文章编号：2096-1472(2021)-06-20-04
DOI:10.19644/j.cnki.issn2096-1472.2021.06.005

基于生成对抗网络的动漫头像生成研究

彭章龙

(长江大学计算机科学学院，湖北 荆州 434100)
✉202072671@yangtzeu.edu.cn

摘 要：在深度学习中，数据是三大核心要素之一。尤其在某些领域，数据的稀有、人工标注造成大量人力的浪费、数据好坏对产出结果的影响，都显现出数据的重要性。鉴于在动漫领域中，人物的制作需要花费大量的人力和时间，所以从动漫头像出发，基于生成对抗网络，结合编码器、残差网络、解码器，经过编码器改变图像的维度，最后利用解码器将提取到的特征数据生成近似于原始图像的数据集。生成对抗网络本身固有的缺点会导致最后的效果并不是很好，于是尝试对生成对抗网络进行深度卷积的改进，再加上WGAN的梯度惩罚思想来优化自编码器基础上的生成对抗网络。

关键词：深度学习；生成对抗网络；数据生成；深度卷积
中图分类号：TP391 **文献标识码**：A

Research on Animation Profile Picture Generation based on Generative Adversarial Network

PENG Zhanglong

(*School of Computer Science*, *Yangtze University*, *Jingzhou* 434100, *China*)
✉202072671@yangtzeu.edu.cn

Abstract: In deep learning, data is one of the three core elements. Especially in some fields, scarcity of data, manpower waste caused by manual labeling, and the impact of data quality on the output results all show the importance of data. As in animation field, production of characters takes a lot of time and manpower, this paper starts from animation profile picture and combines encoder, residual network and decoder based on Generative Adversarial Network. After the encoder changes the dimension of the image, the decoder is used to generate a dataset similar to the original image with extracted feature data. The inherent shortcomings of the Generative Adversarial Network itself will lead to an unideal final effect, so the author tries to improve the Generative Adversarial Network by deep convolution, coupled with the gradient penalty idea of WGAN (Wasserstein Generative Adversarial Network) to optimize the Generative Adversarial Network based on the autoencoder.

Keywords: deep learning; Generative Adversarial Network; data generation; deep convolution

1 引言(Introduction)

生成对抗网络(Generative Adversarial Network，GAN)于2014年被在蒙特利尔读博士的Ian Goodfellow提出，在之后的几年，一直都处于火热研究对象的状态之中[1]，且于2016年席卷AI领域顶会，深度学习三大马车之一的Yann LeCun曾形容它为“20年来深度学习领域最酷的构想”。生成对抗网络被广泛应用于图像生成[2]、图像转换[3]、图像修复[4]，在目标检测[5]、行人识别[6]等方面也有着重要的辅助作用。大量的研究者希望将生成对抗网络应用于各个领域，例如有在医学领域的研究者希望借助生成对抗网络的学习方式及其学习能力来生成药学分子结构和合成新材料的配方。

2 生成对抗网络(Generative adversarial network)

2.1 网络结构与原理展示

生成对抗网络(Generative Adversarial Network，GAN)因为有着极好的生成能力以及效果而得到了广泛的认知，其网络结构如图1所示。

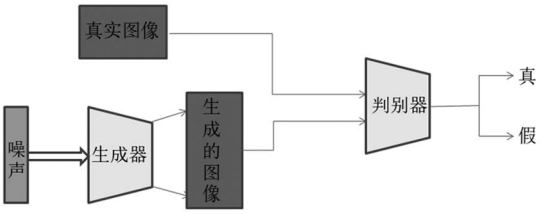

图1 生成对抗网络的网络结构

Fig.1 The network structure of generative adversarial network

它同时包含了判别式模型和生成式模型。生成式模型是为了产生与来自真实数据相似的数据，让判别式模型误以为是真实数据，而判别式模型是为了极力地判断出其数据并不是真实的数据，将其与真实的数据分别开来。

在判别式模型和生成式模型两者互相对抗学习的情况下，判别式模型的判断会让生成式模型逐渐产生逼近于真实的数据，同时生成式模型在生成近似于真实数据的时候，判别式模型的判别能力也会增强，努力找寻两者数据之间的差距，将两者区别开来。到最后，生成式模型会拥有生成真实数据分布的能力，判别式模型会因为生成式模型能力的增强而增强，对生成样本判断为虚假样本的性能增强。

整个过程与画家成长过程相似，画家不断学习自己的画与名画之间的差距，画出的画更接近名画来干扰鉴画师，而鉴画师也会不断学习鉴别假画与真画之间差距的能力。

生成对抗网络采用博弈论中零和博弈游戏的思想，以期达到纳什均衡点。

生成式模型不断生成数据分布，判别式模型判断数据是否为真实数据，两者相互对抗，到最后两者都学习到最优状态。

生成式模型损失函数为：

$$G_{loss}=(1-y)\log\left(1-D\left(G(z)\right)\right) \tag{1}$$

在$G(z)$逼近于真实数据分布时，$D\left(G(z)\right)$判别结果会为1，y是数据的标签，为1或者0。

判别式模型损失函数为：

$$D_{loss}=y\log\left(D(x)\right)-(1-y)\log\left(1-D(G(z))\right) \tag{2}$$

为了让判别式模型判断为假，需使$D\left(G(z)\right)$尽量趋于0，而使真实数据$D(x)$尽量趋于1。

将生成对抗网络优化目标形式化为：

$$\min_G\max_D L(G,D)=E_{x\sim p_{data}(x)}[\log D(x)]+E_{z\sim p_z(z)}\left[\log\left(1-D\left(G(z)\right)\right)\right] \tag{3}$$

G就是生成式模型，D就是判别式模型，G的目标是通过学习，使得目标优化损失函数最小化；D的目标是通过学习，使得目标优化损失函数最大化。通过不停地迭代，双方能力逐渐提升，最后整个网络模型训练完成，使得$D(x)$判别真实数据的分布为1/2，达到真实数据与生成数据形成的数据分布各占一半的结果。

2.2 存在问题

在GAN训练的过程中会有一些问题[7]：在把判别式模型训练得很好的时候，会出现生成式模型完全学不动的情况(*loss*降不下去)。但是，判别式模型学得不好，又会出现生成式模型梯度不准的情况。只能把判别式模型训练得不好不坏才行，而这个状态很难把握，甚至在同一轮训练的前后不同阶段，这个状态都可能不一样，所以GAN很难训练。

公式(3)等价于：

$$\min_G\max_D L(G,D)=E_{x\sim p_{data}(x)}[\log D(x)]+E_{x\sim p_g(x)}\left[\log\left(1-D(x)\right)\right] \tag{4}$$

当令其导数为0时，可以拿到最优判别式模型为：

$$D^*(x)=\frac{P_{data}(x)}{P_{data}(x)+P_g(x)} \tag{5}$$

代入公式(4)，进行变换之后可以得到：

$$E_{x\sim p_{data}(x)}\log\frac{p_{data}(x)}{\frac{1}{2}[P_{data}(x)+P_g(x)]}+E_{x\sim p_g(x)}\log\frac{p_g(x)}{\frac{1}{2}[P_{data}(x)+P_g(x)]}-2\log 2 \tag{6}$$

这里有两个重要的相似度衡量指标，即KL散度(Kullback−Leibler Divergence)和*JS*散度(Jensen−Shannon Divergence)：

$$KL(P_{data}||P_g)=E_{x\sim P_{data}}\log\frac{P_{data}}{P_g} \tag{7}$$

$$JS(P_{data}||P_g)=\frac{1}{2}KL(P_{data}||\frac{P_{data}+P_g}{2})+\frac{1}{2}KL(P_g||\frac{P_{data}+P_g}{2}) \tag{8}$$

最后拿到了公式(6)的变形：

$$2JS(P_{data}||P_g)-2\log 2 \tag{9}$$

可以看出在最优判别式模型的情况下，将*loss*转换为最小化真实分布P_{data}和生成分布P_g之间的*JS*散度。判别式模式不停得到训练，它的能力就会得到提升，最小化生成式模型的*loss*就会越来越接近最小化P_{data}和P_g之间的*JS*散度。*JS*散度越小越好，所以优化*JS*散度可达到以假乱真。这种情况是建立在两个分布有重叠的时候才成立，但是一开始因为噪声的随机分布，与真实的数据分布相差甚远，导致两个分布可能一点重叠的部分都没有，*JS*散度就会是一个常数log2，判别式模型很快便能区分真实数据与生成的假数据，达到最优判别式模型，这就导致了梯度为0的情况。在使用梯度下降法时，对于最优判别式模型的情况，会出现生成式模型拿不到梯度，就算接近最优判别式模型，生成式模型也很有可能拿不到梯度。

对于公式(7)，将KL散度变换为D^*的形式：

$$KL(P_g||P_{data})=E_{x\sim p_g(x)}\log[1-D^*(x)]-E_{x\sim p_g(x)}\log D^*(x) \tag{10}$$

对于生成式模型的改进：

$$G_{loss}=E_{x\sim p_g(x)}[-\log D(x)] \tag{11}$$

利用公式(9)、公式(10)、公式(11)拿到最小化目标的等价变形：

$$E_{x\sim p_g(x)}[-\log D^*(x)]=KL(P_g||P_{data})-2JS(P_{data}||P_g)+2\log 2+E_{x\sim p_{data}}[\log D^*(x)] \tag{12}$$

因为$2\log 2 + E_{x\sim p_{data}}[\log D^*(x)]$不依赖生成器模型，所以最终的问题就是最小化生成的数据分布与真实的数据分布的KL散度，最大化两个分布之间的JS散度，最后可以拿到最小化目标。但是，一个在放大，一个在缩小，在数值上就形成了不稳定的梯度，也就意味着生成器模型会生成重复的样本，而且不会生成多样性复杂的样本。最终，损失优化目标在数值上的不可能，加上不稳定的梯度，再加上缺乏多样性，就会导致模型崩塌。

3 梯度惩罚(Gradient penalty)

用Wasserstein距离(又叫EM距离)替代JS散度：

$$W(P_{data},P_g)=\inf_{\gamma\sim\prod(P_{data},P_g)}E_{(x,y)\sim\gamma}[||x-y||] \tag{13}$$

在$\prod(P_{data},P_g)$中，P_{data}和P_g是全部分布的边缘分布，而$\prod(P_{data},P_g)$表示P_{data}和P_g组合的联合分布集合，γ表示其中可能的联合分布，$||x-y||$表示真实样本和生成样本的距离，EM距离代表在所有的γ下对距离的期望值取下界。与KL散度和JS散度的或大或小相比，平滑的EM距离使得两个分布即使没有重叠，EM距离依旧能够反映它们的差距，所以在使用梯度下降法来优化参数的时候，EM距离能够提供更具优势的梯度。

但是在EM距离替代JS散度的时候，对距离的期望值取下界很难实现，所以通过对偶可以变换形式：

$$W(P_{data},P_g)=\frac{1}{K}\sup_{||f||_L\leqslant K}E_{x\sim P_{data}}[f(x)]-E_{x\sim P_g}[f(x)] \tag{14}$$

常数$K\geqslant 0$，且使得条件里的所有x_1、x_2都可以得到$|f(x_1)-f(x_2)|\leqslant K|x_1-x_2|$，也就相当于函数$f$的导函数不超过$K$。这个函数也就是K−Lipschitz连续函数，它的作用就是加上一个限制条件，通过它得到了存在连续函数里最大局部变动幅度的限制。

公式(14)也就是f函数的Lipschitz常数$||f||$不超过K的时候，$E_{x\sim P_{data}}[f(x)]-E_{x\sim P_g}[f(x)]$取上界除以$K$，对公式(14)进行转换：

$$KW(P_{data},P_g)\approx\max_{w:|f_w|_L\leqslant K}E_{x\sim P_{data}}[f_w(x)]-E_{x\sim P_g}[f_w(g_\theta(z))] \tag{15}$$

原始GAN的判别式模型加入K−Lipschitz连续函数后，新的判别式模型的优化目标为：

$$L=E_{x\sim p_{data}}[f_w(x)]-E_{x\sim p_g}[f_w(x)] \tag{16}$$

两个判别式模型对比，两者的任务由二分类判断真假变成了回归任务去近似拟合Wasserstein距离，所以新的判别式模型需要拿掉sigmoid层，接着生成式模型需要最小化Wasserstein距离。因为Wasserstein距离是平滑的，所以在生成式模型方面梯度问题可以忽略，WGAN的生成式模型损失函数可以表示为：

$$G_{loss}=-E_{x\sim P_g}[f_w(x)] \tag{17}$$

判别式模型的损失函数为：

$$D_{loss}=E_{x\sim P_g}[f_w(x)]-E_{x\sim P_{data}}[f_w(x)] \tag{18}$$

可以看出公式(16)与公式(18)的相反关系，可用来表示进程的训练，损失值越小，Wasserstein距离也就越小，生成的数据分布就接近真实的数据分布，GAN训练得就好，也就代表图片的质量越高。

为了使判别式模型不过度区分相似的样本，需要在判别式模型进行参数更新后，对参数进行检查，如果超过了阈值则调回设定的区域。但是判别式模型要尽可能地区分出样本的差距，所以最终将导致参数向阈值边缘靠近。梯度惩罚，先求判别式模型的梯度，再建立与K之间的二范数实现损失函数，可以用来解决这种不停地需要调参的问题，最终拿到比较好的梯度。

4 网络结构(Network structure)

4.1 自编码器

自编码器(AutoEncoder，AE)[8]，无监督学习一类，在输入这一端有编码器，用来提取输入数据的特征，经过解码器，将得到的特征数据重新构造成输出，达到输出与输入一致。而这样的结果是没有意义的，所以可以在编码器与解码器的中间加入隐藏空间添加限制条件，使其具有其他属性。

加入隐藏层之后的自编码器将得到新的编码器和解码器，编码器由输入层与隐藏层构成，解码器由隐藏层和输出层构成。编码器将输入变成另外一种表示形式，即高维或低维，也就是将输入数据进行扩充或者是压缩，扩充后的数据具有更多的信息成分，压缩后的数据将只具备最为突出的特征。

自编码器提取数据特征的编码公式和解码公式如下：

$$h=f(W_ex+b_e) \tag{19}$$

$$\hat{x}=g(W_dh+b_d) \tag{20}$$

通过编码器将输入编码成h，再使用解码器将h重构成$\hat{x}$。

4.2 残差网络

在残差网络(Residual Network，ResNet)之前，对于那些浅层网络而言，通过叠加网络层数，可以使得模型在训练集和测试集上的性能变得更好。层数增加，性能更强，但是当网络层数到达一定数量的时候会出现“退化”的情况，层数增加，性能下降，因为梯度很难传递到前面去，这个优化问题最终导致了神经网络的“退化”。

残差块中有两条路径，输入为x，输出为$\hat{x}=H(x)$，所以$F(x)+x=H(x)$。残差学习的就是$F(x)=H(x)-x$，因为短接(shortcut)，所以保证了此处通过残差块的性能不会低于没有通过残差块的性能，也就解决了“退化”问题，如图2所示。

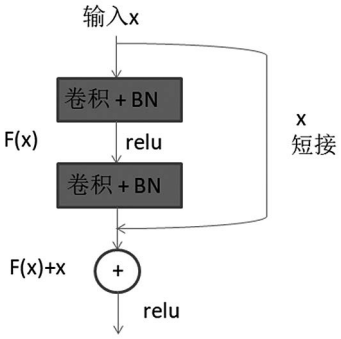

图2 残差块

Fig.2 Residual block

4.3 生成式模型结构

生成式模型由Encoder、ResNet、Decoder组成，如图3所示。

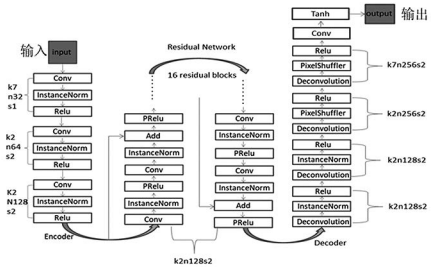

图3 生成式模型结构

Fig.3 Generator framework

编码器部分用来提取数据特征，卷积之后进行实例归一化(Instance Normalization，IN)，激活函数用Relu。

ResNet部分用来结合特征，同样使用IN层，将PRelu用作激活函数，保留特征，丢失最小的特征信息，组成新的具有相似信息的特征组合。

解码器部分将经过ResNet部分的特征数据转换成特征图。

4.4 判别式模型

判别式模型与生成式模型形成竞争关系，判别式模型的鉴别能力对生成式模型生成的图片存在质量影响。考虑到在梯度传播过程中，参数在更新的时候梯度会出现消失的情况，在卷积操作之后加入批量归一化的操作处理方式，以此来保证判别式模型的稳定训练，之后再加入Leaky-Relu激活函数(参数化修正线性单元)，保证Relu优点的时候，不会出现Dead Relu的问题。

4.5 结果

生成结果图如图4所示。

(a)原始图像　　(b)GAN生成结果

图4 原始图像与GAN生成样本的部分结果

Fig.4 Partial results of original image and samples generated by GAN

两个模型的学习率都是0.002。学习率增加，收敛速度也增加的同时会出现生成式模型倒塌，而将学习率降为0.002时，生成式模型倒塌的次数也会下降，相对稳定。

训练时间较长，生成的图像有属于失败的情况，但是对于大多数图像来说属于良好的情况。

5 结论(Conclusion)

本文将已有的自编码器与ResNet在生成对抗网络上进行动漫头像的生成，本着无监督学习在深度学习方向上对智能发展的引导，在自编码器学习特征的过程中加入判别信息，经过惩罚传递判别信息，中间残差网络提高数据特征，以此重构具有相似特征的动漫图像。

在将来如果有更好的条件将弥补本文的不足，使用性能更强的GPU增加网络参数，扩大网络规模，以期生成分辨率更高的图像，解决复杂度以及多样性问题。

参考文献(References)

[1] CRESWELL A, WHITE T, DUMOULIN V, et al. Generative adversarial networks: An overview[J]. IEEE Signal Processing Magazine, 2018, 35(1):53-65.

[2] HUANG H, YU P S, WANG C H. An introduction to image synthesis with generative adversarial nets[DB/OL]. [2018-03]. https://arxiv.org/pdf/1803.04469.pdf.

[3] 朱海峰,邵清.基于深度学习的图像风格转换研究[J].软件,2020,41(3):102-106,117.

[4] DEMIR U, UNAL G. Patch-based image inpainting with generative adversarial networks[DB/OL]. [2018-03].https://arxiv.org/pdf/1803.07422.pdf.

[5] PRAKASH C D, KARAM L J. It GAN DO better: GAN-based detection of objects on images with varying quality [DB/OL]. [2019-12]. https://arxiv.org/pdf/1912.01707.pdf.

[6] GE Y X, LI Z W, ZHAO H Y, et al. Fd-gan: Pose-guided feature distilling gan for robust person re-identification[DB/OL]. [2018-10]. https://arxiv.org/pdf/1810.02936.pdf.

[7] 张恩琪,顾广华,赵晨,等.生成对抗网络GAN的研究进展[J].计算机应用研究,2021,38(04):968-974.

[8] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7553):436-444.

作者简介：

彭章龙(1995-)，男，硕士生.研究领域：图像处理.