

A Brief Review on Cycle Generative Adversarial Networks

Miaomiao Zhu^{a,*}, Shengrong Gong^b, Zhenjiang Qian^b, Lifeng Zhang^a

^aDepartment of Electrical and Electronic Engineering, Kyushu Institute of Technology, Kitakyushu, 8048550, Japan

^bSchool of Computer Science and Engineering, Chang Shu Institute of Technology, Changshu, 215500, China

*Corresponding Author: zhu.miaomiao234@mail.kyutech.jp

Abstract

Image-to-image translation is an important topic in the field of computer vision. It aims to learn the mapping between input image and output image by training the datasets, and finally translates the image style from one domain to another. In terms of the form of translation, it can be divided into the translation between two domains and multiple domains from different datasets. And it is also divided into pairs and unpaired by the training datasets. As a successful representation of the translation of an unpaired image between two domains, the CycleGAN model is of great significance to the research and application. Starting from the application background of the CycleGAN model, this paper compares it with the basic deep learning model GAN, the paired image translation model pix2pix, other unpaired image-to-image translation models and multiple domains translation models separately, and gives their own model structure diagram. Then, the further application of the CycleGAN in other fields of computer vision is analyzed, especially in the field of Person Re-recognition and face change. Finally, according to the actual running results of each model, the existing problems are analyzed and summarized. The further research work is given.

keywords: GAN, pix2pix, CycleGAN, PTGAN, StarGAN.

1. Introduction

Since the advent of Convolutional Neural network (CNN), deep learning as an implementation algorithm of machine learning, has been widely used in Computer Vision field, from the beginning of the recognition of handwritten, object tracking, speech recognition, face recognition, and through training of fruit images dataset, help person determine the type of fruit when given image about it, until Alpha Go, it is the first computer program to defeat a professional human Go player, the first program to defeat a Go world

champion, and arguably the strongest Go player in history, making the deep reinforcement learning attracted more and more researcher's attention.

At the same time, Goodfellow *et al.*⁽⁵⁾ proposed the famous model GAN (Generative Adversarial Networks) in 2014, which is considered to be the coolest idea in the field of machine learning in the past 20 years and bring new breakthroughs to the deep learning model. The key to GAN's success is the idea of an adversarial loss that forces the generated images to be, in principle, indistinguishable from real photos. This loss is particularly powerful for image generation tasks, as this is exactly the objective that many of computer graphics aims to optimize.

Based on the perfect combination of CNN and GAN, there are many problems of image processing have been further studied and solved in recent years.^{(1), (2), (3), (4), (5), (6), (7), (8), (9)} UC Berkeley, Jun-Yan Zhu *et al.*⁽¹⁰⁾ extended the application of GAN to the Image-to-image translation, the models pix2pix and CycleGAN have been proposed, and apply their method to a wide range of applications, including collection style transfer, object transfiguration, season transfer and photo enhancement,^{(2), (3)} and then led to further studies in other fields by more researchers, such as Person Re-identification,^{(4), (11)} Multi-domain translation,⁽¹⁰⁾ game graphics model, and so on.

2. CycleGAN

In this section, we first briefly review the principle of CycleGAN, and then give a comparison with the GAN and pix2pix and other unpaired image-to-image translation models.

2.1 The background of CycleGAN

The task of image style translation is to change a particular aspect of a given image to another, that is to say, given any two unordered image collections X and Y, the algorithm learns to automatically "translate" an image from one into the other

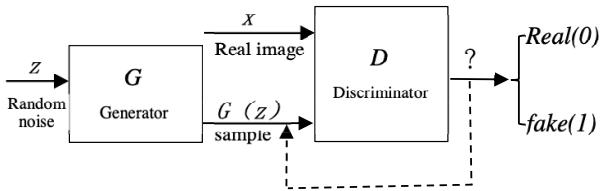


Fig. 1. GAN Model

and vice versa,⁽¹⁾ which can be divided into two situations: one is paired translation, and the other is unpaired translation.

For image translation in pairs, the commonly used models is pix2pix,⁽²⁾ It can verify by example that pix2pix model has a certain universality, but also has certain limitation. First of all, the requirements of the training dataset of training samples must be in pairs, but in real life, if you want to collect a large number of data in pairs, it is a little difficulty. If the data quantity is little, and there will be a fitting, model training effect. Secondly, the image transfer achieved by this model is only limited to the transformation of different styles of the same image, such as color changes. At this time, how to use samples with different styles for training to complete the translation of image style is the basic design idea of CycleGAN model.

2.2 CycleGAN Review

As one of the most interesting and important deep learning achievements in 2017, CycleGAN has many application in the past two years. The original schematic and formula of the CycleGAN paper are as follows.^{(3),(4)}

Jun-Yan Zhun *et al.*⁽⁴⁾ given two datasets $\{x_i\}$ ($i = 1$ to M) and $\{y_j\}$ ($j = 1$ to N), collected from two different domains X and Y , where $x_i \in X$ and $y_j \in Y$, the goal of CycleGAN is to learn a mapping function $G: X \rightarrow Y$, such that the distribution of images from $G(X)$ is indistinguishable from the distribution Y using an adversarial loss.

CycleGAN contains two mapping functions $G: X \rightarrow Y$ and $F: Y \rightarrow X$. Two adversarial discriminators D_X and D_Y are proposed to distinguish whether images are translated from another domain. CycleGAN applies the GAN framework to train the generative and discriminative models jointly. The overall CycleGAN loss function is expressed as follows:

$$\begin{aligned} V(G, F, D_X, D_Y) = & V_{GAN}(D_Y, G, X, Y) \\ & + V_{GAN}(D_X, F, Y, X) \\ & + \lambda V_{cyc}(G, F) \end{aligned} \quad (1)$$

Where $V_{GAN}(D_Y, G, X, Y)$ and $V_{GAN}(D_X, F, Y, X)$ are the loss functions for the mapping functions G and F , and for the discriminators D_Y and D_X is the cycle consistency loss

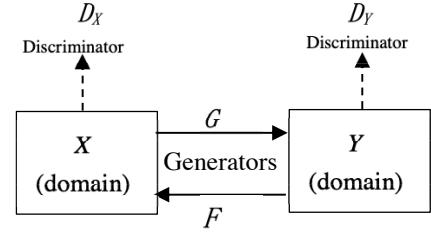


Fig. 2. CycleGAN Model

that forces $F(G(x)) \approx x$ and $G(F(y)) \approx y$, in which each image can be reconstructed after a cycle mapping. λ penalizes the importance between V_{GAN} and V_{cyc} . For more details of CycleGAN, can refer to.⁽⁴⁾

2.3 Comparison with GAN

As show in Fig. 1, GAN model is composed of Generative Nets (G) and Adversarial Nets (D), that is to say, G is Generator and D is Discriminator. The main role of G is to translate the random noise z into the samples $G(z)$, and the role of D , which is equivalent to a binary classifier, is to judge whether the input image is real or fake. If the real output is 1, others is 0, x is the real image.

During training Generative Nets (G) and Adversarial Nets (D) are dynamic game processes for each other. We train D to maximize the probability of assigning the correct label to both training examples and samples from G . We simultaneously train G to minimize $\log(1 - D(G(z)))$, In other words, G plays the following two-player minimax game with the value function $V(G, D)$:⁽⁵⁾

$$\begin{aligned} \min_G \max_D V(D, G) = & \mathbb{E}_{x \sim P_{data}(x)} [\log D(x)] \\ & + \mathbb{E}_{z \sim P_z(z)} [\log (1 - D(G(z)))] \end{aligned} \quad (2)$$

Here x represents the real image, z represents the noise which input to the G , $G(z)$ represents the image generated by the G and $D(x)$ represents the probability that D judge whether the image is true or not.

CycleGAN is based on the GAN model, and is composed of two GANs. Instead of inputting random noise, it inputs the images given by the user, which improves the ability of user control as well as the resolution and image quality.

Combined with the overall CycleGAN loss function expressed in Eq. 1, which have two discriminators and generators, it is a circular structure, so it's called Cycle. The framework as shown in Fig. 2.

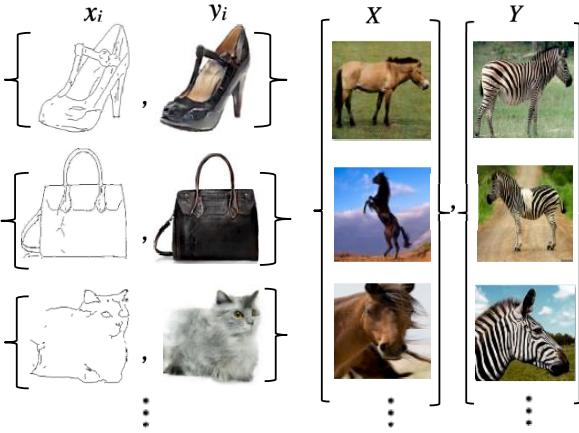


Fig. 3. Paired (left) and unpaired (right) training data

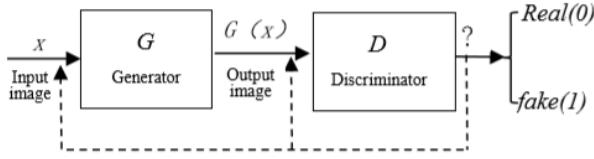


Fig. 4. pix2pix Model

2.4 Comparison with pix2pix

The fundamental similarity between pix2pix and CycleGAN is that they translate one type of image into another. The difference between them is shown in Fig. 3, pix2pix is to learn the mapping between an input image and an output image using a training set of paired data $\{x_i, y_i\}$ ($i = 1$ to N), where the correspondence between x_i and y_i exists.

Similar to the CycleGAN, pix2pix is also based on the GAN model, works by training on pairs of images such as building facade labels to building facades, and then attempts to generate the corresponding output image from any input image you give it.⁽¹⁾ However, instead of inputting random noise, it inputs the image given by the user, whose framework is shown in Fig. 4. Pix2pix uses the conditional GAN, and the objective of it can be expressed as:⁽²⁾

$$L_{cGAN} = \mathbb{E}_{x,y} [\log D(x, y)] + \mathbb{E}_{x,z} [\log (1 - D(x, G(x, z)))] \quad (3)$$

In order to make a comparison, pix2pix needs to train an ordinary GAN simultaneously, only let D judge whether it is a real image. In addition, on the basis of GAN, the conditional feature is added,⁽³⁾ that is, added the $L1$ loss function to the G network and an $L1$ loss was added to make the image of the source domain and the target domain as close as possible.

The overall pix2pix loss function is expressed as:

$$G^* = \arg \min_G \max_D L_{cGAN}(G, D) + \lambda L_{L1}(G) \quad (4)$$

Similar to the Eq. 2, G tries to minimize this objective against

an adversarial D that tries to maximize it.

As described in section 2.2, CycleGAN training with unpaired data, consisting of a source set $\{x_i\}$ ($i = 1$ to M), $x_i \in X$ and a target set $\{y_j\}$ ($j = 1$ to N), $y_j \in Y$, and the input and output of the two images cannot have any relationship. Compared with pix2pix, it has higher scalability and applicability. The overall of its loss function is expressed as the Eq. II.

2.5 Comparison with other unpaired Image-to-Image translation model

With regard to the unpaired Image-to-Image translation, besides the CycleGAN model, there are other models that have been proposed, such as DualGAN,⁽⁴⁾ DiscoGAN,⁽⁵⁾ and so on. All these models tackle the unpaired setting, effectively alleviate the problem of obtaining data pairs and only can learn the relations between two different domains at a time.

Zili Yi *et al.*⁽⁶⁾ have developed a novel DualGAN mechanism, which enables image translators to be trained from two sets of unlabeled images from two domains. In their architecture, the primal GAN learns to translate images from domain U to those in domain V, while the DualGAN learns to invert the task. In order to avoid costly pairing, Taeksoo Kim *et al.*⁽⁷⁾ address the task of discovering cross-domain relations given unpaired data. They have proposed a method based on generative adversarial networks that learns to discover relations between different domains (DiscoGAN).

It can be found that CycleGAN and DiscoGAN preserve key attributes between the input and the translated image by utilizing a cycle consistency loss. But unlike other approaches, CycleGAN does not rely on any task-specific, pre-defined similarity function between the input and output, nor do we assume that the input and output have to lie in the same low-dimensional embedding space. This makes the method a general-purpose solution for many vision and graphics tasks.⁽⁸⁾

3. Applications of CycleGAN

Due to the strong commonality of the CycleGAN framework, since its publication in 2017, it has attracted a lot of attention from image processing amateurs and researchers. In addition to simple applications, CycleGAN has a wide range of important applications in many fields, such as Person Re-Identification,^{(9), (10)} Face Generation,⁽¹¹⁾ Multi-domain image translation,^{(12), (13)} and so on.^{(14), (15)}

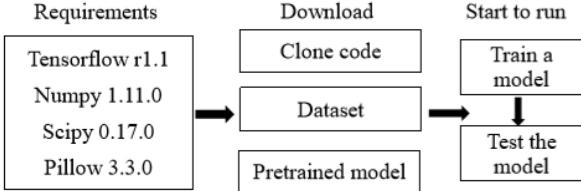


Fig. 5. Tensorflow implementation of CycleGAN

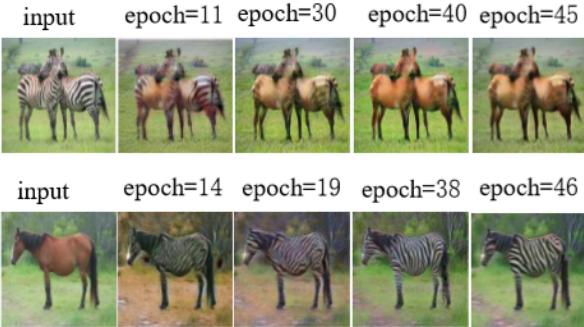


Fig. 6. Translation result on horse2zebra dataset

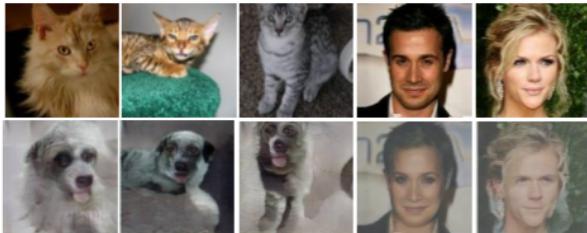


Fig. 7. cat→ dog (left) and female↔ male (right)

3.1 Implementation

In the paper,⁽¹¹⁾ the author given an examples for the translation about horses and zebras, apples and oranges, photos and paintings. In addition to the version developed by the original author, this paper implements the Tensorflow version of xhujoy,⁽¹⁸⁾ whose environmental requirements and running process as shown in Fig. 5.

Another example, as show in Fig. 8, Jack Clark,⁽¹⁶⁾ collected the ancient map of Babylon, Jerusalem and London, then translated it via CycleGAN into a modern Google Map & Satellite View. Instead of the original author's running on GPU, Fig. 8 shows the translation results on the CPU. By observing the training process, we can see the translation results is getting better and better as the number of training increases, and once run 200 epoch as original paper set, it can be find that there are many images which are difficult to distinguish true from fake.



Fig. 8. Map translation

3.2 Simple application

First of all, in addition to the transformation of horses and zebras, apples and oranges, photos and Monet, summer and winter listed in the original paper, there are some peoples who came up with a variety of magical applications about it,⁽¹⁵⁾ as shown in Fig. 7(left).By applying CycleGAN a cat can be translate into a dog, The first row shows the input image while the one below shows the translation results.

In addition, also as shown in Fig. 7 (right), only changing the datasets simply and then directly applying the CycleGAN model, we can translate a male person to female one or female to male (using datasets CelebA, and uniformly scale to 256×256).Of course, only face changing is considered here, and others, for example, hairstyles and clothes are not considered.

3.3 Person Re-Identification(Re-ID)

As a sub-problem of image retrieval, because of its important applications in security and surveillance, Person Re-ID has drawn lots of attention from both academia and industry.⁽¹³⁾

Although the development of deep learning and the emergence of many datasets, Person Re-ID performance has been significantly boosted. For example, the Rank-1 accuracy of single query on Market1501⁽²⁴⁾has been improved from 43.8%⁽²²⁾ to 89.9%.⁽²³⁾The Rank-1 accuracy on CUHK03⁽²⁰⁾ labeled dataset has been improved from 19.9%⁽²⁰⁾ to 88.5%.⁽²⁹⁾

However, there still remain several challenges hindering the applications of Person Re-ID, one of which is the scarcity of datasets. To address this challenge, Person Transfer Generative Adversarial Network (PTGAN) was proposed by Longhui Wei *et al.*⁽¹²⁾in 2018, which is inspired by the CycleGAN. And PTGAN model is propose to bridge the domain gap by transferring persons in dataset *A* to another dataset *B*. The transferred persons from *A*, are desired to keep their identities, meanwhile present similar styles, e.g., backgrounds,



Fig. 9. Comparison of PTGAN and CycleGAN

lightings, etc., with persons in B .

Different from CycleGAN,⁽¹⁾ PTGAN considers extra constraints on the person foregrounds to ensure the stability of their identities during transferring.

As shown some sample results generated by PTGAN and CycleGAN in Fig. 9. Compared with CycleGAN, PTGAN generates person images with substantially higher quality, and person identities are kept and the styles are effectively transformed. Extensive experimental results on several datasets show that PTGAN can effectively reduce the domain gap among datasets.⁽¹²⁾

In addition to the PTGAN, the other challenge of the applications of person Re-ID is camera variations, because of being a cross-camera retrieval task, person Re-Indentation suffers from image style variations caused by different cameras. As described above by PTGAN, using the image translate principle of CycleGAN, labeled training images can be style-transferred to each camera, and along with the original training samples form the augmented training set. However, after analysis and experimentation, it can be found that this method also needs to be improved. In order to solve the problem of over-fitting, increasing the data sample also makes a considerable level of noise.

To address this problem, Zhun Zhong *et al.*⁽⁸⁾ proposed a camera style (CamStyle) adaptation method. It employed CycleGAN to generate new training samples: the styles between different cameras are considered as different domains. Given a Re-ID dataset containing images collected from L different camera views, this method is to learn image-to-image translation models for each camera pair with CycleGAN. To encourage the style transfer to preserve the color consistency between the input and the output, they add the identity mapping loss⁽¹⁾ in the CycleGAN loss function (Eq. II) to enforce the generator to approximate an identity mapping when using

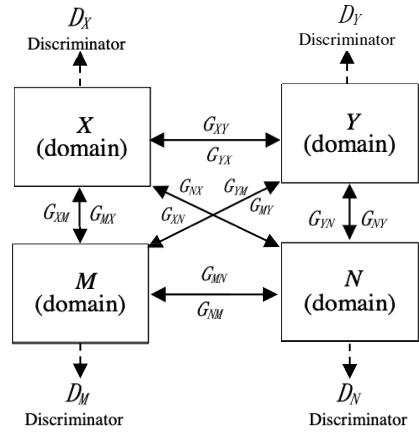


Fig. 10. Multi-domain Model

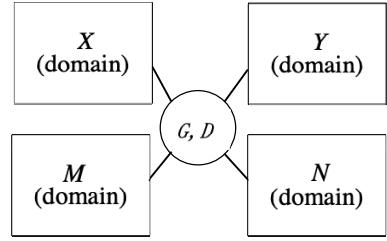


Fig. 11. StarGAN Model

real images of the target domain as input.

Specifically, following the training strategy in,⁽¹⁾ for training images, resize all images to 256×256 , and use CycleGAN to train camera-aware style transfer models for each pair of cameras.

3.4 Multi-domain image translation

For the study of image to-image translation, the CycleGAN model canceled the restriction of paired data, and used two image datasets with different styles to train, completing the image style translation for two domains. However, existing approaches have limited scalability and robustness in handling more than two domains, since different models should be built independently for every pair of image domains.

That is to say, if we want to train multiple domains from different datasets, each domain needs to retrain the model. For example, if there are four styles of images, according to the CycleGAN model, their inefficiency results from the fact that in order to learn all mappings among 4 domains, $4 \times (4-1)=12$ generators have to be trained. As shown in Fig. II, it illustrates how twelve distinct generator networks have to be trained to translate images among four different domains, and the calculation cost is very high.

To address this limitation, Yunjey Choi *et al.*⁽¹⁰⁾ proposed StarGAN in 2018, a novel and scalable approach that can

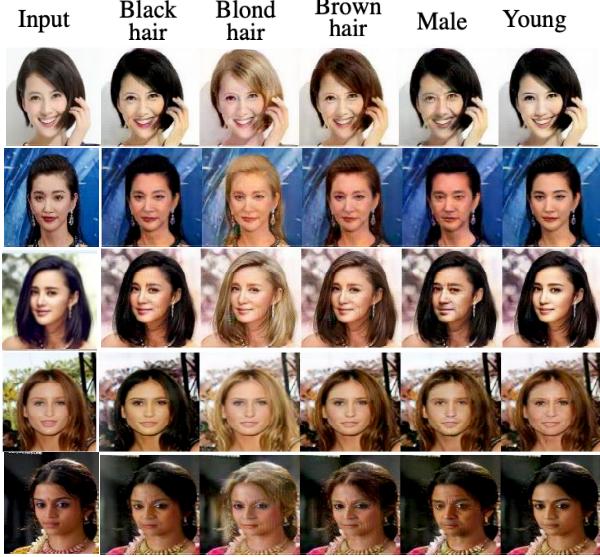


Fig. 12. Multi-domain translation results (A)

perform image-to-image translations for multiple domains using only a single model. Such a unified model architecture of StarGAN allows simultaneous training of multiple datasets with different domains within a single network. This leads to StarGAN’s superior quality of translated images compared with existing models as well as the novel capability of flexibly translating an input image to any desired target domain. They empirically demonstrate the effectiveness of their approach on a facial attribute transfer and a facial expression synthesis tasks.⁽¹⁰⁾

As shown in Fig. 13, the StarGAN model takes in training data of multiple domains, and learns the mappings between all available domains using only one generator and discriminator. As far as our knowledge goes, this work is the first to successfully perform multi-domain image translation across different datasets.⁽¹⁰⁾

The training process of StarGAN is very similar to CycleGAN. The original implementation of StarGAN was the pytorch version.⁽²⁶⁾ Fig. 13 shows the test results by using the simple Tensorflow version.^{(24), (27)} The image I wanted (the first three rows) and the CelebA test images (the last two rows) run simultaneously.

Also as shown in Fig. 13,⁽¹⁰⁾ the first column shows the input images, next three columns show the single attribute transfer results, and rightmost columns show the multi-attribute transfer results. (H: Hair color, G: Gender, A: Aged), which is the facial attribute translation results on the CelebA dataset. It can be found that StarGAN generated images of higher visual quality compared to CycleGAN model. By utilizing a mask vector method that enables StarGAN to control all available domain labels, we can further extend to train multiple do-



Fig. 13. Comparison of CycleGAN and StarGAN



Fig. 14. Multi-domain translation results (B)

mains from different datasets, such as jointly training CelebA and RaFD images to change a CelebA image’s facial expression using features learned by training on RaFD, as shown in Fig. 14.⁽¹⁰⁾

Except as described above the CycleGAN model application of Person re-identification and face change, there are other applications in other fields, such as, Xiaodan Liang *et al.*⁷ proposed the Contrasting GAN, game graphics model, and so on.

4. Discussion and Future Work

Although CycleGAN model have shown remarkable success in image-to-image translation for two domains, it can be found that in addition to the mentioned multi-domain image translation above, there are still some aspects that need to be improved.

First of all, the accuracy of image translation needs to be improved. For example, once a person appears in the input image, CycleGAN will also turn him into a zebra or horse, as shown in Fig. 15, in reality, which is unreasonable. As a solution to such problems, Mask R-CNN⁽¹⁹⁾ has been proposed. It makes an image segmentation first and then translate it into, but this effect is not very satisfactory, because once there are overlaps between people and zebras or horses on the image, for example, if a person rides on a horse, it will be difficult to divide. At the same time, we find that if the translation process

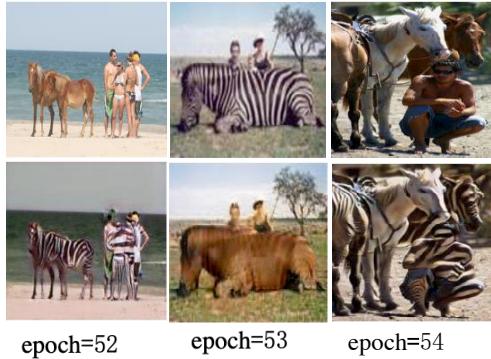


Fig. 15. input image (top) and output image (below)

requires geometric changes, the translation result is also not ideal, as shown in Fig. 7, on the task of translate dog and cat transfiguration. Those problems need further study.

Another problem we observe is that, the calculation cost needs to be improved. In the original program, they set epoch=200, epoch_step=100, and batch_size=1 separately. Instead of running on GPU, the author has trained it on CPU. It takes about six hours for one epoch, and the training time spent in the CPU is too long. Apart from those, there is an important problem that must be paid attention to in image processing field. After analyzing CycleGAN, only 256p/512p low resolution images can be output. For this problem, the pix2pixHD model is well solved.⁽²⁵⁾ Pix2pixHD takes a pyramidal approach: first output low resolution images. The previous output low resolution image is used as input to another network, and then a higher resolution image is generated.

5. Conclusions

The task of image-to-image translation is to change a particular aspect of a given image to another, e.g., it includes translation for two domains and multiple domains. In this paper, we mainly give a brief review on CycleGAN Generative Adversarial Networks, including its application background, basic principles, comparison with GAN, pix2pix, and other unpaired image-to-image translation model, etc. Finally, we give an analysis of its application in the field of computer vision. Our future work will continue to study more effective and efficient image-to-image transfer strategies, especially in the field of Person Re-Identification.

Acknowledgment

This research was supported by “Zhirong Xingjiao” Industry-University Research Innovation Fund Project of Ministry of Education in China under grant No. 2018A01004.

References

- (1) Jun-Yan Zhu, Tinghui Zhou, Phillip Isola, Alexei A. Efros, “Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks”, In International Conference on Computer Vision (ICCV), 2017
- (2) Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A. Efros : “Image-to-image translation with conditional adversarial networks”, In Computer Vision and Pattern Recognition (CVPR), 2017
- (3) Mehdi Mirza, Simon Osindero: “Conditional Generative Adversarial Nets”, arXiv preprint arXiv: 1411.1784, 2014.
- (4) Zili Yi, Hao Zhang, Ping Tan, Minglun Gong: “Dual-GAN: Unsupervised Dual Learning for Image-to-Image Translation, In International Conference on Computer Vision (ICCV), 2017
- (5) IanJ. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio:“Generative Adversarial Nets”, In Conference and Workshop on Neural Information Processing Systems (NIPS), 2014
- (6) Jun-Yan Zhu, Philipp Krähenbühl, Eli Shechtman, and Alexei A. Efros: “Generative Visual Manipulation on the Natural Image Manifold”, In European Conference on Computer Vision (ECCV), 2016
- (7) Alec Radford, Luke Metz, Soumith Chintala: “Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks”, In International Conference on Learning Representations (ICLR), 2016
- (8) Zhun Zhong, Liang Zheng, Zhen Dong, Shaozi Li, Yi Yang: “Camera Style Adaptation for Person Re-identification”, In International Conference on Computer Vision (ICCV), 2017
- (9) Yongyi Lu, Yu-Wing Tai, and Chi-Keung Tang: “Attribute-Guided Face Generation Using Conditional CycleGAN ”, In European Conference on Computer Vision (ECCV), 2018
- (10) Yunjey Choi, Minje Choi, Munyoung Kim, Jung-Woo Ha, Sunghun Kim, Jaegul Choo: “StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation”, In Computer Vision and Pattern Recognition (CVPR), 2018
- (11) David Berthelot, Thomas Schumm, Luke Metz: “BE-GAN: Boundary Equilibrium Generative Adversarial Networks”, In Computer Vision and Pattern Recognition (CVPR), 2018
- (12) Taeksoo Kim, Moonsu Cha, Hyunsoo Kim, Jung Kwon Lee, Jiwon Kim: “Learning to Discover Cross-Domain Relations with Generative Adversarial Networks”, In

- International Conference on Computer Vision (*ICCV*), 2017
- (13) Longhui Wei, Shiliang Zhang, Wen Gao, Qi Tian: “Person Transfer GAN to Bridge Domain Gap for Person Re-Identification”, In Computer Vision and Pattern Recognition (*CVPR*), 2018
 - (14) Augustus Odena, Christopher Olah, Jonathon Shlens1: “Conditional Image Synthesis with Auxiliary Classifier GANs”, In International Conference on Machine Learning (*ICML*), 2017
 - (15) <https://zhuanlan.zhihu.com/p/28342644>, access date: 2019.5.10
 - (16) <https://twitter.com/jackclarkSF>, access date: 2019.5.10
 - (17) <https://affinelayer.com/pixsrv/>, access date: 2019.4.22
 - (18) <https://github.com/XHUJOY/CycleGAN-Tensorflow>, access date: 2019.5.19
 - (19) Kaiming He, Georgia Gkioxari, Piotr Dollár, Ross Girshick: “Mask R-CNN”, In Conference and Workshop on Neural Information Processing Systems (*NIPS*), 2018
 - (20) W. Li, R. Zhao, T. Xiao, and X. Wang: “Deepreid: Deep filter pairing neural network for person re-identification”, In Computer Vision and Pattern Recognition (*CVPR*), 2014
 - (21) L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian: “Scalable person re-identification: A benchmark”, In International Conference on Computer Vision (*ICCV*), 2015
 - (22) S. Liao, Y. Hu, X. Zhu, and S. Z. Li: “Person re-identification by local maximal occurrence representation and metric learning”, In Computer Vision and Pattern Recognition (*CVPR*), 2015
 - (23) L. Wei, S. Zhang, H. Yao, W. Gao, and Q. Tian: “Glad: Global-local-alignment descriptor for pedestrian retrieval”, Proceedings of the 25th ACM international conference on Multimedia (*ACMMM*), 2017
 - (24) <https://github.com/taki0112/StarGAN-Tensorflow#men>, access date: 2019.5.11
 - (25) Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, Bryan Catanzaro: “High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs”, In Computer Vision and Pattern Recognition (*CVPR*), 2018
 - (26) <https://github.com/yunjey/StarGAN>, access date: 2019.4.10
 - (27) <https://github.com/taki0112/StarGAN-Tensorflow>, access date: 2019.6.2
 - (28) M.-Y. Liu, T. Breuel, J. Kautz: “Unsupervised image-to-image translation networks”, In Conference and Workshop on Neural Information Processing Systems (*NIPS*), 2017
 - (29) C. Su, J. Li, S. Zhang, J. Xing, W. Gao, Q. Tian: “Pose driven deep convolutional model for person re-identification”. In International Conference on Computer Vision (*ICCV*), 2017
 - (30) Xun Huang, Ming-Yu Liu, Serge Belongie, Jan Kautz: “Multimodal Unsupervised Image-to-Image Translation”. In European Conference on Computer Vision (*ECCV*), 2018
 - (31) Xiaodan Liang, Hao Zhang, Eric P. Xing: “Generative Semantic Manipulation with Contrasting GAN”, In Conference and Workshop on Neural Information Processing Systems (*NIPS*), 2018