

生成式对抗网络研究与应用进展

柴梦婷, 朱远平

(天津师范大学 计算机与信息工程学院, 天津 300387)

摘 要: 基于零和博弈思想的生成式对抗网络(GAN)可通过无监督学习获得数据的分布,并生成较逼真的数据。基于 GAN 的基础概念及理论框架,研究各类 GAN 模型及其在特定领域的应用情况,从数据相似性度量、模型框架、训练方法 3 个方面进行分析,对 GAN 改进与扩展的相关研究成果进行总结,并从图像合成、风格迁移等应用领域展开讨论,归纳出 GAN 的优势与不足,同时对其应用前景进行展望。分析结果表明,GAN 的学习能力与可塑性强,改进潜力大,应用范围广,但其发展面临的挑战是训练过程不稳定,且缺乏生成数据质量的客观评价标准。

关键词: 生成式对抗网络; 生成式模型; 对抗学习; 深度学习; 人工智能

中文引用格式: 柴梦婷, 朱远平. 生成式对抗网络模型研究与应用进展[J]. 计算机工程, 2019, 45(9): 222-234.

英文引用格式: CHAI Mengting, ZHU Yuanping. Research and application progress of generative adversarial networks[J]. Computer Engineering, 2019, 45(9): 222-234.

Research and Application Progress of Generative Adversarial Networks

CHAI Mengting, ZHU Yuanping

(College of Computer and Information Engineering, Tianjin Normal University, Tianjin 300387, China)

[Abstract] The Generative Adversarial Networks(GAN) based on the zero-sum game idea can obtain the distribution of data through unsupervised learning and generate more realistic data. Based on the basic concepts and theoretical framework of the generated confrontation network, the GAN models and the application results in specific fields are studied, and the data similarity measure, model framework and training method are summarized. The research and related research results of the improvement and expansion are analyzed, and the practical application fields such as image synthesis and style migration are discussed. The advantages and disadvantages of GAN are summarized, and the application prospects are prospected. Analysis results show that the GAN has strong learning ability and plasticity, great potential for improvement and wide application range. However, its development challenges are unstable training process and lack of objective evaluation criteria for generating data quality.

[Key words] Generative Adversarial Networks (GAN); generative model; adversarial learning; deep learning; artificial intelligence

DOI:10.19678/j.issn.1000-3428.0051964

0 概述

深度学习^[1-3]作为机器学习的一个分支,是一种以人工神经网络为架构,对大规模数据进行表征学习的算法。自 2012 年以来,深度神经网络(Deep Neural Networks, DNN)在语音识别^[4]和图像识别^[5]领域被广泛应用,推动了人工智能技术的发展。深度神经网络模型及深度学习模型的提出,也推动了深度学习的全面发展。如文献[6]提出一种生成式对抗网络(Generative Adversarial Networks, GAN)模型,其包括生成式和判别式 2 种模型。生成式模型

是指可以生成数据分布的模型,主要有朴素贝叶斯、K 近邻(KNN)以及混合高斯模型等。生成式模型的学习和计算过程复杂,且一个模型只对应一类数据。而判别式模型由于结构简单灵活,适用于多类别识别。GAN 使得生成式模型受到研究者的关注,其可以被用来生成图像、文字和语音等样本,且效果逼真。

由于 GAN 可以生成随机样本,并通过模型训练和学习得到与真实分布相似的数据,受到学者的广泛关注。同时,研究表明,GAN 的思想与 VAE (Variational Auto-Encoder)^[7]相似,但实际上两者在

基金项目: 国家自然科学基金(61602345, 61703306); 天津市科技计划项目(14RCFGX00847)。

作者简介: 柴梦婷(1992—),女,硕士研究生,主研方向为图像处理、模式识别;朱远平,教授、博士。

收稿日期: 2018-06-29 **修回日期:** 2018-09-04 **E-mail:** zhuyuanping@tjnu.edu.cn

数据的相似性度量方式和学习方式上有较大的不同,VAE模型性能直接依赖预先假设的损失函数和近似分布,预先定义需要一定的经验作基础,因此最终生成的概率分布会受到已定义参数的影响。而GAN的训练过程是无监督学习,其原则上可以逼近任意概率分布,且不需要预先定义参数。GAN在许多领域都有发展,获得较多优秀的研究成果。其中,图像处理仍是GAN研究和应用较广泛的领域之一^[8]。GAN使生成式模型成为人工智能技术的热门研究方向。

本文介绍原始GAN模型框架及原理,归纳和分析改进GAN模型,在此基础上,给出GAN的优势与不足,并对GAN未来发展方向进行展望。

1 生成式对抗网络

GAN由2个网络构成:生成器和判别器。这2个网络在同一时间进行训练,并在极大极小游戏中相互博弈。在训练过程中,生成器捕获真实样本的数据分布,从而将样本数据送入判别器,判别器则估计该样本是来自于训练数据还是生成数据,以保证不被生成器判断错误。最终两者会达到一个平衡,即生成器生成以假乱真的数据样本,判别器输出则逼近一个固定的概率值。

1.1 生成式对抗网络模型基本框架

图1所示为生成式对抗网络模型基本框架,该框架由2个模型构成:生成器 G 和判别器 D ,这2个模型本质都是函数,在GAN定义中通常为多层感知器(Multi-Layer Perception, MLP)。GAN采用神经网络模型,可以看出,首先将一个100维的随机噪声 z 输入生成器,在构成生成器的多层感知器或神经网络中该噪声映射到一个新的数据分布,得到 $G(z)$,将真实数据分布 x 与 $G(z)$ 共同输入到判别器中,判别器对于输入的2个数据分别做出判断,输出一个概率值,真实数据的概率值接近1,生成数据的概率值接近0,这说明此时生成器生成的数据置信度不高,判别器会将生成数据中需要调整的参数反馈给生成器,生成器在接收到调整的梯度信号后进行调整重新生成。如此循环,当判别器无法区分真实数据和生成数据时,就可以认为生成器此时达到最优。这个不断训练的过程,就是生成器 G 与判别器 D 的极大极小博弈过程, D 和 G 的性能在迭代中不断提高,直到最终 $D(G(z))$ 与 $D(x)$ 一致,此时 G 和 D 均已达到最优。

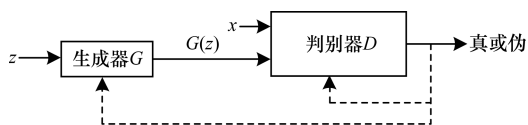


图1 生成式对抗网络模型基本框架

原始GAN是基于无监督学习的生成模型,其输入服从随机噪声分布,期望分布则服从近似但不完全等同于真实数据样本的分布,保证了生成样本的多样性。

1.2 理论公式

文献[6]提出极大极小的相互博弈公式如下:

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

由式(1)可以看出,该公式在 $D(x)$ 和 $D(G(z))$ 的基础上加入求对数和求期望值的操作。对数求解是统计学中常用的方法,目的是降低不可预期的干扰噪声对数据分布的影响和缓解数据分布偏差问题。期望值的计算是使最终生成的数据分布 $p_g(G(z))$ 与真实数据分布 $p_{\text{data}}(x)$ 一致,即希望生成数据尽量地拟合真实数据但不是生成真实数据本身,这样才能保证 G 产生的数据既与真实数据有一定相似性,同时又不同于真实数据。

在式(1)中,右侧第1项是真实数据分布 $p_{\text{data}}(x)$ 的数据通过判别器的熵,判别器的目的是将其最大化为1;第2项是来自随机噪声输入 $P_z(z)$ 的数据通过生成器的熵。生成器生成的样本通过判别器判定其真伪,判别器则尝试将其最大化为0。总体而言,判别器的目标是最大化函数 $V(D, G)$,而生成器的任务完全相反,其试图最小化函数 $V(D, G)$,使真实数据和生成数据之间的差距减小。

关于如何同时优化 D 和 G ,文献[6]只考虑 D 达到最优,使 G 生成的数据过拟合,为在迭代过程中每一次迭代的 D 都可以给出当前状态下的最优值,提出在更新多次 D 后更新一次 G ,可以保证 D 优先 G 一步达到当前最优,如此循环迭代直到两者共同达到最优,即 $p_g = p_{\text{data}}$,对于最终结果的判定,给出一个当 G 固定时的最优判别器公式,如式(2)所示。

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

当 $p_g = p_{\text{data}}$ 时, $D_G^*(x) = \frac{1}{2}$,表示此时 D 和 G 达到最优。

1.3 GAN的优势与不足

在文献[6]提出的原始生成式对抗网络中,为使GAN具有无限建模能力,其输入是一个仅为100维的随机噪声,不需要附加条件或限制,生成结果可以是任何数据分布,即可以拟合所有分布。GAN的模型框架与算法较简单,将生成器与判别器的对抗作为训练准则,使用反向传播算法(Back Propagation, BP)进行训练,在训练过程中不需要使用效率较低的马尔科夫链方法和各种近似推理过程,这些优点可降低生成式模型训练的难度,并提高训练的效率。

GAN的建模能力存在诸多问题,由于对数据没有任何假设,训练过程不可控,生成结果过于自由,

易出现训练过程不稳定、无法收敛和模型崩溃的问题,导致生成的样本过拟合。同时 GAN 对于损失函数的设定也存在问题,由于生成样本与真实数据本质上存在较大的维度差,使得损失函数难以进行计算,进而导致生成器无法调整,使最终的生成结果较难逼近真实样本。

2 GAN 的扩展与改进

2.1 基于数据相似性度量改进

2.1.1 WGAN 模型

文献[9]利用数据相似性度量对 GAN 进行改进,提出 WGAN(Wasserstein GAN)模型,将原始 GAN 中概率分布的距离度量公式替换为 Wasserstein 距离,并对 GAN 算法进行部分调整,优化了原始 GAN 训练过程不稳定(需要小心平衡判别器和生成器的训练程度)、训练后期生成器梯度消失、模型崩溃的问题。

根据文献[9]的 GAN 计算公式,有:

$$C(G) = -\lg 4 + KL\left(p_{\text{data}} \parallel \frac{p_{\text{data}} + p_g}{2}\right) + KL\left(p_g \parallel \frac{p_{\text{data}} + p_g}{2}\right) \quad (3)$$

$$JS(p_{\text{data}} \parallel p_g) = \frac{1}{2}KL\left(p_{\text{data}} \parallel \frac{p_{\text{data}} + p_g}{2}\right) + \frac{1}{2}KL\left(p_g \parallel \frac{p_{\text{data}} + p_g}{2}\right) \quad (4)$$

可以得出:

$$C(G) = -\lg 4 + 2 \cdot JS(p_{\text{data}} \parallel p_g) \quad (5)$$

根据式(5),GAN 模型要计算 $JS^{[10]}$ 散度,前提是 p_{data} 和 p_g 这 2 个分布有重叠。但低维与高维之间完全没有重叠或有微小到可忽略的重叠,这种可能性非常大。因此,生成器的损失为一个常量 $(-\lg 4)$ 而非线性变化,无法继续做出调整,也就无法使生成的样本不断逼近真实样本。因此,使用式(6)的 Wasserstein 距离代替 JS 散度。

$$W(P_r, P_g) = \inf_{Y \in \Pi(P_r, P_g)} E_{(x,y) \sim Y} [\|x - y\|] \quad (6)$$

Wasserstein 距离相比 KL 散度、JS 散度的优越性在于即使 2 个分布没有重叠,Wasserstein 距离仍然能够反映其远近,可同时优化 GAN 原始模型训练过程不稳定和进程指标不恰当的问题。

为更直观验证 WGAN 生成逼真图像的能力,本文使用与文献[9]中相同的数据集进行训练。图 2 给出 WGAN 训练 LSUN 数据集中的 bedroom 数据迭代 10 000 次和 30 000 次的实验结果。为更直接观察样本生成的过程,选取 3 张生成样本在迭代 10 000 次(第 1 行)和迭代 30 000 次(第 2 行)时的效果对比,如图 2(c)所示。可以看出,迭代 30 000 次生成结果较逼真,但放大之后发现样本局部仍与真实效果不符,这与模型在运行过程中出现的梯度消失问题有关。



(a)WGAN训练迭代10 000次实验结果



(b)WGAN训练迭代30 000次实验结果



(c)迭代10 000次与30 000次结果局部对比

图 2 WGAN 在 LSUN 数据集中 bedroom 数据的实验结果

文献[9]的 WGAN 存在梯度消失问题,在 WGAN 中加入梯度惩罚(Gradient Penalty, GP),即 WGAN-GP,使梯度在传播过程中保持平稳,同时该模型还改进了对整个分布进行参数限制的方式,只针对真假样本以及两者之间的过渡部分进行限制,不仅能够提升训练速度,还使生成结果变得多样化^[11]。

与 WGAN 相类似,各种方法被提出,例如:在 f-GAN 中引入 f-divergence 度量样本之间距离^[12],在 LSGAN (Least Squares GAN) 中引入 Pearson 卡方散度对 2 个分布进行相似性度量^[13],LS-GAN (Loss Sensitive GAN)^[14]对生成样本中距离真实数据分布较远的样本进行重点优化,即“按需分配”思想,EBGAN (Energy-based GAN)^[15-17]是从能量模型出发,以能量值作为衡量标准,BEGAN (Boundary Equilibrium GAN)则是在 EBGAN 和 WGAN 的基础上,提出边界均衡 GAN 架构,使用标准的训练步骤实现快速、稳定的收敛^[18]。

2.1.2 EBGAN 模型

与 WGAN 基于样本间的距离度量不同,EBGAN^[15]是从能量模型的角度对 GAN 进行改进。该模型将判别器看作一个能量函数,在真实数据范围内该能量函数的能量值会减小,在非真实数据(即认为是生成数据)范围内能量值会增大。这样生成器的目标就是产生能量值足够小的样本,而判别器则是以对生成的样本赋高的能量值以及对真实样本赋低的能量值为目标。EBGAN 的意义是给予 GAN 一种不同于其他模型通过距离度量定义损失函数的能量模型的定义,即用更宽泛的结构和更多样的损失函数类型来训练 GAN 模型。

2.2 基于模型框架和训练方法的改进

2.2.1 InfoGAN 模型

InfoGAN (Information Maximizing Generative Adversarial Nets) 模型改进了 GAN 的输入信息,使得 GAN 的输入不再只是语义模糊的随机噪声,实现对随机噪声与生成图像特征之间对应关系建模,并且可以通过控制相应维度的变量来控制生成的样本产生相应的变化,同时该模型还提出一个约束函数,称为互信息,可以保证附加的输入信息较好地利用^[19],对 GAN 公式进行修改,如式(7)所示。

$$\min_c \max_b V_I = V(D, G) - \lambda I(c; G(z, c)) \quad (7)$$

其中, c 为具有隐含意义并且可解释的附加信息,而剩下的无法明确描述的信息由 z 提供,这样使生成的结果具有一定的可控性,同时为保证隐变量 c 被有效使用,提出互信息 I ,其计算公式如下:

$$I(X; Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) \quad (8)$$

对于式(8)中的 X 和 Y ,两者之间有较高的互信息才能确保彼此有联系,因此优化的过程就是互信息升高的过程。

将 InfoGAN 真正用于训练时,发现其给出的训练代码目前只有基于 MNIST 数据集。图 3 为由一个 10 维的离散码、2 个 1 维的连续码以及 62 维的噪声变量控制的 InfoGAN 模型训练 50 个 epoch 的实验结果。其中,图 3(a)为离散码控制的数字分类结果,图 3(b)和图 3(c)为 2 个连续码控制数字旋转

和宽窄的结果。与文献[19]的实验结果不同,本文给出的代码无法准确控制数字的单一属性。为使数字旋转效果更明显,文献[19]将连续码范围设置为 $(-2, 2)$,而目前给出的代码只支持 1 维的连续码。因此实验未达到预期结果,数字在旋转的同时宽窄也在改变。



(a)离散码控制数字分类



(b)连续码1控制数字旋转和宽窄



(c)连续码2控制数字旋转和宽窄

图3 InfoGAN 在 MNIST 数据集上的实验结果

对于 InfoGAN,包括其提出的互信息公式不仅可应用于生成样本与隐变量之间的计算,也可与 VAE^[7]模型结合,这样原始信息的处理不仅是对真实数据的一个编码解码再输入的过程,而且可以在原始信息的基础上加入具有隐含意义的条件 c ,在过程中利用互信息进行约束后再输入,得到期望的样本。

与 InfoGAN 相类似的还有 PixelGAN Autoencoders^[20]、CGAN (Conditional Generative Adversarial Nets)^[21]、

Structured GAN^[22]以及 Cat GAN (Categorical Generative Adversarial Networks)^[23],同样是在原始随机噪声的基础上加入条件变量或是在条件变量的基础上加入数据熵判断。文献[20]以潜在代码为条件,通过无监督的学习方式分离图像的样式和内容信息。在 CGAN 中条件变量多为类别标签信息,相当于将原始的无监督 GAN 变为有监督的模型,除利用数据集中的标签生成指定类别的图像,文献[21]将 CGAN 用于图像自动标注的应用,在 MIR Flickr25000 数据集上,以图像特征为条件变量,生成该图像的标注。与 CGAN 密切相关的 StackGAN^[24]其本质就是 CGAN 的堆叠,第 1 层生成一个低分辨率的图像样本,第 2 层则在此基础上进行优化。

2.2.2 DCGAN 模型

原始 GAN 的生成器和判别器都是由最简单的多层感知器构成,训练效率以及生成的样本质量都较低,而 DCGAN (Deep Convolutional Generative Adversarial Networks)^[25]模型将 GAN 中的生成网络和判别网络替换为经过一定程度改进的 CNN (Convolutional Neural Networks)^[26-27],其算法将有监督学习中的 CNN 和无监督学习中的 GAN 相结合,为 GAN 的训练提供较好的网络结构,同时提高了训练过程的稳定性和生成结果的质量。此外,该模型可通过特征学习或是特征向量计算得到一个稳定的向量来进行特定变换。

2.2.3 结构扩展模型

GAN 的优势在于除了可以在生成网络或判别网络上做改进,还可以构建多个神经网络来辅助其训练。Dual GAN^[28]、Disco GAN^[29]、Cycle GAN^[30]及 Triangle GAN^[31]在原始 GAN 结构基础上,增加了一个生成器和一个判别器,通过双生成器双判别器模型更好地学习 2 个域之间的双向条件分布。Dual GAN 基于成对的训练集实现了一个数据分布到另一个数据分布的映射,但该模型的最终目的是为了重建图像与原始图像的高度拟合,来解决高质量匹配图像对的难题。Triangle GAN 则不仅只局限于学习 2 个图像域之间的双向条件分布,还成功学习了图像域与标签域之间的双向分布,最终实现了半监督图像分类和基于属性的图像生成工作。

D2GAN (Dual Discriminator Generative Adversarial Nets)^[32]即双判别器模型,与生成器共同训练,其中一个判别器偏向来自真实样本的数据分布,而另一个判别器则偏向来自生成器生成的数据,以生成高质量和多样化的样本。Triple-GAN^[33]通过构造一个生成器、一个判别器和一个分类器模型实现深度生成模型中最先进的分类结果,模型的生成器和分类器表征图像和标签之间的条件分布,判别器识别伪造的图像标签对以此来实现深度生成模型中最先进的分类结果。SGAN^[34]模型定义了 N 个生成器 G

和判别器 D ,一对全局 GAN 及 $N-1$ 对局部 GAN,训练时全局生成器和判别器分别使用局部生成器和判别器进行训练,这种方法较好地缓解了模式崩溃、收敛时的稳定性并提高了收敛速度。

2.2.4 递归模型

LAPGAN (Laplacian Generative Adversarial Networks)^[35]将原始 GAN 的生成器与判别器定义为 Laplacian Pyramids 框架内的级联卷积网络,以 LAPGAN 的生成模型为例,文中定义多个生成器由粗到精生成样本。GRAN (Generating images with Recurrent Adversarial Networks)^[36]同样是通过重复几次步骤的顺序生成相似的模型,基于 VAE (Variational Auto-Encoder)^[7],分多步对图像编码再解码而不是一步生成,其来自于 DRAW (Deep Recurrent Attentive Writer)^[37],逐步生成一个完整的图像。这种模型的优点是重复产生以前一步状态为基础的输出,通过将复杂数据分布映射到简单的顺序上的方法简化图像的生成问题。Improved GAN^[38]通过改变训练过程中的顺序和方法同样也提高了模型的稳定性。

2.3 GAN 与其他模型的结合

2.3.1 GAN + VAE 模型

GAN 和 VAE 虽不同,但两者并非完全冲突,文献[39-40]将 GAN 与 VAE 结合作为双方的辅助模型,以更好地提升样本生成的质量。文献[39]将 GAN 作为 VAE 的一个辅助,两者结合后可以通过机器学习建立一个公式来进行合理的度量,而不是通过预先人为制定一个适合度量标准问题的方法。将 GAN 的判别器部分作为 VAE 中 decoder 部分的一个辅助,原因在于 GAN 判别器网络隐含着学习图像的丰富的相似性度量的能力,其替换了此前 VAE 采用的像素级误差,虽然实现起来更容易,但其并不适合图像数据。因为对于整幅图来说,细微的变化人们往往不会察觉到,反而是一些关键地方的变化,影响着人们对整幅图像的印象。将像素级误差度量替换为 GAN 判别器的相似性度量,效果会更佳。

除了 GAN 与 VAE 的结合,还有 GAN 与 pixel CNN^[41]以及 pixel RNN^[42]的结合,是将原本对像素的操作改为对整幅图像的衡量与评价,以及文献[43]将通常在原始图像空间的损失度量替换为在特征空间的损失度量,同时在原有 GAN 的对抗训练损失基础上又加入了 2 个损失项,提升了 GAN 生成图像的清晰度和训练时的收敛速度。

2.3.2 GAN + RNN 模型

循环神经网络 (Recurrent Neural Network, RNN)^[44-45],其网络的内部结构可以展示动态时序行为,在自然语言处理、手写识别和语音识别领域应用广泛。

Continuous-RNN-GAN^[46]是 GAN 在音乐生成领域的较早应用,也是 GAN 处理连续序列数据的针对性研究。其模型在 SeqGAN^[47]提出的判别器对于

序列性数据的区分方法的基础上提出了一种循环对抗神经网络架构,针对以往音乐计算研究中使用符号特征(Symbolic Representation)的不足(计算机更容易理解数字表达),以及GAN的优势,将GAN与RNN结合对古典音乐进行学习并模拟其风格生成连续的序列数据。该模型采用Deep LSTM(Long Short-Term Memory)^[48-49]单元构建循环神经网络,每一个LSTM单元都有输入门、遗忘门和输出门,有用的信息被传递给循环神经网络的下一个单元,无用的信息则被遗忘,以保证信息有效的长期记忆。Continuous-RNN-GAN的整个训练过程为生成器产生连续数据序列,按顺序分段传递给每一个LSTM,然后将生成的结果传递给判别器,判别器来区分真实音乐和生成数据。同时该模型还基于音调的变化范围等提出了自己的判别指标,使用韵律学的方式,根据Polyphony(2个音同时弹奏的频率)、Scale consistency(标准音程的比例)、Repetitions(音符组合重复的频率)、Tone span(整段音乐的最低最高音阶差这几个方面进行计算。

本节主要对GAN的扩展与改进工作进行归纳和分析,详略介绍了具有代表性的研究成果。从数据相似性度量、模型框架和训练方法以及与其他模型相结合3个方面展开讨论。对于较为经典的模型如WGAN,InfoGAN等,给出实验结果分析,但由于所训练的数据以及设备的差异都会造成实验结果的不同,因此不能作为研究结果。通过对各类研究成果分析发现,对于GAN的改进与扩展主要集中于训练过程中损失计算方法的改进或对于模型框架结构的扩展,利用该方式达到训练过程快速收敛或优化生成样本质量的目的。

3 GAN应用

近年来,GAN的学习能力和改进潜力受到广泛关注,研究者提出各种形态的GAN并将其应用于各领域,推动了GAN研究与应用的进展。

3.1 图像合成

文献[50]提出Composite GAN,通过多个生成器生成部分图像并合成整幅图像,例如在人脸生成中将背景和人脸分为两部分,由不同的生成器生成放入RNN中最后合成,同样可生成较为逼真的图像。该模型与PixelNN的本质不同在于其结果都是模型通过训练时学习后生成的,生成的结果与原图相类似但并不相同。文献[51]提出LR-GAN,将图像分为前景和背景,以递归的方式分别生成再合成产生完整的图像,其研究成果对于推动GAN在图像合成领域的进一步应用具有重要意义。

3.2 风格迁移

风格迁移(Style Transfer, ST)是深度学习应用中非常有趣的一种,可使用这种方法将一张图片的

风格“迁移”到另一张图片上。深度学习最早是基于CNN框架并尝试进行图像风格迁移^[52]。但该模型有如下2个缺点:一是训练速度太慢,每一次生成都相当于重新训练一个模型;二是对于训练样本要求较高,需要在成对的数据上训练,在实际生活中并没有如此完美的成对训练数据。

基于GAN的自主学习和生成随机样本的特点,与加入条件变量生成特定样本的方法相结合,研究者将GAN改进后能够在无监督条件下,通过不成对的训练数据学习图像风格变换^[53-55]。文献[53]利用一张所需风格图对模型进行无监督学习训练,得到前向生成网络,在测试模型时直接将需要转换的一批图像输入到网络中,即可得到具有预先训练的风格内容图集。文献[56]是苹果公司为训练神经网络并用于用户照片上的人脸和其他物体的检测,通过改进GAN模型生成大量的人物合成图像,未依靠人力采集数据,使训练神经网络的成本更低而且速度也更快。

GAN在图像风格迁移方面的出色表现,为其后在大型数据集的生成工作上奠定了基础,用较小的数据集完成GAN的无监督训练,然后以此来生成更多与数据样本相类似的图像,可节约大量成本。表1列出各类GAN模型在基于Cityscapes数据集上照片到标签实验的性能对比。

表1 基于Cityscapes数据集中不同模型的性能对比

模型	准确率		相关度
	像素级	分类	
GAN ^[6]	0.22	0.10	0.05
DualGAN ^[28]	0.27	0.13	0.06
CycleGAN ^[30]	0.58	0.22	0.16
SimGAN ^[56]	0.47	0.11	0.07
cGAN ^[57]	0.54	0.33	0.19
CoGAN ^[58]	0.45	0.11	0.08

3.3 超分辨率图像

SRGAN(Super Resolution GAN)^[59]以文献[53]中perceptual loss为基础,是一种新的损失函数与GAN本身的损失结合,实现了从低分辨率图像到超分辨率图像的生成。与其他超分图像处理的模型相比,SRGAN的生成图像要比Bicubic算法以及对残差网(Residual Network)^[60]进行改进后得到的SRResNet(super-resolution residual network)^[59]的效果更加清晰,在实验过程中发现,较高的PSNR(peak signal-to-noise ratio)^[61-62]并不能带来较好的视觉感受,因此提出新的损失函数作为评判标准,将生成数据和真实数据分别输入VGG-19^[63]网络,根据得到的feature map的差异来定义损失项,可以使超分辨率效果更能满足人的感官。

SRGAN的实验结果表明,图像看上去比较清晰,

但仔细观察会发现图像虽然清晰了,细节却仍与原图不符。如果能在处理整幅图像的基础上加入一个局部图像的生成和判别模型,效果应该会更好。表 2 列出了各类应用于超分辨率图像领域的 GAN 改进模型实验结果性能对比,其中文献[53]给出基于 2 种不同损失函数下生成的样本,基于单个像素的损失和基于特征重建的损失,基于特征重建的损失为文献[53]提出的损失函数即 feature reconstruction loss (或 perceptual loss)。经过对比同样证实高的 PSNR 和 SSIM 并不代表有更好的视觉观感。

表 2 不同模型 PSNR 及 SSIM 的对比结果

模型	PSNR/dB	SSIM
perceptual loss ^[53]	21.66	0.555 8
SRGAN ^[59]	21.15	0.686 8
SRResNet ^[59]	23.53	0.783 2
SRCNN	22.53	0.652 4
Bicubic	21.69	0.584 0

3.4 视觉概念矢量运算

当前,虽然各类大规模的图像数据集不断出现,但要得到具有指定特征的图像较难。DCGAN^[25]模型可以通过对具有 2 个不同特征图像间的运算提取所需的图像特征,并与第三类图像合成来达到生成特定样本的目的,即图像的视觉概念矢量运算。通过对具有相应属性的图像之间的计算得出所需特定属性的图像。除实现这类单一属性的创建,还通过人脸的左侧和右侧不同角度的图像样本创建“转向”矢量,实现人脸角度的转换。

CoGAN(Coupled Generative Adversarial Networks)^[56]提出的思想是对于训练集不同域中的相应图像属性的联合分布,从而在没有完全相对应属性样本数据的前提下生成所需的图像

上述应用不仅只局限于实现了图像视觉概念的矢量计算,而是当需要某一类图像时,除带着特定的目标去进行数据采样,还可选择通过可用的样本数据来生成所需的数据,降低训练成本。

3.5 人脸修复与识别

人类可以通过局部信息分析整体,例如看到一个人的侧脸,可以大概得到此人本来的面目。因此,将 GAN 与现有的人脸识别和验证方法^[64-65]相结合,应用于人脸识别领域中。

文献[66]提出 TP-GAN(Two-Pathway GAN),通过侧面人脸照片合成正面人脸图像,取得较好的结果。该模型生成网络的特点为双路径生成,2 个路径分别处理局部和全局变换,具体来说,就是一条路径通过侧脸提取的部分五官产生脸部完整五官,另一条路径产生一个模糊的完整人脸。局部生成路径

和全局结构路径都由卷积神经网络构成编码器和解码器,除此之外还应用了 4 个定位标记的补丁网络(Landmark Located Patch Networks,LLPN)处理人脸的五官,通过分工合作,合成一张完整逼真的人脸图像。TP-GAN 的实验结果在相同条件下与其他合成方法相比效果较好,能逼真地还原正脸。

除侧脸到正脸的复原,面部被遮挡的图片研究也较多,在没有参考的情况下,将其生成与现有的背景修复技术相比,难度会更高。文献[67]提出基于 WGAN 的人脸图像生成网络,能够为图像中的遮挡区域生成该人脸的补全图像,目的是为了进行人脸表情的识别,通过建立表情识别任务和身份识别任务的对抗关系来提取表情特征,并推断表情类别。

对于人脸修复,文献[68]提出的修复模型由一个生成器和 2 个判别器以及一个解析网络构成,将被遮挡的人脸图像送入由神经网络构成的编码器和解码器的生成模型中,之后分为两部分判别,一部分为遮盖部分的生成内容与原图遮盖部分进行区分,另一部分为生成的完整人脸图像与原始图像进行区分。最后的解析网络是预先训练好的,进一步确保生成的内容足够逼真,同时也保证生成图像与原始图像的像素一致性。但该模型对于不对称的人脸图像,无法生成特定五官的关键特征,并且无法有效利用相邻像素之间的相关性进行生成。为进一步评估该模型在人脸识别中的性能,提出对一张完整的人脸图像进行 6 种不同的遮挡作为输入数据,将模型生成的样本作为查找标准,在包含原始完整图像的数据集中进行搜寻,以搜索出相同身份的图像为目标,结果显示无论哪种遮挡类型,该模型的生成结果都最接近于原始图像。但不同的遮挡类型对于生成样本最后的识别精度也有较大的影响,遮挡的部分越多,生成的样本越不容易与相同身份的图像进行匹配。最后提出使用基于像素的循环神经网络 PixelRNN^[42]来解决该模型的局限性。

3.6 图像标注

当前基于 RNN 与 CNN 结合的方法根据图像样本生成图像标注已经取得了显著的成果^[69]。RNN 的时序性使其在自然语言生成领域被广泛应用。基于 GAN 的改进方法对自然语言的处理则不需要依赖 RNN。CGAN^[21]在多模态实验中同样实现了图像标注,模型以图像特征为条件变量通过卷积神经网络来生成标签向量的条件分布,实现图像的多标签自动标注。模型 WGAN-GP^[11]将文本视为概率分布向量,不需要 RNN 或是任何基于 RNN 的语言模型,在输入高斯噪声后通过多层反卷积神经网络实现直接一次生成 32 个字符的纯文本生成任务。通

过图像能够生成文本,同样可以利用 GAN 实现从文本生成图像^[70-71]。文献[70]的 GAN 通过文本描述生成图像的工作,其模型训练过程是将描述文本编码后与随机噪声共同输入到生成器中生成与文本相符的图像,同时编码后的文本也作为判别条件输入判别器中用来构建目标函数。该模型解决了文本自动合成图像在人工智能技术具有挑战性的难题,将

文本编码和图像建模技术相结合,并将视觉概念从字符表示转换为像素表示,取得了较好的实验结果,对 GAN 更好地应用于生成特定图像研究有突破性的意义。表 3 列出了 5 个应用领域内 GAN 衍生模型及结构。其中,CGAN 可实现 image to label, WGAN-GP 和 DCGAN 可用于噪声建模, CNN 用于去噪, iVGAN 实现了视频上色、视频修复等。

表 3 GAN 在 5 个应用领域的衍生模型及结构

应用	模型	模型结构主要特征
Image to image	DCGAN ^[25] , DiscoGAN ^[29] , CartoonGAN ^[72]	CNN ^[26-27]
	InfoGAN ^[19]	DCGAN ^[25]
	CycleGAN ^[30,53]	CNN (Instance Normalization)
	DualGAN ^[28]	WGAN ^[9]
	IcGAN ^[73] , GAGAN ^[74] , Triangle GAN ^[31]	CGAN ^[21]
	DA-GAN ^[75]	Autoencoder, VGG ^[63]
	IPCGAN ^[76]	CGAN, LSGAN ^[13]
	StarGAN ^[77]	CGAN, CycleGAN ^[30]
Text to image	CGAN, GAWWN ^[78]	CNN
	GAN-INT-CLS ^[69] , Triangle-GAN ^[31]	CGAN
	StackGAN ^[24]	CNN, encoder-decoder
	AttnGAN ^[71]	CNN, LSTM ^[48-49]
	DA-GAN ^[75]	Autoencoder, VGG
Super-resolution	Perceptual loss ^[53]	CNN
	SRGAN ^[59]	CNN, VGG, SRResNet ^[59]
	Finding tiny faces ^[79]	CNN, VGG
图像增强、图像复原	Stacked CGAN ^[80]	CGAN
	Attentive GAN ^[81]	CNN, LSTM, ResNet ^[60] , Autoencoder
	G CBD ^[82]	WGAN-GP ^[11] , DCGAN ^[25] , CNN
	Dehazing CGAN ^[83]	CGAN, VGG, CNN
	SCGAN ^[84]	U-net, CGAN
视频生成、行为预测	VGAN ^[85]	CNN
	iVGAN ^[86] , TGAN ^[87]	CNN, WGAN ^[9]
	MoCoGAN ^[88]	RNN ^[44-45] , CNN
	时空嵌入式 GAN ^[89]	CNN, LSTM
	MD-GAN ^[90]	3D U-net, encoder-decoder
	Social GAN ^[91]	LSTM, encoder-decoder, MLP

3.7 其他应用

除上述具有代表性领域的应用成果, GAN 在其他领域也开始有所发展。改进后的 GAN 模型用于图像以及文本与图像(无论是动态图像还是静态图像)的研究与当前该领域的常用模型相比,具有较好的性能。此外,国内外学者还将研究重点聚焦于将 GAN 与 RNN 结合生成具有时间序列性质的数据生成工作中。Continuous-RNN-GAN^[46]正是将 GAN 生成的特定数据按时间顺序放入 RNN 的每个单元中,最后生成曲风相似但与原始训练数据不同模型的古典音乐^[72-74]。文献[75-77]基于不同改进方向的 GAN 实现在不同域间的图像转换。由 CNN 构建的 GAN 网络结构模型通过学习可以生成具体的图像绘制指令^[78]。将模糊的小图经过 GAN 后生成高

分辨率人脸图像并进行识别^[79]。通过 GAN 实现图像的去阴影、去雨滴、去雾等图像增强^[80-81]和复原工作^[82-84]。文献[85-87]实现了基于 GAN 的视频生成和行为预测^[88-90],生成结果是很多帧具有序列性的图像^[91]。基于 CNN 的 GAN 通过无监督学习实现了完整且高分辨率的 3D 物体的生成^[92-93]。研究者将 GAN 基于收集到的真实商标图像数据用于商标的设计^[94]。ExGAN^[95]实现了通过闭合双眼的人脸图像生成合理的双眼睁开的人脸图像。通过 GAN 突破单一任务的模仿学习实现了多模态仿真学习,使得机器人可以同时执行多项任务以便于更好地适应现实场景^[96]。文献[97]提出基于模型的生成对抗模仿学习(MGAIL)算法,使用相对较少的专家样本和与环境的相互作用来训练并生成策略样本进行

政策模仿。TextGAN^[98]采用长短期记忆网络作为生成器,卷积神经网络作为判别器实现了生成语法框架合理的语句。GAN 的延伸模型对抗式自编码(Adversarial Auto-Encoders,AAE)^[99]还被应用于医药研制,以 AAE 为基础,加上已知的各类药物成分的医疗特性和有效浓度,将之用在训练神经网络上,将所需化合物相关的信息输入到网络中,得到合成该化合物所需的各成分准确比例。这项工作对于未来基于人工智能的药物研发具有指导意义。

本节主要对 GAN 的应用工作进行梳理和介绍。从图像合成、风格迁移、超分辨率图像、视觉概念矢量运算、人脸修复与识别、图像标注与文本到图像等方面进行具体的描述和讨论。在其他应用中列举了视频生成、文字生成以及利用 GAN 进行多任务的仿真学习等研究成果。与在此之前的深度神经网络更多的是用于对样本的优化和分类不同,GAN 的样本生成能力优于其他模型。GAN 在各领域的广泛应用是对其能力的肯定。

4 未来发展趋势

各类 GAN 不断被提出,应用涉及的领域越来越多,GAN 的不足之处也更加突显。首先,最棘手也最需要解决的就是训练过程不稳定问题,虽然 WGAN 等在 GAN 模型崩溃和训练不收敛的问题上有一定的改进,但无法保证每次的训练结果的收敛速率。GAN 生成的样本质量的优劣没有一个完全适合的判定标准,如直接由人的肉眼来评判,没有将第三方的客观评价作为标准,这就使得模型的判别在一定程度上受主观因素的影响。显然,GAN 还有较多待解决的问题,但这不影响其在人工智能技术中进一步的发展与应用。GAN 未来的发展趋势可分为以下方面:

1) 网络框架及算法的改进

WGAN、DCGAN 等的出现改善了 GAN 模型训练过程的不稳定性与模型崩溃等问题,但未彻底解决此问题,梯度消失、过拟合以及生成样本自由度较高等模型性能问题仍是当前 GAN 发展所面临的阻碍。但 GAN 网络结构设计的自由度较高,能够整合各类损失函数,因此解决上述问题即 GAN 模型的优化是一个重要的研究方向。

2) 与其他学习方法结合

基于 GAN 较强的融合性将其与其他学习方法相结合,例如 DCGAN 将有监督与无监督学习相结合,实现了 GAN 的半监督学习,将半监督学习 GAN 应用于人脸、文字、语音等识别领域。类似地,GAN 也可以与迁移学习、特征学习、强化学习等结合来解

决更加有针对性的实际应用问题,更好地推动其在各领域发展。

3) 样本生成和数据增广

在各行业应用中,获取大量真实样本需要耗费财力、人力、物力,GAN 可通过对有限数据的某个维度稍作改变,生成大量仿真数据,但当前基于 GAN 生成样本的方法其可控程度明显不足,不一定能够获得期望形式、效果和质量的样本。因此,研究精准可控的 GAN 数据生成方法,生成指定形式、效果和质量的样本,是解决数据增广和数据集扩充问题的有效途径和重要研究方向。

4) 评估方法的完善

由于 GAN 目前仍缺乏对于模型性能、准确性以及生成的样本质量等方面的客观评价标准,各项研究工作中的判定方法仅适用于当前研究成果,使得模型的判定在一定程度上受主观因素的影响。因此构建一个更加精确化、标准化、通用化的第三方综合评价标准同样是 GAN 研究领域一个亟待解决的问题。

5 结束语

GAN 将机器学习中的生成式模型与判别式模型相结合,并应用于深度神经网络,根据数据类型构建合适的模型结构。GAN 不仅可以生成逼真的数据,还可通过无监督学习获得数据的分布。分析结果表明,GAN 具有较强的学习能力和可塑性。

参考文献

- [1] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks [J]. Science, 2006, 313(5786): 504-507.
- [2] BENGIO Y. Learning deep architectures for AI [J]. Foundations and Trends in Machine Learning, 2009, 2(1): 1-127.
- [3] SCHMIDHUBER J. Deep learning in neural networks: an overview [J]. Neural Networks, 2015, 61: 85-117.
- [4] 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.
- [5] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks [C]//Proceedings of International Conference on Neural Information Processing Systems. Cambridge, USA: MIT Press, 2012: 1097-1105.
- [6] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets [C]//Proceedings of International Conference on Neural Information Processing Systems. Cambridge, USA: MIT Press, 2014: 2672-2680.
- [7] KINGMA D P, WELING M. Auto-encoding variational

- bayes[EB/OL]. [2018-05-22]. <https://arxiv.org/pdf/1312.6114.pdf>.
- [8] 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): 321-332.
- [9] ARJOVSKY M, CHINTALA S, BOTTOU L. Wasserstein GAN[EB/OL]. [2018-05-22]. <https://arxiv.org/pdf/1701.07875.pdf>.
- [10] FUGLEDE B, TOPSOE F. Jensen-Shannon divergence and Hilbert space embedding [C]//Proceedings of International Symposium on Information Theory. Washington D. C., USA: IEEE Press, 2004: 31.
- [11] GULRAJANI I, AHMED F, ARJOVSKY M, et al. Improved training of wasserstein GANs [EB/OL]. [2018-05-22]. <https://arxiv.org/pdf/1704.00028.pdf>.
- [12] NOWOZIN S, CSEKE B, TOMIOKA R. f-GAN: training generative neural samplers using variational divergence minimization[C]//Proceedings of International Conference on Neural Information Processing Systems. Cambridge, USA: MIT Press, 2016: 271-279.
- [13] MAO Xudong, LI Qing, XIE Haoran, et al. Least squares generative adversarial networks [EB/OL]. [2018-05-22]. <https://arxiv.org/pdf/1611.04076.pdf>.
- [14] QI Guojun. Loss-sensitive generative adversarial networks on lipschitz densities [EB/OL]. [2018-05-22]. <https://arxiv.org/pdf/1701.06264.pdf>.
- [15] ZHAO Junbo, MATHIEU M, LECUN Y. Energy-based generative adversarial network [EB/OL]. [2018-05-22]. <https://arxiv.org/pdf/1609.03126.pdf>.
- [16] DAI Zihang, ALMAHAIRI A, BACHMAN P, et al. Calibrating energy-based generative adversarial networks [EB/OL]. [2018-05-22]. <https://arxiv.org/pdf/1702.01691.pdf>.
- [17] WANG Dilin, LIU Qiang. Learning to draw samples: with application to amortized mle for generative adversarial learning [EB/OL]. [2018-05-22]. <https://arxiv.org/pdf/1611.01722.pdf>.
- [18] BERTHELOT D, SCHUMM T, METZ L. BEGAN: boundary equilibrium generative adversarial networks [EB/OL]. [2018-05-22]. <https://arxiv.org/pdf/1703.10717.pdf>.
- [19] CHEN Xi, DUAN Yan, HOUTHOOFT R, et al. InfoGAN: interpretable representation learning by information maximizing generative adversarial nets [C]//Proceedings of International Conference on Neural Information Processing Systems. Cambridge, USA: MIT Press, 2016: 2172-2180.
- [20] MAKHZANI A, FREY B J. PixelGAN autoencoders [C]//Proceedings of International Conference on Neural Information Processing Systems. Cambridge, USA: MIT Press, 2017: 1972-1982.
- [21] MIRZA M, OSINDERO S. Conditional generative adversarial nets [EB/OL]. [2018-05-22]. <https://arxiv.org/pdf/1411.1784.pdf>.
- [22] DENG Zhijie, ZHANG Hao, LIANG Xiaodan, et al. Structured generative adversarial networks [C]//Proceedings of International Conference in Neural Information Processing Systems. [S. l.]: Curran Associates Inc., 2017: 3902-3912.
- [23] SPRINGENBERG J T. Unsupervised and semi-supervised learning with categorical generative adversarial networks [EB/OL]. [2018-05-22]. <https://arxiv.org/pdf/1511.06390.pdf>.
- [24] ZHANG Han, XU Tao, LI Hongsheng, et al. StackGAN: text to photo-realistic image synthesis with stacked generative adversarial networks [EB/OL]. [2018-05-22]. <https://arxiv.org/pdf/1612.03242.pdf>.
- [25] RADFORD A, METZ L, CHINTALA S. Unsupervised representation learning with deep convolutional generative adversarial networks [EB/OL]. [2018-05-22]. <https://arxiv.org/pdf/1511.06434.pdf>.
- [26] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition [J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [27] LECUN Y, BOSER B, DENKER J S, et al. Backpropagation applied to handwritten zip code recognition [J]. Neural Computation, 1989, 1(4): 541-551.
- [28] YI Zili, ZHANG Hao, PING Tan, et al. DualGAN: unsupervised dual learning for image-to-image translation [EB/OL]. [2018-05-22]. <https://arxiv.org/pdf/1704.02510.pdf>.
- [29] 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. Berlin, Germany: Springer, 2017: 1857-1865.
- [30] ZHU Junyan, PARK T, ISOLA P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks [EB/OL]. [2018-05-22]. <https://arxiv.org/pdf/1703.10593.pdf>.
- [31] GAN Zhe, CHEN Liqun, WANG Weiyao, et al. Triangle generative adversarial networks [C]//Proceedings of International Conference on Neural Information Processing Systems. Cambridge, USA: MIT Press, 2017: 5253-5262.
- [32] NGUYEN T, LE T, VU H, et al. Dual discriminator generative adversarial nets [C]//Proceedings of International Conference on Neural Information Processing Systems. Cambridge, USA: MIT Press, 2017: 2667-2677.
- [33] LI Chongxuan, XU Tun, ZHU Jun, et al. Triple generative adversarial nets [C]//Proceedings of International Conference on Neural Information Processing Systems. Cambridge, USA: MIT Press, 2017: 4091-4101.
- [34] CHAVDAROVA T, FLEURET F. SGAN: an alternative training of generative adversarial networks [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2018: 9407-9415.
- [35] DENTON E, CHINTALA S, FERGUS R. Deep generative image models using a Laplacian pyramid of adversarial networks [C]//Proceedings of International Conference on Neural Information Processing Systems. Cambridge, USA: MIT Press, 2015: 1486-1494.

- [36] IM D J, KIM C D, JIANG Hui, et al. Generating images with recurrent adversarial networks [EB/OL]. [2018-05-22]. <https://arxiv.org/pdf/1602.05110.pdf>.
- [37] GREGOR K, DANIHELKA I, GRAVES A, et al. DRAW: a recurrent neural network for image generation [C]//Proceedings of International Conference on International Conference on Machine Learning. [S. l.]: JMLR. org, 2015:1462-1471.
- [38] SALIMANS T, GOODFELLOW I, ZAREMBA W, et al. Improved techniques for training GANs [C]//Proceedings of International Conference on Neural Information Processing Systems. Cambridge, USA: MIT Press, 2016: 2234-2242.
- [39] LARSEN A B L, SØNDERBY S K, LAROCHELLE H, et al. Autoencoding beyond pixels using a learned similarity metric [C]//Proceedings of International Conference on Machine Learning. [S. l.]: JMLR. org, 2016:1558-1566.
- [40] MESCHEDER L, NOWOZIN S, GEIGER A. Adversarial variational Bayes: unifying variational autoencoders and generative adversarial networks [C]//Proceedings of International Conference on International Conference on Machine Learning. [S. l.]: JMLR. org, 2017:2391-2400.
- [41] SALIMANS T, KARPATY A, CHEN Xi, et al. PixelCNN ++: improving the pixelCNN with discretized logistic mixture likelihood and other modifications [EB/OL]. [2018-05-22]. <https://arxiv.org/pdf/1701.05517.pdf>.
- [42] OORDA V D, KALCHBRENNER N, KAVUKCUOGLU K. Pixel recurrent neural networks [C]//Proceedings of International Conference on International Conference on Machine Learning. [S. l.]: JMLR. org, 2016:1747-1756.
- [43] DOSOVITSKIY A, BROX T. Generating images with perceptual similarity metrics based on deep networks [C]//Proceedings of International Conference on Neural Information Processing Systems. Cambridge, USA: MIT Press, 2016:658-666.
- [44] GRAVES A. Generating sequences with recurrent neural networks [EB/OL]. [2018-05-22]. <https://arxiv.org/pdf/1308.0850.pdf>.
- [45] LEE H Y, TSENG B H, WEN T H, et al. Personalizing recurrent-neural-network-based language model by social network [J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2017, 25 (3): 519-530.
- [46] MOGRENO. C-RNN-GAN: continuous recurrent neural networks with adversarial training [EB/OL]. [2018-05-22]. <https://arxiv.org/pdf/1611.09904.pdf>.
- [47] YU Lantao, ZHANG Weinan, WANG Jun, et al. SeqGAN: sequence generative adversarial nets with policy gradient [C]//Proceedings of Conference on Artificial Intelligence. [S. l.]: AAAI Press, 2017: 2852-2858.
- [48] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. Neural Computation, 1997, 9(8): 1735-1780.
- [49] SAK H, SENIOR A, BEAUFAYS F. Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition [EB/OL]. [2018-05-22]. <https://arxiv.org/pdf/1402.1128v1.pdf>.
- [50] KWAK H, ZHANG B T. Generating images part by part with composite generative adversarial networks [EB/OL]. [2018-05-22]. <https://arxiv.org/pdf/1607.05387.pdf>.
- [51] YANG Jianwei, KANNAN A, BATRA D, et al. LR-GAN: Layered recursive generative adversarial networks for image generation [EB/OL]. [2018-05-22]. <https://arxiv.org/pdf/1703.01560.pdf>.
- [52] GATYS L A, ECKER A S, BETHGE M. A neural algorithm of artistic style [EB/OL]. [2018-05-22]. <https://arxiv.org/pdf/1508.06576.pdf>.
- [53] JOHNSON J, ALAHI A, LI Feifei. Perceptual losses for real-time style transfer and super-resolution [C]//Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2016: 694-711.
- [54] BOUSMALIS K, SILBERMAN N, DOHAN D, et al. Unsupervised pixel-level domain adaptation with generative adversarial networks [EB/OL]. [2018-05-22]. <https://arxiv.org/pdf/1612.05424.pdf>.
- [55] LIU Mingyu, BREUEL T, KAUTZ J. Unsupervised image-to-image translation networks [C]//Proceedings of International Conference on Neural Information Processing Systems. Cambridge, USA: MIT Press, 2017: 700-708.
- [56] SHRIVASTAVA A, PFISTER T, TUZEL O, et al. Learning from simulated and unsupervised images through adversarial training [EB/OL]. [2018-05-22]. <https://arxiv.org/pdf/1612.07828.pdf>.
- [57] ISOLA P, ZHU Junyan, ZHOU Tinghui, et al. Image-to-image translation with conditional adversarial networks [EB/OL]. [2018-05-22]. <https://arxiv.org/pdf/1611.07004.pdf>.
- [58] LIU Mingyu, TUZEL O. Coupled generative adversarial networks [C]//Proceedings of International Conference on Neural Information Processing Systems. Cambridge, USA: MIT Press, 2016: 469-477.
- [59] LEDIG C, THEIS L, HUSZÁR F, et al. Photo-realistic single image super-resolution using a generative adversarial network [EB/OL]. [2018-05-22]. <https://arxiv.org/pdf/1609.04802v1.pdf>.
- [60] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2016: 770-778.
- [61] BOER J F D, CENSE B, PARK B H, et al. Improved signal-to-noise ratio in spectral-domain compared with time-domain optical coherence tomography [J]. Optics Letters, 2003, 28(21): 2067-2069.
- [62] ARDENKJÆR-LARSEN J H, 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.
- [63] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [EB/OL].

- [2018-05-22]. <https://arxiv.org/pdf/1409.1556.pdf>.
- [64] ZHANG Jun, YAN Yong, LADES M. Face recognition: eigenface, elastic matching, and neural nets [J]. *Proceedings of the IEEE*, 1997, 85(9): 1423-1435.
- [65] CHOPRA S, HADSELL R, LECUN Y. Learning a similarity metric discriminatively, with application to face verification [C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Washington D. C., USA: IEEE Press, 2005: 539-546.
- [66] HUANG Rui, ZHANG Shu, LI Tianyu, et al. Beyond face rotation: global and local perception GAN for photorealistic and identity preserving frontal view synthesis [EB/OL]. [2018-05-22]. <https://arxiv.org/pdf/1704.04086.pdf>.
- [67] 姚乃明, 郭清沛, 乔逢春, 等. 基于生成式对抗网络的鲁棒人脸表情识别 [J]. *自动化学报*, 2018, 44(5): 865-877.
- [68] LI Yijun, LIU Sifei, YANG Jimei, et al. Generative face completion [C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Washington D. C., USA: IEEE Press, 2017: 3911-3919.
- [69] KARPATY A, LI Feifei. Deep visual-semantic alignments for generating image descriptions [C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Washington D. C., USA: IEEE Press, 2015: 3128-3137.
- [70] REED S, AKATA Z, YAN Xinchun, et al. Generative adversarial text to image synthesis [C]//*Proceedings of International Conference on International Conference on Machine Learning*. [S.l.]: JMLR.org, 2016: 1060-1069.
- [71] XU Tao, ZHANG Pengchuan, HUANG Qiuyuan, et al. AttnGAN: fine-grained text to image generation with attentional generative adversarial networks [C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Washington D. C., USA: IEEE Press, 2018: 1316-1324.
- [72] CHEN Yang, LAI Yukun, LIU Yongjin. CartoonGAN: generative adversarial networks for photo cartoonization [C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Washington D. C., USA: IEEE Press, 2018: 9465-9474.
- [73] PERARNAU G, WEIJER J V, RADUCANU B, et al. Invertible conditional GANs for image editing [EB/OL]. [2018-05-22]. <https://arxiv.org/pdf/1611.06355.pdf>.
- [74] KOSSAIFI J, TRAN L, PANAGAKIS Y, et al. GAGAN: geometry-aware generative adversarial networks [C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Washington D. C., USA: IEEE Press, 2018: 878-887.
- [75] MA Shuang, FU Jianlong, CHEN Changwei, et al. DA-GAN: instance-level image translation by deep attention generative adversarial networks [C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Washington D. C., USA: IEEE Press, 2018: 5657-5666.
- [76] WANG Zongwei, TANG Xu, LUO Weixin, et al. Face aging with identity-preserved conditional generative adversarial networks [C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Washington D. C., USA: IEEE Press, 2018: 7939-7947.
- [77] CHOI Y, CHOI M, KIM M, et al. StarGAN: unified generative adversarial networks for multi-domain image-to-image translation [C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Washington D. C., USA: IEEE Press, 2018: 8789-8797.
- [78] REED S E, AKATA Z, MOHAN S, et al. Learning what and where to draw [C]//*Proceedings of International Conference on Neural Information Processing Systems*. Cambridge, USA: MIT Press, 2016: 217-225.
- [79] BAI Yancheng, ZHANG Yongqiang, DING Mingli, et al. Finding tiny faces in the wild with generative adversarial network [C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Washington D. C., USA: IEEE Press, 2018: 21-30.
- [80] WANG Jifeng, LI Xiang, HUI Le, et al. Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal [C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Washington D. C., USA: IEEE Press, 2018: 1788-1797.
- [81] QIAN Rui, TAN R T, YANG Wenhan, et al. Attentive generative adversarial network for raindrop removal from a single image [C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Washington D. C., USA: IEEE Press, 2018: 2482-2491.
- [82] CHEN J, CHEN J, CHAO H, et al. Image blind denoising with generative adversarial network based noise modeling [C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Washington D. C., USA: IEEE Press, 2018: 3155-3164.
- [83] LI Ruide, PAN Jinshan, LI Zechao, et al. Single image dehazing via conditional generative adversarial network [C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Washington D. C., USA: IEEE Computer Society, 2018: 8202-8211.
- [84] NGUYEN V, VICENTE T F Y, ZHAO Maozheng, et al. Shadow detection with conditional generative adversarial networks [C]//*Proceedings of IEEE International Conference on Computer Vision*. Washington D. C., USA: IEEE Press, 2017: 4520-4528.
- [85] VONDRICK C, PIRSIIVASH H, TORRALBA A. Generating videos with scene dynamics [C]//*Proceedings of International Conference on Neural Information Processing Systems*. Cambridge, USA: MIT Press, 2016: 613-621.
- [86] BERNHARD K, HUANG Zhiwu, DANDA P P, et al. Improving video generation for multi-functional applications [EB/OL]. [2018-05-22]. <https://arxiv.org/pdf/1711.11453v2.pdf>.

- [87] SAITO M, MATSUMOTO E, SAITO S. Temporal generative adversarial nets with singular value clipping [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA; IEEE Press, 2017: 2830-2839.
- [88] TULYAKOV S, LIU Mingyu, YANG Xiaodong, et al. MoCoGAN: decomposing motion and content for video generation[EB/OL]. [2018-05-22]. <https://arxiv.org/pdf/1707.04993.pdf>.
- [89] 孔德江, 汤斯亮, 吴飞. 时空嵌入式生成对抗网络的地点预测方法[J]. 模式识别与人工智能, 2018, 31(1): 49-60.
- [90] XIONG Wei, LUO Wenhan, MA Lin, et al. Learning to generate time-lapse videos using multi-stage dynamic generative adversarial networks [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA; IEEE Press, 2018: 2364-2373.
- [91] GUPTA A, JOHNSON J, LI Feifei, et al. Social GAN: socially acceptable trajectories with generative adversarial networks [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA; IEEE Press, 2018: 2255-2264.
- [92] WANG Weiyue, HUANG Qiangui, YOU Suya, et al. Shape inpainting using 3D generative adversarial network and recurrent convolutional networks [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA; IEEE Press, 2017: 2298-2306.
- [93] WU Jiajun, ZHANG Chengkai, XUE Tianfan, et al. Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling [C]//Proceedings of International Conference on Neural Information Processing Systems. Cambridge, USA; MIT Press, 2016: 82-90.
- [94] SAGE A, AGUSTSSON E, TIMOFTE R, et al. Logo synthesis and manipulation with clustered generative adversarial networks [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA; IEEE Press, 2018: 5879-5888.
- [95] DOLHANSKY B, FERRER C C. Eye in-painting with exemplar generative adversarial networks [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA; IEEE Press, 2018: 5657-5666.
- [96] HAUSMAN K, CHEBOTAR Y, SCHAAL S, et al. Multi-modal imitation learning from unstructured demonstrations using generative adversarial nets[EB/OL]. [2018-05-22]. <https://arxiv.org/pdf/1705.10479.pdf>.
- [97] BARAM N, ANSCHEL O, CASPI I, et al. End-to-end differentiable adversarial imitation learning [C]//Proceedings of International Conference on International Conference on Machine Learning. [S. l.]: JMLR. org, 2017: 390-399.
- [98] ZHANG Yizhe, GAN Zhe, FAN Kai, et al. Adversarial feature matching for text generation [C]//Proceedings of International Conference on International Conference on Machine Learning. [S. l.]: JMLR. org, 2017: 4006-4015.
- [99] MAKHZANI A, SHLENS J, JAITLY N, et al. Adversarial autoencoders[EB/OL]. [2018-05-22]. <https://arxiv.org/pdf/1511.05644.pdf>.

编辑 赵 辉