Saturated Reflection Detection for Reflection Removal Based on Convolutional Neural Network

TAICHI YOSHIDA¹, (Member, IEEE), ISANA FUNAHASHI¹, NAOKI YAMASHITA, and MASAACKI IKEHARA², (Senior Member, IEEE).

¹Dept. of Computer and Network Engineering, University of Electro-Communications, Chofu, Tokyo 182-8585, Japan (e-mail: t-yoshida@uec.ac.jp, i.funahashi@uec.ac.jp)
²Dept. of Electronics and Electrical Engineering, Keio University, Yokohama, Kanagawa 223-8522, Japan (e-mail: ikehara@tkhm.elec.keio.ac.jp)
Corresponding author: Taichi Yoshida (e-mail: t-yoshida@uec.ac.jp).

ABSTRACT Single image reflection removal is a technique that removes undesirable reflections, which occur due to glass, from images. Various methods of reflection removal have been proposed, but unfortunately, they usually fail to remove reflections particularly with very high pixel values. In this paper, we define these saturated reflections and their characteristics, as well as discuss and propose a removal system. The proposed system detects areas of saturated reflections based on our proposed model of convolutional neural networks and restores them by a conventional method of image estimation. In our experiments, the proposed system shows better peak-signal-to-noise ratio scores and perceptual quality than conventional methods of reflection removal.

INDEX TERMS Single image reflection removal, reflection detection, convolutional neural network, image inpainting.

I. INTRODUCTION

SINGLE image reflection removal has been actively studied, and many methods have been proposed over several decades [1]–[11]. When we take a picture through glass, reflections often occur at the glass surface in the resulting image. The phenomenon is undesirable, and it degrades not only the image quality but also the accuracy in applications of computer vision such as image recognition, object detection, and so on. To remove reflections from images, various methods have been proposed which typically use optimization theory [1]–[11].

Recently, methods based on convolutional neural networks (CNN) have been actively proposed and show better results than the conventional ones [12]–[31]. A method using CNN was firstly proposed in [12]. A method that can be trained with unaligned images is proposed [19]. A method that uses the generative adversarial network and a loss function based on gradient information is proposed [16]. These state-of-the-art methods show better results than conventional ones, both perceptually and objectively.

Unfortunately, state-of-the-art methods usually fail to remove reflections particularly with very high pixel values near the saturation because they cannot recognize them as reflections. In this paper, these particular reflections are called saturated reflections. An example of saturated reflections is shown in the red box of Fig. 1(a) and its ground truth (GT) is shown in Fig. 1(b). Saturated reflections are usually caused by light sources and completely conceal the background information. Fig. 1(c) shows the removed results of (a) by the method [19]. From this figure, it is observed that it fails to remove the saturated reflection. These methods assume that the pixel values of images with reflections are the sum of the background and reflection images. Since this assumption is not valid for saturated reflections, these methods do not recognize them as reflections.

In this paper, we tackle this problem and introduce a removal system using our proposed detection method. First, we discuss the definition, characteristics, and removal procedure of saturated reflections. To remove saturated reflections, we propose a detection method based on CNN. In the proposed method, pixels that have very high values are detected by pre-processing, and then the proposed CNN model classifies resultant pixels into backgrounds and reflections. The proposed CNN model is based on U-Net [32], [33] and uses high-frequency components of the input image as one of the inputs. The proposed system combines the proposed detec-
tion method and a conventional image restoration method for reflection removal. In our experiments, we compare the proposed method with humans in saturated reflection detection and the proposed system with two state-of-the-art methods in reflection removal. The proposed system removes saturated reflections and has better PSNR scores than the state-of-the-art methods, which is perceptually shown in Fig. 1(d).

The contributions of this paper are shown as follows:

- We indicate a new problem of reflection removal and clarify its characteristics, which contributes to the development of reflection removal.
- We propose a method and a system for removing reflections with saturated reflections and show their efficacy in the experiments. The system can be directly used with conventional methods to improve them.
- Through our experiments, we show the efficacy of the proposed CNN using high-frequency components of images for the detection of saturated reflections.

This paper is organized as follows: Section 2 shows the discussion for saturated reflections. The proposed method for detecting saturated reflections and the proposed system using it for reflection removal are explained in Section 3 and 4, respectively. Experiments are denoted in Section 5 and this paper is concluded in Section 6.

II. DISCUSSION FOR SATURATED REFLECTIONS

A. DEFINITION

We define saturated reflections as reflections with very high pixel values near saturation. For example, when images are represented by 8-bit, pixel values of saturated reflections are close to 255 in at least one color channel. By definition, saturated reflections usually saturate and eliminate the background information in their areas. Saturated reflections are typically caused by high-luminance objects such as light sources and hence have white color.

B. CHARACTERISTICS

The state-of-the-art methods based on CNN usually fail to remove saturated reflections as mentioned in Section I [16], [17], [19], [21]. These methods assume that pixel values of input images are the sum of reflection and background images. Unfortunately, since the assumption is not valid for saturated reflections due to their definition, the methods usually recognize saturated reflections as light sources in backgrounds and avoid removing them. Therefore, to remove these saturated reflections, another technique is required.

To remove saturated reflections, it is required to first detect and then estimate the pixel values of the background. Since pixel values of the background in areas of saturated reflections are eliminated as mentioned above, it is required to estimate them using adjacent pixels. Moreover, accurately detecting the saturated areas is required for estimation. Candidate areas can be straightforward detected by thresholding input images, but these areas can belong to either the reflection or the background. Therefore, a technique to classify candidate areas into reflections or backgrounds is required. The above steps, both detection and estimation, are required for the removal of saturated reflections.

We presume that high-frequency components of images are useful for detecting saturated reflections, which is shown in this paper. Since the glass has some thickness, there are two reflection planes and the light from one source reflects several times at the planes [2] as shown in Fig 2. This phenomenon causes the blurring of objects acquired by the reflection and thus they are more blurry than those in the background [2], [17]. Therefore, high-frequency components and gradients are sometimes used for detecting reflections in conventional methods [1], [5], [6], [9], [12], [14], [16]. Specifically, since the light of the saturated reflections has very high-luminance values, we presume that it reflects many times off of the glass and its objects are strongly blurred. Pixel values of saturated reflections are smoothly attenuated towards the circumference, and the gradient information of images is useful for detection.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{(a) Input image, (b) ground truth, (c) removed result by [19], and (d) removed result by the proposed system.}
\end{figure}
C. DETECTION BY HUMANS

An experiment with humans was conducted to measure the human ability to recognize saturated reflections from candidate areas. To reduce the experimental costs for subjects, a simple procedure was conducted as follows: A set of two images was shown side by side, where the left image was a natural image and the right image showed highlighted areas that include pixels near saturation. The highlighted areas are either saturated reflections or light sources in the background. Given an image set, subjects were asked to click the highlighted areas that they recognized as reflections. There was no time limitation or constraint on the number of clicking. Subjects were 20 people, Japanese, twenties and thirties, male and female, and not familiar with the field of reflection removal. 30 images were used from datasets of natural images [3], [5], [15], [19].

Table 1 shows the results of this experiment, where these measures are shown in [34], and \( \mu \) and \( \sigma \) denote mean and standard deviation values for all subjects, respectively. From the recall scores, it is observed that humans sometimes recognize saturated reflections as light sources of backgrounds. From the precision scores, the human ability to recognize reflections is relatively high. Although the procedure of this experiment is simple and was conducted at low cost, humans cannot strictly select saturated reflections from the presented areas. These scores can be recognized as one of the criteria for this theme.

III. DETECTION METHOD FOR SATURATED REFLECTIONS

A. OVERVIEW

Fig. 3 shows an overview of the proposed method for detecting areas of saturated reflections. For pre-processing, a map of candidate areas \( M \) and high-frequency components \( C \) are calculated from an input image \( I \) by thresholding and filtering with various highpass filters. The proposed CNN model \( F_\omega \) produces a map of initially detected areas \( \bar{D} \) from \( I \) and \( C \) as

\[
\bar{D} = F_\omega(I, M, C),
\]

where \( \kappa \) denotes the learnable parameters of the proposed model. \( \bar{D} \) has real values and is binarized. Finally, a resultant map of detected areas \( D \) is calculated via the post-processing from binarized \( \bar{D} \) and \( M \).

B. PRE- AND POST-PROCESSING

The pre-and post-processing of the proposed CNN model, mentioned in Section III-A, are explained here. \( M \) is a binary map and calculated as

\[
m_i = \begin{cases} 
1 & \text{if } y_i > \Delta \\
0 & \text{otherwise}
\end{cases},
\]

where \( m_i \) and \( y_i \) are \( i \)-th elements of \( M \) and the luminance of \( I \), respectively. In experiments of this paper, the luminance is the same as the Y plane of \( I \) in YCbCr color space, and \( \Delta = 0.95 \) when \( y_i \in [0,1] \). To remove outliers, candidate areas whose number of pixels is three or less are eliminated. \( C \) is calculated by separately applying four filters, shown in Table 2, to the luminance plane, where No. denotes the filter type and * means that filter C has no coefficient in diagonals. If more than half of the pixels in a candidate area of \( M \) have pixel values in \( \bar{D} \) greater than \( \delta \), pixel values of the area in \( \bar{D} \) is determined to be 1. In other words, let \( \omega \) and \( |\omega| \) be an index set of the area and the number of its elements, then pixel values of the area in \( \bar{D} \) is defined as

\[
d_{i \in \omega} = \begin{cases} 
1 & \text{if } |\{ j \in \omega | d_j > \delta \}| > |\omega|/2 \\
0 & \text{otherwise}
\end{cases},
\]

where \( d_i \) and \( \bar{d}_i \) are \( i \)-th elements of \( D \) and \( \bar{D} \), respectively. In this paper, \( \delta = 0.5 \) when \( d_i \in [0,1] \). Morphological transformations are applied to \( D \), and the dilation of \( 7 \times 7 \) is applied in this paper.

C. DETECTION CNN

1) Architecture

The architecture of the proposed network is inspired by U-Net [32], [33], and its details are shown as follows: Fig. 4 shows the architecture, where white, blue, green, and red boxes denote convolution, max pooling, and transposed convolution layers and squeeze-and-excitation-block (SE-block) [35], respectively. Some layers have multiple input arrows in Fig. 4. In those layers, input features are concatenated along...
of reflections. We presume that features of because the above property is also necessary for the detection of features without degrading the localization accuracy [32]. The proposed network is constructed based on U-Net image segmentation and can increase the spatial resolution of one of the most efficient and well-known structures for producing a map with accurate localization. U-Net is resolution representations for space and channel direction and then processed. Table 3 shows hyper-parameters of each layer, where Conv. and Pool. denote the convolution and the pooling, and Ch. rate means the multiplication factor from the number of inputs to the number of outputs, respectively. Conv. A is used at the beginning of each layer, Conv. B is used in the encoder side (left-half side in Fig. 4), and Conv. C is used in the rest. Note that in the SE-block, the number of input and output channels is the same. All layers and blocks use the Swish [36] as activation function followed by batch normalization [37].

Thanks to the architecture, the proposed network has high-resolution representations for space and channel direction and produces a map with accurate localization. U-Net is one of the most efficient and well-known structures for image segmentation and can increase the spatial resolution of features without degrading the localization accuracy [32], [33]. The proposed network is constructed based on U-Net because the above property is also necessary for the detection of reflections. We presume that features of C are produced from a different domain of I, and therefore different encoder networks are applied for feature extraction, as shown in the left half of Fig. 4. Moreover, to improve the network representation of the proposed network, SE-block is introduced into features of C at each resolution because SE-block improves the network representation via channel-wise recalibrating features at low computational cost [35].

2) Loss Function

For training, the loss function of the proposed network L is defined as the sum of the cross entropy loss \( L_{\text{bin}} \) and the focal loss \( L_{\text{FP}} \) [38], which is shown as follows:

\[
L = L_{\text{bin}} + \lambda L_{\text{FP}},
\]

where \( \lambda \) is a balancing parameter, \( L_{\text{bin}} \) and \( L_{\text{FP}} \) are defined as

\[
L_{\text{bin}} = \sum_{i=0} \left[-g_i(1 - o_i)^{\gamma} \log(o_i) - (1 - g_i) \alpha_i^{\gamma} \log(1 - o_i)\right],
\]

\[
L_{\text{FP}} = \sum_{i=0} \left[-\hat{m}_i o_i^{\gamma} \log(1 - o_i)\right],
\]

g_i, o_i, and \( \hat{m}_i \) are the \( i \)-th elements of GT, an output of the proposed network, and a mask calculated by eliminating background areas from \( m \) to have only areas of saturated reflections, and \( \gamma \) is the focusing parameter of the focal loss, respectively. \( L_{\text{bin}} \) calculates the pixel-wise difference between \( g_i \) and \( o_i \). \( L_{\text{FP}} \) is introduced to the proposed model to reduce false positives, i.e. detecting light sources of background as reflections. The focal loss is applied to improve the learning speed.

IV. SYSTEM FOR REFLECTION REMOVAL USING THE PROPOSED METHOD

To remove reflections with saturated reflections, we propose a system consisting of previous methods and the proposed method, as shown in Section III. The overview of the proposed system is shown in Fig. 5, where I, \( \hat{I} \), and \( \hat{I} \) denote an input image, the pre-removed image, and the final removed image, respectively. The proposed method is applied to \( \hat{I} \) for detecting saturated reflections and a state-of-the-art method of reflection removal is also applied to remove these reflections without saturated reflections. Saturated reflections of \( \hat{I} \) are removed via estimating pixel values of backgrounds and \( D \) is used as the mask to indicate these areas. We understand that inpainting is one of the most suitable techniques for this estimation, and its method is used in the experiments of this paper. The proposed system can be straightforward improved by developing each technique, and thus encourages that each technique can be studied separately.

| Layer type | Kernel size | Stride | Padding | Ch. rate |
|------------|-------------|--------|---------|----------|
| Conv. A    | 5 × 5       | 1      | 2       | 1        |
| Conv. B    | 5 × 5       | 1      | 2       | 2        |
| Conv. C    | 5 × 5       | 1      | 2       | 1/3      |
| Max Pool.  | 2 × 2       | 2      | 0       | 1        |
| Transposed Conv. | 4 × 4 | 2      | 1       | 1/2      |
γ and λ the loss function explained in Section III-C2, we set every 25. The learning rate is first set to 640.

The momentum is a hyper-parameter: Momentum SGD is used as an optimization method [37]. The momentum is 0.9 and the batch size is 64. The learning rate is first set to 0.001 and multiplied by 0.1 every 25 epochs. The model was trained for 100 epochs. For the loss function explained in Section III-C2, we set λ = 20 and γ = 3.0.

### V. EXPERIMENTS

#### A. TRAINING OF PROPOSED CNN

To train the proposed network shown in Section III-C, we artificially created 56000 images that have light sources of backgrounds and saturated reflections. The Places 365 dataset and high dynamic range (HDR) images [39]–[42] were used as backgrounds and reflections, respectively. To create artificial reflections R, HDR images were converted into low dynamic range (LDR) by a gamma correction [34], and blurred by Gaussian kernels with zero mean and random values of variance in [0, 2] [43]. A natural image and R were added directly and the resultant image X was cropped to have a size of 256 × 256. The GT is a binary map G that indicates areas of saturated reflections in X, and it was calculated as

\[ g_i = \begin{cases} 
1 & \text{if } x_i > 0.95 \text{ and } r_i > 0.75 \\
0 & \text{otherwise}
\end{cases}, \quad (5) \]

where \( g_i \) is the \( i \)-th element of \( G \), \( x_i \) and \( r_i \) are the \( i \)-th elements of luminance values of \( X \) and \( R \) in YCbCr color space, respectively, and pixel values are in [0, 1]. The number of images in the training set was 25000. Random horizontal and vertical flipping were applied to the training set.

The proposed CNN model was trained with following hyper-parameters: Momentum SGD is used as an optimization method [37]. The momentum is 0.9 and the batch size is 64. The learning rate is first set to 0.001 and multiplied by 0.1 every 25 epochs. The model was trained for 100 epochs. For the loss function explained in Section III-C2, we set \( \lambda = 20 \) and \( \gamma = 3.0 \).

#### B. EVALUATION OF SATURATED REFLECTION DETECTION

For the evaluation of the saturated reflection detection, the proposed method was compared with humans as follows: The proposed method was applied to the same 30 images in the experiment with humans shown in Section II-C. Although the proposed method produces a pixel-wise binary map, true and false results are counted area-wise. If more than half of the pixels in a candidate area have 1 in \( D \) without the dilation, we define that the area is recognized as a saturated reflection by the proposed method. Otherwise, it is recognized as an object of the background.

Table 4 shows the detection scores of humans and the proposed method, where Prop. means the proposed method. From Table 4, scores of humans are greatly higher than the proposed method, although humans were not asked to identify the position of the areas. Table 4 shows one of the criteria for this theme and the proposed method needs improvements to achieve more competitive scores. Specifically, since there is the most difference in precision, reducing false detections is an area for improvement.

To visually explain the performance of the proposed method, an example of its results is shown in Fig. 6, where white pixels in blue and green boxes of Fig. 6(b) are light sources of background and the saturated reflections, respectively. Fig. 6 has all cases of detection, true and false positives, true negatives, and false negatives. Comparing Fig. 6(b) and (c), false positives, false negatives, and true negatives are shown in the red, green, and blue boxes, respectively. The proposed method detects saturated reflections and avoids background objects, and the proposed system produces natural images by removing saturated reflections as shown in Fig. 1. However, since the proposed method has lower accuracy compared to ideal detection, it often detects background objects as reflection and misses saturated reflections. Specifically, as mentioned above, reducing false detections is necessary because it leads to the elimination of background objects.

#### C. EVALUATION OF REFLECTION REMOVAL

The proposed system is compared with state-of-the-art methods of reflection removal in this section. The methods [16], [19], [23], [25], [29] are used for comparison, which we refer...
This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

T. Yoshida, et al.: Saturated Reflection Detection for Reflection Removal Based on CNN

FIGURE 7. Input, GT, resultant images of reflection removal.

TABLE 5. PSNR [dB], SSIM, and LPIPS scores that are mean values of 20 images.

|                | GCNet [16] Only | +Prop. | ERRNet [19] Only | +Prop. | IBCLN [23] Only | +Prop. | Kim et al. [25] Only | +Prop. | LRMNet [29] Only | +Prop. |
|----------------|-----------------|-------|------------------|-------|-----------------|-------|----------------------|-------|-------------------|-------|
| PSNR ↑         | 20.14           | 20.32 | 23.43            | 23.55 | 20.25           | 20.35 | 20.56                | 20.88 | 22.19             | 22.56 |
| SSIM ↑         | 0.817           | 0.819 | 0.854            | 0.850 | 0.758           | 0.758 | 0.806                | 0.803 | 0.845             | 0.845 |
| LPIPS ↓        | 0.160           | 0.157 | 0.108            | 0.106 | 0.174           | 0.173 | 0.203                | 0.201 | 0.134             | 0.132 |

to as GCNet [16], ERRNet [19], IBCLN [23], Kim et al. [25], and LRMNet [29] in this paper. Since the proposed system uses a method of reflection removal, each method is compared with the proposed system that uses itself. For this experiment, the proposed system uses a method of inpainting [44] as the Estimation method shown in Fig. 5. 20 sets of images and their GTs are used from datasets of reflection removal [3], [5], [15], [19], and their images have saturated reflections and light sources of background. The peak signal-to-noise ratio (PSNR), structural similarity index measure (SSIM), and learned perceptual image patch similarity (LPIPS) is used as objective measurement [43], [45].

Table 5 shows PSNR, SSIM, and LPIPS scores that are mean values of 20 images, where “Only” means applying only the method to images, whereas “+Prop.” means using the proposed system with it. For each method, the proposed system produces better and comparable images. Unfortunately, the difference is only slight, because the number of pixels in saturated reflections is very small compared to the whole image. Thus, the improvement only slightly influences scores.

Fig. 7-10 show resultant images of reflection removal by the conventional methods and the proposed system. Fig. 9 and 10 show the enlarged images in the area bounded by the red box of corresponding images in Fig. 7 and 8, respectively. These figures show that the conventional methods fail to remove saturated reflections while the proposed system can remove these reflections without degrading the image quality. However, some saturated reflections still remain in the output.
VI. CONCLUSION

In this paper, we discussed the detection and removal of saturated reflections, proposed a detection method based on CNN with several signal processing techniques, and introduced a removal system based on the proposed detection method and conventional removal methods. First, we discussed the definition, characteristics, and removal procedure of saturated reflections, and showed experimental results of humans in recognizing saturated reflections. The proposed method detects candidate areas of saturated reflections by thresholding, and classifies them into reflections and backgrounds using the proposed CNN with high-frequency components of images. The proposed system detects areas of saturated reflections with the proposed method and restores them using a conventional method of image restoration. In experiments using two conventional methods of reflection removal, it is shown that the proposed system has better PSNR scores than the original method. Moreover, the removal of saturated reflections by the proposed system is perceptually shown.

Unfortunately, the proposed method has lower scores of precision and recall in detection than humans. In future work, we hope to improve the proposed method by reducing false detections and achieve similar detection scores to humans.

REFERENCES

[1] A. Levin and Y. Weiss, “User assisted separation of reflections from a single image using a sparsity prior,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 9, pp. 1647–1654, Sep. 2007.
[2] Y. Diamant and Y. Y. Schechner, “Overcoming visual reverberations,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2008, pp. 1–8.
[3] T. Xue, M. Rubinstein, C. Liu, and W. T. Freeman, “A computational approach for obstruction-free photography,” ACM Trans. Graphics, vol. 34, no. 4, pp. 1–11, July 2015.
[4] T. Sirinukulwattana, G. Choe, and I. S. Kweon, “Reflection removal using disparity and gradient-sparsity via smoothing algorithm,” in Proc. IEEE Int. Conf. Image Process., 2015, pp. 1940–1944.
[5] R. Wan, B. Shi, T. A. Hwee, and A. C. Kot, “Depth of field guided reflection removal,” in Proc. IEEE Int. Conf. Image Process., 2016, pp. 21–25.
[6] B. Han and J. Sim, “Reflection removal using low-rank matrix completion,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2017, pp. 3872–3880.
[7] N. Arvanitopoulos, R. Achanta, and S. Süsstrunk, “Single image reflection suppression,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2017, pp. 1752–1760.
[8] R. Wan, B. Shi, L. Duan, A. Tan, W. Gao, and A. C. Kot, “Region-aware reflection removal with unified content and gradient priors,” IEEE Trans. Image Process., vol. 27, no. 6, pp. 2927–2941, June 2018.
FIGURE 9. Enlarged images in red boxes corresponding to Fig. 7.

[9] B.-J. Han and J.-Y. Sim, “Glass reflection removal using co-saliency-based image alignment and low-rank matrix completion in gradient domain,” IEEE Trans. Image Process., vol. 27, no. 10, pp. 4873–4888, Oct. 2018.

[10] Y. Ni, J. Chen, and L.-P. Chau, “Reflection removal on single light field capture using focus manipulation,” IEEE Trans. Comput. Imag., vol. 4, no. 4, pp. 362–572, Dec. 2018.

[11] Y. Yang, W. Ma, Y. Zheng, J.-F. Cai, and W. Xu, “Fast single image reflection suppression via convex optimization,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2019, pp. 8133–8141.

[12] Q. Fan, J. Yang, G. Hua, B. Chen, and D. Wipf, “A generic deep architecture for single image reflection removal and image smoothing,” in Proc. IEEE Int. Conf. Comput. Vis., 2017, pp. 3258–3267.

[13] J. Yang, D. Gong, L. Liu, and Q. Shi, “Seeing deeply and bidirectionally: A deep learning approach for single image reflection removal,” in Proc. European Conf. Comput. Vis., 2018, pp. 675–691.

[14] R. Wan, B. Shi, L.-Y. Duan, A.-H. Tan, and A. C. Kot, “CRRN: Multi-scale guided concurrent reflection removal network,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2018, pp. 4777–4785.

[15] X. Zhang, R. Ng, and Q. Chen, “Single image reflection separation with perceptual losses,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2018, pp. 4786–4794.

[16] R. Abiko and M. M. Ikehara, “Single image reflection removal based on GAN with gradient constraint,” IEEE Access, vol. 7, pp. 148 790–148 799, 2019.

[17] Y. Chang and C. Jung, “Single image reflection removal using convolutional neural networks,” IEEE Trans. Image Process., vol. 28, no. 4, pp. 1954–1966, Apr. 2019.

[18] T. Li and D. P. K. Lun, “Single-image reflection removal via a two-stage background recovery process,” IEEE Signal Process. Lett., vol. 26, no. 8, pp. 1237–1241, 2019.

[19] K. Wei, J. Yang, Y. Fu, D. Wipf, and H. Huang, “Single image reflection removal exploiting misaligned training data and network enhancements,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2019, pp. 8170–8179.

[20] Q. Wen, Y. Tan, J. Qin, W. Liu, G. Han, and S. He, “Single image reflection removal beyond linearity,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2019, pp. 3766–3774.

[21] T. Li and D. P. K. Lun, “Image reflection removal using the wasserstein generative adversarial network,” in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., 2019, pp. 1–5.

[22] M. Heo and Y. Choe, “Single-image reflection removal using conditional GANs,” in Proc. Int. Conf. Electron., Inf., Commun., 2019, pp. 1–4.

[23] J. Li, G. Li, and H. Fan, “Image reflection removal using end-to-end convolutional neural network,” IET Image Process., vol. 14, no. 6, pp. 1047–1058, Apr. 2020.

[24] R. Wan, B. Shi, H. Li, L.-Y. Duan, A.-H. Tan, and A. C. Kot, “CoRNN: Cooperative reflection removal network,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 42, no. 12, pp. 2969–2982, 2020.
[25] S. Kim, Y. Huo, and S.-E. Yoon, “Single image reflection removal with physically-based training images,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2020, pp. 5163–5172.

[26] C. Lei, X. Huang, M. Zhang, Q. Yan, W. Sun, and Q. Chen, “Polarized reflection removal with perfect alignment in the wild,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2020, pp. 1747–1755.

[27] C. Li, Y. Yang, K. He, S. Lin, and J. E. Hopcroft, “Single image reflection removal through cascaded refinement,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2020, pp. 3562–3571.

[28] Y. Li, Q. Yan, K. Zhang, and H. Xu, “Single image reflection removal using convolutional neural networks,” IEEE Trans. Circuits Syst. Video Technol., 2021.

[29] Z. Dong, K. Xu, Y. Yang, H. Bao, W. Xu, and R. W. H. Lau, “Location-aware single image reflection removal,” in Proc. IEEE/CVF Int. Conf. Comput. Vis., 2021, pp. 4997–5006.

[30] Q. Zheng, B. Shi, J. Chen, X. Jiang, L.-Y. Duan, and A. C. Kot, “Single image reflection removal with absorption effect,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2021, pp. 13 390–13 399.

[31] Y. Li, Q. Yan, K. Zhang, and H. Xu, “Image reflection removal via contextual feature fusion pyramid and task-driven regularization,” IEEE Trans. Circuits Syst. Video Technol., vol. 32, no. 2, pp. 553–565, 2022.

[32] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in Proc. Medical Image Comput. Computer-Assisted Intervention, 2015, pp. 234–241.

[33] M. H. Hesamian, W. Jia, X. He, and P. Kennedy, “Deep learning techniques for medical image segmentation: Achievements and challenges,” J. Digit. Imag., vol. 32, no. 4, pp. 582–596, Aug. 2019.

[34] R. Szegedy, Computer Vision: Algorithms and Applications, Springer, 2010.

[35] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, “Squeeze-and-excitation networks,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 42, no. 8, pp. 2011–2023, Aug. 2020.

[36] P. Ramachandran, B. Zoph, and Q. V. Le, “Searching for activation functions,” in Proc. Int. Conf. Learning Representations, 2018.

[37] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016.

[38] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal loss for dense object detection,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 42, no. 2, pp. 318–327, Feb. 2020.

[39] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba, “Places: A 10 million image database for scene recognition,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 40, no. 6, pp. 1452–1464, June 2018.

[40] P. E. Debevec and J. Malik, “Recovering high dynamic range radiance maps from photographs,” in Proc. Annual Conf. Comput. Graphics Interactive Tech., 1997, pp. 369–378.

[41] S. Paris, S. W. Hasinoff, and J. Kautz, “Local laplacian filters: Edge-aware image processing with a laplacian pyramid,” Commun. ACM, vol. 58, no. 3, pp. 81–91, Feb. 2015.
“EMPA HDR images dataset,” this dataset is unavailable now. [Online]. Available: http://empamedia.ethz.ch/hdrdatabase/index.php.

R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Prentice Hall, 2002.

J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. Huang, “Free-form image inpainting with gated convolution,” in *Proc. IEEE/CVF Intl. Conf. Comput. Vis.*, 2019, pp. 4470–4479.

R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, “The unreasonable effectiveness of deep features as a perceptual metric,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 586–595.

TAICHI YOSHIDA received the B.Eng., M.Eng., and Ph.D. degrees in Engineering from Keio University, Yokohama, Japan, in 2006, 2008 and 2013, respectively. In 2014, he joined Nagaoka University of Technology. In 2018, he joined the University of Electro-Communications, where he is currently an Assistant Professor with the Department of Computer and Network Engineering. His research interests include multi-rate signal processing, image processing and computer vision.

ISANA FUNAHASHI received the B.Eng. and M.Eng. degrees from Nagaoka University of Technology, Nagaoka, Japan, in 2017 and 2019, respectively. He is currently pursuing the Ph.D. degree with the Department of Computer and Network Engineering, The University of Electro-Communications, Tokyo, Japan. His research interests include image processing and computer vision.

NAOKI YAMASHITA received the B.Eng. and M.Eng. degrees from The University of Electro-Communications, Tokyo, Japan, in 2019 and 2021, respectively. His research interests include computer vision.

MASAAKI IKEHARA received the B.E., M.E., and Dr. Eng. degrees in electrical engineering from Keio University, Yokohama, Japan, in 1984, 1986, and 1989, respectively. He was appointed as a Lecturer at Nagasaki University, Nagasaki, Japan, from 1989 to 1992. In 1992, he joined the Faculty of Engineering, Keio University, where he is currently a Full Professor with the Department of Electronics and Electrical Engineering. From 1996 to 1998, he was a Visiting Researcher with the University of Wisconsin–Madison and Boston University, Boston, MA, USA. His research interests include multi-rate signal processing, wavelet image coding, and filter design problems.