Painting as your like: Colorization and Neural Style Transfer

Changjie Lu 1129503

College of Science and Technology, Wenzhou-Kean University

Abstract. In the old days of cameras, the pictures we got were colorless. In contrast to today's high definition photographs, the emotions conveyed in grayscale photographs almost disappeared. Although Photoshop has a promising performance to bring the gray photo back to a colorful one, it still takes a long time and needs professional knowledge of color understanding. Nowadays, Deep Convolutional Neural Network (CNN) has become very popular, almost fitting any function in the computer vision task. Given a gray image as the lightness layer, the network can predict color layers a and b. Finally, we converted the Lab to RGB. This algorithm only takes about 1 second on GPU, and even the professional cameraman can not justify the difference between the actual and colorized images. Additionally, young people's demand for photos is not only high-quality color but also the pursuit of art, eager to integrate more elements to express their emotions and tastes. We also introduce the neural style transfer algorithm to activate the image, which can paint the image as the users want.

Keywords: Deep Learning, Neural Style Transfer, Image Colorization

1 Introduction

Memories are invisible. We desperately want to have something to record our history and feel the surprise of our growing up. Since the invention of film in 1888, cameras have become widely popular around the world. Cameras share our memories, and with photos, we can easily tell each other our history. When we look back at our grandfather's life, a lot of information gets lost in the gray picture. It's so hard for us to feel emotion of the past because it was gray. We need more color in photos for we humans can perceive three basic colors: red, green, and blue. Using these three basic colors, we can express most of the visible colors. Therefore, in digital cameras developed in 1975, photos are usually made up of these three channels. In terms of the property of RGB theory, each channel will both control the color and lightness. To express one color correctly, we need to balance all values, which makes the task difficult. In 1976, Commission Internationale d'Eclairage (CIE) adopted the color measurement into L*a*b* channel where L* represents lightness, a* represents the color from green to red, and b* represents the color from blue to yellow. [1–3] The measure of LAB color is relatively simple for neural networks because it separates the tasks expressed

by each layer. In the deep learning coloring task, the input of the image is layer L, and the network needs to predict layer AB and convert it into RGB.[4]

Given an RGB image, we can easily change it to grayscale.[5] Because the information complexity of RGB is higher than that of a grayscale image, which indicates that this process is irreversible. Meanwhile, the complexity of this task makes it difficult to use traditional machine learning models to learn prior knowledge of color images. Therefore, we generally use deep convolutional neural networks.[6] It convolved the original matrix through a sliding window and captured the information of the partial position of the image. Moreover, the neural network can express high dimensional functions by using a nonlinear activation layer. To grab the characteristics of shallow layers, He et al. invented a residual network in 2016, which significantly improved the performance of the network.[7]. In the aspect of end-to-end learning, the generative adversarial network has always been the focus of researchers. [8] Based upon those, another important algorithm for image colorization is Autoencoder(AE), which maps the image into high dimension space to find the relationship of latent information. Then use the decoder to restore them to low dimension, that is Lab dimension.[9]

For colorization networks, most of the previous work used residual networks and autoencoders. [10–12] Base on these, Vitoria et al, introduces the GAN architecture. [13] With the fast development of Vision Transformer[14], Kumar et al, utilized them in image colorization.[15]

To make the images artistic, we introduced the neural style transfer algorithm. Given a target-style image, the network can transfer the style of the original image to the target style. Therefore, we can easily express our feeling and taste in our photos. This algorithm also allows the painting styles of past artists like Van Gogh and Monet to be reintegrated into modern images. [16–19]

2 Model Analysis

In the model analysis part, we choose two representative algorithms[12, 18], respectively, to explain the image colorization and image style transfer. In each part, we will introduce the backbone of the network and its feasibility.

2.1 Colorization Network

Network Workflow This network was proposed by Kumar et al. [12] at the 2020 CVPR conference, which is the state-of-the-art model. The backbone of the network is proposed by Zhang et al. [11]. Given a gray image, firstly, it will be detected by Mask-RCNN[20], which will crop the instance. Then, the Autoencoder part will map the gray image into 512 channels' latent space in which the fusion module communicates the information. After that, the decoder will convert the high dimensional data to 2 channels with a*b* layers, concatenated with the gray image, and calculate the loss. Finally, backpropagation updates the network. This network is trained on ImageNet[21].

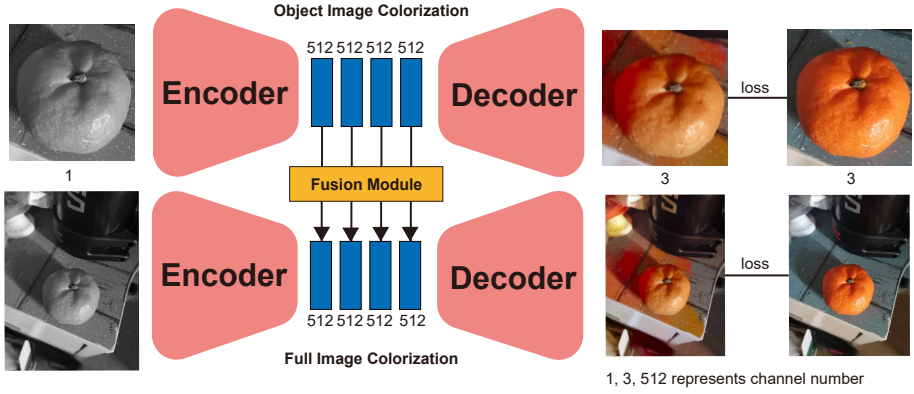

Fig. 1. Model architecture for Instance-aware colorization network[12]

2.2 Neural Style Transfer

Network Workflow Given two images that is the content image and the other one is style image, firstly, the network will put them into a pre-trained VGG19 encoder from which it will output three feature maps. Then the adaptive attention networks learn the feature of the style and content image. After that, they are put into the decoder and output the style transfer image. To calculate the loss, the author uses the same pre-trained parameter VGG19. Finally, backpropagation updates the network. The network is trained on Coco[23].

3 Effect Evaluation

3.1 Colorization Network

This network captures the characteristics of instances well and can learn prior knowledge more competitively. Between the instance and the panorama, the fusion module is designed to make the whole image appear abnormal. However, in the image coloring effect, there are color errors between the instance and the surrounding environment, especially the edge part of the instance. Secondly, the design of the loss function is also relatively simple, although it is easy for the network to learn. Moreover, the effect of the color image needs more constraints, such as image edge, color brightness, noise, etc.

Quantitative/Qualitative Comparison In this section, we compare several of the state-of-the-art models ever made. Through visual comparison, we can see that the network coloring effect of instance-aware is closer to ground truth, especially for the background behind the tennis players and the color of the apple. However, the instance-aware network is not superior in image evaluation index Peak Signal-to-Noise Ratio(PSNR) and Structural Similarity Index

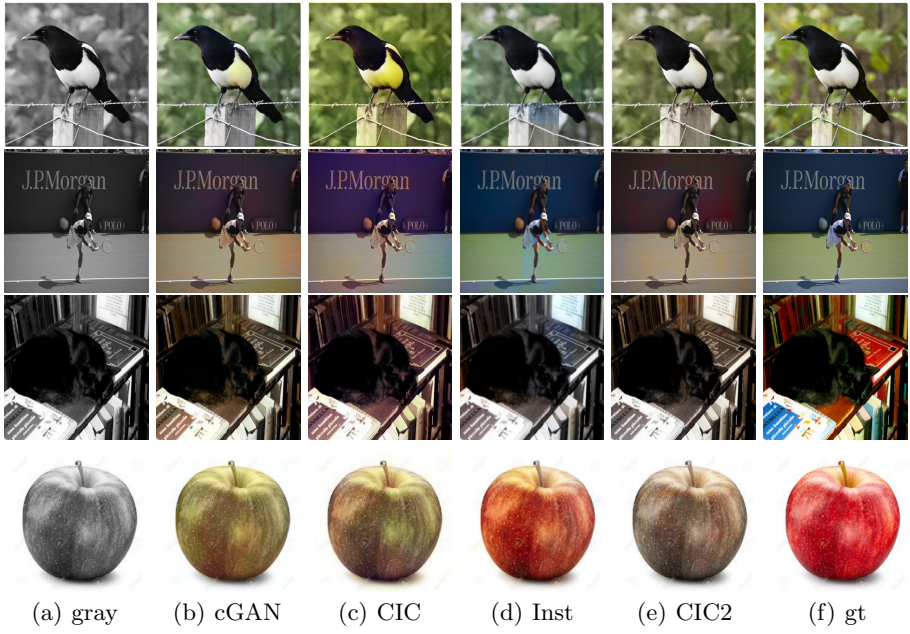

Fig. 2. Visual comparison of several models. From the images, we can see that the instance aware model learn the instance prior well. However, some color in the instance aware are mixed, especially for the edge of the instance. These images are taken from four datasets.[22, 23, 21, 24]

Measure(SSIM).

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2 \tag{1}$$

$$\begin{aligned} PSNR &= 10 \cdot \log_{10}\left(\frac{MAX_I^2}{MSE}\right) \\ &= 20 \cdot \log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right) \\ &= 20 \cdot \log_{10}\left(MAX_I\right) - 10 \cdot \log_{10}(MSE) \end{aligned} \tag{2}$$

where m,n is the weight and weight of the image.

$$\mathrm{SSIM}(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{3}$$

where x,y are two moving window. $c_1 = (k_1 L)^2$, $c_2 = (k_2 L)^2$ $c_2 = (k_2 L)^2$ two variables to stabilize the division with weak denominator; L is the dynamic range of the pixel-values;$k_1 = 0.01$ and $k_2 = 0.03$ $k_2 = 0.03$ by default.

This situation is most likely because the images have color mutations. For

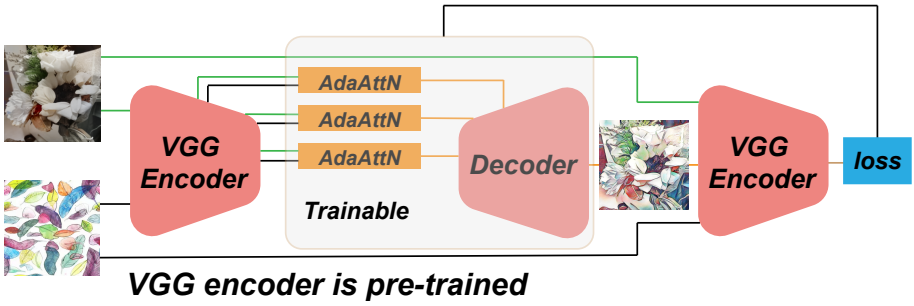

Fig. 3. Model architecture for attention neural style transfer network[18]

example, in the bird picture where the bird's foot is located and in the tennis picture, a distinct noise around the tennis player's body. This shows that the working effect of the fusion module is not very ideal. At the same time, the author also mentioned that there would be a mixture of colors for overlapping objects.

	ChromaGAN[13]		Inst[12]		CIC2[11]		CIC[10]	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
NCD[22]	19.90	0.926	22.45	0.924	21.49	0.908	19.52	0.905
COCO[23]	25.04	0.951	23.52	0.875	23.36	0.870	21.97	0.857
Imagenet1k[21]	25.30	0.950	22.88	0.876	23.48	0.870	21.95	0.857
VOC[24]	25.77	0.955	24.16	0.893	24.49	0.887	22.96	0.876

Table 1. Quantitative Comparison on the image evaluation index of PSNR↑ and SSIM↑. We do this comparison on four datasets, containing over 10k image.

Strength and Weakness This paper cleverly introduces the Mask-RCNN to detect objects in pictures, which solves the poor color performance of instances in the past. In the fusion of object and background, it proposes the fusion module. Comparing the data and pictures, we can see that the effect of this network is more dominant visually. However, it also has some weaknesses. For example, due to the simple design of the loss function(only smooth l1 loss) and the low efficiency of the fusion module, it still can not solve the color anomaly of the object edge. Moreover, the number of parameters of the model is up to 30M. The requirements for hardware equipment are very high.

3.2 Neural Style Transfer

Qualitative Comparison We used several style images to test the effect of the network. The results of the style transfer are very similar to that of the style drawing. At the same time, the rendering does not lose the characteristics of the original object or person. It retains the content of the original picture well while

Fig. 4. Visual results of Neural Style Transfer. The images at the left column are style including color pencil, oil painting, cubism, impressionism, sketch and expressionism. The image at the first row are content. The other image are style transfer image.

doing the conversion of style. However, there are some color inconsistencies in the background of the renderings, which are still a little short of the ideal style transfer, namely the overall harmony of the renderings.

Strength and Weakness This paper introduces an adaptive attention mechanism network using multi-scale feature outputs. Therefore, in the fusion of the original and style images, the network has learned their respective content and style very well. In terms of the network not having panoramic perception, the network is deficient in learning spatial information of images, especially those with a particular color gradient in the background.

4 Conclusions

This paper studies the algorithm of image coloring and style transfer. We analyze the comparison of image coloring models in recent years, focusing on the instance-aware model. For the backbone of the network, they both use Autoencoder and decoder constructs. In the choice of the dimension of intermediate latent space, they use four consecutive 512 channel modules. Since coloring from gray image to colorful image is a process of information gain, the network needs more parameters to remember the original color features of each scene, which is quite different from other image restoration tasks. Because of the extensive network parameters, the image coloring task needs to train on millions of images. In papers before 2020, we also found some shortcomings of the colorization network, such as poor detection of object edges.

Research on style transfer has been prevalent in recent years to satisfy young people's desire to express their taste. This article on adaptive attention mechanisms has excellent effects on the learning and presentation of style maps. However, the authors still need to pay attention to the overall semantics of the style transition diagram. We believe that more loss functions, such as semantic segmentation loss, need to be introduced for future work.

References

1. Papadakis, S.E., Abdul-Malek, S., Kamdem, R.E., Yam, K.L.: A versatile and inexpensive technique for measuring color of foods. Food technology (Chicago) **54**(12) (2000) 48–51
2. Segnini, S., Dejmek, P., Öste, R.: A low cost video technique for colour measurement of potato chips. LWT-Food Science and Technology **32**(4) (1999) 216–222
3. Yam, K.L., Papadakis, S.E.: A simple digital imaging method for measuring and analyzing color of food surfaces. Journal of food engineering **61**(1) (2004) 137–142
4. Leon, K., Mery, D., Pedreschi, F., Leon, J.: Color measurement in l a b units from rgb digital images. Food research international **39**(10) (2006) 1084–1091
5. Kumar, T., Verma, K.: A theory based on conversion of rgb image to gray image. International Journal of Computer Applications **7**(2) (2010) 7–10
6. LeCun, Y., Bengio, Y., et al.: Convolutional networks for images, speech, and time series. The handbook of brain theory and neural networks **3361**(10) (1995) 1995

7. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. (2016) 770–778
8. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. Advances in neural information processing systems **27** (2014)
9. Ng, A., et al.: Sparse autoencoder. CS294A Lecture notes **72**(2011) (2011) 1–19
10. Zhang, R., Isola, P., Efros, A.A.: Colorful image colorization. In: European conference on computer vision, Springer (2016) 649–666
11. Zhang, R., Zhu, J.Y., Isola, P., Geng, X., Lin, A.S., Yu, T., Efros, A.A.: Real-time user-guided image colorization with learned deep priors. arXiv preprint arXiv:1705.02999 (2017)
12. Su, J.W., Chu, H.K., Huang, J.B.: Instance-aware image colorization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. (2020) 7968–7977
13. Vitoria, P., Raad, L., Ballester, C.: Chromagan: Adversarial picture colorization with semantic class distribution. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. (2020) 2445–2454
14. Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., Tang, Y., Xiao, A., Xu, C., Xu, Y., et al.: A survey on visual transformer. arXiv preprint arXiv:2012.12556 (2020)
15. Kumar, M., Weissenborn, D., Kalchbrenner, N.: Colorization transformer. arXiv preprint arXiv:2102.04432 (2021)
16. Huang, X., Belongie, S.: Arbitrary style transfer in real-time with adaptive instance normalization. In: Proceedings of the IEEE International Conference on Computer Vision. (2017) 1501–1510
17. Gatys, L.A., Ecker, A.S., Bethge, M.: A neural algorithm of artistic style. arXiv preprint arXiv:1508.06576 (2015)
18. Liu, S., Lin, T., He, D., Li, F., Wang, M., Li, X., Sun, Z., Li, Q., Ding, E.: Adaattn: Revisit attention mechanism in arbitrary neural style transfer. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. (2021) 6649–6658
19. Hong, K., Jeon, S., Yang, H., Fu, J., Byun, H.: Domain-aware universal style transfer. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. (2021) 14609–14617
20. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision. (2017) 2961–2969
21. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition, Ieee (2009) 248–255
22. Anwar, S., Tahir, M., Li, C., Mian, A., Khan, F.S., Muzaffar, A.W.: Image colorization: A survey and dataset. arXiv preprint arXiv:2008.10774 (2020)
23. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft coco: Common objects in context. In: European conference on computer vision, Springer (2014) 740–755
24. Everingham, M., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A.: The pascal visual object classes (voc) challenge. International journal of computer vision **88**(2) (2010) 303–338