基于样本特征解码约束的 GANs

陈泓佑 [1] 陈 帆 [1] 和红杰 [1] 朱翌明 [1]

摘 要 生成式对抗网络 (Generative adversarial networks, GANs) 是一种有效模拟训练数据分布的生成方法, 其训练的常见问题之一是优化 Jensen-Shannon (JS) 散度时可能产生梯度消失问题. 针对该问题, 提出了一种解码约束条件下的 GANs, 以尽量避免 JS 散度近似为常数而引发梯度消失现象, 从而提高生成图像的质量. 首先利用 U-Net 结构的自动编码机 (Auto-encoder, AE) 学习出与用于激发生成器的随机噪声同维度的训练样本网络中间层特征. 然后在每次对抗训练前使用设计的解码约束条件训练解码器. 其中, 解码器与生成器结构相同, 权重共享. 为证明模型的可行性, 推导给出了引入解码约束条件有利于 JS 散度不为常数的结论以及解码损失函数的类型选择依据. 为验证模型的性能, 利用 Celeba 和 Cifar10 数据集, 对比分析了其他 6 种模型的生成效果. 通过实验对比 Inception score (IS)、弗雷歇距离和清晰度等指标发现, 基于样本特征解码约束的 GANs 能有效提高图像生成质量, 综合性能接近自注意力生成式对抗网络.

关键词 生成式对抗网络, 梯度消失, 特征学习, 自动编码机, 深度学习

引用格式 陈泓佑, 陈帆, 和红杰, 朱翌明. 基于样本特征解码约束的 GANs. 自动化学报, 2022, **48**(9): 2288−2300

DOI 10.16383/j.aas.c190496

开放科学(资源服务)标识码(OSID):

A GANs Model Based on Sample Feature Decoding Constraint

CHEN Hong-You[1] CHEN Fan[1] HE Hong-Jie[1] ZHU Yi-Ming[1]

Abstract Generative adversarial networks (GANs) model is a generative approach for effectively simulating the distribution of training data. One of the common problems in training GANs is the possible vanishing gradient problem while optimizing Jensen-Shannon (JS) divergence. Aiming at the problem, a GANs model under decoding constraint is proposed to avoid JS divergence approximating a constant, thus improving the quality of generated images. Firstly, an auto-encoder (AE) structured under U-Net is utilized to learn the training sample network middle layer feature. It has the same dimension as the random noise used for triggering generative network. Then, the decoding constraint is designed, which shares the same structure and weights as that of the generative network, is used to train decoder before each adversarial training. To prove the feasibility of model, the conclusion is deduced that introducing decoding constraint is beneficial to avoiding JS divergence approximating a constant and the type selection basis of decoding loss function is given. To verify the performance of the model, Celeba and Cifar10 datasets are used to compare and analyze the generated results of other 6 models. By comparing Inception score, Frechet inception distance, clarity and other index via experiment, it is discovered that the novel GANs can improve the quality of generated images, comprehensive performance close to self-attention generation adversarial networks.

Key words Generative adversarial networks, vanishing gradient, feature learning, auto-encoder, deep learning

Citation Chen Hong-You, Chen Fan, He Hong-Jie, Zhu Yi-Ming. A GANs model based on sample feature decoding constraint. *Acta Automatica Sinica*, 2022, **48**(9): 2288−2300

收稿日期 2019-06-29 录用日期 2019-12-02
Manuscript received June 29, 2019; accepted December 2, 2019
国家自然科学基金 (61872303, U1936113) 和四川省科技厅科技创新人才计划 (2018RZ0143) 资助
Supported by National Natural Science Foundation of China (61872303, U1936113) and Technology Innovation Talent Program of Science & Technology Department of Sichuan Province (2018RZ0143)
本文责任编委 王立威
Recommended by Associate Editor WANG Li-Wei
1. 西南交通大学信号与信息处理四川省高校重点实验室 成都 611756
1. Key Laboratory of Signal & Information Processing of Sichuan Province, Southwest Jiaotong University, Chengdu 611756

生成式对抗网络 (Generative adversarial networks, GANs) 是 2014 年 Goodfellow 等[1] 依据零和博弈思想和纳什均衡原理提出的一种数据生成模型, 被广泛应用于图像生成领域. GANs 在网络结构上主要由生成器 G 网络和判别器 D 网络组成[1−3]. G 网络的目的是将随机噪声映射到训练集分布中, 对随机噪声和训练数据的联合概率密度进行建模, 关注于数据生成过程. D 网络的目的是区分出馈入样本的类别问题, 关注于生成数据和训练数据的最优分界面. GANs 的最大特点是对抗学习方式, 训

练过程中 G 网络和 D 网络交替对抗训练, 两者的能力同步提升.

由于 GANs 在图像数据生成上的出色表现, 此后为提高 GANs 生成图像的多样性 (模式坍塌问题) 和质量等, 研究者提出了许多 GANs 衍生模型.

从加入条件变量和图像隐码控制方面进行改进. Mirza 等[4] 提出的条件生成式对抗网络尝试利用训练集样本的某些信息 (如图像类别标签) 来提高随机噪声 z 的可解释性, 使得生成图像质量有所提高. Odena[5] 提出的半监督学习生成式对抗网络将 GANs 进行拓展, 利用半监督学习使得 D 网络分类能力提高, 能够有效提高生成图像质量及收敛速度. Odena 等[6] 提出的辅助分类器生成式对抗网络可实现多分类问题, 输出的则是对应标签的概率值, 有效提高了 GANs 模型模拟多类别、高分辨率数据集的效果. Chen 等[7] 提出的信息极大化生成式对抗网络在 GANs 对抗学习的基础上, 通过引入一个训练集样本对应的隐含信息 (如类别标签, 倾斜度), 使得隐含信息与生成样本具有较高的互信息, 有效提高图像生成质量. Donahue 等[8] 提出双向生成式对抗网络 (Bidirectional generative adversarial networks, BiGANs) 是一种双向结构的对抗模型, 增加了一个训练好的编码器 E 网络用于提取训练样本隐码 c, 在 D 网络的馈入信息是随机噪声 z 与对应生成样本配对或样本隐码 c 与对应的训练样本配对, 在生成实际场景图像上能取得很好的效果. 以上 GANs 中对于需要标签信息的 GANs 模型限制了其在无监督对抗学习中的应用. 双向生成式对抗网络中隐码的引入使得训练样本反复被编码器编码, 而且馈入到 D 网络的数据不仅仅是图像样本, 还有隐码 c, 整个 GANs 网络框架变得更复杂, 增加训练代价.

从 GANs 网络结构或框架设计方面进行改进. Radford 等[9] 提出的深度卷积生成式对抗网络 (Deep convolutional generative adversarial networks, DCGANs) 使用重新设计的卷积神经网络作为 G 和 D 网络, 能够有效提高图像生成质量, 并且成为 GANs 网络结构设计上的标准模型之一. Denton 等[10] 提出的一种拉普拉斯金字塔生成式对抗网络模型, 结合 GANs 和条件 GANs 的一些优点, 使用多个 GANs 逐层地生成高质量自然图像. Brock 等[11] 基于残差网络设计的大型生成式对抗网络能有效生成大尺寸, 高质量的自然图像, 但参数量明显大于一般 GANs 模型, 需要更多的硬件资源和时间成本. Nguyen 等[12] 提出的双判别器生成式对抗网络使用两个 D 网络更细化 GANs 中 D 网络的分类任务, 能使得训练收敛速度变快及提高生成图像的多样性. 张龙等[13] 提出一种协作式结构的 GANs 模型提高生成图像质量, 一定程度避免了模式坍塌现象的发生. GANs 网络结构的设计通常难度较大, 这也是到目前为止, 通过结构设计提升 GANs 能力的经典 GANs 模型很少的主要原因.

从优化目标函数梯度消失方面进行改进. GANs 优化 Jensen-Shannon (JS) 散度时可能导致梯度消失, 使得训练效果相对较差, 多样性不足[14]. 研究者主要是使用其他散度代替 JS 散度. Arjovsky 等[14] 提出沃瑟斯坦距离生成式对抗网络, 利用沃瑟斯坦距离来描述作为两个分布的相似度; 这有效避免了优化 JS 散度容易出现的梯度消失现象, 但对 D 网络权重剪枝比较粗暴. Mao 等[15] 提出的最小二乘生成式对抗网络 (Least squares generative adversarial networks, Least squares GANs 或 LSGANs) 是利用最小二乘原理, 将 G 和 D 网络的损失函数设计成最小二乘形式, 使得 GANs 优化生成数据分布和训练数据分布的 Pearson 散度, 避免梯度消失, 并且损失函数收敛过程更平稳. Berthelot 等[16] 提出的边界平衡生成式对抗网络 (Boundary equilibrium generative adversarial networks, BEGANs) 将一个自编码器作为 D 网络, 设计了 G 和 D 网络的平衡度量方法来优化沃瑟斯坦距离, 进而引入新的超参数来平衡两个网络训练, 以期得到更好的生成图像. Gulrajani 等[17] 提出的梯度惩罚沃瑟斯坦距离生成式对抗网络 (WGANs with gradient penalty, WGANsGP), Wu 等[18] 提出的沃瑟斯坦散度生成式对抗网络均是 WGANs的改进模型, 其中 WGANsGP 通过梯度惩罚的方式替换掉权重剪切, 从而避免因权重剪切导致的权重集中化和调参上的梯度消失问题. 沃瑟斯坦散度生成式对抗网络通过引入沃瑟斯坦散度, 从而去除 WGANs中 D 网络的 Lipschitz 条件, 又能保留沃瑟斯坦距离度量两个分布的良好性质 (如 JS 散度的梯度消失问题). Su[19] 提出的对偶 GANs 模型, 通过引入合理的概率散度并找出它的对偶表达, 再将其转化成极小−极大博弈形式, 从而避免了类似于 WGANs 需要的 Lipschitz 条件和多数 GANs 容易发生梯度消失问题. Zhao 等[20] 提出基于能量的生成式对抗网络是将 D 网络看成能量函数, 提供了一种基于能量解释的 GANs, 并且通过 pull-away term 策略来防止梯度消失问题导致的模式坍塌. 王功明等[21] 等提出一种基于重构误差能量函数的 GANs 模型, 利用深度置信网络作为 G 网络, 能预防网络梯度消失, 在生成效果和网络学习效率上有所提升. 这些方法虽然能有效解决梯度消失问题, 但普遍需要比较多的迭代次数, 特别是优化沃瑟斯坦距离的 GANs, 通常为使得 D 网络满足 1-Lipschitz 条件, 每个批次

的训练中很可能需要对其进行多次训练.

除此之外还有其他的改进途径. Qi[22] 提出的损失敏感型生成式对抗网络主要为了限制 GANs 试图模拟任意训练集分布的能力, 让生成模型能够更偏向于改进真实度不高的样本从而提高图像生成效果. Zhang 等[23] 提出的自注意力生成式对抗网络 (Self-attention generation adversarial networks, SAGANs), 利用注意力机制嵌入 G 网络和 D 网络中, 使得两个网络能更好地学习网络自发关注的训练图像特征提高了生成图像质量和多样性, 但其网络规模和训练迭代次数有所增加.

考虑到优化 JS 散度容易带来的梯度消失问题, 无监督 GANs 模型在训练上更便利的优点. 本文依然将 JS 散度作为主优化目标的前提下, 提出了一种基于训练集样本特征解码损失约束的无监督 GANs 模型. 所设计的模型不仅尽量避免优化 JS散度可能带来的梯度消失问题, 同时也通过改进 GANs 网络拓扑结构, 融入样本本身的特征信息进行训练以提高 GANs 图像生成能力. 首先利用无监督特征学习模型预训练出训练集样本的中间层特征; 然后构建一个与 G 网络结构一致和权重共享的解码器 Dec, 在每次对抗训练前使用本文设计的约束条件进行图像特征解码; 最后再进行优化 JS 散度的 GANs 对抗学习. 为验证所设计的 GANs 性能, 利用 Celeba 和 Cifar10 数据集, 对比分析了几种典型 GANs 模型的生成效果. 实验结果表明, 本文方法能有效提高生成图像的多样性和质量的同时, 还能减少训练所需的 epoch 数.

1 对抗原理

GANs 的典型结构由一个生成器 G 和判别器 D 组成. G 网络的任务是模拟训练集 X 进行数据生成, D 网络的任务是分辨出馈入的样本属于 X 或者 $G(Z)$.

G 网络的每个输入量为一个随机噪声 z, $z \in Z$ 且 $Z \sim F_Z(z)$, 随机噪声 z 的分布函数 $F_Z(z)$ 通常为正态分布或均匀分布. 记训练样本 x, $x \in X$ 且 $X \sim F_X(x)$, 其中 $F_X(x)$ 为训练样本集 X 的分布函数. 那么 D 和 G 网络的损失函数分别为:

$$loss_D = \frac{1}{m}\sum_{i=1}^{m}[\ln D(x_i) + \ln(1 - D(G(z_i)))] \quad (1)$$

$$loss_G = \frac{1}{m}\sum_{i=1}^{m}\ln(1 - D(G(z_i))) \quad (2)$$

式中, m 是每次馈入神经网络样本的个数. 从而整个网络的博弈损失函数为:

$$\min_G \max_D V(G, D) = \mathrm{E}_{X\sim F_X(x)}[\ln D(x)] + \mathrm{E}_{Z\sim F_Z(z)}[\ln(1 - D(G(z)))] \quad (3)$$

式中, $V(G, D)$ 是一个二元极小极大零和博弈函数, $\mathrm{E}(\cdot)$ 为期望函数. 优化损失函数最终目的为使得 $G(Z)$ 的统计分布 $F_G(x)$ 趋近于训练样本集 X 的分布 $F_X(x)$. 为便于以下讨论, 使用概率密度函数代替分布函数来描述分布.

2 解码约束的 GANs

本节先分析优化 JS 散度可能带来的梯度消失问题; 然后提出了本文解决方法, 同时给出了理论推导, 为本文的解决方法提供依据; 最后给出本文方法的训练步骤.

2.1 问题分析

为便于分析和讨论, 先引入 Kullback-Leibler (KL) 散度和 JS 散度的定义.

定义 1[24]**.** 设两个具有相同样本空间 Ω 的随机变量 X 和 G 的概率密度函数分别为 $f_X(x)$ 和 $f_G(x)$. KL 散度定义为:

$$KL(f_X(x)||f_G(x)) = \int f_X(x)\ln\frac{f_X(x)}{f_G(x)}\mathrm{d}x \quad (4)$$

上式定量了 $f_G(x)$ 和 $f_X(x)$ 之间的相似程度, 如果 $f_G(x)$ 与 $f_X(x)$ 越相似, 那么 $KL(f_X(x)||f_G(x))$ 值就越小. $KL(f_X(x)||f_G(x))$ 是非负函数, 当且仅当 $f_G(x) = f_X(x)$ 时取得最小值 0. 它不具有通常距离函数中的对称和三角不等性质. 在信息论中 KL 散度表示的是用 $f_G(x)$ 拟合已知的 $f_X(x)$ 时产生的信息损耗.

定义 2[25]**.** 设两个具有相同样本空间 Ω 的随机变量 X 和 G 的概率密度函数分别为 $f_X(x)$ 和 $f_G(x)$. 它们的 JS 散度定义为:

$$JS(f_X(x)||f_G(x)) = \frac{1}{2}KL\left(f_X(x)||\frac{f_X(x)+f_G(x)}{2}\right) + \frac{1}{2}KL\left(f_G(x)||\frac{f_X(x)+f_G(x)}{2}\right) \quad (5)$$

JS 散度为非负函数, $f_G(x)$ 与 $f_X(x)$ 越相似时 $JS(f_X(x)||f_G(x))$ 越小, 当且仅当 $f_G(x) = f_X(x)$ 时取得最小值 0. $f_G(x)$ 与 $f_X(x)$ 越不相似时 $JS(f_X(x)||f_G(x))$ 越接近常数 1. 它具有距离函数中的对称和三角不等的性质.

式 (3) 给出了 GANs 对抗表达形式, Goodfellow 等[1] 指出 GANs 虚拟训练准则 $C(G)$ 当且仅当 $f_G(x) = f_X(x)$ 时取得全局最小值. 在最小点时, $C(G)$ 的极小值为 $-\ln 4$. $C(G)$ 如下所示:

$$C(G) = -\ln 4 + 2 \cdot JS(f_X(x)||f_G(x)) \quad (6)$$

式 (6) 表明, 式 (3) 的优化目标其实是最小化训练集 X 的概率密度函数 $f_X(x)$ 和生成集 $G(Z)$ 的概率密度函数 $f_G(x)$ 的 JS 散度.

Arjovsky 等[14] 在 WGANs 的分析过程中指出当生成样本集分布 $f_G(x)$ 与训练样本集分布 $f_X(x)$ 的相似度越低, 即当两个分布的交叉区域越小, $JS(f_X(x)||f_G(x))$ 越接近于常数 1. 这可能引发损失函数梯度消失的现象. 在 GANs 训练过程中, $f_G(x)$ 是逐渐拟合 $f_X(x)$ 的过程, JS 散度的固有性质可知, 在 GANs 训练的起步阶段梯度消失现象更明显. 即使 GANs 能够继续通过优化方法进行参数更新, 为使得 $f_G(x)$ 与 $f_X(x)$ 有足够的相交区域, 也需要更多 epoch 数进行训练. 解决这个问题的一般方法是使用 Pearson 散度或沃瑟斯坦距离代替 JS 散度重新设计损失函数.

2.2　特征解码约束的 GANs

由第 2.1 节分析可知, JS 散度为常数而导致梯度消失的一个重要前提是 $f_G(x)$ 与 $f_X(x)$ 的相似度足够低. 那么通过添加约束条件利于 $f_G(x)$相似于 $f_X(x)$ 可以达到尽量避免 JS 散度为常数的目的, 为此本文设计了一种 JS $+\lambda\cdot$KL 混合散度的约束方法. 约束条件 $KL(f_X(x)||f_G(x))$ 的目的是为使得 $f_G(x)$ 与 $f_X(x)$ 的相交区域变大.

如图 1 所示, 本文设计的 GANs 分为 3 个部分: 1) 特征学习部分: 目的是预训练出训练集 X 的特征集 C. 2) 解码学习部分: 目的是先通过本文设计的解码约束条件对特征集 C 进行解码, 完成 $KL(f_X(x)||f_{Dec}(x))$ 约束. 又通过解码器 Dec 与 G 网络结构一致, 参数共享, 以近似达到 $KL(f_X(x)||f_G(x))$ 约束. 最终使得在优化 JS 散度前 $f_G(x)$ 与 $f_X(x)$ 相交区域变大, JS 散度不易为常数, 从而尽量避免出现梯度消失现象. 3) 对抗学习部分: 通过优化 JS 散度使得 $f_G(x)$ 模拟 $f_X(x)$. 其中特征学习部分是预训练, 解码学习和对抗学习部分需要一起动态学习. 与一般含自动编码机 GANs 不同的是, 本文自动编码机主要目的是预训练出可用的隐含特征. 例如, 与双向生成式对抗网络相比, 隐含特征 c 不会馈入 D 网络对其参数更新及直接参与对抗训练, 仅用于解码学习; 与 BEGANs 相比, D 网络的任务仍然是二分类, 无编码功能.

2.2.1　特征学习

在图像特征学习中, 需要提取出图像的隐含信息, 用此表征原始图像. 自编码特征学习是一种有效的图像特征学习方法[26]. 常用的自动编码机较多, 除噪自动编码机[26−28] 经过对训练样本加入噪声并进行降噪的训练过程, 能够强迫网络学习到更加鲁棒的不变性特征, 获得馈入图像的更有效和更鲁棒的表达. 收缩自动编码机[26, 29] 能够较好地重构训练样本, 并且对训练样本一定程度的扰动具有不变性. 稀疏自动编码机[26, 30] 将稀疏编码和自编码机结合, 可以提取馈入样本的稀疏显著性特征. 对于一般任务, 最常用的依然是经典自动编码机模型[26].

由于随机噪声 z 维度相对较低 (如 64 或 100 维), 特征提取任务相对简单, 且为获取更好的重构图像效果. 本文将经典自动编码机结合 U-Net 网络模型[31], 建立了 5 层的全连接类似 U-Net 的自动编码机用于 C 的获取, 并且使得特征 c 的维度与随机噪声 z 的维度是相同的. 图 2 给出了 U-Net 型自动

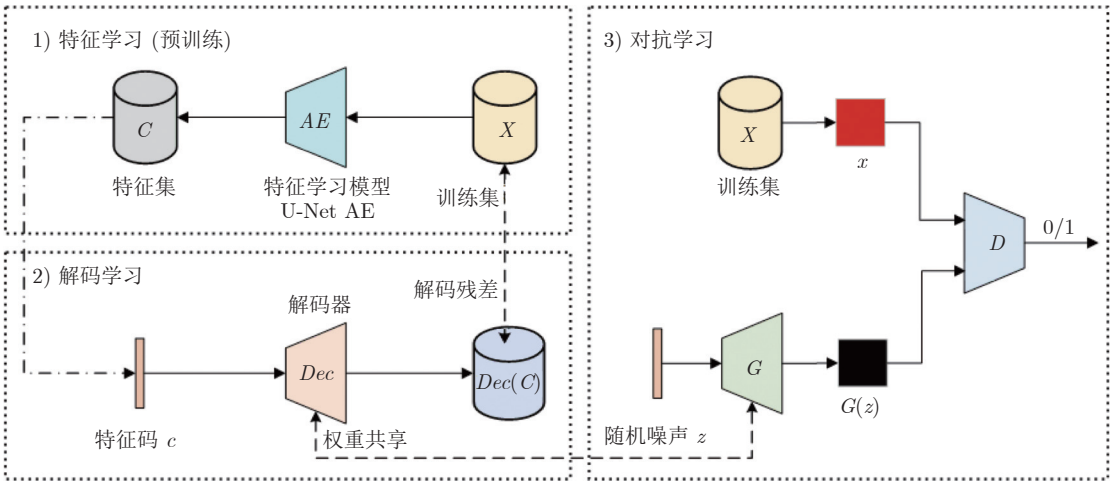


图 1　总体结构示意图

Fig. 1　Overall structure sketch

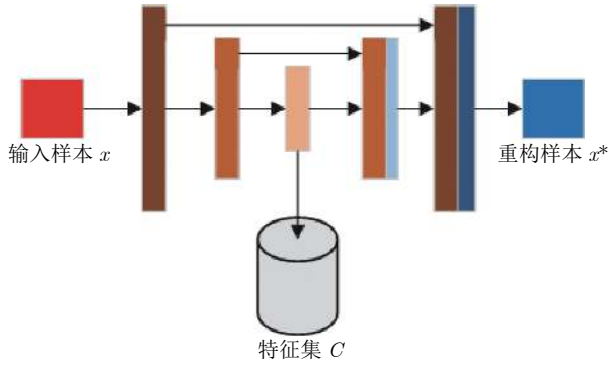


图 2 特征学习网络结构图

Fig. 2 Structure diagram of feature learning network

编码机用于获取 X 的特征集C 的示意图. 该网络由 5 层神经元组成, 第 3 层用于特征提取, 特征图像像素个数与随机噪声 z 维度相一致.

训练过程中, 损失函数选用均方差损失函数:

$$loss_{AE} = \frac{1}{m}\sum_{i=1}^{m}(x_i - x_i^*)^2 \tag{7}$$

式中, x_i^* 是 x_i 对应的重构图像.

2.2.2 解码及对抗学习

设训练样本集 X 对应的特征集为C, 解码器为 Dec, 它与 G 网络共享权重, 网络结构一致. 记 X 的概率密度函数为 $f_X(x)$. 解码集 $Dec(C)$ 的概率密度函数为 $f_{Dec}(x)$. 解码损失函数为:

$$loss_{Dec} = \frac{1}{m}\sum_{i=1}^{m}||x_i - Dec(c_i))|| \tag{8}$$

式中, x_i 为X 中的样本, c_i 为x_i 对应于C 中的样本, m 为样本个数. $||\cdot||$ 为度量两个样本的距离函数, 常用的函数类型有 L1 和 L2 型函数.

在原有的 JS 散度对抗损失函数中引入解码损失函数进行约束, 需要控制解码约束条件对 Dec 网络梯度下降的贡献. 主要原因有以下 3 点: 1) G 网络模拟的是训练集 X 的主要特征, 不需要按像素严格一致. 解码损失函数是按像素严格一致进行图像重构, 因此后者约束更强势. 2) 对抗损失函数是优化 $JS(f_X(x)||f_G(x))$, 解码损失函数是优化 $KL(f_X(x)||f_{Dec}(x))$, 优化后者虽然对避免 $JS(f_X(x)||f_G(x))$ 为常数有益, 但各自的梯度下降方向并不完全一致, 应保证 $JS(f_X(x)||f_G(x))$ 是主优化方向. 3) 优化二元组 $(f_X(x), f_G(x))$相对于优化三元组 $(f_X(x), f_G(x), f_{Dec}(x))$ 难度更低. 当 $f_{Dec}(x) \approx f_G(x)$ 时, 相当于近似优化前者.

为达到以上目的, 可以通过对解码损失函数权重系数, 训练频次及学习率加以控制. 当解码损失函数式 (8) 选用 L2 型函数时, 本文设计的解码损失函数如下:

$$loss_{Dec} = \delta \cdot \lambda \cdot \frac{1}{m}\sum_{i=1}^{m}(x_i - Dec(c_i))^2 \tag{9}$$

式中, δ 是判别函数, 1 表示进行解码训练, 0 表示屏蔽解码训练; λ 是解码损失函数权重系数.

$$\delta = \begin{cases} 1, & (t \bmod r) = 0 \wedge t < l \\ 0, & \text{否则} \end{cases} \tag{10}$$

式中, t 是当前的迭代 epoch 数, r 是控制调用解码约束的频次, l 是控制最后一次解码的控制变量. 每次对抗学习前, 依据条件判别式 (10) 以此来控制解码约束条件的使用总次数和频率.

由此, 最终的对抗网络损失函数为:

$$\min_{G,\,Dec}\max_{D} V(D, G, Dec)|_{f_{Dec}(x)\approx f_G(x)} = V(D, G) + loss_{Dec} \tag{11}$$

由于 D 网络是一个二分类网络, 利用单向标签平滑[32]处理能对分类性能有一定提高, 这有益于降低分类网络的训练难度. 在实际训练操作中可以使用这种方式对式 (1) 进行标签平滑处理.

为使得上面所提供的解决方法有所依据. 分析了以下 3 点: 1) 优化 $\mathrm{JS} + \lambda \cdot \mathrm{KL}$ 混合散度对 JS 散度不为常数的影响. 2) 优化 $\mathrm{JS} + \lambda \cdot \mathrm{KL}$ 混合散度对优化原有 JS 散度相对于分布对 $(f_X(x), f_G(x))$ 的极小值点及单调性的影响. 3) 优化 KL 散度时解码损失函数类型选择的依据. 为此下面 3 个命题进行了讨论分析.

命题 1. 限制解码器 Dec 解码约束条件对 Dec 网络参数更新的梯度贡献, 且使得 $f_{Dec}(x) \approx f_G(x)$. 那么训练过程中引入解码约束条件有利于避免 $JS(f_X(x)||f_G(x)$ 为常数.

证明. 要证明命题结论, 只需要证明引入约束条件后有利于 $f_G(x)$ 相似于 $f_X(x)$ 即可.

记第 t 次解码训练后解码集 $Dec(C, t)$ 对应的概率密度函数为 $f_{Dec}(x, t)$, 第 t 次对抗训练后生成数据集 $G(Z, t)$ 对应的概率密度函数为 $f_G(x, t)$.

由式 (6) 的 $C(G)$ 条件知, G 网络仅仅是使得 $f_G(x)$ 模拟 $f_X(x)$, 并不要求 $G(Z) = X$. 所以优化过程是一个依分布收敛的过程, 即:

$$\lim_{t\to\infty} f_G(x, t) = f_X(x) \tag{12}$$

由式 (8) 可知, 对于解码器 Dec 的理想目标是求解 $C \to X$ 的映射, 使得 $Dec(C) = X$, 即:

$$||x_i - Dec(c_i)|| = 0 \tag{13}$$

式中, x_i 和 $Dec(c_i)$ 分别是各自样本空间中的任意

样本, 且 c_i 是 x_i 的特征码. 故而解码器 Dec 的理想目标是使得 $Dec(\,C,t\,)$ 几乎处处收敛于 X. 但由于训练中, 通常只能达到如下情况:

$$||x_i - Dec(c_i)|| < \varepsilon_1 \tag{14}$$

故而 $Dec(\,C,t\,)$ 是依概率收敛于 X, 即:

$$\mathrm{P}\left\{\lim_{t\to\infty}|Dec(C,t)-X|<\varepsilon_2\right\}=1 \tag{15}$$

其蕴含于

$$\mathrm{P}\left\{\lim_{t\to\infty}||f_{Dec}(x,t)-f_X(x)||<\varepsilon_3\right\}=1 \tag{16}$$

式中, ε 为任意小的正实数.

因为依概率收敛强于依分布收敛 (前者是后者的充分非必要条件). 并且存在条件 $f_{Dec}(x)\approx f_G(x)$. 所以引入约束条件后, 能够使得 $f_G(x)$ 相似于 $f_X(x)$ 的概率变大. □

命题 2. 限制解码约束条件对 Dec 网络参数更新的梯度贡献, 且使得 $f_{Dec}(x)\approx f_G(x)$. 相对于分布对 $(f_X(x),\,f_G(x))$ 引入解码约束条件后可基本不影响新构建的损失函数的单调性及极小值点.

证明. 优化式 (8), 由距离函数的单调性和非负性可知, 当且仅当 $Dec(C)=X$ 时取得极小值 0, 此时 $f_{Dec}(x)=f_X(x)$.

因为解码过程是使解码概率密度函数 $f_{Dec}(x)$ 模拟逼近已知的 $f_X(x)$, 式 (8) 的残差项是解码后信息 $Dec(C)$ 相对原信息 X 的信息损耗. 故而解码的目的为:

$$\min_{Dec} KL(f_X(x)||f_{Dec}(x)) \tag{17}$$

先证明式 (6) 引入 $KL(f_X(x)||f_G(x))$ 条件并不影响新损失函数的单调性和极小值点. 记新的损失函数表达式为:

$$C_1(G)=C(G)+KL(f_X(x)||f_G(x)) \tag{18}$$

因为 JS 散度和 KL 散度对于任意的分布对 $(f_X(x),\,f_G(x))$为非负单调递增函数.

又因为 JS 散度和 KL 散度均为当且仅当 $f_G(x)=f_X(x)$ 时取得极小值 0. 故而式 (18) 当且仅当 $f_G(x)=f_X(x)$ 时取得极小值点 $-\ln 4$.

又因为 $f_{Dec}(x)\approx f_G(x)$. 所以下式的单调性和极值点相对于分布对 $(f_X(x),\,f_G(x))$ 基本不变:

$$C_2(G)=C(G)+KL(f_X(x)||f_{Dec}(x)) \tag{19}$$

故而基本不影响新构建的损失函数的单调性及极小值点. □

命题 3. 当训练集 X 符合正态分布时, 解码器 Dec 应选用 L2 型函数.

证明. 记 X 对应的训练集为 C, 解码集为 $Dec(C)$. $f_X(x|c)$ 为 C 给定时, X 的条件概率密度函数. $f_{Dec}(x|c)$ 为 C 给定时, 解码集 $Dec(C)$ 等于训练集 X 的条件概率密度函数, 那么解码器 Dec 解码的目的可表达为使得 $f_{Dec}(x|c)\approx f_X(x|c)$, 即:

$$||f_{Dec}(x|c)-f_X(x|c)||<\varepsilon \tag{20}$$

式中, ε 是任意小的正实数.

其蕴含于 (由 KL 散度的信息论含义可得):

$$\min_{Dec} KL(f_X(x|c)||f_{Dec}(x|c))= \min_{Dec}\mathrm{E}_{c\in C}\left[\ln f_X(x|c)-\ln f_{Dec}(x|c)\right] \tag{21}$$

因为 $\ln f_X(x|c)$ 为已知的训练集 X 及其对应的特征集 C 表达的信息. 所以其为常数, 在梯度下降优化时不对梯度做贡献. 由此式 (21) 等价于优化下式:

$$\min_{Dec} -\mathrm{E}_{c\in C}\ln f_{Dec}(x|c) \tag{22}$$

又因为

$$-\mathrm{E}_{c\in C}\ln f_{Dec}(x|c)=-\sum_{i=1}^{m} f(c_i)\cdot\ln f_{Dec}(x_i|c_i) \tag{23}$$

式中, m 是馈入神经网络样本的数量.

由于 c_i 在 C 中, 训练过程中 c_i 必然出现. 所以 $f(c_i)=1$.

又因为, X 符合正态分布, $X\sim \mathrm{N}(x;x^*,\sigma^2)$. 其中 x^* 是 x 的估计 (Dec 解码 x 特征 c 的结果, 即 $x^*=Dec(c)$. 从而式 (23) 等于:

$$\begin{aligned}-\sum_{i=1}^{m}\ln f_{Dec}(x_i|c_i)=&\\ -\sum_{i=1}^{m}\ln\left(\frac{1}{\sqrt{2\pi}\sigma}\cdot\exp\left(-\frac{(x_i-x_i^*)^2}{2\sigma^2}\right)\right)=&\\ \frac{m}{2}\cdot\ln(2\pi)+m\cdot\ln\sigma+\sum_{i=1}^{m}\frac{(x_i-x_i^*)^2}{2\sigma^2}&\end{aligned} \tag{24}$$

式中, $x_i^*=Dec(\,c_i\,)$, c_i 为 x_i 的特征. 前 2 项是常数项, 梯度下降过程中对梯度不做贡献, 仅最后一项对梯度下降做出贡献. 均方差损失函数为:

$$loss_{MSE}=\tfrac{1}{m}\sum_{i=1}^{m}(x_i-x_i^*)^2 \tag{25}$$

对比式 (24) 的最后 1 项和式 (25) 可知. 优化式 (24) 等价于优化式 (25). □

由命题 1 可知, 引入解码约束条件当 $f_{Dec}(x)\approx f_G(x)$ 时将有利于 $f_G(x)$ 与 $f_X(x)$ 更相似. 从而达到尽量避免 $JS(f_X(x)||f_G(x))$ 为常数和近似常数的目的, 有利于避免损失函数梯度消失的问题.

由命题 2 可知, 引入解码约束条件基本不影响函数的单调性和极小值点, 表明它们相对于分布对 $(f_X(x),\ f_G(x))$ 的最优解一致, 优化任务的总体目的相近.

由命题 3 可知, 若训练样本 X 符合正态分布, 应选用均方差损失函数. 由于训练集 X 中样本的结构信息 (几何结构量)、颜色信息和清晰度 (与图像纹理相关) 等关键特征信息, 依据三大中心极限定理可知是满足正态分布假设或近似正态分布假设.

2.3 网络训练

通过以上描述, 可以得到整个网络的训练方法, 如下所示:

步骤 1. 依据式 (7) 充分训练 U-Net 型自动编码机, 获取训练集 X 的特征集 C.

步骤 2. 依据式 (10) 计算出判别值 δ, 如果 $\delta = 1$ 则对解码器 Dec (解码器与生成器 G 权重共享, 网络结构一致) 使用均方根传播优化方法进行解码训练. 每次馈入批量尺寸个 x 和对应的特征码 c.

步骤 3. 分别馈入批量尺寸个 x 和 $G(z)$ 到判别器 D 网络, 使用均方根传播优化方法对其进行权重更新.

步骤 4. 馈入批量尺寸个 z 到生成器 G 网络, 使用均方根传播优化方法对其进行权重更新. 连续训练 2 次 G.

步骤 5. 重复步骤 2 ~ 4, 直到达到最大 epoch 数为止.

3 实验及分析

本文实验中, 选取的主要软硬件环境为, Tensor-Flow1.12.0 GPU 版本, CUDA 9.0, cuDNN 7.4, 英伟达 GTX1080, GTX1080Ti, RTX2080Ti 显卡. 实验的其他部分如下.

3.1 评价指标及数据集

为定量对比分析多个生成模型的生成图像效果, 选取 Inception score (IS)[33−34]、弗雷歇距离 (Frechet inception distance, FID)[33−34] 和平均清晰度进行评价. IS 是评价生成图像的质量和模式类别多样性的指标 (对多样性描述更准确一些), 指标值越高越好. FID 也是评价生成图像质量和多样性, 越低越好. 计算 IS 指标不需要训练集做对比, 计算 FID 指标需要训练集做对比, FID 越小表明与训练集的图像质量及多样性越接近. 清晰度是图像重要的视觉质量指标, 越高则有更多纹理结构信息. 清晰度方法选取常用的基于能量梯度表达计算公式:

$$S(x)=\sum_{i=0}^{w-2}\sum_{j=0}^{h-2}(|I(i+1,\ j)-I(i,\ j)|^2+ |I(i,\ j+1)-I(i,\ j)|^2) \tag{26}$$

式中, $I(i,\ j)$ 表示在图像样本 x 中坐标 $(i,\ j)$ 处的像素值大小, w 和 h 分别表示样本图像 x 的宽度和高度. 使用 $S(x)$ 除以图像像素个数以获取平均清晰度.

为验证本文 GANs 模型的生成图像的效果, 选取 Celeba 和 Cifar10 数据集进行测试. 数据集详细信息如下所示.

Celeba 数据集共含有 202 599 张彩色人物上半身图像, 每张图像大小为 178×218 像素. 在实验中选取前 50 000 张图像, 裁剪出 64×64 的人脸图像作为训练集. Cifar10 数据集含有 50 000 张训练集彩色图像和 10 000 张测试集彩色图像. 每张图片大小为 32×32, 10 个类别的图像在训练集和测试集中比例相同. 实验选用 Cifar10 的训练集作为 GANs 的训练集. 图 3 ~ 4 展示了训练集的样本图像.

图 3 Celeba 数据集样本
Fig. 3 Samples of Celeba dataset

图 4 Cifar10 数据集样本
Fig. 4 Samples of Cifar10 dataset

3.2 特征学习实验

在图像特征学习中, 使用类似于 U-Net 的 5 层全连接自编码机用于特征学习, 每层神经元数量为: $w\times h$ 、10×10 、10×10 、10×10 和 $w\times h$ (w 和 h 是图像宽度和高度), 激活函数为 softsign, 使用 Adam 方法进行优化, 学习率为 0.001, 动量因子为 0.9. 每批提取 100 个样本图像的中间层特征, 迭代次数为 7 000. 在 GTX1080Ti 显卡条件下, Cifar10 数据集上所耗时间约为 7 小时, Celeba 数据集上所耗时间约为 18 小时.

图 5 展示了部分训练样本重构效果和提取的特征图. 前后 3 行图像各为一个单元, 每个单元中第 1 行是原图, 第 2 行是重构图, 第 3 行是对应的特征图. 在特征图中, 每 1 个格子对应原特征图的 1 个彩色像素.

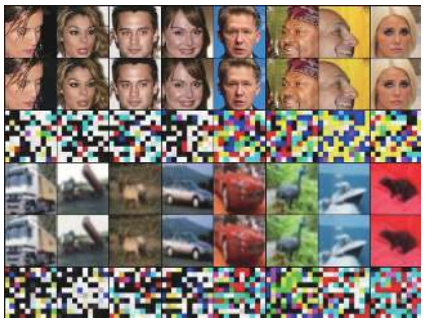


图 5　U-Net 自动编码示例

Fig. 5　Samples of U-Net auto-encoder

由图 5 可以看出, U-Net 结构下的自编码机都有比较好的图像重构视觉效果, 所提取的隐含特征都有比较好的特征表达能力. 从图 5 可以观察出图像颜色和纹理结构越丰富, 隐含特征色彩也越丰富. 反之, 特征的颜色也比较单一. 例如图 5 人脸图像中, 第 1 ~ 4 列头发颜色和背景颜色都偏暗, 面部方向为正面. 第 5 ~ 8 列背景图像, 面部角度及头发颜色都比较鲜明. 与之对应, 它们的特征也有比较明显的颜色区分度, 从而表明所学习到的特征包含了原始图像的一些信息, 如颜色和面部方向. 在 Cifar10 数据集中依然有类似的规律, 能明显看出, 后 4 列图像是颜色鲜明的, 特征也鲜明.

表 1 给出了 Celeba 和 Cifar10 数据集重构样本与训练集的峰值信噪比 (Peek signal to noise ratio, PSNR) 和结构相似度 (Structural similarity, SSIM) 质量评估指标.

表 1　原图像与重构图像的 PSNR 和 SSIM 值统计

Table 1　PSNR & SSIM between original and reconstructed images

数据集	指标	均值	标准差	极小值	极大值
Celeba	PSNR	40.588	5.558	22.990	61.158
	SSIM	0.9984	0.0023	0.9218	1.0000
Cifar10	PSNR	46.219	6.117	28.189	66.779
	SSIM	0.9993	0.0019	0.8180	1.0000

由表 1 可以看出, 在 Celeba 和 Cifar10 数据集上, U-Net 型自动编码机重构的样本在 PSNR 和 SSIM 指标上都有不错的表现. 结合图 5 来看, 其提取的特征具有训练集样本的特征表达能力.

3.3　不同解码实验对比

本节实验为验证样本特征有效性, 限制解码约束条件的必要性及解码函数类型选择的重要性做了如下实验. 1) 正态特征 (每个训练样本对应的特征符合标准正态分布); 2) 均匀特征 (每个训练样本对应的特征符合 [−1, 1] 的均匀分布); 3) L1 解码约束条件. 4) L2 解码约束条件, 但不限制其对损失函数的梯度贡献. 5) 本文方法 (L2 解码约束条件, 限制对损失函数梯度贡献); 除此, 还计算了训练集的指标信息用以对比分析.

所有实验选用均方根传播优化方法, 学习率为 0.0002, 动量因子为 0.9, 批量尺寸为 64, epoch 数为 15. 在第 1 至 3 或 5 组实验中, 式 (9) 选取参数 $\lambda = 1.0 \times 10^{-7}$; 式 (10) 中, $r = 2$, $l = 11$. 在第 4 组实验中, $\lambda = 1.0$, $r = 1$, $l =$ epoch. Celeba 和 Cifar10 实验每组生成 50 000 张图片进行统计分析. 表 2 ~ 3 展示了统计图像数据得到的各项指标结果, 其中上标 * 项是来自不限 L2 约束对损失函数梯度贡献权重实验, 第 5 列是计算与训练集清晰度均值的差距值, 粗体表示最优值.

表 2　Celeba 中不同解码实验结果

Table 2　Results of different decoding experiments in Celeba

对比项	IS ($\sigma \times 0.01$)	FID	清晰度均值	清晰度均值差值
训练集	2.71 ± 2.48	0.00	107.88	0.00
正态特征	1.88 ± 1.25	42.54	121.40	13.52
均匀特征	1.82 ± 1.48	43.04	**123.02**	15.14
L1	1.99 ± 1.53	32.95	120.16	12.28
L2*	1.69 ± 0.97	46.08	96.88	11.00
L2 (本文)	**2.05** ± 1.84	**25.62**	114.95	**7.07**

表 3　Cifar10 中不同解码实验结果

Table 3　Results of different decoding experiments in Cifar10

对比项	IS ($\sigma \times 0.1$)	FID	清晰度均值	清晰度均值差值
训练集	10.70 ± 1.47	0.00	120.56	0.00
正态特征	5.63 ± 0.64	48.21	**139.88**	19.32
均匀特征	5.51 ± 0.79	46.57	137.13	16.57
L1	5.63 ± 0.79	44.53	138.04	17.48
L2*	4.69 ± 0.55	79.10	119.62	**0.94**
L2 (本文)	**5.83** ± 0.70	**42.70**	134.97	14.41

1) 分析对于馈入图像特征 c 的必要性. 对比表 2 ~ 3 中的正态特征, 由均匀特征和 L2 (本文) 表项可知, 本文方法在 IS 和 FID 这两项关键指标上, 均是最优. 特别是在 FID 指标上有显著提升, 表明使用

图像特征 c 进行解码是必要的, 馈入的特征类型是不能随意选取. 对比清晰度, 本文方法的清晰度均值虽不是最大, 但是本文清晰度更接近训练集的清晰度水平, 表明能更合理地模拟训练集高频信息.

2) 分析解码损失函数类型的必要性. 对比表 2 ~ 3 中 L1 和 L2 (本文) 可知, IS 和 FID 指标依然是本文占优. 清晰度均值表项 L1 约束占优表明其生成的图像填充的纹理信息更多, 但本文方法清晰度依然最接近训练集清晰度.

3) 分析限制解码约束条件对梯度贡献的必要性. 对比表 2 ~ 3 中 L2*和 L2 (本文) 可知, L2*的 IS 和 FID 指标明显占劣势, 这表明其多样性和生成图像的指标较差. 对比清晰度指标可以发现不限制 L2 约束条件对梯度的贡献, 会影响生成图像的细节纹理填充. 应注意表 3 中 L2*和训练集表项的清晰度均值相近的原因, 前者是因为纹理细节丢失导致清晰度下降, 后者是因为图像前景或背景本身纹理较少 (如舰船、马匹、汽车、飞机等类别) 导致整体清晰度下降.

通过以上 3 个方面的分析可以发现, 本文方法中使用图像特征进行解码是必要的, 馈入的解码特征类型不具有随意性; 对于解码损失函数使用 L2 效果更优, 具有一定必要性; 限制解码损失函数对梯度的贡献, 使得 $f_{Dec}(x)$ 和 $f_G(x)$ 应近似相等是必要的. 后两点也与模型的理论分析部分一致.

图 6 ~ 11 展示了在 Celeba 和 Cifar10 数据集中, 均匀特征、不限制权重的 L2 约束以及本文方法实验生成样本.

图 6 Celeba 中均匀特征实验样本

Fig. 6 Uniform feature experimental samples in Celeba

图 7 Celeba 中 L2 解码不限制权重实验样本

Fig. 7 L2 decoding with not restrict weight experimental samples in Celeba

图 8 Celeba 中本文方法实验样本

Fig. 8 Experimental samples of our method in Celeba

图 9 Cifar10 中均匀特征实验样本

Fig. 9 Uniform feature experimental samples in Cifar10

图 10 Cifar10 中 L2 解码不限制权重实验样本

Fig. 10 L2 decoding with not restrict weight experimental samples in Cifar10

图 11 Cifar10 中本文方法实验样本

Fig. 11 Experimental samples of our method in Cifar10

由图 3 和图 6 ~ 8 可以看出, 本文方法 (图 8) 生成的图像更细腻, 图像纹理填充主要是填充到头发部分, 视觉效果更好. 而均匀特征生成的图像中 (图 6), 一些纹理信息不仅填充到面部, 而且还填充到背景区域, 这也是表 2 中其清晰度均值偏高的原因. 表明它能够生成更多的纹理细节, 但是填充位置未必合理. 对于 L2* 生成的图像中 (图 7) 能够发现, 生成的样本比较模糊, 纹理信息填充比较差, 影响了视觉效果. 表明限制解码损失函数对梯度下降的贡献是必要的.

由图 9 ~ 11 可以看出, 本文方法 (图 11) 能更明显地生成图像中背景和前景部分. 而均匀特征生成图像 (图 9) 纹理填充得更多. 对于 L2*生成的图像 (图 10) 也能够发现图像相对模糊. 通过以上的数据及生成图像对比分析表明, 在本文方法中, 为生成更好质量的图像, 需要选取合适的解码特征类型, 限制解码约束条件权重以及选取合适的解码函数类型.

3.4　耗时分析

本文 GANs 所使用的 G 和 D 网络内部结构均与 DCGANs 一致, 并且本文将 JS 散度作为主优化目标, 后者将 JS 散度作为优化目标. 为验证模型的所耗时间代价, 在同一台含 GTX1080Ti 显卡的计算机上测试了 DCGANs 和本文 GANs 模型的耗时, 以此对比分析出本文的训练时间代价.

由表 4 可以看出, 在预训练出训练集样本特征前提下, 本文 GANs 总耗时有所下降, 这得益于总的 epoch 数减少. 但单位耗时有所提高, 这源于本文 GANs 在某些 epoch 训练周期内会使用解码约束条件. 由第 3.3 节实验设置可以看出, 解码约束的使用仅在 0 和 0 到 11 之内的偶数训练周期中, 共 6 次. 在特征提取的过程中, 由第 3.2 节可知, 其耗时远大于用于解码和对抗训练耗时. 表明本文 GANs 在特征学习过程中的预训练耗时代价较大. 总耗时的减少为模型的参数调试带来了比较大的便利.

表 4　时间代价测试
Table 4　Test of time cost

数据集	模型	epoch 数	总耗时 (s)	单位耗时 (s)
Celeba	DCGANs[9]	20	3616.03	**180.80**
	本文方法	**15**	**2868.33**	191.22
Cifar10	DCGANs[9]	20	2388.53	**119.43**
	本文方法	**15**	**1859.51**	123.97

3.5　不同 GANs 实验对比

在对抗训练实验中, 本文选取的 G 网络和 D 网络结构与 DCGANs 一致, LSGANs、WGANs 和 WGANsGP 的网络结构处理方法相同. 选取均方根传播优化方法, 学习率和动量因子分别为 0.0002 和 0.9.

BEGANs 和 SAGANs 分别依据文献 [16, 23] 代码单独实验, 关键参数与原文一致, 选用 Adam 优化. 所有实验中批量尺寸为 64. 在 Celeba 和 Cifar10 上每组实验均生成 50 000 张图片进行数据统计, 获得表 5 ~ 6 实验数据. 在表 5 ~ 6 中, SAGANs1 使用 WGANsGP 损失函数 (优化沃瑟斯坦距离), SAGANs2使用 DCGANs 损失函数 (优化 JS 散度); 关于本文所设计 GANs 参数统计, 前半部分是解码及对抗学习模型参数量, 后半部分是 U-Net 自动编码机模型参数量.

对比分析表 5 实验数据可知:

1) 对比前 5 个和本文 GANs 模型. 由 IS 指标可以看出, 本文虽稍好于 LSGANs、WGANs 和 WGANsGP, 但它们之间 IS 指标基本一致; DCGANs 和 BEGANs 较差, 表明两者多样性和质量差于其他方法. 在 FID 指标上, 本文 GANs模型明显优于这 5 个 GANs 模型, 表明本文 GANs模型相对地更能有效模拟训练集分布. 在清晰度指标上, 虽然清晰度均值不是最大, 但是它与训练集之间的清晰度均值差距更小, 表明本文 GANs 对高频细节模拟更合理. 对比模型参数可知, 由于特征学习网络的参数量较少, 所以本文 GANs 模型并没有明显增加参数量. 最后对比 epoch 数可以看出, 本文相对于上述 GANs 模型有明显优势.

2) 对比 SAGANs 和本文 GANs 效果. 从 SAGANs1 和本文 GANs 的实验数据可知, 优化沃瑟斯坦距离的 SAGANs 的综合性能很好, IS 指标与本文相当, FID 指标稍好于本文 GANs; 在清晰度指标

表 5　Celeba 中不同 GANs 对比
Table 5　Comparsion of different GANs in Celeba

GANs 模型	epoch 数	优化项	参数量 ($\times 10^6$)	IS ($\sigma \times 0.01$)	FID	清晰度均值	清晰度均值差值
训练集	—	—	—	2.71 ± 2.48	0.00	107.88	0.00
BEGANs[16]	35	沃瑟斯坦距离	**4.47**	1.74 ± 1.29	46.24	77.58	30.30
DCGANs[9]	20	JS 散度	9.45	1.87 ± 1.58	50.11	**124.82**	16.94
LSGANs[15]	35	Pearson 散度	9.45	2.02 ± 1.63	39.11	122.19	14.31
WGANs[14]	35	沃瑟斯坦距离	9.45	2.03 ± 1.75	40.31	117.15	9.27
WGANsGP[17]	35	沃瑟斯坦距离	9.45	1.98 ± 1.82	37.01	121.16	13.28
SAGANs1[23]	30	沃瑟斯坦距离	10.98	**2.06** ± 1.79	**21.94**	109.94	**2.06**
SAGANs2[23]	30	JS 散度	10.98	1.99 ± 1.79	31.04	99.57	8.31
本文方法	**15**	JS + λ· KL 散度	9.45 + 0.84	2.05 ± 1.84	25.62	114.95	7.07

表 6 Cifar10 中不同 GANs 对比

Table 6 Comparsion of different GANs in Cifar10

GANs 模型	epoch 数	优化项	参数量 ($\times 10^6$)	IS ($\sigma \times 0.1$)	FID	清晰度均值	清晰度均值差值
训练集	—	—	—	10.70 ± 1.47	0.00	120.56	0.00
BEGANs[16]	35	沃瑟斯坦距离	**3.67**	5.36 ± 0.65	107.64	80.89	39.67
DCGANs[9]	20	JS 散度	8.83	5.04 ± 0.27	54.27	139.12	18.56
LSGANs[15]	35	Pearson 散度	8.83	5.70 ± 0.36	43.35	135.80	15.24
WGANs[14]	35	沃瑟斯坦距离	8.83	5.25 ± 0.33	53.88	136.74	16.18
WGANsGP[17]	35	沃瑟斯坦距离	8.83	5.39 ± 0.30	50.60	**139.17**	18.61
SAGANs1[23]	30	沃瑟斯坦距离	8.57	**6.09** ± 0.47	42.90	126.28	**5.72**
SAGANs2[23]	30	JS 散度	8.57	5.37 ± 0.46	53.49	133.54	12.98
本文方法	**15**	JS + λ· KL 散度	8.83 + 0.23	5.83 ± 0.70	**42.70**	134.97	14.41

上, 它能更合理地模拟人脸纹理信息, 虽然参数量两者基本一致, 但其训练 epoch 数明显多于本文 GANs 模型. 再对比 SAGANs2 和本文 GANs 可知, 本文综合效果又较明显优于优化 JS 散度的 SAGANs 模型. 说明当 JS 散度作为优化目标或主优化目标时, 本文 GANs 模型比融入注意力机制和谱归一化优化的 SAGANs 模型表现更佳. 同时, 通过对应地对比 DCGANs 与 WGANs、SAGANs1 与 SAGANs2, 可以看出, 优化 JS 散度模型生成图像质量差于优化沃瑟斯坦距离模型生成图像质量. 这也证明了 WGANs[14] 的分析, JS 散度的确可能带来梯度消失问题, 导致生成图像质量下降.

由表 6 可知, 在 Cifar10 数据集中依然存在上述类似的实验现象, 但从统计的数据来看, 没有单类别数据集那么明显.

通过以上实验数据及分析可知, 本文 GANs 综合性能达到除了优化沃瑟斯坦距离的 SAGANs 外的最优效果. 相对而言, 本文 GANs 在仍以 JS 散度为主优化目标时, 模型综合性能靠近优化沃瑟斯坦距离的 SAGANs, 并且网络结构并没有使用注意力机制和谱归一化优化. 同时在预训练提取出训练特征的前提下, 本文 GANs 模型明显减少 epoch 数.

由图 12 ~ 19 的展示, 可以直观地对比 BEGANs、DCGANs、WGANsGP 和 SAGANs1 的 GANs 生成效果.

图 12 Celeba 中 BEGANs 实验样本

Fig. 12 Experimental samples of BEGANs in Celeba

图 13 Celeba 中 DCGANs 实验样本

Fig. 13 Experimental samples of DCGANs in Celeba

图 14 Celeba 中 WGANsGP 实验样本

Fig. 14 Experimental samples of WGANsGP in Celeba

图 15 Celeba 中 SAGANs1 实验样本

Fig. 15 Experimental samples of SAGANs1 in Celeba

对比分析使用 Celeba 数据集训练 GANs 而生成的图像. 由图 12 可知, BEGANs 虽然能很好对形态特征进行学习, 但的确存在比较严重的高频信息丢失现象, 并且生成的图像出现斑块. 由图 13 ~ 14 可知, DCGANs 和 WGANsGP 生成的图像纹理信息填充区域过多, 比如训练图像面部的高频信息较少, 但是生成图像存在面部填充高频信息的现象,

图 16 Cifar10 中 BEGANs 实验样本

Fig. 16 Experimental samples of BEGANs in Cifar10

图 17 Cifar10 中 DCGANs 实验样本

Fig. 17 Experimental samples of DCGANs in Cifar10

图 18 Cifar10 中 WGANsGP 实验样本

Fig. 18 Experimental samples of WGANsGP in Cifar10

图 19 Cifar10 中 SAGANs1 实验样本

Fig. 19 Experimental samples of SAGANs1 in Cifar10

这也是表 4 对应的清晰度均值项偏高的原因之一. 图 15 能很明显地观察到优化沃瑟斯坦距离的 SAGANs 生成的图像, 在面部形态和纹理等特征更合理, 并且结合图 8 (本文效果), 也能发现更好地生成图像样本, 其形态和纹理等信息都比较协调. 对比 Cifar10 数据集生成的图像, 除图 16 可以明显看出差异外, 难以直接进行视觉评估, 在第 3.5 节和表 6 数据进行了分析.

综上所述, 本文方法 (JS $+\lambda\cdot$ KL 散度) 相对于 DCGANs (JS 散度) 有较明显的提升, 在 IS 指标上也能达到 LSGANs (Pearson 散度)、WGANs (沃瑟斯坦距离) 等 GANs 模型的图像生成效果, 并且在 FID 指标上进一步有所提高. 此外, 本文方法生成的图像效果能逼近优化沃瑟斯坦距离的 SAGANs 图像效果, 并且参数量并没明显增加. 在训练集样本特征预学习完成后, 解码及对抗学习能有效减少训练所需的 epoch 数.

4 结束语

为提高 GANs 图像生成质量, 考虑到 JS 散度可能为近似常数时带来对生成图像效果的不利影响, 本文尝试通过增加样本特征解码约束条件来减弱这些影响. 实验结果表明, 利用样本特征解码约束条件进行对抗训练的约束, 有利于图像生成质量提高和减少 epoch 数. 同时, 本文方法能够更合理地模拟训练集的高频信息部分. 本文方法需对训练样本预学习出样本特征, 虽较少地增加了网络参数量, 但需要较多的特征提取预训练时间. 对于其他特征提取方法, 特征分布与随机噪声分布的关系对生成效果的影响值得进一步研究.

References

1 Goodfellow I J, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. In: Proceedings of the International Conference on Neural Information Processing Systems. Montreal, Canada: 2014. 2672−2680

2 Creswell A, White T, Dumoulin V, Arulkumaran K, Sengupta B, Bharath A A. Generative adversarial networks: An overview. *IEEE Signal Processing Magazine*, 2018, **35**(1): 53−65

3 Hong Y J, Hwang U, Yoo J, Yoon S. How generative adversarial networks and their variants work: An overview. *ACM Computing Surveys*, 2019, **52**(1): 1−43

4 Mirza M, Osindero S. Conditional generative adversarial nets. arXiv preprint, 2014, arXiv: 1411.1784v1

5 Odena A. Semi-supervised learning with generative adversarial networks. arXiv preprint, 2016, arXiv: 1606.01583v2

6 Odena A, Olah C, Shlens J. Conditional image synthesis with auxiliary classifier GANs. In: Proceedings of the International Conference on Machine Learning. Sydney, Australia: 2017.

7 Chen X, Duan Y, Houthooft R, Schulman J, Sutskever I, Abbeel P. InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets. In: Proceedings of the International Conference on Neural Information Processing Systems. Barcelona, Spain: 2016. 2180−2188

8 Donahue J, Krahenbuhl K, Darrell T. Adversarial feature learning. In: Proceedings of the International Conference on Learning Representations. Toulon, France: 2017.

9 Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. In: Proceedings of the International Conference on Learning Representations. San Juan, Puerto Rico: 2016.

10 Denton E, Chintala S, Szlam A, Fergus R. Deep generative image using a laplacian pyramid of adversarial networks. In: Proceedings of the International Conference on Neural Information Processing Systems. Montreal, Canada: 2015. 1486−14944

11 Brock A, Donahue J, Simonyan K. Large scale GAN training for high fidelity natural image synthesis. In: Proceedings of the International Conference on Learning Representations. New Orleans, USA: 2019.

12 Nguyen T D, Le T, Vu H, Phung D. Dual discriminator generative adversarial nets. In: Proceedings of the Proceedings of International Conference on Neural Information Processing Systems. Long Beach, USA: 2017.

13 Zhang Long, Zhao Jie-Yu, Ye Xu-Lun, Dong Wei. Cooperative generative adversarial nets. *Acta Automatica Sinica*, 2018, **44**(5): 804−810
(张龙, 赵杰煜, 叶绪伦, 董伟. 协作式生成对抗网络. 自动化学报, 2018, **44**(5): 804−810)

14 Arjovsky M, Chintala S, Bottou L. Wasserstein generative adversarial networks. In: Proceedings of the International Conference on Machine Learning. Sydney, Australia: 2017. 214−223

15 Mao X D, Li Q, Xie H R, Lau R Y K, Wang Z, Smolley S P. Least squares generative adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy: 2017. 2813−2821

16 Berthelot D, Schumm T, Metz L. BEGAN: Boundary equilibrium generative adversarial networks. arXiv preprint, 2017, arXiv: 1703.10717v4

17 Gulrajani I, Ahmed G, Arjovsky M, Dumoulin V, Courville A. Improved training of Wasserstein GANs. In: Proceedings of the International Conference on Neural Information Processing Systems. Long beach, USA: 2017.

18 Wu J Q, Huang Z W, Thoma J, Acharya D, Gool L V. Wasserstein divergence for GANs. In: Proceedings of the European Conference on Computer Vision. Munich, Germany: 2018. 673−688

19 Su J L. GAN-QP: A novel GAN framework without gradient vanishing and Lipschitz constraint. arXiv preprint, 2018, arXiv: 1811.07296v2

20 Zhao J B, Mathieu M, LeCun Y. Energy-based generative adversarial networks. In: Proceedings of the International Conference on Learning Representations. Toulon, France: 2017

21 Wang Gong-Ming, Qiao Jun-Fei, Qiao Lei. A generative adversarial network in terms of energy function. *Acta Automatica Sinica*, 2018, **44**(5): 793−803
(王功明, 乔俊飞, 乔磊. 一种能量函数意义下的生成式对抗网络. 自动化学报, 2018, **44**(5): 793−803)

22 Qi G J. Loss-sensitive generative adversarial networks on Lipschitz densities. arXiv preprint, 2017, arXiv: 1701.06264v5

23 Zhang H, Goodfellow I, Metaxas D, Odena A. Self-attention generative adversarial networks. In: Proceedings of the International Conference on Machine Learning. Long Beach, USA: 2019

24 Cover T M, Thomas J A. *Elements of Information Theory*. New York: John Wiley & Sons Inc., 2006. 12−49

25 Nielsen F. A family of statistical symmetric divergences based on Jensen′s inequality. arXiv preprint, 2011, arXiv: 1009.4004v2

26 Goodfellow I, Bengio Y, Courville A. *Deep Learning*. Cambridge: MIT Press, 2017. 502−525

27 Alain G, Bengio Y. What regularized auto-encoders learn from the data-generating distribution. *The Journal of Machine Learning Research*, 2014, **15**(1): 3563−3593

28 Vincent P, Larochelle H, Bengio Y, Manzagol P A. Extracting and composing robust features with denoising autoencoders. In: proceedings of the International Conference on Machine Learning. Rhineland, Germany: 2008.

29 Rifai S, Vincent P, Muller X, Glorot X, Bengio Y. Contractive auto-encoders: Explicit invariance during feature extraction. In: proceedings of the International Conference on Machine Learning. Washington, USA: 2011.

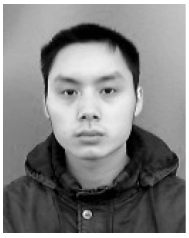
30 Kavukcuoglu K, Ranzato M, LeCun Y. Fast inference in sparse coding algorithms with applications to object recognition. arXiv preprint, 2010, arXiv: 1010.3467v1

31 Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. In: Proceedings of the International Conference on Medical Image Computing and Computer-assisted Intervention. Munich, Germany: 2015. 234−241

32 Salimans T, Goodfellow I J, Zaremaba W, Cheung V, Radford A, Chen X. Improved techniques for training GANs. In: Proceedings of the International Conference on Neural Information Processing Systems. Barcelona, Spain: 2016. 2234−2242

33 Xu Q T, Huang G, Yuan Y, Huo C, Sun Y, Wu F, et al. An empirical study on evaluation metrics of generative adversarial networks. arXiv preprint, 2018, arXiv: 1806.07755v2

34 Shmelkov K, Schmid C, Alahari K. How good is my GAN? In: Proceedings of the European Conference on Computer Vision. Munich, Germany: 2018. 218−234



陈泓佑 西南交通大学信息科学与技术学院博士研究生. 主要研究方向为机器学习和图像处理.
E-mail: chy2019@foxmail.com
(**CHEN Hong-You** Ph.D. candidate at the School of Information Science and Technology, Southwest Jiaotong University. His research interest covers machine learning and image processing.)

陈 帆 西南交通大学信息科学与技术学院副教授. 主要研究方向为多媒体安全和计算机应用.
E-mail: fchen@home.swjtu.edu.cn
(**CHEN Fan** Associate professor at the School of Information Science and Technology, Southwest Jiaotong University. His research interest covers multi-media security and computer applications.)

和红杰 西南交通大学信息科学与技术学院教授. 主要研究方向为图像取证和图像处理. 本文通信作者.
E-mail: hjhe@home.swjtu.edu.cn
(**HE Hong-Jie** Professor at the School of Information Science and Technology, Southwest Jiaotong University. Her research interest covers image forensics and image processing. Corresponding author of this paper.)

朱翌明 西南交通大学信息科学与技术学院硕士研究生. 主要研究方向为深度学习和图像处理.
E-mail: swjtu163zym@163.com
(**ZHU Yi-Ming** Master student at the School of Information Science and Technology, Southwest Jiaotong University. His research interest covers deep learning and image processing.)