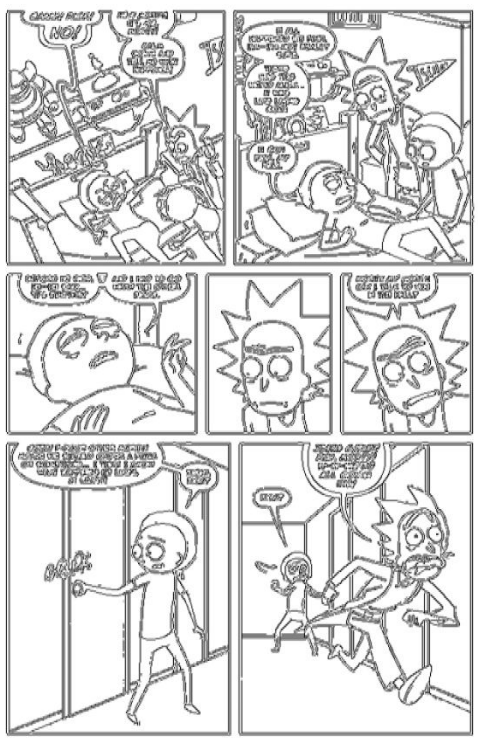
Machine Learning for the Arts
UCSD SPRING 2019
FINAL PROJECT

Manga Colorization



Di Gu
Lihang Gong

Concept:

Human beings possess a great cognitive capability of comprehending black-and-white sketches. For mangaka, after they finished their sketches of the manga, mass production is required to fill all the color on it. Most of them are repetitive works since these mangas follow the same pattern. How to automatically paint the sketch into the color fit their style is a useful application to reduce their workload and can help ordinary people to generate the artist style they like on their own sketch. In this work, we are interested in solving this by employing deep neural networks like generative adversarial networks (GAN) and applying other existing techniques.

Technique:

There exist lots of work on GAN after the first time it is proposed at [1], the general concept of GAN is to minimize the loss between a generator (G) and a discriminator (D) simultaneously by doing the classification of the output image being real or fake.

However, in our case we want output to be conditional on the input. Thus, conditional GAN (cGAN) is studied to fit our purpose. Different from GANs, which learn a mapping from random noise vector z to output image y, conditional GANs learn a mapping from observed image x and random noise vector z. As we found in [2], they proposed method of using cGAN to generate images and it is suitable for image-to-image translation problems such as colorization.

The objective is to get an optimized G, where G tries to minimize this objective against an adversarial D that tries to maximize it. Also, the generator has another task which is to be near the ground truth output. In [2] they used L1 distance rather than L2 as L1 encourages less blurring.

For generator, the 'U-net' architecture with skip connection allows the input and output to share low-level information. For example, in the case of image colorization, the input and output share the location of prominent edges. The U-net structure is as follows:

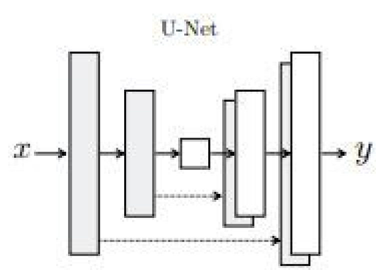


Figure 1. U-net with skip connections

So we use similar network to our proposed solution. In the encoding part, we use U-net 512 as our generator. Each convolutional block is designed with stride equals to 1 and followed by rectified linear unit (ReLU) and batch normalization (BN) layer. The down-sampling in the network is set to be 9. The decoding part is made up of as the symmetric to the encoding part, where the corresponding results of encoding part are concatenated with the upsampling in the decoding part. This operation is like residual neural network (ResNet), where another advantage of using previous parameter is to avoid overfitting. In return, a better performance is presented on our colorization task. Also, the optimizer Adam is used with learning rate equals 0.0002 in our network.

And for discriminator, [2] use a convolutional "PatchGAN" classifier, which only penalizes structure at the scale of image patches. This discriminator tries to classify if each N ×N patch in an image is real or fake.

Process:

After comparison, we decided to use pix2pix model. The first step is to resize train/ test/ validation dataset images to 520*800 or 600*900 since the original sizes are too large. Then concat the monocolor images with colorized images and feed them into pix2pix model. First try we used default U-net 256 as generator, which inputs and outputs 256*256 images. At test phase, The generated results which resized to original size are too blurry to read. And also the results were not so good in color representation because the random crop size is 256, so we lose most of the page areas we guess. So we modified the U-net to 9 layer down-sampling structure which has the ability to inputs and outputs 512*512 images. The results on onepiece pages from grayscale to 3 channel color was quite good. Then we tried Rick & Morty pages and Simpsons pages. They have different performances which will be demonstrated in the next section.

Since black and white manga only have limited grayscale information, most of the time they are pure black or white, we converted above datasets to sketches using edge detection to simulate this case. We tried some post processing methods to make the generated images cleaner, but they didn't work very well, so they were omitted in presentation.

Sample Results:

real_A	fake_B	real_B

Rick & Morty:

Simpsons:

Simpsons to Rick & Morty:

Rick & Morty from sketch:

One piece from sketch:

Reflection:

The results generated from grayscale images achieved our expectation. But when generated from sketches, it was not surprising that many colors were in the wrong places. To improve, the easiest to implement we think is to split the whole pages to blocks to reduce the dimension. Some other solutions choose to manually add 'hints' to get more precise results and cleaner frames. This is a way out, but requires much human works. Maybe the best way is to develop a new architecture which learns a better mapping from sketch to colorized manga pages.

Reference

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In NIPS, 2014.

[2] P. Isola, J.Y. Zhu, T. Zhou, A.A. Efros. Image-to-image translation with conditional adversarial networks. 2016

Code

Github: https://github.com/ucsd-ml-arts/ml-art-final2-dejavu

Results are included.