Deep Demosaicking with Luminance and Chrominance Estimations

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-Honnachi, Showa-ku, Nagoya, Aichi 466-8666, Japan
E-mail: irymt3sc@engs.tamagawa.ac.jp, otake@ieee.org

Abstract A digital camera acquires images using a single electronic sensor with a color filter array (CFA). The raw image contains luminance, defined as a spatial map of intensity, and chrominance, defined as a spatial map of each color information. Since the luminance and chrominance components have different demosaicking complexities, they should be modeled separately. In this paper, we propose a novel convolutional neural network (CNN)-based demosaicking method that separately estimates the luminance and chrominance components. Specifically, we apply two-stage CNNs consisting of a luminance component estimation network and a chrominance component estimation network. The proposed method suppresses artifacts such as false colors and reduces the computational complexity. Experimental results on several benchmark datasets demonstrate that the proposed method provides results that are better or competitive with conventional demosaicking algorithms while reducing the computational complexity.

Keywords: demosaicking, convolutional neural network, luminance estimation

1. Introduction

Most digital still cameras acquire an image by using a single electronic sensor covered with a color filter array (CFA). The CFA is designed in such a way that the sensor corresponding to each pixel samples only one of the three primary colors, red (R), green (G), and blue (B). The most common CFA pattern is the Bayer pattern, which samples on a quincunx grid for G values and on independent rectangle grids for R and B values. To estimate the missing two-color information from the CFA image, an interpolation process called demosaicking is performed. The simplest demosaicking is to apply a linear interpolation, such as bilinear or bicubic interpolation, for each color channel. However, a simple linear interpolation leads to artifacts such as zipper effects and false colors. To improve the restored image quality, many demosaicking approaches first estimate a luminance component, which has the most significant effect on the perceptual quality, and use it as a guide to interpolate the output color values. Alleysson et al. [1] proposed a human visual system-based demosaicking approach according to the observation that the luminance and chrominance signals of a CFA image are well separated in the frequency domain. They estimated the luminance and chrominance components by low-pass and high-pass filtering, respectively, and showed that the sum of these components is equivalent to the color information per pixel.

Recently, convolutional neural network (CNN)-based demosaicking methods have been explored and shown state-of-the-art performance [2–5]. Tan et al. [3] predicted the residuals between the initial interpolated image and the training image by a two-stage network consisting of intermediate G estimation and final RGB estimation. The intermediate G is utilized as a guide to reconstruct the final RGB. Cui et al. [4] extended the above-mentioned network to three stages by adding a middle stage that estimates RG and GB using different networks. These methods exploit the initial interpolated image as an input to estimate the output RGB values. However, the demosaicking noises caused by the initial interpolation make it difficult to analyze features obtained with their network. Moreover, estimating the pixel values with mixed luminance and chrominance components increases the computational complexity.

In this paper, we propose a novel demosaicking method that estimates the luminance and chromi-
nance components independently using CNNs. Specifically, we construct a two-stage network with luminance and chrominance estimation. The first stage estimates the luminance component. The output of the first stage is used as a guide image for the input of the second stage. The second stage estimates the chrominance component, which is defined as the difference between the outputs of the first stage and the training image. Finally, the demosaicking results are obtained by adding the outputs of each stage. The proposed method reduces the computational complexity by separately modeling the demosaicking for the luminance and chrominance components, and suppresses the false colors by exploiting the estimated luminance component as a guide to predict the chrominance component.

In our experiments on several benchmark datasets, the proposed method shows better or competitive demosaicking results compared with state-of-the-art approaches while reducing the computational complexity. For perceptual evaluations, we confirmed that the proposed method reduces artifacts such as false colors that appear in the conventional demosaicking approaches.

2. Proposed Method

2.1 Framework

Our demosaicking framework for separately estimating the luminance and chrominance components using CNNs is shown in Fig. 1. As preprocessing, the input CFA image is packed into a four-channel image with a quarter resolution to make the spatial pattern invariant and improve the spatial resolution of the network input. In the first stage, the network estimates the luminance component from the packed CFA image. The input for the second stage is the difference between the packed CFA image and the estimated luminance component. Therefore, the estimated luminance component is used to guide the chrominance estimation. In the second stage, the network estimates the chrominance component corresponding to each channel. In the proposed method, the luminance component is calculated as \( L = (R + G + B)/3 \), and the chrominance components for each channel are defined by \( R - L \), \( G - L \), \( B - L \). Finally, demosaicked image is obtained by adding the luminance component to each channel of the chrominance component.

2.2 Architecture

To verify the effectiveness of the proposed framework, we constructed the simple network architecture illustrated in Fig. 2. As shown in Fig. 2, our network mainly consists of an initial convolution, a deep convolution, up-sampling, and reconstruction. The initial convolution embeds the input image into the initial features by one-layer convolution. The deep convolution extracts the global and complex features from the initial features by a deep network with simply stacked residual blocks. The residual block consists of a two-layer convolution and an activation function. The extracted initial features are transported to the tail of the residual blocks by skip connections to retain the shallow information and improve the learning efficiency. The extracted deep features are up-sampled to the desired resolution by an upscale module. The upscale module is implemented by one-layer convolution and a pixel shuffler. Finally, the extracted features are reconstructed to the output image by a one-layer convolution. In our networks, the filter size of all convolution layers is \( 3 \times 3 \) and the activation function is a parametric rectified linear unit (PReLU).

The parameters of the proposed network are opti-
Fig. 3 Qualitative comparison of demosaicking results on testing datasets
mized using the loss for the luminance image and the loss for the demosaicked image. The loss function is defined using the $l_1$-norm as

$$\text{Loss}(\Theta) = \frac{1}{N} \sum_{i=0}^{N} \left( \|f_1(I^i, \Theta_1) - L_{GT} \|_1 + \|f_2(I^i, \Theta_1, \Theta_2) + f_3(I^i, \Theta_1) - L_{GT} \|_1 \right) \tag{1}$$

where $\Theta_1$ and $\Theta_2$ indicate the parameters of the first-stage network $f_1$ and second-stage network $f_2$, respectively. $I_{GT}$ indicates the ground truth and $L_{GT}$ indicates the luminance image generated from the ground truth. $N$ is the number of training observations.

3. Experiments

We adopt the Waterloo Exploration Database (WED) [6] dataset for training. The WED dataset consists of 4744 full-color natural images of various scenes. Our dataset is randomly cropped to a patch size of 96 × 96 per epoch and sampled with the Bayer CFA pattern. The mini-batch size is 32. The proposed networks for luminance and chrominance estimations contain 40 and 20 residual blocks, respectively. All features of our networks have 64 channels. The parameters of our network are optimized by the Adam optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 10^{-8}$. The initial learning rate is $10^{-4}$ and, is divided by 2 every 200 epochs. The total number of epochs is 1000. We adopt the Kodak, McMaster, and Urban100 datasets for testing. The Kodak dataset consists of 24 color images with 768 × 512 resolution. The McMaster dataset consists of 18 color images with 500 × 500 resolution. The Urban100 dataset consists of 100 images with various resolutions. The proposed method is compared with the non-CNN-based demosaicking methods of GBFT [7], RI [8], and MLRI [9] and the CNN-based demosaicking methods of JDD [2], 2-stage [3], 3-stage [4], and RNAN [5].

3.1 Qualitative results

The demosaicking results of the conventional demosaicking methods and the proposed method for the testing dataset are shown in Fig. 3. For image19 of the Kodak dataset, the proposed method removed the artifacts such as false colors and false interpolation structures observed with the other demosaicking methods. For image72 of the Urban100 dataset, the proposed method clearly suppresses the false colors compared with the state-of-the-art demosaicking methods. For image5 of the McMaster dataset, containing color-saturated regions and edges, the proposed method recovered a comparable image to the original.

3.2 Quantitative results

We evaluate the proposed method by using the composite peak signal-to-noise ratio (CPSNR) and structural similarity (SSIM) as assessment indices. Table 1 shows the assessment indices for the conventional demosaicking methods and the proposed method on the benchmark datasets and the number of hidden layer parameters. The results for CPSNR and SSIM show that the proposed method provides better or competitive performance compared with RNAN and outperforms the other demosaicking methods. In particular, for the Urban100 dataset, CPSNR and SSIM for the proposed method outperform the second best method by 0.77[dB] and 0.0015, respectively. Furthermore, the comparison of the number of parameters shows that the proposed model requires about half the number of parameters compared with the RNAN model with comparable demosaicking performance. This indicates that the proposed method provides a better tradeoff between the computational cost and the demosaicking performance.

4. Conclusion

A novel deep demosaicking method with luminance and chrominance estimations is proposed. By separately estimating the luminance and chrominance components, the proposed method achieves better or competitive performance compared with the state-of-the-art demosaicking algorithms, while using a network with fewer hidden layer parameters. In a subjective comparison, it was confirmed that the proposed method reduces artifacts such as false colors and false interpolation structures compared with conventional demosaicking methods.

References

[1] D. Alleysson, S. Susstrunk and J. Herault: Linear demosaicing inspired by the human visual system, IEEE Transactions on Image Processing, Vol. 14, No. 4, pp. 439-449, 2005.
[2] M. Gharbi, G. Chaurasia, S. Paris and F. Durand: Deep joint demosaicking and denoising, ACM Transactions on Graphics, Vol. 35, No. 6, pp. 1-12, 2016.
[3] R. Tan, K. Zhang, W. Zuo and L. Zhang: Color image demosaicking via deep residual learning, Proc. 2017 IEEE International Conference on Multimedia and Expo, pp. 793-798, 2017.
[4] K. Cui, Z. Jin and E. Steinbach: Color image demosaicking using a 3-stage convolutional neural network structure, Proc. 2018 25th IEEE International Conference on Image Processing, pp. 2177-2181, 2018.
Table 1  Quantitative comparison by average of assessment indices on testing datasets

|                | Kodak | McMaster | Urban100 |
|----------------|-------|----------|----------|
|                | CPSNR | SSIM     | CPSNR    | SSIM    |
| GBTF           | 40.41 | 0.9855   | 33.94    | 0.9279  |
| RI             | 38.56 | 0.9787   | 36.47    | 0.9604  |
| MLRI           | 39.58 | 0.9837   | 36.02    | 0.9724  |
| JDD            | 41.70 | 0.9883   | 39.14    | 0.9713  |
| 2-stage        | 41.96 | 0.9881   | 38.96    | 0.9702  |
| 3-stage        | 42.18 | 0.9887   | 39.33    | 0.9721  |
| RNAN           | 43.02 | 0.9903   | 39.71    | 0.9725  |
| Ours           | 42.96 | 0.9905   | 39.75    | 0.9741  |

| Hidden layer parameters | |
|-------------------------|---|
| GBTF                    | 35.66 | 0.9708 |
| RI                      | 34.53 | 0.9686 |
| MLRI                    | 34.88 | 0.9749 |
| JDD                     | 38.22 | 0.9825 |
| 2-stage                 | 38.17 | 0.9740 |
| 3-stage                 | 38.49 | 0.9830 |
| RNAN                    | 39.76 | 0.9840 |
| Ours                    | 40.53 | 0.9855 |

[5] Y. Zhang, K. Li, K. Li, B. Zhong, Y. Fu: Residual non-local attention networks for image restoration, Proc. International Conference on Learning Representations, 2019.
[6] 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.
[7] I. Pekkucukksen and Y. Altunbasak: Gradient based threshold free color filter array interpolation, Proc. 2010 IEEE International Conference on Image Processing, pp. 137-140, 2010.
[8] D. Kiku, Y. Monno, M. Tanaka and M. Okutomi: Residual interpolation for color image demosaicking, Proc. 2013 IEEE International Conference on Image Processing, pp. 2304-2308, 2013.
[9] D. Kiku, Y. Monno, M. Tanaka and M. Okutomi: Minimized-Laplacian residual interpolation for color image demosaicking, Proc. International Society for Optical Engineering, Vol. 9023, 2014.

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 at the Graduate School of System Sciences, Tamagawa University, Tokyo, Japan. 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 his 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 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 image processing, bio-inspired computing, and cellular neural networks. He is a member of the RISP, IEICE, and IEEE.
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 theory and application of cellular neural networks. He is a member of the IEEE, IEICE, and RISP.

(Received May 11, 2021; revised August 6, 2021)