Generating Anime Sketches With C-GAN

Ziyuan Wang 1, a
1School of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijing, 100089, China.
awangziyuan@bupt.edu.cn

Abstract. Generating a decent colored anime drawing from black-and-white sketches is not a simple task in the anime industry. For this problem, a novel deep learning model is introduced based on Conditional Generative Adversarial Networks (C-GAN), which can automatically colorize the anime sketches and output satisfactory pictures. By testing the structure of this model, a typical structure is found to have the best performance.

1. Introduction
The last several decades have seen the increasing popularity of anime, cartoons and online games. As a result, the work of coloring black-and-white line arts has always been demanding. Traditionally, the process is implemented by hand or image editing software such as PaintMan and Photoshop, which is time-consuming and laborious, because the workers need to colorize different parts of a line art one by one.

This study aims to develop an alternative line sketchy colorization method which can automatically color the sketches. This is quite challenging from the perspective of computer vision due to the fact that the line arts sketches usually lack greyscale values or some useful information. Moreover, the scarce of sketch colorization pair data sets increases the difficulty.

To design the model, a basic problem to be addressed is image-to-image translation. Previously, some algorithms have been proposed to solve this problem. For example, Neural Style Transfer [1, 2], which aims at generating an image similar to the input image offered by training datasets, has been widely applied in practice[3, 4].

In computer vision, image-to-image translation is a popular area where there are many popular algorithms. One algorithm called Neural Style, Transfer is aimed at generating an image similar to the input image offered by training datasets. This algorithm is also used in some articles [5, 6].

Conditional Generative Adversarial Networks (C-GAN) [7] is another popular solution for colorization to generates new images based on the inputs to the network. The C-GAN model converges when the discriminator and the generator reach a Nash equilibrium [8, 9]. This model can define the condition on the input and generate a corresponding output, which is the colored sketches in this case.

Another problem that worth attention and effort is image colorization. Previous contributions from some research teams can be found in variety of researchers. Many of them focused on greyscale-based sketch colorization [10, 11], style transfer [12], or different ways of colorization [13,14]. Recently, the author of [12] make use of pipeline and introduced a model that can randomly sampled pixels to generate the color of the black-and-white sketches. Nevertheless, very few of existing results have succeeded in generalizing indistinguishable images from human. Additionally, most of the images produced by them have blurry and missing edges, as well as color errors in the exposed skin of the human body.

Based on the existing C-GAN model, a novel C-GAN scheme is developed in this paper, which
introduces a pre-trained network and produces high-quality pictures out of sketches. The effectiveness of the methods is tested by a whole of challenge sketches with a quantities of styles and contents. Due to the particularity of the target, the performance of the model cannot be distinguished by a traditional indicator such as the loss of the model. The quality of the generated pictures often requires people to identify and evaluate. Therefore, a questionnaire survey was conducted through the Internet. Among the 132 participants, 51.3% believed there was no gap between the computer-generated diagram and the hand-drawn diagram. Hence, it can be concluded that the proposed model performs very well.

2. Main results

2.1 Datasets
The anime sketch colorization dataset used in this paper is from Kaggle [15], the dataset of which contains 17769 paired samples. They are randomly split, 14224 pairs are gained for training and 3545, for testing. Each sketch in the datasets is a 512 x 1024 pixel image, as shown in Figure 1.

3. Model
The training image (black-and-white sketch) and the reference(colorized picture by human) are part of the same sketches. They are trained as the input and output, respectively. Then the greyscale normalizes algorithm was used to make the images from the range of [0, 255] to the range of [-1,1] in order to speed up learning. To maximize the usage of the datasets, data augmentation methods, including random cropping and random mirroring, are also implemented. The batch size is set as 128, and the training is shuffled in each 150 epoch.

The basic network in the model is shown in Figure 2. Sketch-to-color model used is C-GAN, which contains the generator model and the discriminator model, and it is often described it as U-net, which is an encoder to decoder model with connections of the layers.
The theoretical loss of this network is
\[
\min_{G} \max_{D} V(D, G) = \mathbb{E}_{x,y} [\log D(x, y)] + \mathbb{E}_{x,z} [\log (1 - D(G(z)))]
\] (1)
where \( z \) refers to random noise. In equation (1), D tries to maximize itself, while G tries to minimize the equation. Then considered the optional solution in equation (2).
\[
G^* = \arg \min_{G} \max_{D} L(G, D)
\] (2)

3.1 Generator model

Instead of using the typically connected layers of encoder-decoder as some previous papers did, this research uses convolution block and convolution transpose block in order to down sample the input from the dataset, which can help avoids information loss when passed through the connection layers in the
model.

![Encoder Section Diagram](image)

**Figure 4.** The encoder section of the program

![Decoder Section Diagram](image)

**Figure 5.** The decoder section of the model

The loss function of the generator is to find the sigmoid cross entropy loss of the output. Therefore, if the outcome from the generator is put into the discriminator and the output value is almost 1, meaning that it is close to a real image.

### 3.2 Discriminator model

As shown in Figure 6, compared to the generator model the discriminator model is kind of a small model due to the functions it has. Its aim is to determine whether the input image is from the real or the generated image pair. The discriminator model is trained to maximize the accuracy of classification.

![Discriminator Structure](image)

**Figure 6.** Structure of discriminator

The discriminator can output a matrix shaped 30*30*1, all of which within the matrix are given the probability of being real or false from the ground truth of the datasets.

### 4. Visual Results

From Figure 7, it can be seen that the program successfully identified the hair, skin and clothes of the characters, and gave them different colors. The boundary between skin and clothes is also very clear.
Figure 7. Visual result

In contrast to the pictures shown in Figure 8, generated by the author of [6] also used the same dataset. It can be observed that the model in this study is more advanced.

Figure 8. Result from other paper

5. Conclusion

In summary, the baseline model of C-GAN is built and produced high-quality outcomes on generating the images from black-and-white sketches, which takes the best performance among the C-GAN networks.

In the next step, a more high-performance platform will be used to build a more accurate model. The original 512x512 images of the datasets will be utilized to build the model instead of composing them into lower pixel images. In addition, different models for testing, such as RasNet, will also be considered for better results. Other topic in this area including style transfer and using pre-trained model from other papers, will be further explored in future research.

Reference

[1] Johnson, J., Alahi, A., & Fei-Fei, L. (2016, October). Perceptual losses for real-time style transfer and super-resolution. In European conference on computer vision (pp. 694-711). Springer, Cham.

[2] Selim, A., Elgharib, M., & Doyle, L. (2016). Painting style transfer for head portraits using convolutional neural networks. ACM Transactions on Graphics (ToG), 35(4), 1-18.

[3] Huang, H., Wang, H., Luo, W., Ma, L., Jiang, W., Zhu, X., … & Liu, W. (2017). Real-time neural style transfer for videos. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 783-791).

[4] Verma, P., & Smith, J. O. (2018). Neural style transfer for audio spectograms. arXiv preprint arXiv:1801.01589.

[5] Zhang, L., Li, C., Wong, T. T., Ji, Y., & Liu, C. (2018). Two-stage sketch colorization. ACM Transactions on Graphics (TOG), 37(6), 1-14.

[6] Yuanzheng Ci, Xinzhu Ma, Zhihui Wang, Haojie Li, and Zhongxuan Luo. 2018. User-Guided Deep Anime Line Art Colorization with Conditional Adversarial Networks. In Proceedings of the
26th ACM international conference on Multimedia (MM '18). Association for Computing Machinery, New York, NY, USA, 1536–1544. DOI:https://doi.org/10.1145/3240508.3240661

[7] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative Adversarial Nets. In Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2 (NIPS’14). MIT Press, Cambridge, MA, USA, 2672–2680. http://dl.acm.org/citation.cfm?id=2969033.2969125

[8] Kevin Frans. 2017. Outline Colorization through Tandem Adversarial Networks. arXiv preprint arXiv:1704.08834 (2017).

[9] Paulina Hensman and Kiyoharu Aizawa. 2017. cGAN-based Manga Colorization Using a Single Training Image. arXiv preprint arXiv:1706.06918 (2017).

[10] Hati, Y., Jouet, G., Rousseaux, F., & Duhart, C. (2019, December). PaintsTorch: a User-Guided Anime Line Art Colorization Tool with Double Generator Conditional Adversarial Network. In European Conference on Visual Media Production (pp. 1-10).

[11] Nazeri, Kamyar, et al. “Image Colorization with Generative Adversarial Networks.” ArXiv:1803.05400 [Cs], vol. 10945, 2018, pp. 85–94. arXiv.org, doi:10.1007/978-3-319-94544-6_9.

[12] Yuanzheng Ci, Xinzhu Ma, Zhihui Wang, Haojie Li, and Zhongxuan Luo. 2018. User-Guided Deep Anime Line Art Colorization with Conditional Adversarial Networks. CoRR abs/1808.03240 (2018). arXiv:1808.03240 http://arxiv.org/abs/1808.03240

[13] Yu, T., & Yang, R. GANime: Generating Anime and Manga Character Drawings from Sketches with Deep Learning.

[14] Chie Furusawa, Kazuyuki Hiroshiba, Keisuke Ogaki, and Yuri Odagiri. 2017. Comicolorization: semi-automatic manga colorization. In SIGGRAPH Asia 2017 Technical Briefs. ACM, 12.

[15] Anime Sketch Colorization Pair. https://kaggle.com/ktaebum/anime-sketch-colorization-pair Accessed 22 Jan. 2020.