IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 30, 2019, accepted February 14, 2020, date of publication February 24, 2020, date of current version March 3, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2975857

Dynamic Range Expansion Using Cumulative Histogram Learning for High Dynamic Range Image Generation

HANBYOL JANG, KIHUN BANG, JINSEONG JANG, AND DOSIK HWANG
School of Electrical and Electronic Engineering, College of Engineering, Yonsei University, Seoul 03722, South Korea
Corresponding author: Dosik Hwang (dosik.hwang@yonsei.ac.kr)

This work was partially supported by Samsung Electronics Co.

ABSTRACT In modern digital photographs, most images have low dynamic range (LDR) formats, which means that the range of light intensities from the darkest to the brightest is much lower than the range that can be perceived by the human eye. Therefore, to visualize images as naturally as possible on devices that display them in high dynamic range (HDR) format, the LDR images need to be converted into HDR images. The aim of this study was to develop an adaptive inverse tone mapping operator (iTMO) that can convert a single LDR image into a realistic HDR image based on artificial neural networks. In contrast to conventional iTMO algorithms, our technique was developed by learning the complicated relationship between various LDR–HDR pair images, which enabled nearly ground-truth HDR images to be generated from various types of LDR images. The novel learning technique is called cumulative histogram-based learning and color difference learning. The superior performance of our technique over conventional methods was assessed through objective evaluations of various types of LDR and HDR images.

INDEX TERMS High dynamic range, inverse tone mapping operator, cumulative histogram, color difference, convolutional neural network.

I. INTRODUCTION

In photographs, the dynamic range refers to the luminance range from the darkest region to the brightest region. It is known that the dynamic range that can be perceived by the human eyes at once is $\sim 10^{3.7}$ ($\cong 5000$) cd/m^2 [1]. However, most dynamic ranges commonly used in current digital displays are below 300 cd/m^2. This is referred to as a low dynamic range (LDR). If the expressed dynamic range is narrow, the differences in brightness that human eyes can distinguish may be displayed as a constant level of brightness. Wide dynamic range scenes, such as sunrises or sunsets, cannot properly be captured or displayed using LDR settings. As such, the current LDR technologies are insufficient to attain the levels of perception made possible by the human eyes. Therefore, techniques that can expand the dynamic range of images as wide as the cognitive range of the human eye are necessary.

High dynamic range (HDR) technologies have recently been developed. These technologies expand the dynamic range represented by conventional LDRs, thereby increasing the contrast ratio of the image. HDR displaying devices that can support 1000 cd/m^2 and 10-bit images are also being made available. Conventional LDR images can represent only 256 color levels, because there are only 8 bits available to express individual pixel intensity. Therefore, because LDR images are not effectively displayed on 10-bit HDR display panels, techniques to obtain HDR images are becoming more important.

To obtain HDR images, it is necessary to combine several LDR images with different exposure values to create a HDR image with a wide dynamic range [2]. However, while acquiring multiple LDR images, movements of the camera or objects to be captured often occur and cause the appearance of artifacts, such as ghost artifacts [3], during the process of combining the images. Several methods have been proposed to eliminate such artifacts, but they have failed to completely remove them [4]. With the development of modern technology, single-shot HDR imaging has become available. There are a variety of methods, including the capture of a single-shot HDR using a spatially varying imaging sensor [5] and the recovery of a HDR image by convolutional sparse coding [6].

The associate editor coordinating the review of this manuscript and approving it for publication was Pasquale De Meo.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see http://creativecommons.org/licenses/by/4.0/

However, these technologies have limitations such as the need to shoot with a specially manufactured camera and the degradation of resolution. Furthermore, large amounts of image data collected in LDR formats need to be converted into HDR formats to be properly visualized on HDR displays. The conversion process is referred to as inverse tone mapping. Several methods using inverse tone mapping operator (iTMO) algorithms have been proposed to expand LDR to HDR based on mathematical models. However, when using such models to convert images, it is difficult to consider the individual characteristics of each image, thus resulting in suboptimal HDR images that are different from the true HDR images. Recently, deep neural networks perform well in computer vision fields for problems that are difficult to solve using traditional algorithms. In the field of iTMO technology, many deep learning based iTMOs have been proposed to overcome the limitations of the existing methods and to increase the adaptability of various types of images as well. In Related work section, we will introduce various iTMOs in detail from the traditional methods to the latest deep learning methods.

We also proposed a new robust technique that converts LDR images into HDR images with an expanded dynamic range that is significantly closer to the true HDR using deep learning. The main differences between our proposed method and conventional learning approaches are as follows: (1) Efficiently learns the difference between LDR and HDR images using the CIE$L^*a^*b^*$ color domain instead of the RGB color domain; (2) Solves small dataset problems using transfer learning; (3) Effectively estimates HDR brightness image using cumulative histograms learning and histogram matching, not image-based learning; (4) Learns not only brightness but also color using color different loss function.

II. RELATED WORK

Over the years, several iTMO methods have been proposed based on various mathematical models and by incorporating knowledge of the human visual system. Akyüz *et al.* [7] proposed the use of a simple linear expanding method based on their psychophysical investigation. They suggested that a simple linear boost of the range of a LDR image to fit a HDR display could equal or even surpass the appearance of a true HDR image. However, over-exposed areas are problematic, particularly when the portion of the over-exposed area is large in images. Therefore, several sophisticated algorithms have been developed. Meylan *et al.* [8] proposed the use of a piecewise linear tone scale function composed of two slopes for LDR–HDR conversion to reduce the large gap between the normal range of luminance (diffuse area) and the very high range of luminance (specular area) in images to recover the natural look of the original HDR scenes. Masia et al. proposed some nonlinear global functions based on a gamma curve [9] to handle the overall luminance differences of individual images and selective tone mapping methods [10] selectively applied to local areas in images, depending on different zones or salient contexts. Banterle *et al.* [11] proposed a general framework to convert LDR images into HDR images based on the inverse of the Reinhard tone mapping operator (TMO) algorithm [12]. They estimated the light sources in images and generated the Expand Map to mitigate abrupt changes in high luminance areas via a linear interpolation between the original LDR and the estimated HDR images. Based on this technique, several problems associated with the direct inversion process were overcome. Recently, Huo *et al.* [13] proposed a physiological inverse tone mapping algorithm that considers the characteristics of the human visual system (HVS). In a manner similar to the local response of the HVS to visual stimuli (local retinal response) [14], they used a mathematical retinal response model that can locally be adapted to different areas depending on the estimated local average luminance to replicate the physiological process that underlies the perception of light. They demonstrated that their method reduced the formation of artifacts and produced HDR images of high visual quality. Although all these methods, including those not mentioned here, can produce appealing HDR images, they are based on a very limited number of mathematical models and partial knowledge of HVS that human researchers can think of and implement in practice. Consequently, these models cannot be used to handle various types of images, and their HDR estimation is significantly different from ground-truth HDR images even though they can produce a good HDR impression. Therefore, there is a need to develop a new method to produce a HDR image from a single LDR image that is significantly identical to the ground-truth HDR image.

In this paper, we propose a new data-driven method to estimate the ground-truth HDR image from a single LDR image using a deep learning method. Using this method, the complicated nonlinear relationship between LDR and HDR images can be learned from various types of LDR–HDR pairs. Very recently, deep learning approaches have been proposed to generate HDR images from single LDR images. Zhang and Lalonde [15] used an autoencoder network to produce HDR panoramas from single LDR panoramas. Although good HDR images were produced, the group's work was limited to outdoor panoramas in which the sun was assumed to be in the same azimuthal position in all images, and the resolution of the generated HDR images was too low (128×64) to clearly show the details of these images. Kalantari and Ramamoorthi [16] generated HDR images from three unfixed LDR images with different exposure values. In general, because of ghost artifacts, the images used for synthesizing must be fixed. However, their method solved the ghost artifact problem by aligning the over- and under-exposure value images based on the middle exposure value image. However, there is a limitation that a single LDR image alone cannot generate a HDR image. Eilertsen *et al.* [17] used a similar autoencoder network with several modifications and logarithmic loss functions. They demonstrated a successful reconstruction of various HDR images. The limitations of their work were the difficulty in recovering dark regions and the underestimation of extreme intensities.

Furthermore, a user-specified parameter for the loss function needed to be tuned for assigning different importance levels to the illuminance and reflection components. Endo *et al.* [18] proposed the use of a deep learning technique to learn several bracketed images with different exposures and subsequently combine them to produce HDR images. They also successfully inferred HDR images from single LDR images. However, the limitations of their work were the difficulty in handling an extremely high dynamic range and the necessity for a large memory owing to the multiple learnings required to generate intermediate multiple exposure images. Some tiling artifacts have been reported, and the selection of appropriate LDR images to combine is yet to be heuristically determined. Lee *et al.* [19] proposed a method of generating images with various exposure values (EVs) from a LDR image with EV 0. Endo's method infers multiple exposure images at once, but they used a chain convolutional neural network (CNN) architecture that sequentially infers EV ± 1, ± 2, ± 3 with a LDR image that has a medium exposure value (EV 0). Marnerides *et al.* [20] proposed multiscale deep learning architecture, which is based on a CNN consisting of three branches that learn local detail, medium-level detail, and higher-level image-wide features. The method successfully inferred HDR images by learning various local and global features. However, LDR images generated from HDR images by TMO and not LDR–HDR pair data were used.

III. METHODS

The overall procedure of our iTMO development is as follows: (1) a training dataset preparation for LDR–HDR image pairs, (2) brightness and chromatic image extraction, (3) deep learning of the relationship between cumulative histograms for LDR and HDR pairs, (4) the brightness of HDR image generation via histogram matching using the estimated HDR cumulative distribution function (CDF), (5) recombination of the estimated brightness and the chromatic of LDR image, and (6) deep learning of the relationship between color in LDR and HDR pairs. The overall framework of the proposed method is shown in Figure 1.

A. DATASET

For a successful development of the data-driven iTMO algorithm via deep learning, a good training dataset with various types of images, structures, contrasts, and properties was essential. However, as it is difficult to collect open-source LDR–HDR pair images, we generated our own set of LDR–HDR pair images. We captured 450 different scene images using a Samsung NX3000 camera in the autoexposure bracket mode (-2, 0, 2 EV) at a resolution of 1024×1024 pixels. The dataset consists of various scenes of the following categories: indoor, outdoor, landscape, objects, buildings, and nights. The number of images for each category is 132, 101, 57, 52, 30, and 78, respectively. We set the aperture value to f/4, the ISO value to 100, and the shutter speed automatically using the autobracketing function. Ground-truth HDR images were generated by merging these three LDR images using the HDR Pro algorithm developed by Debevec and Malik [21] (built in Adobe Photoshop CC 2018). The data used in the experiment was set up with LDR (EV = 0) and HDR images pair images. Our dataset used for the experiment is available at: https://github.com/HanbyolJang/LDR-HDR-pair_Dataset The total images were divided into 5-folds for cross-validation to prevent overfitting [22]. 4 subsets (360 images) were used for training & validation and the remaining subset (90 images) was used for testing. A total of 5 models were trained using the all different combination of subsets and performance was evaluated using a test set not used for learning.

All the HDR images in this paper are tone mapped by Reinhard's TMO [23] to visualize them in a conventional

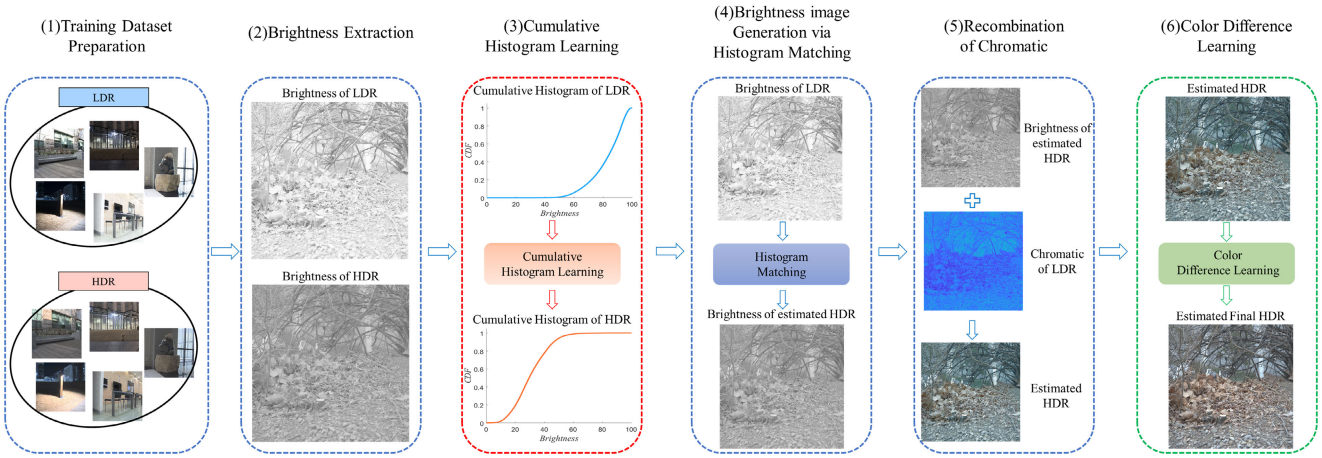


FIGURE 1. Overall framework of the proposed method. Pairs of LDR–HDR images are prepared for the deep learning process. The brightness values are extracted from images, and the relationships between the LDR and HDR cumulative histograms are learned. An estimated HDR brightness image is generated via a histogram matching method using the estimated HDR cumulative distribution function, and the estimated brightness image and the chromatic of the LDR image are recombined. The final HDR image is generated by learning the relationship between LDR and HDR color.

LDR display setting for the readers. No gamma correction or other processes were performed. In Figure 2, examples of LDR–HDR image pairs used in our study are shown. When comparing LDR–HDR images, overexposed bright areas and underexposed dark areas, which are not visible in the LDR images, can be clearly seen in the HDR images.

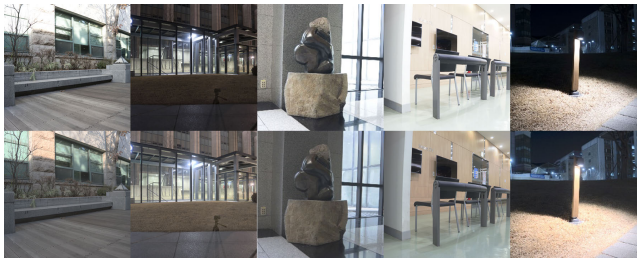


FIGURE 2. Examples of LDR–HDR (tone mapped) pair images. LDR images in the top row contain over- and underexposed areas, where structures and details are not easily observable. However, HDR counterparts in the bottom row clearly show the details in both the dark and bright regions.

B. BRIGHTNESS EXTRACTION

Images can be expressed in several color spaces such as RGB, CMYK, and CIE$L^*a^*b^*$ [24]. As the dominant difference between LDR and HDR images resides in the brightness, we converted images in the RGB space into images in the CIE$L^*a^*b^*$ space and decomposed them into the brightness channel (L^*) and the chromatic channels (a^*, b^*). The color space conversion is provided in Appendix A. Using various images, we calculated the difference between LDR and HDR images in individual channels to identify the channel that manifests the most pronounced difference. This difference was calculated using a normalized root mean square error metric [25], as shown in Eq. (1):

$$NRMSE\ (\%) = \frac{\sqrt{\frac{1}{M \times N} \sum_{x=1}^{M} \sum_{y=1}^{N} [(R\,(x, y) - I(x, y)]^2}}{\max\,(R) - \min\,(R)} \tag{1}$$

where R is the reference image (HDR image), I is the input image (LDR image), (x, y) is the pixel index, $M \times N$ is the image size, $max(R)$ is the maximum pixel intensity of the reference image, and $min(R)$is the minimum pixel intensity of the reference image. In Figure 3, the difference between LDR and HDR at each channel is shown. The difference is large in all individual R (47%), G (47%), and B (46%) channels in the RGB space, while it is large only in the L^* (49%) channel and negligible in the a^* (1.6%) and b^* (1.6%) channels in the CIE$L^*a^*b^*$ space. Therefore, it was more efficient to learn the brightness relationship between the LDR and HDR images than to learn the mixed RGB relationship between these images.

C. CUMULATIVE HISTOGRAM LEARNING

In Figure 4, examples of LDR and HDR images in (a) and (b), respectively are shown. The area inside the orange box of the

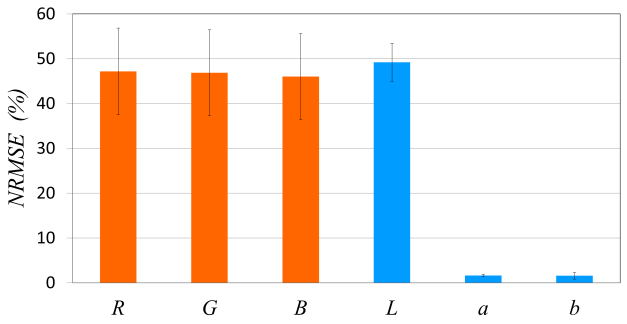


FIGURE 3. Difference (%) between LDR and HDR images at the individual channels: R, G, and B in the RGB color space, and L^*, a^* and b^* in the CIE$L^*a^*b^*$ color space. The difference is large at all three R, G, and B channels, while it is large only at L^* channel and negligible at the a^* and b^* channels.

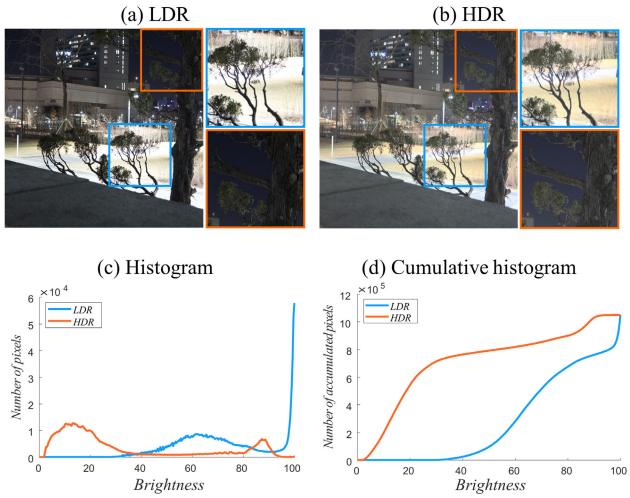


FIGURE 4. Comparison of the LDR and HDR images and their histograms: (a) LDR image indicating dark and bright regions where visibilities of the inside contents are very low. (b) HDR image (tone mapped) with good visibility throughout the image. (c) histograms of LDR and HDR brightness images. (d) cumulative histograms of LDR and HDR brightness images. Cumulative histograms have low complexity, and the difference between the LDR and HDR histograms is low when compared with the original histograms.

LDR image is too dark to indicate the detailed contents in it, whereas the details with good contrast can be clearly seen in the HDR image. A similar observation can be made in the area inside the blue box, which is too bright to indicate the detailed contents in the LDR image, whereas the details of the HDR image are clear. These observations can be explained through histograms of the images. In Figure 4(c) and (d), the histograms and the cumulative histograms of the LDR and HDR brightness images are shown, respectively. The pixels in the LDR image is scarcely present between 0 and 20 and most of the values between 80 and 100 are concentrated, thus resulting in very low contrasts within these compact brightness ranges. However, the pixels in the HDR image are widely distributed throughout the entire brightness range; therefore, good contrasts and visibility can be observed consistently in the HDR images.

There are several methods that change the contrast of an image using a histogram. Conventional histogram equalization algorithms may be applied to distribute the LDR histograms as widely as possible, thus increasing the contrasts of the dark and bright regions to resemble HDR images. However, the resultant images are not generally perceived to be natural by human observers, because they deviate from the natural tones of the original photographs. If we were able to convert LDR histograms into histograms that are nearly identical to the ground-truth HDR histograms, we could generate good-quality HDR images from single LDR images. However, no fixed transfer function that can convert a LDR histogram into a HDR histogram is known because individual images have their own characteristic histograms depending on factors such as type of scene, lighting sources, and structures. The aforementioned model-based iTMO algorithms in Section II were used in an attempt to overcome these obstacles by incorporating various functions such as linear, piecewise linear, nonlinear, zone- or context-adaptive, and local-adaptive response functions; however, the algorithms or these functions could not be generalized for various types of images. Furthermore, their outputs deviate from the ground-truth HDR images.

In this paper, we attempt to overcome this obstacle by using deep learning to investigate the relationship between the LDR and HDR histograms from various types of LDR–HDR image pairs. One problem in this approach is that the shape and features of histograms are too complicated to be learned even when using the current advanced deep learning techniques. Figure 4(c) clearly shows the large differences between the LDR and HDR histograms in terms of shape, density, local variations, etc. In our preliminary studies, the direct learning of the LDR and HDR histograms failed to produce good estimates of the ground-truth HDR histograms for these differences. However, the cumulative histograms of images have learning-favorable properties, as shown in Figure 4(d). Both the LDR and HDR cumulative histograms have smooth curves from the lowest to the highest brightness levels and always end with the same total number of pixels in the image. In addition, local variations have been reduced significantly. Consequently, the overall complexity of the cumulative histograms is very low, and the differences between the LDR and HDR cumulative histograms are also low when compared with the original histograms. Therefore, we applied deep learning to study the relationship between the cumulative histograms corresponding to the LDR and HDR images.

D. CUMULATIVE HISTOGRAM LEARNING ARCHITECTURE

In Figure 5, the deep CNN structure for cumulative histogram learning is shown. This network structure is based on VGG [26] and ResNET structures [27]. As a loss function, we used the L2 norm between the estimated CDF and the ground-truth CDF as described in equation (2).

$$Loss_{CDF} = \sum \left(CDF_{groudn-truth} - CDF_{output}\right)^2 \quad (2)$$

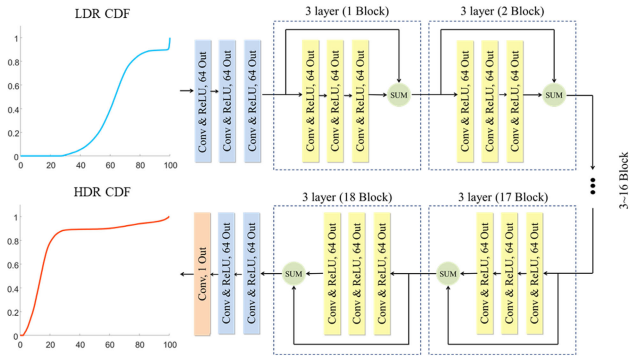


FIGURE 5. Cumulative histogram learning architecture: CNNs with a one-dimensional convolution kernel, ReLU for the activation function, and skip connections to include features extracted from previous layers.

The total number of convolution layers was 60, and the rectified linear unit (ReLU) for the activation function was applied after each layer. After every three layers, skip connections were added to include the features extracted from previous layers. In our preliminary experiments, networks without these skip connections resulted in insufficient estimations. Several intervals between the layers to which the skip connections are to be applied were investigated, and a three-layer interval for the skip connection yielded the best performance. The number of bins for the cumulative histograms was set to 512. And the kernel size of each 2D convolution layer was 11×1 and the number of feature maps was 64, except in the final convolution layer. For the kernel size in the convolution process, 11×1 resulted in the best estimation of HDR cumulative histograms. This is because the receptive field of the network was 601 $[= (11-1)\times 60 + 1]$, which was enough to include 512 bins of cumulative histograms [28]. The total number of parameters of the cumulative histogram learning network is 10,224,660. Our model was implemented in the Python with Keras using Tensorflow library on an Intel Core i7-6700K CPU with a Nvidia GeForce GTX 1080Ti GPU and 32-GB RAM. Adam optimizer was used with an initial learning rate of 1×10^{-5}, and training was performed until validation error saturation occurred. The total training time was 27 minutes, and the test time for an image with 1024×1024 resolution took 8 milliseconds in the cumulative histogram learning model. According to the definition of a cumulative histogram, its curve should be a nondecreasing function. However, the proposed deep learning model may produce slightly decreasing patterns within short intervals in some cases. In such cases, the decreasing portions of the cumulative histograms was post-processed to be automatically corrected using an efficient interpolation function (piecewise cubic hermite interpolating polynomial [29]).

E. TRANSFER LEARNING

Deep learning neural networks, learning the complex relationship between input and output, are generalized with the use of large amounts of data. Therefore, the 450 LDR–HDR pair images that were available for this study were

not sufficient. Generally, to solve the problem of data shortage, augmentation is performed. Augmentation is a method used to avoid overfitting and make better generalizations by increasing the number of data through transformations such as left and right reversal or rotation. However, we propose cumulative histogram learning in this paper. As the cumulative histogram is the sum of histograms for the intensity of the pixels in the image, histograms are the same even when augmentation is performed. For this reason, because augmentation is not applicable, we applied transfer learning using pre-trained weights on large dataset [17]. Transfer learning is a method of subsequent learning with weights extracted from related domains to help learning in a target domain with few data. In other words, fine-tuning the pre-trained weights to the target domain.

To generate a pretrained model from numerous data in the relevant domain, we generated a set of LDR and pseudo- HDR pair images from a single raw image. We collected 8156 high-resolution raw images from an open-source RAISE database [30]. The RAISE dataset contains a wide range of different categories of scenes: ''outdoor,'' ''indoor,'' ''landscape,'' ''nature,'' ''people,'' ''objects,'' and ''buildings.'' The raw images taken with a digital camera without compression or processing are 12- to 14-bit images [31]. We adjusted the exposure values ($-2, -1, 0, +1, +2$) from a single raw image and then generated five different LDR images [32]–[34]. We merged these five LDR images using the HDR Pro algorithm. It was defined as a pseudo-HDR because the resulting image was generated by merging LDR images with different exposure values from one raw image rather than LDR images obtained by taking different exposure values. Pseudo-HDR images are not exactly the same as true HDR images; however, the expansion of the dynamic range is similar, because it is generated from LDR images of various exposure values. Among the 8156 image pairs, 156 images were used for the test set and 8000 images were used for the training and validation set of the pretrained model. In conclusion, our network first learns the relationship between 8000 LDR and pseudo-HDR pair images, and then fine-tunes the learned weights to LDR and true HDR pair images.

F. HDR BRIGHTNESS IMAGE GENERATION VIA HISTOGRAM MATCHING

Figure 6(a) demonstrates that the HDR brightness value can be obtained using the histogram matching technique [35] with the estimated HDR cumulative histogram. The X-axis represents the brightness values (0–100) for the LDR and HDR images, and the Y-axis represents the normalized number of accumulated pixels. The blue and orange curves represent the cumulative distribution functions $f_L(x)$ for LDR and $f_H(x)$ for HDR brightness images, respectively. For an arbitrary brightness value x in an LDR image, its corresponding HDR brightness value x' can be estimated as follows: (1) find the function value of $f_L(x)$ at x and (2) find the HDR brightness value x' that makes $f_H(x') = f_L(x)$. The value of x' becomes

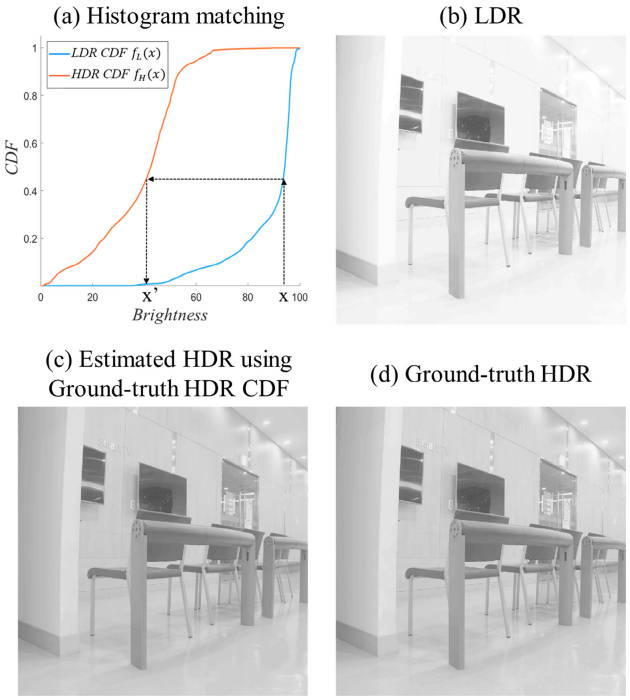


FIGURE 6. Histogram matching and HDR image generation: (a) flow of histogram matching technique, (b) LDR brightness image, (c) estimated HDR brightness image using the ground-truth HDR cumulative histogram, and (d) ground-truth HDR brightness image. (c) and (d) are the tone-mapped brightness images.

the HDR brightness value. During the matching algorithm, a smooth intensity mapping function and cubic hermite polynomial fitting were used to prevent artifacts caused by significant brightness value changes [35]. Finally, a HDR brightness image can be obtained by applying this process to all pixels of the LDR image.

In Figure 6(a), the slope of the CDF in the LDR image increases drastically at values of 90–100, which means that several pixels are gathered to that value. Therefore, the LDR image in Figure 6(b) looks as though the bright regions are clipped to a single value. However, it does not clip all the values in the bright region to 100 but tightly compresses them. Because the area is so tightly saturated, it appears to be clipped, which is indistinguishable to the eye. On the contrary, the HDR image in Figure 6(d) is not saturated, because the CDF for the HDR image is distributed evenly over the whole region. In other words, the histogram matching method is used to spread the brightness of the LDR image, which is densely populated. In Figures 6(c), an example of HDR image generation from an LDR image by the histogram matching technique is shown. If the ground-truth HDR cumulative histogram is perfectly estimated, the LDR image (b) is successfully converted into the HDR image (c) with no noticeable difference between it and the ground-truth HDR image (d). This example suggests that good-quality HDR images can be generated using the histogram matching technique if HDR cumulative histograms can be estimated effectively through our deep learning neural network.

G. COLOR DIFFERENCE LEARNING ARCHITECTURE

In Figure 3, the L^* difference between the LDR image and the HDR image is significantly larger, but there are also differences between the chromatic components a^* and b^*. Even if the HDR brightness may be perfectly predicted by the previous process, if the chromatic components are part of the LDR image, the ground-truth HDR color will not be represented. Figure 7(c) is the estimated HDR image combining L^* of the ground-truth HDR image and a^* and b^* of the LDR image. There is a noticeable color difference between the estimated and the ground-truth HDR images. Therefore, our model is required to learn color as well as brightness. We designed the learning architecture to estimate the color of HDR images from the color of LDR images, as shown in Figure 8. This network structure is based on U-net structures [36]. It is a structure that learns not only high-level features but also low-level features using the 2×2 max pooling and up-sampling layer. The missing information after up-sampling is compensated by skip connection and concat. The network learns local color information using the 3×3 convolution layer, and the rectified linear unit (ReLU) for the activation function was applied after each layer. As a loss function, we used a Euclidean distance in the CIE$L^*a^*b^*$ color space as described in equation (3). It is suitable for color differences learning due to its high correlation with the human color perception [37].

$$Loss_{color} = \sum \sqrt{(L^*_G - L^*_E)^2 + (a^*_G - a^*_E)^2 + (b^*_G - b^*_E)^2} \tag{3}$$

where G is the ground-truth image and E is the estimated image. The total number of convolution layers was 27 and that of feature maps starts at 64 and doubles each time through the pooling layer, resulting in 1024 feature maps at the rightmost layer. The total number of parameters of the color difference learning is 47,049,155. The initial learning rate was 5×10^{-5} and the rest of learning factors were implemented in the same environment as cumulative histogram learning. The total training time was 120 minutes and the test time for an image with 1024×1024 resolution took 0.21 seconds in the color difference learning model.

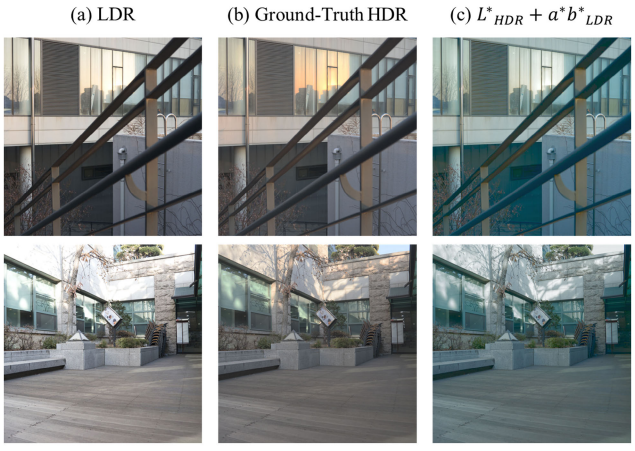


FIGURE 7. Necessity of color learning in HDR image estimation: (a) LDR image, (b) ground-truth HDR image, and (c) estimated HDR image combining L^* of the ground-truth HDR image and a^* and b^* of the LDR image. Even though the brightness image is the same as that of the ground-truth HDR image, the color appearance is different. (b) and (c) are the tone-mapped images.

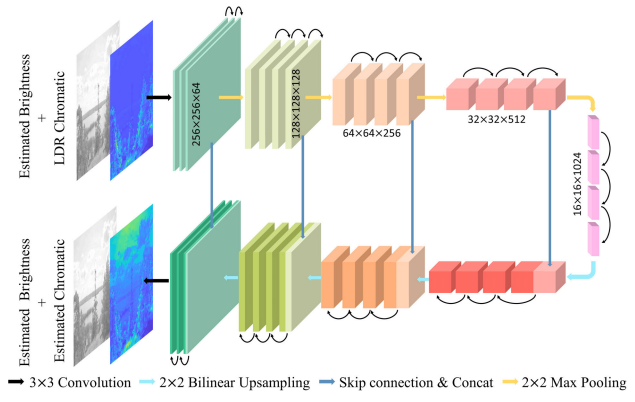


FIGURE 8. Color difference learning architecture.

IV. RESULTS

A. CUMULATIVE HISTOGRAM LEARNING

In Figure 9, the test error graph of the trained model shows the effect of transfer learning. It shows (a) a model trained with 450 LDR and true HDR pair images, (b) a model trained with 8000 LDR and pseudo-HDR pair images, and (c) a model trained with 450 LDR and true HDR pair images from pre-trained weights on 8000 LDR and pseudo-HDR pair images. All cases were tested with 450 LDR and true HDR pair images with 5-fold cross-validation [22]. Among the three models, the test loss of (c) using transfer learning was the lowest. This allowed us to confirm that transfer learning using weights pretrained with pseudo-HDR images with similar features is more effective. For transfer learning, it is common to use a smaller learning rate in order not to distort well-tuned weights too quickly and too much. We used a learning rate of 1×10^{-6}, which is 1/10 of the initial learning rate. Therefore, the fluctuation in the error decreased from the point where transfer learning began.

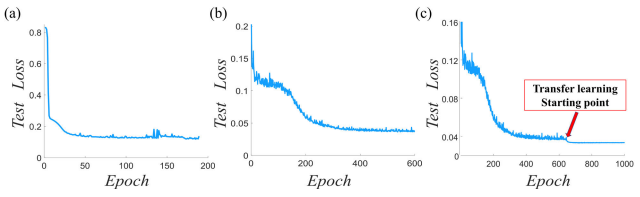


FIGURE 9. Effect of transfer learning: (a) model trained with LDR and true HDR pair images, (b) model trained with LDR and pseudo-HDR pair images, and (c) model trained with transfer learning.

The accuracy of the cumulative histogram learning model was calculated between the CDF estimated by our deep learning model and the ground-truth CDF using three similarity evaluation metrics: Mean squared difference similarity (MSD) [38], Cosine similarity [38], and Pearson similarity [38]. Before calculating the accuracy of the proposed CDF

learning model, it is necessary to calculate the accuracy of the pre-trained model since the pre-trained performance affects transfer learning. The three metric results of the pre-trained model were 99.95%, 99.85% and 99.97%, respectively. These scores mean that the pre-trained model is almost perfectly trained. Table 1 shows the accuracy of the three models used in Figure 9. When using transfer learning, the similarities with the ground-truth CDF were scored the highest. This result is identical to the test loss result in Figure 9, which shows that the proposed model almost perfectly estimates the CDF of the true HDR image. The p-value calculated by the paired t-test [39] for all similarity metrics was less than 0.05, which demonstrated the statistical significance improvement between the models.

TABLE 1. Accuracy results of the three models calculated by the similarity evaluation metrics.

	Trained with LDR & true HDR	Trained with LDR & pseudo-HDR	Transfer learning
MSD Similarity	98.76%	99.21%	99.85%
Cosine Similarity	95.98%	98.78%	99.49%
Pearson Similarity	99.20%	99.43%	99.91%

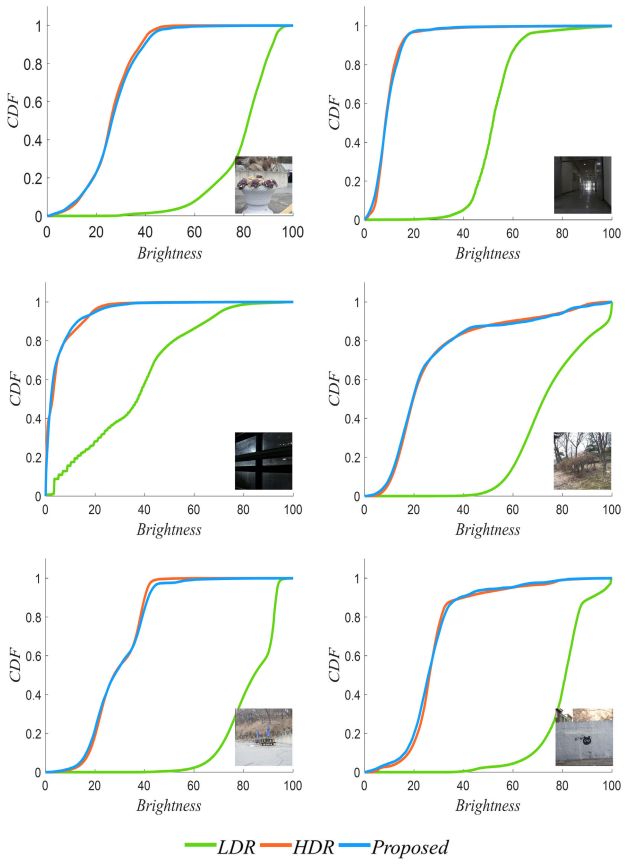


FIGURE 10. Estimates of HDR cumulative histograms from single LDR histograms. The estimates (blue lines) follow the patterns of ground-truth (orange lines) quite well even though there is a large difference between the LDR and the ground-truth HDR cumulative histograms.

In Figure 10, examples of the estimates of the HDR cumulative histogram from a single LDR cumulative histogram obtained using the proposed trained deep learning model are shown. Green lines represent the LDR cumulative histograms. It can be seen that the characteristics of the cumulative histograms are diverse for different individual images in terms of shape, local slopes, curvature, onset, endset, etc. Their ground-truth HDR counterparts (orange lines) also show diverse characteristics. Furthermore, the relationships between the LDR and HDR cumulative histograms are complicated and vary considerably for each case; no fixed transfer function can be estimated using classical mathematical models. Nevertheless, our trained deep learning model was able to effectively estimate the ground-truth HDR cumulative histograms, as shown by the blue lines. In these examples, the estimates (blue lines) follow the patterns of ground-truth (orange lines) quite well even though there are large differences between the LDR and the ground-truth HDR cumulative histograms.

B. HDR BRIGHTNESS IMAGES USING HISTOGRAM MATCHING

In Figure 11, examples of the estimated HDR brightness images using the histogram matching technique with the

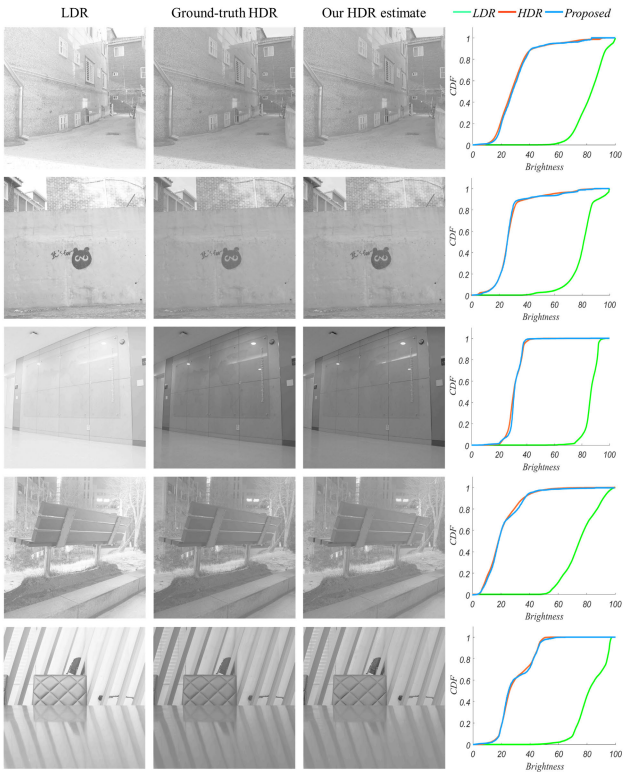


FIGURE 11. Estimates of HDR brightness images in comparison with the LDR and ground-truth HDR brightness images. Ground-truth HDR and our HDR estimate are the tone-mapped brightness images.

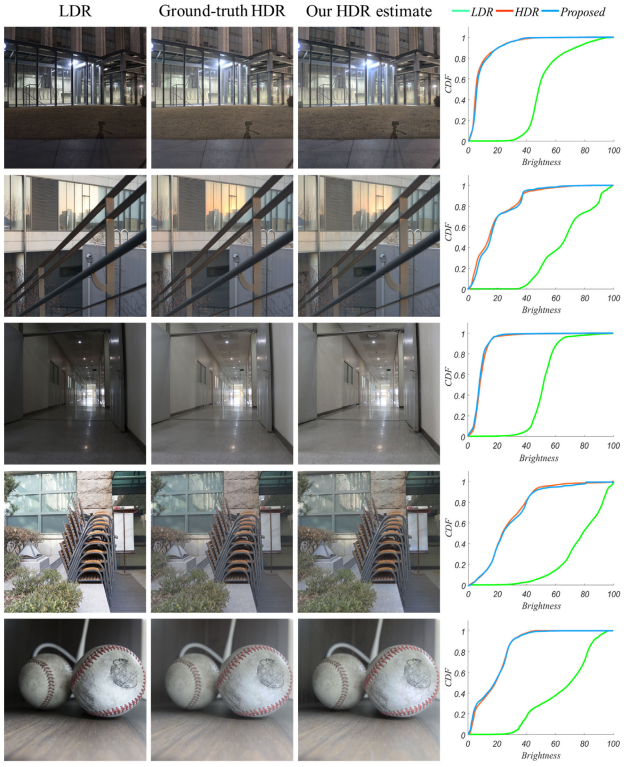


FIGURE 12. Final color HDR image estimates in comparison with LDR and ground-truth HDR images. The dark and bright areas with low visibility in the LDR images were effectively converted into HDR images with improved visibility and local contrasts. All the HDR images are tone mapped.

FIGURE 13. The results from other types of data not used for training. All the HDR images are tone mapped.

inferred HDR cumulative histograms are shown. Global and local contrasts of the estimated HDR images are well matched with the ground-truth HDR brightness images, which are quite different from the LDR brightness images. The accuracy of the brightness images estimated by histogram matching was calculated by the quantitative evaluation metric between it and the HDR brightness images, which are ground-truth. The root mean squared error (RMSE), peak signal-to-noise ratio (PSNR), and structural similarity (SSIM) of the 450 test brightness images estimated by histogram matching scored 2.99, 31.47 and 0.974, respectively. Since RMSE calculates the Euclidean distance between two images, a score of 2.99 in the brightness image with a range of [0-100] is a very small difference. And, the PSNR score above 30 is considered that the difference between two images is invisible to the human eye [40], and the SSIM score above 0.9 is considered to be almost identical between two images [41]. Therefore, the quantitative evaluation results show that brightness images through histogram matching successfully estimate ground-truth HDR brightness images.

Jang *et al.* [34] also experimented with image-to-image learning on brightness images. However, the characteristic of the LDR–HDR pair image was only the change of dynamic range, not the change of structure information in the image. Therefore, image-to-image learning is inefficient because the model also learns the structure information in the image. The RMSE, PSNR, SSIM of the 450 test brightness images estimated by image-to-image learning scored 4.27, 28.94 and 0.934, respectively. This suggests that cumulative histogram learning, which focuses on brightness difference learning, is more effective. In addition, because cumulative histogram learning is one dimensional, the training time is also significantly shorter than for image-to-image learning, which is two dimensional.

C. COLOR DIFFERENCE LEARNING

The CIE$L^*a^*b^*$ color space images obtained through color difference learning are converted into RGB color space images to obtain the final HDR images. Figure 12 shows the LDR images (first column), the ground-truth HDR images (second column), the final estimated HDR images (third column), and the brightness cumulative histogram of the three images (fourth column). The dark and bright areas with low visibility in LDR images were effectively converted into their HDR counterparts with improved visibility and local contrasts. Furthermore, there is no noticeable difference in color or brightness between our HDR estimates and the ground-truth HDR images.

In order to verify general application of our proposed method, we further tested the different types of data not used for training. A total of 120 data were obtained from Eilertsen [17], Fairchild [42], HDR-Eye [43]. Eilertsen provided LDR and HDR pair images collected from various sources. And Fairchild dataset were generated by

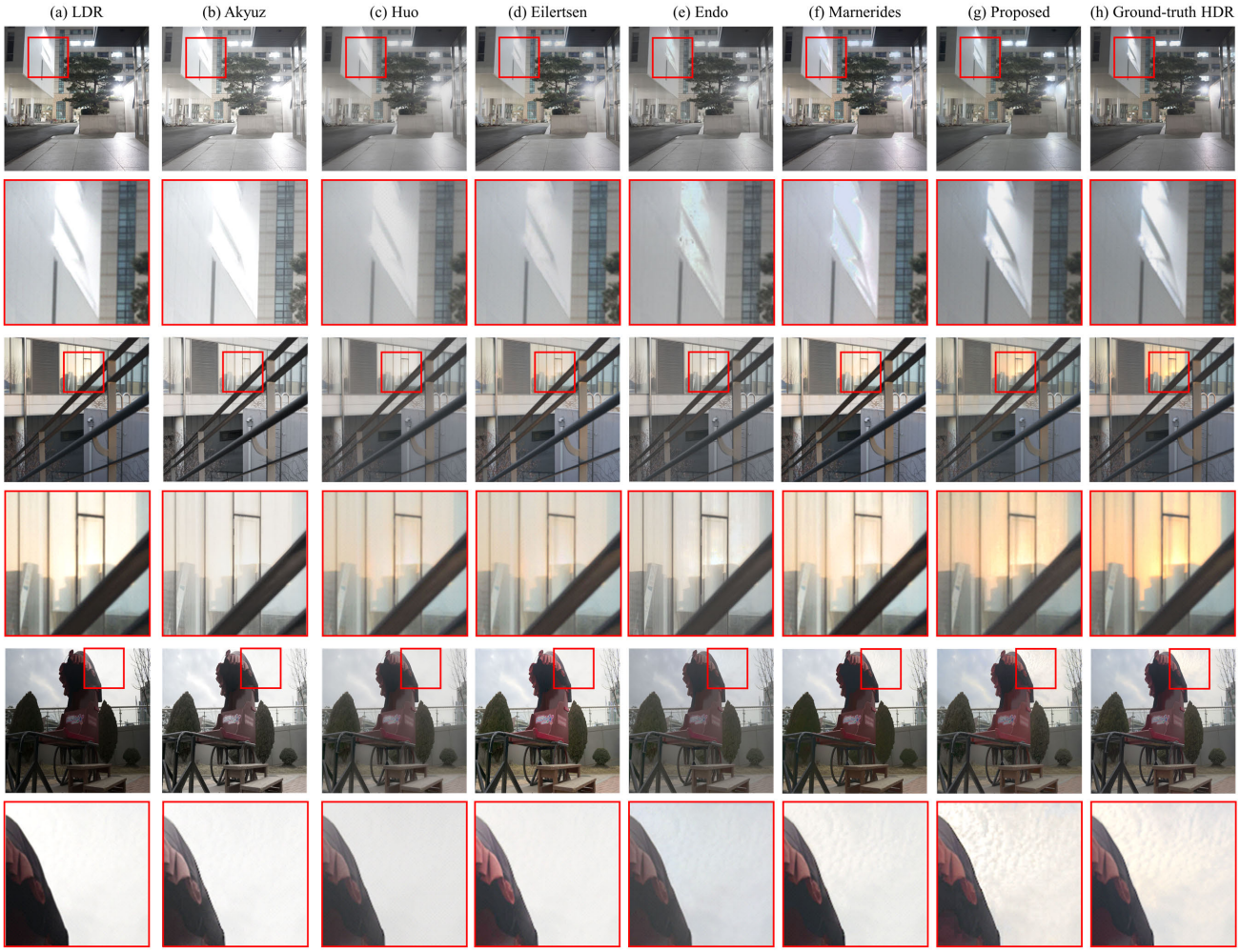


FIGURE 14. Comparison of HDR images generated using the proposed method and other iTMOs. Columns (a) and (h) are LDR and the corresponding ground-truth HDR images, and (b)–(g) are the HDR estimates of Akyüz, Huo, Eilertsen, Endo, Marnerides, and our proposed method, respectively. All the HDR images are tone mapped.

synthesizing nine LDR images with different exposures taken by a Nikon D2x camera and HDR-Eye dataset was generated by several cameras, including Sony DSCRX100 II, Sony NEX-5N, and Sony alpha 6000, in the same process as Fairchild. Figure 13 shows that our proposed method well estimated HDR images for the different types of data not used for training. The slight difference in the brightness cumulative histogram and color between the estimated HDR images and the ground-truth HDR images is due to the difference in camera type, settings, and environment.

D. COMPARISON WITH EXISTING ITMO METHODS

Figure 14 compares the HDR images generated by our proposed method and other existing iTMO methods. Column (a) and (h) are LDR and the corresponding ground-truth HDR images, and (b)–(g) are the HDR estimates of Akyüz *et al.* [7], Huo *et al.* [13], Eilertsen *et al.* [17], Endo *et al.* [18], Marnerides *et al.* [20], and our proposed method, respectively.

In the magnified versions of the first image (second row), the bright region is saturated in the LDR image, making it difficult to distinguish. In ground-truth HDR image, however, this region can clearly be distinguished. Various iTMO methods attempt to distinguish saturated regions by expanding the dynamic range. Since Akyüz's method is a simple linear expanding method, the bright region is excessively brighter, making it more difficult to distinguish. Huo and Eilertsen's estimated HDR images adjust the contrast for overly bright regions; however, the structure is still indistinguishable. Endo and Marnerides's estimated HDR images represent the structure distinguishably. However, artifacts occur in overly bright areas. They also occur in the right side light source of the whole image. In contrast, the estimated HDR image of our proposed method expands the dynamic range like a ground-truth HDR image. Structure information that was not distinguished in the LDR image can be identified, because the difference was effectively expanded through cumulative histogram learning in the region where the intensity difference

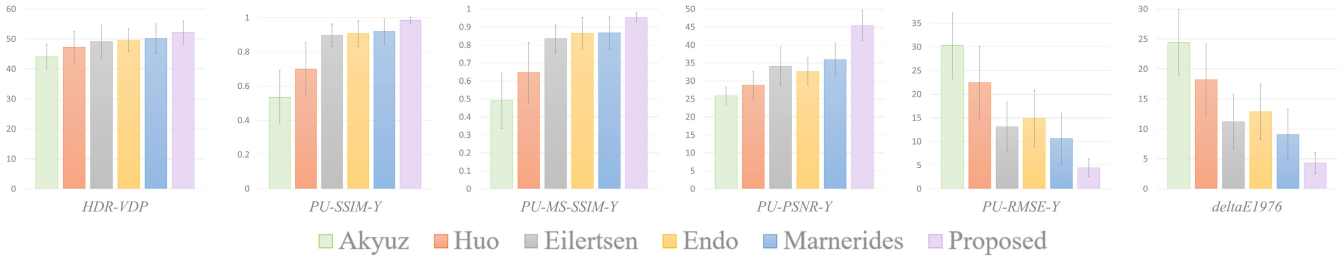


FIGURE 15. Quantitative comparison of the performance of several iTMOs. Six different objective evaluation metrics are calculated on HDR estimates compared with the ground-truth HDR images. All these metrics confirm that our proposed method outperforms the existing iTMOs in estimating the ground-truth HDR images.

was small. In the magnified versions of the second image (fourth row), the sunlight on the window is saturated and white in the LDR image. In the ground-truth HDR image, however, this region becomes distinguishable and displays a red hue. The HDR images estimated using the conventional methods do not estimate colors perfectly even though they expand the dynamic range in this region. In contrast, the estimated HDR image using our proposed method not only makes the shadows of the building more distinct, but also makes the color of sunlight look like that of the ground-truth HDR. This is because our proposed method learns color information separately through color difference learning, unlike other iTMOs. Finally, in the magnified versions of the third image (six row), the cloud appears to be saturated in the LDR image. In ground-truth HDR image, however, the cloud shape is visible. Akyüz, Huo, Eilertsen, and Endo's estimated HDR images not represent the cloud shape. Marnerides' estimated HDR image produces a rough cloud shape but is not clearly distinguishable. In contrast, the HDR image estimated using our proposed method distinguishes clouds from sunlight, thus resembling the cloud shape of the ground-truth HDR image. Overall, our HDR estimates are the most comparable to the ground-truth HDR image in both the global and local sense.

E. QUANTITATIVE COMPARISON WITH OBJECTIVE EVALUATION METRICS

The performance of our proposed method and the other iTMOs (Akyüz *et al.* [7], Huo *et al.* [13], Eilertsen *et al.* [17], Endo *et al.* [18], Marnerides*et al.* [20]) are compared based on the following six different objective evaluation metrics: HDR-VDP-2.2 [44], [45], PU-SSIM(Y) [46], [47], PU-MS-SSIM [48], PU-PSNR, PU-RMSE, and deltaE1976 [37]. These metrics are demonstrated to be well suited for HDR image evaluation [49], [50]. The descriptions for these metrics are provided in Appendix B. The quantitative evaluation metrics were calculated and averaged on a total of 570 images by adding the 120 data with varying types as well as the existing 450 data of ours. Figure 15 and Table 2 show the scores of HDR-VDP-2.2, PU-SSIM, PU-MS-SSIM, PU-PSNR, PU-RMSE, and deltaE1976. All these metrics confirm that our proposed method outperforms the existing iTMOs in estimating the ground-truth HDR images. This is especially apparent in deltaE1976, a metric for evaluating color difference as a Euclidean distance in the CIE$L^*a^*b^*$ color space. This is because deltaE1976 was used as the loss function in our proposed color difference learning method. In other words, compared with the conventional iTMOs, which only expand the dynamic range, the proposed method also considers the change of color.

TABLE 2. Quantitative evaluation of several iTMOs using HDR-specific evaluation metrics for a total of 570 images.

	Akyuz	Huo	Eilertsen	Endo	Marnerides	Proposed
HDR-VDP-2.2	44.03	47.31	49.11	49.67	50.19	52.18
PU-SSIM-Y	0.53	0.69	0.89	0.90	0.92	0.98
PU-MS-SSIM-Y	0.49	0.64	0.83	0.86	0.87	0.95
PU-PSNR-Y	25.84	28.82	34.09	32.68	35.96	45.39
PU-RMSE-Y	30.38	22.46	13.12	14.91	10.60	4.43
deltaE1976	24.46	18.20	11.17	12.85	9.04	4.29

V. CONCLUSION AND FUTURE WORK

HDR image estimation from a single LDR image is challenging task. Several previous iTMOs have focused on creating either linearly or nonlinearly expanded dynamic range images based on mathematical and perceptual models, but they have failed to produce good estimates of the ground-truth HDR images. Our deep-learning-based iTMO can be an effective solution to this challenging task. By learning the complicated relationship between various LDR and HDR images, we were able to produce a good HDR estimate from an arbitrary single LDR image. The objective evaluations confirm that our proposed method is superior to other existing iTMOs that are based on various mathematical models in estimating the ground-truth HDR images.

The main advantage of the proposed method over the existing mathematical-model-based iTMOs is that our deep-learning-based iTMO can produce a good estimate of the ground-truth HDR image itself and not just create an arbitrary

HDR-like image. Conventional iTMOs have been concerned mostly with how to expand Y, the absolute luminance value. They proposed a method of setting the maximum luminance and expanding the luminance accordingly. However, because the maximum luminance to be expanded for each image is different, it is dangerous to handle a static transform function for all images. Therefore, our model generates a transform function appropriately according to the image, and this is the greatest difference from the conventional iTMOs, which apply the static function equally to all images.

The main difference from conventional iTMOs is that our model uses cumulative histogram-based learning, not image-based learning. In other words, the dynamic range, which is the largest difference between LDR and HDR images, was trained with a one-dimensional cumulative histogram instead of a two-dimensional image. Through dimension reduction, the training time was shortened and the accuracy of learning was improved. Another difference from conventional iTMOs is that our model considers the color relationship of the LDR–HDR image. Even if the HDR brightness may have been perfectly predicted in the previous process, if the chromatic components are components of the LDR image, the ground-truth HDR color will not be represented (as demonstrated in Figure 7). Therefore, we performed color learning using deltaE1976, a color difference metric that considers human color recognition, as a loss function. By performing one more specialized learning process on color, it shows much similar color to that of the ground-truth HDR image than the conventional iTMOs.

Our proposed method has a limitation in restoring perfectly lost information. The 8-bit LDR images are so saturated that the intensity is tightly compressed, resulting in clipping or quantization. These regions have small intensity differences that are indistinguishable by the eye, or they are clipped to a single intensity, completely losing the original information. Our method uses histogram matching in the process of making the estimated HDR brightness image using the learned cumulative histogram. However, because histogram matching involves one-to-one mapping for each pixel in the LDR image, it is not possible to restore the perfectly clipped region, although it works effectively to distinguish areas of small intensity differences. We recognized the limitations of histogram matching and designed the method to learn local information through a 3×3 convolution layer when processing additional color difference learning in the two-dimensional image domain. In addition, our model learns some clipped regions through pooling and up-sampling using a U-net structure. However, this may be insufficient because it does not intensively learn only clipped areas. Therefore, the next study will investigate how to estimate clipped regions without artifacts to improve our deep-learning-based iTMO. Another limitation is in noise handling. Recently, many papers have been proposed to study the noise amplification that occurs when dealing with low-light images [51]. However, our proposed method does not currently focus on noise reduction. If the LDR image itself is noisy, histogram matching does not reduce the noise. Also, since the proposed deep learning model is not specialized for noise reduction, it will not completely reduce noise. To solve this problem, we will study the network structure that additionally connects the convolution layer that has been specialized in noise reduction. Finally, the HDR used to train the model is an image merged from LDR images with EVs of -2, 0, and 2. Therefore, the estimated HDR generated by the trained model only represents the dynamic range of HDR merged into LDR images with -2, 0, and 2 EVs. The next step is to build a model that generates HDR images with a wider dynamic range by training with the HDR data representing a wider dynamic range.

APPENDIX

A. CIEL*a*b* COLOR SPACE CONVERSION

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.4124 & 0.3576 & 0.1805 \\ 0.2126 & 0.7152 & 0.0722 \\ 0.0193 & 0.1192 & 0.9505 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{4}$$

$$\begin{aligned} L^* &= 116f\left(\frac{Y}{Y_n}\right) - 16 \\ a^* &= 500\left(f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right)\right) \\ b^* &= 200\left(f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right)\right) \end{aligned} \tag{5}$$

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

B. OBJECTIVE EVALUATION METRICS

1. HDR-VPD-2.2 [44], [45]

$$Q_{hdrvdp} = \frac{1}{F \cdot O} \sum_{f=1}^{F} \sum_{o=1}^{O} w_f \log\left(\frac{1}{I} \sum_{i=1}^{I} D_p^2[f, o](i) + \varepsilon\right) \tag{7}$$

i: pixel index
f: spatial frequency (1 to F)
o: orientation (1 to O)
D_p : noise-normalized difference between the fth and oth of the steerable pyramid for the reference and test images
ε : small constant (10^{-5}) to avoid singularities when D_p is close to 0
I: total number of pixels
w_f : vector of per-band pooling weights

2. PU ENCODING [46]

$$cvi(L, L_a) = (max[CSF(L_a, x)MA(|L - L_a|)])^{-1} \tag{8}$$

$$t(L) = cvi(L, \max(L, L_{a-min})) \tag{9}$$

$$f_i = f_{i-1}(1 + t(f_{i-1})) \quad \text{where f} : L' \to L, \quad i \in [2 \cdots N] \tag{10}$$

$$PU : L \to L' \tag{11}$$

We used the values of f as a lookup table and found the nearest index i for a given luminance value L.

cvi:	contrast versus intensity
CSF:	contrast sensitivity function
L_a :	adapting luminance
L:	background luminance
MA:	estimate of the loss of sensitivity
x:	corresponds to all the parameters (such as spatial frequency, orientation, and stimuli size)
L_{a-min} :	all luminance levels above this value
t:	final estimates of the detection thresholds
i:	luma value
f_i :	gives the luminance value associated with a particular i

3. SSIM [47]

$$Q_{ssim} = \frac{\left(2\mu_x\mu_y + C_1\right)\left(2\sigma_{xy} + C_2\right)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)} \tag{12}$$

μ_x, μ_x:	(mean intensity of image x,y (luminance term)
σ_x, σ_x :	standard deviation of image x,y (contrast term)
σ_{xy} :	$\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu_x)\left(y_i - \mu_y\right)$(structural term)
C_1, C_2 :	small constant to be stable (empirical value)

4. MS-SSIM [48]

$$Q_{msssim} = \frac{1}{M}\sum_{j=1}^{M} SSIM(x_j, y_j) \tag{13}$$

Calculate the mean SSIM index for the whole image.

5. PSNR

$$Q_{psnr} = 10\log_{10}\left[\frac{max\,(E)^2}{\frac{1}{M\times N}\sum_i^M\sum_j^N\left[G\,(i,j) - E(i,j)\right]^2}\right] \tag{14}$$

G:	ground-truth image
E :	estimated image
i, j :	pixel index
$max(E)$:	maximum possible pixel value of the estimated image

6. RMSE

$$Q_{rmse} = \sqrt{\frac{1}{M\times N}\sum_{i}^{M}\sum_{j}^{N}\left[G\,(i,j) - E(i,j)\right]^2} \tag{15}$$

G :	ground-truth image
E :	estimated image
i, j :	pixel index

7. deltaE1976 [37]

$$Q_{deltaE1976} = \sqrt{(\mathrm{L}_G^* - \mathrm{L}_E^*)^2 + (a_G^* - a_E^*)^2 + (b_G^* - b_E^*)^2} \tag{16}$$

Calculate in the CIE$L^*a^*b^*$ color space.

G :	ground-truth image
E :	estimated image

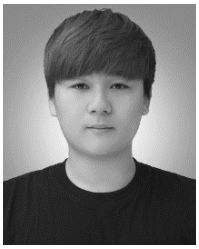
ACKNOWLEDGMENT

The authors would like to thank Inyong Park for discussion and helpful comments on this article.

REFERENCES

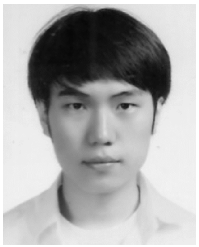
[1] T. Kunkel and E. Reinhard, "A reassessment of the simultaneous dynamic range of the human visual system," in *Proc. 7th Symp. Appl. Perception Graph. Vis. (APGV)*, 2010, pp. 17–24.
[2] S. Mann and R. W. Picard, "On being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures," in *Proc. IST*, May 1995, pp. 442–448.
[3] A. Deepa and S. Muthumariammal, "An overview of ghost artifact in HDRI," *SIJ Trans. Comput. Sci. Eng. Appl.*, vol. 2, no. 1, pp. 1–6, Feb. 2014.
[4] E. A. Khan, A. O. Akyuz, and E. Reinhard, "Ghost removal in high dynamic range images," in *Proc. Int. Conf. Image Process.*, Oct. 2006, pp. 2005–2008.
[5] S. Hajisharif, J. Kronander, and J. Unger, "Adaptive dualISO HDR reconstruction," *EURASIP J. Image Video Process.*, vol. 2015, no. 1, p. 41, Dec. 2015.
[6] A. Serrano, F. Heide, D. Gutierrez, G. Wetzstein, and B. Masia, "Convolutional sparse coding for high dynamic range imaging," *Comput. Graph. Forum*, vol. 35, no. 2, pp. 153–163, May 2016.
[7] A. O. Akyüz, R. Fleming, B. E. Riecke, E. Reinhard, and H. H. Bülthoff, "Do HDR displays support LDR content?: A psychophysical evaluation," *ACM Trans. Graph.*, vol. 26, no. 3, p. 38, 2007.
[8] L. Meylan, S. Daly, and S. Süsstrunk, "The reproduction of specular highlights on high dynamic range displays," in *Proc. Color Imag. Conf.*, 2006, vol. 2006, no. 1, pp. 333–338.
[9] B. Masia, S. Agustin, R. W. Fleming, O. Sorkine, and D. Gutierrez, "Evaluation of reverse tone mapping through varying exposure conditions," *ACM Trans. Graph.*, vol. 28, no. 5, pp. 1–8, Dec. 2009.
[10] B. Masia, R. Fleming, O. Sorkine, and D. Gutierrez, "Selective reverse tone mapping," in *Proc. Congreso Espanol de Inf. Grafical Eurograph.*, 2010, pp. 1–10.
[11] F. Banterle, P. Ledda, K. Debattista, and A. Chalmers, "Inverse tone mapping," in *Proc. 4th Int. Conf. Comput. Graph. Interact. Techn. Australasia Southeast Asia*, 2006, pp. 349–356.
[12] 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.
[13] 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, Sep. 2013.
[14] R. Shapley and C. Enroth-Cugell, "Chapter 9 visual adaptation and retinal gain controls," *Progr. Retinal Res.*, vol. 3, pp. 263–346, Jan. 1984.
[15] J. Zhang and J.-F. Lalonde, "Learning high dynamic range from outdoor panoramas," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4519–4528.
[16] N. K. Kalantari and R. Ramamoorthi, "Deep high dynamic range imaging of dynamic scenes," *ACM Trans. Graph.*, vol. 36, no. 4, pp. 144:1–144:12, 2017.
[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, p. 178, 2017. [Online]. Available: http://github.com/gabrieleilertsen/hdrcnn
[18] 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.
[19] S. Lee, G. H. An, and S.-J. Kang, "Deep chain HDRI: Reconstructing a high dynamic range image from a single low dynamic range image," *IEEE Access*, vol. 6, pp. 49913–49924, 2018.
[20] 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. [Online]. Available: https://github.com/dmarnerides/hdr-expandnet

[21] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Proc. ACM SIGGRAPH Classes (SIGGRAPH)*, 2008, p. 31.
[22] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proc. IJCAI*, 1995, vol. 14, no. 2, pp. 1137–1145.
[23] E. Reinhard, W. Heidrich, P. Debevec, S. Pattanaik, G. Ward, and K. Myszkowski, *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting*. San Mateo, CA, USA: Morgan Kaufmann, 2010.
[24] R. S. Hunter, "Photoelectric color difference meter," *J. Opt. Soc. Amer.*, vol. 48, no. 12, pp. 985–995, Dec. 1958.
[25] H. Khalid, P. Kazemi, L. Perez-Gandarillas, A. Michrafy, J. Szlek, R. Jachowicz, and A. Mendyk, "Computational intelligence models to predict porosity of tablets using minimum features," *Drug Des., Develop. Therapy*, vol. 11, pp. 193–202, Jan. 2017.
[26] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556
[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[28] W. Luo, Y. Li, R. Urtasun, and R. Zemel, "Understanding the effective receptive field in deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 4898–4906.
[29] F. N. Fritsch and R. E. Carlson, "Monotone piecewise cubic interpolation," *SIAM J. Numer. Anal.*, vol. 17, no. 2, pp. 238–246, 1980.
[30] 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. (MMSys)*, 2015, pp. 219–224.
[31] M. Reichmann. *Understanding Raw Files Explained*. [Online]. Available: http://www.luminouslandscape.com/tutorials/understandingseries/uraw-files.shtml
[32] 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.
[33] T.-H. Wang, C.-W. Chiu, W.-C. Wu, J.-W. Wang, C.-Y. Lin, C.-T. Chiu, and J.-J. Liou, "Pseudo-Multiple-Exposure-Based tone fusion with local region adjustment," *IEEE Trans. Multimedia*, vol. 17, no. 4, pp. 470–484, Apr. 2015.
[34] H. Jang, K. Bang, J. Jang, and D. Hwang, "Inverse tone mapping operator using sequential deep neural networks based on the human visual system," *IEEE Access*, vol. 6, pp. 52058–52072, 2018.
[35] R. C. Gonzalez and R. E. Woods, "Object recognition," in *Digital Image Processing*, 3rd ed. London, U.K.: Pearson, 2008, pp. 861–909.
[36] 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.
[37] A. R. Robertson, "The CIE 1976 color-difference formulae," *Color Res. Appl.*, vol. 2, no. 1, pp. 7–11, Jun. 2014.
[38] H. Liu, Z. Hu, A. Mian, H. Tian, and X. Zhu, "A new user similarity model to improve the accuracy of collaborative filtering," *Knowl.-Based Syst.*, vol. 56, pp. 156–166, Jan. 2014.
[39] J. H. McDonald, *Handbook of Biological Statistics*. Baltimore, MD, USA: Sparky House Publishing, 2009.
[40] C.-C. Lee, H.-C. Wu, C.-S. Tsai, and Y.-P. Chu, "Adaptive lossless steganographic scheme with centralized difference expansion," *Pattern Recognit.*, vol. 41, no. 6, pp. 2097–2106, Jun. 2008.
[41] T.-C. Chang, S. S.-D. Xu, and S.-F. Su, "SSIM-based quality-on-demand energy-saving schemes for OLED displays," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 46, no. 5, pp. 623–635, May 2016.
[42] M. Fairchild. (2008). *Fairchild's HDR Photographic Survey*. [Online]. Available: http://rit-mcsl.org/fairchild/HDR.html
[43] P. Korshunov, T. Hanhart, and P. A. Ebrahimi. (2015). *EPFL's Dataset of HDR Images*. [Online]. Available: http://mmspg.epfl.ch/hdr-eye
[44] R. Mantiuk, K. J. Kim, A. G. Rempel, and W. Heidrich, "HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions," *ACM Trans. Graph.*, vol. 30, no. 4, p. 40, 2011.
[45] 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.
[46] T. O. Aydın, R. Mantiuk, and H. P. Seidel, "Extending quality metrics to full luminance range images," *Proc. SPIE*, vol. 6806, p. 68060B, Mar. 2008.
[47] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.
[48] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. 7th Asilomar Conf. Signals, Syst. Comput.*, vol. 2, 2003, pp. 1398–1402.
[49] P. Hanhart, M. V. Bernardo, M. Pereira, A. M. G. Pinheiro, and T. Ebrahimi, "Benchmarking of objective quality metrics for HDR image quality assessment," *EURASIP J. Image Video Process.*, vol. 2015, no. 1, p. 39, Dec. 2015.
[50] E. Zerman, G. Valenzise, and F. Dufaux, "An extensive performance evaluation of full-reference HDR image quality metrics," *Qual. User Exper.*, vol. 2, no. 1, p. 5, Apr. 2017.
[51] X. Guo, "LIME: A method for low-light image enhancement," in *Proc. 24th ACM Int. Conf. Multimedia*, 2016, pp. 87–91.



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. degree 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 a Professor at Yonsei University. His research interests include deep-learning based image processing, medical imaging, and tomographic reconstruction.