Traditional Chinese Painting Creation using CycleGAN

Yanxuan Xu Zihao Chai Chunzhi Guo
School of Electronic Information and Electrical Engineering of Shanghai Jiao Tong University

Abstract

Unsupervised learning with generative adversarial networks (GANs) has made great progress to solve image-to-image translation problem where the goal is to the mapping relationship between input images and output images by using a training set of paired images. However, depending on the task complexity and impracticality of paired training images, we investigate Cycle-Consistent Adversarial Networks (CycleGAN) [5] to translate images from an original domain X to a target domain Y without paired samples. We construct a new dataset of traditional Chinese paintings in the absence of paired examples, which has never been used before. Based on our dataset, similarly, we introduce an adversarial loss [5] into training mapping $G : X \rightarrow Y$ to translate a realistic image from domain X to a tradition Chinese painting from domain Y, so as to make it difficult to distinguish the distirbution of images G(X) from that of real images Y. And we add a cycle-consistency loss into the inverse translating mapping $F : Y \rightarrow X$ to push $F(G(X)) \approx X$. Overall, Our work realize an end-to-end Chinese painting generation network and achieve a good performance.

1. Introduction

Many problems in image processing and computer graphics can be represented as translation problems, which means translating an input image into a cprresponding output image. Recently, Generative Adversarial Networks (GAN) are widely used for image generation and style translation such as transforming seasons in images from summer to winter and winter to summer, and appearance in images from zebras to horse and horse to winter [5] etc. Moreover, a few categories of GANs are popularly applied for artistic tasks turning real photographs into paintings [1]. However, all of them are concentrated on paintings of modern art style, namely western oil painting, such as Monet's, Van Gogh's and Cezanne's. The similarity between them is the realism of oil paintings, which indicates that it's not very hard to translate a real photo into its corresponding oil painting. While comparing with East Asian art, there're big differences.

Firstly, painting themes are diverse. The subject matter of western oil painting is based on characters. Since the Time of Greece, oil painting has already taken characters as its main theme, not to mention the religious paintings of the Middle Ages, which were also mostly about human beings. While Nature is the subject matter of traditional Chinese painting, where the landscape is always the main painting object. The different themes lead to unequal difficulties in image translation.

Secondly, techniques of expression are different. Oil painting emphasizes perspective, which is the representation of solid objects on a plane. Oil painting strive to resemble the real thing, the streets, houses, furniture, utensils, etc. in the painting which all look like the same as the real thing. However, traditional Chinese painting is in the opposite position. It doesn't contain things with obvious three-dimensional appearance, such as streets, houses, furniture and so on. On the contrary, it emphasizes artistic conception, and usually depicts natural objects, like clouds, mountains, trees and rivers, which does not pay attention to perspective techniques. The distinctive techniques also lead to different methods for transformation.

Because of the differences between traditional Chinese painting and oil painting, we need to deal with them differently. Photographs and oil paintings are both realistic, so it is easier to convert each other. But in traditional Chinese painting, some of the scenes are blurred, so we should add fuzzy transforming manipulation for a better style transformation.

As usual, popular GAN-based generation approaches such as style transfer rely on conditional inputs, e.g. annotations and paired images. But in view of the particularity of Chinese painting, we find it im-

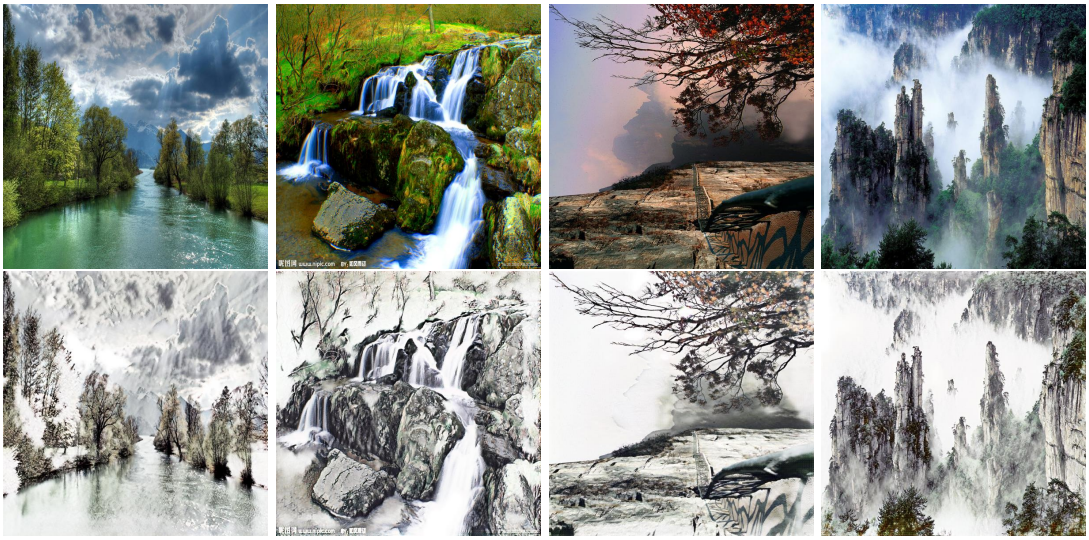

Figure 1. The upper layer contains four arbitrary scene photos and the lower layer contains the corresponding Chinese paintings.

possible to get the paired photographs and Chinese paintings, and it's impractical to label huge amount of images. That means models dependant upon conditional input and paired data are infeasible in this problem. Consequently, we use a generation method which is not based on annotation information and conditional input, which is named as Cycle-Consistend Adversarial Networks, CycleGAN, to complete the style translation between Chinese painting and photographs.

In our model, we preset one given set of images in domain X and another kind of set of images in domain Y. Without paired input-output examples, we train a mapping $G : X \rightarrow Y$ such that the output $\hat{y} = G(x), x \in X$, is indistinguishable from images $y \in Y$ by an adversary trained to classify $\hat{y}$ apart from y, where we introduce an adversarial loss to attain expected effects. For inverse process, we train another mapping $F : Y \rightarrow X$, and introduce a cycle-consistency loss to make the result $F(G(X)) \approx x$. Combining the above two losses, we achieve our full objective for unpaired image-to-image translation between traditional Chinese paintings and western oil paintings.

2. Related Work

Generative Adversarial Networks (GANs) [2] is an important unsupervised learning method in the field of image generation which has achieved significant results. The key point of GANs is the introduction of an adversarial loss that adjusts the generated imgaes to be the real image until it achieves an indistinguishable effect. GANs consists of two adversarial newtorks, one is generator model $\mathcal{G}$ and the other is discriminator network $\mathcal{D}$. Generator $\mathcal{G}$ receives a group of random variables as input, and outputs a corresponding group of generated images. Discriminator $\mathcal{D}$ takes charge of identify images from generator $\mathcal{G}$ and gives a confidence score for each images in a range of zero to one. For generator $\mathcal{G}$, the generated image is expected to be infinitely similar to the real image, whose objective is to fool the discriminator $\mathcal{D}$. However, it is hoped that no matter how similar the generated image from $\mathcal{G}$, $\mathcal{D}$ can always distinguish the fake from the real image. Therefore, GAN is a game process between $\mathcal{G}$ and $\mathcal{D}$, in other word, D and G play a minimax two-player game with the following value function $V(\mathcal{G}, \mathcal{D})$ [2]:

$$\min_{\mathcal{G}} \max_{\mathcal{D}} \mathbb{E}_{x \sim P_{data}}[log\mathcal{D}(x)] + \mathbb{E}_{z \sim P_z}[log(1 - \mathcal{D}(\mathcal{G}(z)))]$$

where x is taken from the real images denoted P_{data}, and z is a hidden vector from probability distribution by the generation $\mathcal{G}$.

Since GAN's commencement, it has been popularly applied as an unsupervised machine learning model for generative tasks, such as image generation and style translation. In the field of style translation, there is two common cases widely studied.

The first one is image-to-image translation, which is often formulated as per-pixel classification or regression [3]. The output space is considered unstructured for this formulation in a sense that each output pixel is supposed conditionally independent from all others given the input image. Instead, conditional GANs [4] introduce some extra information or some kind of auxiliary information, such as class labels of images, then feed these information into both $\mathcal{G}$ and $\mathcal{D}$ as additional

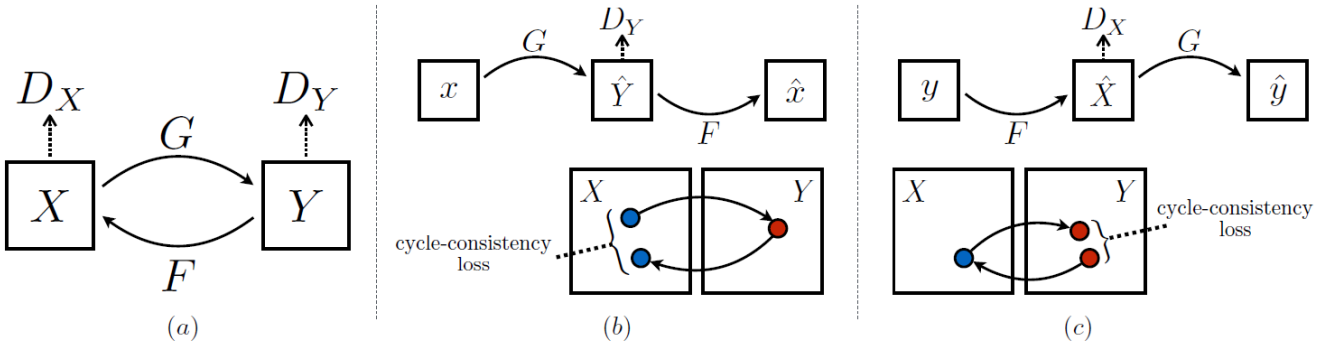

Figure 2. The structure of CycleGAN with cycle-consistency loss.

input layer. Conditional GAN uses a dataset of paired input-output examples to learn a parametric translation function, or a mapping from input to output images.

Conditional GAN (CGAN) usually needs the image class labels as networks input, thus, it is requisite for CGAN to label each training images class, which costs too much manpower and time. However, in the real world, it is too hard to get paired chinese paintings and photographs. Consequently, GAN with labels is not impracticable to our target.

The second is unpaired image-to-image translation. The goal of unpaired translation is to relate two data domains X and Y. CycleGAN does not rely on any predefined similarity function between input and output, instead, it only need unpaired two types of images as input. CycleGAN propose a cycle-consistency loss creatively and combine it with an adversarial loss to train model to get expected effects.

In our algorithm, we use CycleGAN model trained on unpaired data to learn a mapping from real photographs to traditional Chinese painting, and as 'cycle' means, it also learn a inverse mapping from traditional Chinese painting to real photographs. We build a new Chinese painting dataset of 560 images by ourselves for model training.

3. Proposed Method

3.1. Dataset

We find that traditional Chinese painting is mostly based on black and white tones, in other words, our goal is to transform the style of photos into that of Chinese paintings with black and white tones. However, the current open source datasets of Chinese paintings are not appropriate for our goal, where there are mingled with paintings with multiple colors, such as dark yellow, blue and red. And the image quality and quantity are insufficient. For the further research, we build a new dataset of exclusive traditional Chinese paintings with high quality.

Data Collection We scrape 2000 traditional Chinese paintings from Baidu and ivsky [1] by searching key word of Chinese painting artworks.

Data Cleaning We manually take the non-Chinese paintings and Chinese calligraphy, then we filtrate artworks with white and black tones as our final dataset.

Data Pre-processing We adjust the aspect ratio and resize all of them by width and height both to 512 pixels while maintain all feature information.

3.2. Cycle Network Structure

For CGAN, we preserve the features of images in X domain, which are used as constraints or condition to limit the output images in Y domain. So CGAN requires two sets of trained images to match one to one, and the contents of photographs and paintings have to be the same. When we send two completely randomly matched sets of images, CGAN will not learn any connection to retain the image characteristics of X domain. It is unrealistic and meaningless for our machine-creative artworks. However, CycleGAN employs the structure of bidirectional circulation to break the limitation of image corresponding relations, that means, there is no one-to-one correspondence between two groups of images in X domain and Y domain. CycleGAN adopt a wonderful design that instead of the constraints used in CGAN, the generator was restricted from retaining the image characteristics of the original domain by adding a cycle-consistency loss.

Figure 2 [5] shows the structure of CycleGAN. In

[1] Website: https://www.ivsky.com

figure 1(a), CycleGAN contains two mapping generator $G : X \rightarrow Y$ and $F : Y \rightarrow X$, and two corresponding discriminators D_X and D_Y. In figure 1(b), an original image x in X domain is sent into generator G and generate the target image $\hat{Y}$ in Y domain, and in turn, $\hat{Y}$ will be sent into generator F to generate image $\hat{x}$ in X domain. The purpose of above process is to calculate L1 loss with generated image $\hat{x}$ and true image x to align the macro outline of the $\hat{x}$ and x. During this cycle-generation process, generator G and discriminator D_Y can be trained adversarially. But there is an extra reverse generator F on this chain, it is necessary to introduce another discriminator D_X for adversarial training. Similarly, figure (c) shows another way from domain Y to domain X. The two cycle loops construct whole CycleGAN structure.

3.3. Adversarial Loss

CycleGAN apply adversarial losses [2] to both mapping functions. In order to make the generated fake image be almost the same with the real image as possible, adversarial loss function is designed effectively. For the forward mapping G: $X \rightarrow Y$ and its discriminator D_Y, the adversarial loss function [5] is expressed as:

$$\begin{aligned}\mathcal{L}_{GAN}(G, D_Y, X, Y) = \mathbb{E}_{y \sim p_{data}(y)}[log(D_Y(y))] \\ + \mathbb{E}_{x \sim p_{data}(x)}[log(1 - D_Y(y))] \quad (1)\end{aligned}$$

Where G is generator used to realize the process of mapping: $X \rightarrow Y$ and make the G(X) be similar to Y, and discriminator D_Y is used for the distinguishment fake from the real. Therefore, the optimization objective is designed as:

$$\min_{\mathcal{G}} \max_{\mathcal{D}_{\mathcal{Y}}} L_{GAN}(G, D_Y, X, Y)] \quad (2)$$

However, it can't be trained normally with this loss only. The reason is that mapping F can completely map all x to the same picture in Y domain, and the loss is invalidated. Hence, another generator F is introduced to translate images in Y domain into X domain. Similarly, for mapping F: $Y \rightarrow X$ and D_X, the loss function is shown as:

$$\begin{aligned}\mathcal{L}_{GAN}(F, D_X, Y, X) = \mathbb{E}_{x \sim p_{data}(x)}[log(D_X(x))] \\ + \mathbb{E}_{y \sim p_{data}(y)}[log(1 - D_X(x))] \quad (3)\end{aligned}$$

And the objective of this part is:

$$\min_{\mathcal{F}} \max_{\mathcal{D}_{\mathcal{X}}} L_{GAN}(F, D_X, Y, X)] \quad (4)$$

3.4. Cycle-Consistency Loss

It is not enough just to use above two symmetric adversarial losses, because G or F can map the input image to any image in output domain, without relationship with input feature contents. For each image x from X domain, the translation cycle should be able to transform x into y and bring y back to original x as possible, i.e.$x \rightarrow G(x) \rightarrow F(G(x)) \approx x$. CycleGAN add a cycle-consistency loss [5] to incentivize this behavior processing as follows:

$$\begin{aligned}\mathcal{L}_{cyc}(G, F) = \mathbb{E}_{x \sim p_{data}(x)}[\|F(G(x)) - x\|_1] \\ + \mathbb{E}_{y \sim p_{data}(y)}[\|G(F(y)) - y\|_1] \quad (5)\end{aligned}$$

Where the first item on the right is forward cycle consistency and the second item is backward cycle consistency.

3.5. Identity Loss

After above analysis of the principle of CycleGAN, we know that CycleGAN's loss function is composed of adversarial loss and cycle-consistency loss. Exactly, there is another loss function not mentioned, but still used in the implementation of CycleGAN.

That is identity loss [5], which is the icing on the cake. During normal CycleGAN training, the generator G enters x and generates $\hat{y}$. But when calculating the identity loss of generator G, we input y into G and generate $\hat{y}$ and use L1 loss of y and $\hat{y}$ as identity loss of G. Accordingly, identity loss of generator F is also the L1 loss of x and $\hat{X}$. The purpose of employ of identity loss is to maintain the original image tone and avoid excessive translation. The mathematic expression of identity loss is shown as:

$$\begin{aligned}\mathcal{L}_{identity}(G, F) = \mathbb{E}_{y \sim p_{data}(y)}[\|G(y) - y\|_1] \\ + \mathbb{E}_{x \sim p_{data}(x)}[\|F(y) - x\|_1] \quad (6)\end{aligned}$$

3.6. Total Objective

The total loss is:

$$\begin{aligned}\mathcal{L}(G, F, D_X, D_Y) = \mathcal{L}_{GAN}(G, D_Y, X, Y) \\ + \mathcal{L}_{GAN}(F, D_X, Y, X) \\ + \lambda_1 \mathcal{L}_{cyc}(G, F) \\ + \lambda_2 \mathcal{L}_{identity}(G, F)\end{aligned} \quad (7)$$

where λ_1 and λ_2 are hyperparameters that controls the relative importance of the objectives.

The ultimate optimization goal is:

$$G^*, F^* = arg \min_{G,F} \max_{D_X, D_Y} L(G, F, D_X, D_Y) \quad (8)$$

4. Experiments

To optimize our CycleGAN framework, we adjust the hyperparameters and test the robustness and effectiveness of our model. We also scrape 2000 real photographs as anthor dataset. We use GPU V100 to train our CycleGAN model.

4.1. Training Loss

Above all, we have six losses during training, an adversarial loss, a cycle-consistency loss and an identity loss for each domain, where there is two image domain. The following figures from are those training losses. We could find when training to 200 iteration, our CycleGAN network was close to convergence, eight losses all dropped to small values. In the figures, A is behalf of the process from photographs to Chinese paintings and B is behalf of the inverse process, and each process has four losses, two adversarial losses (d_loss and g_loss), one identity loss (idt) and one cycle-consistency loss (cyc).

4.2. Result

After training for 200 iterations, we obtain the final model weights. We apply this model into our experiments. The first part is to input real photographs and desire the corresponding traditional Chinese painting for output. Figure... is from photos to paintings. The next is inverse process, which means the model can translate images from Chinese paintings to real photos, just like figure 4 and 5 show.

We find CycleGAN has an advantageous function for image transalation. No matter from the photo to the traditional Chinese painting or from the traditional Chinese painting to the photo, the performance is very good, and the generated pictures is so similar or even authentic that we can't distinguish them from the human painting of the Chinese painter.

It can be seen from the results that the process from photos to traditional Chinese painting is the process from color pictures to black and white traditional Chinese painting, and the process from traditional Chinese painting to photos is the process of coloring black and white traditional Chinese painting. The colored Chinese painting is full of the atmosphere of fairyland, which is so fascinating.

References

[1] Ahmed Elgammal, Bingchen Liu, Mohamed Elhoseiny, and Marian Mazzone. CAN: Creative Adversarial Networks, Generating "Art" by Learning About Styles and Deviating from Style Norms. arXiv e-prints, page arXiv:1706.07068, June 2017.

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

[3] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-Image Translation with Conditional Adversarial Networks. arXiv e-prints, page arXiv:1611.07004, Nov. 2016.

[4] Mehdi Mirza and Simon Osindero. Conditional Generative Adversarial Nets. arXiv e-prints, page arXiv:1411.1784, Nov. 2014.

[5] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. arXiv e-prints, page arXiv:1703.10593, Mar. 2017.

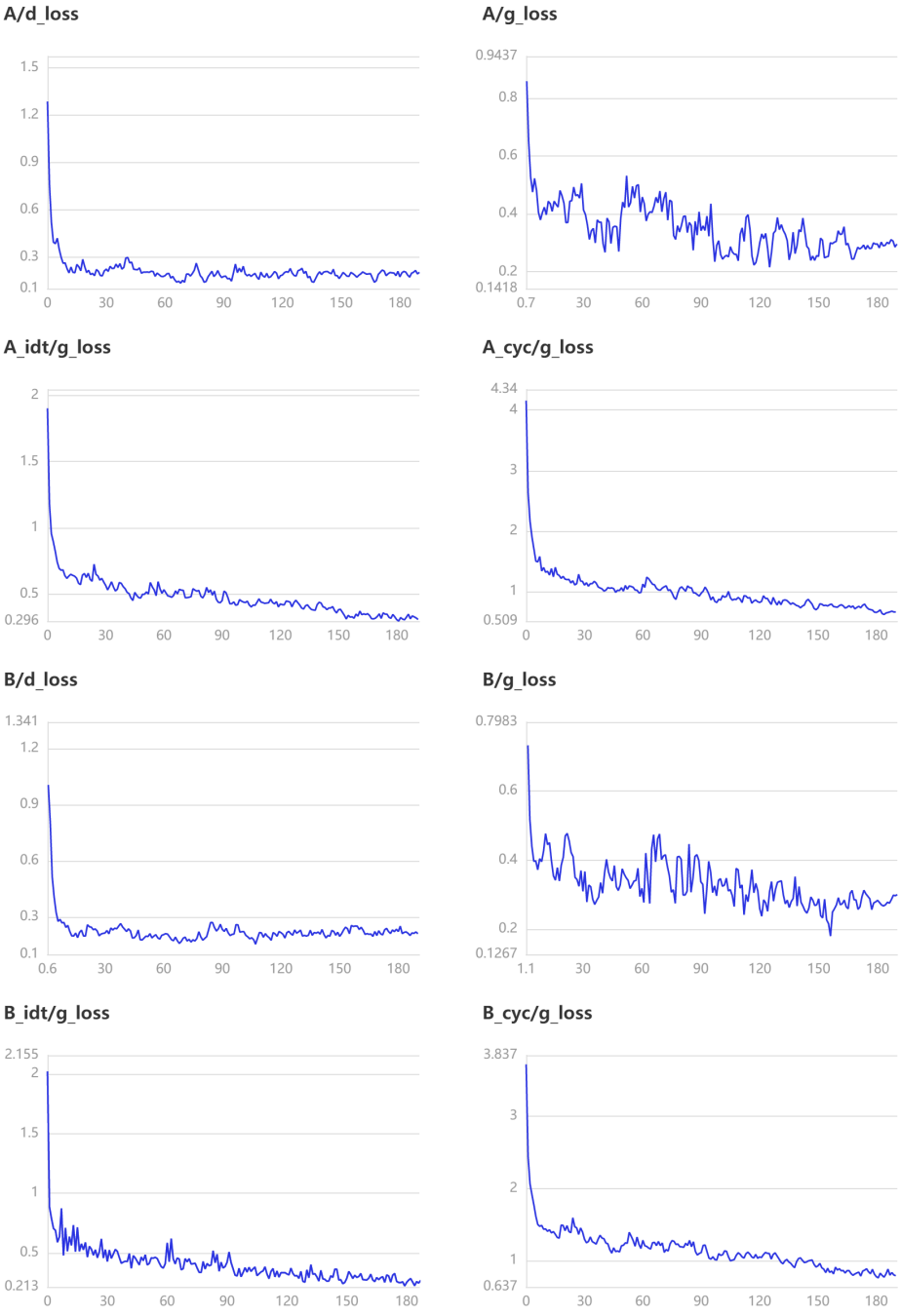

Figure 3. Eight training losses.

Figure 4. The upper layer contains four arbitrary scene photos and the lower layer contains the corresponding Chinese paintings.

Figure 5. The upper layer contains four Chinese paintings and the lower layer contains the corresponding real photos.