

# 基于DCGAN网络算法的华人人脸生成

李怀诚\*, 古 彭\*, 杨诗妍#, 戚馨文, 刘宝镇

西藏大学, 西藏 拉萨

Email: #1296907025@qq.com

收稿日期: 2021年4月25日; 录用日期: 2021年5月20日; 发布日期: 2021年5月27日

## 摘 要

近些年来, 随着深度学习算法的进一步发展, 卷积神经网络进行监督学习在计算机视觉应用上取得了巨大的突破。然而, 在无监督学习方面, 卷积神经网络的关注度并不高。本文将深度卷积增加到无监督学习的生成对抗网络上, 用来生成华人人脸。生成对抗网络, 一方面可以减少复杂的计算, 另一方面生成的图像质量也比较理想。通过使用一万五千张的华人人脸数据集, 本文最终生成了具有多样性的华人人脸照片, 并且最终的判别器损失函数在0.695上下浮动, 生成器损失函数在0.74上下浮动。

## 关键词

深度学习, 生成对抗网络, 计算机视觉

# Chinese Face Generation Based on DCGAN Network Algorithm

Huaicheng Li\*, Peng Gu\*, Shiyang Yang#, Xinwen Qi, Baozhen Liu

Tibet University, Lhasa Tibet

Email: #1296907025@qq.com

Received: Apr. 25<sup>th</sup>, 2021; accepted: May 20<sup>th</sup>, 2021; published: May 27<sup>th</sup>, 2021

## Abstract

In recent years, with the extensive research of deep learning algorithms, convolutional neural networks for supervised learning have made huge breakthroughs in computer vision applications. However, in terms of unsupervised learning, the attention of convolutional neural networks is not

\*共同第一作者。

#通讯作者。

high. This paper adds deep convolution to the unsupervised learning generative confrontation network to generate Chinese faces. On the one hand, generating a confrontation network can reduce complicated calculations, and on the other hand, the quality of the generated images is also ideal. By using 15,000 Chinese face data sets, this paper finally generates diverse Chinese face photos. The final discriminator loss function fluctuates around 0.695, and the generator loss function fluctuates around 0.74.

## Keywords

Deep Learning, Generative Adversarial Network, Computer Vision

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

### 1.1. 研究背景及现状

本次研究依靠高清真实华人人脸数据训练神经网络并结合国外相关前沿算法,生成真实的人脸数据。人脸数据可以应用于两个方向:人脸图像生成和人脸图像分析,两者研究思路相反但又相辅相成互相促进,人脸图像生成促进了人脸检测、人脸识别、人脸属性预测等人脸图像分析的发展,并且人脸图像分析也为人脸图像生成提供了发展动力。人脸图像生成得益于深度学习的不断发展,人脸图像在研究和工业上的应用也越来越广泛。人脸图像生成与分析的关系如图1所示:

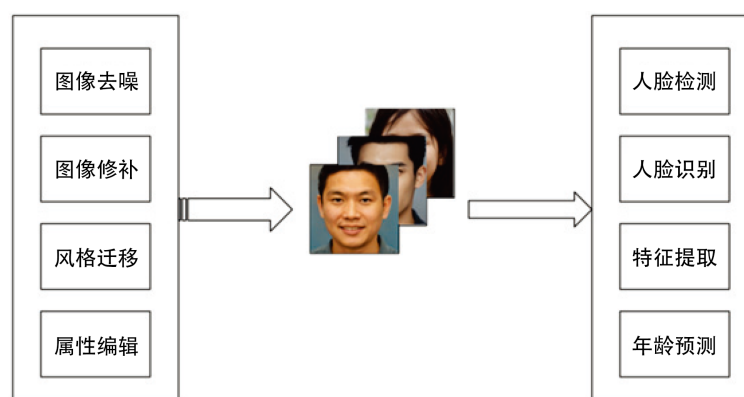


Figure 1. Relation diagram of face generation and analysis

图1. 人脸生成与分析关系图

人脸生成技术可以应用于生活中的各个领域[1],一方面可以直接应用,例如可以通过人脸图像超分辨率技术将画质模糊的人脸图像恢复成高质量的人脸照片,在追查罪犯、安检等安防场景提高人脸检测和识别的精度;运用人脸图像风格转换进行“换脸”、人脸视角转换、人脸姿态估计、人脸画风转换等操作,为互联网用户提供更多娱乐体验,并且可以应用到修图、社交头像等应用上;根据文本信息要求查询数据库中的指定人脸图像,提高检索速度,利用文字或者图像对人脸图像进行属性编辑,控制人脸图像的变化,生成符合要求的人脸图像;另一方面可以通过辅助人脸图像理解间接应用,例如增加数据

库高质量和稀少的特定人脸图像的样本, 解决数据中类别不平衡的问题, 缓解训练样本不足问题, 为图像理解的各个方向提供充足的样本数据, 起到数据增强的作用, 支撑深度学习的模型训练; 通过无监督或者半监督的学习方式, 提取人脸图像的关键特征和抽象特征, 将这些特征应用到人脸图像分析方向中, 有效地提升人脸识别、属性预测等任务的性能。总之人脸图像生成与人脸的研究和应用密切相关, 它可以降低获得人脸数据的成本, 提高人脸分析的效率和性能, 在人脸图像领域内具有非常有意义的研究价值。

## 2. 相关理论基础及算法实现步骤

### 2.1. 生成对抗网络的介绍

2014 年 10 月, 生成对抗网络模型被 Goodfellow 第一次提出。生成对抗网络的灵感来源于博弈论原理[2]。主要体现在生成对抗模型中分为两大主要部分——生成器和判别器。生成器的任务目标主要是将产生的随机数经过神经网络结构输出一张虚假的图片, 而判别器是用来区分真实图片与虚假图片, 此时形成了两个模型动态博弈。这两个模型相互对抗学习, 最终可以得到较可观的输出结果。

生成对抗网络由生成器和判别器两个部分组成, 如图 1 所示。生成器是由  $\theta$  参数化的神经网络实现。生成器的输入是一个服从于某一分布  $p_z$  的随机向量  $z$ , 而生成器的输出可以看成是采样于某一分布  $p_g$  的一个样本  $G(z)$ 。假设真实数据的分布为  $P_{data}$ , 在给定一定量真实数据集的条件下, 对生成对抗网络进行训练, 让生成器学到一个近似于真实数据分布的函数。生成对抗网络中生成器的主要目的是生成类似于真实数据的样本以骗过判别器, 而判别器的输入由真实的样本和生成的样本两个部分组成[3]。判别器的目标就是判断输入的数据是来自于真实的样本还是来自于生成器生成的样本。生成器和判别器经过对抗训练达到一个纳什平衡状态, 即判别器判断不出其输入是来自于真实的样本, 还是来自于生成器生成的样本, 此时就可以认为生成器学习到了真实数据的分布[4]。如下图 2 所示, 在理论上, 假设在生成对抗网络中真实数据分布为  $P_{data}(x)$ , 并且有一个被  $\theta$  参数化的生成分布  $P_G(x; \theta)$  [5]。假使想要让生成器产生的图片和真实数据逼近, 第一要使  $P_G(x; \theta)$  随机采样数量为  $m$  的样本, 通过计算出  $P_G(x^i, \theta)$ , 最终通过最大似然函数:

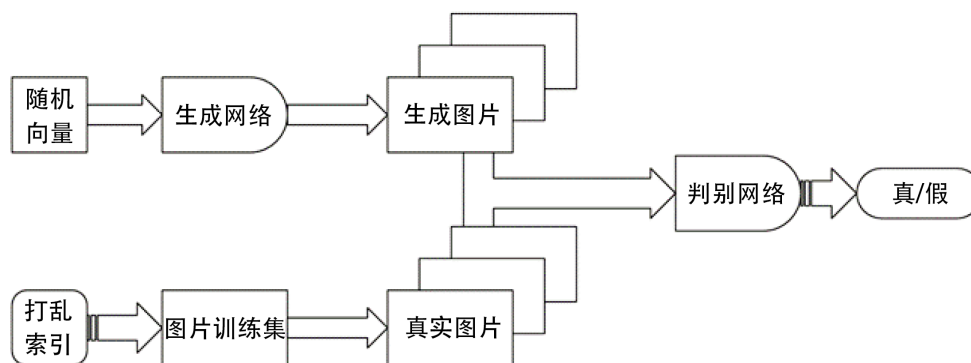


Figure 2. Generate a confrontation network structure diagram

图 2. 生成对抗网络结构图

$$L = \prod_{i=1}^m P_G(x^i, \theta)$$

来求出参数  $\theta$ , 其结果为:

$$\theta^* = \arg \min KL(P_{data}(x) | P_G(x; \theta))$$

通过求导似然函数得到最佳的  $D^*(x) = \frac{P_{data}(x)}{P_{data}(x) + P_G(x)}$ 。

此时似然函数的值最大：此时，将  $D^*(x)$  代入到似然函数中就可以得到生成分布和真实数据之间的 Jensen Shannon 散度[6]，最终的目标是通过一个分布去逼近另一个分布的过程，将两个分布的距离不断减少，转化为最小值问题。基于这样的思想设计构建出生成对抗网络的目标函数，方程如下所示：

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))]$$

其中  $V(G, D)$  是一个二分类的交叉熵函数。该损失函数的最终目标是 최소화 生成分布和真实分布之间的 KL 散度。通过分析对抗网络的目标函数，并且从判别器的角度来看的话，如果判别器的输入是来自于真实样本，判别器将会最大化输出；如果判别器的输入来自于生成器生成的样本，则判别器将会最小化输出[7]；同时生成器想要去骗过判别器，那当生成器生成的样本作为判别器的输入的时候，必须满足最小化损失函数  $V(G, D)$ 。但是当判别器被训练得非常好的时候，它将以很高的置信度直接将来自于生成器的样本判为假。此时  $\log(1 - D(G(z)))$  就会饱和，从而导致梯度为 0，最终参数得不到更新。此时可以将  $\log(1 - D(G(z)))$  换成  $\log D(G(z))$ 。尽管新的目标函数可以提供不同于原始损失函数的梯度，但是仍然存在梯度消失的问题；同时在理论上假设生成器和判别器具有充分的能力去对一个未知的分布进行建模，但实际上这种建模能力是有限的。

## 2.2. 卷积神经网络

### 2.2.1. 归一化层

在机器学习领域，当模型训练的时候，通常会对数据和特征进行归一化的处理，统一量纲，避免数值大小的差异对模型的训练产生影响。而在深度学习领域网络层间也有必要进行特征的归一化处理，若不进行归一化处理，由于前面网络的输出是后面网络的输入，那么当后面的网络层按照前面网络层的学习方式调整后，并且前面的网络层参数也是通过不同批数据不断更新的，此时前面的网络层已发生改变，因此后面的网络层学习到的特征已经滞后，不足以让模型性能有所提升，要重新通过前面网络层的调整而调整，这样反复循环，使得模型训练速度大大降低，这就是内部协变量转移(Internal Covariate Shift, ICS)。而当对每一层的输出进行归一化处理后，不管前面网络参数怎么改变，它们的输出分布都是固定的，后面的网络层无需根据前面网络的更新而不断调整，从而解决了 ICS 的问题。

但是目前存在的问题是，网络层的参数是不断变化的，其输出的特征也是不断变化的，特征图的分布自然也就变了，那么该如何使得每一层的输出特征进行归一化处理，使它们限制在同一个分布中，比如标准正态分布。此时批归一化层(BN)就出现了，其公式如下所示：

$$x_{i,j,k,l}^{(out)} = \gamma_l \times \frac{x_{i,j,k,l}^{(in)} - \mu_l}{\sigma_l + \varepsilon} + \beta_l$$

其中

$$\mu_l = \frac{1}{N} \sum_{i,j,k} x_{i,j,k,l}^{(in)}$$

$$\sigma_l^2 = \frac{1}{N} \sum_{i,j,k} (x_{i,j,k,l}^{(in)} - \mu_l)^2$$

$i, j, k, l$  分别表示数据的批量(batch)、高、宽以及通道数， $\mu_l$  和  $\sigma_l^2$  是均值和方差， $N = \text{batch} \times \text{height} \times \text{width}$ 。 $\beta_l$  和  $\gamma_l$  是可训练的网络参数， $\varepsilon$  是一个整数，作用是使分母不为 0。

BN 层的前向传播过程就是首先计算出特征图的均值和方差,然后对输出特征图去除计算出来的均值与方差,得到标准化的数据,最后通过可学习参数的调整,得到最终的输出结果。注意的是,训练时均值和方差是对一批数据内进行的计算,而推理时是基于所有批次数据计算出的均值和方差。

### 2.2.2. 转置卷积(Deconvolution)

是对一般卷积来说的,它们是互为逆转换的关系。卷积的作用是通过高维度的向量进行低维度特征提取,而转置卷积的作用刚好与之相反,在保证图像的空间关系不变的情况下,将低维度的特征向量通过较小的空间映射到高维特征向量。如下图 3、图 4 转置卷积图所示。转置卷积在神经网络可视化上的成功应用,使得其被越来越多的工作所接受和采纳[8],比如:场景分割、生成模型等。其中转置卷积也有很多其他的叫法[9],比如: Transposed Convolution, Fractional Strided Convolution 等。

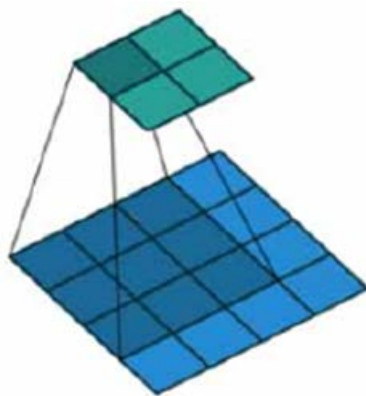


Figure 3. Transposed convolution graph 1  
图 3. 转置卷积图 1

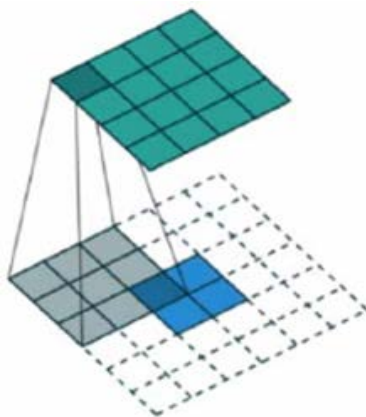


Figure 4. Transposed convolution graph 2  
图 4. 转置卷积图 2

## 3. DCGAN 网络的人脸生成架构

DCGAN 网络是在 GAN 网络的基础上进行改进的,并且该网络引入了图像的卷积操作,其中生成器和判别器用两个 CNN 网络来替代[10]。如下图 5 的 DCGAN 网络结构图所示,对于这两个 CNN 网络,做出一些调整。首先就是使用了一个全卷积网络,用步长卷积来替换掉池化层[11],目的在于可以使网络学习到下采样的方式,在生成器中主要是要使用转置卷积进行上采样,相比于固定的池化层更加灵活。

其次在生成器和判别器中增加批量归一化层，这有助于模型的稳定性、收敛性和防止过拟合。对于深层网络框架去除全连接层，在生成器中除了输出层使用 Tanh 激活函数，其他都是用 ReLU 激活函数，最后在判别器中全部使用 LeakyReLU 激活函数，如图所示。由于 DCGAN 良好的性能，基于此网络结构提出了很多新的方法；例如：pix2pix、cycle-GAN 等；同时受 DCGAN 思想的启发，Daniel 等人使用递归神经网络(RNN)去生成图像。通过交替对抗训练，DCGAN 可以生成质量很高的图像。

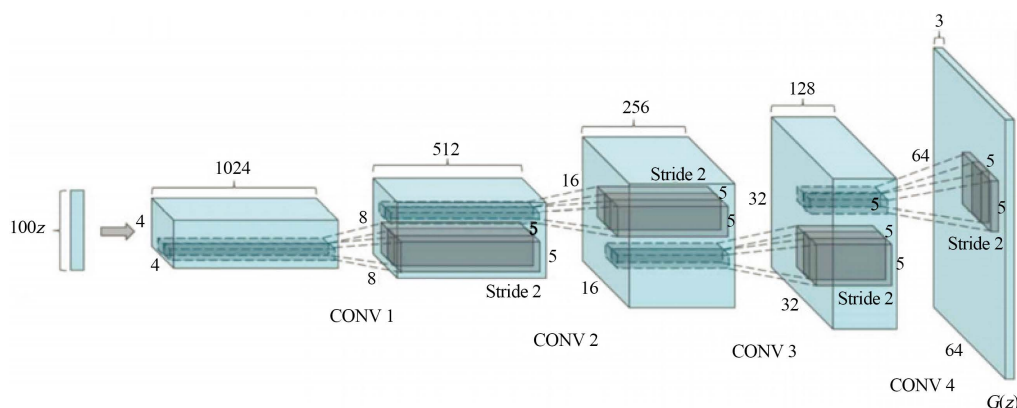


Figure 5. DCGAN network structure diagram

图 5. DCGAN 网络结构图

## 4. 实验及结果分析

### 4.1. TensorFlow 环境框架

#### 4.1.1. TensorFlow 介绍

Google 于 2015 年 11 月正式开源了机器学习和深度学习框架 TensorFlow。TensorFlow 一经开源就受到了极大的关注，在一个月之内便获得了 Github 上超过一万颗星的关注，目前在所有机器学习和深度学习项目中排名第一位，甚至在所有的 Python 项目中也排名第一，可见它在机器学习和深度学习领域发挥了多么巨大的作用。对于机器学习和深度学习的各种算法使用可以在 TensorFlow 的计算框架中获得比较好地支持[12]。

#### 4.1.2. TensorFlow 安装

1) 操作系统：Windows 10。

2) 本文是 Windows 下通过 Anaconda 安装 TensorFlow。Anaconda 是一个用于科学计算的 Python 发行版，内置了数百个 Python 经常使用的库包括 Tensorboard、NumPy、Pandas 等做机器学习或数据挖掘的库，其中可能有一些还是 TensorFlow 的依赖库。Anaconda 为研究者们提供了一个编译好的环境可以对库直接安装，不需要为安装这些库而花费大量时间和精力。

### 4.2. 结果分析

判别器和损失器的损失函数均使用二元交叉熵，变化如下图。开始的时候两个损失函数的跳跃度都很大，这说明最开始时生成对抗网络并没有找到合适的平衡点，随机产生的图片分布也没有逼近真实图像分布，随着迭代次数增加到 8000 次后，生成器和判别器的跳跃幅度降低，当迭代次数到 12,000 次以上，损失函数的值趋于平稳[13]，判别器损失函数在 0.695 上下浮动，生成器损失函数在 0.74 上下浮动。如图 6、图 7 所示：

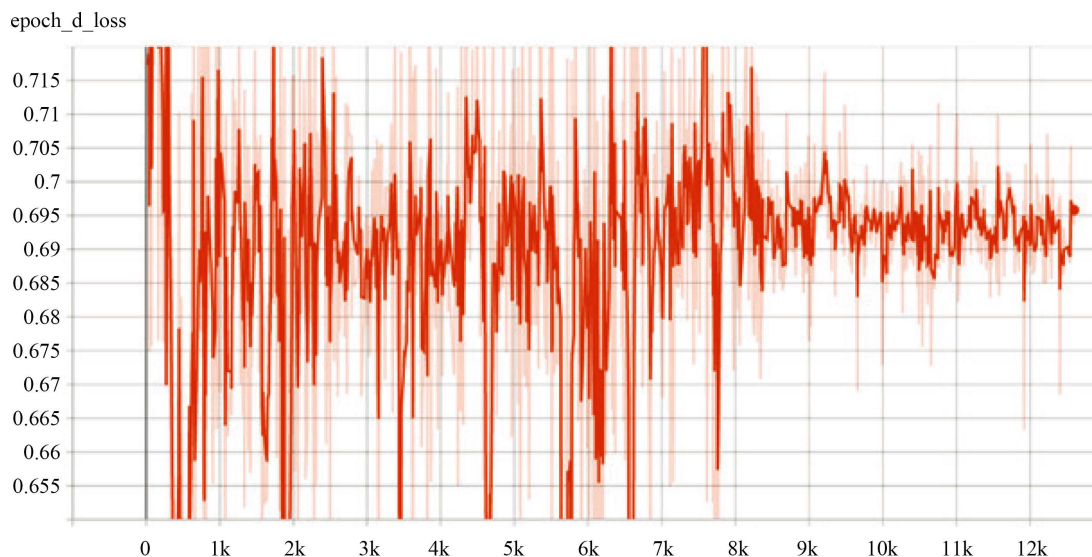


Figure 6. Generator loss function graph

图 6. 生成器损失函数图

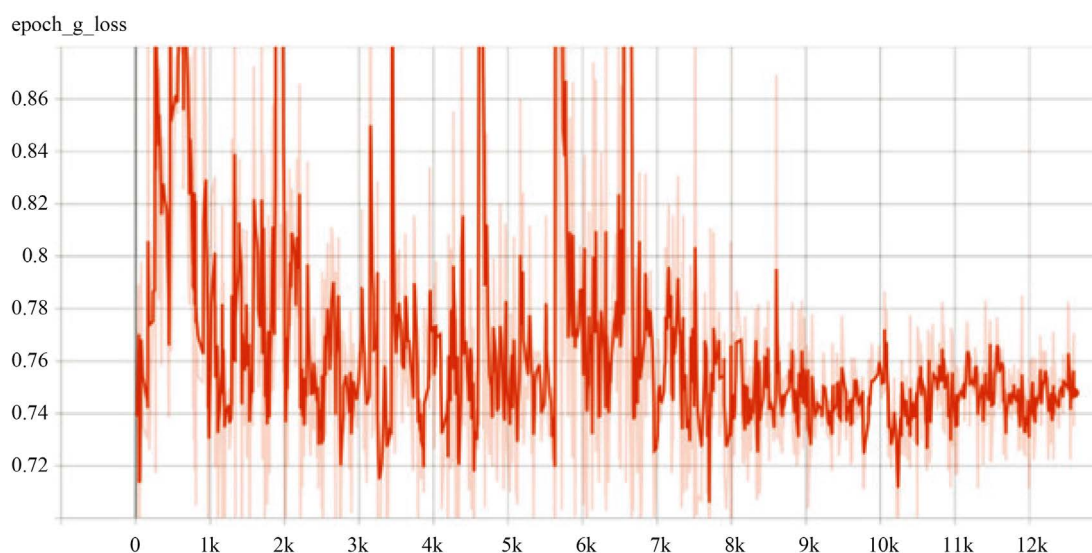
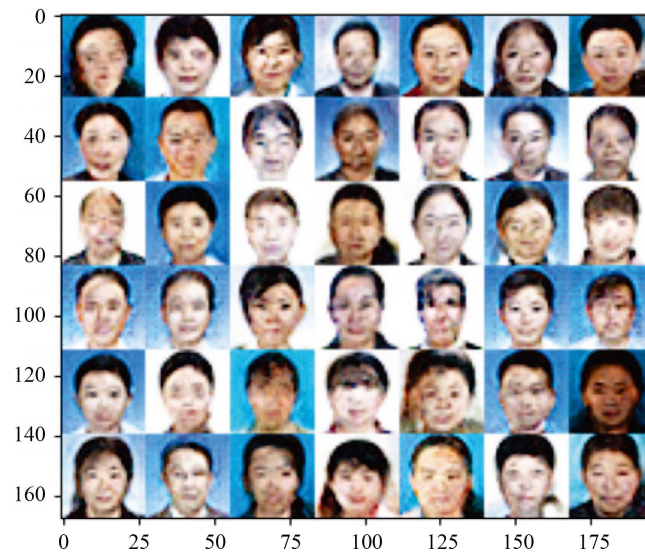


Figure 7. Discriminator loss function graph

图 7. 判别器损失函数图

图 8 和图 9 中的 84 张人脸图像均为  $32 \times 32$  像素大小。在两图中，由上到下，六行分别为训练 300、1200、1800、2100、2400、3000 代(Epoch)的图片，图 8 中的所有图片都是完全随机的输入噪声，图 9 中每一列都对应相同的输入噪声向量。由图 8、图 9 中的结果我们可以看出，随着训练代数的增加，生成出优秀图片的概率在增加；对于同一个输入噪声向量，随着训练代数的增加，图像中的斑块、五官的位置和大小，以及表情的自然和明显程度都在好转。从以上这些客观的表现来看，随着训练的迭代次数增加，网络逐渐优化。

从图 9 的前两列以及图 8 的部分图片可以看出，对于一些输入噪声向量，网络的继续迭代对输出结果并没有太大影响，因为在训练代数较低时，图片的质量就已经相对较好，根据这种现象，我们可以对输入噪声向量进行截断处理，这样有助于提高输出图片的质量。



**Figure 8.** Figure 1 of face generation results

**图 8.** 人脸生成结果图 1



**Figure 9.** Figure 2 of face generation results

**图 9.** 人脸生成结果图 2

## 5. 总结与展望

### 5.1. 总结

生成对抗网络作为一种概率生成模型，其已经被应用于很多视觉任务中，特别是在图像生成方向表现优良。本文首先从工作机理、目标函数、模型结构和训练 GAN 存在的问题以及应对策略等角度对 GAN 进行了一个详细地讨论。其次，华人人脸生成是一个比较具有挑战的项目，在日常生活中也有着非常重要的应用价值。本文使用传统的生成对抗网络思想，基于深度学习的图像处理，分析华人人脸的结构特点，应用了深度卷积生成对抗相关知识。GAN 通过一种间接的方式来对未知的分布进行建模，从而避免无监督学习中难解的推断、难解的归一化常数等问题；所以 GAN 不需要引入下界来近似似然。GAN 可

以并行地生成数据,与自回归模型相比,GAN 生成数据的速度比较快;同时 GAN 生成的图像还比较清晰。在理论上,只要是可微分的函数都能够用于构建生成器和判别器,因而 GAN 能够与深度神经网络结合来构建深度生成式模型。

## 5.2. 展望

本模型虽然在一定程度上生成了华人人脸,但是图像的自然度、清晰逼真度都还不够完善,还有诸多改进之处:可解释性比较差,因为最终生成器学到的数据分布只是一个端到端、黑盒子一样的映射函数,而且没有显式的表达式,在实际应用中 GAN 比较难以训练,由于 GAN 需要交替训练生成器和判别器两个模块,因此两者之间的优化需要很好地同步,可能发生模式崩溃的现象,导致生成器学到的模式仅仅覆盖真实数据中的部分模式,使得生成样本的多样性变低,训练不稳定,神经网络需要良好的初始化,否则可能找不到最优解,导致学到的分布距离真实数据的分布仍然很远,并且无法根据损失函数的值来判断模型的收敛性。优化的目标函数是两个分布上的 Jensen-Shannon 距离,但这个距离存在一个问题,当两个分布的样本空间并不完全重合时,无法进行梯度下降。在本文模型应用和训练过程中,遇到的一些困难是对数据集的选择预处理和对得到结果后的照片比对工作。由于对生成照片的评判不同,最终采样效果也有所不同。

## 基金项目

西藏大学 2020 年国家级大学生创新训练项目“大学生综合能力智能分析系统”(202010694002X)。

## 参考文献

- [1] 唐婷. 人脸识别中基于虚拟样本的稀疏描述研究[D]: [硕士学位论文]. 长沙: 湖南大学, 2014.
- [2] 孙亮, 韩毓璇, 康文婧, 葛宏伟. 基于生成对抗网络的多视图学习与重构算法[J]. 自动化学报, 2018, 44(5): 819-828.
- [3] 张柯, 白富生, 吴至友, 皮家甜, 赵立军. 基于对抗生成网络的人脸照片去网纹技术[J]. 重庆师范大学学报(自然科学版), 2019, 36(6): 110-118.
- [4] 李焯, 许乾坤, 李克东. 面向图像复原的残差密集生成对抗网络新方法[J]. 小型微型计算机系统, 2020, 41(4): 830-836.
- [5] 李浩鹏. 基于生成式对抗网络的图像翻译研究及应用实现[D]: [硕士学位论文]. 武汉: 华中科技大学, 2018.
- [6] 王万良, 李卓蓉. 生成式对抗网络研究进展[J]. 通信学报, 2018, 39(2): 135-148.
- [7] 贺温磊, 王朝立, 孙占全. 基于生成对抗网络的遥感图像超分辨率重建[J/OL]. 信息与控制, 2021(2): 195-203. <https://doi.org/10.13976/j.cnki.xk.2021.0181>, 2021-05-09.
- [8] 徐鲁冰. 基于体素分析的早期阿尔茨海默病患者小脑形态学研究[D]: [硕士学位论文]. 哈尔滨: 哈尔滨工业大学, 2018.
- [9] 刘腊梅, 王晓娜, 刘万军, 曲海成. 融合转置卷积与深度残差图像语义分割方法[J/OL]. 计算机科学与探索: 1-12. <http://kns.cnki.net/kcms/detail/11.5602.tp.20210416.1700.008.html>, 2021-05-09.
- [10] 任飞凯, 邱晓晖. 基于 LBP 和数据扩充的 CNN 人脸识别研究[J]. 计算机技术与发展, 2020, 30(3): 62-66.
- [11] 鲁力, 王洁, 韩要昌, 吴亚晖. 改进的 DCGAN 在 SAR 图像分类中的应用[J]. 控制工程, 2020, 27(3): 561-566.
- [12] 王鑫磊. 基于深度卷积生成对抗网络的图像修复研究与应用[D]: [硕士学位论文]. 重庆: 重庆大学, 2018.
- [13] 朱瑞金, 郭威麟, 龚雪娇. 考虑天然气和电负荷之间相关性的短期电负荷预测[J]. 电力系统及其自动化学报, 2019, 31(8): 27-32.