Potential Improvement of CNN-Based Colorization for Non-Natural Images

Ayumu Shinya, Kazuki Mori
Graduate School of Information Science and Engineering
Ritsumeikan University
Shiga, Japan
is0217rx@ed.ritsumei.ac.jp, is0191kh@ed.ritsumei.ac.jp

Tomohiro Harada, Ruck Thawonmas
College of Information Science and Engineering
Ritsumeikan University
Shiga, Japan
harada@ci.ritsumei.ac.jp, ruck@is.ritsumei.ac.jp

***Abstract*—This paper discusses colorization of grayscale non-natural images using convolutional neural networks (CNNs). We show that current state-of-the-art colorization methods which use deep CNNs trained on natural images do not perform well on non-natural images, such as images of ukiyo-e. We therefore propose how to improve such methods without retraining their networks.**

Keywords—Image colorization; Convolutional neural networks; Deep learning

I. Introduction

Colorization—the process of adding colors to grayscale images or monochrome videos—has been an active research field. Conventionally, optimization methods that optimize every pixel based on user inputs or reference images used to be the mainstream for colorization. However, with the rapid advance of deep learning in recent years, methods that use convolutional neural networks (CNNs) to automatically perform colorization have dominated the current trend. In such methods, deep networks are trained with a large number of images so that they can output a colorized image given a grayscale image.

Existing CNN-based colorization methods have a common problem: the types of images that can be colorized depend strongly on the types of images used for training. For example, a network trained on natural images do not perform well on non-natural images. In this paper, we investigate the performance of a representative CNN-based colorization method by Zhang et al. [1] on non-natural images and propose a potential improvement to it.

II. Related Work

A. Covolutional Neural Networks

CNNs are feed-forward networks mainly used in the field of image recognition. In addition to fully-connected layers seen in normal neural networks, a CNN contains multiple convolutional layers, which extract features of the input image, and pooling layers, which reduce the position sensitivity of the extracted features. Each convolutional layer takes as input the output of the previous layer, and extracts various features of the input image by performing convolution on its input using multiple "small images" called filters. As a result, layers which are closer to the output of a CNN tend to extract higher-level features.

Fig. 1. A sample colorization result by the method of Levin et al. (excerpted from [3]). Left: a grayscale image with user inputs, right: a resulting colorized image.

Fig. 2 A sample colorization result by the method of Gupta et al. Left: a grayscale image, middle: a reference image, right: a resulting colorized image.

B. Existing Colorization Methods

Huang et al. [2] and Levin et al. [3] proposed localized colorization using optimization based on sparse user inputs (hereinafter referred to as "points") for a grayscale image of interest. Each point is specified with a location on the image and an arbitrary color, and then its surrounding region is colorized with the given color. As shown in Fig. 1, these methods require many points to obtain good results. When the user is unfamiliar with inputting of points, accurate colorization cannot be achieved.

Chia et al. [4] and Gupta et al. [5] proposed colorization using reference images instead of user-provided points. Their methods transfer pixel values from a given reference image to a grayscale image of interest using optimization. While these methods can achieve accurate colorization when the reference image is similar to the grayscale image, they do not perform well when the two images have no or few shared contents, as shown in Fig. 2.

Recently, with the advance of deep learning, methods which train CNNs for fully-automatic colorization have been proposed, e.g., Iizuka et al. [6]. This kind of method uses a CNN to extract features from a given grayscale image and colorizes it based on the extracted features. Colorization is done by just inputting the

978-1-5386-2615-3/18/$31.00

Fig. 3. A sample colorization result by the method of Iizuka et al. Left: a grayscale image, right: a resulting colorized image.

Fig. 4. Sample colorization results by the method of Zhang et al. Top: a grayscale image (left) and an automatically-colorized image (right). Bottom: colorized images with user points (insets).

grayscale image into the network. However, unlike the aforementioned methods, it can generate only a single colorized image for a grayscale image since it does not require any user inputs. Moreover, the types of images that can be colorized by this CNN-based method are restricted by the types of images used for training the network. For example, a network trained on natural images can generate an appropriate color image if a given grayscale image is a natural image. By contrast, as shown in Fig. 3, it does not work well on non-natural images such as those of ukiyo-e.

C. *The Method of Zhang et al.*

In [1], Zhang et al. addressed a problem of existing CNN-based colorization methods that they can only generate a single colorized image and do not allow users to specify their preferences. In order to solve this, Zhang et al. proposed a network that can predict the color distribution of a grayscale image given the image and user points. The network has two variants: Global Hints Network and Local Hints Network; both of them use the same main colorization branch and are trained end-to-end. During training, the Global Hints Network uses as inputs color images and grayscale images. Whereas, the Local Hints Network uses as inputs randomly-simulated points and grayscale images, and learns to output colorized images along with predicted color distributions for grayscale images.

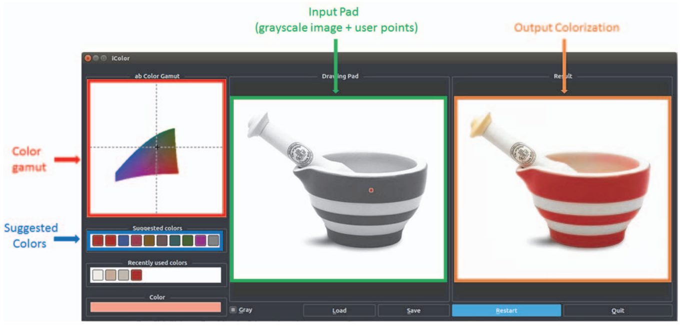

Fig. 5. UI of the system created by Zhang et al. Shown in blue box are suggested colors.

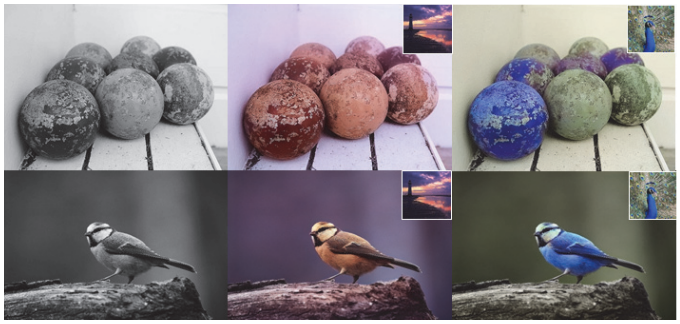

Fig. 6. Sample colorization results by the method of Zhang et al. (excerpted from [1]). Insets are reference images.

The method proposed by Zhang et al. not only can perform fully-automatic colorization as existing CNN-based methods but also allows users to specify their preferences via user points or reference images. Figure 4 (bottom) shows multiple colorization results based on user points for a single grayscale image. In addition, Zhang et al. raised a problem of colorization methods based on user points: they cannot achieve accurate colorization when the user is not familiar with inputting of points. Zhang et al. solved this problem by predicting the color distribution of every pixel in the grayscale image with the Local Hints Network and showing the user a color palette as shown in Fig. 5. The color palette displays the most probable colors for each point specified by the user. By suggesting colors, the system helps novice users quickly learn to create realistic colorizations. On the other hand, the Global Hints Network takes as input a color image rather than user points. It then uses the color image as a reference image for colorizing the grayscale image. Figure 6 shows some examples of colorization using color images without user points. As can be seen, the method can perform colorization properly even when the grayscale images and the reference images have no similar contents.

III. Experiments

By incorporating user points and reference images, the method proposed by Zhang et al. solves the problem of fully-automatic colorization methods that they can only generate a single colorized image for a grayscale image. Moreover, since the system suggests probable colors for each specified point, not only experienced users but also novice users can create appropriate colorizations. However, the problem that the types of images that can be colorized are restricted by the types of images used for training still remains unsolved. In this section, we examine the performance of the Global Hints Network when it takes as inputs non-natural images.

A. Automatic Colorization

We input to the network only grayscale images of ukiyo-e without reference images. Sample colorization results are shown in Fig. 7.

B. Colorization Using Reference Images

In this experiment, we use ukiyo-e for both grayscale images and reference images. Sample colorization results are shown in Fig. 8.

IV. Discussions And Proposed Improvement

As shown in Fig. 7, for images of ukiyo-e, only limited parts of the grayscale images such as kimono are colorized, while human skin or background are not. When we used color images of ukiyo-e as reference images as shown in Fig. 8, some regions containing kimono are also colorized. However, human skin and background are colorized with the same color as a whole, which cannot be considered to be appropriate.

As mentioned earlier, the method proposed by Zhang et al. can also perform colorization using user points. Figure 9 shows colorization results by the Local Hints Network when we added some points to grayscale images of ukiyo-e and selected colors suggested by the system. It can be seen that the network can properly colorize a grayscale image even if it is not a natural image, as long as plausible user points are given. Based on these results, the problem that colorization networks trained on natural images do not perform well on non-natural images can be solved by using user points. Moreover, we propose to generate these points automatically instead of requiring user interactions. This can be done by matching the given grayscale image and its reference image, which has been used in several optimization-based colorization methods.

References

[1] R. Zhang, J.Y. Zhu, P. Isola, X. Geng, A.S. Lin, T. Yu, and A.A. Efros, "Real-time user-guided image colorization with learned deep priors," ACM Transactions on Graphics (TOG), vol. 36, no. 4, ACM, 2017.

[2] Y.C. Huang, Y.S. Tung, J.C. Chen, S.W. Wang, and J.L. Wu, "An adaptive edge detection based colorization algorithm and its applications," Proceedings of the 13th annual ACM international conference on Multimedia, pp. 351-354, ACM, 2005.

[3] A. Levin, D. Lischinski, and Y. Weiss, "Colorization using optimization," ACM transactions on graphics (TOG), vol. 23, no. 3, ACM, 2004.

[4] A.Y.S. Chia, S. Zhuo, R.K. Gupta, Y.W. Tai, S.Y. Cho, P. Tan, and S. Lin, "Semantic colorization with internet images," ACM Transactions on Graphics (TOG), vol. 30, no. 6, ACM, 2011.

[5] R.K. Gupta, A.Y.S. Chia, D. Rajan, E.S. Ng, and H. Zhiyong, "Image colorization using similar images," Proceedings of the 20th ACM international conference on Multimedia, ACM, 2012.

[6] S. Iizuka, E. Simo-Serra, and H. Ishikawa,. "Let there be color!: joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification," ACM Transactions on Graphics (TOG), vol. 35, no. 4, ACM, 2016.

Fig. 7. Automatic colorization of ukiyo-e by the method of Zhang et al. (without reference images). Left: grayscale images, right: colorized images.

Fig. 8. Colorization of ukiyo-e by the method of Zhang et al. (with reference images). First column: grayscale images, first row: reference images.

Fig. 9. Colorization of ukiyo-e by the method of Zhang et al. (with user points). Left: grayscale images, middle: user points, right: colorized images.