文章编号：0253-374X(2020)04-0591-11

DOI：10.11908/j.issn.0253-374x.19204

生成式对抗网络的应用综述

叶　晨，关　玮

（同济大学 嵌入式系统与服务计算教育部重点实验室，上海 201804）

摘要：生成式对抗网络(GAN)是一种优秀的生成式模型，能够不依赖任何先验假设，学习到高维复杂的数据分布。这一强大的性能使得它成为近年来研究的热点，并在诸多应用领域取得了显著的研究成果。首先介绍了生成式对抗网络的基本原理，各种目标函数以及常用的模型结构。然后，详细分析了生成式对抗网络在条件限制下生成图片的各种演进方法。此外，介绍了生成式对抗网络在不同领域的应用，包括高分辨率图像生成、小目标检测、非图像数据生成、医学图像分割等方面的最新研究进展。最后，总结了生成式对抗网络训练过程中的优化技巧。旨在通俗地阐明GAN的基础理论以及发展历程，并从应用角度对未来工作进行了展望。

关键词：生成式对抗网络；条件生成模型；图像生成

中图分类号：TP181　　**文献标志码**：A

A Review of Application of Generative Adversarial Networks

YE Chen, *GUAN Wei*

(Key Laboratory of Embedded System and Service Computing of the Ministry of Education, Tongji University, Shanghai 201804, China)

Abstract: Generative adversarial networks (GAN) is an excellent generative model, which can learn high-dimensional and complex real data distribution without relying on any prior assumptions. This powerful performance makes it a research hotspot in recent years, and remarkable progress has been made in research in many application fields. In this paper, the basic principle of the GAN, various objective functions and common model structures are introduced. Then, the evolutional methods for generating images under the constraints of conditional generative adversarial networks are analyzed in detail. After that, the applications of the GAN in different fields are introduced, including high-resolution image generation, small target detection, non-image data generation, medical image segmentation and so on. Finally, the optimization techniques in the training process of the GAN are summarized. The purpose of this paper is to elucidate the basic theory and development history of GAN, and to forcast the future work from the perspective of application.

Key words: generative adversarial networks(GAN); conditional generative model; image generation

1　GAN基本介绍

生成式对抗网络[1](generative adversarial networks)是一种类似于对抗博弈游戏的训练网络。该训练网络由两个神经网络组成，一个称为生成器，一个称为判别器。生成器尝试生成能够欺骗判别器的虚假样本，而判别器尽可能去判别样本是真实样本，还是生成器生成的虚假样本。就好像假钞专家与造假钞者的博弈。二者在对抗训练下不断优化，最终达到纳什平衡。该网络的目标函数为

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

式中：z是服从高斯分布的随机噪声；G代表生成器；D代表判别器；$P_{\text{data}}(x)$代表真实数据的概率分布；$P_z(x)$代表随机噪声的概率分布；$x\sim P_{\text{data}}$表示从真实数据的分布中随机抽取x；$z\sim P_z$表示从高斯分布的随机噪声中抽取噪声z；$D(x)$和$G(z)$均表示判别器和生成器在接收括号内输入后所输出的向量。对于生成器G来说，通过随机噪声z作为输入，生成器G

收稿日期：2019-05-20

基金项目：国家自然科学基金重点支持项目(U1764261)；同济大学中央高校基本科研业务费专项资金学科交叉类项目(22120180111，22120190200)

第一作者：叶　晨(1980—)，男，工学博士，高级工程师，主要研究方向为智能算法及其应用技术。
E-mail：yechen@tongji.edu.cn

通信作者：关　玮(1994—)，男，硕士生，主要研究方向为计算机视觉、机器学习、深度学习。
E-mail：1833024@tongji.edu.cn

论文拓展介绍

期望自己生成的样本尽可能地欺骗判别器D,所以需要最大化判别概率$D(G(z))$,于是对于生成器G,它的目标函数是最小化$\log(1-D(G(z)))$。对于判别器D,为了尽可能地区分真实样本和虚假的生成样本,它希望最小化判别概率$D(G(z))$的同时,最大化判别概率$D(x)$,其中x是真实样本。于是判别器的目标函数是最大化$\log D(x)+\log(1-D(G(z)))$(可见生成器$G$的目标函数与真实样本无关)。在训练的开始阶段,因为生成器G产生的样本往往较差,判别器D能绝对自信地否定样本,导致梯度较小,所以训练初期一般将目标函数$\log(1-D(G(z)))$改为$\log(D(G(z)))$。

理论证明方面,当生成器固定时,对$V(G,D)$求导,可以得到最优判别器$D^*(x)$

$$D^*(x)=\frac{P_G(x)}{P_G(x)+P_{\text{data}}(x)} \tag{2}$$

式中:$P_G(x)$代表生成器构造的概率分布。把最优判别器代入目标函数,得到生成器G的目标函数等价于优化$P_{\text{data}}(x)$, $P_G(x)$的JS(Jensen-Shannon)散度。而且只有当D是最优判别器时,G的目标函数才等同于JS散度,所以文献[1]提出应该减少更新G参数的次数,多次更新D的参数。可以证明,当训练样本足够多时,模型会收敛,$P_G(x)=P_{\text{data}}(x)$,二者达到纳什均衡。此时判别器$D$对真实样本还是生成样本的判别概率均为1/2,样本达到了难以区分的程度。

2 目标函数的优化

文献[1]提出GAN网络时,对生成器的目标是重构真实分布,使得生成器产生的数据分布与真实分布越接近越好。上文提到,在已达最优判别器的情况下,生成器的优化公式恰好代表了$P_G(x)$与$P_{\text{data}}(x)$之间的JS散度。于是最小化生成器目标函数的任务实质上转变为最小化两个分布间的JS散度。因为只有在判别器是最优的情况下,生成器的目标函数才表示两个分布间的JS散度,所以训练中往往迭代更新多次判别器,再更新一次生成器。这种做法是为了避免在更新生成器之后,$V(G,D)$(式(1))的函数域改变,可能导致再找到的新的判别器反而使得JS散度升高,因此,不要过于频繁地更新生成器。

2.1 *f*-divergence

文献[1]中使用JS散度来度量两个分布之间的距离有一个明显的局限性,就是JS散度有其自身的函数域,从宏观理解上来说,也许会导致$P_G(x)$和$P_{\text{data}}(x)$的分布的可能取值域缺乏重叠,或者$P_G(x)$成为了$P_{\text{data}}(x)$分布的一部分(模式崩塌)。文献[2]提出了一个通用的模式f-divergence,来衡量两个分布的距离。

f-divergence定义为

$$D_f(P\|Q)=\int_x q(x)f\left(\frac{p(x)}{q(x)}\right)\mathrm{d}x \tag{3}$$

式中:P、Q为任意两个不同的分布;$p(x)$和$q(x)$代表从P和Q中采样出x的概率。f可以是各种不同的版本,只要满足它是一个凸函数并且$f(1)=0$。

举例来说,当设置$f(x)=x\log x$,f-divergence即为KL Divergence

$$D_f(P\|Q)=\int_x p(x)\log\left(\frac{p(x)}{q(x)}\right)\mathrm{d}x \tag{4}$$

当设置$f(x)=-\log x$,f-divergence即为Reverse KL Divergence

$$D_f(P\|Q)=\int_x q(x)\log\left(\frac{q(x)}{p(x)}\right)\mathrm{d}x \tag{5}$$

f-divergence可以说是对生成式对抗网络模型的统一,对任意满足条件的f都可以构造一个对应的生成式对抗网络。

该方法的提出是为了解决GAN在训练中模式崩塌的问题,避免$P_G(x)$与$P_{\text{data}}(x)$差异过大。但是在实验当中发现,不同的f-divergence对训练结果并没有改善。

2.2 最小二乘GAN

上面提到,通过JS散度来衡量两个分布的差异,从而拉近$P_G(x)$与$P_{\text{data}}(x)$的距离,存在一些问题。首先,对于图像采样的分布来说,$P_G(x)$与$P_{\text{data}}(x)$是高维复杂空间在低维空间的折叠,有时候两种分布的重叠部分甚至可能不存在。其次,想要完全还原两种分布的重叠部分,需要足够多的数据。从数学计算上来说,两种分布,如果在空间中没有折叠,那么不管它们之间是距离无限远,还是非常接近,只是保持了微妙的不相交,它们的JS散度计算出来都是$\log 2$。那么对于二分类判别器,如果两种分布不重叠,判别器始终都判定生成样本为假,梯度完全消失,无法进行优化。

针对判别器的梯度消失的问题,文献[3]提出LSGAN(最小二乘GAN)的方法。

LSGAN是f-divergence中$f(x)=(t-1)^2$时的特殊情况。LSGAN的损失函数$J(D)$定义如下:

$$\min_D J(D)=\min_D\left[\frac{1}{2}E_{x\sim p_{\text{data}}(x)}[D(x)-a]^2+\frac{1}{2}E_{z\sim p_z(z)}[D(G(z))-b]^2\right] \quad (6)$$

因为sigmoid激活函数在中间部分的梯度较大,而极大或极小部分的梯度近乎消失,当判别器深度置信地判定真实图片与生成图片时,始终处于无梯度状态。于是LSGAN提出使用线性的激活函数,这样梯度便不会消失。实验表明,该方法比传统生成式对抗网络生成图像的效果有较为明显的提升。

2.3 Wasserstein GAN

针对JS散度以及f-divergence散度不能充分表征两个分布之间距离的问题,WGAN[4](Wasserstein GAN)提出了一种全新的衡量两个分布之间距离的方法,称为泥土移动距离。把任意两个分布P和Q当作两堆土堆,移动土堆P的土,使得土堆P与土堆Q的分布完全一致,则最少需要移动的泥土量可以表征两种分布的差异。

可以很明显并且直观地看出,通过泥土移动距离来判别两个分布的距离,不会出现只要$P_G(x)$与$P_{\text{data}}(x)$无重叠部分,不论差距多大,JS散度的值始终是log 2这样的情况。

基于评估泥土移动距离,目标函数变为

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

式中:$D\in$1-Lipschitz表示任意函数$\|f(x_1)-f(x_2)\|\leqslant\|x_1-x_2\|$,即限制判别器$D$足够平滑,因变量不随自变量的变化而变化过快。这种情况可以避免当判别器D训练得过好,真实样本的判别趋近于正无穷,生成样本的判别趋近于负无穷,导致判别器D无法收敛。

但是实际操作中很难做到限制$D\in$1-Lipschitz,于是文献[4]提出的方法是限制权重。设置一个权重上限p和权重下限$-p$,如果更新后的权重$w>p$,则$w=p$;如果更新后的$w<-p$,那么$w=-p$。

2.4 Wasserstein GAN-Gradient Penalty

WGAN-GP[5](Wasserstein GAN-Gradient Penalty)提出,虽然控制$D\in$1-Lipschitz很难,但是可以等效地通过控制每一个样本x的梯度的范式$\leqslant$1来限制判别器,从而得到同样的效果,即

$$D\in\text{1-Lipschitz}\leftrightarrow\left\|\nabla_x D(x)\right\|\leqslant 1 \quad (8)$$

于是引入一个梯度惩罚,只要梯度的范式大于1,就会产生损失。目标函数变为

$$V(G,D)\approx\max_{D\in\text{1-Lipschitz}}\{E_{x\sim p_{\text{data}}}[D(x)]-E_{x\sim p_G}[D(x)]-\lambda\int_x\max(0,\left\|\nabla_x D(x)\right\|-1)\mathrm{d}x\} \quad (9)$$

式中:λ表示人为选定的参数;∇_x表示对x求偏导数;$\|\nabla_x D(x)\|$表示对$D(x)$中的x计算范式。但是对所有的x都判别梯度,计算量过大。文献[5]提出对一个惩罚分布内的x进行梯度惩罚的计算。该范围的选择方法为从$P_{\text{data}}(x)$任选一个点,再从$P_G(x)$任选一个点,把两个点相连,从连线上任取一点,就属于惩罚分布。直觉上可以认为这样选点的意义是连线上的点能够影响$P_G(x)$如何移动到$P_{\text{data}}(x)$。

在实验中发现,$\|\nabla_x D(x)\|$并不是越接近0效果越好,而是越接近1训练效果越好。于是目标函数变为

$$V(G,D)\approx\max_{D\in\text{1-Lipschitz}}\{E_{x\sim p_{\text{data}}}[D(x)]-E_{x\sim p_G}[D(x)]-\lambda E_{x\sim p_{\text{penalty}}}[\left\|\nabla_x D(x)\right\|-1]^2\} \quad (10)$$

式中:$x\sim P_{\text{penalty}}$表示上文提到的从$P_{\text{data}}(x)$一个点和$P_G(x)$一个点的连线上取中间位置的x。实验结果表明,该方法可以显著提高生成图片的效果。

3 常见的GAN模型

传统GAN在目标函数上的优势明显,但是训练困难,效果也不佳。于是人们希望能够将GAN和传统神经网络相结合,生成更真实效果更好的图片。本节列举一些常见的GAN变体。

3.1 条件GAN

由于传统GAN的生成图片除了真实与否完全不受控制,所以条件控制下的GAN(cGAN)应运而生。生成器同时接收随机噪声z和条件变量c作为输入。判别器在接收真实图片x或者虚假图片x'的同时也同时接收条件变量c。下文将会详细介绍这一类GAN的变体。结构图如图1所示。其中x'代表生成器G输出的虚假图片。

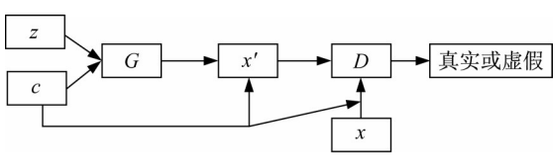

图1　条件GAN结构示意图

Fig.1　Structure of conditional GAN

3.2 增加分类器的GAN

增加一个分类器的GAN,又称为辅助GAN(ACGAN)[6]。在判别器D之外,增加一个分类器C来对图片类别标签进行分类,可以辅助训练。该分类器可以是预训练好的分类器。结构图如图2所示。图中,C代表分类器,c'代表分类器的输出。

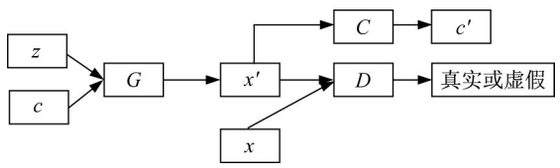

图2 辅助GAN结构示意图

Fig.2 Structure of auxiliary GAN

3.3 与自动编码解码器结合的GAN

自动编码器与解码器是很常见的生成模型训练方法。解码器的功能类似于生成器G,而编码器编码出的向量似乎与隐空间特征有着千丝万缕的联系,常用于特征提取/降维。BiGAN(bidirectional GAN)提出将额外的判别器D加入编码解码器结构中,实现隐空间映射,优化生成结果[7]。真实图片x通过编码器E_n编码为特征向量,随机噪声z通过解码器D_e解码为生成图片,然后把成对数据同时送入判别器D来判断数据对是来自编码器E_n还是解码器D_e。结构图如图3所示。图中,$E_n(x)$和$D_e(z)$分别代表编码器和解码器的输出。

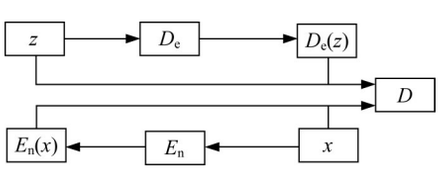

图3 BiGAN结构示意图

Fig.3 Structure of BiGAN

3.4 与变分自动编码器结合的GAN

变分自动编码器(VAE)是自动编码器的一种变体,通过在编码器E_n的输出中加入随机噪声z,再将编码器学习到的方差信息σ的指数输出与z作点乘,加上编码器的输出m后送入解码器D_e,从而削弱输出之间的关联[8]。结构图如图4所示。图中,exp代表取指数,u代表经过一系列运算后传到解码器D_e的输入。

VAE在重构图片时遇到的常见问题是平均了同类别图片的各种不同形态,导致图片模糊。于是VAE-GAN提出将判别器D加入到网络结构中,从而改善图片质量[9]。结构图如图5所示,其中$z\sim E_n(x)$表示噪声z是来自编码器E_n的输出。

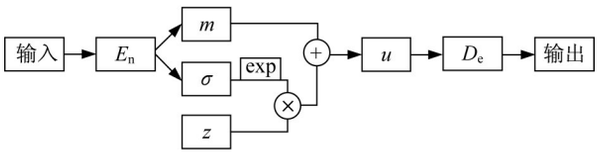

图4 VAE结构示意图

Fig.4 Structure of VAE

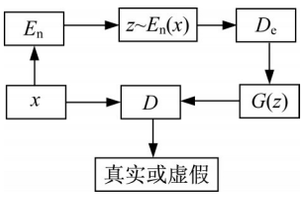

图5 VAE-GAN结构示意图

Fig.5 Structure of VAE-GAN

3.5 深度卷积GAN架构

深度卷积GAN(DCGAN)是一种常见的训练GAN的神经网络架构[10]。该网络自提出后受到大范围的推崇。该网络主要提出了五点更新:①不使用池化层,用步幅卷积层代替;②在生成模型和判别模型时都使用BatchNormalization[11];③不使用全连接层;④生成网络的激活函数除了输出层使用Tanh以外,都使用ReLU;⑤判别网络的激活函数都使用LeakyReLU。该网络对于稳定GAN的训练有许多帮助,实验中往往能取得较为优异的表现。

文献[10]中通过实验来测试DCGAN在特征提取上的能力。在CIFAR-10数据集上,广泛使用的一个基于K-means方法的单层特征提取基准,采用4 800维的特征映射,能够获得80.6%的准确率。该基准方法的多层非监督扩展能够达到82%的准确率。DCGAN只使用512维的特征映射,能够达到82.8%的准确率,超过了所有基于K-means的方法。实验结果见表1[10]。其中可支持向量机简写为SVM。

表1 CIFAR-10分类准确率及特征维度对比

Tab.1 Comparison of accuracy and feature of CIFAR-10

模型	准确率/%	最大特征维度
1层K-means	80.6	4 800
3层K-means+强化学习	82.0	3 200
视图不变K-means	81.9	6 400
DCGAN+L2-SVM	82.8	512

文献[10]还测试了DCGAN在街景房屋数字(SVHN)数据集上对1 000类样本的分类准确率,实验结果表明,DCGAN的表现超过所有现存方法。实验结果见表2[10]。表中,将Stack What-Where

Auto-Encoders模型简称为SWWAE。

表2　SVHN数据集上1 000个类别的分类准确率

Tab.2　SVHN classification accuracy with 1 000 labels

模型	错误率/%
KNN	77.93
TSVM	66.55
SWWAE(不包含dropout)	27.83
SWWAE(包含dropout)	23.56
DCGAN + L2-SVM	22.48

4　条件控制的GAN

传统的GAN网络对生成的模式没有限制,无法控制输出与输入的关系,得不到严格限制条件下的生成样本。通过增加限制条件,可以得到想要的结果。

4.1　标签控制的GAN

在早期的监督学习中,标签控制生成的图片有一个显著的问题,就是相同的物体,可能有不同的远近大小在不同角度的呈现。例如,近距离的动车特写图片和远距离的动车全景图片标签相同。在传统的神经网络训练中,生成图片往往会取相同标签不同场景的平均值,于是产生了模糊的图片。这种多模态生成问题可以通过GAN来解决。

在文献[1]中提到GAN也许可以加上限制条件来进行训练。于是文献[12-13]对类别标签限制进行了尝试。cGAN中将随机噪声z与类别标签c一同送入生成器,同时判别器对生成的样本和类别标签c一同进行判别。于是训练集变成了3种数据对:{真实图片,匹配的标签},{真实图片,不匹配的标签},{生成图片,匹配的标签}。其中只有第一种数据对,判别器D应当判定为正确。目标函数为

$$\min_G \max_D V(G,D)=\min_G \max_D E_{x\sim p_{\text{data}}}[\log D(x|y)]+E_{z\sim p_z}[\log(1-D(G(z|y)))] \quad (11)$$

式中:$x|y$表示在y标签限制下取到x;$z|y$表示在y标签限制下取到z。实验结果表明,cGAN的确能够控制生成模式,产生指定类别的图像,但图像的效果并没有得到改善,说明网络的结构还需要更进一步的优化方法。针对生成高分辨率图片的困难,文献[14]提出了一种方法,利用拉普拉斯金字塔结构[15]来训练网络,从而解决大图片生成模糊的问题。这种策略将最初的问题分解为一系列更易于管理的阶段,在每个图片尺度上都分别使用cGAN方法训练一个基于卷积网络的生成模型。该方法称之为Laplacian GAN(LAPGAN)。具体结构如图6所示。

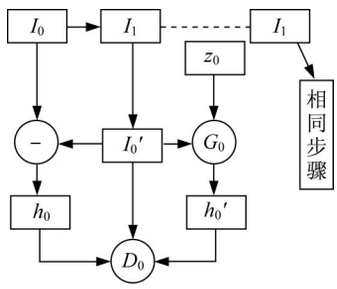

图6　LAPGAN结构示意图

Fig.6　Structure of LAPGAN

训练网络包含一系列的cGAN网络,从64×64的图像I_0开始,每次对图片缩小采样成I_1,再把I_1扩大采样为I_0',得到原始图片相同尺寸的模糊图片。再通过清晰图片和模糊图片得到真实的锐化图片h_0,而模糊图片作为生成器G_0的输入条件,与随机噪声z_0一起生成虚假的锐化图片h_0',让判别器D_0进行判断。缩小的清晰图片I_1送入下一层cGAN进行同样步骤的训练。直到最后一层的图片尺寸变为8×8。该方法能显著改善生成的图片质量。

4.2　文本描述控制的GAN

仅有类别标签限制的样本,并不能完全解决生成问题的多样性需要。人们希望能够仅仅给出语言描述,就产生相关的图片。文献[16]提出通过具体的语言描述,来生成合理的鸟类和花类图片。这类问题可以分解为两个小问题:①学习一个文本特征向量,这个文本特征向量能够捕获重要的视觉细节;②使用这些特征来合成非常真实的图片。将文本编码为特征向量是基于文献[17]中的模型来训练的。给定一组训练数据集$\{(v_n,t_n,y_n):n=1,\cdots,N\}$,文本分类器的损失函数为

$$\frac{1}{N}\sum_{n-1}^{N}\Delta(y_n,f_v(v_n))+\Delta(y_n,f_t(t_n)) \quad (12)$$

式中:Δ表示0-1损失;v_n表示图片;t_n是相对应的描述文本;y_n是类别标签。分类器f_v和f_t的参数化如下:

$$f_v(v)=\arg\max_{y\in Y} E_{t\sim\tau(y)}[\boldsymbol{\alpha}(v)^{\mathrm{T}}\beta(t)] \quad (13)$$

$$f_t(t)=\arg\max_{y\in Y} E_{v\sim\nu(y)}[\boldsymbol{\alpha}(v)^{\mathrm{T}}\beta(t)] \quad (14)$$

式中:$\boldsymbol{\alpha}$是图像编码器;β是文本编码器。对公式(13)来说,给定一个图片v,对v编码为$\boldsymbol{\alpha}(v)$,然后猜测一个文本描述t,这个t是来自类别空间$y\in Y$的某一类别y的某个文本描述,记作$t\sim\tau_{(y)}$。那么通过这个计算,可以知道有某一类别y的文本描述t,可以使

得$\boldsymbol{\alpha}(v)\beta(t)$最大,那么这个$y$就是推测的类别$y$。公式(14)也是同理。最后,使得图像分类器和文本分类器的判断损失最小,获得一个训练好的文本编码器,该文本编码器编码过的文本特征作为条件,传入DCGAN[10]进行训练。

在训练过程中,开创性地区分两种错误来源:不真实的图像与任何文本,以及不匹配的文本的真实图像。即在每个训练步骤中将3种类型的输入馈送到判别器:{真实图像,匹配文本},{真实图像,不匹配文本},{虚假图像,真实文本}。这种训练技术对于生成高质量图像非常重要,因为它不仅告诉模型如何生成逼真的图像,而且还告诉文本和图像之间的对应关系。

在训练文本生成图像的过程中发现,随机噪声z往往与图片风格因素(例如背景色、姿态等)有关。为了获得某种指定风格的图片,可以训练一个风格编码器。风格编码器的损失函数定义为

$$L_{\text{style}}=E_{t,z\sim N(0,1)}\left\|z-S(G(z,\beta(t)))\right\|_2^2 \quad (15)$$

式中:$\beta(t)$代表训练好的文本编码器对t进行编码;S代表风格编码器;损失函数L_{style}最小化随机噪声z和z与文本编码产生的图片经过风格编码器回溯的编码噪声的L_2范数;$t,z\sim N(0,1)$表示t来自于文本样本集,噪声z来自于正态分布。当风格编码器训练好,给定一个指定的图片,想要生成与图片相同风格的另一张图片,只需要编码该图片的风格,作为z传入生成器即可。

通过文本描述只控制生成的图片内容是不够的,还需要能够控制内容生成的具体细节以及具体位置。文献[18]提出一个新模型GAWWN(generative adversarial what-where network),可以控制生成内容和生成位置。该模型把问题分解为确定位置和生成图片两个部分。对于如何确定位置,GAWWN提出的第一种方法是通过空间变换网络学习对象的边界框。空间变换器网络的输出与输入图片大小相同,但是对象边界外的值都置为0。空间变换网络的输出经过几个卷积层将其大小缩减为一维向量,这不仅保留了文本信息,而且还通过边界框提供了对象位置的约束。这种方法的优点是它是端到端的,不需要额外的输入。GAWWN提出的第二种方法是使用指定的关键点来约束图像中对象的不同部分(例如头部、腿部、手臂、尾部等)。为了让关键点包含位置信息,对于每个关键点,生成一个掩码矩阵,其中具有位置信息的关键点置为1,其他为0,该张量放入二进制矩阵中,即1表示存在关键点,0表示不存在需控制的关键点,然后在深度方向上进行复制。虽然这种方法能够对生成的局部特征进行位置约束,但它需要额外的用户输入来指定关键点。实验结果表明,该方法虽然基于文献[16]的架构,但是不仅能指定位置生成图片的内容,还能把清晰图片尺寸从64×64扩展到128×128。不过该方法在人脸生成上效果较差,并且仅适用于具有单个对象的图像。

StackGAN提出,为了模仿人类由粗到精的作图方法,可以使用两个生成器进行文本到图像的合成,而不是只使用一个生成器[19]。第一个生成器由随机噪声和文本描述作为输入,负责生成64×64的粗糙图像,只包含目标图形的一些基础形状和基本颜色,以及背景样式;而第二个生成器获取第一个生成器的输出以及相同的文本描述,来完成整个绘图过程,生成具有更高分辨率和更清晰细节的图像,每个生成器都匹配自己的判别器,并且每个生成器都加入一个KL(Kullback-Leibler)散度参数限制,来使文本特征的分布更接近高斯分布。

StackGAN++在StackGAN的基础上,提出了一个巨有多级生成器的树状结构网络[20]。输入可以看作根节点,而不同的分叉上都能生成不同尺度的图片。在该网络中,监督学习与非监督学习同时进行,并互为补充。

AttnGAN](Attentional GAN)通过在图像和文本特征上使用注意机制进一步扩展了StackGAN++的体系结构[21]。

PPGN(plug & play genarative network)使用了降噪自编码和朗之采样,采用迭代的方法来进行训练[22]。该方法的效果胜于同时期任何的标签限制以及文本限制的图像生成模型。

文本描述控制的GAN有一个共同缺点就是在一个图像中涉及多个复杂对象的情况下,所有现有模型都工作得很糟糕。因为模型只学习图像的整体特征,而不是学习其中每种对象的概念。

4.3 图片控制的GAN

图像到图像的转换就好像一个场景到另一个场景的转换,又或者说是一种风格迁移,比如照片到画家画作的风格迁移。

文献[23]提出将图片作为限制条件的通用GAN训练框架,称为pix2pix。早期的论文研究发现,该损失函数与更传统的损失函数一起训练对训练结果有帮助,例如L_2距离[24],即判别器D的任务保

持不变,但是生成器G的任务不仅是欺骗判别器,而且还要接近L_2意义上的真实输出。该文章使用L_1距离而不是L_2,因为L_1鼓励更少的模糊。在判别网络D方面,设计了一个判别器体系结构,称之为PatchGAN,该网络只能在补丁规模上惩罚结构损失。这个判别器试图分类每个$N\times N$大小的补丁图像是真实的还是假的。该实验可以产生相当优秀的生成样本,尽管添加了Dropout处理噪声[25],但是依然存在网络输出随机性较差的问题。

与此类有监督学习不同的是,无监督学习下的图片风格转换,以不成对的训练集进行训练。CycleGAN[26]、DualGAN[27]以及DiscoGAN[28]可谓有异曲同工之妙。它们的思想类似于一种表现良好的翻译器,可以将中文翻译为英文,该英文再翻译为中文后,和最初的中文无差别。将该思想应用在生成器G上,进行一种循环的双重学习。结构如图7所示。图中,G和F都是生成器,D_X和D_Y都是判别器。

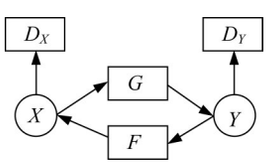

图7　cycleGAN结构示意图

Fig.7　Structure of cycleGAN

以CycleGAN为例,首先要训练两个互为反向的目标函数。对于生成器$G:X\to Y$和它的判别器D_Y,为了满足循环$x\to G(x)\to F(G(x))\approx x$,加上循环一致性损失:

$$L_{\text{cyc}}(G,F)=E_{x\sim p_{\text{data}}(x)}[\|F(G(x))-x\|_1]+E_{y\sim p_{\text{data}}(y)}[\|G(F(y))-y\|_1] \tag{16}$$

在这里,对于来自域X的每个图像x,图像转换周期应该能使x回到原始图像,即$x\to G(x)\to F(G(x))\approx x$,一般称之为前向循环一致性。类似地,对于来自域$Y$的每个图像$y$、$G$和$F$也应该满足后向循环一致性:$y\to F(y)\to G(F(y))\approx y$。所以,完整的训练目标就变成两个对抗损失加上循环一致性损失。

上述模型都旨在实现一对一的场景转换。文献[29]提出一种多场景的通用转换框架,整个网络只使用一对生成器和判别器。生成器除了接收随机噪声外,还需要接收目标领域的信息。判别器除了判别图片是否真实外,还需要判别图片属于哪一个领域。与此同时,为了防止图像翻译过程中内容的改变,需再加上重构损失,以保证内容的完整。

这种无监督学习最明显的缺点就是需要大量的训练数据来学习特征。文献[30]提出通过编码器、解码器来学习特征,从而实现将图片转换到训练阶段从未看过的目标类图像。

上述模型中,pix2pix能够产生最为清晰的图像。

5　GAN在不同领域的应用

由于GAN网络的优点是不需要针对应用的问题去设计专门的损失函数,不需要显式地建模数据分布,所以在图像、文本、语音等诸多领域都有广泛的应用。

5.1　高分辨率图像

高分辨率技术是指从低分辨率图像重建出相应的高分辨率图像,在无法采集到清晰图像时,具有重要应用价值,例如监控设备、医学影像等。然而,传统的基于深度学习的方法是通过卷积网络对原始的图像进行信息采集,生成的高分辨率图像缺乏具体的纹理细节,容易产生模糊。GAN作为一种生成模型,可以巧妙地解决该问题。SRGAN[31](super-resolution GAN)在GAN的原始损失基础上,利用残差网络,加上了感知相似性损失,来生成细节丰富的图像。感知损失重点关注判别器中间层的特征误差,而不是输出结果的逐个像素误差。

高分辨率重建的图像质量往往使用峰值信噪比(PSNR)衡量,PSNR的值越大,图像质量越好,数值大于20 dB符合重建图像标准。文献[31]中采取截断思想进行了实验,在相同网络结构下,通过对抗损失训练的SRGAN产生的图像PSNR达到29 dB以上,在数值上会略低于不采取对抗损失训练的网络产生的图像,然而后者会出现细节模糊,前者则会产生非常细腻的纹理。

5.2　目标检测和变型

图像检测小目标对象存在的问题往往在于小目标对象过低的分辨率,所以直观的解决办法便是将低分辨率图像扩大为高分辨率图像来加强判别能力。文献[32]将判别器分为对抗分支和感知分支。对抗分支负责传统的生成大图像任务,而感知分支负责保证大图像在检测中的效用。SeGAN(segmenting GAN)使用一个分段器、一个生成器和一个鉴别器来复原重构被隐藏对象[33]。

与场景转换相反,对象变形是在背景不变的情况下,用特定条件替换图像中的对象。GeneGAN

(generated GAN)使用了编码解码器结构,编码器将图像分解为背景特征和对象特征,解码器将背景特征和要变形的对象特征重新整合来重构图像[34]。重要的是,为了使特征空间分离,需要两个分离的训练集,一个是具有该对象的图像集,另一个是不具有该对象的图像集。此外,GAN还可以应用于图像混合任务,将一个对象植入另一个图像的背景中。GP-GAN(gaussian-poisson GAN)[35]提出将基于GAN的和传统的基于梯度的图像融合方法相结合。GP-GAN试图通过优化高斯-泊松方程(Gaussian-Poisson equation)[36]生成高分辨率的良好融合图像。

5.3 视频、音乐、语言和语音生成

利用GAN的生成能力,不仅能够生成全新的,具有创造性的视频、音乐或语言类事物,而且还能够通过标签限制、隐空间的条件限制,来对已有的此类事物进行定性或定量的修改、修复和完善,例如视频后期处理、音域修改、人声模仿等,从而大量提升此类产业产品的效率,节省大量人力成本和时间成本。

一般来说,视频由相对静止的背景和动态的运动目标组成。VGAN(video GAN)使用一个两阶段的生成器[37]。3D卷积神经网络生成器负责生成运动前景,2D卷积神经网络生成器负责生成静止的背景。Pose-GAN结合VAE和GAN方法来生成视频,首先,VAE结合当前帧的姿态和过去的姿态特征预测下一帧的运动信息,然后3D卷积神经网络构成的GAN生成后续视频帧[38]。MoCoGAN(motion and content GAN)提出在隐空间对内容部分和运动部分进行分离,使用RNN网络建模运动部分[39]。生成视频内容的一致性往往使用平均内容距离(ACD)衡量。若真实视频的ACD值为最优值0,实验表明,MoCoGAN的ACD平均值能够达到1.79,远低于VGAN的5.02。

在音乐生成方面,C-RNN-GAN(continuous RNN-GAN)[40]将生成器和判别器都建模为一个具有长短时记忆(LSTM[41])的RNN,直接提取音乐的整个序列。但是,音乐这种包括歌词和音符的离散数据,使用GAN生成存在很多问题,缺乏局部一致性。而SeqGAN(sequenceGAN)[42]、ORGAN(object reinforced GAN)[43]则采用了策略梯度算法,不是一次性生成完整的序列。SeqGAN将生成器的输出视为代理的策略,并将判别器的输出作为奖励。就像强化学习一样,生成器选择从判别器那里获得更大的奖励。ORGAN与SeqGAN略有不同,在奖励函数中添加了一个硬编码的目标函数来实现指定的目标。

在语言生成方面,RankGAN[44]用语句生成器和排序器代替传统判别器。类似于传统GAN的对抗,语句生成器努力使生成的虚假语句在排序器中获得较高的排序,而排序器努力把真实语句置于较高的排序。因为生成的语句也是离散的,所以运用了类似SeqGAN和ORGAN的梯度策略算法。

在语音生成方面,VAW-GAN(Variational autoencoding Wasserstein GAN)[45]是一种结合了GAN和VAE框架的语音转换系统。编码器处理语音内容z,而解码器在给定目标说话者信息y的条件下生成语音。在国际标准中,统一使用平均主观意见分(MOS)值来评价语音质量。VAW-GAN在验证集和测试集的得分均超过3分,优于基准VAE,展现了更丰富的频率变化以及更清晰的声音。

5.4 医学图像分割

SegAN(segmentation GAN)提出了一个分割者-批评者结构来分割医学图片[46]。与GAN的结构相似,分割者生成预测的分割图片,而批评者判断分割图片是真实的还是分割者生成的。SeqAN在人脑肿瘤分割挑战BraTS 2013数据集上达到了最优方法的分数,并且在BraTS 2015数据集上,获得了切片分和准确率的最高分。DI2IN(deep image-to-image network)通过对抗训练分割3D CT图像,该方法在颈椎、胸腔、腰椎的CT分割准确率上都超过了已有的方法[47]。SCAN(structure correcting adversarial network)使用GAN方法分割X射线图像,在JSRT数据集上对心脏和肺部的识别准确率以及分割准确率均超过了所有现存方法,并且每张测试时间只需0.84 s,超过原基准时间的26 s[48]。

6 GAN训练问题的改善方法

GAN往往存在较为严重的训练问题,在WGAN中进行了较为详细的介绍。许多论文也致力于解决训练问题,下面列举一些改善训练的方法。

6.1 常见的优化方法

有许多在神经网络中常用的优化方法同样适用于GAN的训练。首先是对输入图片进行预处理,将输入图像规范化到$-1\sim1$之间,并且在训练中使用批量标准化。针对梯度消失和模式崩塌问题,需要避免在训练中引入稀疏矩阵。激活函数尽量采用ReLU等变体,同时使用最大值池化层。模式崩塌时,生成的图片过于相似,可以减小每一批次送入判

别器的图片数量。在对训练过程暂不清晰时,梯度下降的优化可优先使用Adam[49]。针对过拟合问题,可以在判别器的输入中添加随机噪声,Dropout的使用也有极大的帮助,亦或是采取6.3节提到的标签平滑方法。而常用的提前终止训练防止过拟合往往并不需要,因为训练过程很难理想化地达到纳什均衡。

6.2　架构选择

能用DCGAN的时候就选用DCGAN。当不能使用DCGAN而且没有稳定的模型时,可以使用混合模型,例如KL+GAN或者VAE+GAN。在架构选择中,宽度往往比深度重要。在使用VAE时,给生成器与判别器增加一些噪声对训练是有帮助的[50]。

6.3　结果的替换

当判别器D的结果只是真实标签1和虚假标签0时,容易产生过于置信的结果。使用标签平滑方法,用0.7～1.0之间的随机值代替标签1,用0～0.3的随机值代替标签0。

除此之外,可以不评估最后输出的真假标量,而是评估判别器D某一层输出的特征和真实图片在该层输出特征的差异,就像评估编码器编码特征的距离。该方法称为特征匹配。特征匹配引入了随机性,可以减轻模式崩塌,也使得判别器更难过拟合。

4.3节提到的PatchGAN也十分值得尝试使用。它将所有补丁范围内的损失求平均作为最终的损失,可以使生成器生成更加锐利清晰的边缘。

6.4　多个GAN网络的堆叠

当单个GAN不足以有效地处理任务时,可以堆叠多个GAN来将问题模块化。例如,FashionGAN使用两个GAN来执行局部图像翻译[51]。LAPGAN中使用了拉普拉斯金字塔结构的GAN。StackGAN思路同样类似。Progressive GAN (ProGAN)[52]可以生成高质量的高分辨率图像。

7　总结与展望

自从2014年Goodfellow等人首次提出GAN网络模型结构,利用该网络精妙的零和博弈思路来处理各式各样的问题,以及解决该网络训练困难的论文,如雨后春笋般涌现。本文旨在通俗地解释GAN的基础理论,整理解释了较为常用的GAN训练模型,以及在时间线上按照理论发展的逻辑整理出条件限制下GAN在图片生成方面的发展,并且较为宏观地归纳GAN在各种不同场景的应用,同时总结了一些改善训练的技巧。笔者相信,这一技术将会大幅度改进深度学习在一些应用场景中存在的局限性,尤其是在生成模型方面,从零到有所谓创造,一直是难以解决的问题,而GAN恰恰是对生成模型绝佳的改进。目前相关应用仍处于起步阶段,在未来将会有更有效、更广泛的尝试。下面对于GAN在应用上的落地,提出若干展望:

(1) GAN在图像生成和风格转换方面真正体现出创造性。随着社会的发展,文化娱乐所承载的生活比重越来越大,而文娱产业最重要的就是快速且创新。如果能够利用GAN在生成方面的分布广度,创造出一些新颖的文娱产品,从而激发人们的创造力,将会大大改善这个产业的生产力。例如,利用条件控制的GAN,通过给定故事背景,自动生成一系列动画作品,而动画作品的风格也许会别具一格。

(2) GAN在图像修复领域的应用。不仅仅是在图像中的目标移除和填补,图像超分辨率分析等方面,利用深度神经网络对于高维复杂映射具有强大的逼近能力,可以有效地提取图像中的语义。结合语义和纹理可以对艺术品的细节修复提供帮助,例如壁画修补,书法拓片的修复等。

(3) GAN在医学图像生成方面的应用。辅助医生做图像数据分析的人工智能越来越常见,这些人工智能需要的训练数据样本越多样越好。可是相比健康的样本,研究人员难以获取足够多的病变图像样本。在类似各种形态(图像、语音、语言等)的数据增广研究中,GAN拥有广阔的发展前景。

参考文献:

[1] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, *et al*. Generative adversarial nets [C]// Advances in Neural Information Processing Systems. Montreal: Curran Associates, 2014: 2672-2680.

[2] NOWOZIN S, CSEKE B, TOMIOKA R. *f*-gan: training generative neural samplers using variational divergence minimization[C]// Advances in Neural Information Processing Systems. Barcelona: Curran Associates, 2016: 271-279.

[3] MAO X, LI Q, XIE H, *et al*. Least squares generative adversarial networks [C]// Proceedings of the IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 2794-2802.

[4] ARJOVSKY M, CHINTALA S, BOTTOU L. Wasserstein generative adversarial networks[C]// International Conference

on Machine Learning. Sydney: ACM, 2017: 214-223.

[5] GULRAJANI I, AHMED F, ARJOVSKY M, *et al*. Improved training of wasserstein gans [C]// Advances in Neural Information Processing Systems. Long Beach: Curran Associates, 2017: 5767-5777.

[6] ODENA A, OLAH C, SHLENS J. Conditional image synthesis with auxiliary classifier gans[C]// Proceedings of the 34th International Conference on Machine Learning. Sydney: ACM, 2017: 2642-2651.

[7] DONAHUE J, KRÄHENBÜHL P, DARRELL T. Adversarial feature learning[EB/OL]. [2018-05-31]. https://arxiv.org/abs/1605.09782.

[8] KINGMA D P, WELLING M. Auto-encoding variational bayes [EB/OL]. [2018-12-20]. https://arxiv. org/abs/1312.6114.

[9] LARSEN A B L, SØNDERBY S K, LAROCHELLE H, *et al*. Autoencoding beyond pixels using a learned similarity metric [C]// Proceedings of the 33rd International Conference on Machine Learning. New York: Curran Associates, 2016: 1558-1566.

[10] RADFORD A, METZ L, CHINTALA S. Unsupervised representation learning with deep convolutional generative adversarial networks [EB/OL]. [2018-11-19]. https://arxiv.org/abs/1511.06434.

[11] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift[C]// Proceedings of the 32nd International Conference on Machine Learning. Lille: Curran Associates, 2015: 448-456.

[12] MIRZA M, OSINDERO S. Conditional generative adversarial nets[EB/OL]. [2015-11-6].https://arxiv.org/abs/1411.1784.

[13] GAUTHIER J. Conditional generative adversarial nets for convolutional face generation [J]. Class Project for Stanford CS231N Convolutional Neural Networks for Visual Recognition, Winter semester, 2014(5): 2.

[14] 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. Montreal: Curran Associates, 2015: 1486-1494.

[15] BURT P, ADELSON E. The Laplacian pyramid as a compact image code[J]. IEEE Transactions on Communications, 1983, 31(4): 532.

[16] REED S, AKATA Z, YAN X, *et al*. Generative adversarial text to image synthesis [C]// Proceedings of the 33rd International Conference on Machine Learning. New York: ACM, 2016: 1060-1069.

[17] REED S, AKATA Z, LEE H, *et al*. Learning deep representations of fine-grained visual descriptions [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 49-58.

[18] REED S E, AKATA Z, MOHAN S, *et al*. Learning what and where to draw [C]// Advances in Neural Information Processing Systems. Barcelona: Curran Associates, 2016: 217-225.

[19] ZHANG H, XU T, LI H, *et al*. Stackgan: text to photo-realistic image synthesis with stacked generative adversarial networks [C]// Proceedings of the IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 5907-5915.

[20] HAN Z, TAO X, HONGSHENG L, *et al*. StackGAN++: realistic image synthesis with stacked generative adversarial networks [EB/OL]. [2017-10-19]. https://arxiv.org/abs/1710.10916.

[21] XU T, ZHANG P, HUANG Q, *et al*. Attngan: fine-grained text to image generation with attentional generative adversarial networks [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Utah: IEEE, 2018: 1316-1324.

[22] NGUYEN A, CLUNE J, BENGIO Y, *et al*. Plug & play generative networks: conditional iterative generation of images in latent space[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hawaii: IEEE, 2017: 4467-4477.

[23] ISOLA P, ZHU J Y, ZHOU T, *et al*. Image-to-image translation with conditional adversarial networks [C]// Proceedings of the IEEE conference on computer vision and pattern recognition. Hawaii: IEEE, 2017: 1125-1134.

[24] PATHAK D, KRAHENBUHL P, DONAHUE J, *et al*. Context encoders: feature learning by inpainting [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Nevada: IEEE, 2016: 2536-2544.

[25] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, *et al*. Dropout: a simple way to prevent neural networks from overfitting [J]. The Journal of Machine Learning Research, 2014, 15(1): 1929.

[26] ZHU J Y, PARK T, ISOLA P, *et al*. Unpaired image-to-image translation using cycle-consistent adversarial networks [C]// Proceedings of the IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 2223-2232.

[27] YI Z, ZHANG H, TAN P, *et al*. Dualgan: unsupervised dual learning for image-to-image translation [C]// Proceedings of the IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 2849-2857.

[28] 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: ACM, 2017: 1857-1865.

[29] CHOI Y, CHOI M, KIM M, *et al*. Stargan: unified generative adversarial networks for multi-domain image-to-image translation [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Utah: IEEE, 2018: 8789-8797.

[30] LIU M Y, HUANG X, MALLYA A, *et al*. Few-shot unsupervised image-to-image translation [EB/OL]. [2019-05-

05].https://arxiv.org/abs/1905.01723.

[31] LEDIG C, THEIS L, HUSZÁR F, *et al*. Photo-realistic single image super-resolution using a generative adversarial network [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hawaii: IEEE, 2017: 4681-4690.

[32] LI J, LIANG X, WEI Y, *et al*. Perceptual generative adversarial networks for small object detection [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hawaii: IEEE, 2017: 1222-1230.

[33] EHSANI K, MOTTAGHI R, FARHADI A. Segan: segmenting and generating the invisible [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Utah: IEEE, 2018: 6144-6153.

[34] ZHOU S, XIAO T, YANG Y, *et al*. Genegan: learning object transfiguration and attribute subspace from unpaired data [EB/OL]. [2017-05-14].https://arxiv.org/abs/1705.04932.

[35] WU H, ZHENG S, ZHANG J, *et al*. Gp-gan: towards realistic high-resolution image blending [EB/OL]. [2017-03-21].https://arxiv.org/abs/1703.07195.

[36] PÉREZ P, GANGNET M, BLAKE A. Poisson image editing [J]. ACM Transactions on graphics (TOG), 2003, 22(3): 313.

[37] VONDRICK C, PIRSIAVASH H, TORRALBA A. Generating videos with scene dynamics [C]// Advances In Neural Information Processing Systems. Barcelona: Curran Associates, 2016: 613-621.

[38] WALKER J, MARINO K, GUPTA A, *et al*. The pose knows: video forecasting by generating pose futures [C]// Proceedings of the IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 3332-3341.

[39] TULYAKOV S, LIU M Y, YANG X, *et al*. Mocogan: decomposing motion and content for video generation [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Utah: IEEE, 2018: 1526-1535.

[40] MOGREN O. C-RNN-GAN: continuous recurrent neural networks with adversarial training [EB/OL]. [2018-11-29]. https://arxiv.org/abs/1611.09904.

[41] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural computation, 1997, 9(8): 1735.

[42] YU L, ZHANG W, WANG J, *et al*. Seqgan: sequence generative adversarial nets with policy gradient [EB/OL]. [2018-09-22].https://arxiv.org/abs/1609.05473.

[43] GUIMARAES G L, SANCHEZ-LENGELING B, OUTEIRAL C, *et al*. Objective-reinforced generative adversarial networks (ORGAN) for sequence generation models[EB/OL]. [2019-03-30].https://arxiv.org/abs/1705.10843.

[44] LIN K, LI D, HE X, *et al*. Adversarial ranking for language generation [C]// Advances in Neural Information Processing Systems. Long Beach: Curran Associates, 2017: 3155-3165.

[45] HSU C C, HWANG H T, WU Y C, *et al*. Voice conversion from unaligned corpora using variational autoencoding wasserstein generative adversarial networks [EB/OL]. [2019-04-04].https://arxiv.org/abs/1704.00849.

[46] XUE Y, XU T, ZHANG H, *et al*. Segan: adversarial network with multi-scale L1 loss for medical image segmentation[J]. Neuroinformatics, 2018, 16(3/4): 383.

[47] YANG D, XIONG T, XU D, *et al*. Automatic vertebra labeling in large-scale 3D CT using deep image-to-image network with message passing and sparsity regularization[C]// International Conference on Information Processing in Medical Imaging. Boone: Springer, 2017: 633-644.

[48] DAI W, DOYLE J, LIANG X, *et al*. Scan: structure correcting adversarial network for chest x-rays organ segmentation [EB/OL]. [2019-03-26]. https://arxiv. org/abs/1703.08770.

[49] KINGMA D P, BA J. Adam: a method for stochastic optimization [EB/OL]. [2018-12-22]. https://arxiv. org/abs/1412.6980.

[50] ZHAO J, MATHIEU M, LE CUN Y. Energy-based generative adversarial network [EB/OL]. [2018-11-11]. https://arxiv.org/abs/1609.03126.

[51] CUI Y R, LIU Q, GAO C Y, *et al*. FashionGAN: display your fashion design using conditional generative adversarial nets [J]. Computer Graphics Forum, 2018, 37(7): 109.

[52] KARRAS T, AILA T, LAINE S, *et al*. Progressive growing of gans for improved quality, stability, and variation[EB/OL]. [2019-03-27].https://arxiv.org/abs/1710.10196.