基于生成对抗网络的手写数字识别优化

董 飞 周 丽

四川大学锦城学院 计算机与软件学院 四川 成都 611731

【摘 要】在深度学习训练模型时如果训练数据集量少，那么训练出的模型效果通常不会很理想，而生成对抗网络（GAN）可以通过学习原始图片特征生成与其相似的新图片从而扩充数据集解决原始数据集量少的问题。本文主要研究讨论生成对抗网络（GAN）的基本组成结构和实现原理以及它在图像生成与数据扩充上的应用。文中用手写数字集识别的实验来说明GAN生成图像扩充数据集对训练出的模型的泛化能力和准确率的提升效果。

【关键词】生成对抗网络（GAN）；图像生成；手写数字

1 绪论

目前，随着人工智能的快速发展，深度学习也得到了各方面突破。在深度学习发展历程中，诞生了许多模型，在模型训练的过程中往往需要大量的训练数据集作为模型训练的基础条件，而在实际的项目过程中，收集足够量用于训练的数据集是需要大量的时间精力和财力成本的。目前图像生成模型主要有两类，分别是变分自编码器和GAN[1]。GAN是一种不需要很多标注训练数据的生成模型[2]。深度学习方法需要很多的数据集才能取得理想的识别效果[3]。GAN可以通过给定少量的数据集，模仿给定的数据集样本“仿造”与原样本相似的fake样本，这些新样本可以用作训练模型，从而有效节省了收集样本所花费的时间精力和财力成本。对抗样本是一种能够欺骗模型做出错误判断的一类样本，能够触发深度学习模型的缺陷，从而指导模型进化[4]。

2.GAN 基本原理

2.1GAN 结构图

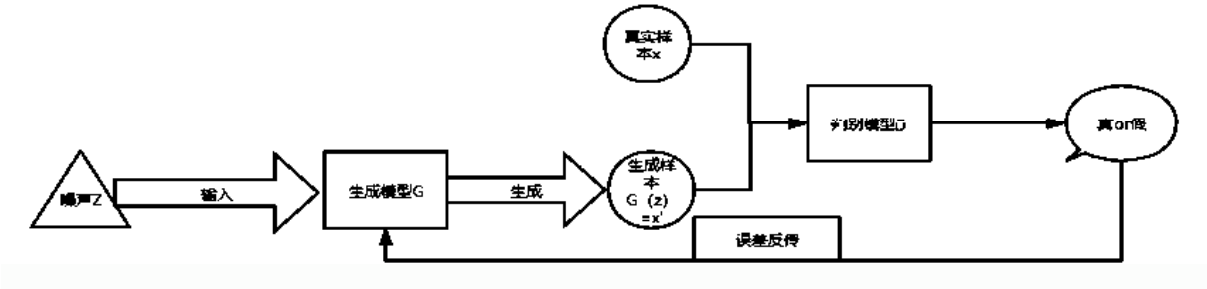

图 1：GAN 结构图

GAN由两个模型构成，判别模型和生成模型，判别模型可用于训练，也可用于测试，但生成模型只能用于测试[6]。生成模型捕捉真实样本的分布，并根据分布生成新的fake样本；判别器是判别输入是真实样本还是fake样本的二分类器[7]。模型G和D通过不断的对抗训练，使D正确判别训练样本来源，同时使G生成的fake样本与真实样本更相像[8]。

生成模型G通过输入随机噪声Z来生成与真实样本集相似的“假”样本X’。其中，生成器G是一个可微分函数，噪声Z可以是随机噪声也可以是符合某种分布的噪声，常见的噪声分布有高斯分布、均匀分布等，所生成的样本X’是与训练数据分布相似的近似分布。通过判别模型D对生成模型G所生成样本X’与真实样本X的相似性也就是真实性的不断反馈，可以使得生成模型G不断的优化改进，生成与真实样本X更相似的“假”样本X’，达到训练生成模型G的目的。

判别模型D接收来自生成模型生成的“假”样本X’或真实样本X，并对输入的样本进行判真，再将得到的结果反馈给生成模型G。其中，判别器D也是可微分函数，它的输出结果是0到1，表示输入样本是真样本数据X而不是生成器G所模拟生成的假样本X’的概率，当输出结果为一个接近0.5的值的时候表明判别模型难以判断输入数据来自生成模型G所生成的假样本X’还是来自真样本X，就可以认为生成模型G有了一个比较好的输出结果，训练结束。在每一次向判别模型D输入样本进行判断的同时，判别模型会从“假”样本X’中学习特征，达到训练判断模型D的目的。

2.2 GAN 的目标函数

GAN是生成网络和判别网络的博弈问题，判别网络D希望真实样本x的概率值越大越好，同时希望判断fake样本G（z）为真实样本的概率值越小越好，而生成网络G希望fake样本G（z）与x越相似越好，让判别网络判断其为真实样本的概率D（G（z））越高越好。两个网络相互博弈可用公式1来表示：

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

（公式 1）

式子中 Pdata（x）表示真实数据分布，Pz（x）表示噪声数据分布，D（x）和 G（z）分别表示真实样本属于真实样本的概率和噪声经过生成器后生成的样本，D（G（z））则表示 fake 样本属于真实样本的概率。D 的目标：在固定 G 的情况下，使目标函数 V 最大； G 的目标：在给定 D 的情况下，使目标函数 V 最小。

3. 实验设计与结果分析

3.1 实验背景

深度学习最常用的数据集就是 MNIST 手写体数字数据集，它包含了七万张 28*28 大小的图片，尽管该数据集很庞大，但是训练出的模型用于识别我们实验者手写的数字准确率仍然不是很高。是因为该数据集没有包含我们实验者的手写体数字，训练时没有训练到实验者手写数字体的特征，对于数据集之外的手写体数字的识别准确率就不高。 由于上述原因，用实验者的手写体数字样本 0 到 9 每个数字三十张照片作为训练样本重新进行训练，发现在测试集上的准确率同样不高。分析原因可能是训练集中实验者手写体数字样本数量过少，导致训练模型过拟合，泛化能力不强，为了增添数据集重新进行训练，选择采用生成对抗网络做图像生成和数据扩充。GAN 通过自身的不断对抗博弈，经过足够的数据训练，能够学到现实世界内在规律[5]。

3.2 实验流程

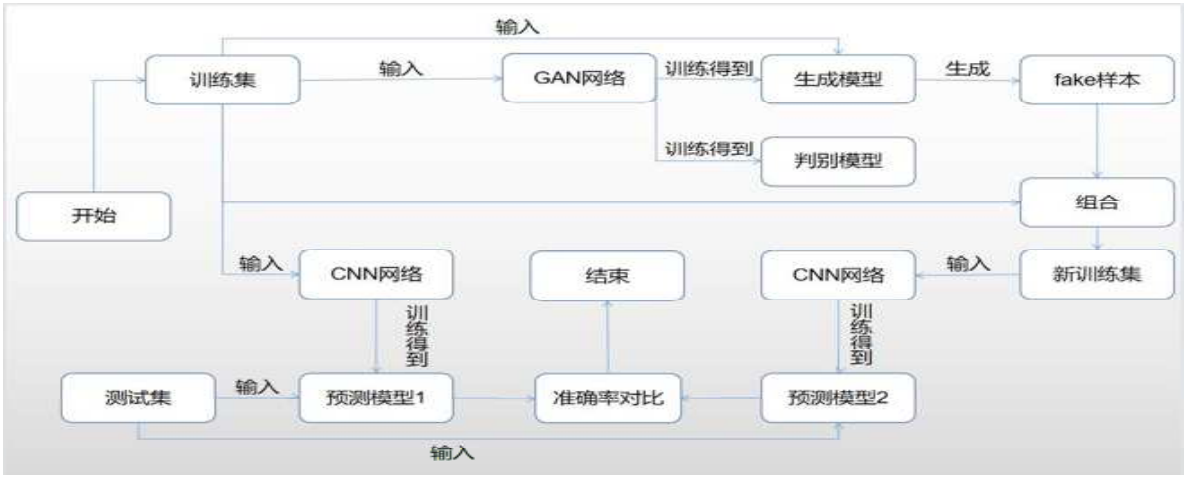

图 2：实验流程图

流程图说明：1. 先把原始训练集中的图片送入 GAN 网络训练得到 GAN 的生成模型和判别模型，同时将原始训练集中的图片送入 CNN 网络中训练得到第一个预测模型 2. 将原始训练集中的图片再送入步骤 1 中训练 GAN 后得到的生成模型中，得到若干输出的 fake 样本 3. 将步骤 2 得到的 fake 样本和原始训练集组合得到一个在原始数据集上加入了 fake 样本进行扩充后的新训练集 4. 将新的训练集送入与步骤 1 相同的 CNN 网络中训练得到第二个预测模型 5. 将测试集的图片送入步骤 1 和步骤 4 得到的两个预测模型中，对比预测准确率得到实验结论。

3.3 实验过程及实验数据

3.3.1 实验者手写数字采集

采集实验者的手写数字 0 到 9 各 30 张作为原始训练集样本，以及 0 到 9 各 20 张作为测试集，如图 3 所示：

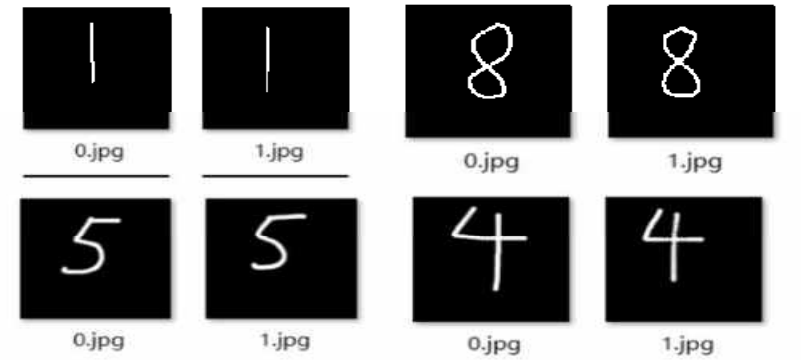

图 3：原始样本图

3.3.2 实验者手写体数字集训练模型

将原始手写体数据集放入 CNN 网络中训练得到第一个预测模型，并将测试集送入第一个预测模型得到且对应的测试集准确率。训练时损失函数选用交叉熵函数，并采用随机梯度下降算法，学习率为 0.05。测试结果如表 1 所示。

epoch	train 准确率	test 准确率
20	10%	10%
80	10%	10%
140	20%	20%
220	63%	61%
300	100%	83%

表 1

3.3.3 GAN 训练模型

将原始手写体数据集放入 GAN 网络中训练得到生成模型和判别模型，损失函数选用二分类交叉熵函数，并在梯度下降算法中加入梯度滑动平均和偏差纠正，学习率为 0.0003，训练过程如下：

GAN的训练过程是生成器和鉴定器交替训练的过程[9]。其优化过程可表示为分别对 D 和 G 进行交互迭代：先固定 G 不变，优化 D，再固定 D 不变，优化 G，直至收敛[10]。

训练 D：先从真实数据集中选择 m 个样本（x1，x2，... ...,xm）然后选择一个分布，随机生成 m 个噪声样本（z1，z2，... ...，zm），再将 m 个噪声样本输入到 G 中，得到 m 个 fake 样本（x1’，x2’，... ...,xm’），其中 xi’=G（zi），接着固定好 G 的参数，防止 G 的参数更新，再将真实样本（x1，x2，... ...，xm）标记为 1，将 fake 样本（x1’，x2’，... ...,xm’）标记为 0，使得 D 能够分辨出输入的数据来自真实样本还是虚假样本。

训练 G：先选择一个分布，随机生成 m 个噪声样本（z1，z2，... ...，zm），然后将 m 个噪声样本输入 G 中，得到 m 个 fake 样本（x1’，x2’，... ...,xm’），其中其中 xi’=G（zi），再将得到的 fake 样本送入 D 中进行判定得到损失进行反馈，接着固定好 D 的参数，防止更新 D 的参数，最后根据 D 中得到的损失更新 G，努力使 D 将 G 生成的数据分辨为真实样本。

3.3.4 用训练好的 GAN 生成模型随机生成新的手写数字

将原始手写体数据集送入训练好的生成模型中得到 fake 样本，fake 样本是根据原始样本的特征生成的，因此它与原始样本相似，但 fake 样本不具有原始样本的真实性，如图 4 所示。

图 4：fake 样本图

3.3.5 图片扩充后的训练模型

将加入了 fake 样本的手写体数据集放入 CNN 网络中训练得到第二个预测模型，并将预测集送入第二个预测模型得到测试准确率。与 3.3.2 节的不同之处在于训练数据集中加入了 GAN 生成模型随机生成的 fake 样本，测试结果如表 2 所示：

epoch	train 准确率	test 准确率
20	18%	22%
80	31%	39%
140	97%	85%
220	99%	96%
300	100%	96%

表 2

3.4 实验结论

由 3.3.2 与 3.3.5 的实验结果对比可以得出结论：在数据集较少的情况下，训练所得模型在训练集上的准确率较高，但是在测试集上的准确率较低，说明在训练集上出现了过拟合的情况，泛化性不强。而利用对抗生成网络用原始少量的数据集新生成更多特征相近的新数据集并将其加入训练集中进行模型训练后，所得模型在训练集上有不错表现的同时在测试集上的准确率也有很大的提高，有效的提高了模型的泛化能力，使得模型的准确率得到了提高。

4 论文总结

本文介绍了 GAN 的基本结构，以及它的目标函数。第三个章节进行了 GAN 的相关实验，用实验者自己手写的数字作为训练集样本进行训练，在加入 GAN 生成新样

本的前后两次训练结果中可以看出 GAN 对于图片的生成和数据集的扩充有非常大的帮助，并且新生成和扩充的图片对于提高模型的泛化能力有非常明显的效果。使用 GAN 的优点在于当训练模型拥有的数据集量较少，可以根据原有数据集扩充相似的新数据集，从而节省重新收集新数据集的时间精力和财力成本，缺点在于 GAN 自身训练需要耗费一定的时间，并且生成的新样本存在真实性不高，与真实样本还是存在一定差距。改进方法：改善 GAN 的网络结构和优化器以及损失函数等来减少训练时间并提高生成样本的真实性，使其生成的样本与真实样本更相似。

【参考文献】

[1] 陈丽芳，芦国军．基于 VAE 和 GAN 融合网络的 mnist 手写体数字图像生成方法 [J]. 廊坊师范学院学报（自然科学版），2019,19(02):25-29.

[2] 奚祥品，陈筱，朱向冰．生成对抗网络介绍及应用 [J/OL]. 无线电通信技术 :1-9[2021-05-16].

[3] 王爱丽，薛冬，吴海滨，王敏慧．基于条件生成对抗网络的手写数字识别 [J]. 液晶与显示，2020,35(12):1284-1290.

[4] 王曙燕，金航，孙家泽 .GAN 图像对抗样本生成方法 [J]. 计算机科学与探索，2021,15(04):702-711.

[5] 曹仰杰，贾丽丽，陈永霞，林楠，李学相．生成式对抗网络及其计算机视觉应用研究综述 [J]. 中国图象图形学报，2018,23(10):1433-1449.

[6] 程显毅，谢璐，朱建新，胡彬，施佺．生成对抗网络 GAN 综述 [J]. 计算机科学，2019,46(03):74-81.

[7] 李超波，李洪均，徐晨．深度学习在图像识别中的应用 [J]. 南通大学学报（自然科学版），2018,17(01):1-9.

[8] 吴少乾，李西明．生成对抗网络的研究进展综述 [J]. 计算机科学与探索，2020,14(03):377-388.

[9] 牛斌，吴鹏，马利，刘景巍．一种基于生成对抗网络的行为数据集扩展方法 [J]. 计算机技术与发展，2019,29(07):43-48.

[10] 杨大为．生成式对抗网络 GAN 及应用 [J]. 信息系统工程，2018(06):81-82.