Deep demosaicking for multispectral filter arrays

Kazuma Shinoda¹, Shoichiro Yoshiba¹ and Madoka Hasegawa¹

Abstract—We propose a novel demosaicking method for multispectral filter arrays based on a deep convolutional neural network. The proposed method first interpolates mosaicked multispectral images utilizing a bilinear approach, then applies a residual network to initial demosaicked images. The residual network consists of various three-dimensional convolutional layers and a rectified linear unit for describing the features of a multispectral data cube. Experimental results reveal that the proposed method outperforms conventional demosaicking methods.

I. INTRODUCTION

Multispectral images (MSIs) have been utilized in many remote sensing and medical applications [1], [2]. A multispectral filter array (MSFA) is one solution for capturing an MSI in a single shot at low cost. The architecture of an MSFA has a different spectral filter for each pixel of an image. Images captured by an MSFA have only one value for each pixel, but a full-resolution MSI can be obtained by recovering the missing spectrum information. This recovering process is referred to as demosaicking.

Various filter array patterns and demosaicking methods [3]–[14] have been proposed for improving demosaicked image quality. Brauers et al. [3] proposed a six-band MSFA arranged in $3 \times 2$ pixels in a straightforward manner for fast linear interpolation. Their demosaicking method first applies bilinear interpolation to each band. Inter-band correlations are then corrected by smoothing inter-band differences. Yasuma et al. [4] designed a seven-band MSFA composed of three primary color filters and four secondary color filters. A simple four-band MSFA was proposed by Aggarwal et al. [5], who presented a versatile demosaicking method based on an $l_1$-norm minimization problem. Jia et al. [8] proposed a Fourier spectral filter array utilizing a Fabry-Perot-type filter to remove aliasing. Monno et al. [11] proposed a five-band MSFA and determined that the sampling density of G-band data was higher than that of the other spectral bands because the human eye is more sensitive to the G-band compared to other spectral bands. Mihoubi et al. [12] proposed a demosaicking method based on the correlation between each spectral band and a pseudo-panchromatic image (PPI), which represents the average image over all bands. In our previous work, we designed an MSFA pattern for medical applications based on training data [7], [14]. Because the number of non-measured pixels in an MSFA is larger than that in an RGB Bayer color filter array, a more accurate demosaicking method is required.

An RGB demosaicking method based on a convolutional neural network (CNN) was proposed in [15] and achieved excellent performance. This method first produces a rough demosaicked image by applying a conventional RGB demosaicking method (Gradient-Corrected Bilinear Interpolation, GCBI [16]), then reduces color artifacts by utilizing deep residual estimation and multi-model fusion.

In this paper, we propose a novel demosaicking method for MSIs based on a deep CNN. The proposed method first interpolates a mosaicked image utilizing a bilinear approach, then removes artifacts in the tentative demosaicked image utilizing a deep residual network (ResNet) [17] and three-dimensional (3D) convolutions [18]. It is expected that the ResNet can reduce the error between the initial demosaicked image and training data, and the 3D-CNN can learn local signal changes in both the spatial and spectral dimensions of feature cubes effectively. We compared the performance of our demosaicking method to that of a bilinear approach and PPI difference (PPID) method [12].

This remainder of paper is organized as follows. In Section II, we describe our proposed demosaicking method. We present experiments to test our method and their results in Section III. Section IV contains our conclusions.

II. PROPOSED DEMOSAICKING METHOD

The proposed method consists of two steps: initial demosaicking and ResNet-based refinement, as shown in Fig. 1. The first step interpolates a mosaicked image utilizing a bilinear approach with an $N \times N$ window ($N$ is the block size of MSFA plus one). This bilinear interpolation is performed for each band independently. Next, the second step refines the initial demosaicked image by utilizing ResNets and 3D-CNN. The proposed network structure is illustrated in Fig. 2. Here, $\oplus$ is the addition operator for each pixel and $(x \times y \times z, n)$ in the convolutional layer is the 3D filter size $(x, y, z)$ in horizontal-vertical-spectral directions, where the number of filter types is $n$. The proposed network consists of six modules, where each module from one to five has a convolutional layer for the 3D-CNN with a rectified linear

Fig. 1: Proposed multispectral demosaicking flow.

Fig. 2: Here, $\oplus$ is the addition operator for each pixel and $(x \times y \times z, n)$ in the convolutional layer is the 3D filter size $(x, y, z)$ in horizontal-vertical-spectral directions, where the number of filter types is $n$. The proposed network consists of six modules, where each module from one to five has a convolutional layer for the 3D-CNN with a rectified linear

The authors are with Graduate School of Engineering, Utsunomiya University, 7-1-2 Yoto, Utsunomiya, Japan

1The authors are with Graduate School of Engineering, Utsunomiya University, 7-1-2 Yoto, Utsunomiya, Japan
The obtained feature map from the proposed convolutional layer does not include a convolutional layer. The number of output features for each module is different from the number of input features. The final demosaicked image is obtained by adding the initial convolutional layer because the number of output features for each module includes a ReLU and shortcut connection. Note that the shortcut connection for each module includes a $1 \times 1 \times 1$ convolutional layer which is expected to speed up training [19]. Each module in the proposed network is based on a ResNet structure. Although ResNets have been applied to comparatively deeper networks for solving the degradation problem [17] by shortcutting input data, wider (rather than deeper) networks have shown similar or better results than those obtained from very deep networks [20]. The proposed network is comparatively shallow, but the shortcut connection may have an effect on solving the degradation problem. The longest shortcut connection has the potential to suppress artifacts in demosaicked images effectively [15].

### III. Experimental Results

In our experiment, we used Python 2.7.13 and Keras 2.0.8 [21] (backend: Tensorflow 1.3.0 [22] with the Adam optimizer [23]) for the proposed method and the CAVE database [4] for test images. The illuminant of the test images is CIE D65. According to the mosaicking method in [12], the mosaicked images were obtained by using a 16-band MSFA, as shown in Fig. 4. We compared the peak signal-to-noise ratios (PSNRs) and appearances of the demosaicked images to those from the bilinear interpolation and PPID [12].

In the proposed method, 8-fold cross validation was used for evaluating PSNR. This validation first divided 32 test images into eight groups evenly, of which seven groups were used for ResNet training 300 times and the remaining group was used for evaluation. Additionally, we divided each image into 16 sub-images (batch size of 8) for training to save memory consumption and computational time.

The PSNRs of the demosaicked images are listed in Table I. The average PSNR of the proposed method was higher than that of both the bilinear interpolation and PPID methods, with a gain of 2.67 dB. This result indicates that the proposed ResNet-based refinement method can improve the quality of the initial images demosaicked via bilinear interpolation, and its PSNR exceeds that of conventional methods.
multispectral demosaicking methods, such as PPID. The number of images for which PPID achieved the highest PSNR was only two (including “fake and real tomatoes demosaicked”) out of 32.

The original and demosaicked images are presented in Figs. 5–8. In Fig. 5 the letters are strongly blurred in the bilinear interpolation results, but readability is significantly improved by the proposed method. Therefore, the blur in the results of the first step of the proposed method was improved in the second step. In Fig. 6 an artifact occurred along the edge in the PPID results, whereas the demosaicked image from the proposed method contains no such artifact. This artifact can be seen also in Fig. 7(c). Therefore, we can conclude that the proposed method suppressed the artifact along the edge. In Fig. 5 a color artifact can be seen in the demosaicked image from the proposed method, but cannot be seen in the image from PPID. Because a strong color artifact occurred in the image resulting from bilinear interpolation, the second step of the proposed method was unable to completely remove the artifact. Fixing this issue will require improving our proposed network in the future.

The second step of the proposed method utilizes shortcut connections as a result of residual estimation. If the shortcut connections in modules one to five are removed, the average PSNR of the proposed method decreases by 1.28 dB. Furthermore, in the case where all shortcut connections are removed, this degradation increases to 5.05 dB. Therefore, the ResNet structure performs effectively in the proposed network. Additionally, we can conclude that the 3D-CNN provides an advantage in terms of feature representation of a multiresolution data cube because the PSNR of the proposed method decreased by 1.19 dB when the convolution process was changed to a 2D process. Based on these results, the proposed method based on ResNets and a 3D-CNN is suitable for multispectral demosaicking.

IV. Conclusions

We proposed a novel demosaicking for MSIs utilizing a deep CNN with ResNets and 3D convolution. The proposed method achieved higher PSNRs than conventional multispectral demosaicking methods and can suppress artifacts. Because some demosaicked images from the proposed method still contained color artifacts, we must improve the network structure (e.g., change from bilinear interpolation to PPID in the first step or increase the number of layers) in the future.

Acknowledgement

This work was supported by JSPS KAKENHI Grant Number 18K11368.
Fig. 5: Demosaicked images of “fake and real peppers.”

Fig. 6: Demosaicked images of “clay.”

Fig. 7: Demosaicked images of “beads.”

Fig. 8: Demosaicked images of “fake and real tomatoes.”