

生成对抗网络研究综述^①

邹秀芳, 朱定局

(华南师范大学 计算机学院, 广州 510631)

通讯作者: 朱定局, E-mail: zhudingju@m.scnu.edu.cn



摘要: 自生成对抗网络 GAN 提出以后, 现这一方向已成为人工智能方向的研究热点. GAN 的思想采用二人零和博弈方法, 由生成器和判别器构成, 生成器负责生成样本分布, 判别器则判别输入是真实样本还是生成样本, 生成器和判别器不断交互优化, 最终达到最优效果. GAN 模型的提出无疑是很新颖的, 但也存在很多缺点, 比如梯度消失问题、模式崩溃等. 随着研究的深入, GAN 不断优化扩展, GAN 的衍生模型也层出不穷. GAN 可应用于不同领域, 主要为计算机图像和视觉领域, 在图像领域有着突出的效果, 能生成高分辨率逼真的图像, 能对图像进行修复、风格迁移等, 也能生成视频并进行预测等. GAN 也能生成文本, 可以进行对话生成、机器翻译、语音生成等. 同时, GAN 在其他领域也有涉及, 比如生成音乐、密码破译等. 但是 GAN 在其他领域的应用效果并不显著, 那么, 如何提高 GAN 在其他领域的应用效果将值得深入研究, 使生成对抗网络在人工智能方面大放异彩.

关键词: 生成对抗网络; 梯度消失; 模式崩溃; 图像领域; 风格迁移; 机器翻译

引用格式: 邹秀芳, 朱定局. 生成对抗网络研究综述. 计算机系统应用, 2019, 28(11): 1-9. <http://www.c-s-a.org.cn/1003-3254/7156.html>

Review on Generative Adversarial Network

ZOU Xiu-Fang, ZHU Ding-Ju

(School of Computer Science, South China Normal University, Guangzhou 510631, China)

Abstract: Since the proposal of generative adversarial networks, GAN has become a research hotspot of artificial intelligence. GAN adopts the method of the zero-sum game between two people, which consists of a generator and a discriminator. The generator is responsible for generating the sample distribution, and the discriminator is responsible for determining whether the input is a real sample or a generated sample. The generator and discriminator constantly interact and optimize to achieve the optimal effect. The model of GAN is undoubtedly very novel, but there are also many shortcomings, such as the problem of gradient disappearance, collapse mode, and so on. With the deepening of research, GAN has been continuously optimized and expanded, and the derivative models of GAN have emerged in endlessly. GAN has been optimized and improved. Also, GAN can be applied in different fields, though it is mainly used in the field of computer image and vision. It has outstanding effects in the field of image. It can generate high-resolution realistic images, repair images, transfer styles, and generate video and prediction. GAN can also generate text, to do some work, such as dialogue generation, machine translation, voice generation, and so on. GAN is also involved in other fields, such as generating music and decoding codes. However, the application effect of GAN in other fields is not significant. Therefore, how to improve its application effect is worthy of further study, which will make the generation of confrontation networks shine in artificial intelligence.

Key words: generative adversarial networks; gradient disappearance; collapse mode; the field of computer image; transfer styles; machine translation

① 基金项目: 国家社会科学基金重大项目 (14ZDB101); 国家自然科学基金 (61105133)

Foundation item: Major Program of National Social Science Foundation of China (14ZDB101); National Natural Science Foundation of China (61105133)

收稿时间: 2019-04-17; 修改时间: 2019-05-16; 采用时间: 2019-05-21; csa 在线出版时间: 2019-11-06

近年来,人们对深度学习研究的热情高涨,深度学习在许多领域取得了不错的成果,各种框架、模型相继被提出,并且不断的被完善、改进。其中,生成对抗网络(GAN)是Goodfellow等^[1]在2014年提出的一种生成模型。GAN的提出受到二人零和博弈(即参与博弈的两人利益之和为零,一方的利益是另一方的损失)的启发,这个模型包含一个生成器和一个判别器。生成器负责捕捉真实数据样本的分布并生成新的数据分布,而判别器则是一个二分器,判别输入的数据是真实的数据还是生成的数据。GAN的优化是“极小极大博弈”问题,使生成器生成的数据分布尽最大的可能接近真实数据分布,从而“迷惑”判别器。

近两年来,GAN已成为一个受欢迎的研究方向,越来越多关于GAN的论文发表(如图1),其中包含对GAN理论的完善,对GAN模型的改进及应用研究。著名学者LeCun Y对GAN模型评价颇高,称其为“有史以来最酷的事情”、“在过去十年机器学习研究中最有趣的点子”。目前,生成对抗性网络主要应用于计算机图像与视觉领域,可以生成逼真的图像,可以生成人脸,可以对目标进行检测,也可以生成真实的场景并应用于无人驾驶场景;还可以根据图像的上下文进行图像修复^[2],对图像进行转换^[3]。此外,GAN还可以应用于文本生成、语音和语言的生成、视频预测等。

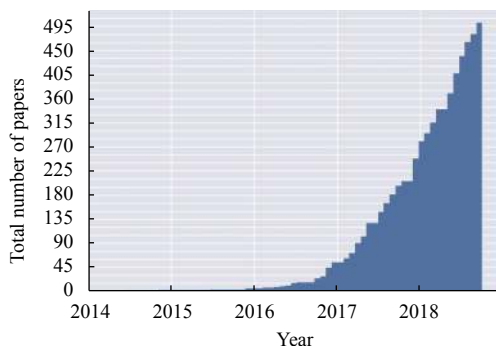


图1 命名为GAN的论文的数量^[4]

1 生成对抗性网络

1.1 GAN的基本思想

GAN的提出受到“二人零和博弈”的启发,由一个生成器和一个判别器构成。生成器根据输入的样本数据分布,生成新的数据分布,新的数据分布需要尽量接近真实数据分布,越接近表示生成的数据就越真实。判别器的作用则是判别输入数据是来自真实数据还是生

成器生成的数据。GAN的模型结构如图2所示。生成器和判别器可以用神经网络或者函数表示,用 G 表示生成器, D 表示判别器。生成器 G 的目的是生成基于数据分布的 p_g , G 的输入为随机噪声 z (例如高斯噪声), z 取样于先验分布 $p_z(z)$ 。判别器对输入的数据进行判别,如果是真实数据则输出1,否则的话输出0。 $D(x)$ 表示输入是真实数据的概率。对于初代生成器 G 和判别器 D 来说,初代 D 很容易能判别出是真实数据还是生成的数据,接下来,初代 G 进行优化产生第二代 G ,初代 D 不能判别第二代 G 生成的数据是假的,那么初代 D 也跟着优化,产生第二代 D ,能判别第二代 G 生成的数据是否是真实数据,接着产生第三代、第四代,直到第 N 代 D 无法正确判别数据是来自真实数据还是生成器 G 生成的。

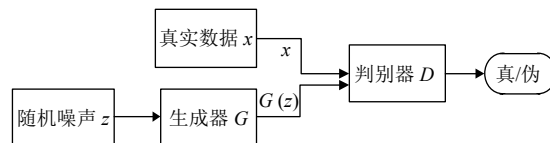


图2 GAN的模型结构

1.2 GAN算法

生成器 G 和判别器 D 之间是互相对抗的,其优化过程是极小极大化过程。GAN的目标函数如式(1)所示。

$$\min_G \max_D F(D, G) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log(1 - D(G(z)))] \quad (1)$$

GAN的优化过程是,先固定生成器 G ,优化判别器 D ,求得最优判别器 D^* ;然后固定 D 为最优 D^* ,优化生成器 G ,求得最优 G 。

当 G 固定的时候,优化判别器 D 就是求得使 $F(D, G)$ 取得最大值的 D^* 。 $F(D, G)$ 可以写为如下形式:

$$\begin{aligned} F(G, D) &= \int_x p_{data}(x) \log(D(x)) dx \\ &\quad + \int_z p_z(z) \log(1 - D(g(z))) dz \\ &= \int_x [p_{data}(x) \log(D(x)) \\ &\quad + p_g(x) \log(1 - D(x))] dx \end{aligned} \quad (2)$$

对于任意非零实数 a, b ,且实数 $y \in [0, 1]$,那么式(3):

$$a \log(y) + b \log(1 - y) \quad (3)$$

在 $\frac{a}{a+b}$ 处取得最大值,故最优判别器 D^* :

$$D_G^*(x) = \frac{p_{data}(x)}{p_{data}(x) + p_g(x)} \tag{4}$$

把最优判别器 D^* 代入式 (1), 令 $H(G) = V(D_G^*, G)$, 可以得到:

$$H(G) = -\log 4 + 2 \cdot JSD(p_{data}||p_g) \tag{5}$$

其中, JSD 是香农散度, 它是 KL 的对称平滑版本, 表示两个分布之间的差异. 当 $p_g = p_{data}$ 时, $H(G)$ 取得最小值 $H^*=-\log(4)$, 此时 G 为最优. GAN 的训练采用的是迭代的数值方法, 为了防止过拟合, 通常是对 D 进行 k 次优化后, 再对 G 进行一次优化.

1.3 GAN 的优缺点

GAN 模型包含生成器 G 和判别器 D , 其模型有许多优点, 但也存在一些问题. 首先, GAN 没有使用复杂的马尔科夫链, 而是用反向传播算法来获得梯度; 其次, GAN 的学习过程不需要过多的推理, 它的模型可以是神经网络, 也可以是多种函数; 再者, GAN 可以表示清晰, 甚至退化的分布, 并且能生成高清的图像.

GAN 基于连续数据, 且假设有无限建模能力. GAN

模型在生成和优化过程中, 存在一些缺点: (1) GAN 在训练过程中, 优化 k 次判别器 D , 同时优化 1 次生成器 G , 需要保持 D 和 G 的同步, 容易造成 GAN 训练的不稳定; (2) GAN 存在模式崩溃 (collapse mode)^[5]现象, 即生成器可能会生成许多它认为是多样的, 实际上却只有细微差别的样本, 这样会造成多样性缺乏; (3) GAN 存在梯度消失问题, 即当真实样本和生成样本之间没有重叠或者重叠可以忽略时, 其优化网络的目标函数的 Jensen-Shannon 散度是一个常数. 这些问题在之后提出的衍生模型中得到了解决.

2 GAN 的优化和衍生模型

随着人工智能的发展, 研究者们对 GAN 的研究热情高涨, 有对 GAN 模型框架、理论的改进, 有对 GAN 模型进行改造, 或结合其他学习方法, 使其能应用于更多场景. GAN 各种不同的衍生模型比较与区别见表 1. 接下来将介绍对 GAN 进行优化和衍生的各种不同模型.

表 1 GAN 各种不同的衍生模型比较与区别

模型	机制	优点	缺点	适用场景
GAN	生成器和判别器	反向传播算法来获得梯度; 能生成高清的图像	训练不稳定; 模式崩溃; 梯度消失	生成图像
BGAN	生成器含有策略梯度, 判别器含有估计差分度量	适用于离散数据	生成效果不是很好	生成文本
DCGAN	生成器和判别器使用卷积神经网络, 并进行细化改进	对 GAN 的体系结构进一步的改进, 提高了 GAN 训练的稳定性	只对结构进行改进, 没有对优化方法进行改进	生成图像
WGAN	生成器和判别器	解决了 GAN 训练不稳定问题和模式崩溃问题	参数设置不合理容易导致梯度消失或者梯度爆炸	无监督学习
LS-GAN	生成器和判别器	限制了 GAN 的无限建模能力	实现时需要考虑很多细节, 容易影响实现效果	无监督、半监督、监督学习
InfoGAN	生成器中增加隐码 c , 判别器	对 GAN 模型信息理论的扩展	需要计算两个数据之间的互信息, 增加计算的负担	生成图像
SGAN	生成器增加类标签, 判别器	加入类标签, 可以对输出结果进行分类	需要对数据进行预处理	半监督学习
CGAN	生成器和判别器, 并在两者的输入层增加额外信息 y	增加了额外信息, 额外信息可以为类标签、辅助信息或者其他数据	额外信息种类较多, 输入层需要处理不同的额外信息	无监督、半监督学习
BiGAN	生成器和判别器, 并增加一个编码器	实现逆映射, 将数据 x 映射到隐变量空间	增加了一个编码器, 使得优化函数计算更复杂	对隐变量空间和数据 x 进行双向映射

2.1 基于离散数据

GAN 模型是基于 Lipschitz 连续空间, 且依赖于生成样本是可微的, 一般对离散数据不起作用. Devon 等^[6]提出一种引入离散数据的 GAN 训练方法——边界寻

找生成对抗网络, 称为 BGAN. BGAN 使用来自鉴别器的估计差分度量来计算生成的样本的重要性权重, 为训练生成器提供一个基于 KL-散度的策略梯度, 且这个策略梯度引入奖励机制, 使用重要权重作为奖励信

号. GAN 应用于离散数据的另一种方法是 Tong 等^[7]提出的最大似然增强的离散生成对抗网络, 其对 GAN 的目标没有直接进行优化, 而是使用遵循对数似然的对应的输出推导出了—种全新的、低方差的目标, 主要是为了解决在离散数据上的反向传播困难的问题. 虽然 GAN 能应用于离散数据空间, 但这几种方法实现的效果不是很好, 不如基于连续空间的效果显著.

2.2 深度卷积生成对抗网络

Radford等^[8]提出了深度卷积生成对抗网络 (DCGAN), 将卷积神经网络 (CNN)^[9]应用到生成对抗网络中, 通过对 GAN 的体系结构更改, 提高了 GAN 的训练的稳定性. 在 DCGAN 中, 对 GAN 的体系结构进行了一些修改: 将空间池化层函数替换为跨卷积; 去除了完全连接层, 能提高模型的稳定性; 除了生成器的输出层和判别器的输入层之外, 对每个单元的输入进行批归—化操作; 在生成器中使用 ReLU 激活函数, 在其输出层使用 Tanh 函数; 在判别器中使用 LeakyReLU 激活函数. DCGAN 具有更强大的生成能力, 训练也更稳定, 生成的样本具有更多的多样性, 因此, 很多对于 GAN 的改进都是基于 DCGAN 的结构. DCGAN 只是对 GAN 模型的结构进行了改进, 对生成器和判别器进一步的细化, 并没有对优化方法进行改进.

2.3 GLS-GAN

GAN 在采用梯度下降训练过程中存在梯度消失问题, 为了解决这个问题, Arjovsky 等^[10]提出 Wasserstein GAN (WGAN). WGAN 使用 Earth-Mover 距离代替 Jensen-Shannon 散度来计算生成样本分布与真实样本分布之间的距离. WGAN 在生成器和判别器的训练中不需要保持平衡, 解决了 GAN 训练不稳定的问题, 同时也解决了模式崩溃现象, 保证了生成样本的多样性, 并且不需要精细设计网络架构就能实现上述功能. WGAN 使用了权重剪枝 (weight clipping) 对批评函数施加 Lipschitz 约束, 但是存在两个问题: (1) weight clipping 独立的限制每一个参数的取值范围, 而判别器的损失函数希望尽量拉大真实样本与生成样本之间的差距, 会导致判别器学习更简单的映射函数; (2) 如果 clipping threshold 设置不合理, weight clipping 会容易导致梯度消失或梯度爆炸. Gulrajani 等^[11]提出了一种代替 weight clipping 的方法—引入梯度惩罚 (gradient penalty), 即对批评函数的输入准则进行处罚, 称为 WGAN-GP. WGAN-GP 只对采样点 x 起作用, 而对真实

数据附近的区域没有检查, 判别器在这个区域可以自由的违反 Lipschitz 连续. 为了增强 WGAN 中批评家的 Lipschitz 约束, Cui 等^[12]提出了结合指数梯度惩罚项和重要性加权策略的算法, 该算法在不增加计算负担的情况下有更快的收敛速度, 并且保持了 WGAN-GP 训练的稳定性.

GAN 的判别器具有无限建模能力, 即不管真实样本和生成样本多复杂, 判别器 D 都能判别, 这样容易造成过拟合问题. 为了限制 GAN 的无限建模能力, Qi^[13]提出损失感知生成对抗网络 (LS-GAN), LS-GAN 引入一个基于 Lipschitz 连续的代价 (cost) 函数, 把最小化目标函数得到的损失函数限定在这个代价函数上. 广义 LS-GAN 称为 GLS-GAN^[13] (如图 3 所示), 是 LS-GAN 的广义形式, 通过定义一个满足一定条件的代价函数, 不同的代价函数可以获得不同的 GLS-GAN. LS-GAN 和 WGAN 是 GLS-GAN 的特例, LS-GAN 是代价函数为 C_0 的 GLS-GAN, 而 WGAN 则为代价函数是 C_1 的 GLS-GAN.

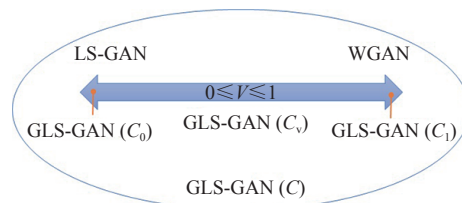


图3 广义 GLS-GAN^[14]

LS-GAN 和 WGAN 的实现都采用了 DCGAN 的网络架构思想, 且对 GAN 的优化方法进行了改进. WGAN 虽然解决了 GAN 训练不稳定问题和模式崩溃问题, 需要掌握对 clipping threshold 的设置, 如果设置不合理也会造成梯度消失或者梯度爆炸问题. WGAN 使用 EM 距离代替 JS 散度, 但是 EM 距离不容易直接优化, 使用的是其共轭函数作为目标函数进行优化, 增加了计算的负担. 广义 LS-GAN 可以通过设置不同的代价函数获得不同的 GLS-GAN, 可以扩宽 GLS-GAN, 不过这个还有待去开发和试验. LS-GAN 在实现时需要考虑很多细节, 容易影响实验效果.

2.4 InfoGAN

信息生成对抗网络 (InfoGAN)^[15]是 GAN 信息理论的一个重要扩展. InfoGAN 相比一般的 GAN, 引入一个隐码 c , c 表示显著结构化隐层随机变量与特定语

义特征之间的关系. 生成器的输入为噪声 z 和隐码 c , 输出为 $G(z, c)$, 在 GAN 中, $P_G(x|c) = P_G(x)$. InfoGAN 使用互信息 $I(c; G(z, c))$ 表示两个数据之间的关联性, 而隐码 c 和生成分布 $G(z, c)$ 之间有高的互信息. InfoGAN 的目标函数如式 (6) 所示:

$$\min_G \max_D F_1(D, G) = F(D, G) - \lambda I(c; G(z, c)) \quad (6)$$

2.5 其他衍生模型

GAN 的训练是无监督的, 也可以结合半监督进行训练. SGAN^[16]是使用半监督训练的生成对抗网络, 它在生成器 G 的输入中加入类标签, 判别器 G 输出相应的类标签. CatGAN^[17]使用无监督和半监督来训练网络, 提出了一种从未标记或部分标记数据中学习判别分类器的方法—使用目标函数来权衡观察到的例子和它们的预测分类分布之间的互信息, 判别器 D 输出类的分布.

条件生成对抗网络 (CGAN)^[18]引入一个额外信息 y , y 可以是任何辅助信息, 类标签或者其他模式的数据, 将 y 加入到生成器 G 和判别器 D 的输入层. 一般的 GAN 生成模型都是把隐层变量分布映射到真实数据分布上, Donahue 等^[19]提出双向生成对抗网络 (BiGAN) 实现逆映射—将数据 x 映射到隐变量空间. BiGAN 增加了一个编码器 E , 输出为 $E(x)$, 其优化目标为 $\min_{G,E} \max_D F(D, E, G)$.

3 GAN 的应用领域

3.1 计算机视觉领域

3.1.1 图像生成

GAN 的最初目的是输入向量生成图像, 开始生成的图像像素都不是很高, 后来随着 GAN 的不断深入研究, 生成图像的质量也在不断提高. Ledig 等^[20]提出一种增加图像分辨率的 GAN—SRGAN, 可以将低分辨率的图像转换为超分辨率图像. SRGAN 的生成器采用了残差网络 (ResNet)^[21], 判别器使用了 VGG^[22]. Karras 等^[23]提出了 Progressive GAN—一种以渐进方式训练的 GAN, 在训练过程中增加一个能处理高分辨率的网络层, 能生成以假乱真的图像. 另外, ProGANsR^[24]也能生成高分辨率的图像, 它采用多尺度渐进式原则, 同时增加上采样 (upsampling) 的重建质量. GAN 可以生成人脸, 根据真实样本可以生成不存在的人脸; 也可以生成动物等其他物体. 除了生成人类、动物等物体,

GAN 还可以生成真实的场景, 可应用于无人驾驶等场景中. 在这些应用中, 对于图像分辨率的要求也将会随着技术的发展越来越高.

3.1.2 图像转换

图像转换也可以叫做图像翻译, 是将图像变成另一种形式的图像, 比如把图像内容移除一个域的属性, 转移到另一个域上. 图像转换的输入是图像, 输出的是转换后的图像. Pix2pix^[25]使用条件生成对抗网络 (CGAN) 来进行图像转换, 使用成对的数据来进行训练. 因为 pix2pix 需要使用成对的数据来进行训练, 但有些情况下成对的数据不容易获得, 因此, Zhu^[3]提出 CycleGAN, 不需要使用成对的数据, 就能把图像从源域 X 转换为目标域 Y . CycleGAN 使用两个判别器 D_x 和 D_y , 分别判别 x 和转换图像 $F(y)$, y 和 $G(x)$, 还引入循环一致性损失来执行 $F(G(x)) \approx x$. 对于上述两种 GAN, 都存在生成图像分辨率低的问题, Wang 等^[26]提出 pix2pixHD, 可以生成分辨率高达 2048×1024 的图像, 同时支持用户交互, 编辑图像中目标外观, 可以合成一些道路场景, 可以结合目标检测及行人识别, 用于无人驾驶中.

3.1.3 风格迁移

风格迁移是指将学习到目标图像的风格应用于源图像, 使得源图像保留内容的同时具有目标图像的风格, 比如把一张图像变成油画风格的图像. Li 等^[27]提出一种基于马尔科夫链的生成对抗网络 (MGAN) 来实现图像的风格迁移, 效果如图 4 所示. Azadi 等^[28]使用 GAN 来对字体进行风格迁移学习, 能产生多种不同风格的字体. 除此之外, 风格迁移还可应用于音乐等其他领域, 只是效果都不如图像的风格迁移好.

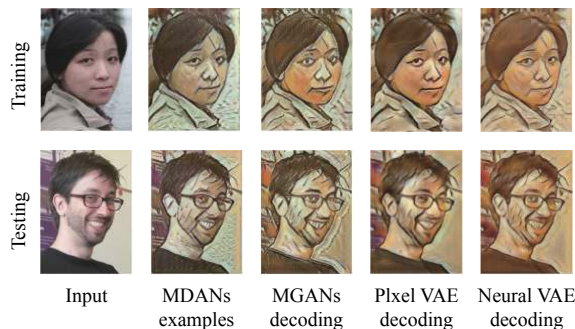


图4 图像风格迁移效果^[27]

3.1.4 图像修复

GAN 与图像修复能很好的结合, 对缺失的图像进行完整的修复. Denten 等^[2]使用半监督方法, 在生成器

中增加一个随机的、能隐藏的补丁,根据上下文信息对图像进行修复. Liu 等^[29]使用部分卷积进行图像修复,修复效果如图5所示.



图5 图像修复效果^[29]

3.1.5 视频生成和预测

在生成图像的基础上, GAN 还可以生成视频. Tulyakov 等^[30]提出一种通过分解动作和内容来生成视频的方法—MoCoGAN, 将一个随机向量的序列映射到一个视频帧序列上, 每个随机向量包含一个动作和一个内容. 视频预测是根据当前的一帧或几帧视频来预测视频接下来的内容, 比如当一个人拿起一个球的时候预测这个人接下来要干什么. Xiong 等^[31]提出一种多阶段动态生成对抗网络来生成延时视频, 首先为每一帧生成现实内容的视频, 然后进行运动建模, 使相邻帧之间的物体运动更加生动, 同时保持内容逼真. 另一种视频预测方法是 Jang 等^[32]提出的方法—通过外观和动作来指定未来, 减少不确定性, 解决了当有多个正确的、等可能的未来时, 模型不知如何选择的问题.

3.2 文本生成领域

除了生成图片之外, GAN 还能生成文本. Zhang 等^[33]用 GAN 来生成文本, 其判别器为卷积神经网络 (CNN). GAN 一般是基于连续空间, 但也可以应用于离散数据. 但是 GAN 用于生成序列存在两个问题, GAN 生成的是连续数据, 难的是直接生成离散序列; 另一个问题是, GAN 只能对整个生成序列打分, 而对于一部分生成的序列, 很难判断它现在生成序列的质量和之后生成整个序列的质量. SeqGAN^[34]能解决上述两个问题, 它结合强化学习, 能生成离散的序列, 使用策略梯度训练生成器 G 来解决输出离散数据时梯度不能返回生成模型的问题, 通过蒙特卡洛搜索得到奖励信号. WGAN 也能推广应用于离散数据空间, 在基于 WGAN 的基础上, 代威^[35]提出一种使用字符级文本构造方法的无监督文本生成方法 Lable GAN, 能适用于多种序列处理

任务, 并且能有效的干预和约束生成数据空间. SeqGAN 可以应用于语音生成、诗词生成、机器翻译、对话生成等. MaskGAN^[36]根据上下文内容来填充缺失的文本, 并引入 actor-critic 架构.

此外, GAN 还可以根据文字描述来生成图文匹配的图像. StackGAN^[37]能够根据输入文本描述生成细化的高分辨率图像. AttnGAN^[38]可以生成更细腻的图像, 通过关注自然语言描述中的相关词汇, 可以在图像的不同子区域合成出精细的细节, 此外, 还提出了一种深度注意多模态相似度模型, 用于计算用于训练生成器的细粒度图像-文本匹配损失.

3.3 其他领域

GAN 除了在视觉领域、文本生成领域外, 还在其他领域有涉及. GAN 能应用于密码破译, CipherGAN^[39]提出了一个能够用于破译密码的框架, 能够高度准确地破译使用移位和 Vigenere 密码的语言数据, 且能支持更多的词汇. GAN 还可以进行行人重识别、重定向等. GAN 除了能生成图像外, 还能生成音乐、3D 物体等.

4 存在的问题及发展趋势

4.1 存在的问题

GAN 模型是一种框架, 可以与其他方法相结合. GAN 模型包含生成器 G 和判别器 D, 在训练过程中需要保持生成器 G 和判别器 D 之间的同步, 因此容易造成训练不稳定. 虽然 WGAN 使用 EM 距离代替 JS 散度, 不需要保持生成器 G 和判别器 D 之间的同步, 解决了训练不稳定的问题, 但是, 在使用 JS 散度的模型的训练稳定性问题还是没有得到很好的解决. 此外, GAN 还存在模式崩溃问题. WGAN 能解决模式崩溃问题, 但是, 如果损失函数的参数设置不合理的话会导致梯度消失或梯度爆炸. 因此, 模式崩溃问题还没有彻底的解决.

同时, GAN 也面临着一个如何评估生成样本质量的问题, 也就是说 GAN 缺乏客观的评估方法. 现有对 GAN 的评价方法都是基于样本层面, 对生成样本与真实样本提取特征, 在特征空间做距离度量. Xu 等^[40]实验结果表明, 在基于样本的评价方法中, 核函数最大平均偏差 (Kernel MMD) 和 1-近邻 (1-NN) 双样本检验方法是最合适的评价方法, 这两种评价指标能够有效的区别生成样本和真实样本, 同时也能识别模式崩溃的

模式下下降问题,检测过度拟合,且具有高效性。但是,对于生成效果和质量,却只能依赖于视觉上的观察,由人眼去判断生成的质量好坏。有些评价指标在数值上虽高,但是实际的生成效果却不好。因此,GAN 很难找到一个比较客观的、可量化的评价指标。

4.2 发展趋势

自从 GAN 被提出以来就受到广泛的关注,近两年来对 GAN 的研究热度也是只增不减。虽然 GAN 能实现的功能巨大,但是这个模型存在的一些问题未能彻底解决,影响其生成效果。在模型的训练稳定性和评价指标方面,仍有待研究改进。GAN 在图像领域有着卓越的应用效果,在其他领域也有着不错的应用。GAN 对于文本方面的应用是基于离散数据,效果不是很理想,在这一方面的改进空间将会很大。随着人工智能的不断发展,对无人驾驶的研究也在不断成熟,GAN 可以结合其对图像领域的生成效果,应用于无人驾驶,这也将会是很热门的一个研究方向。视频与语音在日常生活中是必需的,随着人们需求的增大,GAN 在视频与语音方面的应用也更需要去研究改进。GAN 也可以应用于图像去噪方面,现在也有一些突破性的进展。

4.2.1 图像去噪

人们对图像的反映是通过视觉直观的体现出来的,图像质量差会影响人们对图像细节的掌握,影响图像的视觉效果。噪声是图像干扰的重要原因,实际上一幅图可能会存在各种各样的噪声,所以图像噪声在视觉图像中影响很大。图像噪声一般产生于图像的采集或者传输过程中,比如使用的设备不一样,在传输过程中有压缩等等。

传统的图像去噪最初采用的是使用像素尺度的空间域滤波方法^[41],该方法中的中值滤波和均值滤波都是使用图像信号和滤波模板结合来进行滤波。空间域滤波方法在使用简单的卷积核的时候容易导致图像信息丢失,针对这一缺点,一些学者提出基于变换域的去噪方法,包含傅里叶变换域,小波域,脊波域等等^[42,43]。随着深度学习的发展,深度卷积网络在图像去噪的应用有着不错的效果。去噪卷积网络(DnCNN)^[44]第一次结合残差网络应用于图像去噪领域,使得去噪效果大大提升。Wang^[45]采用空洞卷积代替卷积核,提高了训练运行速度,效果不亚于去噪卷积网络。

4.2.2 结合 GAN 的图像去噪

现有的去噪方法大多是基于已知噪声,例如高斯

噪声,GAN 可以学习复杂的分布,可以在 GAN 的生成器中输入含有已知噪声的图片,然后判别器输出对应无噪声的图像。张元祺^[46]结合生成对抗网络,对彩色图像进行去噪,通过神经网络将含噪图像变换到特征域进行处理,并利用多尺度特征保留图像丰富的纹理细节。

那么对于图像里包含的未知噪声信息,即对图像进行盲去噪,GAN 拥有强大的学习能力,可以学习复杂的分布,可以使用 GAN 来构建无噪-有噪图像的训练数据集。在有未知噪声图像的情况下,一种建立成对训练数据集的方法是训练 GAN 的生成网络,学习从无噪图像到相似噪声图像的映射。生成网络可以训练学习真实噪声图像的分布,生成具有相似噪声的图像,但不能保证原始图像的纹理细节等不改变。因此,Chen 等^[47]提出一种盲去噪方法——使用 GAN 来对未知噪声图像进行噪声提取并建模,生成网络生成类似的是噪声而不是噪声图像,生成相似分布的噪声,然后使用深度卷积网络进行图像去噪。

对于噪声已知的图像,可以直接作为生成器的输入,判别器进行去噪,输出对应的无噪图像,在这个过程中,需要保留图像原始的内容,即不能损失图像的纹理细节等。如何在去噪的过程中保留图像内容也值得去研究。对于图像盲去噪,可以通过 GAN 来建模获得相似噪声分布,然后再使用深度卷积网络或者其他网络进行去噪。在这个过程中,需要使用两个或两个以上模型,如何使用 GAN 模型进行图像的盲去噪将会是一个值得研究的方向。

5 结论

生成对抗性网络 GAN 包含生成器和判别器,有很强的生成能力,在不进行限制的情况下,具有无限建模的能力,主要应用于图像视觉领域。现今 GAN 能够产生超分辨率图像,能够结合半监督、强化学习、特征学习等,在图像、视觉、文本方面有广泛应用。GAN 在图像领域的应用效果显著,虽然在其他领域也应用广泛,但实现的效果并没有图像的好。GAN 如何更好的应用于更多领域,在应用的同时不断优化性能,并发挥更大的优势,将会是有趣的研究。随着人工智能的发展,人们追求的是更加智能化,如何结合深度学习等不同技术来提高 GAN 的应用效果,使 GAN 更智能化,是 GAN 未来发展需要提升的。GAN 可以和图像去噪相结合,利用 GAN 的强学习能力,实现噪声已知的图

像去噪,也可以建模生成相似的噪声分布,有助于图像盲去噪. GAN 对于图像去噪的应用将有待于更进一步的研究.

参考文献

- 1 Goodfellow IJ, Pouget-Abadie J, Mirza M, *et al.* Generative adversarial nets. Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal, QC, Canada. 2014. 2672–2680.
- 2 Denton E, Gross S, Fergus R. Semi-supervised learning with context-conditional generative adversarial networks. arXiv preprint arXiv: 1611.06430, 2014.
- 3 Zhu JY, Park T, Isola P, *et al.* Unpaired image-to-image translation using cycle-consistent adversarial networks. Proceedings of 2017 International Conference on Computer Vision. Venice, Italy. 2017. 2223–2232.
- 4 Hindupur A. The-gan-zoo. <https://github.com/hindupuravinash/the-gan-zoo>.
- 5 Goodfellow I. NIPS 2016 tutorial: Generative adversarial networks. arXiv preprint arXiv: 1701.00160, 2016.
- 6 Hjelm RD, Jacob AP, Che T, *et al.* Boundary-seeking generative adversarial networks. arXiv preprint arXiv: 1702.08431v4, 2018.
- 7 Che T, Li YR, Zhang RX, *et al.* Maximum-likelihood augmented discrete generative adversarial networks. arXiv preprint arXiv: 1702.07983, 2017.
- 8 Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv: 1511.06434, 2016.
- 9 Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe, NV, USA. 2012. 1097–1105.
- 10 Arjovsky M, Chintala S, Bottou L. Wasserstein GAN. arXiv preprint arXiv: 1701.07875, 2017.
- 11 Gulrajani I, Ahmed F, Arjovsky M, *et al.* Improved training of Wasserstein GANs. Proceedings of the 30th Neural Information Processing Systems. Long Beach, CA, USA. 2017. 5767–5777.
- 12 Cui SB, Jiang Y. Effective Lipschitz constraint enforcement for Wasserstein GAN training. Proceedings of the 2nd IEEE International Conference on Computational Intelligence and Applications. Beijing, China. 2017. 74–78.
- 13 Qi GJ. Loss-sensitive generative adversarial networks on Lipschitz densities. arXiv preprint arXiv: 1701.06264, 2017.
- 14 Qi GJ. Generalized LS-GAN (GLS-GAN): Now LS-GAN and WGAN are special cases of this supermodel. <https://zhuanlan.zhihu.com/p/25580027>.
- 15 Chen X, Duan Y, Houthoofd R, *et al.* InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets. Proceedings of 2016 Neural Information Processing Systems. Barcelona, Spain. 2016. 2172–2180.
- 16 Odena A. Semi-supervised learning with generative adversarial networks. arXiv preprint arXiv: 1606.01583, 2016.
- 17 Springenberg JT. Unsupervised and semi-supervised learning with categorical generative adversarial networks. arXiv preprint arXiv: 1511.06390, 2015.
- 18 Mirza M, Osindero S. Conditional generative adversarial nets. arXiv preprint arXiv: 1411.1784, 2014.
- 19 Donahue J, Krähenbühl P, Darrell T. Adversarial feature learning. arXiv preprint arXiv: 1605.09782, 2016.
- 20 Ledig C, Theis L, Huszár F, *et al.* Photo-realistic single image super-resolution using a generative adversarial network. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA. 2017. 105–114.
- 21 He KM, Zhang XY, Ren SQ, *et al.* Deep residual learning for image recognition. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. 2016. 770–778.
- 22 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv: 1409.1556, 2014.
- 23 Karras T, Aila T, Laine S, *et al.* Progressive growing of GANs for improved quality, stability, and variation. arXiv preprint arXiv: 1710.10196v3, 2018.
- 24 Wang YF, Perazzi F, McWilliams B, *et al.* A fully progressive approach to single-image super-resolution. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Salt Lake City, UT, USA. 2018. 977–986.
- 25 Isola P, Zhu JY, Zhou TH, *et al.* Image-to-image translation with conditional adversarial networks. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA. 2017. 5967–5976.
- 26 Wang TC, Ming YL, Zhu JY, *et al.* High-resolution image synthesis and semantic manipulation with conditional GANs. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA.

2018. 8798–8807.
- 27 Li C, Wand M. Precomputed real-time texture synthesis with Markovian generative adversarial networks. Proceedings of the 14th European Conference on Computer Vision. Amsterdam, The Netherlands. 2016. 702–716.
- 28 Azadi S, Fisher M, Kim V, *et al.* Multi-content GAN for few-shot font style transfer. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. 2018. 7564–7573.
- 29 Liu GL, Reda FA, Shih KJ, *et al.* Image inpainting for irregular holes using partial convolutions. Proceedings of the 15th European Conference on Computer Vision. Munich, Germany. 2018. 89–105.
- 30 Tulyakov S, Liu MY, Yang XD, *et al.* MoCoGAN: Decomposing motion and content for video generation. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. 2018. 1526–1535.
- 31 Xiong W, Luo WH, Ma L, *et al.* Learning to generate time-lapse videos using multi-stage dynamic generative adversarial networks. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. 2018. 2364–2373.
- 32 Jang Y, Kim G, Song YL. Video prediction with appearance and motion conditions. Proceedings of the 35th International Conference on Machine Learning. Stockholm, Sweden. 2018. 3496–3510.
- 33 Zhang YZ, Gan Z, Carin L. Generating text via adversarial training. Proceedings of the 2016 Conference on Neural Information Processing Systems. Barcelona, Spain. 2016. 1–6.
- 34 Yu LT, Zhang WN, Wang J, *et al.* SeqGAN: Sequence generative adversarial nets with policy gradient. Proceedings of the 31st AAAI Conference on Artificial Intelligence. Phoenix, AZ, USA. 2016. 2852–2858.
- 35 代威. 基于生成式对抗网络的文本生成问题的研究[硕士学位论文]. 长春: 吉林大学, 2018.
- 36 Fedus W, Goodfellow I, Dai AM. MaskGAN: Better text generation via filling in the _____. arXiv preprint arXiv: 1801.07736, 2018.
- 37 Zhang H, Xu T, Li HS, *et al.* StackGAN: Text to photo-realistic image synthesis with stacked generative adversarial networks. Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy. 2017. 5908–5916.
- 38 Xu T, Zhang PC, Huang QY, *et al.* AttnGAN: Fine-grained text to image generation with attentional generative adversarial networks. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. 2018. 1316–1324.
- 39 Gomez AN, Huang SC, Zhang I, *et al.* Unsupervised cipher cracking using discrete GANs. arXiv preprint arXiv: 1801.04883, 2018.
- 40 Xu QT, Huang G, Yuan Y, *et al.* An empirical study on evaluation metrics of generative adversarial networks. arXiv preprint arXiv: 1806.07755, 2018.
- 41 Liu XW, Huang LH, Guo ZY. Adaptive fourth-order partial differential equation filter for image denoising. Applied Mathematics Letters, 2011, 24(8): 1282–1288. [doi: [10.1016/j.aml.2011.01.028](https://doi.org/10.1016/j.aml.2011.01.028)]
- 42 Sendur L, Selesnick IW. Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency. IEEE Transactions on Signal Processing, 2002, 50(11): 2744–2756. [doi: [10.1109/TSP.2002.804091](https://doi.org/10.1109/TSP.2002.804091)]
- 43 Luisier F, Blu T, Unser M. A new SURE approach to image denoising: Interscale orthonormal wavelet thresholding. IEEE Transactions on Image Processing, 2007, 16(3): 593–606. [doi: [10.1109/TIP.2007.891064](https://doi.org/10.1109/TIP.2007.891064)]
- 44 Zhang K, Zuo WM, Chen YJ, *et al.* Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. IEEE Transactions on Image Processing, 2017, 26(7): 3142–3155. [doi: [10.1109/TIP.2017.2662206](https://doi.org/10.1109/TIP.2017.2662206)]
- 45 Wang TY, Sun MX, Hu KN. Dilated deep residual network for image Denoising. arXiv preprint arXiv: 1708.05473, 2017.
- 46 张元祺. 基于生成对抗网络的彩色图像去噪方法[硕士学位论文]. 大连: 大连理工大学, 2018.
- 47 Chen JW, Chen JW, Chao HY, *et al.* Image blind denoising with generative adversarial network based noise modeling. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. 2018. 3155–3164.