

# Deep Joint Demosaicing and High Dynamic Range Imaging Within a Single Shot

Yilun Xu<sup>✉</sup>, Ziyang Liu<sup>✉</sup>, Xingming Wu<sup>✉</sup>, Weihai Chen<sup>✉</sup>, Member, IEEE, Changyun Wen<sup>✉</sup>, Fellow, IEEE,  
and Zhengguo Li<sup>✉</sup>, Senior Member, IEEE

**Abstract**—Spatially varying exposure (SVE) is a promising choice for high-dynamic-range (HDR) imaging (HDRI). The SVE-based HDRI, which is called single-shot HDRI, is an efficient solution to avoid ghosting artifacts. However, it is very challenging to restore a full-resolution HDR image from a real-world image with SVE because: a) only one-third of pixels with varying exposures are captured by camera in a Bayer pattern, b) some of the captured pixels are over- and under-exposed. For the former challenge, a spatially varying convolution (SVC) is designed to process the Bayer images carried with varying exposures. For the latter one, an exposure-guidance method is proposed against the interference from over- and under-exposed pixels. Finally, a joint demosaicing and HDRI deep learning framework is formalized to include the two novel components and to realize an end-to-end single-shot HDRI. Experiments indicate that the proposed end-to-end framework avoids the problem of cumulative errors and surpasses the related state-of-the-art methods. Related codes and datasets will be provided at <https://github.com/yilun-xu/SVEHDRI/>.

**Index Terms**—High-dynamic-range imaging, spatially varying exposure, demosaicing, spatially varying convolution, exposure guidance.

## I. INTRODUCTION

THE dynamic range of a natural scene is usually much higher than that of a low-dynamic-range (LDR) image captured using a smartphone or a digital camera via a single shot. Considerable information from the real scene is lost in the LDR image. HDRI technology was introduced to address such a problem [1]–[4]. HDRI has become one of the hottest topics in the fields of image processing and computer vision.

Manuscript received July 9, 2021; revised October 17, 2021; accepted November 16, 2021. Date of publication November 19, 2021; date of current version July 5, 2022. This work was supported in part by the National Natural Science Foundation of China under Project 61620106012, in part by Beijing Municipal Natural Science Foundation under Project 4202042, in part by the Key Research and Development Program of Zhejiang Province under Project 2020C01109, in part by the Foundation Strengthening Program Technology Foundation under Project 2019-JCJQ-JJ-268, in part by the National Natural Science Foundation of China under Project 61573048, and in part by the Macao Science and Technology Development Fund, under Project 0022/2019/AKP. This article was recommended by Associate Editor X. Li. (Yilun Xu and Ziyang Liu are co-first authors.) (Corresponding author: Xingming Wu.)

Yilun Xu, Ziyang Liu, Xingming Wu, and Weihai Chen are with the School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China (e-mail: yilunxu\_buaa@163.com; by1703126@buaa.edu.cn; wxmbuaa@163.com; whchen@buaa.edu.cn).

Changyun Wen is with the School of Electrical and Electronics Engineering, Nanyang Technological University, Singapore 639798 (e-mail: ecywen@ntu.edu.sg).

Zhengguo Li is with the SRO Department, Institute for Infocomm Research, Singapore 138632 (e-mail: ezhgli@i2r.a-star.edu.sg).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TCSVT.2021.3129691>.

Digital Object Identifier 10.1109/TCSVT.2021.3129691

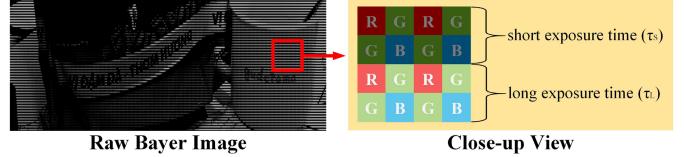


Fig. 1. Raw Bayer image with row-wise varying exposure times [12].

A popular method for HDRI is to sequentially capture multiple LDR images with varying exposures sequentially and then merge them into an HDR image [5], [5]–[7]. This method is called exposure stacking, which is widely adopted in smartphones and digital cameras. Exposure stacking performs well in static scenes. However, moving objects could exist in the shooting scenes, which lead to unavoidable ghosting artifacts in the HDR image [8]–[11]. Many methods were proposed to eliminate ghosting artifacts in the HDR image, but a large amount of computation is often required and these methods may fail in scenes with very complex motion and extreme dynamic range.

Due to the challenge of ghosting removal [9], [10], ghosting artifacts are believed to be the Achilles' heel for the exposure stacking-based HDRI. As such, exposure stacking is unsuitable for capturing HDR videos [13]–[15]. The single-shot HDRI was proposed to capture ghost-free HDR images with varying exposures in a single image [12], [16]–[18]. In the single-shot HDRI, the exposure of pixels varies along with different spatial locations. In general, the raw data obtained by the camera is in a Bayer image. One typical example is given in Fig. 1, where the raw Bayer image is sensed by alternating the exposure time every other rows. This shooting method [12], [19], [20] is referred to as dual-time in this paper. In addition, the single-shot HDRI is also a good candidate to capture HDR videos [21].

Among the single-shot dual-time HDRI algorithms [12], [19], [20], the stages of demosaicing and HDR reconstruction are separated. Thus, the error generated in the previous stage will affect the posterior one, resulting in error accumulation and drift. Recently, the idea of joint demosaicing with other low-level image processing tasks [22]–[26] has appeared in some studies to avoid the cumulative error. However, the existing joint demosaicing algorithm's convolution methods have not been adjusted accordingly to the spatial change of the data pattern. Moreover, when the captured Bayer image with varying exposure times is converted into a Bayer radiance image [12], [20] by using camera response functions (CRFs) [27], the brightness difference caused by

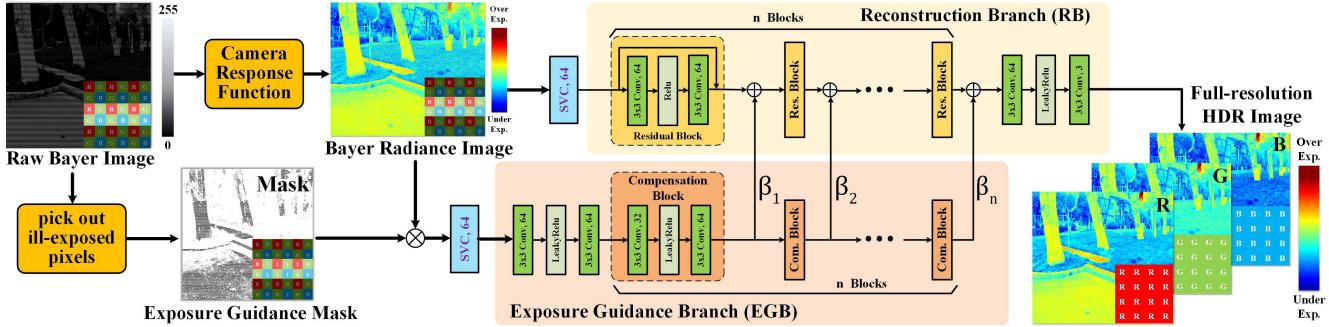


Fig. 2. Overall framework of the proposed algorithm. The proposed CNN includes a reconstruction branch (RB), and two distinctive components: spatially varying convolution (SVC) and exposure-guidance branch (EGB).

different exposure time cannot be completely eliminated. The ill-exposed pixels in the Bayer radiance image will also interfere the convolutional neural network (CNN) [12].

In this paper, a one-stage CNN is proposed for the single-shot dual-time HDRI. As shown in Fig 2. This novel CNN can restore high-quality HDR image at the full resolution from a single Bayer radiance image with varying exposures, by which the problem of cumulative errors in [12], [19], [20] can be avoided. Two distinctive components are introduced into the CNN to handle the challenges in single-shot HDRI. For the challenges of processing image pixels with varying colors (caused by the Bayer pattern) and varying exposures (caused by the dual-time), a novel spatially varying convolution (SVC) is designed and introduced as the first layer of CNN. The proposed SVC is adaptive to both varying colors and varying exposures, and can extract information from the dual-time Bayer image efficiently. The SVC can be easily inserted into other networks to process the Bayer image with or without SVE, and can be redesigned into other flexible variants when the data pattern or SVE changes. For the challenge of processing ill-exposed pixels, a novel exposure-guidance method is proposed, and the method is inspired by the dual-branch network [28], [29]. By exposure-guidance method, the prior knowledge of ill-exposed pixels is exploited, and an exposure-guidance branch (EGB) is proposed to assist the CNN by incorporating this prior. Clearly, both the SVC and the exposure-guidance method can improve the explainability of the proposed one-stage CNN. Moreover, a new HDR dataset is proposed in this paper. The proposed dataset consists of 500 pairs of images with short and long exposures, respectively. The varying exposures are achieved by changing the exposure time. Images with camera shaking or object movement are filtered out artificially. Finally, extensive experimental results demonstrate the effectiveness of the proposed algorithm.

Overall, the four major contributions of this paper are as follows:

- A one-stage CNN with an improved explainability is proposed to address the problems of demosaicing and single-shot dual-time HDRI end-to-end..
- A novel SVC is introduced to process the Bayer image with or without SVE appropriately.
- An exposure-guidance method is proposed to reduce the interference of ill-exposed pixels.

- A new HDR dataset is proposed in this paper. Camera parameters for shooting are provided in detail for other related HDR researchers to use.

The rest of this paper is organized as follows. Relevant works are reviewed in Section II. Details of the proposed algorithm are presented in Section III. Extensive experimental results are provided in Section IV. Lastly, concluding remarks are listed in Section V.

## II. LITERATURE REVIEW

In this section, the relevant works on HDRI and demosaicing are reviewed.

### A. HDRI

1) *Stack-Based HDRI*: The most popular method for generating an HDR image is to capture multiple differently exposed LDR images and merge all the LDR images into one HDR image. Such a method is called stack-based HDRI [30]–[33]. All the captured LDR images are firstly mapped into the corresponding radiance maps through the CRFs [27], and multiple radiance maps are then fused into an HDR image via a weighted average manner.

Exposure stacking-based methods often perform well in static scenes. Nevertheless, in dynamic scenes, the positions of moving objects in the exposure stack are different, resulting in ghost artifacts in HDR images. To remove the ghosting artifacts, one of the input images is selected as the reference image. All moving objects in other images are synchronized with those objects in the reference images. Ghost removal was widely studied, and many interesting algorithms were introduced [4], [8], [11], [34]–[36]. However, when complex motion or extreme dynamic range occurs in the scene, all these algorithms could fail. As indicated in [9], [10], no universal deghosting algorithm is available. Ghosting artifacts are thus believed to be the Achilles' heel for the exposure stacking-based HDRI.

2) *Single-Shot HDRI*: An alternative solution is to obtain multiple exposure information of a scene via a single shot, and this solution is attractive for HDR videos [21] and full light field reconstruction [37]. Nayar and Suda proposed the concept of spatially varying exposure (SVE) in [16], [18], such that diverse pixel values of a single image are differently exposed.

Two types of typical methods can be adopted to achieve SVE.

One type is to change the ISO value at different positions of the sensor. Given an ISO-based SVE image, a few methods were proposed to restore the final HDR image, including the adaptive kernel regression-based method [38], inpainting-based deinterlacing method [39], joint learning-based method [40], dictionary-based method [41], adaptive filter-based method [42], and deep learning-based method [17]. Nevertheless, the increase in ISO for a high exposure will amplify the camera noise, especially in low-lighting conditions.

The other way to achieve SVE is to change exposure times [12], [19], [20], [43], in which higher-quality images than the ISO-based approach can be obtained, as indicated in [27]. For example, an image can be shot with the exposure times varying every other lines, as shown in Fig. 1. Gu *et al.* proposed to adopt the structure of a coded rolling shutter as the readout structure of a CMOS image sensor [19] and introduced several coding schemes and corresponding applications. Cho *et al.* designed a multistage processing flow to restore dual-time SVE Bayer images to HDR images gradually [20]. An and Lee introduced an CNN-based technology to correct Bayer images, and then adopted demosaicing to obtain HDR images [12].

The stages of HDR reconstruction and demosaicing are separated. Thus, the error generated in the previous stage will affect the posterior one, resulting in error accumulation and drift. In addition, to deal with the ill-exposed pixels in the Bayer radiance image, Cho *et al.* [20] completes the image by deleting ill-exposed pixels and then interpolating from neighbors. An *et al.* [12] takes the complete image as input and use CNN to correct ill-exposed pixels. Compared with [20], the CNN is allowed to make full use of the information in the Bayer radiance image [12], [43], but ill-exposed areas will also interfere with the imaging results of well-exposed areas during the calculation process.

### B. Image Demosaicing

A color filter array (CFA) is put in front of a CMOS sensor to use a sensor designed for grayscale images and capture color images [44]. Each pixel on a Bayer image taken in this way has only one of red, green, and blue. The image is subjected to a postprocessing algorithm called demosaicing, which is to complete the missing information in the Bayer image. In images without SVE, no brightness difference caused by the different exposures occurs. Existing demosaicing algorithms are divided into two categories: model-based [45]–[48] and learning-based [22], [22]–[26].

To complete the missing two-thirds of pixels in the Bayer image, conventional model-based demosaicing algorithms [45]–[49] usually adopt different interpolation schemes for different data patterns. In learning-based demosaicing [22]–[26], [50] the missing pixels are interpolated by CNN. However, existing deep learning-based algorithms share a contrary interpolation philosophy with conventional demosaicing algorithms. As indicated by [51], the interpolators should be

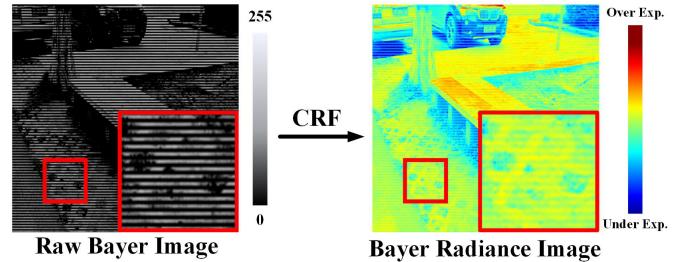


Fig. 3. Brightness differences are partly reduced using CRF. The slight horizontal stripes represent that the brightness differences cannot be completely eliminated.

adaptive to the changing of data pattern. It is unreasonable to adopt a single same interpolation scheme for all data patterns, because missing colors that need to be interpolated vary with different data patterns. Thus, it is the same for deep learning, where different data patterns should be convolved by different convolution kernels, and same patterns by same convolution kernels. But their convolution method [22]–[26] has not been adjusted accordingly due to the spatial change of the data pattern.

Moreover, in the Bayer image with SVE, the CRFs cannot completely eliminate the brightness difference caused by different exposure times. Restoring an HDR image from a dual-time SVE Bayer image brings more challenges to demosaicing.

## III. PROPOSED ALGORITHM

Given a raw Bayer image captured within a single shot, the CRFs are firstly applied to generate the Bayer radiance image, and then the exposure-guidance mask is generated from a raw Bayer image. The Bayer radiance image is restored into an HDR radiance map via a novel CNN end-to-end, and the prior information in the exposure-guidance mask can assist the CNN. Then the CNN includes a reconstruction branch (RB), and two distinctive components: spatial varying convolution (SVC) and exposure-guidance method. Both the proposed SVC and exposure-guidance method make the CNN more explainable. The overall joint learning framework is shown in Fig. 2. It should be pointed out that the proposed algorithm is on top of our previous work [52].

### A. Generation of Bayer Radiance Image

The raw input image  $Z$  captured within a single shot is a  $N$ -bit ( $N$  can be 8, 10, 12, etc.) Bayer image with row-wise varying exposure times [12], [19], [20], as shown in Fig. 1. Let  $\Delta t_{ij}$  be the exposure time of the pixel in the  $i$ th row,  $j$ th column of  $Z$ , then it is given as:

$$\Delta t_{ij} = \begin{cases} \tau_S, & i \bmod 4 = 1 \text{ or } 2 \\ \tau_L, & i \bmod 4 = 3 \text{ or } 0, \end{cases} \quad (1)$$

where  $\tau_S$  and  $\tau_L$  are the short and long exposure times, respectively.

To restore the corresponding HDR image, the Bayer image  $Z$  is firstly converted into a Bayer radiance image  $E$  [12]. The

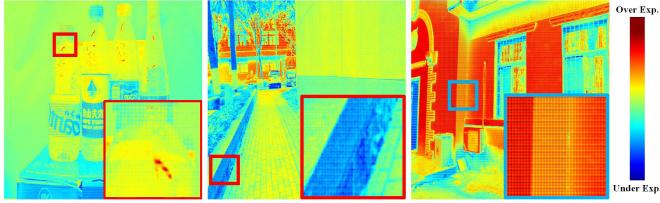


Fig. 4. Since only a single color (R, G, or B) is recorded at each pixel position, the visible grids appear in the Bayer radiance image.

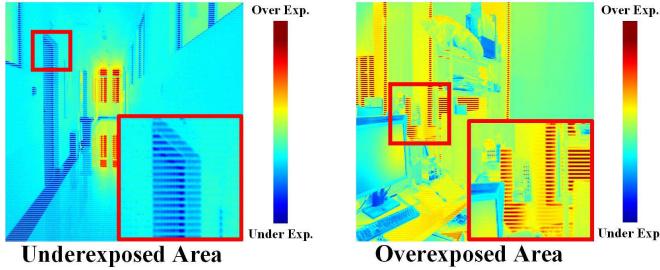


Fig. 5. Horizontal stripes indicating poor information, which is caused by over- and under-exposure.

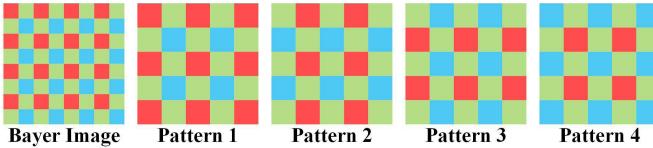


Fig. 6. Illustration of four different patterns in a  $5 \times 5$  receptive field.

pixel  $z_{ij}$  of  $Z$  is normalized according to its exposure time.

$$\ln(e_{ij}) = \ln(f_{c_{ij}}^{-1}(z_{ij})) - \ln(\Delta t_{ij}), \quad (2)$$

where  $c_{ij} \in \{R, G, B\}$  is the color channel of  $z_{ij}$ . The CRF  $f_{c_{ij}}(\cdot)$  can be estimated via the method in [27].  $f_{c_{ij}}^{-1}(\cdot)$  is the inverse function of  $f_{c_{ij}}(\cdot)$ .  $e_{ij}$  represents the converted irradiance pixel. Note that CRFs are assumed to be available, because they can be easily estimated for any digital cameras and smartphones. As shown in Fig. 3, the brightness differences caused by the dual-time are partially reduced.

As indicated above, there are three main challenges in restoring a full-resolution HDR image from a dual-time Bayer radiance image. First, two-thirds of color pixels are missed, resulting in visible grids, as shown in Fig. 4. Second, the remaining pixels are carried with varying exposures and the brightness differences cannot be completely eliminated via CRFs, like the horizontal stripes shown in Fig. 3. Moreover, some of the captured pixels are ill-exposed, which makes the HDR restoration more difficult, as shown in Fig. 5.

### B. Spatially Varying Convolution

There are four color patterns in the Bayer image, as shown in Fig. 6. Conventional demosaicing methods complete pixels via color pattern-oriented interpolation schemes [45], [53], [54], which means that there exist different interpolators corresponding to the four different color patterns.

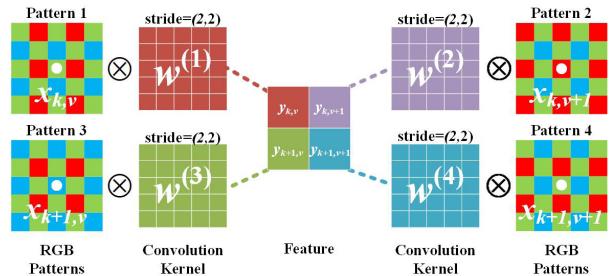


Fig. 7. Illustration of four different RGB patterns in a  $5 \times 5$  receptive field, and the corresponding degraded version of spatially varying convolution (SVC-D).

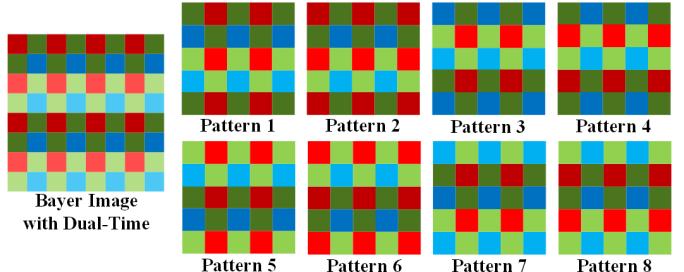


Fig. 8. Illustration of eight different patterns in a  $5 \times 5$  receptive field, where the darker and lighter colors represent the short and long exposed radiation pixels.

However, demosaicing has not been adjusted accordingly across various color patterns in existing deep learning-based methods [23]–[26]. All color patterns in an image are interpolated by a same kernel slidingly. It is difficult to realize adaptive interpolation due to weights sharing in the convolution.

We proposed a novel convolution method to use incomplete weight sharing for sliding kernels, which is called spatial varying convolution (SVC) [52]. In the SVC, the kernel weights are shared across the same patterns but different across different patterns. Considering the four color patterns in Fig. 6, at least four different kernels need to be included in the SVC, as shown in Fig. 7.

Our proposed SVC [52] is extended to a more complicated scenario. In a Bayer image captured by the dual-time shooting, the exposure time also varies along with image pixels. As shown in Fig. 3, the brightness differences, caused by the varying exposure times, cannot be eliminated by converting a Bayer image into a Bayer radiance image via the CRFs. Considering both varying colors and varying brightness, there are totally eight patterns in a  $5 \times 5$  receptive field, as shown in Fig. 8. The darker and lighter colors represent the short and long exposed radiation values, respectively. Therefore, to adapt better to more kinds of patterns, the SVC is modified into a more complicated version, which includes eight interpolation kernels overall, as shown in Fig. 9. The improved SVC is more robust to the varying patterns in a dual-time Bayer image. The details of the improved SVC are given in the following text.

Let  $x_{k,v}$  and  $y_{k,v}$  be the pixels at position  $(k, v)$  in the input and output of SVC, respectively.  $w^{(1)}, w^{(2)}, w^{(3)}, w^{(4)}, w^{(5)}$ ,

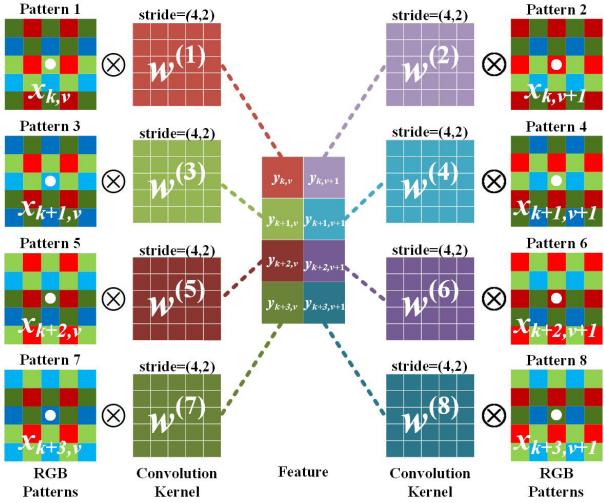


Fig. 9. Illustration of spatially varying convolution (SVC) based on 8 patterns.

$w^{(6)}$ ,  $w^{(7)}$ , and  $w^{(8)}$  represent different convolution kernels. The proposed SVC is then defined as follows:

$$\left\{ \begin{array}{l} y_{k,v} = \sum_{i=-2}^2 \sum_{j=-2}^2 (w_{i,j}^{(1)} \times x_{k+i,v+j}) \\ y_{k,v+1} = \sum_{i=-2}^2 \sum_{j=-2}^2 (w_{i,j}^{(2)} \times x_{k+i,v+1+j}) \\ y_{k+1,v} = \sum_{i=-2}^2 \sum_{j=-2}^2 (w_{i,j}^{(3)} \times x_{k+1+i,v+j}) \\ y_{k+1,v+1} = \sum_{i=-2}^2 \sum_{j=-2}^2 (w_{i,j}^{(4)} \times x_{k+1+i,v+1+j}) \\ y_{k+2,v} = \sum_{i=-2}^2 \sum_{j=-2}^2 (w_{i,j}^{(5)} \times x_{k+2+i,v+j}) \\ y_{k+2,v+1} = \sum_{i=-2}^2 \sum_{j=-2}^2 (w_{i,j}^{(6)} \times x_{k+2+i,v+1+j}) \\ y_{k+3,v} = \sum_{i=-2}^2 \sum_{j=-2}^2 (w_{i,j}^{(7)} \times x_{k+3+i,v+j}) \\ y_{k+3,v+1} = \sum_{i=-2}^2 \sum_{j=-2}^2 (w_{i,j}^{(8)} \times x_{k+3+i,v+1+j}), \end{array} \right. \quad (3)$$

where  $k = 0, 4, 8, \dots, 4n_1 \leq H$ , and  $v = 0, 2, 4, \dots, 2n_2 \leq W$ .  $H$  and  $W$  are the height and width of input or output, respectively. Benefiting from the SVC, the data pattern is the same for each convolution kernel, which can reduce the burden of network learning. **The proposed SVC is one distinctive component of the proposed CNN.**

Note that the difference between the SVC and existing methods is whether it is the convolution with complete weight sharing. Specifically, considering the algorithm in [23], the Bayer image is convolved at stride 1. The convolutional kernel receives four kinds of patterns during sliding, as shown in Figure 6. According to the principle of convolution, kernel weights are shared, by which different data patterns are convolved by the same convolutional kernels. Existing CNN-based demosaicing algorithms adopt two methods. One method is to set the stride as 2 or 4 [24], by which the kernel receives only one kind of data pattern. Another method is to rearrange

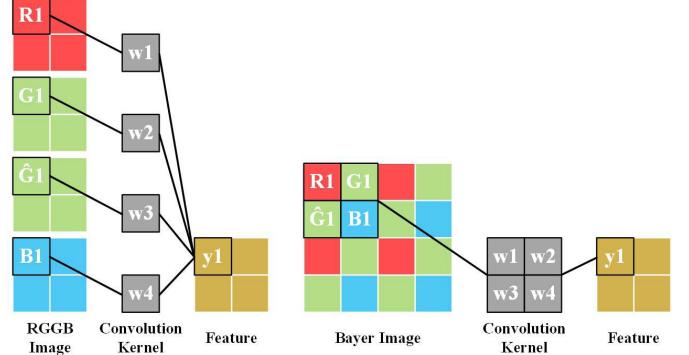


Fig. 10. Illustration of convolving an RGGB image which uses  $1 \times 1$  convolution kernels at stride 1, and convolving a Bayer image which uses  $2 \times 2$  convolution kernels at stride 2. The two operations are equivalent.

the single-channel Bayer image into a four-channel RGGB image, whose length and width are reduced by half. Then the rearranged image is convolved with a stride of 1 [25], [26]. The above two methods are actually equivalent. For example, convolving an RGGB image with a  $1 \times 1$  kernel at stride 1 is equivalent to convolving a Bayer image with a  $2 \times 2$  kernel at stride 2, as shown in Fig. 10. Both methods focus on interpolating only one kind of data pattern, where the other three kinds of patterns are not considered. While, the SVC can be adjusted accordingly to the spatial change of the data pattern.

Moreover the SVC can have flexible variants when the data pattern and SVE changes. The SVC is only used in the first layer of CNN, as shown in Fig. 2. Thus, compared to the weight-sharing convolution method, the SVC improves the imaging ability of CNN via minimal additional parameters.

### C. Reconstruction Branch

The reconstruction branch (RB) shares a similar philosophy to the networks utilized in image-to-image translation tasks, such as image demosaicing [55], super-resolution imaging [56], and image denoising [57]. The RB is composed of several residual blocks, as shown in Fig. 2. Each block is realized by an operation of convolution-ReLu-convolution and an identical mapping. Specifically, the kernel size in each convolutional layer is set to  $3 \times 3$ . The stride and padding are both set to 1 to maintain the feature resolution during propagation. The structure of RB is shown in Fig. 2.

### D. Exposure-Guidance Method

Over- and under-exposed areas exist in the rows with long- and short-exposure times, respectively. These ill-exposed areas are usually dominated by saturation noise. These areas lead to unreliable and poor information in the radiance map  $E$ , which turns into visible boundaries at the corresponding pixel positions, as shown in Fig. 5. These ill-exposed areas can interfere with the CNN and are thus desired to be detected. To overcome this, an exposure-guidance method is proposed, which consists of exposure guidance mask and exposure guidance branch (EGB). **The exposure-guidance method is the other distinctive component of the proposed CNN.**

1) *Exposure-Guidance Mask*: In conventional HDRI algorithms [5], [8], [11], [12], a threshold value is usually pre-defined on a N-bit image to identify the ill-exposed pixels. To simulate this artificial selection in a CNN-based algorithm, a exposure-guidance mask  $M$  is computed using the following equation:

$$m_{ij} = \begin{cases} 0, & \Delta t_{ij} = \tau_L \text{ and } z_{ij} \geq (1 - \alpha)(2^N - 1) \\ 0, & \Delta t_{ij} = \tau_S \text{ and } z_{ij} \leq \alpha(2^N - 1) \\ 1, & \text{otherwise,} \end{cases} \quad (4)$$

where the  $m_{ij}$  is the value at position  $(i, j)$  in  $M$ . N is the number of bits for the input image and  $\alpha$  is a percentage constant. Note that  $M$  has the same size as  $Z$  and  $E$ . Then,  $M$  can be easily fed into the network along with the radiance map  $E$ , which can be considered as a prior of knowledge to help the network know which area is ill-exposed. Based on the rule of thumb in [12] and the sensitivity analysis of  $\alpha$  in Section IV-C, threshold value  $\alpha$  is empirically selected as 3.92%.

2) *Exposure-Guidance Branch*: The mask can be concatenated or element-wise multiplied with a Bayer radiance image as the network input, providing auxiliary prior information. However, the features of the radiance map are not guided by the mask in a deep-level manner because of this early fusion. Inspired by [28], [29], an EGB is proposed to guide the HDR reconstruction as an auxiliary branch, as shown in Fig. 2. Prior information can be extracted by this branch, which makes the CNN explainable.

During propagation, the exposure-guidance mask is first multiplied with the Bayer radiance image  $E$ . Thus, the poor information from the ill-exposed area will be filtered out, such that the EGB will tend to utilize the features extracted from the well-exposed area. Then, the features of the EGB are embedded into the RB in a multilevel manner to realize compensation, in which information is fused in a deeper level. Finally, the network can pay considerable attention to the well-exposed areas, which provide accurate HDR information.

The proposed EGB consists of  $n$  blocks, and each block includes two convolutional layers and an activation layer. To reduce the computational cost, the output channels of the first convolutional layer are compressed in each block.

$E$  can be simply multiplied by  $M$  as the input of RB to make the network focus on the well-exposed areas only. The experimental results in Section IV-C show that this method does not work as expected. The possible reason is that the raw Bayer pattern is destroyed, and the irregular data makes the network difficult to deal with.

#### E. Loss Function

The  $L_1$  loss function is widely used in deep learning-based low-level image processing [58], and it is formalized as

$$\mathcal{L}_{l1} = \frac{1}{3 \times H \times W} \sum_{i=1}^H \sum_{j=1}^W |\mathbf{r}_{ij} - \tilde{\mathbf{r}}_{ij}|, \quad (5)$$

where  $\mathbf{r}_{ij}$  and  $\tilde{\mathbf{r}}_{ij}$  represent two 3D ( $R, G, B$ ) vectors at position  $(i, j)$  in the image. The resolution of the ground truth  $R$  and generated HDR image  $\tilde{R}$  is  $H \times W$ .

To reduce the color deviation between  $R$  and  $\tilde{R}$ , the color loss  $\mathcal{L}_c$  is introduced as follows [59]:

$$\mathcal{L}_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W (1 - \cos(\mathbf{r}_{ij}, \tilde{\mathbf{r}}_{ij})), \quad (6)$$

where  $\cos(\mathbf{r}_{ij}, \tilde{\mathbf{r}}_{ij})$  represents the cosine similarity of  $\mathbf{r}_{ij}$  and  $\tilde{\mathbf{r}}_{ij}$ .  $\mathcal{L}_c$  is sensitive to color difference. The overall loss function is given as

$$\mathcal{L}_{total} = \mathcal{L}_{l1} + \lambda \mathcal{L}_c, \quad (7)$$

where  $\lambda$  is a constant, and its value is selected as 0.1.

## IV. EXPERIMENTAL RESULTS

### A. Implementation Details

1) *Datasets Description*: Experiments are conducted on two datasets, VETHDR-Nikon dataset and VETHDR-Canon dataset, where the different exposures are achieved by varying the exposure times (VET) instead of changing the ISO. 500 pairs of images are included in each dataset. Each pair consists of two full-resolution images  $I_L$  and  $I_S$ , with long and short exposure times, respectively. The exposure time ratio and ISO are fixed as 16 and 800, during shooting respectively. All the images are resized to  $480 \times 480$ . Among each of the dataset, 300 pairs are used for training, 100 pairs for validation and 100 pairs for test. To simulate the input dual-time Bayer image  $Z$ , the pixels on every two rows are alternatively sampled from  $I_S$  and  $I_L$ . The ground-truth HDR image  $Y$  is obtained by merging the full-resolution images  $I_S$  and  $I_L$  via the method of [27].

The VETHDR-Nikon dataset is from [60], [61], collected by a Nikon 7200 camera. It consists of original images in 8-bit color JPEG files format. The corresponding CRF [27] is recorded in the dataset. The Bayer Radiance Image can be calculated from the raw input image  $Z$  through the nonlinear CRF.

The VETHDR-Canon dataset is collected by a Canon 5D4 camera in this paper. It contains 16-bit color images with the original digital counts for each of the RGB channels, which are generated from the 16-bit RAW image files by the method in [62]. Since the RAW files typically have a nearly linear CRF, the Bayer Radiance Image can be calculated from the raw input image  $Z$  through a simple linear transformation.

2) *Comparison Description*: In terms of qualitative comparison, the results are visualized by sequentially executing the tone mapping algorithm [63] and the white balance algorithm [64] on the radiance map. In terms of quantitative comparison, we choose HDR-MAE, HDR-MSE, HDR-VDP, HDR-PSNR-RGB, HDR-SSIM-RGB, HDR-PSNR-Y, and HDR-SSIM-Y as the evaluation metrics. These metrics are measured on the radiance maps. For the HDR-MAE and HDR-MSE, lower is better. For the HDR-VDP, HDR-PSNR-RGB, HDR-SSIM-RGB, HDR-PSNR-Y, and HDR-SSIM-Y, higher is better. Among them, a perceptually uniform (PU) encoding [65] is utilized to enable PSNR and SSIM to be used to evaluate the quality of HDR images. In addition, the model parameters (P) and the floating-point operations (FLOPs) are

TABLE I  
SUMMARY OF VARIOUS CONVOLUTIONAL LAYERS FOR BAYER IMAGES IN EXISTING ALGORITHMS

Method	Opt-base	Opt-2-2 [24]	Opt-4-2 [25]	Opt-4-4 [25]	Opt-RGGB [26], [27]
<b>Input Shape</b>	$h \times w \times 1$	$\frac{h}{2} \times \frac{w}{2} \times 4$			
<b>Kernel Size</b>	$3 \times 3$	$2 \times 2$	$4 \times 4$	$4 \times 4$	$3 \times 3$
<b>Stride</b>	(1,1)	(2,2)	(2,2)	(4,4)	(1,1)
<b>Upsampling Layer</b>	Unused	2	2	4	2
<b>Output Shape</b>	$h \times w \times 64$				

TABLE II

QUANTITATIVE COMPARISON AMONG THE BASELINE, EXISTING OPERATIONS, AND OUR SVC. THE BEST RESULTS ARE SHOWN IN BOLD, AND THE SECOND-BEST RESULTS ARE SHOWN IN BLUE. THE RESULTS COME FROM THE VETHDR-NIKON TEST SET

Method	HDR-MAE	HDR-MSE	HDR-VDP	HDR-PSNR-RGB	HDR-SSIM-RGB	HDR-PSNR-Y	HDR-SSIM-Y	P [ $10^6$ ]	FLOPs [ $10^{11}$ ]
Opt-base	2.982	53.874	65.59	41.38	0.9786	43.99	0.9860	<b>1.221</b>	<b>2.813</b>
Opt-2-2	2.911	41.480	<b>65.72</b>	41.72	0.9787	44.12	0.9859	<b>1.222</b>	<b>2.813</b>
Opt-4-2	2.962	42.596	65.55	41.61	0.9782	44.02	0.9856	1.225	2.814
Opt-4-4	2.983	41.916	65.57	41.62	0.9783	44.00	0.9856	1.238	2.814
Opt-RGGB	2.903	41.159	65.68	41.79	0.9788	44.20	0.9859	1.230	2.817
SVC-D	<b>2.889</b>	<b>40.744</b>	<b>65.74</b>	<b>41.84</b>	<b>0.9790</b>	<b>44.21</b>	<b>0.9860</b>	1.227	2.816
SVC	<b>2.881</b>	<b>40.052</b>	65.69	<b>41.91</b>	<b>0.9791</b>	<b>44.33</b>	<b>0.9861</b>	1.234	2.816

provided as reference. All evaluation metrics are calculated on the HDR radiance image, and which are explained as follows:

- **HDR-MAE:** The average absolute value based on the PU encoding, which calculates the difference between the test image and the reference image.
- **HDR-MSE:** The average square value based on the PU encoding, which calculates the difference between the test image and the reference image.
- **HDR-VDP:** A high dynamic range visible difference predictor in version 2.2.2. Its quality correlate score can be used to evaluate image differences [66].
- **HDR-PSNR-RGB:** Peak signal-to-noise ratio based on the PU encoding, which calculates the difference of the image in the R, G, and B channels.
- **HDR-SSIM-RGB:** Structural similarity [67] based on the PU encoding, which calculates the difference of the image in the R, G, and B channels.
- **HDR-PSNR-Y:** Peak signal-to-noise ratio based on the PU encoding, which calculates the difference of the image in the Y channel.
- **HDR-SSIM-Y:** Structural similarity [67] based on the PU encoding, which calculates the difference of the image in the Y channel.
- **P:** Number of parameters in CNN.
- **FLOPs:** The number of floating-point operations required for network to generate an image.

3) *Training Details:* The RB and EGB are realized using 16 blocks. We randomly sample  $128 \times 128$  patch from each input during training. We set the batch size and the number of iterations to 16 and  $2 \times 10^5$ , respectively. We set  $\lambda$  in Equation (7) to 0.1. The Adam optimizer with  $\beta_1 = 0.9$  and  $\beta_2 = 0.99$  is used for optimization. The learning rate is initially set as  $2 \times 10^{-4}$  and finally decreased to  $1 \times 10^{-7}$  through a cosine annealing schedule. Each model is trained on

an NVIDIA GTX 1080Ti GPU for approximately two days. All the experiments are implemented using PyTorch.

### B. Analysis of SVC

To demonstrate the effectiveness of the SVC, we compare the SVC with many other alternative methods qualitatively and quantitatively. We also perform a quantitative comparison of SVCS with different convolution kernel sizes.

1) *Alternative Methods:* The SVC is a flexible solution for processing a dual-time Bayer image, which ensures that pixels of different color patterns are convolved using different kernels. When the varying exposures caused by the dual-time are not considered, the SVC can be degraded to SVC-D [52]. We further increase the kernel size of SVC-D to  $5 \times 5$ , as shown in Fig. 7. The SVC-D is equivalent to adding the following conditions to Equation (3).

$$\begin{cases} w_{i,j}^{(1)} = w_{i,j}^{(5)} & w_{i,j}^{(2)} = w_{i,j}^{(6)} \\ w_{i,j}^{(3)} = w_{i,j}^{(7)} & w_{i,j}^{(4)} = w_{i,j}^{(8)}, \end{cases} \quad (8)$$

In addition to the proposed SVC and SVC-D in this paper, a few special convolutional layers in deep learning have been utilized to process a special Bayer image in advance [23]–[26]. All these layers/methods are summarized in Table I. The first column indicates the method characteristics, which are explained as follows:

- **Input Shape:** The shape of the input Bayer image, including the height, width, and number of channels.
- **Kernel Size:** The size of kernel in the first convolutional layer.
- **Stride:** The stride during kernel sliding in the first convolutional layer, including the vertical and horizontal strides.
- **Upsampling Layer:** The upsampling layer is realized by the sub-pixel convolution [68]. 2 and 4 indicate the scaling factor.

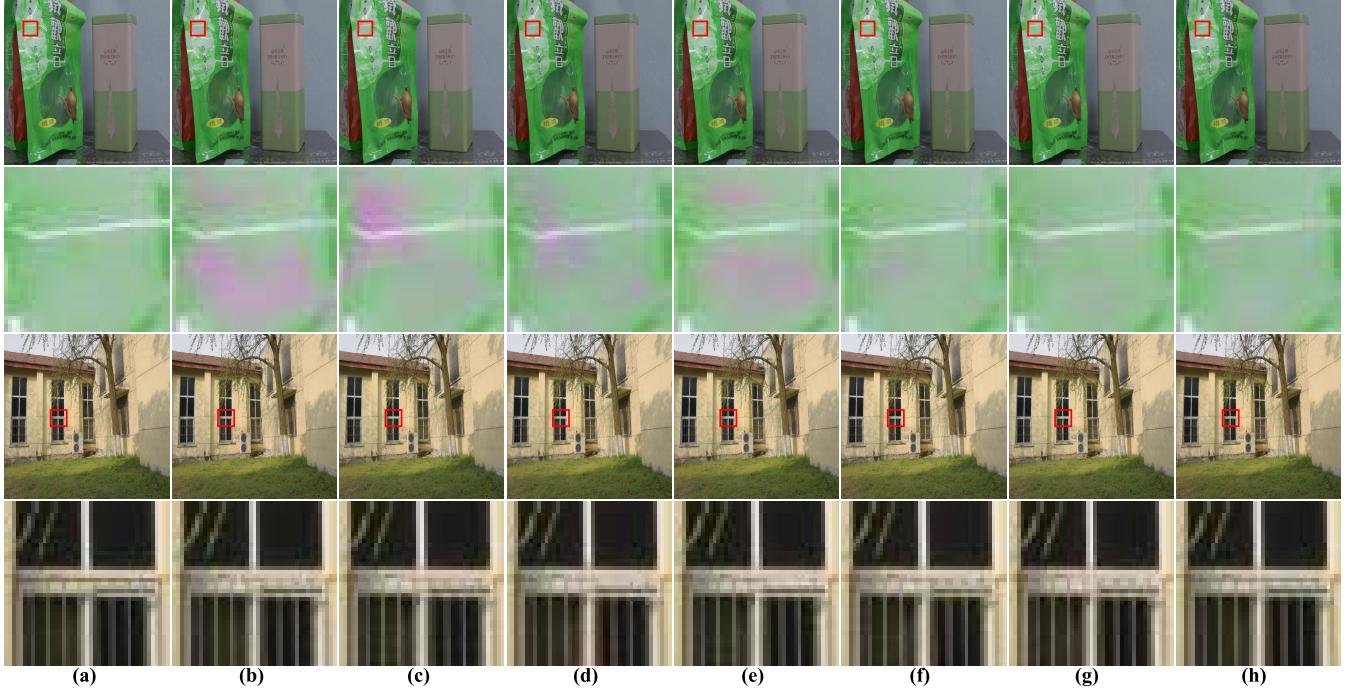


Fig. 11. Qualitative comparison among the baseline, existing convolutional layers, and our SVC. (a) Ground truth, (b) Opt-base, (c) Opt-2-2, (d) Opt-4-2, (e) Opt-4-4, (f) Opt-RGGB, (g) SVC-D, (h) SVC. The results come from the VETHDR-Nikon test set.

TABLE III

QUANTITATIVE COMPARISON AMONG SVC-3, SVC-5, AND SVC-7. THE BEST RESULTS ARE SHOWN IN BOLD, AND THE SECOND-BEST RESULTS ARE SHOWN IN BLUE. THE RESULTS COME FROM THE VETHDR-NIKON TEST SET

Method	HDR-MAE	HDR-MSE	HDR-VDP	HDR-PSNR-RGB	HDR-SSIM-RGB	HDR-PSNR-Y	HDR-SSIM-Y	P [10 <sup>6</sup> ]	FLOPs [10 <sup>11</sup> ]
Baseline	2.982	53.874	65.59	41.38	0.9786	43.99	0.9860	<b>1.221</b>	<b>2.813</b>
SVC-3	2.941	40.809	<b>65.70</b>	41.80	0.9788	44.19	0.9859	<b>1.225</b>	<b>2.813</b>
SVC-5	<b>2.881</b>	<b>40.052</b>	<b>65.69</b>	<b>41.91</b>	<b>0.9791</b>	<b>44.33</b>	<b>0.9861</b>	1.234	<b>2.816</b>
SVC-7	<b>2.902</b>	<b>40.548</b>	65.68	<b>41.87</b>	<b>0.9790</b>	<b>44.28</b>	<b>0.9860</b>	1.246	2.819

TABLE IV

QUANTITATIVE COMPARISON BETWEEN THE BASELINE AND FOUR DIFFERENT METHODS OF USING EXPOSURE-GUIDANCE MASKS. THE BEST RESULTS ARE SHOWN IN BOLD, AND THE SECOND-BEST RESULTS ARE SHOWN IN BLUE. THE RESULTS COME FROM THE VETHDR-NIKON TEST SET

Method	HDR-MAE	HDR-MSE	HDR-VDP	HDR-PSNR-RGB	HDR-SSIM-RGB	HDR-PSNR-Y	HDR-SSIM-Y	P [10 <sup>6</sup> ]	FLOPs [10 <sup>11</sup> ]
Baseline	2.982	53.874	65.59	41.38	0.9786	43.99	0.9860	<b>1.221</b>	<b>2.813</b>
Multiplication	3.314	87.445	64.74	40.42	0.9757	43.24	0.9843	1.886	4.345
Concatenation	2.904	50.018	<b>65.75</b>	41.50	0.9790	44.11	0.9862	1.886	4.346
RB+EACB [53]	<b>2.900</b>	<b>48.224</b>	65.68	<b>41.54</b>	<b>0.9790</b>	<b>44.12</b>	0.9862	2.072	4.773
RB+EGB	2.861	47.835	<b>65.83</b>	<b>41.72</b>	<b>0.9794</b>	<b>44.32</b>	<b>0.9865</b>	<b>1.850</b>	<b>4.263</b>

- Output Shape:** The shape of the output produced by the first convolutional layer, including the height, width and number of channels.

Table I presents that two types of convolutional layers can be used for processing a Bayer image. One type is that a convolution kernel of even size extracts the color information with a stride of even. Then, the downsampled output is resized to the original resolution by using the upsampling layer, such as Opt-2-2 [23], Opt-4-2 [24], Opt-4-4 [24], and Opt-RGGB [25], [26]. Overall, all these convolutional layers are designed toward the same goal, i.e., to make the color pattern

convolved using the kernel be the same. Opt-base represents the general convolutional layer with a kernel size of  $3 \times 3$ .

The convolutional layers in Table I are combined with the RB to compare their performance. The results of 7 evaluation metrics on the radiance map are reported in Table II. The number of parameters is also provided as reference. The SVC and SVC-D surpass other convolutional layers, which demonstrates that the network learning can be affected by the varying color patterns and the SVC is a more robust solution. Moreover, the leading performance of SVC over SVC-D can prove the necessity of further consideration of brightness difference in a dual-time Bayer image.

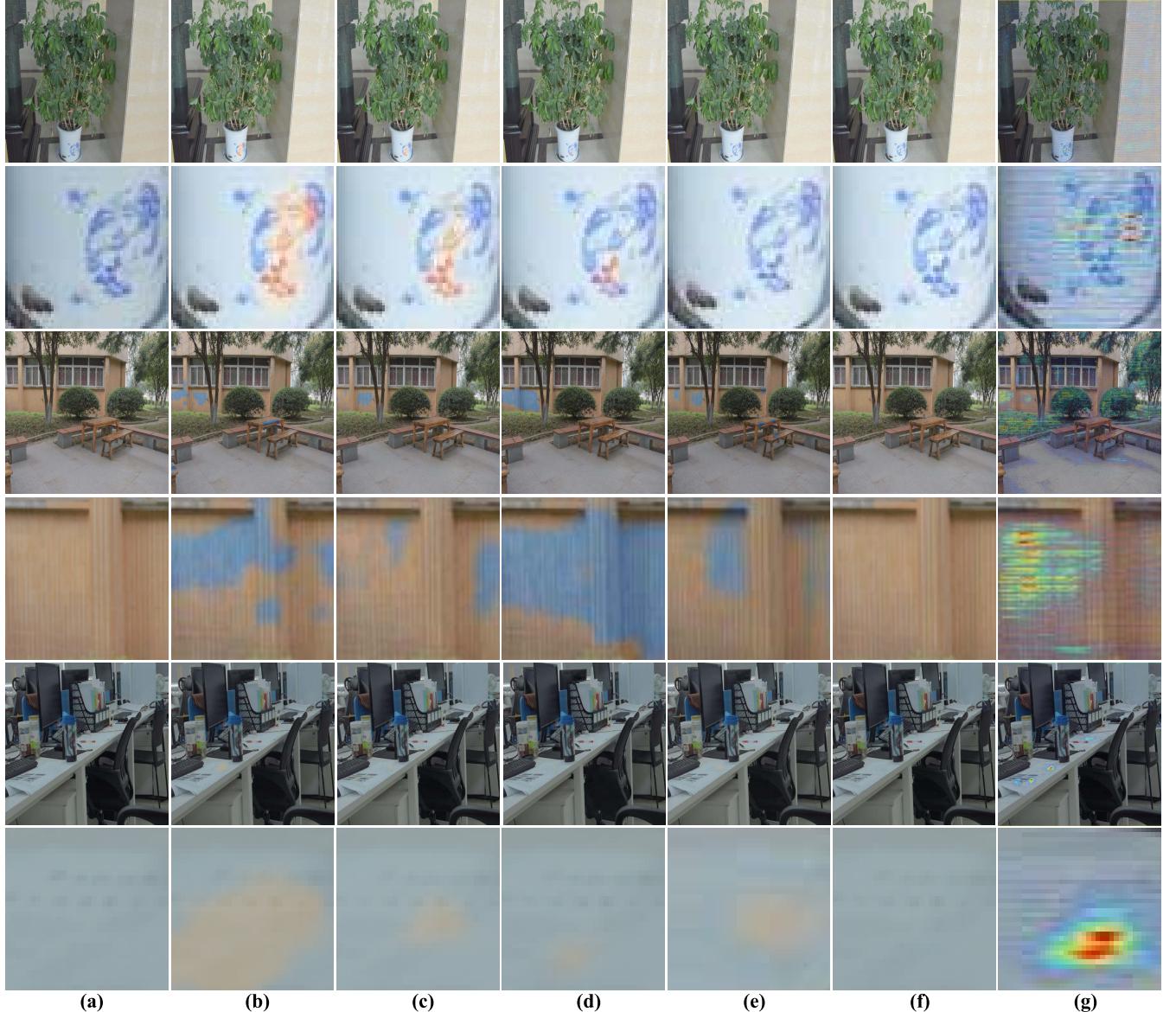


Fig. 12. Qualitative comparison between the baseline and three different methods of using exposure-guidance masks. (a) Ground truth, (b) baseline, (c) the method of using multiplication, (d) the method of concatenation, (e) RB+EACB [52], (f) RB+EGB, (g) overlay of EGB’s feature output and tone-mapped ground truth. The results come from the VETHDR-Nikon dataset.

TABLE V  
 $\beta$  IN THE TRAINED RB+EGB

$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$	$\beta_6$	$\beta_7$	$\beta_8$
0.6463	0.3757	0.6446	0.8103	0.7389	0.6234	0.4541	0.6906
$\beta_9$	$\beta_{10}$	$\beta_{11}$	$\beta_{12}$	$\beta_{13}$	$\beta_{14}$	$\beta_{15}$	$\beta_{16}$
0.5762	0.6027	0.6895	0.7473	0.8965	0.9381	0.9597	0.9734

The qualitative results are shown in Fig. 11. In the first row, (b), (c), (d), (e), and (f) fail in estimating correct colors. In the second row, clearer edges are recovered using (h), whereas other results have different degrees of blurring on edges. The SVC can make the network restore correct colors and textures.

2) *Kernel Size*: The SVCs with different kernel sizes are also experimented. In Table III, SVC-3, SVC-5, and SVC-7

indicate a kernel size of  $3 \times 3$ ,  $5 \times 5$ , and  $7 \times 7$ , respectively. All the SVCs outperform the original convolution (Opt-base). It can be seen that simply increasing the kernel size of SVC brings little gains on performance. Thus, the power of SVC lies in the design philosophy, not the increasing parameters.

### C. Analysis of Exposure Guidance

To demonstrate the effectiveness of the exposure-guidance method, we compare the exposure-guidance method with many other alternative methods qualitatively and quantitatively. We also perform a quantitative comparison of EGBs with different threshold values  $\alpha$ . Moreover, we analyze the working mechanism of EGB through feature map and  $\beta_i$ .

1) *Alternative Methods*: The exposure-guidance mask is necessary to incorporate into the main network RB by adopting

TABLE VI

QUANTITATIVE COMPARISON BASELINE AND RB+EGB WITH THREE DIFFERENT WEIGHTS. THE BEST RESULTS ARE SHOWN IN BOLD, AND THE SECOND-BEST RESULTS ARE SHOWN IN BLUE. THE RESULTS COME FROM THE VETHDR-NIKON TEST SET

Method	HDR-MAE	HDR-MSE	HDR-VDP	HDR-PSNR-RGB	HDR-SSIM-RGB	HDR-PSNR-Y	HDR-SSIM-Y	P [10 <sup>6</sup> ]	FLOPs [10 <sup>11</sup> ]
Baseline	2.982	53.874	65.59	41.38	0.9786	43.99	0.9860	<b>1.221</b>	<b>2.813</b>
Robertson's function [70]	2.926	<b>42.686</b>	<b>65.86</b>	41.66	<b>0.9791</b>	44.26	<b>0.9863</b>	1.850	4.263
Debevec's function [28]	<b>2.879</b>	<b>45.644</b>	<b>65.85</b>	<b>41.67</b>	0.9790	<b>44.35</b>	0.9863	1.850	4.263
Our function	<b>2.861</b>	47.835	65.83	<b>41.72</b>	<b>0.9794</b>	<b>44.32</b>	<b>0.9865</b>	1.850	4.263

TABLE VII

QUANTITATIVE COMPARISON OF MODELS WITH DIFFERENT COMPONENTS. THE BEST RESULTS ARE SHOWN IN BOLD, AND THE SECOND-BEST RESULTS ARE SHOWN IN BLUE. THE RESULTS COME FROM THE VETHDR-NIKON TEST SET

Method	HDR-MAE	HDR-MSE	HDR-VDP	HDR-PSNR-RGB	HDR-SSIM-RGB	HDR-PSNR-Y	HDR-SSIM-Y	P [10 <sup>6</sup> ]	FLOPs [10 <sup>11</sup> ]
RB	2.982	53.874	65.59	41.38	0.9786	43.99	0.9860	<b>1.221</b>	<b>2.813</b>
RB+SVC	2.881	<b>40.052</b>	65.69	<b>41.91</b>	0.9791	<b>44.33</b>	0.9861	<b>1.234</b>	<b>2.816</b>
RB+EGB	<b>2.861</b>	47.835	<b>65.83</b>	41.72	<b>0.9794</b>	44.32	<b>0.9865</b>	1.850	4.263
RB+2xSVC+EGB	<b>2.777</b>	<b>38.713</b>	<b>66.02</b>	<b>42.15</b>	<b>0.9797</b>	<b>44.56</b>	<b>0.9865</b>	1.912	4.352

the EGB. Besides the EGB, there are three other ways in incorporating the exposure-guidance mask. The first method is to multiply the exposure-guidance mask and the Bayer radiance image to obtain an input matrix with a size of  $h \times w \times 1$ , in which the value of ill-exposed pixels becomes 0. This matrix is directly fed into the main network RB. The second method is to concatenate the exposure-guidance mask with the Bayer radiance image to generate a  $h \times w \times 2$  matrix, which is then fed into the RB. The third method is to use the exposedness-aware compensation branch (EACB) to achieve deep-level fusion of masks and features [52]. To make a fair comparison, for the first two method, we set the number of residual blocks in the RB as 25. For the third method and the EGB, we set both the number of blocks in RB and EGB as 16.

The 7 evaluation metrics on the test set are reported in Table IV. Among them, the result of the multiplication-based method is worst. This implies that the information is not utilized after resetting the ill-exposed pixels to 0. The raw RGB pattern in the Bayer image is destroyed, and the irregular data make the network difficult to learn. The concatenation-based method is improved compared with the baseline on the 7 evaluation metrics, indicating that the awareness of ill-exposed pixels can improve the performance. However, the prior of the exposure-guidance mask is not incorporated into the network in a deep-level way, in which the performance improvements are limited. The EACB [52] is able to incorporate the prior knowledge of ill-exposed pixels in the feature level. The proposed EGB is a modification of EACB and compensates the RB in a both multi-level and feature-level way. Since the EACB-based method [52] only compensates RB once at the deepest level, the lack of compensation times make the performance improvement less obvious than the concatenation-based method. The method of using the EGB achieves the best results on all four indicators, which strongly proves the effectiveness of fusing the prior in a deep- and multi-level way. The EGB can focus only on the well-exposed area, and can exploit accurate information for HDR reconstruction. This network structure design is also more explainable.

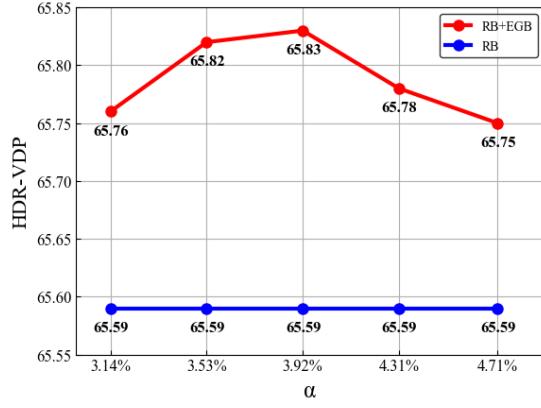


Fig. 13. The HDR-VDP score of the models under different  $\alpha$ . The results come from the VETHDR-Nikon test set.

Moreover, the original information of the dual-time Bayer image is preserved. The EGB also has the highest computational efficiency.

The qualitative results are shown in Fig. 12. The EGB-based method performs evidently better on the four sets of images, whereas the other methods cause different degrees of unnatural color distortions.

2) *Threshold Selection*: In order to observe the effect of different  $\alpha$  on the EGB, we chose 5 values for testing. The selected values are all around the empirical value 3.92% given in [12]. The HDR-VDP scores of RB+EGB and RB under different  $\alpha$  are shown in Figure 13. It can be observed from the figure that the performance of EGB is less affected by the changing of  $\alpha$ , and all the scores of RB+EGB significantly surpass the baseline RB. This proves the robustness of the exposure-guidance method. Since the HDR-VDP score of RB+EGB is the highest when  $\alpha = 3.92\%$ , 3.92% is selected as the threshold.

3) *Choice of Weight Function*: Equation (4) is compared with two commonly used weight functions, Debevec's function [27] and Robertson's function [69]. The results of the baseline and RB+EGB with three different weight functions

TABLE VIII

QUANTITATIVE COMPARISON OF MODELS WITH DIFFERENT COMPONENTS. THE BEST RESULTS ARE SHOWN IN BOLD, AND THE SECOND-BEST RESULTS ARE SHOWN IN BLUE. THE RESULTS COME FROM THE VETHDR-CANON TEST SET

Method	HDR-MAE	HDR-MSE	HDR-VDP	HDR-PSNR-RGB	HDR-SSIM-RGB	HDR-PSNR-Y	HDR-SSIM-Y	P [10 <sup>6</sup> ]	FLOPs [10 <sup>11</sup> ]
RB	2.290	26.957	70.41	42.01	0.9889	43.70	0.9922	<b>1.221</b>	<b>2.813</b>
RB+SVC	2.280	25.408	70.43	42.20	0.9890	43.75	0.9921	<b>1.234</b>	<b>2.816</b>
RB+EGB	<b>2.151</b>	<b>24.382</b>	<b>70.67</b>	<b>42.47</b>	<b>0.9895</b>	<b>44.06</b>	<b>0.9925</b>	1.850	4.263
RB+2xSVC+EGB	<b>2.146</b>	<b>23.572</b>	<b>70.64</b>	<b>42.57</b>	<b>0.9897</b>	<b>44.12</b>	<b>0.9926</b>	1.912	4.352

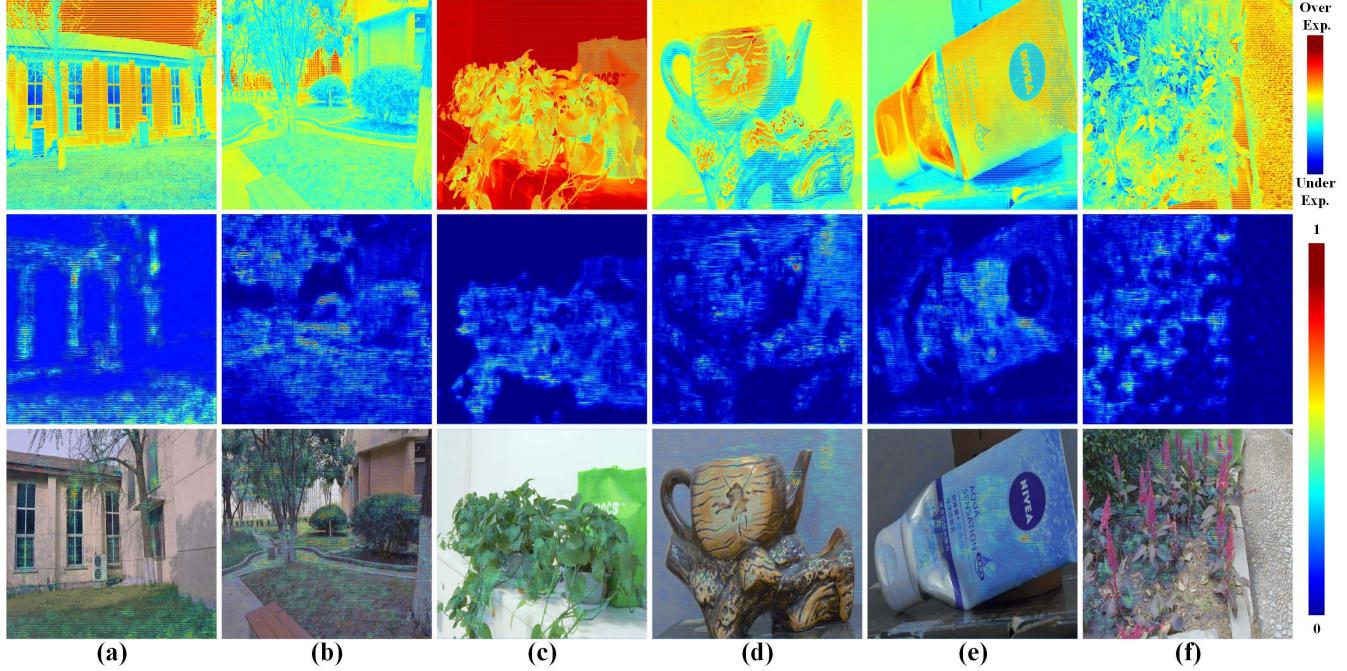


Fig. 14. Jet color map. The first row shows the input dual-time Bayer radiance image. The second row shows the output feature from the last layer of EGB. The third row shows the overlay of the EGB’s feature output and tone-mapped ground truth. The results come from the VETHDR-Nikon test set.

TABLE IX

THE FLOPs AND RUNNING TIME REQUIRED TO PROCESS IMAGES AT DIFFERENT RESOLUTIONS WITH RB+2xSVC+EGB

Input Resolution	FLOPs [10 <sup>11</sup> ]	Time(ms)
120 × 120	0.272	17.943
240 × 240	1.088	71.772
360 × 360	2.448	161.486
480 × 480	4.352	287.086
600 × 600	6.800	448.573
720 × 720	9.793	645.945
840 × 840	13.329	879.202
960 × 960	17.409	1148.346

are shown in Table VI. The RB+EGB improves significantly from the baseline no matter what weight functions are adopted. This demonstrates that the gains are mainly coming from the design of EGB. Compared with the weight functions in [27], [69], the Equation (4) achieves comparable performances. The simple and concise Equation (4) is selected as the final weight function.

4) *Working Mechanism*: The learned parameter  $\beta$  in Fig. 2 is reported in Table V. All values are between 0.3757 and 0.9734, which means that the EGB provides useful information

for the restoration of full-resolution HDR images. The  $\beta$  tends to increase as the network deepens, indicating that the compensated information from the EGB is important in a deep level.

The EGB’s output is also visualized. Sixty-four output feature maps are clustered into one feature map through principal component analysis (PCA). Then, the clustered feature map is normalized into 0 to 1 and visualized using a jet color map, as shown in the second row of Fig. 14.

The first row shows the input dual-time Bayer radiance image, where the horizontal stripes are caused by ill exposing. The third row demonstrates the overlay of the EGB’s output and tone-mapped ground truth. The EGB’s features are concentrated on the well-exposed areas. For the ill-exposed areas, e.g., the sky in column (a) and the building in the distance in column (b), few activations occur. Therefore, the EGB can focus on extracting features from the well-exposed areas to compensate for the main branch, which avoids the effect of saturation noise caused by ill exposing.

More visualization examples are shown in the last column of Fig. 12. All the other methods lead to unnatural color distortions. On the contrary, the compensation information from the EGB enables accurate HDR restoration. This result proves

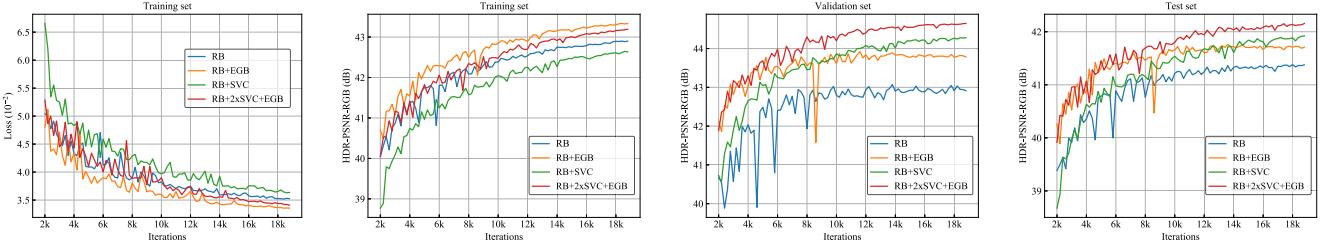


Fig. 15. The training plot on the training set, validation set, and test set. The results come from the VETHDR-Nikon test set.

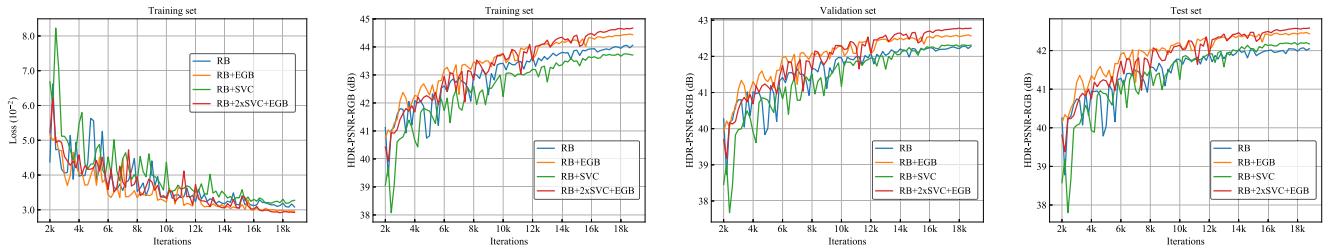


Fig. 16. The training plot on the training set, validation set, and test set. The results come from the VETHDR-Canon dataset.

TABLE X

QUANTITATIVE COMPARISON BETWEEN EXISTING METHODS AND OUR COMPLETE MODEL. THE BEST RESULTS ARE SHOWN IN BOLD, AND THE SECOND-BEST RESULTS ARE SHOWN IN BLUE. THE RESULTS COME FROM THE VETHDR-NIKON TEST SET

Method	HDR-MAE	HDR-MSE	HDR-VDP	HDR-PSNR-RGB	HDR-SSIM-RGB	HDR-PSNR-Y	HDR-SSIM-Y	P [ $10^6$ ]	FLOPs [ $10^{11}$ ]
Gharbi <i>et al.</i> [24]	3.699	56.522	64.47	39.79	0.9703	42.06	0.9794	<b>0.561</b>	0.330
Xu <i>et al.</i> [26]	3.191	48.28	65.31	40.89	0.9759	43.31	0.9841	<b>1.021</b>	0.467
An <i>et al.</i> [13]	4.167	71.635	61.64	38.95	0.9659	41.43	0.9776	9.480	4.616
An <i>et al.</i> [44]	6.050	92.197	63.84	37.20	0.9676	38.95	0.9781	53.631	13.196
Akyuz <i>et al.</i> [18]	3.414	61.856	61.90	39.79	0.9692	42.46	0.9803	1.553	3.579
Suda <i>et al.</i> [19]	6.707	334.793	56.60	34.83	0.9428	37.84	0.9596	—	—
Hajisharif <i>et al.</i> [43]	23.745	1446.367	56.58	25.93	0.8305	27.24	0.8584	—	—
Serrano <i>et al.</i> [42]	49.531	4041.573	56.70	22.99	0.9239	23.59	0.9326	—	—
Xu <i>et al.</i> [53]	<b>2.855</b>	<b>40.88</b>	<b>65.81</b>	<b>41.89</b>	<b>0.9790</b>	<b>44.32</b>	<b>0.9861</b>	2.073	4.773
Ours	<b>2.777</b>	<b>38.713</b>	<b>66.02</b>	<b>42.15</b>	<b>0.9797</b>	<b>44.56</b>	<b>0.9865</b>	1.912	4.352

TABLE XI

QUANTITATIVE COMPARISON BETWEEN EXISTING METHODS AND OUR COMPLETE MODEL. THE BEST RESULTS ARE SHOWN IN BOLD, AND THE SECOND-BEST RESULTS ARE SHOWN IN BLUE. THE RESULTS COME FROM THE VETHDR-CANON TEST SET

Method	HDR-MAE	HDR-MSE	HDR-VDP	HDR-PSNR-RGB	HDR-SSIM-RGB	HDR-PSNR-Y	HDR-SSIM-Y	P [ $10^6$ ]	FLOPs [ $10^{11}$ ]
Gharbi <i>et al.</i> [24]	2.529	36.102	69.77	40.81	0.9864	42.04	0.9899	<b>0.561</b>	0.330
Xu <i>et al.</i> [26]	<b>2.310</b>	28.142	70.37	41.83	0.9886	43.31	0.9918	<b>1.021</b>	0.467
An <i>et al.</i> [13]	3.774	60.669	57.79	38.10	0.9744	39.80	0.9813	9.480	4.616
An <i>et al.</i> [44]	5.547	59.729	69.31	37.76	0.9856	38.30	0.9893	53.631	13.196
Akyuz <i>et al.</i> [18]	3.353	58.502	57.78	38.35	0.9754	40.27	0.9823	1.553	3.579
Suda <i>et al.</i> [19]	9.732	269.536	57.27	31.85	0.9392	34.23	0.9481	—	—
Hajisharif <i>et al.</i> [43]	10.982	416.847	61.07	29.72	0.8542	31.32	0.8980	—	—
Serrano <i>et al.</i> [42]	69.954	7468.299	52.54	19.06	0.8725	19.44	0.8841	—	—
Xu <i>et al.</i> [53]	2.366	<b>25.157</b>	<b>70.59</b>	<b>42.23</b>	<b>0.9895</b>	<b>43.78</b>	<b>0.9925</b>	2.073	4.773
Ours	<b>2.146</b>	<b>23.572</b>	<b>70.64</b>	<b>42.57</b>	<b>0.9897</b>	<b>44.12</b>	<b>0.9926</b>	1.912	4.352

that the compensation information can effectively improve the imaging quality of the RB.

#### D. Ablation Study

Experiments on different models are conducted to validate the necessity of each part in our proposed framework. We use an RB with 16 blocks as the baseline and compare 3 models with it. When the SVC is not used, it is replaced with Opt-base. The first one has SVC in the network without EGB. The second one has EGB in the network without SVC. The third

one is our proposed complete model, which has two SVCs and one EGB, the SVC is in front of the RB and EGB respectively.

The test results are shown in Table VII and Table VIII. They present that RB+SVC and RB+EGB have significant improvements concerning the RB from the 7 evaluation metrics points of view, which demonstrates the effectiveness of the proposed SVC and EGB. In terms of the complete model, RB+2xSVC+EGB surpasses others in image quality.

In order to verify the effectiveness of SVC and EGB more comprehensively, the training plots of RB, RB+SVC,



Fig. 17. Qualitative comparison between existing methods and our complete model. (a) Ground truth, (b) An et al. [43], (c) Akyuz et al. [17], (d) Suda et al. [18], (e) Hajisharif et al. [42], (f) Serrano et al. [41], (g) Ours. The results come from the VETHDR-Nikon test set and VETHDR-Canon test set.

RB+EGB, and RB+2xSVC+EGB are shown in Fig. 15 and Fig. 16. We can see that though RB has better convergence on training sets, it is not as robust as RB+SVC and RB+EGB on

validation and test sets. Thus, both SVC and EGB improve the generalization ability of CNN. In addition, RB+2xSVC+EGB performs the best across both validation and test sets.

### E. Speed Evaluation

In order to evaluate the running speed of our complete model RB+2xSVC+EGB when dealing with different resolutions, Table IX lists the test results of the RB+2xSVC+EGB on GTX 1080Ti. Note that the running time of this model are proportional to the FLOPs and image resolution.

### F. Comparison With Existing Algorithms

The proposed algorithm is compared with seven single-shot HDRI algorithm [12], [17], [18], [41]–[43], [52] and two one-stage joint demosaicing algorithms [23], [25] qualitatively and quantitatively on the two datasets. The denoising and super-resolution in [23], [25] are disabled because they are not required by the single-shot dual-time HDRI.

The results about 7 evaluation metrics are reported in Table X and Table XI. With the fourth least number of parameters (P), our method achieves the best performance in HDR-MAE, HDR-MSE, HDR-VDP, HDR-PSNR-RGB, HDR-SSIM-RGB, HDR-PSNR-Y, and HDR-SSIM-Y on the test set. This superiority is mainly benefited from the proposed SVC and EGB for the single-shot HDRI.

For a qualitative comparison, the tone-mapping algorithm [63] is applied to compress the generated radiance maps for display. The synthesized results and their detailed parts are shown in Fig. 17. In the first row, the color of flowers is synthesized incorrectly in (b) and (d). The texture of the flower is blurred in (e). The images in (c) and (f) are generally slightly darker than the ground truth. In the second row, the luminance of images in (c) and (e) is different from the ground-truth, and the color of images in (b) and (d) is abnormal. Evident horizontal stripes can also be observed from the results in (e) and (f).

In a nutshell, the results in (e) and (f) are likely to have horizontal stripes, which are caused by the brightness differences in the dual-time Bayer image. The results in (c) tend to have an inappropriate luminance, whereas the results in (b) and (d) have unnatural color. Different from these methods, the proposed algorithm is robust to most scenes, providing accurate HDR restoration. The overall comparison demonstrates the effectiveness of the proposed EGB and SVC.

## V. CONCLUSION REMARKS AND DISCUSSION

In this paper, a novel CNN-based method is proposed to restore high-quality HDR images at full resolution for single-shot HDRI. The proposed CNN includes two distinctive components, the SVC and the exposure-guidance method by which the CNN is more explainable. Experimental results demonstrate that the proposed algorithm outperforms a few existing algorithms. The proposed algorithm focuses on HDR images. The idea of joint demosaicing and HDRI within a single shot can avoid cumulative errors. Since single-shot HDRI has the advantage of not having to take multiple shots, the proposed algorithm can also be extended for HDR videos [21].

Note that the two distinctive components can be extended to other low-level image processing tasks. The SVC can be easily

inserted into other networks to process Bayer image with or without SVE, and can be redesigned into flexible variants when the data pattern or SVE changes. The exposure-guidance method can be used to study other HDRI problems, such as [60], [61], allowing CNN to reduce the interference of ill-exposed pixels. It is worth noting that the proposed algorithm can be further improved through the related technologies of augmentation and transfer learning. All these problems will be studied in our future research.

## REFERENCES

- [1] F. Banterle, A. Artusi, K. Debattista, and A. Chalmers, *Advanced High Dynamic Range Imaging: Theory and Practice*, 2nd ed. Natick, MA, USA: AK Peters,, Jul. 2017.
- [2] J. Han *et al.*, “Neuromorphic camera guided high dynamic range imaging,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 1730–1739.
- [3] C. LeGendre *et al.*, “Learning illumination from diverse portraits,” in *Proc. SIGGRAPH Asia Tech. Commun.*, 2020, pp. 1–4.
- [4] K. R. Prabhakar, S. Agrawal, D. K. Singh, B. Ashwath, and R. V. Babu, “Towards practical and efficient high-resolution HDR deghosting with CNN,” in *Computer Vision*. Glasgow, U.K.: Springer, Aug. 2020, pp. 497–513.
- [5] Q. Wang, W. Chen, X. Wu, and Z. Li, “Detail-enhanced multi-scale exposure fusion in YUV color space,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 8, pp. 2418–2429, Aug. 2020.
- [6] F. Kou, Z. Li, C. Wen, and W. Chen, “Multi-scale exposure fusion via gradient domain guided image filtering,” in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2017, pp. 1105–1110.
- [7] J. Cai, S. Gu, and L. Zhang, “Learning a deep single image contrast enhancer from multi-exposure images,” *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 2049–2062, Apr. 2018.
- [8] Z. Li, J. Zheng, Z. Zhu, and S. Wu, “Selectively detail-enhanced fusion of differently exposed images with moving objects,” *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4372–4382, Oct. 2014.
- [9] A. Srikantha and D. Sidibe, “Ghost detection and removal for high dynamic range images: Recent advances,” *Signal Process., Image Commun.*, vol. 27, no. 6, pp. 650–662, 2012.
- [10] O. T. Tursun, A. O. Akyuz, A. Erdem, and E. Erdem, “The state of the art in hdr deghosting: A survey and evaluation,” in *Computer Graphics Forum*, vol. 34, no. 2. Hoboken, NJ, USA: Wiley, 2015, pp. 683–707.
- [11] J. Zheng, Z. Li, Z. Zhu, S. Wu, and S. Rahardja, “Hybrid patching for a sequence of differently exposed images with moving objects,” *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 5190–5201, Dec. 2013.
- [12] V. G. An and C. Lee, “Single-shot high dynamic range imaging via deep convolutional neural network,” in *Proc. Asia-Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA ASC)*, Dec. 2017, pp. 1768–1772.
- [13] E. François, C. Fogg, Y. He, X. Li, A. Luthra, and A. Segall, “High dynamic range and wide color gamut video coding in HEVC: Status and potential future enhancements,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 1, pp. 63–75, Jan. 2016.
- [14] Y. Zhang, M. Naccari, D. Agrafiotis, M. Mrak, and D. R. Bull, “High dynamic range video compression exploiting luminance masking,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 5, pp. 950–964, May 2016.
- [15] Z. Wei, C. Wen, and Z. Li, “Local inverse tone mapping for scalable high dynamic range image coding,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 2, pp. 550–555, Feb. 2018.
- [16] S. Nayar and T. Mitsunaga, “High dynamic range imaging: Spatially varying pixel exposures,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 1, Jun. 2000, pp. 472–479.
- [17] U. Cogalan and A. O. Akyuz, “Deep joint deinterlacing and denoising for single shot dual-ISO HDR reconstruction,” *IEEE Trans. Image Process.*, vol. 29, pp. 7511–7524, 2020.
- [18] T. Suda, M. Tanaka, Y. Monno, and M. Okutomi, “Deep snapshot HDR imaging using multi-exposure color filter array,” in *Proc. Asian Conf. Comput. Vis.*, Nov. 2020, pp. 1–18.
- [19] J. Gu, Y. Hitomi, T. Mitsunaga, and S. Nayar, “Coded rolling shutter photography: Flexible space-time sampling,” in *Proc. IEEE Int. Conf. Comput. Photogr.*, Mar. 2010, pp. 1–8.
- [20] H. Cho, S. J. Kim, and S. Lee, “Single-shot high dynamic range imaging using coded electronic shutter,” *Comput. Graph. Forum*, vol. 33, no. 7, pp. 329–338, 2014.

- [21] N. K. Kalantari and R. Ramamoorthi, "Deep HDR video from sequences with alternating exposures," in *Computer Graphics Forum*, vol. 38, no. 2, Hoboken, NJ, USA: Wiley, 2019, pp. 193–205.
- [22] Y. Lee, K. Hirakawa, and T. Q. Nguyen, "Joint defogging and demosaicing," *IEEE Trans. Image Process.*, vol. 26, pp. 3051–3063, Jun. 2017.
- [23] M. Gharbi, G. Chaurasia, S. Paris, and F. Durand, "Deep joint demosaicing and denoising," *ACM Trans. Graph.*, vol. 35, no. 6, pp. 1–12, Nov. 2016.
- [24] X. Xu, Y. Ye, and X. Li, "Joint demosaicing and super-resolution (JDSR): Network design and perceptual optimization," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 968–980, 2020.
- [25] X. Xu, Y. Ma, and W. Sun, "Towards real scene super-resolution with raw images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1723–1731.
- [26] G. Qian *et al.*, "Rethinking the pipeline of demosaicing, denoising and super-resolution," 2019, *arXiv:1905.02538*.
- [27] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Proc. Conf. Comput. Graph. Interact. Techn.*, 1997, pp. 369–378.
- [28] X. Wang, K. Yu, C. Dong, and C. Change Loy, "Recovering realistic texture in image super-resolution by deep spatial feature transform," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 606–615.
- [29] C. Ma, Y. Rao, Y. Cheng, C. Chen, J. Lu, and J. Zhou, "Structure-preserving super resolution with gradient guidance," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 7769–7778.
- [30] P. Sen, N. K. Kalantari, M. Yaeoubi, S. Darabi, D. B. Goldman, and E. Shechtman, "Robust patch-based hdr reconstruction of dynamic scenes," *ACM Trans. Graph.*, vol. 31, no. 6, pp. 1–203, 2012.
- [31] N. D. Bruce, "Expoblend: Information preserving exposure blending based on normalized log-domain entropy," *Comput. Graph.*, vol. 39, pp. 12–23, Apr. 2014.
- [32] T.-H. Oh, J.-Y. Lee, Y.-W. Tai, and I. S. Kweon, "Robust high dynamic range imaging by rank minimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 6, pp. 1219–1232, Jun. 2014.
- [33] Y. Niu, J. Wu, W. Liu, W. Guo, and R. W. H. Lau, "HDR-GAN: HDR image reconstruction from multi-exposed LDR images with large motions," *IEEE Trans. Image Process.*, vol. 30, pp. 3885–3896, 2021.
- [34] T. Grosch *et al.*, "Fast and robust high dynamic range image generation with camera and object movement," *Vis., Model. Visualizat.*, RWTH Aachen, Aachen, Germany, Tech. Rep., 2006, vol. 277284.
- [35] O. Gallo, A. Troccoli, J. Hu, K. Pulli, and J. Kautz, "Locally non-rigid registration for mobile HDR photography," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2015, pp. 49–56.
- [36] J. Hu, O. Gallo, K. Pulli, and X. Sun, "HDR deghosting: How to deal with saturation?" in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 1163–1170.
- [37] E. Miandji, J. Unger, and C. Guillemot, "Multi-shot single sensor light field camera using a color coded mask," in *Proc. 26th Eur. Signal Process. Conf. (EUSIPCO)*, Sep. 2018, pp. 226–230.
- [38] H. Takeda, S. Farsiu, and P. Milanfar, "Kernel regression for image processing and reconstruction," *IEEE Trans. Image Process.*, vol. 16, no. 2, pp. 349–366, Feb. 2007.
- [39] R. G. Rodríguez and M. Bertalmío, "High quality video in high dynamic range scenes from interlaced dual-iso footage," in *Proc. IS&T Int. Symp. Electron. Imag. Sci. Technol.*, 2016, pp. 1–7.
- [40] I. Choi, S.-H. Baek, and M. H. Kim, "Reconstructing interlaced high-dynamic-range video using joint learning," *IEEE Trans. Image Process.*, vol. 26, no. 11, pp. 5353–5366, Nov. 2017.
- [41] A. Serrano, F. Heide, D. Gutierrez, G. Wetzstein, and B. Masia, "Convolutional sparse coding for high dynamic range imaging," *Comput. Graph. Forum*, vol. 35, no. 2, pp. 153–163, 2016.
- [42] S. Hajisharif, J. Kronander, and J. Unger, "Adaptive dualISO HDR reconstruction," *EURASIP J. Image Video Process.*, vol. 2015, no. 1, p. 41, Dec. 2015.
- [43] A. G. Vien and C. Lee, "Single-shot high dynamic range imaging via multiscale convolutional neural network," *IEEE Access*, vol. 9, pp. 70369–70381, 2021.
- [44] S. Farsiu, M. Elad, and P. Milanfar, "Multiframe demosaicing and super-resolution of color images," *IEEE Trans. Image Process.*, vol. 15, no. 1, pp. 141–159, Jan. 2005.
- [45] R. Ramanath, W. E. Snyder, and G. L. Billbro, "Demosaicing methods for Bayer color arrays," *J. Electron. Imag.*, vol. 11, no. 3, pp. 306–315, 2002.
- [46] R. Lukac, K. Martin, and K. N. Plataniotis, "Demosaiced image postprocessing using local color ratios," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 6, pp. 914–920, Jun. 2004.
- [47] S.-C. Pei and I.-K. Tam, "Effective color interpolation in CCD color filter arrays using signal correlation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 6, pp. 503–513, Jun. 2003.
- [48] X. Chen, L. He, G. Jeon, and J. Jeong, "Multidirectional weighted interpolation and refinement method for Bayer pattern CFA demosaicing," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 8, pp. 1271–1282, Aug. 2014.
- [49] K. Hirakawa and T. W. Parks, "Adaptive homogeneity-directed demosaicing algorithm," *IEEE Trans. Image Process.*, vol. 14, no. 3, pp. 360–369, Mar. 2005.
- [50] O. Kapah and H. Z. Hel-Or, "Demosaicing using artificial neural networks," *Proc. SPIE*, vol. 3962, pp. 112–120, Apr. 2000.
- [51] R. Lukac, K. N. Plataniotis, and D. Hatzinakos, "Color image zooming on the Bayer pattern," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 11, pp. 1475–1492, Nov. 2005.
- [52] Y. Xu, Z. Liu, X. Wu, W. Chen, and Z. Li, "Restoration of HDR images for SVE-based HDR10 via a novel DCNN," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2021, pp. 1–6.
- [53] R. Lukac, K. N. Plataniotis, D. Hatzinakos, and M. Aleksic, "A novel cost effective demosaicing approach," *IEEE Trans. Consum. Electron.*, vol. 50, no. 1, pp. 256–261, Feb. 2004.
- [54] W. Lu and Y.-P. Tan, "Color filter array demosaicing: New method and performance measures," *IEEE Trans. Image Process.*, vol. 12, no. 10, pp. 1194–1210, Oct. 2003.
- [55] D. Verma, M. Kumar, and S. Eregala, "Deep demosaicing using resnet-bottleneck architecture," in *Proc. Int. Conf. Comput. Vis. Image Process.* Jaipur, India: Springer, 2019, pp. 170–179.
- [56] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. pattern Recognit.*, Jul. 2017, pp. 4681–4690.
- [57] H. Ren, M. El-Khamy, and J. Lee, "DN-ResNet: Efficient deep residual network for image denoising," in *Proc. Asian Conf. Comput. Vis.* Perth, WA, Australia: Springer, 2018, pp. 215–230.
- [58] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Trans. Comput. Imag.*, vol. 3, no. 1, pp. 47–57, Mar. 2016.
- [59] D. Marnerides, T. Bashford-Rogers, J. Hatchett, and K. Debattista, "ExpandNet: A deep convolutional neural network for high dynamic range expansion from low dynamic range content," *Comput. Graph. Forum*, vol. 37, no. 2, pp. 37–49, May 2018.
- [60] C. Zheng, Z. Li, Y. Yang, and S. Wu, "Exposure interpolation via hybrid learning," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 2098–2102.
- [61] C. B. Zheng, Z. G. Li, Y. Yang, and S. Q. Wu, "Single image brightening via multi-scale exposure fusion with hybrid learning," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 8, pp. 1–10, Jul. 2021.
- [62] B. Funt and L. Shi, "The effect of exposure on MaxRGB color constancy," *Proc. SPIE*, vol. 7527, Feb. 2010, Art. no. 75270Y.
- [63] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 267–276, Jul. 2002.
- [64] C.-C. Weng, H. Chen, and C.-S. Fuh, "A novel automatic white balance method for digital still cameras," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2005, pp. 3801–3804.
- [65] T. O. Aydin, R. Mantiuk, and H.-P. Seidel, "Extending quality metrics to full luminance range images," *Proc. SPIE*, vol. 6806, Feb. 2008, Art. no. 68060B.
- [66] R. Mantiuk, K. J. Kim, A. G. Rempel, and W. Heidrich, "HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions," *ACM Trans. Graph.*, vol. 30, no. 4, pp. 1–14, Jul. 2011.
- [67] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [68] W. Shi *et al.*, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. pattern Recognit.*, Jun. 2016, pp. 1874–1883.
- [69] M. A. Robertson, S. Borman, and R. L. Stevenson, "Estimation-theoretic approach to dynamic range enhancement using multiple exposures," *J. Electron. Imag.*, vol. 12, no. 2, pp. 219–228, Apr. 2003.



**Yilun Xu** received the B.Eng. degree from Northwestern Polytechnical University, Xi'an, Shaanxi, China, in 2019. He is currently pursuing the M.Eng. degree with the School of Automation Science and Electrical Engineering, Beihang University, Beijing, China. His current research interests include deep learning, computer vision, image/video processing, and high-dynamic range imaging.



**Ziyang Liu** received the B.Eng. degree from Beijing Jiaotong University, Beijing, China, in 2017. He is currently pursuing the Ph.D. degree with the School of Automation Science and Electrical Engineering, Beihang University, Beijing. His main research interests include computer vision and image/video processing.



**Xingming Wu** received the B.Eng. and M.Eng. degrees from Zhejiang University, China, in 1981 and 1985, respectively. He is currently a Professor with the School of Automation Science and Electronic Engineering, Beihang University, China. His main research interests include image processing, scene understanding, visual mapping, autonomous mobile robots, and embedded systems.



**Weihai Chen** (Member, IEEE) received the B.Eng. degree from Zhejiang University, China, in 1982, and the M.Eng. and Ph.D. degrees from Beihang University, Beijing, China, in 1988 and 1996, respectively. He is currently a Professor with the School of Automation Science and Electrical Engineering, Beihang University. He has published more than 200 technical papers in refereed journals and conference proceedings and filed more than 20 patents. His research interests include computer vision, image processing, automation, bio-inspired robotics, and precision mechanism. He is a member of the Technical Committee on Manufacturing Automation of the IEEE Robotics and Automation Society. He was a recipient of the 2017 Best Paper Award for the IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS and the 2018 IET Premium Award for the *IET Intelligent Transport Systems*.



**Changyun Wen** (Fellow, IEEE) received the B.Eng. degree from Xi'an Jiaotong University, China, in 1983, and the Ph.D. degree from the University of Newcastle, Australia, in 1990.

From August 1989 to August 1991, he was a Post-Doctoral Fellow at the University of Adelaide, Australia. Since August 1991, he has been with Nanyang Technological University, Singapore, where he is currently a Full Professor. His main research interests include control systems and applications, cyber-physical systems, smart grids, complex systems, and networks. Some of his publications in these areas are available in <https://publons.com/researcher/2816818/changyun-wen/publications/>.

Dr. Wen was a member of the IEEE Fellow Committee from January 2011 to December 2013. He was a recipient of a number of awards, including the Prestigious Engineering Achievement Award from the Institution of Engineers, Singapore, in 2005, and the Best Paper Award of the IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS in 2017. He has been actively involved in organizing international conferences playing the roles of the General Chair (including the General Chair of IECON 2020 and IECON 2023), and a TPC Chair (e.g. the TPC Chair of Chinese Control and Decision Conference since 2008). He is also the Co-Editor-in-Chief of the IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, an Associate Editor of *Automatica* (since February 2006), and an Executive Editor-in-Chief of *Journal of Control and Decision*. He also served as an Associate Editor for the IEEE TRANSACTIONS ON AUTOMATIC CONTROL from 2000 to 2002, the IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS from 2013 to 2020, and *IEEE Control Systems Magazine* from 2009 to 2019. He was a Distinguished Lecturer of IEEE Control Systems Society from 2010 to 2013.



**Zhengguo Li** (Senior Member, IEEE) received the B.Sci. degree in applied mathematics and the M.Eng. degree in automatic control from Northeastern University, Shenyang, China, in 1992 and 1995, respectively, and the Ph.D. degree in automatic control from Nanyang Technological University, Singapore, in 2001.

He is currently with the Agency for Science, Technology and Research, Singapore. He has coauthored one monograph, more than 200 journal articles/conference papers, including about 50 IEEE TRANSACTIONS, and 11 granted U.S. patents, including normative technologies on scalable extension of H.264/AVC and HEVC. He has been actively involved in the development of H.264/AVC and HEVC since 2002. He had three informative proposals adopted by the H.264/AVC and three normative proposals adopted by the HEVC. His current research interests include computational photography, mobile imaging, video processing and delivery, and switched and impulsive control.

Dr. Li served as a TPC Chair for IEEE IECON 2023 and 2020, a Special Session Chair for IEEE ICASSP 2022, a Technical Brief Co-Founder of SIGGRAPH Asia, and a leading Workshop Chair for IEEE ICME in 2013. He was an Associate Editor of the IEEE SIGNAL PROCESSING LETTERS from 2014 to 2016, the IEEE TRANSACTIONS ON IMAGE PROCESSING from 2016 to 2020, and the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY during 2020–2024; and a Senior Area Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING during 2020–2024.