

中山大学本科学位论文

融合了半监督机制的生成式对抗网络

Semi-supervise Generative Adversarial Network

学位申请人： 吴佳卫

指导老师： 饶洋辉 副教授

专业名称： 软件工程

答辩委员会主席（签名）： _____

答辩委员会成员（签名）： _____

中山大学数据科学与计算机学院

二零一八年四月一日

表一：毕业论文（设计）开题报告
Form 1: Research Proposal of Graduation Thesis (Design)

论文（设计）题目 Thesis (Design) Title: 基于迁移学习对 GAN 进行优化预训练
（简述选题的目的、思路、方法、相关支持条件及进度安排等） （Please briefly state the research objective, research methodology, research procedure and research schedule in this part.）
<h2 style="margin-top: 0;">一、 论文选题意义及创新点</h2> <p>近年来生成式对抗神经网络（GAN）是目前深度学习研究中最活跃的领域之一，因为其算法生成的结果非常的真实。GAN 的应用领域比较广泛，较为熟知的有图像和视觉领域[1]、语音和语言领域[3]、信息安全、棋类比赛等。GAN 仍然有着巨大的研究前景与更多的应用空间。</p> <p>GAN 作为一种生成式算法，有效的解决了可建立自然性解释的数据的生成难题[3]。GAN 不仅可以生成处理高维度的数据，还大大改善了生成式模型的训练难度和训练效率。</p> <p>目前主流有很多基于 GAN 改进和优化的算法在不同的领域都有着很好的表现。主要有以下几个大的方面：条件生成式对抗网络[4]、双向生成式对抗网络[5]、自编码生成式对抗网络[6]、组合生成式对抗网络[7]等。</p> <p>但是根据我自己的实践，利用 tensorflow 单机多 GPU 框架对 GAN 训练对计算机的 GPU 有着相当高的要求。如果不适用多 GPU 进行训练则会导致所需要的时间倍数地增加。因此我希望可以引入有关生成式对抗神经网络的优化预训练机制，使一些较低端的 GPU 也可以运行生成式对抗神经网络。</p> <p>至于如何进行对生成式对抗神经网络的优化预训练，在这里希望引入迁移学习（Transfer Learning）的机制，共有下面几种可行的方案：样本迁移（Instance based）、特征迁移（Feature based）、模型迁移（Parameter based）以及关系迁移（Realtion based）。</p> <p>本文提出新模型主要创新点如下：</p> <ol style="list-style-type: none"> 1. 可以提高 GAN 的训练效率； 2. 可以使用通用化的模型来训练更加具有个性化的数据。 <h2 style="margin-top: 20px;">2.1 二、国内、外相关文献掌握程度和研究方法的可靠性</h2> <h3 style="margin-top: 10px;">2.1. 国内外研究现状</h3> <p>首先 Yosinski J 和他的团队在论文[14]中探讨了神经网络的可迁移性，文中讨论了对于深层神经网络那些属性是可以被迁移的，论文中分析了最佳的迁移策略以及可能遇到的问题。</p> <p>VGGNet[15]是牛津大学计算机视觉组研发的的深度卷积神经网络，其中就引入了迁移学习的概念。VGGNet 探索了卷积神经网络的深度与其性能之间的</p>

关系, 可用来在特定的图像分类任务上进行再训练(相当于提供了非常好的初始化权重)。

Anderson 在他们的文章[16]中则是引入了一种通过学习预测状态动态来进行预训练的方法。

Abtahi 和他的团队提出了一种使用非监督学习来学习网络中潜在特征的预训练方法[17]。

2.2. 研究方法的可靠性

已经有很多迁移学习对神经网络与卷积神经网络优化的案例[11]。而生成式对抗神经网络的本质就是两个互相博弈的神经网络, 因此与之前的研究是有相通点的。

2.2 三、论文整体设计的科学性

3.1. 理论基础

主要依赖的是迁移学习的理论基础。

3.2. 实验依据

可以从两个方面来衡量算法:

第一个方面是时间效率的提高, 也就是面对同一个任务从零开始训练生成式对抗神经网络与通过一定的迁移借助已经训练好的网络来训练生成式对抗神经网络所需要的时间。

第二个是 GAN 本身的表现, 由于现在没有一个很好的标准可以来衡量 GAN 的表现, 有一些模型会引入通过帕尔森窗口法对 GAN 进行评估[12][13]。

2.3 四、存在的主要不足

1. 因为迁移学习机制的引入很有可能会影响 GAN 本身的表现;
2. 方法也许并没有足够的鲁棒性。

2.4 五、进度安排

1. 确定研究方向: 阅读文献, 确定研究方向, 提交毕业论文开题报告。(2018 年 11 月)
2. 前期准备: 整理和收集数据集, 确定实验思路。(2018 年 12 月)
3. 中期过程:
 - i. 根据思路, 实现模型, 做实验。(2018 年 12 月)
 - ii. 根据各方面实验结果, 总结不足, 改进模型。(2019 年 1 月)
 - iii. 确定最终模型和实验结果, 开始撰写论文。(2019 年 2 月)
 - iv. 中期检查, 完成初稿。(2019 年 3 月)
4. 后期工作: 修改论文, 定稿。(2019 年 4 月)

参考文献:

- [1] Ledig C, Wang Z, Shi W, et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network[C]// Computer Vision and Pattern Recognition. IEEE, 2017:105-114.
- [2] Li, Jiwei, Monroe, Will, Shi, Tianlin, et al. Adversarial Learning for Neural Dialogue Generation[J]. 2017.
- [3] 王坤峰,苟超,段艳杰,林懿伦,郑心湖,王飞跃.生成式对抗网络 GAN 的研究进展与展望[J].自动化学报,2017,43(03):321-332.
- [4] Mirza M, Osindero S. Conditional Generative Adversarial Nets[J]. Computer Science, 2014:2672-2680.
- [5] Donahue J, Krähenbühl P, Darrell T. Adversarial Feature Learning[J]. 2016.
- [6] Larsen A B L, Larochelle H, Winther O. Autoencoding beyond pixels using a learned similarity metric[J]. 2015:1558-1566.
- [7] Wang Y, Zhang L, Joost V D W. Ensembles of Generative Adversarial Networks[J]. 2016.
- [8] Li Y, Shen L. cC-GAN: A Robust Transfer-learning Framework for HEp-2 Specimen Image Segmentation[J]. IEEE Access, 2018, PP(99):1-1.
- [9] Kim M., Zuallaert J. and De Neve W. (2017). Towards Novel Methods for Effective Transfer Learning and Unsupervised Deep Learning for Medical Image Analysis . In Doctoral Consortium - DCBIOSTEC, (BIOSTEC 2017) ISBN , pages 32-39
- [10] Daniel De Freitas Adiwardana, Akihiro Matsukawa, Jay Whang. Using Generative Models for Semi-Supervised Learning
- [11] Sinno Jialin Pan, Qiang Yang. A Survey on Transfer Learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(10):1345-1359.
- [12] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]// International Conference on Neural Information Processing Systems. MIT Press, 2014:2672-2680.
- [13] Mirza M, Osindero S. Conditional Generative Adversarial Nets[J]. Computer Science, 2014:2672-2680.
- [14] Yosinski J , Clune J , Bengio Y , et al. How transferable are features in deep neural networks?[J]. Eprint Arxiv, 2014, 27:3320-3328.

- [15] Karen Simonyan, Andrew Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition
- [16] Anderson C W , Lee M , Elliott D L . Faster reinforcement learning after pretraining deep networks to predict state dynamics[C]// International Joint Conference on Neural Networks. IEEE, 2015.
- [17] Abtahi F , Fasel I . Deep belief nets as function approximators for reinforcement learning[C]// Aaai Conference on Lifelong Learning. AAAI Press, 2011.

Student Signature:

Date:

指导教师意见

Comments from Supervisor:

1.同意开题

2.修改后开题

3.重新开题

1.Approved()

2. Approved after Revision ()

3. Disapproved()

Supervisor Signature:

Date:

表二：毕业论文（设计）过程检查情况记录表
Form 2: Process Check-up Form

指导教师分阶段检查论文的进展情况（要求过程检查记录不少于 3 次）
The supervisor should check up the working process for the thesis (design) and fill up the following check-up log. At least three times of the check-up should be done and kept on the log.

第 1 次检查 (First Check-up) :

学生总结
Student Self-summary:

指导教师意见
Comments of Supervisor:

第 2 次检查 (Second Check-up):

学生总结
Student Self-summary:

指导教师意见
Comments of Supervisor:

第 3 次检查 (Third Check-up):

学生总结

Student Self-summary:

指导教师意见

Comments of Supervisor:

第 4 次检查

Fourth Check-up

学生总结

Student Self-summary:

指导教师意见 (Comments of Supervisor):

学生签名 (Student Signature):

日期 (Date):

指导教师签名 (Supervisor Signature):

日期 (Date):

总体完成情况

(Overall
Assessment)

指导教师意见 Comments of Supervisor:

- 1、按计划完成，完成情况优 (Excellent): ()
- 2、按计划完成，完成情况良 (Good): ()
- 3、基本按计划完成，完成情况合格 (Fair): ()
- 4、完成情况不合格 (Poor): ()

指导教师签名 (Supervisor Signature):

日期 (Date):

论文原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

学位论文作者签名：

日期： 年 月 日

学位论文使用授权声明

本人完全了解中山大学有关保留、使用学位论文的规定，即：学校有权保留学位论文并向国家主管部门或其指定机构送交论文的电子版和纸质版；有权将学位论文用于非赢利目的的少量复制并允许论文进入学校图书馆、院系资料室被查阅；有权将学位论文的内容编入有关数据库进行检索；可以采用复印、缩印或其他方法保存学位论文；可以为存在馆际合作关系的兄弟高校用户提供文献传递服务和交换服务。

保密论文保密期满后，适用本声明。

学位论文作者签名：

日期： 年 月 日

导师签名：

日期： 年 月 日

摘要

作为近年来被提出的最有前景的无监督深度学习方法之一，生成式对抗网络算法（Generative Adversarial Networks，简称 GAN）广受学者们的关注。因为其网络可以生成出非常真实的结果，因此 GAN 在计算机视觉和图像、语音和语言、信息安全等领域都有着巨大的研究前景。

而 CycleGAN 算法作为近年来在图像处理领域被提出的颇有影响力的生成式对抗网络算法之一，不仅继承了传统 GAN 算法结构清晰效果真实的优点，更是提供了直接将一类图片转换为另一类图片而不需要建立一一对应关系的强大功能。

然而，传统的 CycleGAN 算法由于网络结构庞大，训练成本和开销都非常巨大。对于普通电脑配置的研究者来说自己完整训练一个 CycleGAN 网络是难以实现的。与此同时 CycleGAN 作为一种无监督的深度学习方式，其生成结果是不可控的。CycleGAN 网络在生成图片的过程中没有办法区分任务的主体对象，因此往往也会造成生成结果的背景部分被改变的情况。

目前解决第一个问题的思路包括直接使用预训练好的 CycleGAN 模型或者自主降低数据集的规模。第二个问题的解决思路是在一些特定的任务中引入注意力的机制。但是我们发现这些解决方案都没有办法很好的从根本上解决普通计算机用户希望自主设计训练一个 CycleGAN 网络的问题，使用预训练模型和降低数据规模都是与实际需要背道而驰的，而针对任务设置注意力机制复杂且增加了训练成本和难度。

本文提出了一种引入了半监督学习机制的 CycleGAN 训练方案，通过对数据集的分割和分类处理让网络可以专注于学习任务主体目标的特征，在更有效地利用计算机资源的同时为解决生成结果不可控的问题提供了一种解决思路。

关键词：生成式对抗网络，半监督学习，图像转化

Abstract

Generative Adversarial Networks(GAN), being regarded as one of the most promising structure in un-supervised learning area, has being drawing attentions continuously from scholars in the recent years. Because of the fact that the model can generate realistic results, GAN has great research prospects in the studies of computer vision, voice and language, information security and other fields.

As one of the most influential GAN algorithms proposed in images processing filed these years, GycleGAN not only has the advantages of generating realistic images as GAN, but also can transform one kind of images into another without one-to-one correspondence.

However, because traditional CycleGAN has a relatively large and complex network structure, the cost and requirement of training it is correspondingly high. It is hard for researchers with a normal computer to train a CycleGAN model. Meanwhile, as a un-supervised learning model, the generating result of CycleGAN is uncontrollable. That is to say, a traditional CycleGAN model cannot distinguish the background and target in an image in training and generating processes. As the result of it, it is common that the generating results of CycleGAN contain images which background is obviously changed.

By now the common way to solve the first problem is to use pre-trained versions of CycleGAN models or simply lower the size of datasets. And the solution to the second problem is to use attention mechanism in specific tasks. However, we find out that the solutions mentioned above cannot solve the basic requirement that enabling the user to design and train a CycleGAN model using their personal computers. Using pre-trained model or lower the size of the datasets will not help and using attention mechanism would make the model more complex, which would increase the training difficulty.

This paper proposes a method using semi-supervised mechanism to train a CycleGAN model. By cutting up images into pieces containing certain information, the method enables the network to learn from the target of the tasks. With this method, we find a way to lower the training requirement as well as improving the performance

in a certain way.

Keywords: Generative Adversarial Networks, semi-supervised learning, image converting

目录

摘要	9
第一章. 引言	13
1.1 研究背景和意义.....	13
1.2 本文的工作.....	14
1.3 论文结构简介	15
第二章. 相关工作综述	16
2.1 生成式对抗网络.....	16
2.2 生成式对抗网络优化训练.....	16
第三章. 融合了半监督机制的生成式对抗网络	18
3.1 问题定义.....	18
3.2 模型描述.....	18
3.3 参数设置.....	25
第四章. 实验结果与分析	27
4.1 数据集介绍.....	27
4.2 运行模型.....	28
4.3 实验结果及分析.....	33
第五章. 总结与展望	39

第一章. 引言

1.1 研究背景和意义

随着计算机计算能力的不断提高,近年来深度学习领域发展迅速。其中图像生成作为计算机视觉领域一个重要的研究方向一直广受关注。今年备受瞩目的变分自编码器 (Variational Auto-Encoder, 简称 VAE) [7]和生成式对抗网络 (Generative Adversarial Networks, 简称 GAN) [1]是图像生成模型中最有发展潜力的两个模型。两个模型都希望找到或接近目标样本的分布,将一个已知可控的分布通过某种方法映射到位置的目标分布上,从而就可以无限制的生成该样本。

而 CycleGAN[4]作为 GAN 大框架下的一种模型可以将来自两个不同域之间的模型进行互相转化,并且不要求训练集两个域的图像样本有一一对应的关系。这个特性使得 CycleGAN 可以被用在图像转化(猫狗图片转化),风格迁移(风景画转换莫奈画风),自定义图像生成(建筑物图像生成),图像语义分割等方面。虽然 CycleGAN 模型的提出广为称道,但是我们不难看到其中的一些问题。

1. CycleGAN 模型中一共有四个卷积神经网络,包括两个图像的生成器以及两个判别器,训练他们需要消耗相当大量的计算机资源。因此对于一些配置比较普通的计算机来说,按照自己的期望训练一个 CycleGAN 网络通常是非常困难的。
2. CycleGAN 作为 GAN 框架下的一种无监督图像生成算法,不可避免地会有生成结果不可控的问题。再加上 GAN 算法本身的约束条件非常少,使这个问题更加地明显。

因此本文希望借助图像分割的技术以及半监督学习[8]的手段来对以上的两个问题提供一个解决思路。

1.2 本文工作

本文主要专注于解决 CycleGAN 模型的两个问题：第一、训练成本过高，在一般配置计算机上实现困难。第二、生成结果存在不可控的现象，没有办法专注于主体目标的训练。

为了针对上述两个问题提供可行的解决方案，本文提出了一种融合图像分割思想和的半监督思想 CycleGAN 模型优化方案。本文首先是使用了一种针对数据集的图像分割及融合方法，将训练集和测试集的图像分割成若干比较小的图像便于后续模型的处理。使原本对于计算机来说比较庞大的数据集图片可以被分割成便于处理的小块图像来经行训练和生成。再最后将训练集对应结果的小块图像通过图像融合的手段还原较大尺寸的图像，从而获取最终的结果。值得一提的是，这种分割训练集图像的做法也同时起到了扩充训练数据集规模的作用，使 CycleGAN 模型的训练拥有更多可用的样本。其次，我们针对数据集训练了一个基于卷积神经网络的分类器，可以有效地分辨出经过分割后的图像中是否含有对 CycleGAN 网络有用的信息。在 CycleGAN 网络的训练过程中，只有分类器认为有价值的小块图像才会被用于训练以及计算损失函数。在 CycleGAN 网络运行测试集生成最终结果的过程中，只有分类器认为属于主体目标的小块图像才会被输入生成器网络中生成最终的结果。这个分类器的引入精简了训练集与测试，使得 CycleGAN 网络可以以相对较高的效率完成训练以及测试，节约了时间与计算机资源。与此同时分类器也帮助 CycleGAN 网络判断测试集图像中的哪些部分属于主体目标，使其学会通过修改一张图片中的关键部分而并非整张图像来完成任务。因此网络的生成结果也会变得更加可控。

总体来说本文一共有以下三点主要贡献：

1. 提出了一种训练 CycleGAN 的方法，使 CycleGAN 的训练成本大大降低。
2. 通过训练一个分类器，使 CycleGAN 可以通过忽略图片背景的方式专注于对前景（目标）的训练，从而更有效率的学习到目标特征。
3. 分割原始数据集的方法一定程度上也起到了扩充训练集的作用。

1.3 论文结构

本文接下来的结构安排如下：第二章中，我们将概述国内外相关领域的研究现状；第三章中，我们将详细描述我们模型的框架与细节；第四章中，我们将对实验结果进行展示与分析；第五章中，我们将进行总结与展望。

第二章. 相关工作综述

在这一章中，我们将首先概述与生成式对抗神经网络有关的一些工作，尤其是在图像生成领域的一些重要文章。其次我们会归纳一些优化或者加速生成式对抗网络的方法并且列举这些方法下的一些相关研究。

2.1 生成式对抗网络

生成式对抗网络（GAN）的框架于 2014 年被 Goodfellow 等人提出[1]，该框架通过使两个互相对抗的神经网络在训练的过程中互相博弈从而可以最终得到非常理想的生成效果。作者通过在一些传统图像生成任务（MNIST 手写数字集，Toronto 人脸数据集等）上的实验，证明了 GAN 的框架可以在图像生成任务中取得很好的结果。

在随后的研究中，越来越多基于 GAN 框架针对图像生成任务的算法被提出。Conditional GAN[9]算法通过在输入每一张图片时拼接一个额外信息 y ，给 GAN 算法添加了约束，我们可以通过控制 y 的值来约束网络的生成结果。Invertible Conditional GANs[10]算法在这个基础上将原本生成器的噪声输入改为了一个编码器的生成结果。作者认为这个编码器可以从相似的图片中提取出带有图片特征的向量表达，从而代替原本的噪音作为生成器的输入。StarGAN[11]则是一个能够学习多个域之间映射的生成对抗网络。在 StarGAN 中，只使用一个生成器和一个鉴别器就可以学习多个域之间的映射，做到同一个网络对多个属性的生成和判别。

最后是我们模型基于的 CycleGAN[4]算法。CycleGAN 区别于传统的 GAN，通过训练两个判别器和两个共用的生成器，实现了两个域的图片互相生成。与此同时，这种算法引入了循环一致性损失，因此一次训练中不需要两个域的图片严格一一对应。

2.2 生成式对抗网络优化训练

由于传统的生成式对抗网络框架在训练的时候可能会产生的种种问题，一些

针对于解决这些不同方面的问题的优化方案或者模型也被研究者们陆续提出，大致可以分为以下几个方面。

第一个大方向是加速生成器和判别的两个神经网络的收敛速度。例如 Wasserstein GAN[12]算法通过重新定义 GAN 的损失函数解决了原始 GAN 框架训练过程中模式崩坏(mode collapse)以及难以收敛的问题。在此之后, LSGAN[13]借助了最小二乘的思想, 使生成器可以更快地向真实分布的方向收敛。在 Improved Techniques for Training GANs[14]这篇文章中, 作者提出了几种思路: 优化目标函数、使用 minibatch 方法训练、添加惩罚项、标签平滑和 virtual batch normalization 方法。

第二个大方向是优化生成器的生成结果。DCGAN[15]算法将 CNN 与 GAN 结合起来, 将 GAN 作为特征的提取器并且结合鉴于学习领域的方法, 从而让 GAN 的生成器和判别器都学习到了丰富的层次表达。而发表于 2018 年的 GANimation[6]算法, 通过引入解剖学中动作单元 (Action Unit) 的概念并训练一个注意力机制, 可以做到更加柔和的表情生成过程, 同时也避免了原始输入图片除了表情以外的成分 (例如背景, 头发等) 被算法修改, 在表情生成这一领域达到了非常好的效果。Improved Techniques for Training GANs 文章的作者也认为传统 GAN 在图像生成任务下缺乏一个明确的目标函数, 因此提出了 MTurk (邀请人来判别) 和使用 inception score 指标两种思路。

第三个大方向是引入半监督机制。Tim Salimans[14]等人将生成器的输出和真实样本理解为不同的类, 把判别器替换为分类器, 从而引入了监督学习的成分, 把 GAN 转换为了一种半监督学习。

最后不可忽略的是, 随着预训练正在深度学习领域中发挥着越来越大的作用, 一些预训练的 GAN 模型也被公布出来。这些模型不需要用户花费大量的时间和计算力来训练模型就可以直接观察到最终模型的生成结果。比如最近 Hugging Face 团队公布的预训练的 Google BigGAN[16]模型, 让对于普通人也可以接触到对他们来说难以训练的 BigGAN 模型。

第三章. 融合了半监督机制的生成式对抗网络

在本章中，我们首先明确模型需要完成的任务。然后再分步详细介绍本文提出的融合了半监督机制的生成式对抗网络模型。最后，我会对模型训练过程中涉及到的参数取值进行定义。

3.1 问题定义

假设我们有来自两个比较相似的域的图片（例如苹果和橙子、马和斑马等），模型通过学习来自两个域的图像的特征，最终的目标是生成器可以将来自一个域的图像转换为属于另一个域的图像。以苹果转换为橙子的任务（以下简称为 `apple2orange`）为例，最终的模型需要读入来自测试集的苹果的图片，并输出将该苹果转换为橙子后的图像。在整个模型的训练以及转换的过程还有以下两点需要满足。

首先，这个模型是一般配置的计算机可以运行的。本文使用 `python` 语言以及 `tensorflow` 单机多 GPU 框架来实现这个项目。模型的实验将在常用的笔记本电脑（显卡配置为英伟达 GTX970M 3G）上进行。如果模型是可以完整运行的，我们就认为一般配置的计算机也具有运行该模型的能力。第二个需求是，模型在转化图片的过程中需要区分主体和背景。以 `apple2orange` 任务为例，当模型将一张图片中的苹果转化为橙子的时候需要保证图片中的其他元素最大程度的不变。

3.2 模型描述

本小节中，我将重点介绍我提出的融合了半监督机制的生成式对抗网络。该模型的主体框架如图 3-1 所示，它主要有四个部分组成：图片的分割，CNN[17] 分类器，CycleGAN 网络以及图像的拼接与融合。输入模型的原始图片会先经过分割变成众多小块的图像，随后分类器会将包含任务目标的小块图像输入 CycleGAN 网络。最后 CycleGAN 网络的输出结果会和包含背景的小块图像拼接融合得到最后的结果。接下来我会对模型的四个部分进行详细的描述。

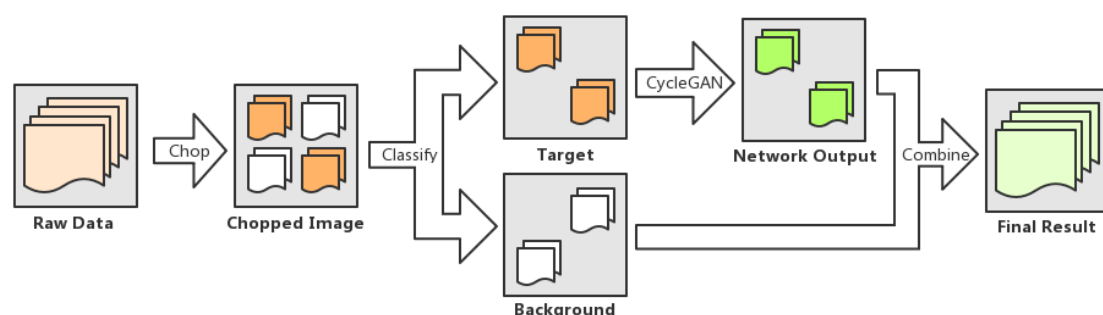


图 3-1 模型主体框架图

i. 图像分割

为了方便将原始数据转换为后续算法可以处理的大小，模型的第一步是将原始数据集的每一张图片均匀分割。分割算法会将一张图片分为相同大小的 16 张图片，由于考虑到最后在图片拼接和融合的过程中需要相邻的两张图片有相同的部分，因此每一张图片的四个方向上都额外再取像素的图像信息。连续的两次选取过程如图 3-2 所示，深红色与深灰色的框表示我们感兴趣的图像部分，浅红色和浅灰色的框表示实际选择的图像部分。

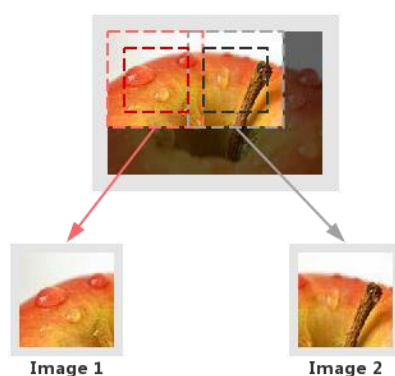


图 3-2 连续的两次图像分割

记我们感兴趣的部分大小为 $d \times d$ 像素，每一张分割得到的图片四个方向额外取 $\frac{d}{4}$ 像素的内容，那么最后实际取出的图像大小为 $d + 2 \cdot \frac{d}{4}$ 即 $\frac{3d}{2}$ 。注意到虽然图像的边缘不需要考虑融合的问题，但是为了保持一致在这里我们也做相同的

处理。假设分割前输入图像的大小为 $h \times h$ 像素，那么我们很容易找到 d 与 h 之间的关系，即 $4 \cdot d + 5 \cdot \frac{d}{4} = h$ ，解得：

$$d = \frac{4h}{21}, \quad (3.1)$$

为了保证在图像拼接及合成这一步骤可以合成出完成的图像，测试集图像在分割的时候要求完整连续，即连续的两张图片（包括上下连续和左右连续）之间必须要有 $\frac{d}{4}$ 像素的重合部分。与之相对的是，由于为后续的神经网络分割训练集的时候只需要保证分割的结果有足够的信息给网络学习，而不考虑还原的过程，因此在分割训练集的时候可以直接使用多次随机从图像中获取长宽均为 $\frac{4h}{21}$ 像素的图像即可，这样做也保证了训练集会有更加丰富的图像信息。经过分割的操作后，我们将图像转换为了包含原始图像部分信息的小块图像。由于分割出的每一块图像粒度都相对较小，因此我们可以做这样的假设：分割得到的每一块小图像只含有主体和背景的其中之一的信息。

ii. CNN 分类器

基于在上一步骤中提出的假设，我们可以认为分割原始数据集图片后得到的图片集合内只有两种类型的图片，即只包含主体内容的图片和只包含背景信息的图片。因此在这里我们训练一个基于 CNN 的图像分类器来判断上一步分割后的图片属于哪一类。图 3-3 展示了我们搭建的 CNN 图像分类器的结构。其中左侧输入为一张长宽均为 48 像素的三通道图像，输出为一个二维数组，用于预测输入图片所属类别。

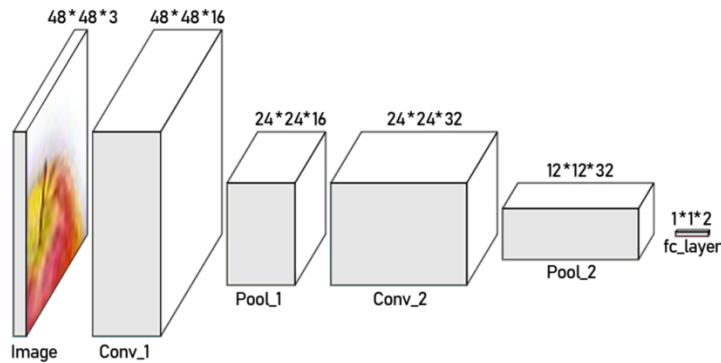


图 3-3 图像分类器网络结构

图片在网络中会分别经过两个卷积层和两个池化层，以此提取出图像的特征，最后经过一个全连接层得到预测的结果。关于网络中各层的说明与参数见表 3-4。

表 3-4 图像分类器网络结构

名称	类型	大小	特征图	卷积核	步长	激活函数
In	输入	48×48	3	-	-	-
Conv_1	卷积	48×48	16	4×4	1	-
Pool_1	最大池化	24×24	16	2×2	2	-
Conv_2	卷积	24×24	32	4×4	1	Relu
Pool_2	最大池化	12×12	32	2×2	2	-
fc_layer	全连接	1×1	2	-	-	-

在 CNN 的网络中，卷积层的作用是提取出上一层输入的图像特征。当图像中存在某个特征时，相应识别这种特征的神经元就会产生一个明显的输出。最大池化层的引入是为了减少卷积核的尺寸，同时保留图像中的特征。最后的全连接层是为了将之前网络中学习到的特征组合起来，便于最后的分类。

最后，被判定包含主体目标的图片会被输入下一步骤中的 CycleGAN 网络进行转化，而被判定为背景的图片会保持不变。

iii. CycleGAN 网络

上一步骤中的 CNN 分类器为我们输出了包含两个域的主体目标的图像，因此在这一步中，我们将训练一个 CycleGAN 网络来学习如何将两个域之间的图像互相转化。

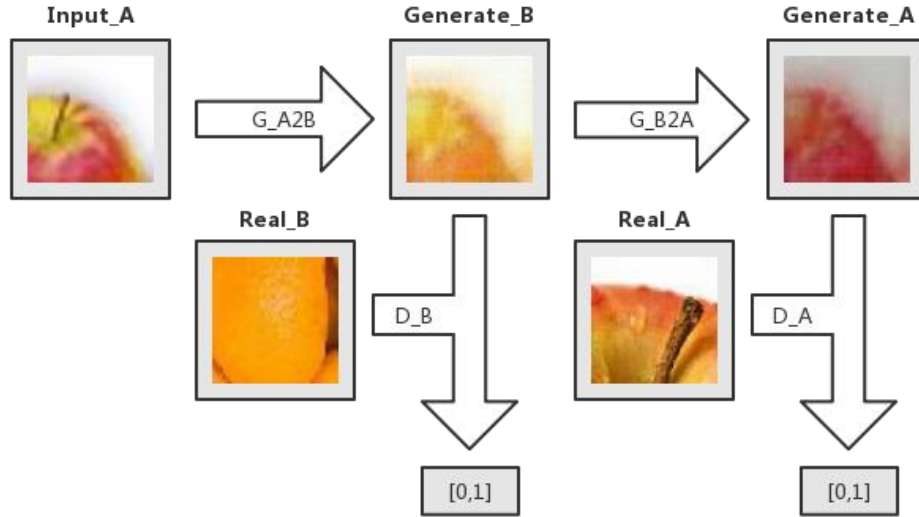


图 3-5 CycleGAN 网络结构 (X to Y)

图 3-5 展示的是 CycleGAN 由 X 类图片生成 Y 类图片的网络结构图（以苹果转换为橙子为例）。网络中一共有四个神经网络, 包括两个生成器和两个判别器。左侧的输入是 X 类的图片输入, 记为 x 。 x 经过第一个生成器 $G: X \rightarrow Y$ 后将输入的图片转换为属于 Y 类的图片, 生成结果为 $G(x)$ 。生成的结果会被输入一个判别器 D_Y , 判断其生成结果是否满足 Y 类图片的特征。接下来作者引入了图像重构的操作, 将上一步骤生成的结果输入第二个生成器 $F: Y \rightarrow X$ 得到第二次的生成结果 $F(G(x))$ 。经过以上的步骤, 我们可以认为重构后的图像 $F(G(x))$ 是接近输入的原始图像 x 的, 可以理解为 $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$ 。重构得到的图片 $F(G(x))$ 最后会被输入第二个判别器 D_X , 判断生成结果的好坏。从 Y 类图片生成 X 类图片的原理与 X 类生成 Y 类图片是一样的, 两个过程共享相同的生成器和判别器。

CycleGAN 的损失函数一共由三个损失函数相加组成, 其表达式如下:

$$\mathcal{L}(G, F, D_X, D_Y) = \mathcal{L}_{GAN}(G, D_Y, X, Y) + \mathcal{L}_{GAN}(F, D_X, Y, X) + \lambda \mathcal{L}_{cyc}(G, F), \quad (3.2)$$

其中 $\mathcal{L}_{GAN}(G, D_Y, X, Y)$ 对应的是将 X 类的图片转换为 Y 类图片的过程中生成器 $G: X \rightarrow Y$ 和判别器 D_Y 的损失函数, $\mathcal{L}_{GAN}(G, D_X, Y, X)$ 对应的是将 Y 类的图片转换为 X 类图片的过程中生成器 $G: Y \rightarrow X$ 和判别器 D_X 的损失函数, $\mathcal{L}_{cyc}(G, F)$ 则对应的是重构图像的损失函数。三个损失函数的具体的表达式如下:

$$\mathcal{L}_{GAN}(G, D_Y, X, Y) = \mathbb{E}_{y \sim p_{data}(y)} [\log D_Y(y)] + \mathbb{E}_{x \sim p_{data}(x)} [\log(1 - D_Y(G(x)))], \quad (3.3)$$

$$\mathcal{L}_{GAN}(F, D_X, Y, X) = \mathbb{E}_{x \sim p_{data}(x)} [\log D_X(x)] + \mathbb{E}_{y \sim p_{data}(y)} [\log(1 - D_X(F(y)))], \quad (3.4)$$

$$\mathcal{L}_{cyc}(G, F) = \mathbb{E}_{x \sim p_{data}(x)} [\|F(G(x)) - x\|_1] + \mathbb{E}_{y \sim p_{data}(y)} [\|G(F(y)) - y\|_1], \quad (3.5)$$

其中, 式 3.2 基于交叉熵公式, 使用期望的形式表示了生成器 $G: X \rightarrow Y$ 及判别器 D_Y 的损失函数。在训练这部分网络的时候先固定生成器 $G: X \rightarrow Y$, 尽可能地让判别器 D_Y 能够最大化地判别出样本来自于真实数据 y 还是生成的数据 $G(x)$ 。然后固定判别器 D_Y 训练生成器 $G: X \rightarrow Y$, 要求生成器 $G: X \rightarrow Y$ 能够最小化真实样本 y 与生成样本 $G(x)$ 的差异。训练这一部分网络就是对判别器最大化损失函数而对生成器最小化损失函数, 记为 $\min_G \max_{D_Y} \mathcal{L}_{GAN}(G, D_Y, X, Y)$ 。式 3.3 与 3.2 原理相同, 我们通过 $\min_F \max_{D_X} \mathcal{L}_{GAN}(F, D_X, Y, X)$ 来训练生成器 $F: Y \rightarrow X$ 以及判别器 D_X 。式 3.4 计算的是图像以两个方向通过网络并重构的损失, 即 X 类图像 x 先后通过 $G: X \rightarrow Y$ 、 $F: Y \rightarrow X$ 后与 x 的差与 Y 类图像 y 先后通过 $F: Y \rightarrow X$ 、 $G: X \rightarrow Y$ 后与 x 的差。通过计算循环一致性损失可以保证图片在转换的同时, 图中物体的几何形状和空间关系不发生变化。

iv. 图像的拼接与融合

原始图像经过分割后, 包含主体目标的图像被输入 CycleGAN 网络进行处理, 而包含背景的图像则保持不变。最后一步, 我们将这些图像拼接并融合得到最后的结果。

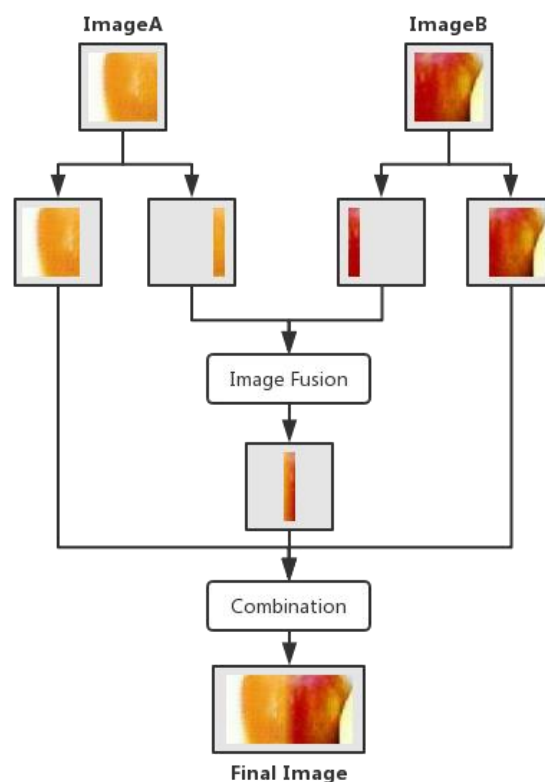


图 3-6 图像拼接与融合算法

图 3-6 展示的是连续两张图像拼接的方法。由于在图像分割的步骤我们在每一张分割后的图像的上下左右四个方向额外取了 $\frac{h}{21}$ （其中 h 指代原始图像的长宽）像素的图像，因此在拼接时连续两张图像（包括上下连续和左右连续）有 $\frac{h}{21}$ 像素的相同部分。我们希望用这样的原理来融合两张图像：重叠部分的图像由第一张逐渐过渡到第二张。

因此我们定义表 3-7 与表 3-8 中的图像融合代码，输入为重叠部分的两张图像，输出为融合的结果。当拼接左右两张连续图像时，第一张图像每一列的像素逐渐过渡到第二张图像；当拼接上下两张连续图像时，第一张图像每一行的像素逐渐过渡到第二张图像。

表 3-7 图像融合伪代码 1

Algorithm 1 Image Fusion Horizontal

Input: Image1, Image2

Output: Fusion_Result

1: **for** $col \in [0, h/21]$ **do**

2: $\alpha = 21 \cdot col / h$ $\alpha = 21 \cdot col / h$

3: $Fusion_result[col] = (1 - \alpha) \cdot Image1[col] + \alpha \cdot Image2[col]$

4: **end for**

表 3-8 图像融合伪代码 2

Algorithm 2 Image Fusion Vertical

Input: Image1, Image2

Output: Fusion_Result

1: **for** $row \in [0, h/21]$ **do**

2: $\alpha = 21 \cdot row / h$

3: $Fusion_result[row] = (1 - \alpha) \cdot Image1[row] + \alpha \cdot Image2[row]$

4: **end for**

为了证明这种融合算法的有效性，我们选取了两张连续但是差异较大的图像来做实验。见图 3-8，左侧的图像时使用了图像融合算法的拼接结果，右侧的图像是没有使用融合算法的拼接结果。不难看出融合算法的使用可以使两张图像之间的边缘变得更加平滑，在一定程度上增加了生成结果的真实性。



图 3-8 两种方式对比

3.3 参数设置

本文的模型中，共有 6 个参数需要调节。首先是输入图像的大小 h ，在这里我们设置 h 的值为 168。因此每一张分割后的图像大小为 48×48 像素，包括 32×32 像素的目标区域与四个方向额外取的 48×8 像素图像。第二是 CNN 分类

器相关的参数，在训练过程中我们设置 CNN 分类器的 $batch_size=80$ ， $learning_rate=1\times 10^{-4}$ 。最后是训练 CycleGAN 过程中的参数，训练的 $batch_size=1$ ， $learning_rate=2\times 10^{-4}$ ，损失函数中的 $\lambda=15$ 。

第四章. 实验结果与分析

在这一章中，我们将运行模型进行实验并且对最终的实验结果进行分析。首先，我们会介绍实验中使用到的数据集以及其基本信息；其次，我们运行实验并且观察模型每个部分的运行结果；最后，我们设置对照试验，并且对实验结果进行分析，从而论证提出的模型的有效性。

4.1 数据集介绍

为了验证模型可行性，我们选取了两组在图像风格迁移领域常用的数据集进行实验，有关于两组数据集的介绍如下：

apple2orange 与 summer2winter 都是常用于图像分类及迁移任务的数据集，apple2orange 包含苹果与橘子两类图像数据；而 summer2winter 包含夏天与冬天两类图像数据。关于两个数据集具体的参数以及样例见表 4-1 至表 4-4。

表 4-1 数据集整体特征

数据集名称	数据类型	属性数	实例数	相关任务
apple2orange	图像数据	N/A	2528	分类、图像风格迁移
summer2winter	图像数据	N/A	2740	分类、图像风格迁移

表 4-2 数据属性描述

数据集名称	文件格式	宽	高	分辨率	色彩通道
apple2orange	JPG	128 像素	128 像素	96 dpi	RGB
summer2winter	JPG	128 像素	128 像素	96 dpi	RGB

表 4-3 apple2orange 数据实例





















Apple					
Orange					

表 4-4 summer2winter 数据实例

Summer					
Winter					

4.2 运行模型

运行模型一共分为四个步骤：分割数据集，训练并运行 CNN 图像分类器，训练并运行 CycleGAN 图像转换器，图像拼接与融合。在接下来的部分我会以 apple2orange 任务为例分别对这些步骤进行说明。

i. 分割数据集

原始数据集由四个部分组成，分别是两类图像的训练集与测试集。对于训练集我们只要按照第三章中的图像分割规则随机划分出足够多的数据集即可。在实验过程中我设置的是从一张图像中随机分割 6 次。经过分割后我们不仅获得了想要的小块图像，还将训练集的规模扩充为原来的 6 倍。而对于测试集需要严格遵守连续分割的规则，因此一张图片被均匀分割为 16 份，等待后续的处理。表 4-5 与表 4-6 分别展示了两类图像分割后得到的图像。

表 4-5 Apple 类图像分割实例







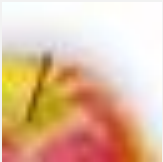
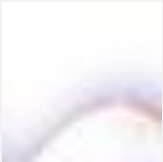
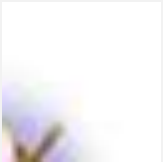


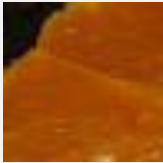



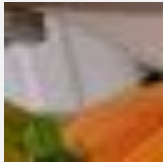
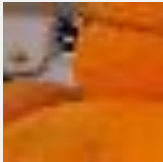

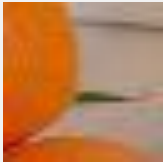

Train					
Test					

表 4-6 Orange 类图像分割实例

Train					
Test					

ii. 训练并运行 CNN 分类器

在这一步中训练 CNN 分类器的目标是使这个网络可以判断出分割得到的小块图像中是否包括 CycleGAN 需要的主体信息，因此我们需要先为 CNN 分类器手动划分出训练集。同时需要注意的是由于两类的图像会在特征信息上有着较大的差距，所以我们要针对两类训练出两个图像分类器。

我们分别从从上一步分割后的两类训练集图片中随机取出 1000 张图像并手动分割出 300 张包含主体目标的图片和 300 张不包含主体目标的图片用于训练 CNN 分类器。两个分类器模型在训练过程中分类正确率的变化如图 4-7 和 4-8 所示。

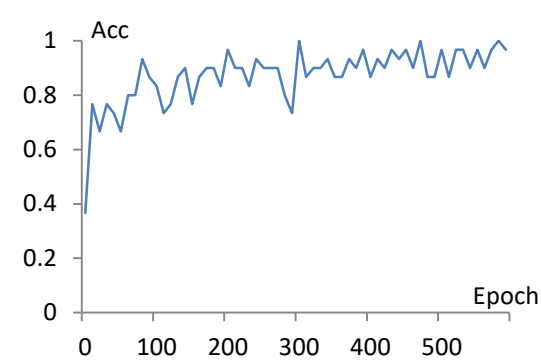


图 4-7 Apple 分类器准确率

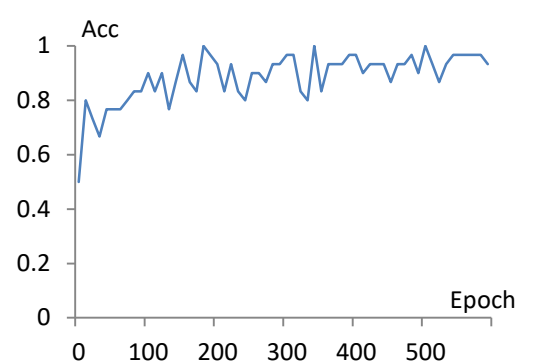


图 4-8 Orange 分类器准确率

随着迭代次数的增加，两个分类器的准确率逐渐从 0.5 上升至 0.9，并且在可接受范围内波动，因此我们可以认为模型经过迭代已经趋于收敛。

训练完成分类器后，我们将第一步分割得到的剩余的测试集以及全部的训练集导入分类器进行分类，表 4-9 和 4-10 展示了两类图像分类的部分结果。我们可以通过观测分类结果的准确性从而证明分类器的训练是成功的。

表 4-9 Apple 类图像分类实例

主体					
背景					

表 4-10 Orange 类图像分类实例

主体					
背景					

iii. 训练并运行 CycleGAN 转换器

我们取上一步 CNN 分类器认为是主体目标的两个类别的图像作为 CycleGAN 的训练集输入到模型中，并开始模型的训练过程。注意到 CycleGAN 框架中一共有四个神经网络，因此在图 4-11 到图 4-14 中我们会展示训练过程中四个神经网络损失函数的变化情况。

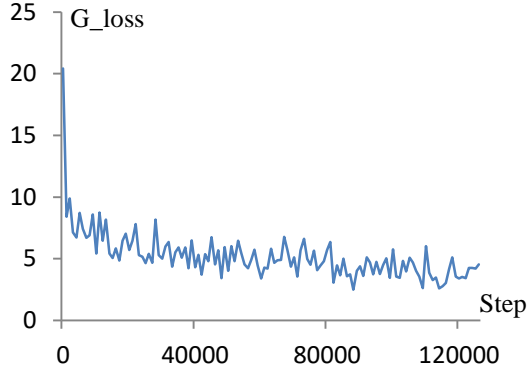


图 4-11 $G: X \rightarrow Y$ 损失函数

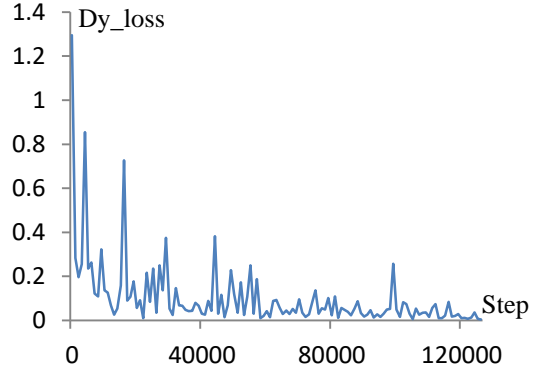


图 4-12 D_Y 损失函数

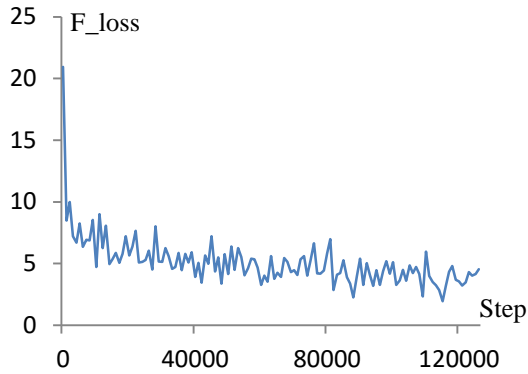


图 4-13 $F: Y \rightarrow X$ 损失函数

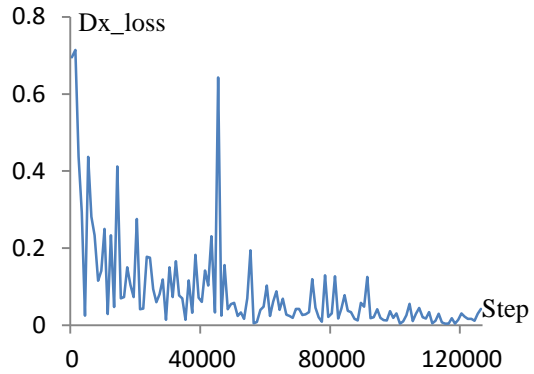






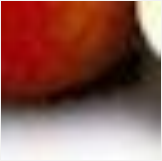




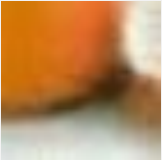
图 4-14 D_X 损失函数

我们观察四个损失函数随着训练次数增加的收敛情况，不难发现两个判别器 D_X 和 D_Y 的损失很快就收敛到 10^{-2} 级别，但是两个生成器 G 和 F 的损失函数则是最后收敛在 2.0 ~ 3.9 的范围内波动。可以看到生成网络和判别网络的训练是缺乏平衡性的，这也是目前生成式对抗网络模型存在的共同问题。我们可以得出结论，通过这种方法训练的 CycleGAN 可以以较快的速度收敛，且两个判别器 D_X 和 D_Y

可以达到比较理想的效果。

因此我们认为 CycleGAN 已经收敛，并且将测试集被分类为主体目标的图片输入网络进行转换，表 4-15 展示了部分的转换结果。

表 4-15 Apple 类图像转换结果实例

输入					
输出					

我们可以看到由于生成式对抗网络本身生成结果不可控的原因，同一张图像分割出的图片转换出的图片在色调和亮度上出现了一定的差距。但是因为 CycleGAN 框架中重构损失函数的引入，图片转换这个步骤没有改变图像本身的组成，甚至在一些部分保留了原有的细节。（表 4-15 中图像中原始图像的反光以及阴影都被很好地保留了下来）

iv. 图像拼接与融合

在最后一步，我们使用拼接与融合算法，将 CycleGAN 输出的包含主体目标的图片的转换结果与仅包含背景图片的图片拼接融合得到最终的图片。由于生成的图像中存在一定的噪点，因此引入一个简单的滤波器进行简单的降噪。表 4-16 展示了一些生成的结果。

表 4-16 Apple 类图像转换结果实例



4.3 实验结果及分析










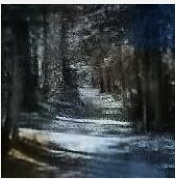
首先我们为了证明模型的有效性,为了对比得出结论我们一共运行了四组实验,关于每一组实验的模型配置与数据集选用详情见表 4-17。由于 summer2winter 数据集的转换目标是风景图片,因此不需要经过 CNN 分类器区分主体目标与背景。而对于 apple2orange 任务,我们分别进行三次独立的实验。实验二中我们使用完整的模型,即“图像分割-CNN 分类-CycleGAN 网络-图像融合”步骤来处理数据集;实验三中我们不经过图像分割与融合的操作,而是直接将数据集中的所有图像压缩为 48×48 像素的图片输入 CycleGAN 网络进行训练和测试;实验四中我们不使用 CNN 分类器,而是将经过分割的图像直接输入 CycleGAN 网络进行训练和测试。四组饰演的迭代次数与参数设置均相同。

表 4-17 实验配置

	数据集	经过图像分割	经过 CNN 分类器	迭代次数
实验一	summer2winter	是	否	126000
实验二	apple2orange	是	是	126000
实验三	apple2orange	否	否	126000
实验四	apple2orange	是	否	126000








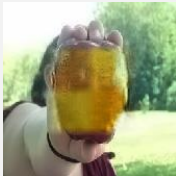







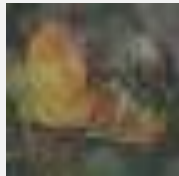
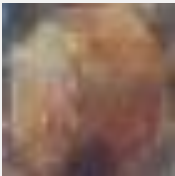
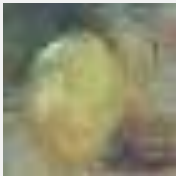


在接下来的部分,我们逐一进行实验并对转换的结果进行分析,表 4-18 与表 4-19 展示的是两个数据集下四个实验的转化结果。

表 4-18 summer2winter 模型转换结果实例

输入					
实验一					

分析表 4-18 中得到的结果，证明在 **summer2winter** 这一类不区分主体与背景的数据集下，模型的转换效果是令人满意的。但是在第三张图像的转换过程中还是可以看到拼接的痕迹，这是目前的算法难以避免的。

表 4-19 apple2orange 模型转换结果实例

输入					
实验二					
实验三					
实验四					

对表 4-19 的结果进行分析，首先我们可以看到实验二的转换效果也是令人满意的。但是实验三三，由于我们对该任务没有使用 CNN 分类器区分主体和背景，显然算法没有办法很好地将苹果转换成橘子。只有在处理第五张没有主体和背景区分的图像时，实验三才取得了不错的结果。将实验一，实验二与实验三进行比较，我们不难得出以下的结论：当我们的算法在进行有明确主体和背景区分的转换任务（例如 **apple2orange** 任务）时，引入 CNN 分类器的机制是非常有必要的；而对于没有明确主体和背景区分的转换任务（例如 **summer2winter** 任务）时，CNN 分类器并不是必须的。进一步地，我们单独比较实验二与实验三在训练过程中损失函数 $G: X \rightarrow Y$ 随迭代数的变化，见图 4-20。

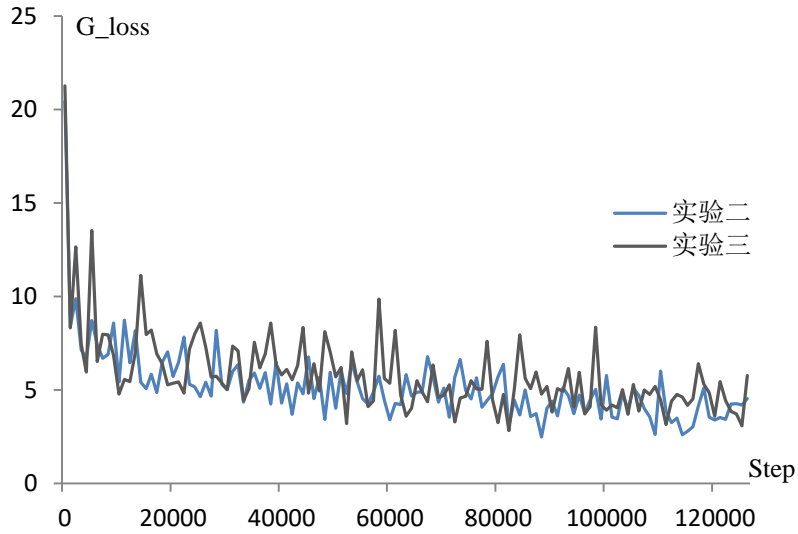


图 4-20 模型一与模型二损失函数 $G: X \rightarrow Y$

可以观测到在实验的过程中实验二的损失函数 $G: X \rightarrow Y$ 的收敛过程与实验三比是更加稳定的。我们可以认为 CNN 分类器的引入可以帮助 CycleGAN 模型更有针对性的学习到主体目标的特征，提高了学习的效率。

最后分析实验四的结果，实验四是我们直接压缩原始图片大小并输入 CycleGAN 网络的对比实验。通过实验四我们可以发现 CycleGAN 网络在处理低像素图像时没有办法很好地从训练集中学习到足够多的细节，因此在测试的图像转化过程中也会出现大量的细节丢失。通过实验二与实验四的比较我们不难总结出本文模型提出的“分割-融合”思想既可以很好地保存原始图像中的信息，也可以显著降低数据集的大小，降低训练 CycleGAN 网络的要求。

通过以上四组实验，我们可以证明出模型提出的“分割-融合”以及“CNN 分类器”两个机制的有效性。最后我们通过实验二 apple2orange 任务的转换结果，并结合引言中提出的目标，从三个方面对模型进行评估：训练一个新的模型对计算机的要求是否较低；图像转化的结果是否令人满意；模型在改变主体目标的同时是否最大限度避免改变背景中的内容。

i. 降低训练成本

我们使用 python 语言以及 tensorflow 框架实现了文中的模型，并且使用单机多 GPU 框架 tensorflow-gpu 来加快模型的训练速度，我们的个人计算机配置见表 4-21。

表 4-21 个人计算机配置

系统	Windows10 家庭版
CPU	I7-4720HQ 四核处理器
GPU	NVIDIA GTX970M (显存 3G)
运行内存	12GB

我们尝试过使用这台计算机运行作者开源的 CycleGAN 代码 (同样基于 tensorflow-gpu 架构), 但是在运行模型训练的时候会由于显存溢出导致中途崩溃。虽然我们可以通过减少训练过程中的 *batch_size* 或者每次只保留少量的生成器生成模型供判别器学习, 但是这种做法无疑会使 CycleGAN 模型的收敛速度大大降低。因此通过本文提出的这种训练策略, 可以在保证模型在以较快速度收敛的情况下仍然可以由一般配置的计算机运行。

ii. 图像生成结果

表 4-22 与表 4-23 分别展示了 apple2orange 以及 summer2winter 任务一组图像的转换结果。

表 4-22 apple2orange 模型转换结果实例





















输入					
输出					

表 4-23 summer2winter 模型转换结果实例

输入					
输出					


分析表 4-22 的结果，可以看到虽然生成的图片仍有比较明显的拼接痕迹，但是从结果来说转换是比较令人满意的。图像中的主体和背景被很好地区分开来，并且转换过程中一些细节也被最大程度地保留。分析表 4-23 的结果，可以看出转换的结果也是比较不错的。再将夏天的图像转换成冬天时，算法在树梢和岩石上都覆盖上了一层积雪，但是也很好地保留了树木和岩石本体的特征。

iii. 对背景内容的处理

通过对表 4-22 及 4-23 中图像的分析，我们不难发现模型对背景部分的内容进行了很好的保留。这得益于模型特殊的生成方式，因为分割-判别器的引入，模型主动选择出了图像中属于主体目标的部分并进行转化。

接下来我们将本文提出的模型与原始的 CycleGAN 模型进行比较，由于我们的设备没有办法直接运行原始的 CycleGAN 训练程序，因此我们直接选用了 CycleGAN 文章作者公布的预训练模型来代替训练的步骤。原始模型的 CycleGAN 参数均为作者使用的参数，使用的数据集为 apple2orange 数据集。在表 4-24 中我们挑选了一些典型案例进行分析。

表 4-23 apple2orange 模型转换结果实例比较

输入					
原始模型					
本文模型					

通过对实验结果的比较，我们可以发现本文提出的模型在主体目标的转换结果上是比原始模型要差一些的。造成这个现象的原因是本文提出的模型中 CycleGAN 网络输入的是大小为 48×48 像素的图像，而原始模型训练使用的是 256×256 像素的图像，这也就导致了原始模型的网络可以学习到更多的图像信息（例如橙子表面的纹理以及橙子内部果肉的纹理）。但是我们也不难发现原始模型在改变主体目标的时候也将图像的背景进行了比较显著的改变，使得生成图像的真实性大大降低。而本文提出的模型在图像转换的过程中很好地区分了主体和背景，因此在背景处理方面是超过原始模型的。

第五章. 总结与展望

近年来基于生成式对抗网络框架提出的模型层出不穷，模型的生成的结果越来越逼真，精度也越来越高。然而生成对抗网络的问题也一直存在，研究提出目前的生成式对抗网络框架还存在着以下几个缺陷：对训练集的样本依赖过大，生成结果缺乏多样性；生成结果不可控，训练过程中没有约束条件；通过对抗关系生成的两个网络难以平衡和同步。我们提出的融合了半监督机制的生成式对抗网络模型不仅让生成式对抗网络本身可以被更多一般配置的计算机运行，同时也提出了一种将有监督学习和无监督学习融合起来的思路。通过这种手段，我们成功地让网络生成的结果更加的可控。

当然我们提出的模型还存在一些缺陷可以进一步改善。由于模型对数据集引入了分割融合的操作，必然会导致最后的生成结果存在一定的缺陷。我们希望在未来找到更加合适的机制来解决这个问题。其次，由于我们在模型中引入了一个 CNN 图像分类器，相当于在整个模型中一共有五个神经网络，而这些网络的误差叠加起来也是不容忽视的。因此如何在保证生成生成式对抗网络的较好的生成结果的前提下引入无监督学习的机制也是一个很值得研究的课题。

参考文献:

- [1] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]// International Conference on Neural Information Processing Systems. 2014.
- [2] Ledig C, Wang Z, Shi W, et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network[C]// Computer Vision and Pattern Recognition. IEEE, 2017:105-114.
- [3] Li, Jiwei, Monroe, Will, Shi, Tianlin, et al. Adversarial Learning for Neural Dialogue Generation[J]. 2017.
- [4] Zhu J Y, Park T, Isola P, et al. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks[J]. 2017.
- [5] 王坤峰,苟超,段艳杰,林懿伦,郑心湖,王飞跃.生成式对抗网络GAN的研究进展与展望[J].自动化学报,2017,43(03):321-332
- [6] Pumarola A, Agudo A, Martinez A M, et al. GANimation: Anatomically-Aware Facial Animation from a Single Image[J]. 2018.
- [7] Kingma D P, Welling M. Auto-Encoding Variational Bayes[J]. 2013.
- [8] Chapelle O, Lkopf B S, Zien A. Semi-Supervised Learning[M]// Handbook on Neural Information Processing. 2013.
- [9] Mirza M, Osindero S. Conditional Generative Adversarial Nets[J]. Computer Science, 2014:2672-2680.
- [10] Perarnau G , Joost V D W , Raducanu B , et al. Invertible Conditional GANs for image editing[J]. 2016.
- [11] Choi Y , Choi M , Kim M , et al. StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation[J]. 2017.
- [12] Arjovsky M, Chintala S, Bottou L. Wasserstein GAN[J]. 2017
- [13] Mao X, Li Q, Xie H, et al. Least Squares Generative Adversarial Networks[J]. 2016.
- [14] Salimans T , Goodfellow I , Zaremba W , et al. Improved Techniques for Training GANs[J]. 2016.
- [15] Radford A, Metz L, Chintala S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks[J]. Computer Science, 2015.

- [16] Brock A , Donahue J , Simonyan K . Large Scale GAN Training for High Fidelity Natural Image Synthesis[J]. 2018.
- [17] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]// International Conference on Neural Information Processing Systems. 2012.