The Implement of Cartoon GAN

1102-CSC0032 Artificial Neural Network Project2

LIU,TZU-HUNG
Department of Computer Science and Information Engineering
National Taiwan Normal University
Taipei, Taiwan
40747031s@gapps.ntnu.edu.tw

***Abstract*—In this assignment, I will implement the cartoon GAN,one of Generative Adversarial Network, propose by the paper "CartoonGAN: Generative Adversarial Networks for Photo Cartoonization" [1],witch can convert a realistic scenes to a cartoon or animation style.**

***Index Terms*—Artificial intelligence,Neural networks**

I. Introduction

In many cases, animators draw pictures based on realistic scenes.If we can convert a realistic photo into a cartoon or animation style by using a program,then hand it over to the animators for modification.It will greatly reduce their workload.So it is a valuable but also challenging task in computer vision and computer graphics.With the rapid development of artificial intelligence, the GAN proposed by Ian J. Goodfellow in 2014 [2] seems to be able to be used very well for this task, and in this GAN-themed assignment, I will also use CartoonGAN,the extend of GAN to complete the task and assignment.

II. Related Work

Gan, Generative Adversarial Network, was originally a machine learning method for unsupervised learning, but it turned out to have the great effect in other fields,too.Yann Le Cun called it as the most interesting idea of machine learning in the past 20 years. The main structure are two Neural network.One is Generative Network whitch produce the results we need, and then it will be verified by the Discriminating Network. They fought with each other to continuously evolve themselves,and made them be well trained.

III. Cartoon GAN

The structure of GAN is mainly composed of two CNNs. One is Generator G, which is used to convert real picture P into cartoon picture C, and the other is Discriminator D, which is used to distinguish whether it is a real cartoon image or virtual image generated by G, and uses the loss provided by D to Train G to reach the goal.

A. CartoonGAN architecture

- In the Generator, the input image needs to be mapped to the cartoon manifold.The structure used in the paper began with a two-layer Down-convolution for compression and encoding, and under that are 8 residual blocks for extracting features and transforming.Finally, the output cartoon style images are reconstructed by two up-convolution blocks Generator architecture.
- In Discriminator, it is responsible for judging whether the image is real or not. Due to the low task requirements, a patch-level network is used. After the flat layer, there are only two chicken nuggets for compression and encoding.

B. Loss function

This model uses two new losses created by the authors, Adversarial loss and Content loss.

- Adversarial loss is used in Discriminator, the main reason is that cartoon images have two big features: clear edges and smooth shadows, but if only the real cartoon images and generated cartoon images are discriminated.It will deceive the Discriminator as long as the shadow of the generated image is correct because of the clear edges take up a small proportion of the image. So,we will input the real cartoon image and its blurred edges version into the Discriminator,making it more sensitive to the edges.
- Content loss is used in Generator because cartoonized pictures also need to retain the content of real pictures. Therefore, pre-trained VGG will be used to extract the feature maps of real images and generate cartoon pictures, and calculate L1 loss between the feature map to compare and ensure the content of generated cartoon pictures

C. Initialization phase

In Initialization, it is very likely that the training is extremely inefficient at the beginning because GAN is randomly initialized. So, the author only used Content loss to train G at the beginning. We can see the results of the initial training from the above figure, which allowed Gan to skip the chaotic phase at first.

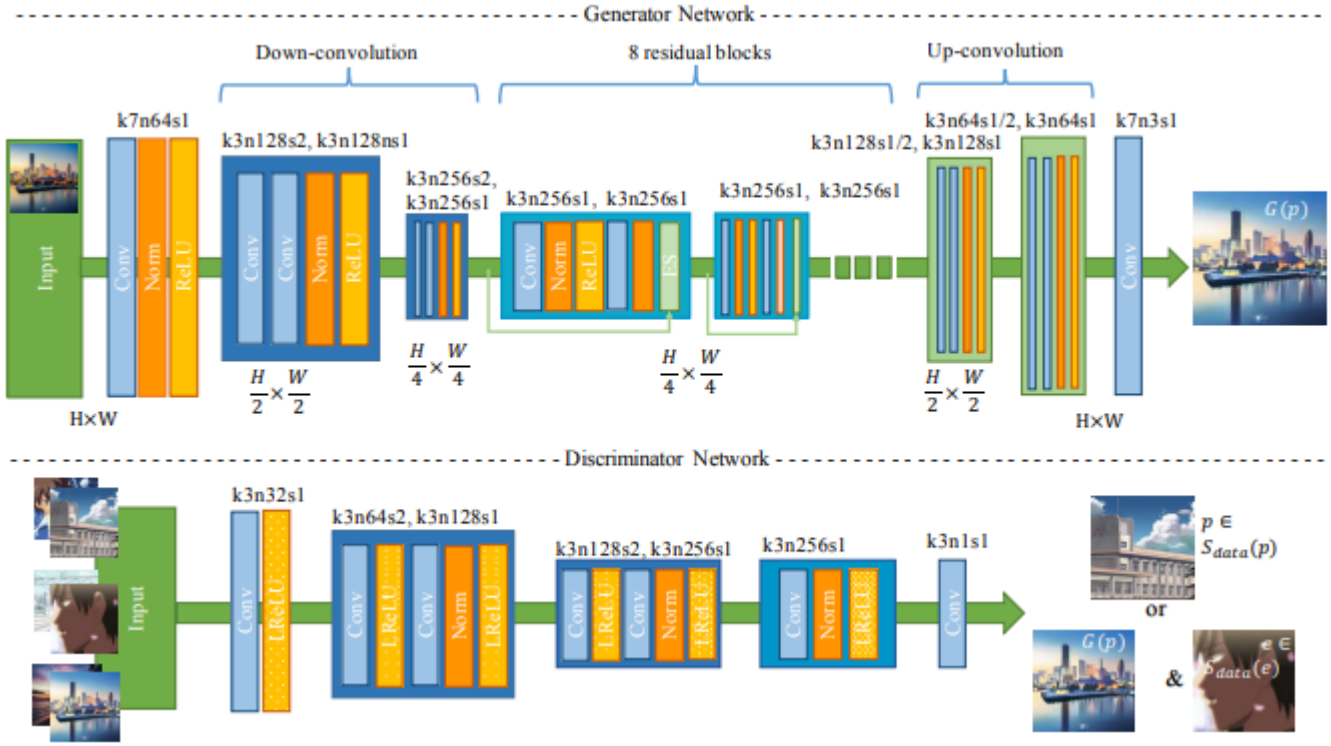

Fig. 1. CartoonGAN architecture

IV. Experiments

A. Different input

Since the inputs of Discriminator in this method areH not image pairs,real images and generated animations,but real animations and generated animations.If a single style of animation is used for training, the results will also tend to be that style, so that this model can also be specialized for different animators, such as hayao style animation version or shinkai style animation version shown as below.

(a) Hayao style

(b) Shinkai style

Fig. 2. different style of animation

B. Comparison with state of the art

Fig3 are comparison between NST, CycleGAN and CartoonGAN

C. Roles of components in loss function

Fig4 are the evaluation with initialization and different loss using ablation experiment.

V. Conclusion

This paper mainly proposes different training methods and two new losses as a breakthrough in generating animated pictures by realistic scenes. Although there are also advanced models based on cartoon gan, such as AnimeGan [3], AnimeGanV2, AnimeGanV3, and so on. I think cartoon Gan is challenging enough for me,a DL beginner, especially when it comes to the process of gan training, which requires constant adjustment of parameters.Fortunately, there are many resources and implementations shared on the Internet, which helped me to complete this assignment and gained me a lot of knowledge.

References

[1] Chen, Y., Lai, Y.K., Liu, Y.J.: CartoonGAN: generative adversarial networks for photo cartoonization. In: Proceedings 31st Meeting of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT,United States, pp. 9465–9474 (2018)

[2] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, Generative Adversarial Networks,arXiv:1406.2661 ,Jun 2014.

[3] Jie Chen, Gang Liu and Xin Chen , "AnimeGAN: A Novel Lightweight GAN for Photo Animation" ISICA 2019: Artificial Intelligence Algorithms and Applications pp 242–256.

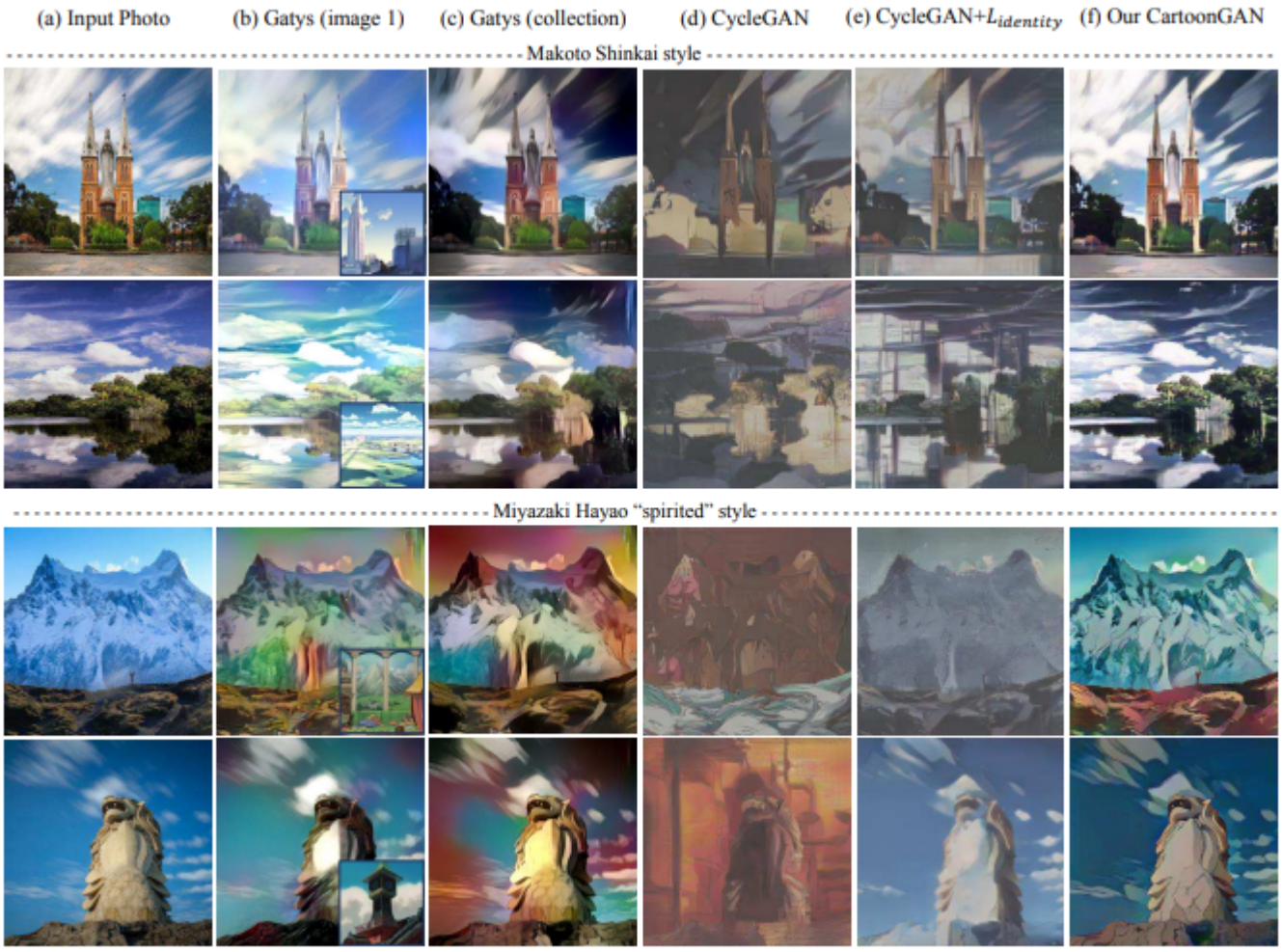

Fig. 3. Results of NST,CycleGAN and CartoonGAN

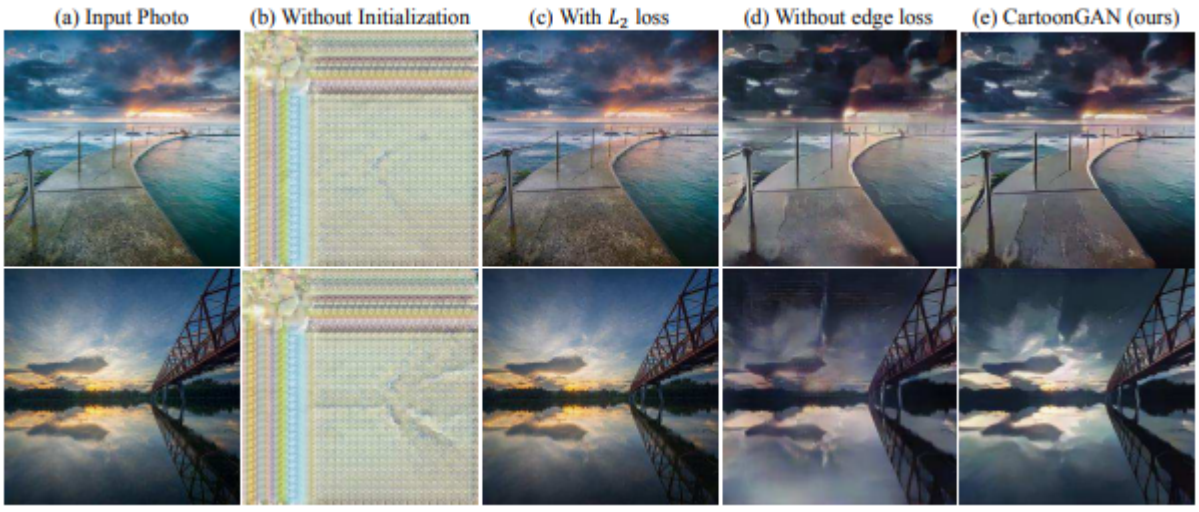

Fig. 4. choice from initialization and different loss