



多尺度特征融合的双判别器残差生成对抗网络

管凤旭, 路斯棋, 郑岩

引用本文:

管凤旭,路斯棋,郑岩. 多尺度特征融合的双判别器残差生成对抗网络[J]. 智能系统学报, 2023, 18(5): 917–925.

GUAN Fengxu, LU Siqi, ZHENG Yan. Dual discriminator residual generation adversarial network with multiscale feature fusion[J]. *CAAI Transactions on Intelligent Systems*, 2023, 18(5): 917–925.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202207005>

您可能感兴趣的其他文章

基于生成对抗网络的人脸口罩图像合成

Masked face image synthesis based on a generative adversarial network

智能系统学报. 2021, 16(6): 1073–1080 <https://dx.doi.org/10.11992/tis.202012010>

基于风格转换的无监督聚类行人重识别

Clustering approach based on style transfer for unsupervised person re-identification

智能系统学报. 2021, 16(1): 48–56 <https://dx.doi.org/10.11992/tis.202012014>

利用残差密集网络的运动模糊复原方法

Image restoration with residual dense network

智能系统学报. 2021, 16(3): 442–448 <https://dx.doi.org/10.11992/tis.201912002>

基于生成式对抗网络的道路交通模糊图像增强

Enhancement of blurred road-traffic images based on generative adversarial network

智能系统学报. 2020, 15(3): 491–498 <https://dx.doi.org/10.11992/tis.201903041>

基于生成对抗网络的机载遥感图像超分辨率重建

Super-resolution reconstruction of airborne remote sensing images based on the generative adversarial networks

智能系统学报. 2020, 15(1): 74–83 <https://dx.doi.org/10.11992/tis.202002002>

基于小样本学习的LCD产品缺陷自动检测方法

An automatic small sample learning-based detection method for LCD product defects

智能系统学报. 2020, 15(3): 560–567 <https://dx.doi.org/10.11992/tis.201904020>

DOI: 10.11992/tis.202207005

网络出版地址: <https://kns.cnki.net/kcms2/detail/23.1538.TP.20230516.1840.004.html>

多尺度特征融合的双判别器残差生成对抗网络

管凤旭, 路斯棋, 郑岩

(哈尔滨工程大学 智能科学与工程学院, 黑龙江 哈尔滨 150001)

摘要: 生成对抗网络 (generative adversarial networks, GANs) 作为一类基于深度学习的无监督生成模型, 无需对数据分布进行建模便可以生成真实且质量较高的图像。标准的 GANs 往往训练困难, 常出现梯度消失、梯度爆炸或者模式崩溃等问题, 限制模型的性能。为解决模式崩溃问题, 本文提出一种双判别器结构来提高模型生成图像的多样性。另外, 本文改进了生成器模型和判别器模型, 提出一种基于残差网络和多尺度特征融合的生成器和基于多尺度特征融合的判别器, 在提高生成图像质量的前提下解决深层网络出现的梯度消失、梯度爆炸的问题。将其应用于 MNIST、LSUN、CelebA 数据集上, 训练结果稳定且生成图像质量较高, 取得了令人满意的 FID 和 IS 值。

关键词: 生成对抗网络; 深度学习; 无监督模型; 模式崩溃; 梯度爆炸; 梯度消失; 多尺度特征融合; 训练稳定性
中图分类号: TP18 **文献标志码:** A **文章编号:** 1673-4785(2023)05-0917-09

中文引用格式: 管凤旭, 路斯棋, 郑岩. 多尺度特征融合的双判别器残差生成对抗网络 [J]. 智能系统学报, 2023, 18(5): 917-925.

英文引用格式: GUAN Fengxu, LU Siqi, ZHENG Yan. Dual discriminator residual generation adversarial network with multiscale feature fusion[J]. CAAI transactions on intelligent systems, 2023, 18(5): 917-925.

Dual discriminator residual generation adversarial network with multiscale feature fusion

GUAN Fengxu, LU Siqi, ZHENG Yan

(School of Intelligent Science and Engineering, Harbin Engineering University, Harbin 150001, China)

Abstract: Generative adversarial networks (GANs), a class of unsupervised generative models based on deep learning, have shown promising results in generating realistic and high-quality images without modeling the data distribution. However, standard GANs are often difficult to train and often suffer from gradient disappearance, gradient explosion, or mode collapse, which can limit the overall performance of the model. To address the mode collapse problem and improve the variety of the generated images, this paper proposes a dual discriminator structure. Furthermore, this paper improves the generator and discriminator model and proposes a generator based on residual network and multiscale feature fusion and a discriminator based on multiscale feature fusion, which effectively solves the problem of gradient disappearance and gradient explosion that occurs in deep networks while improving the quality of generated images. The proposed approach is applied to various datasets, including MNIST, LSUN, and CelebA. The training results reveal that the stability and quality of the generated images are high, achieving satisfying FID and IS values.

Keywords: generative adversarial networks; deep learning; unsupervised model; mode collapse; gradient explosion; gradient disappearance; multiscale feature fusion; training stability

传统的近似法^[1-2]、马尔可夫链法^[3-4]等生成模型学习到的数据具有人类可以理解的分布规律, 但是这种分布规律不能被计算机很好地理解。

此外, 由于数据分布的复杂性, 在对其进行数据建模时往往存在一定的困难, 导致生成图像质量不高。2014 年 Goodfellow 等^[5]借鉴了零和博弈中二人博弈的思想, 提出一种基于对抗思想的无监督生成模型——生成对抗网络 (generative adversarial networks, GANs), 并持续受到相关学者们

收稿日期: 2022-07-06. 网络出版日期: 2023-05-17.

基金项目: 国家自然科学基金项目 (62101156).

通信作者: 管凤旭. E-mail: guanfengxu@hrbeu.edu.cn.

的关注。由于无需对复杂数据进行建模便可以取得质量较高的生成效果, GANs 及其衍生模型近年来被应用于各个领域, 并且取得了巨大的成功^[6-12]。除此之外, GANs 能在深度学习领域引起如此巨大的轰动, 还因为 GANs 在理论方面具有许多引人思考的地方^[13], 如 Casanova 等^[14] 从核密度估计技术中获得启发, 提出一种对复杂数据集进行建模的生成对抗网络 (instance-conditioned generative adversarial networks, IC-GAN)。IC-GAN 不仅获得了更优秀的生成效果, 而且也为迁移学习的发展提供重要的贡献。

但是, 标准的 GANs 在训练过程中是通过判别器 D 的辅助, 寻找生成器和判别器两个模型损失函数的最小值点, 这相比于传统的只训练一个网络的情况来说会存在训练上的困难^[15]。标准的 GANs 网络在训练的过程中, 由于判别器相较于生成器收敛速度更快, 通常很快就达到收敛点, 此时判别器将很难继续正确地指导生成器。而相比于判别器模型, 生成器模型往往具有更深的网络结构, 由于缺少了有效指导, 使得生成器往往会出现梯度消失现象^[16]。GANs 面临的另一个严重问题是模式崩溃问题, 表现为生成器所生成的结果单一, 缺乏多样性。这是因为传统的判别器能力较弱, 在小批次训练中往往不能正确地分数数据来源^[17], 从而导致在反向传播过程中无法给予生成器有意义的指导, 使生成器更倾向于生成能够获得判别器更高评分的分布结果。正是由于这些错误的“指导”, 导致生成器生成自认为正确的分布来欺骗判别器, 输出质量极差的生成结果, 同时产生模式崩溃现象。同时, 由于训练初期判别器的性能较弱, 这也为模式崩溃现象的发生提供了可能。对于生成器来说, 由于生成器在图像生成过程中每步独立进行生成, 下一层无法知晓上一层所学习到的特征, 导致生成过程中存在一定程度上的信息丢失, 生成的图像具有极大相似性且导致生成图像具有不连续的语义结构、质量较差。

针对标准 GANs 存在的模式崩溃等一系列问题, 学者们从不同的角度提出改进方案。而在众多的改进方案中, 主流思路有两种。一是对模型拓扑结构进行改进, 优化生成器、判别器模型, 如 Radford 等^[18] 提出的深度卷积生成对抗网络 (deep convolution generative adversarial networks, DCGAN), 通过引入深度卷积来提高模型对特征的学习能力, 增强训练的稳定性, 提高生成图像的质量; Karras 等^[19] 提出的渐进式增长的生成对抗网络 (progressive growing of GAN, proGAN) 采用逐层渐

进式生成方法, 随着训练轮次的提高不断增加生成模型的结构来逐步提高高分辨率图像的生成质量以及多样性, 解决模型训练稳定性的问题; 史彩娟等^[17] 提出的残差生成对抗网络 (residual generative adversarial networks, Re-GAN) 模型在生成模型中引入残差网络思想进行特征融合, 有效提高了训练稳定性, 缓解梯度消失现象。Mirza 等^[20] 提出的条件生成对抗网络 (conditional generative adversarial networks, cGAN)、Odena 等^[21] 提出的辅助分类器生成对抗网络 (auxiliary classifier generative adversarial networks, AC-GAN) 以及 Zhou 等^[22] 提出的 Omni-GAN 等模型, 通过增加辅助信息来控制生成类别, 解决模式崩溃等问题; Nguyen 等^[23] 提出双判别器模型 (dual discriminator generative adversarial networks, D2GAN) 来解决模式崩溃问题, D2GAN 中一个判别器会给符合分布的数据样本给予高奖励, 而另外一个鉴别器却更喜欢生成器生成的数据。DCGAN 等模型虽然在一定程度上提高了生成图像的质量, 减轻了模式崩溃等问题, 但是没有从根本上解决梯度问题。比如当 DCGAN 的层数增加时, 会造成梯度消失的情况。

另一种是通过改进损失函数、增加惩罚项、进行不同正则化等方法来提高 GANs 训练稳定性, 解决模式崩溃问题。如 Arjovsky 等^[24] 提出的瓦瑟斯坦生成对抗网络 (Wasserstein generative adversarial networks, WGAN) 将原始 GANs 模型中的损失函数替换为 W-距离, 在一定程度上解决了模式崩溃问题; 带梯度惩罚的 WGAN (WGAN-GP)^[25] 通过加入梯度惩罚项, 有效地解决了 WGAN 中存在的梯度爆炸问题; Miyato 等^[13] 提出的光谱归一化生成对抗网络 (spectral normalization generative adversarial networks, SN-GAN) 通过谱范数归一化方法, 有效解决了 WGAN 对判别器限制条件过强而无法满 1-Lipschitz 性质的问题, 提高了训练的稳定性; Bhaskara 等^[26] 提出的分段梯度归一化生成对抗网络 (piecewise gradient normalization for generative adversarial networks, GraN-GAN) 通过对输入进行梯度正则化处理, 使在应用 W-距离作为损失函数的 GANs 模型训练中训练更加稳定。但是, WGAN 及其变体所采用的 BN 模型在小批量训练中无法使用^[17]。

受 Re-GAN 的启发, 本文将多尺度特征融合引入到生成器模型, 在转置卷积中应用不同尺寸感受野以及步长, 通过通道融合方式进行特征融合, 避免图像生成过程中的特征丢失。受 D2GAN 的启发, 引入双判别器模型, 来解决训练前期由

于判别器能力不足引起的模式崩溃现象。同时采用多尺度特征融合思想, 提高判别器对不同特征的感知能力, 以提高其对输入图像来源的判断能力, 在反向传播时为生成器提供更多有意义的信息。通过对生成器和判别器结构的改进, 本文提出基于多尺度特征融合的双判别器残差生成对抗网络 (dual discriminator residual generation adversarial network with multi-scale feature fusion, DDRM-GAN) 模型。

1 生成对抗网络

生成对抗网络 (GANs) 是一类无需对复杂数据进行建模, 便可以从数据集中学习到数据分布的无监督生成类网络模型。简单来说, GANs 由生成器 G 和判别器 D 两个模块构成。生成器 G 从真实数据样本中捕捉数据潜在的分布规律, 生成一组由随机数构成的伪样本, 判别器 D 尽可能地区分输入样本来自于真实分布还是伪分布。通过生成器 G 与判别器 D 的相互对抗, 不断提高生成器的生成样本能力, 最终达到纳什均衡。因此, 生成对抗网络整体的训练目标是减少伪样本与真实样本之间的差异。传统 GANs 的网络模型结构如图 1 所示。

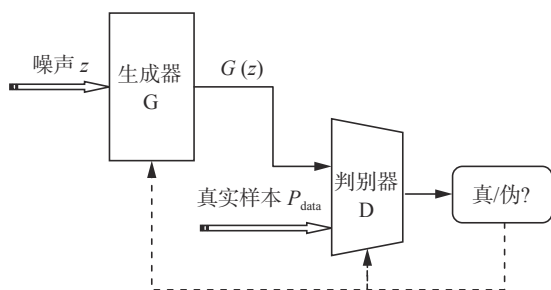


图 1 生成对抗网络结构

Fig. 1 Structure of generative adversarial network

2 DDRM-GAN

本文提出的基于多尺度特征融合的双判别器残差生成对抗网络 (DDRM-GAN) 与传统的 GANs

模型相比, 具有以下两个方面的优势。一方面, 将 ResNet^[27] 与多尺度特征融合思想引入到生成器模型, 使生成模型在图像生成过程中所学习到的特征传递过程更加连续, 在一定程度上避免了由于学习独立性所引起的特征丢失, 使生成图像语义更加连续、种类更加多样。另一方面, 改进判别器网络结构, 使判别器能够更好地分辨输入图像来源, 同时使用双判别器模型, 为生成器提供更有价值的指导, 避免模式崩溃、梯度消失等问题的发生。

2.1 DDRM-GAN 的生成器模型

为了应对生成器各层独立学习这一特点, 提高模型对生成图像的全局把控并减小图像生成过程中的特征丢失, 本文根据残差网络思想采用双尺寸卷积核的转置卷积来进行特征传递, 利用 DenseNet^[28] 结构中的通道 concatenate 操作进行特征融合, 可以更好地利用不同尺度特征图的语义信息, 通过增加 channel 的方式实现较好的性能。图 2 为 DDRM-GAN 的生成器模型在生成 64 像素×64 像素尺寸图像时的拓扑结构。

在 64 像素×64 像素尺寸的图像生成任务中, DDRM-GAN 的生成器中包含 5 个卷积核尺寸为 2×2 的转置卷积层、3 个卷积核尺寸为 4×4 的转置卷积层以及 1 个卷积核尺寸为 1×1 的转置卷积层, 进行了 3 次特征叠加操作, 将输入的 (100,1) 维高斯随机噪声映射为 (64,64,3) 维的 RGB 图像或者 (64,64,1) 的灰度图像。在整个生成模型中, 利用转置卷积对图像进行上采样操作, 在转置卷积操作之后应用批标准化操作 (batch normalization) 来防止梯度爆炸以及梯度消失现象, 同时可以加快网络收敛, 防止过拟合情况发生。在对通道维度进行改变时, 采用卷积核尺寸和步长均为 1 的卷积操作进行, 保证网络可以学习到对应特征的权重, 保留有用的特征, 在一定程度上避免模式崩溃现象发生。在对激活函数的选择上, 中间层采用 Leaky-Relu 函数, 输出层采用 Tanh 函数, 以此提高生成图像的质量, 同时一定程度上保证了训练的稳定性。

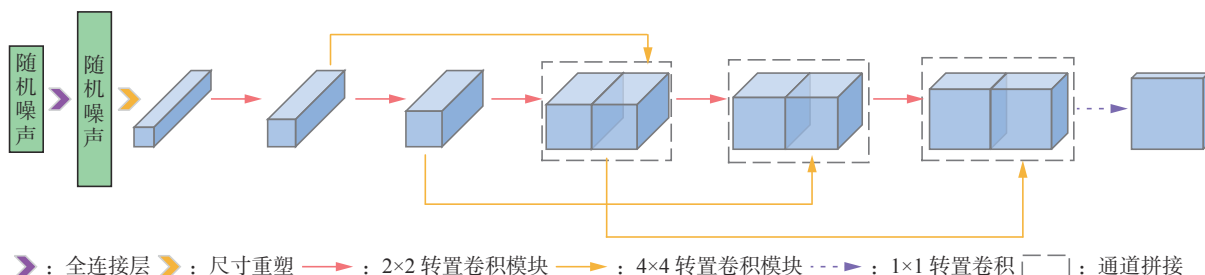


图 2 DDRM-GAN 的生成器结构

Fig. 2 Structure of the generator of DDRM-GAN

2.2 DDRM-GAN 的判别器模型

传统的生成对抗网络中,因为在训练初期判别器的能力不强,可能会对生成器进行错误的引导,从而导致模式崩溃等现象发生。为了解决这一问题,本文使用双判别器模型,共同对生成器进行指导。

在双判别器模型中,其中一个判别器用于对原数据集以及生成器生成的数据进行判定。而另一个判别器的输入则是生成器图像与原数据集经过同一图像简单处理后(如上下镜像翻转等)的图像。理想情况下,伪图像与真实图像经过同一简单处理后,判别器无法判别其来自哪个分布。

为了增加判别网络对不同尺度特征的提取,提高判别器的判别能力,本文采用多尺度特征融合结构来代替传统的多层卷积网络,网络拓扑如图 3 所示。判别器首先对输入图像进行一次卷积操作,提取图像特征。随后将此特征分别用 3×3 、 5×5 以及 7×7 3 种不同尺寸的卷积核进行特征提取。将三路特征进行特征融合,将融合后的特征进行一次卷积操作后输入到全连接层,输出判别概率。实验证明,多尺度特征提取网络不仅可以提高对输入来源判断的准确率,在反向传播过程中

给予生成器更有意义的指导信息,并且在生成更高分辨率图像的时候也表现出令人满意的效果。

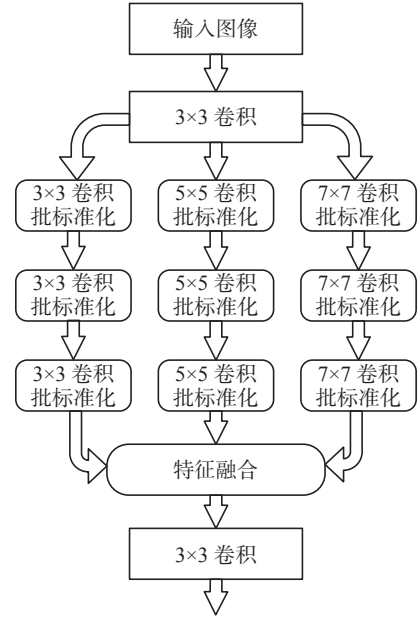


图 3 DDRM-GAN 的判别器模型结构

Fig. 3 Structure of the discriminator of DDRM-GAN

与标准的 GANs 网络不同,DDRM-GAN 由两个判别器网络与一个生成器网络共同组成。DDRM-GAN 的整体拓扑结构如图 4 所示。

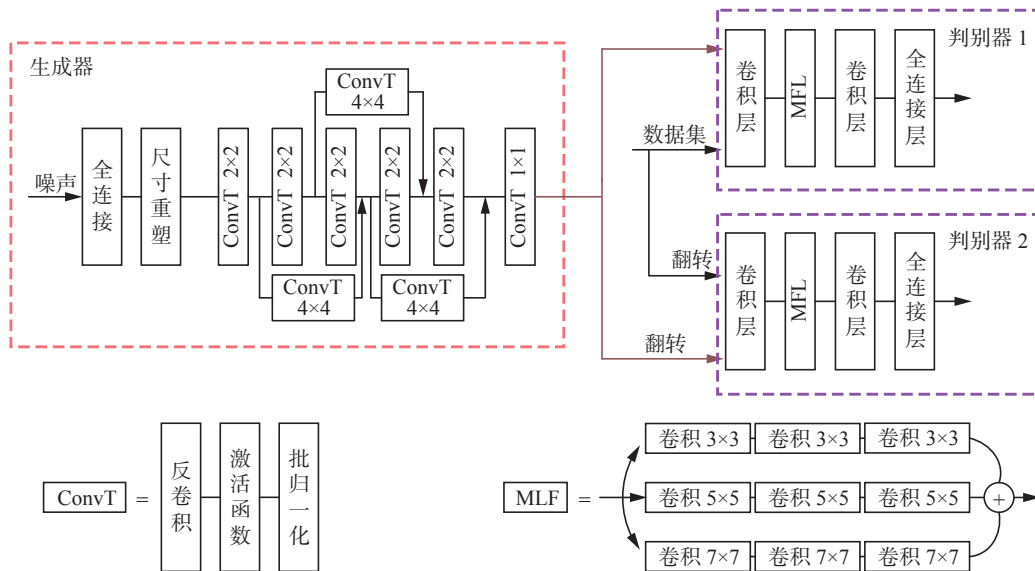


图 4 DDRM-GAN 的网络结构

Fig. 4 Network structure of DDRM-GAN

2.3 DDRM-GAN 的损失函数

判别器与生成器的损失函数采用标准的对抗损失^[5],不同的是生成器的损失函数在原始 GAN 的基础上增加了 D_2 项。DDRM-GAN 中判别器的损失函数如下所示:

$$\max_{D_1} L(D_1) = E_{x \sim p_{\text{data}}} [\log D_1(x)] + E_{z \sim p_z(z)} [\log (1 - D_1(G(z)))] \quad (1)$$

$$\max_{D_2} L(D_2) = E_{x \sim p'_{\text{data}}} [\log D_2(x)] + E_{z \sim p_z(z)} [\log (1 - D_2(G'(z)))] \quad (2)$$

式中: $G'(z)$ 为 $G(z)$ 经过镜像翻转所得图像, p'_{data} 为

真实图像数据经过镜像翻转后的数据分布。

生成器的损失函数如下:

$$\min_G L(G) = E_{z \sim p_z(z)} [\log(1 - D_1(G(z)))] + E_{z \sim p_z(z)} [\log(1 - D_2(G'(z)))] \quad (3)$$

3 DDRM-GAN 的生成实验

3.1 数据集和评估指标

为了验证 DDRM-GAN 的有效性, 分别在 MNIST、CelebA 及 LSUN 数据集上进行了测试。在测试中所应用的评估指标分别为: (fréchet inception distance, FID) 和 (inception score, IS)。

MNIST 数据集是一个具有 70 000 张手写数字的图像数据集, 其中包含 60 000 张训练集图片和 10 000 张测试集图片, 是计算机视觉领域最基础、应用最广泛的数据集之一。CelebA 数据集是一个包含 20 多万张人脸图像的大型数据集, 被广泛应用于图像生成、人脸检测等领域的训练。LSUN 是一个场景理解图像数据集, 主要包含了卧室、固房、客厅、教室等场景图像, 是计算机视觉领域的重要数据集之一。此次实验采用 LSUN 数据集中的 tower 数据集, 共 70 万张不同尺寸、不同场景的训练集数据。

IS 从信息熵的角度对生成图像进行评估, 它可以对生成图像的质量以及多样性进行衡量。但是由于 IS 指标并没有考虑真实的数据影响, 因此不能反应生成图像是否逼近真实图像, 在某些角度来说没有意义。

FID 是图像生成领域对生成图像进行评估时最常用的指标, 它可以度量两个图片数据集的相似度。而且, FID 对噪声更加鲁棒, 更适合描述 GANs 模型的多样性。

3.2 训练过程

此次实验中, 为证明 DDRM-GAN 的有效性, CelebA、LSUN 以及 MNIST 数据集的图像在不同模型中均被转换成 64×64 分辨率进行训练。设定

输入噪声为 100 维服从高斯分布的随机序列, 学习速率设置为 0.0005, 设置 batch 大小为 64, 迭代次数为 500 000 轮。

训练时, 将数据集图像与生成器生成图像送入到正向判别器 D_1 , 将数据集数据与生成器生成数据分别进行上下翻转操作后送入到翻转判别器 D_2 , 每一轮依次对翻转判别器 D_2 、正向判别器 D_1 、生成器 G 进行训练, 并且对每个模型进行参数更新。训练后, 取最优权重进行最终的单幅图像输出。算法的步骤如下:

首先, 初始化 G 、 D_1 、 D_2 权重

其次, 设置迭代轮次并循环执行以下操作:

- 1) 从 $z \sim N(0, 1)$ 中采样, 利用 G 生成伪图像数据 x_i ;
 - 2) 将 x_i 进行图像上下翻转操作, 得到翻转图像 x'_i ;
 - 3) 从 χ 中采样真实数据 x_j ;
 - 4) 从 χ' 中采样翻转的真实数据 x'_j ;
 - 5) 将 x_i 和 x_j 送入到 D_1 , 并且计算 D_1 的损失函数 L_{D1} ;
 - 6) 将 x'_i 和 x'_j 送入到 D_2 , 并且计算 D_2 的损失函数 L_{D2} ;
 - 7) 根据式 (3), 计算生成器 G 的损失函数 L_G ;
 - 8) 利用 Adam 算法依次更新 D_1 、 D_2 、 G 的权重。
- 循环结束后, 选择最优权重并保存模型。

3.3 实验结果

为了对比 DDRM-GAN 的性能, 将 DDRM-GAN 与 GAN、WGAN、WGAN-GP、DC-GAN 分别在 MNIST、LSUN 以及 CelebA 数据集上进行实验。采用相同的训练次数, 取训练过程中最优的生成器权重, 生成 20 000 张伪图像并且与原始数据集进行 FID 以及 IS 指标计算。

不同的模型在不同的训练集上训练后, 生成结果的 FID 与 IS 指标如表 1 所示, 表中黑体表示每一列中的最优结果。

表 1 不同 GANs 在 MNIST、LSUN 以及 CelebA 数据集的测试结果
Table 1 Test results of MNIST, LSUN and CelebA datasets on different GANs

基线模型	MNIST		LSUN		CelebA	
	FID	IS	FID	IS	FID	IS
GAN	88.636	2.582±0.043 2	79.623	2.582±0.032 2	95.316	2.376±0.022 8
WGAN	83.254	2.641±0.012 5	77.252	2.207±0.015 2	83.286	2.537±0.032 4
WGAN-GP	79.624	2.952±0.026 3	73.075	3.025±0.021 3	66.472	2.317±0.045 2
DC-GAN	73.684	2.836±0.022 6	77.037	2.407±0.026 5	73.087	2.255±0.065 3
DDRM-GAN	62.427	3.216±0.025 4	57.925	3.682±0.025 1	40.243	2.878±0.002 9

表 1 的数据显示不同模型的性能, 本文的模型 (DDRM-GAN) 在 3 个数据集上都取得了最好的成绩。在 MNIST 手写数据集上, DDRM-GAN 的 FID 指标相比于 DC-GAN 降低了 15.28%, IS 指标相较于 WGAN-GP 提高了 8.94%。在 LSUN 数据集上, DDRM-GAN 的 FID 指标相比于 WGAN-GP 降低了 20.73%, IS 指标提高了 21.72%。在 CelebA 数据集上, DDRM-GAN 的 FID 指标相

比于 WGAN-GP 降低了 39.46%, IS 指标相较于 WGAN 提高了 13.44%。

图 5~7 分别给出了不同模型在 CelebA、MNIST 和 LSUN 数据集上的生成图像。可以看出, 在 3 个数据集上, 与传统的 GAN、WGAN 等模型相比, DDRM-GAN 模型生成的图像质量最高, 可分辨性最强而且生成的图像种类更加丰富。直观来看, DDRM-GAN 的性能更胜一筹。



图 5 不同模型在 CelebA 数据集上的实验结果

Fig. 5 Experimental results of different models on the CelebA dataset

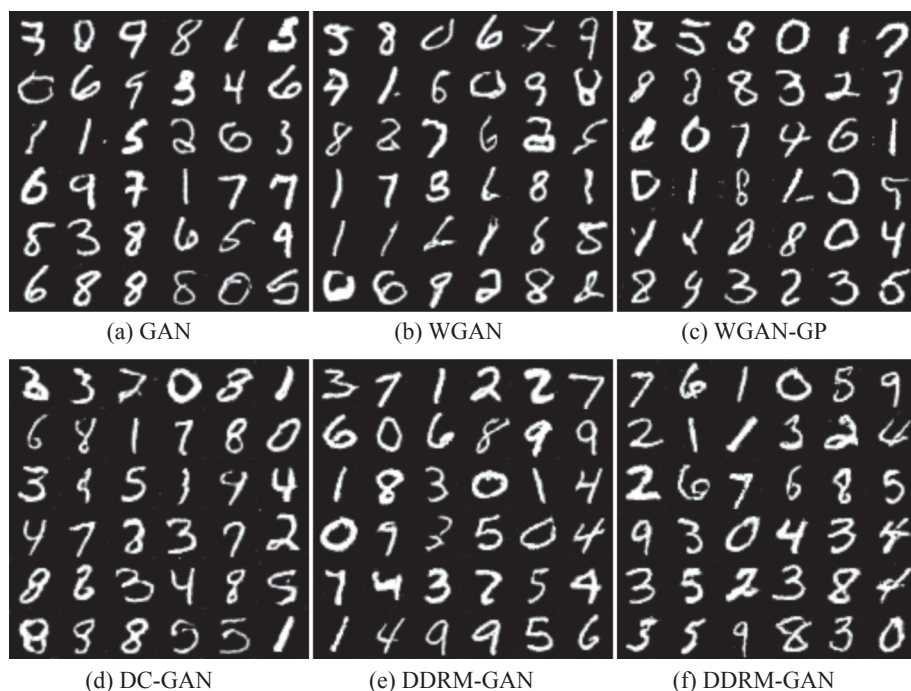


图 6 不同模型在 MNIST 数据集上的实验结果

Fig. 6 Experimental results of different models on the MNIST dataset



图 7 不同模型在 LSUN 数据集上的实验结果

Fig. 7 Experimental results of different models on the LSUN dataset

3.4 消融实验

本文在 CelebA 数据集上进行消融实验, 进一步证明提出的 DDRM-GAN 有效提高了生成图像的质量和多样性。消融实验的内容包括:

1) 完整模型去除判别器中的多尺度特征融合结构 (Full-mf);

2) 完整模型去除双判别器结构, 使用单一判别器进行生成实验 (Full-dd);

3) 完整模型去除生成器中的复杂连接 (Full-cc)。

表 2 给出了消融实验的结果。为解决生成对抗网络的模式崩溃问题, 提高生成图像的多样性, 本文引入双判别器模型。相比于完整的模型, 在去除双判别器结构之后, 在 CalabA 数据集上, FID 指标提升了 74.02%, IS 指标下降了 4.54%。为提高模型的生成质量, 本文引入了多尺度特征融合的判别器结构和基于 ResNet 的多尺度特征融合生成器结构。

表 2 消融实验结果对比

Table 2 Comparison of ablation experiment results

基线模型	FID	IS
Full	40.243	2.878±0.0029
Full-mf	52.016	2.326±0.0013
Full-dd	70.032	2.753±0.0016
Full-cc	51.074	2.125±0.0023

在完整模型去除多尺度判别器模型, 使用单一卷积结构的判别器模型时, FID 指标上升了 29.25%, IS 指标下降了 23.73%; 在去除生成器复杂连接后, FID 指标上升了 26.91%, IS 指标下降了 35.44%。总的来说, DDRM-GAN 在保证生成图像质量的条件下, 有效地提升了生成伪图像的多样性, 缓解了模型的模式崩溃现象。

4 结束语

为解决标准 GANs 存在的训练困难——如模式崩溃、梯度消失、梯度爆炸等问题, 本文提出了基于多尺度特征融合的双判别器残差生成对抗网络 DDRM-GAN。将 ResNet 思想与 DenseNet 的特征融合方法应用于生成器, 搭建了可以掌握图像全局特征、减少特征丢失的生成器, 解决在生成图像过程中特征丢失的问题。将多尺度特征融合思想应用于判别器网络, 并且将数据集与生成图像进行简单图像处理提出双判别器模型, 共同完成对生成器的指导工作, 缓解了模式崩溃以及梯度消失等问题, 提高了模型的鲁棒性。将模型在 MNIST、LSUN 以及 CelebA 数据集上进行测试, DDRM-GAN 均表现出良好的性能, 取得了令人满意的 FID 以及 IS 值。但是随着双判别器以及多尺度融合思想的引入, 模型参数量提高, 训练速度下降, 这也是 DDRM-GAN 需要改进的地方。

参考文献:

- [1] KINGMA D P, WELLING M. Auto-encoding variational Bayes[C]//2nd International Conference on Learning Representations. Banff: ICLR, 2014: 1–14.
- [2] REZENDE D J, MOHAMED S, WIERSTRA D. Stochastic backpropagation and approximate inference in deep generative models[C]//2014 International Conference on Machine Learning. SAN DIEGO: JMLR, 2014: 1278–1286.
- [3] HINTON G E, SEJNOWSKI T J, ACKLEY D H. Boltzmann machines: constraint satisfaction networks that learn[J]. *Cognitive science*, 1994, 9: 1–40.
- [4] ACKLEY D H, HINTON G E, SEJNOWSKI T J. A learning algorithm for boltzmann machines[J]. *Cognitive science*, 1985, 9(1): 147–169.
- [5] GOODFELLOW I, POUGET-ABADIE J, MIRZA M. Generative adversarial nets[C]//Proceedings of the 2014 Conference on Advances in Neural Information Processing Systems 27. Montreal: Curran Associates, 2014. 2672–2680.
- [6] RUAN Congcong, YUAN Liuchun, HU Haifeng, et al. Image translation with dual-directional generative adversarial networks[J]. *IET computer vision*, 2021, 15(1): 73–83.
- [7] CHEN Xiacong, LI Yun, YAO Lina, et al. Generative adversarial U-net for domain-free medical image augmentation[EB/OL]. (2021–01–12)[2022–07–06]. <https://arxiv.org/abs/2101.04793>.
- [8] KONG Jiahui, SHEN Haibin, HUANG Kejie. DualPathGAN: facial reenacted emotion synthesis[J]. *IET computer vision*, 2021, 15(7): 501–513.
- [9] 张冀, 曹艺, 王亚茹, 等. 融合 VAE 和 StackGAN 的零样本图像分类方法 [J]. *智能系统学报*, 2022, 17(3): 593–601.
- ZHANG Ji, CAO Yi, WANG Yaru, et al. Zero-shot image classification method combining VAE and StackGAN[J]. *CAAI transactions on intelligent systems*, 2022, 17(3): 593–601.
- [10] 王凯旋, 任福继, 倪红军, 等. 基于循环互相关系数的 CGAN 温度值图像扩增 [J]. *智能系统学报*, 2022, 17(1): 32–40.
- WANG Kaixuan, REN Fuji, NI Hongjun, et al. Image amplification for temperature value image based on cyclic cross-correlation coefficient CGAN[J]. *CAAI transactions on intelligent systems*, 2022, 17(1): 32–40.
- [11] 姜义, 吕荣镇, 刘明珠, 等. 基于生成对抗网络的人脸口罩图像合成 [J]. *智能系统学报*, 2021, 16(6): 1073–1080.
- JIANG Yi, LYU Rongzhen, LIU Mingzhu, et al. Masked face image synthesis based on a generative adversarial network[J]. *CAAI transactions on intelligent systems*, 2021, 16(6): 1073–1080.
- [12] 尹诗, 侯国莲, 胡晓东, 等. 基于 AC-GAN 数据重构的风电机组主轴温度监测方法 [J]. *智能系统学报*, 2021, 16(6): 1106–1116.
- YIN Shi, HOU Guolian, HU Xiaodong, et al. Temperature monitoring method of the main bearing of wind turbine based on AC-GAN data reconstruction[J]. *CAAI transactions on intelligent systems*, 2021, 16(6): 1106–1116.
- [13] MIYATO T, KATAOKA T, KOYAMA M, et al. Spectral normalization for generative adversarial networks [EB/OL]. (2018–02–16)[2022–07–06]. <https://arxiv.org/abs/1802.05957>.
- [14] CASANOVA A, CAREIL M, VERBEEK J, et al. Instance-conditioned GAN[C]//35th Conference on Neural Information Processing Systems. North torrey: NIPS, 2021.
- [15] CHAVDAROVA T, FLEURET F. SGAN: an alternative training of generative adversarial networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 9407–9415.
- [16] KANG M, SHIM W J, CHO M, et al. Rebooting ACGAN: auxiliary classifier GANs with stable training[C]//Neural Information Processing Systems. Virtual: NeurIPS, 2021: 1–13.
- [17] 史彩娟, 涂冬景, 刘靖祎. Re-GAN: 残差生成式对抗网络算法 [J]. *中国图象图形学报*, 2021, 26(3): 594–604.
- SHI Caijuan, TU Dongjing, LIU Jingyi. Re-GAN: residual generative adversarial network algorithm[J]. *Journal of image and graphics*, 2021, 26(3): 594–604.
- [18] RADFORD A, METZ L, CHINTALA S. Unsupervised representation learning with deep convolutional generative adversarial networks[EB/OL]. (2016–01–07)[2022–07–06]. <https://arxiv.org/abs/1511.06434>.
- [19] KARRAS T, AILA, LAINE S, et al. Progressive growing of GANs for improved quality, stability, and variation [EB/OL]. (2018–02–26)[2022–07–06]. <https://arxiv.org/abs/1710.10196>.
- [20] MIRZA M, OSINDERO S. Conditional generative adversarial nets[EB/OL]. (2014–11–06)[2022–07–06]. <https://arxiv.org/abs/1411.1784>.
- [21] ODENA A, OLAH C, SHLENS J. Conditional image synthesis with auxiliary classifier GANs[C]// 34th International Conference on Machine Learning. San Diego: JMLR, 2017.

- [22] ZHOU Peng, XIE Lingxi, NI Bingbing, et al. Omni-GAN: on the secrets of cGANs and beyond[C]//2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2022: 14041–14051.
- [23] NGUYEN T D, LE Trung, VU Hung, et al. Dual discriminator generative adversarial nets[J]. Advances in neural information processing systems, 2017, 30(12): 2671–2681.
- [24] ARJOVSKY M, CHINTALA S, BOTTOU L. Wasserstein Gan[EB/OL]. (2017–12–06)[2022–07–06]. <https://arxiv.org/abs/1701.07875>.
- [25] Gulrajani I, Ahmed F, Arjovsky M, et al. Improved training of wasserstein GANs[J]. Advances in neural information processing systems, 2017, 30(12): 5768–5778.
- [26] BHASKARA V S, AUMENTADO-ARMSTRONG T, JEPSON A, et al. GraN-GAN: piecewise gradient normalization for generative adversarial networks[C]//2022 IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa: IEEE, 2022: 2432–2441.
- [27] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 770–778.
- [28] HUANG Gao, LIU Zhuang, VAN DER MAATEN L, et al.

Densely connected convolutional networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 2261–2269.

作者简介:



管凤旭, 副教授, 博士, 主要研究方向为无人系统自主控制、机器视觉目标检测与跟踪、计算机控制及应用。获授权发明专利近 20 项, 出版教材 5 部, 发表学术论文 40 余篇。



路斯棋, 硕士研究生, 主要研究方向为水下图像处理、计算机视觉。



郑岩, 讲师, 博士, 主要研究方向为计算机视觉和机器学习。