

生成对抗网络 GAN 综述*

梁俊杰¹, 韦舰晶¹⁺, 蒋正锋²

1. 湖北大学 计算机与信息工程学院, 武汉 430062

2. 广西民族师范学院 数学与计算机科学学院, 广西 崇左 532200

+ 通信作者 E-mail: wjj_huda2017@sina.com

摘要:生成对抗网络(GAN)作为一种新的无监督学习算法框架得到越来越多研究者的青睐,已然成为当下的一个研究热点。GAN受启发于博弈论中的二人零和博弈理论,其独特的对抗训练思想能生成高质量的样本,具有比传统机器学习算法更加强大的特征学习和特征表达能力。目前GAN在计算机视觉领域尤其是在样本生成领域取得显著成功,每年有大量GAN相关研究的论文产出。针对GAN这一热点模型,首先介绍了GAN的研究现状;接着介绍了GAN的理论、框架,详细分析了GAN在训练过程中存在梯度消失和模式崩溃的原因;然后讨论了一些典型的GAN的改进模型,总结了它们理论的改进之处、优点、局限性、应用场景以及实现成本,同时还将GAN与VAE、RBM模型进行比较,总结出GAN的优势和劣势;最后展示了GAN在数据生成、图像超分辨率、图像风格转换等方面的应用成果,并探讨了GAN目前面临的挑战以及未来的研究方向。

关键词:机器学习;无监督学习;生成对抗网络(GAN);梯度消失;模式崩溃

文献标志码:A **中图分类号:**TP391.41

梁俊杰, 韦舰晶, 蒋正锋. 生成对抗网络 GAN 综述[J]. 计算机科学与探索, 2020, 14(1): 1-17.

LIANG J J, WEI J J, JIANG Z F. Generative adversarial networks GAN overview[J]. Journal of Frontiers of Computer Science and Technology, 2020, 14(1): 1-17.

Generative Adversarial Networks GAN Overview*

LIANG Junjie¹, WEI Jianjing¹⁺, JIANG Zhengfeng²

1. School of Computer Science and Information Engineering, Hubei University, Wuhan 430062, China

2. School of Mathematics and Computer Science, Guangxi Normal University for Nationalities, Chongzuo, Guangxi 532200, China

Abstract: As a new unsupervised learning algorithm framework, generative adversarial networks (GAN) has been favored by more and more researchers, and it has become a research hotspot. GAN is inspired by the two-person zero-sum game theory in game theory. Its unique confrontation training idea can generate high-quality samples and has more powerful feature learning and feature expression ability than traditional machine learning algorithms. At present, GAN has achieved remarkable success in the field of computer vision, especially in the field of sample generation. Each year, there are a large number of GAN-related research papers. For the hotspot model of GAN,

* The National Natural Science Foundation of China under Grant No. 61300125 (国家自然科学基金).

Received 2019-09-30, Accepted 2019-11-20.

CNKI 网络出版: 2019-11-28, <http://kns.cnki.net/KCMS/detail/11.5602.TP.20191128.1738.002.html>

firstly this paper introduces the research status of GAN; then introduces the theory and framework of GAN, and analyzes the reasons why GAN has gradient disappearance and mode collapse during training; then discusses some typical models of GAN. This paper summarizes the improvement, advantages, limitations, application scenarios and implementation costs of the theory. At the same time, this paper compares GAN with VAE (variational autoencoder) and RBM (restricted Boltzmann machine) models, and summarizes the advantages and disadvantages of GAN. Finally, the application results of GAN in data generation, image super-resolution, image style conversion, etc. are presented, and the challenges and future research directions of GAN are discussed.

Key words: machine learning; unsupervised learning; generative adversarial networks (GAN); gradients disappearing; collapse mode

1 引言

近年来,随着计算能力的提高,数据量的积累以及对动物神经网络的研究及成果^[1],使得人工智能领域发展迅速,尤其在机器学习方面最为突出。依据数据集是否有标记,机器学习任务被分为有监督学习、无监督学习和半监督学习。目前机器学习方法,特别是深度学习方法在有监督学习任务中取得令人振奋的成绩,如图像识别^[2-3]、语音合成^[4-5]、机器翻译^[6-7]等。

有监督学习依赖带标记的数据,然而大量带标记数据的获取代价昂贵,在数据生成、策略学习等学习任务中,这些标记数据的获取甚至不可行。无监督学习更符合智能的思想,研究者们普遍认为,无监督学习将会是人工智能未来重要的发展方向之一。

生成模型是无监督学习任务中的关键技术,早期的生成模型有深度信念网络(deep belief network, DBN)^[8]、深度玻尔兹曼机(deep Boltzmann machines, DBM)^[9]等网络结构,它们将受限玻尔兹曼机(restricted Boltzmann machine, RBM)^[10]、自编码器(autoencoder, AE)^[11]等生成模型融合其中,形成了效果不错的生成式模型,但泛化能力却不强。

生成对抗网络(generative adversarial networks, GAN)是Goodfellow等人^[12]在2014年提出的一种新的生成式模型。GAN独特的对抗性思想使得它在众多生成器模型中脱颖而出,被广泛应用于计算机视觉(CV)、机器学习(ML)、语音处理(AS)等领域。在arXiv上以generative adversarial networks、generative adversarial nets和adversarial learning为关键词的论文发文量总体呈逐年上升趋势(如图1),并且GAN被应用的学科领域很广,如图2显示的是arXiv上GAN论文所属学科领域。这些数据说明了对GAN这一模型研究的火热程度,也说明了该方法在人工

智能等领域的重要性。

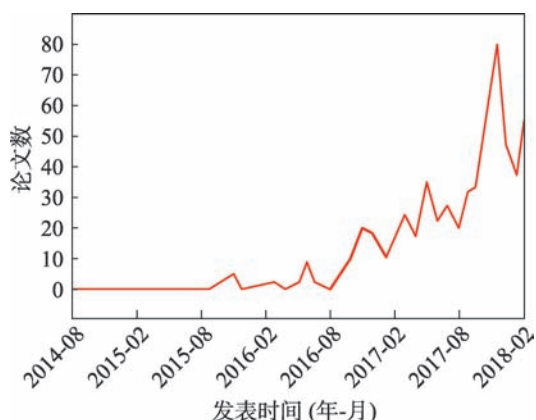


Fig.1 Trend of the number of GAN papers published on arXiv

图1 arXiv上GAN论文数量的变化趋势

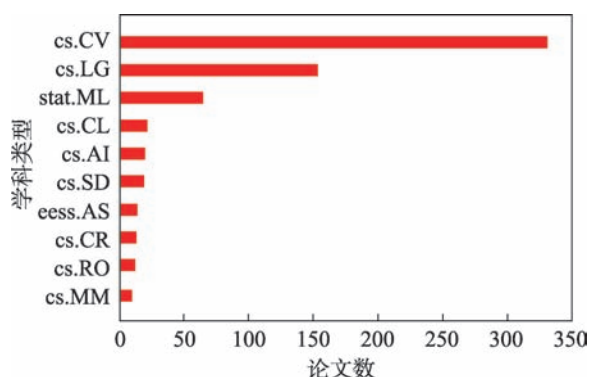


Fig.2 Top 10 subject categories of GAN papers published on arXiv

图2 arXiv上GAN论文所属的TOP10学科领域

2 原始生成对抗网络

2.1 GAN网络结构

GAN的网络结构由生成网络和判别网络组成,

模型结构如图3所示。生成器 G 接收随机变量 z , 生成假样本数据 $G(z)$ 。生成器的目的是尽量使得生成的样本和真实样本一样。判别器 D 的输入由两部分组成, 分别是真实数据 x 和生成器生成的数据 $G(x)$, 其输出通常是一个概率值, 表示 D 认定输入是真实分布的概率, 若输入来自真实数据, 则输出 1, 否则输出 0。同时判别器的输出会反馈给 G , 用于指导 G 的训练。理想情况下 D 无法判别输入数据是来自真实数据 x 还是生成数据 $G(z)$, 即 D 每次的输出概率值都为 1/2 (相当于随机猜), 此时模型达到最优。在实际应用中, 生成网络和判别网络通常用深层神经网络来实现。

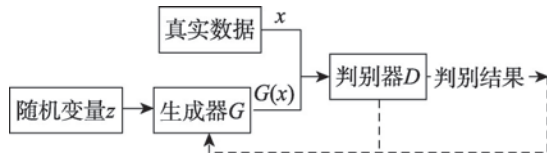


Fig.3 GAN network model structure diagram

图3 GAN网络模型结构示意图

GAN的思想来自于博弈论中的二人零和博弈^[13], 生成器和判别器可以看成是博弈中的两个玩家。在模型训练的过程中生成器和判别器会各自更新自身的参数使得损失最小, 通过不断迭代优化, 最终达到一个纳什均衡状态, 此时模型达到最优。GAN的目标函数定义为:

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

2.2 生成网络

生成器本质上是一个可微分函数, 生成器接收随机变量 z 的输入, 经 G 生成假样本 $G(z)$ 。在GAN中, 生成器对输入变量 z 基本没有限制, z 通常是一个100维的随机编码向量, z 可以是随机噪声或者符合某种分布的变量。生成器理论上可以逐渐学习任何概率分布, 经训练后的生成网络可以生成逼真图像, 但又不是和真实图像完全一样, 即生成网络实际上是学习了训练数据的一个近似分布, 这在数据增强应用方面尤为重要。

2.3 判别网络

判别器同生成器一样, 其本质上也是可微分函数, 在GAN中, 判别器的主要目的是判断输入是否为真实样本, 并提供反馈以指导生成器训练。判别器

和生成器组成零和游戏的两个玩家, 为取得游戏的胜利, 判别器和生成器通过训练不断提高自己的判别能力和生成能力, 游戏最终会达到一个纳什均衡状态, 此时生成器学习到了与真实样本近似的概率分布, 判别器已经不能正确判断输入的数据是来自真实样本还是来自生成器生成的假样本 $G(x)$, 即判别器每次输出的概率值都是 1/2。

3 原始 GAN 模型存在的问题分析

原始的GAN并不成熟, 存在着诸多问题, 其中梯度消失和模式崩溃 (collapse mode) 问题严重限制GAN的发展。只有了解问题发生的本质, 才能做出相应的改进, 本章主要对GAN在训练中存在梯度消失和模式崩溃的原因进行分析。

3.1 梯度消失问题分析

梯度消失即是利用误差反向传播 (back propagation, BP) 算法^[14]对深度神经网络进行训练时, 梯度后向传播到浅层网络时基本不能引起数值的扰动, 最终导致神经网络收敛很慢甚至不能收敛。GAN存在梯度消失的问题, 并且在判别器训练得越好的时候, 生成器梯度消失得越严重^[15]。最优判别器如式(2):

$$D_G^*(x) = \frac{p_{\text{data}}(x)}{p_{\text{data}}(x) + p_g(x)} \quad (2)$$

在最极端的情况下, 当判别器达到最优时, 此时生成器模型如式(3):

$$C(G) = \max_D V(G, D) = E_{x \sim p_{\text{data}}} \left[\ln \frac{p_{\text{data}}(x)}{p_{\text{data}}(x) + p_g(x)} \right] + E_{x \sim p_g} \left[\ln \frac{p_g(x)}{p_{\text{data}}(x) + p_g(x)} \right] \quad (3)$$

原始 GAN 模型使用 KL 散度 (Kullback-Leibler divergence)^[16] 和 JS 散度 (Jensen-Shannon)^[17] 衡量两个分布之间的差异, 即是:

$$C(G) = -\ln 4 + KL \left(p_{\text{data}} \parallel \frac{p_{\text{data}} + p_g}{2} \right) + KL \left(p_g \parallel \frac{p_{\text{data}} + p_g}{2} \right) = -\ln 4 + 2JS(p_{\text{data}} \parallel p_g) \quad (4)$$

其中: $KL(p_1 \parallel p_2) = E_{x \sim p_1} \left(\ln \frac{p_1}{p_2} \right)$

$$JS(p_1 \parallel p_2) = \frac{1}{2} KL \left(p_1 \parallel \frac{p_1 + p_2}{2} \right) + \frac{1}{2} KL \left(p_2 \parallel \frac{p_1 + p_2}{2} \right)$$

训练 GAN 网络需要极小化 $C(G)$, 即是要求 $\min(JS(p_{\text{data}} \parallel p_g))$, JS 散度的值越小表示两个分布之间

越接近,这符合生成器的优化目标,即是要生成以假乱真的样本(两个样本之间的概率分布很接近)。当两个分布有重叠的时候,优化JS散度是可行的,而在两个分布完全没有重叠部分,或者重叠的部分可以忽略时,JS散度是一个常数,此时梯度为0,即生成器在训练的过程中得不到任何的梯度信息,出现梯度消失的现象。下面从两个分布是否有重叠部分分析梯度消失的原因。

对于两个分布 $p_{\text{data}}(x)$ 和 $p_g(x)$, 不难得出如下四种情况,如图4所示。

情况1 $p_{\text{data}}(x)=0$ 且 $p_g(x)=0$

情况2 $p_{\text{data}}(x) \neq 0$ 且 $p_g(x) \neq 0$

情况3 $p_{\text{data}}(x)=0$ 且 $p_g(x) \neq 0$

情况4 $p_{\text{data}}(x) \neq 0$ 且 $p_g(x)=0$

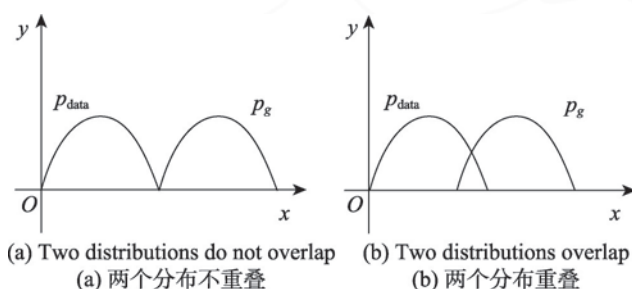


Fig.4 Sample probability distribution

图4 样本分布图

(1)两个分布没有重叠部分。对于情况1,由于 p_{data} 和 p_g 的取值都在函数的支撑集(support)之外,因此情况1对JS散度无贡献。情况2由于两个分布的支撑集没有交集,因此对最后的JS散度也无贡献。对于情况3,将 $p_{\text{data}}(x)=0$ 且 $p_g(x) \neq 0$ 带入式(4),得到式(5)。

$$JS(p_{\text{data}} \| p_g) = \frac{1}{2} KL \left(0 \parallel \frac{0 + p_g}{2} \right) + \frac{1}{2} KL \left(p_g \parallel \frac{0 + p_g}{2} \right) = 0 + \sum_{x \sim X} p_g(x) \times \ln 2 \quad (5)$$

由于 $\sum_{x \sim X} p_g(x) = 1$, 有 $JS(p_{\text{data}} \| p_g) = \ln 2$ 。同理,对于情况4,有 $JS(p_{\text{data}} \| p_g) = \ln 2$ 。说明此时生成器没有得到任何梯度信息,梯度消失。

(2)两个分布有重叠部分。文献[18]指出,两个分布在高维空间是很难相交的,即使相交,其相交部分其实是高维空间中的一个低维流形,其测度为0,

这说明两个分布相交部分可以忽略不计,此时JS散度的值和(1)讨论的一致,换言之, p_{data} 和 p_g 两个分布只要它们没有重叠部分或者重叠部分可以忽略,那么JS散度就固定是常数 $\ln 2$,这对于梯度下降法而言意味着梯度为0。文献[19]证明了若存在一个判别器 D 无限接近于最优解时,即 $\|D - D^*\| < \varepsilon$, 有如下关系:

$$\|\nabla_{\theta} E_{z \sim p(z)} [\ln(1 - D(g_{\theta}(z)))]\|_2 < M \frac{\varepsilon}{1 - \varepsilon} \quad (6)$$

这说明了当判别器训练得越好,无限接近于最优判别器 D_c^* 时,生成器的梯度消失越严重,即是:

$$\lim_{\|D - D^*\| \rightarrow 0} \nabla_{\theta} E_{z \sim p(z)} [\ln(1 - D(g_{\theta}(z)))] = 0 \quad (7)$$

3.2 模式崩溃问题分析

GAN模式崩溃(mode collapse)是指GAN生成不了多样性的样本,而是生成了与真实样本相同的样本,该缺陷在数据增强领域中是致命的。为解决GAN梯度消失问题,Goodfellow等人重新定义了损失函数,如式(8)。

$$\min_G \max_D V(D, G) = E_{x \sim p_{\text{data}}(x)} [\ln D(x)] + E_{z \sim p_g(z)} [-\ln D(x)] \quad (8)$$

在最优判别器 $D_c^*(x)$ 下,优化生成器的目标函数如式(9)(具体计算结合式(4)、式(8))。

$$\min(C(G)) = KL(p_g \| p_{\text{data}}) - 2JS(p_{\text{data}} \| p_g) \quad (9)$$

要优化该目标函数,则要求同时满足式(10)、式(11)。

$$\min(KL(p_g \| p_{\text{data}})) \quad (10)$$

$$\max(2JS(p_{\text{data}} \| p_g)) \quad (11)$$

式(10)要求 p_g 和 p_{data} 的概率分布一样,而式(11)则要求 p_g 和 p_{data} 的概率分布不一样,这样会产生矛盾,使得生成器无法稳定训练。放宽约束,只要求满足式(10)同样不可行。如下:

(1)当 $p_g(x) \rightarrow 0$ 且 $p_{\text{data}}(x) \rightarrow 1$ 时, $p_g(x) \ln \frac{p_g(x)}{p_{\text{data}}(x)} \rightarrow 0$, 此时的 $KL(p_g \| p_{\text{data}})$ 趋近于0。该情况说明了生成器生成了与真实样本相似的样本。

(2)当 $p_g(x) \rightarrow 1$ 且 $p_{\text{data}}(x) \rightarrow 0$ 时, $p_g(x) \ln \frac{p_g(x)}{p_{\text{data}}(x)} \rightarrow \infty$, 此时的 $KL(p_g \| p_{\text{data}})$ 趋近于正无穷。该情况说明生成器生成了不真实的样本。

情况(1)生成的样本缺乏多样性,惩罚较小,生成器宁可生成一些重复但很安全的样本,也不愿意生成多样性的样本,该现象就是模式崩溃。情况(2)

是生成器生成了不真实的样本,样本缺乏准确性,惩罚较大。

4 改进 GAN 模型

研究者们针对 GAN 存在训练困难等问题,通过不断探索,最终提出了很多基于 GAN 的变体。本章主要对一些典型的 GAN 变体进行讨论。

4.1 CGAN

原始 GAN 对于生成器几乎没有任何约束,使得生成过程过于自由,这在较大图片的情形中模型变得难以控制。CGAN (conditional GAN)^[20] 在原始 GAN 的基础上增加了约束条件,控制了 GAN 过于自由的问题,使网络朝着既定的方向生成样本。CGAN 的网络结构如图 5 所示,CGAN 生成器和判别器的输入多了一个约束项 y , 约束项 y 可以是一个图像的类别标签,也可以是图像的部分属性数据。CGAN 的目标函数变成如下式(12)。

$$\min_G \max_D V(D, G) = E_{x \sim p_{\text{data}}(x)} [\text{lb } D(x|y)] + E_{z \sim p_z(z)} [\text{lb}(1 - D(G(z|y)))] \quad (12)$$

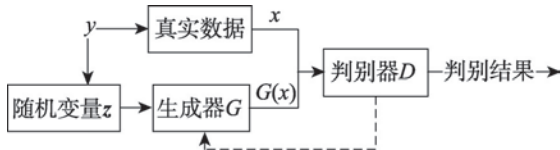


Fig.5 CGAN model architecture diagram

图 5 CGAN 模型架构图

CGAN 的缺点在于其模型训练不稳定,从损失函数可以看到,CGAN 只是为了生成指定的图像而增加了额外约束,并没有解决训练不稳定的问题。

4.2 LAPGAN

LAPGAN^[21] 基于 CGAN 进行改进,它能够生成高品质的图片,如图 6 是 LAPGAN 模型的学习过程。LAPGAN 的创新点是将图像处理领域中的高斯金字塔^[22]和拉普拉斯金字塔^[23]的概念引入 GAN 中,高斯金字塔用于对图像进行下采样,拉普拉斯金字塔用于上采样以此重建图像。高斯金字塔被定义为 $\wp(I)=[I_0, I_1, \dots, I_K]$, 其中 $I_0=I$ 是原始图像, I_{k+1} 是 I_k 下采样的图像, K 是尺度变换的层数,拉普拉斯金字塔由式(13)表示。

$$h_k = \wp_k(I) - u(\wp_{k+1}(I)) = I_k - u(I_{k+1}) \quad (13)$$

其中, $u(\cdot)$ 表示上采样,该式表达了原始图像与经过

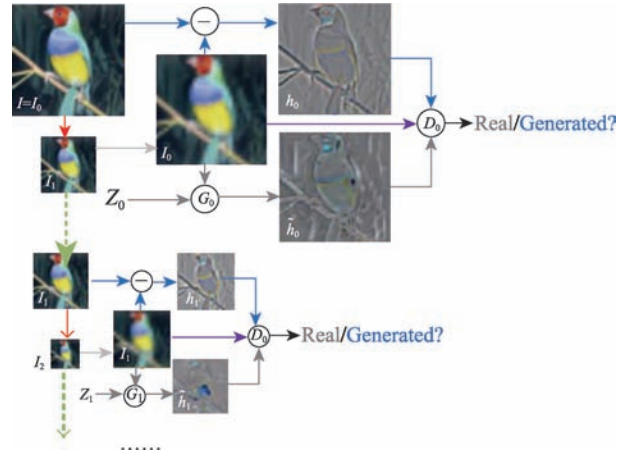


Fig.6 Learning process of LAPGAN model

图 6 LAPGAN 模型学习过程

变换之后图像的残差,即是 $h_k = I_k - u(d(I_k))$, 其中 $d(\cdot)$ 表示下采样。LAPGAN 实际是学习样本和生成样本之间的残差,如式(14)。

$$\tilde{I}_k = u(\tilde{I}_{k+1}) + \tilde{h}_k = u(\tilde{I}_{k+1}) + G_k(z_k, u(\tilde{I}_{k+1})) \quad (14)$$

4.3 DCGAN

DCGAN (deep convolutional GAN)^[24] 的提出对 GAN 的发展有着极大的推动作用,它将卷积神经网络 (convolutional neural networks, CNN) 和 GAN 结合起来,使得生成的图片质量和多样性得到了保证。其网络模型结构如图 7 所示。DCGAN 使用了一系列的训练技巧,如使用批量归一化 (batch normalization, BN) 稳定训练,使用 ReLU 激活函数降低梯度消失风险,同时取消了池化层,使用步幅卷积和微步幅卷积有效地保留了特征信息。DCGAN 虽然能生成多样性丰富的样本,但是生成的图像质量并不高,而且也没有解决训练不稳定的问题,在训练的时候仍需要小心地平衡 G 和 D 的训练进程。

4.4 CycleGAN

传统的不同两个域中的图片要实现相互转化,一般需要两个域中具有相同内容的成对图片作为训练数据。比如 pix2pix^[25],但是这种成对的训练数据往往很难获得。CycleGAN^[26] 可以让两个域的图片互相转化并且不需要成对的图片作为训练数据。CycleGAN 是一个互相生成的网络,其网络是个环形结构,如图 8 所示。

x 表示 X 域的图像, y 表示 Y 域的图像, G 和 F 是两个转换器, D_X 和 D_Y 是两个判别器。如图 8(a), X 域的图像 x 经转换器 G 转换成 Y 域的图片 $G(x)$, 并

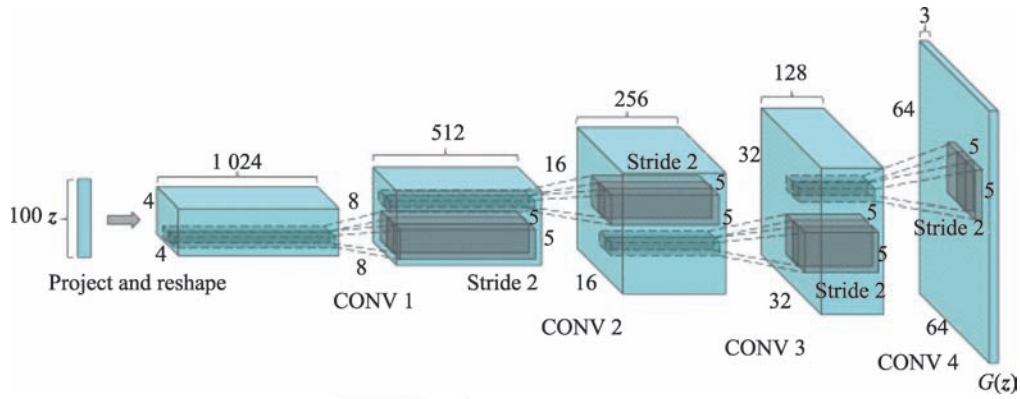


Fig.7 DCGAN model architecture diagram

图7 DCGAN生成模型架构图

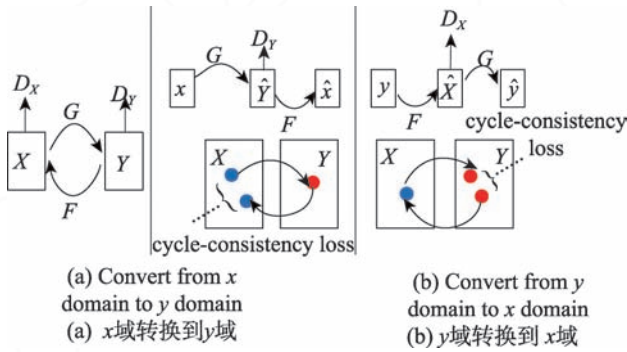


Fig.8 CycleGAN structure diagram

图8 CycleGAN 结构图

由判别器 D_y 判别它是否是真实图片。同理, Y 域的图像 y 经转换器 F 转换成 X 域的图片 $F(y)$, 再由判

别器 D_x 判别它是否是真实图片。为了避免转换器将域内的所有图片都转换成另一个域内的同一张图片, CycleGAN使用循环一致性损失(cycle consistency loss)做约束。最终的损失函数由三部分组成,如式(15)。

$$L = L_{GAN}(G, D_y, X, Y) + L_{GAN}(F, D_x, Y, X) + \lambda L_{cyc}(G, F) \quad (15)$$

CycleGAN的缺点是其循环机制能保证成像不会偏离太远,但是循环转换中会造成一定的信息丢失,使得生成图像质量不高。

4.5 StackGAN

根据文字描述人工生成对应高质量图片一直是计算机视觉领域的一个挑战, StackGAN^[27]是第一个根据文字描述生成图像分辨率达到 256×256 的网络模型。其网络模型结构图如图9所示。

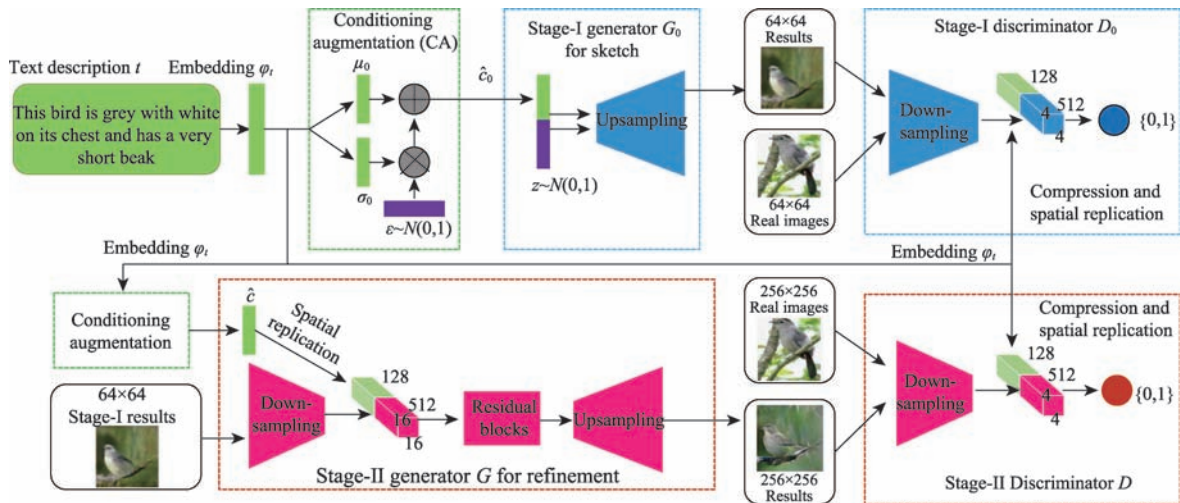


Fig.9 StackGAN network model structure

图9 StackGAN网络模型结构图

StackGAN 的训练分为两个阶段,它们分别被定义为 Stage-I 和 Stage-II。Stage-I 阶段是根据给定的文字描述,学习到初始的形状和色彩,生成低分辨率(64×64)的图片;Stage-II 阶段根据 Stage-I 生成的低分辨率图像以及原始文字描述,生成具有更多细节的高分辨率(256×256)图片。在 Stage-I 阶段中,词嵌入向量 φ_i 经过全连接层生成高斯分布 $\mathcal{N}(\mu_0(\varphi_i), \sigma_0(\varphi_i))$ 中的 μ_0 和 σ_0 ,最后与 z 噪声向量拼接作为 G_0 的输入,通过一组上采样生成 64×64 的图像。Stage-II 阶段以高斯隐含变量 \hat{c} 以及 Stage-I 生成器的输出作为输入,来训练生成器和判别器。但是这种分段式模式可能会出现每个任务找不到重点,最终导致生成失败。

StackGAN 及其改进版 StackGAN++^[28]都是用于文本生成图像的网络,目前在该领域比较新的突破性工作是由斯坦福大学李飞飞教授小组发表的文献[29],该研究引入了场景图的概念,相比于 StackGAN 和 StackGAN++ 方法,它能够处理更加复杂的文本,并且取得了不错的效果。

4.6 InfoGAN

GAN 强大的学习能力最终可以学习到真实样本的分布,但对输入噪声信号 z 和数据的语义特征之间的对应关系不清楚。一个理想的情况是清楚它们之间的对应关系,这样就能通过控制对应的维度变量来达到相应的变化。比如对于 MNIST 手写数字识别项目,在知道其对应关系的情况下,可以控制输出图像的光照、笔画粗细、字体倾斜度等。InfoGAN^[30]解决了这个问题,它将输入噪声 z 分成两部分,一部分是噪声信号 z ,另一部分是可解释的有隐含意义的信号 c 。InfoGAN 模型结构图如图 10 所示。

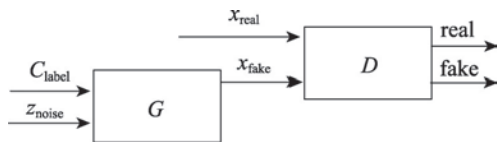


Fig.10 InfoGAN model structure

图 10 InfoGAN 模型结构

生成器的输入多了一个隐含变量 $C(c_1, c_2, \dots, c_L)$,它代表的是如上面提到的图像的光照、笔画粗细、字体倾斜度等图像的语义特征信息。InfoGAN 对目标函数进行了约束,即 c 和 $G(z, c)$ 之间的互信息 $I(c; G(z, c))$,如式(16)。

$$\min_c \max_D V_I(D, G) = V(D, G) - \lambda I(c; G(z, c)) \quad (16)$$

实验证明了 InfoGAN 确实学到了一些可解释的语义特征,通过控制这些特征可以生成想要的数。如图 11,通过控制角度、宽度,生成形状不一样的数据。

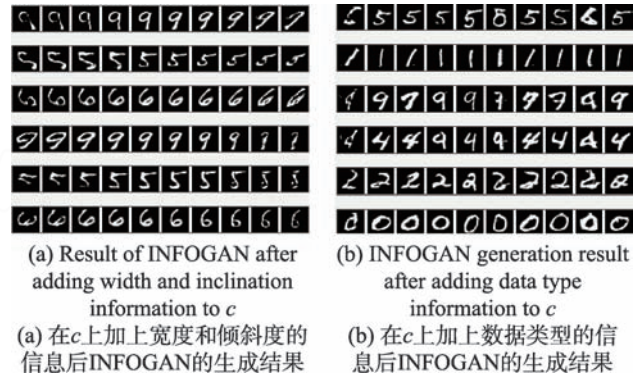


Fig.11 Feature learning of angle or width of image

图 11 图像的角度或宽度的特征学习

4.7 WGAN

原始 GAN 训练出现梯度消失和模式崩溃的主要原因使用 JS 距离来衡量两个分布, WGAN (Wasserstein GAN)^[31]使用 Wasserstein 距离(又称 EM (Earth-mover) 距离)代替 JS 距离对真实样本和生成样本之间的距离进行度量。

$$W(P_r, P_g) = \frac{1}{k} \sup_{\|f\|_L \leq k} E_{x \sim P_r}[f(x)] - E_{x \sim P_g}[f(x)] \quad (17)$$

Wasserstein 距离在两个分布没有重叠或者重叠部分可以忽略的情况下仍然能够很好地度量它们之间的距离,同时 Wasserstein 距离是平滑的,这在理论上能解决梯度消失/爆炸的问题。但在实验过程中发现, WGAN 往往存在梯度消失/爆炸的情况,其主要原因是为了满足 Lipschitz 限制条件,直接用 weight clipping 方式将参数 clip 到 $[-c, c]$ 范围内,在最优策略下,参数往往会走极端,即参数要么取最大值(如 0.01),要么取最小值(如 -0.01),这意味着拟合能力差,同时参数 c 很难确定,稍有不慎就会引起梯度消失/爆炸,如图 12 所示。

4.8 WGAN-GP

WGAN-GP^[32]是针对 WGAN 为满足 Lipschitz 连续性条件强行将权重剪切到一定范围,如 $[-0.01, 0.01]$ 之后造成网络拟合能力差、容易发生梯度消失/爆炸等问题而进行改进的一种 GAN 模型。作者抛弃

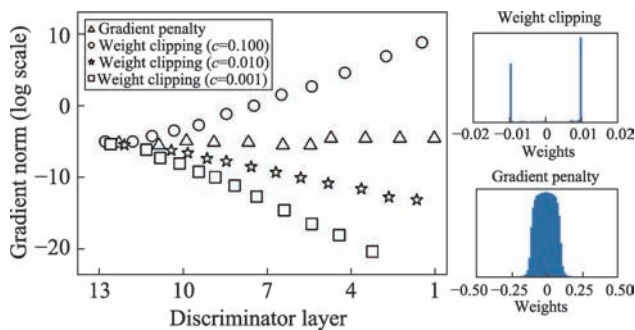


Fig.12 WGAN gradient disappearance/explosion analysis

图 12 WGAN 梯度消失/爆炸分析

了 WGAN 中的 weight clipping, 使用一种称作梯度惩罚 (gradient penalty) 的方式来满足 Lipschitz 连续性条件。具体而言, 它增加了一个额外的 loss 函数来实现判别器梯度与 K 之间的联系 (Lipschitz 限制是要求判别器梯度不超过 K), 如下:

$$E_{\hat{x} \sim P_g} [(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2] \quad (18)$$

这里 $K=1$ 。

WGAN-GP 解决了梯度消失/爆炸的问题, 同时能生成更高质量的样本。但在实验中发现该方法收敛速度慢, 在同一数据集下, WGAN-GP 需要训练更多次数才能收敛。

4.9 EBGAN

EBGAN (energy-based GAN)^[33] 从能量的角度诠释 GAN, 它将判别器 D 视为一个能量函数, 该能量函数将靠近真实样本域附近视为低能量区域, 远离真实样本区域则视作高能量区域。因此 EBGAN 在训练中, 生成器会尽可能生成能量最小的样本, 而判别器则会对这些生成的样本赋予高能量, 其模型结构如图 13 所示。

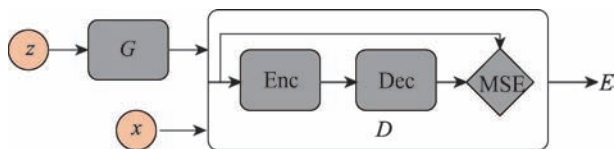


Fig.13 EBGAN model structure

图 13 EBGAN 模型结构

EBGAN 将判别器设计成能量函数, 这种改进的优点是可以使用更多、更广泛的损失函数来训练 GAN 结构, 但 EBGAN 同样存在缺点, 在实验中发现其收敛速度很慢, 其他模型都已经能够生成大致表达图像的轮廓, 它生成的图像依然杂乱无章。

4.10 BEGAN

传统的 GAN 是优化生成数据分布与真实数据分布之间的距离, 认为两者的分布越相似, G 的生成能力越好, 很多 GAN 的改进也是基于该思路。BEGAN (boundary equilibrium GAN)^[34] 并没有直接去估计生成分布 p_g 和真实分布 p_x 的距离, 而是估计两者分布误差的距离。作者认为, 分布之间的误差分布相近的话, 也可以认为 p_g 和 p_x 是相近的。BEGAN 将判别器 G 设计成自编码器用来重构分布误差, 并优化分布误差之间的距离, 如下式:

$$\begin{cases} L(x, D) = \|x - D(x)\| \\ L(G(z), D) = \|G(z) - D(G(z))\| \end{cases} \quad (19)$$

BEGAN 提出一种均衡概念, 用以平衡 G 和 D 的训练, 使 GAN 即使使用很简单的网络, 不加如 BN、minibatch 等训练技巧也能得到很好的训练效果。同时还提出了一种能够在样本多样性和样本质量上均衡的超参数以及衡量模型收敛性的方法。实验中发现 BEGAN 收敛很快, 并且 G 和 D 训练平衡, 但超参数的选取比较考量经验。

4.11 LSGAN

LSGAN (least squares GAN)^[35], 又叫最小二乘生成对抗网络, 具体工作是将传统 GAN 的交叉熵损失函数换成最小二乘损失函数, 该做法有效改善了传统 GAN 生成图片质量不高, 训练不稳定的问题。最小二乘损失与交叉熵损失相比, 优势在于生成样本在欺骗判别器的前提下同时让生成器把距离决策边界比较远的生成图片拉向决策边界, 这样保证了生成高质量的样本。作者认为以交叉熵作为损失, 会使得生成器不会再优化那些被判别器识别为真实图片的生成图片, 即使这些生成图片距离判别器的决策边界仍然很远, 也就是距离真实数据比较远, 因为此时的交叉熵损失已经很小, 生成器完成了为它设计的目标。LSGAN 的缺陷在于它并没有解决当判别器足够优秀时生成器发生梯度弥散的问题, 因为按照 WGAN 的理论 (式 (17)), 最小二乘函数并不满足 Lipschitz 连续性条件, 它的导数没有上界。

4.12 BigGAN

BigGAN^[36] 是目前为止生成图像质量最好的 GAN 模型, 它在 ImageNet (128×128 分辨率) 训练下, IS (inception score) 高达 166.3, 比之前最好的 52.52 (SAGAN (self-attention generative adversarial networks)^[37])

提升 2 倍多,同时 FID(frechet inception distance)下降至 9.6。BigGAN 通过将模型扩大得到性能的提升,它在训练中使用了很大的 Batch(2 048),并且在卷积层的设计上使用了更大的通道数(channel)。同时文章还使用截断技巧(truncation trick),使模型性能得

到提升,并且训练变得更加平稳。为了在样本多样性和保真度上进行平衡,将正交正则化应用到生成器上,同时对判别器的梯度做适当惩罚。BigGAN 的缺点也很明显,BigGAN 模型很大,参数多,训练成本高。

表 1 是典型 GAN 模型的对比。

Table 1 Comparisons of typical GAN models

表 1 典型 GAN 模型对比

| GAN 模型 | 机制 | 优势 | 缺点 | 适用范围 | 实现成本 |
|--------------------------|--------------------------------------|------------------------------------|--------------------------------|--------------------------------|---------------------------------------|
| CGAN ^[20] | 对 G 和 D 增加一个条件变量 c 约束模型,指导数据生成 | 有效约束了 GAN 过于自由的问题,能生成指定目标图像 | 对数据集要求高,需要有标签数据,训练不平衡,生成图片质量不高 | 适合生成指定目标图像的场景 | 在 CPU 上即可实现,但需要对应的标签数据,在数据收集上成本较高 |
| DCCGAN ^[24] | 将 GAN 与 CNN 结合,使用 BN 等技巧训练模型 | 生成图像多样性丰富 | 生成图片质量不高,模型训练不稳定 | 适用大部分图像生成场景 | 只需自然图像即可训练,可运行在 CPU 上,实现成本较低 |
| LAPGAN ^[21] | 使用高斯金字塔对图像下采样,使用拉普拉斯金字塔上采样重建高质量图像 | 对残差的学习和逼近较容易,收敛速度快,生成的样本分辨率高 | 必须在有监督下训练 | 生成高品质的自然图像 | 在 CPU 上即可实现,但同时训练多个 GAN 可能带来时间和内存成本增大 |
| InfoGAN ^[30] | 对随机噪声 z 施加可解释的隐含变量,控制图像生成 | 加大了对 GAN 模型的可解释性,能控制图像生成 | 因加入可解释性信息控制图像生成,造成多样性差 | 适用于生成指定目标样本的场景 | 在 CPU 上即可实现,成本较低 |
| WGAN ^[31] | 使用 Wasserstein 距离代替传统 GAN 中的 JS 散度 | 训练更加稳定,理论上解决了模式崩溃和梯度消失/爆炸的问题 | 对权值不恰当的剪枝会造成梯度弥散 | 适用于 GAN 模型不收敛,模式崩溃的情况 | 对截断参数 c 的选取需要足够丰富的经验,实现成本较高 |
| WGAN-GP ^[32] | 使用梯度惩罚机制来满足 Lipschitz 连续条件 | 训练稳定,训练过程中不再需要精心平衡 G 和 D 的训练程度 | 收敛速度慢,生成样本多样性不足 | 模型参数不确定情况下 | 需要更大的时间成本才能使训练达到收敛 |
| EBGAN ^[33] | 从能量角度诠释 GAN,对真实样本赋予低能量,生成样本赋予高能量 | 能使用更多、更广泛的网络和损失函数训练 | 收敛速度慢,同一个数据集往往要训练多次才收敛 | 适用大部分图像生成场景 | 需要更大的时间成本才能使训练达到收敛 |
| BEGAN ^[34] | 估计生成数据分布与真实数据分布之间的误差,并优化误差分布 | 不需担心模型崩溃和训练不平衡问题,生成图片质量高 | 对平衡训练超参数的选取较考量经验,需要多次调参 | 适用于 GAN 模型不收敛,模式崩溃,需要生成高品质图像场景 | 在 CPU 上即可实现,但额外训练一个自编码网络可能带来更多的内存消耗 |
| LSGAN ^[35] | 使用最小二乘损失函数代替传统交叉熵损失函数 | 生成高质量的样本 | 训练时易发生梯度消失/爆炸情况 | 适用于生成高质量图片 | 只需自然图像即可训练,可运行在 CPU 上,实现成本较低 |
| BigGAN ^[36] | 扩大模型规模,使用截断、正交正则化等技巧训练 | 模型训练稳定,能生成超清晰图像 | 参数量大,训练困难 | 适用于生成超清晰图像 | 庞大的参数需要 TPU 这样的硬件进行训练,成本高 |
| CycleGAN ^[26] | 提出循环机制,实现两个域转换和约束 | 对数据要求低,不需要一对一配对图像,只需提供不同风格图像 | 生成图像质量较低,分辨率不高 | 大部分风格转移场景,如艺术作品风格迁移等 | 对数据要求不高,在 CPU 上即可实现,成本较低 |
| StackGAN ^[27] | 将两个 GAN 叠加,施行分段式训练,提高图像分辨率 | 生成高分辨率的图像 | 分段式模式可能出现每个任务找不到重点,导致生成失败 | 从文本到清晰图像的生成 | 同时训练两个 GAN,可能带来更多的内存消耗 |

机制有效充分运用了两个域的信息,同时其对训练数据要求不高,只需提供不同风格数据图像就可以实现风格转换,实现成本低,但循环转换会造成一定的信息丢失,使得生成样本质量较低;StackGAN 应用在文本到图像生成的领域,它能生成高分辨率的图像,但是其分段式模式也往往造成各个分任务找不到重点,导致生成失败。

5 GAN 与 VAE、RBM 比较

变分自编码器(variational autoencoder, VAE)和受限玻尔兹曼机(RBM)是除 GAN 之外的两个优秀的生成式模型,在 MNSIT 数据集上对比了 GAN 与 VAE、RBM 的性能,最后总结 GAN 对比于这两大生成式模型存在的优势与劣势。

从图 15 可以看到 VAE 和 RBM 生成的图像相对于 GAN 而言较模糊,这说明了 GAN 能生成更加清晰的图像。图 16 中图(a)、(b)分别为 GAN 生成器和判别器的损失,图(c)、(d)分别为 KL 散度变化以及 VAE 损失变化的过程,可以看到 GAN 是比较难训练的。总体而言,GAN 的优势在于:

- (1)从实际结果看,GAN 能产生更好的生成样本。
- (2)GAN 框架可以训练任何生成网络,在模型设计上不需要遵循任何种类的因子分解,所有的生成器和鉴别器都可以正常工作。
- (3)相比 RBM,GAN 没有变分下界,也没有棘手的配分函数,样本是一次生成的,并没有重复地应用 Markov Chain 来生成。

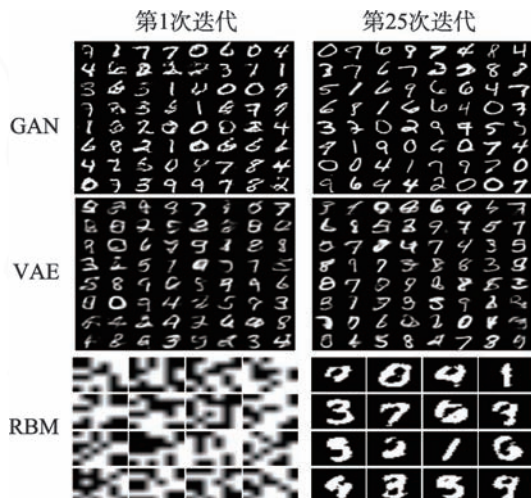


Fig.15 3 generation models generate results on MNIST
图 15 3 种生成模型在 MNIST 上生成结果

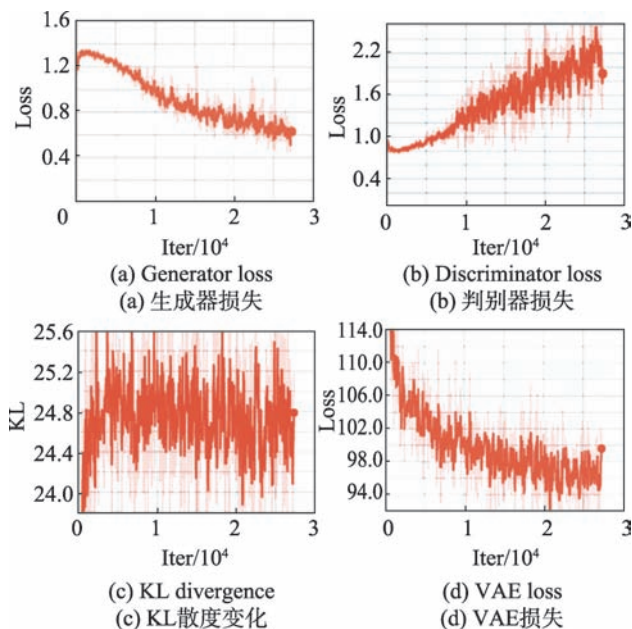


Fig.16 GAN and VAE training
parameters change process

图 16 GAN 和 VAE 训练参数变化过程

(4)相比于 VAE, GAN 没有引入任何决定性偏置(deterministic bias), VAE 由于优化的是对数似然的下界,而不是对数似然本身,因此引入了决定性偏置,这也导致了其生成的实例比 GAN 更模糊。

GAN 的劣势体现在:

- (1)训练 GAN 需要达到纳什均衡,目前没有找到很好的达到纳什均衡的方法,甚至是否存在纳什均衡仍然没有很好的理论证明,因此 GAN 相对于 VAE 而言其训练是不稳定的。
- (2)GAN 对离散数据的生成效率很低,如文本数据。
- (3)可解释性差,生成模型的分布没有显式的表达;极易受到对抗样本的干扰,导致模型错误的输出。

6 GAN 的典型应用

6.1 生成数据

目前,数据缺乏仍是限制深度学习发展的重要因素之一,而 GAN 能够从大量的无标签数据中无监督地学习到一个具备生成各种形态(图像、语音、语言等)数据能力的函数(生成器)。以生成图像为例,GAN 能够生成百万级分辨率的高清图像^[38],如 BigGAN、WGAN, WGAN-GP 等优秀的模型。但是 GAN 并不是单纯地对真实数据的复现,而是具备一

定的数据内插和外插作用,因此可以达到数据增广的目的。比如用GAN来生成一张红颜色的汽车图像,GAN生成的汽车图像颜色可能变成了黑色,或者在汽车的轮胎、车窗上和原图像有差别,这样一张图像就可以生成出很多张不同的图像,与经过剪切、旋转生成的图像相比,使用GAN生成的图像数据更符合模型训练的需要。不同GAN模型生成图像结果如图17所示。

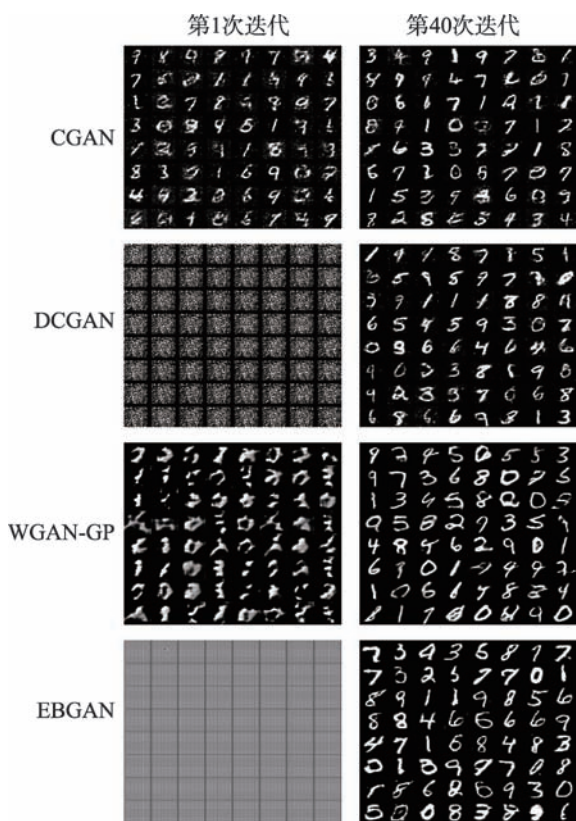


Fig.17 Image results generated by different GAN models

图17 不同GAN模型生成图像结果

6.2 图像超分辨率

图像超分辨率一直是计算机视觉领域的一个研究热点,SRGAN(super-resolution generative adversarial network)^[39]是GAN在图像超分辨率应用上的一个成功案例。SRGAN基于相似性感知方法^[40]提出了一种新的损失函数,有效解决了恢复后图像丢失高频细节的问题,并使人能有良好视觉感受。在此之前的其他超分辨率图像处理模型如Residual Network^[41]以及在其基础上改进而来的SRResNet^[42]都是追求高的PSNR(peak signal-to-noise ratio),但在实验中发现较

高的PSNR并不能带来良好的视觉效果^[43-44]。SRGAN从特征上定义损失,它将生成样本和真实样本分别输入VGG-19网络,然后根据得到的特征图的差异来定义损失项,最后将对抗损失、图像平滑项(生成图像的整体方差)和特征图差异这3个损失项作为模型的损失函数,得到了很好的效果。图18是应用在图像超分辨率上的网络模型实验的结果,可以看到SRGAN的PSNR虽然不是最高,但生成的图像细节、纹理更明显。



Fig.18 Super-resolution image generation

图18 超分辨率图像生成

6.3 图像翻译和风格转换

图像风格迁移是将一张图片的风格“迁移”到另一张图片上。深度学习最早是使用CNN框架来实现的^[45],但这样的模型存在训练速度慢,对训练样本要求过高等问题。由于GAN的自主学习和生成随机样本的优势,以及降低了对训练样本的要求,使得GAN在图像风格迁移领域取得了丰硕的研究成果^[46-52]。

6.4 其他

GAN除了应用在数据生成、图像超分辨率、图像翻译和风格迁移等领域外,在其他领域也有着很好的表现,如与人体相关的人体姿态估计^[53]、行为轨迹追踪^[54]、人体合成^[55];与人脸相关的人脸检测、人脸合成^[56]、人脸表情识别^[57]、预测年龄^[58]等。此外GAN还在图像分割^[59]、图像复原^[60]、图像补全^[61]中表现出色,文献[60]使用生成对抗网络思想的对抗损失提出了一种端到端的模糊图像复原方法;文献[61]提出一种基于生成式对抗网络的图像补全方法,生成器用于对图像缺失区域进行补全,判别器对补全效果判别并指导生成器学习。GAN对时间序列数据同样表现出强大的学习能力,如音乐生成^[62]、视频预测与生成^[63-64]。文献[65]使用一种对偶对抗学习机制,经过训练后能够准确预测下一帧或多帧的视频;文献[64]

采用将随机噪声向量映射到视频帧的方法来实现未来帧的预测;不仅如此,GAN 在重症监护室的 ICU 记录生成^[66]、检测恶意代码攻击^[67]、医疗影像分割^[68]、目标跟踪检测^[69-71]、自动驾驶^[72]等领域都有令人非常惊艳的表现。

7 面临的挑战和未来的研究方向

7.1 GAN 面临的挑战

(1)性能与成本权衡的问题。SAGAN、BigGAN 等都证明了庞大的 GAN 模型能得到性能上的提升,如 BigGAN 生成了超清晰的图像,但是庞大的模型需要使用 TPU 或者多个 GPU 训练,成本高。一个可行的方法是保证性能的同时将模型压缩,如使用模型量化^[73]等方法,但这仍是一个严峻的挑战。

(2)GAN 在使用类别特定数据集(如人脸、卧室)进行训练时已经能生成逼真、高质量样本,但建模具有多个类别,高度多样化的数据集(如 ImageNet)的分布仍然是一项重大的挑战,通常需要根据另一种输入信号来调节生成或为特定任务训练模型,此外 GAN 在这种高可变性的数据集中难以生成全局一致的高分辨率样本。

(3)GAN 的全局收敛性未得到真正证明。GAN 的训练方式是交替优化生成器和判别器,这和其他神经网络不同。目前比较常用的是基于博弈理论原理进行 GAN 的训练,但这一过程是在资源限制的情况下完成的,下一步应该要研究怎样减少这样的资源限制。或者试图分析常规神经网络来解答 GAN 收敛性的问题,因为生成器模型和判别器模型的参数均为非凸性损失函数,这与绝大多数神经网络有相似之处。

7.2 GAN 未来的研究方向

虽然目前 GAN 的理论结构仍不太成熟,限制了 GAN 在相关领域的应用与发展,但 GAN 仍然是一种很有发展前景的网络模型,未来 GAN 将在以下方面取得进一步的发展:

(1)理论突破。GAN 的理论基础是博弈论中的二人零和博弈,目前还没能在理论上真正证明均衡点的存在性,同时 GAN 训练的不稳定性如梯度消失、模式崩溃、过拟合和生成样本自由度较高等模型性能问题仍是当前 GAN 所面临的问题。从理论层面上进行突破,找出导致以上问题的根本原因并做出改

进是 GAN 未来的一个重点研究方向。

(2)算法拓展。GANs 在图像合成、图像超分辨率等连续样本上表现良好,但是在离散样本如文本处理上的表现还低于基于似然的语言模型。对 GAN 的算法拓展,使 GAN 这一个优秀模型的应用范围更加广泛,是迫切需要解决的问题。目前的一个方向是 GAN 吸收机器学习中最新的理论与研究成果并与之相结合,如 GAN 与强化学习结合。

(3)解决对抗样本带来的困扰。分类器极易受到对抗样本的影响,往往会因为故意添加的一些细微干扰就会导致模型判断错误。比如判别器 D 判别一个生成样本 $G(z)$ 为假,却将加了扰动的生成样本 $G(z)+p$ 判定为真。对抗样本影响 GAN 的性能,如何解决这个问题,使得模型的精度和稳定性有所提升,是未来一个充满前景的研究方向。

(4)证明 GAN 可以为什么样的分布建模。研究者们发现,GAN 在 MNIST 数据集上训练比在 Imagenet、CIFAR-10 等数据集上训练更为容易,原因是 MNIST 数据集的类别相对较少。这说明 GAN 对数据集的规模存在偏好。GAN 可以对一些数据建模但效率很低,或者甚至存在一些 GAN 无法学习的分布。如何建立一套理论和标准来判断一个数据集是否适合使用 GAN 来训练,将是 GAN 领域未来的一个重要的研究方向。

(5)完善评测体系。目前 GAN 模型没有一个科学的、统一的评估标准,各项研究工作中无非是提出适用于当前研究成果的判定方法,或是直接由人的肉眼来评判。因此提出一个更精确的评价指标,采用统一的标准,构建标准化、通用化的科学评估体系是亟待解决的问题。

(6)针对 GAN 可解释性差进行改进。GAN 一般是输入一个噪声信号 z 进行训练,但是对输入的噪声信号和数据的语义特征之间的对应关系不清楚,使得很难控制 GAN 生成需要的样本,虽然 InfoGANs 等模型对此做了相应的一些改进,但这方面仍缺乏系统的研究。解决 GAN 解释性差的问题,将会对 GAN 未来的发展有着重大的意义。

8 结束语

本文对 GAN 这一模型在原理、应用、存在挑战和未来研究方向等方面进行了介绍和总结,GAN 模型

的提出为人工智能领域注入了新的活力,特别在无监督学习上,GAN的对抗性思想提供了一种新的算法框架,促进了这一领域的发展。虽然GAN目前还存在诸多未解决的难题,但对GAN的继续探索,相信未来一定能突破这些难题,打造出更优秀的模型。

References:

- [1] Mantziaris C, Bockemühl T, Holmes P, et al. Intra- and intersegmental influences among central pattern generating networks in the walking system of the stick insect[J]. *Journal of Neurophysiology*, 2017, 118(4): 2296-2310.
- [2] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]//Proceedings of the 25th Annual Conference on Neural Information Processing Systems, Lake Tahoe, Dec 3-7, 2012. Cambridge: MIT Press, 2012: 1106-1114.
- [3] Wu J, Qian X Z. Application of compact deep convolutional neural network in image recognition[J]. *Journal of Frontiers of Computer Science and Technology*, 2019, 13(2): 275-284.
- [4] Dou G C, Cai R Z M, Nan C J, et al. Tibetan speech synthesis based on neural network[J]. *Journal of Chinese Information Processing*, 2019, 33(2): 75-80.
- [5] Hinton G, Deng L, Yu D, et al. Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups[J]. *IEEE Signal Processing Magazine*, 2012, 29(6): 82-97.
- [6] Collobert R, Weston J, Bottou L, et al. Natural language processing (almost) from scratch[J]. *Journal of Machine Learning Research*, 2011, 12: 2493-2537.
- [7] Sutskever I, Vinyals O, Le Q V. Sequence to sequence learning with neural networks[C]//Proceedings of the 27th Annual Conference on Neural Information Processing Systems, Montreal, Dec 8-13, 2014. Cambridge: MIT Press, 2014: 3104-3112.
- [8] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets[J]. *Neural Computation*, 2006, 18(7): 1527-1554.
- [9] Salakhutdinov R, Hinton G E. Deep Boltzmann machines [C]//Proceedings of the 12th International Conference on Artificial Intelligence and Statistics, Clearwater Beach, Apr 16-18, 2009: 448-455.
- [10] Smolensky P. Information processing in dynamical systems: foundations of harmony theory[M]//Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Cambridge: MIT Press, 1986.
- [11] Hinton G E, Zemel R S. Autoencoders, minimum description length and Helmholtz free energy[C]//Proceedings of the 6th International Conference on Neural Information Processing Systems, Denver, Nov 29-Dec 2, 1993. San Francisco: Morgan Kaufmann Publishers Inc, 1994: 3-10.
- [12] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]//Proceedings of the 27th International Conference on Neural Information Processing Systems, Montreal, Dec 8-13, 2014. Cambridge: MIT Press, 2014: 2672-2680.
- [13] Mertens J F, Zamir S. The value of two-person zero-sum repeated games with lack of information on both sides[J]. *International Journal of Games Theory*, 1971, 1(1): 39-64.
- [14] Rumelhart D E, Hinton G E, Williams R J. Learning representations by backpropagating errors[J]. *Nature*, 1986, 323(6088): 533-536.
- [15] Arjovsky M, Leon B. Towards principled methods for training generative adversarial networks[J]. arXiv:1701.04862, 2017.
- [16] Raiber F, Kurland O. Kullback-Leibler divergence revisited [C]//Proceedings of the ACM SIGIR International Conference on Theory of Information Retrieval, Amsterdam, Oct 1-4, 2017. New York: ACM, 2017: 117-124.
- [17] Fuglede B, Topsøe F. Jensen-Shannon divergence and Hilbert space embedding[C]//Proceedings of the 2004 IEEE International Symposium on Information Theory, Chicago, Jun 27-Jul 2, 2004. Piscataway: IEEE, 2004.
- [18] Zheng H B. Awesome Wasserstein GAN[EB/OL]. (2017-04-02) [2018-01-20]. <https://zhuanlan.zhihu.com/p/25071913>.
- [19] Browne M, Ghidary S. Convolutional neural networks for image processing: an application in robot vision[C]//LNCS 2903: Australasian Joint Conference on Artificial Intelligence, Perth, Dec 3-5, 2003. Berlin, Heidelberg: Springer, 2003: 641-652.
- [20] Mirza M, Osindero S. Conditional generative adversarial nets[J]. arXiv:1411.1784, 2014.
- [21] Denton E, Chintala S, Szlam A, et al. Deep generative image models using a Laplacian pyramid of adversarial networks[C]//Proceedings of the 28th International Conference on Neural Information Processing Systems, Montreal, Dec 7-12, 2015. Cambridge: MIT Press, 2015: 1486-1494.
- [22] Olkkonen H, Pesola P. Gaussian pyramid wavelet transform for multiresolution analysis of images[J]. *Graphical Models & Image Processing*, 1996, 58(4): 394-398.

- [23] Burt P J, Adelson E H. The Laplacian pyramid as a compact image code[J]. *IEEE Transactions on Communications*, 1983, 31(4): 532-540.
- [24] Zhu W, Miao J, Qing L, et al. Unsupervised representation learning with deep convolutional generative adversarial networks computer science[J]. *arXiv:1511.06434*, 2015.
- [25] Isola P, Zhu J Y, Zhou T H, et al. Image-to-image translation with conditional adversarial networks[C]//*Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, Jul 21-26, 2017. Washington: IEEE Computer Society, 2017: 5967-5976.
- [26] Zhu J Y, Pakr T, Isola P, et al. Unpaired image-to-image translation using cycleconsistent adversarial networks[C]//*Proceedings of the 2017 IEEE International Conference on Computer Vision*, Venice, Oct 22-29, 2017. Washington: IEEE Computer Society, 2017: 2242-2251.
- [27] Zhang H, Xu T, Li H S, et al. Stackgan: text to photo-realistic image synthesis with stacked generative adversarial networks[C]//*Proceedings of the 2017 IEEE International Conference on Computer Vision*, Venice, Oct 22-29, 2017. Washington: IEEE Computer Society, 2017: 5908-5916.
- [28] Zhang H, Xu T, Li H S, et al. StackGAN++: realistic image synthesis with stacked generative adversarial networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019, 41(8): 1947-1962.
- [29] Johnson J, Gupta A, Li F F. Image generation from scene graphs[C]//*Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, Jun 18-22, 2018. Washington: IEEE Computer Society, 2018: 1219-1228.
- [30] Chen X, Duan Y, Houthoofd R, et al. InfoGAN: interpretable representation learning by information maximizing generative adversarial nets[C]//*Proceedings of the 2016 International Conference on Neural Information Processing Systems*, Barcelona, Dec 9, 2016. Cambridge: MIT Press, 2016: 1-8.
- [31] Arjovsky M, Chintala S, Bottou L. Wasserstein GAN[J]. *arXiv:1701.07875*, 2017.
- [32] Gulrajani I, Ahmed F, Arjovsky M, et al. Improved training of Wasserstein GANs[J]. *arXiv:1704.00028*, 2017.
- [33] Zhao J B, Mathieu M, LeCun Y. Energy based generative adversarial network[J]. *arXiv:1609.03126*, 2016.
- [34] Berthelot D, Schumm T, Metz L. BEGAN: boundary equilibrium generative adversarial networks[J]. *arXiv:1703.10717*, 2017.
- [35] Mao X D, Li Q, Xie H R, et al. Least squares generative adversarial networks[C]//*Proceedings of the 2017 IEEE International Conference on Computer Vision*, Venice, Oct 22-29, 2017. Washington: IEEE Computer Society, 2017: 2813-2821.
- [36] Brock A, Donahue J, Simonyan K. Large scale GAN training for high fidelity natural image synthesis[J]. *arXiv:1809.11096*, 2018.
- [37] Zhang H, Goodfellow I, Metaxas D, et al. Self attention generative adversarial networks[J]. *arXiv:1805.08318*, 2018.
- [38] Karras T, Aila T, Laine S, et al. Progressive growing of GANs for improved quality, stability, and variation[J]. *arXiv:1710.10196*, 2017.
- [39] Ledig C, Theis L, Huszar F, et al. Photorealistic single image superresolution using a generative adversarial network[J]. *arXiv:1609.04802*, 2016.
- [40] Alexey D, Brox T. Generating images with perceptual similarity metrics based on deep networks[J]. *arXiv:1602.02644*, 2016.
- [41] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//*Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, Jun 27-30, 2016. Washington: IEEE Computer Society, 2016: 770-778.
- [42] Ledig C, Theis L, Huszar F, et al. Photorealistic single image superresolution using a generative adversarial network [C]//*Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, Jul 21-26, 2017. Washington: IEEE Computer Society, 2016: 105-114.
- [43] De Boer J F, Cense B, Park B H, et al. Improved signal-to-noise ratio in spectraldomain compared with timedomain optical coherence tomography[J]. *Optics Letters*, 2003, 28(21): 2067-2069.
- [44] Ardenkjer-Larsen J, Fridlund B, Gram A, et al. Increase in signal-to-noise ratio of >10,000 times in liquid-state NMR [J]. *Proceedings of the National Academy of Sciences*, 2003, 100(18): 10158-10163.
- [45] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks[J]. *arXiv:1511.06434*, 2015.
- [46] Kim T, Cha M, Kim H, et al. Learning to discover cross-domain relations with generative adversarial networks[C]//*Proceedings of the 34th International Conference on Machine Learning*, Sydney, Dec 18-21, 2017: 1730-1741.
- [47] Taigman Y, Polyak A, Wolf L. Unsupervised cross domain image generation[J]. *arXiv:1611.02200*, 2016.

- [48] Chen Y, Lai Y K, Liu Y J. CartoonGAN: generative adversarial networks for photo cartoonization[C]//Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, Jun 18-22, 2018. Washington: IEEE Computer Society, 2018: 9465-9474.
- [49] Chang H, Lu J, Yu F, et al. PairedCycleGAN: asymmetric style transfer for applying and removing makeup[C]//Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, Jun 18-22, 2018. Washington: IEEE Computer Society, 2018: 40-48.
- [50] Choi Y, Choi M, Kim M, et al. StarGAN: unified generative adversarial networks for multidomain image-to-image translation[C]//Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, Jun 18-22, 2018. Washington: IEEE Computer Society, 2018: 8789-8797.
- [51] Shuang M, Jian L F, Chang W C, et al. DA-GAN: instance-level image translation by deep attention generative adversarial networks[C]//Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, Jun 18-22, 2018. Washington: IEEE Computer Society, 2018: 5657-5666.
- [52] Azadi S, Fisher M, Kim V, et al. Multicontent GAN for few-shot font style transfer[C]//Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, Jun 18-22, 2018. Washington: IEEE Computer Society, 2018: 7564-7573.
- [53] Chen Y S, Wei W, Liang W, et al. Multistage adversarial losses for posebased human image synthesis[C]//Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, Jun 18-22, 2018. Washington: IEEE Computer Society, 2018: 118-126.
- [54] Gupta A, Johnson J, Li F F, et al. Social GAN: socially acceptable trajectories with generative adversarial networks[C]//Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, Jun 18-22, 2018. Washington: IEEE Computer Society, 2018: 2255-2264.
- [55] Li Q M, Qian R S, Georgoulis S, et al. Disentangled person image generation[C]//Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, Jun 18-22, 2018. Washington: IEEE Computer Society, 2018: 99-108.
- [56] Jian M B, Dong C, Fang W, et al. Towards openset identity preserving face synthesis[C]//Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, Jun 18-22, 2018. Washington: IEEE Computer Society, 2018: 6713-6722.
- [57] Guo Z P, Shang F W. Weakly supervised facial action unit recognition through adversarial training[C]//Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, Jun 18-22, 2018. Washington: IEEE Computer Society, 2018: 2188-2196.
- [58] Hong Y Y, Di H, Yun H W, et al. Learning face age progression: a pyramid architecture of GANs[C]//Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, Jun 18-22, 2018. Washington: IEEE Computer Society, 2018: 31-39.
- [59] Zi Z Z, Lin Y, Ye F Z. Translating and segmenting multimodal medical volumes with cycle and shape-consistency generative adversarial network[C]//Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, Jun 18-22, 2018. Washington: IEEE Computer Society, 2018: 9242-9251.
- [60] Sang L, Gao S, Yin Z S. Motion blurred image restoration based on generating confrontation network[J]. Computer Engineering and Applications, 2019, 55(6): 173-177.
- [61] Shi C, Pan B, Guo X M, et al. Application of generating confrontation network in image completion[J]. Journal of Frontiers of Computer Science and Technology, 2019, 13(8): 1402-1410.
- [62] Yang L C, Chou S Y, Yang Y H. MidiNet: a convolutional generative adversarial network for symbolic-domain music generation[J]. arXiv:1703.10847, 2017.
- [63] Mathieu M, Couprie C, LeCun Y. Deep multiscale video prediction beyond mean square error[J]. arXiv:1511.05440, 2015.
- [64] Tulyakov S, Liu M Y, Yang X D, et al. MoCoGAN: decomposing motion and content for video generation[C]//Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, Jun 18-22, 2018. Washington: IEEE Computer Society, 2018: 1526-1535.
- [65] Liang X D, Lee L, Dai W, et al. Dual motion GAN for future-flow embedded video prediction[C]//Proceedings of the 2017 IEEE International Conference on Computer Vision, Venice, Oct 22-29, 2017. Washington: IEEE Computer Society, 2017: 1762-1770.
- [66] Esteban C, Hyland S L, Rtsch G. Realvalued (medical) time series generation with recurrent conditional GANs[J]. arXiv:1706.02633, 2017.
- [67] Hu W W, Tan Y. Generating adversarial malware examples for blackbox attacks based on GAN[J]. arXiv:1702.05983,

- 2017.
- [68] Xue Y, Xu T, Zhang H, et al. SegAN: adversarial network with multiscale L1 loss for medical image segmentation[J]. arXiv:1706.01805, 2017.
- [69] Yun H S, Rong R J, Sheng C Z, et al. Generative adversarial learning towards fast weakly supervised detection[C]// Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, Jun 18-22, 2018. Washington: IEEE Computer Society, 2018: 5764-5773.
- [70] Yi B S, Chao M, Xiao H W, et al. VITAL: visual tracking via adversarial learning[C]//Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, Jun 18-22, 2018. Washington: IEEE Computer Society, 2018: 8990-8999.
- [71] Xiao W, Cheng L L, Bin L, et al. SINT++: robust visual tracking via adversarial positive instance generation[C]// Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, Jun 18-22, 2018. Washington: IEEE Computer Society, 2018: 4864-4873.
- [72] Ghosh A, Bhattacharya B, Chowdhury S B R. SADGAN: synthetic autonomous driving using generative adversarial networks[J]. arXiv:1611.08788, 2016.
- [73] Jacob B, Kligys S, Chen B, et al. Quantization and training of neural networks for efficient integer-arithmetic-only inference[J]. arXiv:1712.05877, 2017.

附中文参考文献：

- [3] 吴进, 钱雪忠. 紧凑型深度卷积神经网络在图像识别中的应用[J]. 计算机科学与探索, 2019, 13(2): 275-284.
- [4] 都格草, 才让卓玛, 南措吉, 等. 基于神经网络的藏语语音合成[J]. 中文信息学报, 2019, 33(2): 75-80.
- [60] 桑亮, 高爽, 尹增山. 基于生成对抗网络的运动模糊图像复原[J]. 计算机工程与应用, 2019, 55(6): 173-177.
- [61] 时澄, 潘斌, 郭小明, 等. 生成式对抗网络在图像补全中的应用[J]. 计算机科学与探索, 2019, 13(8): 1402-1410.



LIANG Junjie was born in 1974. She received the Ph.D. degree from Huazhong University of Science and Technology in 2007. Now she is a professor at School of Computer Science and Information Engineering, Hubei University. Her research interests include multimedia database, high dimensional indexing, machine learning and deep learning.

梁俊杰(1974—),女,湖北武汉人,2007年于华中科技大学获得博士学位,现为湖北大学计算机与信息工程学院教授,主要研究领域为多媒体数据库、高维索引、机器学习、深度学习。在国内外重要期刊和会议上发表38篇学术论文。我国第一个数据库管理系统、第一套数据库管理系统国家标准的领衔研制者和编制者,国家“十一五”重大专项“核高基”课题大型通用数据库管理系统与套件研发及产业化项目的主要参与人员。



WEI Jianjing was born in 1994. He is an M.S. candidate at School of Computer Science and Information Engineering, Hubei University. His research interests include deep learning and computer vision.

韦舰晶(1994—),男,广西南宁人,湖北大学计算机与信息工程学院硕士研究生,主要研究领域为深度学习、计算机视觉。



JIANG Zhengfeng was born in 1979. He received the M.S. degree from Guangxi Normal University in 2004. Now he is a lecturer at School of Mathematics and Computer Science, Guangxi Normal University for Nationalities. His research interests include deep learning, machine learning, speech recognition and algorithm optimization.

蒋正锋(1979—),男,广西南宁人,2004年于广西师范大学获得硕士学位,现为广西民族师范学院数学与计算机科学学院讲师,主要研究领域为深度学习、机器学习、语音识别、算法优化。