IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 13, 2018, accepted August 17, 2018, date of publication September 10, 2018, date of current version September 28, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2868246

Deep Chain HDRI: Reconstructing a High Dynamic Range Image from a Single Low Dynamic Range Image

SIYEONG LEE, GWON HWAN AN, AND SUK-JU KANG, (Member, IEEE)
Department of Electronic Engineering, Sogang University, Seoul 04107, South Korea

Corresponding author: Suk-Ju Kang (sjkang@sogang.ac.kr)

This work was supported in part by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) and the Ministry of Trade, Industry & Energy (MOTIE) of the Republic of Korea under Grant 20161210200560 and in part by the Korea Electric Power Corporation under Grant R17XA05-28.

ABSTRACT Recently, high dynamic range (HDR) imaging has attracted much attention as a technology to reflect human visual characteristics owing to the development of the display and camera technology. This paper proposes a novel deep neural network model that reconstructs an HDR image from a single low dynamic range (LDR) image. The proposed model is based on a convolutional neural network composed of dilated convolutional layers and infers LDR images with various exposures and illumination from a single LDR image of the same scene. Then, the final HDR image can be formed by merging these inference results. It is relatively simple for the proposed method to find the mapping between the LDR and an HDR with a different bit depth because of the chaining structure inferring the relationship between the LDR images with brighter (or darker) exposures from a given LDR image. The method not only extends the range but also has the advantage of restoring the light information of the actual physical world. The proposed method is an end-to-end reconstruction process, and it has the advantage of being able to easily combine a network to extend an additional range. In the experimental results, the proposed method shows quantitative and qualitative improvement in performance, compared with the conventional algorithms.

INDEX TERMS High dynamic range imaging, image restoration, computational photography, convolutional neural network.

I. INTRODUCTION

Image restoration is an important field in image processing and computer vision. This task restores an original image by using prior knowledge on the degradation phenomena. Unlike image enhancement, the main purpose of image restoration is to restore a latent clean image $\mathbf{x}$ from a corrupted image $\mathbf{y} = H(\mathbf{x}) + \mathbf{d}$, where H is the degradation function and $\mathbf{d}$ is the additive noise.

Recently, several approaches for inferring missing information through deep learning have been proposed [1]–[5]. As a function approximator that infers the unknown mapping between input and output image sets, deep neural networks have advanced performance in the image restoration field such as super resolution [1], [4], deblurring, and denoising [5].

Similarly, efforts to acquire original images close to those in the actual physical world, called high dynamic range imaging (HDRI), have also continued [6]–[8]. Debevec and Malik [9] proposed an HDRI method to expand a narrow (or low) dynamic range owing to the limitations of the camera sensor. This method estimates the camera response function (CRF) from images with different exposures and derives the radiance information using the estimated CRF, which makes it possible to obtain image information close to the information in the real world. In addition, because of the considerable improvements in display technology, it is possible to express larger luminance ranges than in the past. As a result, interest in generating images with a high quality that is close to that of the real world has increased, and image restoration optimized for high dynamic range (HDR) displays from existing low dynamic range (LDR) images has also become important. Typically, existing images have an LDR that has lost specific information about the captured image owing to camera limitations, and it is impossible to

retake these images. Therefore, recovering the dynamic range is an ill-posed problem and can be considered as an image restoration problem. To solve this problem, inverse tone mapping (ITM) [10] has been proposed. However, conventional ITM algorithms [11]–[20] do not infer physical brightness information, but focus on adjusting the brightness values in specific areas such as the highlight regions [18] to create a perceptually high-quality image. In addition, because an image with a narrow range is enlarged to an image with a wide range, it is difficult to find an appropriate relationship between two spaces with different ranges.

To find the missing information from a single LDR image, we propose a method of restoring an HDR that is close to the real-world range using a deep neural network. The proposed method is a novel ITM method that produces a multiple exposure image stack from a single LDR image using a deep neural network with a chain structure.

II. RELATED WORKS

A. HIGH DYNAMIC RANGE IMAGING (HDRI)

Because of the physical limitations of a charge-coupled device sensor, a digital camera captures a single LDR image with limited dynamic range scene information. This image has a certain exposure value (EV) [9], which is the amount of light that reaches the sensor of the digital camera. It is controlled by the aperture, shutter speed, and sensor sensitivity, and is defined as follows:

$$EV = 2\log_2 F - \log_2 S + \log_2(\frac{ISO}{100}) \tag{1}$$

where F is the relative aperture (F-number), S is the exposure time (=1/shutter speed), and ISO is the sensor sensitivity [21]. When displaying an image with a specific EV, there is a difference between the image and the scene observed by a human because the range of the image representation in the camera is smaller than the human perception range. To solve this problem, Debevec and Malik [9] proposed an HDRI technique that estimates the CRF from multiple LDR images with different exposure levels and extracts an omnidirectional HDR radiance map of the physical world. The relationship between the physical brightness and image pixel level is modeled as follows [22]:

$$Z_i = f([g_{cv}(C_i + D_i) + N_{reset}]g_{out} + N_{out} + Q \tag{2}$$

where f is the CRF, g is the gain of the camera, Z_i is the pixel value, C_i is the number of photons, D_i is the dark shot noise, N is additional noise, and Q is the uniformly distributed quantization error, which occurs during the conversion from analog voltage values to digital quantized values. These methods [9], [22] are limited because it is difficult to capture multiple exposure LDR images for a given scene at the same time. Even if the multiple exposure LDR images are taken and merged to create an HDR image, the method is sensitive to changes caused by moving objects or illumination, thereby degrading the HDR image quality. To solve these problems, several methods have been proposed [23]–[26] to enhance image quality.

To overcome the disadvantage of requiring multiple shots for HDRI, Tocci *et al.* [47] and Kronander *et al.* [48] developed several optical architectures with multiple sensors, and Wahab *et al.* [49] proposed an HDRI method using the light field (or plenoptic) camera. Unlike conventional standard cameras, multiple images with different exposures can be acquired with a single shot using these techniques. However, these techniques still have the limitation that multiple images are needed to make an HDR image. In contrast to conventional methods requiring multiple images, we propose a process of reconstructing an HDR image from a single LDR image, which can be acquired using standard cameras, without any assumption.

B. INVERSE TONE MAPPING (ITM)

An HDR display has the expanded dynamic range to represent bright and dark areas better than an LDR display. Therefore, displaying the large number of existing LDR images on an HDR display has become a critical issue. For LDR images, there is no information in the saturated regions and dark regions owing to the limitations of the dynamic range. Hence, when an HDR image is generated from a single LDR image, the corresponding regions are difficult to restore. The method for solving the LDR-to-HDR conversion problem is called ITM, and several algorithms have been proposed [11]–[20]. Banterle *et al.* [12] proposed an ITM method that detects highlight regions and extends the range of those regions. Masia *et al.* [16], [17] proposed an exponential expansion method, and Meylan *et al.* [18] proposed a piecewise linear mapping function that further increases the range for the bright portions of the image. Rempel *et al.* [19] used a brightness enhancement map to linearly increase the contrast of the intermediate range. Further, they restored the saturated pixel values using an edge stopping function. Kovaleski and Oliveira [15] also used a bilateral grid to broaden the dynamic range. Unlike existing methods, Wang *et al.* [20] proposed a region-based enhancement of the pseudo-exposures to generate an HDR image. Although these algorithms change an LDR image into an HDR image with a wide range, the expansion [11], [16] is not correctly performed for an image if the parameters are not set appropriately for a given input. Other algorithms [13], [15], [19] cause contour artifacts on bright objects because of boosting the brightness of the saturated areas, and image quality degrades as a negative consequence of the additional processing required to remove artifacts. To solve these problems, Huo *et al.* [14] proposed a method that considers the human visual system, using perceptual brightness rather than absolute brightness. However, because it is based on the local adaptive response of the retina, it is difficult to obtain HDR images that match the actual brightness. Recently, methods for restoring lost dynamic range using deep learning have been proposed [27]–[29].

C. CONVOLUTIONAL NEURAL NETWORK (CNN)

A CNN automatically extracts a feature map by using a loss function defined by the designer when the data set is given

and minimizes the error between the inferred and reference values. Because of these advantages, CNNs have made many improvements in the field of image restoration. Specifically, in the cases of ResNet [2] and DenseNet [3], it is possible to learn deeper structures through the skip-connections between low-layer information and high-layer abstract information. The VDSR approach [4] obtained good results by using residual blocks. Zhang *et al.* [30] also restored colors from grayscale images using CNNs.

In recent years, several methods have been presented to convert a single LDR image into an HDR image by using a deep neural network. Zhang and Lalonde [29] proposed a method for converting an LDR panoramic image into an HDR image through deep learning, but the input LDR image has a resolution of 64 × 128 pixels, and the method is more suitable for finding the light source position. Endo *et al.* [27] proposed a method of creating an exposure stack from a single LDR image, and this network generates images of different exposure levels with the same depth. Conversely, Eilerstsen *et al.* [28] proposed a neural network for converting an LDR image to an HDR image, which is not an end-to-end neural network structure, and mainly focuses on restoring the saturated region for underexposed LDR inputs. Unlike Zhang and Lalonde [29], the methods proposed by Endo *et al.* [27] and Eilertsen *et al.* [28] are relatively free of pixel resolution, but not perfect.

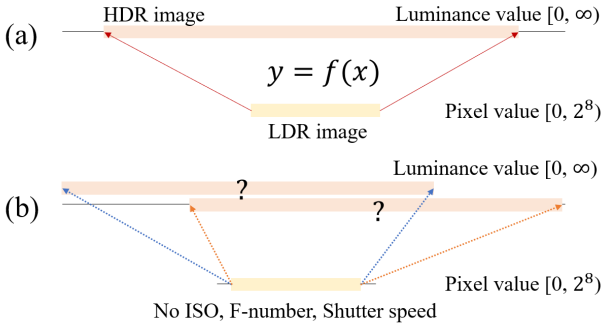

FIGURE 1. ITM problem. Two problems arise when generating an HDR image from a single LDR image. First, as the range becomes wider, as in (a), there is a lack of mapping information and it becomes difficult to map. Second, if the metadata for the input image does not exist, as in (b), it is impossible to accurately estimate the HDR luminance because the pixel value may be the same depending on the EV, even though it is a different scene.

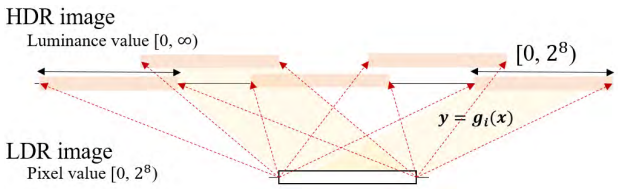

FIGURE 2. Proposed multiple exposure image stack. To obtain an HDR image using the proposed method, several subnetworks generate LDR images with various exposure levels rather than inferring the entire part at once.

III. DIFFICULTIES OF DIRECT LDR-TO-HDR MAPPING

Before describing the details of the proposed neural network architecture, we first explain the feasibility of the neural network structure, which generates a multiple exposure image stack from a single LDR image rather than direct LDR-to-HDR mapping. In terms of restoring the information for an ill-posed problem, a neural network would be an ideal problem solver if it directly extracts the actual scene radiance values from a single LDR image. However, in Fig. 1(a), as the range required for restoration is widened, it is difficult to infer the relationship between the two image sets. In addition, it is impossible to simply expand the dynamic range, as shown in Fig. 1(b), and hence, the metadata (e.g., sensor sensitivity, F-number, and shutter speed) of the input LDR image are required to infer the radiance values of the actual scene. Generally, most existing LDR images do not have EV information, and hence, there is a distinct limitation to restore the actual scene radiance values. We assume that the existing LDR images are well-captured or properly exposed because the appropriate (or optimal) exposure value will have been typically selected by humans. The single input LDR image of the proposed network is defined as a middle exposure (EV 0) image. (A description of middle exposure in the proposed neural network is given in Section IV-D.) By assuming a middle exposure and using a multiple exposure image stack, it is possible to train and infer ±1, ±2, and ±3 EV LDR images that contain higher or lower exposure information, as shown in Fig. 2. Even if the EV value is not known for the existing LDR image, we can obtain a tone-mapped HDR image that is well-fitted to the wider range display. In addition, if the EV for the input LDR image is known, the scene radiance of the HDR can be inferred through Debevec and Malik's method [9]. Therefore, we propose a neural network that infers the multiple exposure image stack from a single LDR image to find the relationship between the HDR image from the LDR image, which is defined as a middle exposure image.

IV. PROPOSED DEEP CHAIN HDRI

A. OVERALL NETWORK ARCHITECTURE

To generate bracketed images for HDRI reconstruction, the proposed method produces multiple images with different exposures gradually from the intermediate image. The structure of the proposed neural network, which consists of six subnetworks, is shown in Fig. 3. This network produces images with the top three exposures and the bottom three exposures from the input LDR image with the middle exposure. In other words, the proposed method produces a stack through multiple outputs, whereas the conventional method [27] produces bracketed images at once.

To produce a stack with wider exposure, the entire network of [27] needs to be re-trained. However, the proposed method is more scalable because the network structure can easily generate additional images by training only a relatively small subnetwork and connecting it to the existing network. Therefore, the proposed network constructs the sequential

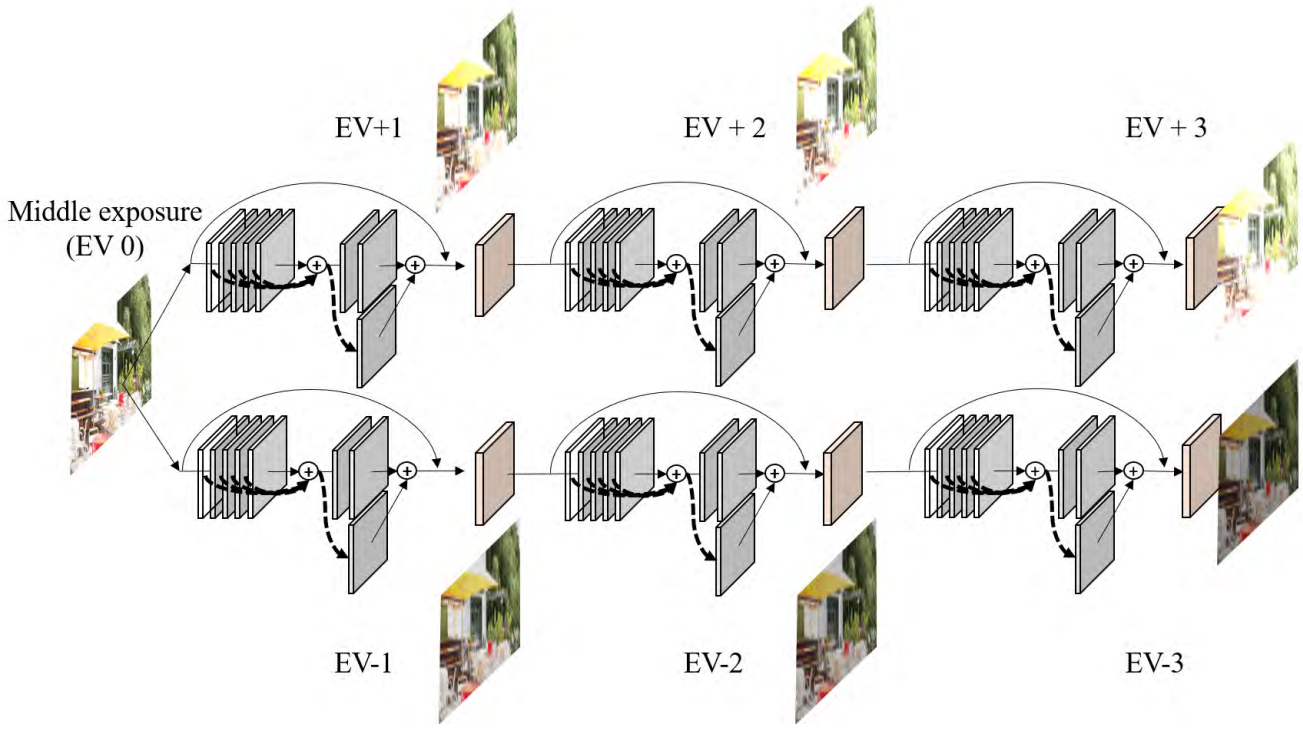

FIGURE 3. Proposed deep chain HDRI architecture. Given an LDR image with a middle exposure value (EV 0), the EV ±1, ±2, ±3 images are inferred sequentially through the network. When the EV of the inferred image is far from the middle exposure value, the structure depth goes deeper than that of the image that has less exposure difference to infer the mapping relation more accurately. After finishing the process through the proposed network, a total of six LDR images are inferred to generate the LDR image stack. Then, an HDR image is generated using the HDRI technique.

learning process. With this overall network structure, it is possible to generate patches that are farther from the input by increasing the depth of the neural network.

B. SUBNETWORK ARCHITECTURE

As shown in Fig. 4, we use a CNN-based subnetwork to create an LDR image from a given EV i to EV j. The subnetwork is affected by the DCSCN architecture [31]. It uses a 64×64 LDR patch in the image with an EV of i as an input and produces a 64×64 LDR patch with an EV of j as the output. The subnetwork is divided into two parts. The front part consists of a total of seven feature extraction blocks, and the rear part consists of four reconstruction blocks. Each block consists of a (dilated) convolution layer, a batch normalization layer, and an activation layer. We use PReLU [32] as an activation function for a block that infers brighter images and MPReLU, which is first proposed in this paper, as an activation function for a block that infers darker images. (Additional description of MPReLU is given in Section V-B.) The feature extraction network operates as an abstraction of feature extraction for a given input. Different from DCSCN [31], the feature extraction network in each subnetwork does not concatenate feature maps from a previous layer but adds them after extracting multi-scale feature maps. By considering both the low-level and the high-level features, it is possible to generate a high-quality image in the reconstruction process.

The reconstruction network consists of path 1 (r_1, r_2), path 2 (r_3), and r_4, and reconstructs the image using the extracted features, as shown in Fig. 4. By using a relatively large kernel compared with path 2, path 1 can be restored by taking into consideration a wider range of information. Unlike the layers of other reconstruction networks, r_4 uses the *tanh* function as an activation function to enable the representation of the residual.

Additionally, when extracting feature maps from a CNN, the receptive field is significant. Therefore, we set the dilation parameters of the convolution layer of the feature extraction block to 1, 1, 2, 3, 5, 8, and 13 respectively, to set the feature extraction network's receptive field to be larger than the input patch size. In other words, as the size of the receptive field is 67, we can consider all the information in the patch. Each layer in the feature extraction network has 32 kernels of 3×3 size. In the reconstruction network, each layer consists of 32 kernels. The kernel size of r_1, r_2, r_3, and r_4 are 1×1, 3×3, 1×1, and 3×3, respectively.

C. TRAINING

To train the proposed model, we divided the entire network into subnetworks and trained each subnetwork instead of the entire network at once. As we want the subnetwork N_j to estimates the image with EV j from EV i, the input of this network was set to the ground truth image with EV i (or the image inferred from the previous subnetwork), and

TABLE 1. Detail architecture of subnetwork.

		Layer	Activation size
Feature extraction network		Input	64×64×3
	f_1	3×3×32, stride 1, dilated 1	64×64×32
	f_2	3×3×32, stride 1, dilated 1	64×64×32
	f_3	3×3×32, stride 1, dilated 2	64×64×32
	f_4	3×3×32, stride 1, dilated 3	64×64×32
	f_5	3×3×32, stride 1, dilated 5	64×64×32
	f_6	3×3×32, stride 1, dilated 8	64×64×32
	f_7	3×3×32, stride 1, dilated 13	64×64×32
Reconstruction network	r_1	1×1×32, stride 1, dilated 1	64×64×32
	r_2	3×3×32, stride 1, dilated 1	64×64×32
	r_3	1×1×32, stride 1, dilated 1	64×64×32
	r_4	1×1×3, stride 1, dilated 1	64×64×3

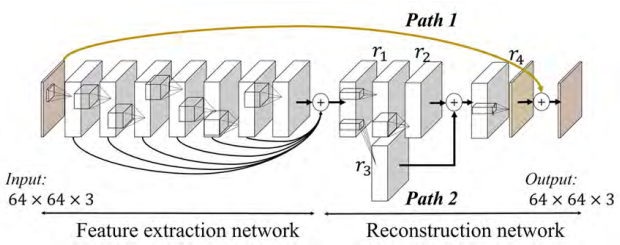

FIGURE 4. Subnetwork architecture.

the output was matched to the ground truth image with EV j. The optimization proceeds in the direction of minimizing these losses. To optimize the weights and biases of the neural network, we used the Adam optimizer [33] with a learning rate of 0.001 and momentum parameter β_1 of 0.9. At the time of training, the batch size was set to one.

Each subnetwork uses the batch normalization to produce regularization effects and prevent the vanishing gradients that can occur during learning. In addition, by separating the entire network into subnetworks and training them, vanishing gradients, which can occur in the deep structure for EV ±3 learning, are prevented. Each subnetwork was trained using a Nvidia GeForce 1080 Ti GPU with 100 epochs for about 48 hours.

1) LOSS FUNCTION

The proposed CNN structure is a network structure with multiple outputs from one input image. Given the output $\hat{y}_i$ of EV i and the target image y_i, we set the loss function as follows.

$$\mathbb{L}_{all}(\hat{y}_i, y_i) = l_{pixel}(\hat{y}_i, y_i) + \lambda l_{tv}(\hat{y}_i) \tag{3}$$

where l_{pixel} is the pixel loss and l_{tv} is the total variation regularization. Experimentally, we set the relative weights of each loss to $\lambda = 10^{-4}$. Therefore, the entire network trains to minimize the loss of each of y_i and $\hat{y}_i$.

Pixel loss The pixel loss is an L1 norm between an output $\hat{y}_i$ and a target y_i. If both have shape $w \times h \times c$, then the pixel loss is defined as

$$l_{pixel}(\hat{y}_i, y_i) = \frac{1}{whc} \sum_{u,v} \left\| \hat{y}_i(u, v) - y_i(u, v) \right\|_1 \tag{4}$$

where w is the width of the image, h is the height of the image, c is the channel of the image, and u and v are the pixel coordinates.

Total variation regularization To improve the smoothness of the inferred image and prevent it from overfitting to a specific pattern, we add total variation regularization.

$$l_{tv}(\hat{y}_i) = \sum_{u,v} \sqrt{(\hat{y}_i(u, v+1) - \hat{y}_i(u, v))^2 + (\hat{y}_i(u+1, v) - \hat{y}_i(u, v))^2} \tag{5}$$

2) PATCH-BASED LEARNING

The brightness caused by light radiation in an image has a localized property. Hence, we trained the proposed network to infer the relationship between patches with different exposures. Furthermore, unlike [27] and [28], the inputs of the proposed method are patches extracted from the input image, instead of the entire image. Therefore, the network transforms an input image into patches with different exposure.

We make a set of 64 × 64 patches $p \in [0, 255]^{64\times64}$ from the image $I \in [0, 255]^{w\times h}$ corresponding to EV i with a stride of 10. As a result, each subnetwork that infers an EV j image from an EV i image is trained with about 180,000 patch pairs.

3) RESIDUAL LEARNING

In the case of image transformations that learn the relationship between different images, a neural network often loses the morphological information from a given input image while optimizing the loss function. Therefore, to avoid the loss of spatial information, the neural network is designed to learn the residual image from the ground truth image.

D. DATASET

1) MIDDLE EXPOSURE

The correct exposure is difficult to define because it reflects a subjective viewpoint. However, we assume that a correctly exposed image depicts all parts of the image in detail. In other words, pixels are uniformly distributed throughout the grayscale range. This is defined as a middle exposure from a technical point of view. Therefore, when a multiple exposure image stack is obtained by the auto bracketing function of the camera, which changes the EV automatically when capturing images, we can determine the middle exposure image, which is the image with the most evenly distributed histogram of the images of the stack. In addition, we define the EV of the corresponding image as the middle exposure value (EV 0).

2) NEW DATA SET OF MULTIPLE EXPOSURE IMAGE STACK

For training each subnetwork, we need seven multiple exposure ground truth images satisfying EV 0, ±1, ±2, ±3 for static scenes. Generally, many datasets are required to train

a neural network, but only five stacks satisfy this condition among the existing HDR datasets [34], [35]. Therefore, we captured 96 different scene image stacks using a Nikon D700 with a resolution of 4256 × 2832 pixels (672 images; outdoor: 504, indoor: 168 images) to train and test the proposed network. We used a tripod to minimize image blur, set the aperture value to $f/4$, and automatically adjusted the sensor sensitivity and shutter speed using the auto bracketing function. Then, EV ±1, ±2, ±3 images from the middle exposure image were stored as shown in Fig. 5. Each stack consists of seven images with different exposure values for one scene. As shown in Fig. 6, the new dataset consists of various types of images: indoor and outdoor with an artificial light source and natural light, modern buildings, and wooded grounds. (The dataset will be updated publicly after published.) We shuffled and split the dataset into a training set, validation set, and test set. The ratio of each set is 7 : 3 : 10, respectively. All images used in the training process were resized to a resolution of 912×608 pixels with an 8-bit depth.

3) HDRI RECONSTRUCTION

For a given image $I \in [0, 255]^{w \times h}$ with EV 0, we slice I into the 64×64 LDR patches with a stride of 1. After transforming the image into a tensor shape, we infer a patch with the upper (or lower) EV through the neural network and reconstruct $I_{ev\pm1} \in [0, 255]^{w \times h}$ using the reconstruction process. Then, the process is repeated to generate LDR images with other EVs and construct the multiple LDR image stack. After that, an HDR image is merged using the HDRI method proposed by Debevec and Malik [9].

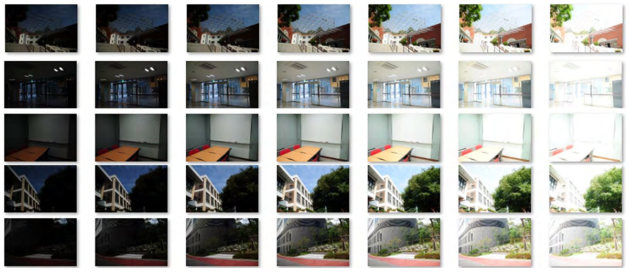

FIGURE 5. Sample images for the new dataset consisting of multiple exposure image stacks. The new dataset consists of a total of 96 stacks (outdoor: 72, indoor: 24). Each stack is composed of seven images with different exposure.

FIGURE 6. VDS dataset: the dataset contains 96 scenes that cover a wide variety of content, e.g., natural scenes (both indoor and outdoor), wooded grounds, buildings, etc.

V. UNDERSTANDING THE PROPERTIES OF THE PROPOSED METHOD

In this section, we analyze the reasons for the validity of the proposed neural network architecture and the characteristics of the network. First, we examine the necessity for having a chain structure as the exposure difference increases. Second, we analyze the settings of the activation function and its properties to solve the problems that may arise through residual learning for various exposures in the proposed method.

A. WHY NEED THE DEEP CHAIN STRUCTURE?

The average peak signal-to-noise ratios (PSNRs) between the ground truth images with different EVs of the same scene are calculated in Table 2. This difference arises because the information in the amount of light entering the camera gradually changes as the EV is changed. This indicates that the greater the exposure value, the farther away from the input image space, as shown in Fig. 7. Therefore, as the difference in the exposure value increases, it becomes increasingly difficult to restore the detail of the image while changing the brightness. It can be assumed that a relatively deeper neural network is required when inferring the relationship between two such images. Hence, we designed a neural network structure not just simply by increasing the depth structure, but by also proposing a chain structure, which infers the EV sequentially. For example, EV ±3 images can be made after inferring EV ±1 and ±2 images sequentially from an input image (EV 0). To validate the chain structure, we compared the proposed method with a relatively shallow network that is skip-connected between three convolution layers and three deconvolution [36] layers for the problem of inferring the relationship between EV 0 and EV +3. The results are shown in Table 3. The overall result shows that when inferring the relationship between images with a large distance, the deeply structured neural network is better. Therefore, it can be concluded that it is good to have a deeper structure when the exposure difference increases between the images. In addition, as the depth of the structure increases, the gradient vanishing effect, where the error cannot be delivered to the end during backpropagation, may occur. Therefore, a loss term is added to the intermediate results, which correspond to the EV ±1 and ±2 images during the process. As a result, the proposed neural network architecture infers EV ±3 images more accurately.

B. RESIDUAL LEARNING AND ACTIVATION FUNCTION

In the deep neural network structure, neuron activation is determined by a nonlinear function to describe the nonlinear relationship between the input and output. Based on the

TABLE 2. Average PSNR between the middle exposure image and various EV images.

EV -3	EV -2	EV -1	EV 0	EV +1	EV +2	EV +3
8.7	*10.87*	*15.74*	*∞*	*15.88*	*10.40*	*7.62*

TABLE 3. Validity of the chain structure neural network.

Method	PSNR(dB)		SSIM		MS-SSIM	
	m	*σ*	*m*	*σ*	*m*	*σ*
Ours	*28.18*	*2.77*	*0.953*	*0.065*	*0.983*	*0.015*
[36]	*14.52*	*4.83*	*0.913*	*0.077*	*0.966*	*0.048*

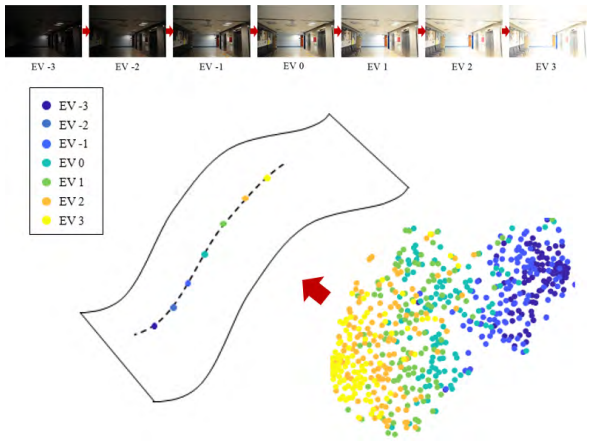

FIGURE 7. Two-dimensional distribution for the image dataset with different exposure values in the image manifold space: for images labeled with the corresponding exposure value, we visualized the image space by two-dimensional reduction using t-distributed stochastic neighbor embedding [50]. Images with different exposure levels for the same scene can be considered to be lying on a low dimensional manifold in high-dimensional image space. when the difference in the exposure value between the images is large, they are far from each other on the manifold.

output of each neuron, the sigmoid and ReLU [37] functions have a non-negative output. In contrast, ELU [38] and SELU [39] have a gradual slope toward negative infinity, and PReLU [32] changes the slope in the negative domain. The proposed neural network consists of a residual learning process that learns the difference between the input and reference. For the residuals of the image with a lower EV than the input LDR image, images with EVs of -1, -2, -3 decrease in pixel value. This means that the weight and bias values become more negative. Because of this perspective, when the EV of the image decreases, it is difficult to find a relation with functions such as the ReLU function. Therefore, the proposed neural network requires an activation function that can reflect a negative value such as PReLU. However, in the negative domain of PReLU, the error does not flow easily owing to gradient variation, and it has the relatively smaller slope than in the positive domain, which flows consistently with the backpropagation.

Accordingly, we propose a new activation function for the HDRI method: MPReLU, which is an extension of the existing PReLU. It is defined as follows.

$$\text{minus PReLU}(x) = \begin{cases} ax, & \text{if } x \geq 0 \\ x, & \text{if } x < 0 \end{cases} \tag{6}$$

TABLE 4. Comparison of MPReLU and PReLU.

		PSNR(dB)		SSIM		MS-SSIM	
		m	*σ*	*m*	*σ*	*m*	*σ*
EV -1	*MPReLU*	***29.01***	*3.83*	***0.953***	*0.065*	***0.980***	*0.017*
	PReLU	*28.28*	*2.96*	*0.931*	*0.053*	*0.977*	*0.015*
EV -2	*MPReLU*	***26.72***	*4.54*	***0.952***	*0.029*	***0.974***	*0.021*
	PReLU	*24.98*	*3.92*	*0.910*	*0.050*	*0.962*	*0.028*
EV -3	*MPReLU*	***24.33***	*4.57*	***0.919***	*0.036*	***0.948***	*0.037*
	PReLU	*22.58*	*4.46*	*0.913*	*0.075*	*0.933*	*0.046*

The comparison results are listed in Table 4. These results show that MPReLU is effective for learning the residuals between a given input image and an image with a darker exposure. Hence, in the proposed neural network architecture, PReLU is used for images with a higher exposure (brighter images), and MPReLU is used for images with a lower exposure (darker images) to train the network.

VI. EXPERIMENTAL RESULTS

The results of the proposed network are divided into three parts: (1) a comparison between the ground truth LDR image stack and the inferred LDR image stack and (2) a comparison between the ground truth HDR images and inferred HDR images. (3) a comparison with state-of-the-art methods. The conventional ITM algorithms of Huo *et al.*'s method [14], Kovaleski and Oliveira's method [15] and Masia *et al.*'s method [17] were used. In addition, we used the state-of-the-art deep learning-based ITM methods of Endo *et al.*'s method [27] and Eilertsen *et al.*'s method [28]. To tone-map the HDR images, representative tone mappers, Reinhard *et al.*'s method [40] and Kim and Kautz's method [41] were used. In the experiments, 48 image stacks among our dataset and 40 image stacks among HDR-eye dataset [42] were used for the comparisons, and the HDR Toolbox [43] and pfstools [44] were used to compare the HDR images with those obtained by the conventional ITM algorithms.

A. COMPARISON BETWEEN THE GROUND TRUTH LDR IMAGE STACK AND INFERRED LDR IMAGE STACK

To determine the similarity between the inferred LDR images and the ground truth LDR images, PSNR, structural similarity (SSIM), and multi-scale SSIM (MS-SSIM) were used. The comparison results are shown in Table 5 and Fig. 8. The similarity between the inferred LDR images and the ground truth images decreased as the exposure difference from EV 0 increased. In addition, the images with a brightness that is darker than that of the single input LDR image were less similar to the ground truth than the images with a brightness that is brighter than the input single LDR image. It is assumed that it is more difficult to infer relative darkness from an input LDR image.

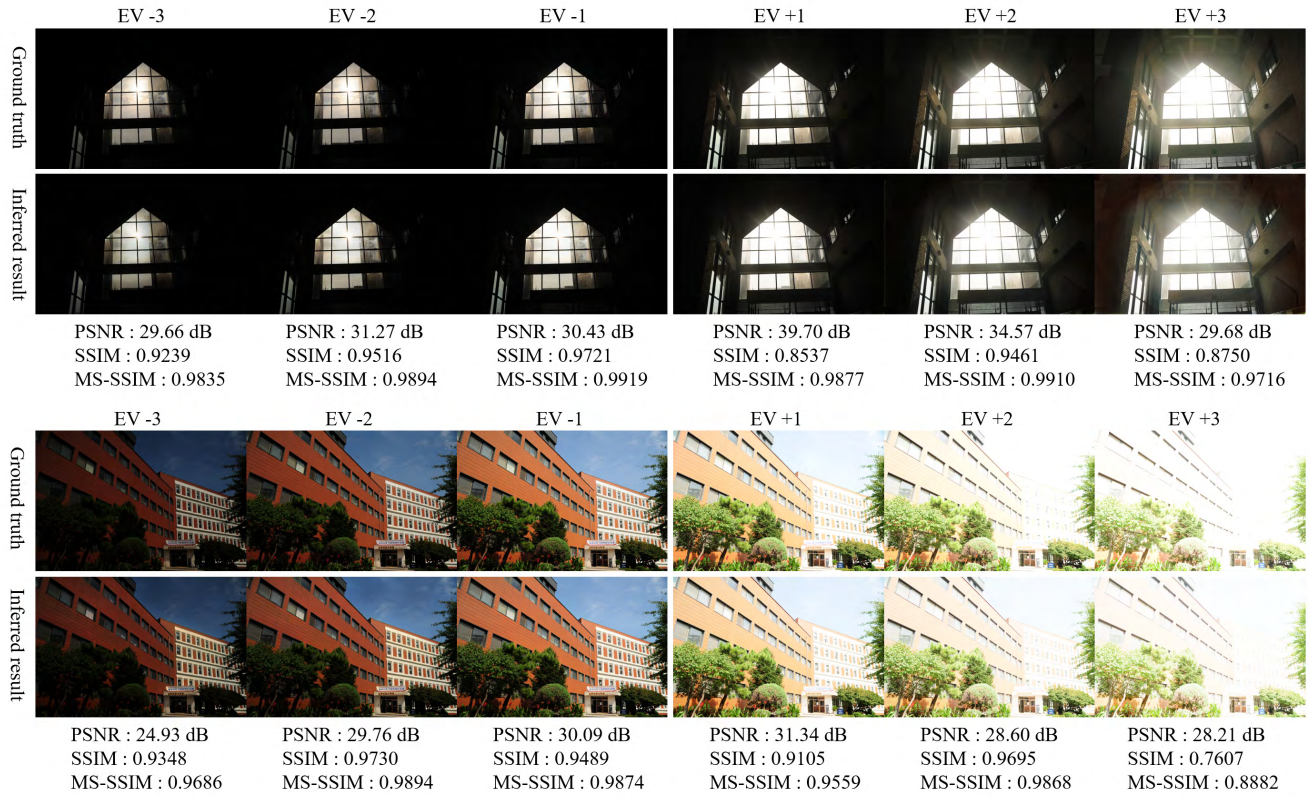

FIGURE 8. Comparison of the ground truth LDR image stack and inferred LDR image stack. The proposed neural network follows not only the actual variation trends of the exposure, but also the properties of the light source.

B. COMPARISON BETWEEN GROUND TRUTH HDR IMAGES AND INFERRED HDR IMAGES

HDR images were generated using the ground truth LDR image stack and the inferred LDR image stack. In the process of HDRI with the LDR image stacks, the CRF was estimated based on Debevec and Malik's method [9]. The PSNR among the tone-mapped HDR images was used for a quantitative comparison of the HDR images. The HDR images were evaluated through HDR-VDP-2 [45] and DRIM [46], which are based on the human visual system. The evaluation used the input parameters of a 24-inch display, 0.5 m viewing distance, 0.0025 peak contrast, and 2.2 gamma. The evaluation results are shown in Table 6 and Fig. 9. The PSNR among the tone mapped HDR images and the VDP-Quality score among the HDR images quantitatively show how much more closely to the ground truth image the proposed method infers than the other methods. In the conventional methods, it is difficult to infer the actual scene luminance because the single LDR image information is simply expanded to fit the target dynamic range. Therefore, there are artifacts such as brightness boosting and loss of details in some areas. In addition, the HDR-VDP-2 and DRIM results also showed that the proposed method approximated the actual scene luminance better than the conventional algorithms. In the result image of HDR-VDP-2, the proposed method had more blue-color pixels than the existing methods. A result image with colors close to 0 (blue) means that the observer cannot recognize the difference from the ground truth HDR image. The DRIM showed the contrast reversal, loss of visible contrast, and amplification of invisible contrast through red, green, and blue points, respectively, to represent the differences between HDR images. Because the proposed method aims to infer the actual scene radiance, it can be seen that the best-inferred result is when there are no red, green, and blue points in the DRIM result image. From this point of view, the proposed neural network inferred the actual scene luminance from a single LDR image better than the conventional methods.

C. COMPARISON WITH STATE-OF-THE-ART METHODS

We additionally provided quantitative and qualitative comparisons for Endo *et al.*'s method [27] and Eilerstsen *et al.*'s method [28], which infer an HDR image through the deep neural network. The results of the quantitative evaluation can be seen in Table 7. In addition, we conducted qualitative comparison on each method to compare the restoration result of the saturated area, and the result is shown in Fig. 10. Eilerstsen *et al.*'s method [28] mainly focused on restoring the saturated region for underexposed LDR inputs, and hence, the performance of the typical images was lower than those of Endo *et al.*'s method [27] and the proposed method. In addition, the proposed method generated a relatively accurate exposure image stack, which shows higher performance than

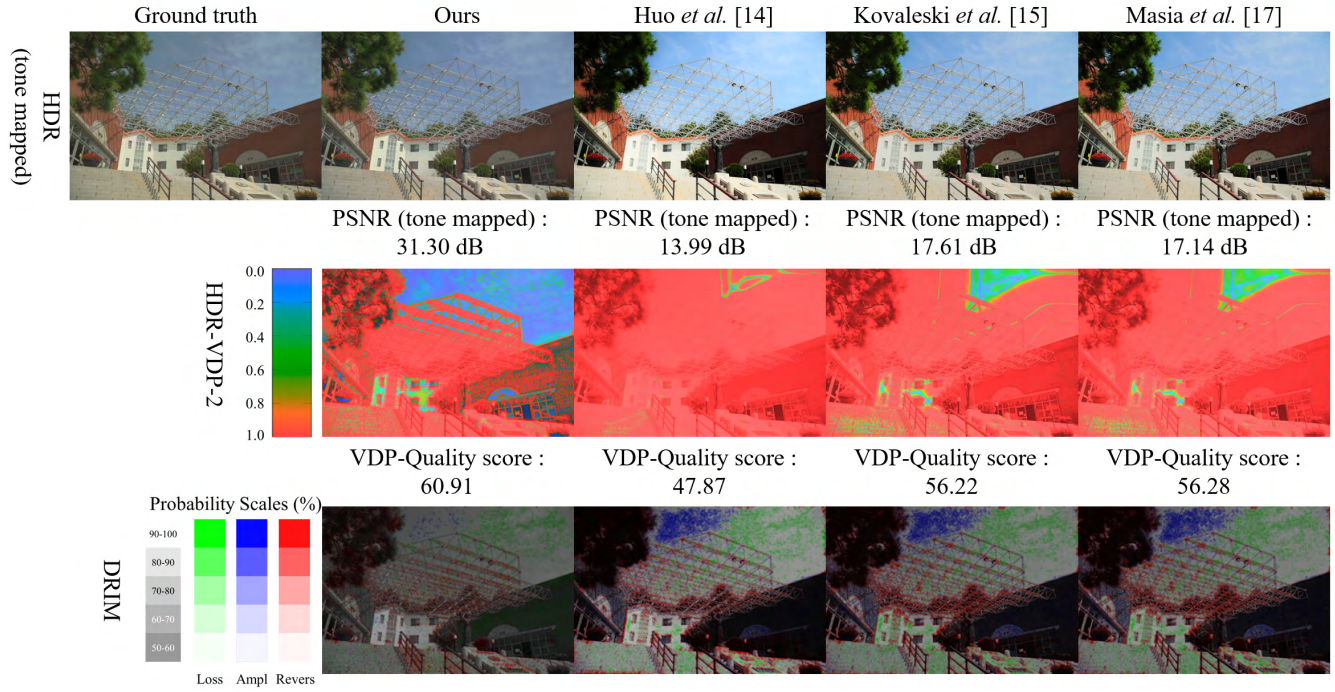

FIGURE 9. Comparison of the ground truth HDR image with HDR images inferred by the proposed and conventional methods.

TABLE 5. Comparison of the ground truth LDR image stack and inferred LDR image stack.

EV	PSNR(dB)		SSIM		MS-SSIM	
	m	σ	*m*	σ	*m*	σ
EV +3	*28.18*	*2.77*	*0.953*	*0.065*	*0.983*	*0.015*
EV +2	*29.65*	*3.06*	*0.959*	*0.065*	*0.986*	*0.016*
EV +1	*31.90*	*3.43*	*0.969*	*0.039*	*0.992*	*0.008*
EV -1	*29.01*	*3.83*	*0.935*	*0.056*	*0.980*	*0.017*
EV -2	*26.72*	*4.54*	*0.952*	*0.029*	*0.974*	*0.021*
EV -3	*24.33*	*4.57*	*0.919*	*0.036*	*0.948*	*0.037*

TABLE 6. Comparison of the ground truth HDR image with HDR images inferred by the proposed and conventional methods.

Method	PSNR(dB) Reinhard's TMO [40]		PSNR(dB) Kim and Kautz's TMO [41]		VDP-Quality score	
	m	σ	*m*	σ	*m*	σ
Ours	***30.86***	*3.36*	***24.54***	*4.01*	***56.36***	*4.41*
[14]	*18.43*	*3.04*	*13.27*	*3.29*	*50.00*	*5.86*
[15]	*21.76*	*2.81*	*14.26*	*2.94*	*50.28*	*4.98*
[17]	*20.13*	*2.21*	*10.74*	*2.16*	*51.24*	*5.67*

TABLE 7. Comparison of the ground truth HDR image with HDR images inferred by [27], [28] and ours.

Dataset	Method	PSNR(dB) Reinhard's TMO [40]		PSNR(dB) Kim and Kautz's TMO [41]		VDP-Quality score	
		m	σ	*m*	σ	*m*	σ
Ours dataset	*Ours*	***30.86***	*3.36*	***24.54***	*4.01*	***56.36***	*4.41*
	[27]	*25.49*	*4.28*	*21.36*	*4.50*	*54.33*	*6.27*
	[28]	*17.97*	*2.17*	*13.16*	*2.72*	*34.25*	*3.37*
HDR-Eye [42]	*Ours*	***25.77***	*2.44*	***22.62***	*3.39*	***49.80***	*5.97*
	[27]	*23.68*	*3.27*	*19.97*	*4.11*	*46.49*	*5.81*
	[28]	*16.36*	*1.35*	*13.41*	*2.17*	*37.08*	*4.62*

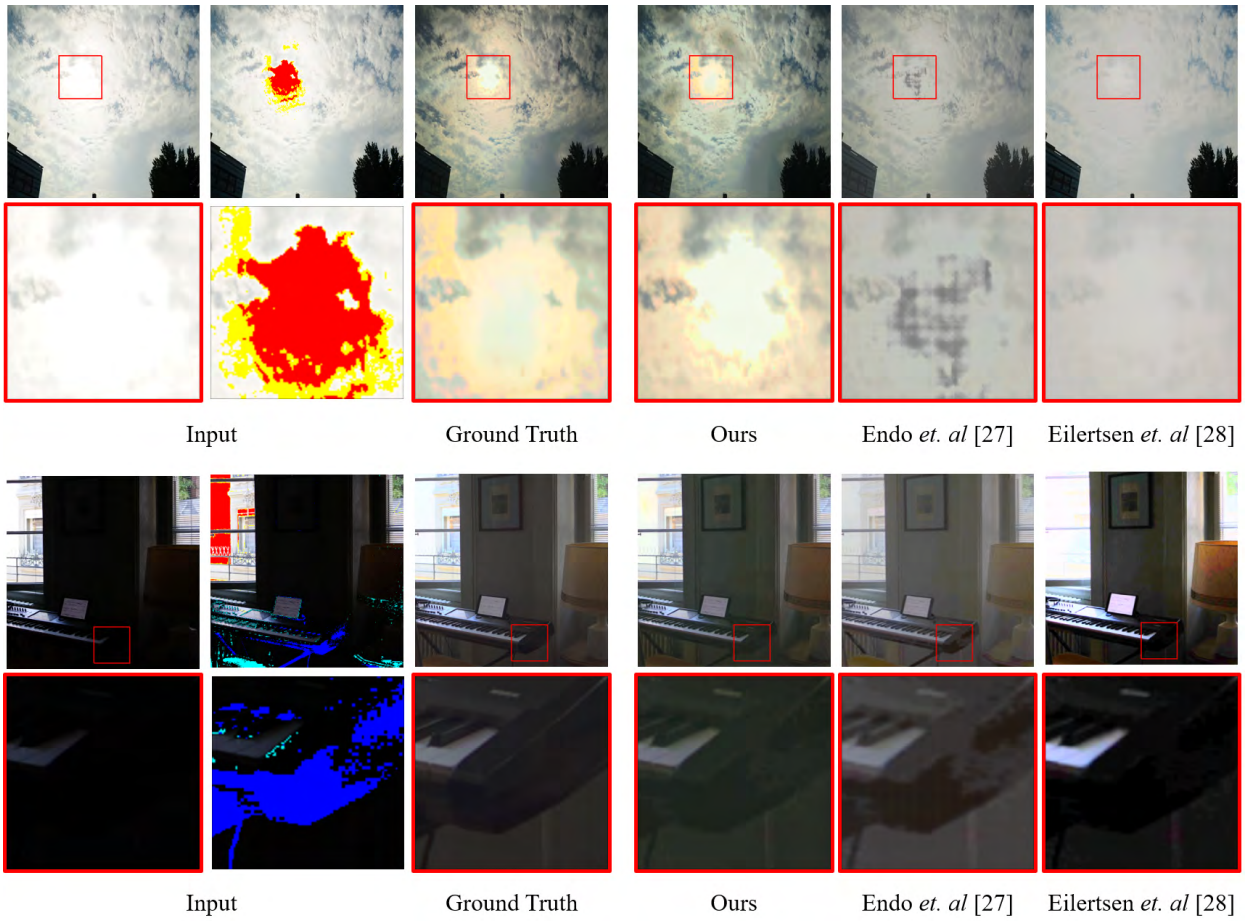

FIGURE 10. Comparison of the ground truth HDR image with HDR images inferred by [27], [28] and the proposed method (ours). To visualize the clipped region in the input images, pixels are painted in red/yellow if they are completely under/overexposed or in blue/cyan if one or two channels are under/overexposed, respectively.

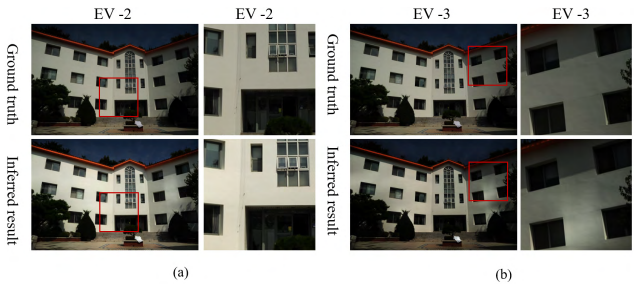

FIGURE 11. Examples of failed cases of the proposed method: (a) The brightness of the outer wall of the building is not estimated correctly, and (b) A patch with nonuniform brightness was created in the plain texture.

the Endo *et al.*'s method [27]. Specifically, the PSNR value and VDP-quality score of the proposed method were higher than those of the conventional deep neural network algorithms as shown in Table 7.

VII. CONCLUSION

In this paper, we proposed a novel artificial neural network structure that infers an HDR image from a single LDR image. By inferring the suitable image luminance for the scene when the given LDR image is captured, the proposed network structure not only widens the dynamic range of the LDR image, but also generates an image that is closer to the ground truth image than the previously proposed methods. Moreover, the proposed network trains the residuals in the image pair, which contains the morphological information and changes in illumination, from the given training set. The proposed subnetwork serves as a dictionary that contains the brightness information for each image with the desired exposure level by rearranging the images sequentially in the exposure space.

As a result, the proposed network is able to solve problems such as ghosting and tearing, which appear in conventional HDRI. Furthermore, the proposed network is scalable in that it can be further extended to obtain a far wider dynamic range. In addition, because patch-based learning has been carried out, it is less restricted to the image resolution for restoring HDR images.

Limitations and Future Work As shown in the experimental results, the proposed network has the limitation in restoring the lost information according to the context. Fig. 11 shows two examples where a bright region could

not be restored with the plain texture. In Fig. 11(a), as the proposed network could not determine whether a given patch is a light source or not, it didn't create a content-dependent patch, which is brighter than the ground truth. In addition, in Fig. 11(b), the proposed method did not generate patches with uniform brightness in the region with the plain texture. To solve these problems, we will further study the network structure by using conditional generative adversarial networks (c-GAN) [51] that can impose additional constraints for HDR reconstruction. It will eliminate unexpected artifacts and enhance the image quality by generating an image that cannot be distinguished from the ground truth by the discriminator.

REFERENCES

[1] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 184–199.

[2] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

[3] G. Huang, Z. Liu, K. Q. Weinberger, and L. van der Maaten. (2016). "Densely connected convolutional networks." [Online]. Available: https://arxiv.org/abs/1608.06993

[4] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1646–1654.

[5] K. Zhang, W. Zuo, S. Gu, and L. Zhang. (2017). "Learning deep CNN denoiser prior for image restoration." [Online]. Available: https://arxiv.org/abs/1704.03264

[6] B. Gu, W. Li, M. Zhu, and M. Wang, "Local edge-preserving multiscale decomposition for high dynamic range image tone mapping," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 70–79, Jan. 2013.

[7] M. Le Pendu, C. Guillemot, and D. Thoreau, "Local inverse tone curve learning for high dynamic range image scalable compression," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5753–5763, Dec. 2015.

[8] Z. Li, Z. Wei, C. Wen, and J. Zheng, "Detail-enhanced multi-scale exposure fusion," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 1243–1252, Mar. 2017.

[9] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Proc. ACM SIGGRAPH Classes*, 2008, p. 31.

[10] 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.

[11] 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.

[12] F. Banterle, P. Ledda, K. Debattista, and A. Chalmers, "Inverse tone mapping," in *Proc. 4th Int. Conf. Comput. Graph. Interact. Techn. Australasia Southeast Asia*, New York, NY, USA, 2006, pp. 349–356.

[13] F. Banterle *et al.*, "High dynamic range imaging and low dynamic range expansion for generating HDR content," *Comput. Graph. Forum*, vol. 28, no. 8, pp. 2343–2367, 2009.

[14] 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.

[15] R. P. Kovaleski and M. M. Oliveira, "High-quality reverse tone mapping for a wide range of exposures," in *Proc. 27th SIBGRAPI Conf. Graph., Patterns Images (SIBGRAPI)*, 2014, pp. 49–56.

[16] 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, p. 160, 2009.

[17] B. Masia, A. Serrano, and D. Gutierrez, "Dynamic range expansion based on image statistics," *Multimedia Tools Appl.*, vol. 76, no. 1, pp. 631–648, 2017.

[18] L. Meylan, S. Daly, and S. Süsstrunk, "The reproduction of specular highlights on high dynamic range displays," in *Proc. Color Imag. Conf.*, 2006, pp. 333–338.

[19] A. G. Rempel *et al.*, "Ldr2Hdr: On-the-fly reverse tone mapping of legacy video and photographs," *ACM Trans. Graph.*, vol. 26, no. 3, p. 39, 2007.

[20] T. H. Wang *et al.*, "Pseudo-multiple-exposure-based tone fusion with local region adjustment," *IEEE Trans. Multimedia*, vol. 17, no. 4, pp. 470–484, Apr. 2015.

[21] S. F. Ray, "Camera exposure determination," in *The Manual of Photography: Photographic and Digital Imaging*. Oxford, U.K.: Focal Press, 2000.

[22] C. Aguerrebere, J. Delon, Y. Gousseau, and P. Musé, "Best algorithms for HDR image generation. A study of performance bounds," *SIAM J. Imag. Sci.*, vol. 7, no. 1, pp. 1–34, 2014.

[23] E. A. Khan, A. O. Akyuz, and E. Reinhard, "Ghost removal in high dynamic range images," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2006, pp. 2005–2008.

[24] T.-H. Min, R.-H. Park, and S. Chang, "Histogram based ghost removal in high dynamic range images," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jun. 2009, pp. 530–533.

[25] P. Sen, N. K. Kalantari, M. Yaesoubi, S. Darabi, D. B. Goldman, and E. Shechtman, "Robust patch-based HDR reconstruction of dynamic scenes," *ACM Trans. Graph.*, vol. 31, no. 6, p. 203, 2012.

[26] A. Srikantha and D. Sidibé, "Ghost detection and removal for high dynamic range images: Recent advances," *Signal Process., Image Commun.*, vol. 27, no. 6, pp. 650–662, 2012.

[27] Y. Endo, Y. Kanamori, and J. Mitani, "Deep reverse tone mapping," *ACM Trans. Graph.*, vol. 36, no. 6, p. 177, 2017.

[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] J. Zhang and J.-F. Lalonde. (2017). "Learning high dynamic range from outdoor panoramas." [Online]. Available: https://arxiv.org/abs/1703.10200

[30] R. Zhang, P. Isola, and A. A. Efros, "Colorful image colorization," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 649–666.

[31] J. Yamanaka, S. Kuwashima, and T. Kurita. (2017). "Fast and accurate image super resolution by deep CNN with skip connection and network in network." [Online]. Available: https://arxiv.org/abs/1707.05425

[32] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1026–1034.

[33] D. Kingma and J. Ba. (2014). "Adam: A method for stochastic optimization." [Online]. Available: https://arxiv.org/abs/1412.6980

[34] *EMPA Media Techonlogy Dataset*. Accessed: Nov. 21, 2017. [Online]. Available: http://empamedia.ethz.ch/hdrdatabase/index.php

[35] *High Dynamic Range Imaging*. Accessed: Sep. 1, 2018. [Online]. Available: http://pages.cs.wisc.edu/~csverma/CS766_09/HDRI/hdr

[36] X.-J. Mao, C. Shen, and Y.-B. Yang. (2016). "Image restoration using convolutional auto-encoders with symmetric skip connections." [Online]. Available: https://arxiv.org/abs/1606.08921

[37] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. ICML*, 2013, p. 30.

[38] D.-A. Clevert, T. Unterthiner, and S. Hochreiter. (2015). "Fast and accurate deep network learning by exponential linear units (ELUs)." [Online]. Available: https://arxiv.org/abs/1511.07289

[39] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter. (2017). "Self-normalizing neural networks." [Online]. Available: https://arxiv.org/abs/1706.02515

[40] 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.

[41] M. H. Kim and J. Kautz, "Consistent tone reproduction," in *Proc. 10th IASTED Int. Conf. Comput. Graph. Imag. (CGIM)*, Innsbruck, Austria, 2008, pp. 152–159.

[42] *HDR-Eye: Dataset of High Dynamic Range Images With Eye Tracking Data*. Accessed: Sep. 1, 2018. [Online]. Available: https://mmspg.epfl.ch/hdr-eye

[43] F. Banterle, A. Artusi, K. Debattista, and A. Chalmers, *Advanced High Dynamic Range Imaging*. Boca Raton, FL, USA: CRC Press, 2017.

[44] R. Mantiuk, G. Krawczyk, R. Mantiuk, and H.-P. Seidel, "High dynamic range imaging pipeline: Perception-motivated representation of visual content," *Proc. SPIE*, vol. 6492, p. 649212, Feb. 2007.

[45] 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.

[46] T. O. Aydin, R. Mantiuk, K. Myszkowski, and H.-P. Seidel, "Dynamic range independent image quality assessment," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 69-1–69-10, Aug. 2008.
[47] M. D. Tocci, C. Kiser, N. Tocci, and P. Sen, "A versatile HDR video production system," *ACM Trans. Graph.*, vol. 30, no. 4, p. 41, 2011.
[48] J. Kronander, S. Gustavson, G. Bonnet, and J. Unger, "Unified HDR reconstruction from raw CFA data," in *Proc. IEEE Int. Conf. Comput. Photograp. (ICCP)*, Apr. 2013, pp. 1–9.
[49] A. Wahab, M. Z. Alam, and B. K. Gunturk, "High dynamic range imaging using a plenoptic camera," in *Proc. 25th Signal Process. Commun. Appl. Conf. (SIU)*, 2017, pp. 1–4.
[50] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.
[51] M. Mirza and S. Osindero. (2014). "Conditional generative adversarial nets." [Online]. Available: https://arxiv.org/abs/1411.1784

SIYEONG LEE received the B.S. degree in computer science and engineering from Sogang University, South Korea, in 2017, where he is currently pursuing the M.S degree in electronic engineering. His current research interests include computer vision, deep learning, and probabilistic graphical model.

GWON HWAN AN received the B.S. degree in electronic engineering from Sogang University, South Korea, in 2017, where he is currently pursuing the M.S. degree in electronic engineering. His current research interests include realistic display application, high-dynamic-range imaging, and deep learning.

SUK-JU KANG (S'08–M'11) received the B.S. degree in electronic engineering from Sogang University, South Korea, in 2006, and the Ph.D. degree in electrical and computer engineering from the Pohang University of Science and Technology, South Korea, in 2011. From 2011 to 2012, he was a Senior Researcher with LG Display, South Korea, where he was a Project Leader for resolution enhancement and multi-view 3-D system projects. From 2012 to 2015, he was an Assistant Professor of electrical engineering with Dong-A University, Busan, South Korea. He is currently an Associate Professor of electronic engineering with the Sogang University, Seoul, South Korea. His current research interests include image analysis and enhancement, video processing, multimedia signal processing, and deep learning.