生成对抗网络研究综述

于文家，樊国政，左昱昊，陈怡丹
（西安航空工业计算技术研究所，西安 710064）

摘　要：生成对抗网络（GAN）是一种基于深度学习的强大生成模型，目前广泛应用于计算机视觉、自然语言处理、半监督学习等领域，并取得了显著的成果。随后研究人员通过对GAN的生成器、判别器做结构上的改进或对目标函数等进行优化，提出了更多种的GAN。首先介绍GAN的研究进展和基本思想，其次对一些经典的GAN，如深度卷积生成对抗网络（DCGAN）、条件生成对抗网络（CGAN）、WGAN和超分辨率生成对抗网络（SRGAN）等进行综述，最后对GAN的相关工作进行总结与展望。

关键词：深度学习；生成对抗网络；生成器；判别器
DOI:10.16184/j.cnki.comprg.2023.05.030

1　概述

近年来，人工智能[1]与深度学习[2]在机器学习领域得到快速发展，各种新型的神经网络不断出现，如受限玻尔兹曼机（RBM）[3]、卷积神经网络[4]和自动编码器[5]等。现有的深度学习模型大致分为两类：判别器与生成器。2014年，Ian Goodfellow将RBM与自动编码器结合并引入极大值、极小值双边博弈思想，提出了由判别器与生成器两部分组成的GAN[6]这一全新的生成模型。但是GAN的调参难度较大，处理比较复杂的数据集、训练过程有时会出现模式崩溃、生成图片效果较差、梯度消失等问题。随着研究人员对GAN的研究逐渐深入，GAN在许多领域与其他模型进行了结合，均取得了良好的应用效果。对几种经典的GAN进行综述，大致介绍了GAN目前的发展趋势。

2　GAN模型的基本原理

作为一种强大的生成器，GAN是一种包含两个由多层感知机（MLP）组成的网络的深度神经网络结构，即生成器和判别器。生成器负责生成尽可能逼真的数据以便成功“欺骗”判别器，而判别器则需要尽可能准确地区分出真实数据与生成数据。

真实数据与生成数据之间的关系如图1所示。

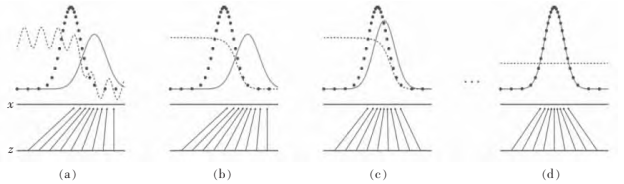

图1　真实数据与生成数据之间的关系

在图1中，真实数据的分布由点间距较大的虚线表示，判别器判断出的结果分布由点间距较小的虚线表示，生成器生成数据的分布由实线表示。其中，z表示噪声，而z到x之间的线段表示由生成器生成的数据分布与真实分布之间的对应情况。GAN的目的是使用生成样本分布（实线）去拟合真实的样本分布（点间距较大的虚线），进而生成尽可能真实的样本。在训练过程的初期，由生成器生成的数据分布与真实的样本分布区别较大，判断数据的真实性需要对判别器进行训练。在经过一定的训练次数后，GAN达到图1（b）状态，此时判别器可以较好地指示样本的来源。随后通过对生成器的训练，提升生成数据的真实性，以“欺骗”判别器，进而达到图1（c）状态。最后通过多次的迭代训练后，GAN达到图1（d）状态，生成的数据几乎完全拟合于真实的数据分布情况，此时判别器判断数据来源的结果约为1/2。

设z为随机噪声，x为真实数据，生成器和判别器分别为G和D。其中，D是一个二分类器，用于判断数据的来源是生成器还是真实数据。GAN的损失函数计算公式如公式（1）所示：

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

其中，第1项中的$\log D(x)$表示判别器对样本数据的判断结果；第2项则表示对数据的合成与判断。基于极大值、极小值双边博弈，分别对生成器和判别器进行优化并进行交替训练，直至达到纳什均衡。对于GAN的目标函数，在生成器G的参数固定时，可以得到最优的判别器D。对于一个来自真实分布或生成分布的数据，它对判别器损失函数的贡献如公式（2）所示：

$$-p_r(x)\log D(x) - p_g(x)\log[1-D(x)] \tag{2}$$

其中，$p_r(x)$为真实分布；$p_g(x)$为生成分布。令其关于$D(x)$的导数为零，可以得到全局最优解，如公

式（3）所示：

$$D(x)=\frac{p_r(x)}{p_r(x)+p_g(x)} \tag{3}$$

若将生成器固定，则将目标函数中的数学期望按照定义展开，如公式（4）所示：

$$\begin{aligned}V(G,D)&=\int p_r(x)\log(D(x))\mathrm{d}x+\int p_z(z)\log\left(1-D(g(z))\right)\mathrm{d}z\\&=\int P_r(x)\log(D(x))\mathrm{d}x+P_g(x)\log(1-D(x))\mathrm{d}x\end{aligned} \tag{4}$$

当生成器固定时，公式（4）中的$p_r(x)$与$p_g(x)$均表示常数，此时$V(G,D)$表示$D(x)$的函数。令$y=D(x)$，$a=p_r(x)$，$b=p_g(x)$，构造函数计算公式如公式（5）所示：

$$F(y)=a\log y+b\log(1-y) \tag{5}$$

对$F(y)$求导，并令其导数为零，所得公式如公式（6）所示：

$$D(x)=y=\frac{a}{a+b} \tag{6}$$

将该最优判别器的值代入目标函数中并消去$D(x)$，得到关于G的目标函数，如公式（7）所示：

$$\begin{aligned}C(G)&=\max_D V(D,G)\\&=E_{Z\sim p_g(z)}\left[\log\left(1-D(G(z))\right)\right]\\&=-\log 4+\mathrm{KL}\left\{p_r(x)\|\frac{p_r(x)+p_g(x)}{2}\right\}\\&\quad+\mathrm{KL}\left\{p_g(x)\|\frac{p_r(x)+p_g(x)}{2}\right\}\\&=-\log 4+2\mathrm{JS}D(p_r(x)\|p_g)\end{aligned} \tag{7}$$

KL散度与JS散度均为非负，并且当且仅当两个分布相等时取值为0。

由公式（7）可得，当且仅当$p_r(x)=p_g(x)$时，$C(G)$取得最小值$-\log 4$。

当生成对抗网络训练过程的迭代次数足够多时，$p_r(x)$与$p_g(x)$无限接近，可看作近似相等。此时$D(x)$的最优解近似为0.5，即判别器无法判断样本数据的来源，而生成器生成的数据与真实的样本数据完全一致。

3 GAN的改进

3.1 DCGAN

与GAN中生成器和判别器所使用的多层感知机相比，卷积神经网络具有更强大的数据拟合与表达能力，并且其在判别式模型中取得了一定的成功。因此，Alec等人使用卷积神经网络对原始GAN的生成器和判别器所采用的多层感知机进行了替换，提出了DCGAN。从本质上来说，DCGAN是对GAN的训练过程进行指导。DCGAN在强调隐藏层分析和可视化计数的同时，通过使用卷积层取代了全连接层、去除池化层并采用批标准化等技术，将判别器的训练结果作为输入回传到生成器中。

DCGAN的出现极大地增强了GAN的数据生成质量。在理论上，DCGAN虽然没有带来解释性，但是其生成图像的强大能力使得众多研究人员逐渐开始对生成对抗网络进行研究。同时，DCGAN证明了将神经网络引入生成对抗网络的可行性。另外，DCGAN的网络结构也被用以评价不同目标函数的GAN生成能力。

3.2 CGAN

随着研究的深入，越来越多带有附加信息的数据成为了深度学习的研究对象。Mehdi Mirza提出了CGAN将额外信息引入到GAN之中，通过使用条件概率对目标函数进行替换并生成带有标签的数据。CGAN将附加信息作为输入的一部分引入到GAN中，使其能够指导数据的生成过程，从而使GAN可以处理带有类别标签的数据。

CGAN与GAN的区别在于其目标函数中的概率是条件概率。CGAN的目标函数计算公式如公式（8）所示：

$$\begin{aligned}\min\max V(D,G)&=E_{x\sim p_{\mathrm{data}}(x)}[\log D(x|y)]\\&\quad+E_{Z\sim p_z(Z)}\left[\log\left(1-D(G(z|y))\right)\right]\end{aligned} \tag{8}$$

其中，y表示额外信息。CGAN将额外信息y作为生成器和判别器输入的一部分。在生成器中，先验输入噪声$p_z(z)$和额外信息y共同构成生成器的隐藏层输入。

CGAN以GAN为基础，通过将类别标签之类的额外信息作为生成器与判别器输入的一部分，实现GAN处理带有额外信息数据的功能。原始GAN的生成器输入是随机噪声，而CGAN的生成器可以将类别标签与随机噪声组合后的数据作为隐藏层输入。原始GAN判别器的输入是图片数据，CGAN的判别器的输入是类别标签和图片数据拼接以后的数据，并将其作为判断是生成器生成的数据还是实际数据的依据。

3.3 WGAN

2017年，马丁·阿约夫斯基（Martin Ayovsky）等用Wasserstein距离代替GAN中的JS散度，以解决生成对抗网络中两种分布不重叠的梯度消失问题，提出了WGAN。Wasserstein距离的数学表达式如公式（9）所示：

$$W(P_r,P_g)=\inf_{\gamma\sim\prod(P_r,P_g)}E_{(x,y)\sim\gamma}[\|x-y\|] \tag{9}$$

当f_w为判别器时，WGAN的目标函数如公式（10）所示：

$$L=E_{x\sim p_r}[f_w(x)]-E_{x\sim p_g}[f_w(x)] \tag{10}$$

与之前的判别器不同，WGAN不再需要使用判别器作为0~1分类来将其值限制在[0,1]。f_w的值越大，WGAN

的生成分布就越接近真实的分布；反之，则WGAN的生成分布越接近生成的分布。此外，由于利普希茨（Lipschitz）常数是1，显然Lipschitz连续在鉴别器中是难以实现的。为了将Lipschitz连续表示为权重剪枝，需要参数$w\in[-c,c]$，其中，c表示一个常数。判别器的损失函数如公式（11）所示：

$$L_D=\max_{f_w}E_{x\sim p_r}[f_w(x)]-E_{z\sim p_z}[f_w(G(z))] \tag{11}$$

与此同时，生成器的损失函数如公式（12）所示：

$$L_G=\min_{f_w}-E_{z\sim p_z}[f_w(G(z))] \tag{12}$$

WGAN在理论上解释了因生成器的梯度消失而导致的不稳定训练的原因，并用Wasserstein距离代替了JS散度，从理论上解决了梯度消失的问题。

3.4 SRGAN

2017年，Christian Ledig等首次将生成对抗网络用于图像超分辨率领域，提出了SRGAN[10]并取得了良好的效果。SRGAN是由对抗损失和内容损失共同组成的感知损失，用来取代传统GAN的损失函数。

SRGAN的生成器使用的是SRResNet，因为当损失函数从判别器反向传播返回到生成器时，需要经过多层网络，而在经过多层网络的过程中势必会出现梯度弥散。通过使用残差连接，可以有效地保证生成对抗网络训练的鲁棒性。

SRGAN的生成器损失函数如公式（13）~（16）所示：

$$l^{\mathrm{SR}}=\lambda_{\mathrm{msc}}\times l^{\mathrm{SR}}_{\mathrm{MSE}}+\lambda_{\mathrm{perc}}\times l^{\mathrm{SR}}_{\mathrm{perc}}=\lambda_{\mathrm{msc}}\times l^{\mathrm{SR}}_{\mathrm{MSE}}+\lambda_{\mathrm{perc}}\times l^{\mathrm{SR}}_{X}+\lambda_{\mathrm{gen}}\times l^{\mathrm{SR}}_{\mathrm{Gen}} \tag{13}$$

$$l^{\mathrm{SR}}_{\mathrm{MSE}}=\frac{1}{r^2WH}\sum_{x=1}^{rW}\sum_{y=1}^{rH}\left(I^{\mathrm{SR}}_{x,y}-G_{\theta_G}(I^{\mathrm{LR}})_{x,y}\right)^2 \tag{14}$$

$$l^{\mathrm{SR}}_{X}=l^{\mathrm{SR}}_{\mathrm{VGG}/i,j}=\frac{1}{W_{i,j}H_{i,j}}\sum_{x=1}^{W_{i,j}}\sum_{y=1}^{H_{i,j}}\left(\phi_{i,j}(I^{\mathrm{SR}})_{x,y}-\phi_{i,j}\left(G_{\theta_G}(I^{\mathrm{LR}})\right)_{x,y}\right)^2 \tag{15}$$

$$l^{\mathrm{SR}}_{\mathrm{Gen}}=\sum_{n=1}^{N}-\log D_{\theta_D}\left(G_{\theta_G}(I^{\mathrm{LR}})\right) \tag{16}$$

其中，$l^{\mathrm{SR}}_{\mathrm{MSE}}$表示重构损失，即生成图像与真实图像的逐像素点MSE损失；$l^{\mathrm{SR}}_{\mathrm{perc}}$表示感知损失，由对抗损失和内容损失组成；$l^{\mathrm{SR}}_{X}$表示内容损失；VGG网络中第$i$层第$j$个卷积核输出的特征映射的MSE损失为$l^{\mathrm{SR}}_{\mathrm{VGG}/i,j}$；$l^{\mathrm{SR}}_{\mathrm{Gen}}$表示对抗损失。

SRGAN的判别器损失函数如公式（17）所示：

$$l_D=-y\times\log D(x^{\mathrm{HR}})-(1-y)\times\log\left(1-D(x^{\mathrm{HR}})\right),\quad y=\begin{cases}1, x^{\mathrm{HR}}\in X^{\mathrm{HR}}_{\mathrm{Real}}\\ 0, x^{\mathrm{HR}}\in X^{\mathrm{HR}}_{\mathrm{Gen}}\end{cases} \tag{17}$$

SRGAN提出了新的生成器——SRResNet，以及全新的损失函数，在通过对抗损失提升生成图像的真实感的同时，通过内容损失获取高分辨率图像和生成图像的感知相似性，而不只是像素级相似性，这些改进使得SRGAN可以更好地捕捉图像的感知细节。

4 结语

GAN的强大生成能力使其在深度学习领域得到了快速发展。目前国内外众多学者已经将新的研究方向放在了训练指标、模式坍塌及生成能力的可解释性上。在实际应用场景中，如何去除噪声、提升生成图像的质量及如何将这一强大的深度学习模型应用于自然语言处理领域，也是亟待解决的问题。

参考文献

[1] NILSSON N J. Principle of artificial intelligence [J]. IEEE Intelligent Systems, 1982, 29 (2): 2–4.

[2] LECUN Y, BENGIO Y, HINTON G. Deep learning [J]. Nature, 2015, 521 (7553): 436.

[3] 张健，丁世飞，张楠，等. 受限玻尔兹曼机研究综述 [J]. 软件学报，2019，30 (7)：2073–2090.

[4] LAWRENCE S, GILES C L, TSOI A C, et al. Face recognition: a convolutional neural–network approach [J]. IEEE Transactions on Neural Networks, 1997, 8 (1): 98–113.

[5] 邓俊锋，张晓龙. 基于自动编码器组合的深度学习优化方法 [J]. 计算机应用，2016，36 (3)：697–702.

[6] LEDIG C, WANG Z, SHI W, et al. Photo–realistic single image super–resolution using a generative adversarial network [C] //IEEE. IEEE Conference on Computer Vision and Pattern Recognition, Puerto Rico. NewYork: IEEE, 2017: 1–10.