RESEARCH NOTE

Predictive Filter Flow Network for Universal Demosaicking

Daiki Arai¹, Taishi Iriyama¹, Masatoshi Sato¹, Hisashi Aomori² and Tsuyoshi Otake¹

¹Graduate School of Engineering, Tamagawa University
6-1-1 Tamagawagakuen, Machida, Tokyo 194-8610, Japan

²Department of Electrical and Electronic Engineering, Chukyo University
101-2 Yagoto-Honmachi, Showa-ku, Nagoya, Aichi 466-8666, Japan
E-mail: th5t21232@stu.tamagawa.ac.jp, otake@ieee.org

Abstract Demosaicking is an image reconstruction process for restoring full-color images from color filter array (CFA) data. In recent years, many deep convolutional neural network (CNN)-based demosaicking methods have been reported, and state-of-the-art accuracy has been achieved. In this paper, we propose a novel demosaicking method using the predictive filter flow (PFF) network for various CFA patterns. The PFF is a model that predicts a spatial variant linear filter that transforms an input image into a target image. To incorporate the PFF into demosaicking, the proposed network synthesizes the filter flow corresponding to each channel by means of a network trained by integrating RGB channels. Our model, designed to apply demosaicking with the PFF to various CFA patterns, provides versatility and extensibility. Experimental results demonstrate that the proposed method provides better or competitive results compared with several state-of-the-art deep-CNN-based demosaicking algorithms.

Keywords: universal demosaicking, predictive filter, convolutional neural network

1. Introduction

Digital images are obtained by converting light into electrical signals using image sensors built into digital image devices such as digital cameras and smartphones. In a general digital camera, a color filter array (CFA) is mounted on an image sensor to acquire mosaic data with only one-color information per pixel. The Bayer CFA, which consists of RGB color filters, is the most commonly used CFA pattern. To recover a full-color image from mosaic data, it is necessary to obtain the information of the two missing color by interpolation [1]. This process is called demosaicking or CFA interpolation. When each color channel is estimated separately using standard spatially invariant interpolation techniques, the reconstructed image loses information at high frequencies, resulting in artifacts such as false colors and zipper effects. Many demosaicking algorithms have been proposed to reduce these artifacts.

In recent years, deep convolutional neural network (CNN) techniques have provided excellent performance in many image processing tasks such as super-resolution, classification, deblurring, and denoising. The CNN exploits many convolutional layers

with only localized connections to analyze the entire image and enables complex feature extraction or prediction.

Although certain CFA demosaicking methods, such as the Bayer CFA, have been relatively well researched and optimized, various CFA demosaicking methods have not yet been fully and systematically considered. In general, a particular demosaicking method is specific to a particular CFA pattern and cannot be applied to other CFA patterns. It would be very useful to have a universal demosaicking method that can be applied to various CFA patterns. In this paper, we propose a novel demosaicking method that can be applied to multiple CFA patterns using the predictive filter flow (PFF) network. The PFF is one of the network frameworks applied in CNN prediction [2]. Given an input image, the PFF directly predicts the filter flow and reconstructs the desired output. The PFF is modeled as a linear mapping to transform the input image to the target image, so that each output pixel is reconstructed using only local information of the input image.

The contributions of this paper are summarized as follows: (1) A universal demosaicking method using the PFF is proposed and a novel framework for incor-

Fig. 1 Framework of proposed method

porating the PFF into the demosaicking is presented. (2) Experimental results on the Kodak and McMaster datasets show that the proposed method provides better or competitive results compared with state-of-the-art deep-CNN-based demosaicking algorithms.

2. Predictive Filter Flow (PFF)

The filter flow is an image transformation that uses a space-variant filter to linearly combine with the input image to reconstruct the target image [3]. Unlike the general filtering process, the filter flow is designed so that each filter corresponding to the spatial location has a different weight. The PFF is a framework for synthesizing an optimal filter flow by analyzing the input signal with a deep neural network.

The filter flow $\mathbf{T}$ can be obtained by optimizing the parameter θ of the network $F_\theta()$ from the given input image $\mathbf{I}_1$ and the target image $\mathbf{I}_2$.

$$\mathbf{I}_2 \approx \mathbf{T}\mathbf{I}_1, \mathbf{T} \equiv F_\theta(\mathbf{I}_1) \quad (1)$$

This model is called the PFF. The network $F_\theta()$ is learned by minimizing the l_1 -norm loss between the predicted image and the target image with the Adam optimizer [4],

$$loss = \sum_{i=0}^N |F_\theta(\mathbf{I}_1^i) \cdot \mathbf{I}_1^i - \mathbf{I}_2^i|_1 + R(F_\theta(\mathbf{I}_1^i)) \quad (2)$$

where N is the number of learning observations. In practice, the softconstraints due to the regularization term $R()$ are also added.

3. Feature Extraction

Feature extraction consists of a two-stream network architecture (Fig. 1). The first stream is a shallow network consisting of a convolution layer that keeps the original resolution of the input image. The shallow CNN extracts local features and retains the spatial information of various CFA images. The second stream is a deep CNN that extracts global

and complex features. In the proposed method, the residual-in-residual (RIR) structure proposed in [5] is adopted as the deep CNN. Fig. 2 shows a deep CNN architecture. The RIR consists of multiple residual groups (RGs) with long skip connections (Fig. 2(a)), and each RG has short skip connections connecting multiple residual blocks (RBs) (Fig. 2(b)). The proposed network consists of four RGs, and each RG consists of 20 RBs.

3.1 Universal demosaicking via PFF network

To incorporate the PFF into the demosaicking process for arbitrary CFA pattern images, the network starts by downsampling the CFA image into a three-channel image. Then, each CFA image is converted to the first estimated RGB image with a bilinear interpolation method adapted to the CFA pattern. We have designed the filters to efficiently remove various noises due to each CFA pattern by using the proposed deep CNN based on the RIR structure.

(a) Deep CNN architecture

(b) Residual block

Fig. 2 Deep CNN architecture

Fig. 3 Synthesis network

3.2 Filter flow synthesis

Using the features obtained from network composed of shallow and deep CNNs the filter flow for each RGB channel is synthesized by a three-layer network consisting of 3×3 convolution and PReLU layers (Fig. 3). Since PReLU improves the performance with nearly zero extra computational cost, we utilize PReLU instead of ReLU in this study [6]. These convolution layers output a 3D filter with filter flows corresponding to the RGB channels. The softmax function is applied instead of PReLU only for the last layer to obtain the coefficients of a normalized space-variant filter. The predicted image is obtained by multiplying the synthesized filter flow with the first estimated RGB image and summing the elements in the channel direction.

4. Experiments

For the experimental evaluation of our method, we use the Waterloo Exploration Database (WED) dataset [7] for network training. The WED dataset contains 4744 color images of nature. 64×64 patches are extracted randomly from the first estimated RGB image of the training dataset for each iteration. The hyperparameters in the training of our proposed net-

work are set as follows: the filter size is 9×9 , the batch size is 16, the number of epochs is 1000, and the initial learning rate of the Adam optimizer is 10^{-4} . We train our network on a single NVIDIA GeForce 2070, SUPER GPU.

4.1 Quantitative and qualitative comparisons

We quantitatively and qualitatively compare the proposed method with the state-of-the-art CNN-based demosaicking methods of the three-stage CNN structure (3-Stage) [8] and CDM CNN structure (CDM) [9]. As an objective evaluation, Table 1 shows the results for the color peak signal-to-noise ratio (CPSNR) and structural similarity (SSIM) measures. The CPSNR and SSIM are computed on the Kodak dataset comprising 24 images and the McMaster dataset comprising 18 images. We also add the average CPSNR and SSIM for all methods. As shown in Table 1, the proposed method shows higher CPSNR and SSIM than the other demosaicking methods for both datasets.

As a visual comparison, Fig. 4 and Fig. 5 respectively show part of the demosaicking results obtained by 3-Stage, CDM, and the proposed method for the Kodak and McMaster datasets. As shown in Fig. 4 and Fig. 5, the proposed method restores object structures more accurately than the other methods.

Table 1 Quantitative comparison of CPSNR and SSIM

Kodak													
No.	3-Stage		CDM		Proposed		No.	3-Stage		CDM		Proposed	
	CPSNR	SSIM	CPSNR	SSIM	CPSNR	SSIM		CPSNR	SSIM	CPSNR	SSIM	CPSNR	SSIM
01	41.70	0.9931	41.31	0.9925	41.70	0.9939	13	37.51	0.9901	37.23	0.9892	38.22	0.9918
02	41.95	0.9826	41.49	0.9800	42.06	0.9842	14	40.30	0.9912	40.03	0.9903	40.55	0.9918
03	39.15	0.9844	39.25	0.9848	45.32	0.9918	15	41.60	0.9854	41.34	0.9840	41.93	0.9871
04	43.22	0.9884	42.60	0.9869	43.82	0.9897	16	45.62	0.9921	45.07	0.9917	43.59	0.9917
05	40.93	0.9940	40.56	0.9936	41.24	0.9945	17	42.98	0.9900	42.75	0.9900	43.57	0.9908
06	42.19	0.9916	42.10	0.9910	42.12	0.9918	18	39.00	0.9853	38.62	0.9847	40.01	0.9874
07	44.99	0.9924	44.71	0.9923	44.76	0.9930	19	42.84	0.9881	42.35	0.9878	43.14	0.9893
08	37.69	0.9892	37.09	0.9883	39.53	0.9920	20	43.28	0.9811	42.76	0.9805	43.43	0.9820
09	44.53	0.9861	43.92	0.9860	44.20	0.9870	21	41.61	0.9849	41.32	0.9854	42.01	0.9870
10	44.21	0.9876	43.67	0.9875	44.41	0.9887	22	40.52	0.9833	40.24	0.9826	40.98	0.9850
11	42.43	0.9910	42.02	0.9900	43.02	0.9919	23	45.07	0.9879	44.90	0.9876	45.47	0.9892
12	45.17	0.9897	44.93	0.9893	45.00	0.9901	24	37.20	0.9901	36.85	0.9891	38.54	0.9913
Kodak Ave.								41.90	0.9883	41.55	0.9877	42.44	0.9897
McMaster													
No.	3-Stage		CDM		Proposed		No.	3-Stage		CDM		Proposed	
	CPSNR	SSIM	CPSNR	SSIM	CPSNR	SSIM		CPSNR	SSIM	CPSNR	SSIM	CPSNR	SSIM
01	31.58	0.9440	31.14	0.9398	32.25	0.9531	10	41.21	0.9801	40.86	0.9788	41.79	0.9823
02	36.30	0.9562	36.11	0.9550	37.12	0.9624	11	41.88	0.9800	41.47	0.9776	42.35	0.9815
03	36.67	0.9810	36.25	0.9794	37.43	0.9933	12	41.73	0.9722	41.25	0.9706	42.10	0.9728
04	41.10	0.9925	40.20	0.9915	42.00	0.9934	13	42.50	0.9624	42.08	0.9606	42.56	0.9629
05	37.30	0.9722	36.73	0.9695	37.96	0.9757	14	40.42	0.9667	40.09	0.9650	40.89	0.9690
06	41.07	0.9796	40.51	0.9775	41.69	0.9816	15	40.67	0.9673	40.26	0.9647	41.00	0.9705
07	41.56	0.9849	41.20	0.9844	41.76	0.9860	16	37.25	0.9758	36.40	0.9705	38.12	0.9815
08	41.34	0.9824	40.96	0.9821	41.69	0.9847	17	36.94	0.9769	36.30	0.9734	37.88	0.9812
09	40.30	0.9735	39.95	0.9719	40.68	0.9757	18	38.50	0.9723	37.74	0.9710	38.73	0.9763
McM Ave.								39.33	0.9733	38.86	0.9713	39.89	0.9763

Fig. 4 Visual comparison for Kodak dataset

Fig. 5 Visual comparison for McMaster dataset

4.2 Other CFA pattern results

To verify the universality of the proposed method, we demonstrate demosaicking for various RGB CFA patterns, Bayer, Lukac, Yamanaka, and Modified Bayer, illustrated in Fig. 6. Table 2 shows the average measurement results for each pattern. The experimental results objectively demonstrate our method outperforms the other conventional methods in several evaluation metrics on the Kodak and McMaster benchmark datasets.

5. Conclusions

In this paper, we propose a novel universal demosaicking method using the framework of the PFF. Experiments on benchmark data objectively show that our method outperforms other conventional methods in several evaluation metrics. In addition, the proposed method can be applied to various CFA patterns with the same framework.

Fig. 6 CFA patterns

Table 2 Quantitative comparison of CPSNR and SSIM for various CFA patterns

Dataset	CFA	CPSNR	SSIM
Kodak	Bayer	42.44	0.9897
	Lukac	42.76	0.9896
	Yamanaka	42.44	0.9891
	Modified	42.46	0.9893
McMaster	Bayer	39.89	0.9763
	Lukac	39.85	0.9759
	Yamanaka	39.45	0.9748
	Modified	37.54	0.9731

References

- [1] R. Ramanath, W. E. Snyder, G. L. Bilbro and W. A. Sander III: Demosaicking methods for Bayer color arrays, *Journal of Electronic Imaging*, Vol. 11, No. 3, pp. 306–316, 2002.
- [2] S. Kong and C. Fowlkes: Image reconstruction with predictive filter flow, *arXiv:18 11.11482*, 2018.
- [3] S. M. Seitz and S. Baker: Filter flow, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 143–150, 2009.
- [4] D. P. Kingma and J. Ba: Adam: A method for stochastic optimization, *arXiv:1412.6980v9*, 2017.
- [5] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong and Y. Fu: Image super-resolution using very deep residual channel attention networks, *Proc. European Conference on Computer Vision*, pp. 286–301, 2018.
- [6] K. He, X. Zhang, S. Ren and J. Sun: Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification, *arXiv:1502.01852v1*, 2015.
- [7] K. Ma, Z. Duanmu, Q. Wu, Z. Wang, H. Yong, H. Li and L. Zhang: Waterloo exploration database: New challenges for image quality assessment models, *IEEE Transactions on Image Processing*, Vol. 26, No. 2, pp. 1004–1016, 2017.
- [8] K. Cui, Z. Jin and E. Steinbach: Color image demosaicking using a 3-stage convolutional neural network structure, *Proc. IEEE International Conference on Image Processing*, pp. 2177–2181, 2018.
- [9] R. Tan, K. Zhang, W. Zuo and L. Zhang: Color image demosaicking via deep residual learning, *Proc. IEEE International Conference on Multimedia and Expo*, 2017.

Daiki Arai received his B.E. degree from the Department of Software Science, Tamagawa University, Tokyo, Japan, in 2021. Currently, he is studying on the master's course in the Graduate School of Engineering, Tamagawa University. His research interests include image and video processing, and machine learning.

Taishi Iriyama received his B.E. and M.E. degrees in electronic information engineering from Tamagawa University, Tokyo, Japan, in 2017 and 2019, respectively. Currently, he is pursuing his Ph.D. degree in system sciences at Tamagawa University. His research interests include image and video processing, and machine learning.

Masatoshi Sato received his B.E. and M.E. degrees in electrical and electronics engineering and Ph.D. degree in information and communication sciences from Sophia University, Tokyo, Japan, in 2006, 2008, and 2011, respectively. In 2011, he became a research assistant with the Department of Information and Communication Sciences of Sophia University. From 2013 to 2017, he was an assistant professor with the Faculty of System Design of Tokyo Metropolitan University. In 2017, he joined the College of Engineering of Tamagawa University as an assistant professor. Since 2018, he has been an associate professor at Tamagawa University. His research interests are in the optimization of nonlinear networks, nonlinear circuit analysis, image processing using graph cuts, and convolutional neural networks. He is a member of the IEICE and IEEE.

Hisashi Aomori received his B.E., M.E., and Ph.D. degrees in electrical and electronics engineering from Sophia University, Tokyo, Japan, in 2002, 2004, and 2007, respectively. In 2007, he became an assistant professor with the Department of Electrical and Electronics Engineering of Sophia University. From 2009 to 2012, he was an assistant professor with the Department of Electrical Engineering of Tokyo University of Science. In 2012, he joined the Department of Information System Technology of Chukyo University as a junior associate professor. Since 2017, he has been an associate professor at the Department of Electrical and Electronic Engineering of Chukyo University. His research interests are in image processing, bio-inspired computing, and cellular neural networks. He is a member of the IEEE, IEICE, and RISP.

Tsuyoshi Otake received his B.E., M.E., and Ph.D. degrees in electrical and electronics engineering from Sophia University, Tokyo, Japan, in 1991, 1993, and 1997, respectively. In 1997, he became an assistant professor with the Department of Electrical and Electronics Engineering of Sophia University. In 2004, he joined the Department of Media-Network Science at Tamagawa University, Tokyo, Japan, where he is currently a professor in the Department of Software Science. From September 2009, he spent a year in the Faculty of Information Technology at Pazmany Peter Catholic University, Budapest, Hungary, as a visiting professor. His research interests include image and video processing, machine learning, and the theory and application of cellular neural networks. He is a member of the IEEE, IEICE, and RISP.

(Received May 11, 2021; revised August 14, 2021)