



# Deep Recursive HDRI: Inverse Tone Mapping Using Generative Adversarial Networks

Siyeong Lee, Gwon Hwan An, and Suk-Ju Kang<sup>(✉)</sup>

Department of Electronic Engineering, Sogang University, Seoul, South Korea  
{siyeong,ghan,sjkang}@sogang.ac.kr

**Abstract.** High dynamic range images contain luminance information of the physical world and provide more realistic experience than conventional low dynamic range images. Because most images have a low dynamic range, recovering the lost dynamic range from a single low dynamic range image is still prevalent. We propose a novel method for restoring the lost dynamic range from a single low dynamic range image through a deep neural network. The proposed method is the first framework to create high dynamic range images based on the estimated multi-exposure stack using the conditional generative adversarial network structure. In this architecture, we train the network by setting an objective function that is a combination of  $L1$  loss and generative adversarial network loss. In addition, this architecture has a simplified structure than the existing networks. In the experimental results, the proposed network generated a multi-exposure stack consisting of realistic images with varying exposure values while avoiding artifacts on public benchmarks, compared with the existing methods. In addition, both the multi-exposure stacks and high dynamic range images estimated by the proposed method are significantly similar to the ground truth than other state-of-the-art algorithms.

**Keywords:** High dynamic range imaging · Inverse tone mapping  
Image restoration · Computational photography  
Generative adversarial network · Deep learning

## 1 Introduction

Most single low dynamic range (LDR) images cannot capture light information for infinite levels owing to physical sensor limitations of a camera. For the too bright or dark area in the image, the boundary with surrounding objects does not appear. However, a high dynamic range (HDR) image containing various brightness information by acquiring and combining LDR images having different

**Electronic supplementary material** The online version of this chapter ([https://doi.org/10.1007/978-3-030-01216-8\\_37](https://doi.org/10.1007/978-3-030-01216-8_37)) contains supplementary material, which is available to authorized users.

exposure levels does not encounter this problem. Owing to this property, interests on HDR imaging have been increasing in various fields. Unfortunately, creating an HDR image from multiple LDR images requires multiple shots, and HDR cameras are still unaffordable. As a result, alternative methods are needed to infer an HDR image from a single LDR image.

Generating an HDR image with only a single LDR image is referred to as an inverse tone mapping problem. This is an ill-posed problem, because a missing signal not appearing in a given image should be restored. Recently, studies have been conducted on an HDR image application using deep learning technique [1–3]. Endo *et al.* [1], Lee *et al.* [2], and Eilertsen *et al.* [3] successfully restored the lost dynamic range using deep learning. However, a disadvantage is that it requires additional training to generate additional LDR images or fails to restore some patterns.

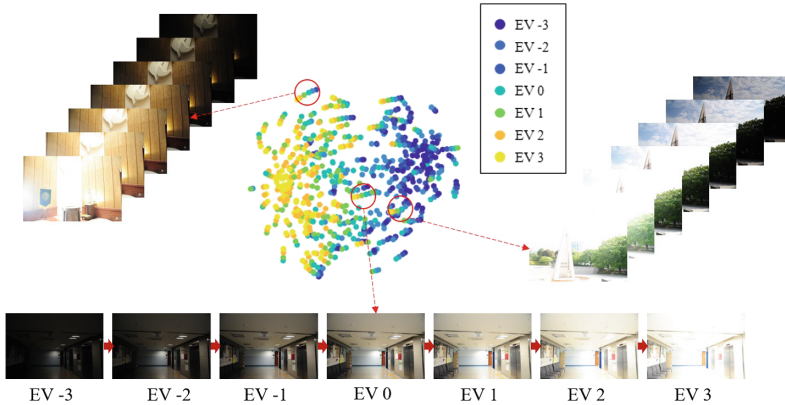
Deep learning is a method of processing information by deriving a function that connects two domains that are difficult to find relation as a function approximator. Deep neural networks demonstrate noteworthy performance for real-world problems (image classification, image restoration, and image generation) that are difficult to be solved by the hand-crafted method. Deep learning, which has emerged in the field of supervised learning that requires labeled data during the learning process, has recently undergone a new turning with the stabilization of the generative adversarial network (GAN) structure [4–8].

We propose a novel method for inverse tone mapping using the GAN structure. This paper has the following three main contributions:

1. The GAN structure creates more realistic images than a network trained with a simple pixel-wise loss function because a discriminator represents a changeable loss that includes the global and local information in the input image during the training process. Thus, we use the structural advantages of the GAN to infer natural HDR images that extend the dynamic range of a given image.
2. We propose a novel network architecture that reconfigures the deep chain HDRI network structure [2], which is a state-of-art method for restoring the lost dynamic range. The reconfigured network can be significantly simplified in scale compared with the existing network, while the performance is maintained.
3. Unlike the conventional deep learning-based inverse tone mapping methods [1, 2] that produce a fixed number of images with different exposure values, we represent the relationship between images with relative exposure values, which has the advantage of generating images with the wider dynamic range without the additional cost.

## 2 Related Works

**Deep Learning-Based Inverse Tone Mapping.** As with other image restoration problems, inverse tone mapping involves the issue of restoring the



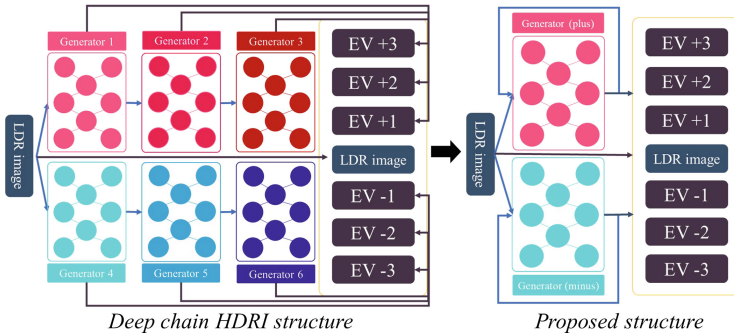
**Fig. 1.** Three-dimensional distribution for the image dataset with different exposure values in the image manifold space: for images labeled with the corresponding exposure value, we visualized the image space by three-dimensional reduction using t-distributed stochastic neighbor embedding [9]. Images having the same scene gradually change in the space. In addition, when the difference in the exposure value between the images is large, they are far from each other on the manifold. (Color figure online)

lost signal information. To solve this problem, the conventional hand-craft algorithms in this field deduce a function to infer the pixel luminance based on the lightness and relations between spatially adjacent pixels of a given image [10, 11], create a pseudo multi-exposure image stack [12], or merge optimally exposed regions of LDR red/green/blue color components for generating an HDR image [13]. By contrast, methods using deep learning [1–3] are included in the example-based learning and successfully applied to restore the lost dynamic range of LDR images. In other words, these types of deep neural networks estimate a function mapping from the pixel brightness to the luminance from a given train set and generate HDR images of given LDR images. Endo *et al.*'s method [1] creates a multi-exposure stack for a given LDR image using a convolutional neural network (CNN) architecture which consists of three-dimensional convolutional layers. Similarly, Lee *et al.*'s method [2] constructs a multi-exposure image stack using a CNN-based network that is designed to generate images through a deeper network structure as the difference in exposure values between the input and the image to be generated increases. By contrast, Eilerstsen *et al.*'s method [3] determines a saturated region using a CNN-based network for an underexposed LDR image and produces the final HDR image by combining the given LDR image and estimated saturated region. These methods require further networks (or parameters) that generate additional images for creating the final HDR image with a wider dynamic range.

**Deep Learning and Adversarial Network Architecture.** Because AlexNet [14] has garnered considerable attention in image classification, deep learning is

used in various fields, such as computer vision and signal processing, to demonstrate significant performance than conventional methods have not reached. For training deep neural networks, techniques such as residual block [15] and skip connection [16] have been introduced. These techniques smooth the weight space and make these networks easy to train [17]. Based on these methods, various structures of neural networks have been proposed. Thus, generating a high-quality image using neural networks in the image restoration is possible.

The GAN structure proposed by Goodfellow *et al.* [4] is a new type of neural network framework that enables highly efficient unsupervised learning than conventional generative models. However, there is a problem that GAN training is unstable. Hence, various types of min-max problems have been proposed for stable training recently: WGAN [18], LSGAN [19], and f-GAN [20]. In addition, by extending the basic GAN structure, recent studies have shown the remarkable success in the image-to-image translation for two different domains [6–8]. Ledig *et al.* [21] proposed a network, SRGAN, capable of recovering the high-frequency detail using the GAN structure and successfully restored the photo-realistic image through this network. Isola *et al.* [6] demonstrated that it can be successful in image-to-image translation using a simple combination of the modified conditional GAN loss [22] and  $L1$  loss.



**Fig. 2.** The structural relationship between a deep chain HDRI [2] and proposed network: the proposed network has a structure of folding sub-networks, which can be interpreted as a structure in which each network shares weight parameters.

### 3 Proposed Method

We first analyze the latest algorithms based on deep learning that focuses on the stack restoration and attempted to determine problems of these algorithms. As a solution, we propose novel neural networks by reconstructing a deep chain HDRI structure [2]. Figure 2 shows the overall structure of the proposed method.

### 3.1 Problems of Previous Stack-Based Inverse Tone Mapping Methods Using Deep Learning

The purpose of the inverse tone mapping algorithm to reconstruct the HDR image from the estimated multi-exposure stack is to generate images with different exposure values. When producing images with different exposure values, previous methods [1, 2] generate LDR images with a uniform exposure differences  $T$  for a given input image (i.e.,  $T = 1$  or  $0.7$ ). In this case, generating  $2M$  images with different exposure values from a given image requires  $2M$  sub-networks, because each sub-network represents the relationship between input images and images with the difference of exposure value  $i \times T$ , for  $i = \pm 1, \pm 2, \dots, \pm M$ . Hence, these methods have the disadvantage that the number of additional networks increases linearly to widen the dynamic range. In addition, different datasets and optimization process are needed to train additional networks. Moreover, these fail to restore some patterns by creating artifacts that do not exist. To solve this problem, we define two neural networks  $G^{plus}$  and  $G^{minus}$  considering the direction of change in the exposure value (plus or minus). In addition, these networks are constrained to generate images considering adjacent pixels using conditional GAN [22]. Then, using these networks, we infer images with relative exposure  $+T$  and  $-T$  for a given image.

### 3.2 Training Process Using an Adversarial Network Architecture

The conditional GAN based architecture that is constrained by input images produces higher-quality images than the basic GAN structure [6]. Therefore, we design the architecture conditioned on the exposure value of the given input using a conditional GAN structure. In other words, to convert to images with a relative exposure value  $+T$  (or  $-T$ ), we define a discriminator network  $D^{plus}$  (or  $D^{minus}$ ) that outputs the probability to determine whether a given pair of images is real or fake.

The proposed architecture determines the optimal solution in the min-max problem of Eqs. (1) and (2):

$$G^{plus}, D^{plus} = \min_G \max_D \{ \mathbb{E}_{I^{EV_{i+1}}, I^{EV_i}} [\log D(I^{EV_{i+1}}, I^{EV_i})] + \mathbb{E}_{I^{EV_i}, z} [1 - \log D(G(I^{EV_i}, z), I^{EV_i})] \}, \quad (1)$$

$$G^{minus}, D^{minus} = \min_G \max_D \{ \mathbb{E}_{I^{EV_{i-1}}, I^{EV_i}} [\log D(I^{EV_{i-1}}, I^{EV_i})] + \mathbb{E}_{I^{EV_i}, z} [1 - \log D(G(I^{EV_i}, z), I^{EV_i})] \}, \quad (2)$$

where  $I^{EV_i}$  is an image with  $EV$   $i$ ,  $z$  is a random noise vector, and  $\mathbb{E}$  is the expectation function. For  $D^{plus}$ , we set the pair  $(I^{EV_{i+1}}, I^{EV_i})$  as a real and the pair  $(G(I^{EV_i}, z), I^{EV_i})$  as a fake.

### 3.3 Structure of the Proposed Neural Network Architecture

We verified the specific network settings of the generator and discriminator through the supplementary document (Fig. 3).

**Generator: U-Net [23] Structure.** We adopt an encoder-decoder model as the generator structure. When the data goes to the next layer, the size of the feature map is reduced by one-half, vertically and horizontally, and conversely doubled. Then, the abstracted feature map is reassembled with the previous feature maps for creating the desired output through a structure that increases the width and height of the feature map. In this structure, we add skip-connections between encoder layers and decoder layers, so that the characteristics of low-level features are reflected in the output. The downsampling block consists of a convolutional layer, one batch normalization layer, and one parametric ReLU (PReLU) [24]. And, the upsampling block contains an upsampling layer, one convolutional layer, one batch normalization layer, and one PReLU. The upsampling layer doubles the feature map size using the nearest-neighbor interpolation. As with the deep chain HDRI, we used PReLU for the network inferring relative  $EV + 1$  and MPReLU [2] for the opposite direction.

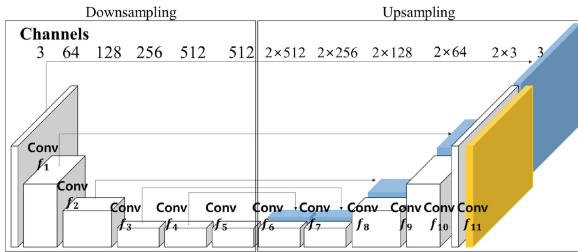


Fig. 3. Structure of proposed generators  $G^{plus}$ ,  $G^{minus}$ .

**Discriminator: Feature Matching.** The neural network of the GAN structure is difficult to train [4, 5, 18–20]. In particular, the problem that the discriminator does not distinguish clearly between the real and fake leads to the difficulty in determining the desired solution in the min-max problem. To solve this problem, we use the method training the generator to match the similarity of features on an intermediate layer of the discriminator in the basic GAN [5]. Therefore, the proposed discriminator is similar to the Markovian discriminator structure [6, 25]. This discriminator generates feature maps that consider the neighboring pixels in an input through convolutional layers. Hence, this network outputs the probability whether each patch in an input image is real or not. Unlike pixel-wise loss, the loss function expressed by the discriminator network represents the structured loss such as the structural similarity, feature matching, and conditional random field [26]. In other words, the loss produced by this

discriminator allowed the generator to create natural images that reflect in the relationship between adjacent pixels. The proposed discriminator is composed of convolution blocks, including one convolution layer, one batch normalization layer, and one leaky ReLU layer [27]. The activation function of the last convolution block is a sigmoid function. In addition, there is no batch normalization layer for the first and last layers (Fig. 4).

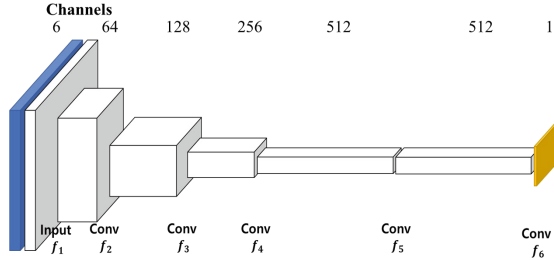


Fig. 4. Structure of proposed discriminators  $D^{plus}$ ,  $D^{minus}$ .

### 3.4 Loss Functions

For  $G^{plus}$  and  $G^{minus}$ , we set an objective function that combined the following two losses for the training. We set the relative weights of each loss to  $\lambda = 100$  through the experimental procedure. the final objective is:

$$G^{plus} = \arg \min_G L_{LSGAN}(G) + \lambda L_{L1}(G) \text{ for training pairs } (I^{EV1}, I) \text{ and } \quad (3)$$

$$G^{minus} = \arg \min_G L_{LSGAN}(G) + \lambda L_{L1}(G) \text{ for training pairs } (I^{EV-1}, I), \quad (4)$$

where  $I$  is an input image,  $I^{EV1}$  (or  $I^{EV-1}$ ) is an image with the relative exposure difference 1 (or  $-1$ ) for a given  $I$ .

**GAN Loss.** As the basic GAN structure [4] is unstable in the training process, we use LSGAN [19] to determine the optimal solution of the min-max problem. For an input image  $x$ , a reference image  $y$ , and random noise  $z$ ,

$$L_{LSGAN}(D) = \frac{1}{2} \mathbb{E}_{x,y} [(D(y, x) - 1)^2] + \frac{1}{2} \mathbb{E}_{x,z} [(D(G(x, z), x))^2], \quad (5)$$

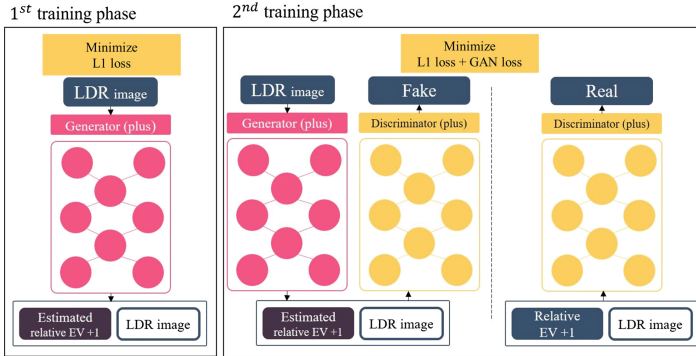
$$L_{LSGAN}(G) = \mathbb{E}_{x,z} [(D(G(x, z), x) - 1)^2], \quad (6)$$

where  $G$  and  $D$  are training networks. We divide the loss of the discriminator by half compared with the generator process to make the overall learning stable by delaying the training of the discriminator.

**Content Loss.** The pixel-wise mean absolute error (MAE) loss  $L_{L1}$  is defined as:

$$L_{L1}(G) = \mathbb{E}_{x,y,z} [||y - G(x, z)||_1]. \tag{7}$$

A method to calculate the pixel-wise difference between two images through  $L2$  norm generates a blurred image relative to  $L1$  norm for image restoration [28]. Therefore, we use  $L1$  loss as a term of the objective function to recover low-frequency components.



**Fig. 5.** The training process of proposed network architecture: we trained the generators to minimize  $L1$  loss and defeat discriminator networks. The discriminator distinguishes the pair (reference, input) from the pair (estimated image, input) as the training progresses.

### 3.5 Optimization Process

The proposed architecture is trained through two steps, as shown in Fig. 5. In the first training phase, we used only  $L1$  loss, and in the second training phase, we additionally used GAN loss. We set the two training phases epoch with the same ratio (1:1). In the second training phase, the discriminator and generator alternated one by one to minimize each objective function. We used the Adam optimizer [29] with 0.00005 of the learning rate, and momentum parameters were  $\beta_1 = 0.5$  and  $\beta_2 = 0.999$ . We set the batch size to one. The dropout noise is added during training.

### 3.6 Inference

First, we generated images  $\hat{I}^{EV1}$  and  $\hat{I}^{EV-1}$  from the given LDR image, as shown in Fig. 6, using  $G^{plus}$ ,  $G^{minus}$ . In the next phase, we obtained  $\hat{I}^{EV2}$ ,  $\hat{I}^{EV-2}$  by using  $\hat{I}^{EV1}$  and  $\hat{I}^{EV-1}$  as the input of  $G^{plus}$  and  $G^{minus}$ , respectively. We recursively repeated this process for creating a multi-exposure stack. Figure 6 shows an example of outputting the multi-exposure stack up to  $EV \pm 3$ .



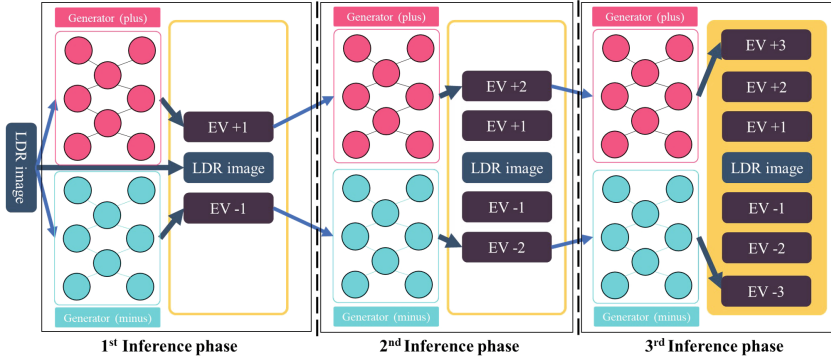


Fig. 6. The multi-exposure stack generation process of the proposed structure.

### 4 Experimental Results

For a dataset, we used 48 stacks of VDS dataset [2] for training, and other 48 stacks of VDS dataset and 41 stacks of HDREye dataset [30] for testing. VDS database is composed of images taken with Nikon 7000, and HDREye consists of images taken with Sony DSC-RX100 II, Sony NEX-5N, and Sony  $\alpha$ 6000. Both the VDS and HDREye datasets consists of seven images, each of which has uniformly different exposure levels. We set the unit exposure value  $T$  to exposure value one at  $ISO - 100$  like the deep chain HDR [2]. By using Debevec *et al.*'s algorithm [31], we synthesized the generated stack with a target HDR image, and we generated the tone-mapped images by using Reinhard *et al.*'s [32] and Kim and Kautz's methods [33] through HDR Toolbox [34]. For the image pair with the exposure value difference, we set the image with low exposure value as an input image and set the other image as a reference when training  $G^{plus}$ . ( $G^{minus}$  was done in the opposite way.) We randomly cropped the sub-images with the  $256 \times 256$  pixel resolution from the training set, which contained adequate information about the entire image rather than patches, thereby providing 20,700 training pairs. We set epochs of the first and second phases to 10 for training. First, to verify that the images were generated successfully, we compared them with the ground truths through the peak signal-to-noise ratios (PSNR), structural similarity (SSIM), and multi-scale SSIM (MS-SSIM) on test images with  $512 \times 512$  pixel resolution. Second, we compared our method with the state-of-the-art algorithms using deep learning [1-3]. Finally, we confirmed the performance of the proposed method by testing the different loss functions with two cases:  $L1$  loss and  $L1 + GAN$  Loss.

#### 4.1 Comparison Between the Ground Truth LDR and Inferred LDR Image Stacks

Table 1 and Fig. 7 show the several results and comparisons between estimated and ground truth stacks. In addition, we compared it to the deep chain HDR

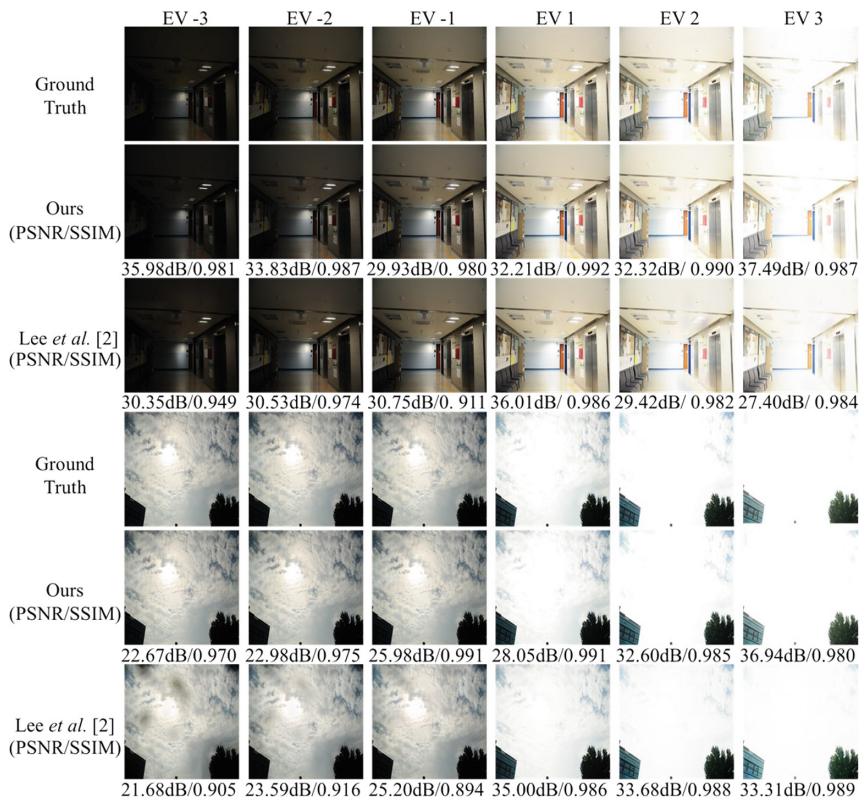
method [2] that estimated a stack with the same unit exposure value  $T = 1$ . In the proposed method, the similarity between the inferred LDR and reference images was reduced as the difference of exposure value increased. This is because the artifacts were amplified as the input image passed recursively through the network to generate an image with the high exposure value. However, the proposed method used the GAN structure, where the discriminator evaluated the image quality by considering adjacent pixels, and generated inferred images, thereby increasing the similarity with the ground truth compared with the deep chain HDR method.

**Table 1.** Comparison of the ground truth LDR and inferred LDR image stacks.

		PSNR(dB)		SSIM		MS-SSIM	
		$m$	$\sigma$	$m$	$\sigma$	$m$	$\sigma$
$EV +3$	Proposed	28.97	2.92	0.944	0.044	0.981	0.014
	[2]	28.18	2.77	0.953	0.065	0.983	0.015
$EV +2$	Proposed	29.43	2.85	0.952	0.039	0.986	0.010
	[2]	29.65	3.06	0.959	0.065	0.986	0.016
$EV +1$	Proposed	32.02	2.85	0.969	0.026	0.992	0.006
	[2]	31.90	3.43	0.969	0.039	0.992	0.008
$EV -1$	Proposed	31.22	3.69	0.951	0.031	0.986	0.09
	[2]	29.01	3.83	0.935	0.056	0.980	0.017
$EV -2$	Proposed	31.08	3.07	0.948	0.041	0.986	0.014
	[2]	26.72	4.54	0.952	0.029	0.974	0.021
$EV -3$	Proposed	29.15	4.75	0.910	0.061	0.966	0.025
	[2]	24.33	4.57	0.919	0.036	0.948	0.037

## 4.2 Comparisons with State-of-the-art Methods

For quantitative comparisons with the state-of-the-art methods, we compared PSNR, SSIM, and MS-SSIM with the ground truth for tone-mapped HDR images. Also, we used HDR-VDP-2 [35] based on the human visual system for evaluating the estimated HDR images. We set the input parameters of HDR-VDP-2 evaluation as follows: a 24-inch display, a viewing distance of 0.5 m, peak contrast of 0.0025, and gamma of 2.2. To establish a baseline, we reported the comparison with HDR images inferred by Masia *et al.*'s method [36] using the exponential expansion. Table 2 and Fig. 8 show the evaluation results. In addition, to verify the physics-based reconstruction, we performed to convert an LDR image of a color-checker into an HDR image. LDR and HDR image pairs including a color checker board [30] were used in the experiment. The results of the verification are shown in Fig. 9.



**Fig. 7.** Comparison of the ground truth LDR and inferred LDR image stacks.

The proposed method exhibited similar performance to the deep chain HDR [2]. Moreover, the average PSNR of the tone-mapped images was 3 dB higher than that of Endo *et al.* [1], and the average of 10 dB was higher than Eilertsen *et al.* [3]. For HDREye dataset, which consists of images with different characteristics from the training set, the proposed method was almost better than other methods [1–3] in the HDR VDP Q-score. The reconstructed images of the proposed method were more similar to the ground truth than others in the overall tone and average brightness, as shown in Fig. 8. In addition, the dark and saturated regions of the input image were restored.

### 4.3 Comparison of the Different Loss Functions

To evaluate the effect of the GAN loss term, we compared images generated by the proposed method with training results using only  $L1$  loss. When using only the  $L1$  loss, we trained the network for 20 epochs. Table 3 presents the results of the quantitative comparison. For tone-mapped images by Reinhard’s TMO [32], the average PSNR of the proposed method with  $L1 + \text{GAN}$  was

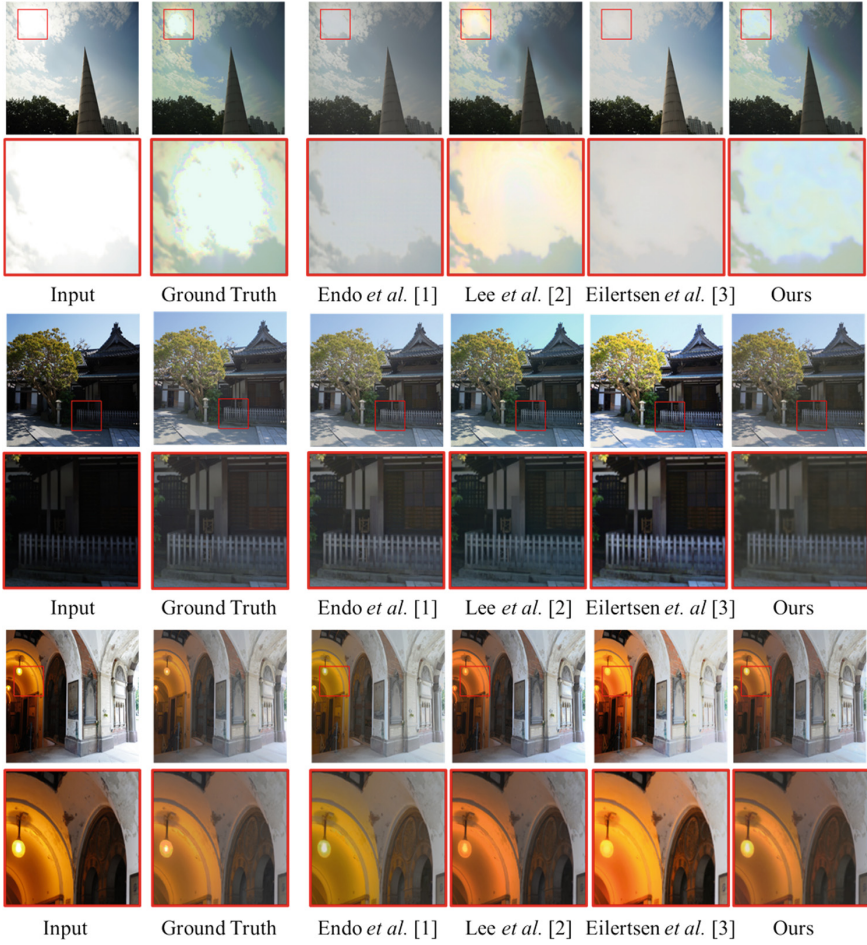


Fig. 8. Comparison of the ground truth HDR images with HDR images inferred by [1–3], and the proposed method (ours).

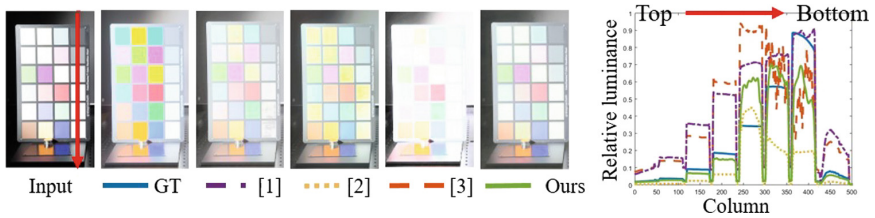


Fig. 9. Comparison of ground truth HDR with HDR images inferred by [1–3], and the proposed method (ours) about physical luminance.

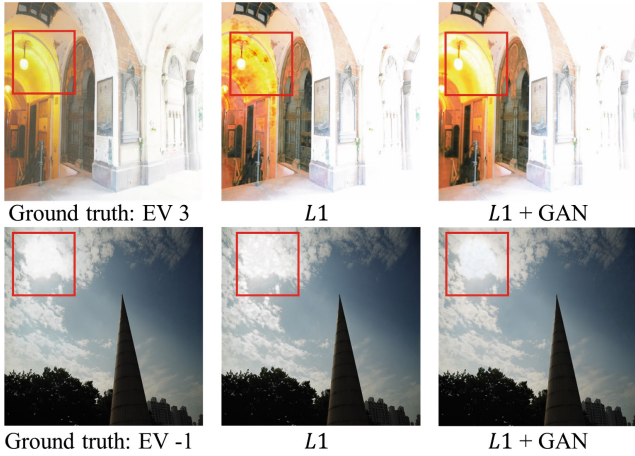
**Table 2.** Comparison of the ground truth HDR images with HDR images inferred by [1–3, 36] and ours. *Red* color indicates the best performance and *blue* color indicates the second best performance.

		PSNR (dB)		PSNR (dB)		VDP quality	
		Reinhard’s TMO		Kim and Kautz’s TMO		Score	
		$m$	$\sigma$	$m$	$\sigma$	$m$	$\sigma$
VDS	<i>Proposed</i>	<b>32.14</b>	3.53	<b>25.36</b>	4.11	<b>57.28</b>	5.17
	[1]	25.49	4.28	21.36	4.50	54.33	6.27
	[2]	<b>30.86</b>	3.36	<b>24.54</b>	4.01	<b>56.36</b>	4.41
	[3]	17.97	2.17	13.16	2.72	34.25	3.37
	[36]	20.13	2.21	10.74	2.16	51.24	5.67
HDREye	<i>Proposed</i>	<b>26.71</b>	2.78	<b>22.31</b>	3.20	<b>48.85</b>	4.91
	[1]	23.68	3.27	19.97	13.41	46.49	5.81
	[2]	<b>25.77</b>	2.44	<b>22.62</b>	3.39	<b>49.80</b>	5.97
	[3]	16.36	1.35	13.41	4.12	37.08	4.62
	[36]	17.18	1.89	9.89	1.94	45.74	5.69

2.27 dB higher than the other. For images generated by Kim and Kautz’s TMO [33], the proposed method had an average PSNR of 1.29 dB higher. Figure 10 shows the tone-mapped HDR images generated by the proposed method using the Reinhard’s TMO. The network trained by setting  $L1$  loss as an objective function generated images that prominently contained artifacts. By contrast, the network architecture with GAN loss did not generate it.

**Table 3.** Average values of image quality metrics PSNR and VDP quality score on the testing dataset for different cost functions.

		PSNR (dB)		PSNR (dB)		VDP-quality	
		Reinhard’s TMO		Kim and Kautz’s TMO		Score	
		$m$	$\sigma$	$m$	$\sigma$	$m$	$\sigma$
VDS	$L1$	28.12	2.51	23.41	3.67	55.78	4.87
	$L1+GAN$	32.14	3.53	25.36	4.11	57.28	5.17
HDREye	$L1$	26.19	2.14	21.69	3.20	49.00	5.19
	$L1+GAN$	26.71	2.78	22.31	3.20	48.85	4.91



**Fig. 10.** Comparison of the ground truth HDR images with HDR images inferred by  $L1$  and  $L1 + GAN$ . The proposed method generates fewer artifacts in the image than the network with  $L1$ .

## 5 Conclusion

We proposed the deep neural network architecture based on the GAN architecture to solve the inverse tone mapping problem, reconstructing missing signals from a single LDR image. Moreover, we trained this CNN-based neural network to infer the relation between relative exposure values using a conditional GAN structure. Therefore, the proposed method generated an HDR image recovered in a saturated (or dark) region of a given LDR image. This network differed from existing networks [1,2], in that it converted an LDR image into a non-linear LDR image corresponding to +1 or -1 exposure stops. This property led the architecture to generate images with varying exposure levels without additional networks and training process. In addition, we constructed a relatively simple network structure by changing the deep structure effect of deep chain HDRI into a recursive structure.

**Acknowledgements.** This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2018R1D1A1B07048421) and Korea Electric Power Corporation. (Grant number R17XA05-28). We thank Yong Deok Ahn and members of the Sogang Vision and Display Lab. for helpful discussions.

## References

1. Endo, Y., Kanamori, Y., Mitani, J.: Deep reverse tone mapping. *ACM Trans. Graph. (TOG)* **36**(6), 177 (2017)



2. Lee, S., An, G.H., Kang, S.J.: Deep chain HDRI: reconstructing a high dynamic range image from a single low dynamic range image. arXiv preprint [arXiv:1801.06277](https://arxiv.org/abs/1801.06277) (2018)
3. Eilertsen, G., Kronander, J., Denes, G., Mantiuk, R.K., Unger, J.: HDR image reconstruction from a single exposure using deep CNNs. *ACM Trans. Graph. (TOG)* **36**(6), 178 (2017)
4. Goodfellow, I., et al.: Generative adversarial nets. In: *Advances in Neural Information Processing Systems*, pp. 2672–2680 (2014)
5. Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., Chen, X.: Improved techniques for training GANs. In: *Advances in Neural Information Processing Systems*, pp. 2234–2242 (2016)
6. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks
7. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. arXiv preprint (2017)
8. Kim, T., Cha, M., Kim, H., Lee, J.K., Kim, J.: Learning to discover cross-domain relations with generative adversarial networks. arXiv preprint [arXiv:1703.05192](https://arxiv.org/abs/1703.05192) (2017)
9. Maaten, L.V.D., Hinton, G.: Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**(Nov), 2579–2605 (2008)
10. Rempel, A.G.: Ldr2Hdr: on-the-fly reverse tone mapping of legacy video and photographs. *ACM Trans. Graph. (TOG)* **26**, 39 (2007)
11. Meylan, L., Daly, S., Süsstrunk, S.: The reproduction of specular highlights on high dynamic range displays. In: *Color and Imaging Conference*, vol. 2006, pp. 333–338. Society for Imaging Science and Technology (2006)
12. Wang, T.H., et al.: Pseudo-multiple-exposure-based tone fusion with local region adjustment. *IEEE Trans. Multimed.* **17**(4), 470–484 (2015)
13. Hirakawa, K., Simon, P.M.: Single-shot high dynamic range imaging with conventional camera hardware. In: *2011 IEEE International Conference on Computer Vision (ICCV)*, pp. 1339–1346. IEEE (2011)
14. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems*, pp. 1097–1105 (2012)
15. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778 (2016)
16. Mao, X.J., Shen, C., Yang, Y.B.: Image restoration using convolutional auto-encoders with symmetric skip connections. arXiv preprint [arXiv:1606.08921](https://arxiv.org/abs/1606.08921) (2016)
17. Li, H., Xu, Z., Taylor, G., Goldstein, T.: Visualizing the loss landscape of neural nets. arXiv preprint [arXiv:1712.09913](https://arxiv.org/abs/1712.09913) (2017)
18. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein GAN. arXiv preprint [arXiv:1701.07875](https://arxiv.org/abs/1701.07875) (2017)
19. Mao, X., Li, Q., Xie, H., Lau, R.Y., Wang, Z., Smolley, S.P.: Least squares generative adversarial networks. In: *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2813–2821. IEEE (2017)
20. Nowozin, S., Cseke, B., Tomioka, R.: f-GAN: training generative neural samplers using variational divergence minimization. In: *Advances in Neural Information Processing Systems*, pp. 271–279 (2016)
21. Ledig, C., et al.: Photo-realistic single image super-resolution using a generative adversarial network. In: *CVPR*, vol. 2, p. 4 (2017)

22. Mirza, M., Osindero, S.: Conditional generative adversarial nets. arXiv preprint [arXiv:1411.1784](https://arxiv.org/abs/1411.1784) (2014)
23. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
24. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1026–1034 (2015)
25. Li, C., Wand, M.: Precomputed real-time texture synthesis with markovian generative adversarial networks. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9907, pp. 702–716. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-46487-9\\_43](https://doi.org/10.1007/978-3-319-46487-9_43)
26. Lafferty, J., McCallum, A., Pereira, F.C.: Conditional random fields: probabilistic models for segmenting and labeling sequence data (2001)
27. Maas, A.L., Hannun, A.Y., Ng, A.Y.: Rectifier nonlinearities improve neural network acoustic models
28. Zhao, H., Gallo, O., Frosio, I., Kautz, J.: Loss functions for image restoration with neural networks. *IEEE Trans. Comput. Imaging* **3**(1), 47–57 (2017)
29. Kingma, D., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980) (2014)
30. Nemoto, H., Korshunov, P., Hanhart, P., Ebrahimi, T.: Visual attention in LDR and HDR images. In: 9th International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM), Number EPFL-CONF-203873 (2015)
31. Debevec, P.E., Malik, J.: Recovering high dynamic range radiance maps from photographs. In: ACM SIGGRAPH 2008 classes, p. 31. ACM (2008)
32. Reinhard, E., Stark, M., Shirley, P., Ferwerda, J.: Photographic tone reproduction for digital images. *ACM Trans. Graph. (TOG)* **21**(3), 267–276 (2002)
33. Kim, M.H., Kautz, J.: Consistent tone reproduction. In: Proceedings of the Tenth IASTED International Conference on Computer Graphics and Imaging (CGIM 2008), Innsbruck, Austria, pp. 152–159. IASTED/ACTA Press (2008)
34. Banterle, F., Artusi, A., Debattista, K., Chalmers, A.: Advanced High Dynamic Range Imaging. CRC Press, Boca Raton (2017)
35. Mantiuk, R., Kim, K.J., Rempel, A.G., Heidrich, W.: HDR-VDP-2: a calibrated visual metric for visibility and quality predictions in all luminance conditions. *ACM Trans. Graph. (TOG)* **30**, 40 (2011)
36. Masia, B., Agustin, S., Fleming, R.W., Sorkine, O., Gutierrez, D.: Evaluation of reverse tone mapping through varying exposure conditions. *ACM Trans. Graph. (TOG)* **28**(5), 160 (2009)