Advances in Applied Mathematics 应用数学进展, 2025, 14(5), 660-669
Published Online May 2025 in Hans. https://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2025.145291

Hans汉斯

防止生成对抗网络模式坍塌及梯度消失的一种有效方法

熊　梅[1]，陈嘉迪[1]，陈龙伟[2,3*]，余益民[2,3]

[1]云南财经大学统计与数学学院，云南　昆明
[2]云南财经大学云南省服务计算重点实验室，云南　昆明
[3]云南财经大学智能应用研究院，云南　昆明

收稿日期：2025年4月29日；录用日期：2025年5月23日；发布日期：2025年5月31日

摘　要

针对深度学习中的生成对抗网络GAN的模式坍塌和梯度消失问题，通过Wasserstein距离和梯度惩罚的方法，缓解模式坍塌和梯度消失。主要方法包括使用Wasserstein距离代替JS散度改进损失函数，同时防止梯度消失；利用梯度惩罚法强制判别器的连续性约束，确保梯度稳定，平衡生成器与判别器的训练速度；采用从简单到复杂的训练策略优化，平滑过渡，防止判别器过强，同时也防止梯度消失。对传统GAN和改进GAN进行了对比，其生成器和判别器的Loss曲线显示改进后较为稳定。

关键词

生成对抗网络，模式坍塌，梯度消失，正则化，梯度惩罚

An Effective Method to Prevent the Generative Adversarial Network Mode Collapse and Gradient Disappearance

Mei Xiong[1], Jiadi Chen[1], Longwei Chen[2,3*], Yimin Yu[2,3]

[1]School of Statistics and Mathematics, Yunnan University of Finance and Economics, Kunming Yunnan
[2]Yunnan Key Laboratory of Service Computing, Yunnan University of Finance and Economics, Kunming Yunnan
[3]Institute of Intelligence Applications, Yunnan University of Finance and Economics, Kunming Yunnan

Received: Apr. 29th, 2025; accepted: May 23rd, 2025; published: May 31st, 2025

*通讯作者。

文章引用: 熊梅, 陈嘉迪, 陈龙伟, 余益民. 防止生成对抗网络模式坍塌及梯度消失的一种有效方法[J]. 应用数学进展, 2025, 14(5): 660-669. DOI: 10.12677/aam.2025.145291

Abstract

Aiming at the problem of pattern collapse and gradient disappearance of generative adversarial network GAN in deep learning, Wasserstein distance and gradient penalty are used to alleviate the pattern collapse and gradient disappearance. The main methods include using Wasserstein distance instead of JS divergence to improve the loss function, while preventing the gradient from disappearing. The gradient penalty method is used to force the continuity constraint of discriminator to ensure the stability of gradient and balance the training speed of generator and discriminator. Smooth transitions are optimized using training strategies from simple to complex, preventing the discriminator from being too strong and also preventing the gradient from disappearing. A comparison is made between the traditional GAN and the improved GAN, the loss curves of its generator and discriminator show that the improved one is more stable.

Keywords

Generative Adversarial Network, Mode Collapse, Gradient Disappears, Regularization, Gradient Punishment

Copyright © 2025 by author(s) and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

1. 引言

1.1. 生成对抗网络的模式坍塌问题

生成对抗网络 GAN (Generative Adversarial Networks)是深度学习中的一种重要方法，其模式坍塌(Mode Collapse)是训练过程中常见的问题之一[1]。其含义是生成器倾向于生成相似或重复的样本，无法覆盖真实数据分布的多样性。主要原因有如下几个方面：

对抗训练的博弈失衡：生成器(Generator)与判别器(Discriminator)的对抗可能导致生成器“走捷径”，仅仅生成能欺骗当前判别器的少数样本，并非学习真实数据分布的全貌。另一方面是生产器可能通过极小化单一模式而非全分布来降低损失。

损失函数的局限性：传统 GAN 使用 JS (Jensen-Shannon Divergence)，当生成分布与真实分布不重叠时，梯度消失(亦称梯度真空)问题。生成器可能通过极小化单一模式而非全分布来降低损失。

数据分布的复杂性：真实数据分布可能包含多个子模式，比如不同类别的图像，生成器难以捕捉所有模式。

优化过程的局部最优：生产器陷入局部最优，反复生成同一类样本。

模式坍塌给 GAN 带来严重影响，主要表现在生成样本的多样性低，如生成人脸时性别单一或肤色单一；模型的实用性下降，无法用于多样性任务，如数据增强；造成训练不稳定，生成器与判别器的对抗产生震荡或崩溃。

1.2. 生成对抗网络的梯度消失

生成对抗网络中断梯度消失的现象是训练过程中常见的问题之一，尤其是判别器(Discriminator)过于强大时，来自判别器的反向传播梯度非常小，生成器(Generator)可能无法获得有效的梯度信号，导致训练

停滞。损失函数存在设计缺陷，因为损失函数依赖于 JS 散度，当真实分布与生成分布之间没有重叠时，JS 散度的梯度会趋近于零，从而无法提供有效的学习信号。此外，模式坍塌也会让梯度消失[2]。

2. 解决模式坍塌的主要方法

本文主要考虑通过 Wasserstein 距离和梯度惩罚缓解模式坍塌。

改进损失函数：WGAN 方法(Wasserstein GAN)，使用 Wasserstein 距离代替 JS 散度，缓解梯度消失问题，通过梯度剪裁或梯度惩罚(WGAN-GP)进一步稳定训练。LSGAN (最小二乘 GAN)方法，用最小二乘损失替代交叉熵，缓解模式坍塌。合页损失法(Hinge Loss)增强判别器的鲁棒性，防止生成器过度优化单一模式。

正则化约束：可以采用梯度惩罚(Gradient Penalty)在 WGAN-GP 中强制判别器 Lipschitz 连续性约束，确保梯度稳定，平衡生成器与判别器的训练速度。谱归一化(Spectral Normalization)对叛别器权重进行谱归一化，稳定训练动态。多样性正则化(Diversity Regularization)强制生成样本间的差异[3]。

训练策略优化：逐渐增加生成任务的复杂度；将判别器的标签平滑，防止判别器过强，同时还可以防止梯度消失；在判别器输入中混合真实样本和数据与生成数据，防止判别器过度自信。

两步时间的更新，为生成器和判别器设置不同的学习率，可以使生成器的学习率略高于判别器。

3. 理论推导

从理论上来说，对 GAN 的优化可以视为对目标函数的极大与极小化，生成对抗网络的生成器与判别器的损失函数为：

$$\text{Loss}_{G_1} = E_{z\sim Pz}\left(\log\left(1-D(G)(z)\right)\right)$$

$$\text{Loss}_{G_2} = -E_{z\sim Pz}\left(\log\left(D(G)(z)\right)\right)$$

$$\text{Loss}_{\text{D}} = E_{x\sim P_{\text{data}}}\left(\log\left(D(x)\right)\right) + E_{z\sim P_z}\left(\log\left(1-D(G)(z)\right)\right)\text{Loss}_{G_1}$$

其中，Loss_{G_1} 为生成器网络处于饱和状态的损失函数，Loss_{G_2} 是生成器网络处于非饱和状态的损失函数。为避免出现梯度的消失，并使得 GAN 网络能在训练中趋于稳定，通常使用饱和状态的损失函数。

定理 1 由原始 GAN 可知，当生成器固定时[4]，有：

$$D_G^*(x) = \frac{P_{\text{data}}(x)}{P_{\text{data}}(x) + P_g(x)}$$

定理 2 设 $C(G) = \max V(D,G)$，如果判别器网络达到最优时，那么 $C(G) = 2\text{JSD}\left(P_{\text{data}} \| P_g\right) - \log 4$，当 $P_{\text{data}} = P_g$，则 $C(G) = -\log 4$ 取到最小值[4]。式中，$\text{JSD}\left(P_{\text{data}} \| P_g\right)$ 表示 JS 散度，用于刻画真实数据分布 P_{data} 与生成数据分 P_g 之间的距离。

定理 2 表明，当 D 取到最优判别器时，如果要最小化目标函数 $C(G)$，则相当于最小化 $\text{JSD}\left(P_{\text{data}} \| P_g\right)$，所以训练生成对抗网络的目标是使得 JS 散度最小化，使得生成数据样本的分布不断地趋近于原始数据的分布，最终使得生成样本接近于真实样本。但是判别器 D 只能在理想状态下取得最优化判别器，在 GAN 的实际训练过程中，GAN 只能通过不断地训练判别器，使得 D 不断地趋近于最优化判别器 D_G^*。大多数情况下，判别器并不能达到最优化。此时对生成器的目标函数进行最小化相当于最小化 JS 散度与对数函数。

通过定理 1 可以计算出 D_G^*，但是 P_{data} 和 P_g 没有显性的计算公式，因此没有办法直接求得 D_G^*。当且仅当生成数据分布与原始数据分布重合时，可得出 D_G^* 为 1/2。在实际的训练过程中，$P_{\text{data}} = P_g$ 也只有在理想状态下才成立，因此 D_G^* 也只能在理想状态下才能得到。

在训练生成对抗网络时，较难判断 D 有没有达到 D_G^*。

因此，设 $D(x)=\frac{1}{a}D_G^*(x)$ ，其中 $a>D_G^*(x)$。因此，可以用 $\frac{1}{a}$ 来刻画 D 和 D_G^* 的接近程度，$\frac{1}{a}$ 也可以用来表征 D 辨别真假样本的能力。当 a 越接近于 1，代表判别器越趋近于最优化判别器。生成器损失原始形式生成器试图最小化：

$$C(G)=E_{x\sim P_g}\left[\log\left(1-D(x)\right)\right]$$

将 $D(x)=\frac{1}{a}D^*(x)$ 代入：

$$\begin{aligned}C(G)&=E_{x\sim P_g}\left[\log\left(1-\frac{1}{a}\cdot\frac{P_{\text{data}}(x)}{P_{\text{data}}(x)+P_g(x)}\right)\right]\\&=E_{x\sim P_g}\left[\log\left(\frac{P_{\text{data}}(x)+P_g(x)-\frac{1}{a}P_{\text{data}}(x)}{P_{\text{data}}(x)+P_g(x)}\right)\right]\\&=E_{x\sim P_g}\left[\log\left(\frac{\left(1-\frac{1}{a}\right)P_{\text{data}}(x)+P_g(x)}{P_{\text{data}}(x)+P_g(x)}\right)\right].\end{aligned}$$

由定理 2 当判别器最优时 $C(G)=2\text{JSD}\left(P_{\text{data}}\|P_g\right)-\log 4$，引入参数 a 的影响：

$$\begin{aligned}C(G)&=E_{x\sim P_g}\left[\log\left(\frac{P_g(x)+\left(1-\frac{1}{a}\right)P_{\text{data}}(x)}{P_{\text{data}}(x)+P_g(x)}\right)\right]\\&\overset{\text{积分展开}}{=}\int P_g(x)\log\left(\frac{P_g(x)+\left(1-\frac{1}{a}\right)P_{\text{data}}(x)}{P_g(x)+P_{\text{data}}(x)}\right)\text{d}x\\&=\int P_g(x)\log\left(\frac{P_g(x)}{P_g(x)+P_{\text{data}}(x)}\right)\text{d}x\\&\quad\underset{\text{JSD项}}{}\\&\quad+\int P_g(x)\log\left(1+\frac{\left(1-\frac{1}{a}\right)P_{\text{data}}(x)}{P_g(x)}\right)\text{d}x.\end{aligned}$$

由 $\text{JSD}\left(P\|Q\right)=\frac{1}{2}\text{KL}\left(P\|\frac{P+Q}{2}\right)+\frac{1}{2}\text{KL}\left(Q\|\frac{P+Q}{2}\right)$，进一步整理：

$$\begin{aligned}C(G)&=2\text{JSD}\left(P_{\text{data}}\|P_g\right)-\log 4\\&\quad+E_{x\sim P_g}\left[\log\left(1+\frac{\left(1-\frac{1}{a}\right)P_{\text{data}}(x)}{P_g(x)}\right)\right].\end{aligned}$$

令 $\lambda = 1 - \frac{1}{a}$，$C(G) = 2\mathrm{JSD}(P_{\text{data}} \| P_g) - \log 4 + \mathbb{E}_{x\sim P_g}\left[\log\left(1 + \lambda \cdot \frac{P_{\text{data}}(x)}{P_g(x)}\right)\right]$。

由推论可知，当判别器未能达到最优时即 $a \neq 1$ 时，生成器的优化目标等效于对 $E_{x\sim P_g}\left[\log\left(1 + \lambda \cdot \frac{P_{\text{data}}(x)}{P_g(x)}\right)\right]$ 的联合优化。为进一步分析其影响，定义生成器网络参数为 θ，并令 $r(x) = \frac{P_{\text{data}}(x)}{P_g(x;\theta)}$ 以简化表达，则生成器梯度可分解为：

$$\nabla_\theta C(G) = \nabla_\theta (2\mathrm{JSD} - \log 4) + E_{x\sim P_g}\left[\frac{\lambda r(x)}{1 + \lambda r(x)} \cdot \left(-\frac{\nabla_\theta P_g(x;\theta)}{P_g(x;\theta)}\right)\right]$$

当 $r(x) \to \infty$ 时，$\frac{\lambda r(x)}{1 + \lambda r(x)} \to \lambda$，梯度简化为：

$$\nabla_\theta C_{\text{correction}} \approx -\lambda \cdot E_{x\sim P_g}\left[\frac{\nabla_\theta P_g}{P_g}\right]$$

若 $\lambda > 0$，即 $a > 1$ 时，判别器保守，梯度强制生成器降低低概率区域的分布，加剧模式坍塌；若 $\lambda < 0$，即 $a < 1$ 时，判别器激进，梯度方向逆转为提升 P_g，但与 JSD 项的目标产生对抗，导致参数震荡。

当 $r(x) \to 1$ 时，修正项退化为 $\log(1 + \lambda)$，梯度贡献小，此时 JSD 项主导训练，但若判别器无法稳定至最优 $a \neq 1$，该平衡点将被扰动向非理想状态偏移。

考虑修正项内层函数 $f(r) = \log(1 + \lambda r)$，其二阶特性显示其极值点敏感性，令 $f'(r) = \frac{\lambda}{1 + \lambda r} = 0$，则极值仅在 $\lambda = 0$ 时存在，此时退化为 JS 散度；当 $\lambda \neq 0$ 时，$f(r)$ 无自然极值，导致梯度动态完全由比值 $r(x)$ 驱动。若对参数 θ 进行约束或引入正则化，虽可抑制梯度爆炸，但会削弱分布的拟合能力。

在训练初期，判别器因快速收敛至次优状态(a 远离 1)使 λ 显著偏离平衡值。

例如，当 $a \to \infty$ 时(判别器过早过拟合)，$\lambda \to 1$，修正项梯度幅值放大为：

$$\|\nabla_\theta C_{\text{correction}}\| \propto E\left[\frac{\nabla_\theta P_g}{P_g} \cdot \frac{1}{1 + P_{\text{data}}/P_g}\right]$$

此时若 $P_g(x)$ 在真实数据区域未充分覆盖 $(P_g \ll P_{\text{data}})$，则分母 $1 + P_{\text{data}}/P_g$ 极小，梯度剧烈发散，导致参数更新失稳甚至数值溢出。

上述分析揭示，生成器梯度动力学的脆弱性本质源于 $r(x)$ 的非对称动力学特征。当生成分布 $P_g(x)$ 与真实分布 $P_{\text{data}}(x)$ 在局部区域发生显著偏离时，传统梯度机制对 $r(x)$ 的极端值高度敏感：在低密度区域 $(P_g(x) \to 0,\ r(x) \to \infty)$，梯度权重 $\nabla_\theta \log r(x)$ 呈现无界增长，促使参数更新方向陷入剧烈高频震荡；在高置信区域 $(r(x) \approx 1)$，线性优化策略却无法充分捕获分布的细微结构差异。这种强区域过敏感、弱区域欠驱动的矛盾，直接导致模式坍塌、训练发散等典型失衡现象，暴露出基于传统概率散度的优化框架的刚性缺陷，即其对概率空间的变化难以实现非线性自适应响应。为解决这一根本矛盾，必须引入一种能够解耦梯度幅值与方向动态的几何约束。在此背景下，闵可夫斯基距离(Lp 范数)的数学特性可以通过参数 p 的调节，构造一种非线性梯度压缩机制。

4. 损失函数的改进

本文引入闵可夫斯基距离更新原始的 GAN 的损失函数，其中闵可夫斯基距离表达式如公式所示：

$$D(x,y)=\left(\sum_i |x_i-y_i|^p\right)^{\frac{1}{p}}$$

其中， D 为距离， p 为距离的阶数， x,y 为向量。引入闵可夫斯基距离后，判别器和生成器的损失函数可以写成：

$$\text{Loss}_D=\frac{1}{p}E_{x\sim P_{\text{data}}}\left(D(x)-1\right)^p+\frac{1}{p}E_{z\sim P_Z}\left(D(G(z))\right)^p$$

$$\text{Loss}_G=\frac{1}{p}E_{Z\sim P_Z}\left(D(G(z))-1\right)^p$$

在式子前面加上系数 $\frac{1}{p}$ 是为了在求导时使得前面的系数为 1：

$$\begin{aligned}pC(G)&=E_{x\sim P_{\text{data}}}\left(D_G^*(x)-1\right)^p+E_{x\sim P_g}\left(D_G^*(x)-1\right)^p\\&=E_{x\sim P_{\text{data}}}\left(\frac{P_{\text{data}}(x)}{P_{\text{data}}(x)+P_g(x)}-1\right)^p+E_{x\sim P_g}\left(\frac{P_{\text{data}}(x)}{P_{\text{data}}(x)+P_g(x)}-1\right)^p\\&=\int P_{\text{data}}(x)\left(\frac{-P_g(x)}{P_{\text{data}}(x)+P_g(x)}\right)^p\mathrm{d}x+\int P_g(x)\left(\frac{-P_g(x)}{P_{\text{data}}(x)+P_g(x)}\right)^p\mathrm{d}x\\&=\int P_{\text{data}}(x)\frac{\left(-P_g(x)\right)^p}{\left(P_{\text{data}}(x)+P_g(x)\right)^{p-1}}\mathrm{d}x\end{aligned}$$

为使得积分取得正值，p 只能取偶数，且为可调参数。

5. 实验对比

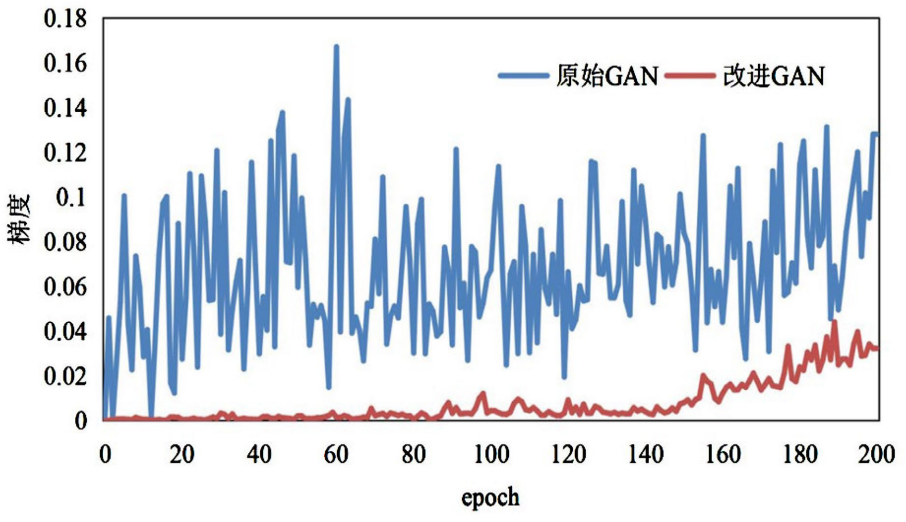

Figure 1. Gradient variations of initial GAN and improved GAN
图 1. 原始 GAN 与改进 GAN 的梯度变化

为了验证本文提出方法的有效性，本文将原始 GAN 与改进损失函数过后的 GAN 在不同数据集上进行对比试验。在 CIFAR-10 数据集上，原始 GAN 和改进过后的 GAN 的判别器网络的梯度变化如图 1 所

示。由图 1 可知，改进过后的 GAN 在训练的过程中相比于原始 GAN，其梯度的波动较小。原始的 GAN 的判别器网络在训练的过程中，梯度处于不断的震荡，说明原始 GAN 的训练中判别器网络的梯度变化极不稳定，这也是 GAN 训练不稳定的重要原因。改进过后的损失函数加入到 GAN 中，能有效地防止 GAN 出现训练不稳定。

为了便于观察生成器和判别器训练过程的变化情况，输出原始 GAN 与判别器的生成器以及判别器的 Loss 曲线，如图 2 和图 3 所示。红色的曲线代表改进的 GAN，蓝色的曲线表示原始 GAN。在对原始 GAN 进行改进损失函数后，生成器和判别器的 Loss 曲线变得较稳定，没有像原始 GAN 一样出现较大的波动。由此可以得出，本文改进了损失函数之后，使得 GAN 的训练稳定性得到提升。

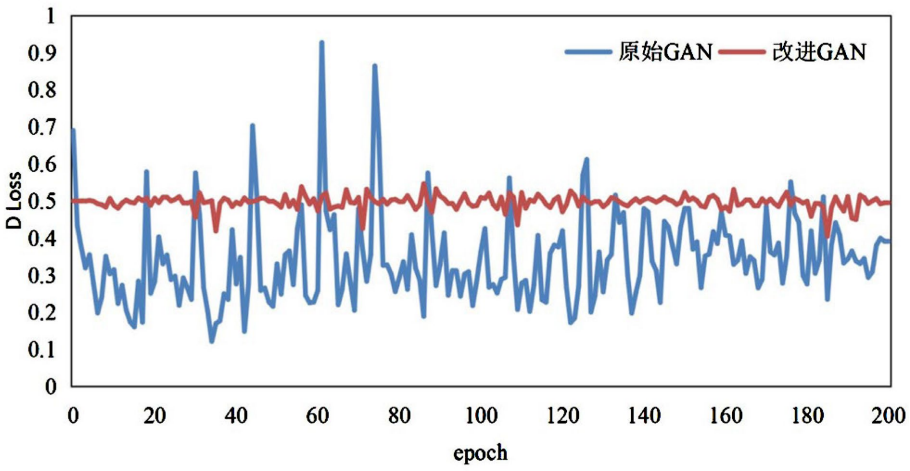

Figure 2. Discriminator Loss curve of initial GAN and improved GAN
图 2. 原始 GAN 与改进 GAN 的判别器 Loss 曲线

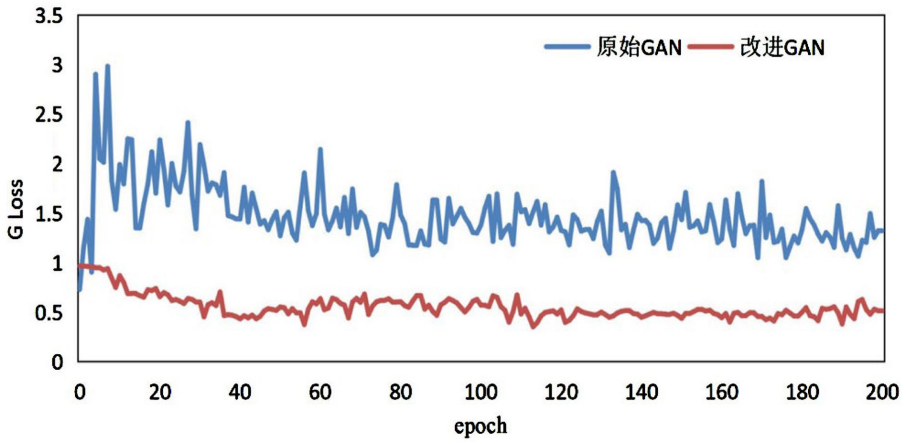

Figure 3. Generator Loss curve of initial GAN and improved GAN
图 3. 原始 GAN 与改进 GAN 的生成器 Loss 曲线

为验证本文提出方法的有效性，我们将其与原始 GAN 在标准数据集 CIFAR-10 上进行对比。两种网络的判别器采用梯度裁剪策略，权重裁剪区间为$[-0.01, 0.01]$，均使用 Adam 优化器，学习率 $2e-4$, $\beta_1 = 0.5$, $\beta_2 = 0.999$，训练轮数 200 epochs，batch size 设置为 64。图 4 和图 5 为原始 GAN、WGAN 与改进 GAN 在 CIFAR-10 数据集上的生成效果对比，由图 4 和图 5 可以看出，改进 GAN 的生成效果在图像的质量上与多样性上明显地要优于原始 GAN 和 WGAN。通过实验验证了本文提出方法的有效性。

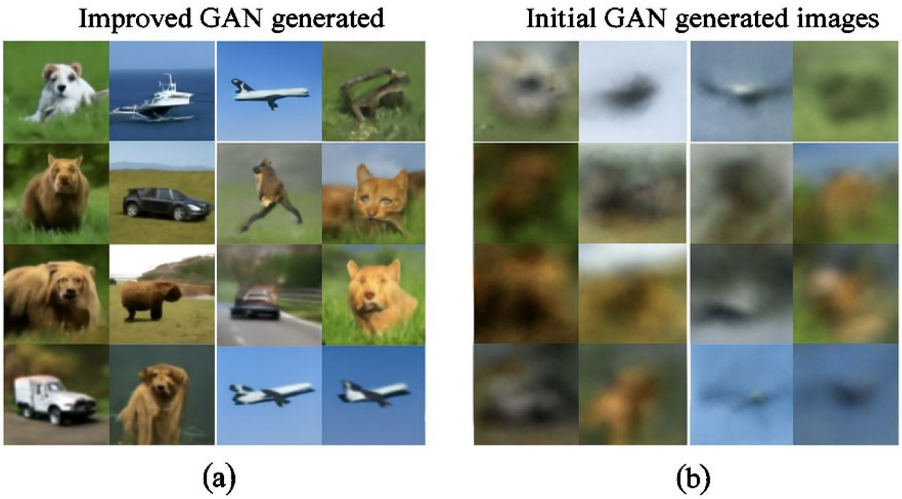

Figure 4. Figure comparison of improved GAN (left) and initial GAG (right) on CIFAR-10 dataset: (a) Improve GAN discriminator figure; (b) Initial GAN discriminator figure
图 4. 改进 GAN (左图)与原始 GAN(右图)在 CIFAR-10 数据集上的图片比较：(a) 改进的 GAN 判别图；(b) 原始 GAN 判别图

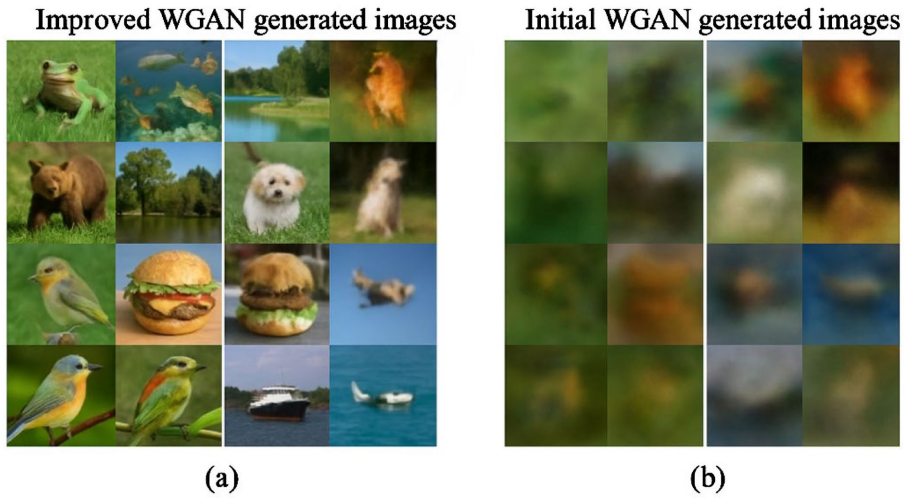

Figure 5. Figure comparison of improved GAN (left) and initial WGAG (right) on CIFAR-10 dataset: (a) Improve GAN discriminator figure; (b) Initial WGAN discriminator figure
图 5. 改进 GAN (左图)与原始 WGAN (右图)在 CIFAR-10 数据集上的图片比较：(a) 改进的 GAN 判别图；(b) 原始 WGAN 判别图

FID 作为 GAN 模型的常用评价指标，在一定程度上能代表生成图片的质量与多样性。FID 分数越低，表示 GAN 生成图片的质量越好，生成图片的多样性越高，生成模型的生成能力越好。本文通过对比实验，对原始 GAN、WGAN 与改进的 GAN 在 CIFAR-10 数据集上进行训练。为了使得实验的结果公正客观，本文实验中 WGAN、原始 GAN 与改进的 GAN 的所有参数都一样，具体的参数设置如表 1 所示。

Table 1. Table of experiment parameter setting
表 1. 实验参数设置表

参数名称	训练轮数	batchsize	学习率	随机噪声维度
参数值	200	64	0.0002	100

在三个 GAN 网络训练好之后，分别利用原始 GAN、WGAN (Wasserstein 距离)与改进的 GAN 的生成器分别生成 5000 张图片，并分别将各自的 5000 张生成图片与原始数据集的 5000 张图片计算 FID 分数。具体的实验结果如表 2 所示。

Table 2. FID values of three GAN on CIFAR-10
表 2. 三种生成对抗网络在 CIFAR-10 上的 FID 值

模型	原始 GAN	WGAN	改进 GAN
FID	197.8	210.3	99.6

由实验结果可得，本文提出改进 GAN 在 FID 上明显低于原始 GAN，说明了本文的 GAN 通过改进原始的损失函数，提高了 GAN 的训练稳定性，并且提高了生成样本的质量和多样性，通过实验证明了本文提出的改进的损失函数的有效性。由于改进的损失函数中存在着可调参数 p，为了获得最佳的损失函数，那么就要找到最佳的参数 p。因此，本文进行了消融实验，分别将参数 p 设定为 $1,2,3,4,5,6,7,8$ 进行对比实验，并继续利用 FID 分数对 GAN 的生成效果进行评价，不同 p 值的实验结果如表 3 所示。

Table 3. Results of experiment on different p
表 3. 不同 p 的实验结果

p	1	2	3	4	5	6	7	8
FDI	128.4	123.6	105.8	99.6	108.5	114.8	137.4	149.6

通过对比实验的结果，可以得出当 $p=4$ 时，生成器网络的生成效果最好。在 p 的值从 1 到 8 进行变化时，对应的 FID 值有较明显的变化，这也表明了超参数 p 对于实验的结果有较大的影响，本文通过实验，发现当 $p=4$ 时，GAN 的训练具有更好的效果。而当 $p=4$ 时，判别器和生成器的损失函数可以写成：

当 $p=4$ 时，损失函数对四阶矩的惩罚强度最大。

证明：闵可夫斯基损失函数定义为：

$$C(G)=E\left[\|\varepsilon\|_p^p\right]=E\left[\sum_{i=1}^{d}|\varepsilon_i|^p\right]$$

其中，$\varepsilon=y-x$ 为误差向量。

假设误差单维度分析，在基准点 $\delta>0$ 处展开 $|\varepsilon|^p$ 的泰勒级数：

$$|\varepsilon|^p\approx\sum_{k=0}^{\infty}\binom{p}{k}\varepsilon^k\delta^{p-k}$$

忽略高阶小项后，第 k 阶矩的权重近似为：

$$w_k(p)\propto\binom{p}{k}\delta^{p-k}$$

固定阶数 $k=4$，由斯特林公式展开组合数：

$$\ln\binom{p}{4}=\ln\Gamma(p+1)-\ln\Gamma(5)-\ln\Gamma(p-3)$$

其中，$\Gamma(x)$ 为伽玛函数。对 p 求导并寻找极值：

$$\frac{\mathrm{d}}{\mathrm{d}p}\ln\binom{p}{4}=\psi(p+1)-\psi(p-3)$$

这里，$\psi(x)=\dfrac{\mathrm{d}}{\mathrm{d}x}\ln\Gamma(x)$是双伽玛函数。当$p=4$时，

$$\psi(5)-\psi(1)=\left(\ln 5-\frac{1}{10}\right)-(-\gamma)\approx 1.609+0.577\approx 2.186>0$$

在p的值从1到8进行变化时，对应的FID值有较明显的变化，这也表明了超参数p对于实验的结果有较大的影响，而当$p=4$附近时导数为零，确认组合数在$p=4$时取得极大值，即$p=4$对四阶矩的惩罚权重最大，即此时损失函数最敏感于分布的四阶特征，生成分布的峰度、高阶协方差更接近真实数据，实验上表现为FID分数越低，GAN的训练具有更好的效果。当$p=4$时，判别器和生成器的损失函数可以写成：

$$\mathrm{Loss}_D=\frac{1}{4}E_{x\sim \mathrm{P}_{\mathrm{data}}}\left(D(x)-1\right)^4+\frac{1}{4}E_{z\sim P_z}\left(D\left(G(z)\right)\right)^4$$

$$\mathrm{Loss}_G=\frac{1}{4}E_{z\sim P_z}\left(D\left(G(z)\right)-1\right)^4$$

6. 小结

本文详细地分析了生成对抗网络模式坍塌与梯度消失的原因，从理论上分析了对抗网络的不稳定性，并将闵可夫斯基距离加入到GAN的损失函数，用于解决GAN训练不稳定的问题，提出了改进的损失函数。并通过进行对比实验，验证了本文提出的损失函数的有效性。模式坍塌本质问题是生成器与判别器对抗过程中的博弈失衡造成的，需从损失函数、模型结构、训练策略等多角度协同优化。但在复杂场景多模态数据生成中仍然需要进一步研究。

致　谢

该项研究得到智慧口岸关键技术与创新应用研究(云南省重大科技专项，202402AD080005)、云南省跨境贸易与金融区块链国际联合研发中心、云南省科技厅建设面向南亚东南亚科技创新中心专项(202203AP140010)资助，在此表示真诚的谢意。

参考文献

[1] GAN模型存在的问题分析(梯度消失、模式崩溃) [EB/OL]. 腾讯云开发社区. https://blog.csdn.net/Drifter_Galaxy/article/details/125123222, 2022-06-04.

[2] 雷娜, 顾险峰. 最优传输理论与计算[M]. 北京: 高等教育出版社, 2021: 393-394.

[3] 曹文明, 王浩. 深度学习理论与实践[M]. 北京: 清华大学出版社, 2024: 106-106.

[4] Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., *et al*. (2014) Generative Adversarial Networks. *Advances in Neural Information Processing Systems*, **3**, 2672-2680.