Deep Snapshot HDR Imaging
Using Multi-Exposure Color Filter Array

Takeru Suda, Masayuki Tanaka, Yusuke Monno, and Masatoshi Okutomi

Tokyo Institute of Technology, Tokyo, Japan
tsuda@ok.sc.e.titech.ac.jp; {mtanaka,ymonno,mxo}@sc.e.titech.ac.jp

Abstract. In this paper, we propose a deep snapshot high dynamic range (HDR) imaging framework that can effectively reconstruct an HDR image from the RAW data captured using a multi-exposure color filter array (ME-CFA), which consists of a mosaic pattern of RGB filters with different exposure levels. To effectively learn the HDR image reconstruction network, we introduce the idea of luminance normalization that simultaneously enables effective loss computation and input data normalization by considering relative local contrasts in the "normalized-by-luminance" HDR domain. This idea makes it possible to equally handle the errors in both bright and dark areas regardless of absolute luminance levels, which significantly improves the visual image quality in a tone-mapped domain. Experimental results using two public HDR image datasets demonstrate that our framework outperforms other snapshot methods and produces high-quality HDR images with fewer visual artifacts.

1   Introduction

The dynamic range of a camera is determined by the ratio between the maximum and the minimum amounts of light that can be recorded by the image sensor at one shot. Standard digital cameras have a low dynamic range (LDR) and only capture a limited range of scene radiance. Consequently, they cannot capture a bright and a dark area outside the camera's dynamic range simultaneously. High dynamic range (HDR) imaging is a highly demanded computational imaging technique to overcome this limitation, which recovers the HDR scene radiance map from a single or multiple LDR images captured by a standard camera.

HDR imaging is typically performed by estimating a mapping from the sensor's LDR outputs to the scene radiance using multiple LDR images which are sequentially captured with different exposure levels [1]. Although this approach works for static situations, it is not suitable for dynamic scenes and video acquisition since the multiple images are taken at different times. Recent learning-based methods [2–4] have successfully reduced ghost artifacts derived from target motions between input LDR images. However, those methods are limited to small motions and the artifacts remain apparent for the areas with large motions, as shown in Wu's method [3] of Fig. 1.

Some studies have used only a single LDR image to realize one-shot HDR imaging [5–7]. They essentially inpaint or hallucinate missing over- and under-exposed areas by exploiting an external database of LDR-HDR image pairs.
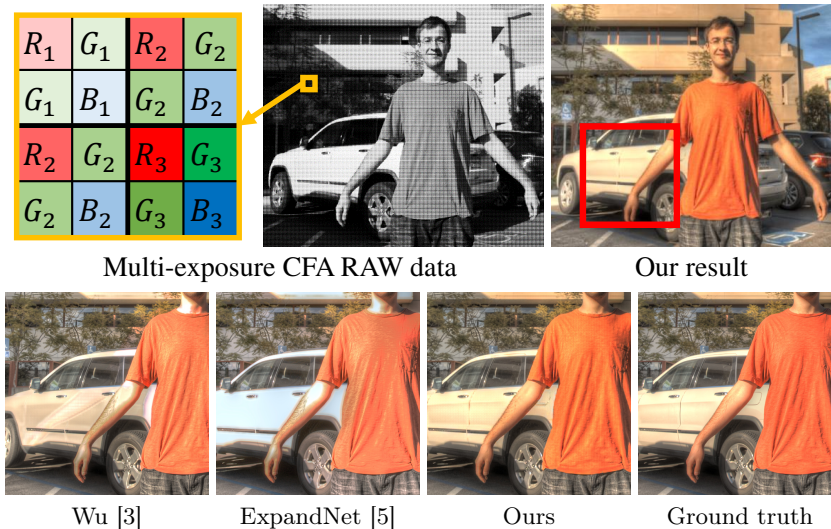
<div align="center">Multi-exposure CFA RAW data   Our result</div>

<div align="center">Wu [3]   ExpandNet [5]   Ours   Ground truth</div>

Fig. 1. Top: Examples of a multi-exposure CFA RAW data and an HDR image result by our framework. Bottom: Comparisons with state-of-the-art HDR imaging methods.

Although this approach is free from ghost artifacts, it generates inpainting or hallucination artifacts for largely missing areas, as shown in ExpandNet [5] of Fig. 1, due to the lack of information in the missing areas.

As another one-shot approach, the methods based on a snapshot HDR sensor have also been investigated [8–11]. One way to realize a snapshot HDR sensor is to use a single image sensor with spatially varying exposure levels [10, 11]. This can be achieved by using what we call a multi-exposure color filter array (ME-CFA), which consists of a mosaic pattern of RGB filters combined with neutral density filters with different attenuation levels. (see Fig. 1 for an example with three levels). The snapshot HDR sensor has the advantage of capturing multi-exposure information at one shot. Thus, it has great potential for HDR imaging of dynamic scenes and also HDR video acquisition without ghost and inpainting artifacts. However, HDR image reconstruction from the snapshot measurement includes two challenging problems: demosaicking (i.e., interpolation of missing RGB values) and HDR reconstruction (i.e., scene radiance estimation from LDR measurements). As we will show later, a simple combination of existing demosaicking/interpolation and HDR reconstruction methods cannot produce satisfactory results for this joint problem.

In this paper, we propose a novel deep snapshot HDR imaging framework that can effectively reconstruct an HDR image from the RAW data captured using an ME-CFA (see Fig. 1). In our framework, we introduce the key idea of luminance normalization and propose luminance-normalized network (LN-Net) that enables effective learning of HDR images in the luminance-normalized domain (LN domain).

In the training phase, we first train the network to generate tentatively interpolated LDR images, from which tentative HDR luminance is estimated. Then, we normalize the input ME-CFA RAW data and the corresponding ground-truth HDR image by the tentative HDR luminance. Finally, we train our main network (LN-Net) to reconstruct HDR images based on the pair of the RAW and the HDR image data in the LN domain. In the application phase, the HDR image is reconstructed through the learned two networks.

Our LN-Net mainly has two benefits. The first one is effective loss computation in the LN domain. The standard mean squared error (MSE) loss in the linear HDR domain has a problem of neglecting the errors in dark areas because they are quite small compared with the errors in bright areas. However, those errors in dark areas significantly affect the visual quality in a tone-mapped domain [12], which is commonly used to display HDR images. Based on this, some studies have computed the loss in a transformed domain, such as a log domain [6] and a global tone-mapped domain [2]. However, such signal-independent transformations do not reflect an actual signal component of each image. In contrast, by computing the loss in the LN domain, we can equally handle the errors in bright and dark areas by considering the actual luminance of each image.

The second benefit of LN-Net is effective input data normalization. In deep learning, the normalization of input data is important to extract effective features. Since a diverse range of scene radiance information is simultaneously encoded in the ME-CFA RAW data, we need to consider relative local contrasts, rather than absolute differences. Otherwise, features such as edges and textures in dark areas are prone to be ignored. In our LN-Net, by normalizing the input RAW data by the tentative luminance, we can naturally consider the relative local contrasts in both bright and dark areas regardless of absolute luminance.

Through the experiments using two public HDR image datasets, we validate the effectiveness of our framework by comparing it with other snapshot methods and current state-of-the-art HDR imaging methods using multiple LDR images. **Main contributions** of this paper are summarized as below.

- We propose a novel deep learning framework that effectively solves the joint demosaicking and HDR reconstruction problem for snapshot HDR imaging.
- We propose the idea of luminance normalization that enables effective loss computation and input data normalization by considering relative local contrasts of each image.
- We demonstrate that our framework can outperform other snapshot methods and reconstruct high-quality HDR images with fewer visible artifacts.

2   Related Work

Multiple-LDR-images-based methods have been studied for years. Their approaches include inverse radiometric function estimation [1], exposure fusion [13–15], patch-based [16, 17], rank minimization-based [18, 19], image alignment-based [20, 21], and learning-based [2–4, 22–24] methods. Although their performance has continuously been improved (see [17, 25] for reviews), it is essentially

difficult for multi-LDR-images-based methods to handle dynamic scenes with target or camera motions, resulting in ghost artifacts. Some methods have exploited a multi-camera/sensor system [26, 27] for one-shot acquisition of multi-LDR images. However, they require image or sensor alignment, which is another challenging task.

Single-LDR-image-based methods, also called inverse tone-mapping, have been actively studied in recent years. They train a mapping from a single-LDR image to an HDR image directly [5, 6, 28–30] or train a mapping to multiple LDR images intermediately, from which the HDR image is derived [7, 31, 32]. Although the single-LDR-image-based approach realizes one-shot HDR image acquisition, it is essentially difficult to reconstruct high-quality HDR images because there are no measurements obtained from different exposure levels.

Snapshot methods are based on a snapshot HDR imaging system with spatially varying exposure levels [10, 11]. Several hardware architectures or concepts have been proposed to realize a snapshot system, such as a coded exposure time [8, 33, 34], a coded exposure mask [35–37], a dual-ISO sensor [9, 38–40], and what we call an ME-CFA, which consists of the mosaic of RGB filters combined with neutral density filters with different attenuation levels [10, 11, 41–45]. The snapshot systems have great potential for HDR imaging in dynamic situations since it enables one-shot acquisition of multi-exposure information. However, HDR image reconstruction from the snapshot measurements is very challenging due to the sparse nature of each color-exposure component.

Some existing snapshot methods based on an ME-CFA first convert the snapshot LDR measurements to the sensor irradiance domain. By doing this, the problem reduces to the demosaicking problem in the sensor irradiance domain, for which several probability-based [41, 44] or learning-based [43] approaches have been proposed. However, this combined approach could not necessarily produce satisfactory results because the errors in the first step are propagated by the demosaicking step. Although some joint approach has also been proposed [10, 11, 42, 45], it is limited to a specific ME-CFA pattern [10, 11, 45] or it only performs a limited over- and under-exposed pixels correction [42].

In this paper, different from existing methods, we propose a general and high-performance framework exploiting deep learning to jointly solve the demosaicking and the HDR reconstruction problems for snapshot HDR imaging using an ME-CFA.

3   Proposed Deep Snapshot HDR Imaging

3.1   Framework overview

In this paper, we assume the ME-CFA pattern shown in Fig. 1, which consists of 4×4 regular pattern with three exposure levels, assuming the mosaic of RGB
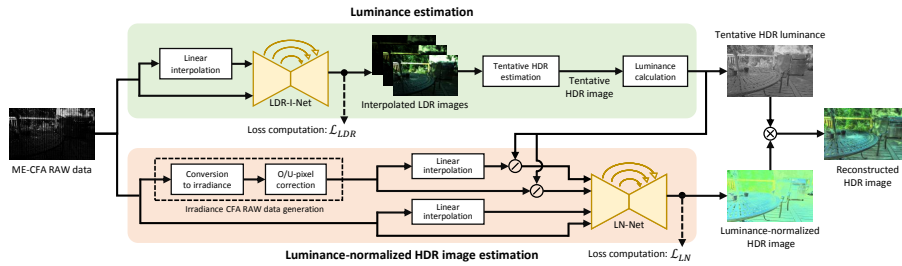
Fig. 2. Overview of our deep snapshot HDR imaging framework. It first estimates tentative HDR luminance and then estimates the HDR image in the luminance-normalized domain. The idea of luminance normalization enables to consider relative local contrasts in both bright and dark areas, regardless of the absolute luminance levels.

filters combined with neutral density filters with three attenuation levels. Although our framework is general and not limited to this pattern, we use this because (i) it is based on the common Bayer pattern [46], similar to existing ME-CFA patterns [10, 39], and (ii) it consists of three exposures, which are commonly used in recent HDR imaging studies [2, 3]. Those two conditions make it possible to experimentally compare our framework with standard Bayer demosaicking methods [47, 48] and also state-of-the-art HDR imaging methods using three LDR images [2, 3].

Figure 2 shows the overview of our framework, which mainly consists of two parts: (i) luminance estimation and (ii) luminance-normalized HDR image estimation. The first part estimates tentative HDR luminance based on the interpolated LDR images by the learned LDR-interpolation-network (LDR-I-Net). Then, based on the tentative luminance, the second part estimates the luminance-normalized HDR image by the learned LN-Net. Each part is detailed in subsection 3.2 and 3.3, respectively. Finally, the HDR image is reconstructed by multiplying the tentative luminance and the estimated luminance-normalized HDR image.

Throughout this paper, we use the term "irradiance" or "sensor irradiance [1]" to represent the irradiance of the light reaching the image sensor after going through the camera's optical elements such as a lens. Because we assume a linear optical system as in [1, 49], the sensor irradiance is assumed to be the same as scene radiance in this paper. We also assume linear responses between the sensor irradiance and pixel values because we process the RAW data, which typically have linear camera responses.

3.2   Luminance estimation

In the luminance estimation, we first interpolate the missing RGB pixel values to generate interpolated LDR images. For this purpose, we train LDR-I-Net, for which we adopt the U-Net architecture [50]. The inputs of LDR-I-Net are the

ME-CFA RAW data and the linearly interpolated each color-exposure sample of the ME-CFA raw data. The loss function for LDR-I-Net is described as

$$\mathcal{L}_{\mathrm{LDR}} = \sum_{k=1}^{N} ||\boldsymbol{f}_k\left(\boldsymbol{y};\boldsymbol{\theta}\right) - \boldsymbol{z}_k||_2^2$$
$$+ \lambda_{\mathrm{LDR}} ||\nabla \boldsymbol{f}_k\left(\boldsymbol{y};\boldsymbol{\theta}\right) - \nabla \boldsymbol{z}_k||_2^2, \tag{1}$$

where $\boldsymbol{y} = [\boldsymbol{x};\boldsymbol{h}(\boldsymbol{x})]$ is the network input, $\boldsymbol{x}$ is the sub-mosaicked representation of the ME-CFA RAW data, as used in [51] (e.g., sparse 16-channel data for the 4×4 regular pattern), $\boldsymbol{h}()$ represents the linear interpolation for the sparse data, $\boldsymbol{f}_k(\boldsymbol{y};\boldsymbol{\theta})$ is the output of LDR-I-Net for k-th exposure LDR image, $\boldsymbol{\theta}$ represents the network weights, $\boldsymbol{z}_k$ is true k-th exposure LDR image, N is the number of exposure levels in the ME-CFA, ∇ represents the horizontal and the vertical derivative operators, and λ_{LDR} is a hyper-parameter. The second term, which we call the gradient term, evaluates the errors in the image gradient domain. In this paper, we empirically set to $\lambda_{\mathrm{LDR}} = 1$ for the hyper-parameter. We will show the effectiveness of the gradient term by the ablation study in subsection 4.2. Note that the loss function of Eq. (1) evaluates the MSE in the LDR image domain, not in the HDR image domain. Thus, we use the standard MSE for loss computation.

Once we have interpolated LDR images, we apply Debevec's method [1] to the LDR images for tentative HDR image estimation. Then, tentative HDR luminance is derived as the maximum value of the RGB sensor irradiance values, which corresponds to the value (V) in the HSV color space. The tentative luminance can be formulated as

$$\hat{L}_i = \max_c \tilde{E}_i^{(c)}, \tag{2}$$
$$\tilde{\boldsymbol{E}} = \boldsymbol{R}(\boldsymbol{f}(\boldsymbol{y};\boldsymbol{\theta})), \tag{3}$$

where $\boldsymbol{f}(\boldsymbol{y};\boldsymbol{\theta})$ is the interpolated LDR images, $\boldsymbol{R}()$ represents the Debevec's HDR image estimation operation [1], $\tilde{\boldsymbol{E}}$ is the tentative HDR image estimated from the interpolated LDR images, $\tilde{E}_i^{(c)}$ is the estimated tentative sensor irradiance of c-th channel at i-th pixel in $\tilde{\boldsymbol{E}}$, and $\hat{L}_i$ is the tentative luminance of i-th pixel, where max operation is performed in a pixel-by-pixel manner.

3.3   Luminance-normalized HDR image estimation

In the luminance-normalized HDR image estimation, we train LN-Net to reconstruct the HDR image in the luminance-normalized domain. The inputs of LN-Net are the sub-mosaicked representation of the ME-CFA RAW data, its sensor irradiance version as we will explain below, and linearly interpolated versions of them. The irradiance data is normalized by the tentative luminance to consider relative local contrasts regardless of the absolute luminance levels. We detail each process to train LN-Net below.
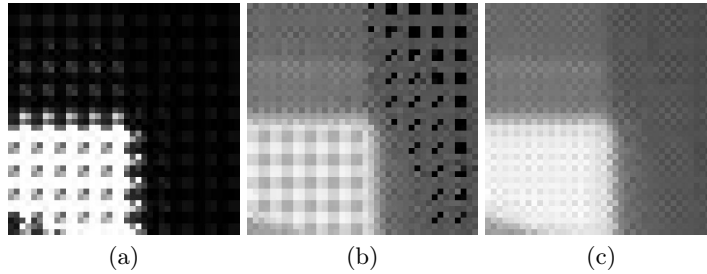
$$(a) \qquad\qquad (b) \qquad\qquad (c)$$

Fig. 3. Examples of (a) the ME-CFA RAW data in the LDR domain, (b) the irradiance CFA RAW data converted by Eq. (4), and (c) the irradiance CFA RAW data after the O/U-pixel correction.

We first convert the ME-CFA RAW data to the sensor irradiance domain as

$$\xi_{k,i} = \frac{x_i}{\rho_k \Delta t} \, , \tag{4}$$

where x_i is i-th pixel value of the ME-CFA RAW data, ρ_k is the attenuation factor for k-th exposure, Δt is the exposure time, and $\xi_{k,i}$ is the converted sensor irradiance of i-th pixel corresponding to k-th exposure. In the snapshot case using an ME-CFA, the attenuation factor ρ_k varies for each pixel according to the ME-CFA pattern, while the exposure time is constant for all pixels. Thus, in what follows, we set to $\Delta t = 1$, without loss of generality. Also, we call the irradiance data converted by Eq. (4) "irradiance CFA-RAW data", in which different exposure levels are already corrected by converting the ME-CFA RAW data to the sensor irradiance domain. Fig. 3(a) and 3(b) show the examples of the ME-CFA RAW data and the converted irradiance CFA RAW data.

In the original ME-CFA RAW data, many pixels are over-exposed (saturated) or under-exposed (black-out) depending on the exposure level of each pixel compared to the scene radiance. Such over- or under-exposed pixels (what we denote as "O/U pixels") have no meaningful irradiance information even after the conversion by Eq. (4). Thus, we present an effective O/U-pixel correction method, which replaces the irradiance of an O/U pixel with the linearly interpolated value using adjacent lower- or higher-exposure irradiance samples. For example, the irradiance of an over-exposed pixel is corrected as

$$\hat{\xi}_{k,i} = \begin{cases} \xi_{k,i} & (\xi_{k,i} \le \tau_{O,k}) \\ h_i(\boldsymbol{\xi}_{k-1}) & (\xi_{k,i} > \tau_{O,k}) \end{cases} \, , \tag{5}$$

where the suffix k represents k-th exposure, the suffix i represents i-th pixel, $\hat{\xi}_{k,i}$ is the irradiance after the over-exposed pixel correction, $\tau_{O,k}$ is the over-exposure threshold, $\boldsymbol{\xi}_{k-1}$ is the one-step lower-exposure sparse irradiance samples in the irradiance CFA RAW data, and $h_i(\boldsymbol{\xi}_{k-1})$ is i-th pixel value of the linearly interpolated irradiance of $\boldsymbol{\xi}_{k-1}$. We empirically set $0.995/(\rho_k \Delta t)$ to the over-exposure threshold $\tau_{O,k}$, where the range of the irradiance CFA RAW data is [0, 1]. This over-exposure correction is applied from the lower exposure data to the higher

exposure data. The under-exposed pixel correction is performed in the same manner, where the under-exposure threshold is set to $\tau_{U,k} = 0.005/(\rho_k \Delta t)$. Figure 3(b) and 3(c) show the examples of the irradiance CFA RAW data before and after the O/U-pixel correction, respectively.

We then apply a linear interpolation to the corrected irradiance CFA RAW data to prepare the network input. The corrected irradiance CFA RAW data, $\hat{\boldsymbol{\xi}}$, and its linearly interpolated version, $\boldsymbol{h}(\hat{\boldsymbol{\xi}})$, can be considered as the HDR domain data, in which local contrasts in dark areas are very low compared with those in bright areas. Thus, we normalize the HDR domain data by the estimated tentative HDR luminance. This luminance normalization converts the absolute local contrasts to the relative local contrasts. We also use the LDR domain data of the sub-mosaicked ME-CFA RAW data $\boldsymbol{x}$ and its linearly interpolated version $\boldsymbol{h}(\boldsymbol{x})$. For these LDR domain data, we do not perform the luminance normalization because the range of the absolute local contrasts is limited. The input to LN-Net $\boldsymbol{\eta}$ is described as

$$\boldsymbol{\eta} = \left[\boldsymbol{x}, \boldsymbol{h}(\boldsymbol{x}), \hat{\boldsymbol{\xi}}/\hat{\boldsymbol{L}}, \boldsymbol{h}(\hat{\boldsymbol{\xi}})/\hat{\boldsymbol{L}}\right] \,, \tag{6}$$

where $\hat{\boldsymbol{L}}$ is the tentative luminance, and the divide operation is performed in a pixel-by-pixel manner.

We then estimate the luminance-normalized HDR image by LN-Net, where we adopt the U-Net like network architecture [50]. The loss function of LN-Net is the MSE in the luminance-normalized HDR domain as follows.

$$\begin{aligned}
\mathcal{L}_{\mathrm{LN}} = &\|\boldsymbol{g}(\boldsymbol{\eta}; \boldsymbol{\psi}) - \boldsymbol{E}_{\mathrm{LN}}\|_2^2 \\
&+ \lambda_{\mathrm{LN}}\|\nabla \boldsymbol{g}(\boldsymbol{\eta}; \boldsymbol{\psi}) - \nabla \boldsymbol{E}_{\mathrm{LN}}\|_2^2 \,,
\end{aligned} \tag{7}$$

where $\boldsymbol{g}(\boldsymbol{\eta}; \boldsymbol{\psi})$ represents LN-Net which estimates the luminance-normalized HDR image, $\boldsymbol{\psi}$ is the weights for the network, $\boldsymbol{\eta}$ is the network input defined in Eq. (6), $\boldsymbol{E}_{\mathrm{LN}}$ is the luminance-normalized true HDR image, ∇ represents the horizontal and vertical gradient operator, and λ_{LN} is a hyper-parameter. In this paper, we empirically set to $\lambda_{\mathrm{LN}} = 1$ for the hyper-parameter. The luminance-normalized true HDR image is defined by

$$\boldsymbol{E}_{LN} = \boldsymbol{E}/\hat{\boldsymbol{L}} \,, \tag{8}$$

where $\boldsymbol{E}$ is the true HDR image and the divide operation is perform in a pixel-by-pixel manner. We finally reconstruct the final HDR image as

$$\hat{\boldsymbol{E}} = \hat{\boldsymbol{L}} \times \boldsymbol{g}(\boldsymbol{\eta}; \boldsymbol{\psi}) \,, \tag{9}$$

where the multiplying operation is performed in a pixel-by-pixel manner.

By substituting Eqs. (8) and (9) into Eq. (7), one can find that the loss function for LN-Net corresponds to the MSE normalized by the tentative luminance. In the proposed luminance-normalized HDR image estimation, the input and the output of the network are normalized by the luminance, which enables us to consider the relative local contrasts, rather than the absolute differences.

3.4   Network architecture

In our framework, we use two networks: LDR-I-Net and LN-Net. In this paper, we adopt the U-Net architecture [50] for both networks, where the depth of the U-Net is five. Though, one can use any network architectures.

As mentioned above, the network inputs are a pair of the sparse sub-mosaicked data and the dense interpolated data. To adapt the data sparseness difference, we insert RAW data adaptation blocks as follows. The RAW data adaptation for the sparse data consists of a convolution layer with the ReLU activation whose kernel size is 7×7. The adaptation for the interpolated data is a convolution layer with the ReLU activation whose kernel size is 3×3. The outputs of both adaptations are concatenated and then fed into the network. We validate the effectiveness of the RAW data adaptation blocks in subsection 4.2. The detailed network architecture of the U-Net and the RAW data adaptation blocks are included in the supplemental material.

4   Experimental Results

4.1   Setups

Datasets. We used two public HDR image datasets for evaluating our framework: Funt's dataset [52] and Kalantari's dataset [2].

Funt's dataset consists of static scenes. We used this dataset for comparison with other snapshot methods. Although other static HDR datasets are available as listed in [29], we used Funt's dataset because it contains relatively a large number of HDR images (224 images) generated using the same camera. Each HDR image is generated by Debevec's method [1] using 9 LDR images with the exposure value (EV) set of {-4, -3, -2, -1, 0, 1, 2, 3, 4}.

Kalantari's dataset is recently used for evaluating dynamic scenes mainly with human motions. We used this dataset for comparison with state-of-the-art HDR imaging methods using a single or multiple LDR images. The dataset contains 89 dynamic scenes. For each scene, the ground-truth HDR image is generated using static LDR images taken with a reference human pose. In contrast, test LDR images are taken with a human motion including the reference pose. The EV set of {-2, 0, 2} or {-3, 0, 3} is used for the LDR image acquisition, where EV=0 is used to take the test LDR image with the reference pose.

ME-CFA data simulation. We simulated the ME-CFA data from the above-mentioned HDR data, which is 32-bit RGBE image format. We first normalized each HDR data to [0, 1] and scaled the normalized HDR data as [0, 1], [0, 4], and [0, 16], according to the three exposure levels corresponding to the assumed EV set of {-2, 0, 2}. Then, we clipped each scaled data by [0, 1] and quantized the clipped data by 8-bit depth. Finally, we sampled the quantized data according to the ME-CFA pattern to generate the mosaic ME-CFA data. By this data generation process, quantization errors were properly taken into account.

Table 1. Ablation study.

	CPSNR	G-CPSNR	L-CPSNR	HDR-VDP-2	LN-MSE
Ours (full version)	**48.57**	**41.94**	**40.38**	**80.57**	**0.0585**
without luminance normalization	46.70	30.24	29.29	77.97	0.1289
without O/U-pixel correction	42.30	36.60	34.45	78.51	0.0913
without gradient term in the loss	41.76	36.14	35.76	65.41	0.4364
without RAW data adaptation	46.19	39.45	37.77	78.16	0.0688

Training setups. For Funt's dataset, we used randomly selected 13 images for testing and the rest 211 images for training. For Kalantari's dataset, we used the provided 15 test and 74 training images. In the training phase, we randomly sampled $32{\times}32$-sized patches from each training image set and randomly applied each of a horizontal flip, a vertical flip, and a swapping of horizontal and vertical axes (transpose) for data augmentation. The used optimizer is Adam [53], where the learning rate was set to 0.001 and the parameters (β_1, β_2) were set to $(0.9, 0.999)$. We performed 3,000 times mini-batch updates, where the mini-batch size was set to 32.

Evaluation metrics. We used the following five metrics: color PSNR (CP-SNR) in the linear HDR domain, CPSNR in the global tone-mapped domain (G-CPSNR), CPSNR in the local tome-mapped domain (L-CPSNR), HDR-VDP-2 [54], and luminance-normalized MSE (LN-MSE), which is the MSE normalized by the true luminance. We used the same global tone-mapping function as [2] for G-CPSNR and the MATLAB local tone-mapping function for L-CPSNR. For each dataset, the average metric value of all test images is presented for comparison. For subjective evaluation, we used a commercial software, Photomatix[1], to apply local tone-mapping for effective visualization.

4.2   Validation study of our framework

We first evaluate the validity of our framework using Funt's dataset.

Ablation study. Table 1 shows the result for the ablation study. We can observe that the case without the normalization by luminance (second row) presents much lower G-PSNR and L-CPSNR values compared with the cases with the normalization. This is because the cases without the normalization tend to neglect local contrasts in dark areas, which decreases the performance of both the subjective and the objective evaluation in the tone-mapped domains, where dark areas are significantly enhanced. We can also observe that the other proposed components certainly contribute to the performance improvements in all evaluated metrics.

---

[1] https://www.hdrsoft.com

Table 2. Loss domain comparison.

Loss domain	CPSNR	G-CPSNR	L-CPSNR	HDR-VDP-2	LN-MSE
Linear HDR domain	47.63	38.42	36.25	78.96	0.0749
Global tone-mapped HDR domain	47.56	41.57	39.57	79.73	0.0609
Our luminance-normalized HDR domain	**48.57**	**41.94**	**40.38**	**80.57**	**0.0585**

Loss domain comparison. Table 2 shows the comparison of loss computation domains. The loss in the standard linear HDR domain presents lower G-CPSNR and L-CPSNR values because it tends to disregard the errors in dark areas. The loss in the global tone-mapped domain [2] improves the G-CPSNR and the L-CPSNR performance, respectively. The loss in our proposed luminance-normalized domain provides further better performance by considering the relative local contrasts.

The above two studies demonstrate the effectiveness of our framework with the luminance normalization that enables effective local contrast consideration.

4.3   Comparison with other methods

Compared methods. To the best of our knowledge, there is no publicly available source code that is directly workable and comparable for the considered snapshot HDR imaging problem. Thus, we compared our framework with two combination frameworks as follows.
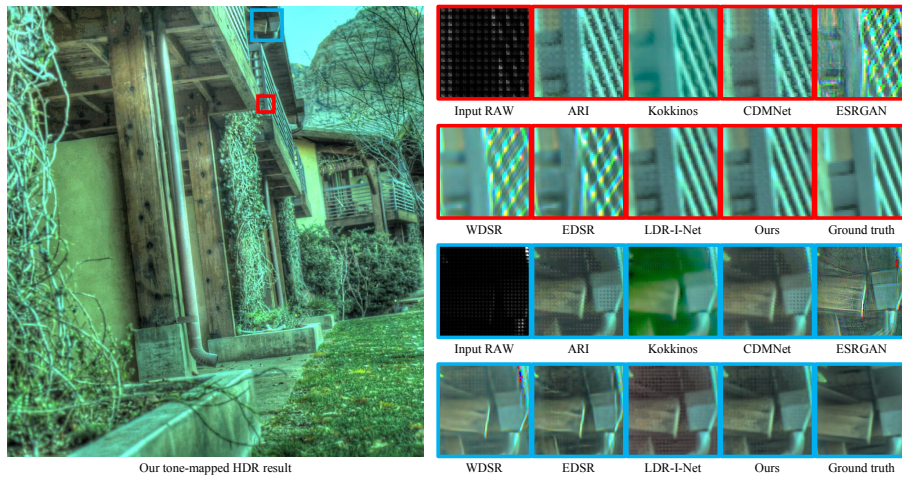
The first one is the demosaicking-based framework. It first converts the ME-CFA RAW data to the irradiance CFA RAW data and then applies an existing Bayer demosaicking method to the irradiance CFA RAW data. To generate the irradiance CFA RAW data, we applied the same processes as in subsection 3.3, including our proposed O/U-pixel correction since it was confirmed that our pixel correction significantly improves the numerical performance of existing methods. We used state-of-the-art interpolation-based (ARI [55]) and deep learning-based (Kokkions [48] and CDMNet [47]) demosaicking methods for comparison.

The second one is the LDR-interpolation-based framework. It first interpolates (up-samples) the sub-mosaicked ME-CFA RAW data by an existing super-resolution (SR) method with the scaling factor of 4 and then performs HDR reconstruction from the interpolated LDR images. We used existing competitive SR methods (ESRGAN [56], WDSR [57], and EDSR [58]) and our LDR-I-Net for SR and Debevec's method [1] for HDR reconstruction.

Results. Table 3 and Fig. 4 show the numerical and the visual comparisons using Funt's dataset. Table 3 demonstrates that our framework can provide the best performance in all metrics. In Fig. 4, we can observe that the demosaicking-based methods generate severe zipper artifacts (ARI and CDMNet) or over-smoothed results (Kokkions), while the LDR-interpolation-based methods (ES-RGAN, WDSR, and EDSR) generate severe aliasing artifacts for the high-frequency area of the red box. Although LDR-I-Net provides comparable results

Table 3. Comparison with two combination frameworks for snapshot HDR imaging.

Framework	Demosaicking/SR	CPSNR	G-CPSNR	L-CPSNR	HDR-VDP-2	LN-MSE
Demosaicking-based framework:	ARI [55]	46.14	38.15	36.69	75.68	0.0712
Irradiance CFA RAW data generation	Kokkinos [48]	41.06	26.27	26.65	69.32	0.1840
→ Demosaicking	CDMNet [47]	46.32	38.37	37.12	58.00	0.0713
LDR-interpolation-based framework:	ESRGAN [56]	30.66	25.21	21.87	53.55	0.2720
LDR interpolation by SR	WDSR [57]	35.75	30.97	29.32	60.16	0.3796
→ HDR reconstruction	EDSR [58]	39.19	32.57	29.95	66.04	0.1190
	LDR-I-Net	43.38	35.64	34.54	76.30	0.1030
Our deep snapshot HDR imaging framework		**48.57**	**41.94**	**40.38**	**80.57**	**0.0585**

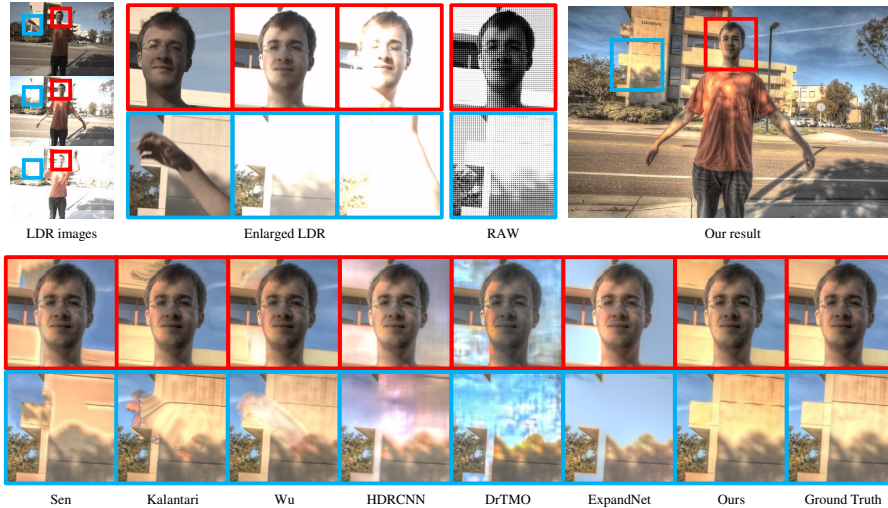**Fig. 4.** Visual comparisons on Funt's dataset.

for bright areas, it generates severe quantization artifacts in the dark area of the blue box, because it only learns LDR interpolation, but does not perform any learning for HDR reconstruction. In contrast, our framework can produce a better result with fewer visible artifacts, though slight artifacts still appear in the dark area. More results can be seen in the supplemental material.

4.4 Comparison with state-of-the-art methods using a single or multiple LDR images

Compared methods. Using Kalantari's dataset, we next compare our snapshot framework with state-of-the-art HDR imaging methods using multiple LDR images (Sen [17], Kalantari [2], and Wu [3]) or a single LDR image (HDRCNN [6], DrTMO [31], and ExpandNet [5]). We used all three LDR images for the multi-LDR-based methods and the second-exposure LDR image for the single-LDR-based methods.

Table 4. Comparison with state-of-the-art HDR imaging methods.

Input sources	Methods	CPSNR	G-CPSNR	L-CPSNR	HDR-VDP-2	LN-MSE
Multiple LDR images	Sen [17]	38.05	40.76	36.13	61.08	0.0389
	Kalantari [2]	41.15	**42.65**	**38.22**	64.57	**0.0306**
	Wu [3]	40.88	42.53	37.98	65.60	0.0338
Single LDR image (Second exposure)	HDRCNN [6]	12.92	14.13	34.80	54.48	4.1082
	DrTMO [31]	18.23	14.07	25.32	56.78	8.7912
	ExpandNet [5]	22.09	22.37	28.03	57.34	1.2923
ME-CFA RAW data	Ours	**41.43**	38.60	35.23	**66.59**	0.0832

LDR images   Enlarged LDR   RAW   Our result

Sen   Kalantari   Wu   HDRCNN   DrTMO   ExpandNet   Ours   Ground Truth

Fig. 5. Visual comparisons on Kalantari's dataset.

Results. Figure 5 shows the visual comparison. We can observe that the multi-LDR-based methods (Sen, Kalantari, and Wu) generate severe ghost artifacts in the red and the blue box areas, which are due to the head and the arm motions between input LDR images. The single-LDR-based methods (HDR-CNN, DrTMO, and ExpandNet) generate severe inpainting artifacts for the over-exposed areas in the input second-exposure LDR image. In contrast, our framework can reconstruct a visually pleasing result with much fewer visible artifacts. More results can be seen in the supplemental material.

Table 4 shows the numerical comparison. Our framework provides the highest score in CPSNR and HDR-VDP-2. In contrast, the multi-LDR-based methods present better performance for the other metrics. This is because these methods have the benefit of having all three-exposure information for each pixel, and thus should provide better performance for static regions, which are dominant in each scene of Kalantari's dataset. However, as shown in the visual comparison, these methods are very susceptible to ghost artifacts, which significantly
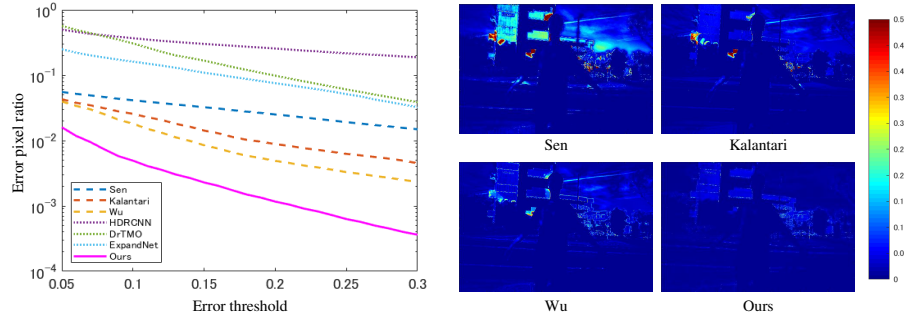
Fig. 6. Left: Comparison of the error pixel ratio. Right: Comparison of the error maps.

disturbs visual perception and makes the perceptual HDR-VDP-2 score lower. To quantitatively evaluate such significant artifacts, in Fig. 6, we evaluate the ratio of error pixels whose MSE of RGB irradiance values is larger than the threshold of the horizontal axis. From the result, we can clearly observe that our snapshot framework can generate the HDR image with much fewer error pixels. We also show the comparison of error maps, where the multi-LDR-based methods generate significant errors in the dynamic regions around the head and the arm.

4.5   Limitation

In our results, zipper artifacts still remain in some areas. This is because of the very challenging nature of snapshot HDR reconstruction with very sparse sampling of each color-exposure component and many saturated/blacked-out pixels. Furthermore, in the snapshot HDR problem, zipper artifacts may occur even for uniform areas without textures because of the differences of the quantization levels of three exposure images, meaning that converted sensor irradiance values in the uniform area do not match completely among the three exposure levels. To reduce the remaining zipper artifacts is our future work.

5   Conclusion

In this paper, we have proposed a novel deep learning-based framework that can effectively address the joint demosaicking and HDR reconstruction problem for snapshot HDR imaging using an ME-CFA. We have introduced the idea of luminance normalization that simultaneously enables effective loss computation and input data normalization to learn the HDR image reconstruction network by considering relative local image contrasts. Experimental results have demonstrated that our framework can produce HDR images with much fewer visual artifacts compared with other snapshot methods and also state-of-the-art HDR imaging methods using multiple LDR images. We provide the source code to reproduce our results at http://www.ok.sc.e.titech.ac.jp/res/DSHDR/index.html.

References

1. Debevec, P., Malik, J.: Recovering high dynamic range radiance maps from photographs. In: Proc. of SIGGRAPH. (1997) 1–10
2. Kalantari, N.K., Ramamoorthi, R.: Deep high dynamic range imaging of dynamic scenes. ACM Trans. on Graphics **36** (2017) 144:1–12
3. Wu, S., Xu, J., Tai, Y.W., Tang, C.K.: Deep high dynamic range imaging with large foreground motions. In: Proc. of European Conf. on Computer Vision (ECCV). (2018) 121–135
4. Yan, Q., Gong, D., Shi, Q., van den Hengel, A., Shen, C., Reid, I., Zhang, Y.: Attention-guided network for ghost-free high dynamic range imaging. In: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). (2019) 1751–1760
5. Marnerides, D., Bashford-Rogers, T., Hatchett, J., Debattista, K.: ExpandNet: A deep convolutional neural network for high dynamic range expansion from low dynamic range content. Computer Graphics Forum **37** (2018) 37–49
6. Eilertsen, G., Kronander, J., Denes, G., Mantiuk, R.K., Unger, J.: HDR image reconstruction from a single exposure using deep CNNs. ACM Trans. on Graphics **36** (2017) 178:1–15
7. Lee, S., An, G.H., Kang, S.J.: Deep chain HDRI: Reconstructing a high dynamic range image from a single low dynamic range image. IEEE Access **6** (2018) 49913–49924
8. Cho, H., Kim, S.J., Lee, S.: Single-shot high dynamic range imaging using coded electronic shutter. Computer Graphics Forum **33** (2014) 329–338
9. Choi, I., Baek, S.H., Kim, M.H.: Reconstructing interlaced high-dynamic-range video using joint learning. IEEE Trans. on Image Processing **26** (2017) 5353–5366
10. Narasimhan, S.G., Nayar, S.K.: Enhancing resolution along multiple imaging dimensions using assorted pixels. IEEE Trans. on Pattern Analysis and Machine Intelligence **27** (2005) 518–530
11. Nayar, S.K., Mitsunaga, T.: High dynamic range imaging: Spatially varying pixel exposures. In: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). (2000) 1–8
12. Eilertsen, G., Mantiuk, R.K., Unger, J.: Real-time noise-aware tone mapping. ACM Trans. on Graphics **34** (2015) 198:1–15
13. Ma, K., Duanmu, Z., Yeganeh, H., Wang, Z.: Multi-exposure image fusion by optimizing a structural similarity index. IEEE Trans. on Computational Imaging **4** (2017) 60–72
14. Ma, K., Li, H., Yong, H., Wang, Z., Meng, D., Zhang, L.: Robust multi-exposure image fusion: a structural patch decomposition approach. IEEE Trans. on Image Processing **26** (2017) 2519–2532
15. Mertens, T., Kautz, J., Van Reeth, F.: Exposure fusion: A simple and practical alternative to high dynamic range photography. Computer Graphics Forum **28** (2009) 161–171
16. Hu, J., Gallo, O., Pulli, K., Sun, X.: HDR deghosting: How to deal with saturation? In: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). (2013) 1163–1170
17. Sen, P., Kalantari, N.K., Yaesoubi, M., Darabi, S., Goldman, D.B., Shechtman, E.: Robust patch-based HDR reconstruction of dynamic scenes. ACM Trans. on Graphics **31** (2012) 203:1–11

18. Lee, C., Li, Y., Monga, V.: Ghost-free high dynamic range imaging via rank minimization. IEEE Signal Processing Letters **21** (2014) 1045–1049

19. Oh, T.H., Lee, J.Y., Tai, Y.W., Kweon, I.S.: Robust high dynamic range imaging by rank minimization. IEEE Trans. on Pattern Analysis and Machine Intelligence **37** (2014) 1219–1232

20. Hasinoff, S.W., Sharlet, D., Geiss, R., Adams, A., Barron, J.T., Kainz, F., Chen, J., Levoy, M.: Burst photography for high dynamic range and low-light imaging on mobile cameras. ACM Trans. on Graphics **35** (2016) 192:1–12

21. Kalantari, N.K., Ramamoorthi, R.: Deep HDR video from sequences with alternating exposures. Computer Graphics Forum **38** (2019) 193–205

22. Prabhakar, K.R., Arora, R., Swaminathan, A., Singh, K.P., Babu, R.V.: A fast, scalable, and reliable deghosting method for extreme exposure fusion. In: Proc. of IEEE Int. Conf. on Computational Photography (ICCP). (2019) 170–177

23. Ram Prabhakar, K., Sai Srikar, V., Venkatesh Babu, R.: DeepFuse: A deep unsupervised approach for exposure fusion with extreme exposure image pairs. In: Proc. of IEEE Int. Conf. on Computer Vision (ICCV). (2017) 4724–4732

24. Yan, Q., Gong, D., Zhang, P., Shi, Q., Sun, J., Reid, I., Zhang, Y.: Multi-scale dense networks for deep high dynamic range imaging. In: Proc. of IEEE Winter Conf. on Applications of Computer Vision (WACV). (2019) 41–50

25. Tursun, O.T., Akyüz, A.O., Erdem, A., Erdem, E.: The state of the art in HDR deghosting: A survey and evaluation. Computer Graphics Forum **34** (2015) 683–707

26. Ogino, Y., Tanaka, M., Shibata, T., Okutomi, M.: Super high dynamic range video. In: Proc. of Int. Conf. on Pattern Recognition (ICPR). (2016) 4208–4213

27. Tocci, M.D., Kiser, C., Tocci, N., Sen, P.: A versatile HDR video production system. ACM Trans. on Graphics **30** (2011) 41:1–9

28. Yang, X., Xu, K., Song, Y., Zhang, Q., Wei, X., Lau, R.W.: Image correction via deep reciprocating HDR transformation. In: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). (2018) 1798–1807

29. Moriwaki, K., Yoshihashi, R., Kawakami, R., You, S., Naemura, T.: Hybrid loss for learning single-image-based HDR reconstruction. arXiv preprint 1812.07134 (2018)

30. Kim, S.Y., Kim, D.E., Kim, M.: ITM-CNN: Learning the inverse tone mapping from low dynamic range video to high dynamic range displays using convolutional neural networks. In: Proc. of Asian Conf. on Computer Vision (ACCV). (2018) 395–409

31. Endo, Y., Kanamori, Y., Mitani, J.: Deep reverse tone mapping. ACM Trans. on Graphics **36** (2017) 177:1–10

32. Lee, S., Hwan An, G., Kang, S.J.: Deep recursive HDRI: Inverse tone mapping using generative adversarial networks. In: Proc. of European Conf. on Computer Vision (ECCV). (2018) 613–628

33. Gu, J., Hitomi, Y., Mitsunaga, T., Nayar, S.: Coded rolling shutter photography: Flexible space-time sampling. In: Proc. of IEEE Int. Conf. on Computational Photography (ICCP). (2010) 1–8

34. Uda, S., Sakaue, F., Sato, J.: Variable exposure time imaging for obtaining unblurred HDR images. IPSJ Trans. on Computer Vision and Applications **8** (2016) 3:1–7

35. Alghamdi, M., Fu, Q., Thabet, A., Heidrich, W.: Reconfigurable snapshot HDR imaging using coded masks and inception network. In: Proc. of Vision, Modeling, and Visualization (VMV). (2019) 1–9

36. Nagahara, H., Sonoda, T., Liu, D., Gu, J.: Space-time-brightness sampling using an adaptive pixel-wise coded exposure. In: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW). (2018) 1834–1842

37. Serrano, A., Heide, F., Gutierrez, D., Wetzstein, G., Masia, B.: Convolutional sparse coding for high dynamic range imaging. Computer Graphics Forum **35** (2016) 153–163

38. Go, C., Kinoshita, Y., Shiota, S., Kiua, H.: Image fusion for single-shot high dynamic range imaging with spatially varying exposures. In: Proc. of Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC). (2018) 1082–1086

39. Hajisharif, S., Kronander, J., Unger, J.: Adaptive dualISO HDR reconstruction. EURASIP Journal on Image and Video Processing **2015** (2015) 1–13

40. Heide, F., Steinberger, M., Tsai, Y.T., Rouf, M., Pająk, D., Reddy, D., Gallo, O., Liu, J., Heidrich, W., Egiazarian, K., Kautz, J., Pulli, K.: FlexISP: A flexible camera image processing framework. ACM Trans. on Graphics **33** (2014) 231:1–13

41. Aguerrebere, C., Almansa, A., Delon, J., Gousseau, Y., Musé, P.: A Bayesian hyperprior approach for joint image denoising and interpolation, with an application to HDR imaging. IEEE Trans. on Computational Imaging **3** (2017) 633–646

42. Aguerrebere, C., Almansa, A., Gousseau, Y., Delon, J., Muse, P.: Single shot high dynamic range imaging using piecewise linear estimators. In: Proc. of IEEE Int. Conf. on Computational Photography (ICCP). (2014) 1–10

43. An, V.G., Lee, C.: Single-shot high dynamic range imaging via deep convolutional neural network. In: Proc. of Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC). (2017) 1768–1772

44. Rouf, M., Ward, R.K.: High dynamic range imaging with a single exposure-multiplexed image using smooth contour prior. In: Proc. of IS&T Int. Symposium on Electronic Imaging (EI). (2018) 440:1–6

45. Cheng, C.H., Au, O.C., Cheung, N.M., Liu, C.H., Yip, K.Y.: High dynamic range image capturing by spatial varying exposed color filter array with specific demosaicking algorithm. In: Proc. of IEEE Pacific Rim Conf. on Communications, Computers and Signal Processing (PACRIM). (2009) 648–653

46. Bayer, B.E.: Color imaging array, US patent 3,971,065 (1976)

47. Cui, K., Jin, Z., Steinbach, E.: Color image demosaicking using a 3-stage convolutional neural network structure. In: Proc. of IEEE Int. Conf. on Image Processing (ICIP). (2018) 2177–2181

48. Kokkinos, F., Lefkimmiatis, S.: Deep image demosaicking using a cascade of convolutional residual denoising networks. In: Proc. of European Conf. on Computer Vision (ECCV). (2018) 317–333

49. Grossberg, M.D., Nayar, S.K.: What is the space of camera response functions? In: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). (2003) 1–8

50. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Proc. of Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI). (2015) 234–241

51. Henz, B., Gastal, E.S., Oliveira, M.M.: Deep joint design of color filter arrays and demosaicing. Computer Graphics Forum **37** (2018) 389–399

52. Funt, B., Shi, L.: The rehabilitation of MaxRGB. In: Proc. of Color and Imaging Conference (CIC). (2010) 256–259

53. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint 1412.6980 (2014)

54. Mantiuk, R., Kim, K.J., Rempel, A.G., Heidrich, W.: HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. ACM Trans. on Graphics **30** (2011) 40:1–13
55. Monno, Y., Kiku, D., Tanaka, M., Okutomi, M.: Adaptive residual interpolation for color and multispectral image demosaicking. Sensors **17** (2017) 2787:1–21
56. Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., Change Loy, C.: ESRGAN: Enhanced super-resolution generative adversarial networks. In: Proc. of European Conf. on Computer Vision Workshops (ECCVW). (2018) 1–16
57. Fan, Y., Yu, J., Huang, T.S.: Wide-activated deep residual networks based restoration for BPG-compressed images. In: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition workshops (CVPRW). (2018) 2621–2624
58. Lim, B., Son, S., Kim, H., Nah, S., Mu Lee, K.: Enhanced deep residual networks for single image super-resolution. In: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW). (2017) 1132–1140