

生成对抗网络 GAN 综述

程显毅^{1,2} 谢璐² 朱建新^{2,3} 胡彬² 施佳²

(硅湖职业技术学院 江苏 昆山 215300)¹ (南通先进通信技术研究院(南通大学) 江苏 南通 226019)²
(武汉理工大学信息工程学院 武汉 430010)³

摘要 人能够理解事物运动的方式,因此对事物未来发展的预测比机器准。不过,作为一种新的深度神经网络系统,GAN(Generative Adversarial Network)生成的数据非常逼真,连人也无法辨别数据是真实的还是生成的。从某种意义上讲,GAN 为指导人工智能系统完成复杂任务提供了一种全新的思路,让机器成为了一个专家。首先,讨论了 GAN 的基本模型和一些改进的 GAN 模型;然后,展示了 GAN 在超分辨率生成、由文本描述生成图像、艺术风格图像生成和短视频生成方面的应用成果;最后,探讨了 GAN 在理论、架构和应用方面所面临的问题和其未来的研究方向。

关键词 人工智能,深度学习,生成对抗网络,生成器,判别器

中图分类号 TP181 文献标识码 A DOI 10.11896/j.issn.1002-137X.2019.03.009

Review of Generative Adversarial Network

CHENG Xian-yi^{1,2} XIE Lu² ZHU Jian-xin^{2,3} HU Bin² SHI Quan²

(Silicon Lake College, Kunshan, Jiangsu 215300, China)¹

(Nantong Research Institute for Advanced Communication Technologies(Nantong University), Nantong, Jiangsu 226019, China)²

(School of Information Engineering, Wuhan University of Technology, Wuhan 430010, China)³

Abstract Humans can understand the way of movement, so they can predict the future development of things more accurately than machines. But GAN (Generative Adversarial Network) is a new neural Network system, its data are very lifelike, even people can't identify whether the data are real or generated. In a sense, GAN provides a brand new thought for guiding the artificial intelligence system to accomplish complex tasks, and makes the machine a specialist. In this paper, first of all, the basic model and some improvements model of GAN were discussed. Then, some application achievements of GAN were shown, such as the images generated by the super resolution, by a text description, by the artistic style and short video generated. Finally, some problems of theory, architecture, and application in the future research were discussed

Keywords Artificial intelligence, Deep learning, GAN, Generator, Discriminator

深度学习(Deep Learning, DL)在图像识别、图像分类、语音处理和自然语言处理等方面非常成功,但无法完成图像生成工作,即输入一只狗的图像时无法输出另外一只狗的图像^[1-2]。

最近很热门的生成对抗网络模型 GAN 是深度学习未来的发展方向。尽管早在 2014 年 Goodfellow 等就提出了 GAN^[3],但直到 2016 年研究者才发现 GAN 的巨大潜力。GAN 突破了之前限制深度学习发展的瓶颈^[4-5]。

图 1 比较了深度学习与 GAN 之间的差异。在图 1 中,三角形底边表示“图像”,与底边对应的顶点表示“特征”,自下而上为“输入→输出”的过程。因此,深度学习在训练时实现“图像→特征”,在测试时实现“特征→图像”。

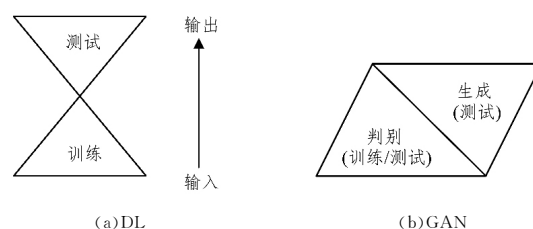


图 1 深度学习和 GAN 之间的对比

Fig. 1 Comparison between deep learning and GAN

GAN 由两个子网构成,一个是判别网,另一个是生成网,判别网既可用于训练,又可用于测试,但生成网只能用于测试,因此,GAN 测试可实现“图像→特征”“特征→图像”“图

到稿日期:2018-02-12 返修日期:2018-06-09 本文受国家自然科学基金项目(61771265,61340037),江苏省现代教育技术研究课题(2017-R-54131),南通大学-南通智能信息技术联合研究中心开放课题(KFKT2016B06)资助。

程显毅(1956—),男,教授,主要研究方向为机器学习、自然语言处理,E-mail:xycheng@ntu.edu.cn(通信作者);谢璐(1990—),女,硕士生,主要研究方向为深度学习;朱建新(1976—),男,博士生,副教授,主要研究方向为大数据技术;胡彬(1980—),男,博士,主要研究方向为图像处理;施佳(1973—),男,教授,主要研究方向为智能信息处理。

像→图像”的过程。图2说明了DL到GAN的演化^[6]。从图2可以看出,自编码器(encoder)和生成器(generator)构成了DL模型,生成器和判别器(discriminator)构成了GAN模型。其中,判别器用于替换DL求精重构误差的计算,因此GAN模型可以看作是对自编码器的改进^[7]。

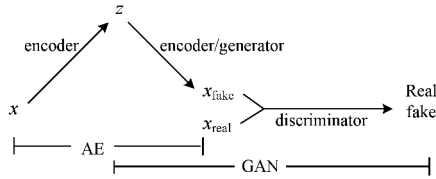


图2 深度学习到GAN的演化

Fig. 2 Evolution from deep learning to GAN

1 基本 GAN 模型

1.1 GAN 的诞生

受二人零和博弈的启发,2014年Goodfellow等在NIPS发表了生成对抗网络的开创性文章^[3],该文章的主要观点是:生成对抗网络背后的基本思想是学习训练样本的概率分布。

GAN让两个网络(生成网络G和判别网络D)相互竞争,G不断捕捉训练集里真实样本 x_{real} 的概率分布,然后通过加入随机噪声将其转变成赝品 x_{fake} 。D观察真实样本 x_{real} 和赝品 x_{fake} ,判断这个 x_{fake} 到底是不是 x_{real} 。其过程是,首先让D观察(机器学习)一些真实样 x_{real} ,当D对 x_{real} 有了一定的认知之后,G尝试用 x_{fake} 来愚弄D,让D相信 x_{fake} 是 x_{real} 。有时候G能够成功骗过D,但是随着D对 x_{real} 了解的加深(即学习的样本数据越来越多),G发现越来越难以愚弄D,因此G也在不断提升自己仿制赝品 x_{fake} 的能力。如此往复多次,不仅D能精通 x_{real} 的鉴别,G对 x_{real} 的伪造技术也大为提升。

1.2 GAN 模型结构

如图3所示,当G输出真实样本 x_{real} 时,期待D输出高概率(接近1);当D的输入是一些服从某一简单分布(例如高斯分布)的随机噪声 z 时,输出是与 x 逼真的生成样本 $G(z)$,再将其输入到判别器D,对于D来说,此时期望输出低概率(接近0),G要尽量欺骗D,使D输出高概率(把 $G(z)$ 误判为 x_{real})。对于D来说,遇到 x_{real} 则输出1,遇到 $x_{\text{fake}}=G(z)$ 则输出0。从而,G和D形成竞争与对抗,通过该过程,一旦G对训练样本的分布有了足够的了解,它就可以生成有类似特性的新样本。D想要实现它的目标,就会不断优化自己,防止被 x_{fake} 欺骗。反之,G也要优化自己,以生成非常逼真的 x_{fake} ,尽可能地让D难辨真伪。最后,G所生成的无论是 x_{fake} 还是 x_{real} ,在大多数时间里,D的正确判别概率是稳定的。

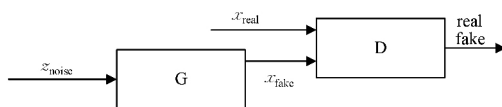


图3 GAN模型结构

Fig. 3 GAN model structure

若对GAN直接建模,则可以得到如下目标函数:

$$\min_G \max_D (D(x) - D(G(z)))$$

文献^[3]把上式目标函数修改为:

$$\min_G \max_D E_{h, \ell \sim p_{\text{Data}}(h, 1)} [\log D(h, \ell)] + E_{z \sim p_{\text{Noise}}(z), \ell \sim p_{\ell}(1)} [\log (1 - D(G(z, \ell), \ell))]$$

其中, $p_{\ell}(1)$ 是类别的先验分布,模型允许G的输出被条件变量1控制。

1.3 GAN 的特点

GAN具备很多优势:

- (1)能学习真实样本的分布,探索样本的真实结构;
- (2)具有更强大的预测能力;
- (3)样本的脆弱性在很多机器学习模型中都普遍存在,而GAN对生成样本非常鲁棒;

(4)通过GAN生成以假乱真的样本,缓解了小样本机器学习的困难;

(5)为指导人工智能系统完成复杂任务提供了一种全新的思路;

(6)与强化学习相比,对抗式学习更接近人类的学习机理;

(7)GAN与传统神经网络的一个重要区别是传统神经网络需要人工精心设计和建构一个损失函数,而GAN可以学习损失函数;

(8)GAN解决了先验概率难以确定的难题。

2 改进的 GAN 模型

GAN虽然具备很多优势,但却存在致命的弱点,即网络难以收敛。针对这一问题,出现了种类繁多的GAN变形模型。

2.1 CGAN 模型

条件生成式对抗网络(Conditional Adversarial Nets, CGAN)^[8]模型对G和D都增加了额外信息 C_{class} 作为条件,用于指导样本的生成。如果条件变量 C_{class} 是类别标签,那么CGAN是把无监督的GAN变成了有监督的GAN。CGAN的模型结构如图4所示。

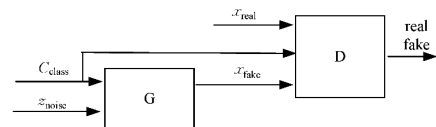


图4 CGAN结构

Fig. 4 CGAN structure

CGAN被证明非常有效,并广泛用于后续的相关研究工作中。

2.2 LAPGAN 模型

LAPGAN模型的主要操作是上采样和下采样,其优点是每次只考虑样本和生成样本之间的残差^[9]。从某种程度上讲,LAPGAN类似于残差网络,其学习过程如图5所示。

LAPGAN由一个带通图像集合构成,形成一个尺度空间,加上一个低频残差,具体步骤如下。

(1)生成高通图像:令 $I_0 = I$ 是原始图像, I_k 是第 k 次下采样的图像, k 是尺度变换的层数。

$$h_k = I_k - L_k(I)$$

(2)生成低频残差:从 I_{k+1} 开始循环(令 $I_{k+1} = 0$),然后给 G_k 层输入噪声向量 z_k ,生成 I_k 的残差图像 $\tilde{I}_k = G_k(z_k)$ 。

(3) 构造生成模型: 从右边的噪声样本 z_3 和生成模型 G_3 生成的 \tilde{I}_3 开始, 上采样 I_2 (空心三角形箭头) 和条件变量 (圆点箭头), 形成下一层生成模型 G_2 , 依次类推。

(4) 训练: 等概率地为判别器 D_0 选择一个真实样本或生

成样本; 把 L_0 和噪声 z 输入生成模型 G , 输出高通图像 (残差) $\tilde{h} = G(z_0, L_0)$ 。

D_0 的输入实现了 G 学习真实高频结构 (与 L_0 一致)。 D_1, D_2, D_3 的训练过程类似。

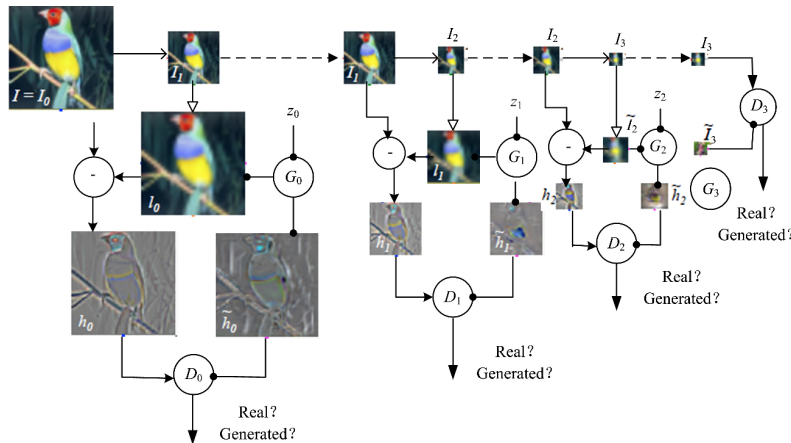


图 5 LAPGAN 模型的学习过程

Fig. 5 Learning process of LAPGAN model

通过尺度变换的方式减少了每一次 GAN 需要学习的参数, 增强了 GAN 的学习能力。值得注意的是, 图 5 的每一步都是独立训练得到的生成式卷积网络。利用多尺度变换, 构建一系列生成式模型。每个生成式模型处理特定尺度的图像结构。这一策略将生成对抗过程转变成一系列更加可控的步骤。

2.3 DCGAN 模型

卷积神经网络 CNN 在有监督学习的各项任务中均表现出色, 但在无监督学习领域却不尽人意。DCGAN (Deep Convolution GAN) 模型将 CNN 和 GAN 结合到一起^[10], GAN 模型在结构上需做如下变化。

(1) 将池化层用卷积层替代, 判别器用步幅卷积 (strided convolutions) 替代, 生成器用反卷积 (fractional-strided convolutions) 替代。

(2) D 和 G 都使用 BN (batchnorm) 策略解决初始化敏感的问题, 防止 G 把所有样本都视为同一类样本。

(3) 直接将 BN 应用到所有层会导致样本震荡和模型的不稳定, 在 G 输出层和 D 输入层不采用 BN 可以防止这一问题。

(4) 移除全连接的全局池化层增加了模型的稳定性, 但会影响收敛速度。

(5) G 中除了输出层外的所有层都使用了 ReLU, 输出层采用 tanh。D 的所有层使用 LeakyReLU。

DCGAN 的生成器结构如图 6 所示, 其中深色的圆表示输入, 浅色的圆表示隐藏层。

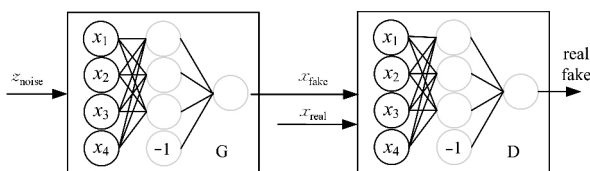


图 6 DCGAN 的生成器结构

Fig. 6 DCGAN generator structure

2.4 f-GAN 模型

文献[11]证明了 GAN 是在 f-散度取特定度量时的特殊情况。其中, f-散度是指概率分布的距离, 如 KL 散度、Pearson 散度等。具体来说, 作者将 GAN 的优化步骤进行分解, 将真实样本分布的估计问题转化为 f-散度的最小化问题, 即 f-GAN 模型。文献[11]利用 GAN 模型框架, 结合不同散度进行图像生成。其中, 在选择 KL 散度的度量方式时, 对比结果如图 7 所示, 从图中可以看出两者的效果相差并不大。



图 7 GAN 与 KL 散度图像生成对比

Fig. 7 Contrast to GAN and KL divergence images

2.5 EBGAN 模型

文献[12]从能量模型的角度对 GAN 进行了扩展, 提出了 EBGAN 模型。该模型把 D 看作一个能量函数, 这个能量函数在真实样本域附近区域中的能量值较小, 而在非真实样本区域拥有较高的能量值。因此, EBGAN 中给予 GAN 一种能量模型的解释, 即 G 是以生成能量最小的样本为目的, 而 D 则对这些生成的样本赋予较高的能量为目的。

EBGAN 模型的优点是可以使用更多、更宽泛的结构和损失函数来训练 GAN 结构。EBGAN 模型结构如图 8 所示。

在模型的稳定性方面, EBGAN 优于 GAN, 能生成更加清晰、逼真的图像^[13]。

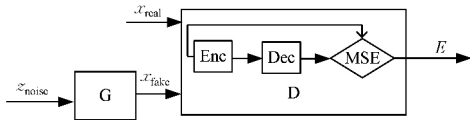


图8 EBGAN 模型结构

Fig. 8 EBGAN model structure

2.6 InfoGAN 模型

GAN 模型通过对抗学习,最终可以得到一个与真实样本分布一致的模型分布,此时虽然模型相当于已经学到了样本的语义特征,但输入的噪声信号 z 与数据的语义特征之间的对应关系并不清楚,比如 z 中的哪些维度对应于光照变化是不明显的,哪些维度对应于姿态变化是不明确的。而 InfoGAN 不仅能对这些对应关系建模,同时可以通过控制相应维度的变量来达到相应的变化,如姿态的变化或光照的变化^[14],其基本思想是把输入噪声 z 分成两个部分:不可压缩的噪声信号 z 和可解释的有隐含意义的 c 。InfoGAN 模型结构如图 9 所示。

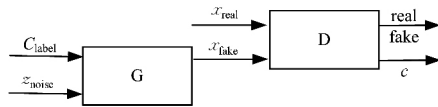


图9 InfoGAN 模型结构

Fig. 9 InfoGAN model structure

对于 MNIST 手写数字识别项目来说,隐藏代码 c 可以对应于图像光照、笔画粗细、字体倾斜度等,用 c_1, c_2, \dots, c_L 表示,而把剩下的不知道怎么描述的或者是不能明确描述的信息称为 z ,此时 G 的输出就从原来的 $G(z)$ 变成了 $G(z, c)$;在学习过程中,为了避免学到一些不重要的隐藏代码而忽略了重要的代码,文献^[14]对 GAN 目标函数增加了一个约束,即 c 和 $G(z, c)$ 之间的互信息 $I(c; G(z, c))$ 。因此,InfoGAN 的目标函数为:

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

InfoGAN 希望能学到比较重要的、有意义的隐藏代码 c ,在具体的优化过程中,引入变分分布来逼近真实分布,通过改变隐藏代码 c 中的维度来观察生成样本的变化。实验证明其确实学到了一些可解释的语义特征,如图像的角度和光照等,其效果如图 10 所示^[15]。

(a) Varying c_1 on InfoGAN
(Digit type)(b) Varying c_2 on regularGAN
(NO clear meaning)(c) Varying c_2 from -2 to 2 on
InfoGAN(Rotation)(d) Varying c_3 from -2 to 2 on
InfoGAN(Width)

图10 图像的角度或光照的特征学习

Fig. 10 Feature learning of image angle or light

图 10 使用了 3 个隐含编码, c_1 用 10 个离散数字进行编码,每个类别的概率都是 0.1, c_2, c_3 为连续编码,是 -2 到 2 的均匀分布。通过实验发现,如果使用 infoGAN 模型,表示的语义是数字的类别,如图 10 所示,分类的错误率为 5%,图 10(a)中第二行将 7 识别为 9。如果使用 reglarGAN 模型,生成的数字没有太多的语义,如图 10(c)所示。图 10(c)和图 10(d)都是使用 infoGAN 模型,从图 10(c)可以发现 c_2 表示生成数字的旋转的角度,小的 c_2 值表示数字向左偏,大的 c_2 值表示数字向右偏。从图 10(d)可以发现 c_3 表示生成数字的宽度, c_3 的值越大,生成的数字越宽,所以 infoGAN 模型具有表示语义的能力。

2.7 CatGAN 模型

在判别式的聚类方法中,一般是以某种距离作为度量准则,而文献^[16]提出的 CatGAN 模型则是采用样本熵来作为衡量准则构建 GAN。具体来说,对于真实的样本, CatGAN 希望 D 能具有较大的置信度将其划分为真实样本,但将生成样本划分到哪一类是不确定的。因此, G 的目标即为生成那些“将其划分到某一类别中去”的确信度较高的样本,尝试骗过 D 。将该置信度目标与 GAN 的真伪鉴别的优化目标结合,得到 CatGAN 的目标。CatGAN 模型结构如图 11 所示。

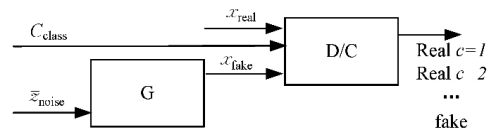


图11 CatGAN 模型结构

Fig. 11 CatGAN model structure

用 CatGAN 模型进行目标识别与分类时,性能有了比较明显的提升。从实验效果(见图 12)可以看到,对于典型的环形样本, CatGAN 可以较好地找到两者的分类面,实现无监督聚类功能^[17]。

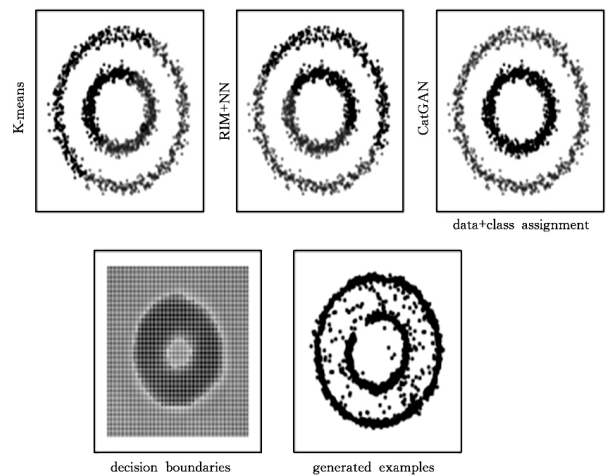


图12 环形样本分类的对比

Fig. 12 Comparison of ring sample classification

图 12 从左到右依次为:K-means 分类结果、RIM+NN 分类结果、CatGAN 分类结果、判别的边界(decision boundaries)、生成的例子(generated examples)。

3 GAN 相关问题的研究

3.1 假设太过自由的问题

与其他生成式模型相比, GAN 模型的最大优势是不再要求一个假设的样本分布, 而是通过采样达到逼近真实样本的目的。其缺点是太过自由, 如果图像较大, GAN 的收敛便不可控^[18]。为了解决这个问题, 一种很自然的想法是给 GAN 添加一些约束, CGAN 便应用而生。另一种思路源于 DeepMind 的 DRAW 项目^[19], 即不让 GAN 一次完成全部任务, 而是分多次生成一张完整的图片。DRAW 使用了一种有序的 VAE 模型, 让机器一点点“写”出一个个数字。LAPGAN^[7]模型就采用了这种思想, 值得关注。

3.2 空间关系表达问题

图 13 显示了 Radford 等^[20]在 MNIST 上使用 DCGAN 生成模型生成的手写数字的效果。



图 13 生成模型生成的手写数字

Fig. 13 Handwritten numbers generated by generate model

图 13 中, 除了个别数字之外, 大多数的数字生成得较为逼真。

图 14 给出 G 和 D 在训练过程中的损失函数^[21], 其中, 黑色是 G 的损失函数, 灰色是 D 的损失函数。可以看出, 两个网络的损失函数都存在一定程度的抖动, 可以看作是对抗过程中的此消彼长。

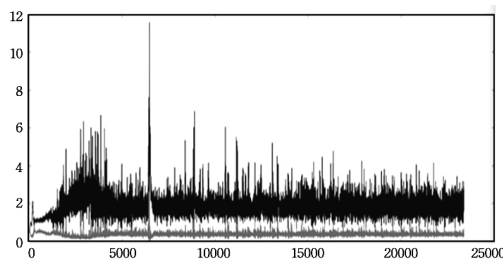


图 14 G 和 D 在训练过程中的损失

Fig. 14 Loss of two networks during training

DCGAN 能够把相似的模式放在一起。例如, 在人脸图像数据集中, G 不理解什么是微笑, 但是它能发现人类微笑图片的相似性, 并将它们分为一组^[22]。

简单的算术运算可以揭示出空间中的丰富关系, 如图 15 所示。

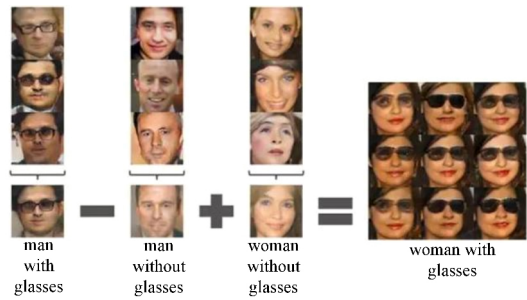


图 15 图像合成^[23]

Fig. 15 Image synthesis^[23]

3.3 训练过程的稳定性问题

虽然 GAN 模型引起了极大的关注, 但是由于不需要先验知识, 过于自由的训练方式使得 GAN 对初始参数极其敏感, 造成训练过程极其不稳定, 并存在严重的模式忽略 (missing modes) 问题。模式忽略是指训练样本中某一类模式的样本如果占比很小, 则很容易在训练时被忽略掉, 使得 GAN 只能学会大样本模型的表达, 从而使生成的样本集中在某几个模式上缺乏多样性。当模式忽略出现时, 更容易加剧 GAN 训练的不稳定性。

为此, 文献^[24]通过引入模式正则项来对 GAN 的训练过程进行控制。具体来讲, G 直接从噪声向量 z 映射到样本空间, 即 $z = \text{encoder}(X)$, 然后输出 $G(\text{encoder}(X))$ 。这样做的优点是, 重构过程会增加额外的信息, 使得 G 生成的图片不再容易被 D 识别。另一方面, $\text{encoder}(X)$ 保证了 X 与映射后的样本空间的对应关系, 从而保证了 G 能覆盖所有样本空间的样本模式, 从理论上避免了模式忽略问题。

在此基础上, 多样扩散 (Manifold-Diffusion GAN, MDGAN) 模型将重构正则项分成两步训练。第一步的多样化就是用 X 训练 G, 即 $G(\text{Enc}(X))$, 减小 X 和 $G(\text{Enc}(X))$ 两者之间的差别; 第二步的扩散是对 $G(\text{Enc}(X))$ 和 $G(z)$ 再做拉近, 可以将该过程形象地比喻为: 先把两个分布的“形状”调整好, 再把两个分布的距离拉近。

重构正则项分成两步训练的思想提高了 GAN 训练的稳定性, 而且增加了模式的多样性, 同时使得生成的图片的质量不受任何影响。

受文献^[24]最小批判别 (minibatch discrimination) 训练方法的启发, 在每次 D 进行判别时, 不是基于单个样本, 而是基于一批样本。文献^[25]解决稳定性的基本思路是不让 GAN 每次用一个 D 来判别一个生成样本, 而是同时训练一组判别网络 D_1, D_2, \dots, D_k 和一组生成网络 G_1, G_2, \dots, G_k , 然后让 D_i 交替着去判别其他 G_j 的生成样本。也就是 D_1, D_2, \dots, D_k 一起训练, D_1 可能只见过模式 A, D_2 见过模式 B, 那就用 D_2 来判别只会生成模式 B 的生成器 G_{i1} , 用 D_1 来判别只会生成模式 A 的生成器 G_{i2} , 这样判别就更加鲁棒, 更容易让学习出来的生成样本的分布逼近真实样本的分布。可惜的是, 文献^[19]中更多的实验是停留在“玩具”层面。

文献^[26]给出的代理损失函数 (surrogate loss) 不是通过改变模型结构来引入的, 而是引入了一种新的指标。如果过

去进行梯度下降时只改变梯度,其他模型参数保持不变,那么代理损失函数就是基于一阶梯度改变其他模型参数,用改变后的模型参数再去改变二阶梯度,依次类推。这种做法并不常见,因为在大部分神经网络训练中,都只会用到一阶导数信息。然而,代理损失函数可以利用二阶导数或者高阶导数不断地展现精确解。为了避免告诫导数计算困难的问题,代理损失函数的策略是把告诫导数作为一种中间态。代理损失函数的表达式如下:

$$\frac{df_K(\theta_G, \theta_D)}{d\theta_G} = \frac{\partial f(\theta_G, \theta_D^K(\theta_G, \theta_D))}{\partial \theta_G} + \frac{\partial f(\theta_G, \theta_D^K(\theta_G, \theta_D))}{\partial \theta_D^K(\theta_G, \theta_D)} \frac{d\theta_D^K(\theta_G, \theta_D)}{d\theta_G}$$

代理损失函数把损失函数的表达式拆成两项,其中第二项就是当 G 向不好的方面改变时,此项就会给出较大的惩罚。因此,基于代理损失函数的 G 在生成样本时就要考虑到 G 改变时 D 会发生什么变化,比原始 GAN 的 G 考虑的信息更多。

引入代理损失函数的贡献就是,强调了高阶导数信息在神经网络训练中的重要性。受此启发,代理损失函数也能与其他模型和其他任务带来一些新的解决方法。

文献[27]的思路与文献[26]存在一定的相似性,讨论了特征匹配的延伸和泛化。在训练 GAN 时使用先验知识有助于提高训练的稳定性。用 D 学到的特征来做辅助,让 G 生成尽可能匹配 D 的一些统计特征。该过程容易理解,即既然 D 能通过一些特征判断 G 的生成样本是假的,那么 G 就可以把 D 的特征直接复制过来而变成自己的一种能力,让 D 不能再利用这个特征来判别出 G 是假的。

3.4 精准预测的研究

文献[28]提出了一种对抗梯度差损失预测器(AGDL),它通过几帧桌球游戏视频构造模型,预测球被击中后下一帧球的运动情况(见图16)。

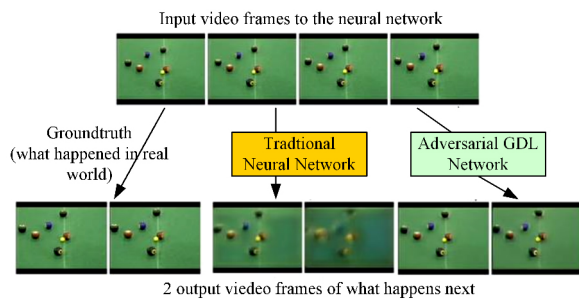


图16 台球预测

Fig. 16 Billiard prediction

4 应用

4.1 超分辨率图像的生成

文献[29]基于相似性感知方法^[30]提出了一种新的损失函数与 GAN 本身的损失函数相结合的方法——SRGAN,实现了超分辨率图像的生成。SRGAN 的基本原理是:

(1)用生成图像的整体方差作为图像空间的损失约束项,

以保证图像的平滑性;

(2)SRGAN 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。其与 GAN 的主要区别在于加入了规则化的特征图差异损失,而不是直接累加求和。

$$l_{VGG/i,j}^{SR} = \frac{1}{W_{ij}H_{ij}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\phi_{i,j}(I^{HR})_{x,y} - \phi_{i,j}(I^{LR})_{x,y})^2$$

其中, $\phi_{i,j}$ 表示在 VGG-19 网络中第 i 个池化层前面的第 j 个卷积层, W_{ij} 和 H_{ij} 为特征图的宽和高。

(3)将对抗损失、图像平滑项、特征图差异这 3 个损失项作为 SRGAN 模型的损失函数,能够生成比其他方法效果更好得多的超分辨率图像,其对比如图 17 所示。



图17 超分辨率图像生成

Fig. 17 Super-resolution image generation

4.2 文本描述生成图像

基于噪声数据生成图像的任务是“噪声→图”的问题;而给定属性生成图像的任务是“图→图”的问题;文献[31]则另辟蹊径,实现“文本→图”,这对于推动 GAN 以及 G 在实际中的应用具有一定的积极影响。

文献[32]将文本“This white and yellow flower have thin white petals and round yellow stamen”编码后的特征与随机噪声信息作为 G 的输入,G 的输出是图像。从图 18 中看出,大部分 G 生成出来的图像与文本意义基本相符。



图18 文本描述生成图像

Fig. 18 Generated images from text description

4.3 视频帧预测

文献[33]给出了视频处理的一个应用,可以同时生成和预测下一视频帧。

为了生成视频帧,该模型的 G 将动态前景部分和静态背景部分分开建模和生成,构建双向生成器,然后将生成的前景和背景进行组合后作为 G 生成的视频;D 的主要任务是识别出视频帧间的行为,从而指导 G 生成视频。

G 和 D 互相竞争,生成了越来越真实的视频。当让人从两种视频中挑选真实的视频时,人挑选了机器合成而非真实视频的比例只有 20%。

4.4 艺术风格的迁移

若一个卷积网络足够深,则其可以在高层表示图像的高级抽象特征,如果把这些高级抽象特征应用到另外一个图上,那么另外一个图也可以继承到这些高级特征。文献[34]用格拉姆(gram)矩阵来描述艺术特征的转化。

艺术风格迁移的最初形式计算的是原图和艺术图的每个像素的损失函数,格拉姆矩阵很大,需要通过一个逻辑回归更新每一层的参数,耗时较长,一般一幅图的转换需要几十秒。文献[27]使用相似性感知损失函数,其创新点是把图输入到迁移网络,通过相似性感知损失函数更新迁移网络,并最小化这个损失,使得迁移网络学习到损失网络定义的艺术风格并输出转化好的艺术图像。

这种相似性感知损失函数需要输入一系列预先训练好的艺术图,实验效果如图 19 所示。

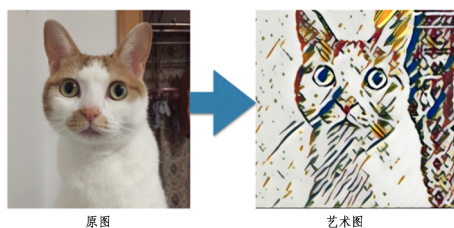


图 19 艺术风格图像的生成

Fig. 19 Generation of art style image

5 未来的研究方向

文献[3]只是通过实验验证了 GAN 应用的有效性,而对其中的许多细节并没有进行探讨,比如输入信号只是随机噪声,G 和 D 的建模都是基于最简单的 MLP。因此,GAN 的未来研究方向应该集中在以下几个方面。

(1)用 CGAN 模型替代 GAN。关键是对输入条件的研究,即什么样的场景适合什么条件。

(2)改进输入的随机噪声 z 。可根据真实样本 x 用其他网络学习一个 z 。

(3)基于先验知识的相关建模研究工作出现较晚,直到 2016 年才出现^[35],有许多问题值得探讨。

(4)半监督学习问题。纯无监督的学习存在收敛慢、稳定性差的问题,现有的许多 GAN 研究都表明通过加入少量类别标签,引入有标签数据的类别损失度量,不仅在功能上能实现半监督学习,而且也有助于 GAN 的稳定训练^[36,38]。

(5)提升 GAN 的训练效率。由于训练的稳定性问题还没有得到很好的解决,因此对训练效率的研究目前并不深入。相比于其他生成式模型,GAN 的速度也不算很慢^[39]。

(6)将 GAN 与其他模型结合,综合利用各模型的优点来完成样本生成任务是一条很好的思路。这方面已出现了许多成果,如多个 GAN 进行组合的模型^[40]、GAN 与 CNN 组合的模型^[41]、GAN 与 RNN 的组合模型^[42]等。

(7)是否存在更好的学习模型取代深度网络?最近,南京大学周志华教授提出的“gcForest 模型(深度森林)”^[43],引发了很多研究者的关注。实验表明,该模型的性能与深度神经

网络之间有很强的竞争力;另外,在几乎完全一样的超参数设置下,gcForest 具有如下优点:

1)gcForest 天然适用于并行的部署 gcForest,因此训练过程效率高且可扩展,在处理不同领域的不同数据时,也能达到极佳的性能。

2)训练成本低,在一台 PC 上的训练时间与在 GPU 上运行的深度神经网络差不多。

3)gcForest 在小样本训练数据的情况下也能照常运转,而深度神经网络需要大量的训练数据。

结束语 GAN 是深度学习领域中比较重要的一个模型,为无监督学习模型提供了一个强大的计算框架,也是人工智能研究的一个重要工具。GAN 的一个很重要的特性就是能够像人类一样理解周围复杂的世界。在 GAN 这个方向上继续探索,有可能成功打造出比识别层次更高的、能够理解世界的机器学习模型。

参 考 文 献

- [1] GOODFELLOW I, BENGIO Y, COURVILLE A. Deep Learning [M]. Cambridge, UK: MIT Press, 2016: 23-34.
- [2] LIU Q, ZHAI J H, ZHANG Z Z, et al. A Survey on Deep Reinforcement Learning[J]. Chinese Journal of Computers, 2017, 40(1): 1-28. (in Chinese)
- [3] 刘全, 翟建伟, 章宗长, 等. 深度强化学习综述[J]. 计算机学报, 2017, 40(1): 1-28.
- [4] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]// Proceedings of the 2014 Conference on Advances in Neural Information Processing Systems 27. Montreal, Canada: Curran Associates, 2014: 2672-2680.
- [5] YU L T, ZHANG W N, WANG J, et al. SeqGAN: sequence generative adversarial nets with policy gradient[J/OL]. <https://arxiv.org/abs/1609.05473>.
- [6] SHAKIR M, LAKSHMINARAYANAN B. Learning in Implicit Generative Models[J/OL]. <https://openreview.net/pdf?id=B16Jem9xe>.
- [7] CHENG C. Interpretation of the GAN and its progress in 2016 [EB/OL]. <https://zhuanlan.zhihu.com/p/25000523?refer=dlclass>.
- [8] HU W W, TAN Y. Generating adversarial malware examples for black-box attacks based on GAN[J/OL]. <https://openreview.net/pdf?id=7xes>.
- [9] MIRZA M, OSINDERO S. Conditional Generative Adversarial Nets[J]. Computer Science, 2014, 27(8): 2672-2680.
- [10] DENTON E L, CHINTALA S, FERGUS R. Deep Generative Image Models using a Laplacian Pyramid of Adversarial Networks[C]// Advances in Neural Information Processing Systems. 2015: 1486-1494.
- [11] RAVANBAKHS S, LANUSSE F, MANDELBAUM R, et al. Enabling Dark Energy Science with Deep Generative Models of Galaxy Images[J/OL]. <https://arxiv.org/abs/1609.05796>.
- [12] SEBASTIAN N, CSEKE B, TOMIOKA R. f-GAN: Training Generative Neural Samplers using Variational Divergence Mini-

- mization[J/OL]. <https://arxiv.org/abs/1606.00709>.
- [12] ZHAO J B, MICHAEL M, YANN L C. Energy-based Generative Adversarial Network[J/OL]. <https://arxiv.org/abs/1609.03126>.
- [13] 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 (CVPR). Las Vegas, NV, USA: IEEE, 2016: 770-778.
- [14] CHEN X, DUAN Y, HOUTHOOFT R, et al. InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets[J/OL]. <https://arxiv.org/abs/1606.03657>.
- [15] GANIN Y, USTINOVA E, AJAKAN H, et al. Domain adversarial training of neural networks[J]. Journal of Machine Learning Research, 2016, 17(59): 1-35.
- [16] TOBIAS J. Unsupervised and semi-supervised learning with categorical generative adversarial networks[C]//ICLR-2016, Springer, 2016: 876-884.
- [17] CHEN W Z, WANG H, LI Y Y et al. Synthesizing training images for boosting human 3D pose estimation[C]//Proceedings of the 2016 Fourth International Conference on 3D Vision (3DV). Stanford, CA, USA: IEEE, 2016: 479-488.
- [18] PROBST M. Generative Adversarial Networks in Estimation of Distribution Algorithms for Combinatorial Optimization [J/OL]. <https://arxiv.org/abs/1509.09235>.
- [19] GREGOR K, DANIHELKA I, GRAVES A, et al. DRAW: A recurrent neural network for image generation[J/OL]. <https://arxiv.org/abs/1502.04623>.
- [20] RADFORD A, METZ L, CHINTALA S. Unsupervised representation learning with deep convolutional generative adversarial networks[J/OL]. <https://arxiv.org/abs/1511.06434>.
- [21] AUGUSTUS O, OLAH C, SHLENS J. Conditional Image Synthesis With Auxiliary Classifier GANs[J/OL]. <https://arxiv.org/abs/1610.09585>.
- [22] HANOCK K, ZHANG B T. Generating Images Part by Part with Composite Generative Adversarial Networks[J/OL]. <https://arxiv.org/abs/1607.05387>.
- [23] KURAKIN A, GOODFELLOW I, BENGIO S. Adversarial examples in the physical world[J/OL]. <https://arxiv.org/abs/1607.02533>.
- [24] ANTONIA C, BHARATH A A. Task Specific Adversarial Cost Function [J/OL]. [2017-01-17]. <http://Creswell.com/caa?arXiv:1609.08661>.
- [25] CHE T, LI Y, JACOB A P, et al. Mode Regularized Generative Adversarial Networks[C/OL]. [2017-01-30]. <https://openreview.net/pdf?id=HJKkY35le>.
- [26] IM D J, MA H, KIM C D, et al. Generative Adversarial Parallelization[C/OL]. <https://openreview.net/pdf?id=Sk8J83oe>.
- [27] METZ L, POOLE B, PFAU D, et al. Unrolled Generative Adversarial Networks[C/OL]. <https://openreview.net/pdf?id=BydrOIcle>.
- [28] WARDE-FARLEY D, BENGIO Y. Improving Generative Adversarial Networks with Denoising Feature Matching[C/OL]. <https://openreview.net/pdf?id=S1X7nhsxl>.
- [29] CHRISTIAN L. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network[J/OL]. <https://arxiv.org/abs/1609.04802>.
- [30] ALEXEY D, BROX T. Generating images with perceptual similarity metrics based on deep networks[J/OL]. <https://arxiv.org/abs/1602.02644>.
- [31] REED S, AKATA Z, YAN X, et al. Generative adversarial text to image synthesis[C]//International Conference on International Conference on Machine Learning. JMLR.org, 2016: 1060-1069.
- [32] LARSEN A B L, SØNDERBY S K, WINTHER O. Autoencoding beyond pixels using a learned similarity metric[J/OL]. [2015-11-02]. <https://arxiv.org/abs/1512.09300>.
- [33] VONDRICK C, PIRSIYAVASH H, TORRALBA A. Generating Videos with Scene Dynamics[C]//NIPS-2016. Stanford, CA: IEEE, 2016: 562-570.
- [34] SPRINGENBERG J T. Unsupervised and Semi-supervised Learning with Categorical Generative Adversarial Networks[J/OL]. <https://arxiv.org/abs/1511.06390>.
- [35] LEON A. Gatys, Alexander S. Ecker, Matthias Bethge. A Neural Algorithm of Artistic Style [J/OL]. <https://arxiv.org/abs/1508.06576>.
- [36] ZHU H, LI Q M, LI D Q. Facial Multi-landmarks Localization Based on Single Convolution Neural Network[J]. Computer Science, 2018, 45(4): 273-279. (in Chinese)
朱虹, 李千目, 李德强. 基于单个卷积神经网络的面部多特征点定位[J]. 计算机科学, 2018, 45(4): 273-279.
- [37] REN J, HU X F, LI N. Transfer Prediction Learning Based on Hybrid of SDA and SVR[J]. Computer Science, 2018, 45(1): 281-286. (in Chinese)
任俊, 胡晓峰, 李宁. 基于 SDA 与 SVR 混合模型的迁移学习预测算法[J]. 计算机科学, 2018, 45(1): 281-286.
- [38] WANG K F, GOU C, DUAN Y J, et al. Generative Adversarial Networks: The State of the Art and Beyond[J]. Acta Automatica Sinica, 2017, 43(3): 321-333. (in Chinese)
王坤峰, 苟超, 段艳杰, 等. 生成式对抗网络 GAN 的研究进展与展望[J]. 自动化学报, 2017, 43(3): 321-333.
- [39] ODENA A. Semi-Supervised Learning with Generative Adversarial Networks[J/OL]. <https://arxiv.org/abs/1508.06576>.
- [40] WANG X, GUPTA A. Generative Image Modeling using Style and Structure Adversarial Networks [J/OL]. <https://arxiv.org/abs/1603.05631>.
- [41] DENTON E L, CHINTALA S, FERGUS R. Deep Generative Image Models using a Laplacian Pyramid of Adversarial Networks[C]//Advances in Neural Information Processing Systems. 2015: 1486-1494.
- [42] EDWARDS H, STORKEY A. Censoring Representations with an Adversary [J/OL]. [2015-01-26]. <https://arxiv.org/abs/1511.05897>.
- [43] ZHOU Z H, FENG J. Deep Forest: Towards An Alternative to Deep Neural Networks [J/OL]. <https://arxiv.org/abs/1702.08835>.