◎生成对抗网络专题◎

生成式对抗网络研究综述

孙书魁[1,2],范 菁[1,2],曲金帅[1,2],路佩东[1,2]
1.云南民族大学 电气信息工程学院,昆明 650000
2.云南民族大学 云南省高校信息与通信安全灾备重点实验室,昆明 650000

摘 要:生成式对抗网络(GAN)凭借其强大的对抗学习能力受到越来越多研究者的青睐,并在诸多领域内展现出巨大的潜力。阐述了GAN的发展背景、架构、目标函数,分析了训练过程中出现模式崩溃和梯度消失的原因,并详细介绍了通过架构变化和目标函数修改而提出GAN衍生模型,对一些用来评估生成图像质量和多样性的标准进行了小结,总结了GAN在不同领域的广泛应用,总结全文并对该领域未来的研究方向提出一些展望。

关键词:机器学习;生成式对抗网络;图像生成;无监督学习
文献标志码:A **中图分类号:**TP181;TP183 doi:10.3778/j.issn.1002-8331.2205-0097

Survey of Generative Adversarial Networks

SUN Shukui[1,2], FAN Jing[1,2], QU Jinshuai[1,2], LU Peidong[1,2]
1.School of Electrical and Information Technology, Yunnan Minzu University, Kunming 650000, China
2.University Laboratory of Information and Communication on Security Backup and Recovery in Yunnan Province, Yunnan Minzu University, Kunming 650000, China

Abstract: With its strong adversary learning ability, generative adversarial networks(GAN) is favored by more and more researchers in many fields. This paper expounds the development background, framework and objective function of GAN, analyzes the causes of pattern collapse and gradient disappearance in the training process, and introduces in detail the GAN derived model proposed through the change of architecture and the modification of objective function. Then, it summarizes some standards used to evaluate the quality and diversity of generated images, and summarizes the wide application of GAN in different fields, Finally, this paper summarizes and puts forward some prospects for the future research direction in this field.

Key words: machine learning; generative adversarial networks; image generation; unsupervised learning

近年来,随着不同行业领域海量数据的涌现以及硬件设备的算力不断增强,人工智能的身影开始出现在各个领域。其中机器学习是人工智能的核心应用,它关注的是计算机学习能力所依据的程序和算法的改进与优化。根据有无监督,机器学习[1]分为监督学习和非监督学习。在监督学习中,人工标记的数据即昂贵又耗时;另外,自动收集数据也较繁杂。在深度学习中,解决这个问题的关键技术之一是数据扩充,将这种方法应用于模型可以提高模型的能力,并减少泛化误差。通过对图像进行旋转、裁剪、缩放和其他简单变换等操作,创造出新的、可接受训练的样本集,从而实现数据扩充。然而,使用这种方法获得的数据是有限的。最先进的数据扩充方式是通过生成模型生成高质量的样本。考虑到生成模型具有生成大规模数据的能力,标签数据短缺的问题将得到大幅度缓解。

生成模型通常基于马尔科夫链、最大似然估计和近似推理。受限玻尔兹曼机[2]及其开发模型、深度信念网络[3]以及深度玻尔兹曼机[4]都是基于最大似然估计的,这些模型存在一些严重的缺陷,泛化能力不强。为解决这些缺陷,Goodfellow等人在2004年提出生成式对抗网络(generative adversarial networks,GAN)[5]。自首次提出,GAN就被研究学者誉为“深度学习中最重要的创新

基金项目:国家自然科学基金(61540063);云南省教育厅项目(2020J0655)。
作者简介:孙书魁(1996—),男,硕士研究生,CCF会员,研究方向为计算机视觉,E-mail:shukuisun@163.com;范菁(1976—),通信作者,女,博士,教授,CCF会员,研究方向为机器学习和计算机视觉。
收稿日期:2022-05-06 **修回日期:**2022-06-20 **文章编号:**1002-8331(2022)18-0090-14

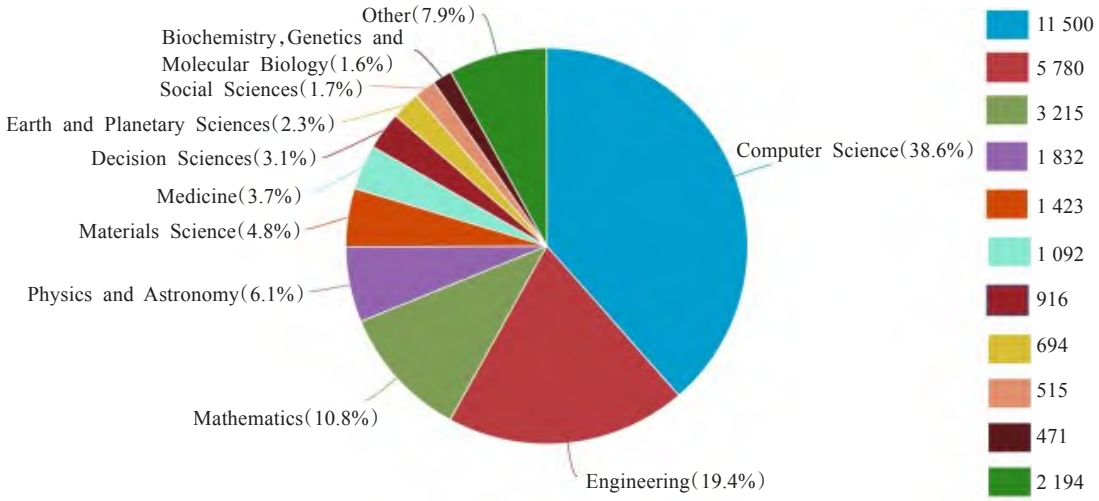


图1 Scopus上GAN论文所属Top 10学科领域

Fig.1 Top 10 disciplines of GAN papers on Scopus

之一”。作为深度学习领域领军人物的LeCun曾表示,GAN及其变体是“过去20年来深度学习中最酷的想法”。

如今,GAN已经渗透到各个领域,例如视频语音、计算机视觉以及诸如医学物理等学科领域。图1显示了Scoups上至2021年以来GAN在不同学科领域所发表论文数量。图2表示近年来Scoups上GAN论文的数量变化趋势。从这些数据可以得知GAN是人工智能中不可多得的技术,并且有非常大的应用前景等待人们去发掘。

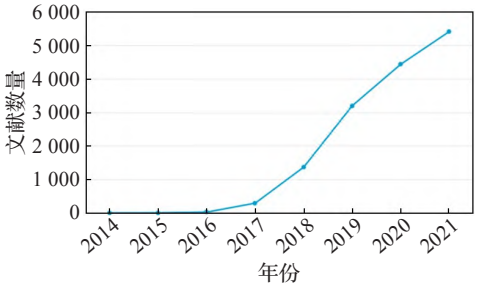


图2 Scopus上GAN论文数量的变化趋势

Fig.2 Change trend of GAN papers on Scopus

本文阐述了GAN在架构、训练、目标函数、面临挑战和评估指标等方面的最新研究进展,然后对GAN在图像、序列数据和半监督学习等领域中的应用进行了梳理,最后进行了总结并对其下一步研究方向进行展望。

1 生成式对抗网络

主要介绍了GAN的网络架构、训练过程以及目标函数,并且讨论了GAN如今所面临的挑战。

1.1 GAN网络架构

GAN是一种类似于二人博弈的网络模型,该模型由生成网络 G 和判别网络 D 组成。GAN的架构如图3所示。X_{data} 和 $G(z)$ 分别表示真实数据样本和生成器 G 生成的伪数据样本,判别器 D 判断输入数据的真伪。在GAN中,生成器 G 以固定长度的随机噪声向量 z(均匀分布或高斯分布)作为输入,生成器的目的是尽量使生成数据分布近似于真实数据分布。鉴别器 D 的输入有两部分:X_{data} 和 $G(z)$,其输出为概率值,表示 D 认为输入数据是真实样本的概率,同时输出会反馈给 G,用于指导 G 的训练。理想情况下 D 无法判别输入数据是来自真实数据还是生成数据,即 D 每次的输出概率值都是1/2,此时模型达到最优[6]。

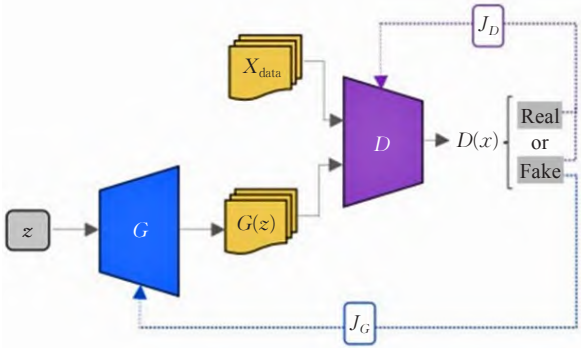


图3 GAN架构

Fig.3 Architecture of GAN

1.2 GAN训练过程

GAN是一组非常复杂且富有挑战性的网络,因为生成和判别网络是以对抗方式同时进行训练。GAN的核心是两个网络之间的平衡。图4显示了GAN的训练过程。在图4(a)中,通过更新判别分布(蓝色虚线)使其能够区分输入是来自真实分布(黑色虚线)还是生成数据分布(绿色实线)。在图4(b)中,判别器经过训练可以区分真假数据,并且很容易完成任务。在图4(c)中,固定判别器,只训练生成器,使其生成假数据的分布更接近真实数据分布。更新一直持续到判别器无法区分为止(图4(d))[7]。值得注意的是,训练过程并不像图4所

示的这么简单。理想情况下,假数据分布与真实的数据分布完全重叠,但实践中存在各种挑战[8]。

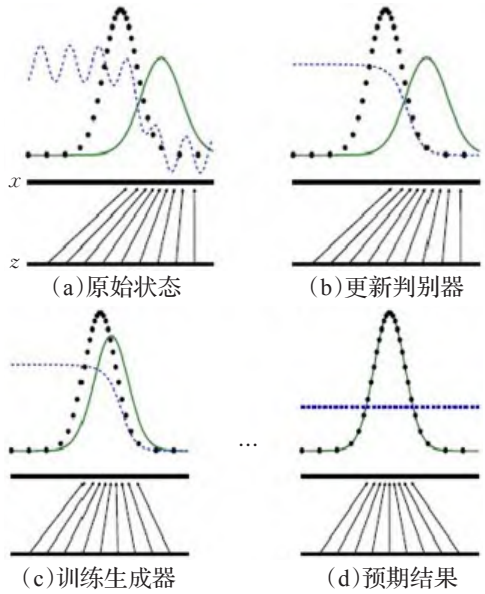


图4　GAN训练过程

Fig.4　Training process of GAN

1.3　目标函数

原始GAN使用两个目标函数:(1) D 最小化二元分类的负对数似然;(2) G 最大化生成样本为真实的概率。D 参数为 θ_D,G 参数为 θ_G,θ_G 和 θ_D 分别最小化和最大化目标函数,因此这是一场零和博弈。等式(1)中的 $p_{\text{data}}(x)$ 和 $p_z(z)$ 分别表示数据空间 x 中定义的真实数据概率和隐藏空间定义的 z 的概率分布。

$$\min_{\theta_G}\max_{\theta_D}V(G,D)=\min_{\theta_G}\max_{\theta_D}E_{x\sim p_{\text{data}}(x)}[\ln D(D)]+E_{z\sim p_z(z)}[\ln(1-D(G(z)))] \tag{1}$$

$V(G,D)$ 为二元交叉熵函数,常用于二元分类问题,可以注意到 G 将 z 从 Z 中映射到 X 的元素中,而 D 接收输入 X,并判断 X 是真实数据还是由 G 生成的假数据。为了更新各自的模型,G 和 D 的训练是通过它们各自的模型执行反向传播来实现的。从 D 的角度来看,如果样本来自真实数据样本,D 将其最大化,若样本来自生成器生成的样本数据,D 则将其最小化输出。与此同时,G 试图欺骗 D,所以当一个假样本输入 D 时,它试图最大化 D 的输出[9]。因此,D 试图最大化 $V(G,D)$,而 G 试图最小化 $V(G,D)$,从而形成了等式(1)的极大极小关系。等式(1)是通过交替执行两个梯度更新来求解的:

$$\theta_D^{t+1}=\theta_D^t+\lambda^t\nabla_{\theta_D}V(D^t,G^t)$$

$$\theta_G^{t+1}=\theta_G^t+\lambda^t\nabla_{\theta G}V(D^{t+1},G^t)$$

λ 是学习率,t 是迭代次数。实际上,在等式(1)的第二项 $\ln(1-D(G(z))$ 梯度饱和并使梯度不饱和的流向 G 时,即梯度变小,停止学习。为了克服梯度消失,可以将等式(1)的目标函数分解成等式(2):

$$\begin{cases}\max\limits_{\theta_D}E_{x\sim p_{\text{data}}(x)}[\ln D(x)]+E_{z\sim p_z(z)}[\ln(1-D(G(z)))]\\ \min\limits_{\theta_G}E_{z\sim p_z(z)}[\ln(D(G(z)))]\end{cases} \tag{2}$$

这两个单独目标函数的 G 梯度具有相同的固定点,且总是在相同的方向上训练。在等式(2)中的损失计算出来后,利用反向传播更新参数。若有充足训练,G 能够将简单的隐式分布 p_g 转换为更复杂的分布,即 p_g 收敛于 p_{data}。

1.4　GAN面临的挑战

原始GAN生成的样本缺乏多样性,生成器在存在多个可能输出类别的情况下,却一直生成单一类别输出[10],即模式崩溃。模式崩溃是GAN训练过程中常见的问题,其原因和解决方法尚未被完全理解。GAN在训练过程中 G 和 D 也会发生振荡,而不是固定点收敛。当一个玩家比另一个玩家强大的时候,网络可能无法学习并受梯度消失影响。在本节中,将讨论GAN训练过程中所面临的挑战。

1.4.1　模式崩塌

由于GAN的max-min的解决方案与min-max的解决方案的工作方式不同,所以可能会导致模式崩溃。因此,在 $G^*=\min\limits_{\theta_G}\max\limits_{\theta_D}V(G,D)$ 中,G^* 从数据分布中生成样本。如果 $G^*=\max\limits_{\theta_G}\min\limits_{\theta_D}V(G,D)$,$G$ 将每个 z 值映射到 D 认为它们是真实的单个 x 坐标。同时梯度下降并没有明显地使min-max优于max-min。

通常来说,模式崩溃[11]是泛化能力差的结果,这类崩溃大致可以分为两种:(1)输入数据中的大部分模式在生成的数据中不存在;(2) G 只学到了特定模式的子集,对于一些修改 D 的目标函数[12-13]和修改 G 的目标函数[14]的GAN的变体而言,不合适的目标函数也可能会导致模式崩溃。在这些变体中,G 处于平衡状并且能够学习整个数据分布,但在实际中收敛常是难以捉摸的。为了解决这个问题,最近几项研究引入了具有新目标函数或替代训练方案的新型网络架构。图5显示了GAN在玩具集上的模式崩溃,其中目标分布是二维空间中的高斯分布。

1.4.2　非收敛性和不稳定性

在原始GAN中,G 使用的两个损失函数是 $E_z[\ln D(G(z))]$ 和 $E_z[\ln(1-D(G(z)))]$。当 D 可以轻松区分真假样本时,前一个损失函数 $E_z[\ln D(G(z))]$ 可能是梯度消失问题的原因。对于最优 D,G 损失最小化类似于真实图像分布和生成图像分布之间的JSD(Jensen-Shannon divergency,JSD)最小化。在这种情况下,JSD将是2ln2。这允许最优 D 将概率0分配给假样本,将1分配给真实样本,并导致 G 损失函数梯度接近于0,这称为 G 上的梯度消失。图6显示,随着 D 的优化,G 的梯度逐渐消失[15]。

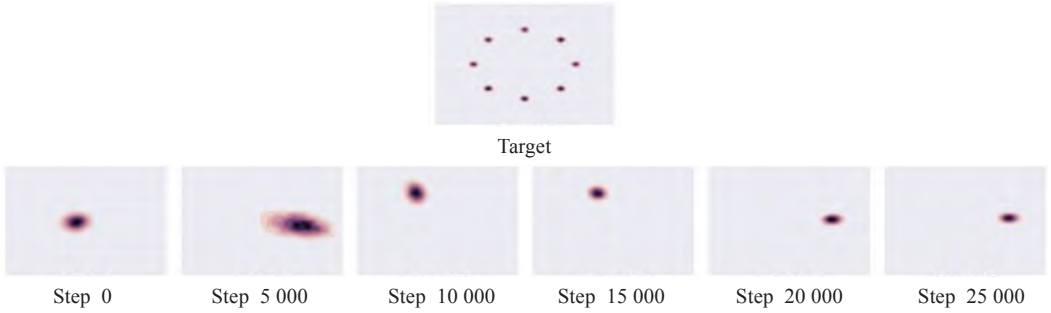


图5 2D玩具数据集上的模式崩溃示例

Fig.5 Example for mode collapse problem on 2D toy dataset

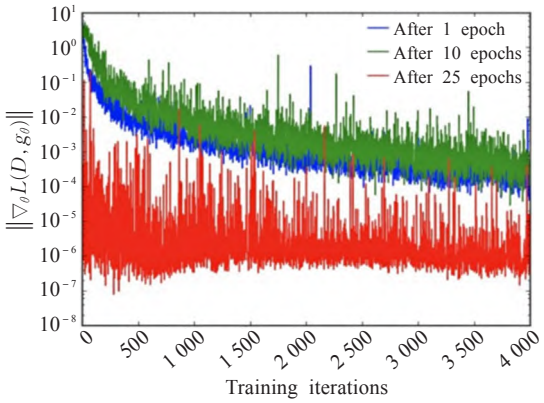


图6 原始代价函数的梯度

Fig.6 Gradient of original cost function

在GAN中,D试图最小化交叉熵,而G则尝试最大化。当D的置信度很高时,D拒绝G生成的样本,然后G的梯度消失。缓解此问题的第一种方案是反转用于构建交叉熵成本的目标。第二个被认为$-\ln D$方式。G的损失函数$E_z[\ln(1-D(G(z)))]$的最小化等于最小化$D_{KL}(p_g\|p_{\text{data}})-2(D_{JS}(p_{\text{data}}\|p_g))$,这会导致梯度不稳定,因为它会同时最小化KL散度(Kullback-Leibler divergence)和最大化JSD。这种情况称为G的梯度更新不稳定。图7显示了G的梯度正在快速增长。且还显示了梯度的方差在增长,更新梯度将导致产生低质量样本[15]。

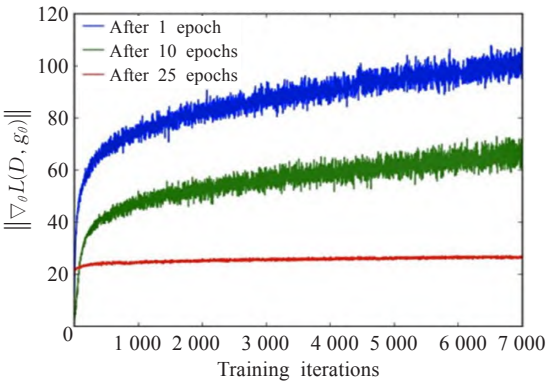


图7 $-\ln D$代价函数的梯度

Fig.7 Gradient of $-\ln D$ cost function

如上所述,为了应对GAN所面临的挑战,文献[16]提出一种新的、鲁棒性更强的算法去寻找二人博弈的纳什均衡,该方法能够在各种结构和散度度量上稳定地训练GAN,并能够达到局部收敛,但全局收敛性还有待研究。文献[17]对如今的训练方法进行分析比较,并指出通过添加一致优化正则项和zero centered gradient可以较好地实现收敛,这为以后的研究指明了方向。

1.4.3 评估指标

GAN模型已被广泛应用于很多的领域,且衍生出较多的变体。但不同模型的评估仍存在大量的分歧。目前虽然已经有评估GAN的性能的措施和方案,但这些评估方案是定性的,而且这些评估耗时、主观且无法捕获分布特征。由于选择合适的模型对于获得良好的应用性能至关重要,因此选择合适的评价指标对于得出正确的结论也至关重要。为了设计更好的GAN模型,需要通过开发或使用适当的定量度量来克服定性度量的局限性。

2 GAN的发展及其衍生模型

2.1 GAN的分类

表1展示一种基于GAN设计和优化方案的新型GAN分类。近年来,基于两种主要技术:重组网络架构和新型损失函数,人们提出了很多解决方案来更好地设计和优化原始GAN。重组网络架构侧重于对原始GAN架构进行重新组合和创新[18-20],新型损失函数涵盖对GAN损失函数的修改和重新设计[21-22]。对于每种技术,相关的研究从未停止且已经提出了相应的解决方案来解决上述GAN所面临的挑战[12]。

2.2 基于架构优化的GAN

生成和判别网络的架构会极大地影响GAN的训练稳定性。如上文讨论,目前有很多学者对架构进行分析和优化,并尝试与其他的模型结合,以结合彼此的优势。这些基于架构优化的模型大致可以分为三类:条件、卷积和自动编码器,现在的绝大多数变体都是基于以上三种经典模型进行开发的。在下文中,将详细介绍这三种经典架构。

2.2.1 基于条件优化的GAN

如上所述,在原始GAN中,把随机噪声向量z输入到G中,G从噪声z中输出一个样本。假设GAN的训

表1　GAN衍生变体分类

Table 1　Classification of GAN derived variants

技术	方案(S)		变体
重组网络架构(S_1)	条件生成(S_{11})		cGAN,FCGAN,IRGAN,GRGANs,LAPGAN,SGGAN,IcGAN,BiCoGAN,MatAN,Self-Conditioned GANs,AC-GANs,TripleGAN
	生成-鉴别网络对(S_{12})	训练单个G($S_{12(i)}$)	DCGAN,ProgressGAN,PacGAN,BayesianGAN,CapsNets,QuGANs,SAGAN
		训练多个G($S_{12(ii)}$)	cGAN,AdaGAN,MAD-GAN,MGAN,MPMGAN,FictitiousGAN,MIX+GAN
		训练多个D($S_{12(iii)}$)	D2GAN,GMAN,StabGAN,Dropout GAN,MicroBatchGAN,SGAN
	联合架构(S_{13})	数据空间自动编码器($S_{13(i)}$)	VAE-GAN,AAE,AVB,ASVAE,MDGAN,Dist-GAN,α-GAN
		潜在空间自动编码器($S_{13(ii)}$)	ALI,BiGAN,DALI,CV-BiGAN,MV-BiGAN,HAlI,AGE,VEEGAN,MGGAN
	改进D(S_{14})		EBGAN,BEGAN,MAGAN,Max-Boost-GAN
	记忆网络(S_{15})		MemoryGAN
	潜在空间工程(S_{16})		DeLiGAN,NEMGAN,MultiplicativeNoise,DeGAN,InfoGAN
新的损失函数(S_2)	新的概率距离和散度(S_{21})		WGAN,LS-GAN,RWGAN,f-GAN,χ^2-GAN,OT-GAN,LSGAN,SoftGAN,GANRL,GoGAN,IGAN,McGAN,MMDGAN,MMGAN
	正则化(S_{22})		WGAN-GP,BWGAN,CT-GAN,SN-GAN,FisherGAN,UnrolledGAN,DRGAN

练集有很多类样本,由于原始GAN对于生成器几乎没有任何约束限制,因此无法控制生成特定类的样本,在生成内容复杂图像的情形时模型会变得更加难以控制。因此,文献[23]提出了条件生成对抗网络(conditional generative adversarial network,cGAN),在原始GAN的基础上添加条件信息有y,通过y用户可以让G具体输出某一类别的数据样本,D则以真实图像和附加信息y作为输入。使网络朝着既定的方向生成样本[24]。cGAN的目标函数是:

$$\min_G \max_D V(D,G)=E_{x\sim p_{\text{data}}(x)}[\ln D(x|y)]+E_{z\sim p_z(z)}[\ln(1-D(G(z|y)))] \tag{3}$$

为实现等式(3),将one-hot编码y连接到x作为判别器的输入,同时将y连接到噪声z作为生成器的输入。因此,判别器和生成器的输入层被放大以接受连接的输入,这样用户就可以让生成器输出特定类别的数据[25]。

现有的研究表明,cGAN只是为了使生成的图像更具有控制性而增加了控制信息,并没有解决训练不稳定问题。因此,cGAN不能很好地执行监督任务,如语义分割、实例分割、直线检测等。可能的原因是G通过最小化并不直接依赖于真实数据标签的损失函数来优化。为了解决上述问题,文献[26]还提出了一种类条件GAN,该GAN不需要人工标注类标签。而标签是通过在D的特征空间中应用聚类自动导出的。聚类操作会自动发现不同的模式,并要求G明确地覆盖它们。在后面的介绍中,很多的应用都直接或间接地用到cGAN,或者是对cGAN的进一步改进,但使控制信息应用到GAN的思想是不变的。因此,条件生成可以说是GAN应用中极为重要的一部分。

2.2.2　基于卷积优化的GAN

卷积神经网络(convolutional neural network,CNN)[27]作为如今最有效和应用最多的学习模型,近几年在计算机视觉领域开始变得举足轻重。在原始GAN中,多层感知机(multilayer perceptron,MLP)用于生成和鉴别网络。由于MLP的训练不稳定且具有挑战性,并且CNN在特征提取方面比MLP具有更好的性能,因此提出了深度卷积GAN(deep convolutional GAN,DCGAN)[28]架构。在这个体系架构中,通过对CNN进行了一些更改以便可以将其应用于生成和判别网络。这些更改主要是通过在架构、设置和训练上的大量实验和错误中获得的。表2显示了对CNN所做的更改。DCGAN可以很好地生成高分辨率图像,也可以说是设计和训练可持续GAN模型的最关键步骤之一,大多数GAN模型都是基于这种架构。

表2　在GAN中应用CNN架构

Table 2　Application of CNN architecture in GAN

层级结构	标准CNN	CNN(生成器)	CNN(鉴别器)
降维层	池化层	微步卷积	跨步卷积
批归一化	不必要	必要	必要
激活函数	各种激活函数	所有层的ReLU和最后一层的Tanh	所有层中的Leaky ReLU
全连接层	有	没有	没有

在DCGAN架构中,生成器捕获潜在空间中的随机点作为输入并生成图像,所提出的方法是通过使用转置卷积层来实现这一目标。换句话说,步幅为2会产生相反的效果,即在标准卷积层中将使用上采样操作而不是下采样操作。图8显示了发生器的结构。鉴别器是一个标准的卷积网络,它捕获图像作为输入并显示二进制分类(真或假)作为输出。在标准模式下,深度卷积网络利用池化层减少深度网络的输入维数和特征图。对于DCGAN,并没有使用池化层而是使用跨步卷积降维。

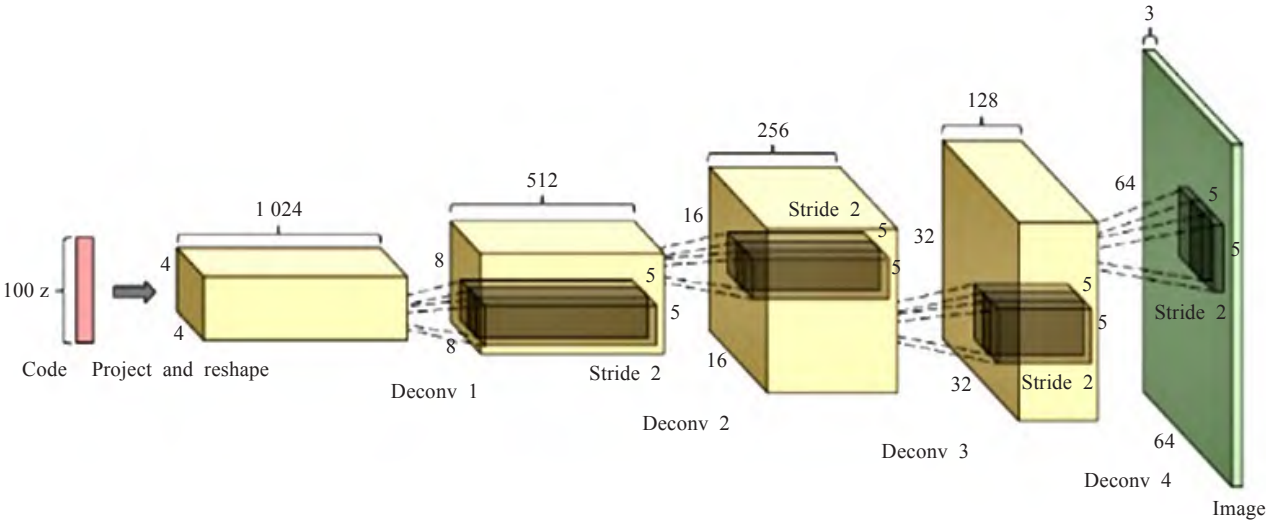


图8 DCGAN的生成器

Fig.8 Generator of DCGAN

虽然实验证明DCGAN训练更加的稳定,但是伴随训练的时间延长,滤波子集有时会折叠为单一振荡模型,对于这一不稳定现象,需进一步研究。

2.2.3 基于自编码器优化的GAN

自编码神经网络是一种用于提取特征和重建操作的无监督神经网络模型。这个网络由两部分组成:编码器 $z=f(x)$ 和解码器 $\hat{x}=g(z)$。编码器通过输入降维过程将 x 转换为潜层 z。同时,解码器通过接收来自潜在层 z 的代码来重构输入 x 以输出。在过去的几年中,自动编码器网络已被用于深度生成模型中。自编码器网络的缺点之一就是编码器产生的潜层在指定空间上没有均匀分布,导致分布中存在大量间隙。

因此,提出了对抗式自动编码(adversarial autoencoder, AAE)[29],这是对抗式网络与自动编码器的组合。在这种方法中,先前的任意分布被施加在编码器获得的潜在层分布上,以确保不存在间隙,从而解码器可以从其每个部分重建有意义的样本。AEE架构如图9所示,在这种架构中,潜在代码 z 表示假信息,z' 由指定分布的 $p(z)$ 表示,两个输入都充当判别器。在完成训练过程后,编码器可以学习期望的分布,而解码器则可以生成根据所需分布重构的样本。表3为基于架构优化的衍生体对比分析。

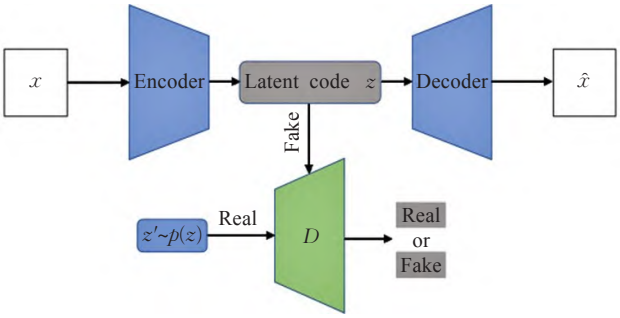


图9 自动编码器(AAE)的结构

Fig.9 Structure of automatic encoder (AAE)

表3 cGAN、DCGAN和AAE的对比

Table 3 Comparison of cGAN, DCGAN and AAE

网络	学习类型	网络架构	梯度更新	方法论	性能指标	机制	优势	局限性	应用场景
cGAN	监督学习	多层感知机	K 步 D 和1步 D 的SGD	以额外信息为条件,最小化 G,最大化 D	对数似然估计	利用条件信息控制输出类别	控制输出类别	生成图像质量较差 训练不稳定 不适合处理离散数据	图像生成 图像编辑 风格迁移 图像修复 语音增强 视频生成
DCGAN	无监督学习	卷积网络	带有Adam优化器的SGD	在 G 和 D 中学习从对象部分到场景的表示层次结构	准确率 错误率	G 和 D 采用深度卷积神经网络	训练稳定 参数较少 无监督学习	延长训练时间,滤波子集可能会折叠为单一振荡模式 无法控制输出类别	图像生成 图像分类 语音合成 视频生成
AAE	监督 半监督 无监督	自动编码器	具有重建和正则化步骤的SGD	通过将自动编码器的隐藏代码向量的后验与先验分布匹配进行推理	对数似然估计 错误率	自动编码器	无监督学习	不适合图像压缩 无法控制输出类别 生成图像质量较差	图像生成 数据降噪 异常检测 可视化降维

2.3 基于目标函数优化的GAN

文献[5]提出GAN网络时,对生成器的目标是最小化 p_{data} 和 p_G 分布距离[30]。在本节中,将讨论如何使用各种距离和从这些距离导出的目标函数来测量 p_{data} 和 p_G 之间的差异。

2.3.1 f-divergence

文献[5]中使用 JS 散度来衡量两个分布之间的差异,但是 JS 散度有自身的函数域,这可能会造成 P_G 和 P_{data} 分布的可能取值域缺乏重叠。理论上来说,真实的数据分布通常是一个低维流形,数据并不具备高维特性,而是嵌入在高维度的低维空间中,而实际当中,数据维度的空间远远不止3维,甚至上百维,这样的话,数据就更难以重合。为解决上述问题,文献[15]提出通用模式 f-divergence来衡量分布之间的差异。使用两个分布的比率,P_{data} 和 P_G 与函数 f 的 f-divergence定义如下:

$$D_f(P_{\text{data}}\|P_G)=\int_x p_G(x)f\left(\frac{p_{\text{data}}(x)}{p_G(x)}\right)\mathrm{d}x \tag{4}$$

上式中,P_{data} 和 P_G 是两个不同的分布;$p_{\text{data}}(x)$ 和 $p_G(x)$ 代表着分别从 P_{data} 和 P_G 中采样出 x 的概率。f 可以是任意类型的收敛,只需满足它是一个凸函数同时 $f(1)=0$。

当令 $f(x)=x\ln x$,f-divergence转换为KL-divergence:

$$D_f(P_{\text{data}}\|P_G)=\int_x p_G(x)\ln\left(\frac{p_{\text{data}}(x)}{p_G(x)}\right)\mathrm{d}x \tag{5}$$

当令 $f(x)=-\ln x$,f-divergence转换为Reverse KL Divergence:

$$D_f(P_{\text{data}}\|P_G)=\int_x p_G(x)\ln\left(\frac{p_G(x)}{p_{\text{data}}(x)}\right)\mathrm{d}x \tag{6}$$

f-divergence根据任意凸函数 f 下的 f 散度来构建目标函数。由于不知道确切的分布,方程(2)应该通过易于处理的形式来估计,例如期望形式。通过使用共轭 $f(u)=\sup_{t\in\text{dom}f^*(t)}(tu-f^*(t))$,等式(2)可以被重新整理如下:

$$D_f(P_{\text{data}}\|P_G)=\int_x p_G(x)\sup_{t\in\text{dom}f^*}\left(t\frac{p_{\text{data}}(x)}{p_G(x)}-f^*(t)\right)\mathrm{d}x \tag{7}$$

$$D_f(P_{\text{data}}\|P_G)\geqslant\sup_{T\in\Gamma}\left(\int_x T(x)p_{\text{data}}(x)-f^*(T(x))p_G(x)\right)\mathrm{d}x \tag{8}$$

$$D_f(P_{\text{data}}\|P_G)=\sup_{T\in\Gamma}(E_{x\sim p_{\text{data}}}[T(x)]-E_{x\sim p_G}[f^*(T(x))]) \tag{9}$$

其中 f^* 是凸函数 f 的Fenchel共轭,$\text{dom}f^*$ 表示了 f^* 域。

等式(7)源于最大值的总和大于总和的最大值,并且 Γ 是满足 $\chi\to R$ 的任意函数类。可以注意到将方程(6)中的 t 替换为方程(7)中的 $T(x):\chi\to\text{dom}f^*$ 使 t 参与到 f_χ。如果将 $T(x)$ 表示为满足 $a(\cdot):\mathbb{R}\to\text{dom}f^*$ 的 $T(x)=a(D_\omega(x))$ 和 $D_\omega(x):\chi\to\mathbb{R}$,可以将 $T(x)$ 解释为具有特定激活函数 $a(\cdot)$ 的判别器。

f-divergence统一了GAN模型,对任何满足条件的 f 都有一个对应的GAN。解决训练过程中模式崩塌的问题,但是在训练的过程中,不同的 f-divergence对训练结果并没有改善。

2.3.2 最小二乘GAN

如上所述,通过 JS 散度衡量两个分布的差异进而拉近 $P_G(x)$ 与 $P_{\text{data}}(x)$ 的距离,由于 JS 散度自身函数域的影响可能会导致 P_G 与 P_{data} 缺乏重叠。从数学的角度来看,无论两种分布多么接近,只要没有相交,那么它们的 JS 散度都是一个常数 $\ln 2$。因此如果生成的样本被判别器分类为真实样本,那么即使生成的样本远离真实数据分布,也没有理由更新生成器。

针对上述问题,文献[20]提出了LSGAN(least squares GAN,LSGAN)的方法。原始GAN的鉴别器使用sigmoid交叉熵损失去判断输入是真还是假。通过上面分析,知道sigmoid交叉损失无法将生成的样本推向真实数据分布,因为它已经实现了分类的作用。受此启发,用更加平滑和非饱和的最小二乘损失代替了sigmoid交叉熵损失。通过 D 将 G 生成的样本拖到真实数据流形中。与公式(1)相比,LSGAN解决了以下问题:

$$\min_D V_{\text{LSGAN}}(D)=\min_D\frac{1}{2}E_{x\sim p_{\text{data}}(x)}[(D(x)-b)^2]+E_{z\sim p_z(z)}[(D(G(z))-a)^2] \tag{10}$$

$$\min_G V_{\text{LSGAN}}(G)=\min_G E_{z\sim p_z(z)}[(D(G(z))-c)^2] \tag{11}$$

其中 a、b 和 c 是判别器的基线值。

等式(10)和(11)使用最小二乘损失,在此情形下,判别器被迫分别具有真实样本和生成样本的指定值(a、b 和 c),而不是真实或假样本的概率。因此最小二乘损失不仅可以对真实样本和生成样本进行分类,还可以将生成的样本拉近真实数据分布。此外,LSGAN还可以连接到 f-divergence框架。

2.3.3 Wasserstein GAN

如上所述,为了较好地测量 P_G 和 P_{data} 分布的差异的问题,文献[11]提出了用Wasserstein距离(也称was距离、EM距离),以便较好地测量两个分布之间的差异,从而解决 JS 和 f-divergence散度不能充分体现两个分布之间差异的弊端。Wasserstein距离定义如下:

$$W(p_{\text{data}},p_G)=\inf_{\gamma\in\Pi(p_{\text{data}},p_G)}E_{(x,y)\sim\gamma}[\|x-y\|] \tag{12}$$

其中,$\Pi(p_{\text{data}},p_G)$ 表示联合分布的集合,$\gamma(x,y)$ 的边缘分布为 $p_{\text{data}}(x)$ 和 $p_G(x)$[31]。

由于方程(12)中的inf项是难以处理的,它通过具有Lipschita函数类[27-28]的Kantorovich-Rubinstein对偶性转换为易于处理的方程,即 $f:X\to R$,满足 $d_R(f(x_1),f(x_2))\leqslant 1\times d_X(x_1,x_2),\forall x_1,x_2\in X$ 其中 d_X 表示域 X 的度量距离。等式(12)的对偶性如下:

$$W(p_{\text{data}},p_G)=\sup_{|f|\leqslant 1}E_{x\sim\text{data}(x)}[f(x)]-E_{x\sim p_G(x)}[f(x)] \quad (13)$$

因此,如果将带有 w 的函数 f 参数化1-Lipschitz函数,则该公式变成了一个极大极小问题,因为首先通过公式(13)中的最大值来训练 f_w 以逼近 $W(p_{\text{data}},p_G)$,并且通过优化生成器 g_θ 来最小化这种近似距离。为了保证 f_ω 是Lipschitz函数,对 w 的每次更新进行权重裁剪,以确保 w 的参数空间位于紧凑空间中。值得注意的是,在WGAN中输出的是was距离,原始GAN的输出结果在(0,1)区间之内,但是was距离是没有上下界的,意味着随着训练的进行,判别器将永远无法收敛,虽然文献[11]中做了权重裁剪处理,但是这个方法并没有让 D 真的限制在Lipschitz函数内,所以WGAN严格意义上说并没有给出was距离的计算方法。

2.3.4 Wasserstein GAN-Gradient Penalty

具有梯度惩罚的WGAN(Wasserstein GAN-Gradient Penalty,WGAN-GP)[32]指出:WGAN并未能将 D 限制在Lipschitz函数内,在训练WGAN时对判别器进行权重裁剪会导致判别器的不良行为,并建议添加梯度惩罚而不是权重裁剪。它表明,通过权重裁剪来保证判别器的Lipschitz条件将判别器限制在所有Lipschitz函数的一个非常有限的子集中;这使判别器偏向于简单的功能。权重裁剪还会产生梯度问题,因为它将权重推到了裁剪范围的极端。为了通过直接约束判别器的梯度来实现Lipschitz条件,建议在方程(13)中添加一个梯度惩罚项,而不是权重裁剪[33]。

Loss sensitive GAN(LS-GAN)[33]也使用Lipschitz约束,但方法不同。它学习损失函数 L_θ 而不是判别器,这样真实样本的损失应该比生成样本小一个与数据相关的边距,导致关注边际高的假样本。此外,LS-GAN假设真实样本 $p_{\text{data}}(x)$ 的密度是Lipschitz连续的,因此附近的数据不会突然变化。采用Lipschitz条件的原因与WGAN的Lipschitz条件无关。关于LS-GAN的文献讨论了Goodfellow等人提出的模型应该无限容量的非参数假设。文献[5]的条件过于苛刻,即使对于深度神经网络也无法满足,并导致训练中出现各种问题;因此它将模型限制在Lipschitz连续函数空间中,而WGAN的Lipschitz条件来自kantorovich-Rubinstein对偶,并且只有判别器受到限制。此外,LS-GAN使用权重衰减正则化技术将模型的权重强加在有界区域内,以确保Lipschitz函数条件。

虽然WGAN-GP很好解决了WGAN遗留的问题,但是WGAN-GP依然存在改进的空间,因为它并没有保证每一个 x 梯度的模都小于或等于1,对于这一问题,在文献[34]中通过频谱范数正则化得到了很好的解决。表4总结分析了基于损失函数优化的衍生体。

表4 基于损失函数优化的衍生体比较

Table 4 Comparison of variants based on loss function optimization

模型	改进	优势	局限性	应用场景
f-GAN	根据任意凸函数下的 f-散度来构建GAN的目标函数	增强了GAN的稳定性	更多的是一种推论,具有较大的不确定性	—
LSGAN	用最小二乘损失函数代替传统GAN模型中的交叉熵损失函数	解决了梯度消失的问题,训练更加稳定	可能会降低生成样本的多样性,训练时生成器可能会发生梯度弥散问题	生成高质量图像
WGAN	使用was距离来计算真实数据的概率分布与生成数据的相似度	解决了训练不稳定的问题,梯度消失的问题	产生的数据样本质量较低,甚至在收敛过程中有时会失败	适合GAN模型不收敛,模式崩溃时使用
WGAN-GP	使用了一个新的惩罚项来实施Lipschitz约束	相比于WGAN有更好的稳定性和更多的多样性	收敛速度慢	模型参数不确定时使用
WGAN-LP	使用了一个新的惩罚项来实施Lipschitz约束	提高网络训练的稳定性	梯度惩罚增加了计算复杂度	模型参数不确定时使用
ACGAN	附加类别标签输入生成器,判别器给出两个概率输出	生成多样丰富和高分辨的图像	训练数据较少时,多样性不足	无监督学习;半监督学习
SNGAN	使用频谱归一化的方式让判别器满足Lipschitz约束	训练更加稳定,保留原始信息	需计算两次梯度下降,训练速度较慢	生成高品质图像
BigGAN	对生成器应用正则化使其可以使用简单的阶段技巧进行训练	使得模型训练稳定,且能使得生成图像的品质变好	模型大,参数多,成本较高	生成高品质图像
OT-GAN	使用Sinkhorn距离和广义能量距离的小批量能量距离	即使停止训练critic,小批量能量距离仍然是一个有效的训练目标	需要大量的计算和内存	生成高质量图像
McGAN	Critic函数的均值和协方差度量	训练稳定,模式梯度下降减少	裁剪的使用限制了模型的能力,难以扩展到高阶的矩阵	生成高品质图像
MMDGAN	引入对抗核学习技术,作为原始GMMN中固定高斯核的替换	具有弱拓扑的优势,可以通过梯度下降以相对较小的批大小进行优化	随着样本的增加,MMDGAN的复杂度也增加	模型参数不确定时使用
CramerGAN	提出具有无偏样本梯度的Cramer距离	在数据流形中间接测量能量距离,并具有变换函数 h	—	生成高质量图像

3 GAN模型评价指标

GAN衍生体目前种类繁多,如果仅仅凭借人工去评测生成样本的优劣,将会消耗大量的人力以及时间成本,并且极易受到主观因素影响。考虑到定性评估存在以上的内在缺陷,因此定量评估就显得十分重要,本节以下内容对GAN的定量评价指标进行了全面的概述。

3.1 Inception分数

文献[35]于2016年提出Inception Score(IS),它对每个生成的图像使用Inception模型来获得条件标签分布 $p(y|x)$。IS指标常用来评估生成图像的质量,采用熵的形式体现了量化的概念。生成图像的多样性与熵的大小成正比,熵值越大意味着样本越丰富。鉴于考虑图像质量和多样性的情形,以互信息形式设计GAN评价指标。为了简化计算添加了指数项,因此IS表达式如下:

$$\exp(\frac{1}{N}\sum_{i=1}^{N}D_{KL}(p(y|x)\|p(y))) \tag{14}$$

其中结果取幂是为了方便比较。较高的IS表明生成样本多样性且高质量。然而,IS也有缺点,如果生成模型陷入模式崩溃,则IS可能仍然很好,而实际情况则非常糟糕。

3.2 FID距离

如上所述,由于IS指标在GAN发生模式崩溃时并不能较好的工作且只考虑生成数据的分布 p_g 而忽略真实数据的分布 p_{data},于是文献[36]在2017年提出了Fréchet Inception Distance(FID),FID计算了在特征空间高斯分布中真实数据与生成数据的弗雷歇距离。FID需要先选取一个特征函数 φ(默认是Inception网络的卷积特征)。FID将 $\varphi(p_{\text{data}})$ 和 $\varphi(p_g)$ 建模为具有均值 μ_r、μ_g 和协方差 C_r、C_g 的高斯随机变量并计算:

$$FID(p_{\text{data}},p_g)=\left\|\mu_r-\mu_g\right\|+\text{tr}(C_r+C_g)-2(C_rC_g)^{1/2} \tag{15}$$

FID距离与GAN的性能成反比,且反映了两个分布之间的亲疏关系。FID数值越大,两个分布相差越大。然而,IS和FID均不能很好地处理过拟合问题。为了缓解这个问题,文献[32]提出了内核初始距离(KID)。总的来说,FID还是相对有效的,其不足之处是高斯分布的假设只存在于理想情况下。

3.3 最大均值差异(MMD)

MMD(maximum mean discrepancy)[37]主要用来测量两个不同但相关的分布的距离。MMD计算了在希尔伯特空间中两个分布的距离,是一种核学习方法。受到FID启发,将求解FID距离的方法替换成MMD,数据分布之间的距离可以作为GAN性能指标。MMD距离与GAN的性能成反比,MMD值越低,则两种分布越相似。实验证明MMD可以较好地识别模式崩塌,尽管是有偏的,但仍推荐使用[38]。

3.4 多尺度结构相似性(MS-SSIM)

SSIM(structural similarity)[39]主要用来度量图像之间的相似性。与单尺度相比,多尺度结构相似性(multi-scale structural similarity,MS-SSIM)[40]是用来评估多尺度图像质量。它通过尝试预测人类感知相似度判断进而定量评估图像的相似度。MS-SSIM值得范围在0和0.1之间,较低的MS-SSIM值通常意味感知上更不相似的图像。文献[41]建议MS-SSIM应与FID和IS指标一起考虑去测量样本的多样性。

3.5 1-最近邻分类器(1-NN)

1-NN(1-nearest neighbor classifier)[42]是如今评价GAN的完美指标,且具备其他指标的所有优势,其输出分数还在[0,1]区间,类似于分类问题中的准确率/误差[38]。1-NN比较训练数据集与生成数据集的概率分布,如果两者的结果一样,则说明GAN的性能良好,反之较差,这类的方法通常采用准确率作为评价指标。1-NN分类器进行评估时,通过留一(leave-one-out,LOO)的准确率去评估 P_r 和 P_g 的差异。假设 T_r~P_r^n、T_g~P_g^m 分别是来自两个概率分布 P_r^n 和 P_g^m 的样本,同时样本的数量一样,即 $|T_r|=|T_g|$,如果两个分布完全匹配的话,则LOO的准确率即为50%。

上述可知,不同的GAN评估指标侧重点不同。在ImageNet上预训练ResNet的卷积空间中,MMD和1-NN在判别力、鲁棒性和效率都是优秀的指标,而IS、FID和MMD不适合于ImageNet差异较大的数据集。尽管人们广泛认为GAN对训练数据过拟合,但这只在训练样本很少的情况下才会发生[38]。考虑到应用场景多样化的情形,评价指标的设计也应该多样化,同时最大化保持模型性能。

4 GAN在不同领域的应用

如上所述,GAN是一个非常神奇的生成模型,它可以根据任意噪声向量 z 生成逼真的数据,且无需知道数据真实分布,也无需任何数学假设。这些特性使得GAN身影遍布诸多领域。本章将讨论GAN在几个领域中的应用。

4.1 图像

4.1.1 图像翻译

图像到图像的翻译是一种无监督学习,用于将图像从一种表示转换成另一种表示。传统方式是利用每个像素的分类或回归。在这些方法中,每一个输出像素都是针对整个输入图像进行预测,并独立于先前输入图像的所有像素,导致图像的大部分语义丢失。为了解决上述问题,文献[43]提出了基于cGAN架构的像素到像素的构架(pix2pix),该体系结构可以在图像到图像翻译应用程序中生成真实的高分辨率图像。与旧的GAN模型

相比,它还允许创建更高分辨率图像(256×256)。图10展示了pix2pix的性能。在pix2pix架构中,生成器和鉴别器分别受到U-Net[44]和PatchGAN[45]的启发。此体系结构中的两个网络使用的都是卷积神经网络。在PatchGAN中,对所有图像进行分类是一步进行的,而在pix2pix的鉴别器中,取而代之的是每个图像首先被分成$n\times n$个patch。然后,对于每个patch分别预测图像是真的还是假的。最终,通过平均所有响应来执行最终分类。

图10 pix2pix生成的图像

Fig.10 Image generated by pix2pix

4.1.2 图像合成

人脸生成以及人脸识别被广泛应用,尽管已有研究提出基于数据的深度学习来达到目的,但是该领域仍然具有一定的挑战性。由于人类视觉对面部畸形和变形很敏感,因此生成逼真的面部图像并非易事。GAN已被证明能够生成具有精致纹理的高质量人脸图像。

双路径GAN(two-pathway generative adversarial network,TP-GAN)[46]可以使用侧面轮廓图像生成高分辨的正面图像(如图11)。这种技术可以像人类一样考虑局部和全局信息,该方法生成的人脸图像很好地保留了个人身份的很多特征。它还可以处理不同模式和光照下的多幅图像并且具有双路径架构。训练全局生成器生成面部标记(标记点)周围的细节。

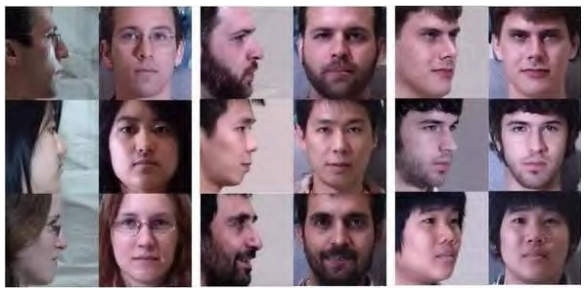
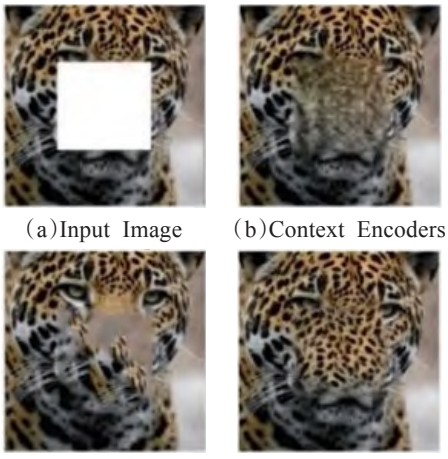


图11 TP-GAN基于侧面轮廓合成正面人脸图像

Fig.11 TP-GAN synthesizes front face image based on side contour

4.1.3 图像修复

图像修复试图重建图像中丢失的部分,使读者无法察觉重建的区域。这种方法通常用于从图像中删除不需要的对象或恢复图像中损坏的部分。在传统的技术中,图像中的孔洞是通过复制原始图像的像素或图像库来填补的。基于深度学习的方法在恢复图像中丢失的区域取得了良好的结果。这些方法可以创建可接受图像结构和纹理。其中一些使用卷积网络的技术,在用正确的特征填补空白方面表现不佳。因此,开发了生成模型以找到训练过程中已知的正确特征。第一种基于GAN的图像恢复方法是上下文编码器[47]。该方法基于编码器-解码器架构进行训练,以根据图像语义推断图像中任意缺失的大区域。尽管如此,在这种方法中,全连接层无法存储准确的空间信息。上下文编码器有时会创建与孔周围区域成比例的模糊纹理。文献[48]将"风格迁移"的思想和上下文编码相结合。但是,该模型不足以用复杂的结构填充缺失的区域。图12显示了该方法和上下文编码器的示例结果。



(a)Input Image (b)Context Encoders

(c)PatchMatch (d)Our Result

图12 Context encoder与PatchGAN生成图像比较

Fig.12 Comparison of images generated by Context encoder and PatchGAN

文献[49]提出了一种基于GAN的图像恢复方法,该方法与全局和局部环境兼容。输入是带有附加二进制掩码的图像,以显示缺失的孔。恢复图形的输出具有相同的分辨率。生成器采用编码器-解码器架构和扩展卷积层而不是标准卷积层来支持更大的空间[50]。并且由两个判别器,一个能够捕获整个图像作为输入的全局判别器和一个能够覆盖小区域的孔作为输入的局部判别器。这两个判别器网络确保生成的图像在"全局"和"局部"尺度上都是兼容的,这使它能够恢复任意孔洞的高分辨率图像。

4.2 序列数据

生成离散值的GAN变体大多借用了RL(reinforcement learning,RL)的策略梯度算法,以规避离散值的直接反向传播。为了输出离散值,生成器作为一个函数,需要将潜在变量映射到元素不连续的域中。然而,如果将反向传播作为另一个连续值的生成过程,判别器会稳定地引导生成器生成类似真实的数据,而不是突然跳到目标离散值。因此,生成器的这种微小变化不能轻易地寻找有限的真实离散数据源[51]。此外,在生成音乐或语言等

序列时,需要逐步评估部分生成的序列,衡量生成器的性能,这也可以通过策略梯度算法来解决。

4.2.1 音乐

当生成音乐时,需要一步一步地生成音乐的音符和音调,而这些元素不是连续的值。一种简单而直接的办法是连续RNN-GAN(C-RNN-GAN)[52],它将生成器和判别器建模为具有长-短期记忆(long short term memory)[53]的RNN,直接提取整个音乐序列。然而,如上所述,只能评估整个序列,而不是评估部分的序列此外,它的结果不是非常令人满意,因为它不考虑音乐元素的离散属性。

相反,序列GAN(sequence generative adversarial network,SeqGAN)[51]和Lee等人[54]使用了策略梯度算法,无需一次生成整个序列。它们将生成器的输出作为代理的策略,并将判别器的输出视为奖励。选择带有鉴别器的奖励是一种常规操作,因为生成器的作用是从判别器中获得大的奖励,类似于代理学习在强化学习中获得的较大奖励。

4.2.2 语言和语音

RankGAN(ranking generative adversarial network)[55]提出了语言(句子)生成方法和排名器而不是传统鉴别器。在自然语言处理中,除了真实性之外,还需要考虑自然语言的表达能力。因此,RankGAN在生成句子和人工编写的参考句子之间采用了相对排名概念。生成器尝试将其生成的语言样本排名靠前,而排名器评估人类编写句子的排名分数高于机器编写的句子。由于生成器输出离散符号,它同样采用类似于SeqGAN和ORGAN(objective reinforced generative adversarial network)[56]的策略梯度算法。在RankGAN中,生成器可以解释为预测下一步符号的策略,并且可以将等级分数视为给定过去生成序列的值函数。

可变自动编码Wasserstein-GAN(variatio-nalautoencoding Wasserstein GAN,VAW)[57]作为一种语音转换系统,结合了VAE和GAN框架。其编码器推断出源语音的语音内容 z,解码器则根据目标说话人的信息 y 合成目标语音,类似于条件VAE。由于高斯分布的假设过于简单,VAE会产生尖锐的结果。为了解决这个问题,VAW-GAN与VAEGAN[12]合并了WGAN[11]。通过将解码器分配给生成器,它的目标是在给定说话人表示的情况下重建目标语音。

4.3 半监督学习

半监督学习在既有标签数据又有未标记数据的情况下,通常使用只有一小部分带有标签的数据集来训练模型。

基于GAN的半监督学习方法[35]演示了如何在GAN框架上使用未标记数据和生成的数据。生成的数据分配给一个 $K+1$ 类,而1至 K 类用于标记真实数据。对于标记的真实数据,鉴别器对其正确的标签进行分类(1到 K)。对于未标记的真实数据和生成的数据,它们使用GAN极大极小游戏进行训练。它们的训练目标可以表示如下:

$$L = L_s + L_{us} \tag{16}$$

$$L_s = -E_{x,y\sim p_{\text{data}(x,y)}}[\text{lb}p_\theta(y|x, y<K+1)] \tag{17}$$

$$L_{us} = -E_{x,y\sim p_{\text{data}(x)}}[1-\text{lb}p_\theta(y=K+1)] + E_{x\sim G}[\text{lb}p_\theta(y=K+1)] \tag{18}$$

其中,L_s 和 L_{us} 分别代表标记数据和未标记数据的损失函数。因为只有生成的数据被归为 $K+1$ 类,所以可以将 L_{us} 视为标准的极小极大博弈。未标记的数据和生成的数据用于告知模型真实数据所在的空间。

4.4 其他应用

4.4.1 信息安全

随着互联网的大规模普及,人们在上网的同时自身的隐私数据也时刻面临巨大的威胁。近年来,由于深度学习在各个领域都取得了不俗的成绩,很多学者将GAN应用于安全检测和系统防护之中。文献[58]基于GAN提出一种新的数据异常检测方法,该方法不需要较多的异常样本数据,只要达到数据平衡就能很好地检测到系统中的异常数据。还有恶意代码生成器MalGAN[59]和识别恶意网页的WS-GAN[60]。MalGAN通过模拟一种可控的透明的攻击方式进而提升自身系统的防御能力。WS-GAN可以有效地识别恶意网页,避免自身的数据泄露。

4.4.2 持续学习

深层生成重放[61]将GAN框架扩展到持续学习领域。持续学习解决多项任务并不断积累新知识。深度神经网络中的持续学习会遭受灾难性遗忘:当新任务被学习时过去获得的知识被遗忘。受大脑机制启发,灾难性遗忘可以通过深层生成重放的GAN框架来解决。深层重放模型是一个有深层生成模型(生成器)和任务解决模型(求解器)组成的合作双模型架构。在这个架构中,该模型通过重发生成的伪数据来保留以前学习过的知识。被称为scholar模型的生成器-求解器对可以针对需求生成假数据和所需的目标对,当出现新任务时,这些目标对会与新数据共同更新生成器和求解网络。因此,scholar模型既可以学习新任务而不忘记自己的知识,还可以在子网络配置不同的情况下用生成的输入-目标对训练其他模型。

4.4.3 疫情防控

自COVID-19大流行以来,已经对公众的健康构成严重的威胁,传统的CT扫描虽然对于疫情防控有很好的效果,但是会花费大量的时间[62]。文献[63]提出一种新颖的算法——CCS-GAN,该算法只需要较少的样本数据就能高效快速地识别CT扫描分类,大大提高了诊断效率。类似的还有数据同化预测的GAN(DA-PredGAN)[64]

和使用辅助分类器进行数据增强的GAN(CovidGAN)[65]。DA-PredGAN利用生成模型具有的模拟时间向前和向后的能力,对COVID-19在传播预防和流控具有很好的效果。CovidGAN生成的合成图像可用于增强COVID-19的检测性能,大大加快检测速度。

5 总结与展望

自从2014年Goodfellow等人把GAN网络架构模型首次引入深度学习之中,GAN便凭借其强大的对抗学习能力,受到广大科研人员的喜爱,GAN的应用也开始蔓延至各行各业,尤其在计算机视觉方向更是层出不穷。本文首先从架构、目标函数和面临的挑战以及应对策略等角度进行了一系列探讨,为了应对GAN面临的挑战,从架构和目标函数两个优化角度进行讨论。然后,对目前工作中GAN性能评估的方法做了详细的汇总和分析。最后,根据不同的应用场景对GAN的应用做了详细介绍。尽管GAN在很多领域取得了令人鼓舞的成就,但是GAN模型本身以及训练算法还是有很多的优化空间,实际上GAN在各领域的应用才真正开始,还有很多值得探索和深耕的未知领域。下面对于GAN的探索提出若干展望:

(1)GAN网络轻量化。研究人员已经设计出大量优秀的大型神经网络算法模型。但这些模型都有一个共同的缺点,就是所需的计算量庞大和参数繁多,对于硬件的配置提出了极高的要求,无疑加大了GAN的应用门槛。这时候GAN网络轻量化显得尤为重要。

(2)GAN与其他模型相结合。随着GAN的不断飞速发展,优点被不断扩大的同时,其弊端也变得不容忽视,如果可以把GAN与其他模型相结合,结合二者优点,克服自身缺点,生成鲁棒性更强的模型,这将是不错的选择。例如:DCGAN是GAN和卷积神经网络结合而成的,这种模型训练更加稳定;还有GAN与ViT结合的ViTGAN[66],与NAS结合的autoGAN[67],都是未来值得关注的方向。

(3)GAN理论突破。GAN源自二人博弈理论,由于自身理论缺陷的原因,GAN的发展一直受到模型崩溃训练不稳定的困扰,虽然学者们提出了很多方案去解决,但是依然基于这一理论,所以无法从根本上解决这一问题。这个时候思考源头本身,从理论出发做出突破或许是一个不错的解决方式,例如,文献[68]提出的三人博弈提论值得借鉴学习。

(4)GAN伪造鉴别。随着GAN的能力被大家所熟知,其生成的数据甚至于达到以假乱真的地步,如何能够保证GAN技术永远用在正确的地方,而不是被别有用心的人利用,这个值得严肃对待,同时也亟需一个答案。

参考文献:

[1] ALPAYDIN E.Machine learning[M].[S.l.]:MIT Press,2021.

[2] SMOLENSKY P.Information processing in dynamical systems:foundations of harmony theory[R].Colorado University.Boulder Department of Computer Science,1986.

[3] HINTON G E,OSINDERO S,TEH Y W.A fast learning algorithm for deep belief nets[J].Neural Computation,2006,18(7):1527-1554.

[4] HINTON G E,SALAKHUTDINOV R.A better way to pretrain deep boltzmann machines[J].Advances in Neural Information Processing Systems,2012,25.

[5] GOODFELLOW I,POUGET-ABADIE J,MIRZA M,et al.Generative adversarial nets[J].Advances in Neural Information Processing Systems,2014,27.

[6] 梁俊杰,韦舰晶,蒋正锋.生成对抗网络GAN综述[J].计算机科学与探索,2020,14(1):1-17.
LIANG J J,WEI J J,JIANG Z F.Generative adversarial networks GAN overview[J].Journal of Frontiers of Computer Science and Technology,2020,14(1):1-17.

[7] 夏萌霏,叶子鹏,赵旺,等.几何视角下深度神经网络的对抗攻击与可解释性研究进展[J].中国科学(信息科学),2021,51(9):1411-1437.
XIA M F,YE Z P,ZHAO W.Adversarial attack and interpretability of the deep neural network from the geometric perspective[J].Science in China(Information Sciences),2021,51(9):1411-1437.

[8] BROWNLEE J.Generative adversarial networks with python:deep learning generative models for image synthesis and image translation[M]//Machine learning mastery,2019.

[9] 陈佛计,朱枫,吴清潇,等.生成对抗网络及其在图像生成中的应用研究综述[J].计算机学报,2021,44(2):347-369.
CHEN F J,ZHU F,WU Q X,et al.A survey about image generation with generative adversarial nets[J].Chinese Journal of Computers,2021,44(2):347-369.

[10] CHEN X,DUAN Y,HOUTHOOFT R,et al.Infogan:interpretable representation learning by information maximizing generative adversarial nets[J].arXiv:1606.03657,2016.

[11] ARJOVSKY M,CHINTALA S,BOTTOU L.Wasserstein generative adversarial networks[C]//Proceedings of the 34th International Conference on Machine Learning,2017:214-223.

[12] LARSEN A B L,SØNDERBY S K,LARO-CHELLE H,et al.Autoencoding beyond pixels using a learned similarity metric[C]//International Conference on Machine Learning,2016:1558-1566.

[13] TRAN N T,BUI T A,CHEUNG N M.Dist-GAN:an improved gan using distance constraints[C]//Proceedings of the European Conference on Computer Vision,2018:370-385.

[14] GUO X,HONG J,LIN T,et al.Relaxed Wasserstein with applications to GANs[C]//2021 IEEE International Con-

ference on Acoustics,Speech and Signal Processing,2021:3325-3329.

[15] NOWOZIN S,CSEKE B,TOMIOKA R.f-GAN:training generative neural samplers usingvariational divergence minimization[J].Advances in Neural Information Processing Systems,2016,29.

[16] MESCHEDER L,NOWOZIN S,GEIGER A.The numerics of GANs[J].Advances in Neural Information Processing Systems,2017,30.

[17] MESCHEDER L,GEIGER A,NOWOZIN S.Which training methods for GANs do actually converge?[C]//International Conference on Machine Learning,2018:3481-3490.

[18] UEHARA M,SATO I,SUZUKI M,et al.Generative adversarial nets from a density ratio estimation perspective[J].arXiv:1610.02920,2016.

[19] LIU M Y,TUZEL O.Coupled generative adversarial networks[J].Advances in Neural Information Processing Systems,2016,29.

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

[21] MISHRA D,PRATHOSH A P,ARAVIND J,et al.Unsupervised conditional generation using noise engineered mode matching GAN[C]//Proceedings of International Conference on Learning Representations,2018.

[22] LIN M.Softmax GAN[J].arXiv:1704.06191,2017.

[23] MIRZA M,OSINDERO S.Conditional generative adversarial nets[J].arXiv:1411.1784,2014.

[24] 马永杰,徐小冬,张茹,等.生成式对抗网络及其在图像生成中的研究进展[J].计算机科学与探索,2021,15(10):1795-1811.
MA Y J,XU X D,ZHANG R,et al.Generative adversarial network and its research progress in image generation[J].Journal of Frontiers of Computer Science and Technology,2021,15(10):1795-1811.

[25] GHOJOGH B,GHODSI A,KARRAY F,et al.Generative adversarial networks and adversarial autoencoders:tutorial and survey[J].arXiv:2111.13282,2021.

[26] ADLER J,LUNZ S.Banach Wasserstein GAN[J].Advances in Neural Information Processing Systems,2018,31.

[27] 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.

[28] RADFORD A,METZ L,CHINTALA S.Unsupervised representation learning with deep convolutional generative adversarial networks[J].arXiv:1511.06434,2015.

[29] MAKHZANI A,SHLENS J,JAITLY N,et al.Adversarial autoencoders[J].arXiv:1511.05644,2015.

[30] 叶晨,关玮.生成式对抗网络的应用综述[J].同济大学学报(自然科学版),2020,48(4):591-601.
YE C,GUAN W.A review of application of generative adversarial networks[J].Journal of Tongji University(Natural Science),2020,48(4):591-601.

[31] 魏丙财,张立晔,孟晓亮,等.基于深度残差生成对抗网络的运动图像去模糊[J].液晶与显示,2021,36(12):1693-1701.
WEI B C,ZHANG L Y,MENG X L,et al.Motion image deblurring based on depth residual generative adversarial network[J].Chinese Journal of Liquid Crystals and Displays,2021,36(12):1693-1701.

[32] GULRAJANI I,AHMED F,ARJOVSKY M,et al.Improved training of wasserstein gans[J].Advances in Neural Information Processing Systems,2017,30.

[33] QI G J.Loss-sensitive generative adversarial networks on lipschitz densities[J].International Journal of Computer Vision,2020,128(5):1118-1140.

[34] MIYATO T,KATAOKA T,KOYAMA M,et al.Spectral normalization for generative adversarial networks[J].arXiv:1802.05957,2018.

[35] SALIMANS T,GOODFELLOW I,ZAREMBA W,et al.Improved techniques for training GANs[J].Advances in Neural Information Processing Systems,2016,29.

[36] SOLOVEITCHIK M,DISKIN T,MORIN E,et al.Conditional frechet inception distance[J].arXiv:2103.11521,2021.

[37] DZIUGAITE G K,ROY D M,GHAHRAMANI Z.Training generative neural networks via maximum mean discrepancy optimization[J].arXiv:1505.03906,2015.

[38] XU Q,HUANG G,YUAN Y,et al.An empirical study on evaluation metrics of generative adversarial networks[J].arXiv:1806.07755,2018.

[39] DOSSELMANN R,YANG X D.A comprehensive assessment of the structural similarity index[J].Signal,Image and Video Processing,2011,5(1):81-91.

[40] WANG Z,SIMONCELLI E P,BOVIK A C.Multiscale structural similarity for image quality assessment[C]//The Thrity-Seventh Asilomar Conference on Signals,Systems & Computers,2003:1398-1402.

[41] KURACH K,LUČIĆ M,ZHAI X,et al.A large-scale study on regularization and normalization in GANs[C]//International Conference on Machine Learning,2019:3581-3590.

[42] SATHE S,AGGARWAL C C.Nearest neighbor classifiers versus random forests and support vector machines[C]//2019 IEEE International Conference on Data Mining,2019:1300-1305.

[43] 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,2017:1125-1134.

[44] SIDDIQUE N,PAHEDING S,ELKIN C P,et al.U-net

and its variants for medical image segmentation:a review of theory and applications[J].IEEE Access,2021.

[45] CHANDALIYA P K,NAIN N.Child face age progression and regression using self-attention multi-scale patch GAN[C]//International Joint Conference on Biometrics (IJCB 2021),2021:1-8.

[46] YU W,CHEN F,CHOI J.Multi-pose face recognition based on TP-GAN[C]//International Conference on Intelligent and Fuzzy Systems,2021:725-732.

[47] ZENG Y,FU J,CHAO H,et al.Aggregated contextual transformations for high-resolution image inpainting[J]. IEEE Transactions on Visualization and Computer Graphics,2022.

[48] YANG C,LU X,LIN Z,et al.High-resolution image inpainting using multi-scale neural patch synthesis[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2017:6721-6729.

[49] IIZUKA S,SIMO-SERRA E,ISHIKAWA H.Globally and locally consistent image completion[J].ACM Transactions on Graphics(ToG),2017,36(4):1-14.

[50] YU F,KOLTUN V.Multi-scale context aggregation by dilated convolutions[J].arXiv:1511.07122,2015.

[51] YAN Y,SHEN G,ZHANG S,et al.Sequence generative adversarial nets with a conditional discriminator[J].Neurocomputing,2021,429:69-76.

[52] MOGREN O.C-RNN-GAN:continuous recurrent neural networks with adversarial training[J].arXiv:1611.09904, 2016.

[53] HOCHREITER S,SCHMIDHUBER J.Long short-term memory[J].Neural Computation,1997,9(8):1735-1780.

[54] LEE S,HWANG U,MIN S,et al.A seqgan for polyphonic music generation[J].arXiv:1710.11418,2017.

[55] LIN K,LI D,HE X,et al.Adversarial ranking for language generation[J].Advances in Neural Information Processing Systems,2017,30.

[56] GUIMARAES G L,SANCHEZ-LENGELING B,OUTEIRAL C,et al.Objective-reinforced generative adversarial networks(ORGAN) for sequence generation models[J].arXiv: 1705.10843,2017.

[57] HSU C,HWANG H T,WU Y C,et al.Voice conversion from unaligned corpora using variational autoencoding wasserstein generative adversarial networks[J].arXiv:1704.00849,2017.

[58] 庄跃生,林珊玲,林志贤,等.生成对抗网络在数据异常检测中的研究[J].计算机工程与应用,2022,58(4):143-149.
ZHUANG Y S,LIN S L,LIN Z X,et al.Study on generative adversarial network for data anomaly detection[J]. Computer Engineering and Applications,2022,58(4):143-149.

[59] MOTI Z,HASHEMI S,KARIMIPOUR H,et al.Generative adversarial network to detect unseen Internet of things malware[J].Ad Hoc Networks,2021,122:102591.

[60] 万梦翔,姚寒冰.面向恶意网页训练数据生成的GAN模型[J].计算机工程与应用,2021,57(6):124-130.
WAN M X,YAO H B.GAN model for malicious Web training data generative[J].Computer Engineering and Applications,2021,57(6):124-130.

[61] SHIN H,LEE J K,KIM J,et al.Continual learning with deep generative replay[J].Advances in Neural Information Processing Systems,2017,30.

[62] 吴辰文,梁雨欣,田鸿雁.改进卷积神经网络的COVID-19 CT影像分类方法研究[J].计算机工程与应用,2022,58(2):225-234.
WU C W,LIANG Y X,TIAN H Y.Research on COVID-19 CT image classification method based on improved convolutional neural network[J].Computer Engineering and Applications,2022,58(2):225-234.

[63] MENON S,MANGALAGIRI J,GALITA J,et al.CCS-GAN: COVID-19 CT-scan classification with very few positive training images[J].arXiv:2110.01605,2021.

[64] SILVA V L S,HEANEY C E,LI Y,et al.Data assimilation predictive GAN(DA-PredGAN):applied to determine the spread of COVID-19[J].arXiv:2105.07729,2021.

[65] WAHEED A,GOYAL M,GUPTA D,et al.CovidGAN:data augmentation using auxiliary classifier GAN for improved Covid-19 detection[J].IEEE Access,2020,8:91916-91923.

[66] LEE K,CHANG H,JIANG L,et al.Vitgan:training GANs with vision transformers[J].arXiv:2107.04589,2021.

[67] GONG X,CHANG S,JIANG Y,et al.Autogan:Neural architecture search for generative adversarial networks[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision,2019:3224-3234.

[68] LI C,XU T,ZHU J,et al.Triple generative adversarial nets[J].Advances in Neural Information Processing Systems,2017,30.