GAN的原理及应用研究

高楚晴 PB18111688 中国科学技术大学

**摘 要**：

**关键词**：生成对抗网络；图像生成；语义分割

**Research on the Principle and Application of Generative Adversarial Networks**

GAO Chu-qing

University of Science and Technology of China

**Abstract：**

**Key word:** Generative adversarial networks; Image generation; Semanteme division

**1 GAN的基本原理**

**1.1 综述**

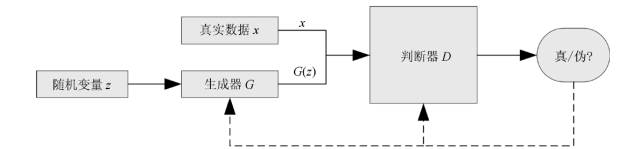
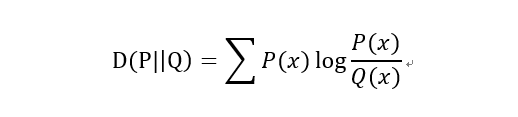


图1：GAN基本模型示意

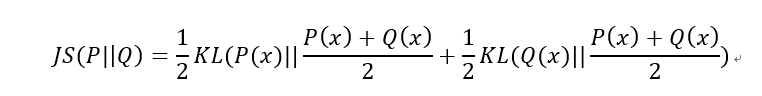
GAN（Generative Adversarial Networks）是一种通过对抗的方式，去学习数据分布的生成式模型（图1）。所谓的对抗，指的是生成网络和判别网络的互相对抗。生成网络尽可能生成逼真样本，判别网络则尽可能去判别该样本为真实数据还是生成器生成的数据。

**1.2 数学补充**

**KL散度(Kullback-Leibler divergence)：**又称KL距离，相对熵。当P(x)和Q(x)的相似度越高，KL散度越小。



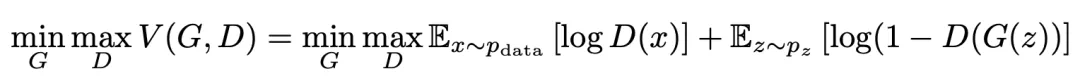
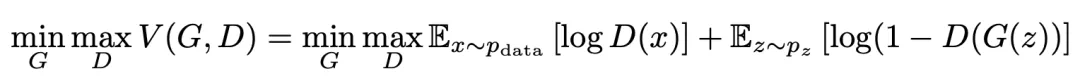
**JS散度(JSD, Jenson Shannon Divergence)：**为解决KL散度所存在的关于P、Q的不对称性问题，同时进一步以一个0-1之间的数值更直观地表现P和Q之间的相似度，引入JS散度

****

**1.3 GAN数学模型[1]**

向生成器G输入隐变量 z（通常为服从高斯分布的随机噪声）获得对应的生成数据，判别器D负责判别输入的数据是生成样本还是真实样本。

对于生成器G来说，其目的为尽可能的生成与真实数据相近的样本来欺骗Discriminator，因此其优化目标为

即最大化生成样本的判别概率D(G(z))，即最小化(1-D(G(z)))，（容易知道D(x)仅与判别器D有关，在G的优化过程中可忽略），其中V(G,D)称为关于G与D博弈过程中的交叉熵损失。

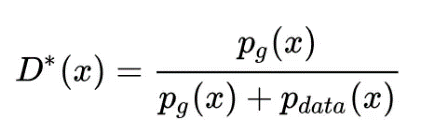
对于判别器D来说，其目的为尽可能准确地判定输入数据的来源，因此优化目标为



即最小化生成器G得到的最大值。

生成器与判别器两者之间不断博弈最后达到纳什均衡时，即在判别器的判别能力达到很高程度时，生成器仍能生成使判别器难以分辨真伪的数据时，即达到了GAN最终的训练目的。

实际训练时，生成器和判别器采取交替训练，由于该过程需要真实数据启动，故先训练判别器再训练生成器，并不断重复该过程。根据上文中提到的对于生成器，其最小化的是V(D,G) 的最大值。为了保证 V(D,G) 取得最大值，所以我们通常会训练迭代k次判别器，然后再迭代1次生成器。当生成器 G 固定时，我们可以对 V(D,G) 求导，求出最优判别器 D\*(x)：

****

把得到的最优判别器代入上述目标函数，可以进一步求出在最优判别器下，生成器的目标函数等价于优化 Pdata(x) , Pg(x) 的 JS 散度（见1.2）。

值得一提的是，后人对于原始GAN模型进行训练稳定性优化时，许多研究者采用了不同于JS散度的距离度量方式，也因此得到了许多不同的性能各异的模型，举例如下（图2）

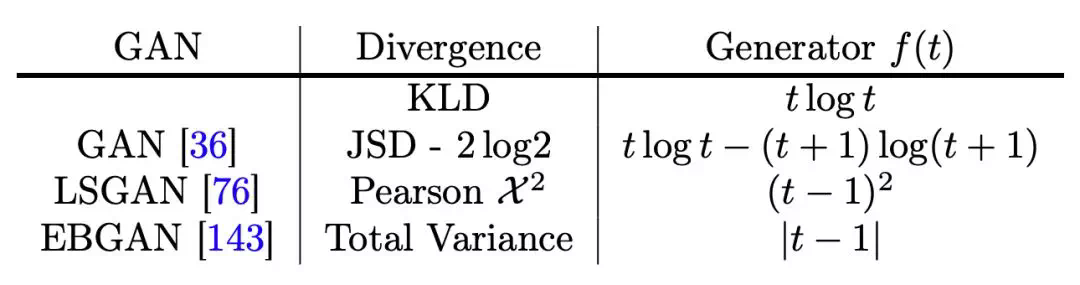


图2[2]

可以证明，当 G，D 二者的样本容量均足够时，模型会收敛，二者将达到纳什均衡。此时，Pdata(x) = Pg(x)，无论是对于真实样本，还是生成器由随机噪声产生的生成数据，判别器的预测概率均为 1/2，即生成样本与真实样本达到了难以区分的地步，此时我们达到了生成训练器的目的。

上述过程有比较直观的图像理解如下（图3）：

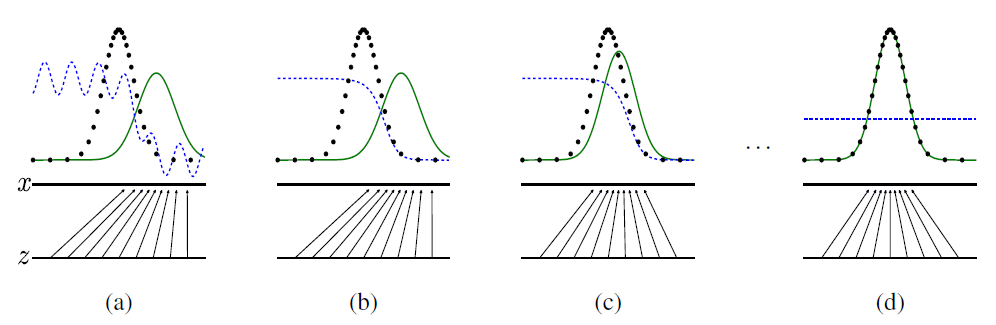


图3[1]

从上述图像可以看出，在训练初期，生成样本的分布情况与真实数据相差仍较大，而判别器输出的生成样本判别概率分布也波动较大（如图3(a)），对判别器进一步迭代训练，得到较光滑的判别概率分布（图3(b)），此时迭代生成器，使生成样本的数据分布与真实样本的拟合度更高（图3(c)），将此过程不断重复最终博弈达到纳什均衡，此时真实样本与生成样本基本吻合，判别器对两者的判别概率均为0.5，即无法分辨生成样本的真伪（图3(d)）。

下面基于python语言模拟上述逐步收敛过程，并绘制图像作以直观理解。

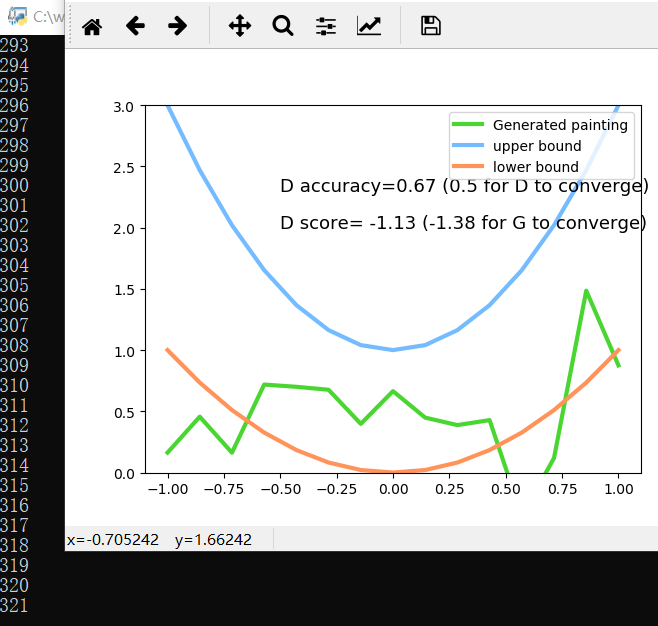


图4：固定k（判别器迭代次数与生成器迭代次数之比）为1时，迭代到第321次

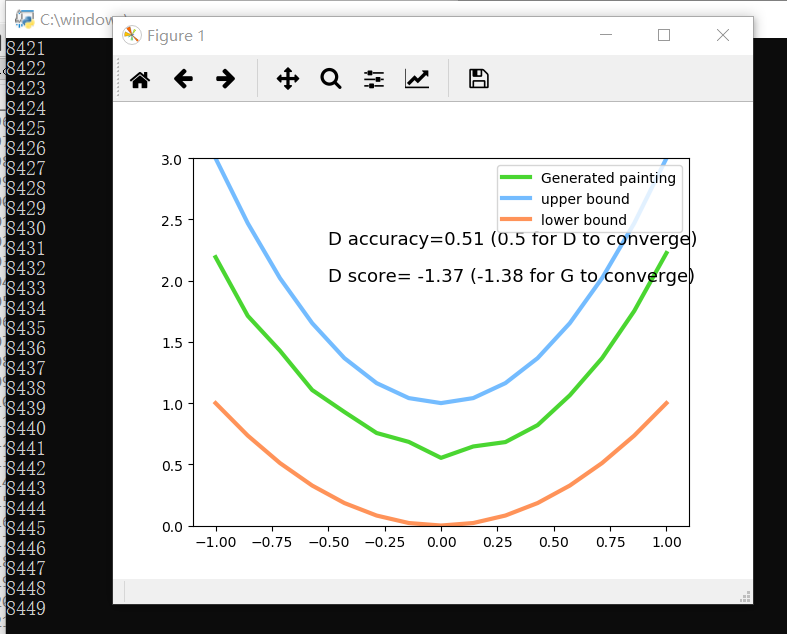


图5：固定k为1时，迭代到第8449次

本次模拟以10000次迭代为终止点，由前文数学推导可知，收敛情况较好时应有生成样本判别概率为0.5，交叉熵函数目标值

图4可以看到迭代321次时生成数据的分布情况还不很理想，生成数据分布均匀度较差，判别准确率约为0.67，与目标值0.5距离还较远，对应的交叉熵函数值为-1.13，与-1.38有较大出入。而由图5看到，迭代到第8449次时，判别准确率为0.51，对应交叉熵函数-1.37，生成样本曲线的光滑度也较好，此时已基本达到收敛。

**2 GAN的应用**

**2.1 图像生成**

**2.1.1 模型简介**

当前随着人工智能、大数据等技术逐步渗透进我们的生活，机器在训练时对原始数据的需求也更加旺盛，举个最简单的例子，在智能城市的研究中，车牌识别是一项基本的技术支持。但在识别器的训练过程中需要大量的原始车牌图像，但原始数据的收集存在一定困难。例如，如果大规模收集合肥当地的车牌图像信息，那么数据集中“皖A”的车牌比例显然远高于理想数据集，若在全国各地收集信息实际操作工作量又过大，因此此时利用GAN批量生成车牌图像便是一个很好的解决方案。

图像生成的实际模型要比上文中提到的原始GAN具有更高的复杂度。图像的符合度并不能仅仅通过一个简单的损失函数来衡量，因此我们希望判别器具有更多维的判断方式，更能够自顶向下地把握全局的信息。同时，原始GAN中，无法控制要生成的内容，因为输出仅依赖于随机噪声。我们可以将条件输入*c*添加到随机噪声*z*中，以便生成的图像由*G*(*c*,*z*) 定义。这就是CGAN[3]（图6），通常条件输入矢量*c*与噪声矢量z直接连接即可，条件*c*可以是图像的类，对象的属性或嵌入想要生成的图像的文本描述等。这样我们得到的GAN就更适合于图像判别及生成了。

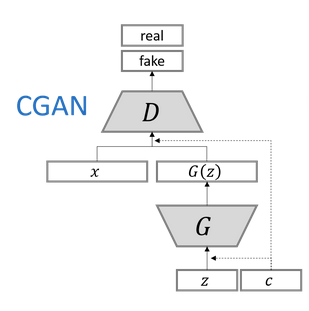


图6：CGAN的基本结构

一、下面介绍两种图像处理方面的GAN改进模型。

1.Apple提出的模型[4]

Apple的模型改进之处主要为

①使用Simulated + Unsupervised模式用无标签的真实数据精细化生成数据（图7）

②引入额外自正则项使生成图像与精细化图像的绝对值误差最小化，保留更多的标签信息（在图7中为眼睛的注视方向）

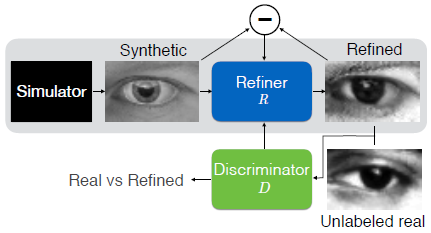


图7：使用无标签真实数据对生成数据精细化，并与原生成图像之间绝对值误差引入额外自正则项

③对输入图像分块（图8），并分别对各个小块进行判别，避免过于强调某些特定的图像特征而导致的伪差。提供一些更锐利的细节，使得生成结果具有更丰富的信息（图9）。



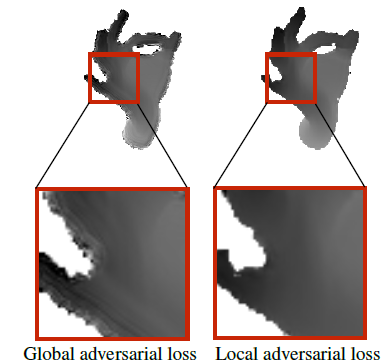
图8：对原图像进行精细化处理 

图9：全局损失与局部损失处理分别得到的图像

④使用一系列历史精细化图像替代最近一次生成数据来优化判别器。

2、阿里提出的模型[5]

该模型的主要改进为分别编码示例图像的内容特征和风格特征，综合这两个特征解码得到新图像（图10）。

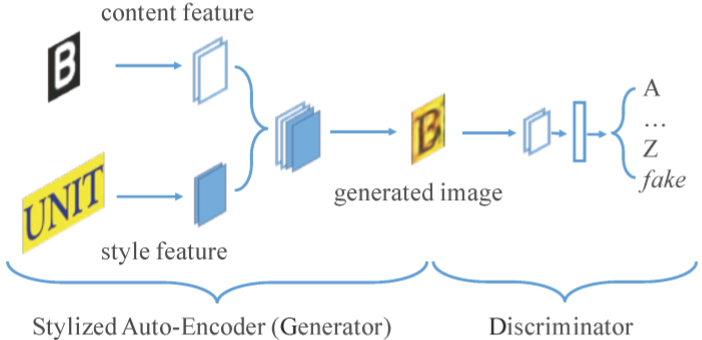


图10：隐变量被拆分成内容特征和风格特征两部分，分别用GAN

进行训练，再对生成数据进行综合得到目标生成样本

该版本使用多类分类器作为鉴别器，更好地建模生成的图像的变化情况，有效地迫使生成网络生成更具真实感的结果。

图像生成技术现在已经比较成熟，在将召开的ICLR 2019参会论文得分排第三的文章[6]提出的模型性能远优于前人的工作，同时图像真实度与精细化程度也很高（图11）。



图11：最左侧为GAN合成的图像

二、图像生成的拓展——根据文本生成图像

容易联想到，该项技术成熟之后，可以在根据文本描述生成罪犯面容上有较好应用。笔者在这里仅以举例（图12），不做拓展说明。

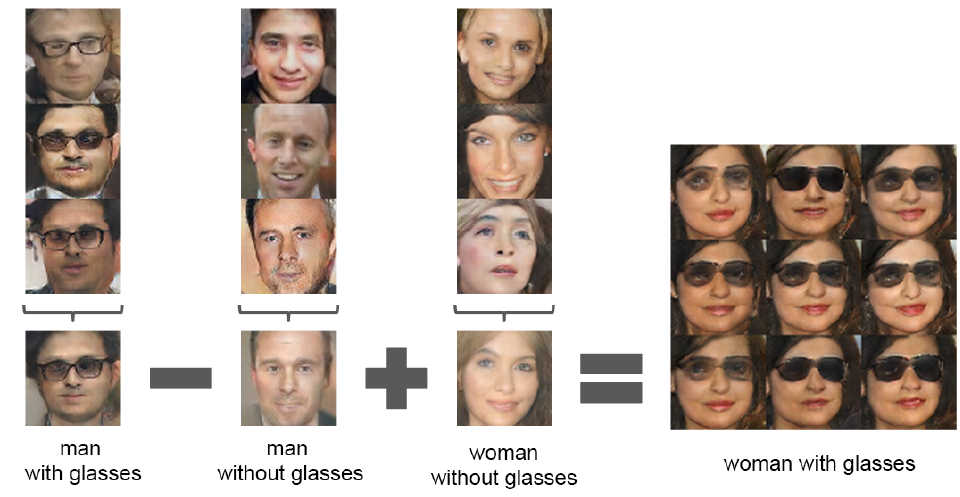


图12：根据文本生成图像的示意，训练时需要带有文本标记的图像数据集

**4 引用**

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In Proc.NIPS, 2014.

[2] Yongjun Hong, Uiwon Hwang, Jaeyoon Yoo and Sungroh Yoon. How Generative Adversarial Networks and Their Variants Work: An Overview. arXiv preprint arXiv:1711.05914,2019.

[3] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784, 2014.

[4] Ashish Shrivastava, Tomas Pfister, Oncel Tuzel, Josh Susskind, Wenda Wang, Russ Webb. Learning from Simulated and Unsupervised Images through Adversarial Training, CVPR 2017.

[5] Yiru Zhao, Bing Deng, Jianqiang Huang, Hongtao Lu, Xian-Sheng Hua:Stylized Adversarial AutoEncoder for Image Generation. ACM Multimedia 2017: 244-251

[6] AndrewBrock, JeffDonahue and KarenSimonyan. Large Scale GAN Training for High Fidelity Natural Image Synthesis . ICLR 2019.