \log E(x, y)]^2} \quad (8)$$

where  $G$  and  $E$  are the ground-truth HDR image and the estimated HDR image, respectively,  $(x, y)$  is the pixel index, and the image size is  $M \times N$ .

However, the log function gives larger differences for much brighter displays, such as HDR displays. Therefore, Narwaria *et al.* [52] proposed PU encodings, which transforms the luminance values into perceptually uniform values for evaluating the quality of HDR images. The evaluation procedure is as follows: 1) Convert pixel values of both the ground-truth HDR image and estimated HDR image to luminance, 2) re-encode the luminance of both images by using PU encoding, and 3) perform quality assessments, such as SSIM, PSNR and RMSE.

The look-up table and Matlab code for the PU encoding are provided at [http://www.mpii.mpg.de/resources/hdr/fullhdr\\_extension/](http://www.mpii.mpg.de/resources/hdr/fullhdr_extension/).

The SSIM metrics are evaluated per equation (9). The images are divided into a total of  $I^{\text{th}}$  patches with window size 11, and  $i$  is the index of the patch. The indexes  $g$  and  $e$  are the patches of the PU-encoded ground-truth HDR image

and estimated HDR image, respectively,  $\mu_g$  and  $\mu_e$  are the mean intensities of patches g and e, respectively,  $\sigma_g$  and  $\sigma_e$  are standard deviation of patches g and e, respectively,  $\sigma_{ge}$  is a structural term that is evaluated per equation (10), j is the index of the pixel in the patch, and N is the total number of pixels in the patch. We set  $C_1 = 0.01$ ,  $C_2 = 0.03$ .

$$Q_{SSIM} = \frac{1}{I} \sum_{i=1}^I \frac{(2\mu_g\mu_e + C_1)(2\sigma_{ge} + C_2)}{(\mu_g^2 + \mu_e^2 + C_1)(\sigma_g^2 + \sigma_e^2 + C_2)} \quad (9)$$

$$\sigma_{xy} = \frac{1}{N} \sum_{j=1}^N (g_j - \mu_g)(e_j - \mu_e) \quad (10)$$

The PSNR metrics are evaluated per equation (11).

$$Q_{PSNR} = 10 \log_{10} \left[ \frac{\max(E)^2}{\frac{1}{M \times N} \sum_i^M \sum_j^N [G(x, y) - E(x, y)]^2} \right] \quad (11)$$

where G and E are the PU-encoded ground-truth HDR image and estimated HDR image, respectively, (x, y) is the pixel index, the image size is  $M \times N$ ,  $\max(E)$  is the maximum possible pixel value of the estimated HDR image, and if the estimated HDR image has n bits, then  $\max(E) = 2^n - 1$ .

The RMSE metrics are evaluated per equation (12).

$$Q_{RMSE} = \sqrt{\frac{1}{M \times N} \sum_i^M \sum_j^N [G(x, y) - E(x, y)]^2} \quad (12)$$

where G and E are the PU-encoded ground-truth HDR image and estimated HDR image, respectively, (x, y) is the pixel index, and the image size is  $M \times N$ .

3. Multi-exposure PSNR [mPSNR; *High Dynamic Range Texture Compression for Graphics Hardware*, Jancob 2006] is designed to evaluate HDR images at various exposure levels. The HDR image is tone mapped to a number of different exposure levels. Then, the image quality is evaluated with PSNR using the mean of all MSEs. We used a simple gamma correction after exposure compensation in equation (13) as a tone mapping operator.

$$I_{TM} = \text{Clamp}_{0,255} \left[ 255 \times (2^i \times I_{HDR})^{\frac{1}{\gamma}} \right] \quad (13)$$

where  $I_{HDR}$  and  $I_{TM}$  are the HDR image and tone mapped image, respectively,  $i$  is the exposure compensation in f-stops, and we set  $\gamma = 2.2$ .  $\text{Clamp}_{0,255}$  indicates clamping to the integer ranges [0, 255]. Finally, mPSNR is evaluated per equation (14),

$$Q_{mPSNR} = 10 \log_{10} \left[ \frac{3 \times 255^2}{\frac{1}{M \times N \times 3} \sum_i^M \sum_j^N [(\Delta R_{ij})^2 + (\Delta G_{ij})^2 + (\Delta B_{ij})^2]} \right] \quad (14)$$

$$\Delta R_{ij} = I_{TM,R}(G) - I_{TM,R}(E)$$

where G and E are the ground-truth HDR image and estimated HDR image, respectively, and  $\Delta R_{ij}$  is the difference in the red components between tone mapped G and E at pixel i, j (similar for green and blue component). The Matlab code is provided at [https://github.com/banterle/HDR\\_Toolbox/blob/master/source\\_code/Metrics/mPSNR.m/](https://github.com/banterle/HDR_Toolbox/blob/master/source_code/Metrics/mPSNR.m/).

#### 4. deltaE1976

The deltaE1976 metric is calculated in the CIEL\*a\*b\* domain. The RGB color space converts to the XYZ color space first per equation (15). Then, the XYZ color space can be converted to the CIEL\*a\*b\* color space per equation (16).

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.4124 & 0.3576 & 0.1805 \\ 0.2126 & 0.7152 & 0.0722 \\ 0.0193 & 0.1192 & 0.9505 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (15)$$

$$L^* = 116f\left(\frac{Y}{Y_n}\right) - 16$$

$$a^* = 500 \left( f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right) \right)$$

$$b^* = 200 \left( f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right) \right)$$

$$f(t) = \begin{cases} \sqrt[3]{t} & \text{if } t > \delta^3 \\ \frac{t}{3\delta^2} + \frac{4}{29} & \text{otherwise} \end{cases} \quad \text{where } \delta = \frac{6}{29} \quad (16)$$

deltaE1976 is computed as the root means square error between the ground-truth HDR image, G, and the estimated HDR image, H, in the CIEL\*a\*b\* color space, which is expressed by equation (17).

$$\text{deltaE1976} = \sqrt{(L_G^* - L_E^*)^2 + (a_G^* - a_E^*)^2 + (b_G^* - b_E^*)^2} \quad (17)$$

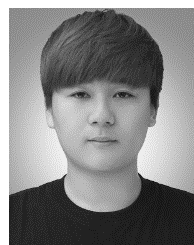
## ACKNOWLEDGMENT

(Hanbyol Jang and Kihun Bang contributed equally to this work.)

## REFERENCES

- [1] D. C. Hood and M. A. Finkelstein, "Sensitivity to light," in *Handbook of Perception and Human Performance*, vol. 1. New York, NY, USA: Wiley, 1986.
- [2] T. Kunkel and E. Reinhard, "A reassessment of the simultaneous dynamic range of the human visual system," in *Proc. 7th Symp. Appl. Perception Graph. Vis.*, 2010, pp. 17–24.
- [3] C. Chinnock. (2016). Dolby Vision and HDR10. Insight Media. [Online]. Available: [http://www.uhd-forum.ch/fileadmin/user\\_upload/temp/\\_Dolby\\_Vision\\_and\\_HDR10.pdf](http://www.uhd-forum.ch/fileadmin/user_upload/temp/_Dolby_Vision_and_HDR10.pdf)
- [4] Dolby Lab., Inc. San Francisco, CA, USA. (2016). *Dolby Vision for the Home*. [Online]. Available: <https://www.dolby.com/us/en/technologies/dolby-vision/dolby-vision-white-paper.pdf>
- [5] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Proc. 24th Annu. Conf. Comput. Graph. Interact. Techn.* New York, NY, USA: ACM, 1997, pp. 369–378.
- [6] S. Mann and R. W. Picard, "Being undigital with digital cameras," MIT, Cambridge, MA, USA, Tech. Rep., 1944, vol. 1, no. 2.
- [7] A. S. Deepa and S. Muthumariammal, "An overview of ghost artifact in HDRI," *SIJ Trans. Comput. Sci. Eng. Appl.*, vol. 2, no. 1, pp. 1–6, 2014.
- [8] E. A. Khan, A. O. Akyuz, and E. Reinhard, "Ghost removal in high dynamic range images," in *Proc. ICIP*, Oct. 2006, pp. 2005–2008.
- [9] N. K. Kalantari and R. Ramamoorthi, "Deep high dynamic range imaging of dynamic scenes," *ACM Trans. Graph.*, vol. 36, no. 4, 2017, Art. no. 144.

- [10] J. Dai, K. He, and J. Sun, "BoxSup: Exploiting bounding boxes to supervise convolutional networks for semantic segmentation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1635–1643.
- [11] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2015, pp. 234–241.
- [12] H. M. Bui, M. Lech, E. Cheng, K. Neville, and I. S. Burnett, "Object recognition using deep convolutional features transformed by a recursive network structure," *IEEE Access*, vol. 4, pp. 10059–10066, 2016.
- [13] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2015.
- [14] J. Kim, J. K. Lee, and K. Mu Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1646–1654.
- [15] M. Ni, J. Lei, R. Cong, K. Zheng, B. Peng, and X. Fan, "Color-guided depth map super resolution using convolutional neural network," *IEEE Access*, vol. 5, pp. 26666–26672, 2017.
- [16] C. Dong, Y. Deng, C. C. Loy, and X. Tang, "Compression artifacts reduction by a deep convolutional network," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 576–584.
- [17] S. Iizuka, E. Simo-Serra, and H. Ishikawa, "Let there be color!: Joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification," *ACM Trans. Graph.*, vol. 35, no. 4, p. 110, 2016.
- [18] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. (2017). "Image-to-image translation with conditional adversarial networks." [Online]. Available: <https://arxiv.org/abs/1611.07004>
- [19] A. O. Akyüz, R. Fleming, B. E. Riecke, E. Reinhard, and H. H. Bühlhoff, "Do HDR displays support LDR content?: A psychophysical evaluation," *ACM Trans. Graph.*, vol. 26, no. 3, 2007, Art. no. 38.
- [20] B. Masia, S. Agustín, R. W. Fleming, O. Sorkine, and D. Gutierrez, "Evaluation of reverse tone mapping through varying exposure conditions," *ACM Trans. Graph.*, vol. 28, no. 5, 2009, Art. no. 160.
- [21] L. Wang, L.-Y. Wei, K. Zhou, B. Guo, and H.-Y. Shum, "High dynamic range image hallucination," in *Proc. 18th Eurograph. Conf. Rendering Techn.*, 2007, pp. 321–326.
- [22] P. Didyk, R. Mantiuk, M. Hein, and H. P. Seidel, "Enhancement of bright video features for HDR displays," *Comput. Graph. Forum*, vol. 27, no. 4, pp. 1265–1274, 2008.
- [23] F. Banterle, P. Ledda, K. DeBattista, and A. Chalmers, "Inverse tone mapping," in *Proc. 4th Int. Conf. Comput. Graph. Interact. Techn. Austral. Southeast Asia*, 2006, pp. 349–356.
- [24] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 267–276, Jul. 2002.
- [25] P. Debevec, "A median cut algorithm for light probe sampling," in *Proc. ACM SIGGRAPH*, 2006, Art. no. 6.
- [26] A. G. Rempel et al., "LDR2HDR: On-the-fly reverse tone mapping of legacy video and photographs," *ACM Trans. Graph.*, vol. 26, no. 3, 2007, Art. no. 39.
- [27] R. P. Kovaleski and M. M. Oliveira, "High-quality brightness enhancement functions for real-time reverse tone mapping," *Vis. Comput.*, vol. 25, nos. 5–7, pp. 539–547, 2009.
- [28] 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, p. 178, 2017.
- [29] T. Smith and J. Guild, "The C.I.E. colorimetric standards and their use," *Trans. Opt. Soc.*, vol. 33, no. 3, p. 73, 1931.
- [30] E. F. Schubert, "Human eye sensitivity and photometric quantities," in *Light-Emitting Diodes*, 2nd ed. New York, NY, USA: Cambridge Univ. Press, 2006, ch. 16, pp. 275–291.
- [31] W. D. Wright, "A re-determination of the trichromatic coefficients of the spectral colours," *Trans. Opt. Soc.*, vol. 30, no. 4, p. 141, 1929.
- [32] J. Guild, "The colorimetric properties of the spectrum," *Philos. Trans. Roy. Soc. London A, Math. Phys. Sci.*, vol. 230, nos. 681–693, pp. 149–187, 1931.
- [33] R. S. Hunter, "Accuracy, precision, and stability of new photoelectric color-difference meter," *J. Opt. Soc. Amer.*, vol. 38, no. 12, p. 1094, 1948.
- [34] R. S. Hunter, "Photoelectric color difference meter," *J. Opt. Soc. Amer.*, vol. 48, no. 12, pp. 985–995, 1958.
- [35] D. Pascale, "A review of RGB color spaces... from xyY to R'G'B," BabelColor Company, Montreal, QC, Canada, Tech. Rep., 2003.
- [36] K. Chiu, M. Herf, P. Shirley, S. Swamy, C. Wang, and K. Zimmerman, "Spatially nonuniform scaling functions for high contrast images," in *Proc. Graph. Interface*, Toronto, ON, Canada, 1993, pp. 245–253.
- [37] M. Ashikhmin, "A tone mapping algorithm for high contrast images," in *Proc. 13th Eurograph. Workshop Rendering*, Pisa, Italy, 2002, pp. 145–155.
- [38] M. D. Fairchild and G. M. Johnson, "Meet iCAM: A next-generation color appearance model," in *Proc. IS&T/SID 10th Color Imag. Conf. (IS&T)*, Scottsdale, AZ, USA, 2002, pp. 33–38.
- [39] S. N. Pattanaik, J. A. Ferwerda, M. D. Fairchild, and D. P. Greenberg, "A multiscale model of adaptation and spatial vision for realistic image display," in *Proc. Conf. SIGGRAPH*, 1998, pp. 287–298.
- [40] H. Tsutsui, S. Yoshikawa, H. Okuhata, and T. Onoye, "Halo artifacts reduction method for variational based realtime Retinex image enhancement," in *Proc. IEEE Asia Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA ASC)*, Dec. 2012, pp. 1–6.
- [41] A. Odena, V. Dumoulin, and C. Olah, "Deconvolution and checkerboard artifacts," *Distill*, vol. 1, no. 10, p. e3, 2016.
- [42] W. S. Mokrzycki and M. Tatol, "Color difference  $\Delta E$ : A survey," *Mach. Graph. Vis.*, vol. 20, no. 4, pp. 383–411, 2012.
- [43] A. R. Robertson, "The CIE 1976 color-difference formulae," *Color Res. Appl.*, vol. 2, no. 1, pp. 7–11, 1977.
- [44] D.-T. Dang-Nguyen, C. Pasquini, V. Conotter, and G. Boato, "RAISE: A raw images dataset for digital image forensics," in *Proc. 6th ACM Multimedia Syst. Conf.*, 2015, pp. 219–224.
- [45] M. Reichmann. *Understanding Raw Files Explained*. [Online]. Available: <http://www.luminous-landscape.com/tutorials/understanding-series/u-raw-files.shtml>
- [46] M. D. Grossberg and S. K. Nayar, "What is the space of camera response functions?" in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2, Jun. 2003, p. II-602.
- [47] M. Fairchild. (2008). *Fairchild's HDR Photographic Survey*. [Online]. Available: <http://rit-mcs.l.org/fairchild/HDR.html>
- [48] P. Korshunov, T. Hanhart, and P. A. Ebrahimi. (2015). *EPFL's Dataset of HDR Images*. [Online]. Available: <http://mmspg.epfl.ch/hdr-eye>
- [49] E. Reinhard, *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting*. San Mateo, CA, USA: Morgan Kaufmann, 2010.
- [50] Y. Huo, F. Yang, L. Dong, and V. Brost, "Physiological inverse tone mapping based on retina response," *Vis. Comput.*, vol. 30, no. 5, pp. 507–517, 2014.
- [51] P. Hanhart, M. V. Bernardo, M. Pereira, A. M. Pinheiro, and T. Ebrahimi, "Benchmarking of objective quality metrics for HDR image quality assessment," *EURASIP J. Image Video Process.*, vol. 39, no. 1, pp. 1–18, 2015.
- [52] 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, p. 010501, 2015.
- [53] G. Valenzise, F. De Simone, P. Lauga, and F. Dufaux, "Performance evaluation of objective quality metrics for HDR image compression," in *Proc. SPIE 9217, Appl. Digit. Image Process. XXXVII*, 2014, p. 92170C.
- [54] S.-K. Yip, O. O. Au, and H.-M. Wong, "PI-preserve data hiding for halftone image," in *Proc. Int. Symp. IEEE Intell. Signal Process. Commun. Syst. (ISPACS)*, Dec. 2005, pp. 125–128.
- [55] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, document ITU-R Rec. BT.500-10, 2000.
- [56] R. McDonald and K. J. Smith, "CIE94—A new colour-difference formula," *J. Soc. Dyers Colourists*, vol. 111, no. 12, pp. 376–379, 1995.
- [57] M. R. Luo, G. Cui, and B. Rigg, "The development of the CIE 2000 colour-difference formula: CIEDE2000," *Color Res. Appl.*, vol. 26, no. 5, pp. 340–350, 2001.



**HANBYOL JANG** received the B.S. degree in electrical and electronic engineering from Yonsei University, Seoul, South Korea, in 2016, where he is currently pursuing the Ph.D. degree in electrical and electronic engineering. His research interests include image processing, computer vision, machine learning, and medical imaging.





**KIHUN BANG** received the B.S. and M.S. degrees in electrical engineering from Yonsei University, Seoul, South Korea, in 2016 and 2018, respectively. His research interests include image processing, computer vision, machine learning, and medical imaging.



**JINSEONG JANG** received the B.S. degree in electrical and electronic engineering from Yonsei University, Seoul, South Korea, in 2013, where he is currently pursuing the Ph.D. in electrical and electronic engineering. His research interests include medical image processing, machine learning, and image quality assessment.



**DOSIK HWANG** received the B.S. and M.S. degrees in electrical engineering from Yonsei University, Seoul, South Korea, in 1997 and 1999, respectively, and the Ph.D. degree in bioengineering from the University of Utah, Salt Lake City, UT, USA, in 2006. He is currently an Associate Professor with Yonsei University. His research interests include deep-learning based image processing, medical imaging, and tomographic reconstruction.

• • •