《数据挖掘导论》大作业

Image Style Conversion using Cycle-Consistent Adversarial Networks

院系: 数据科学与计算机学院

专业: 通信软件

姓名: 周芙蓉

学号: 16341023

中山大学 • 中国广州

Image Style Conversion using Cycle-Consistent Adversarial Networks

ABSTRACT

Image style conversion refers to the conversion of image content from one domain to another. This tasks generally require paired pictures with the same con- tent in the both domains as training data. Such as pix2pix[7], but this pair of training data is difficult to obtain. The innovation of CycleGAN is that can migrate picture content from the source domain to the target domain without paired training data. The Generated Confrontation Network (GAN) consists of two subnetworks: a generator and a recognizer. The input to the generator is a random noise or condition vector and the output is the target image. The recognizer is a classifier, the input is an image, and the output is whether the image is a real image. In the training process, the generator and the recognizer enhance their ability through continuous mutual game.

Key Words: Computer Vision , Pattern Recognition, computer science

目录

第 1 章　绪论

1.1　选题背景与意义

Handling another style of scene with different styles is a time-consuming and laborious task and requires a lot of professional drawing skills. In order to obtain a high quality picture, the original author must carefully draw each line and paint each color area of the target scene. It appears that existing art editing software and algorithms with standard features do not produce satisfactory comic effects. Therefore, if professional technology can automatically convert one style of photo into another, it is very helpful for the artist: it can save them a lot of time and let them focus on More meaningful and creative work The goal of picture migration is to use a paired picture training data set to learn the mapping between the input picture and the output picture, such as pix2pix[1], however, there are no paired training data sets for many items. The innovation of CycleGAN[2] is that can migrate picture content from the source domain to the target domain without paired training data.

（1） When CycleGAN[2] is training, it only needs to input the image of the source domain and the image of the target domain. This does not require the source domain to match the image content of the target domain.

（2） Based on cyclegan's CariGAN[3], you can automatically convert live-action photos into formally exaggerated comics without paired images.

（3） With CycleGAN, not only convert between two types of images, but also convert between two objects, such as translating one person into another.

1.2　国内外研究现状和相关工作

随着国内外研究者们对图像迁移越来越关注，出现了一系列的 Cyclegan 算法。在这一小节，我们简短的介绍四条相关工作（Generative Adversarial Networks

(GANs)、Image-to-Image Translation、Unpaired Image-to-Image Translation、Cycle Consistency）和其国内外研究现状。

1.2.1 Generative Adversarial Networks (GANs)

Generative Adversarial Networks (GANs)[4, 5] have achieved impressive results in image generation[6, 7], image editing [8], and representation learning[7, 9]. The key to the success of GAN is the concept of confrontational loss, which forces the generated image to be indistinguishable from the real image in principle.

1.2.2 Image-to-Image Translation

The idea of image to image conversion is to use a non-parametric texture model[10] on a single input-output training image pair. The latest method uses the data set of input and output checks to learn the parameter conversion function through CNN[11]. Our approach is based on the ”pix2pix”[1] framework of Isola et al[12]. It uses conditional generation of the network to learn the mapping from input to output images.

1.2.3 Unpaired Image-to-Image Translation

The similarity function between the predefined inputs and outputs that does not depend on any particular task, nor does it assume that we assume that the inputs and outputs must be in the same low-dimensional embedded space. This makes our approach a universal solution for many visual and graphical tasks.

1.2.4 Cycle Consistency

In target tracking, language translation, 3D shape registration and so on. There are also applications in CNN. A similar loss was introduced to push G and F to agree with each other.

1.3 国内外研究现状

（1） In 2018, researchers from Tsinghua University, City University of Hong Kong and Microsoft recently proposed CariGAN[3], which can automatically convert live-action photos into formally exaggerated cartoons without paired images.

（2）

第 2 章　技术背景知识

2.1　model

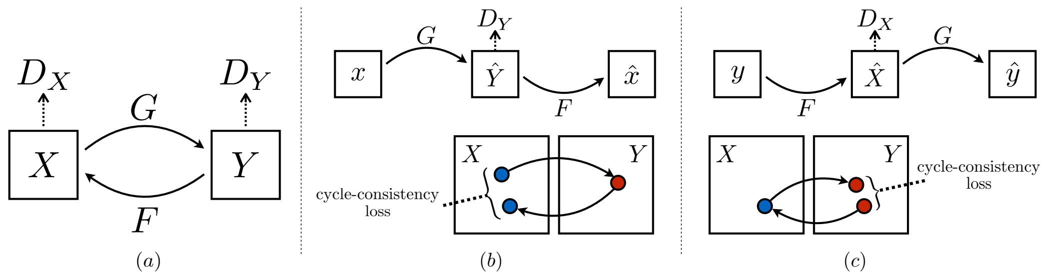


图 2-1: CycleGAN is actually an A→B one-way GAN plus a B→A one-way GAN. The two GANs share two generators, each with a discriminator, so adding up to a total of two discriminators and two generators. A one-way GAN has two losses, and CycleGAN adds up to a total of four losses.

2.1.1　表示

Adversarial Loss：For the mapping function G : X → Y and its dis- criminator DY , we express the objective as

$$\begin{aligned} \mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) = \ & \mathbb{E}_{y\sim p_{\text{data}}(y)}[\log D_Y(y)] \\ & + \mathbb{E}_{x\sim p_{\text{data}}(x)}[\log(1 - D_Y(G(x))] \end{aligned} \tag{2-1}$$

$$\begin{aligned} \mathcal{L}_{\text{GAN}}(G, D_X, X, Y) = \ & \mathbb{E}_{x\sim p_{\text{data}}(y)}[\log D_X(y)] \\ & + \mathbb{E}_{y\sim p_{\text{data}}(x)}[\log(1 - D_X(G(x))] \end{aligned} \tag{2-2}$$

Cycle Consistency Loss. The so-called Cycle consistency is to ensure

Forward consistent: x->G(x)->F(G(x)) x
Backward agreement: y->F(y)->G(F(y)) y
We incentivize this behavior using a cycle consistency loss:

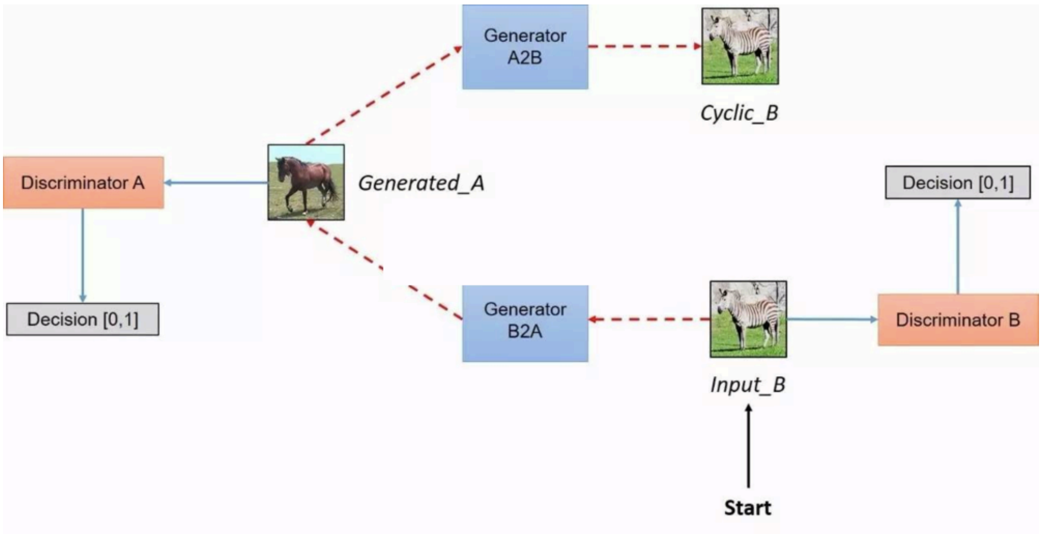
$$\begin{aligned} \mathcal{L}_{\text{cyc}}(G, F) = mathbbE_{x \sim p_{\text{data}}(x)}[[F(G(x)) - x]_1] \\ + \mathbb{E}_{y \sim p_{\text{data}}(y)}[[G(F(y)) - y]_1] \end{aligned} \tag{2-3}$$

full objective loss function is:

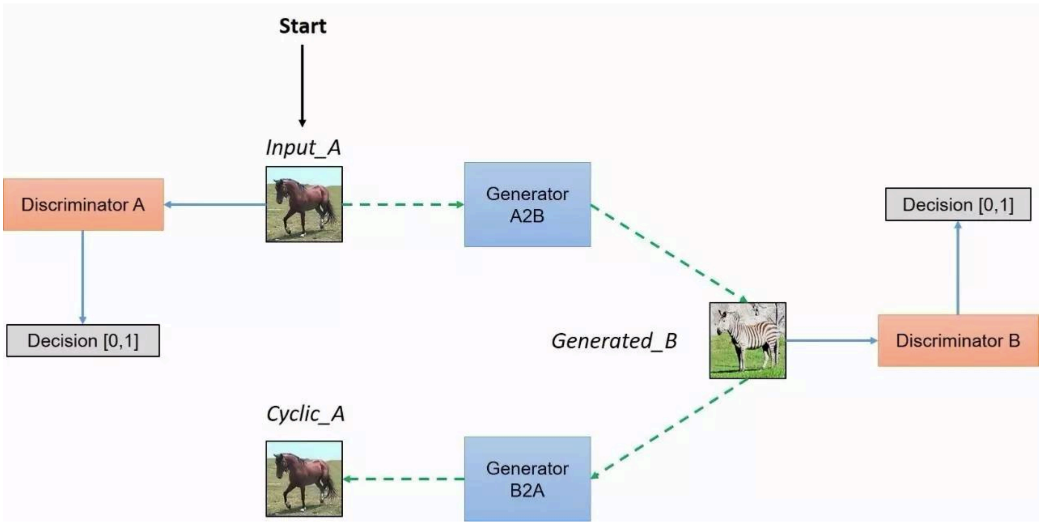
$$\begin{aligned} \mathcal{L}(G, F, D_X, D_Y) = \mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) \\ + \mathcal{L}_{\text{GAN}}(F, D_X, Y, X) \\ + \lambda \mathcal{L}_{\text{cyc}}(G, F) \end{aligned} \tag{2-4}$$

2.2 深度生成模型

The network architecture of CycleGAN is shown in the figure2-2:



(a)



(b)

图 2-2: The model acquires an input image from the domain DA, which is passed to the first generator GeneratorA→B, whose task is to convert a given image from the domain DA to an image in the target domain DB. This newly generated image is then passed to another generator, GeneratorB→A, whose task is to convert back to the image CyclicA in the original domain DA, which can be compared to the autoencoder. This output image must be similar to the original input image and is used to define meaningful mappings that did not exist in the unpaired data set.

第 3 章　非配对图片的图像迁移算法

3.1　算法流程

Algorithm 1: k, is a hyperparameter.We used k = 1, the least expensive option, in our experiments

Require: 图片 A 数据集 $\mathbf{R}$, 图片 B 数据集 $\mathbf{X}$，迭代次数 n

for $i = 1$ to n **do**

　for $i = 1$ to k **do**

　　Sample minibatch of m noise Samples from noise prior Pg(z)

　　Sample minibatch of n examples from data generationg districution Pdata(x)

　　Update the discriminator by ascending its stochastic gradient:

$$\bigtriangledown_{\theta_d} 1/M \sum_{n=1}^{M} [\log D(x^i) + \log(1 - D(G(z^i)))] \tag{3-1}$$

　end for

　Sample minibatch of m noise Samples from noise prior Pg(z)

　Update the discriminator by decending its stochastic gradient:

$$\bigtriangledown_{\theta_g} 1/M \sum_{n=1}^{M} \log(1 - D(G(z^i))) \tag{3-2}$$

end for

3.2　实验数据集

3.2.1　数据集来源

本次实验数据集来源为原作者提供的数据集，数据集来源:`https://people.eecs.berkeley.edu/~taesung_park/CycleGAN/datasets/`

3.3 实验结果

The experimental results are shown in the figure3-23-13-4.

3.4 不同图像迁移算法性能比较

As shown in the table3-1, Cyclegan basically realizes the conversion between unpaired pictures, but the image conversion effect is slightly worse than other image migration algorithms. The results obtained from the experiment do not affect the intuitive judgment of the human eye.

表 3-1

	Map to Photo	Photo to Map
	Error rate	Error rate
CoGAN[13]	1.1%	1.4%
SimGAN[14]	1.6%	2.8%
CycleGAN(our)	23.6%	22.4%

3.5 算法参数的影响

When obtaining the optimal parameters of the model, we use the small epoch to continuously test the optimal parameters of the approximation model. Finally, the learning rate is best around 0.002.

图 3-1

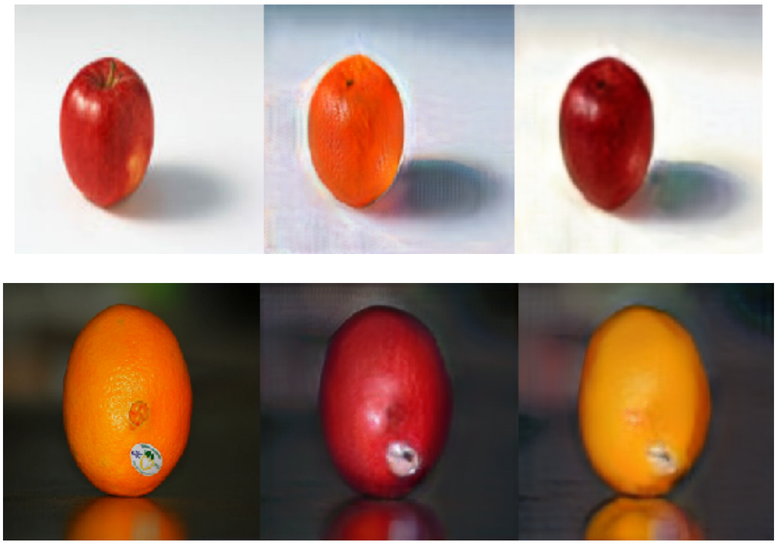


图 3-2: Basically, the conversion between unpaired images and images has been implemented, but the conversion effect is not very good, and the converted image is not smooth enough. In the pixel processing, our model still needs improvement.

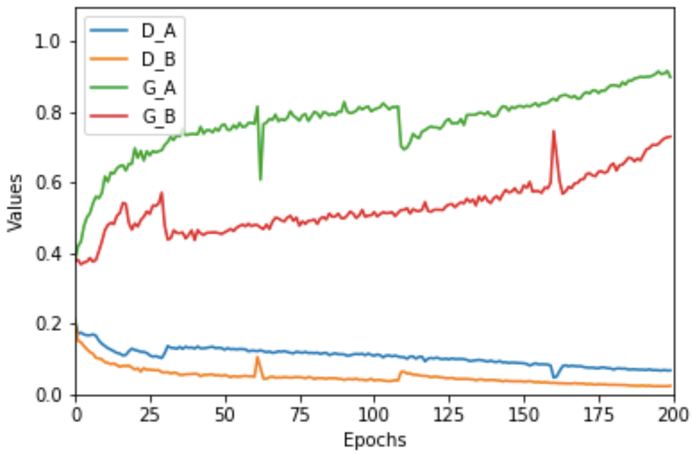


图 3-3: As the epoch continues to increase, the accuracy of GA and GB keep rising and gradually stabilizes. the stable values of D_A is about 0.83, the stable values of D_B is about 0.76.

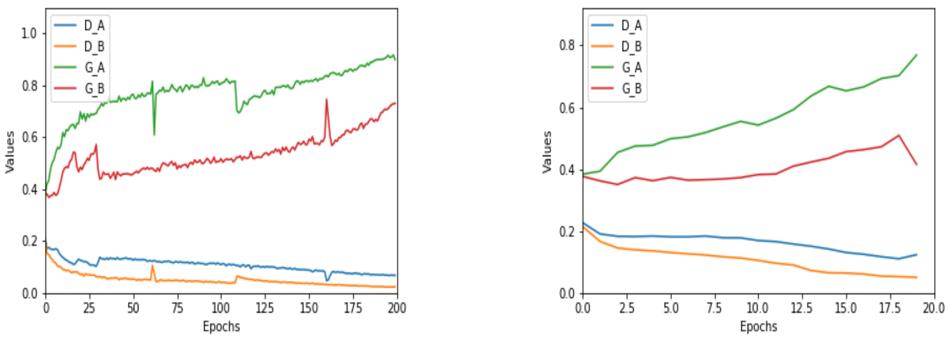


图 3-4: The epoch value increases and the value increases when the accuracy reaches stability.

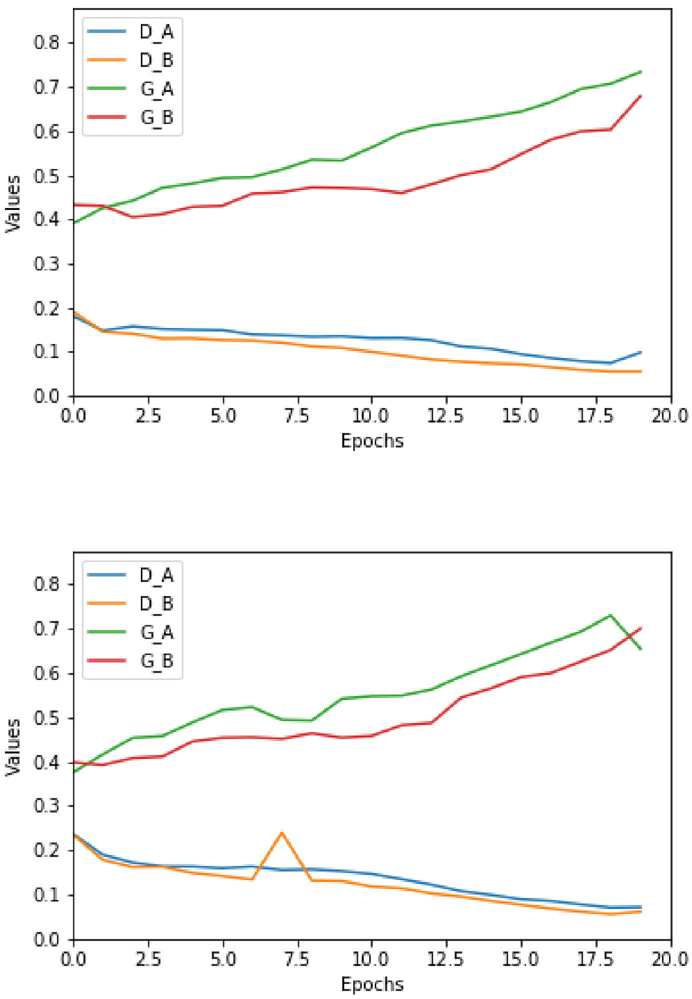


图 3-5: When the learning rate is equal to 0.003, there is an over-fitting phenomenon. When the learning rate is 0.0015, the image is slightly worse than the learning rate of 0.002.

第 4 章 总结

4.1 总结

In this report, we implement the code of “Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks”. The success of our cyclegan model is due to the standing of the work of our predecessors. This recurring experiment gives us a deeper understanding of GAN and encourages us to focus on our deep learning.

Code implementation is a fine-tuning, referring PyTorch-GAN and pytorch-CycleGAN-and-pix2pix

Since some code lines are too long, the cvpr template cannot display the code properly. The code item address is given below. `https://github.com/zfr0411/CycleGAN`

参考文献

[1] JPIsola, Zhu JY, TZhou, and Efros AA. Image-to-image translation with conditional adversarial networks. *CVPR*, 2017

[2] Zhu JY, Taesung, Isola PP, and Efros AA. Unpaired image-to-image translation using cycle- consistent adversarial networks. *CVPR*, 2017

[3] Cao K, Liao J, and Yuan L. CariGANs:Unpaired Photo-to-Caricature Translation. *SIGGRAPH Asia 2018*, 2018

[4] IGoodfellow, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, and Ben-gio Y. Generative adversarial nets. *NIPS*, 2014

[5] Zhao J, Mathieu M, and LeCun Y. Energy-based generative adversarial network. *ICLR,*, 2017

[6] Denton EL, Chintala S, Fergus R, and et a. Deep gen- erative image models using a laplacian pyramid of ad- versarial networks. *NIPS,*, 2015

[7] Radford A, Metz L, and Chintala S. Unsupervised representation learning with deep convolutional gen- erative adversarial networks. *ICLR*, 2016

[8] Zhu JY,¨henbu¨hl PK, Shechtman E, and Efros AA. Generative visual manipulation on the natural image manifold. *ECCV*, 2016

[9] Salimans T, Goodfellow I, Zaremba W, Cheung V, Radford A, and Chen X. Improved techniques for training GANs. *NIPS*, 2016

[10] Efros AA and Leung TK. Texture synthesis by non-parametric sampling. *ICCV*, 1999

[11] Long J, Shelhamer E, , and Darrell T. Fully convolu- tional networks for semantic segmentation. *CVPR*, 2015

[12] Isola P, Zhu JY, Zhou T, and Efros AA. Image- to-image translation with conditional adversarial net- works. *CVPR*, 2017

[13] Liu MY and Tuzel O. Coupled generative adversar- ial networks. *NIPS*, 2016

[14] Shrivastava A, Pfister T, Tuzel O, Susskind J, Wang W, and Webb R. Learning from simulated and unsupervised images through adversarial training. *CVPR*, 2017