Photo Style Transfer Based on Improved Cycle Generative Adversarial Network

WEIGUANG RAN

School of Computing Science
Sir Alwyn Williams Building
University of Glasgow
G12 8QQ

A dissertation presented in part fulfilment of the requirements of the Degree of Master of Science at The University of Glasgow

07th December 2021

Abstract

Image style conversion is a method of converting images with different attribute types (gender, landscape, etc.) (e.g. from cartoon style to pencil style). In recent years, with the rapid development of generative adversarial networks (GAN) in the field of deep learning, the application of GAN in the field of image style conversion has received more and more attention. In particular, CycleGAN has the features of not requiring paired training samples and high accuracy of feature extraction in GAN models. However, these algorithms still have disadvantages such as difficulty in obtaining paired training data and poor conversion effect of generated images. In this paper, an improved cycle consistent generative adversarial network CycleGANplus is proposed on the basis of CycleGAN. The new algorithm focuses on how to optimise recurrent networks and loss functions, attempts to use classification loss instead of recurrent consistency loss and achieves image to image conversion without feature mapping of training data. We present experimental results which demonstrates that our algorithm performs well compared with some classical algorithms of GAN, and the results demonstrate that the improved algorithm has improved in terms of accuracy and fitting time.

Education Use Consent

I hereby give my permission for this project to be shown to other University of Glasgow students and to be distributed in an electronic format. **Please note that you are under no obligation to sign this declaration, but doing so would help future students.**

Name: _____ Signature: _____

Acknowledgements

"Without a lot of personal help and support, this paper is impossible. First of all, I would like to thank my family, especially my dear parents, as an international student, without the help of parents in all aspects, I have no chance and can not continue until graduation, they have done everything they can to greatly help me. Without your continuous help, support and guidance, this will never be possible. Secondly, I would like to think about my supervisor — Professor Alice Miller, she has given me a lot of help in the thesis and professional theory, thank you very much for your help during the thesis, especially in the revision of the dissertation .

At last, I would particularly like to thank my friend Yuli Yang - (A master Student at the University of Glasgow) for his constructive criticism and helpful comments, which greatly helped me write this paper and changed many grammar mistakes."

Contents

Chapter 1

Introduction

1.1 Research Background

I was fortunate enough to visit a free exhibition at the design museum in London a few months ago on the theme of Ai-Da robot art, and I was intrigued by the introduction of the link between art and computing. It has long been assumed that the two fields could not possibly intersect, but both art robots and the recent rise of image style migration technology are breaking down the barriers between the two. Because of this, I wanted to see if I could contribute to the field of image style migration with a focus on deep learning, and with some research I settled on the CycleGAN model, which has become better known in recent years.

Image style migration, also known as image style conversion, is a method of converting the style of an input image to a specified image or images. The algorithm must guarantee the original image structure and convert only the image styles, so that the final output composite image presents a perfect combination of content and style with the input image. Image style migration can be traced back to Image Analogies published by Hertzmann et al [1], which implemented style migration in a single input-output paired training image using a non-parametric texture model. Such methods have been slow to develop over the years due to the limitations of non-parametric texture models that can only extract underlying features rather than higher level abstract features and the processing of different types of images. In recent years, with the development of deep learning-like algorithms, supervised learning-like algorithms based on Convolutional Neural Network (CNN) have been widely used, but the limited size of manually created libraries of paired training images resulted in poor image generation. It was not until Goodfellow et al [2] proposed Generative Adversarial Network (GAN) that the field took on a new life. A typical GAN model consists of two modules: a generator, which generates fake images and makes them indistinguishable from real images, and a discriminator, which learns to distinguish between real and fake images.

With the introduction of GAN, deep neural networks based on GAN ideas have been gradually applied in the field of image style migration [3-5]. A typical example is the ”pix2pix” [6] method, which uses supervised learning for training based on cGANs [7] and combines adversarial loss with L1 loss to propose a unified framework for solving image migration problems. As supervised learning-like algorithms suffer from the common problem of requiring large amounts of pairwise data to be processed manually and trained in advance, an unsupervised learning-based framework

for image style transformation was proposed. rosales et al [8] devised a Bayesian framework for computing a priori Markov random fields from source images and obtaining likelihood terms from multiclass images. unit [9] combined CoGAN [10] with Variational AutoEncoders [11] (VAEs) by combining two generators trained jointly in the correlation domain with shared weights. Cycle-consistent Generative Adversarial Network (CycleGAN) [12], DiscoGAN [13] and DualGAN [14] use cyclic consistent loss to preserve key information between the input and transformed images, enabling image mismatching in the case of data style migration. To improve the quality of the generated images, scholars have attempted to share information about different styles of image features between the input and output [15- 17]. These methods use generative adversarial networks and combine with other studies to make the output approximate the input in a predefined metric space, such as class label space [15], image pixel space [16] and image feature space [17].

CycleGAN, a classical image style migration algorithm for the GAN class using two mirror-symmetric GANs, so that key information of the image is preserved and can be trained without paired data. However, due to lack of a priori information about the target domain, the generated images still do not achieve a more satisfactory The results are still not satisfactory due to the lack of a priori information in the target domain. Meanwhile, CycleGAN serves as a classical image style migration algorithm for GAN-like methods. Using two mirror-symmetric GANs, key information of the image is preserved and can be trained without paired data. However, due to the lack of a priori information about the target domain, the generated images still do not achieve more satisfactory results. In this paper, we improve on CycleGAN by eliminating the cyclic network design and using a single group GAN for training. The inclusion of target domain feature information in the generative network improves the fidelity of the generated images and increases the conversion accuracy. The original loss function is improved by combining adversarial loss, cross-entropy-like loss and reconstruction loss to make the network better adaptable.

1.2 Research Question and Research Objectives

This dissertation investigates how to improve the result of CycleGan algorithm in photo style transfer. The Specific research objectives are as below:

- Can performance be improved by eliminating the ring network and cascading the a priori information from the target and source domains with the corresponding images in the image generation phase.
- Can the loss function be optimised using categorical loss instead of cyclic consistent loss.

1.3 Research Significance

Based on the explosion of the image style conversion field in recent years, many new algorithms have started to be applied in this field, among which CyleGAN, as one of the more applied algorithms in recent years, breaks the corresponding limitation of the training set image data by adding cyclic consistency loss to the GAN network.

This study hopes to confirm through conjecture and empirical research that CycleGAN can also

improve performance and accuracy in both data structure and loss function. First, a conjecture is made to determine how the grid structure and loss function of CycleGan can be improved, after which an empirical study is conducted to verify the conjecture results and a final conclusion is drawn based on the developed baseline.

1.4 Undergraduate Project Work

For my bachelor degree, I studied at Edinburgh Napier University and majored in computing. During this time I was very interested in the machine learning and ai aspects and chose to take some relevant courses as well as use my time outside of school to study on my own. For my final project in my senior year, I chose the field of machine learning as my research course.

The final title of my undergraduate project is: ”Neural Speech Recognition - English to Chinese Speech Recognition System Framework and models Based on Deep Learning”. models Based on Deep Learning”. The main focus of the research is to improve a speech recognition model for English into a Chinese speech recognition model and compare the performance of the improved model with the previous one.

Although the titles of these two projects are not related, there are similarities in the areas of research and experimental methods. Firstly, both projects study the field of deep learning. Although the models and loss functions used are different, the structure of the models and the optimisation process is similar. Secondly, both projects use other models to study and improve the performance of the original model, and they all try to improve the algorithm of the original model and optimise the loss function to get better results.

Through my undergraduate project, I gained an understanding of deep learning content such as feature extraction and loss functions. More importantly, I have gained an initial understanding of how to improve the performance of the model, which has provided me with many ideas and thoughts for my current postgraduate project.

1.5 Structure of the Article

This dissertation focuses on the Image-to-image Translation and analyze how to improve the Cycle-consistent Generative Adversarial Network. This dissertation will be divided into 5 chapters.

The first chapter is the introduction. The second chapter is literature review, in this part there are four key information will be discussed, which are GAN, CycleGAN model, loss function and Pix2Pix model.

The third chapter is the methodology section, describes the choice of database, how to improve the algorithm and how to change the loss function of the model.

In chapter 4 we describe our experiments and present our results. We show how to set up the environment and how to improve the CycleGan model.

In chapter 5 we present our conclusions and evaluate the project, giving suggestions for future work.

Chapter 2

Literature Review

This chapter we describe some related work and give some background material related to this project. Firstly, the process and characteristics of the development of GAN will be noted. Secondly, what is the CycleGan model and the main components will be presented. The third part of the literature review will focus on how to improve the loss function in the model. Finally, Pix2Pix model will be discussed.

2.1 Generative Adversarial Network

Generative Adversarial Network(GAN) is an adversarial network originally proposed by Goodfellow in 2014 [18]. The network framework consists of two parts, a generative model that is called a "forger" and a discriminative model that is called a "cop". The main goal of the generative model is to deceive the discriminative model by constructing fake data, while the main goal of the discriminative model is to detect whether the data comes from real sample data or fake data constructed by forgers. Both models can improve their capabilities through continuous deep learning, i.e. the generative model is trained to generate more real fake data in an attempt to deceive the discriminative model, and the discriminative model is constantly learning how to more accurately identify such fake data from the generative model.

$$\min_G \max_D V(D,G) = E_{X\sim pdata(x)[logD(x)]} + E_{z\sim pz(z)}[log(1-D(G(Z)))] \tag{2.1}$$

The formula maxD denotes maximizing discriminator D, minG denotes minimizing generator G, Pdata is the real image sampled in the sample, D(x) is the probability of the real image; P z means generating a copy of random noise z; G(z) is the image generated by noise z through the generator, D(G(z)) is the probability that this generated image is the real image.

In practice, however, we do not usually train G directly to minimise log(1-D(G(z))), as it saturates early in the learning process, so we usually maximise log(D(G(z)).

Many of the problems with GAN have been raised, including by Goodfellow himself [19], and a number of improvements have been made, resulting in a large number of GAN variants. These kind of problems include the JS scatter problem [20], the Mode Collapse problem [21], etc. These specific issues will not be discussed here.

2.2 Image style migration model - CycleGAN

CycleGAN, or cyclic generative adversarial network, is a model for image style migration. The original image style migration model learns the feature mapping relationship between the input image set and the output image set by training on two sets of one-to-one matching images, thus enabling the migration of features from the input image to the output image, for example, the striped appearance of a zebra in set A to a normal horse in set B. However, the required two sets of one-to-one matching images for training are often difficult to obtain. However, the one-to-one correspondence between the two sets of training images required for training is often difficult to obtain, and CycleGAN breaks the one-to-one correspondence limit by adding a cyclic consistency loss to the GAN network. [12]

2.2.1 Principle of the CycleGan algorithm

CycleGAN, published by Zhu et al [12] in ICCV2017, uses two unidirectional GANs to form a cyclic GAN, which solves the problem of needing paired training data for image style migration. The network structure of the algorithm is shown in Figure 1. Firstly, domain A real image IA is transformed into domain B false image by generator GX , then the image is reconstructed by generator GY to get the reconstructed image IRA, so that the original picture information is preserved. Then the same image as the real CycleGAN is a combination of two one-way GANs, and the training network is supervised by the cyclic consistent loss [6]. The related formulas is:

$$Ł_{cyc}(G,F) = E_{X\sim pdata(x)}[||F(G(X)) - X||_1] + E_{y\sim pdata(y)}[||G(F(y)) - y||_1] \tag{2.2}$$

Whereas in Cycle-GAN in order to be able to not rely on one-to-one image correspondence, one needs to ensure that the learned mapping cannot map different a1, a2,... mapping to the same b, although this b can indeed achieve false results. To prevent this Cycle-GAN introduces the concept of cycle, which simply means mapping a false b back to A, generating a false a1, and determining whether this false a1 is close to the true a1, in such a way as to ensure that the model actually learns a one-to-one mapping rather than a one-to-many. Two transformations G and F are defined in the model to represent the transformation and inverse transformation from the input domain X to the target domain Y, respectively, and Dx and Dy are used to discriminate between the effects of F and G. While using Dx and Dy, two cycle processes are introduced, the first cycle process is to use the true x to generate an estimated y, and then use this estimated y to perform an inverse transformation to generate an estimated x The first cycle process involves using the true x to produce an estimated y and then using this estimated y to produce an estimated x by inversion, at which point the difference between the true x and this estimated x is evaluated, and the same cycle process is introduced for the y to x process.

The network structure of the algorithm is shown in the diagram figure2.1:

The two GANs share two generators, each with a discriminator. Thus the whole loop is trained by forming a joint loss by four loss functions GAN network, such as in equation below:

$$Ł(G,F,D_X,D_Y) = Ł_{GAN}(G,D_Y,X,Y) + Ł_{GAN}(F,D_X,Y,X) + \lambda Ł_{cyc}G,F \tag{2.3}$$

where X,Y represent the two data domains respectively, are the sample data in the two data domains,

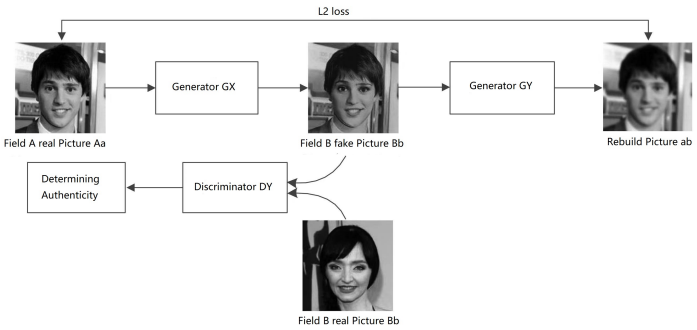

Figure 2.1: Structure of the unidirectional GAN network in CycleGAN

G is the mapping function from X to Y, F is the mapping function from Y to X, DX, DY are discriminators and used to control the the weights of the cyclic consistent loss function.

2.2.2 The Disadvantages of CycleGAN

The CycleGAN algorithm has a deep network structure due to the use of two one-way GANs to form the network, which increases the computational complexity and reduces the training rate, resulting in poor applicability and real-time performance. Moreover, as the generator does not have a priori information about the target transform domain image, the generated image is less effective. In some domains of image transformation, the test images will be blurred, vignetted and poorly transformed.

2.2.3 CycleGan Training Notes

In order to ensure that the trained model is more stable, CycleGan uses two techniques: first, it changes the GAN loss from a non-negative likelihood calculation to a least square loss based on the results of other research [23], and second, it uses a ”memory” technique for GAN training, i.e., it uses the data stored earlier when updating the Discriminator instead of the data just generated by the Generator [24]. Discriminator using the data stored earlier rather than the data just generated by Generator [5]. These two techniques have made the CycleGan model very stable and one of the most used models in the field of image style migration today.

2.3 Loss function in photo style transfer

The loss function is something that is often noticed during model optimisation. Every time the training model gets data, it has to compare its output with the real result, and if it deviates a lot, it has to give feedback on this deviation and find ways to reduce it, i.e. learn more features until it learns how to make the correct judgement on the input data. By using different loss functions (softmax, Mean Squared Error, etc.) and different optimisation methods (Batch Gradient Descent, Stochastic Gradient Descent, Mini-Batch Gradient Descent, etc.) it is possible to achieve very different optimization effects.

2.3.1 Loss functions commonly used for image style migration

Style loss function

Style conversion is the process of converting the semantic content of an image into a different style. The goal of a style conversion model is to generate an output image containing the content of C and the style of S, given a content image © and a style image (S). Here, we briefly discuss one of the simplest implementations of the content-style loss function, which is used to train this style conversion model. A number of variants of the content-style loss function have been used in later studies.

CNNs capture information about content at a higher level, while lower levels are more concerned with individual pixel values. Therefore, we use one or more CNN top layers to compute the activation map for the original content image© and predict the output§. The algorithm is implemented as follows:

$$Loss_{content} = \frac{\Sigma_{i,j}||A_{i,j^l(C)} - A_{i,j^l(P)}||_2^2}{2} \tag{2.4}$$

Texture loss function

Gatys et al (2016) [25]first introduced a style loss component for image style conversion a few years ago. Texture loss is an introduced loss function that is an improvement on perceptual loss and is particularly suitable for capturing the style of an image.Gatys et al found that we can extract a stylistic representation of an image by looking at the spatial correlation of the values within the activation or feature map (from the VGG network). This is achieved by calculating the Gram matrix. The Gram matrix (for layer l of the VGG network) is the inner product of the vectorised feature mappings Fi and Fj (at layer l), which captures the tendency for features to appear simultaneously in different parts of the image.

Content-Style Loss

Content-Style Loss is a loss function in style migration. Style conversion is the process of converting the semantic content of an image to a different style. The goal of the style migration model is that given a content image (C) and a style image (S), an output image containing the content of C and the style of S is generated. In style migration, content loss is similar to perceptual loss in that it compares the similarity between the generated image and the feature map of the content image at a particular layer of the pre-trained network, i.e. the L2 distance between the two feature maps is calculated.

Topological Perceputal Loss

Topological perceptual loss is an extension of perceptual loss, and the perceptual loss function is more applicable when comparing two different images that look similar, such as the same image with only one pixel point shifted or the same image with different resolutions. Mosinska et al. propose that pixel-wise loss is generally used in the field of image segmentation because the essence

of the image segmentation task is still the classification of pixel points, known as dense prediction, so it is natural to use a cross-entropy loss function for classification problems. The pixel-wise loss relies only on local measures and does not take into account features of the topology in the image (such as the shape of connections, or the number of holes), which leads to the problem that traditional classification models can be prone to misclassification for shallow structures. To improve pixel loss, Mosinska introduced a penalty term that is based on the feature maps generated by the pre-trained model (VGG-19 network), similar to perceptual loss.

Wasserstein GAN loss function

Martin Arjovsky has proposed that the goal of a traditional GAN is to minimise the distance between the actual and predicted probability distributions of the real and generated images[26], the so-called Kullback-Leibler (KL) scatter, using a Min-Max loss that is exactly equivalent to the JS scatter, due to a serious problem with the JS scatter: the two distributions do not overlap, the JS scatter is zero, whereas in the early stages of training, there is a very high probability that the JS scatter will be zero. So if D is trained too strongly, the loss will often converge to very small values without a gradient.

2.3.2 Relationship between loss function and gradient descent

According to the definition of the loss function, the smaller the value of the loss function, the better the robustness of the model, often referred to as least squares. Gradient descent, on the other hand, is an iterative method that can be used to solve least-squares problems (both linear and non-linear). The relationship between gradient descent and the loss function is that gradient descent is an optimisation algorithm that solves the loss function.

2.4 Pix2Pix model

Many problems in image processing, computer graphics and computer vision can be thought of as the 'translation' of an input image into a corresponding output image. The term "translation" is often used for translating between languages, for example between Chinese and English. Image translation, however, means conversion between images in a different form. Traditional image translation processes use specific algorithms for specific problems; the essence of these processes is to predict from pixels to pixels based on pixel information and function.

2.4.1 Background of pix2pix

Perhaps CycleGAN shone so brightly that the Pix2Pix model was stolen by another new model, CycleGAN, soon after it was officially adopted. Although pix2pix has not yet had a formal name, this is not because CycleGAN is a better model than pix2pix in every way, and is a direct and complete replacement for the 'pixel migration' Pix2Pix.

In fact, apart from the advantages of CycleGAN, such as "day photo to night", "image colouring"

and ”blueprint to street view”, Pix2Pix also has its own Pix2Pix also has its own unique techniques. For example, a Pix2Pix model trained on a natural landscape photo can render a hand-drawn sketch into a corresponding landscape photo in real time. Pix2Pix's work has also inspired more specific applications, such as SketchyGAN, which specialises in hand-drawn photos, and DeepFaceDrawing, a model for hand-drawn human faces. Also, Pix2Pix – Pix2PixHD – Vid2Vid is a good route for development. Imagine how much time could be saved by game designers by simply building structural models of game characters and scenes, and then having the machine automatically render the characters and scenes in the style they were trained for.

2.4.2 cGAN

Many researchers have previously used GAN to achieve impressive results in many areas, but it soon became apparent that the biggest problem with GAN is that it has to be application-specific, which can waste a lot of time and effort in engineering. To address this problem, researchers proposed the Pix2pix framework, which differs in that pix2pix is not application-specific and it differs from previous work in several architectural choices for generators and discriminators. For the generator, pix2pix uses a ”U-Net” based architecture; for the discriminator, it uses a convolutional ”PatchGAN” classifier, which features a penalty for structure only within the context of image patches.

Pix2pix is based on the idea of cGAN. cGAN does not only input noise to the G-network, but also a condition, and the fake images generated by the G-network are influenced by the specific condition. If an image is taken as a condition, the generated fake images will correspond to the condition images, thus realising an Image-to-Image Translation process.

2.4.3 Advantages and disadvantages of pix2pix

The Pix2pix model has many advantages, such as clever use of the GAN framework to provide a general framework for the ”Image-to-Image translation” class of problems, the use of U-Net to improve detail, and the use of PatchGAN to handle high frequency parts of the image.

The biggest disadvantage of the Pix2pix model is that it requires a one-to-one mapping between axes, and this one-to-one mapping has very limited application, as the results generated are likely to be meaningless when our input data and the training set are very different, which requires our dataset to cover as many types as possible.

Chapter 3

Methodology

3.1 Hypothesis

H1: CycleGan plus improved accuracy of images in AMT perception study results compare with other algorithm include CycleGan.

H2: CycleGan plus improved accuracy of images in FCN perception study results compare with other algorithm include CycleGan.

3.2 Algorithm

To address the problems of CycleGan, we redesigned the structure of the network: the new model eliminates the loop structure. We also replaced the cyclic consistency loss with a classification loss to improve the generated images, resulting in an improved algorithm, named CycleGANplus.

3.2.1 Network structure of CycleGANplus

In this paper, a one-way GAN is redesigned to import domain information using longitudinal cascading and adding cross-entropy classification loss as a constraint. Longitudinal cascading is a method of fusing the original domain image with the target domain information using a concatenation operation. For example, the image and target domain information are both 3-channel tensors, the input to generator A through the longitudinal cascade can result in a 6-channel tensor, achieving This provides a priori knowledge in the training process of the generator and shortens the training time. This provides a priori knowledge during the training of the generator, shortens the training time and improves the accuracy of the generated image. The algorithm structure is shown in Figure 3.1

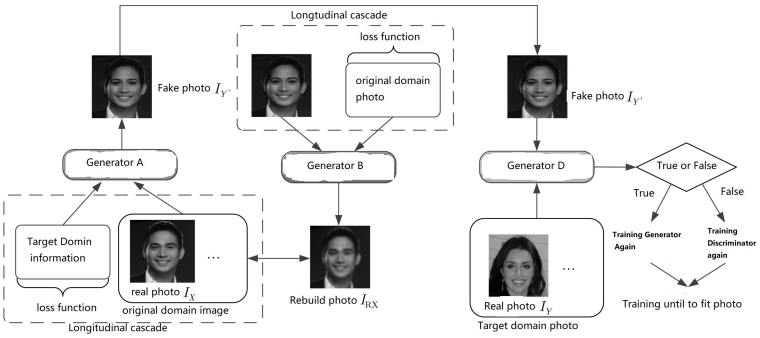

Figure 3.1: the network structure of CycleGANplus

3.2.2 Neural network design for generators and discriminators

As shown in Figure 3.2, the generative network of CycleGANplus consists of three parts: downsampling, residual block and upsampling. The downsampling part consists of 3 layers of convolution with a step size of 2, the residual block contains 6 layers of convolution with residual connections, and the upsampling consists of two layers of deconvolution and one layer of convolution. Instance normalization is used for the generator during the training process. Instance normalization processes the data of a single image and serves to integrate global information, which can be applied to image transformation tasks with good results. PatchGANs [18] are used to implement a discriminative network to determine the authenticity of the generated images, as shown in Figure 3.2 and Figure 3.3.

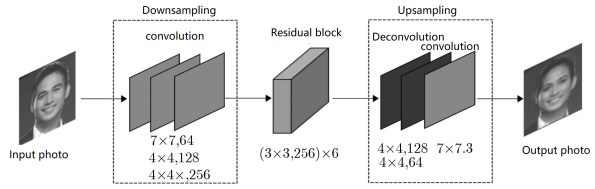

Figure 3.2: Generative networks for CycleGANplus

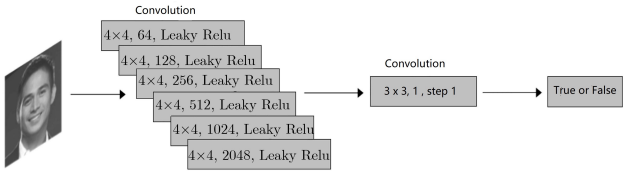

Figure 3.3: Discriminative networks for CycleGANplus

3.2.3 Improvement of the correlation loss function

The total loss function of CycleGANplus consists of a 3-part loss, first adding an adversarial loss function to make the generated image indistinguishable from the real image, as follows:

$$Ł_{adv} = E_x[lnD_{src}(x)] + E_{x,c}[ln(1 - D_{src}(G(x,c)))] \tag{3.1}$$

where G generates an image G(x; c) conditioned on the input image x and the target domain label c, while D attempts to discriminate the authenticity of the image. Dsrc(x) is called the source domain probability distribution given by D. G attempts to minimize the target, and D aims to do the opposite.

Then a category loss is added, which is divided into two categories, training the discriminative network D with real images under a supervised signal in the source domain, and training the generative network G with generated images in the visual domain. trained with a source-domain supervised signal. In contrast, the generative network G is trained using the generated images under a supervised signal in the target domain. The loss function for training D is:

$$Ł_{\frac{r}{cls}} = E_{x,c'}[-lnD_{cls}(c^{'}|x)] \tag{3.2}$$

For a given input image x and target domain label c, the objective is to convert x to an output image y and be correctly classified to the target domain c. An auxiliary classifier is chosen to be added to D. b A classification loss is added when optimizing D and G. The loss function for training G is:

$$Ł_{\frac{f}{cls}} = E_{x,c}[-lnD_{cls}(c|G(x,c))] \tag{3.3}$$

The last part of the loss function is the reconstruction loss, the main purpose of which is to ensure that image content information is preserved during image style migration to enable the use of unpaired data to train the network and improve the results of the generated images. The L-1 parametric function is used as the reconstruction function, with the metric is given by:

$$Ł_{rec} = E_{x,c,c'}[||x - G(G(x,c),c^{'})||_1] \tag{3.4}$$

So the total loss of the generating and discriminating network can be obtained as follows:

$$Ł_G = L_{adv} + \lambda cls Ł_{cls}^{f} + \lambda_{rec} + Ł_{rec} \tag{3.5}$$

$$Ł_D = -L_{adv} + \lambda cls Ł_{cls}^{r} \tag{3.6}$$

Using a redesigned loss function and GAN network for CycleGAN is improved to solve its poor image generation and problem of slow convergence of the training process.

CycleGAN was improved using a redesigned loss function and GAN network to solve the problems of its poor image generation and slow convergence of the training process. The following figure shows the comparison between the training process of this algorithm and the classical CycleGAN. It can be seen that CycleGANplus reaches convergence after 16 h when the loss value stabilises. In contrast, CycleGAN takes about 40 h to reach convergence due to the complex structure of the network.

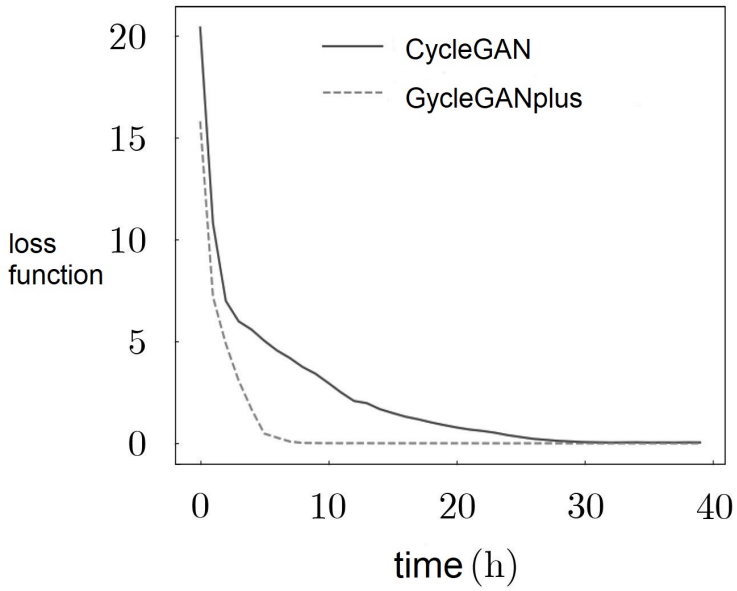

Figure 3.4: Comparison of the training process of CycleGAN and CycleGANplus

3.3 Dataset

The datasets in this project's experiments will be conducted using two open source data, CelebA[27] and Cityscapes[28]. CelebA includes 200,000 178 x 218 pixel images of people's faces. This dataset was divided into two categories, male and female, to achieve style conversion between male and female images. The original images were first cropped to 178×178 pixel by face centre cropping and then scaled down to 128×128 pixel. 2000 images were randomly selected as the test set and all remaining images were used as the training set. The cityscape dataset is an image segmentation dataset containing street scenes from 50 cities with different specific scenarios (different weather, different city sizes), backgrounds and seasons, with a total of 5000 images and 30 classes of objects. It is divided into 2975 training images, 500 validation images and 1525 test images.

Chapter 4

Experimental set up and results

The experiments were conducted using an Ubuntu 16.04 system, and an Nvidia GTX 2060 11 GB single graphics card. The experimental control algorithms were DIAT, which first come up with Li Mu etc. 2016[20], IcGAN, which improved by Perarnau G etc. 2016[21], GoGAN and CycleGAN, CycleGAN and other four classical image style migration algorithms.

4.1 Baseline

The experimental approach uses the AMT perceptual studies (Amazon Mechanical Turk perceptual studies) test and the FCN score (Full-Convolutional Network score) to evaluate the performance of the algorithm. Both of these metrics are used in papers such as CycleGAN [12] and DualGAN [14] to measure the performance of algorithms. One of the AMT perception studies was to evaluate the authenticity of output images by posting a ”discern image authenticity” task on the Amazon labour platform, where pairs of authentic images were made available for task participants to view and click on to select the image they believed to be authentic, and finally aggregating the results for each person to produce experimental data.

4.1.1 Fully Convolutional networks and Architecture

Full Convolutional Networks (FCNs) can train an end-to-end, point-to-point network for semantic segmentation, reaching state-of-the-art. training end-to-end FCNs can be used for pixel-level prediction; the method of training FCNs is supervised training.

In the case of FCN evaluation, the file score.py is the python file used in FCN to test the test set/validation set and output the corresponding four metrics of pixel accuracy, average accuracy, mean IU and frequency weighted intersection ratio (frequency weighted IU). These metrics can be used to compare the training results in image style migration and are widely used to test the results of validation experiments.

4.1.2 Amazon Mechanical Turk perceptual studies

Amazon Mechanical Turk (AMT) is a web service application programming interface (API) that allows developers to integrate human intelligence into remote procedure calls (RPCs), which Amazon makes for tasks that are difficult for computers to perform but can be easily done by "human AI". Tasks. AMT is also a good crowdsourcing platform in the non-commercial sector, and in some studies of human-computer interaction, the results obtained from AMT are often more reliable.

This paper has just started to prepare experiments with AMT on the CelebA and Cityscapes datasets together, and FCN scores are only used to evaluate experiments conducted on the Cityscapes dataset. While AMT perception studies may be the gold standard for assessing image authenticity, it is also particularly important to use automated quality detection methods that do not require human experience. To this end, FCN scores were used to evaluate a style migration task on the Cityscapes dataset, where a fully convolutional network evaluates the algorithm against a standard semantic segmentation model, generating semantic labels for the images associated with the input images. To evaluate the performance of the style migration task, this paper initially intends to evaluate the performance of the style migration task using standard metrics from the Cityscapes dataset, which include per-pixel accuracy, per-class accuracy and IoU classification average.

However, as the experiment progressed, when we were done with the FCN Score, we found that we did not have time to do the AMT, and we found that the AMT platform was not suitable for students to test large datasets, and the platform was not cheap to test if someone wanted to do a large dataset and wanted a little more people to join the experiment, so we finally decided not to use the AMT. however, it was very useful to do this test, and this was one of the things that didn't end up in this project.

4.2 Experimental procedure of CycleGANplus

4.2.1 Improvements to the network structure

The relevant structures have been described earlier in Figure 3.1 and Figure 3.2, and the main vertical cascade is used here. For statistical purposes, the CycleGAN with only the improved network structure is noted as CycleGAN+. The experimental results of the FCN scores are shown in Table 1.

The results in the table show that with the improved network structure, CycleGAN shows a small improvement in all aspects of pixel accuracy, class accuracy and class IoU.

Table 4.1: FCN Store results of CycleGAN+ versus the original algorithm

method	Accuracy per pixel	Accuracy per class	Class IoU
CycleGAN	0.52	0.17	0.11
CyleGAN+	0.60	0.21	0.16

4.2.2 Improve the loss function

After the network structure was changed, the cycle consistent function of the original algorithm was further eliminated and finally the classification loss function was added and the target domain information was introduced to improved loss function's conversion accuracy. The CycleGAN+ with the added improved loss function is noted as CycleGANplus. the FCN score test results are shown in Table 2.

Table 4.2: Comparison of FCN score results for CycleGANplus and CycleGAN+

method	Accuracy per pixel	Accuracy per class	Class IoU
CycleGAN+	0.60	0.21	0.16
CyleGANplus	0.69	0.27	0.23

4.2.3 Comparison of algrithms

In this paper, the improved algorithm CycleGANplus is obtained and achieves higher accuracy in the FCN score compared to the original. In order to verify the effectiveness of this algorithm, the algorithm was compared with other 4 classical algorithms for experiments, as shown in Table 3. It can be found that CycleGANplus achieves a higher conversion accuracy compared to the other 4 algorithms.

Table 4.3: Comparison of FCN score results by algorithm

method	Accuracy per pixel	Accuracy per class	Class IoU
CycleGAN	0.52	0.17	0.11
IcGAN(ref)	0.43	0.11	0.07
CoGAN(ref)	0.40	0.10	0.06
DIAT(ref)	0.68	0.24	0.21
CycleGANplus	0.69	0.27	0.23

4.2.4 Dataset Testing Result

The visualisation of the data is compared with the examples in the figure below. The analysis in figure4.1 and figure4.2 shows that in the CelebA dataset, Figure (a)-Figure (c) shows the conversion of males to females and Figure (d)-Figure (f) shows the conversion of females to males. In the Cityscapes dataset in figure4.3, Figure (a)-Figure (c) shows the conversion of photos to semantic tags, and Figure (d)-Figure (f) shows the conversion of semantic tags to photos. As seen from the results, CycleGANplus in this paper outperforms the original algorithm in terms of facial detail processing, character face shape optimisation, background colour processing, semantic segmentation accuracy and clarity.

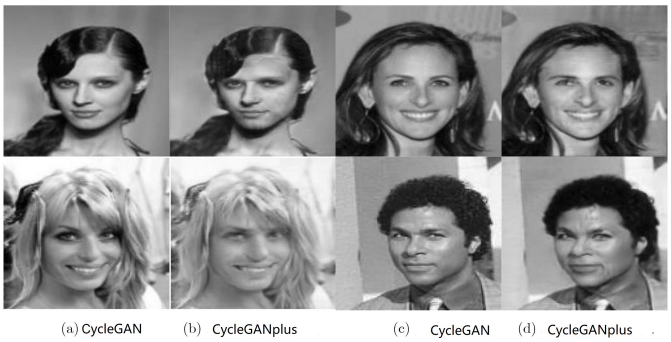

Figure 4.1: Comparison of the training process of CycleGAN and CycleGANplus

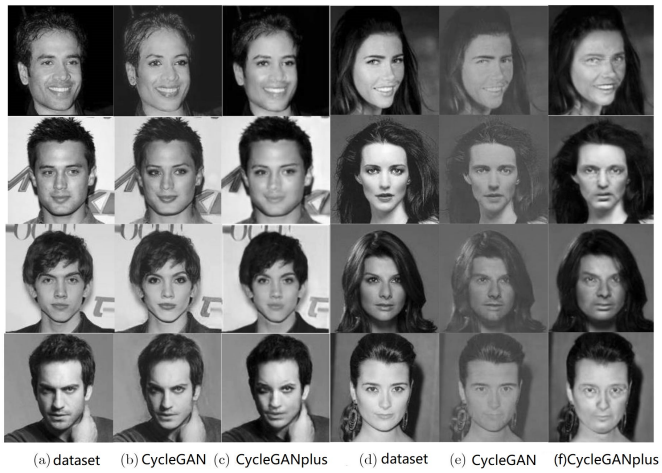

Figure 4.2: Comparison of CycleGANplus with the original algorithm under the CelebA test set

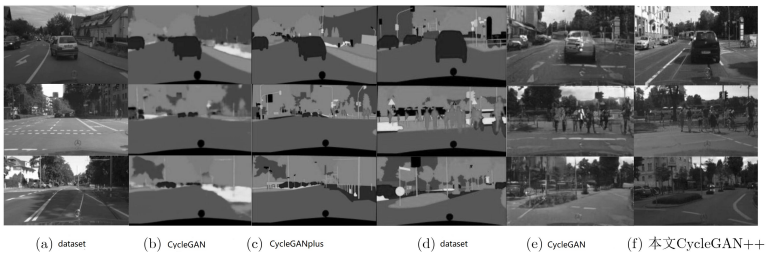

Figure 4.3: Comparison of CycleGANplus with the original algorithm under the Cityscapes test set

4.3 Discussion of the Experiment Result

CycleGAN introduces a cyclic consistency loss function on top of the traditional GAN one-way mapping, which avoids model collapse to a certain extent through bi-directional mapping. The improved CycleGANplus proposed in this paper introduces the same mapping loss and perceptual loss function, which makes the image style migration process more stable and less prone to collapse than the original model, and the convergence speed of CycleGANplus is faster and the stylised image quality is better when comparing the metrics of FCN Score.

Chapter 5

Conclusion

5.1 Overview

Based on CycleGAN, this report proposes an improved algorithm-CycleGANplus that improves the conversion accuracy of the generated images by adding target domain information to the one-way GAN for training. The classification and reconstruction loss functions are optimised so that the predicted values are closer to the true values and the image generation results are more realistic. Compared with classical GAN algorithms such as CycleGAN, CoGAN and DIAT, CycleGANplus achieves a higher level of performance in terms of FCN scores, Figure 3.4 demonstrates that the new model and the original model are faster and more efficient in terms of fitting time to the loss function, Table 4.3 demonstrates through FCN tests that the new model is more accurate in terms of pixel recognition accuracy and class recognition accuracy and other The new model has more satisfactory results compared with other models. However, the current improved model can only achieve one-to-one modal migration, and how to achieve one-to-many modal transformation is a future work to be improved.

5.2 Future Work

As a three month long postgraduate project, there are some areas where this experiment could be improved due to the small amount of time and some problems encountered during the project. The first is the processing of the data set. This experiment used ready-made data directly, but it was not classified and pre-processed; it would have been better if a simple pre-processing of the data was done before the experiment started. Secondly, although the algorithm and loss function have been improved, it does not change the drawback of CycleGAN itself - it can only perform 1-to-1 image conversion, and how to perform 1-to-many or many-to-many image conversion will be one of the main tasks in the future. Thirdly, this experiment was validated using only FCN validation due to time issues, eliminating the AMT validation that was expected to be performed. The single validation was not very good in terms of accuracy and lacked multiple comparisons.

5.3 Personal Statement

I am delighted to have participated and spent time completing this honours project. I found that I learned many valuable points as I completed this paper of mine.

For example, I learned more about the application of deep learning in the area of image style migration and mastered how to discover useful macro information in the data. Mastered how image style migration renders the semantic content of images in different styles. Understood and practised how to increase the speed of convergence and improve the evaluation score value using the perceptual loss function.

In completing my dissertation, I also read a large number of papers and reports, which greatly improved my academic analysis and article reading skills. Some of the ideas in the articles I have applied to this thesis. For example, the idea for this paper was initially based on the CycleGAN paper published by zhu et al. in 2016, and I have been inspired by many other people's experiments and data as the project has progressed, such as using the FCN and AMT platforms to test the experimental results.

Bibliography

[1] HERTZMANN A, JACOBS C E, OLIVER N, et al. Image analogies[C]. The 28th Annual Conference on Computer Graphics and Interactive Techniques, New York, USA, 2001: 327–340. doi: 10.1145/383259.383295.

[2] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]. The 27th International Conference on Neural Information Processing Systems, Montreal, Canada, 2014: 2672–2680.

[3] RADFORD A, METZ L, and CHINTALA S. Unsupervised representation learning with deep convolutional generative adversarial networks[EB/OL]. https://arxiv.org/abs/ 1511.06434, 2015.

[4] ARJOVSKY M, CHINTALA S, and BOTTOU L. Wasserstein GAN[EB/OL]. https://arxiv.org/abs/1701. 07875, 2017.

[5] GULRAJANI I, AHMED F, ARJOVSKY M, et al. Improved training of wasserstein GANs[C]. The 31st International Conference on Neural Information Processing Systems, Red Hook, USA, 2017: 5769–5779.

[6] ISOLA P, ZHU Junyan, ZHOU Tinghui, et al. Image-to image translation with conditional adversarial networks[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 5967–5976. doi: 10.1109/CVPR.2017.632.

[7] MIRZA M and OSINDERO S. Conditional generative adversarial nets[EB/OL]. https://arxiv.org/abs/1411.1784, 2014.

[8] ROSALES R, ACHAN K, and FREY B. Unsupervised image translation[C]. The 9th IEEE International Conference on Computer Vision, Nice, France, 2003: 472–478. doi: 10.1109/ICCV.2003.1238384.

[9] LIU Mingyu, BREUEL T, KAUTZ J, et al. Unsupervised image-to-image translation networks[C]. The 31st Conference on Neural Information Processing Systems, Long Beach, USA, 2017: 700–708.

[10] LIU Mingyu and TUZEL O. Coupled generative adversarial networks[C]. The 30th Conference on Neural Information Processing Systems, Barcelona, Spain, 2016: 469–477.

[11] KINGMA D P and WELLING M. Auto-encoding variational bayes[EB/OL]. https://arxiv.org/abs/1312.6114, 2013.

[12] ZHU Junyan, PARK T, ISOLA P, et al. Unpaired image-to image translation using cycle-consistent adversarial networks[C]. 2017 IEEE International Conference on Computer Vision, Venice, Italy, 2017: 2242–2251. doi: 10.1109/ICCV.2017.244.

[13] KIM T, CHA M, KIM H, et al. Learning to discover cross domain relations with generative adversarial networks[C]. The 34th International Conference on Machine Learning, Sydney, Australia, 2017: 1857–1865.

[14] YI Zili, ZHANG Hao, TAN Ping, et al. DualGAN: Unsupervised dual learning for image-to-image translation[C]. 2017 IEEE International Conference on Computer Vision, Venice, Italy, 2017: 2868–2876. doi: 10.1109/ICCV.2017.310.

[15] BOUSMALIS K, SILBERMAN N, DOHAN D, et al. Unsupervised pixel-level domain adaptation with generative adversarial networks[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 95–104. doi: 10.1109/CVPR.2017.18.

[16] SHRIVASTAVA A, PFISTER T, TUZEL O, et al. Learning from simulated and unsupervised images through adversarial training[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 2242–2251. doi: 10.1109/CVPR.2017.241.

[17] TAIGMAN Y, POLYAK A, and Wolf L. Unsupervised cross-domain image generation[EB/OL]. https:// arxiv.org/abs/1611.02200, 2016.

[18] Goodfellow, Ian, et al. “Generative adversarial nets.” Advances in neural information processing systems. 2014

[19] Goodfellow, Ian. “NIPS 2016 tutorial: Generative adversarial networks.” arXiv preprint arXiv:1701.00160 (2016).

[20] Arjovsky, Martin, Soumith Chintala, and Léon Bottou. “Wasserstein gan.” arXiv preprint arXiv:1701.07875 (2017).

[21] LI Chuan and WAND M. Precomputed real-time texture synthesis with markovian generative adversarial networks[C]. The 14th European Conference on Computer Vision, Amsterdam, The Netherlands, 2016: 702–716. doi: 10.1007/978-3-319-46487-943.

[22] Zhu, Jun-Yan, et al. ”Unpaired image-to-image translation using cycle-consistent adversarial networks.” Proceedings of the IEEE international conference on computer vision. 2017.

[23] Multiclass generative adversarial networks with the l2 loss function. arXiv preprint arXiv:1611.04076, 2016.

[24] A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, and R. Webb. Learning from simulated and unsupervised images through adversarial training. arXiv preprint arXiv:1612.07828, 2016.

[25] Ulyanov, Dmitry, et al. ”Texture networks: Feed-forward synthesis of textures and stylized images.” ICML. Vol. 1. No. 2. 2016.

[26] Gulrajani, Ishaan, et al. ”Improved training of wasserstein gans.” arXiv preprint arXiv:1704.00028 (2017).

[27] LIU Ziwei, LUO Ping, WANG Xiaogang, et al. Deep learning face attributes in the wild[C]. 2015 IEEE International Conference on Computer Vision, Santiago, Chile, 2015: 3730–3738. doi: 10.1109/ICCV.2015.425.

[28] KINGMA D P and BA J. Adam: A method for stochastic optimization[EB/OL]. https://arxiv.org/abs/1412.6980, 2014.

[29] LI Mu, ZUO Wangmeng, and ZHANG D. Deep identity aware transfer of facial attributes[EB/OL]. https:// arxiv.org/abs/1610.05586, 2016.

[30] PERARNAU G, VAN DE WEIJER J, RADUCANU B, et al. Invertible conditional GANs for image editing[EB/OL]. https://arxiv.org/abs/1611.06355, 2016.