



计算机工程与应用
Computer Engineering and Applications
ISSN 1002-8331,CN 11-2127/TP

《计算机工程与应用》网络首发论文

题目: 改进生成对抗网络在图片数据生成中的应用
作者: 孟辰, 曾志高, 朱艳辉, 朱文球, 易胜秋, 董丽君
网络首发日期: 2021-04-12
引用格式: 孟辰, 曾志高, 朱艳辉, 朱文球, 易胜秋, 董丽君. 改进生成对抗网络在图片数据生成中的应用. 计算机工程与应用.
<https://kns.cnki.net/kcms/detail/11.2127.TP.20210412.1410.025.html>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

改进生成对抗网络在图片数据生成中的应用

孟辰^{1,2}, 曾志高^{1,2,*}, 朱艳辉^{1,2}, 朱文球^{1,2}, 易胜秋^{1,2}, 董丽君^{1,2}

1. 湖南工业大学 计算机学院, 湖南 株洲 412008

2. 湖南省智能信息感知及处理技术重点实验室, 湖南 株洲 412008

摘要：图片数据生成旨在根据现有的图片数据，产生与原始图片数据分布相似的图片数据。当前主流的生成对抗网络模型（Generative Adversarial Networks, GAN）产生的图片数据质量较差，模型的训练总是遇到调试困难、训练不稳定、梯度消失、模式崩溃等一系列问题。根据稀疏表达结构和残差结构组合而成的生成器，残差结构组成的判别器，提出了一种能够生成高质量图片的 GAN 模型。根据分支网络模型构成的生成器，设计了多种类图片数据生成模型，可以使用一个模型同时训练生成多种类型的图片数据。为了更好地对数据进行训练，设计了一种动态匀速下降学习率，能够根据运行时间对学习率的衰减进行指导。在各个数据集上的实验结果表明，改进模型结构在图像数据生成上比其他算法更加稳定、鲁棒，能够生成更高质量的图片数据。

关键词：图片数据生成；生成对抗网络（GAN）；残差结构；动态学习率

文献标志码：A **中图分类号：**TP391 **doi：**10.3778/j.issn.1002-8331.2012-0429

Application of Improved Generative Adversarial Networks in Image Data Generation

MENG Chen^{1,2}, ZENG Zhigao^{1,2,*}, ZHU Yanhui^{1,2}, ZHU Wenqiu^{1,2}, YI Shengqiu^{1,2}, DONG Lijun^{1,2}

1. School of Computer, Hunan University of Technology, Zhuzhou, Hunan 412008, China

2. Hunan Key Laboratory of Intelligent Information Perception and Processing Technology, Zhuzhou, Hunan 412008, China

Abstract: Image data generation aims to generate image data with similar distribution to original image data according to existing image data. At present, the image data produced by the mainstream Generative Adversarial Networks (GAN) model is of poor quality, and the training of the model always encounters a series of problems such as debugging difficulties, unstable training, gradient disappearance, mode collapse, and so on. According to the generator composed of sparse expression structure and residual structure, and the discriminator composed of residual structure, a GAN model capable of generating high-quality pictures is proposed. According to the generator composed of the branch network model, a generation model of various kinds of picture data is designed, and one model can be used to train and generate various types of picture data at the same time. To train the data better, a dynamic learning rate with uniform decline is designed, which can guide the attenuation of the learning rate according to the running time. Experimental results on various data sets show that the improved model structure is more stable and robust than other algorithms in image data generation, and can generate higher quality image data.

基金项目：科技创新 2030—“新一代人工智能”重大项目（2018AAA0100400）；中文湖南省自然科学基金（2018JJ2098，2020JJ6089）；湖南省教育厅重点项目（19A133）。

作者简介：孟辰(1998-),男,硕士在读,研究方向为机器学习、图像处理。E-mail:mengchenmc@outlook.com; 曾志高(1973-),通信作者,男,博士,教授,研究方向为机器学习、智能信息处理, E-mail:zzgzzg99@163.com。

Key words: image data generation; generative adversarial networks (GAN);residual structure; dynamic learning rate

近年来深度学习的发展解决了很多人工无法解决的问题，但在有监督学习中，人工标注训练集需要大量的成本。直到 2014 年 Ian Goodfellow 提出了无监督学习 GAN，它能够自动完成这个过程，且拥有较低的成本和较高的效率，之后各种 GAN 的模型被相继提出^[1]。GAN 能够完成一些使用传统方法很难完成的任务。在高分辨率图片生成领域，渐进式增长 GAN(Progressive Growing of GANs,ProGAN)^[2]和条件 GAN (Analyzing and Improving the Image Quality of StyleGAN, StyleGAN2)^[3]等架构表现出了出色的效果，在文本到图像合成任务上，StackGAN^[4]也表现优秀，研究者社区已经对 GAN 及其应用广泛接纳。

GAN 的网络架构由一个生成器网络和一个判别器网络组成，生成器和判别器两个网络之间进行多次迭代，互相对抗，试图战胜对方，从而得到训练。通过生成随机的一组一维向量，经过生成器网络，可以得到所需的图像、文本、音频、视频等数据，而判别器网络需要对输入的真实数据以及生成数据进行区分。通常，GAN 中的判别网络是一个二分类的网络，将真实数据判别为 1，将生成数据判别为 0。训练开始时，判别网络可以很轻松地将生成数据和真实数据区分开来，然而生成网络经过不断地训练，判别网络越来越难将真实数据和生成数据区分开来，于是它就会训练自己的网络，使得判别能力提升，区分出真实数据和生成数据。在两个网络不断训练博弈的情况下，最终达到纳什均衡的状态^[5]——判别器的准确率到达 50%，即送来的真实数据和生成数据有一半的几率被误判，则此时的 GAN 就完成了训练。

在 GAN 的发展历程中,存在训练不稳定的问题,甚至有时完全不收敛,生成网络和判别网络的梯度在训练路径上发散。研究者们为了稳定 GAN 训练花费了很大的精力。本文中所提出的网络结构可以很好地使 GAN 的训练变得稳定,在图片数据生成中得到高质量的图片。

在本文中，我们做出了如下贡献：

(1) 改进了原始 GAN 的结构，使得网络结构具有更高的稳定性、鲁棒性，它在训练中可以有效避免模式崩塌的问题

(2) 设计了一种多分类图片数据生成的网络结构，能够满足一个网络训练多种类图片数据生成的任务。

(3) 设计了一种随时间匀速更新学习速率的算法，并将其应用在了本文的网络结构中

1 研究背景

1.1 生成对抗网络 (GAN)

GAN 的博弈过程可以看作是一位数据制假者和一位数据鉴别者的博弈，数据制假者代表了 GAN 中的生成器，数据鉴别者代表了 GAN 中的判别器。在博弈过程中，数据制假者会一直提高数据制假能力，使生成数据达到以假乱真的地步，成功骗过数据鉴别者。而数据鉴别者的任务是提高自己的鉴别能力，找出真假数据间的区别，成功判别出数据的真假，从而给真实数据打上真的标签，给生成数据打上假的标签。

博弈的最开始，生成数据和真实数据会很容易被鉴别器区分开来，在生成器网络进行训练更新之后，生成数据的质量得到提高，让判别器难以辨别出生成数据与真实数据之间的区别。同时，判别器也得到训练，使得生成数据难以通过判别。在两个神经网络不断训练进步下，生成器的生成数据的数据分布逐渐与真实数据趋于拟合，判别器无法区分真实数据与生成数据，对送来的真假数据都有一半的误判几率，这个博弈过程，如 (1) 式所示：

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

在 (1) 式中， G 代表了生成器； D 代表了判别器； V 为式子的价值函数， $V(D, G)$ 相当于表示真实数据和生成真实的差异程度， $\max_D V(D, G)$ 意思是固定住生成器 G ，尽可能地让判别器 D 能够最大

化地判别出样本来自于生成数据还是真实数据。

$\min_V(D, G)$ 表示固定住判别器 D ，训练生成器 G ，生成器 G 要将真实样本和生成样本的差异最小化，如图 1 所示。

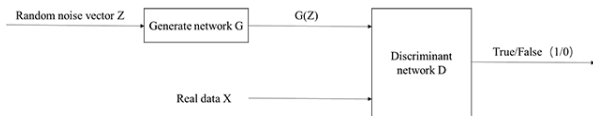


图 1 GAN 的基本结构
Fig.1 Basic structure of GAN

通过上述的 \min_G 、 \max_D 的博弈，理想情况下，生成分布会拟合于真实分布。

1.2 改进的生成对抗网络模型架构

原始 GAN 的网络层使用了全连接层进行设计，训练生成的图片数据质量很差。DCGAN^[6]使用了卷积层代替了 GAN 的全连接层，使用了带步长的卷积代替上采样层。DCGAN 虽然有着不错的架构，但是没有解决 GAN 训练的稳定性问题，训练时仍需要小心平衡生成器与鉴别器的训练进程，如图 2 所示。

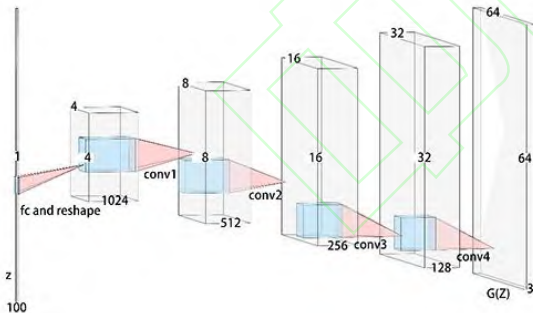


图 2 DCGAN 的网络结构
Fig.2 Network structure of dcgan

WGAN^[7]对原始 GAN 的损失函数做了改进，使得网络模型在全连接层上也能得到很好的效果，并且提出了使用沃瑟斯坦 (Wassertein) 距离作为优化方式训练 GAN，但为了使用 Wassertein 距离需要满足利普希茨 (lipschitz) 连续性，所以进行了权重限制，导致理论证明漂亮，但是实际的训练效果并不好。改进后的 WGAN-GP^[8]提出了使用梯度惩罚的方式以满足连续性条件，解决了梯度消失和梯度爆炸的问题。

改进的生成对抗网络采用了稀疏表达结构^[9]和残

差结构^[10]组合的生成器网络模型，在生成器网络中，通过两个全连接层构建了 $4*4*64$ 和 $24*24*64$ 两种尺度的特征，在其中的 $4*4*64$ 结构中构建了稀疏连接，引入了多尺度感受野 ($3*3, 24*24$) 与多尺度融合，使得网络的表征能力上升。为了防止网络梯度在反向传播时产生梯度爆炸或梯度弥散，引入了残差结构，在梯度得不到更新时，梯度会通过短接层进行更新。稀疏表达结构和残差结构组合的策略会导致模型结构复杂，但是能获得生成更高图像质量的网络模型，生成器网络结构如图 3 所示。

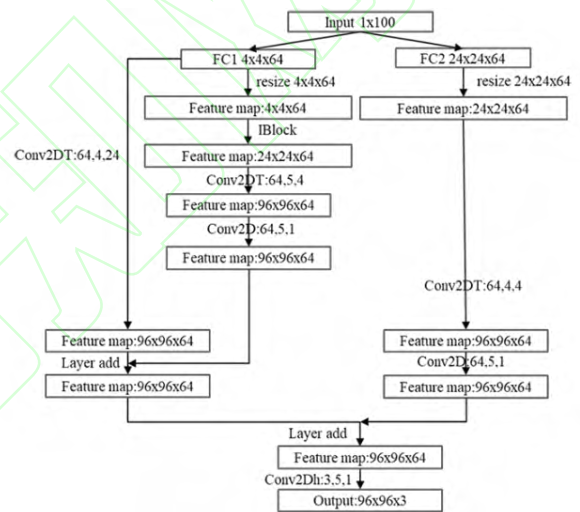


图 3 改进的 GAN (IGAN) 生成器网络结构

Fig.3 Improved network structure of GAN (IGAN) generator

在最开始，随机生成的 100 维向量，通过两个全连接层分别生成一个 $4*4*64$ 和 $24*24*64$ 大小的向量。 $4*4*64$ 的向量重塑成一个通道数量为 64 的 $4*4$ 的矩阵，在此时， $4*4$ 的矩阵经过多次卷积与反卷积生成通道数量为 64 的 $96*96$ 的矩阵。同时，它还经过了一次反卷积生成了另一个通道数量为 64 的 $96*96$ 的矩阵，生成的两个矩阵在通道数量维度进行相加，在后续梯度得不到更新的时候，将会通过残差层跳过中间层更新梯度。在右边，通道数量为 64 的 $24*24$ 的矩阵也通过反卷积生成通道数量为 64 的 $96*96$ 的矩阵，然后与左边生成的通道数量为 64 的 $96*96$ 的矩阵在通道数量维度相加，生成一个通道数量为 64 的 $96*96$ 的矩阵。最后，通道数量为 64 的 $96*96$ 的矩阵经过一个卷积层，

生成需要得到的 $96 \times 96 \times 3$ 的图像。

改进的生成对抗网络构建了一组过完备特征集合。在原始的 GAN 中, 100 维特征向量经过全连接生成了特征集合 $F(X) = \{x_1, x_2, x_3 \dots, x_{(4 \times 4 \times 64)}\}$, 原始特征集合 $F(X)$ 经过网络模型不断地进行反卷积生成了输出特征集合 $F(Y) = \{y_1, y_2, y_3 \dots, y_{(96 \times 96 \times 3)}\}$, 此时 $length_{F(X)} < length_{F(Y)}$ 。改进的生成对抗网络在最初通过两个全连接层生成了两个特征集合, 一个为 $F(X) = \{x_1, x_2, x_3 \dots, x_{(4 \times 4 \times 64)}\}$, 另一个特征集合为 $F(X') = \{x'_1, x'_2, x'_3 \dots, x'_{(24 \times 24 \times 64)}\}$, 此时, $length_{F(X)} + length_{F(X')} > length_{F(Y)}$, 它比之原始的 GAN 抽象出了更多的基础特征, 这也是第二个全连接层取 $24 \times 24 \times 64$ 的原因。

改进的生成对抗网络采用了多种不同的卷积核 ($24 \times 24, 4 \times 4$), 使得网络拥有了不同的感受野, 最后进行的两次拼接意味着多种尺度特征的融合。网络层中的短接层, 使得训练后期网络仍然能够得到更新。

在每一个卷积 (Conv2D) 层和反卷积^[11] (Conv2Dtranspose) 层的后面接了一个批规范化层^[12] (BatchNormalization, BN) 层和激活函数层 (Leaky Relu^[13]) 层, 最后一个 Conv2D 层后面接一个 BN 层和激活函数层 (tanh) 层, 如图 4 所示。

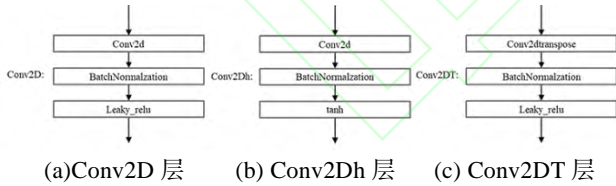


图 4 模块名称

Fig.4 Module name

在实际任务中, 有时需要训练生成多种类图像数据, 若每次只训练一种数据, 则花费的时间会比较漫长。在多种类图像数据生成中, 需要完成的任务比较复杂, 需要使用更深的网络进行学习, 故沿用 IGAN 的网络结构进行拓展, 在每一层反卷积结束后添加一层通道数相同步长为 1 的卷积, 增加网络的表征能力, 并在 $24 \times 24 \times 64$ 的模块后连接相应分类数量分支网络, 并与另一模块生成的分支模块进行连接, 得到如图 5 所示的二分类图像数据生成网络结构。

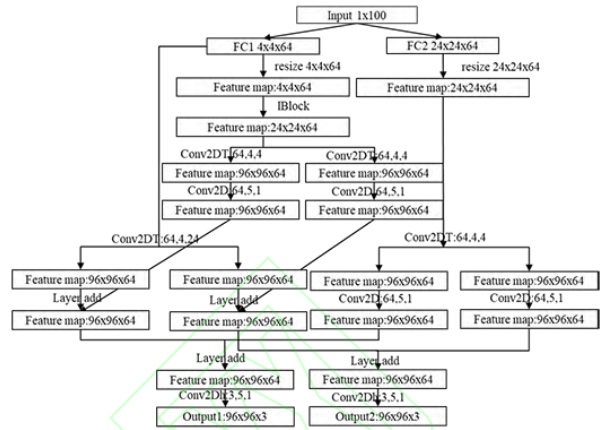


图 5 二分类图像数据生成网络结构(IGAN_V2)

Fig.5 Binary image data generation network structure (IGAN_V2)

网络中的 IBlock 结构由两个上卷积层构成, 如图 6 所示。

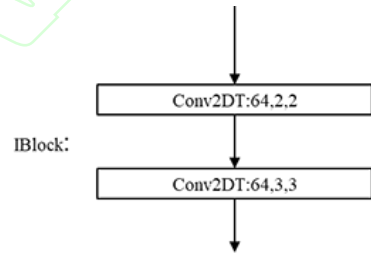


图 6 IBlock 网络结构

Fig.6 IBlock network structure

由此基本结构可以拓展出更多图片种类数据生成的网络结构, 图 7 为 N 分类图像数据生成网络。

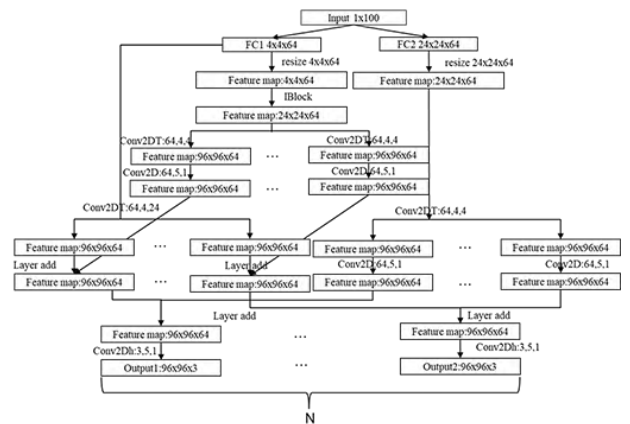


图 7 N 分类图像数据生成网络结构(IGAN_VN)

Fig.7 N classification image data generation network structure (IGAN_VN)

判别器采用如图 8 的网络结构, 它将输入的图片

数据经过一个残差网络进行分类，判断输入数据为生成数据还是真实数据。在多分类图像数据生成中，生成几分类图片数据，就使用多少个判别器：

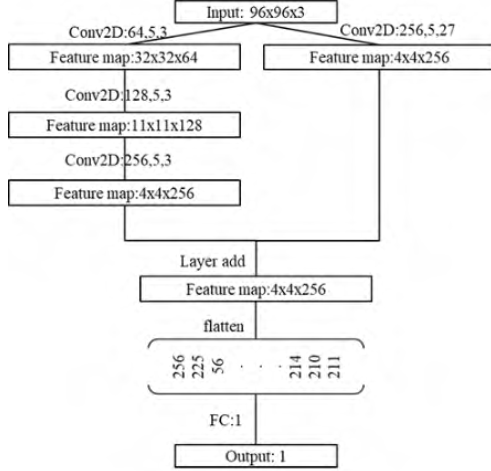


图 8 IGAN 判别器网络结构

Fig.8 Network structure of IGAN discriminator

可以看到，判别器网络结构比生成器网络结构要简单，其原因是因为判别数据的难度要比生成数据的难度要小的多，如果使用复杂的判别网络，判别损失会迅速下降到一个很小的数值，使得网络不再更新，同时也为了降低对显卡的要求。如图 9 所示，网络中的每个卷积块包含一个卷积层和一个批归一化层以及 LeakyRelu 层。在之后还会采取一些调参技巧（trick）来平衡网络大小不平衡的问题。

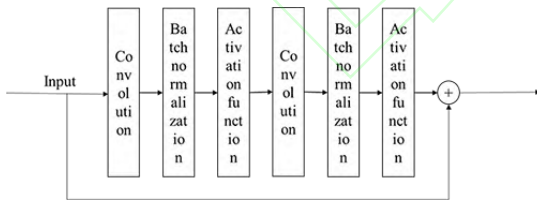


图 9 残差块基本结构

Fig.9 Basic structure of residual block

2 研究方法

2.1 实验数据准备

本文使用的数据集为 Fashion-MNIST, MNIST, CelebA, 以及 LSUN-bedrooms 数据集, 实验中的 LSUN 数据集以及 CelebA 数据集都经由程序处理成统一的 96*96 的大小, 可将数据预处理或在训练中实时处理, 为了在程序运行时, 可以减小程序资源的开销, 节省时间, 本文采用了预先处理。

2.2 网络中的参数设置

2.2.1 使用动态学习速率

在 GAN 的训练中, 生成器最开始拟合的数据分布距离真实数据分布较远, 需要使用较大的学习速率进行训练, 加快网络的收敛^[14]。随着生成的数据分布不断接近真实分布, 学习率需要随着训练的进行不断减小, 提高网络的稳定性, 避免学习率过大跳过最小值。

在学习率的选择上, 一般采取以下四种方式:

1、使用固定的学习速率。

$$init_lr = 0.01 \quad (2)$$

2、每迭代 step 次后学习率减少 gamma 倍。

$$lr = lr \times gamma \quad (3)$$

3、学习率呈多项式曲线下降。

$$LR(t) = init_lr \times \left(\frac{t}{T}\right)^{power} \quad (4)$$

4、学习率随迭代次数增加而下降。

$$LR(t) = init_lr \times (1 + gamma \times t)^{power} \quad (5)$$

实验中采用了第四种策略的变体。

$$LR(t) = end_lr + \frac{1}{(init_set + gamma * t)} \quad (6)$$

此时:

$$end_lr = 0.0002$$

$$init_set = 3000$$

$$gamma = 10$$

在 (6) 式中, end_lr 为最后的学习速率,

$end_lr + \frac{1}{init_set}$ 为最初的学习速率, $gamma$ 为学

习速率衰减率, 通过控制 end_lr 、 $init_set$ 、 $gamma$ 可以动态地控制学习速率, 图 10 为学习速率衰减曲线。

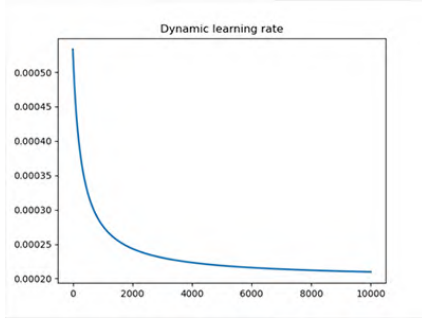


图 10 动态学习速率曲线

Fig.10 Dynamic learning rate curve

2.2.2 单侧标签平滑

在原始 GAN 的训练中,分类器的标签只使用 0 或者 1,0 代表生成数据,1 代表真实数据。标签平滑是指将标签处理成 α (真), β (假),而不是采取 0(假)和 1(真)的处理方式。此时, GAN 的损失函数变为式 (7):

$$L = E_{p_{data}}(x)\{alb[D(x)]\} + E_{p_{z \sim N(0,1)}}(z)\{\beta lb[(1 - D(z))]\} \quad (7)$$

在实际训练中,我们去掉了 \log 函数,此时, GAN 的损失函数变为式 (8):

$$L_{label} = E_{p_{data}}(x)\{\alpha lb[D(x)]\} + E_{p_{z \sim N(0,1)}}(z)\{\beta lb[(1 - D(G(z)))]\} \quad (8)$$

其最优解 $D(x)$ 变为:

$$D(x) = \frac{\alpha p_{data}(x)}{\alpha p_{data}(x) + \beta p_g(x)} \quad ((0 \leq \alpha \leq 0.1), (0.9 \leq \beta \leq 1)) \quad (9)$$

真实数据和生成数据都需要进行平滑处理,有助于 GAN 进入纳什均衡状态、损失函数得到收敛、增加程序的鲁棒性。

2.2.3 批归一化 (Batch Normalization)

批归一化技术是将特征向量归一化,把非线性函数映射后向取值区间饱和和区靠拢的输入分布,调整至均值为 0, 方差为 1 的比较标准的正态分布,使得非线性变换函数的输入值落入对输入比较敏感的区域,避免梯度消失。该技术可提高学习过程的稳定性,以及解决权重值初始化效果差的问题。将该技术作为预处理步骤应用于神经网络的隐藏层,有助于缓解内部协变量转移的问题,如图 11 所示。

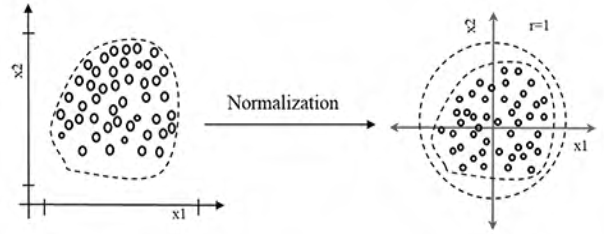


图 11 数据经过 BN 层后数据分布示意图

Fig.11 Diagram of data distribution after data passing through batch normalization (BN) layer

批归一化需要对所有隐藏层应用,而不仅仅是输入层使用。

2.2.4 工程手段

在实际应用中,一些调参技巧对图像质量的改进拥有不错的效果,在本文实验中,采取了如下调参技巧。

- 在判别器进行数据输入时,将输入图像的数值范围限制在 $[-1,1]$,在最后输出图像时再还原成 $[0,255]$,归一化限制了输入向量的最大值跟最小值不超过隐层跟输出层函数的限定范围。

- 在生成数据以及真实数据送入判别器之前,给其分别加入一个高斯噪声^[15],这样可以有效提高程序的鲁棒性。

- 给所有的随机数设置一个随机数种子,以便于之后的实验复现。

- 由于网络复杂度的不平衡,在训练一次生成器之后训练三到五次判别器,可以有效平衡两个网络。

- 在反卷积时使用 PixelShuffle 和转置卷积进行上采样^[16],避免较大化池用于下采样,使用带步长的卷积,减少梯度稀疏。

- 给生成器与判别器设置不同的学习速率^[17],通常生成器的学习速率设置的大一点,实验中将生成器学习速率设置为判别器的 1.5 倍。

- 激活函数选择使用 Leaky_Relu 代替 Relu 避免梯度稀疏^[18]:

$$\sigma_{Relu}(x) = \max(0, x)$$

$$\sigma_{Leaky_relu} = \begin{cases} x, & \text{若 } x > 0 \\ 0.02x, & \text{否则} \end{cases} \quad (10)$$

激活函数示意图如图 12 所示:

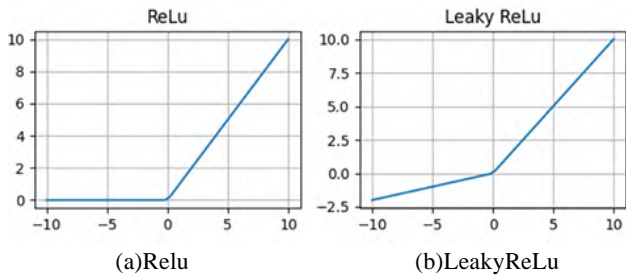


图 12 激活函数示意图

Fig.12 Activation function diagram

•选择 ADAM 作为优化器加快训练。

•实际代码中用反转标签来训练 G 更有效^[19]，即让生成图片使用真实的标签来训练，而真实数据使用生成数据的标签进行训练。

此时，GAN 的损失函数变为：

$$L_{label} = E_{p_{data}}(x) \{ \alpha \ln[1 - D(x)] \} + E_{p_{z \sim N(0,1)}}(z) \{ \beta \ln[D(G(z))] \} \quad (11)$$

((0 ≤ α ≤ 0.1), (0.9 ≤ β ≤ 1))

•在发现损失值不变时，或者生成图像质量不好时，不要终止训练，GAN 的训练非常耗时，多等等会得到好的结果。但是在判别器的损失快速接近 0 的时候，可以终止训练，调整网络，重新训练。

•加入一个惩罚项来惩罚那些和历史平均权重相差过多的权重值，本文采用的是 WGAN-GP 中的梯度惩罚，梯度惩罚损失函数如式 (12) 所示，其中 \tilde{x} 为真实样本， x 为生成样本。

$$L_{gp} = \lambda E_{\tilde{x} \sim p_z} \left[\left(\nabla_{\tilde{x}} D(\hat{x})_2 - 1 \right)^2 \right] \quad \left(\nabla_{\tilde{x}} = \frac{\partial D}{\partial x} \right)$$

$$\hat{x} = t\tilde{x} + (1-t)x \quad (0 \leq t \leq 1) \quad (12)$$

•使用随机失活技术—DropOut，也是为了引入一定的噪声，在生成器的某几层中使用 DropOut，增加网络的鲁棒性^[20]，如图 13 所示。

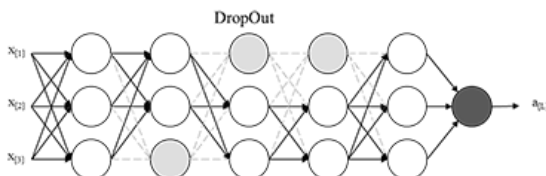


图 13 DropOut 示意图

Fig.13 Dropout diagram

3 实验

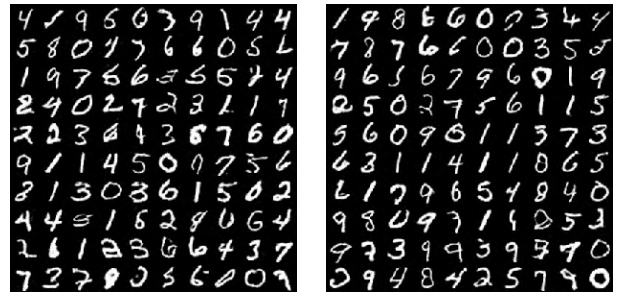
3.1 实验配置

生成网络和判别网络架构由 1.2 节所示。损失函

数 $L = L_{label} + L_{gp}$ ， L_{gp} 中惩罚因子 λ 设置为 5。一轮训练大小设置为 64，因为数据集大小过大，一个轮次设置为一个批次数据的学习时间，而不是跑完整个数据集所用的时间。学习速率设置为动态学习速率，每一百轮保存一次权重以及可视化输出结果查看训练情况，每 30 轮保存一次损失值。所用机器为 NVIDIA 2070s，内存 32G，使用的深度学习的框架为 tensorflow2.0。

3.2 简单数据集训练情况

MNIST（手写数字数据集）包含了 60000 个用于训练的示例和 10000 个用于训练的示例，这些数字已经经过标准化并位于图像中心，为值 0 到 1 的 28*28 大小的图像。Fashion-MNIST（服饰数据集）包含了 70000 张 28*28 的灰度图像，其中包含了 60000 个图片的训练集以及 10000 个图片的测试集，共包含了 10 类图像。在 MNIST（手写数字识别数据集）和 Fashion-MNIST（服饰数据集）等简单数据集上，IGAN 和 IGAN_v2 都能取得不错的结果，图 14 为在 IGAN 网络和 IGAN_v2 网络上 MNIST 的生成结果，图 15 为在 IGAN 网络和 IGAN_v2 网络上 Fashion-MNIST 生成结果。

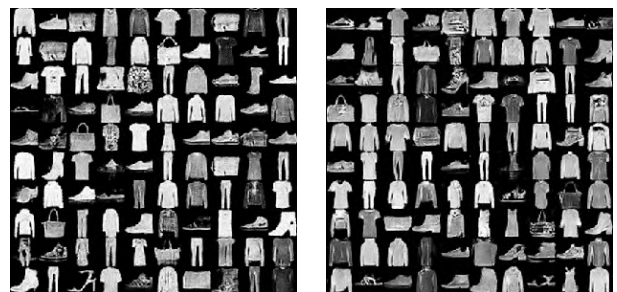


(a)IGAN

(b)IGAN_v2

图 14 模型在 MNIST 数据集上生成结果

Fig.14 The model generates results on the MNIST dataset



(a)IGAN

(b)IGAN_v2

图 15 模型在 Fashion-mnist 数据集上生成结果

Fig.15 The model generates results on fashion MNIST dataset

在训练这类简单数据集的时候，如直接使用 0.0002 的学习速率，判别网络的损失将会锐减至很小，导致不能指导生成网络的参数进行更新，所以我们需要将最开始的学习速率调成 0.002，然后逐步衰减至 0.0002，可使用 2.2.1 节中的动态学习速率。

我们可以看到,经过 IGAN 与 IGAN_v2 的网络，生成的结果可以清楚地判别出它的类别，生成质量优秀。

3.3 名人人脸属性数据集（CelebA）训练情况

在 GAN 中，应用最多和最全的便是人脸数据生成,数据集我们采用了名人人脸属性数据集—CelebA，其中包含了 10177 个名人身份的 202599 张人脸图片，该数据集经过预处理成 96*96*3 大小后分别送入 IGAN 与 IGAN_v2 网络进行训练，得到了如图 16、图 17 的高质量人脸图片。



图 16 IGAN 在 CelebA 数据集上生成结果

Fig.16 IGAN generates results on CelebA dataset



图 17 IGAN_v2 在 CelebA 数据集上生成结果

Fig.17 IGAN_v2 generates results on CelebA dataset

在图中可以发现，人物的肤色光影可以很好地生成出来，这说明图像数据生成的质量和多样性能得到保证。而且在训练中，网络生成的背景趋于同一种颜色，保证了生成图像的真实性。

3.4 大规模图像数据集（LUSN）训练情况

LUSN Dataset 是一个大规模图像数据集，包含了 10 个场景类别和 20 个对象类别，共计约 100 万张标记图像，本文使用了 LUSN Dataset 数据集中的 bedding-room 类别的训练部分数据集。在这些数据集中，能否生成优秀的图像是一个大的挑战。

我们将图片预处理成 96*96*3 大小的数据集，然后经过 IGAN 与 IGAN_v2 神经网络的训练，分别得到如图 18、图 19 所示的生成样本。



图 18 IGAN 在 LUSN 数据集上生成结果

Fig.18 IGAN generates results on LUSN dataset



图 19 MGAN_v2 在 LUSN 数据集上生成结果

Fig.19 IGAN_v2 generates results on LUSN dataset

可以看到，图像中的窗户、门、床，甚至是挂饰也可以比较不错地生成，但是其中也会出现一些图片

训练的有些失败，这也是在训练 GAN 时必然会碰到的问题，即图像越复杂，生成难度越高，生成质量越差。

3.5 实验评估

3.5.1 评估方式介绍

在 GAN 进行了图像生成任务后，需要一个度量指标来衡量生成的图像是否优秀，在进行一项任务的度量时，不能完全靠主观判断对图像进行是否优秀的判断，我们从以下两方面来考虑：

（1）图片自身的质量。（是否逼真，是否清晰，内容是否完整）

（2）多样性。在一项任务中，生成网络所产生的图片是多种多样的，如果只产生一种或者几种类型的图像，将会出现模式崩溃（Mode collapse）的现象。

在进行图像评价时，一般使用如下两种常用的评价指标，IS（Inception Score）和 FID（Fr é chet Inception Distance）^[21]。

但是 IS 评价标准的 Inception V3 权重是在 ImageNet 下训练出来的，而本文中我们使用了其他四个数据集生成图像，这种跨数据集地计算 IS 是错误的，所以在利用 ImageNet 以外的数据集计算 IS 时，得出的结果没有意义。

为了克服 IS 评价的缺点，本实验中使用 FID 进行评价。FID 可以计算真实图像和生成图像的特征向量距离的一种度量，特征向量由 Inception V3 Network 所得到，它使用的是 Inception V3 Network 分类网络倒数第二个全连接层的输出的 1*1*2048 维图像特征向量进行距离度量。

这个距离可以衡量真实图像和生成图像之间的相似程度，FID 值越小，相似程度越高，在 FID=0 时，两个图像相同。

使用式（13）计算数据的 FID：

$$d^2((\mu_r, \Sigma_r), (\mu_g, \Sigma_g)) = \|\mu_r - \mu_g\|^2 + \text{Tr}(\Sigma_r + \Sigma_g - 2\left(\Sigma_r \Sigma_g\right)^{\frac{1}{2}}) \quad (13)$$

其中，Tr 表示矩阵对角线上元素之和——迹（trace）。均值为 μ 协方差为 Σ 。此外 r 表示真实图片数据，g 表示生成图片数据。

FID 对模式崩溃更加敏感。相比较 IS 来说，它的

对比用到了生成数据和真实数据，比 IS 更加灵活，而且可以使用不同的数据集进行训练。

3.5.2 评分结果

受限于设备和数据集的大小，不能将整个数据集与生成的图像进行对比，为保证公平性，我们将抽取数据集中的一部分样本与生成样本进行 FID 计算，抽取样本的大小为 10000，实验结果如表 1 所示。

表 1 中文
Table 1 English

名称	MNIST	Fash-ion-MNIST	CelebA	LUSN-bedroom
IGAN	6.56	14.16	32.19	71.90
IGAN_v2	9.77	27.48	31.22	76.43
WGAN-GP	7.80	14.95	34.51	140.94
DCGAN	7.61	16.31	35.08	176.20

由表 1 可以看出，IGAN 在简单以及复杂的数据集上表现都很优异，而 IGAN_v2 在简单数据集上的效果并不理想，因为它复杂的网络在小数据集上很难得到收敛，而在复杂数据集上，它的表现和单分类生成同样优秀。

3.6 主观评价

在图像生成领域最直观的还是主观评价，客观量化至今还存在着各种问题。图 20、图 22 为 IGAN 生成图像，图 21、图 23 为 IGAN_v2 生成图像，图 24、图 25 为 WGAN-GP 生成图片，图 26、图 27 为 DCGAN 生成图像。



图 20 IGAN CelebA 数据集生成图像

Fig.20 IGAN generates results on CelebA dataset



图 21 IGAN_v2 CelebA 数据集生成图像

Fig.21 IGAN_v2 generates results on CelebA dataset



图 22 IGAN LUSN-bedroom 数据集生成图像

Fig.22 IGAN generates results on LUSN dataset



图 23 IGAN_v2 LUSN-bedroom 数据集生成图像

Fig.23 IGAN_v2 generates results on LUSN dataset



图 24 WGAN-GP CelebA 数据集生成图像

Fig.24 WGAN_GP generates results on CelebAdataset



图 25 WGAN-GP LUSN-bedroom 数据集生成图像

Fig.25 WGAN_GP generates results on LUSN dataset



图 26 DCGAN CelebA 数据集生成图像

Fig.26 DCGAN generates results on CelebAdataset



图 27 DCGAN LUSN-bedroom 数据集生成图像

Fig.27 DCGAN generates results on LUSN dataset

从上图可以看到,四组生成结果之中,IGAN 与 IGAN_v2 网络的生成效果最佳,DCGAN 中还出现了模式崩溃的问题,生成了模糊重复的图片。在客观量化手段还不能完全替代主观判断的今天,主观判断仍然是图像生成质量评价的一大手段。

3.7 结论

本节实验在两种类型的数据集之间进行了对比,分别是简单数据集——手写数字数据集(MNIST)、服饰数据集(Fashion-MNIST)、以及复杂数据集——名人人脸属性数据集(CelebA)、大规模图像数据集(LUSN)。在简单数据集上,对比于其他的网络模型,IGAN 与 IGAN_v2 几乎不需要进行调参,而其他的 GAN 需要经过精心的调参才能达到接近的效果。在复杂的数据集上,IGAN 和 IGAN_v2 达到了最佳生成质量效果,这和它复杂强大的网络模型是分不开的。在进行真实实验场景应用时,如果需要生成简单数据集数据且没有相应的计算资源时,可以使用传统 GAN 架构。在进行复杂图像任务,并且拥有计算资源的情况下,推荐使用 IGAN 的架构,它更适合复杂的图像任务。

4 未来研究方向

在完成了现有的工作后,可以发现,GAN 在很多领域具有广阔的应用前景,甚至在一些领域还是一片蓝海。

•在进行有监督学习的时候,神经网络需要大量的

数据,而获取这些数据的成本却十分高昂,耗时耗力。在只有一个小规模数据集时,一旦训练好一个 GAN,就可以使用 GAN 来生成大量的数据,用来进行训练,可以节省很多精力。

•在设计领域,设计过程不仅耗时耗力,而且需要具备专业技能。借助人工智能和 GAN,我们可以进行各种各样的设计,或者生成初始数据供设计师启发灵感,可以节省大量的时间成本。

•在多分类图像生成网络结构中,由于生成器结构复杂,且每一分类的图像生成对应一个不同的判别器,导致特别耗费 GPU 资源,在之后可以对网络结构进行优化,使得计算资源的使用更加合理。

近些年来,研究者们研发出了数以百计的 GAN 架构,GAN 的领域不断变得丰富,我将继续研究,并设计新的架构,不断完善 GAN 的能力,争取早日使自己设计的 GAN 的应用落地。

参考文献:

- [1]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]. Computer science and exploration, 2020, 14 (1): 1-17
- [2] Mahalingaiah K, Matichuk B. ProDeblurGAN: Progressive Growing of GANs for Blind Motion Deblurring in Face Recognition[C]//International Conference on Smart Multimedia. Springer, Cham, 2019: 439-450.
- [3] Karras T, Laine S, Aittala M, et al. Analyzing and improving the image quality of stylegan[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 8110-8119..
- [4]Huang X, Li Y, Poursaeed O, et al. Stacked generative adversarial networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 5077-5086.
- [5] Oliehoek F A, Savani R, Gallego J, et al. Beyond local nash equilibria for adversarial networks[C]//Benelux Conference on Artificial Intelligence. Springer, Cham, 2018: 73-89.
- [6] Radford A, Metz L, Chintala S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks[J]. Computer ence, 2015.
- [7] Gulrajani I, Ahmed F, Arjovsky M, et al. Improved training of wasserstein gans[J]. Advances in neural information processing systems, 2017, 30: 5767-5777.
- [8]Wei X, Gong B, Liu Z, et al. Improving the improved training of wasserstein gans: A consistency term and its dual effect[J]. arXiv e-prints, 2018, arXiv:1803.01541.
- [9] Barratt S, Sharma R. A note on the inception score[J]. arXiv preprint arXiv:1801.01973, 2018.
- [10] Wu Z, Shen C, Van Den Hengel A. Wider or deeper: Re-

-
- visiting the resnet model for visual recognition[J]. Pattern Recognition, 2019, 90: 119-133.
- [11] Sage D, Donati L, Soulez F, et al. DeconvolutionLab2: An open-source software for deconvolution microscopy[J]. Methods, 2017, 115: 28-41.
- [12] Santurkar S, Tsipras D, Ilyas A, et al. How does batch normalization help optimization?[J]. Advances in neural information processing systems, 2018, 31: 2483-2493.
- [13] Dubey A K, Jain V. Comparative Study of Convolution Neural Network's ReLu and Leaky-ReLu Activation Functions[M]//Applications of Computing, Automation and Wireless Systems in Electrical Engineering. Springer, Singapore, 2019: 873-880.
- [14] De Z, Guozhang L, Huaiguang W, et al. Improvement of adaptive learning rate for pose estimation[J]. Journal of Electronic Measurement and Instrumentation, 2019.
- [15] Sønderby C K, Caballero J, Theis L, et al. Amortised map inference for image super-resolution[J].arXiv e-prints, 2016, arXiv:1610.04490.
- [16] Wenzhe Shi, Jose Caballero, Ferenc Husz á r, Johannes Totz, Zehan Wang. Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network[J].10.1109/CVPR.2016.207.
- [17]Heusel M, Ramsauer H, Unterthiner T, et al. Gans trained by a two time-scale update rule converge to a local nash equilibrium[J]. arXiv e-prints, 2017, arXiv:1706.08500.
- [18]卢倩雯, 陶青川, 赵娅琳,等. 基于生成对抗网络的漫画草稿图简化[J]. 自动化学报, 2018, 44(005):840-854
- Lu Q W, Tao Q C, Zhao Y L. Sketch Simplification Using Generative Adversarial Networks[J]. 自动化学报, 2018, 44(005):840-854
- [19]Guo T, Xu C, Huang J, et al. On positive-unlabeled classification in GAN[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 8385-8393.
- [20] Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: A Simple Way to Prevent Neural Networks from Overfitting[J]. Journal of Machine Learning Research, 2014, 15(1):1929-1958.
- [21]Xu Q, Huang G, Yuan Y, et al. An empirical study on evaluation metrics of generative adversarial networks[J]. arXiv e-prints, 2018: arXiv: 1806.07755.