https://doi.org/10.2352/CIC.2023.31.1.26


# An HDR Image Database Construction and LDR-to-HDR Mapping for Metallic Objects

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

## Abstract

*We consider a method for reconstructing the original HDR image from a single LDR image suffering from saturation for metallic objects. A deep neural network approach is adopted for directly mapping from 8-bit LDR image to an HDR image. An HDR image database is first constructed using a large number of objects with different shapes and made of various metal materials. Each captured HDR image is clipped to create a set of 8-bit LDR images. The whole pairs of HDR and LDR images are separated and used to train and test the network. Next, we design a deep CNN in the form of a deep auto-encoder architecture. The network was also equipped with skip connections to keep high image resolution. The CNN algorithm is constructed using MATLAB's machine-learning functions. The entire network consists of 32 layers and 85,900 learnable parameters. The performances of the proposed method are examined in experiments using a test image set. We also compare our method with other methods. It is confirmed that our method is significantly superior in reconstruction accuracy and the good histogram fitting.*

## Introduction

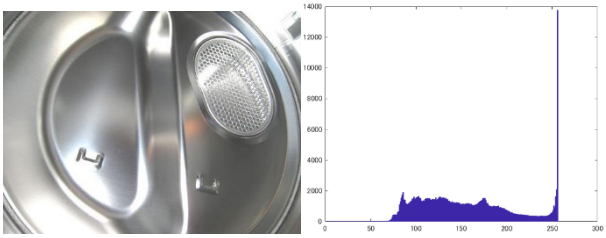
Digital cameras can only capture a limited range of luminance level in real-world scenes due to sensor constraints. High-quality cameras for high dynamic range (HDR) imaging are sometimes unaffordable. However, most existing image content has a low dynamic range (LDR), and the majority of legacy content is predominantly 8-bit LDR images. Objects in real scenes do not always have matte surfaces, and often have surfaces with strong gloss or specular highlights. In such a case, pixel values in the captured images are saturated and clipped due to a limited dynamic range of image sensors, so physical information is missing in the saturated image regions.

Metals are typical object materials that easily saturate, where the luminance of the reflected light from a metal object has an extensive dynamic range from matte surface reflection component to highlight specular reflection component. Figure 1 demonstrates an example from the image data belonging to the material category of metal in the Flicker Material Database (see [1]-[2]), where the database is divided into 10 material categories, such as metal, plastic, fabric, foliage, and so on. All of which consist of 8-bit images. Figure 1A shows the color image named metal_moderate_002_new. Figure 1B shows the luminance histogram in the 8-bit range. It can be seen that a wide area of the metal object surface is saturated. The color and shading information in the saturated image area is entirely incorrect due to the missing physical details. As a result, such attempts as appearance reproduction, gloss perception, and appearance modeling fail for this object.

Therefore, a method is needed to infer the original HDR image from a single LDR image suffering from saturation, often called inverse tone mapping problems [3]. This is an ill-posed problem because a missing signal not appearing in a given LDR image should be restored [4]. So far, this problem has been mainly dealt with in the field of computer graphics [5]-[10] and also partly in the field of computer vision [4], [11]. The target images are natural scenes, not material objects. Therefore, the captured images contain not only objects but also the sky and various light sources.

This paper considers a method for reconstructing the original HDR image from a single LDR image suffering from saturation for metallic objects. A deep neural network approach is adopted for directly mapping from 8-bit LDR image to an HDR image. We note that there is no publicly available HDR image dataset although a few LDR datasets are in widespread use like the flicker material database. Therefore, we first construct an HDR mage database specialized for metallic objects. A large number of objects with different shapes and made of various metal materials are collected for this purpose. These objects are photographed under a general lighting environment so that strong gloss or specular reflection can be observed. Each captured HDR image is clipped to create a set of 8-bit LDR images. Pair of the created LDR images and the original HDR images in the database are used to train and test the network.

We propose a LDR-to-HDR mapping method to predict information that has been lost in saturated areas of the LDR images. We design a convolutional neural network (CNN) in the form of a deep auto-encoder architecture. We also equip the network with skip connections to make optimal use of high resolution image details in the construction. In experiments, the performances of the proposed method are examined in detail and compared with those of other methods. The accuracy of the reconstructed HDR images and the superiority in comparison with other methods are shown based on validations of numerical error and histogram reconstruction.




(A) (B)

*Figure 1 Example from the image set belonging to the metal category in the Flickr Material Database: (A) Color image named "metal moderate 002 new". (B) Luminance histogram of the image in 8-bit range.*

## HDR Image Database for Metallic Objects

We collected a large number of object with different shapes and made of different materials. The material set collected consists of a wide range of metal materials such as iron, copper, zinc, nickel, brass, aluminum, stainless steel, gold, silver, and metal plating. Painted metal objects are excluded. The object shapes are not only flat plates but also mostly various complicated

curved surface. Figure 2 shows 150 metal objects collected in this way. It is known that light reflection from a metallic object consists of mostly specular reflection different from diffuse reflection [12]. The color appearing on an object surface is a metal color, coincident with the gloss/highlight color. We note that the color at gloss/highlight areas is not white but the metal colors seen in Figure 2.

The metal objects were photographed using a camera of iPhone 8. The camera's depth is 12 bits, and the details including the spectral sensitivity functions are shown in Ref. [13]. The camera images were captured in a lossless raw image format in Adobe digital negative (DNG) format. The dark response was measured and discarded from the camera output.

The lighting environment at the time of capturing was both a LED ceiling lamp and natural daylight through a window, and the capturing was devised so that the surface of the metal object included glosses or highlights. The images were taken by adjusting the shutter speed and lighting conditions to avoid saturation in the one-shot mode, and the images were then processed as HDR. In a sense, the images were shot with multiple exposures.

Such bright areas in the captured images are highly dependent on the position of the object and the camera. Therefore, by shifting their positions, we additionally photographed multiple objects with different shading. Thus, a set of 191 original images of metal objects was constructed, where the backgrounds of the target objects were erased.

The original image was resampled to a size of 256×256 pixels. For data augmentation, each original image was geometrically varied in such a way as (1) image horizontal flipping, (2) zoom using the three factors of 1.0, 1.3, and 1.5, and (3) rotations with the 13 angles of -90, -75, -60, -45, -30, -15, 0, 15, 30, 45, 60, 75, 90 degrees. That is, each original image had 78 modifications.

Processes for creating HDR and LDR are summarized as.

**1) HDR creation**: The pixel values of the captured original image are normalized so that the pixel value of the white standard is 1 (8 bits). Then, the inverse gamma transformation is applied to compress the normalized images.

**2) LDR creation**: The LDR images are created after clipping the HDR images and adjusting the final format to 8 bits.

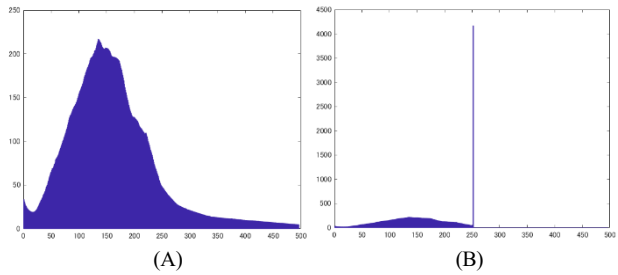
The captured images take relative values based on a white reference standard. The white reference standard (Minolta, CR-A43) was photographed with a target object, and then the object camera values were normalized using the white reference values. If the luminance level of the object is the same as the white reference, the pixel value $x = 1.0$.

To compress the dynamic range for convenience of data processing, we applied a non-linear transformation of inverse gamma correction to the pixel values $x$

$$y = x^{1/\gamma}, \qquad (1)$$

where the $\gamma$ value of 2.0 was used. Furthermore, the pixel values were converted with $255 \times y$ to fit the 8-bit LDR range [0, 255]. Pixel values above this range were saturated into HDR. When the number of saturated pixels was small, we regarded these as noises. We also supposed that the saturated areas were not large enough to cover the entire object because, in such a case, we could not recover the saturated pixels from the single LDR image. Based on this consideration, we calculated the ratio $R$ of the saturated area to the total object area in each image. Then, the saturated HDR image set that was effective for the present study was adopted by satisfying the conditions $0.04 \le R \le 0.40$. The total number of HDR and LDR pairs in the image database created in this way was 9,855. Figure 3A plots the average luminance histogram in the HDR image database. The RGB pixel values range very widely. The maximum value of HDR images is 2010. Figure 3B shows the average luminance histogram in the corresponding LDR image database suffered from saturation by clipping into the 8-bit range with maximum of 255.

*Figure 2 Set of material objects collected from different shapes and made of different materials.*



(A) (B)

*Figure 3 Average luminance histograms of the created image database: (A) Average HDR image histogram. (B) The average LDR image histogram suffered from saturation by clipping into the 8-bit range.*

## Method for LDR-to-HDR Mapping

We consider a deep-learning approach to automatically predict a plausible HDR image from a single LDR image. Supervised learning using a deep CNN is performed based on the image database previously created. We design the network in the form of a deep auto-encoder architecture. The entire network

designed in this paper is shown in Figure 4. The LDR input image is transformed by an encoder network to produce a compact feature representation of the image. The encoded image is then provided to an HDR decoder network to reconstruct an HDR image. Furthermore, the network is equipped with skip connections in order to make optimal use of high resolution image details in the construction, which is inspired by U-net architecture [14]. The green dotted arrows in Figure 4 represent the skip connections.

We used MATLAB's machine learning functions for constructing the above designed network. The layers are constructed as

```
layers = [imageinputLayers, encodingLayers,
decordingLayer].
```

The part of encoding layers is described as

```
encodingLayers = [convolution2dLayer(3, 8, …),
reluLayer, maxPooling2dLayer(2, … ,'Stride', 2),
convolution2dLayer(3, 16, …), reluLayer,
maxPooling2dLayer(2, … ,'Stride', S),
convolution2dLayer(3,32, …), reluLayer,
maxPooling2dLayer(2, … ,'Stride', S),
convolution2dLayer(3, 64, …), reluLayer,
maxPooling2dLayer(2, … , 'Stride', S)],
```

where **convolution2dLayer**(N, M, ...) applies M sliding convolutional filters with size [N, N] to a 2D input image, and **reluLayer** performs a threshold operation on each element of the input, setting it to zero if the value is less than zero, and **maxPooling2dLayer**(N, ..., 'Stride', S) performs downsampling by dividing the input into rectangular pooling regions with size [N, N] and stride [S, S], and computing the maximum value of each region.

The decoding part is descried in the basic form as

```
decodingLayers =
[transposed_Conv2dLayer(2,64,Stride=2),
reluLayer, transposed_Conv2dLayer(2,32,
Stride=2), reluLayer,
transposed_Conv2dLayer(2,16,Stride=2),reluLayer,
transposed _Conv2dLayer(2,8,Stride=2), reluLayer,
convolution2dLayer(1, 3,…),
clippedReluLayer(1023.0), regressionLayer],
```

where **transposed_Conv2dLayer**(N, M, 'Stride', S) performs upsampling to a 2D feature map with M filters of [N, N] size, and stride [S, S], **clippedReluLayer**(T) performs clipping with upper limit T, and **regressionLayer** computes the half-mean squared error loss for a regression task.

The loss function is defined as

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

where $\{t_i\}$ are the target values, $\{y_i\}$ are the predicted values by the network, and $R$ is the total number of observations, that is $R = 256 \times 256 \times 3$ in our case. The entire network consists of 32 layers, and the total number of learnable parameters is 85,900.

The network training is performed in the form as

```
net = trainNetwork(ds_train, net_Layers, opts),
```

where ds_train indicates the training dataset consisting of LDR and HDR pairs and opts specifies several options, including the learning algorithm and learning rates. We use the stochastic gradient descent algorithm with a momentum term (SGDM) algorithm [15] for network training. A parameter $\theta_t$ at the $t$-th step is updated as

$$\theta_{t+1} = \theta_t - \alpha\nabla E\left(\theta_t\right) + \beta\left(\theta_t - \theta_{t-1}\right), \tag{3}$$

where $\nabla E\left(\theta_t\right)$ is the gradient of the loss function $E$, $\alpha$ is the learning rate, and the third term is the momentum, and $\beta$ represents contribution from the past.

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

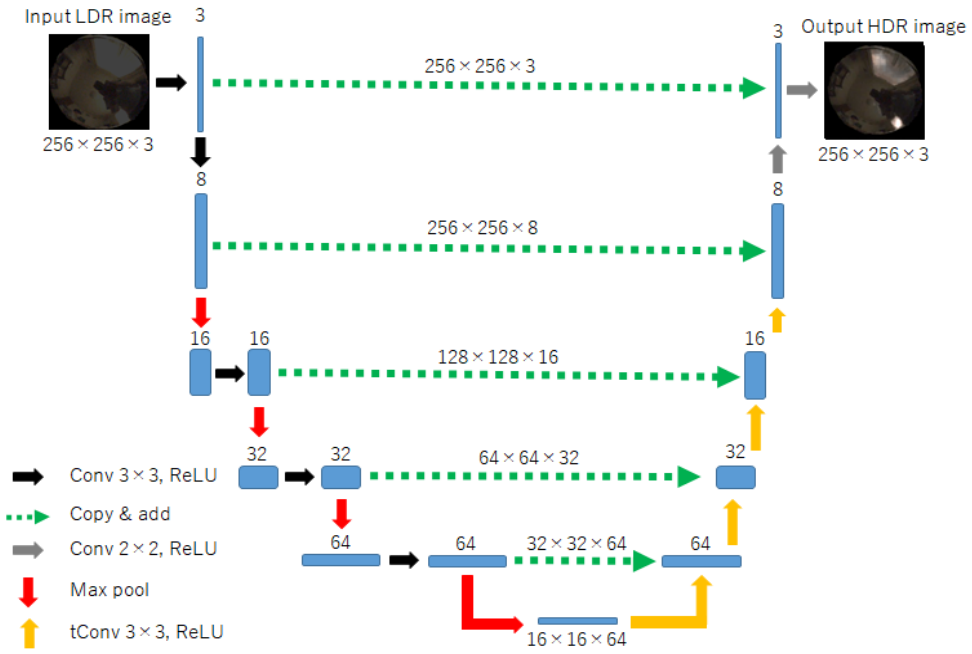



*Figure 4 Entire network designed in this paper. The abbreviations of Conv, ReLU, Max pool, and tConv represent the respective operations of convolution, rectified linear unit, max pooling, and transposed convolution. The green dotted arrows represent the skip connections.*

```
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*

We randomly selected 100 pairs of HDR and LDR images to validate the proposed method from the original database, which consists of 9,855 pairs of HDR and LDR images. The remaining 9,755 image pairs were used as the data for the network training. Each pair was presented to the network input and output. One period of presenting the entire training data is defined as an epoch. The training was iterated for as many epochs as necessary to decrease the mean-square-error to an acceptable level. The root-mean-square error (RMSE) was 19.73 for 1,450 epochs.

Figure 5 shows the test dataset consisting of 100 images. The average RMSE was 26.95 only for the saturated areas in these test images. To visually clarify the reconstruction results by the proposed method, we further selected three samples from the test image set. Figure 6 compares the input LDR image (left), the predicted HDR image (middle), and the original HDR image (right) for each sample. Each image is displayed in a 16-bit tiff to avoid saturation. The LDR images are saturated in the 8-bit range, so all LDR images are very dark, and the highlight areas with saturation appear gray.

We note that not only matte areas on the surface, but also gloss and highlight areas have the same object color, which is metallic. This characteristic essentially differs from dielectric materials such as plastic. For instance, gloss/highlight areas on the predicted images for the first and third samples appear the metallic colors of copper and gold. Thus, we can see that the predicted HDR images from the LDR images are well recovered close to the target HDR images.

In addition to the validation using the loss function of RMSE, we investigated the histogram distributions of RGB pixel values. The sample images are the same as shown in Figure 6. We note that if the pixel values in the predicted HDR image are not saturated, the pixel values have the same as the LDR image, and also the same as the pixel values without saturation in the original HDR image. Therefore, we should compare the histogram distributions in the saturated areas only between the predicted images from the LDR images and the target images of the original HDR image database. Figure 7 plots the RGB histograms of the predicted and target images in the range of [200, 800] suffering from the saturation. When noticing that the LDR histograms are saturated similarly to Figure 3(B), the RGB histograms of the predicted HDR images are well recovered for the histograms of the original HDR images.

*Figure 5 Test dataset consisting of 100 images used for validation.*

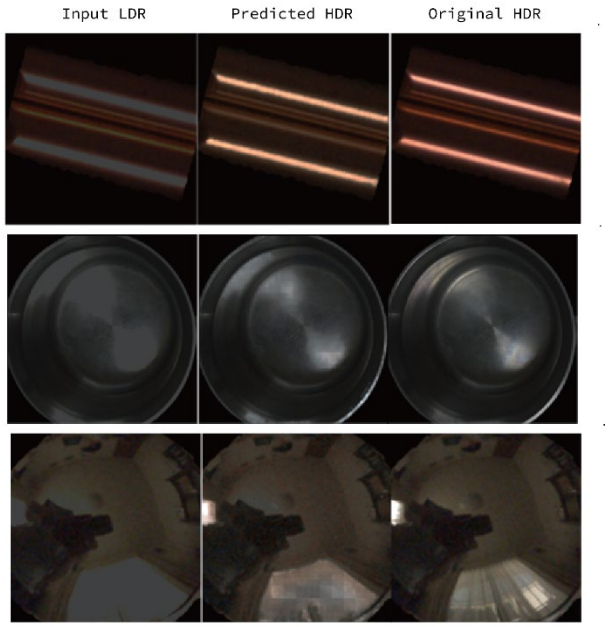



*Figure 6 Comparisons of the input LDR image (left), the predicted HDR image (middle), and the original HDR image (right) for each of the selected samples.*

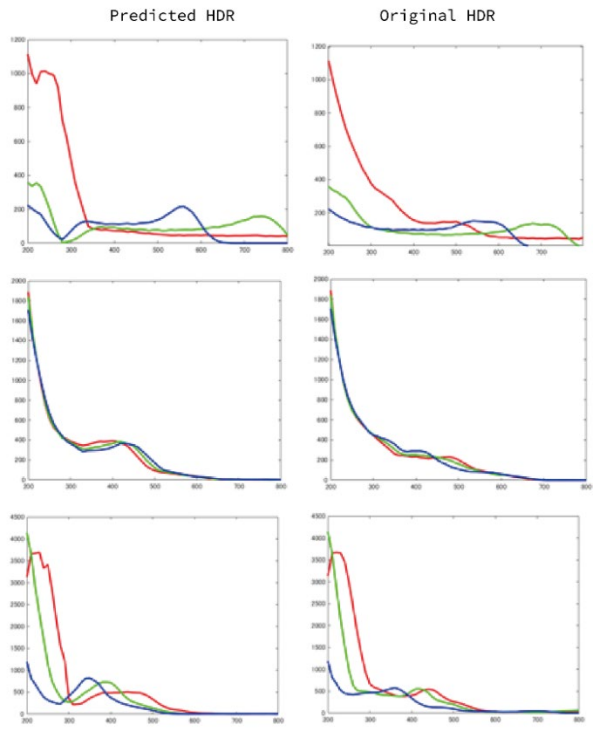



*Figure 7 Comparisons of RGB histograms between the predicted HDR image (left), and the original HDR image (right) for each selected sample. The respective RGB histograms correspond to the respective sample images shown in Figure 6.*

### *(2) Comparisons with other methods*

We randomly selected another test dataset consisting of 10 images from the test dataset shown in Figure 5. Figure 8 shows a set of HDR images used to compare with other methods.

The following five methods with algorithms open to public use were selected for comparison. These methods were proposed for natural scenes, not limited to metallic objects.

(1) G. Eilertsen, et al. [7]
(2) D. Marnerides, et al. [8],
(3) Y.-L. Liu, et al. [11]
(4) M. S. Santos, et al. [9]
(5) B. Masia, et al. [5].

We input the LDR images with saturation and executed the respective algorithms to reconstruct the HDR images. Figure 9 compares the reconstruction results for the first test sample, where from left to right, the input LDR image, the proposed method, other methods (1) to (5), and the ground truth image are arranged in order. The RMSEs for the first test data are 22.31 for ours, 67.50 for (1), 63.81 for (2), 82.08 for (3), 78.79 for (4), and 64.57 for (5). Table 1 compares the average RMSEs over the whole test samples shown in Figure 8. Thus, it can be seen that the proposed method with the RMSE 20.76 is remarkably superior in reconstruction accuracy compared to other methods.

*Figure 8 Test dataset of 10 images selected for the comparison.*

**Table 1 Comparisons between the average RMSEs over the whole test samples shown in Figure 8.**

| Ours | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| 20.76 | 57.96 | 47.95 | 51.08 | 60.55 | 51.96 |

Histogram reconstruction is a kind of performance evaluation. Figure 10 compares the RGB histograms of the reconstructed images between our method and the other methods (1)-(5) in the range of [200, 800]. Compared with the other methods, we can see that the histograms of the proposed method is smooth and close to the ground truth. The goodness-fitting coefficient (GFC) is useful for evaluating the histogram distributions [16] numerically. This measure is a kind of correlation coefficient between the predicted and target histogram curves. Let $\mathbf{h}_{true}$ be a 61-dimensional (D) column vector representing the histogram of the target HDR image in the range of [200, 800] at 10 steps, and $\mathbf{h}_{pred}$ be a 61-D column vector representing the histogram of the predicted HDR image in the same range.

Then GFC is defined as

$$GFC = \frac{\mathbf{h}_{true}^{t} .* \mathbf{h}_{pred}}{\left\|\mathbf{h}_{true}\right\|\left\|\mathbf{h}_{pred}\right\|}, \quad (4)$$

where $\mathbf{h}^{t}$ and $\left\|\mathbf{h}\right\|$ indicates the matrix transposition and the norm of $\mathbf{h}$, respectively, the symbol (. *) represents element-wise multiplication. The GFCs for the first test data are 0.999 for ours, 0.730 for (1), 0.850 for (2), 0.767 for (3), 0.702 for (4), and 0.635 for (5). Table 2 compares the average GFCs over the whole test samples. Thus, the histograms in the proposed method fit the original ones.

**Table 2 Comparisons between the average GFC over the whole test samples.**

| Ours | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| 0.984 | 0.772 | 0.795 | 0.811 | 0.792 | 0.712 |

## Conclusions

In this paper, we have considered a method for reconstructing the original HDR image from a single LDR image suffering from saturation for metallic objects. A deep neural network approach was adopted for mapping from 8-bit LDR image directly to an HDR image. We first constructed an HDR mage database specialized for metallic objects. A large number

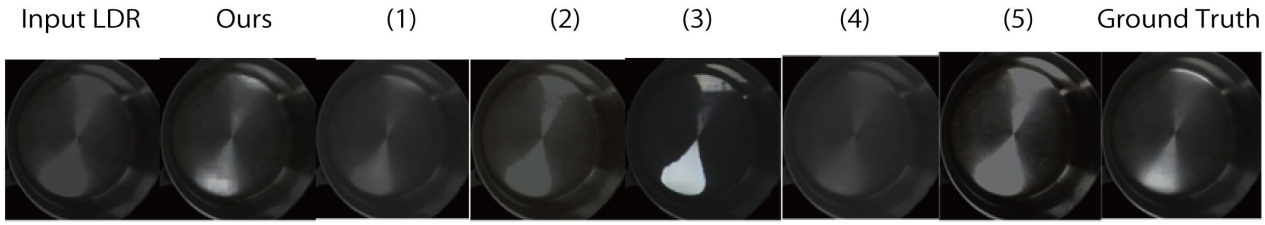



*Figure 9 Comparison of the reconstruction results for the first test sample.*

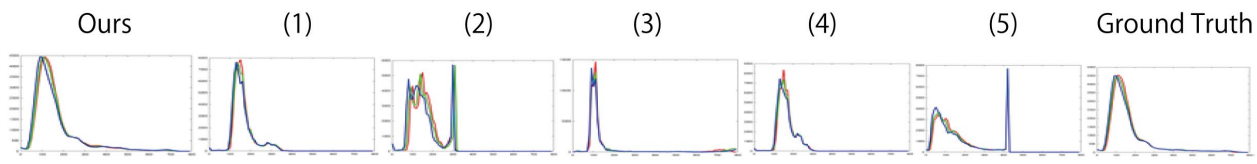



*Figure 10 Comparison of the RGB histograms between our method and the other methods (1)-(5) in the range of [200, 800].*

of objects with different shapes and made of various metal materials were collected for this purpose. We photographed these objects under a general lighting environment so that strong gloss or specular reflection could be observed. Each of the captured HDR images was clipped to create a set of 8-bit LDR images. The HDR and LDR images were represented with 256×256 pixels. The total number of HDR and LDR pairs in the created image database was 9,855, separated and used to train and test the network.

An LDR-to-HDR mapping method was proposed to predict information that was lost in saturated areas of the LDR images. We designed a deep CNN in the form of a deep auto-encoder architecture. The LDR input image was transformed by an encoder network to produce a compact feature representation of the image. The encoded image was then provided to an HDR decoder network to reconstruct an HDR image. The network was also equipped with skip connections to keep high image resolution. The entire network algorithm was constructed using MATLAB's machine-learning functions. The entire network consisted of 32 layers, and the total number of learnable parameters was 85,900.

In experiments, we examined the performances of the proposed method using a set of test images for validation. The predicted HDR images from the LDR images were well recovered close to the target HDR images. We showed the RMSE values and the RGB histogram distributions. We also compared the performances with other methods whose algorithms were open to public use. Our method was significantly superior to the others in reconstruction accuracy and excellent histogram fitting.

## Acknowledgments

This work was supported by JSPS KAKENHI Grant Number JP 20K11893.

## 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).

[5] 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).

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

[7] 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).

[8] 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).

[9] 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).

[10] P. Hanji, R. K. Mantiuk, G. Eilertsen, S. Hajisharif, and J. Unger, Comparison of single image HDR reconstruction methods — the caveats of quality assessment, SIGGRAPH '22 Conference Proceedings (2022).

[11] 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).

[12] S. Tominaga, Dichromatic reflection models for a variety of materials, *Color Research and Application,* Vol. 19, No. 4, pp.277 – 285 (1994).

[13] S. Tominaga, S. Nishi, and R. Ohtera, Measurement and estimation of spectral sensitivity functions for mobile phone cameras, *Sensors*, Vol. 21, No. 15, Article 4985 (2021).

[14] 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 Springer (2015).

[15] K. P. Murphy, Machine Learning: A Probabilistic Perspective. The MIT Press (2012).

[16] 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*