International Journal of Advanced Smart Convergence Vol.11 No.2 13-21 (2022)
http://dx.doi.org/10.7236/IJASC.2022.11.2.13

IJASC 22-2-3

Joint Demosaicing and Super-resolution of Color Filter Array Image based on Deep Image Prior Network

Edwin Kurniawan† and Suk-Ho Lee††

†*PhD Candidate, Dept. Computer Engineering, Dongseo University, Korea*
††*Professor, Dept. Computer Engineering, Dongseo University, Korea*
E-mail: petrasuk@gmail.com

Abstract

In this paper, we propose a learning based joint demosaicing and super-resolution framework which uses only the mosaiced color filter array(CFA) image as the input. As the proposed method works only on the mosaicied CFA image itself, there is no need for a large dataset. Based on our framework, we proposed two different structures, where the first structure uses one deep image prior network, while the second uses two. Experimental results show that even though we use only the CFA image as the training image, the proposed method can result in better visual quality than other bilinear interpolation combined demosaicing methods, and therefore, opens up a new research area for joint demosaicing and super-resolution on raw images.

Keywords: *Super-resolution, Color filter array, Deep Image Prior, Demosaicing, Deep Learning,*

1. Introduction

Nowadays, digital camera systems include Image Processing Units(IPU) which perform several color image processing like demosaicing, denoising, white balancing, and super-resolution of the sensed color images. Normally, the tasks of demosaicing and super-resolution are independent, i.e., first demosaicing is performed on the Color filter Array (CFA) image, then the super-resolution is performed on the demosaiced image. With the recent success of deep learning on image processing, many learning-based single image super-resolution(SISR) have been proposed [1][2][3][4][5]. However, treating the tasks of demosaicing and SISR independently can result in undesired artifacts as blurry edges, etc. Even though there are a few learning based approaches which use a large dataset for joint demosaicing and super-resolution, there is no work which uses a learning-based approach with only the mosaiced CFA image as the data for joint image demosaicing and super-resolution. Therefore, for the first time in this field, we propose a DIP based joint demosaicing, denoising, and super-resolution framework, which uses only the sensed CFA image as the input. As will be shown in the experimental sections, our proposed framework achieves better visual quality both in quantitative and qualitative measures than the approaches which deal the demosaicing and super-resolution as separate problems.

Manuscript Received: March. 7, 2022 / Revised: March. 11, 2022 / Accepted: March. 13, 2022
Corresponding Author: petrasuk@gmail.com
Tel: +82-51-320-1744, Fax: +82-51-327-8955
Professor, Department of Computer Engineering, General graduate school, Dongseo University, Korea

2. Preliminaries

The following preliminaries have to be understood to understand the proposed method.

2.1 Demosaicing problem

The problem of demosaicing is to restore the color image from the sensed color filter array (CFA) image, which contains only partial information of the full color information, i.e., to restore the original color vector $\boldsymbol{I}_{orig}[k] = (\boldsymbol{I}_R[k], \boldsymbol{I}_G[k], \boldsymbol{I}_B[k])$ from $I_s[\mathrm{k}]$ which is the k-th pixel of the sensed image I_s. The relationship between $I_s[\mathrm{k}]$ and $\boldsymbol{I}_{orig}[k]$ is established via the CFA color filter triplet $\mathbf{c}[\mathbf{k}] = (\boldsymbol{c}_R[k], \boldsymbol{c}_G[k], \boldsymbol{c}_B[k])$, i.e.,

$$I_s[\mathrm{k}] = \mathbf{c}[\mathbf{k}] \cdot \boldsymbol{I}_{orig}[k].$$

where $\cdot$ denotes the inner product operation. The CFA color filter triplet $\mathbf{c}[\mathbf{k}]$ differs for different CFAs. In [6], a white-dominant RGBW CFA is proposed, for which $\mathbf{c}[\mathbf{k}]$ becomes

$$\begin{aligned}
&\boldsymbol{c}_R[k] = 1, \quad \boldsymbol{c}_G[k] = 0, \quad \boldsymbol{c}_B[k] = 0 \quad if \quad k \in S_R \\
&\boldsymbol{c}_R[k] = 0, \quad \boldsymbol{c}_G[k] = 1, \quad \boldsymbol{c}_B[k] = 0 \quad if \quad k \in S_G \\
&\boldsymbol{c}_R[k] = 0, \quad \boldsymbol{c}_G[k] = 0, \quad \boldsymbol{c}_B[k] = 1 \quad if \quad k \in S_B \\
&\boldsymbol{c}_R[k] = \alpha_R, \ \boldsymbol{c}_G[k] = \alpha_G, \ \boldsymbol{c}_B[k] = \alpha_B \quad if \quad k \in S_W,
\end{aligned}$$

where, α_R, α_G and α_B are the ratio of the R, G, and B components in the white pixel. For example, with the VEML6040 sensor, $\alpha_R = 0.2936$, $\alpha_G = 0.4905$, and $\alpha_B = 0.2159$, which ratio is used in [6]. With the proposed method, we use the same RGBW CFA as in [6] for the joint demosaicing and super-resolution problem.

2.2 Deep Image Prior Network

The Deep Image Prior (DIP) network is a network which restores an image $I \in \mathrm{R}^{w \times h \times c}$, where w, h, and c are the width, height, and number of channels, respectively, by training a network with parameter θ, to produce $\boldsymbol{g}_\theta(\boldsymbol{z})$, the restored image. Here, $\boldsymbol{z} \in \mathrm{R}^{w \times h \times c}$ is a noise image, where each pixel value is drawn from a Gaussian distribution of zero mean and standard deviation of 1. The loss function includes the image I which should be restored, so that the DIP is trained only on I, eliminating the need for a large dataset:

$$L_{DIP} = ||M \odot (\boldsymbol{g}_\theta(\boldsymbol{z}) - I)||^2,$$

Here, M is a mask matrix which has a value of 1 at the pixels positions where I has non-zero values, and a value of 0 at the positions where I has zero values. The DIP is actually an encoder-decoder type neural network which has a structure shown in Fig. 1(a). In this paper, we modify the network structure as shown in Fig. 1(b), since the output should be a super-resolved image with double the spatial size of the input image. It has been shown in [6] and [7] that the DIP can also work well for joint denoising and demosaicing applications. To denoise the demosaiced image, the input noise is constituted in [6] and [7] as a combination of a composed of a static noise $\boldsymbol{z}_c$ and a varying noise $\boldsymbol{z}_v$, i.e, $\boldsymbol{z} = \boldsymbol{z}_c + \boldsymbol{z}_v$, where $\boldsymbol{z}_c$ and $\boldsymbol{z}_v$ both have the same size as I. Here, the static noise $\boldsymbol{z}_c$ is generated only once in the beginning of the training of the DIP, while the varying noise $\boldsymbol{z}_v$ is generated newly at each iteration of the training. In this work, we further extend the work of [7] so that it can be applied to the application of joint demosaicing and super-resolution.

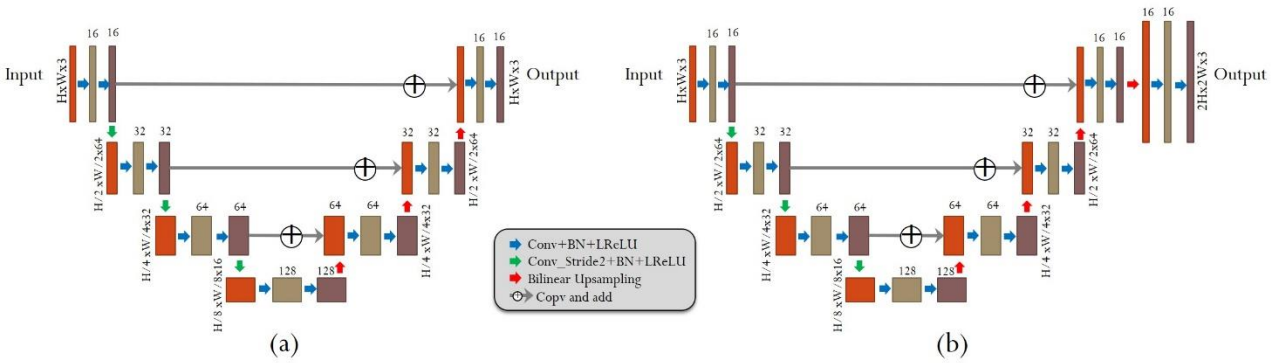

Figure 1. The DIP network structure (a) used in [6] and [7] (b) used in the proposed work.

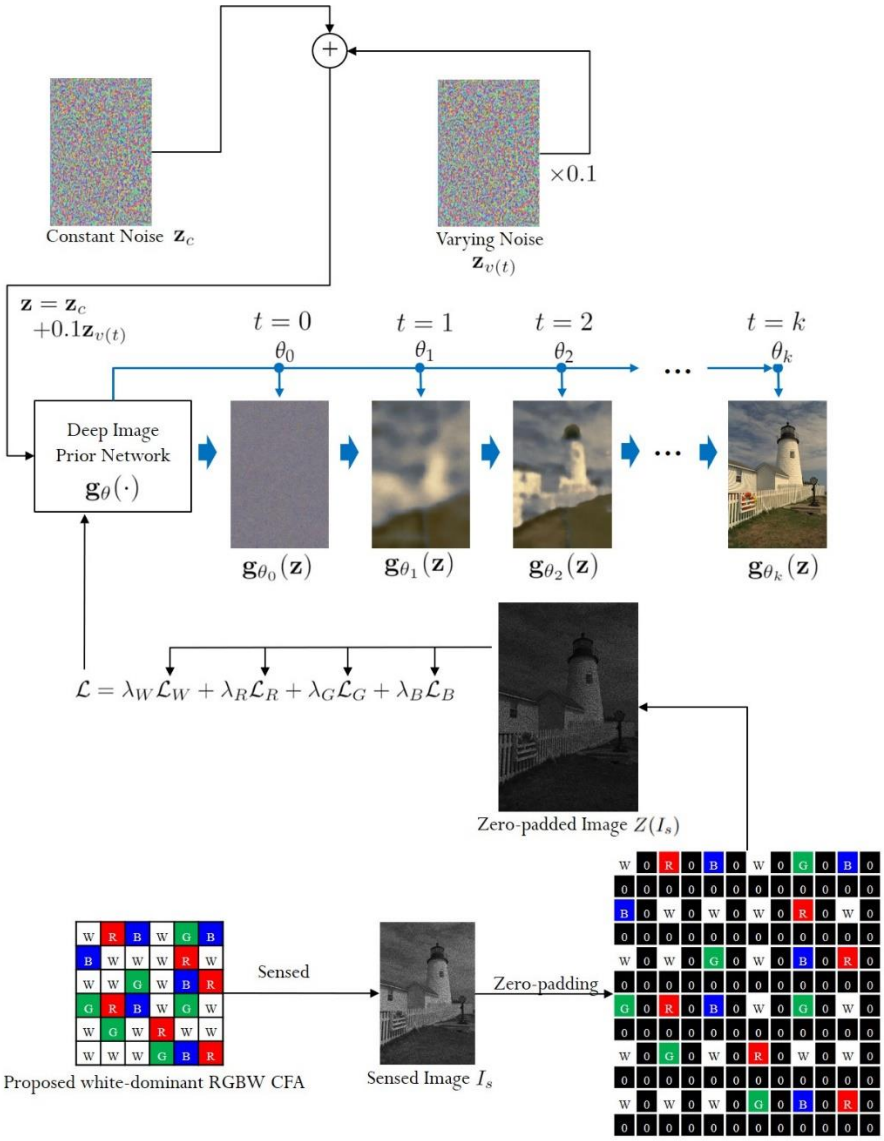

Figure 2. Overall System Diagram of the proposed one DIP based joint demosaicing and super-resolution

3. Proposed Joint Demosaicing and Super-resolution Method

As explained in the introduction, the joint super-resolution and demosaicing method is an important cost-effective alternative to more expensive hardware-based solutions like optical zooming. In this section, we propose a model that does simultaneous denoising, demosaicking, and superresolution on the raw CFA image. In the case of superresolution, we want to super-resolve the sensed CFA image I_s.

To resize the sensed CFA image, we just double the size of the grid, and then put the image values of I_s at every other pixel in the doubled size grid. At the pixels which have no values, we put just zero values in them. Figure 1 shows the resizing and zero-padding of I_s, which results in a new image $Z(I_s)$, where $Z(\cdot)$ denotes the resizing and zero-padding function. We will undergo the same resizing and zero-padding operation on the output of the DIP, i.e., change $\mathrm{F}(\mathbf{c} \odot \boldsymbol{g}_\theta(\boldsymbol{z}))$ to $\mathrm{Z}(\mathrm{F}(\mathbf{c} \odot \boldsymbol{g}_\theta(\boldsymbol{z})))$. The sizes of the zero-padded images $Z(I_s)$ and $\mathrm{Z}(\mathrm{F}(\mathbf{c} \odot \boldsymbol{g}_\theta(\boldsymbol{z})))$ depend on the resizing factor, i.e., are $\times$ n larger than I_s, where n is the resizing factor.

The loss function which does the joint superresolution and demosaicing becomes:

$$\mathrm{L} = \lambda_W L_W + \lambda_R L_R + \lambda_G L_G + \lambda_B L_B$$

where L_W is a loss function related to white pixels and L_R, L_G, and L_B are loss functions related to Red, Green, and Blue pixels, respectively. Furthermore, L_W, L_R, L_G, and L_B are defined with respect to the resized and zero-padded images $Z(I_s)$ and $\mathrm{Z}(\mathrm{F}(\mathbf{c} \odot \boldsymbol{g}_\theta(\boldsymbol{z})))$, and become:

$$L_W = || M_W \odot [\mathrm{Z}(\mathrm{F}(\mathbf{c} \odot \boldsymbol{g}_\theta(\boldsymbol{z}))) - Z(I_s)]||^2, \quad (1)$$

$$L_R = || M_R \odot [\mathrm{Z}(\mathrm{F}(\mathbf{c} \odot \boldsymbol{g}_\theta(\boldsymbol{z}))) - Z(I_s)]||^2, \quad (2)$$

$$L_G = || M_G \odot [\mathrm{Z}(\mathrm{F}(\mathbf{c} \odot \boldsymbol{g}_\theta(\boldsymbol{z}))) - Z(I_s)]||^2, \quad (3)$$

and

$$L_B = || M_B \odot [\mathrm{Z}(\mathrm{F}(\mathbf{c} \odot \boldsymbol{g}_\theta(\boldsymbol{z}))) - Z(I_s)]||^2. \quad (4)$$

In (1), M_W has values of 1 at white pixel position and zero at non-white pixels including the padded zero values. Also, M_R, M_G, and M_B have values of 1 in their respective pixel color position and zeros at other pixel locations also including the padded zero values. As in [6], we apply different values for λ_W, λ_R, λ_G and λ_B for different iteration numbers such as

$$\begin{aligned}
&\lambda_W = 1,\ \lambda_R = 1,\ \lambda_G = 0,\ \lambda_B = 0 && \text{if} \quad t \geq 500 \\
&\lambda_W = 1,\ \lambda_R = 0,\ \lambda_G = 1,\ \lambda_B = 0 && \text{if} \quad 500 < t \leq 1000 \\
&\lambda_W = 1,\ \lambda_R = 1,\ \lambda_G = 0,\ \lambda_B = 0 && \text{if} \quad 1000 < t \leq 1500 \\
&\lambda_W = 1.5,\ \lambda_R = 0.5,\ \lambda_G = 0.5,\ \lambda_B = 0.5 && \text{if} \quad 1500 < t
\end{aligned} \quad (5)$$

Figure 2 shows the overall system diagram of the joint denoising, demosaicing, and super-resolution method. It should be noted that the DIP network has a different output size as that used in the joint denoising and demosaicing method, i.e., the output has a size of $\times$ n of that of the input, where n is the resizing factor.

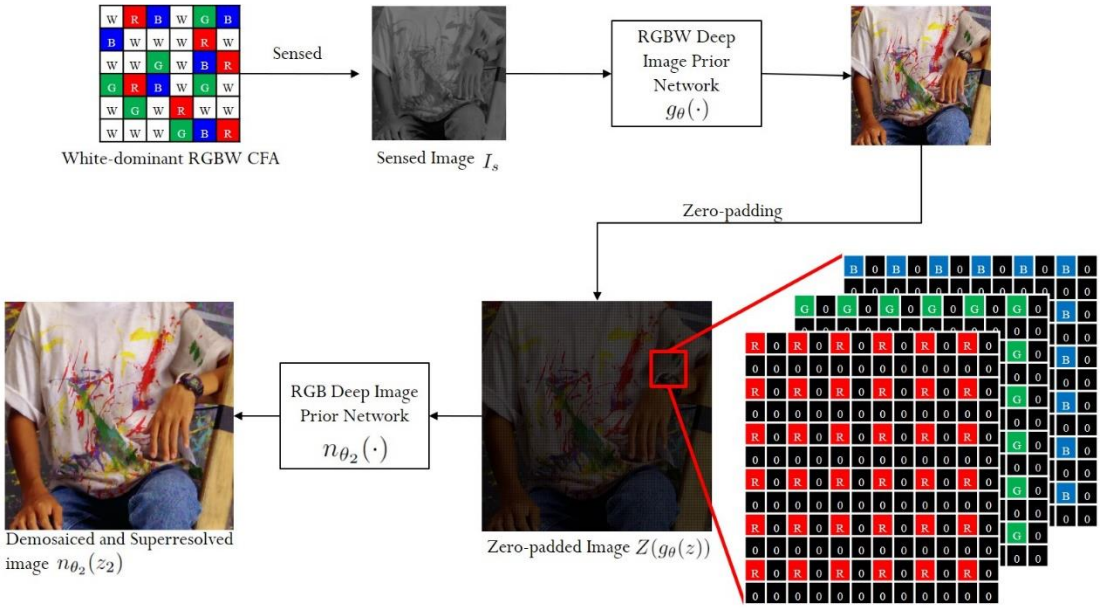

Figure 3. Overall Procedure Diagram of the proposed two DIP based joint demosaicing and super-resolution

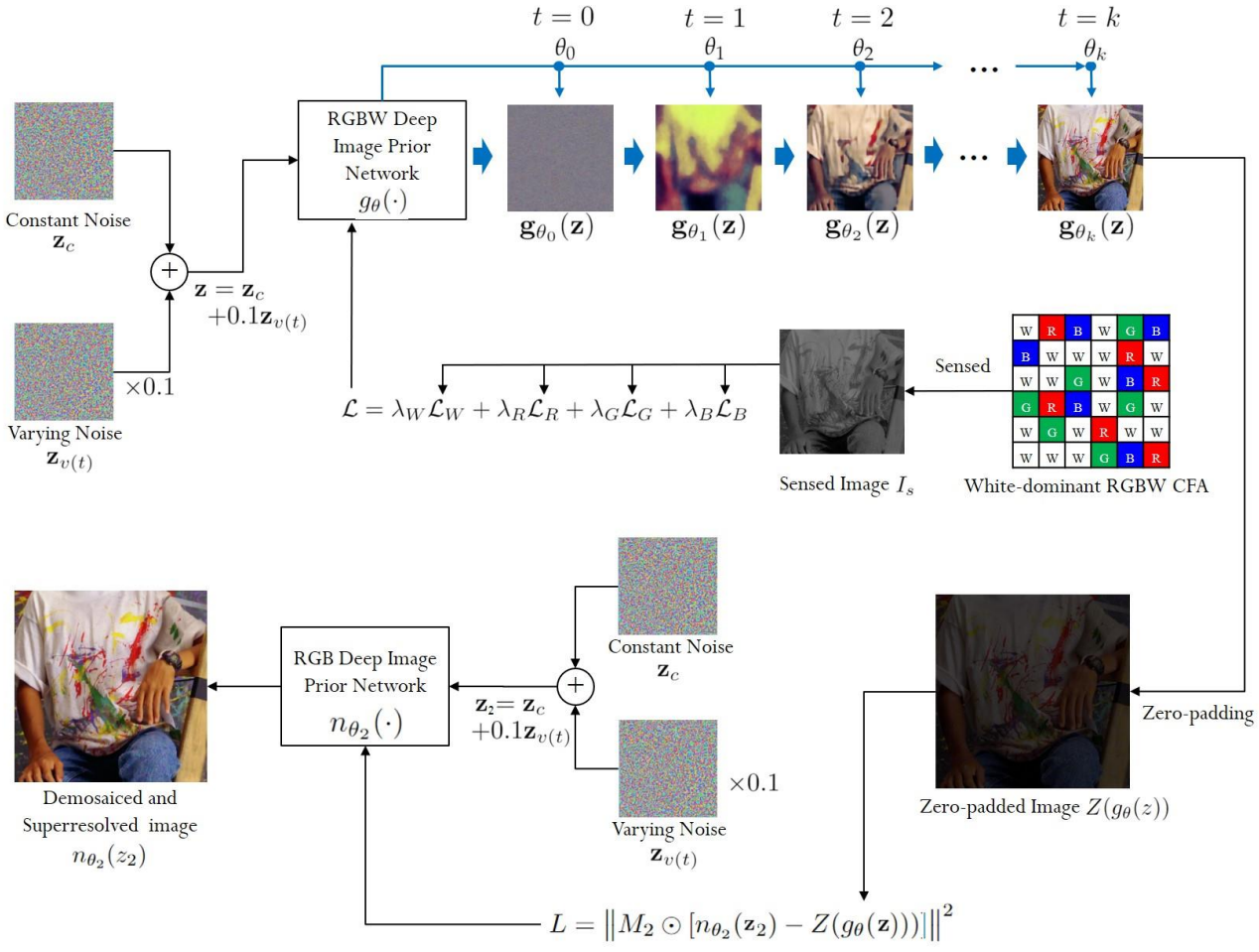

Figure 4. Overall System Diagram of the proposed joint demosaicing and super-resolution

However, this rather straightforward method suffers from over-blurring, since the zero padding is directly done on the raw sensed CFA image. As in the case of many deep learning based image generation, a better approach would be if we use a hierarchical structure. We do this by a three-step procedure: first, we do the demosaicing with the RGBW DIP, which demosaics the small-sized RGBW CFA image, then resize the output of the demosaiced image to the super-resolved image size by zero-padding, and last, put the zero-padded RGB image in a second DIP, which interpolates the zero-padded RGB image to the size of the superresolved image.

Figure 3 shows the overall working process of the joint demosaicing and super-resolution with two DIP networks. First, the sensed RGBW CFA image is put into the first DIP network $\boldsymbol{g}_\theta(\cdot)$, which is the same network as proposed for the One DIP based method. This results in a demosaiced image $\boldsymbol{g}_\theta(\boldsymbol{z})$. Then, we apply a zero-padding on $\boldsymbol{g}_\theta(\boldsymbol{z})$, to resize it to the size of the desired super-resolved image. The zero-padded image $\boldsymbol{Z}(\boldsymbol{g}_\theta(\boldsymbol{z}))$ is then used in the loss function of the second DIP $\boldsymbol{n}_{\theta_2}(\cdot)$, which interpolates the zero-valued pixel values to obtain the desired demosaiced and super-resolved result image $\boldsymbol{n}_{\theta_2}(\boldsymbol{z}_2)$, where $\boldsymbol{z}_2$ is the noise which goes as the input into $\boldsymbol{n}_{\theta_2}(\cdot)$. As with the DIP proposed in section 3.2, the noise $\boldsymbol{z}_2$ is also composed of a static noise $\boldsymbol{z}_c$ and a varying noise $\boldsymbol{z}_v$. By letting M_2 be the mask matrix of the second DIP, which contains values of 1 at the position where $\boldsymbol{Z}(\boldsymbol{g}_\theta(\boldsymbol{z}))$ has non-zero values and values of 0 where $\boldsymbol{Z}(\boldsymbol{g}_\theta(\boldsymbol{z}))$ has non-zero values, the loss function for training the second DIP $\boldsymbol{n}_{\theta_2}(\cdot)$ becomes:

$$\mathrm{L} = ||M_2 \odot [\boldsymbol{n}_{\theta_2}(\boldsymbol{z}_2) - Z(\boldsymbol{g}_\theta(\boldsymbol{z}))]||^2. \tag{6}$$

Figure 4 shows the overall system diagram of the joint demosaicing and super-resolution with two DIP networks, and Table 1 shows the overall algorithm pseudo code of the two DIP method.

Table 1. Showing the Pseudo code of training the two DIP method

Algorithm 1 Training of the two DIP method

Sample z_c from Zero Mean Gaussian Distribution
$I_S \leftarrow$ sensed CFA image
$TrainingEndTime \leftarrow 2000$
while $Iteration \neq TrainingEndTime$ **do**
  Sample z_v from Zero Mean Gaussian Distribution
  Update the RGBW DIP by minimizing:
  $\mathcal{L} = \lambda_W \mathcal{L}_W + \lambda_R \mathcal{L}_R + \lambda_G \mathcal{L}_G + \lambda_B \mathcal{L}_B$ where
  $\mathcal{L}_W = \|M_W \odot [\mathrm{Z}(\mathrm{F}(\mathbf{c} \odot \boldsymbol{g}_\theta(\mathbf{z}))) - Z(I_S)]\|^2$
  $\mathcal{L}_R = \|M_R \odot [\mathrm{Z}(\mathrm{F}(\mathbf{c} \odot \boldsymbol{g}_\theta(\mathbf{z}))) - Z(I_S)]\|^2$
  $\mathcal{L}_G = \|M_G \odot [\mathrm{Z}(\mathrm{F}(\mathbf{c} \odot \boldsymbol{g}_\theta(\mathbf{z}))) - Z(I_S)]\|^2$
  $\mathcal{L}_B = \|M_B \odot [\mathrm{Z}(\mathrm{F}(\mathbf{c} \odot \boldsymbol{g}_\theta(\mathbf{z}))) - Z(I_S)]\|^2$
end while
return $g_\theta(\mathbf{z})$ ▷ RGBW DIP training finished
Sample z_c from Zero Mean Gaussian Distribution
$TrainingEndTime \leftarrow 2000$
while $Iteration \neq TrainingEndTime$ **do**
  Sample z_v from Zero Mean Gaussian Distribution
  Update the RGB DIP by minimizing:
  $\mathcal{L} = \|M_2 \odot [n_{\theta_2}(\mathbf{z}_2) - Z(g_\theta(\mathbf{z}))]\|^2$
end while
return $n_{\theta_2}(\mathbf{z}_2)$ ▷ RGB DIP training finished

4. Experimental Results

Figure 5 compares the results of different demosaicing and super-resoluton methods on the McMaster No. 5 image. As there are almost no joint demosaicing and super-resolution methods which work on the raw CFA directly, we compare our method with the SEM+BI and the LCNN+BI methods, where BI denotes the normal bilinear interpolation. In other words, we first perform the conventional SEM and LCNN demosaicing on the small-sized (250 × 250) CFA image, and then resize the results to the original sizes (500 × 500) by bilinear interpolation.

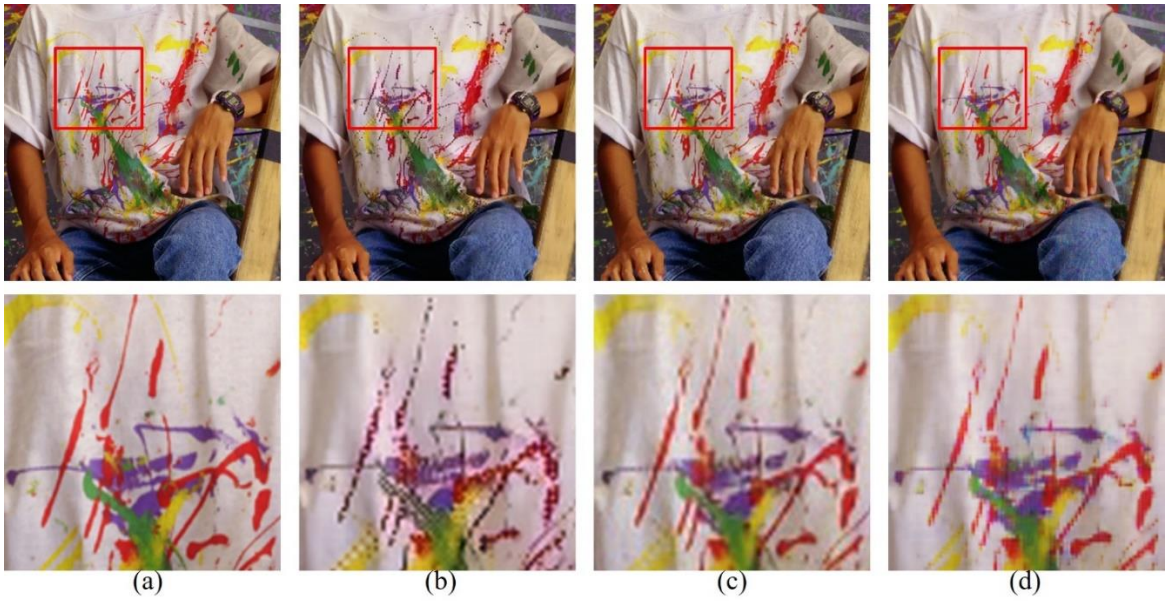

Figure 5. Comparison between the different demosaicing and super-resolution methods on the McMaster no. 5 image: (a) original, (b) SEM+BI (c) LCNN+BI (d) proposed

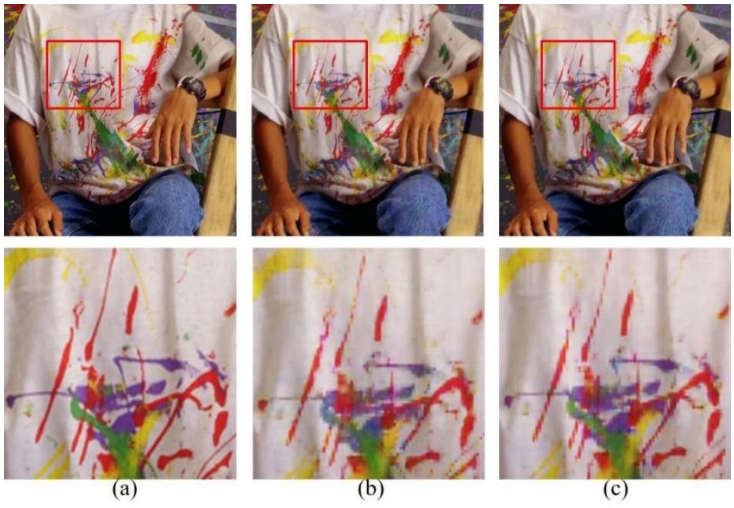

Figure 6. Comparison between the one DIP based and the two DIP based demosaicing and super-resolution methods on the McMaster no. 5 image: (a) original, (b) One DIP based result (c) Two DIP based result

As can be observed in Fig. 5(b), the SEM+BI method cannot reconstruct well all the colors, and the colors appear to be faded. The LCNN+BI method also shows some color fading and blurring artifacts. In comparison, the proposed two DIP based method faithfully reconstructs the colors while showing less blurring artifacts.

We also compare the results of the proposed one DIP based method with the proposed two DIP based method in Fig. 6. As can be seen, compared to the two DIP based method, the one DIP based method results sometimes in false colors. However, the speed with the one DIP based method is faster than the two DIP based one. The One DIP based method takes about 2000 iterations for the training of the DIP, which takes about 843 seconds on a PC with Intel i7-9700K CPU, 16Gb RAM memory, and GPU 2080Ti. The two DIP based method takes 2000 iterations for the training of the RGBW DIP and 2000 for the RGB DIP, which in total take about 1724 seconds. Therefore, in the future, the improvement of the one DIP based joint demosaicing and superresolution method can be a possible further study topic. Table 2 compares the results of the LCNN+BI, the, SEM+BI, and the proposed methods in quantitative measures on the Kodak and the McMaster datasets. It can be seen from Table 2, that the proposed method results in higher PSNR and SSIM values than the conventional deep learning based methods.

Table 2. Quantitative Comparison between the SEM+BI , the LCNN+BI and the proposed method

Dataset	Method	PSNR	SSIM
Kodak	LCNN + BI	24.7625	0.9802
	SEM + BI	23.1468	0.9780
	Ours (One DIP)	**26.1944**	**0.9880**
	Ours (Two DIP)	**26.7666**	**0.9911**
McMaster	LCNN + BI	25.3408	0.9860
	SEM + BI	24.8714	0.9842
	Ours (One DIP)	**27.0954**	**0.9896**
	Ours (Two DIP)	**28.5659**	**0.9922**

5. Conclusion

In this paper, we proposed a DIP based joint demosaicing and super-resolution framework, and implemented this framework with a One DIP based method and a Two DIP based method. The proposed framework does not require a large dataset but only the mosaiced CFA image itself in the training of the DIP networks. Experimental results show that the proposed framework can better reconstruct a super-resolved demosaiced

image from the raw CFA image. One of the drawbacks of the proposed method is that the DIP have to be re-trained for every new CFA image. Therefore, the process of a single CFA image takes a long time. The reduce in the computational cost is one of the major future topics. Nonetheless, the proposed framework opens a possibility for future works on joint demosaicing and super-resolution of raw CFA images with better performance.

Acknowledgement

This work was supported by the Technology development Program(S2840023) funded by the Ministry of SMEs and Startups(MSS, Korea).

References

[1] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a Deep Convolutional Network for Image Super-Resolution," in Proc. ECCV, pp. 184–199, Sep. 6-12, 2014. DOI: https://doi.org/10.1007/978-3-319-10593-2

[2] C. Dong, C. C. Loy, and X. Tang, "Accelerating the Super-Resolution Convolutional Neural Network," in Proc. ECCV, pp. 391–407, Oct. 11-14, 2016. DOI: https://doi.org/10.1007/978-3-319-46475-6

[3] J. K. Lee, J. Kim, and K. M. Lee, "Deeply-Recursive Convolutional Network for Image Super-Resolution," in Proc. CVPR 2016, pp. 1637-1645, June 27-30, 2016. DOI: 10.1109/CVPR.2016.181

[4] J. Kim, J. K. Lee, and K. M. Lee, "Accurate Image Super-Resolution Using Very Deep Convolutional Networks," in Proc. CVPR 2016, pp. 1646-1654, June 27-30, 2016. DOI: https://doi.org/10.1109/CVPR.2016.181

[5] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced Deep Residual Networks for Single Image Super-Resolution", in Proc. CVPR 2017 Workshops, pp. 136-144, July 21-26, 2017. DOI: https://doi.org/10.1109/CVPRW.2017.151

[6] Y. Park, S. Lee, B. Jeong, J. Yoon, "Joint Demosaicing and Denoising Based on a Variational Deep Image Prior Neural Network," Sensors, Vol. 20, No. 2970, pp. 1-14, May 2020. DOI: https://doi.org/10.3390/s20102970

[7] E. Kurniawan, Y. Park, and S. Lee, "Noise-Resistant Demosaicing with Deep Image Prior Network and Random RGBW Color Filter Array," Sensors, Vol. 22, No. 1767, pp. 1-13, Feb. 2022. DOI: https://doi.org/10.3390/s22051767