High Reflection Removal Using CNN with Detection and Estimation

Isana Funahashi*, Naoki Yamashita*, Taichi Yoshida*, and Masaaki Ikehara†
*University of Electro-Communications, Dept. of Computer and Network Engineering, Tokyo Japan
†Keio University, Dept. of Electronics and Electrical Engineering, Kanagawa Japan

***Abstract*—In this paper, we propose a method of reflection removal that reduces high intensity reflection for single image. Various methods of reflection removal have been proposed, but they fail to reduce the high reflections due to their assumption. To tackle this issue, the proposed method detects the target areas with high reflections by the proposed convolutional neural network (CNN) model and estimates their background information by inpainting. It is observed that the reflection is strongly blurred because of its physical property, and hence the proposed CNN model utilizes edge features of pixels for the detection. In simulation, we compare state-of-the-art methods of reflection removal with and without the proposed method for natural images, and the proposed method improves peak signal-to-noise ratio and perceptual quality.**

***Index Terms*—Single image reflection removal, convolutional neural network, image inpainting**

I. INTRODUCTION

To reduce undesirable reflections in images due to glass, various methods of reflection removal have been proposed [1]–[18]. When we take a picture through glass, the resultant image often captures undesirable objects behind us by reflection. It causes negative effects degrading not only the image quality but also accuracy in computer vision techniques such as image recognition, object detection, and so on. To avoid it, methods that use the optimization technique have been proposed [1]–[6], [8]–[10], [14], [19]. Wan et al. have proposed a method that uses the structural information around reflection areas [8], and Arvanitopoulos et al. have proposed a method that uses the Laplacian and L_0 smoothing [6]. Recently, machine-learning-based methods have been actively studied [7], [11]–[13], [15]–[18]. Fan et al. have proposed a first method using convolutional neural networks (CNNs) [7], and Wei et al. have proposed a method that can be trained with an unaligned dataset and produces perceptually efficient results [17].

Unfortunately, it is difficult for previous methods to detect and remove reflections with high intensity, for example reflections of light sources as shown in Fig. 1. Fig. 1 (a) is an input image, which has the reflections inside the red square, and (b) and (c) respectively show the removed results by the state-of-the-art [17] and the proposed method. Note that areas of the reflections are called high reflection in this paper. Fig. 1 (b) shows that the conventional method fails to remove the high reflection. The conventional methods assume that an image with reflections is defined as the linear combination of reflection and background layers, which is discussed in detail in Sec. II. Since the assumption does not hold for high reflection due to saturation of pixel values, a different approach is necessary for removing them.

(a) Input (b) [17] (c) Proposed

Fig. 1. Results of reflection removal.

In this paper, we introduce the different approach and propose a method removing high reflection by detecting high reflection and estimating their information in the background layer. For the detection, we realize a CNN model using edge features, and use methods of image inpainting for the estimation. The CNN model is based on U-net, and extracts multi-resolution features from RGB values and edge features. In experiments for real images, the proposed method shows higher scores of peak signal-to-noise ratio (PSNR) and perceptually superior images compared with state-of-the-art methods.

Contributions of this paper are summarized as follows:

- This study introduces a new problem, high reflection removal, in reflection removal and achieves clearer images via tackling it.
- This study shows the possibility of edge features for detecting high reflection and an effective method for removing them.

II. DISCUSSION

The hypothesis of images with reflection in conventional methods is not valid for high reflection, and hence they are slightly effective for removing them. Images with reflections are defined as the linear combination of background and reflection layers B and R [5], [7], [8], [10], [13]–[15], [18]. However, this definition is not valid for high reflection, because components at there in R, which are very high intensities, induce saturation in images and they erase the information in B. Conventional methods use the definition as one of the constraints in optimization and incorporate it into the training of CNN models, and hence they fail to remove high reflections.

To tackle this issue, we introduce a procedure of reflection removal targeting not only normal reflections, which obey the definition, but also high reflection. Since their physical

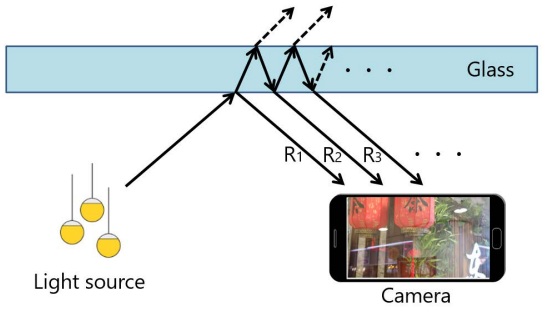


Fig. 2. Multi reflections from one high intensity source.

properties are quite different as discussed above, the procedure consists of two methods in parallel. One is to remove normal reflections, and the knowledge of previous methods is used. Another is to tackle high reflection, the detection and estimation for them are necessary. Since saturated pixels of images with reflection, which have high intensity values, are caused by, it is necessary that Saturated pixels of images with reflection, which have high intensity values, are caused by not only reflections but also light in the background. Therefore, it is necessary that saturated pixels caused by only reflections, i.e. areas of high reflection, are detected from them. Since true information of high reflection areas in B is lost due to saturation, it should be estimated from the information of neighboring pixels after removing reflection by conventional methods. Details of the procedure are explained in the next section.

To detect high reflection, we propose the CNN model and utilize edge features of pixels. Components in R are blurred due to reflection and refraction at glass, which is also mentioned in previous papers [15], [20], and this phenomenon is particularly valid for high reflection. Since light sources have high intensity and remain through multiple reflections, a camera captures multiple reflections of one light source as high reflection, which is simply shown in Fig. 2. Therefore, components at high reflections in R are more strongly blurred compared with normal reflections. Edge features of pixels are calculated and results are used by the proposed CNN model as input.

III. PROPOSED METHOD

A. Overview

Fig. 3 shows the overview of the proposed method. First, the proposed CNN model is applied to an input image, and its high reflections are detected. Second, in the morphology process, the 3×3 erosion and the 10×10 dilation are applied to the detection result to erase its outliers and smooth its residual areas [21]. Finally, the information of B in areas of high reflections is estimated by applying image inpainting into the detected them of reflection removed image by the conventional method.

B. Proposed CNN model

The proposed CNN model detects high reflections of an input image using its RGB values, saturation mask, and edge information. The saturation mask m is calculated as

$$m_k = \begin{cases} 1 & \text{if } x_k > \Delta \\ 0 & otherwise \end{cases}, \tag{1}$$

where m_k and x_k are the k-th values of m and intensity of input images, respectively. We set $\Delta = 245$ in this paper, empirically. Since pixels with $m_k = 1$ denote areas of high reflection and light in the background, the proposed CNN model is realized to detect only high reflection from them via learning. The edge information is obtained by four types of high-pass filters, which are 4 neighborhood 3×3 Laplacian-filter, 8 neighborhood 3×3 Laplacian-filter, 5×5 Laplacian-filter, and 3×3 Sobel filter.

The architecture of the proposed CNN model is based on U-Net shown in Fig. 4, where values denote the number of channels. U-Net is one of CNN architectures that show the highest performance in image segmentation [22], [23]. The architecture of the proposed CNN model consists of two encoder networks and one decoder network. Each encoder network has its own input layers. One uses RGB values of the input image and the saturation mask m_k, and another uses results of the filtering as edge information. The encoder network for the edge information combines convolutional layers with SE-blocks. The SE-block is an additional structure of networks that adds weight to each channel of feature maps [24]. By combining the SE-block, the importance of each channel is learned and hence the learning efficiency of models can be improved [24]. Therefore, the proposed model learns the importance of each edge information and improves the detection accuracy. The decoder network detects high reflections using features extracted by the two encoders. The proposed model uses the swish as activation functions, the batch normalization after each activation function, and 5×5 kernels for each convolutional layer [25], [26].

C. Loss function of proposed CNN model

For training of the proposed CNN model, the loss function L_{all} is a binary cross entropy loss with the focal-loss as,

$$L_{\text{all}} = L_{\text{bin}} + 20 \times L_{\text{satu}}, \tag{2}$$

$$\begin{aligned} L_{\text{bin}} = \sum_{k=0} [& -g_k (1-o_k)^{\gamma} \log(o_k) \\ & - (1-g_k)(o_k)^{\gamma} \log(1-o_k)], \\ L_{\text{satu}} = \sum_{k=0} [& -b_k (o_k)^{\gamma} \log(1-o_k)], \end{aligned}$$

where g_k, o_k, and b_k are k-th values of ground truth (GT), the output of the model, and mask that is the same as m with substituting B by the input image, respectively. L_{bin} is a binary cross entropy, which calculates pixel-wise errors between g_k and o_k. L_{satu} is a loss function comparing o_k and b_k. L_{satu} imparts an additional loss to the model for reducing false positives of saturation areas in B. Furthermore,

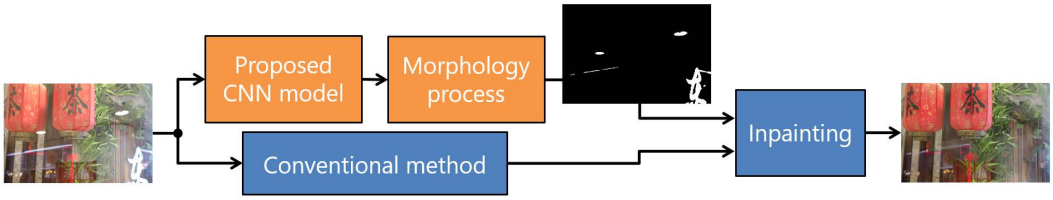


Fig. 3. Overview of proposed method.

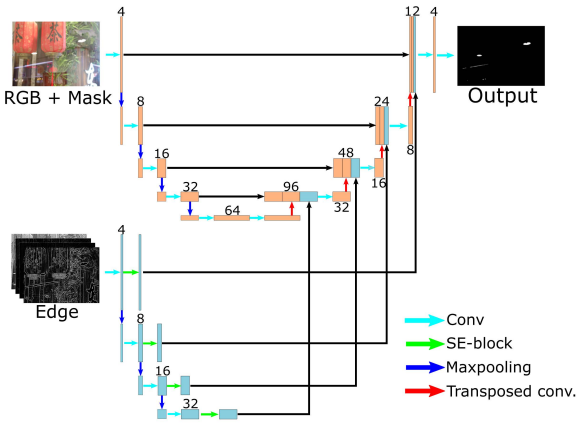


Fig. 4. Architecture of proposed CNN model.

we applied the focal-loss to L_{bin} and L_{satu} for improving learning speed. The focal-loss is expressed by the following formula.

$$FL = -(1-p)^{\gamma}\log(-p), \tag{3}$$

where p and γ are the probability of the model's estimation and the focusing parameter, respectively.

IV. SIMULATION

A. Training of proposed CNN model

56482 images that have high reflection are created artificially for training. The Places 365 dataset and high dynamic range (HDR) images [27]–[30] are used as background and reflection layers, respectively. As pre-processing, HDR images are converted into low dynamic range (LDR), and blurred by Gaussian kernel with random variance values in $[0, 2]$. For training, 256×256 images that have high intensity areas are cropped from the composited images. The number of training sets becomes 256761. Random horizontal flipping and vertical flipping are applied to the training sets.

The proposed CNN model is trained with the following conditions: Momentum SGD is used as the optimization method. The momentum is 0.9 and the batch size is 64. The learning late is firstly set to 0.001 and multiplied by 0.1 for every 25 epochs. The model has trained 100 epochs. It has taken about 38 hours with two Nvidia(R) Geforce(R) GTX 1080Ti. The γ for the loss function is set to 3.0.

TABLE I
PSNR SCORES OF SIMULATION.

	[17]		[18]	
	Only	+proposed	Only	+proposed
Image 1	13.22	**13.32**	12.98	**13.05**
Image 2	15.58	**15.59**	15.22	**15.32**
Ave. of 23 sets	21.83	**21.94**	18.93	**19.09**

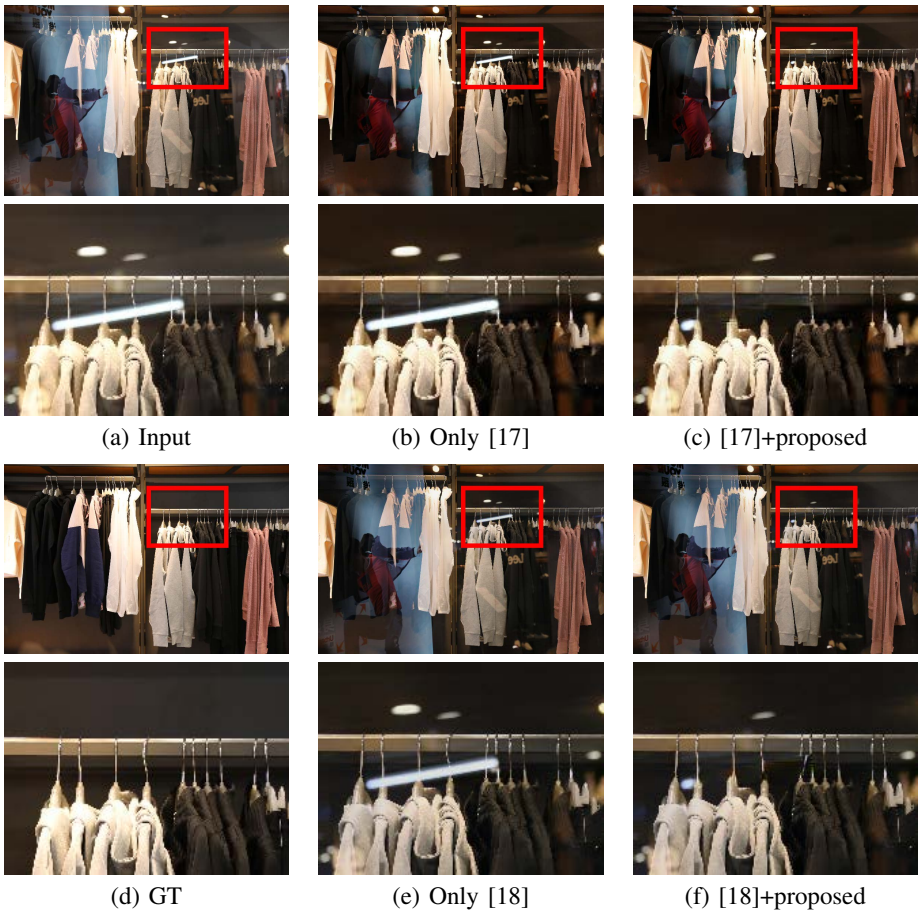
B. Evaluation of proposed method

To verify the effectiveness of the proposed method, it is compared with state-of-the-art methods for real image datasets. The inpainting method of Yu et al. is used in this experiment [31]. For a test dataset, 23 sets of images that have high reflection are selected from four datasets [3], [5], [13]. The peak signal-to-noise ratio (PSNR) is used as an objective metric. As compared methods, two state-of-the-art methods are used [17], [18].

Fig. 5, Fig. 6, and Table I show some of the resultant images and PSNR, respectively. Fig. 5 and 6 (a)-(f) show input images, removal results by only the method [17], the method [17] with the proposed one, only the method [18], the method [18] with the proposed one, and GT images. The "only" and "+proposed" means results of conventional methods and with the proposed method as additional processing, and "Ave." denotes mean PSNR of 23 sets, respectively. From the figures, in conventional methods, it is observed that high reflection still remains, whereas the proposed method can remove their reflections and generates visually better results. From the table, it is observed that the proposed method improves PSNR. This experiment shows that the proposed method is perceptually and objectively effective for reflection removal of images that have high reflections.

V. CONCLUTION

In this paper, we proposed a method that removes reflections in high reflection. The proposed method is additionally applied to conventional methods of reflection removal for tackling high reflection. The proposed method detects areas of the reflection by the proposed CNN model that utilizes edge features and estimates B information in there by an image inpainting method. Through simulation, the proposed method detects and restores high reflection, and improves results of conventional



(a) Input (b) Only [17] (c) [17]+proposed

(d) GT (e) Only [18] (f) [18]+proposed

Fig. 5. Results of reflection removal for Image 1.

methods for images that have the reflections, perceptually and objectively.

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] C. Simon and I. K. Park, "Reflection removal for in-vehicle black box videos," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 4231–4239.

[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] N. Arvanitopoulos, R. Achanta, and S. Süsstrunk, "Single image reflection suppression," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1752–1760.

[7] 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.

[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.

[9] B. Han and J. 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. Chau, "Reflection removal on single light field capture using focus manipulation," *IEEE Trans. Computational Imaging*, vol. 4, no. 4, pp. 562–572, Dec. 2018.

[11] 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.

[12] R. Wan, B. Shi, L. Duan, A. 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.

[13] 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.

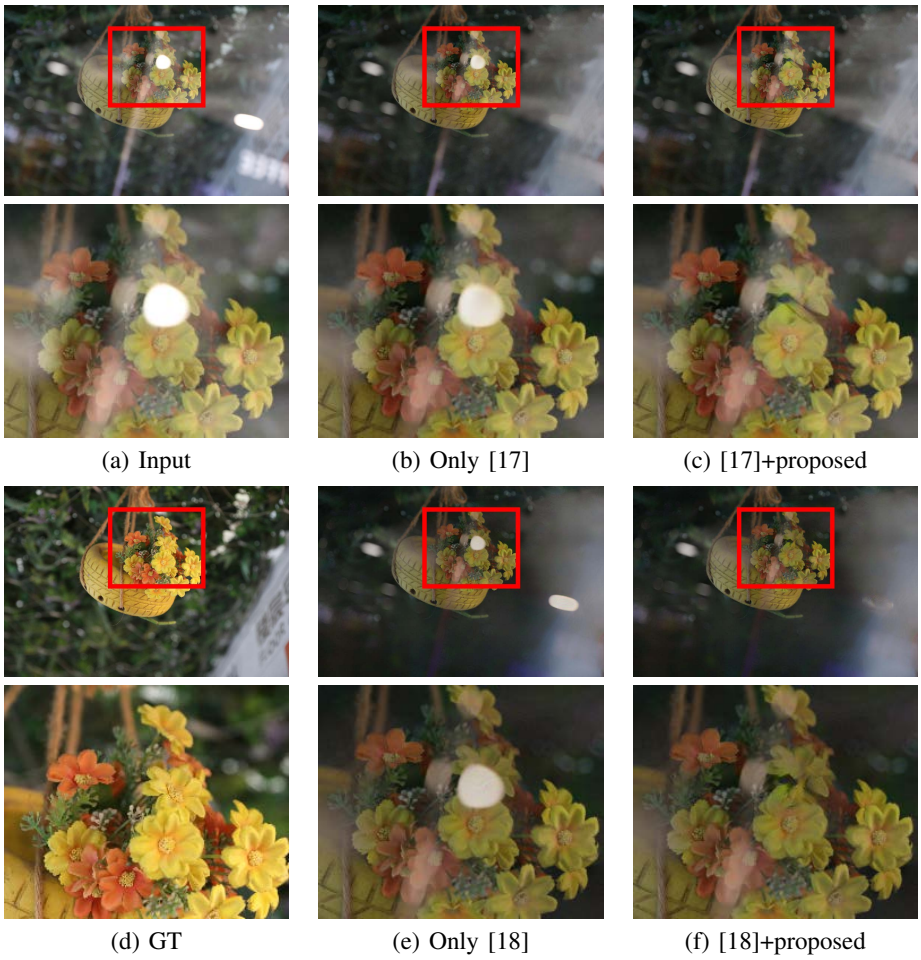
[14] Y. Yang, W. Ma, Y. Zheng, J. Cai, and W. Xu, "Fast single image reflection suppression via convex optimization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 8133–8141.

[15] 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.

[16] 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.

[17] 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.

[18] R. Abiko and M. Ikehara, "Single image reflection removal based



(a) Input (b) Only [17] (c) [17]+proposed

(d) GT (e) Only [18] (f) [18]+proposed

Fig. 6. Results of reflection removal for Image 2.

on GAN with gradient constraint," *IEEE Access*, vol. 7, pp. 148790–148799, 2019.

[19] B. Han and J. Sim, "Reflection removal using low-rank matrix completion," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 3872–3880.

[20] Y. Diamant and Y. Y. Schechner, "Overcoming visual reverberations," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2008, pp. 1–8.

[21] R. Szeliski, *Computer Vision: Algorithms and Applications*, Springer, 2010.

[22] 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.

[23] M. H. Hesamian, W. Jia, X. He, and P. Kennedy, "Deep learning techniques for medical image segmentation: Achievements and challenges," *J. Digital Imaging*, vol. 32, no. 4, pp. 582–596, Aug. 2019.

[24] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141.

[25] R. Prajit, Z. Barret, and V. L. Quoc, "Searching for activation functions," in *Proc. Int. Conf. Learning Representations*, 2018.

[26] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.

[27] 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.

[28] "EMPA HDR images dataset," this dataset is unavailable now. [Online]. Available: http://empamedia.ethz.ch/hdrdatabase/index.php.

[29] 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.

[30] 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.

[31] 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.*, Oct. 2019, pp. 4470–4479.