https://doi.org/10.2352/CIC.2025.33.1.39

HDR Image Reconstruction from Saturated LDR images of Dielectric Objects

Shoji Tominaga; Norwegian University of Science and Technology, Gjøvik, Norway / Nagano University, Ueda, Japan
Takahiko Horiuchi; Chiba University, Chiba, Japan

Abstract

This paper presents a method for reconstructing the original high dynamic range (HDR) image from a saturated low dynamic range (LDR) image with missing physical information, specifically for single dielectric objects. A deep neural network approach is employed to map an 8-bit LDR image directly to its corresponding HDR representation. We begin by analyzing the reflection and saturation characteristics of dielectric materials and then construct an HDR image database using a diverse set of dielectric objects. Each HDR image is clipped to generate a set of 8-bit LDR images. All HDR-LDR image pairs are normalized to a fixed resolution and used for training and validation. A deep convolutional neural network (CNN) is designed in the form of an autoencoder architecture with skip connections. The entire network is implemented using MATLAB's machine learning toolbox, with the ADAM optimizer employed for training. The performance of the proposed method is evaluated using a separate validation set. Comparative experiments with existing methods demonstrate that our approach achieves significantly higher reconstruction accuracy and better histogram fitting.

Introduction

Digital camera sensors have limited dynamic ranges, meaning that only a restricted range of luminance levels can be captured when imaging real-world scenes. Additionally, since most image display devices used in daily life support only 8-bit output, the majority of existing image content is in 8-bit low dynamic range (LDR) format. Most legacy image archives are similarly stored in 8-bit LDR format. For example, the Flickr Material Database (see [1]-[2]), a widely recognized dataset used for studying human material categorization, consists entirely of 8-bit LDR images.

Images of scenes with brightness levels exceeding the dynamic range of the camera sensor become saturated and suffer from clipped whites. This commonly occurs when capturing objects with strong gloss or specular highlights. Ideally, the original scene should be captured as a high dynamic range (HDR) image by taking multiple photos at different exposure levels and merging them into a single image. However, in the LDR case, pixel values exceeding the sensor's dynamic range are clipped, resulting in the loss of physical information in saturated regions.

This study addresses the challenge of reconstructing the original HDR image containing missing physical information from a single saturated LDR image of a single-material object. Metallic objects are typical examples of materials prone to saturation, as the luminance of reflected light spans a wide dynamic range—from diffuse matte reflections to intense specular highlights. Consequently, large portions of metallic object surfaces often appear saturated. The task of estimating the original HDR image from a saturated LDR image is commonly referred to as the inverse tone mapping problem [3]. This is an ill-posed problem because it involves restoring missing signal information that is not present in the observed LDR image [4]. The problem has been extensively addressed in the field of computer graphics [5]–[9], and to a lesser extent in computer vision [4], [10]. Many proposed methods rely on convolutional neural networks (CNNs), while others use perceptual-based techniques, such as an extended gamma transformation [11]. In a previous study [12], we proposed a deep neural network approach for reconstructing HDR images from saturated LDR inputs specifically for metallic objects, demonstrating superior reconstruction accuracy compared to existing methods.

In contrast, the present study focuses on reconstructing HDR images from single LDR images of dielectric objects. Dielectric materials comprise many everyday objects, including plastics, paints, ceramics, vinyls, tiles, fruits, leaves, and wood (see [13]). Unlike metallic objects, which are limited to the inherent colors of metals such as copper, gold, silver, aluminum, and iron, dielectric materials exhibit a broad range of vivid colors. Light reflection from homogeneous materials like metals arises mainly from interface reflection at the air-metal boundary, resulting in a monochromatic appearance. In contrast, reflection from inhomogeneous dielectric materials is typically dichromatic, comprising both specular (interface) and diffuse (body) reflection components. The surfaces of dielectric objects often produce strong specular highlights that can saturate the dynamic range. Notably, due to their dichromatic nature, both highlight and body color regions may be subject to saturation.

This paper proceeds as follows: First, we describe the reflection and saturation characteristics of dielectric objects. Second, we present the construction of an HDR image database using a large number of dielectric objects with varied shapes and material types. Third, we propose a deep neural network architecture for direct mapping from an 8-bit LDR image to its HDR counterpart. This network is designed in the form of a deep auto-encoder using a CNN. Finally, we evaluate the performance of the proposed method through validation tests and compare the results with existing approaches.

Reflection and Saturation Characteristics for Dielectric Objects

Light reflection at the surface layers of dielectric objects is modeled using inhomogeneous materials, as illustrated in Figure 1 (see [14]). The reflection is generally decomposed into two components: specular reflection and diffuse reflection. Specular reflection occurs at the interface between the object's surface and the surrounding air. This component typically retains the color of the illumination light source. In contrast, diffuse reflection results from light scattering within the subsurface pigment layer. Consequently, the observed light from the surface is a linear combination of these two reflection components.

The spatial distributions of these two components differ significantly. Light reflected through specular reflection is

confined to a narrow angular range, similar to mirror reflection. Conversely, light scattered by diffuse reflection is distributed uniformly in all directions.

Figure 2 illustrates an example HDR image obtained by capturing a plastic ball using a mobile phone camera. Figure 2(A) shows the original HDR image. The body color resulting from diffuse reflection appears yellow, while the highlight from specular reflection appears white due to the application of white balance. Figure 2(B) presents the corresponding pixel distribution in the (R, G) plane of the RGB color space. The distribution exhibits two distinct linear clusters: one corresponding to the specular highlight and the other to the diffuse body color. The image clearly extends beyond the 8-bit (0–255) range, confirming its HDR nature.

Figure 3 presents the corresponding LDR image obtained by clipping the HDR image from Figure 2(A) into the 8-bit range. In Figure 3(A), if any of the RGB values exceeds 255, the pixel is considered saturated and is highlighted in red. A substantial portion of the plastic object's surface is visibly saturated. Figure 3(B) shows the pixel distribution in the 8-bit RGB space. Comparing Figure 3(A) with Figure 2(A), it is evident that saturation due to clipping occurs not only in the specular highlight regions but also in areas representing the body color caused by diffuse reflection.

Figure 4 shows the result of applying the previously proposed method [12], originally developed for metallic objects, to reconstruct the original HDR image from the saturated LDR image. Compared to Figure 3(A), it is clear that the reconstruction of the saturated regions was unsuccessful. The color and shading information in these areas are inaccurate. Therefore, applying a method developed for metallic objects fails to achieve accurate appearance reconstruction for dielectric objects.

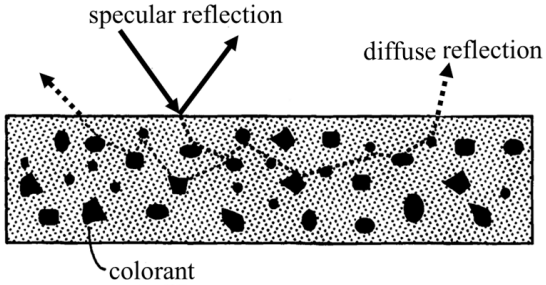
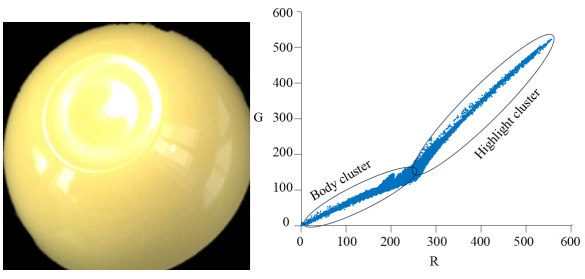
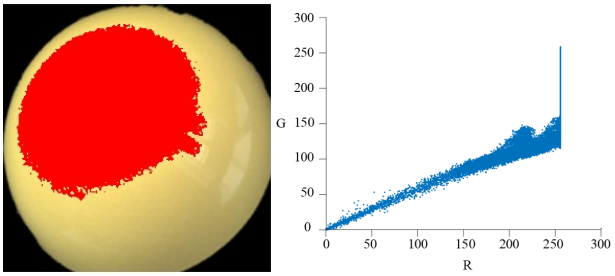


Figure 1 Reflection model in an inhomogeneous dielectric material.



(A) (B)

Figure 2 Example of HDR images by capturing a plastic ball: (A) Original image. (B) Pixel distribution on (R, G) plane in the RGB color space. Note that the image in Figure 2 (A) is not clipped.



(A) (B)

Figure 3 LDR image obtained from the HDR image in Figure 2 (A): (A) LDR image clipped in the 8-bit range, where the saturated area is pained in red. (B) Pixel distribution in the 8-bit range.

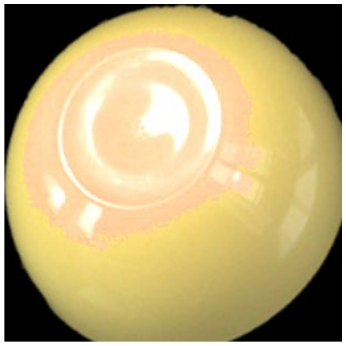


Figure 4 Resulting image by applying the previous method [12] developed for metallic objects to reconstruct the original HDR image from the LDR image suffering from saturation.

HDR Image Database for Dielectric Objects

A large number of objects with various colors, shapes, and materials were collected to construct an HDR image database. The material set includes a wide range of dielectric substances such as plastics, ceramics, painted woods, and leaves. The objects are not limited to flat plates but predominantly feature complex curved surfaces. Figure 5 shows the 267 objects collected for this purpose. Due to the dichromatic reflection properties of dielectric materials, the colors observed on their surfaces result from a combination of diffuse reflection (representing the object color) and specular reflection (representing the illumination color).

All dielectric objects were photographed using an iPhone 8 camera. The camera supports 12-bit depth, and the images were captured in a lossless raw format using Adobe Digital Negative (DNG). The camera's dark response was measured and subtracted from the output. The lighting conditions included both an LED ceiling lamp and natural daylight from a window. Care was taken to ensure that the surfaces of the objects exhibited gloss or highlights. The shutter speed and lighting conditions were adjusted to avoid saturation in a single-shot mode, and the image with the widest pixel distribution within the 12-bit range was selected for each object.

The captured images were recorded as relative values based on a white reference standard. The reference (Konica Minolta, Inc., CR-A43) was photographed alongside the target object, and the object's pixel values were normalized with respect to this reference. A pixel value of $x = 1.0$ corresponds to the same luminance level as the white standard. To compress

Figure 5 Set of dielectric objects collected from different colors and shapes, and made of different materials.

the dynamic range for efficient data processing, an inverse gamma correction was applied to the pixel values x:

$$y = x^{1/\gamma}, \tag{1}$$

where the γ value of 2.0 was used. Subsequently, the pixel values were scaled by 255 × y to fit the 8-bit LDR range [0, 255]. Thus, the original HDR image was normalized so that the white reference value equaled 1.0 in 8-bit terms, followed by gamma compression. The LDR images were then created by clipping the HDR images and formatting them into 8-bit representations.

Each original image was resampled to a resolution of 256 × 256 pixels. For data augmentation, the dataset was expanded through geometric transformations, including: (1) horizontal flipping, (2) zooming at scales of 1.0, 1.3, and 1.5, and (3) rotation at 13 angles: -90°, -75°, -60°, -45°, -30°, -15°, 0°, 15°, 30°, 45°, 60°, 75°, and 90°. As a result, each original image was augmented with 78 variants. Consequently, the total number of HDR and LDR image pairs in the resulting database was 20,826.

Method for Reconstructing HDR Image from Saturated LDR Image

A deep learning approach is employed to automatically predict a plausible HDR image from a single saturated LDR input. Supervised learning is conducted using the previously constructed dichromatic image database and a CNN.

(1) Network Structure

The network is designed in the form of a deep auto-encoder architecture. The full architecture used in this study is illustrated in Figure 6. It follows a U-Net-like structure [15], where the LDR input image is first processed by an encoder to produce a compact feature representation. This encoded representation is then passed to a decoder that reconstructs the corresponding HDR image.

To preserve high-resolution details during reconstruction, the network incorporates skip connections between corresponding layers in the encoder and decoder. These skip connections, indicated by green dotted arrows in Figure 6, help retain spatial information that might otherwise be lost during downsampling.

The network was implemented using MATLAB's machine learning toolbox [16]. It consists of 32 layers in total, with 85,900 learnable parameters—representing the sum of the weights and biases across all layers.

(2) Learning Procedure

Network training was performed using

net = trainNetwork(ds_train, net_Layers, opts),

where **ds_train** indicates the training dataset consisting of LDR and HDR pairs, **net_Layers** indicates the network layers, and **opts** specifies several options, including the learning algorithm and learning rates. The loss function is defined as follows:

$$E(\boldsymbol{\theta}) = \frac{1}{2L}\sum_{i=1}^{K}\left(t_i - y_i(\boldsymbol{\theta})\right)^2, \tag{2}$$

where the $\{t_i\}$ are the pixel values of the target HDR image, and $\{y_i\}$ are the reconstructed values predicted by the network. The vector $\boldsymbol{\theta}$ is a large learnable parameter vector with 85,900-demensions; K is the total number of observations, $K = 256 \times 256 \times 3$, indicating the product of image size and RGB channels, and L is the mini-batch size, representing the number of samples used for training.

We use the adaptive moment estimation (ADAM) algorithm [17] as an optimization algorithm in deep learning,

which is an improvement over stochastic gradient descent. A parameter $\boldsymbol{\theta}_t$ at the t-th step is updated as follows:

$$\begin{aligned}
\mathbf{g}_t &= \nabla E(\boldsymbol{\theta}_t) \\
\mathbf{m}_t &= \beta_1 \mathbf{m}_{t-1} + (1-\beta_1)\mathbf{g}_t \\
\mathbf{v}_t &= \beta_2 \mathbf{v}_{t-1} + (1-\beta_2)\mathbf{g}_t^2 \\
\hat{\mathbf{m}}_t &= \mathbf{m}_t / (1-\beta_1^t) \\
\hat{\mathbf{v}}_t &= \mathbf{v}_t / (1-\beta_2^t) \\
\boldsymbol{\theta}_t &= \boldsymbol{\theta}_{t-1} - \alpha \hat{\mathbf{m}}_t \left(\sqrt{\hat{\mathbf{v}}_t} + \varepsilon\right)
\end{aligned} \tag{3}$$

where $\mathbf{g}_t (= \nabla E(\boldsymbol{\theta}_t))$ is the gradient of the loss function E, α is the learning rate, $\mathbf{m}_t$ is the first moment vector, $\mathbf{v}_t$ is the second moment vector, and β_1 , $\beta_2 \in [0,1]$ represent exponential decay rates for moment estimates, and ε is a constant.

The prediction of an HDR image from an input LDR image is performed using the trained network in the form as

y = predict(net, ds_validation),

where **net** is the network trained above and **ds_validation** indicates the test dataset of LDR images for validation.

Experiments

(1) Performances of the proposed method

The original database contained 20,826 pairs of HDR and LDR images, which were randomly shuffled. Of these, 20,000 pairs were used for training the network, while the remaining 826 pairs were reserved for validation. Each pair was fed to the network as input and output during training. A single presentation of the entire training dataset is defined as one epoch. Training was repeated for multiple epochs until the mean squared error (MSE) reached an acceptable threshold. The training parameters in the trainNetwork function were configured as follows:

opts = trainingOptions('adam', 'InitialLearnRate', 1.0e-04, 'LearnRateSchedule', 'piecewise', 'LearnRateDropFactor', 0.98, 'LearnRateDropPeriod', 25, 'MaxEpochs', ***, 'MiniBatchSize', 64);

At 500 epochs, the average root mean square error (RMSE) across the 826 validation test images was 15.47. For the saturated regions only, the average RMSE was 34.55.

To visually demonstrate the reconstruction performance of the proposed method, three samples were selected from the test set. Figure 7 compares the input LDR image (left), the predicted HDR image (middle), and the original HDR image (right) for each sample. To facilitate comparison between LDR and HDR images, brightness levels were adjusted for display. Since the LDR images are clipped to the 8-bit range, they appear darker, and saturated highlight regions are displayed in gray. The first sample is a ceramic object, while the second and third are plastic. As discussed in Section 2, saturation due to clipping affects not only specular highlight areas but also body color regions resulting from diffuse reflection. The predicted HDR images generated by the proposed method visually match well with the original HDR images, indicating successful reconstruction.

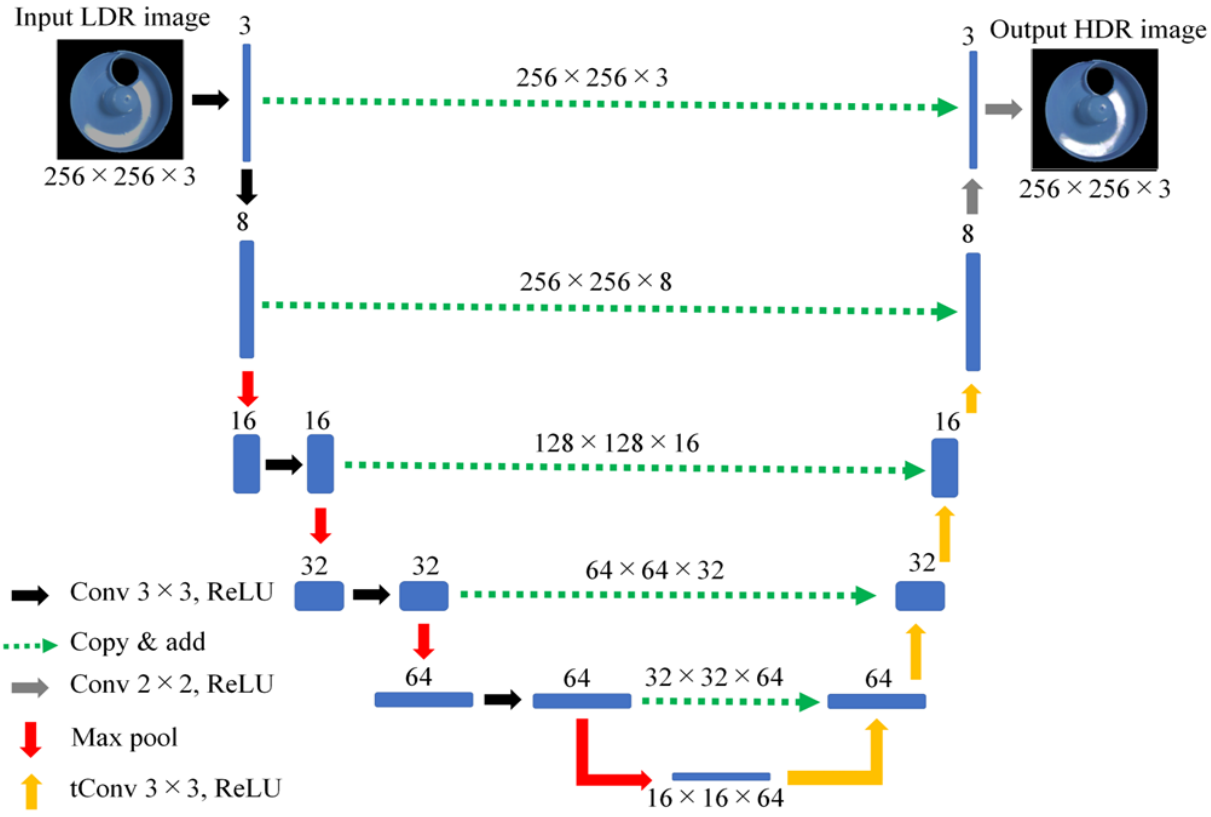


Figure 6 Entire network designed in this paper. The abbreviations of Conv, ReLU, Max pool, and tConv represent the respective operations of convolution, rectified linear unit, max pooling, and transposed convolution. The green dotted arrows represent the skip connections.

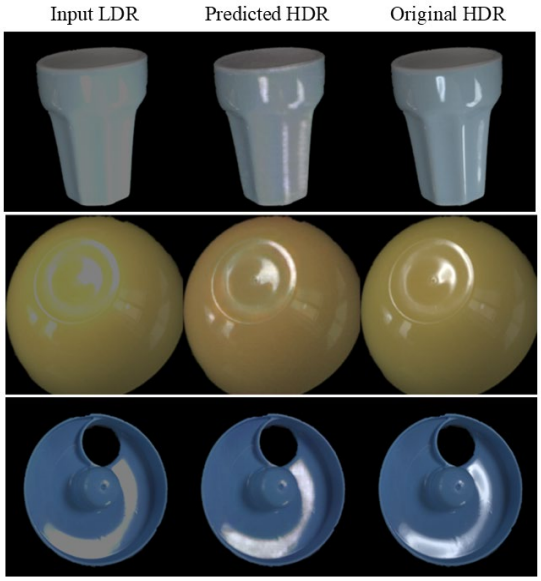


Figure 7 Comparisons of the input LDR image (left), the predicted HDR image (middle), and the original HDR image (right) for each of the selected samples.

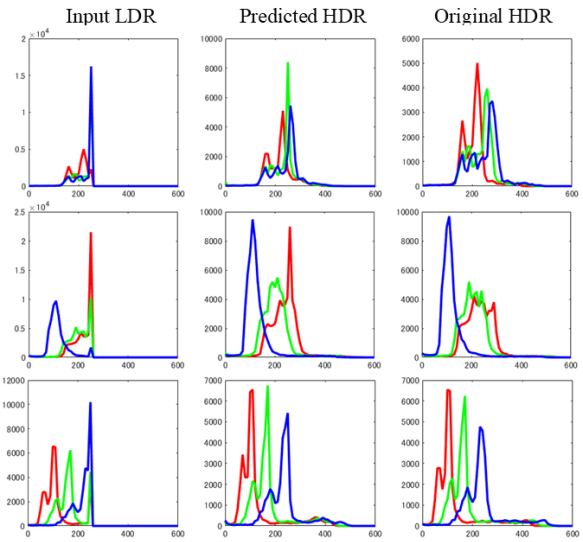


Figure 8 Comparisons of RGB histograms between the input LDR image (left), the predicted HDR image (middle), and the original HDR image (right) for each selected sample. The respective RGB histograms correspond to the respective sample images shown in Figure 7.

In addition to evaluating performance using the RMSE loss function, we also examined the RGB histogram distributions of the pixel values. The sample images used are the same as those in Figure 7. Figure 8 compares the RGB histograms of the input LDR (left), predicted HDR (middle), and original HDR images (right) over the range [1, 600] for each sample. In regions where the pixel values are not saturated, the predicted HDR values are consistent with both the LDR input and the corresponding values in the original HDR images. Therefore, the histogram portions within the range [1, 254] are largely identical across all three. While the LDR histograms exhibit clipping, the RGB histograms of the predicted HDR images closely approximate those of the original HDR images.

(2) Performance of neural networks trained on mixed data of metallic and dielectric objects

A neural network trained exclusively on metallic objects was found to perform poorly when applied to predicting HDR images from LDR images of dielectric objects. However, since dielectric objects with strong highlights can often be mistaken visually for metallic objects, we constructed a mixed image database containing both dielectric and metallic objects. This database was used to train the same network architecture as illustrated in Figure 6.

The mixed training database contained a total of 20,826 HDR-LDR image pairs, with half of the samples representing dielectric materials and the other half metallic. The training parameters for the **trainNetwork** function were identical to those used previously. All 826 test images used for evaluation were selected from dielectric object samples.

The average RMSE across the test images was 21.96 at 500 epochs. For the saturated areas alone, the average RMSE increased to 54.04. These results indicate a significant degradation in prediction accuracy when using a mixed-material training set. Therefore, it is essential to use a training dataset composed exclusively of dielectric objects for reconstructing HDR images of dielectric materials.

(3) Comparison with other methods

The performance of the proposed method was evaluated in comparison with five publicly available algorithms, all originally developed for natural scenes and not specifically for dielectric objects:

(1) G. Eilertsen, et al. [6]
(2) D. Marnerides, et al. [7],
(3) Y.-L. Liu, et al. [10]
(4) M. S. Santos, et al. [8]
(5) B. Masia, et al. [11].

Test dataset of 25 dielectric images were selected for this comparison. In these other methods, pixel values are typically represented on a linear scale, and the dynamic range varies across methods. To standardize the comparison, we applied an inverse gamma correction to the HDR outputs of each method and normalized the average pixel values to match those produced by our proposed method.

Each algorithm was applied to the same set of saturated LDR images to reconstruct HDR outputs. Table 1 presents the average RMSE values across the entire test set. The proposed method achieved the lowest RMSE of 14.48, demonstrating superior reconstruction accuracy compared to all other methods. Table 2 lists the average RMSE values computed only for the saturated regions in the test images.

In addition to RMSE, we evaluated histogram similarity using the goodness-of-fit coefficient (GFC), which quantifies the correlation between the predicted and original histogram curves [18]. For this analysis, the RGB histograms were represented as 61-dimensional vectors sampled at intervals of 10 across the range [1, 600]. Table 3 summarizes the average GFC values across all test samples. The proposed method achieved the highest GFC, indicating that the predicted HDR histograms closely match the original distributions.

Table 1 Comparisons between the average RMSEs over the whole test samples.

Ours	(1)	(2)	(3)	(4)	(5)
14.48	44.32	84.49	104.63	38.90	32.55

Table 2 Comparisons between the average RMSEs only for the saturated areas in the test images.

Ours	(1)	(2)	(3)	(4)	(5)
33.38	84.95	200.73	195.78	69.89	81.34

Table 3 Comparisons between the average GFC over the whole test samples.

Ours	(1)	(2)	(3)	(4)	(5)
0.935	0.698	0.230	0.221	0.653	0.514

Conclusions

In this study, we proposed a method for recovering the original HDR image from a saturated LDR image with missing physical information, specifically focusing on single dielectric objects. A deep neural network approach was adopted to map 8-bit LDR images directly to HDR images. We first examined the reflection and saturation properties of dielectric materials. Subsequently, an HDR image database was constructed using a wide variety of dielectric objects with different shapes and material properties.

Each object was photographed under general lighting conditions to induce strong gloss and specular reflections. The image data were captured in 12-bit RAW format. Although in the strong glossy areas of the HDR image, the luminance levels exceeded 8 bits, the overall maximum range was within 12 bits. The resulting HDR images were clipped to produce corresponding 8-bit LDR images. The database comprises approximately 20,000 HDR-LDR image pairs, each with a resolution of 256 × 256 pixels, which were used for training and validation. To estimate the lost information in saturated regions of LDR images, a deep CNN was developed with an autoencoder architecture incorporating skip connections. The network was implemented using MATLAB's machine learning framework, and the ADAM optimizer was used for training.

Performance evaluation was conducted using a validation set. Reconstruction accuracy was assessed using RMSE values, and histogram fidelity was evaluated through RGB histogram comparisons. The predicted HDR images closely matched the original HDR references. We further validated the reliability of our image database by training the network with mixed data comprising both metallic and dielectric objects. Additionally, comparisons with publicly available methods demonstrated that our approach significantly outperformed existing techniques in terms of reconstruction accuracy and histogram fidelity.

In this study, we focused on dielectric objects. The other objects with a highly glossy surface are metallic, and have different reflection properties than inhomogeneous dielectrics materials. Generalization to the case where a scene consists of different glossy materials, such as inhomogeneous dielectrics and metals, is left for future research.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number JP 24K15014.

References

[1] https://people.csail.mit.edu/celiu/CVPR2010/FMD/

[2] L. Sharan, R. Rosenholtz, and E. H. Adelson, Accuracy and speed of material categorization in real-world images, *Journal of Vision*, Vol. 14, No. 9, article 12 (2014).

[3] E. Reinhard, W. Heidrich, G. Ward, S. Pattanaik, P. Debevec, and K. Myszkowski, High Dynamic Range Imaging, 2nd Ed.: Acquisition, Display, and Image-Based Lighting, Morgan Kaufmann Publisher (2010).

[4] S. Lee, G. H. An, S-J. Kang, Deep recursive HDRI: Inverse tone mapping using generative adversarial networks, *ECCV 2018*, pp. 1-16 (2018).

[5] Y. Endo, Y. Kanamori, and J. Mitani, Deep reverse tone mapping, *ACM Trans. Graph.*, Vol. 36, No. 6, Article 177 (2017).

[6] 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, Article 178 (2017).

[7] 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, *Computer Graphics Forum*, Vol.37, No. 2 (2018).

[8] M. S. Santos, T. I. Ren, and N. K. Kalantari, Single image HDR reconstruction using a CNN with masked features and perceptual loss, *ACM Trans. Graph.*, Vol. 39, No. 4, Article 80 (2020).

[9] P. Hanji, R. K. Mantiuk, G. Eilertsen, S. Hajisharif, and J. Unger, Comparison of single image HDR reconstruction methods — the caveats of quality assessment, *Proc. SIGGRAPH '22*, Article No. 1, pp. 1-8 (2022).

[10] Y.-L. Liu, W-S. Lai, Y.-S. Chen, Y.-L. Kao, M.-H. Yang, Y.-Y. Chuang, and J.-B. Huang, Single-image HDR reconstruction by learning to reverse the camera pipeline, *CVPR 2020*, pp. 1651-1660 (2020).

[11] 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, Article 160 (2009).

[12] S. Tominaga and T. Horiuchi, An HDR image database construction and LDR-to-HDR mapping for metallic object, Proc. Color and Imaging Conference (CIC31), pp. 138–143 (2023).

[13] S. Tominaga, Surface identification using the dichromatic reflection model, IEEE Trans. PAMI, Vol. 13, No. 7, pp. 658-670 (1991).

[14] S. Tominaga, Dichromatic reflection models for rendering object surfaces, Journal of Imaging Science and Technology, Vol. 40, No. 6, pp. 549-555 (1996).

[15] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, *Proc. Medical Image Computing and Computer-Assisted Intervention (MICCA)*, LNCS, Vol. 9351, pp.234-241 (2015).

[16] https://jp.mathworks.com/discovery/machine-learning-models.html

[17] D.P. Kingma and L.B. Jimmy, Adam: A method for stochastic optimization, Proc. ICLR., arXiv: 412.6980 (2014).

[18] J. Romero, A. Garcıa-Beltran, and J. Hernandez-Andres, Linear bases for representation of natural and artificial illuminants, J. OSA-A, Vol. 5, No. 5, pp.1007-1014 (1997).

Author's Biography

Shoji Tominaga received the B.E., M.S., and Ph.D. degrees in electrical engineering from Osaka University, Japan. He was a Professor (2006-2013) and Dean (2011-2013) at Graduate School in Chiba University. He is now an Adjunct Professor, Norwegian University of Science and Technology and also a Visiting Researcher, Nagano University. His research interests include multispectral imaging, and material appearance. He is a Fellow of IEEE, IS&T, SPIE, and OSA