

参赛队员姓名：蒙永翔 陈文妍

中学：广州市第六中学

省份：广东省

国家/地区：中国

指导教师姓名：梁靖韵

论文题目：基于 GAN 的条件式人脸图像生成和调整

本参赛团队声明所提交的论文是在指导老师指导下进行的研究工作和取得的研究成果。尽本团队所知，除了文中特别加以标注和致谢中所罗列的内容以外，论文中不包含其他人已经发表或撰写过的研究成果。若有不实之处，本人愿意承担一切相关责任。

参赛队员：蒙永翔 陈文妍

指导老师：梁靖韵

2018年9月15日

基于 GAN 的条件式人脸图像生成和调整

作者：蒙永翔 陈文妍

指导老师：梁靖韵

广州市第六中学，广州，中国

目录

摘要.....	6
Abstract.....	7
1 研究背景与意义.....	8
1.1 文本到图像的研究.....	8
1.2 多种面部特征调整的研究.....	9
2 研究目标及创新点.....	10
2.1 研究目标.....	10
2.2 研究创新点.....	10
3 模型设计.....	11
3.1 编码器和解码器.....	12
3.2 生成式对抗网络.....	12
3.3 自编码网络.....	15
4 实验设置与结果.....	16
4.1 实验设置.....	16
4.1.1 实验环境.....	17
4.1.2 数据集.....	17
4.1.3 实验项目.....	17

4.1.4 训练细节.....	17
4.2 实验结果与分析.....	17
5 讨论.....	21
5.1 应用.....	21
5.2 不足与改进.....	21
6 总结.....	22
7 参考文献.....	23
致谢.....	25

摘要

目前, 人脸图像生成和人脸特征调整是一个有价值的课题。然而, 当前人脸的生成和调整仍然两种模型。本文提出的条件式人脸图像生成和调整模型目的在于: 使用一个模型进行生成和调整从而得到理想的人脸图像。该模型在 DCGAN[1]的基础上, 增添了自编码网络, 将生成的人脸结合新的调整条件, 在图像空间中对图像进行调整。实验证明, 该模型可以有效根据人脸特征条件生成真实可信的人脸图像, 也能够根据调整的条件进行合理调整。

关键词: 人脸图像生成、人脸图像调整、GAN

Abstract

Nowadays, generating faces and face image editing have been an attractive topic. However, there are still two models for face generation and adjustment. In this work, we develop a model based on DCGAN[1], added Auto-Encoder, which can generate facial images and adjust the target attributes according to the updated description in the image space. In our experiments, it is demonstrated that our model can generate plausible images of faces from attributes and be adjusted by the modify condition.

key words: facial images generating, facial attribute editing, GAN

1 研究背景与意义

近几年机器学习研究已运用到社会之中,人脸图像生成与调整便是其中一个典例。人脸作为人类重要的特征,在人与人的互动交流中起着非常重要的作用。人脸图像生成,即通过模型和算法生成人脸图像;生成的人脸图像结合人脸识别技术可以确定人物身份。这一技术关键应用在身份认证。身份认证在打击犯罪中可运用在对失踪人口或目标人物的追踪上;在社交中可应用于智能交互场景的角色身份识别,如校门口的安全检测、企业出入口的查点

目前搜集目标人物图像的方式部分局限于根据目击者口头描述,经过与作画者的合作得到其画像,进而在已有的图像数据库中进行搜索定位,从而完成对目标人物身份的确认和识别。我们认为在这一途径有很大的弊端,一是作画者的主观意识会影响画像的生成,二是对作画者的绘画素质要求较高、成本较高,三是特定场合下对图像生成的速度有较高要求。或是根据提供的目标人物照片,则也可能出现拍照角度不佳或是拍照时间久远的问题,使得目标人物的人脸并不准确,我们所做的更改人脸图像属性的研究便试图解决这一问题。而尤其在需要目标人物图像的领域,如刑侦方面,人物图像的获取尤其重要,因此,面对准确定位目标人物,破解案件的难题,本研究具有一定的必要性。

为了实现高效生成高质量符合人们需求的人脸的目的,一方面需要合理处理输入,并在多次调整人脸图像指定属性之后仍能保持原图像的部分特征;另一方面则是对人脸进行训练得到更有效的模型。

随着机器学习的发展,现在计算机研究中已有针对人脸合成的相关研究,下面我们对国内外的相关工作进行简要介绍。

1.1 文本到图像的研究

许多生成图像的模型对人脸合成有极大的启发。

ScottReed 等较早提出从文本到图片的生成[2]。他们基于 DCGAN,通过文本的描述生成

低分辨率的图像，再经过卷积，运用 RNN 做图片生成叙述，并与原先文本描述进行对比更新。

StackGAN[3]中作者通过文本描述初步生成物体边框，再进一步结合原先的文本描述生成高分辨率的图像。其中他们应用多个生成器生成多个分辨率的图像，并将每个分辨率对应一个鉴别器从而提高图像的质量；此外，他们还用到了非条件的 loss，从而保留原先描述中的属性。

上述两篇论文都是将生成图像的任务分解成两个方面。首先获取文本特征表示，然后用这个特征合成一张较好的图片。我们的模型也有类似的思想，但不同的是，我们是连续生成，并且生成的是人脸图像。

1.2 多种面部特征调整的研究

我们的第二步生成图像即调整人脸图像的过程。其中已经有相关研究。StarGAN[4]通过跨数据集进行训练，从而能够使得模型可以学习到来自多个数据集的属性和特征，能够指定更多属性从而生成图像。

AttGAN[5]则对输入的图像进行隐含向量的提取。在 AttGAN[5]实验中，先提取现有图像的隐含向量，然后结合更改后的属性生成图像。这一过程中的提取隐含向量过程会不可避免地造成信息丢失，因此新图像中将会丢失部分原有图像的特征。

目前关于人脸的研究中人脸图像生成及人脸图像属性调整一般是分离的，需要使用不同的多个模型才能完成，因此我们希望使用单个模型完成人脸图像生成及调整的任务，这样不仅可以共享提取图像特征的网络层参数、来减少训练模型中时间和算力的消耗，还能够使得图像的生成和调整过程中，其属性参数等统一，避免不同模型的转换间，带来的额外消耗。

2 研究目标及创新点

为了给人脸图像缺失的问题提供新的解决途径, 本文将对基于生成式对抗网络(GAN)的人脸图像生成和调整模型进行研究。

2.1 研究目标

- (1) 为了给人脸图像缺失的问题提供新的解决途径, 我们使用大量人脸数据训练模型, 探索利用机器学习技术提出可行、便捷、高效的方法, 弥补诸如手绘和人工利用软件修改图像等的传统手段及其他技术手段的不足。
- (2) 结合具体问题及前沿研究成果, 在 DCGAN[1]的基础上加以改进, 提出一步生成人脸图像并调整的模型, 实现在多种属性信息的环境下对人脸的生成与调整。
- (3) 在生成人脸图像的基础上, 结合调整条件在图像空间内完成对图像的调整, 并保留原先的。
- (4) 使用公开数据集 CelebA[6], 通过收敛速度、损失函数的对比, 验证模型的可行性。

2.2 研究创新点

- (1) 本研究提出基于 GAN 的条件式人脸图像生成和调整的模型, 可实现使用单一模型来完成生成人脸图像和调整人脸图像的两个任务。能够有效的减少模型训练量, 提高人脸图像生成速率
- (2) 我们的模型创新性地结合 U-NET[7]网络, 根据调整在图像空间内调整图像, 减小调整图像过程中原图像信息的损失。
- (3) 本研究结合了人脸图像生成和人脸图像属性调整的方法。在刑侦领域, 人脸图像是确定目标任务的重要途径。人脸图像的获取来源则主要分为图像的采集和画手根据目击者描述的

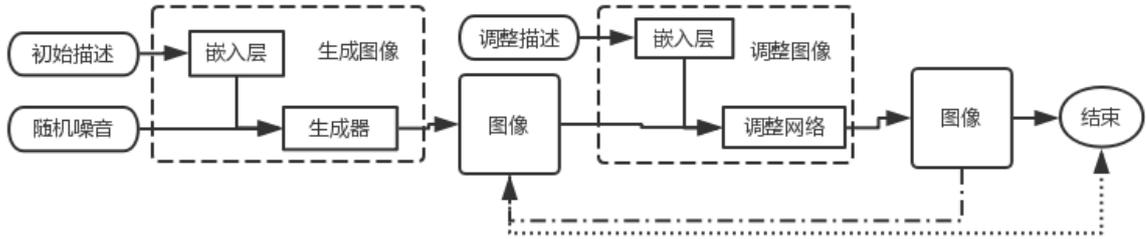
绘画, 但这类图像信息在角度、清晰度或是多方理解转述偏差等方面利用率低。本研究拓展了人脸图像的应用, 为图像应用方式从被动的判断搜寻转为主动的生成调整, 应用领域上从人工的获得调整到智能的生成调整提供了一个可行的范例, 具有实时性与实效性。在操作上, 对专业人员没有要求, 且相较于原有的人工方式(素描), 生成的图像更为真实。

3 模型设计

模型概述

我们提出的基于 GAN 的条件式人脸图像生成和调整模型, 是一种基于生成式对抗网络的图像生成和调整的方法, 重要的改进在于使用一个模型完成对人脸图像的生成和调整两个任务。

该模型包含一个解码器和一个编码器, 这两个组件构成一个生成式对抗网络和一个类似于 U-Net[7]结构的自编码网络, 两个网络共享编码器的网络参数。在本模型中, 通过生成器将初始噪声和条件生成原始图像, 并结合调整描述通过调整网络得到新的图像, 完整的图像生成和调整过程如下:



图表 1 基于 GAN 的条件式人脸图像生成和调整过程

3.1 编码器和解码器

在我们的网络中，编码器主要用于从图像中提取特征。我们采用卷积的方法，从 $128 \times 128 \times 3$ （高 \times 宽 \times 通道数）的图像开始，连续进行带步长的卷积操作，实现提取特征同时降低图像维度，并在每次卷积操作之后按顺序进行标准化、激活和 Dropout[8]，从全局到局部提取特征。

其中卷积层使用卷积核大小为 5×5 ，过滤器从 64 个起、每层以上一层 2 倍的数量递增、直至 512 个封顶，步长设为 2，因此在提取特征的同时降低图像维度，降低网络计算量。由于我们的训练中使用到了 WGAN-GP[9]中提出的梯度惩罚，因此我们的标准化层选择了实例标准化(Instance Normalization)[10]。激活函数使用 $\alpha=0.2$ 的 Leaky ReLU。为了减轻过拟合的发生，训练中我们使用了 Dropout 随机隐藏部分神经元，减少过滤器间的相互作用。

解码器主要用于从特征向量生成图像，采用转置卷积的方法。与编码器对称地使用相同的过滤器数量，步长设为 2，卷积核大小同样设为 5×5 ，每次卷积操作后按顺序进行标准化和激活。标准化曾和激活函数与编码器相同，使用实例标准化和 Leaky ReLU。

3.2 生成式对抗网络

在生成式对抗网络中，模型分为生成器和判别器。生成器由顶部的全连接层和解码器组成，人脸特征条件参数（下文简称“条件”）结合随机噪音通过转置卷积的方式生成图像；判别器由编码器和底部的全连接层组成，通过卷积提取特征并结合条件判断图像是否来自真实世界以及是否与所给条件吻合。

生成器顶部是全连接网络，将输入的人脸特征条件和符合正态分布的噪音（隐含向量）共同连接成 $4 \times 4 \times n$ 层的特征图，在每层的卷积之间，掺入由条件参数连接成的特征图，然后通过由解码器逐层由全局到细节生成图像，最后进行一次步长为 1，过滤器为 3 的卷积操作并使用 \tanh 函数激活，得到 $128 \times 128 \times 3$ 大小的彩色图像。

判别器顶部是编码器，将输入图像提取出 $4 \times 4 \times n$ 大小的特征向量。然后特征向量通过全连接到维度为 1 的向量，表示图像来自真实世界的可能性。特征向量再结合输入条件并进行全连接，得到图像是否符合输入条件的可能性。

对于判别器和生成器，为保证损失函数过大或过小，导致模型失去平衡，我们使用将其约束到 0~1 之间。由于我们的判别器不仅输出判断图像是否来自真实世界，还包含图像是否与输入的条件参数一致，因此包含 GAN 损失和条件判断损失。我们期望判别器对真实的图像判别结果尽可能接近 1，来自生成器的结果尽量接近 0。同样，希望判别器对与输入特征一致的图像判别结果接近 1，不符合的接近 0，因此损失函数设计为如下：

$$Loss_{D_1}(G, D) = E_{y \sim P_{data}} [||1 - D(y)||] + E_{z, c \sim P_g} [||D(G(z, c))||]$$

公式 1 判别器的 GAN 损失

$$Loss_A(A) = E_{y, c \sim P_{data}} [||1 - A(y, c)||] + E_{y, c \sim P_{fake}} [||A(y, c)||]$$

公式 2 条件判断损失

$$Loss_D(G, D, A) = Loss_{D_1}(G, D) + Loss_A(A)$$

公式 3 判别器的总损失

我们在生成器的损失函数中增加了平均绝对误差, 期望模型在具有 GAN 的生成结果多样性的情况下使生成图像整体色彩和布局更加真实。我们期望生成器的结果尽量使得判别器判别结果为 1, 与真实图像的平均绝对误差为 0, 因此生成器的损失函数分为 GAN 损失以及平均绝对误差损失如下:

$$Loss_{G1}(G, D, A) = E_{z,c \sim P_g} [||D(G(z, c))||] + E_{z,c \sim P_g} [||A(G(z, c), c)||]$$

公式 4 生成器的 GAN 损失

$$Loss_{L1}(G) = E_{z,c \sim P_g, y \sim P_{data}} [||y - G(z, c)||]$$

公式 5 生成器的绝对误差损失

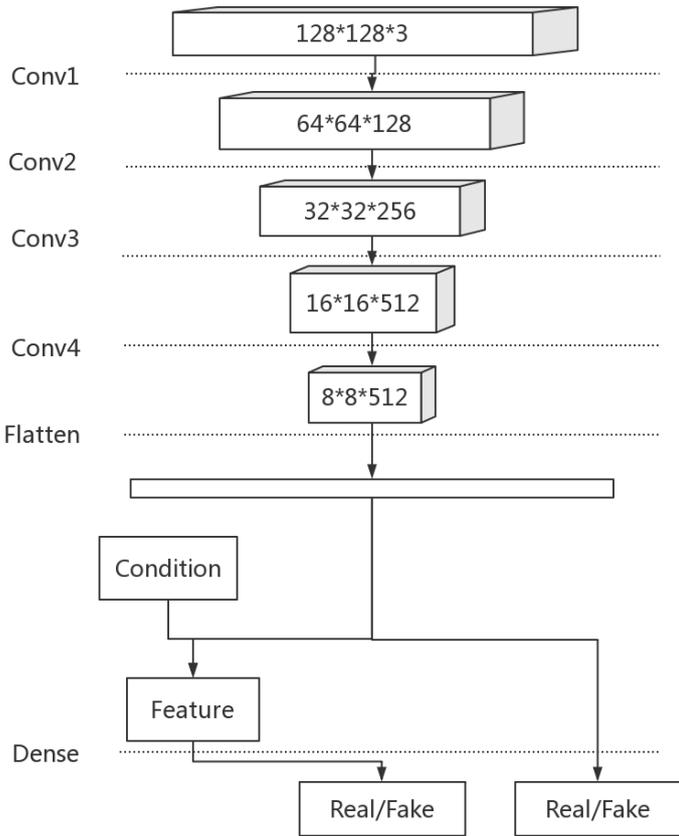
$$Loss_G(G, D, A) = Loss_{G1}(G, D, A) + \lambda Loss_{L1}(G)$$

公式 6 生成器的总损失

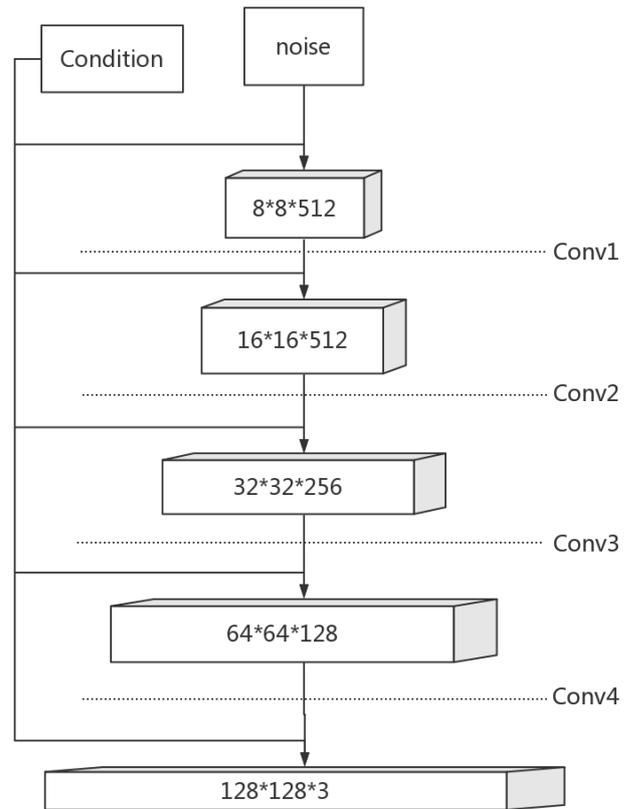
我们的目标是使得生成器和判别器的损失函数同时最小, 因此生成式对抗网络的优化目标为:

$$GAN^* = \arg \min_{D, A} \max_G Loss_G(G, D, A) + Loss_D(G, D, A)$$

公式 7 生成式对抗网络的优化目标



图表 2 判别器网络



图表 3 生成器网络

3.3 自编码网络

自编码网络由解码器、编码器、残差层、多个捷径通道以及调整条件的合并通道共同构成。

网络使用编码器将输入图像从全局到细节逐层提取特征，并将每层特征通过捷径通道传输至解码器网络，使得每层网络都不需要承载图像的所有信息。

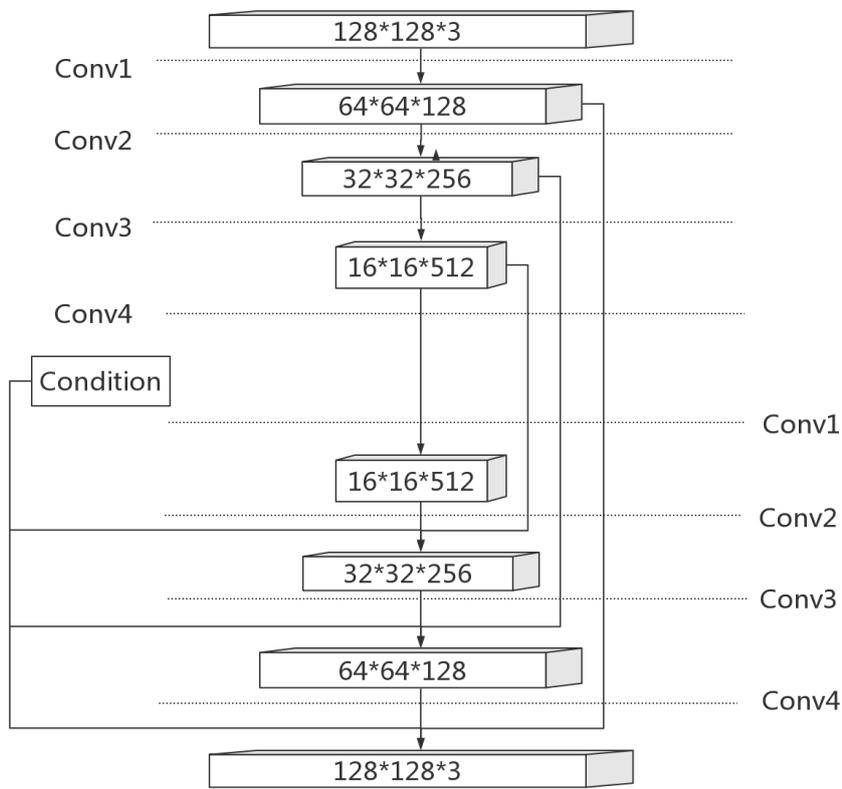
网络使用解码器将图像从细节到全局逐步将图像原有特征和调整条件合并、生成图像。在编码器和解码器之间，我们添加了残差层，使得图像分辨率增大、网络层数增加后自编码网

络仍能正常工作。

自编码网络中，解码器的参数与 GAN 中的解码器参数独立，编码器参数与 GAN 中的共享。我们使用 GAN 损失和均方误差作为自编码网络的损失函数。

$$Loss_U(U, D, A) = E_{x, c \sim P_u} [||D(U(x, c))||] + E_{z, c \sim P_u} [||A(U(x, c), c)||] + \lambda E_{y \sim P_{real}, x, c \sim P_u} [||y - U(x, c)||]$$

公式 8 自编码网络的损失函数



图表 4 自编码网络

4 实验设置与结果

4.1 实验设置

4.1.1 实验环境

本文实验中使用 Ubuntu 16.04 x64 操作系统, 使用 Python 编程语言版本 3.6.6, 基于 Tensorflow 1.10.1 和 Keras 2.2.2 框架。模型训练在 NVIDIA TITAN X (Pascal) 上进行。

4.1.2 数据集

我们使用公开数据集 CelebA[6], 包含总计 202,599 张人脸图像, 并且对每张照片人脸的 40 个属性进行了标注。我们将所有图像中的人脸对齐并裁剪为 128×128 像素的 RGB 三通道图像, 然后按照 8: 1: 1 的比例切分为训练集、测试集、验证集。对于数据集中 40 个属性, 我们选择了其中的 5 个标记: '性别', '年龄', '戴眼镜', '皮肤颜色', '发色', 属性来验证模型的可行性。

4.1.3 实验项目

为了验证我们的模型能够实现根据输入条件生成真实度高的人脸图像, 我们设置了模型生成并调整人脸图像的实验。

4.1.4 训练细节

我们使用 Adam[11] 优化器对网络进行优化, 判别器和生成器的学习率设为 0.0001, 自编码网络的学习率设为 0.0002, β_1 设为 0.5, β_2 设为 0.9。训练自编码网络时将编码器网络固定。自编码网络和生成器中的 λ 均设为 0.2, 使用 100 维的隐含向量输入, 每批使用随机的 32 张训练集图像, 进行训练, 一共使用训练集进行 15 轮训练。每一批次中先用生成器生成图像, 然后用判别器辨别图像, 用自编码网络调整图像, 用以上计算结果梯度惩罚, 计算判别器、生成器、自编码网络的损失, 使用各自的优化器分别进行反向传播、优化网络权重参数。

4.2 实验结果与分析

我们的模型在第一轮的训练中已经能够产生清晰、辨识度较高的人脸图像，相较我们的基线模型 DCGAN，可以明显地看出我们模型生成的图像更为锐利和清晰，整体在色彩上也更加接近真实图像。

为了更好的度量我们的模型与 DCGAN 收敛速度，我们使用 Tensorboard 可视化工具生成训练过程中网络的损失变化图。通过对比可发现，相比 DCGAN, 我们的网络的 loss 变化更为平稳，抖动较小，而且在稳步下降，而 DCGAN 会经常出现拟合导致 loss 突然增大的现象。



Figure 7 我们的模型第 1 轮的生成图像



Figure 7 DCGAN 第 1 轮的生成图像

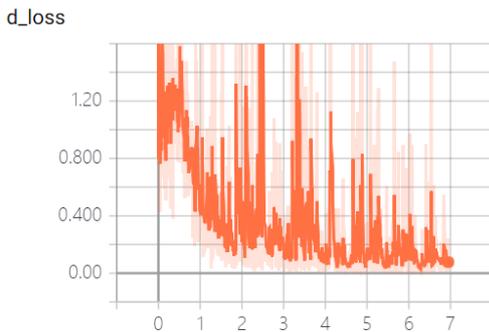


Figure 8 DCGAN 的判别器损失变化

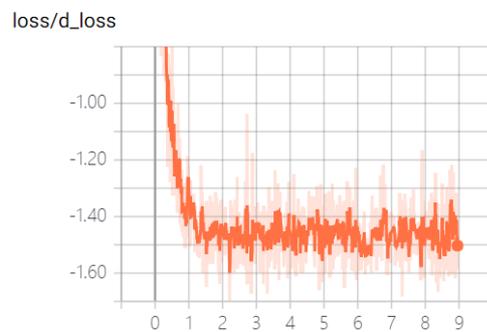


Figure 9 我们的判别器损失变化

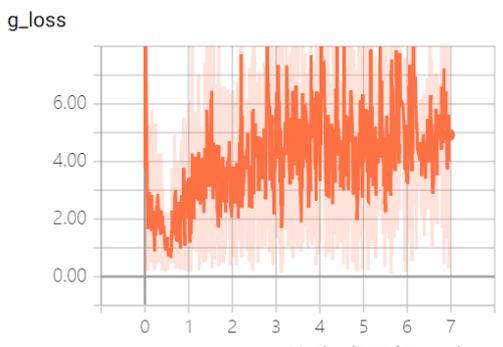


Figure 5 DCGAN 的生成器损失变化

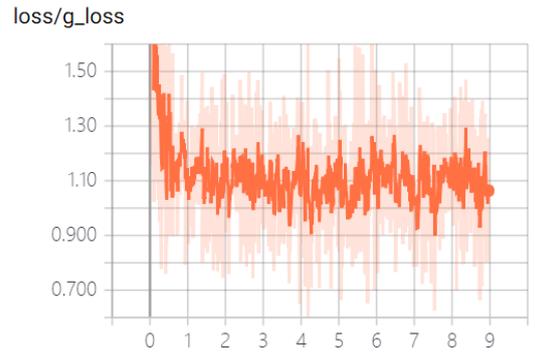
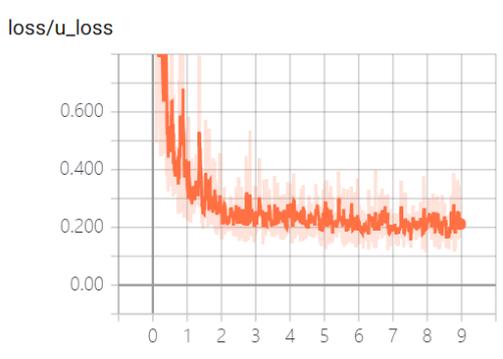


Figure 6 我们的生成器损失变化



图表 5 我们的自编码网络损失变化

大约在第 15 轮时网络已经能够输出非常真实的人脸图像。

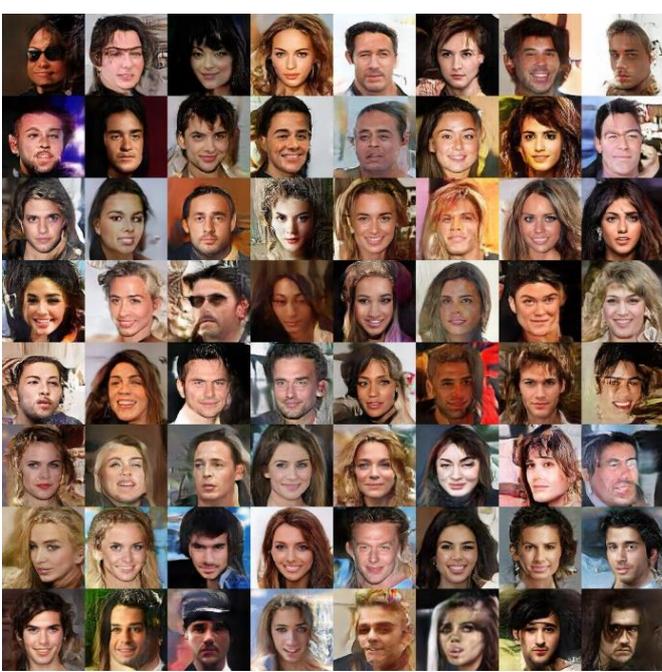


Figure 10 我们的模型 第 15 轮的生成图像



Figure 11 DCGAN 第 15 轮的生成图像

下面是我们进行的人脸图像的生成并调整的测试结果,可以看到图像在大致一致的情况下,某一单一特征与其他图像出现了不同,证明我们的网络能够对人脸图像的制定特征进行调整:



Figure 11 我们的模型第 15 轮 进行人脸单一属性调整测试

5 讨论

5.1 应用

5.1.1 对目标人物的搜寻

在刑侦领域，出于成熟的人脸识别技术及完善的人脸数据库，目标人物的图像较容易被捕捉到，但由于角度及清晰程度等问题，捕捉到的图像可能与实物出现偏差，通过我们的模型可以进一步改进这一问题。

5.1.2 虚拟形象生成

虚拟空间中，无论是游戏中的角色生成，还是 VR，在个人方面或是企业方面上对于自制角色的需求都越来越大。我们认为结合我们的模型能够更方便对人物进行个性化更改。

5.2 不足与改进

5.2.1 多领域训练

由于数据集的标注较少，缺少一些关键属性，以及某些属性标记过于宽泛，导致我们输入模型的人脸描述不够全面，无法真正做到通过任意描述对人连进行调整。因此，我们考虑到可以参照 StarGAN[4]中提出的方法，即在多领域、多数据集之间进行训练。

5.2.2 更高清的图像

由于数据集限制，在我们的实验中，我们用于训练的图像分辨率较低，因此，我们生成的图像的清晰度不够高。在之后的改进中，我们希望使用高清的标记人脸数据集，如 CelebA-

HQ[12]并借助 PG-GAN[12]的方法, 逐层训练网络层以得到生成更高分辨率的人脸。

5.2.3 使用自然语言作为条件输入

由于我们暂时没有关于人脸属性的自然语言描述数据集, 因此我们未能训练从自然语言生成图像的模型。我们希望创建一个人脸属性的自然语言描述数据集, 借此训练一个从自然语言生成人脸图像的模型, 更好的达到我们希望普通人通过描述快速得到准确的人脸图像的期望。

6 总结

我们在 DCGAN[1]的基础上, 提出了一种能生成人脸图像并调整人脸图像属性的模型。其中我们通过增加 U-Net[7]网络, 结合 WGAN-GP[9], 在 DCGAN[1]的基础上进行了创新。我们的创新点在于建立一个生成人脸图像与调整人脸的模型, 其中只在图像空间内对人脸进行调整。这减小了调整过程中生成图像与原图像之间的信息损失。

我们会对该模型做进一步的调整。并且相信本研究可以达成我们最开始的目标, 即让普通人通过描述快速得到准确的人脸图像, 便捷人们的生活。

7 参考文献

- [1] A. Radford, L. Metz, and S. Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks,” *CoRR*, vol. abs/1511.06434, 2015.
- [2] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee, “Generative adversarial text-to-image synthesis,” in *Proceedings of The 33rd International Conference on Machine Learning*, 2016.
- [3] H. Zhang, T. Xu, H. Li, S. Zhang, X. Wang, X. Huang, and D. Metaxas, “Stackgan++: Realistic image synthesis with stacked generative adversarial networks,” *arXiv: 1710.10916*, 2017.
- [4] Y. Choi, M. Choi, M. Kim, J. Ha, S. Kim, and J. Choo, “Stargan: Unified generative adversarial networks for multi-domain image-to-image translation,” *CoRR*, vol. abs/1711.09020, 2017.
- [5] Z. He, W. Zuo, M. Kan, S. Shan, and X. Chen, “Arbitrary facial attribute editing: Only change what you want,” *arXiv preprint arXiv:1711.10678*, 2017.
- [6] Z. Liu, P. Luo, X. Wang, and X. Tang, “Deep learning face attributes in the wild,” in *Proceedings of International Conference on Computer Vision (ICCV)*, 2015.
- [7] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” *CVPR*, 2017.
- [8] K. R. Konda, X. Bouthillier, R. Memisevic, and P. Vincent, “Dropout as data augmentation,” *CoRR*, vol. abs/1506.08700, 2015.
- [9] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, “Improved training of wasserstein gans,” *CoRR*, vol. abs/1704.00028, 2017.
- [10] D. Ulyanov, A. Vedaldi, and V. S. Lempitsky, “Improved texture networks: Maximizing quality

and diversity in feed-forward stylization and texture synthesis,” *CoRR*, vol. abs/1701.02096, 2017.

[11]D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *CoRR*, vol. abs/1412.6980, 2014.

[12]T. Karras, T. Aila, S. Laine, and J. Lehtinen, “Progressive growing of gans for improved quality, stability, and variation,” *CoRR*, vol. abs/1710.10196, 2017.

致谢

在本次研究中,我们先是各自提出对研究方向的想法,在结合老师的意见及实际情况下,我们选择了人脸图像生成与调整这一课题。本文论文构架由我们共同探讨得出。在本次实验中,我们根据观察出的问题,共同设计模型、进行实验,并在实验中相互交流探讨,对模型进行调整。

其中蒙永翔主要负责代码的编写,主要撰写论文的模型设计、实验设置与结果部分;

陈文妍主要负责背景调查,实验结果的检测与对比,撰写论文的背景部分和模型结构图。

本论文在广州市第六中学的梁靖韵老师的悉心指导下完成,她严谨的学术作风为我们的研究起到了很好的模范作用。

感谢中山大学的郑伟诗教授及他的研究生团队,他们为我们的模型测试提供了实验室和必要的设备,并为我们的论文提供了必要的指导。

最后感谢我们的父母,他们在我们研究过程中提供的理解与支持成为我们的动力。

队员简历

蒙永翔:男,广州市第六中学高中生。初高中多次参加与电脑制作相关的比赛,成绩突出。18年的作品《xWriter 基于人工智能 NLP 的写作助手》获省二等奖。熟悉电脑技术和机器学习。擅长编写程序、网站搭建、服务器管理等。带领并组织学校社团的活动,并多次于校园开放日展示的作品。个人主页:<https://github.com/ix64>

陈文妍:女,广州市第六中学高中生。关注科技发明,特别是人工智能方向。带领并组织学

校社团的活动, 并多次于校园开放日展示游戏作品。兴趣广泛。积极参加各种比赛, 并分别在辩论、英语、民族舞等比赛中获得较好的成绩。高二曾作为校篮球队的队员参加区篮球比赛, 取得第二的成绩。课余, 爱好绘画, 钢琴。

指导老师:

梁靖韵, 中学信息技术高级教师, 华南师范大学教育硕士, 高级程序员, 广州六中教学能手。

长期承担信息学(计算机)奥林匹克竞赛与科技创新、电脑作品比赛的辅导工作。历年来所指导的学生有 10 人次以上获提高组信息学奥赛全国一等奖, 2015-2018 年指导学生参加宋庆龄少年儿童发明奖、中小学电脑制作活动等获全国一、二等奖共 4 人次。另有多名学生获信息学奥赛、电脑制作活动、科技创新活动的省、市级各项奖励。

郑伟诗博士, 中山大学数据科学与计算机学院教授。他主要面向大规模智能视频监控, 展开视频图像信息与信号的处理研究, 并开展大规模机器学习的算法和理论研究。他目前的主要研究应用领域是: 视频监控下的行人身份识别与行为信息理解。面向大规模监控网络下的行人追踪问题, 他在国内外较早和持续开展跨视域行人重识别的研究, 发表一系列以跨视域度量学习为主线的研究工作, 他提出的基于相对比较思想建模思路在行人重识别中被广泛深入研究。他已发表 100 余篇主要学术论文, 其中 70 余篇发表在图像识别和模式分类 IEEE TPAMI、IEEE TIP、IEEE TNN、PR、IEEE TCSVT、IEEE TSMC-B 等国际主流权威期刊和 ICCV、CVPR、AAAI、IJCAI 等计算机学会推荐 A 类国际学术会议。担任 Pattern Recognition 等期刊的编委, 担任 AVSS 2012、ICPR 2018、BMVC 2018 Area Chair/Associate Editor 等。获国家优秀青年科学基金、英国皇家学会牛顿高级学者基金、广东省自然科学杰出青年基金和广东省创新领军人才项目支持。个人主页: <http://isee.sysu.edu.cn/~zhwshi/>