

生成对抗网络图像处理综述

朱秀昌,唐贵进

(南京邮电大学 江苏省图像处理与图像通信重点实验室 江苏 南京 210003)

摘要: 2014 年提出的生成对抗网络(Generative Adversarial Networks, GAN) 是近年来神经网络领域中为数不多的一项新锐技术。GAN 在常见生成模型的基础上增加了一个判别模型,以形成巧妙的对抗学习机制,使它能够产生更高质量的图像。近年来各种改进型 GAN 在图像处理领域得到广泛应用,不但覆盖了几乎所有传统图像处理领域,还包括一些新应用,如图像编辑、图像翻译、风格转移等,普遍取得了胜过传统方法的良好结果。文中在简要分析 GAN 的系统结构、对抗生成和网络训练的基础上,重点介绍了为提高 GAN 性能、克服现存缺陷和满足不同应用而出现的多种改进型 GAN,如 DC-GAN、W-GAN、Big-GAN 等。尽管如此,目前 GAN 尚处于初始发展阶段,将来的前途不可估量。

关键词: 深度学习; 生成对抗网络; 图像处理; 生成模型; 判别模型

中图分类号: TN919.8 **文献标志码:** A **文章编号:** 1673-5439(2019)03-0001-12

A survey on generative adversarial networks in image processing

ZHU Xiuchang, TANG Guijin

(Jiangsu Province Key Lab on Image Processing & Image Communication, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

Abstract: Generative adversarial networks (GAN) proposed in 2014 is one cutting-edge technique of several deep learning technologies in the neural networks in the past few years. A discriminative model is added to familiar generative model by GAN to form an adversarial competitive strategy in subtle way in order to produce higher quality images. Recently, many improved GANs are widely applied in image processing, not only covering nearly all applications of traditional image processing, but also including other new applications, such as image editing, image translation, style transfer, etc., and have gotten better performance than that the traditional methods got. Based on brief analysis on GAN system architecture, adversarial generating and networks training, this paper stresses on introducing several schemes of improved GANs, such as DC-GAN, W-GAN, Big-GAN, and so on, whose target is to promote GAN's performance, overcome GAN's defects, and adapt the environment requirements for deferent applications. Nowadays, GAN techniques are still in initial stage of vigorous development, and have a bright future for us.

Keywords: deep learning; generative adversarial networks; image processing; generative model; discriminative model

收稿日期: 2019-04-02 本刊网址: <http://nyzr.njupt.edu.cn>

作者简介: 朱秀昌,男,教授,博士生导师, zhuxc@njupt.edu.cn

引用本文: 朱秀昌,唐贵进. 生成对抗网络图像处理综述[J]. 南京邮电大学学报(自然科学版) 2019, 39(3): 1-12.

2014年由加拿大 Goodfellow 博士提出的“生成对抗网络”(generative adversarial networks, GAN)^[1]是一种基于概率和统计理论,借鉴博弈论的对抗思想,用深度学习网络来生成数据样本的方法,已迅速成为人工智能研究和应用领域中一种非常重要的模型和工具。

GAN 网络提供了一种高效数据生成的深度学习方法,不需要或需要很少的标注数据,通过生成网络和判别网络之间的竞争来获得高质量的生成数据。GAN 学习模型可适用于不同类型的信号处理,图像处理是其中一个重要的领域,例如图像合成、语义图像编辑、风格迁移、超分辨率重建、图像分类等。

本文首先简述生成模型和判别模型,然后分析基本 GAN 的原理、组成和特点,在此基础上重点介绍为改进基本 GAN 缺陷、适应具体应用的多种改进型 GAN,并由此介绍 GAN 在图像处理中常见的几种应用,最后对当前 GAN 中存在的开放问题和今后发展趋势进行了简单的小结。

1 生成模型和判别模型

要理解生成对抗网络,首先要了解人工神经网络中的生成模型(generative model)和判别模型(discriminative model),这两个模型都涉及到明确或隐含的数据分布问题。

1.1 数据分布

统计信号处理的核心问题之一就是密度估计,即获得真实世界中数据的一种表示,隐含的或明显的、参数的或非参数的表示,这也是 GAN 发展的关键动因。在 GAN 中的生成数据分布一般指它们的潜在概率或概率密度函数。GAN 正是通过计算生成数据潜在分布和相应真实数据分布之间的某种相似性来进行学习的。

对于图像数据分布问题,可将图像看作是一个高维空间,具体的一幅图像则是该高维空间的一个点(或样本),图像数据的分布则是高维空间的数据分布。例如,所有 64×64 尺寸、256 灰度级图像组成了一个 $64 \times 64 = 4\,096$ 维空间,任意一幅 256 灰度级 64×64 的图像只是该高维空间的一个点。不难理解,实际有意义(可以理解的)的某一类图像(如人脸)只占极少数的点。从理论上说,某一类真实图像的分布仅是高维空间的一个潜在的低维流形(manifold)。那么我们要求生成的图像也位于这个流形上,其分布也应尽可能和真实

数据分布一致。

1.2 判别模型

在机器学习中,学习得到的判别模型的作用是对输入数据 x 进行判断,它到底是属于哪一类或哪几类数据。例如 GAN 的判别模型,需判断输入的是真实的图像还是由模型生成的图像。判别模型一般可以通过决策函数 $y = f(x)$ 或者条件概率 $p(y|x)$ 来判断输入 x 是否属于 y 类。

决策函数判别模型是一种确定性的判别方法,可通过训练图像集(有标签)来确定特征空间中不同类别的分界线。此后输入图像特征落在哪个类别的范围内,它就属于此类别。典型的方法有支持向量机(SVM)、感知机(perceptron)、 k -近邻(k -nearest neighbor)等。

条件概率判别模型是一种基于统计特性的判别方法,通过训练图像(特征)获得条件概率分布,即样本 x 属于类别 y 的概率 $p(y|x)$ 分布。此后计算输入图像(特征)属于各个类别的条件概率,概率最大的那个类别就是该图像所属的类别。

决策函数判别方法较为简单,只要寻找不同类别数据之间的差异,不需要知道这些数据的概率分布。但这种方法往往难以对付复杂的类别划分,因此 GAN 中的判别器多采用条件概率判别模型。

1.3 生成模型

在机器学习中,生成模型学习到的是给定输入 x 产生输出 y 的生成关系,即联合概率分布 $p(x, y)$ 。根据贝叶斯定理,在得知 $p(x)$ 的情况下,还可以由 $p(x, y)$ 得到相应的条件概率 $p(y|x)$ 。因此生成模型不仅可以用于图像的生成,还可以和判别模型一样用于图像的分类。对给定的训练图像,生成模型可以自动学习到其内部分布,能够解释给定的训练图像,同时生成新的类似于训练图像的样本,这就是所谓的数据“生成”。在训练过程中,生成模型会尽力挖掘数据背后更为简单的统计规律,从而能够生成类似的数据。

如上所述,两种模型各有其用,而生成模型更具优势,即使没有标签,也有可能理解和解释输入数据的基本结构。这是非常有用的,因为标记数据是非常不易的,而未标记数据相对是非常丰富的。

2 GAN 基础

传统的概率生成模型,如最大似然估计、马尔可夫链等方法在实现时至少存在两点困难:一是

对真实数据进行建模需要大量先验知识(如大量的样本),建模的好坏直接影响生成模型的性能;二是真实数据往往非常复杂,拟合模型所需计算量往往非常庞大。因此,相对于判别模型来说,生成模型研究的进展较为缓慢,性能也不够理想。针对上述两点困难,生成对抗网络巧妙地使用对抗训练机制对生成网络和判别网络进行轮番训练、优化,学习样本数据内在统计规律,并基于此生成酷似样本的新数据。

2.1 GAN 的结构

基本 GAN 主要由生成模型和判别模型两个网络组成,如图 1 所示。这两个模型可由不同形式的神经网络实现,只要它们可以实现函数功能,将数据从一个空间映射到另一个空间。

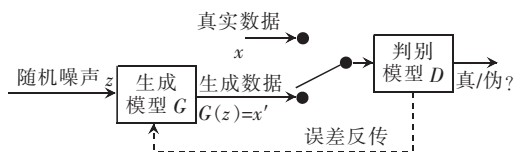


图 1 GAN 结构

GAN 运作的核心思想来源于博弈论(Game Theory)的纳什均衡(Nash equilibrium),两个模型之间经不断竞争、优化而逐渐达到平衡。打个形象的比喻,生成器(generator, G)是一个图像伪造者,尽量去学习真实图像的分布,目的是制造赝品图像。判别器(discriminator, D)是一个图像鉴别者,接收赝品和真实图像,其目的是将它们区分开。

GAN 工作开始时生成模型产生一幅图像去“欺骗”判别模型,然后判别模型判断此图像是真是假,将判别结果的误差反传给这两个模型,各自改进网络参数,提高自己的生成能力和判别能力。如此反复,不断优化,两个模型的能力越来越强,最终达到纳什均衡状态:生成器所生成的图像达到以假乱真的程度,判别器已经分辨不出。

2.2 GAN 的目标函数

参照图 1 的 GAN 结构,设输入到生成器的随机数据 z 服从 $p_z(z)$ 分布,生成器的输出数据 $G(z) = x'$ 服从 $p_g(x)$ 分布。输入到判别器的可能是生成数据 $G(z)$,也可能是服从 $p_{data}(x)$ 分布真实数据 x ,判别器输出是 x 的函数 $D(x)$,通常表示 x 为真实数据的概率,由此建立 GAN 优化的目标函数,它是 D 和 G 的函数 $V(G, D)$,即

$$V(G, D) = E_{x \sim p_{data}} \log D(x) + E_{x \sim p_g} \log(1 - D(x)) \quad (1)$$

其中 E 表示数学期望。目标函数的最优解即求在

$V(G, D)$ 取得极小的极大值时的 D^* 和 G^* 的参数:

$$G^* = \arg \min_G [\max_D V(G, D)] \quad (2)$$

如果 D 能够正确判断,将真实数据尽量判为 1,生成数据尽量判为 0,会使函数 $V(G, D)$ 值尽量增大;反之,如果 D 出现错判,则会使 $V(G, D)$ 值迅速减小。因此,调整 D 的参数,希望 $V(G, D)$ 值越大越好,达到极大值时的 $D(x) = D^*(x)$,即

$$D^*(x) = \arg \max_D V(G, D) = \frac{p_{data}(x)}{p_{data}(x) + p_g(x)} \quad (3)$$

将 D^* 带入 $V(G, D)$ 后用 $C_g(x)$ 表示:

$$C_g(x) = V(G, D^*) = KL[p_{data}(x) \parallel \frac{p_{data}(x) + p_g(x)}{2}] + KL[p_g(x) \parallel \frac{p_{data}(x) + p_g(x)}{2}] - \log 4 \quad (4)$$

其中 $KL(p(x) \parallel q(x))$ 表示同一变量 x 的两个概率分布的差异程度,又称之为 KL 散度(Kullback-Leibler divergence)。

当 $D(x)$ 固定后, $V(G, D)$ 只和式(1)最右边的那一项有关。通过调整 $G(x)$,使得 $D(G(x))$ 的输出越大才越能“骗过” D ,等效于希望 $1 - D(G(x))$,或 $V(G, D^*)$ 或 $C_g(x)$ 的值越小越好,让 D 分不清真假数据。表示 $C_g(x)$ 的式(4)中 KL 值总是大于 0 的,只有在两个分布相同时,即 $p_{data}(x) = p_g(x)$, KL 值到达最小值 0, $C_g(x)$ 取得极小值为 $-\log 4$,再由式(3)得 $D^*(x) = 1/2$,表明判别器输出为 0.5,分不清输入数据的真假,生成器的性能达到最佳。

进一步的研究证明, GAN 的目标函数是存在全局最优解的,最优化迭代是可以收敛的,当生成数据与真实数据分布完全一致时,目标函数的最优解是 $V(G, D)$ 曲面上的一个鞍点(saddle point)。

2.3 GAN 的训练

GAN 一般采用交替的对抗训练的方式,在一轮训练中:先固定生成器 G ,训练判别器 D ,使得目标函数 $V(G, D)$ 最大化;然后固定判别器 D ,训练生成器 G ,使得目标函数最小化。如此反复多轮训练,直至 p_{data} 等于或约等于 p_g 时达到全局最优解。

在 GAN 的训练中一般采用误差反传、梯度下降的迭代方式来调整生成器和判别器的网络参数。在同一轮参数更新训练中,并不是对所有数据进行,而是对数据集里的一小批(mini-batch)数据进行处理。而且,一般在一轮训练中需对 D 的参数更新 k 次再对 G 的参数更新 1 次。整个 GAN 训练常常需成千

上万轮,甚至更多。

在训练中,对输入的数据或网络运行中层间传送的数据可进行批量规范化(batch-normal)处理,就是将输入到每一层或某几层的批量数据标准化为0均值、单位方差的规范化数据。这样,有助于解决初始化不良导致的训练问题,有助于“拉大”网络深层处的梯度数据的幅度,防止梯度信息在反传中逐渐消失。

2.4 GAN 的优势和不足

和传统的生成模型比较,生成对抗网络 GAN 有多项突出的优势,主要如下:

(1) GAN 不要求对数据预设分布,直接进行采样和推断,理论上可生成完全逼近真实分布的数据。在很多情况下,数据的分布是不可知的,这时 GAN 就显得格外有用,非常适用于无监督和半监督学习任务。

(2) 由于 GAN 内部引入了非常有效的对抗训练机制,使得生成器 G 的参数更新驱动不是直接来自数据样本,而是来自判别器 D 的误差反传,其优化根据和路径都很明确。

(3) GAN 框架非常灵活,各种类型的损失函数和约束条件都可以整合到此框架中,有利于针对不同任务设计出不同类型的损失函数和优化方法。

(4) GAN 可以和现有的卷积神经网络(CNN)、循环神经网络(RNN)等深度网络结合使用,逼近任何可微函数,形成参数化生成模型和判别模型。

处于初始阶段的 GAN 技术还存在若干不足:

(1) GAN 的可解释性较差,生成模型的分布 $p_g(x)$ 没有显式表达。

(2) 训练比较困难, D 与 G 之间需要很好协同,常常出现训练的不稳定或模式坍塌(model collapse)现象。

(3) 不能保证 GAN 训练都能达到纳什均衡。

(4) 对生成质量的评价尚无统一有效的准则。

3 几种改进的 GAN

近年来针对 GAN 不足之处的改进一直在不断进行当中。到目前为止,在 GitHub 网站(开源软件等托管平台)上已公布 400 多种改进型 GAN,这里介绍的只是几种比较认可的改进类型。

3.1 C-GAN 和 Info-GAN

3.1.1 C-GAN

基本 GAN 系统的图像生成过程过于自由,尤其是对于较大的图像,难以控制。为了解决这个问题,

Mirza 等^[2]提出了一种条件生成对抗网络(C-GAN)的改进方法,给 GAN 加上约束条件,让生成的样本符合预期。这个条件可以是图像的类别标签,也可以是图像的语义等。C-GAN 的结构和 GAN 基本一致,如图 2 所示。反映到目标函数中,就是在式(1)中用条件概率 $D(x|C)$ 替代原有的无条件概率 $D(x)$ 。

C-GAN 对模型增加约束条件,引导数据生成过程。如果条件变量 C 是类别标签,可以看作 C-GAN 是把无监督的 GAN 变成有监督模型的一种改进。Mirza 等的工作是在 MNIST 手写数字数据集上以类别标签为条件变量,生成指定类别的图像。这个简单、直接的改进非常有效,所生成的图像质量明显优于 GAN 的生成图像,并启示后续的 GAN 改进工作,如下面的 Info-GAN。

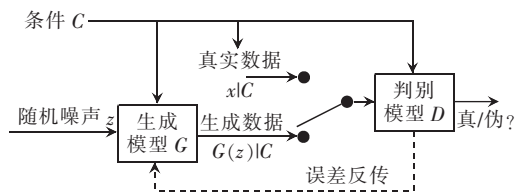


图 2 C-GAN

3.1.2 Info-GAN

在基本 GAN 中,生成器输入的随机噪声 z 的维度和特征之间没有明显的对应关系,呈现为一种高度混杂(entangled)方式。例如,在手写数字的生成过程中,我们无法得知什么噪声 z 可以生成数字 1,什么 z 生成数字 2,……。 z 的这种无序性常常严重影响了 GAN 训练的稳定和实际应用。

为了解决这一问题,Chen 等^[3]提出了信息最大化生成对抗网络(Info-GAN),其结构和 C-GAN 类似,如图 3 所示。Info-GAN 在生成器输入中除了原先的噪声 z 外,还增加了一个隐含码(latent code) c 。 c 通常由若干个子码组成,例如在生成 MNIST 数据时,使用了三个子码串接起来形成隐含码 c : 子码 c_1 用 10 个离散码对应数字 0,1,……,9,每个类别的概率都是 0.1;子码 c_2 和 c_3 为连续码,服从 $[-2, 2]$ 之间的均匀分布,分别表示数字的倾斜角度和笔画的粗细。我们希望 c 和生成数据之间密切相关,这样就可以达到通过 c 控制输出 $G(z)$ 的目的。

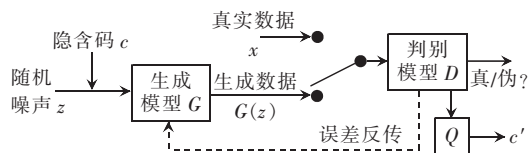


图 3 Info-GAN

生成数据 x 与隐含编码 c 之间关联程度的大小可用它们之间的互信息 (mutual information) 表示, 为了使 x 与 c 之间关联密切, 所以我们需要最大化它们的互信息 $I(c; x)$, 据此 Info-GAN 在基本 GAN 代价函数上加了一个互信息的正则化项, 即

$$\min_G \max_D V_I(D, G) = V(D, G) - \lambda I(c; G(z, c)) \quad (5)$$

其中 $\lambda I(c; G(z, c))$ 为互信息正则化项, 参数 λ 控制正则化强度。困难在于实际计算互信息 $I(c; x)$ 时需要用到难以获得的后验分布 $P(c|x)$, 文献[3]引入一个辅助分布 $Q(c|x)$, 它的下界可以用作 $P(c|x)$ 的近似。在图 2(b) 中, Q 模块表示辅助分布 $Q(c|x)$ 的计算, 可与 D 共享卷积层, 以期减少额外的计算负担。

3.2 DC-GAN

最早 GAN 的生成器和鉴别器使用的是全连接神经网络, 这类结构适用于比较简单的图像数据集。由于 CNN 特别适合处理图像数据, 对于较复杂的图像生成任务, 研究人员自然将目光从全连接网络转向卷积网络。在这一思想指导下, Radford 等^[4]提出了深度卷积 GAN (deep convolutional GAN, DC-GAN), 它的改进主要在网络结构上, 将 CNN 与 GAN 结合起来, 把基本 GAN 中的生成器 G 和判别器 D 换成了两个深度 CNN, 并对 CNN 结构作了一些改变, 以提高生成样本的质量和收敛的速度。

DC-GAN 的判别器和生成器的结构大致对称, 和 CNN 比较, 网络中没有池化 (pooling) 层和全连接层。生成器使用分步幅 (fractional-strided) 卷积实现上采样, 判别器采用普通的整步幅卷积实现下采样, 以此增加训练的稳定性。生成器和判别器中几乎每一层都对数据进行了批量规范化处理, 有利于解决初始化差别问题, 有助于梯度信息传播, 防止生成器把所有的样本都收敛到同一个点。

DC-GAN 虽然改善了生成器和判别器的网络结构, 但是对 GAN 训练稳定性来说并没有从根本上解决问题。训练的时候仍然是先训练 D 多次, 再训练 G 一次, 需要小心地平衡 G 、 D 的训练进程。

3.3 W-GAN 和 W-GAN-GP

3.3.1 W-GAN

前述式 (4) 用 KL 散度表示的 G 优化的目标函数 C_g 也可以用类似的 JS (Jensen-Shannon) 散度表示:

$$C_g = 2\text{JS}(P_{\text{data}} \| P_g) - 2\log 2 \quad (6)$$

在基本 GAN 的训练中, 如果判别器训练得很好, 应该对生成器的提高有很大作用, 但实际常常会出现判别器训练得越好, 生成器的梯度消失越严重的现象。Arjovsky 等^[5]认为生成器随机初始化分布与真实分布都是高维空间的低维流形, 互不重叠或重叠部分可忽略的概率非常高, 而代价函数中衡量分布距离的 JS 散度不合理, 在这种情况下 JS 散度都等于常数 $\log(2)$, 导致梯度为 0, 生成器得不到优化所必须梯度信息。

为了解决上述问题, Arjovsky 等^[6]提出了 W-GAN (Wasserstein GAN), 用 W (Wasserstein) 距离代替普通 GAN 中的 JS 散度, 并由此导出 W-GAN 的优化目标函数为:

$$V(G, D) = \max_{D \in 1\text{-Lipschitz}} \{ E_{x \sim p_{\text{data}}} [D(x)] - E_{x \sim p_g} [D(x)] \} \quad (7)$$

这个表达式的求解结果就是 p_g 与 p_{data} 之间的 W 距离 (证明见文献[6])。值得注意的是, 式 (7) 中对 D 加上了 1-Lipschitz 限制, 限制了 W 距离的上下界。K-Lipschitz 限制就是要求可微函数 $D(x)$ 的增量不大于自变量 x 增量的 K 倍, 限制函数 $D(x)$ 的增长速率。由此可见, W 距离相比 JS 散度的优越性在于, 即便 p_g 与 p_{data} 分布没有重叠, W 距离仍然能够反映它们的远近, 具有优越的平滑过度特性。

3.3.2 W-GAN-GP

虽然 W-GAN 相对于 GAN 有了很大的改进, 但仍然存在权重限制 (weight clipping), 未能将 D 完全限制在 1-Lipschitz 内, 导致训练时参数集中化和参数调整中的梯度爆炸 (更新变化太大)、梯度消失 (无更新信息) 等问题。

由于 $D \in 1\text{-Lipschitz}$ 是和 D 的梯度模 $\|\nabla_x D(x)\| < 1$ 是等价的, 于是 Gulrajani 等^[7]提出一种梯度惩罚 W-GAN (W-GAN-Gradient Penalty, W-GAN-GP) 方案, 在原来 W-GAN 的 D 优化函数中增加一个“梯度惩罚”项来满足判别器的梯度模不超过 1 这个要求, 如式 (8) 第 3 项所示:

$$V(G, D) \approx \max_D \{ E_{x \sim p_{\text{data}}} [D(x)] - E_{x \sim p_g} [D(x)] - \lambda E_{x \sim p_{\text{penalty}}} [\|\nabla_x D(x)\| - 1]^2 \} \quad (8)$$

由于惩罚项对所有 x 有效, 会让惩罚的计算量变得非常高。考虑到整个 W-GAN 的目的是让 p_g 渐

渐向 p_{data} 靠拢,位于 p_g 和 p_{data} 之间的区域会对判别器产生实质性的影响。因此将惩罚项中 x 的范围缩小为介于 p_g 和 p_{data} 之间的区域,以减少大量不必要的计算。

3.4 f-GAN

如前所述,当 p_g 与 p_{data} 无重合时,即使两者的差距在缩小,但是它们的 JS 散度始终为 $\log(2)$,缺少梯度信息,没有更新 G 参数的动力,生成器得不到改进。因此,解决 GAN 这个问题的关注点仍然在于改进度量两个分布散度或距离的方法。Nowozin 等^[8]从这里着手,提出一种 f-散度概念和相应的 f-GAN 系统。f-散度不是某一种具体的散度,而是对散度概念的泛化。这样, f-GAN 的训练就不再局限于 JS 散度,可以由最小化 f-散度得到。

设 $p(x)$ 和 $q(x)$ 是同一空间两个不同的分布,则它们的 f-散度定义为:

$$D_f(p \parallel q) = \int_x q(x) f(p(x)/q(x)) dx \quad (9)$$

其中,要求 $f(x)$ 是凸函数,同时 $f(1) = 0$ 。根据凸函数的性质可证明 D_f 非负。假设对于所有的 x 来说,都有 $p(x) = q(x)$,则有 $D_f(p, q) = 0$,这意味着两个相同分布的 f-散度为 0, D_f 的最小值。也就是说,只要两个分布稍有不同,就能通过 D_f 的值反映出来。显然, f-散度随着采用不同的 $f(x)$ 而变化,如当 $f(x) = x \log(x)$ 时, f-散度即为前述的 KL-散度。

在 f-GAN 的目标函数中需用到凸函数 $f(x)$ 的 Fenchel 共轭函数 $f^*(t)$ 的概念:

$$f^*(t) = \max_{x \in \text{dom}(f)} \{xt - f(x)\} \quad (10)$$

其中, $x \in \text{dom}(f)$ 表示 x 属于函数 f 的定义域。又因为 Fenchel 共轭函数是相互的,所以 $f^*(t)$ 的 Fenchel 共轭函数就是 $f(x) = \max_{t \in \text{dom}(f^*)} \{tx - f^*(t)\}$ 。

在 $f(x)$ 中,令 $x = p(x)/q(x)$, $t = D(x)$,再由前面的 f-散度定义经推导可得 f-GAN 生成器 G 的更新:

$$G^* = \arg \min_G \max_D \{E_{x \sim p_{data}} [D(x)] - E_{x \sim P_G} [f^*(D(x))] \} = \arg \min_G (P_{data} \parallel P_G) \quad (11)$$

可见,更新生成器 G 的过程,实际就是不断地减小 f-散度的过程。在普通 GAN 的训练中,每一步优化处理都包括判别器 D 的多次更新和生成器

G 的一次更新。而在 f-GAN 中,只需要一步即可更新 D, G 。而且可以选任意一种 f-散度,或者说任何满足上述条件的 f-散度都可以放到 f-GAN 的架构中。

3.5 Big-GAN

Brock 等^[9]提出一种新颖的 Big-GAN (Large Scale GAN) 模型,训练了迄今为止规模最大的生成性对抗网络,生成的图像十分清晰、逼真、多样,人眼已经难以辨别其真伪了。Big-GAN 系统主要改进在以下三方面:

3.5.1 增大 batch 尺寸

为了产生逼真、精细的图片,需要提升 GAN 模型的规模,如增加 batch 尺寸和卷积的通道数。生成模型的实验表明,当 batch 尺寸增加 8 倍时,生成图像的性能起始分 (Inception Score, IS) 提高近一倍。分析认为增加 batch 尺寸引起质量改善可能是每批次覆盖更多模式的结果,为生成网络和判别网络提供了更好的梯度信息。当然,增大 batch 尺寸会使模型训练的稳定性下降,因而并不能一味地增加 batch 的尺寸,需要适量考虑。

3.5.2 采样截断和多层输入

Big-GAN 的采样截断是一种简单的采样技巧,可以对生成样本的多样性与保真性进行适当的调节。这种方法通过对 $N(0, 1)$ 分布的输入 z 采样设置阈值的方式来截断采样,其中超出范围的值被重新采样以落入阈值范围内。这个阈值可以根据生成质量、生成多样性等指标决定。

Big-GAN 采用了分层隐含空间输入技术,将噪声向量 z 送到 G 的多个层,而不是如普通 GAN 仅将 z 直接输入到生成网络的初始层。

3.5.3 模型稳定性的控制

文献[9]发现并分析了 Big-GAN 特有的不稳定性不仅来自 G 或 D ,还来自它们在抗性训练过程的相互作用。为了控制模型的稳定性,对生成器和判别器增加适当的正则化^[10],在训练期间监测网络的一系列权重、梯度和损失统计数据,以寻找可能预示训练崩溃开始的指标。当然,训练稳定性的获得牺牲了模型的部分性能。

在 128×128 分辨率的 ImageNet 数据库训练后, Big-GAN 取得了优秀的成绩,与以往最好的成绩相比较, IS 从 52.52 提升到 166.3。图 4 中是由 Big-GAN 生成的图像,其背景和纹理都生成得如此逼真。

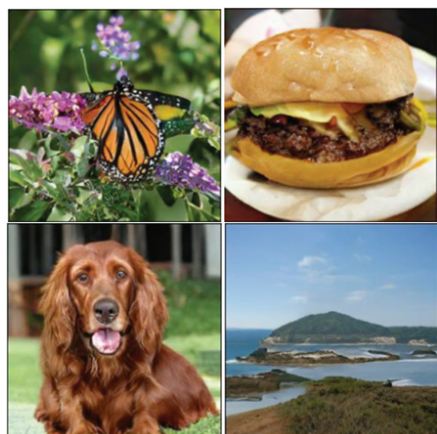


图4 Big-GAN 生成的图像

3.6 EB-GAN

Zhao 等^[11]提出的基于能量的生成对抗网络(Energy Based GAN, EB-GAN)的结构如图5所示。它和普通GAN的主要不同之处在于它的判别器 D 主要由自编码器(auto-encoder)构成,不再依靠输入图像分布的差异来分辨真实数据还是生成数据,而是依靠重构图像质量的高低来判别输入图像的真伪。

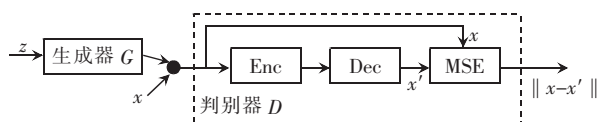


图5 EB-GAN 结构

EB-GAN 的判别器的网络参数可以预先用真实图像训练好。此后如果输入是来自真实数据集的图像,自编码器的输出和输入图像非常相似,即重构图像质量较高;否则,如果输入是来自生成器的图像,输出图像就不会和输入图像相似,即重构图像的质量较低。我们就能用输入图像 x 与输出图像 x' 差值模 $\|x - x'\|$,即重构误差能量的大小作为评判输入图像 x 质量高低的标,能量越低意味着输入图像越有可能是真实图像,反之则很可能是生成图像。EB-GAN 的实验表明,在训练期间比常规GAN收敛更快,稳定性更好,生成图像的分辨率更高。

3.7 SR-GAN 和 ESR-GAN

3.7.1 SR-GAN

GAN 图像超分辨率技术是指从一幅低分辨率(Low Resolution, LR)图像用GAN网络生成对应的高分辨率(High Resolution, HR)图像,相当于在LR图像上采样时用训练模型推断实际HR图像的细节。由Ledig等^[12]提出的超分辨率生成对抗网络(Super Resolution GAN, SR-GAN)是直接针对这一

领域应用,用GAN网络生成高分辨率图像的方法。SR-GAN 的本质就是通过卷积神经网络,用一系列低分辨率图像和对应的高分辨率图像作为训练数据,学习一个从低分辨率图像到高分辨率图像的映射函数。

SR-GAN 采用的常规GAN结构,其生成网络和判别网络都是由深度卷积网络构成。生成网络包含5个残差块(residual block),每个残差块包含两个 3×3 、64特征的卷积层,后接批量规范化(Batch Normalization, BN)层和PReLU激活函数。残差块后面两个2倍的分步幅卷积层用来增大特征尺寸。判别网络部分包含8个卷积层,随着网络层数加深,特征个数不断增加,特征尺寸不断减小,最终通过两个全连接层和一个sigmoid激活函数得到判断为自然图像的概率。

SR-GAN 的损失函数为感知损失(perceptual loss)函数,它包括加权的两部分,一部分是内容损失(content loss)函数,另一部分是对抗损失(adversarial loss)函数。内容损失函数采用特征空间的最小均方差来表示。对抗损失函数和判别器输出的概率有关。文献[12]在4倍放大时重建图像显得细节丰富、真实感很强。和以往的超分辨率重建方法比较,SR-GAN 得到的PSNR和SSIM值并不是最高,但主观评价分(Mean Opinion Score, MOS)最高。

3.7.2 ESR-GAN

尽管SR-GAN模型重建的HR图像质量已经不错,但和真实图像比较仍有差距。为此,Wang等^[13]对SR-GAN进行了改进,称为增强SR-GAN(Enhanced SR-GAN, ESR-GAN)。ESR-GAN 主要在网络结构、判别器和感知损失等三方面进行了改进。

文献[13]参考SR-GAN网络,将超分辨率残差网络(SRResNet)作为基本结构。对生成器 G ,ESR-GAN 去掉所有的批量规范化(BN)层,以防止训练中图像不时出现的“伪影”,减少计算复杂度和内存占用;将原来的基本残差单元更换为残差密集块中残差(Residual in Residual Dense Block, RRDB)单元。ESR-GAN 借鉴Ra-GAN(Relativistic average GAN)的方法,将判别器 D 换成相对平均判别器(Relativistic average Discriminator, RaD)^[14],让判别器估计真实图像相对生成图像更逼真的概率。ESR-GAN 采用一种更适合于超分辨的感知域损失,这样可以为重建图像提供更尖锐的边缘,有利于克服重建图像与真实图像亮度不一致等问题。

4 GAN 的典型应用

在图像处理领域, GAN 作为一种新型的生成模型, 主要用于真实数据分布的建模和生成, 包括对图像的分类、生成、转换、预测等处理。GAN 可以凭借对抗生成机制, 产生大量的数据样本, 以弥补机器学习中的训练数据不足。时下各种 GAN 应用如雨后春笋不断涌现, 难以一一列举, 这里介绍的仅是当前 GAN 在图像处理中应用中较为成熟的几个方面。

4.1 图像分类

GAN 网络必定包含一个判别器, 实际上就是一个天生的分类器, 再加上生成器的辅助, 实现的是对原始图像和生成的图像的联合分类, 而不是对图像单独分类, 使得 GAN 可以完成比一般统计分类更加复杂的分类任务。GAN 能够生成和原始图像分布相同的模拟样本, 在分类过程中, 通过添加这些模拟样本, 等价于人为增加训练样本集的数量, 有助于提高分类器的判别能力。

例如, 使用适当数量的“图像-标签”对, 加上简单的线性模型, 判别器的卷积层输出可用作特征提取器^[15-16]。在一个 DC-GAN 网络内, 可以用正则化 L2-SVM 分类器对训练后判别器中提取出的特征矢量来进行评价。使用这种方法, 在全监督和半监督数据集上已取得良好的分类效果。

当有标注训练数据提供有限时, 对抗训练还可以用于合成更多的训练数据。Shrivatava 等^[17]使用 GAN 来提炼合成图像, 同时维持它们的标注信息, 通过仅在 GAN 提炼的合成图像(即不是真实的图像)上进行训练, 在姿态和注视估计(pose-and gaze-estimation) 实验中取得了优异的性能。类似地, 使用时空 GAN 结构^[18], 注视估计和预测也获得了很好的结果。在某些实际应用情况下, 合成数据上训练的模型并不能很好地泛化。Busmalis 等^[19]提出解决这一问题的方法, 即用对抗训练的调适方法, 将来自源域的合成样本匹配到目标域。

4.2 图像合成

图像合成是 GAN 的一项核心功能, 并且当生成的图像受到一些限制时特别有用。LAP-GAN 模型^[20]在 Laplacian 金字塔框架中引入串接的卷积网络, 扩展了 GAN 模型的条件, 它的 G 和 D 网络都接收附加的标注信息作为输入, 以从粗到细的“金字塔”方式生成图像。这种条件 GAN 的想法

后来由 Reed 等^[21]扩展到从文字表述来合成图像。例如, 给出一个有关“鸟”的文字“一只只有粉色胸和冠的小鸟”, 训练好的 GAN 能够生成若干和此描述匹配的图像, 如图 6 所示。

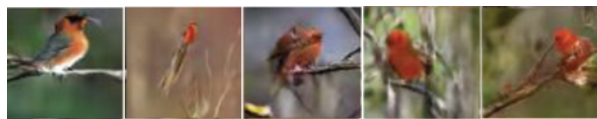


图 6 按文本合成的图像

条件 GAN 不仅允许我们用特殊的限制来合成新的样本, 它还允许我们开发新的工具去直观地编辑图像。例如, 改变图像中一个人的头发的式样, 编辑图像使他们显得更加年轻, 或者使他们佩戴眼镜。GAN 对图像合成的另外应用可参考 Zhu^[22]和 Brock 等^[23]的工作。

4.3 图像翻译

条件生成对抗网络(C-GAN) 非常适合将一幅输入图像“翻译”为一幅对应的不同形式的输出图像, 这是一种计算机图形、图像处理和计算机视觉领域常见的需求。文献[24-25]用一种多模式的混合 GAN 模型, 即 Bicycle-GAN 网络来实现“图像-图像的翻译”, 可以将语义标注图、灰度图或边缘图等作为输入, 那么 GAN 能够输出和输入图一致的真实的图像。图 7 给出了从航拍图到地图、从边缘图到实体图翻译的一例。



图 7 图像翻译实例

4.4 风格转换

图像的风格转移就是将一幅图像转换为另一种风格的图像, 如将一幅油画图像经过 GAN 处理, 自动转化为照片风格的图像。如果输入和输出的训练图像是同一场景, 只是表现风格不同, 这是一种“配对”训练。文献[26]提出利用循环 GAN(cycle-GAN) 进行无监督地从一种风格图像转化到另一种风格的图像, 是一种无配对(unpaired) 训练方法: 输入和输出的训练图像不仅风格不同, 而且内容也不同。这种方法比配对训练方法难度更大, 但训练数据的来源也大大扩展了。

没有配对的情况下 cycle-GAN 训练将一幅图像从源域 X 转换到目标域 Y 。训练的目标就是学习 $G: X \rightarrow Y$ 的映射, 利用对抗损失函数, 使得来自 $G(X)$ 的图像分布和来自 Y 的分布没有区别。设 Y

到 X 的映射为 $F: Y \rightarrow X$, cycle-GAN 引入了两个“循环一致”(cycle consistent)判断: 前向循环一致 $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$ 和后向循环一致 $y \rightarrow F(y) \rightarrow G(F(y)) \approx y$ 。训练的目标就是使前向和后向循环内容尽量保持一致,也就是使两个约等于号尽量趋近于等号。

4.5 图像修复

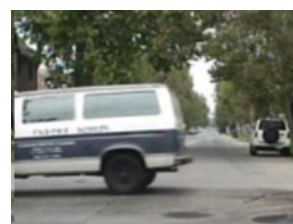
图像修复(image inpainting)是图像处理中的一个经典的应用领域,GAN 技术的引入提高了修复后图像的质量。在各种有损图像的修复中,人脸图像的修复是一种相对比较困难,但有广泛应用需求的技术。文献[27]提出了用于人脸补全(face completion)的一种生成对抗网络,主要由一个生成器、两个判别器和一个解析网络(parsing network)组成。生成器实际上是一个自编码器,输入为有损的整个脸部图像。一个判别器鉴别输入的整个真实图像和生成图像的真伪,另一个判别器鉴别输入的遮挡部分的生成图像和真实图像的真伪。生成图像的非遮挡部分用真实图像的对应像素替代。解析网络将脸部分割为11个语义部位(如嘴、鼻、眼、……),确定每个像素所属的部位。在人脸图像数据库 CelebA^[28]上的实验展示了这种方法对关键部位缺失的人脸图像的高质量补全的结果。

4.6 视频预测

由于自然场景的内容复杂和运动变化,对视频中将来帧(future frame)的预测是一项具有挑战性的任务,也是无监督视频表示学习的关键技术之一。现存的预测方法致力于对像素值的直接估计,和实际的将来帧有一定的差距,容易形成模糊的预测结果。为此,Liang 等^[29]提出了一种双重运动 GAN(dual motion GAN)结构。为了使合成的将来帧及将来流(future flow)与真实帧没有区别,这种 GAN 通过双重对抗训练机制,迫使预测的将来帧和视频帧中像素流(pixel-wise flows)保持一致。为了更好地预测,最初的将来帧预测和双重的将来流预测形成一个闭环,相互之间传递反馈信号。实验显示,双重运动 GAN 模型可广泛适应不同的可视场景,图8为其视频预测帧的一个实例,图8(b)的预测帧和图8(a)的真实帧几乎没有什么区别。

Yu 等^[30]提出另一种基于 GAN 的生成式视频预测网络。这个网络由运动编码器、帧生成器和帧判别器组成。运动编码器接收多帧帧差作为输入,输出一个全局视频运动表示。帧生成器产生将来帧。帧判别器是一个3D卷积网络,判断输入的帧

是否来自将来帧的分布。在两个著名图像数据库 UCF-1012 和 Ms. Pac-Man 上的实验表明,这种方法的主要预测指标高于当前同类方法。



(a) 真实帧



(b) 预测帧

图8 dual motion GAN 的预测帧

4.7 其他应用

GAN 的应用对传统的图像处理领域几乎是全覆盖的,而且还创造了不少新的应用。除了上面介绍的应用外,GAN 还在图像去噪^[31]、图像压缩^[32]、图像分割^[33]、图像编辑^[34]、图像增强^[35]、3D 图像表示^[36]等应用场合展现了良好的处理效果。以3D应用为例,作为2维图像的合成的扩展,Wu 等^[37]提出一种 GAN,它能够使用行矩阵卷积合成新的3D数据样本,例如椅子、桌子和小汽车等。另外,他们还提出了一种方法,可将目标2维图像映射为对应的3D物体版本。

5 GAN 研究的挑战和发展

GAN 技术无论是在理论、技术或应用领域取得长足进展的同时,尚存在不少开放问题亟待解决,仍面临严峻的挑战。

5.1 问题与挑战

(1) 模型训练的不稳定性

不稳定、甚至模式坍塌是 GAN 训练中的常见问题,使得训练难以收敛,生成器只产生一系列类似的样本或简单地产生单一的样本。在对 GAN 的种种改进中,最大的一部分就是针对这些训练问题提出的,如采用不同的分布距离的度量方式,对数据的批量规范化处理,设计更加有效的损失函数,判别器和生成器的协调训练等。

(2) 目标函数的优化

GAN 目标函数的优化包括两方面的内容:其一

是对数据的不同概率分布的距离或差距的度量方法,如 KL-散度、JS-散度、W-散度以及综合的 f-散度等。目前,很多基本 GAN 的改进方案都是针对这一关键测度进行的。其二是目标函数最优解的搜寻。GAN 的优化训练在于发现目标函数的一个最小的最大值,即“马鞍点”,而不是局部最小点。收敛到 GAN 的马鞍点需要较好的初始化, Lee 等^[38]证明,如果随机地选择优化器的初始点,梯度下降将不会以概率 1 收敛到马鞍点。

(3) 生成模型的性能评价

如何评价和比较由生成模型合成样本的性能尚未有定论。文献[39]声称,使用不同的方法评估 GAN 有可能导致合成样本质量的相矛盾的结论,决定选择一种方法胜过另一种方法还取决于应用。寻找有效的生成模型的性能评价方法尚在探究中。

(4) 生成对抗的可解释性

和很多人工神经网络雷同, GAN 也存在工作过程的“可解释性”问题^[40]。这是一个比较根本性的问题,不论 GAN 取得多么灿烂的成果,如果始终处于不可解释状态,就不能保证不出意外情况。希望随着我们对网络行为的解释能力的增强, GAN 技术的透明性会逐渐显现。附带提一下,对于 GAN 机制还存在些许质疑的声音^[41-42]。例如,认为判别器训练好以后,只要按照一定的公式,就可以由随机噪声来生成图像。

5.2 成因和性能评价

我们知道, GAN 的成功应该归功于两个方面,一是数据科学的流形分布规律:同类高维数据一般分布在低维流形附近,并具有特定概率分布。二是深度学习网络强大的逼近非线性映射的能力,能够以任意精度逼近连续函数。因此 GAN 技术的基本任务就在于从数据中学习流形结构,建立流形的参数表达和变换概率分布,将学习到的映射规律隐含在神经元的权重之中。

Lucic 等^[43]指出,在深度学习中, GAN 是生成器模型中的一个强有力的子类。尽管导致各种 GAN 算法的研究工作非常活跃,但是要定量地评价一种生成算法性能是很困难的。论文对模型评估方法进行了多方面研究后发现,大多数模型能够通过足够多的超参数优化和随机重启来获得类似的得分。他们还初步设计了几种特殊的数据集,用于估计精确度(precision) 和召回率(recall)。

6 结束语

从上述分析可知, GAN 研究和应用进展的驱动

力不仅在于它的深度学习方式、从隐空间到数据空间的高度非线性映射和反映射的潜能,而且还在于它可利用大量无标注图像数据的潜能^[44]。我们看到了在基本 GAN 结构基础上衍生出来的多种性能优良的 GAN 新品种,说明 GAN 之树正值枝叶繁茂、茁壮成长之际,同时也说明 GAN 技术无论从理论角度还是从应用角度都还缺少系统的理论基础和稳定的应用场景。在 GAN 的奥妙里,有许多机会去进行理论探索和技术开发,这里蕴育大量创新应用的机会。

参考文献:

- [1] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets [C] // Advances Neural Information Processing Systems Conference. 2014: 2672 - 2680.
- [2] MIRZA M, OSINDERO S. Conditional generative adversarial nets [EB/OL]. [2019-03-16]. <https://arxiv.org/pdf/1411.1784.pdf>.
- [3] CHEN X, DUAN Y, HOUTHOOFT R, et al. InfoGAN: interpretable representation learning by information maximizing generative adversarial nets [C] // Advances Neural Information Processing Systems Conference. 2016: 2172 - 2180.
- [4] RADFORD A, METZ L, CHINTALA S. Unsupervised representation learning with deep convolutional generative adversarial networks [C] // International Conference on Learning Representations Workshop Track. 2016.
- [5] ARJOVSKY M, BOTTOU L. Towards principled methods for training generative adversarial networks [C] // Neural Information Processing Systems Conference. 2016.
- [6] ARJOVSKY M, CHINTALA S, BOTTOU L. Wasserstein GAN [C] // 34th International Conference on Machine Learning. 2017: 214 - 223.
- [7] GULRAJANI I, AHMED F, ARJOVSKY M, et al. Improved training of Wasserstein GANs [C] // Proceedings of Advances Neural Information Processing Systems Conference. 2017.
- [8] NOWOZIN S, CSEKE B, TOMIOKA R. F-GAN: training generative neural samplers using variational divergence minimization [C] // Advances Neural Information Processing Systems Conference. 2016: 271 - 279.
- [9] BROCK A, DONAHUE J, SIMONYAN J. Large scale GAN training for high fidelity natural image synthesis [C] // International Conference on Learning Representations (ICLR). 2019.
- [10] MESCHEDER L, GEIGER A, NOWOZIN S. Which training methods for GANs do actually converge? [C] // International Conference on Machine Learning (ICML). 2018.
- [11] ZHAO J, MATHIEU M, LECUN Y. Energy-based generative adversarial network [C] // International Conference on

- Learning Representations (ICLR). 2017.
- [12] LEDIG C, THEIS L, HUSZÁR F, et al. Photo-realistic single image super-resolution using a generative adversarial network [C] // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017: 4681 – 4690.
- [13] WANG X, YU K, WU S, et al. ESRGAN: enhanced super-resolution generative adversarial networks [C] // European Conference on Computer Vision (ECCV Workshops). 2018.
- [14] JOLICOEUR-MARTINEAU A. The relativistic discriminator: a key element missing from standard GAN [EB/OL]. [2019-03-21]. <https://arxiv.org/abs/1807.00734>.
- [15] ZHU Lin, CHEN Yushi, GHAMISI P, et al. Generative adversarial networks for hyperspectral image classification [J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 56(9): 5046 – 5063.
- [16] SALIMANS T, GOODFELLOW I, ZAREMBA W, et al. Improved techniques for training GANs [C] // Advances Neural Information Processing Systems Conference. 2016: 2226 – 2234.
- [17] SHRIVASTAVA A, PFISTER T, TUZEL O, et al. Learning from simulated and unsupervised images through adversarial training [C] // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016: 2107 – 2116.
- [18] ZHANG M, MA K T, LIM J H, et al. Deep future gaze: gaze anticipation on egocentric videos using adversarial networks [C] // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017: 4372 – 4381.
- [19] BOUSMALIS K, SIBERMAN N, DOHAN D, et al. Unsupervised pixel-level domain adaptation with generative adversarial networks [C] // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016: 3722 – 3731.
- [20] DENTON E L, CHINTALA S, SZLAM A. Deep generative image models using a Laplacian pyramid of adversarial networks [C] // Advances Neural Information Processing Systems Conference. 2015: 1486 – 1494.
- [21] REED S, AKATA Z, YAN X, et al. Generative adversarial text to image synthesis [C] // 33rd International Conference on Machine Learning (ICML). 2016.
- [22] ZHU J Y, KRÄHENBÜHL P, SHECHTMAN E, et al. Generative visual manipulation on the natural image manifold [C] // European Conference on Computer Vision. 2016: 597 – 613.
- [23] BROCK A, LIM T, RITCHIE J M, et al. Neural photo editing with introspective adversarial networks [C] // 6th International Conference on Learning Representations (ICLR). 2017: 24 – 26.
- [24] LI Chengcheng, WANG Zi, QI Hairong. Fast-converging conditional generative adversarial networks for image synthesis [C] // IEEE International Conference on Image Processing (ICIP). 2018: 2132 – 2136.
- [25] ISOLA P, ZHU J Y, ZHOU T, et al. Image to image translation with conditional adversarial networks [C] // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017: 5967 – 5976.
- [26] ZHU J Y, PARK T, ISOLA P, et al. Unpaired image-to-image translation using cycle consistent adversarial networks [C] // IEEE International Conference on Computer Vision (ICCV). 2017: 2242 – 2251.
- [27] LI Yijun, LIU Sifei, YANG Jimei, et al. Generative face completion [C] // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017: 5892 – 5900.
- [28] LIU Z, LUO P, WANG X, et al. Deep learning face attributes in the wild [C] // IEEE International Conference on Computer Vision (ICCV). 2015: 3730 – 3738.
- [29] LIANG Xiaodan, LEE Lisa, DAI Wei, et al. Dual motion GAN for future-flow embedded video prediction [C] // IEEE International Conference on Computer Vision (ICCV). 2017: 1762 – 1770.
- [30] YU Tingzhao, WANG Lingfeng, GUA Huxiang, et al. Deep generative video prediction [J]. Pattern Recognition Letters, 2018, 110: 58 – 65.
- [31] CHEN Jingwen, CHEN Jiawei, CHAO Hongyang, et al. Image blind denoising with generative adversarial network based noise modeling [C] // IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2018: 3155 – 3164.
- [32] CHENG Zhengxue, SUN Heming, TAKEUCHI M, et al. Performance comparison of convolutional auto encoders, generative adversarial networks and super-resolution for image compression [C] // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2018: 2613 – 2616.
- [33] SOULY N, SPAMPINATO C, SHAH M. Semi supervised semantic segmentation using generative adversarial network [C] // IEEE International Conference on Computer Vision (ICCV). 2017: 5689 – 5697.
- [34] WU Xian, XU Kun, HALL P. A survey of image synthesis and editing with Generative adversarial networks [J]. Tsinghua Science and Technology, 2017, 22(6): 660 – 674.
- [35] PATRICIA L S, SAPP A D, VINTIMILLA B X, et al. Deep learning based single image dehazing [C] // IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). 2018: 1250 – 12507.
- [36] ZHU Junyan, ZHANG Zhoutong, ZHANG Chengkai, et al. Visual object networks: image generation with disentangled 3D representation [C] // 32nd Conference on Neural Information Processing Systems (NeurIPS). 2018.
- [37] WU J, ZHANG C, XUE T, et al. Learning a probabilistic latent space of object shapes via 3D generative-adversarial modeling [C] // Advances Neural Information Processing Systems Conference. 2016: 82 – 90.

- [38] LEE J D ,SIMCHOWITZ M ,JORDAN M I ,et al. Gradient descent only converges to minimizers [C] // Proceedings of Conference on Learning Theory. 2016: 1246 – 1257.
- [39] THEIS L ,OORD A ,BETHGE M. A note on the evaluation of generative models [C] // International Conference on Learning Representations (ICLR) . 2018.
- [40] ZHOU Bolei ,KHOSLA A ,LAPEDRIZA A ,et al. Learning deep features for discriminative localization [C] // IEEE Conference on Computer Vision and Pattern Recognition (CVPR) . 2016: 2921 – 2929.
- [41] LEI N ,LUO Z ,YAU S T ,et al. Geometric understanding of deep learning [EB/OL]. [2019-03-22]. <https://arxiv.org/pdf/1805.10451.pdf>
- [42] LEI Na ,SU Kehua ,CUI Li ,et al. A geometric view of optimal transportation and generative model [J]. Computer Aided Geometric Design 2019 68: 1 – 21.
- [43] LUCIC M ,KURACH K ,MICHALSKI M ,et al. Are GANs created equal a large-scale study [C] // 32nd Conference on Neural Information Processing Systems (NIPS) . 2018.
- [44] CRESWELL A ,WHITE T ,DUMOULIN V ,et al. Generative adversarial networks: an overview [J]. IEEE Signal Processing Magazine 2018 35(1) : 53 – 65.