Advances in Social Science, Education and Humanities Research, volume 664
Proceedings of the 2022 8th International Conference on Humanities and Social Science Research (ICHSSR 2022)

New Evaluation Method of Animation Image Transfer

Dixin Li

Beijing Normal University - Hong Kong Baptist University United International College
**Corresponding author. Email: 2232840862@qq.com*

ABSTRACT
Today, many animation industries need to use landscape migration technology. However, the evaluation metrics after landscape transfer are not particularly accurate, which leads to some failed landscape transfer images being used in the final animation pages. People are reluctant to see that phenomenon. Therefore, it is necessary to find a judgment standard suitable for landscape indicators, and select qualified pictures after landscape migration. Therefore, we will use Cycle-Gan as a method to convert graphics by using the Python programming language and then adjust the weights of psnr and ssin as the latest evaluation indicators in order to mitigate the uncorrected transferred images that appear at the final result. We finally use a new metric as a judging criterion to get high-quality pictures after landscape migration.

Keywords: *Machine Learning, Cycle-Gan, Python, evaluation.*

1. INTRODUCTION

In order to stimulate China's economic growth, the author needs to improve the IP capacity and animation production capabilities to attract overseas investors to invest in Chinese animation production companies.

The current evaluation criteria are too single and one-sided. This will cause some unqualified landscape migration photos to be used in the animations. However, the author has developed an evaluation model indicator that combines multiple aspects (vision, pixels) to make a more accurate judgment on its results. This saves time in selecting qualified photos and improves the efficiency of animation production. The significance of the study is to help people to use technology appropriately, the animation industry may provide better images in the future.

2. QUOTATION OF CHINESE ANIMATION MARKET

At present, as domestic animation becomes more and more exquisite, there are more and more animation lovers.

Up to date, the global animation market is still firmly dominated by the United States and Japan. The global animation output value approximates USD250 billion, and records as much as USD500 billion or so if peripheral products are taken into account.

Also, China's animation industry has been progressing apace over the recent years, with its output value surging from RMB88.2 billion in 2013 to RMB174.7 billion in 2018 and expectedly out numbering RMB200 billion in 2019 and standing at RMB375 billion by 2025.

Figure 1 Output Value and Growth Rate of China's Animation Industry

However, the excellent 2D animation production still needs the hand drawing of the producer frame by frame, which has caused huge production costs and indirectly hindered the development of Chinese animation. The research on style transfer of machine learning may solve this problem, so the author will do research on directly changing natural landscape images into animation style images.

Figure 2 The transfer from the natural landscape images to animation one

The author used Cycle-GAN, and evaluated the Cycle-GAN method, hoping to provide a new possibility for the development of Chinese animation.

3. METHOD SELECTING

3.1. VGG19

VGG is another Convolution Neural Network (CNN) architecture devised in 2014, the 16 layer version is utilized in the loss function for training this model.

Simply put, in VGG, 3 3x3 convolution cores were used instead of 7x7 convolution cores, and 2 3x3 convolution cores were used instead of 5 x 5 convolution cores[1].

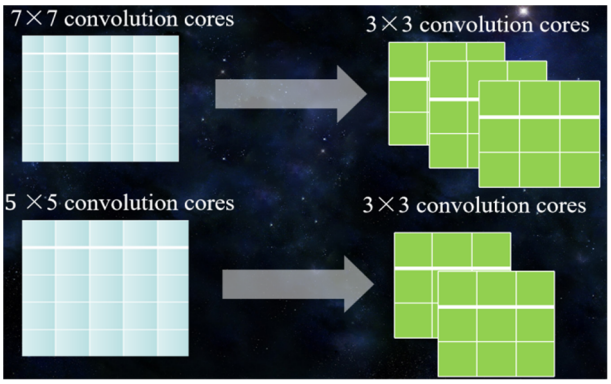

Figure 3 The convolutional cores for VGG19

The main purpose of which was to improve the depth of the network and, to a certain extent, the effect of the neural network under the condition of ensuring the same perceived wildness.

But the author did not use its method, because it wastes more computing resources (contains 3 fully-connected convolutions) and L2 loss of image pixel space.

3.2. Pix2Pix

It is one of the methods of GAN(generative adversarial networks), Image-to-image translation is an important application direction of GAN[2], in fact, based on an input image to get the desired output image process, can be seen as a mapping between the image and the image (mapping, our common image repair, ultra-resolution is actually an example of image-to-image translation).

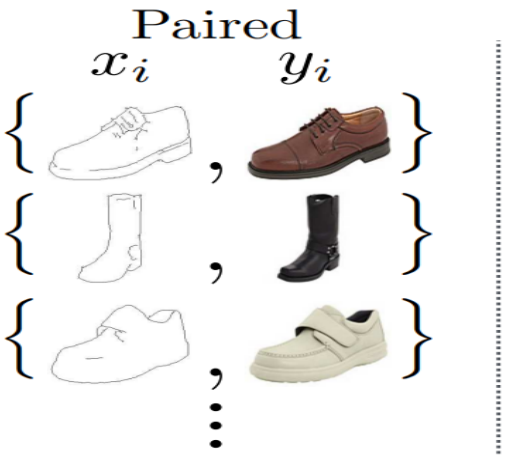

Figure 4 Paired training data

{xi, yi} Ni=1, where the yi that corresponds to each xi is given

However, for the Pix2pix must have a pair picture. If not, it cannot get the correctly transfer pictures.

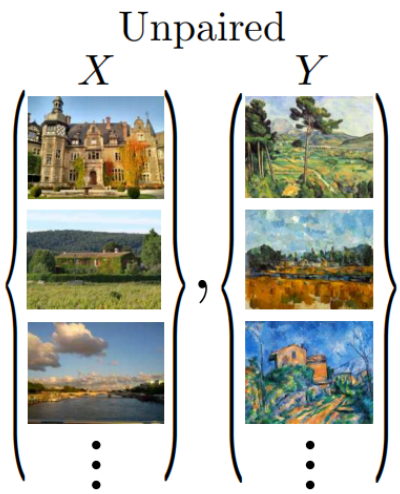

Figure 5 Consisting of a source set

{xi} Ni=1 ∈ X and a target set {yj} Mj=1 ∈ Y, with no information provided as to which xi matches which yj.

So, the author did not choose this method.

3.3. Cycle-GAN

Cycle-GAN is essentially two mirror-symmetrical GANs that form a ring network[3]. The two GAN share two generators and each brings a different generator, i.e. two discriminators and two generators. One one-way GAN two loss, two is a total of four loss.

There are generators G and Discriminators in the network. There are two data fields that are X, Y. G is responsible for taking the data in the X domain and desperately imitating it into real data and hiding it in real data, while D is desperately trying to separate the fake

data from the real data. After the game of the two, G's counterfeiting technology has become more and more powerful, and D's identification technology has become more and more powerful. Until D can no longer tell whether the data is real or G-generated data, this adversarial process reaches a dynamic equilibrium.

The flowchart can explain the process of the Cycle-GAN:

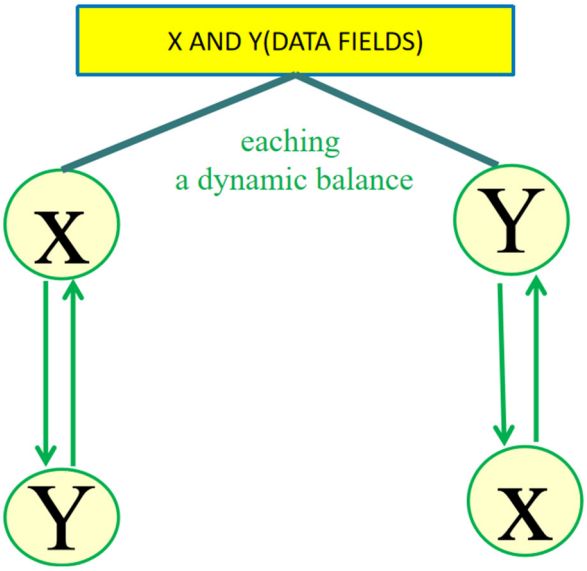

Figure 6 CycleGAN is e ssentially two mirrors symmetrical GANs that form a ring network

And the loss function for the Cycle-GAN[5]

$$\mathcal{L}_{GAN}(G, D_Y, X, Y) = E_{y\sim Pdata(y)}[\log D_Y(y)] + E_{x\sim Pdata(x)}[1 - D_Y(G(x))]$$

The cyclic loss of the dual network is divided into forward cyclic loss and backward cyclic loss.

$$x \rightarrow G(x) \rightarrow F(G(x)) \approx x$$

$$y \rightarrow F(y) \rightarrow G(F(y)) \approx y$$

Using the L1 loss:

$$\mathcal{L}_{cyc}(G, F) = E_{y\sim Pdata(y)}[||G(F(y) - y||_1] + E_{x\sim Pdata(x)}[||F(G(x) - x||_1]$$

So, the total loss:

$$\mathcal{L}(G, F, D_X, D_Y) = \mathcal{L}_{GAN}(G, D_Y, X, Y) + \mathcal{L}_{GAN}(G, D_X, Y, X) + \lambda\mathcal{L}_{cyc}(G, F) \quad (\lambda=10)$$

4. DATA PREPROCESSING

For the data-set was downloaded from Kaggle, but there were many low-resolution images and some images that could not be opened properly in the downloaded data-set, then removed the low resolution and unopenable images, and at the same time, the author unified the image size to 256×256 due to the need of model training. Finally, the author got the dataset.

The train set: It is divided into two parts, one part is 1315 images about the natural landscape and 1180 images about Japanese animation style.

The test set: It only has 1170 images of the natural landscape.

5. CODE

Module.py

There are consists of a discriminator and two generators

The generator divided into three parts:

The first one is encoder, that contains three layers convolution, and use the method of instance_norm, that means the individual channels of a single image are individually Normalized.

In the converter, since the author used an image size of 256*256, then the author used 9 residual blocks. In addition to reducing the disappearance of gradient, other blocks can also adjust the depth of layer by themselves. This can make the network deeper and smoother, and make the training of deep network possible.

The last one is decoder, that contains two layers of convolution.

Model.py

It is for model building, training, testing and saving

It contains four methods for Model building, training, saving and testing.

Ops.py

The main function is Batch_Norm, that is a regularization of the whole batch, which is to prevent gradients between the same batch from canceling each other.

And the second one is instance_norm that is subtracts the mean from the input in the depth direction and divides it by the standard deviation, which increases the number of times for the network can be trained faster.

Main.py

This section is for the training and testing of the model, and the author used the argparse module in order to support command line operations.

Then the functions are executed, either train or test, the author used a computer with a GPU to train the model, and the entire training time of the model was about 30 hours.

The result for test the model: can see the images in the result is one-to-one.

6. MODEL OPTIMIZATION

6.1. Model Optimization——Quantitative Measure

The result obtained by style transfer is a conversion of the original image style, so it is very different from the original image, so when the author evaluates the model effect, the author cannot directly compare the trained image with the original image.

Considering the problem, after reviewing a lot of literature, the author decided to use the feature of Gan model itself - mutual conversion.

First, the author takes advantage of the nature that Gan can be inter-transformed and put the set of anime images tested from natural landscape images as the training set into the model training once again to get a new test set of natural landscape images. By such a transformation, the author can get a one-to-one correspondence with the training set of the original test set of the natural scenery images transformed by the model.

In this way, the author can get the paired images and calculate the similarity.

The first column of pictures is the original picture, the second column of pictures is the picture converted into animation style, and the third column of pictures is the picture converted into natural scenery again. the author uses the first column of pictures and the third column of pictures to evaluate the model.

By reviewing the information, the author chose three methods for quantitative model metrics calculation, namely psnr, ssim and ms-ssim[6]

psnr: Typically, after image compression, the output image will differ to some degree from the original image. To measure the quality of the processed image, the author usually refers to the PSNR value to measure whether particular processing is satisfactory or not.

ssim: Many experimental results have shown that the PSNR score cannot be exactly the same as the visual quality seen by the human eye, and it is possible that those with higher PSNR may look worse than those with lower PSNR instead, so the author introduced ssim. It models distortion as a combination of three different factors: luminance, contrast and structure[7].

ms-ssim: In practical applications, it is usually possible to chunk the images using sliding windows and calculate ssim with average weighting. but since it requires the size of images after handling greater than 160 and the image dimensionality is not so large, the author does not use this method.

To evaluate a model, the author should consider both the difference between the two pictures (psnr) and the effect of the human eye (ssim)[8]. Because the author did not find a formula that can include the two after consulting the data, the author combined the two to get a new evaluation formula.

Considering that after consulting the data, the author found that the better range of psnr is 10-20, The general range of ssim is 0.7-1[9]. Considering the order of magnitude difference, the author final formula:

$$50 \times ssim + psnr$$

The author expect its value to be 45-70. I think the model is acceptable.

6.2. Model Optimization——Choose Parameters

For Selection of model parameters, the author mainly tested epoch. My standards are as follows:

Firstly, the accuracy of the model is 45-70

Epoch should not waste too much space and time

The author calculated the quantitative indexes when the epoch is 150, 200 and 220 respectively[10]. The results are as follows

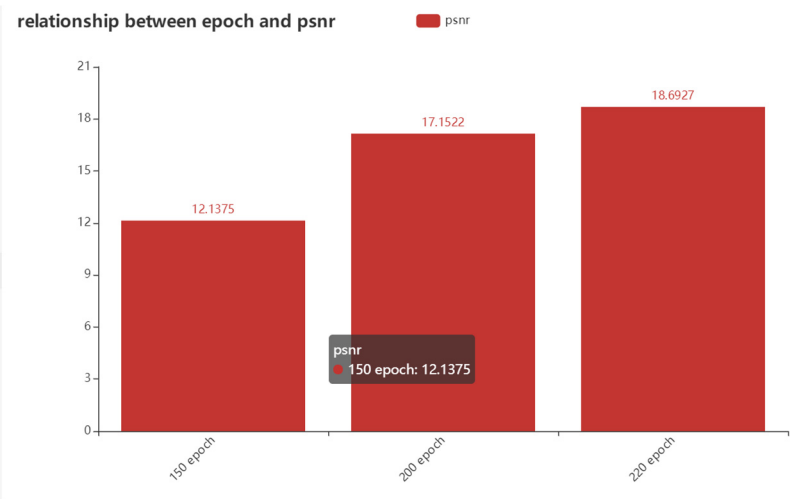

Figure 7 Relationship between epoch and psnr

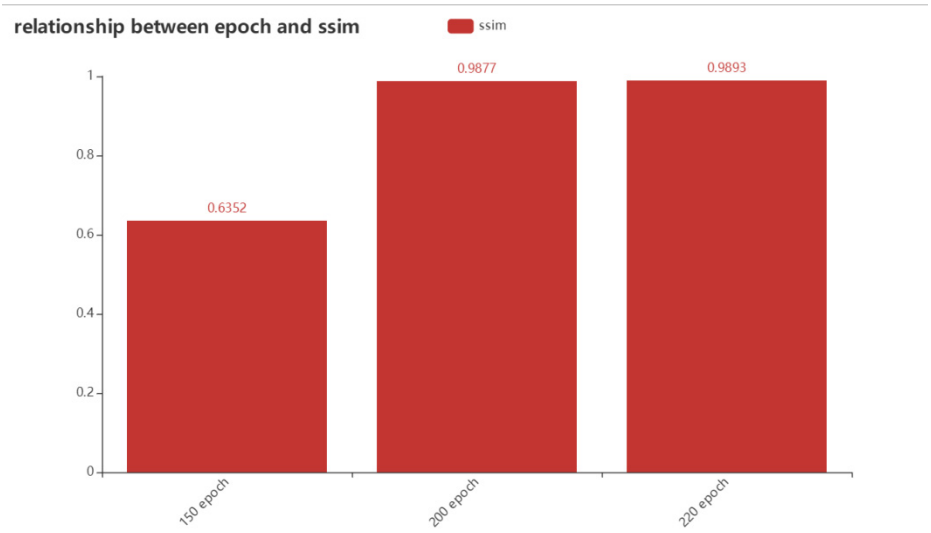

Figure 8 The relationship between epoch and ssim

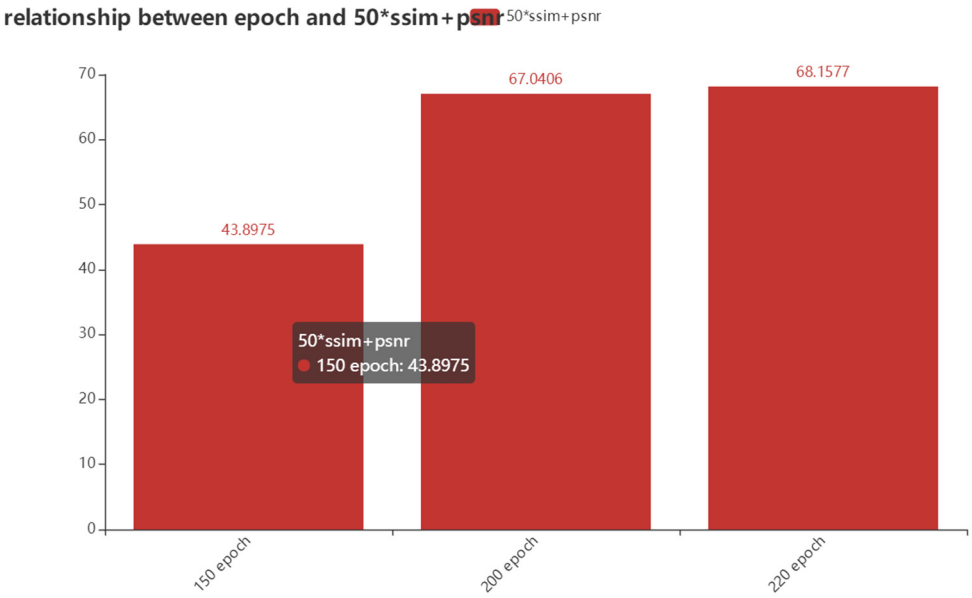

Figure 9 The relationship between epoch and 50*ssim+psnr

From these three figures, the author can see that when epoch = 200, the model training effect is significantly stronger than that when epoch = 150. Although the model training result is slightly stronger than that when epoch = 220, it takes more time and space. Considering comprehensively, the author choose epoch = 200 and get the accuracy[11].

	quantification evaluation
psnr	17.152241
ssim	0.997769
combination(psnr and ssim)	67.040676

Figure 10 The final result of the accuracy

7. CONCLUSION

Firstly, the author compared different kinds of models and selected the better model (Cycle-GAN) for the training. During the training, the author found that the evaluation method of Cycle-GAN could be improved, so the author built a new function by myself to evaluate the result of the Cycle-GAN. The researcher used this new model criterion(50*ssim+psnr) to evaluate the accuracy of landscape transfer can better filter out high-quality landscape transferred pictures. The author hopes that the result of the cycle-GAN may be more accurate in practical use, so it can help my country's animation market become more booming.

REFERENCES

[1] Y.Aytar, L.Castrejon, C.Vondrick,H. Pirsiavash, andA. (2016). Torralba. Cross-modal scenene tworks. arXiv preprint arXiv: 1610.09003.

[2] K.Bousmalis, N.Silberman,D. Dohan, D.Erhan, andD.Krishnan. (2016). Unsupervised pixel-level domain adaptation with generative adversarial networks. arXiv preprint arXiv: 1612.05424.

[3] R.W.Brislin. (1970). Back-translation forcross-culturalre search. Journalofcross-cultural psychology.

[5] E.L.Denton,S. (2015). Deepgen erative image models using alaplacian pyramid of adversarial networks.

[6] J.Donahue. (2016). Adversarial feature learning.

[7] V.Dumoulin&I.Belghazi. (2016). Adversarially learn edinference.

[8] A.A.EfrosandT.K.Leung. (1999). Texture synthesis by non-parametric sampleling. 22(2). 1033–1038.

[9] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. (2014). Courville, and Y. Bengio. Generative adversarial nets.

[10] N. Sundaram, T. Brox, and K. Keutzer.(2010). Dense point trajectories by gpu-accelerated large displacement optical flow. jectories by gpu-accelerated large displacement optical. 438-451.

[11] M. Mathieu, C. Couprie, and Y. LeCun. (2016). Deep multi scale video prediction beyondmean square error.