Advances in Applied Mathematics 应用数学进展, 2023, 12(3), 1367-1373
Published Online March 2023 in Hans. https://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2023.123139

Hans汉斯

基于梯度范数差值的一种正则化方法

吴天宝，徐　芳，张云轩*

西南石油大学理学院，四川　成都

收稿日期：2023年2月27日；录用日期：2023年3月24日；发布日期：2023年3月31日

摘　要

生成对抗网络(GANs)在学习从给定数据集指定的分布中采样方面非常成功，特别是给定数据集的数据量远大于其维度时。当数据有限时，经典的生成对抗网络生成的图像的质量会有显著降低，而输出正则化、数据增强、使用预训练模型和修剪等策略已被证明可以改善这种情况。然而这些方法常受限于特定的设置，例如预训练模型受限于数据的类型等。相比之下，本文提出的正则化方法通过优化鉴别器在真实图像与生成样本的梯度范数之间的差值来增强现有的生成对抗网络，并且具有很强的兼容性，适用于大多数现有的生成对抗网络。在数据有限的情况下显著的改善了训练成果。

关键词

生成对抗网络，图像生成，正则化，梯度范数

A Regularization Method Based on Gradient Norm Difference

Tianbao Wu, Fang Xu, Yunxuan Zhang*

School of Science, Southwest Petroleum University, Chengdu Sichuan

Received: Feb. 27th, 2023; accepted: Mar. 24th, 2023; published: Mar. 31st, 2023

Abstract

Generative adversarial networks (GANs) are very successful at learning to sample from a specified distribution of a given dataset, especially when the amount of data in a given dataset is much larger than its dimensions. Classical generative adversarial networks struggle when data is limited, while strategies such as output regularization, data augmentation, using pre-trained models, and pruning have been shown to bring improvements. However, these methods are often limited by

*通讯作者。

文章引用: 吴天宝, 徐芳, 张云轩. 基于梯度范数差值的一种正则化方法[J]. 应用数学进展, 2023, 12(3): 1367-1373.
DOI: 10.12677/aam.2023.123139

specific settings. For example, pre-trained models are limited by the type of data. In contrast, the regularization method proposed in this paper enhances the existing generative adversarial network by optimizing the difference between the discriminator between the real image and the gradient norm of the generated sample, and has strong compatibility applicable to most existing generative adversarial networks. Training outcomes were significantly improved when data were limited.

Keywords

Generate Adversarial Network, Image Generation, Regularization, Gradient Norm

1. 引言

生成式对抗网络(Generative adversarial network，简称 GANs)自 Ian Goodfellow [1]等人提出后，越来越受到学术界和工业界的重视。随着对抗生成网络的发展，其在图像与视频的生成[2] [3] [4]、图像翻译[5]、图像修复[6]等领域都取得了巨大的成功。这些成功引起了人们对 GANs 广泛应用的兴趣，从数据增强[7]和领域适应[8]到图像转换[9]和照片编辑[10]。GANs 的成功很大程度上依赖于大型数据集的可用性。

在实践中，常常遇到数据维度很高且数据量较少的情况。这种情况会致使 GANs 的性能显著降低，例如使用 StyleGAN 端对端生成图片时，使用 20%的 FFHQ (1024*1024 分辨率)数据集得到的 FID 值为 18.6，使用 10%的 FFHQ (1024*1024 分辨率)数据集时得到的 FID 值为 25.6 (FID 越小越好)，性能明显的下降。为了解决 GAN 性能下降的问题，最近提出了各种策略，包括使用预训练模型[11]、剪枝[12]和数据增强[13]。然而，尽管改善了结果，这些策略也都有限制。如果数据域保持相似，预训练模型的使用效果最好。剪枝需要进行多轮训练，以增加神经结构的稀疏性，然而这提高了训练成本。数据增强可以增强结果，但由于数据不足，其收益有限(见表 3)。正则化是一种廉价且潜在有效的方法，Tseng 等人[14]最近的工作采用了这种方法，控制判别器对真实图像的预测与生成图像之间的距离。

在本文中，我们研究了一种新的正则化方法来增强有限数据下的 GANs 训练。

2. 相关工作

生成对抗网络。已经提出了许多 GAN 变体来稳定训练并提高生成结果的感知质量。主要分为三个方向：1) 研究了不同的损失函数。2) 设计新的体系结构[15]。3) 各种归一化技术[16]。除此之外还设计了一些技术来产生更多样化的样本[17]并提高收敛性[18]。

GANs 的正则化，正则化技术被广泛用于稳定训练，其中最具有代表性的就是 WGAN。WGAN 中最小化了真实分布和生成分布的 Wasserstein 距离，鉴别器(Discriminator, D)的正则项为 1-Lipschitz，通过在真实数据和生成数据之间进行插值，惩罚鉴别器相对于输入数据点的梯度。Roth 等人[19]鼓励鉴别器在真实数据和生成数据上的梯度范数为零。除梯度范数外，约束鉴别器是另一种常用的机制[20]，权值惩罚也是 GANs 常用的正则化方法[21]。

数据不足导致 GAN 训练变得更具挑战性。已经提出了一些方法来提高用有限数据训练的 GANs 的

性能。较为常见的方法是使用数据增强，Jiang 等人[22]使用生成的数据作为对真实数据的“增强”，而其他人则在真实实例上进行增强。Chen 等人[23]利用修剪神经网络来提高性能。使用预训练模型也是一个不错的方法，使用与目标数据集相似度较高且数据量足够的数据集先进行训练，然后再训练目标数据集。本文的方法与这些方法的不同之处在于，主要考察梯度之间的范数差。设计的正则项主要是考虑真实数据在鉴别器上的梯度范数，与生成的数据在鉴别器上的梯度范数二者之间的差值(见图 2)。且该正则项与绝大多数 GANs 兼容可同时使用。

本文的主要贡献有 3 个方面：

1) 本文设计了一种新的正则项，该正则项具有很强的兼容性，适用于多种模型，本文的正则项几乎不增加计算成本。

2) 使用新的正则项，有效地改善了图像的质量。在有限的数据集上对于图像质量有明显的提升，在数据量足够的数据集上生成的图像包含更多的细节。本文的正则项几乎不增加计算成本。

3. 方法

3.1. 生成对抗网络简述

生成对抗网络(GANs)由一个生成器(Generator, G)和一个鉴别器(Discriminator, D)组成，它们相互竞争。生成器 $G(z;\theta)$ 由参数 θ 的调整使输入的一个简单的低维分布 $p(z)$ (例如高斯分布)向包含有高纬度数据的 χ 域学习，使二者最终形成一个复杂的映射。鉴别器 $D(x)$ 被训练区分真实数据 $x_R \sim \chi$ 与合成数据 $x_F = G(z;\theta)$。生成器与鉴别器之间博弈过程可以由两个损失函数表示：

$$\begin{aligned} L_G &= \mathrm{E}_{z\sim p(z)}\left[\iota_G\left(-D\left(G\left(z;\theta\right)\right)\right)\right], \\ L_D &= \mathrm{E}_{z\sim p_{data}(x)}\left[\iota_G\left(-D\left(x\right)\right)\right] + \mathrm{E}_{z\sim p(z)}\left[\iota_G\left(D\left(G\left(z;\theta\right)\right)\right)\right]. \end{aligned} \tag{1}$$

对于不同的生成对抗模型，使用的损失函数有所不同，例如 $\iota_G(t) = \iota_D(t) = \log(1+\exp(t))$ 或者 $\iota_G(t) = t, \iota_D(t) = \max(0, 1+t)$。

3.2. 问题提出

Karras 等人和 Tseng 等人在实验中发现数据量越少模型生成图像的质量越低，当数量低于一定的值时，模型无法收敛。具体来说，当使用 100%、20%、10%的 Flickr FaceHQ (FFHQ)数据集分别来训练 DCGAN，在使用 10%或者 20%的数据集时能够发现 FID 值(越低越好)会呈现急速的上升。

本文认为，数据量较少时模型的损失函数无法获取合理的梯度指导，导致函数始终在局部最优点处徘徊或者直接错过了最优点。通过实验发现在不同的数据使用比例下，使用真实数据与合成数据，它们在判别器上梯度的范数之间的差值存在明显的差异。具体来说，数据量越小时二者的梯度的范数差值越大(见图 1)。

3.3. 正则化方法

在数据量不足时，常见的解决方法有三种：数据扩充，模型架构改进，添加正则项。这三种方法中添加正则项相较于模型架构的改进，其优势在于可以节省更多的算力，因此添加正则项是成本更低的一种选择。

为了使模型在小样本下也能获取更加合理的梯度指导，能够获得更高质量的产出，本文提出一种新的正则化方法。梯度的范数差的数学表达式为：

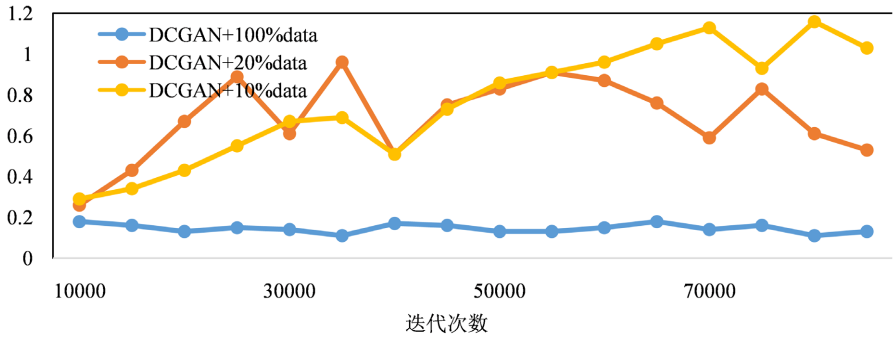

Figure 1. Gradient norm difference
图 1. 梯度范数差

$$R\left(D,x_F,x_R\right)=\left\|\frac{\partial D}{\partial x_R}\right\|_2-\left\|\frac{\partial D}{\partial x_F}\right\|_2 \tag{2}$$

其中 x_F 来自于生成器合成的数据即 $x_F=G\left(z;\theta\right)$， x_R 是来自于真实的数据。为了表述方便，我们称梯度的范数差为梯度差。

$$\begin{aligned}&\tilde{L}_D=L_D+\lambda R_t^3\\&R_t=\begin{cases}\left(\left\|\frac{\partial D}{\partial x_R}\right\|_2-\left\|\frac{\partial D}{\partial x_F}\right\|_2\right)_t & \text{if } t=1,2\\\frac{\sum_{t-2}^{t}\left(\left\|\frac{\partial D}{\partial x_R}\right\|_2-\left\|\frac{\partial D}{\partial x_F}\right\|_2\right)_t}{3} & \text{if } t>2\end{cases}\end{aligned} \tag{3}$$

在(3)式中 λR_t^3 为判别器的正则项，λ 是一个非负的超参数，数据量越少该参数大。t 为模型的迭代次数，考虑到梯度的突增或者突减而导致的误差，本文使用均值的手段来缓解误差所带来的影响。具体来说，就是考虑其前两次迭代的梯度值求和然后求其均值。$\partial D/\partial x_R$ 是真实图像的梯度值，$\partial D/\partial x_F$ 是来自生成图像的梯度值。具体流程图如图 2。

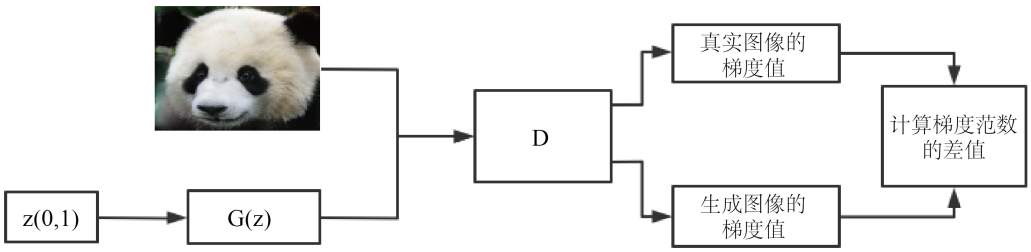

Figure 2. Gradient extraction flowchart
图 2. 梯度提取流程图

4. 实验

4.1. 实验环境

本文配置的环境 python3.7 + pytorch1.8.0 + cuda11，使用一张 RTX-TITANX 显卡(为了满足大多数模型的显存需求)。

数据集：在单一种类生成实验中，在 256 × 256 分辨率下，测试了动物脸狗和猫、100 镜头奥巴马、熊猫和脾气暴躁的猫。在 1024 × 1024 分辨率下，测试了 Flickr FaceHQ (FFHQ)、牛津花、WikiArt 的艺术画、Unsplash 的自然景观照片、Pokemon、动漫脸、头骨和贝壳。这些数据集旨在涵盖具有不同特征的图像：写实照片、图形插图和类似艺术的图像。在多样性实验中使用 CIFAR-100 数据集。

除开 CIFAR-100 数据集外，模型主要使用 256*256 与 1024*1024 这两个分辨率的数据集。其原因主要有一下几点 1)：首先这两种分辨率是常用的图片分辨率。2) 使用 256*256 作为先验数据集验证模型的改动是否有效，有效后在使用高分辨率的图片训练模型可以有效的减少计算量。3) 1024*1024 分辨率代表了图像生成的绝对质量，是对模型生成能力最大考验(图像超分辨率下除外)。

评估指标：1) 我们采用 Fréchet 起始距离(FID)来度量模型生成的图片的质量，FID 量化生成图像和真实图像的分布之间的距离。对少于 1000 张图像的数据集(大多数只有 100 张图像)，我们让 G 生成 5000 张图像，并计算合成图像和整个训练集之间的 FID (FID 值越低表示生成的图像质量越高)。2) 使用 IS (inception score)值，IS 值可以体现模型生成图像的质量的好坏与多样性的丰富程度(IS 值越高越好)。

4.2. 实验

在该实验中选取具有代表性的三个模型：1) 最先进的(SOTA)无条件模型 StyleGAN2，2) BIGGAN 模型。3) WGAN 模型。在选取的三个模型中使用正则项 λR_t^3 进行对比。所有的模型均迭代 100,000 次，批次大小为 16。

通过表 1 的对比试验，容易发现在添加了正则项的模型与未添加正则项的模型在数据量较少的部分提升是显著的。当数据量达到一定的规模时，提升的效果逐渐减弱，这是符合预期的结果。在数据量充足时，有足够多的数据支撑模型原本的损失函数去求其梯度值，得到的梯度值是较为稳定使得模型的能有较为稳定的输出。

Table 1. FID comparison of some datasets at 1024*1024 resolution
表 1. 1024*1024 分辨率部分数据集的 FID 比较

	Art Paintings			FFHQ			Nature Photograph		
图片数量	10% data	20% data	100% data	10% data	20% data	100% data	10% data	20% data	100% data
WAGN	80.6	61	59.3	36.8	29.6	23.2	78.3	69.2	60.1
WGAN + λR_t^3	42.3	41.3	39.9	19.1	18.3	16.4	49.3	45.1	41.2
StyleGAN	70.6	50.69	43.1	25.6	18.6	8.31	75.3	56.3	42.1
StyleGAN + λR_t^3	48.1	41.32	34.3	16.8	9.41	6.32	53.6	48.6	31.9
BIGGAN	75.6	68.3	50.6	34.1	29.6	15.6	78.6	59.1	47.1
BIGGAN + λR_t^3	59.1	53.8	30.1	28.3	25.1	9.4	62.1	51.1	32.1

在上述的实验中，对比的模型本身就是较为优秀的模型。为了进一步体现本文的正则项 λR_t^3 的效果，将使用更少的数据量(几百张)来进行消融实验(见表 2)。

由表 2 的结果可知，λR_t^3 的在数据量越少的情况下展现的效果是越明显的。值得注意的，在数据量较少的情况下对于 λ 的选择就比较重要了。例如，在表 3 的实验中取定 $\lambda = 100$。这样取值主要是本文认为一般的损失函数以及正则项不能很好的反应出梯度变化的趋势，更多的是保证模型的收敛以及稳定性。本文提出的正则项可以直观的体现出数据量带来的影响，那么在数据量较少的情况下提升其在损失函数中的占比是有助于模型更快更好的收敛的。

Table 2. FID comparison of a few sample data sets at 256*256 resolution
表 2. 256*256 分辨小数据集的 FID 比较

	Animal Face - Dog	Animal Face - Cat	Obama	Panda	Grumpy-cat
图片数量	389	160	100	100	100
WAGN	61.03	46.07	35.75	34.5	29.34
WGAN + λR_t^3	48.31	35.96	29.44	27.2	24.61
StyleGAN	60.23	45.06	47.14	40.03	26.65
StyleGAN + λR_t^3	44.32	33.21	37.26	35 .32	23.13
BIGGAN	48.32	34.82	39.26	30.12	25.82
BIGGAN + λR_t^3	36.15	29.13	31.79	21.32	17.32

CIFAR-100 数据更具挑战性，因为它包含 100 个类别，每个类别的图像更少。在表 3 中，由 IS 值作为主要的多样性指标，就结果来看增加正则项对于 IS 值是有正面影响的。

Table 3. FID value and IS value at 64*64 resolution
表 3. 64*64 分辨率下的 FID 值与 IS 值

图片数量	CIFAR-100					
	10% data		20% data		100% data	
	FID	IS	FID	IS	FID	IS
WAGN	49.61	5.98	35.12	7.56	23.41	8.12
WGAN + λR_t^3	28.44	6.51	23.41	8.12	17.31	8.71
BIGGAN	75.91	5.42	38.12	8.61	13.82	12.44
BIGGAN + λR_t^3	31.81	8.05	26.71	9.21	12.61	10.15

5. 结论

本文提出的正则化方法是一种基于梯度之间的范数差来设计的，该方法的优势在于它能匹配模型本身已经设计好的损失函数(正交)不需要修改原函数。其次，该正则化方法能够有效的缓解由数据量不足导致的模型性能的显著下降，并且该方法几乎不会增加计算的成本。不足之处在于，本文试图从理论的角度来分析该正则化方法，但是，由于对数据批量处理的执行方式，对所研究的损失进行理论分析极具挑战性，本文找不到一个具有严密逻辑性的数学解释。

参考文献

[1] Goodfellow, I., Pouget-Abadie, J., Mirza, M., *et al*. (2014) Generative Adversarial Nets. *NIPS*'14: *Proceedings of the 27th International Conference on Neural Information Processing Systems*, Volume 2, 2672-2680.

[2] Karras, T., Aittala, M., Laine, S., *et al*. (2021) Alias-Free Generative Adversarial Networks. *Advances in Neural Information Processing Systems* 34: *Annual Conference on Neural Information Processing Systems* 2021, *NeurIPS* 2021, 6-14 December 2021, 214-233.

[3] Wang, Z.W., She, Q. and Ward, T.E. (2021) Generative Adversarial Networks in Computer Vision: A Survey and Taxonomy. *ACM Computing Surveys* (*CSUR*), **54**, 1-38. https://doi.org/10.1145/3439723

[4] Wang, L., Ho, Y.-S. and Yoon, K.-J. (2019) Event-Based High Dynamic Range Image and Very High Frame Rate Video Generation Using Conditional Generative Adversarial Networks. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, 15-20 June 2019, 10081-10090.

https://doi.org/10.1109/CVPR.2019.01032

[5] Tov, O., Alaluf, Y., Nitzan, Y., Patashnik, O. and Cohen-Or, D. (2021) Designing an Encoder for StyleGAN Image Manipulation. *ACM Transactions on Graphics*, **40**, 1-14. https://doi.org/10.1145/3476576.3476706

[6] Liu, H.Y., Wan, Z.Y., Huang, W., Song, Y.B., Han, X.T. and Liao, J. (2021) Pd-GAN: Probabilistic Diverse GAN for Image in Painting. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, 20-25 June 2021, 9371-9381.

[7] Mao, X., Li, Q., Xie, H., *et al*. (2017) Least Squares Generative Adversarial Networks. *Proceedings of the IEEE International Conference on Computer Vision*, 2794-2802.

[8] Choi, J., Kim, T. and Kim, C. (2019) Self-Ensembling with GAN-Based Data Augmentation for Domain Adaptation in Semantic Segmentation. 2019 *IEEE/CVF International Conference on Computer Vision* (*ICCV*), Seoul, 27 October-2 November 2019, 6830-6840. https://doi.org/10.1145/3476576.3476706

[9] Zhu, J.-Y., Park, T., Isola, P. and Efros, A.A. (2017) Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. 2017 *IEEE International Conference on Computer Vision* (*ICCV*), Venice, 22-29 October 2017, 2242-2251. https://doi.org/10.1109/ICCV.2017.244

[10] Zhuang, P., Koyejo, O. and Schwing, A.G. (2021) Enjoy Your Editing: Controllable GANs for Image Editing via Latent Space Navigation. *International Conference on Learning Representations*, *ICLR* 2021, Vienna, 4 May 2021, 124-136. https://doi.org/10.1109/CVPR52688.2022.01039

[11] Kumari, N., Zhang, R., Shechtman, E. and Zhu, J.-Y. (2022) Ensembling Off-the-Shelf Models for GAN Training. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), New Orleans, 18-24 June 2022, 10641-10652.

[12] Chen, T., Cheng, Y., Gan, Z., Liu, J. and Wang, Z. (2021) Data-Efficient GAN Training beyond (Just) Augmentations: A Lottery Ticket Perspective. 35*th Conference on Neural Information Processing Systems* (*NeurIPS* 2021), 6-14 December 2021, 256-263.

[13] Zhang, H., Zhang, Z., Odena, A. and Lee, H. (2020) Consistency Regularization for Generative Adversarial Networks. *International Conference on Learning Representations*, Addis Ababa, 26-30 April 2020, 21655-21667.

[14] Tseng, H.-Y., Jiang, L., Liu, C., Yang, M.-H. and Yang, W. (2021) Regularizing Generative Adversarial Networks under Limited Data. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, 20-25 June 2021, 7921-7931. https://doi.org/10.1109/CVPR46437.2021.00783

[15] Karras, T., Laine, S. and Aila, T. (2019) A Style-Based Generator Architecture for Generative Adversarial Networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Long Beach, 15-20 June 2019, 4401-4410. https://doi.org/10.1109/CVPR.2019.00453

[16] Ba, J.L., Kiros, J.R. and Hinton, G.E. (2016) Layer Normalization. Preprint. arXiv: 1607.06450.

[17] Xiao, C., Zhong, P. and Zheng, C. (2018) Bourgan: Generative Networks with Metric Embeddings. *NeurIPS* 2018, Montreal, 3-8 December 2018, 136-152.

[18] Yazıcı, Y., Foo, C.-S., Winkler, S., Yap, K.-H., Piliouras, G. and Chandrasekhar, V. (2019) The Unusual Effectiveness of Averaging in GAN Training. 7*th International Conference on Learning Representations*, *ICLR* 2019, New Orleans, 6-9 May 2019, 25-31.

[19] Roth, K., Lucchi, A., Nowozin, S. and Hofmann, T. (2017) Stabilizing Training of Generative Adversarial Networks through Regularization. *Advances in Neural Information Processing Systems* 30: *Annual Conference on Neural Information Processing Systems* 2017, Long Beach, 4-9 December 2017, 30-36.

[20] Xie, T., Du, Y.M., Wang, T., *et al.* (2020) Structure Improved Conditional Spectral Normalization Generative Adversarial Networks for Image Recognition. 2020 *Chinese Automation Congress* (*CAC*), 842-846.

[21] Brock, A., Donahue, J. and Simonyan, K. (2018) Large Scale GAN Training for High Fidelity Natural Image Synthesis.

[22] Jiang, L., Dai, B., Wu, W. and Change Loy, C. (2021) Deceive d: Adaptive Pseudo Augmentation for GAN Training with Limited Data. *NeurIPS* 2021, 6-14 December 2021, 256-263.

[23] Graves, A., Fernández, S. and Schmidhuber, J. (2007) Multi-Dimensional Recurrent Neural Networks. *Artificial Neural Networks-ICANN* 2007: 17*th International Conference*, Porto, 9-13 September 2007, 549-558.