

摘要

风格迁移是指将一幅原始图像转换为一种特定的风格，如油画、水彩和某个大师的风格。此前风格迁移已经被成功地应用在西方油画和一些代表性画家的作品上，而没有应用在水彩画和中国水墨画这类具有晕染效果、笔刷顺滑和留白特征的艺术风格上。生成对抗网络自提出以后，经常应用于艺术创作领域。本文的主要贡献在于，首先构建了水彩画数据集，然后应用最新提出的用于非配对数据集的基于平方势力散度的循环生成对抗网络模型(CycleGAN-qp)，模型中包含对抗损失函数、循环一致性损失函数和身份损失函数，由此模型实现了由自然花图像向水彩画图像的风格迁移，并将风格迁移效果与基于卷积神经网络的风格迁移效果进行定性的视觉效果质量比较，得出本文所使用的基于生成对抗网络的方法在生成图像的背景晕染效果和色彩创造性上表现得更好，另外本文最后对模型进行了消融实验及调优并提出了模型改进方向。

关键词：风格迁移；生成对抗网络；卷积神经网络；水彩画；循环一致性

Abstract

Style transfer is the conversion of an original image into a specific style, such as oil painting, watercolor, and the style of a master. Previous style transfer has been successfully applied to Western oil paintings and some representative painters' style, but not to the artistic styles of watercolors and Chinese ink paintings with smudged effects, smooth brush and blank area. After the proposition of the generative adversarial network, it is often used in the field of artistic creation. The main contribution of this paper is to first construct a watercolour dataset, and then apply the newly proposed cycle generative adversarial network model based on quadratic potential divergence (CycleGAN-qp) for unpaired datasets. The model contains the adversarial loss, the cycle consistency loss and the identity loss. The model realizes the style transfer from the natural flower image to the watercolor image, and compares the style transfer effect with the qualitative visual quality of the convolutional neural network based style transfer effect. The method based on generative adversarial network is better in the background smudge effect and color creativity of the generated image. In addition, this paper also conducts ablation experiments and tuning on CycleGAN-qp. Finally, this paper proposes the direction of model improvement.

Key words: style transfer; generative adversarial network; convolutional neural network; watercolor; cycle consistency

目录

摘要.....	1
Abstract.....	1
1. 导论.....	1
1.1 研究背景.....	1
1.2 研究思路.....	1
1.3 研究框架.....	2
2. 生成对抗网络研究进展.....	4
2.1 生成对抗网络(GAN).....	4
2.2 图像风格迁移.....	6
2.3 基于 CNN 的风格迁移.....	7
2.4 基于 GAN 的风格迁移.....	7
2.5 循环一致性(Cycle Consistency).....	8
3. CycleGAN 损失函数.....	10
3.1 对抗性损失(Adversarial loss).....	10
3.2 循环一致性损失(Cycle consistency loss).....	10
3.3 CycleGAN 损失函数.....	11
3.4 CycleGAN-qp 损失函数.....	12
4. 数据集及训练细节.....	13
4.1 数据集.....	13
4.2 优化器和学习率.....	13
4.3 伪代码.....	14
4.4 训练细节.....	14
5. 实验.....	16
5.1 CNN 与 CycleGAN 的效果对比.....	16
5.2 FID-Score.....	18

5.3	消融实验	18
5.4	调优结果	19
6.	结论	21
6.1	创新与不足	21
6.2	方法改进	22
	参考文献	23
	附录	26

1. 导论

1.1 研究背景

生成对抗网络^[2](Generative adversarial networks, 简称 GAN)自从 2014 年被提出以来,其在艺术创作领域的表现尤为突出,特别是在风格迁移中的应用研究,如在 2018 年末大火的“换脸”程序 DeepFake 原始版本基础模型使用的是卷积神经网络^[25](Convolutional neural networks, 简称 CNN),而改进版本中引入了 GAN,使得训练速度更快,生成的人脸图像也更清晰。我们希望能够不断提高 GAN 对于画作的理解能力,满足人们对于各种各样艺术性的追求,比如张择端、莫奈等的绘画风格,油画或是水彩画的风格。GAN 能够更深入地理解画作的空间结构、笔法技巧、颜色搭配和晕染效果,后期如果在 GAN 的训练过程中融入艺术家的协助,那么艺术创作将会更上一层楼。另外现在也开发出了创造性对抗网络(CAN)^[3],它的目标是创造的作品要与自己学习的数据集的东西有天壤之别。这种艺术创作无法被人们接受的现状也在改变,比如由法国艺术团体 Obvious 创作的 AI 画作 *Edmond de Belamy* 在佳士得以 35 万美元落锤,这幅画同样是用 GAN 生成的,另外由 AI 作曲家写的交响乐也被许多科技大厂用作发布会开场曲,它也能深刻地调动现场观众的情绪。所以我们现在要解决的问题就是,如何让 GAN 能够充分实现用户的个性化风格图像需求,毕竟技术和硬件都已经相当成熟了,进而创造出高质量的风格化图像。

1.2 研究思路

在 GAN 提出后不久,对其在图像艺术创作领域大多集中在某张图风格迁移为艺术家的某幅作品,即只涉及两幅图像,而不是大量学习这位艺术家的所有作品,不能使艺术家的整体绘画风格更自然地融入图像中,且此前绘画风格也多集中在西方油画,本文尝试 CNN 和 GAN 两种方法,将自然图像转换为水彩画风格,本文最后选择了花这一种物象进行风格迁移尝试。仅利用 CNN

进行风格迁移时，每次只能将一张图像转换为特定的风格，也就是不具备学习艺术家或者是某种绘画技巧整体风格的能力。当运用 GAN 进行风格迁移时，给定风格数据集，GAN 能学习到数据集的分布，例如本文给定的风格数据集水彩画，并把我们的原始图像映射到这个风格分布产生一个具有风格的假图像，在对抗训练过程中，不断降低判别器的判别能力，以致于最后判别器不能区分映射产生的假图像和风格数据集中的真实图像，最终使得风格迁移效果更自然。

本文主要利用 CycleGAN^[4]模型实现水彩画风格迁移，评价通过双向的对抗训练、图像风格映射与原始图像重构，是否能让 GAN 更好地理解水彩画的笔触、色彩和晕染效果以及重构效果。

在本文实验过程中着重利用 CycleGAN-qp^[5, 6]模型实现了水彩画风格迁移，由水彩画和自然图像训练出的生成器，对色彩和晕染效果的风格迁移表现良好，且在经过消融实验后提高了模型的训练速度，定量指标 FID-Score 也下降了。

1.3 研究框架

本文的整体研究框架，主要包括以下几个方面：

第一章导论部分，阐述了本文的研究背景和研究思路，提出本文要解决的问题是水彩画风格迁移和主要采用的方法是 CycleGAN。

第二章为文献综述。主要介绍了生成对抗网络的提出、应用以及其主要变体。随后介绍了图像风格迁移的概念，再分别介绍了目前基于卷积神经网络的风格迁移和基于 GAN 的风格迁移的相关研究。最后介绍了本文要使用的 CycleGAN 模型的基本原理。

第三章为主要模型分析部分。本章详细阐述了基础 CycleGAN 的损失函数构成以及本文所使用的 CycleGAN 模型变体 CycleGAN-qp。

第四、五章为实验过程。第四章提出了本文训练数据集的来源以及预处理，介绍了基于 CNN 和 GAN 的两种方法的训练细节，同时展示了 CycleGAN-qp 模型的代码逻辑和针对算法的调优过程。第五章为正式实验过程，定性地比较评价了两种方法的水彩画风格迁移的效果，CycleGAN-qp 方法的重构效果，另外对 CycleGAN 模型进行了消融实验，利用 FID-Score 定量地分析了不同方法的实验效果，最后展示了调优结果。

第六章为全文的结论。本章提出了本文的创新点与不足，讨论了在运用 CycleGAN 进行风格迁移时损失函数可以进行的改进。

2. 生成对抗网络研究进展

2.1 生成对抗网络(GAN)

生成对抗网络是 Goodfellow 等^[2]在 2014 年提出的一种生成式模型。GAN 在结构上受博弈论中的二人零和博弈的启发, 由一个生成模型和一个判别模型构成, 生成模型利用真实数据样本, 首先假设数据服从某个分布, 采用极大似然估计方法估测分布参数, 并依靠此分布生成“无限”新数据, 由判别模型这个二分类器判别之后输入的数据是真实数据样本还是生成的样本, 通过交互迭代进行优化, 优化过程是一个极大极小博弈(Minimax game), 即训练判别模型的目标是最大化判别能力, 而生成模型的目标是让判别模型无法区分真实数据样本和生成样本, 整个优化过程最终达到纳什均衡, 使生成模型逐步收敛到数据样本的真实分布。

GAN 的基本工作原理为: 首先生成模型利用随机数生成假样本, 与此同时判别模型接收到假样本和已经标注好的真实样本, 最开始生成模型生成的假样本与真实样本差距很大, 而判别模型也并不能区分出真假样本, 在根据已经标注的样本对判别模型的判别进行打分后, 判别模型区分出真假样本的能力不断提高, 同时为了“迷惑”判别模型, 生成模型根据随机数生成的假样本也会越来越接近真实样本, 生成模型生成能力也不断提高。最终, 使得输入随机变量经过生成模型生成为真实数据的分布, 得到最优的生成器。

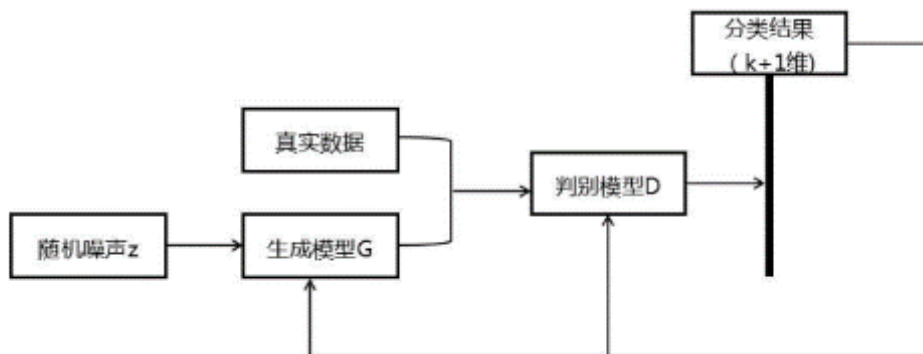


图 2-1 GAN 工作示意图

从 2014 年提出 GAN 以来，GAN 经历了损失函数演变^[7]，适应数据类型扩展^[10]以及改善优化器提高模型稳定性等过程^[8]，产生了一系列变体。

表 2-1 部分 GAN 的损失函数

Name	Value Function
GAN	$L_D^{GAN} = E[\log(D(x))] + E[\log(1 - D(G(z)))]$ $L_G^{GAN} = E[\log(D(G(z)))]$
LS-GAN ^[9]	$L_D^{LSGAN} = E[(D(x) - 1)^2] + E[D(G(z))^2]$ $L_G^{LSGAN} = E[(D(G(z)) - 1)^2]$
W-GAN ^[10]	$L_D^{WGAN} = E[D(x)] - E[D(G(z))]$ $L_G^{WGAN} = E[D(G(z))]$ $W_D \leftarrow clip_by_value(W_D, -0.01, 0.01)$
C-GAN ^[11]	$L_D^{CGAN} = E[\log(D(x, c))] + E[\log(1 - D(G(z), c))]$ $L_G^{CGAN} = E[\log(D(G(z), c))]$
Info-GAN ^[12]	$L_{D,Q}^{infoGAN} = L_D^{GAN} - \lambda L_I(c, c')$ $L_G^{infoGAN} = L_G^{GAN} - \lambda L_I(c, c')$
EB-GAN ^[13]	$L_D^{EBGAN} = D_{AE}(x) + \max(0, m - D_{AE}(G(z)))$ $L_G^{EBGAN} = D_{AE}(G(z)) + \lambda \cdot PT$
BE-GAN ^[14]	$L_D^{BEGAN} = D_{AE}(x) - k_t D_{AE}(G(z))$ $L_G^{BEGAN} = D_{AE}(G(z))$ $k_{t+1} = k_t + \lambda(\gamma D_{AE}(x) - D_{AE}(G(z)))$

图像和视觉、语音和语言是 GAN 研究和应用最广泛的领域，自从提出 GAN 以来，各种针对 GAN 存在问题的改进、GAN 的应用以及与强化学习等方法的结合不断衍生出来。目前 GAN 在各个领域的应用百花齐放，李飞飞^[15]团队在前人的基础上利用 GAN 改进了 text-to-image 即将文本翻译为图像的应用，新的方法中将句子文本先转换为 Scene Graph 使 GAN 更能理解输入文本的语义。在数据生成与增强领域，GAN 已经被应用在自动驾驶中的仿真场景生成^[19]，脑部 CT 图像生成^[16]以及图像像素丢失数据填补^[17]等方向，王坤峰等人^[1]的工作提出利用仿真图像和真实图像作为训练样本来实现人眼检测，只需要少量真实样本再通过 GAN 对仿真图像的修饰来实现。王飞跃也将仿真图像生成应用在平行视觉中^[1]。另外基于 GAN 的图像生成功能，早期很多研究都集中在 GAN 的艺术创作功能上，GAN 在风格迁移和风格转换中的应用尤为突出，如 Samneh 等人^[18]利用条件生成对抗网络实现由少量几个同类型英文字母训练生成同风格的余下字母，中科院的团队利用 GAN 实现针对中文描述的图像生成。由于现在硬件和算法都已经非常成熟了，GAN 生成的图像分辨率很高，在 2018 年末提出的 BigGAN^[19]，以及在此基础上变身后的 DeepGAN 表现出了极强的高分辨率真实图像生成能力，而且训练速度也更快，只是硬件要求很高。

2.2 图像风格迁移

图像风格迁移大体上分为配对图像风格迁移和非配对图像风格迁移两类，配对图像风格迁移每次实现过程中，只能有一幅特定输入图像作为迁移对象，一幅风格图像作为风格参考对象，因此输出的图像只能借鉴到该图像的风格，因此在风格迁移实践过程中，面临着大量图像需要配对标记，且每幅图像只能转换成一种特定风格，而不能自然的学习到某一类风格，例如梵高作品的整体风格，局限性较大。对于非配对性风格迁移来说，目标是关联两个不同的域，两个域中各包含着某种类型的图像，如水彩画和自然景观图像，而不仅仅是两幅图像，非配对图像风格迁移的实现过程中，能够先学习一组图像的整体风格，得到一个整体分布，再将此风格映射到迁移对象中，就能得到一个风格更加自然的结果，对具体输入图像没有要求，不仅仅是简单地复刻一幅风格图像。

2.3 基于 CNN 的风格迁移

基于卷积神经网络^[20](Convolutional neural networks, 简称 CNN)的风格迁移模型目标是根据预训练的卷积神经网络提取出图像的语义特征, 结合输入图像内容和参考图像的风格, 定义风格损失函数和内容损失函数, 加权后作为总的损失函数。利用 CNN 新生成一幅图像, 即是针对两张特定图像, 而不是两个不同域的图像数据集。最开始 Gatys^[20]等人运用 CNN 训练学习出图像的特征, 实现了在自然图像上再现著名绘画风格的功能, 如将梵高的《星空》这样著名画作的风格迁移到某一幅自然图像上。Johnson 和李飞飞^[15]团队在此基础上改进损失函数, 获得了细节更丰富的超分辨率结果。然而这些方法面临训练速度慢和计算成本高的问题, 而且之前的方法大多选择西方油画作为绘画风格, 并没有全面将风格迁移泛化与所有艺术风格, 例如有大量留白的中国画和水彩画。

2.4 基于 GAN 的风格迁移

基于 GAN 的方法与之前基于卷积神经网络的方法比较起来, 就是当 GAN 在应用于风格迁移时, 引入了对抗性损失^[2], 另外 GAN 一般能针对一个特定的域, 即一个含有某种风格的图像数据集, 在训练过程中, 得到这个图像数据集的整体风格分布, 最后再将这个分布映射到需要进行风格迁移的图像上, 或者直接是 end-to-end 的风格迁移, 即将自然风景图像转化为油画风格。

Zhu 等^[4]利用 Cycle-GAN^[7]实现对图像风格的转换, 包括可以实现将自然风景图像转换为梵高等艺术家风格的图像, 图像季节转换, 将图像中的物体转换为另一种物体以及转换图像情感等功能。Samanesh^[18]基于 GAN 字体风格迁移^[18]论文中, 利用 C-GAN^[7]实现对从少量(5 个左右)的任意类型艺术字体中泛化至全部 26 个字母, 采用两段式架构两次使用 C-GAN, 首先生成 26 个同类型字体字母, 再进行颜色和艺术的点缀。

此前清华的未来实验室高峰等人^[21]进行了中国画风格迁移的尝试,开发了“道子智能绘画系统”,此系统学习大量国画名家的作品后,可以创作出极其风格的国画作品,如可以输入真实马的图片从而绘画出徐悲鸿风格的马,他们为此系统开发了 Chip-GAN。道子系统的原理是将一副绘画图像的风格与内容剥离成两个独立元素,进行量化,其中自然图像是“内容”,画家个人的风格习惯是“风格”,由内容和风格最终产生了了根据自然图像艺术风格话后的水墨图像。针对于中国化对留白、笔墨和线条三方面做了专门的深入优化设计,使得道子系统能够更懂中国水墨画的技法。



图 2-1 Cycle-GAN 风格迁移演示



图 2-2 道子系统风格迁移演示

2.5 循环一致性(Cycle Consistency)

循环一致性^[4]来源于对偶学习^[22],运用对偶学习中涉及两个互为对偶的任务,例如机器翻译中从中文翻译到英语和从英语翻译到中文互为对偶、自然图像转化为优化风格和油画图像转化为自然图像互为对偶,这些互为对偶的任务可以形成一个闭环,如在风格迁移中,先尝试将自然图像迁移为油画风格,再将得到的这张图像重构为自然景观风格,对偶学习的目的就是要使这样产生的自然图像尽

可能地接近最开始输入的原始自然图像，这样可以达到在没有标注的数据中学习的目的。在 GAN 中引入循环一致性损失，能够降低对数据集的要求，至少不再需要特定的配对标记数据集，实用性更强。

3. CycleGAN 损失函数

3.1 对抗性损失(Adversarial loss)

生成对抗网络的核心是对抗性损失，而在 CycleGAN 中同时包含两个方向的对抗性损失。给定两个非配对的不同域训练集 X 和 Y ，模型中包含一个双映射： $G: X \rightarrow Y$ 和 $F: Y \rightarrow X$ ，对于生成器 G 和判别器 D_Y ，对于生成器 G 对抗性损失^[28]为：

$$L_{GAN}(G, D_Y, X, Y) = E_{y \sim p_{data}(y)}[\log D_Y(y)] + E_{x \sim p_{data}(x)}[\log(1 - D_Y(G(x)))] \quad (1)$$

其中 $p_{data}(y)$ 和 $p_{data}(x)$ 为两个真实数据分布，此处的损失函数为对数损失函数， $D_Y(y)$ 表示判别器 D_Y 判别真实数据 y 来自于真实数据集 Y 而不是由生成器 G 生成的假数据集的概率，由于 $D_Y(y) < 1$ 时，取对数为负数，最大化这一项的意义就在于，令判别器当 y 来自于真实数据集 Y 而不是由 G 生成的假数据时能准确地判断出 $D_Y(y) = 1$ 。同时 $D_Y(G(x))$ 也表示判别器 D_Y 判别由生成器 G 生成的假数据来自于真实数据集 Y 的概率，以此来训练生成器。整个对抗性损失对所有样本取期望，目标是最大化损失函数 L_{GAN} ，其中概率的衡量由散度决定，章节 3.4 中将介绍 CycleGAN-qp 模型对散度的改进。

与基本的 GAN 模型不同的是，CycleGAN 模型对生成器的输入数据不再是随机数，而是变成了两个不同域数据集的样本。

生成器 G 在训练过程中生成与训练集 Y 中真实图像风格相似的假图像，而判别器 D_Y 作为分类器，负责判别假图像与真实图像是否风格一致。训练中的目标是最大化判别器 D_Y 判别输入数据是否来自于真实数据集的能力，生成器 G 负责降低判别器的区分能力： $\min_G \max_{D_Y} L_{GAN}(G, D_Y, X, Y)$ ，对于生成器 F 和判别器 D_X ，对抗性损失函数类似 $\min_F \max_{D_X} L_{GAN}(F, D_X, Y, X)$ 。

3.2 循环一致性损失(Cycle consistency loss)

将循环一致性损失运用在生成对抗网络^[4]中，即从数据集 X 中给定一个真实样本 x ，把它的风格经过生成器 G 向前迁移为数据集 Y 中的整体风格，得到一个假样本 \hat{y} ，再用生成器 F 重构为数据集 X 中的风格，得到重构图像 $F[G(x)]$ ，目的是让此重构假图像尽可能地接近原始图像： $x \rightarrow G(x) = \hat{y} \rightarrow F[G(x)] \approx x$ 。另外循环一致性中包含两个方向的重构损失，因此对于数据集 Y 中给定每张真实图像也包含一个方向的循环损失约束： $y \rightarrow F(y) = \hat{x} \rightarrow G[F(y)] \approx y$ 。最终循环损失函数定义为双映射各自映射回原始图像的 L1 范数，即两个方向的重构损失：

$$\begin{aligned} L_{cycle}(G, F, X, Y) = & E_{x \sim p_{data}(x)} [||F(G(x)) - x||_1] \\ & + E_{y \sim p_{data}(y)} [||G(F(y)) - y||_1] \end{aligned} \quad (2)$$

3.3 CycleGAN 损失函数

CycleGAN^[4]最终的损失函数使对抗性损失和循环一致性损失的线性组合：

$$\begin{aligned} L(G, F, D_X, D_Y) = & L_{GAN}(G, D_Y, X, Y) \\ & + L_{GAN}(F, D_X, Y, X) \\ & + \lambda L_{cycle}(G, F, X, Y) \end{aligned} \quad (3)$$

训练的目标是解出最优的风格迁移生成器：

$$G^*, F^* = \operatorname{argmin}_{G, F} \max_{D_X, D_Y} L_{GAN}(G, F, D_X, D_Y) \quad (4)$$

其中超参数 λ 为循环一致性损失系数，控制循环损失函数在生成对抗网络训练中的贡献。

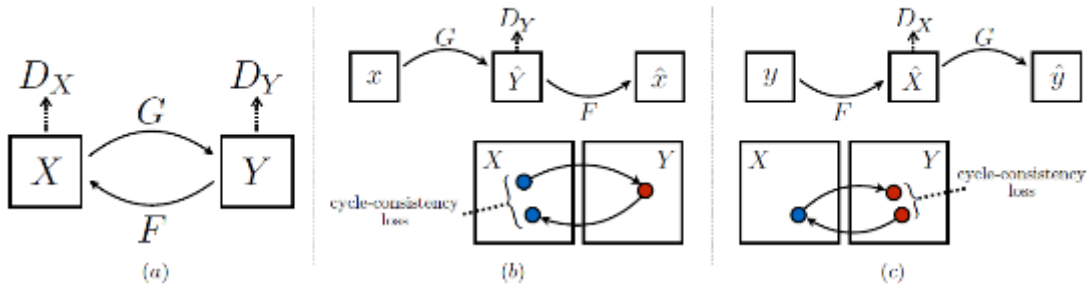


图 3-1 基本 CycleGAN 的语义结构

3.4 CycleGAN-qp 损失函数

本文在进行水彩画风格迁移训练时使用的是^[5, 6]最新提出的模型 CycleGAN-qp, 模型核心是定义了新散度, 即平方势散度(quadratic potential divergence, 以下简称 qp-div):

$$D_{qp}(x, \hat{x}) = a - \frac{a^2}{2\lambda\|x - \hat{x}\|_1} \quad (5)$$

从而改变了章节 3.1 中提到的对抗性损失, 其中 $a = D(x, \hat{x}) - D(\hat{x}, x)$, D 为生成对抗网络中的判别器, 负责对生成器生成的假样本和真实样本进行判别, 超参数 $\lambda > 0$ 。

散度衡量的是真实样本分布和生成的假样本分布之间的差异, 生成对抗网络的目的就是要使解出的生成器散度达到最优。Zhu^[4]等人最开始提出 CycleGAN 的时候是以 Least SquaresGAN^[9]为基础, 使用最小二乘法设计损失函数, 用皮尔森卡方散度代替原始 GAN^[2]中的 JS 散度。苏剑林提出的新散度 qp-div 解决了 WGAN^[10]需要额外的 L1 约束和 LS-GAN 存在梯度消失风险的问题。

同时 CycleGAN-qp 作者在构建模型时, 引入了身份损失(Identity Loss^[23])使对抗训练过程中, 生成器生成的假图像在整体结构上更接近原始图像。在本文第五章实验第一节中将比较运用 CNN 的水彩画风格迁移和以 CycleGAN-qp 为基础的风格迁移效果, 第二节将对 CycleGAN-qp 模型进行消融实验, 分析身份损失函数对风格迁移效果的影响。总损失函数:

$$\begin{aligned} L(G, F, D_X, D_Y) = & L_{GAN}(G, D_Y, X, Y) \\ & + L_{GAN}(F, D_X, Y, X) \\ & + \alpha L_{cycle}(G, F, X, Y) \\ & + \beta L_{ID}(G, F, X, Y) \end{aligned} \quad (6)$$

其中 α 和 β 分别为循环一致性损失函数和身份损失函数的在总损失函数中的权重。

4. 数据集及训练细节

4.1 数据集

本文数据集全部运用 flickr.walk 从图片分享网站 Flickr 上下载，用搜索标签 `text = watercolor flower`, `sort = relevance` 下载 1500 幅水彩画风格的花朵图像构成水彩画数据集。用搜索标签 `text = flower`, `sort = relevance` 下载 1000 幅自然花朵图像，另外用 `text = flower in vase` 下载 500 幅含花瓶和花两种物象的图像，共同构成含 1500 张图像的花数据集，因为初步下载完水彩画数据集后，发现包含一部分含花瓶的水彩画，而仅用搜索标签 `flower` 下载出的图像基本不包含有花瓶的图像。取得 `watercolor` 和 `flower` 两个数据集后，删除了其中的黑白图像、包含人脸的图像以及包含文字的图像，以免训练中学习这些特征，导致进行风格迁移时出现阴影。

由于下载图像时使用了 `sort = relevance` 需要随机划分训练集和测试集，两个数据集按照训练集:测试集 = 9:1 的比例随机划分，最终结果为：训练集中 `watercolor` 数据集包含 1311 张图像，`flower` 数据集中包含 1285 张图像；测试集中 `watercolor` 数据集包含 145，`flower` 数据集中包含 168 张图像。另外由于 GPU 训练内存的限制，调整了所有图像的分辨率至 256×256 。另外从 Flickr 上分别下载了 5 张水彩画风格图像和自然花图像用于基于 CNN 的风格迁移，同时自然花图像也用于基于 CycleGAN 风格迁移的输入图像。

4.2 优化器和学习率

出于减少训练时间的目的本文采用了不同的优化器和学习率进行尝试。优化器选择了 CycleGAN-qp 原文作者使用的 Adam 优化器^[24]和 2019 年年初由 Liangchen Luo 等人^[25]提出的 Adabound 优化器。针对学习率在使用 Adam 优化器的情况下使用固定学习率 0.0002，在使用 Adabound 优化器的情况下分别尝试了 0.001 和 0.0002 的学习率。

4.3 伪代码

下面是本文基于 CycleGAN-qp^[5, 6]方法实现水彩画风格迁移的算法伪代码:

Algorithm: CycleGAN-qp, both training and inference(style transfer)

```
1: if train = TRUE then
2:   X_set = real_flower, Y_set = watercolor;
3:    $\lambda = 10$ , CYC_WEIGHT = 10.0, ID_WEIGHT = 0.5;
4:   Learning_rate = 2e-4, BETA1 = 0.5, BETA2 = 0.999; #for Adam optimizer
5:   # Learning_rate = 1e-3, final_lr = 0.1; #for AdaBound optimizer
6:   BEGIN_ITER = 0, END_ITER = 15000;
7:   for i in range(BEGIN_ITER, END_ITER + 1):
8:     for j in range(1):
9:       total_loss = x_loss + y_loss; # QP-div losses
10:      total_loss += CYC_WEIGHT * (x_cyc_loss + y_cyc_loss);
          # Cycle-consistency loss
11:      total_loss += ID_WEIGHT * (x_id_loss + y_id_loss); # Identity loss
12:      update the networks(Generator, Discriminator);
13:    end for
14:  end for
15: else
16:  INFERENCE_ITER = 15000;
17:  input image;
18:  output = Generator(image, INFERENCE_ITER);
19: end if
```

4.4 训练细节

本文所有训练均在 Google Cloud 提供的 Colabratory 上完成,使用 Tensorflow 和 Pytorch 平台,开发语言为 Python3.6,硬件加速器选择 Google 提供的 GPU。在 Tensorflow 平台上运用深度卷积神经网络中的已经训练好的 imagenet-vgg-verydeep-19.mat 进行风格迁移实现,优化器是 Adam Solver,风格损失函数权重 $\alpha = 100$ 以及内容损失函数权重 $\beta = 5$,这样设置控制优化更趋于风格而不是原始内容。迭代次数 = 2000,在 Colabratory 上训练一次风格迁移,即一张输入图像和一张风格参考图像,大约花费 20 分钟。CycleGAN 的风格迁移训练在 Pytorch 上完成,运用 CycleGAN-qp 训练 watercolor 和 flower 两个训练集,迭代次数为 15000 次,在 Colabratory 上总计约花费 37 个小时,即约 1.5 天,整个训练过程

中，循环损失函数的权重 $\text{cyc_weight} = 10$ ，身份损失权重 $\text{Identity Loss_weight} = 0.5$ ，平方势散度系数 $\lambda = 10$ ，使用 Adam 优化器，学习率为 0.0002。由于 Identity Loss 的存在，导致训练过程中无法学习到水彩画的特有结构和晕染效果，因此在第五章第二节中进行了消融实验，训练过程中设置身份损失权重为 0，剔除影响重新进行训练，训练过程仍然迭代 15000 次，累计约花费 33 个小时，比保留身份损失函数时略快。另外在章节 5.4 中展示了基于 CycleGAN-qp 的风格迁移方法使用不同优化器和学习率的效果。

5. 实验

本次实验目的在于实现对自然花图像进行水彩画风格迁移,对基于卷积神经网络和生成对抗网络的两种方法进行生成风格迁移效果对比,得出基于生成对抗网络 CycleGAN-qp 的方法在对水彩画的晕染效果和色彩丰富性方面比基于卷积神经网络的方法能处理得更好。另外针对基于 CycleGAN-qp 的方法,本文进行了针对损失函数的消融实验和针对算法的调优,在一定程度上减少了训练时间、提升了风格迁移质量,对此方法同时进行了生成效果的定性和定量分析。

5.1 CNN 与 CycleGAN 的效果对比

基于卷积神经网络的风格迁移只针对两张图像,而基于生成对抗网络的风格迁移,本文所使用的 CycleGAN-qp 进行风格迁移训练时,使用两个不同风格的数据集即水彩画和自然图像,因此本节在对比两种方法的效果时,更重要的是需要进行定性的图像风格迁移质量分析。

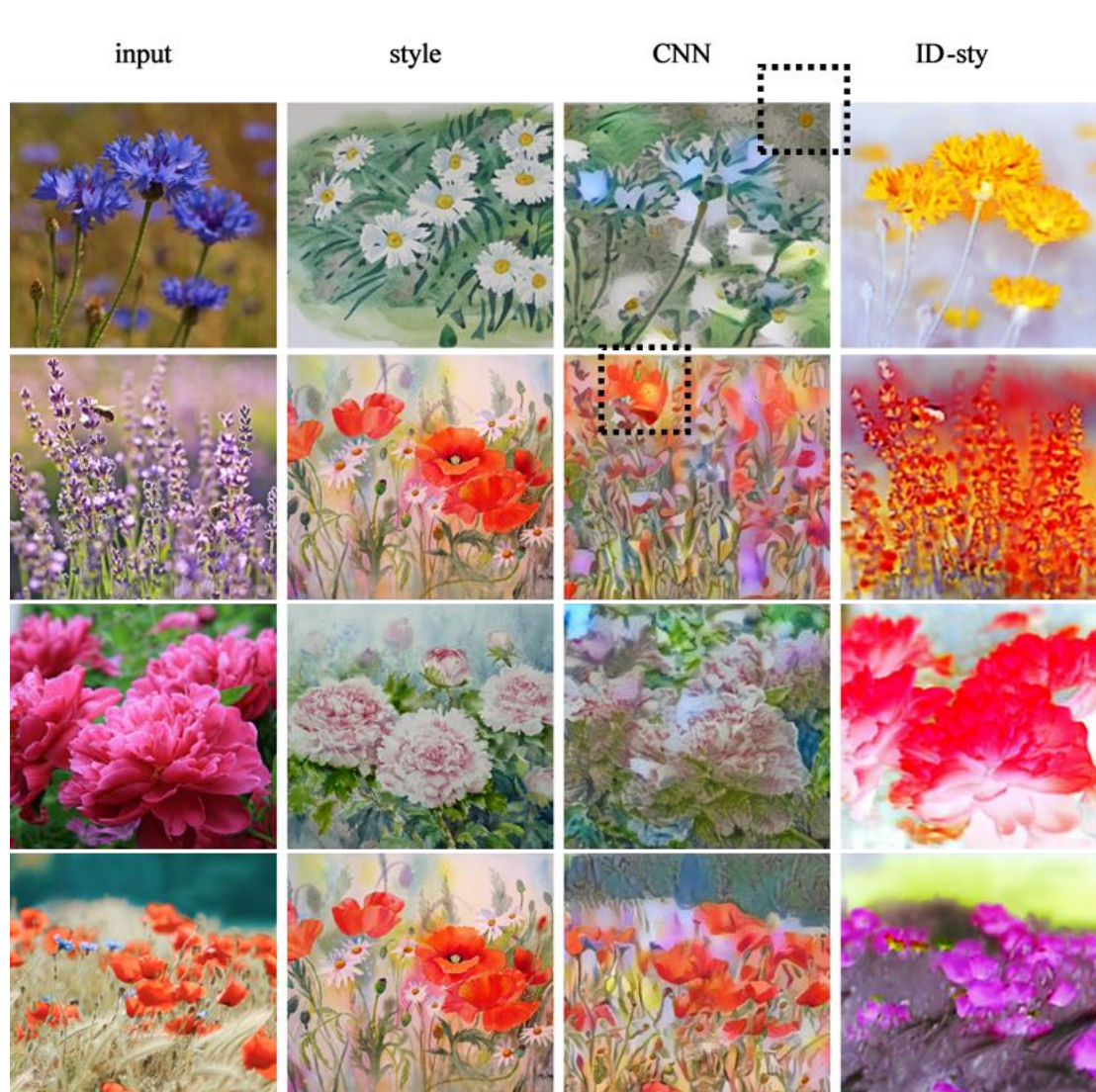


图 5-1 CNN 与 CycleGAN-qp(包含身份损失函数)，从左往右依次是：作为原始输入图像的自然花图像、用于 CNN 方法进行风格迁移时的水彩画风格参考图像、CNN 风格迁移效果、CycleGAN-qp 风格迁移效果。

在图 5-1 中，可以看出运用 CNN^[25]的风格迁移方法，使得输出图片学习出了不应该有的特征，如第一行矢车菊图中虚线框所画出的地方，CNN 风格迁移出的图片出现了只有风格参考图片中才有的黄色花蕊，第二行薰衣草图片经过 CNN 风格迁移也出现了不应该有的橘黄色花朵。因此运用 CNN 进行风格迁移时，为了让输出图像不产生多余的特征，水彩画风格参考图像的选择必须尽可能地与原始输入图像相似，如第三行芍药图的表现，由于原始图像与风格参考图像属于同一种类的不同风格，且图像花和叶特征区分度大，所以风格迁移效果较好。

CycleGAN-qp^[29]风格迁移是借鉴的整个水彩画数据集的风格，因此生成的水彩画风格图像在背景的晕染效果上更接近水彩画风格，而图像的主体花朵部分保留了原始图像的结构，也不会学习出原始输入图像没有的多余的特征。两种方法比较来看，基于 CycleGAN-qp 的方法生成图像的背景，更能表现出水彩画作品中特有的大色块晕染效果。另外基于 CNN 的风格迁移，生成图像的色彩只能来源于参考图像，色彩创造性缺乏，而基于 CycleGAN-qp 的风格迁移生成图像，色彩更富有创造性。

5.2 FID-Score

表 5-1 FID-Score

	ID-Loss	Learning rate	Optimizer	FID-Score	Training time
CycleGAN-qp	√	0.0002	Adam	317.80682	37h
	×	0.0002	Adam	310.93190	33h
	×	0.0002	Adabound	408.65767	
	×	0.001	Adabound	402.36352	28h

FID-Score^[35](Fréchet Inception Distance Score, 以下简称 FID-Score)是由 Heusel 提出的用于评价 GAN 的方法，衡量的是真实数据分布与生成数据分布之间的距离，指标数值越低表示生成数据分布越接近真实数据分布。在训练时同样迭代 15000 次的情况下，在本文中使用 Adam 优化器、在消融实验中剔除身份损失函数以及设置 Learning rate = 0.0002 的情况下 FID-Score 最低，可以认为此种情况下风格迁移效果最好。

5.3 消融实验

由苏剑林^[29]提出的 CycleGAN-qp 模型，主要包含对抗损失函数、循环一致性损失函数和身份损失函数，由于身份损失函数的存在是为了使得生成器生成的假图像在整体结构上尽可能保持与原始输入图像一致，这一效果对于我们进行水彩画风格迁移来说是不利的，因为水彩画风格迁移的目的是要使得生成器能够在输入图像的基础上重现水彩画丰富的色彩和富有晕染效果的笔触风格，而身份损失函数的存在使得风格迁移产生的图像过分地遵循原始图像的线条感觉，造成向

水彩画风格迁移效果不理想。在此节中对损失函数进行了消融实验(Ablation experiment), 目的是分析身份损失函数对水彩画风格迁移的影响, 在实验训练中, 将身份损失函数权重设置为 0 重新训练。

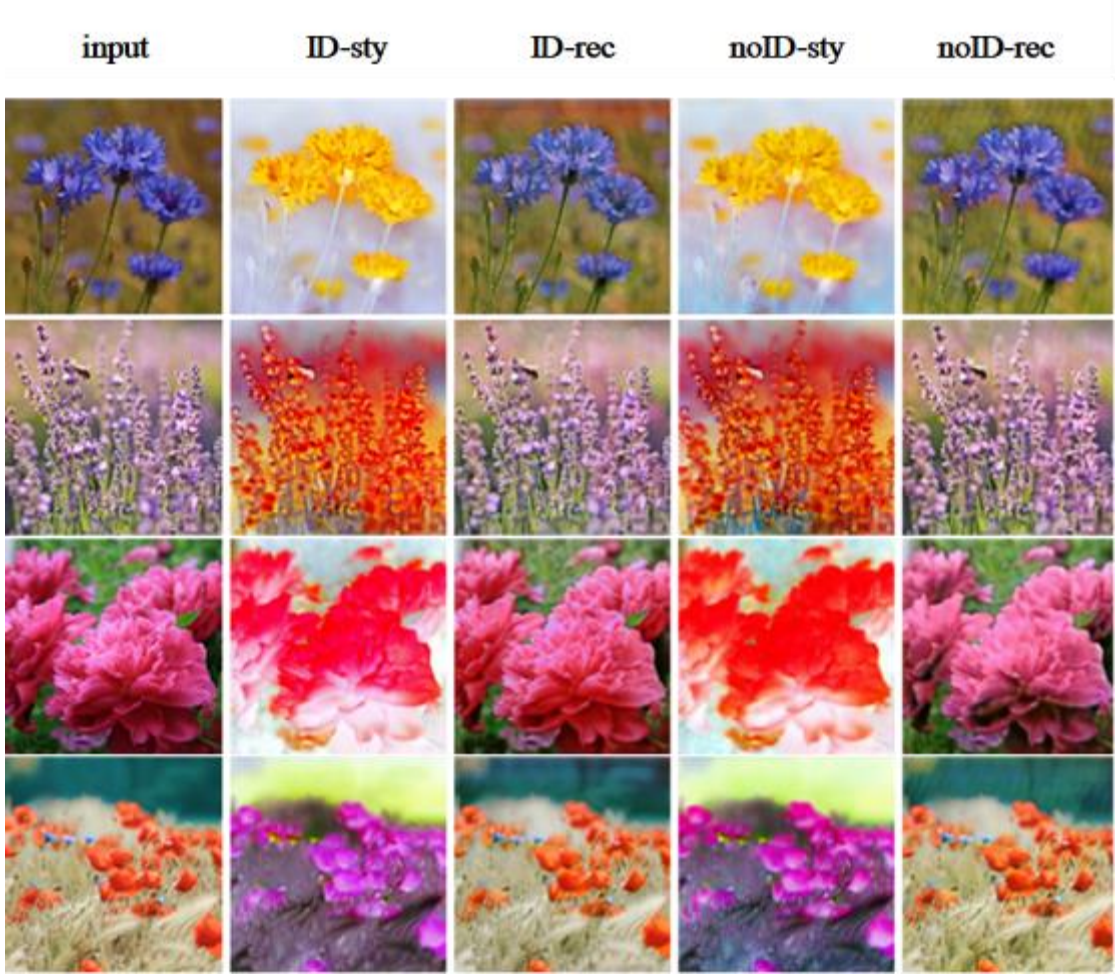


图 5-2 消融实验, 从左往右依次是: 原始输入图像、含身份损失函数模型风格迁移效果、含身份损失函数重构图像效果、不含身份损失函数风格迁移效果、不含身份损失函数重构图像效果。

从图 5-2 来看, 两种模型重构图像对原始输入图像的还原度都很高, 含身份损失函数的模型会加重重构图像的细节感, 另外除了亮度和色彩上两种模型生成的风格图像有差异外, 再无明显差异, 说明身份损失函数在此水彩画风格迁移中其实并没有多大作用, 是可以直接剔除, 简化模型, 从表 5-1 中可以看出剔除身份损失函数还能减少训练时间。

5.4 调优结果

由表 5-1 可以看出使用新提出的 Adabound 优化器虽然训练时间由 37h 减少到 28h, 但是 FID-Score 却由 310 提高到了 402, 另外从生成图像的质量上来看, 在图 5-3 中可以看出, 使用不同学习率, 无论是风格迁移图像还是图像重构效果, 使用 Adabound 优化器的情况下图像质量都很差。

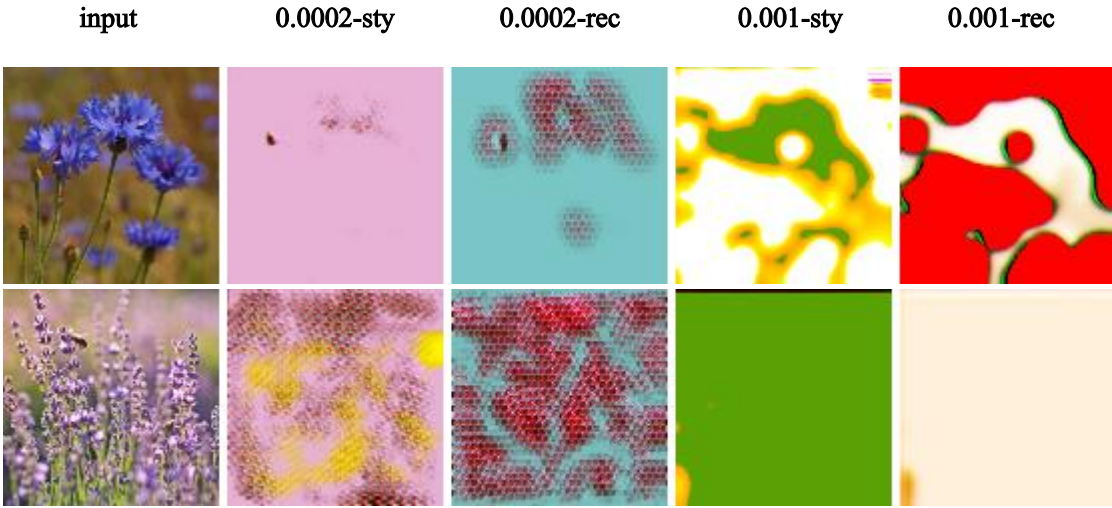


图 5-3 使用 Adabound 优化器以及不同学习率情况下的 CycleGAN-qp 风格迁移和图像重构效果, 从左往右依次是: 原始输入图像、学习率为 0.0002 的风格迁移及重构效果、学习率为 0.001 的风格迁移及重构效果。

6. 结论

6.1 创新与不足

本文借助生成对抗网络^[2]的思想，在总结思考苏剑林^[29]、高峰^[21]等人工作的前提下，为解决自然景观图像艺术化处理的问题，考虑到基于 CycleGAN-qp 模型的方法，最后实现了基于 CycleGAN 的水彩画风格迁移。具体来说最大的贡献在于构建了水彩画数据集，实现了由自然花图像向水彩画图像进行风格迁移，而且水彩画风各迁移效果在图像背景晕染效果和色彩创造性上都表现得很好，而且一般不会学习出原始图像没有的多余特征。在调优过程中得出，基于 CycleGAN-qp 的模型在剔除身份损失函数的情况下，综合来看风格迁移图像生成质量和训练速度比其他方法更好。

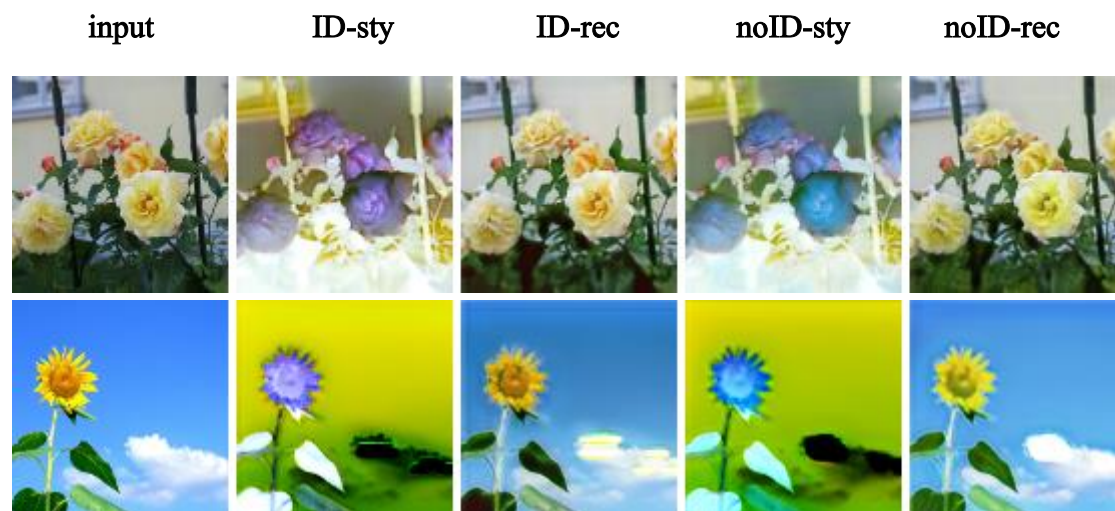


图 6-1 CycleGAN-qp 失败例子

但是利用 CycleGAN-qp 进行的风格迁移存在问题：一是除了背景的晕染和色彩向水彩画风格靠近外，作为图像主体的花朵并没有在线条效果上向水彩画的笔刷效果靠近；二是在实验过程中出现了一些失败的例子，向水彩画进行风格迁移时，色彩映射异常，产生了不能被大家接受的色彩组合，如图 6-1 所示，黄色月季被风格迁移为深蓝色，云朵被风格迁移为黑色。前一个问题出现可能是

由模型的损失函数并不能完美地适应水彩画风格迁移导致的,使得生成器在训练过程中没有学习到水彩画的笔刷特征。第二个问题原因可能是本文数据集不够全面,在构建水彩画数据集时,考虑到训练的时间成本,只选择了 1500 张水彩画图像,而且只选择了花朵这一物象,当现在训练结果出来后,发现进行水彩画风格迁移的目的是学习水彩画的笔刷特征、色彩和晕染效果,也许水彩画数据集并不需要限定在花这一种物象上,现在觉得不限定训练效果可能会更好。

6.2 方法改进

由 6.1 节中提到的问题,可以对本文提出的水彩画风格迁移做以下改进:首先在模型中应该引入笔刷损失函数,使得生成的风格图像在笔刷特征上更接近水彩画的笔刷特征,要做到这一点首先应该在原始输入图像中识别出花朵这一物象,因此此损失函数只针对输入数据集,而不针对水彩画数据集,可以设计新模型如下:

$$\begin{aligned} L(G, F, D_X, D_Y) = & L_{GAN}(G, D_Y, X, Y) + L_{GAN}(F, D_X, Y, X) \\ & + \alpha L_{cycle}(G, F, X, Y) + \gamma L_{brush}(G, X) \end{aligned} \quad (7)$$

另外,应该完善数据集,一是增加数据集的全面性,二是增大数据集,使得原始输入图像不再限制于花朵这一个单一物象以及生成器能够更好地把握水彩画的色彩、晕染和笔刷三个特征,实现对任意输入的自然景观图像进行水彩画风格迁移,同时使生成的图像更接近水彩画的风格,提高生成器的泛化能力。

参考文献

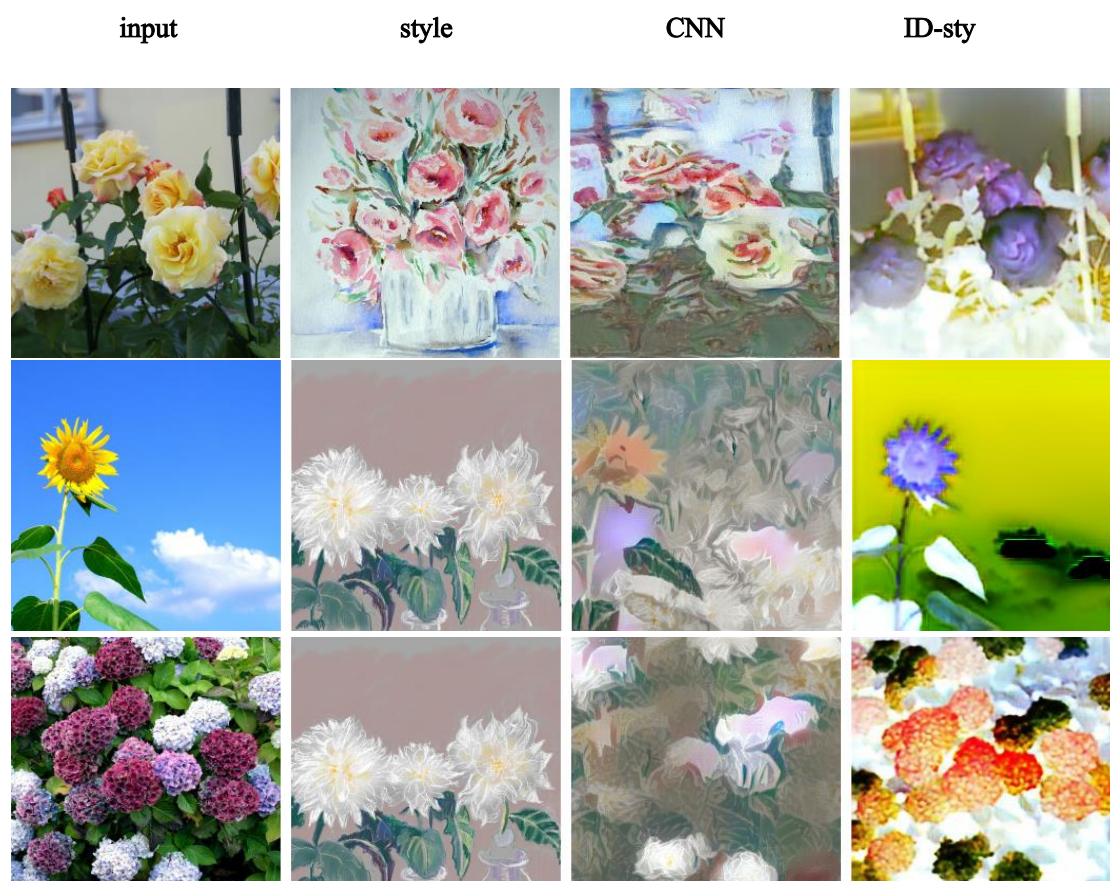
- [1] 王坤峰, 苟超, 王飞跃. 平行视觉: 基于 ACP 的智能视觉计算方法. 自动化学报, 2016, 42(10): 1490-1500.
- [2] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative adversarial nets. In: Proceedings of the 2014 Conference on Advances in Neural Information Processing Systems 27. Montreal, Canada: Curran Associates, Inc, 2014. 2672-2680.
- [3] Arjovsky M, Chintala S, Bottou L. Wasserstein GAN. arXiv preprint arXiv: 1701.07875, 2017.
- [4] Xudong M, Qing L, Haoran X, Raymond Y. Least Squares Generative Adversarial Networks. arXiv preprint arXiv: 1611.04076, 2017.
- [5] Jae H L, Jong C Y. Geometric GAN. arXiv preprint arXiv: 1705.02894, 2017.
- [6] Mehdi M, Simon O. Conditional Generative Adversarial Nets. arXiv preprint arXiv: 1411.1784, 2014.
- [7] Chen X, Duan Y, Houthoofd R, Schulman J, Sutskever I, Abbeel P. InfoGAN: interpretable representation learning by information maximizing generative adversarial nets. In: Proceedings of the 2016 Neural Information Processing Systems. Barcelona, Spain: Department of Information Technology IMEC, 2016. 2172-2180.
- [8] Zhao J B, Mathieu M, LeCun Y. Energy-based generative adversarial network. arXiv preprint arXiv: 1609.03126, 2016.
- [9] David B, Thomas S, Luke M. BEGAN: Boundary Equilibrium Generative Adversarial Networks. arXiv preprint arXiv: 1703.10717, 2017.
- [10] Yu L T, Zhang W N, Wang J, Yu Y. SeqGAN: sequence generative adversarial nets with policy gradient. arXiv preprint arXiv: 1609.05473, 2016.
- [11] Anders B, Søren K, Hugo L, Ole W. Autoencoding beyond pixels using a learned similarity metric. arXiv preprint arXiv: 1512.09300, 2016.
- [12] Takeru M, Toshiki K, Masanori K, Yuichi Y. SPECTRAL NORMALIZATION

- FOR GENERATIVE ADVERSARIAL NETWORKS. arXiv preprint arXiv: 1802.05957, 2018.
- [13] Han Z, Goodfellow I, Metaxas D, Odena A. Self-Attention Generative Adversarial Networks. arXiv preprint arXiv: 1805.08318, 2018.
- [14] Brock A, Donahue J, Simonyan K. LARGE SCALE GAN TRAINING FOR HIGH FIDELITY NATURAL IMAGE SYNTHESIS. arXiv preprint arXiv: 1809.11096, 2018.
- [15] Wang T C, Liu J Y, Tao A, Kautz J, Catanzaro. High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs. arXiv preprint arXiv: 1711.11585, 2018.
- [16] Ma S, Fu J, Chen C W, Mei T. DA-GAN: Instance-level Image Translation by Deep Attention Generative Adversarial Networks. arXiv preprint arXiv: 1802.06454, 2018.
- [17] Zhu J Y, Park T, Isola P, A.Efros A. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. arXiv preprint arXiv: 1703.10593, 2018.
- [18] Samaneh A, Fisher M, Kim V, Wang Z, Shechtman E, Darrell T. Multi-Content GAN for Few-Shot Font Style Transfer. arXiv preprint arXiv: 1712.00516, 2017.
- [19] Santana E, Hotz G. Learning a driving simulator. arXiv preprint arXiv: 1608.01230, 2016.
- [20] Gou C, Wu Y, Wang K, Wang K F, Wang F Y, Ji Q. A joint cascaded framework for simultaneous eye detection and eye state estimation. *Pattern Recognition*, 2017, 67: 23-31.
- [21] ChipGAN: A Generative Adversarial Network for Chinese Ink Wash Painting Style Transfer. ACM2018.
- [22] Frid-Adar M, Diamant I, Klang E, Amitai M, Goldberger J. GAN-based Synthetic Medical Image Augmentation for increased CNN Performance in Liver Lesion Classification. arXiv preprint arXiv: 1803.01229, 2018.
- [23] Huang J B, Kang S B, Ahuja N, Kopf J. Image completion using planar structure guidance. *ACM Transactions on Graphics*, 2014, 33(4): Article No. 129.

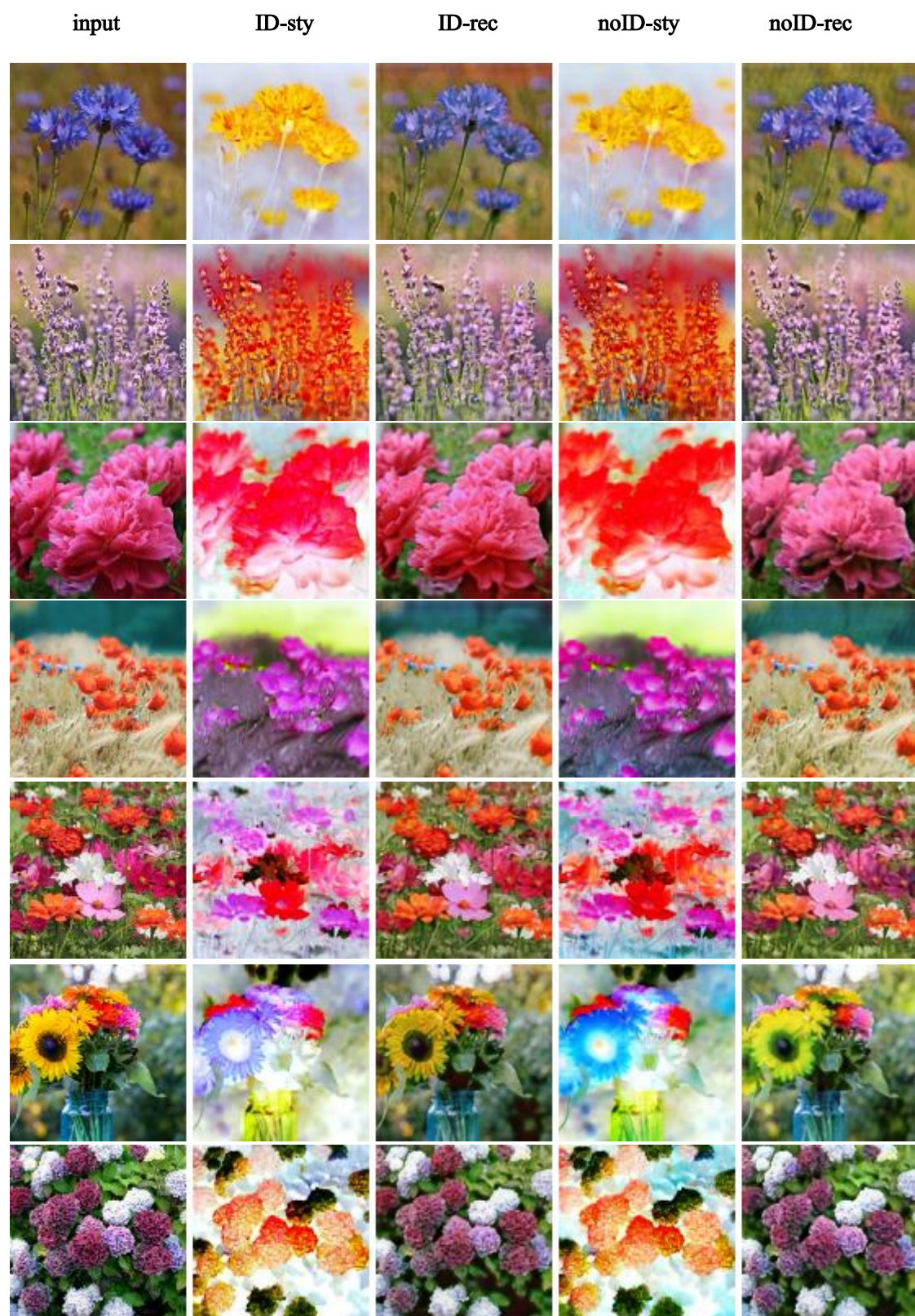
- [24] Johnson J, Gupta A, Li Fei-Fei. Image Generation from Scene Graphs. arXiv preprint arXiv: 1804. 01622, 2018.
- [25] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. In Proc. Advances in neural information processing systems. 1097 – 1105.
- [26] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. 2015. A neural algorithm of artistic style. arXiv:1508.06576 , 2015.
- [27] A. Hertzmann, C. E. Jacobs, N. Oliver, B. Curless, and D. H. Salesin. Image analogies. In *SIGGRAPH*, pages 327–340. ACM, 2001. 2
- [28] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. 2017. Image to image translation with conditional adversarial networks. arXiv:1611.07004 (2017).
- [29] Rahul Bhalley, Jianlin Su. Artist Style Transfer Via Quadratic Potential. arXiv:1902.11108v2 [cs.CV] 5 Mar 2019.
- [30] Ahmed E, Bingchen L, Mohamed E, Marian M. CAN: Creative Adversarial Networks Generating “Art” by Learning About Styles and Deviating from Style Norms. arXiv:1706.07068v1 [cs.AI] 21 Jun 2017.
- [31] Ashish Shrivastava, Tomas Pfister, Oncel Tuzel, Joshua Susskind, Wenda Wang, and Russell Webb. Learning from simulated and unsupervised images through adversarial training. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 2107–2116, 2017.
- [32] Jianlin Su. Gan-qp: A novel gan framework without gradient vanishing and lipschitz constraint. arXiv preprint arXiv:1811.07296, 2018.
- [33] Diederik P.Kingma, Jimmy Lei Ba. ADAM: A METHOD FOR STOCHASTIC OPTIMIZATION. arXiv:1412.6980v9 [cs.LG] 30 Jan 2017.
- [34] Liangchen Luo, Yuanhao Xiong, Yan Liu, Xu Sun. ADAPTIVE GRADIENT METHODS WITH DYNAMIC BOUND OF LEARNING RATE. In Proceedings of the ICLR 2019 Conference.
- [35] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, Sepp Hochreiter. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. arXiv:1706.08500v6 [cs.LG] 12 Jan 2018.

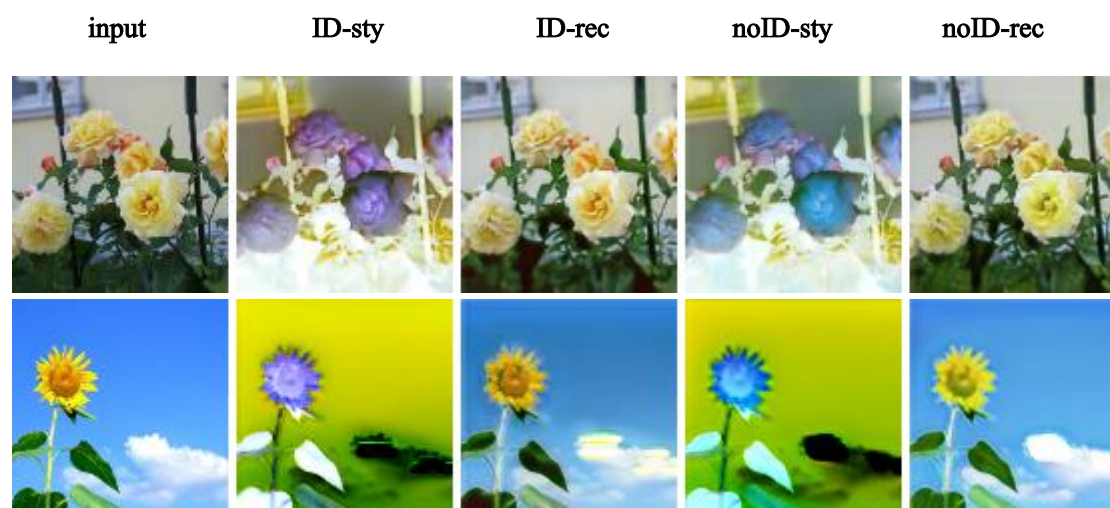
附录



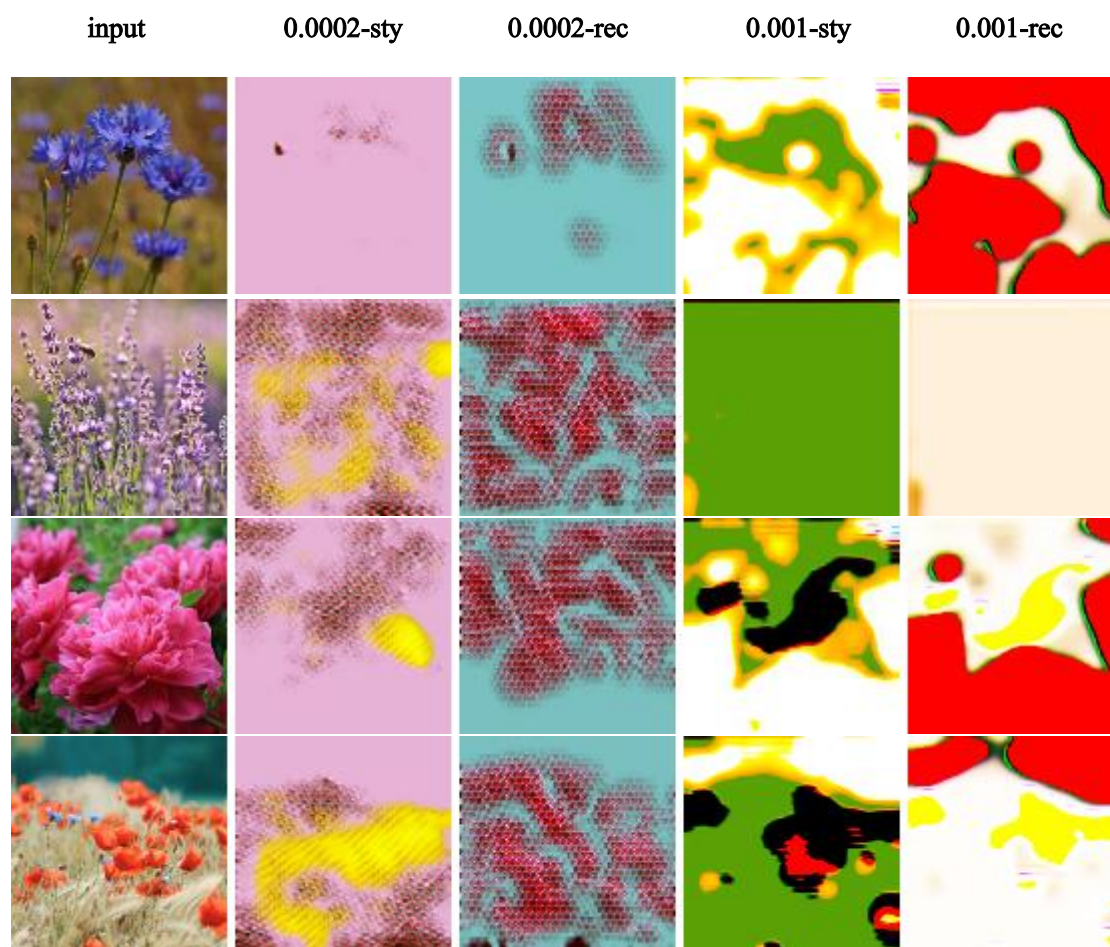


附图 1 CNN 与 CycleGAN-qp(包含身份损失函数), 从左往右依次是: 作为原始输入图像的自然花图像、用于 CNN 方法进行风格迁移时的水彩画风格参考图像、CNN 风格迁移效果、CycleGAN-qp 风格迁移效果。





附图 2 消融实验，从左往右依次是：原始输入图像、含身份损失函数模型风格迁移效果、含身份损失函数重构图像效果、不含身份损失函数风格迁移效果、不含身份损失函数重构图像效果。



附图 3 优化器 Adabound 不同学习率学习效果，从左往右依次是：原始输入图像、学习率为 0.0002 的风格迁移效果、学习率为 0.0002 的重构图像效果、学习率为 0.001 的风格迁移效果、学习率为 0.001 的重构图像效果，由图中结果可知，无论是风格迁移还是重构效果，使用新的优化器 Adabound 效果都很差。



附图 4 不包含身份损失函数的 CycleGAN-qp 风格迁移效果, 输入图像为西南财经大学微信公众号提供的西财壁纸, 突显出了训练出来的模型分割迁移规则里把绿色映射成了蓝色。