Interactive Anime Sketch Colorization with Style Consistency via a Deep Residual Neural Network

Ru-Ting Ye,Wei-Li Wang,Ju-Chin Chen and Kawuu W. Lin
Computer Science and Information Engineering
National Kaohsiung University of Science and Technology, NKUST
Kaohsiung, Taiwan
1106308114@nkust.edu.tw, F107151105@nkust.edu.tw, jc.chen@nkust.edu.tw, linwc@nkust.edu.tw

Anime line sketch colorization is to fill a variety of colors the anime sketch, to make it colorful and diverse. The coloring problem is not a new research direction in the field of deep learning technology. Because of coloring of the anime sketch does not have fixed color and we can't take texture or shadow as reference, so it is difficult to learn and have a certain standard to determine whether it is correct or not. After generative adversarial networks (GANs) was proposed, some used GANs to do coloring research, achieved some result, but the coloring effect is limited. This study proposes a method use deep residual network, and adding discriminator to network, that expect the color of colored images can consistent with the desired color by the user and can achieve good coloring results.

Keywords：Deep Learning、Colorization

I. INTRODUCTION

In the field of animation drawing, it usually creates grayscale line sketch at first and then coloring, processing and adjustment. This part of the work will take a lot of time and effort. So, wonder if there is a way to make the sketch automatically coloring or coloring by given the color, reducing working hours and labor costs. In recent years, deep learning has flourished, some method has been proposed which automatic coloring or according to the user's reference hint coloring by using neural networks.

The color of anime images is very varied, some things are imaginary no reality thing can be reference, and the line sketch of anime do not necessarily have shadow and texture as input information, it is difficult that how to make network learn correct coloring. After generative adversarial networks (GANs) [4] proposed, several sketch automatic coloring systems by using GANs, like Paintschainer [8], Style2paints [7] have achieved some good results.

This study proposes an interactive colorization method that uses the deep residual network and quotes the concept of GAN. Expect that the color of the coloring image can be consistent with the color hint given by the user.

II. RELATED WORK

In the field of animation drawing, the painter coloring color at various parts in lines sketch, then drawing light, shadow and texture. In paper [2] pointed out that color prediction is a diversity problem, an object can be different color; in the field of anime drawing, the problem of color diversity is more obvious.

The coloring method of anime line sketch in the deep learning field [5][7][8][9][19][22], usually let the network learn how to color by itself, like the skin of character will be colored with the skin color; or color by according user given reference color hint or reference color image, like giving a pink color point at hair of character then hair should be colored pink; giving the image that eyes are red, then the character's eyes of sketch should be colored red.

A. Anime sketch coloring with swish-gated residual u-net[5]

This paper [5] proposed swish-gated residual u-net (SGRU), a method of automatic coloring for anime sketch. This paper proposed the swish layer and swish-gated residual blocks (SGB), can filter feature information effectively and passing it in the network, also speeds up network convergence, and then replaced skip connection and residual block of residual u-net with them. And used the VGG [2] network to do the perceptual network to extract the features of the image to calculate perceptual loss. But in fact, the color of the anime sketch coloring is very diverse as mentioned above, so this paper used the diversity loss in CRNs [6] to let network produce nine different colors images in once. The experiment result of this paper proved SGRU can generates a more vivid and saturated color image than u-net and residual u-net.

B. Paintschainer[8]

Paintschainer [8] is a colorization system, where the generator network is U-net, and because the condition discriminator makes it easy for the generator to pay too much attention to the relationship between sketches and images maps, so that images construction ignored some extent, leads to overfitting, the content of the sketch will be distorted. Therefore, the Paintschainer uses an unconditional discriminator instead of the condition discriminator. Paintschainer can automatic coloring, in addition, users can also add or use color hints, to make coloring results better.

978-1-7281-4666-9/19/$31.00 ©2019 IEEE

C. Style2paints[7]

Style2paints [7] combines residual u-net with the auxiliary classification GAN (ACGAN) [26] to apply the style color to anime sketch , making the network to categorize hair, eyes, skin and clothing, and anime sketch can be colored according to the color of the given reference style image. This paper proposed adding two guide decoders at the entrance and exit of the residual u-net intermediate layer and input the reference style image to VGG16/19 to get the fc1 layer 4096 output and then be as the global style to the middle layer of the residual u-net.

The author of this paper published "Two stage sketch colorization [19]" in 2018, which is an anime sketch coloring system with reference color hint. The coloring is divided into two stages. The first stage network for drafting, will be generated a slightly rough color image; the second stage network for refinement, identify the wrong place of the image and to refine the details, and finally the refined color image of the output is better than first stage network output.

D. Related network

In this study, several neural networks or their concepts are used. For example, the most basic body of the coloring network is u-net [3], and the concept of GAN [4]is adopted, constructed a discriminator in the network.

1) U-net[3]

U-net [3] is proposed by Ronneberger et al. in 2015 and has achieved good results in the ISBI 2015 Cell Tracking Challenge. U-net is a network of encoder-decoder structures, it is often used to process image segmentation, many image segmentation tasks use it as a comparative standard, and also image generation and more. U-net is composed of two similar branches, can extract the features in the image. The branch on the left side of U-net is the encoder, used to encode the input and then reduce the resolution of the feature map and obtain the image features. The branch on the right is the decoder, increases the resolution of the feature map to create a bottom-to-top image output.

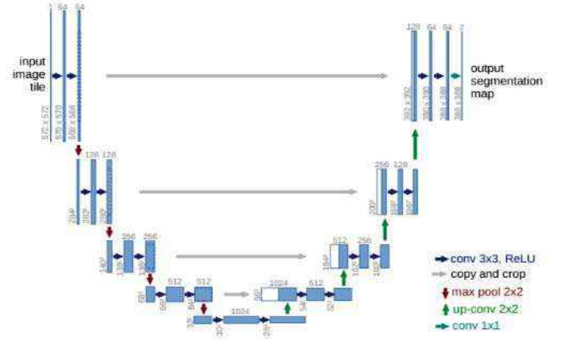

Figure 1.U-net network architecture

Used the residual block proposed in Resnet [10]to improve u-net, when increase the depth of the network can make the network more stable, the network be called residual u-net. In the residual u-net, the information integrity is protected by connecting the input information to the output and paralleling them together, also conveys more information so that the final result can be better.

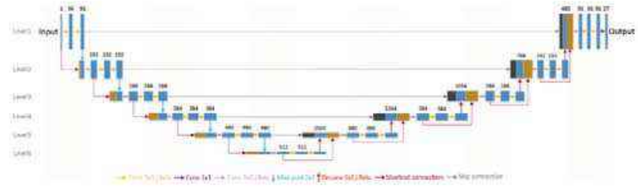

Figure 2. Residual U-net [5]

2) Generative Adversarial Networks[4]

Generative Adversarial Networks (GANs) [4]was proposed by Ian et al. in 2014. The GAN concept is competition, the generator that using the known information plus noise or random noise to generate a fake image and the discriminator that recognizing image is real or fake, both compete with each other, finally expect the generated image of generator can trick the discriminator.

GAN is mainly used in the field of image vision, such as generating numbers, faces and super resolution image, etc., recently, the GANs has been used for automatic coloring or coloring with user reference input [5][7][8][9][19][22], the generator of colorization usually based on u-net or residual u-net, and achieved some good result.

III. SYSTEM STRUCTURE

Given anime line sketch and corresponding color hints, we train a deep residual neural network can let anime line sketch be colored by according color hints. In Section A, introduce the objective of our colorization network. Section B the training dataset processing, and Section C describe architecture of the network.

A. Objective of the network

The result of automatic colorization of anime sketch sometimes does not match the user expected color, and may also color the sketch with a strange, non-vivid color. So, we proposed a colorization network to expect network can generate a color image that is consistent and vivid with the color that the user wants.

B. Pre-processing dataset

In train, each color image in the training dataset is scaled to 256*256, using opencv to extract the corresponding line sketch, and then the real color image T and the line sketch S as input for the colorization network training data.

When loading the training data to the colorization network, in order to train the colorization network to know which color is to be in which position, we used the real color image to randomly generate the color hints. The real color image will be 4 times downsampled, and create a (w/4)*(h/4)*1 random numbers matrix as a hint mask which the value between 0 and 1, when the value in hint mask greater than the threshold value, the mask is set to 1 to indicate that the position has hint, else is set to 0, then hint mask multiply by the real color image of 4 times downsampled, next, concatenate it to hint mask, the output is (w/4)*(h/4)*4 matrix and as the color hints H for training, then also input network.

C. Architecture of the network

Our colorization network is based on SGRU_H[5], the automatic coloring system proposed by Gang Liu et al., its base architecture is a residual u-net, but the direction of the skip connections is replaced by the swish layer them proposed, the swish layer contains the 3*3 convolutional layer and the swish activation function [16], can be considered as the swish-inspired adaptive gating mechanism. The swish layer in the direction of the skip connections can accelerate the convergence and improve the performance of the network. Our colorization network which modified the structure of SGRU_H and be as the generator G.

In order to allow the network learning to color according to the color hint given by the user, we created a discriminator D to learn identify generated image is good or bad, and we added color hint input in 3st level of the generator G. The architecture of colorization generator is shown in Figure 4. Input anime line sketches and the random color hints of the real color images to the generator G, to generate a coloring image.

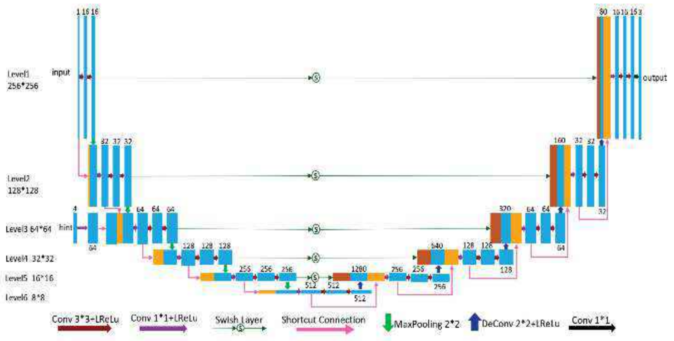

Figure 3. The architecture of colorization generator[5]

Afterwards, in order for our network to learn to generate a good coloring images, we used the pretrained VGG19 as the perceptual network to extract the perceptual feature to calculate perceptual loss with the real color image and coloring image output by colorization generator, also calculates pre-pixel loss between them.

The loss function of colorization generator is represented as:

$$L_g = w_p * L_p + w_h * L_h \qquad (1)$$

L_p is perceptual loss with the real color image and coloring image output by colorization generator, w_p is weight of perceptual loss, L_p is represented as:

$$L_p = \sum_l \lambda_l \| \varphi_l(G(S,H)) - \varphi_l(T) \|_1 \qquad (2)$$

φ_l represent the feature maps from the lth layer in VGG19, λ_l is the hyperparameters of the lth layer in VGG19.

L_h is huber loss with the real color image T and coloring image output G(S, H) by colorization generator, w_h is weight of huber loss.

Furthermore, used the sigmoid cross entropy as loss function to make the discriminator D to learn identify image. The real color image and coloring image output by colorization generator will input to the discriminator D to calculate loss, when backward for discriminator D, the label of sigmoid cross entropy of the real color image will be 1 , of the coloring image will be 0, on the other hand when backward for generator G the label of sigmoid cross entropy of coloring image will be 1, expect that the network learning is colored according to the given color hint. The loss of discriminator D is represented as:

$$L_D = -\log(D(T)) + -\log(1 - D(G(S,H))) \qquad (3)$$

The total loss of colorization generator is represented as:

$$L_G = w_p * L_p + w_h * L_h + w_g * -\log(D(G(S,H))) \qquad (4)$$

w_g is weight of sigmoid cross entropy loss when backward for generator G.

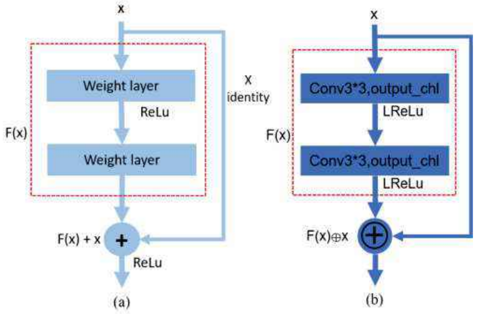

Figure 5. (a) the residual block[10],
(b)Residual block structure in our network

The residual block in colorization generator is minor difference with the residual block proposed by Resnet[10], output of the residual block in colorization generator is F(x) ⊕x, F(x) is concatenated with x, not add.

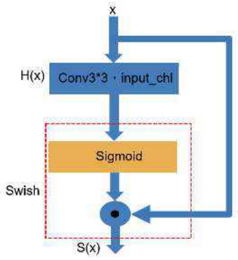

Figure 6. Swish Layer structure[5]

The swish layer[5] contains a 3*3 convolutional layer and the swish activation function [16].

IV. EXPERIMENT

A. *Dataset and setttings*

In experiment, the dataset used was Danbooru 2017, 2018 [1]. Danbooru is a large public source and labeled anime illustrator dataset. In training, 2500 images of Danbooru 2017 are used as training data. In test, 4 different types images are selected in Danbooru 2017 and 2018 as a test data and given color hint that different colors and amount.

During training, the generator G input is D={S, T, H}, where S is a 256*256 anime line sketch, T is the 256*256 the real color anime image corresponding to S, H is the 64*64 color hints that gerented by 4 times downsampled T and randomly selects the color points in T. In training, the learning rate is 0.0001, and the optimizer adopts Adam Optimizer, wherein the hyperparameter β1 is 0.9 and β2 is 0.99, the epoch is 251 and because the GPU memory size, the batch size was 1.

When testing, the generator G input is D= {S, H}, S is the anime line sketch of 256*256 size, H is the color hint map given by the user, the size is 256*256, then generate color image.

B. *Evaluation method*

The general standard evaluation method is like Peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) [15], not suitable estimate the generated quality of the coloring image is good or bad, because anime image different from person to person and diversified, even the same style, the images drawn by different artists will still have different differences, so a reliable method to evaluate the quality of the colored images result is usually judged by human observers' perception experiment, but because this evaluation method experiment is very subjective, the score does not fully represent the quality of the coloring system, because people's sense are not necessarily the same.

In experiment, 10 human observers were invited to evaluate the color images that generated by the study and the Paintschainer generated color images. The questions that the observers evaluated were, 1.Whether the color is the same?,2.Whether the color overflows the range of coloring?, 3.The overall look feels, 4. Is the color of the place without the hint good?, the score of each question is 5~0 points, 5 points means excellent; 0 points means bad.

Table 1. Coloring result rating

Hint	**System**	**Question1**	**Question2**	**Question3**	**Question4**
No	*Ours*		*3.3*	*3.6*	
	Paintschainer		3.3	3.1	
Have	*Ours*	3.3	2.8	3.1	3.2
	Paintschainer	3.25	2.4	2.7	2.5

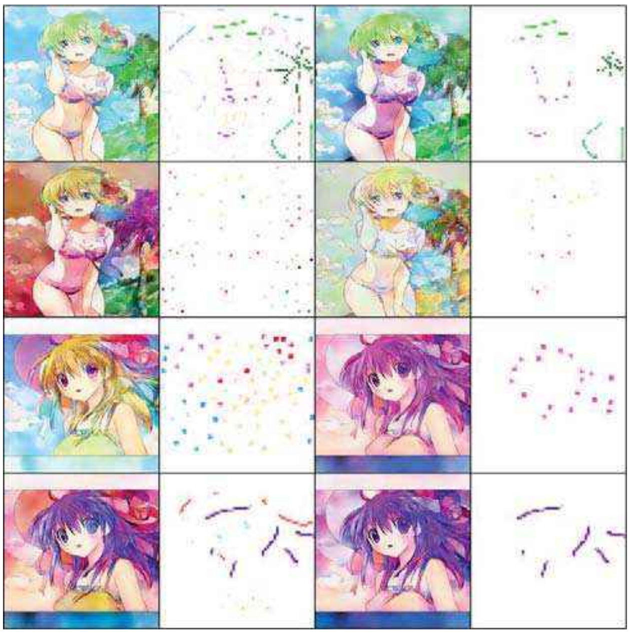

Figure 7. results of hint many or less

C. *Experimental results*

This experiment uses 4 line sketches as the test data, and user draws the color hint corresponding to the line sketch. The case where there is no hint and have hint is tested. The type of the hint is divided into lines and points, and the number of points is also divided. The various test results are shown in the following figures.

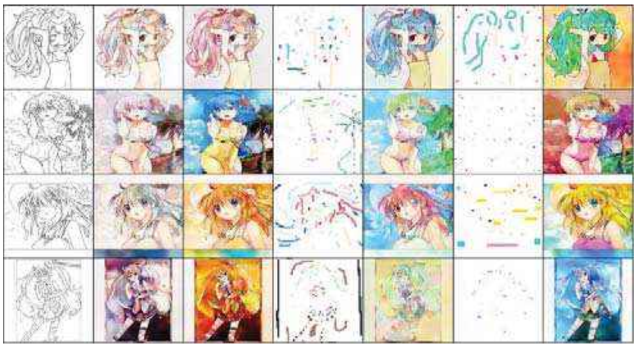

Figure 8. Hint results

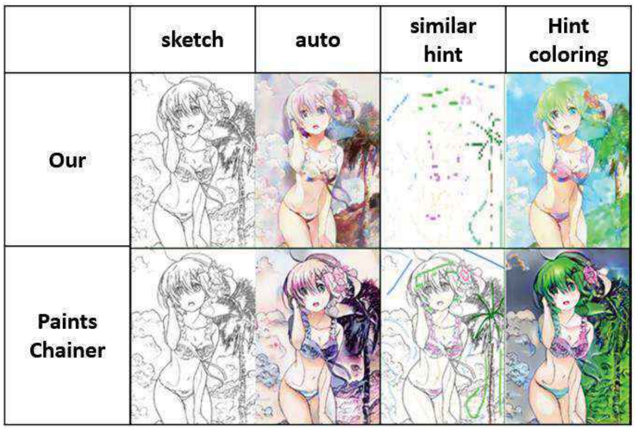

Figure 9. The coloring result comparison with paintschainer

V. CONCLUSION

In the field of animation drawing, it takes a lot of time and effort to complete an image work. This study colorization network modified the architecture of SGRU_H, employed the concept of GAN, adds the discriminator to the system network. And add a color hint to colorization network, expect the network can generate coloring image according the color hints given by the user, and color will be consistency with the color hint. This study used various loss functions, different parameter values to detect what effect does it have on the results of generated color image. Overall, this study generated a color image that was consistency or similar to the color hint given by the user, however, the quality of color required to be improved in the future as well as the better consistency with the color hint given by the user.

REFERENCES

[1] Anonymous, The Danbooru Community, Branwen, G., Gokaslan, Danbooru2018: A Large-Scale Crowdsourced and Tagged Anime Illustration Dataset. https://www.gwern.net/Danbooru2018

[2] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *International Conference on Learning Representations,* 2015.

[3] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” *International Conference on Medical Image Computing and Computer-Assisted Intervention*, p. 234–241, 2015.

[4] IJ. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative Adversarial networks,” *Conference and Workshop on Neural Information Processing Systems,* p. 2672–2680, 2014.

[5] G. Liu, X. Chen, and Y. Hu, “Anime Sketch Coloring with Swish-Gated Residual U-Net,” *Computational Intelligence and Intelligent Systems,* p. 190-204, 2019.

[6] Q. Chen, and V. Koltun, ”Photographic image synthesis with cascaded refinement networks, “ *Proceedings of International Conference on Computer Vision*, 2017.

[7] L. Zhang, Y. Ji, X. Lin, “Style transfer for anime sketches with enhanced residual U-net and auxiliary classifier GAN,” *Proceedings of Asian Conference on Pattern Recognition*, 2017.

[8] Taizan Yonetsuji. “Paintschainer,” *github.com/pfnet/Paintschainer*, 2017.

[9] Y. Liu, Z. Qin, T. Wan, and Z. Luo, ”Auto-painter: cartoon image generation from sketch by using conditional Wasserstein generative adversarial networks,” *Neurocomputing 311*, p. 78–87, 2018.

[10] K. He, X. Zhang, S. Ren, and J. Sun, ”Deep residual learning for image recognition,” *Proceedings of 29th IEEE Conference on Computer Vision and Pattern Recognition*, p. 770–778 , 2016.

[11] Z. Zhang, Q. Liu, and Y. Wang, ”Road extraction by deep residual U-net,” *IEEE Geoscience and Remote Sensing Letters*, p. 749–753, 2018.

[12] D.P. Kingma, and J. Ba, ” Adam: a method for stochastic optimization, “ *Proceedings of the 3rd International Conference for Learning Representations*, 2015.

[13] P. Isola, J.Y. Zhu, T. Zhou, and A.A. Efros, ”Image-to-image translation with conditional adversarial networks,“ *IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[14] J.Y. Zhu, T. Park, P. Isola, and A. A. Efros ,“Unpaired image-to-image translation using cycle-consistent adversarial networks,” *IEEE International Conference on Computer Vision*, 2017.

[15] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, ”Image quality assessment: from error visibility to structural similarity,” *IEEE Transactions on Image Processing*, p. 600–612, 2004.

[16] P. Ramachandran, B. Zoph, and Q.V. Le, ”Searching for activation functions,” *CoRR* abs/1710.05941, 2017.

[17] J.L. Ba, J.R. Kiros, and G.E. Hinton, “Layer normalization,” *CoRR* abs/1607.06450, 2016

[18] Z. Cheng, Q. Yang, and B. Sheng, “Deep colorization,” *IEEE International Conference on Computer Vision*,2015.

[19] L. Zhang, C. Li, T.T. Wong, Y. Ji, and C. Liu, “Two-stage sketch colorization,” *ACM Transactions on Graphics (SIGGRAPH Asia 2018 issue)*, p. 261:1-261:14 , 2018.

[20] R. Zhang, P. Isola, and A. A. Efros, “Colorful image colorization,” *European Conference on Computer Vision*, 2016.

[21] E. S. L. Gastal, and M. M. Oliveira. “Domain transform for edge-aware image and video processing,” *ACM SIG International Conference on Computer Graphics and Interactive Techniques*, 2011.

[22] H. Heo, and Y. Hwang, “Automatic Sketch Colorization using DCGAN,” *18th International Conference on Control, Automation and Systems*,2018.

[23] L. Fang,, L. Wang, G. Lu, D. Zhang, and J. Fu, “Hand-drawn grayscale image colorful colorization based on natural image,” *The Visual Computer.* ,2018.

[24] L.A. Gatys, A.S. Ecker, and M. Bethge, “A neural algorithm of artistic style,” *arXiv preprint* arXiv:1508.06576 ,2015.

[25] L.A. Gatys, A.S. Ecker, and M. Bethge, “Image style transfer using convolutional neural networks,” *IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[26] A. Odena, C. Olah, and J. Shlens. “Conditional image synthesis with auxiliary classifier gans,” *arXiv preprin*t arXiv:1610.09585, 2016.

[27] J. Johnson, A. Alahi, and F.F. Li,” Perceptual Losses for Real-Time Style Transfer and Super-Resolution,” arXiv:1603.08155, 2016.

[28] A. Radford, L. Metz, and S. Chintala. “Unsupervised representation learning with deep convolutional generative adversarial networks,” *arXiv preprint* arXiv:1511.06434, 2015.

[29] M. Mirza, and S. Osindero. “Conditional generative adversarial nets,” CoRR, abs/1411.1784, 2014