Journal of Semiconductor Engineering Vol. 2, Issue 1, March, 2021
https://doi.org/10.22895/jse.2020.0101
Received 24 November 2020, Accepted 22 January 2021 ISSN(On-line) 2733-6719

GAN-based Color Palette Extraction System by Chroma Fine-tuning with Reinforcement Learning

Sanghyuk Kim[1] and Suk-Ju Kang[2]

[1] QCT WT RF SW Applications Engineering, Qualcomm CDMA Technologies
[2] Department of Electronic Engineering, Sogang University

Corresponding Author Suk-Ju Kang (sjkang@sogang.ac.kr)

ABSTRACT

As the interest of deep learning, techniques to control the color of images in image processing field are evolving together. However, there is no clear standard for color, and it is not easy to find a way to represent only the color itself like the color-palette. In this paper, we propose a novel color palette extraction system by chroma fine-tuning with reinforcement learning. It helps to recognize the color combination to represent an input image. First, we use RGBY images to create feature maps by transferring the backbone network with well-trained model-weight which is verified at super resolution convolutional neural networks. Second, feature maps are trained to 3 fully connected layers for the color-palette generation with a generative adversarial network (GAN). Third, we use the reinforcement learning method which only changes chroma information of the GAN-output by slightly moving each Y component of YCbCr color gamut of pixel values up and down. The proposed method outperforms existing color palette extraction methods as given the accuracy of 0.9140.

KEY WORDS

Deep neural network, generative adversarial network, reinforcement learning

1. INTRODUCTION

Currently, a methodology for automating color extraction in various fields is being studied. In particular, interest in this is increasing in related fields such as automobiles and construction where accurate color analysis is required. Color palette extraction from an image is to understand the information in an image with image processing and computer vision technologies. The technologies related to color have the inherent problem of color recognition from images. It is that various light sources make impossible to analyze color precisely because of mutable aspect of color itself. Conventional methods approach this problem by downsizing color range [1] or clustering pixel values [2]. Also, with the development of DNNs, there has been many trials to extract colors close to the ground truth by supervised learning expecting end-to-end system. However, not like object classification, color classes are not distinguishable because of their small gap between them. That means classifiers are only able to classify colors not that specific and accurate. Another problem is that the largest number of pixels do not always represent the image. Therefore, a novel methodology is required to extract the accurate color.

In order to solve this problem, this paper proposes a novel generative adversarial network (GAN)-based [3] color palette extraction system by chroma fine-tuning with reinforcement learning [4]. The processes of the proposed system are mainly three steps as follows. First, we use 4-channel (RGBY) images to create feature maps by transferring the backbone network with the well-trained model-weight verified at super resolution convolutional neural networks (SRCNNs) [5]. Among trained models, SRCNNs contain a lot of hidden information for each pixel since this model needs to be upscaled. Second, feature maps are trained for the color-palette generation with a GAN. Through the interaction of a generator and discriminator, it

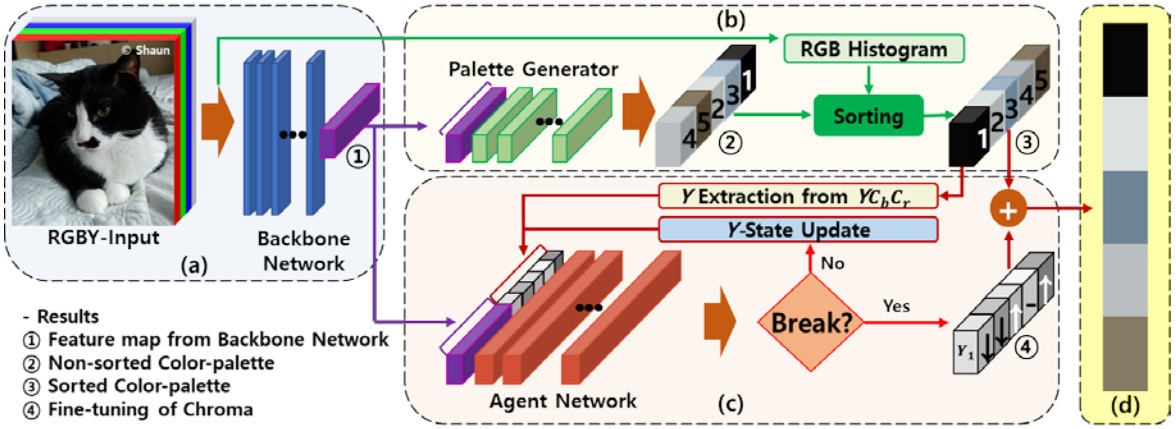


Figure 1: Overall flowchart of the proposed color palette extraction: (a) feature map generation from RGBY input image, (b) color palette generation with GAN, (c) fine-tuning of chroma with reinforcement learning and (d) output (color-palette).

demonstrates the color combination that best represent an input image. Third, with RL, the proposed system changes chroma information of the color palette from the GAN process by gradually moving each Y component of YCbCr color gamut up and down.

Overall, there are three main contributions as follows. First, a supervisor, discriminator in GAN, is placed when training the model so that there is no heterogeneity than the learning results of a simple single convolutional neural network (CNN) model. Second, when conventional color extraction cannot avoid the effect of various light sources, we construct customized database based on paintings and movies. These sources ensure that the chroma of an entire image is not biased to one side. Third, gradual fine-tuning with RL was made at the reference point, making the color subtle. By dramatically reducing the number of classes from 28 × 28 × 28 to only 11, the system performance is improved. We posted experimental results by comparing color palette from both conventional method and ours to figure out how useful searching colors in limited class-number.

The remainder of this paper is organized as follows. Section 2 introduces the details of the proposed method. Section 3 presents experimental results, and the final section presents the conclusion.

2. PROPOSED METHOD

In this section, we introduce the operation process of our color palette extraction system in detail. Figure 1 shows an overall flowchart of the proposed system. Mainly, the operation process of it is three parts: feature map generation, color palette generation, and fine-tuning of chroma. The detailed process is as follows.

2.1 Feature Map Generation

The input image has 4 channels, RGB and Y from YCbCr gamut. The reason is during the fine-tuning process as shown in Figure 1. (c), the process handles Y component. Therefore, in the backbone network, each hidden layer has Y information as well as RGB information; it results in the feature maps including Y information, naturally. The backbone network is from SRCNN with transfer learning [6]. There are three reasons. First, the network in the task of super resolution (SR) should consist of elements of layers for extracting pixelwise information. Therefore, the network is appropriate to the generated feature map for encompassing pixel information of an entire image. Second, using a validated training model in similar image processing tasks can yield better performance than when learning whole layers. In particular, the distribution of elements in each layer is evenly distributed, which can lead to more hidden pixelwise information. Third, we do not need to train every layer, which saves training time.

We only transfer the model just before upscaling or deconvolution. This is because we do not reconstruct the image, but the output is only a part of the information representing the image. After transferred deep neural network (DNN) model, one fully connected (FC) layer is arranged to generate 1024 feature maps, which unify the size of the output feature map for the input of color palette generation and fine-tuning of chroma process.

2.2 Color Palette Generation

Figure 2 shows the overall training flowchart of color palette generation. There are primary two process, generator and discriminator. The generator as shown in Figure 2(a) consists of the backbone network including transferred DNN model and one FC layer with a ReLU layer and two FC layers with ReLU layers. The last three FC layers are trained in this process. The FC layer in the backbone network is trained in only this process, so the layer is frozen in fine-tuning process as shown in Figure 2(a). We train the generator before training with the discriminator.

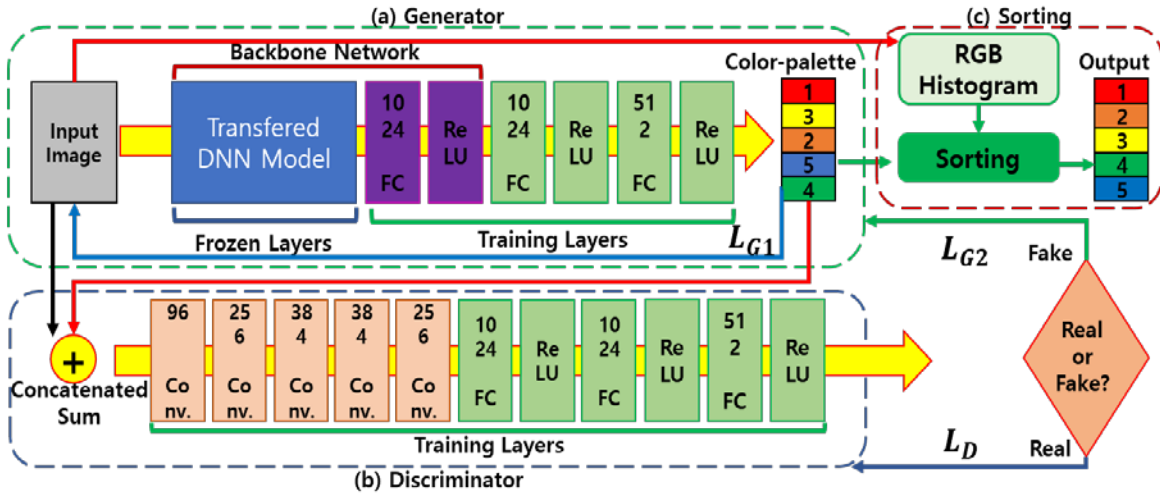


Figure 2: Overall training flowchart of color palette generation with GAN: (a) generator, (b) discriminator and (c) sorting.

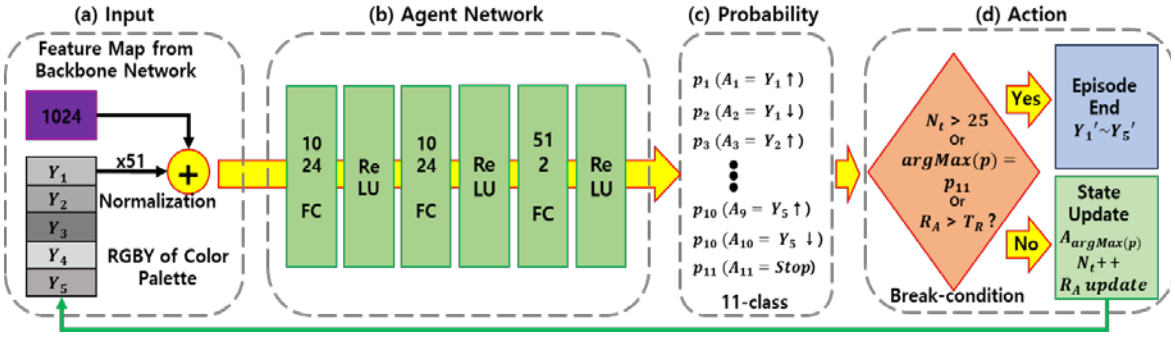


Figure 3: Overall feed-forward process of fine-tuning of Y-component with reinforcement learning: (a) Input, (b) agent network, (c) probability of each action-class and (d) action.

The reason is that if we train two networks together without separate training process, the output of the generator is not converged or overfit the majority color. If this is the case, the discriminator loses proper judgment, and the whole system behaves abnormally. Therefore, instead of using an initialized generator, the generator that has been trained to the backpropagation of the generator L_{G1} is used for training two DNN together.

The input of the discriminator in Figure 2(b) is to concatenate the image of the original RGB-input and the image by simply increasing each pixel information about five pixels to the size of the input image of discriminator. Thus, the input channel of the discriminator is six. The DNN of discriminator is based on AlexNet [7] which is a shallow network. It identifies two things; the color palette can represent the input image and its color combination is not too overfitted on a specific color despite the different colors in the image. Therefore, the loss of the generator, $LG2$, and loss of the discriminator, LD, apply in the case of fake and real decision respectively in backpropagation. The generated color palette is ranked by the sorting process based on RGB histogram as shown in Figure 2(c). The RGB histogram is binning through clustering 16 each channel. The output colors are ranked by the closest cluster.

2.3 Fine-tuning of chroma

Figure 3 shows the overall feed-forward process of fine-tuning of Y-component. We can create a color palette to some extent with the previous generation process, but we do a fine-tuning process to make the color more vivid, especially chroma. The input of fine-tuning process is divided into two, feature maps from backbone network and the flat-shape RGBY of the output color palette enlarged by 51 times to balance. All input information is normalized between zero to one. Agent network consists of 3 FC layers with ReLU layers.

The output of agent network control Y-component of the palette up and down. The number of the output of agent net is only 11 which has revolutionarily simplified the problem. As applying general CNN methods, the exact solution class of one output pixel is 16,777,216 cases, which is the possibility 24-bit RGB-channel 28×28×28. Therefore, too many classes result

Table 1: Accuracy Comparison of Proposed Methods with Different Numbers of Correct Color-patch

The Number of Correct Color-patch	3	4	5
K-mean Clustering	0.6214	0.5824	0.5203
ColorMind	0.7321	0.6532	0.6252
Backbone-EDSR without Fine-tuning	0.9252	0.8834	0.8585
Backbone-EDSR with Fine-tuning	0.9354	0.8922	0.8712
Backbone-MDSR without Fine-tuning	0.9504	0.9024	0.8801
Backbone-MDSR with Fine-tuning	0.9578	0.9126	0.8968
Backbone-RDN without Fine-tuning	0.9702	0.9223	0.9013
Backbone-RDN with Fine-tuning	0.9745	0.9272	0.9180

Figure 4: Results of the proposed system and other conventional methods: (a) input, (b) ground truth, (c) k-mean, (d) ColorMind, (e) proposed method without fine-tuning and (f) proposed method with fine-tuning.

in the convergence of DNN. To solve these limitations, our proposed method approaches fine-tuning. We define starting point from the output of color palette generation process, and then as agent of RL model changes only Y-channel, color palette is gradually heading to the target RGB color palette. Our approach reduces the class of model which only has 11-class as shown in Figure 3(c).

The break conditions of our RL model as shown in Figure 3(d) are three. First, the number of progresses is limited as 25. Second, when $p11$, the probability of stop action, $A11$ after softmax [8], is the highest, the episode is finished Third, the accumulated reward, RA, is over the threshold of max reward, TR (0.9). The reward occurring in t-th, Rt, is as below:

$$R_t = \left\{\sum_{i=1}^{5}(5 - |p_{2i} - p_{2i+1}|) + p_{11}\right\} / 6, \quad (1)$$

$$R_A = \sum_{i=1}^{t}\{(0.5)^{1+t-i} \times R_i\}, \quad (2)$$

where R_t is related to the movement of each Y component. A similar probability for a specific Y value means that the Y value is close to convergence. Since the current motion is much import than the past motion, the sum that reduced the importance of previous reward by half is used as RA. Otherwise, the RGBY state is update and the RL process continues until end.

3. EXPERIMENTAL RESULTS

We trained the model by 200 epochs. For the deep learning library, Tensorflow and Keras in Anaconda 3.0 environment was used. The laptop has Intel core i7-6700HQ (2.60Hz), 960M of NVIDIA GPU and 16 GB of RAM. The dataset is constructed by around 100,000 paintings, and 100,000 movie-sequence. We randomly applied crop to the input data and used it as training data. This is to ensure that the model does not recognize that the largest number of pixels always represent the image.

Figure 4 shows the result of our proposed method and conventional methods for the evaluation qualitatively. K-means algorithm [2] was clustering method of machine learning, and [9] was based on generative adversarial network (GAN). We used color quantization, ground truth [10]. Other conventional methods represent image-itself which means they focus on the number of pixels. Therefore, there is a tendency to represent invisible color as representative color visually. Also, conventional methods did not handle the balance of color combination. In the same vein, they frequently focused on the background or the main object region. Especially, the flower image in the second column of Figure 4 had green leaves and cyan-tone background, but conventional methods could not two features together. The proposed method extracted both information. In addition, conventional methods were hard to extract color palette from the image has a variety color tones like fourth column of Figure 4, but our method could extract it well.

Table 1 showed that the accuracy of all methods. For quantitative evaluation, we used the error rate ± 15 of the RGB-channel in each color patch of ground truth to determine accuracy. We measured the accuracy differently depending on how many number of the five fitted. We also applied three backbone networks [11], [12] with and without our fine-tuning process to see how each network affects accuracy. The accuracy of conventional methods were not over even 0.75 every case. Regardless of the backbone network,

our methods were over 0.85.In addition, when the fine-tuning process was used, the performance was improved by 0.0102 on average.

4. CONCLUSION

In this paper, we proposed a novel color palette extraction system by chroma fine-tuning with reinforcement learning. First, we used RGBY images for creating feature maps by transferring the backbone network. Second, feature maps were trained to GAN-based color-palette generation. Third, we used the reinforcement learning method to change chroma information. The experiments have shown successful color palette generation in both quantitative and qualitative terms with the accuracy of 0.9140.

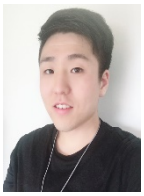
ACKNOWLEDGMENT

This research was supported by a grant (19PQWO-B153369-01) from Smart road lighting platform development and empirical study on test-bed Program funded by the Ministry of the Interior and Safety of Korean government, and the Ministry of Science and ICT (MSIT), Korea, under the Information Technology Research Center (ITRC) support program (IITP-2021-2018-0-01421) supervised by the Institute for Information & Communications Technology Promotion (IITP).

REFERENCES

[1] W. K. Leow, and R. Li, " The analysis and applications of adaptive-binning color histograms," Computer Vision and Image Understanding, vol. 94. No. 1-3, pp.67-91, 2004.
[2] A. Ahmad, and L. Dey, “A k-mean clustering algorithm for mixed numeric and categorical data," Data & Knowledge Engineering, vol. 63, no. 2, pp. 503-527, 2007.
[3] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative Adversarial Nets,” Proceedings of Advances in Neural Information Processing Systems 27 (NIPS 2014), pp. 2672-2680, 2014.
[4] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing Atari with Deep Reinforcement Learning," arXiv, 2013.
[5] C. Dong, C. C. Loy, K. He and X. Tang, "Image Super-Resolution Using Deep Convolutional Networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, no. 2, pp. 295-307, 2015.
[6] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A Survey on Deep Transfer Learning," arXiv, 2018.
[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," Proceedings of Advances in Neural Information Processing Systems 25 (NIPS 2012), pp. 1097-1105, 2012.
[8] W. Liu, Y. Wen, Z. Yu, and M. Yang, "Large-Margin Softmax Loss for Convolutional Neural Networks," Proceedings of International Conference on Machine Learning (ICML), vol. 2, no. 3, p. 7, 2016.
[9] Colormind, http://colormind.io/
[10] D. Bloomberg, "Color Quantization Using Modified Median Cut," http://leptonica. com/papers/mediancut. pdf, 2008.
[11] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced Deep Residual Networks for Single Image Super-Resolution," arXiv, 2017.
[12] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual Dense Network for Image Super-Resolution," arXiv, 2018.

AUTHOR BIOGRAPHIES



Sanghyuk Kim received the B.S. and M.S. degrees from Sogang University, Rep. of Korea, in 2017 and 2019 respectively. From Jan. 2019 to July 2019, he was a project manager for AI service Dev. team. He is currently a customer engineer for RF SW Applications Engineering in Qualcomm CDMA. His current research interests include deep learning systems, financial engineering, 5G telecommunication.

Suk-Ju Kang received a B.S. degree in Electronic Engineering from Sogang University, Rep. of Korea, in 2006 and a Ph.D. degree in Electrical and Computer Engineering from Pohang University of Science and Technology in 2011. From 2011 to 2012, he was a senior researcher at LG Display, where he was a project leader for resolution enhancement and multi-view 3D system projects. From 2012 to 2015, he was an assistant professor of Electrical Engineering at the Dong-A University, Busan. He is currently an associate professor of Electronic Engineering at the Sogang University. His current research interests include image analysis and enhancement, video processing, multimedia signal processing, circuit design for display systems, and deep-learning systems