arXiv:2303.01559v2 [cs.CV] 8 Apr 2023

通过特征空间收缩改进 GAN 训练

Haozhe Liu[1†]，Wentian Zhang[2†]，Bing Li[1✉]，Haoqian Wu[3],Nanjun He[2]，
Yawen Huang[2]Yuexiang Li[2✉]，Bernard Ghanem[1],YefengZheng[21] 人工智能倡议，
阿卜杜拉国王科技大学（KAUST），[2]Jarvis 实验室，腾讯，[3]优图实验室，腾讯{*haozhe.liu;bing.li;bernard.ghanem*}*@kaust.edu.sa*；*zhangwentianml@gmail.com*；{
linuswu;yawenhuang;nanjunhe;vicyxli;yefengzheng}*@tencent.com*

摘要

由于具备卓越的数据生成能力，生成对抗网络（*GAN*）在无监督学习领域引起了广泛关注。然而，生成对抗网络的训练十分困难，因为对于判别器而言，训练分布是动态变化的，从而导致不稳定的图像表征。在本文中，我们从一个全新的视角研究生成对抗网络的训练问题，即，鲁棒图像分类。受鲁棒图像表征相关研究的启发，我们为生成对抗网络提出了一个简单而有效的模块，即 *AdaptiveMix*。该模块收缩判别器图像表征空间中的训练数据区域。鉴于直接约束特征空间难以实现，我们提出构造困难样本，并缩小困难样本与易样本之间的特征距离。困难样本通过混合一对训练图像构造而成。我们使用广泛采用且最先进的生成对抗网络架构评估 *AdaptiveMix* 的有效性。评估结果表明，*AdaptiveMix* 能够促进生成对抗网络的训练，并有效提升生成样本的图像质量。我们还表明，通过结合最先进的方法，*AdaptiveMix* 可以进一步应用于图像分类和分布外（*OOD*）检测任务。在七个公开数据集上开展的大量实验表明，我们的方法能够有效提升基线方法的性能。代码公开于*https://github.com/WentianZhang-ML/AdaptiveMix*。

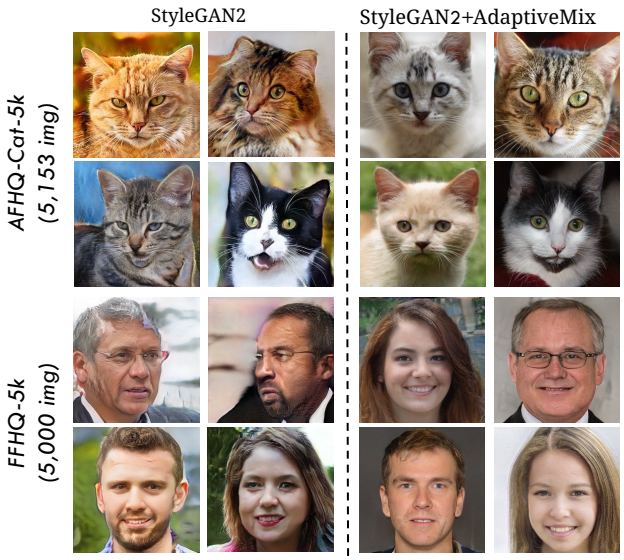


图 1. StyleGAN-V2 [20]和我们的方法（StyleGAN-V2 + AdaptiveMix）在 AFHQ-Cat 和 FFHQ-5k 上生成的结果。我们提出了一个简单而有效的模块 AdaptiveMix，可用于辅助无监督生成对抗网络的训练。在少量数据上训练时，由于训练不稳定，StyleGAN-V2 生成的图像会出现伪影。然而，AdaptiveMix 能有效提升 StyleGAN-V2 的图像质量。

1. 引言

人工好奇心 [40, 41] 和生成对抗网络（GAN）因其在图像生成方面的卓越表现而备受关注。标准的生成对抗网络由一个生成器和一个判别器网络组成，其中判别器经过训练以区分真实样本和生成样本，而生成器则旨在生成能够欺骗判别器的样本。然而，生成对抗网络的训练困难且不稳定，导致生成样本质量低下 [23, 34]。

人们为改进生成对抗网络的训练付出了诸多努力（例如。 [1, 5, 12, 25, 27, 34, 39]）。此前的研究[37] 尝试协同设计生成器和判别器的网络架构，以平衡迭代训练。

†共同贡献。通讯作者：Bing Li 和 Yuexiang Li。

沿着这一研究方向，PG-GAN [16] 按照课程学习范式，采用渐进式增长架构逐步训练生成对抗网络。最近，基于数据增强的方法（如 APA [15], ADA [17],以及向生成器中添加噪声 [20], ）也被提出，以进一步稳定生成对抗网络的训练。一些研究从判别器方面着手解决这一问题。例如，WGAN [1]提出通过权重裁剪施加Lipschitz约束。与之不同，WGAN-GP [6]直接对判别器梯度的范数进行惩罚。这些方法表明，改进判别器可以取得令人鼓舞的效果。然而，改进生成对抗网络的训练仍然是一个尚未解决且颇具挑战的问题。

本文考虑到判别器对生成对抗网络的训练至关重要，从一个全新的角度，即鲁棒图像分类，来解决生成对抗网络训练问题。具体而言，判别器可以被视为执行区分真实/虚假样本的分类任务。我们的见解是，受鲁棒图像分类研究的启发，通过控制判别器的图像表征（即特征提取器）可以改进生成对抗网络的训练 [28, 44]。更具体地说，近期关于鲁棒图像表征的研究 [44]提出了富有启发性的观察：训练数据在原始分类网络的学习空间中较为分散；因此，对于偏离训练数据底层流形的样本，网络会错误地赋予其较高的置信度。这一现象也导致了生成对抗网络的脆弱性，即判别器无法专注于学习真实数据分布。因此，我们提出在判别器的图像表征空间中收缩训练数据所在的区域。

不同于现有工作 [15, 17], ，我们探讨了一个关于生成对抗网络的问题：如果我们明确缩小判别器在图像表征空间中所支持的训练数据区域，生成对抗网络的训练稳定性是否会得到提升？为此，我们提出了一个名为 AdaptiveMix 的模块，用于缩小由特征提取器构建的潜在空间中的训练数据区域。然而，直接捕捉特征空间的边界并非易事，且颇具挑战性。对此，我们的见解是，可以通过缩小潜在空间中困难样本与易样本之间的距离来收缩特征空间；其中，困难样本指分类网络难以区分或分类的样本。为此，AdaptiveMix 将一对训练图像混合以构造困难样本，然后缩小这些混合图像与特征提取器表征的易训练样本之间的距离，从而收缩特征空间。我们在最先进的生成对抗网络架构上评估了 AdaptiveMix 的有效性，包括 DCGAN [37] 和 StyleGAN-V2 [20],。结果表明，所提出的 AdaptiveMix 有助于生成对抗网络的训练，并有效提升生成样本的图像质量。

除图像生成外，为其配备合适的分类器后，我们的 AdaptiveMix 还可应用于图像分类 [9,52] 和分布外（OOD）检测 [11, 14, 49] 任务。为了展示 AdaptiveMix 的应用方式，我们将其与近期分布外检测领域最先进工作 [51]中的正交分类器相结合。大量实验结果表明，我们的 AdaptiveMix 简单而有效，在多个数据集上的鲁棒图像分类和分布外任务中，均能持续提升 [51]的性能。

简而言之，本文的贡献可总结如下：

- 我们提出了一种名为 AdaptiveMix 的新模块，用于改进生成对抗网络的训练。AdaptiveMix 简单、有效且即插即用，有助于生成对抗网络生成高质量图像。

- 我们证明，通过收缩判别器所支持的图像表征空间中的训练数据区域，可以稳定、高效地训练生成对抗网络。

- 我们展示了 AdaptiveMix 不仅可以应用于图像生成，还可以应用于分布外和鲁棒图像分类任务。大量实验表明，在七个广泛使用的数据集上，对于四种不同的任务（例如，分布外），AdaptiveMix 均能稳定提升基线方法的性能。

2. 相关工作

插值方法综述。 基于正则化的方法能够在几乎不增加计算开销的情况下，同时提升泛化能力和鲁棒性，因此日益受到学界关注[7, 50, 53]。从某种程度上说，Mixup [53]是首个将样本插值策略引入卷积神经网络（CNN）模型正则化的研究。通过对成对样本进行线性插值生成的虚拟训练样本，能够平滑网络预测。沿着这一方向，研究者通过改变插值形式提出了许多变体[7, 21, 29, 44, 50]。本文重新审视基于插值的策略，并将混合样本视为用于收缩特征的生成困难样本，从而为混合操作的应用提供了新的视角。

生成对抗网络中的正则化。 近年来，各种方法 [1, 15, 34, 39] 尝试对判别器进行正则化，以稳定生成对抗网络的训练。以往的研究为判别器参数设计了多种惩罚项，例如权重裁剪 [1]和梯度惩罚 [6] 。由于这可能会降低判别器的能力，因此人们提出了谱归一化，通过权重归一化进一步稳定训练。然而，谱

归一化需要引入额外的结构，这可能会限制其在任意网络架构中的应用。作为一种更灵活的途径，近年来提出了对抗训练 [58],C-Reg [55], LC-Reg [43], Transform-Reg [35], ADA [17],和 APA [15] ，其作用方式与数据增强类似。这些方法容易受到增强数据强度的影响，因此必须使用自适应超参数 [15, 17]。本文尝试从一个新思路解决这一问题，即通过收缩特征空间来训练鲁棒的判别器。在不牺牲表征能力的前提下，所提出的方法可扩展至任意网络，并能轻松结合现有正则化方法。此外，该方法无需设置过多超参数，便能持续构造困难样本用于训练，因此也可用于各种其他任务，例如分布外检测 [24, 26, 32, 47, 51]和图像分类 [4, 22]。

3. 方法

本文研究如何改进生成对抗网络的训练。我们首先提出一种名为 AdaptiveMix 的新模块，用于收缩判别器图像表征空间中训练数据所占的区域。随后，我们证明 AdaptiveMix 能够促进利普希茨连续性，从而提升生成对抗网络的性能。最后，我们将 AdaptiveMix 与 [51] 中最先进的分布外方法所用的正交分类器结合，以展示如何使用我们的模块执行分布外检测和图像识别任务。

3.1. AdaptiveMix

我们的目标是通过控制判别器来改进生成对抗网络的训练，这可以从鲁棒图像分类的角度进行表述。不失一般性，设判别器由一个特征提取器 $\mathcal{F}(\cdot)$和一个分类器头 $\mathcal{J}(\cdot)$组成，其中，特征提取器用于从图像中提取特征，而分类器用于对提取的特征进行分类。我们的见解是，受鲁棒图像表征研究 [44, 59] 的启发，我们可以通过改进特征提取器 $\mathcal{F}$ 的表征来改进生成对抗网络的训练。正如 [44]，所观察到的，原始分类网络会将训练数据分散在其特征空间中，导致分类器不当地对偏离训练数据底层流形的样本赋予较高置信度。类似地，采用这种表征时，判别器难以学习真实数据分布
数据。因此，我们提出通过收缩由判别器的特征提取器在图像表征空间中支持的训练数据区域，来改进生成对抗网络的训练。

我们提出了一个名为 AdaptiveMix 的模块，用于收缩特征提取器 $\mathcal{F}$ 所表征空间中的训练数据区域。然而，直接捕捉特征空间中的训练数据区域是难以实现的。

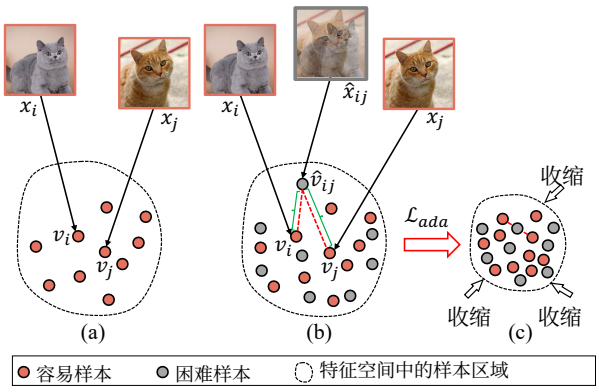


图2。AdaptiveMix 的示意图。(a) 某个类别的易样本x_i 和 x_j 通过特征提取器 $\mathcal{F}(\cdot)$映射到特征空间，其中 $v_i = \mathcal{F}(x_i)$。(b) 困难样本$\hat{x}_{ij}$ 由一对训练样本 x_i 和x_j线性组合生成，并映射到特征空间。(c) AdaptiveMix 通过缩小容易样本与困难样本之间的特征距离，收缩特征空间中的该类别区域。

给定某一类别 c的训练样本，我们的直觉是，可以通过减小特征空间中困难样本与易样本之间的距离，来收缩该类别在特征空间中的区域。其中，困难样本被视为网络难以分类的样本。换言之，我们认为，在由某一类别的所有训练样本形成的特征空间中，大多数困难样本比易样本更处于外围，这会导致决策边界扩大类别内距离，以覆盖困难样本。因此，对于类别 c，如果将困难样本向易样本拉近，则可以收缩类别 c 训练样本在特征空间中的区域。因此，所提出的 AdaptiveMix 包含两个步骤。首先，AdaptiveMix 从训练数据中生成困难样本。其次，我们的 AdaptiveMix 减小困难样本与易样本之间的距离。

难样本生成。寻找困难样本的一种朴素方法是使用经过训练的网络评估训练样本，将网络低置信度预测的样本视为困难样本。然而，这会带来新的问题。例如，这需要训练良好的网络，而这类网络并非总是可用。为此，受基于 Mixup 的图像增强方法 [7, 13, 53]所展现出的良好性能启发，我们提出了一种简单的困难样本生成方法。近期，人们提出了各种基于 Mixup 的方法，将多张图像混合成一张新图像。为简便起见，我们在此采用 Mixup 的原始版本 [53] 来生成困难样本。令 $\mathcal{X} = \{x_i\}_{i=1}^{N}$ 表示 N 训练图像，其中x_i 是第 i张训练图像。AdaptiveMix 按照 Mixup 的方式混合一对训练图像，生成一个困难样本：

$$\hat{x}_{ij} = g(x_i, x_j, \lambda) = \lambda x_i + (1-\lambda)x_j, \qquad (1)$$

其中，$g(\cdot,\ \cdot,\ \lambda)$ 是一个对 x_i 和 x_j 进行线性组合的函数，

λ 是从贝塔分布中采样得到的超参数 $\lambda \in \mathbb{B}(\alpha, \alpha)$。

如图 2所示，与原始训练样本 x_{ij} 和 x_i 相比，混合样本 x_j对网络来说更令人困惑，也更难以区分。在不影响泛化能力的前提下，我们将原始训练样本 x_i 称为易样本，而将 $\hat{x}_{ij}$ 称为困难样本。

AdaptiveMix 损失。我们的 AdaptiveMix 会在由特征提取器 $\mathcal{F}(\cdot)$ 表示的表征空间中，缩小困难样本 $\hat{x}_{ij}$ 与其对应的易样本x_i 和 x_j 之间的距离，从而收缩类别 x_i 或 x_j 所属训练数据的区域。需要注意的是，我们提出的是一种软损失，因为困难样本 $\hat{x}_{ij}$ 并不完全属于 x_i 或 x_j所属的类别。我们根据线性组合中 x_{ij} 里 x_i 所占的比例，在特征空间中缩小 $\hat{x}_{ij}$与 x_i 之间的距离：

$$\mathcal{L}_{ada}=\sum_i \sum_j \mathbb{D}_v(\lambda\mathcal{F}(x_i)+(1-\lambda)\mathcal{F}(x_j), \mathcal{F}(\hat{x}_{ij})+\sigma), \tag{2}$$

其中 σ 是从高斯分布中采样的噪声项，用于防止过拟合；$\mathbb{D}_v(\cdot, \cdot)$ 表示用于评估距离的度量，例如 L1 范数、L2 范数。请注意，我们的 AdaptiveMix 不需要训练图像的标签；然而，它能够在特征空间中收缩每个类别的训练数据区域（见图 2），因为易样本 x_i及其对应的困难样本 $\hat{x}_{ij}$ 属于同一类别。

3.2. 与利普希茨连续性的联系

为了进一步探究我们方法的优越性，我们从理论上分析所提出的 AdaptiveMix 与利普希茨连续性之间的关系。

预备知识。 在所提出的方法中，特征提取器$\mathcal{F}(\cdot)$连接输入空间 $\mathcal{X}$ 与嵌入空间$\mathcal{V}$。给定分别定义在 $\mathcal{X}$ 和 $\mathcal{V}$ 上的两个评估度量 $\mathbb{D}_x(\cdot, \cdot)$ 和 $\mathbb{D}_v(\cdot, \cdot)$，如果存在实常数 K，使得所有 $x_i, x_j \in \mathcal{X}$都满足以下条件，则 $\mathcal{F}(\cdot)$ 满足利普希茨连续性：

$$K\mathbb{D}_x(x_i, x_j) \geq \mathbb{D}_v(\mathcal{F}(x_i), \mathcal{F}(x_j)). \tag{3}$$

命题。 基于 [3, 44], <style id='6}>的分析，平坦的嵌入空间，尤其是具有利普希茨连续性的嵌入空间，是应对不稳定训练和对抗攻击的理想解决方案。因此，通过证明 AdaptiveMix 与K-利普希茨连续性之间的等价性，可以论证所提出方法的有效性。

定理。 对于利普希茨连续性的任意 K，AdaptiveMix 在 L1 范数度量空间下是一种近似解。

证明。 给定从 $\mathcal{X}$中采样的 x_i 和 x_j，可通过 $g(x_i, x_j, \lambda)$ 得到基于线性组合的枢轴点 $\hat{x}_{ij}$。

由于 $\hat{x}_{ij}$ 可视为 $\mathcal{X}$中的一个样本，因此我们可以将利普希茨连续性（公式（3））转换为

$$\begin{aligned} K\mathcal{B} =& K(\mathbb{D}_x(\hat{x}_{ij}, \lambda x_i) + \mathbb{D}_x(\hat{x}_{ij}, (1-\lambda)x_j)) \\ \geq& \mathbb{D}_v(\hat{v}_{ij}, \lambda v_i) + \mathbb{D}_v(\hat{v}_{ij}, (1-\lambda)\lambda v_j) \\ \geq& \mathbb{D}_v(\hat{v}_{ij}, g(v_i, v_j, \lambda)) \end{aligned} \tag{4}$$

其中 $\mathbb{E}[\lambda] = 0.5$ 和 $\mathbb{E}[x_i] = \mathbb{E}[x_j]$。因此，可以通过小批量训练将上界 $\mathcal{B}$ 估计为 0。由于 $\mathbb{D}_v(\cdot, \cdot)$ 是 L1 范数距离，因此 $\mathbb{D}_v(\cdot, \cdot)$ 应不小于 0。因此，在利普希茨连续性下，我们可以得到 $\mathbb{D}_v(\hat{v}_{ij}, g(v_i, v_j, \lambda))$ 的下界和上界：

$$\mathbb{E}[0] \leq \mathbb{E}[\mathbb{D}_v(\hat{v}_{ij}, g(v_i, v_j, \lambda)))] \leq \mathbb{E}[K\mathcal{B}] = 0. \tag{5}$$

因此，如果 $\mathcal{F}$ 满足利普希茨连续性，则 $\mathbb{D}_v(\hat{v}_{ij}, g($ $v_i, v_j,${v13))) 应为零，而 AdaptiveMix 的最优结果与 $\mathbb{D}_v(\hat{v}_{ij}, \Lambda_\lambda(v_i, v_j))$ 相同。因此，通过最小化 AdaptiveMix，可以确保 K-利普希茨连续性。

直观解释。 这一理论结果也与直觉一致。我们的总体思路是缩小特征空间，以获得鲁棒表征。由于利普希茨连续性要求嵌入空间中的距离小于图像空间中的距离，因此，缩小特征空间应当是近似确保利普希茨连续性的一种合理方式。

3.3. 基于AdaptiveMix的图像生成

基于前述分析，所提出的模块有助于稳定生成对抗网络的训练。本文展示了如何将 AdapativeMix 与两种最先进的图像生成方法结合，从而将其应用于图像生成：WGAN [1], 和 StyleGAN-V2 [20]。我们主要在正文中详细阐述与 WGAN 的结合方式，而 StyleGAN-V2 的相关内容则见补充材料。得益于 AdaptiveMix 即插即用的特性，我们可以轻松地为 WGAN 配备 AdaptiveMix。具体而言，我们将 AdaptiveMix 应用于 WGAN 的判别器，该判别器由特征提取器 $\mathcal{F}(\cdot)$ 和分类头 $\mathcal{J}(\cdot)$ 组成。随后，我们重写学习目标，将 AdaptiveMix 纳入 WGAN 的训练中：

$$\begin{aligned} \min_G \max_{\mathcal{F}, \mathcal{J}} & \mathbb{E}_{x\sim p_r}[\mathcal{J}(\mathcal{F}(x))] - \mathbb{E}_{z\sim p_z}[\mathcal{J}(\mathcal{F}(G(z)))] \\ & + \min_{G, \mathcal{F}} \mathbb{E}_{x\sim p_r, p_g}[\mathcal{L}_{ada}] \end{aligned} \tag{6}$$

其中，z 是生成器 $G(\cdot)$的噪声输入；L2 范数被用作衡量 $\mathcal{L}_{ada}$的指标，$\mathcal{J}(\cdot)$的输出是一个标量，用于估计给定样本的真实性。为简化 $\mathcal{J}(\cdot)$的结构，我们直接采用平均算子作为 $\mathcal{J}(\cdot)$。在本文中，AdaptiveMix通过将真实样本与生成器生成的虚假样本进行线性组合来生成困难样本。这种混合是一种特征平滑方式，能够促使决策边界

判别器变得平滑，从而提高生成对抗网络训练的稳定性。伪代码见补充材料。请注意，传统的基于混合的方法**不**适用于各种Wasserstein生成对抗网络，因为WGAN进行的是动态极小极大博弈，其判别器的输出范围为$(-\infty, +\infty)$，而我们的方法能够改善WGAN的训练。

3.4. 基于 AdaptiveMix 的图像分类

除了图像生成之外，所提出的 AdaptiveMix 还可应用于图像分类。下面说明如何将 AdativeMix 应用于这一任务。与图像生成任务不同，图像分类要求分类模型的特征提取器 $\mathcal{F}(\cdot)$ 提取的特征尽可能具有判别性。我们的 AdaptiveMix 会收缩特征空间中训练数据所在的区域；如果单独应用 AdaptiveMix，特征在一定程度上会变得平滑。不过，采用能够确保不同类别的特征彼此可分的适当分类器，就能轻松解决这个问题。

受图像分类方法的启发 [51]，，我们采用~一个正交分类器 $\mathcal{J}(\cdot)$ 来确保特征空间中的类别感知分离，其中正交分类器~$\mathcal{J}(\cdot)$ 由若干个权重向量 $w_k \in \mathcal{W}$组成，且 w_k对应于 k第 - 个类别。具体而言，我们将基于 CNN 的分类~模型的最后一个全连接层替换为正交分类器 $\mathcal{J}(\cdot)$。因此，给定 x_i，第 y^k 个 k类别的预测分数计算如下：

$$y_i^k = \frac{w_k^T v_i}{||w_k||\,||v_i||}. \tag{7}$$

其中，$v_i = \mathcal{F}(x_i)$表示分类模型的特征提取器 $\mathcal{F}(\cdot)$提取的 x_i的特征。通过 Softmax 层计算 x_i 属于第 k类的概率p_i^k：

$$p_i^k = \frac{exp(y_i^k)}{\sum\limits_{1\le l\le n} exp(y_i^l)} \tag{8}$$

其中，集合 $\mathcal{P}_i = \tilde{\mathcal{J}}(v_i) = \{p_i^k | 1 \le k \le n\}$ 构成了基于CNN的模型在 n类识别任务中的最终输出。

通过移除最后一层中的偏置和激活函数，分类模型将 x 映射到允许的范数球空间，从而确保不同类别对应的特征可以分离。为进一步增强类别感知分离，我们引入正交约束来初始化 $\mathcal{W}$，其定义如下：

$$\prod_{w_k, w_l \in \mathcal{W}, k\neq l} w_k^T w_l = 0. \tag{9}$$

此外，除了 AdaptiveMix 损失外，我们还可以在图像

表1. 使用我们的 AdaptiveMix 所带来的改进汇总，其中**提升幅度** 指我们相较于基线方法的改进。我们的方法 AdaptiveMix 在七个广泛使用的数据集上的四项任务中，提升了六种基线方法的性能。详细的对比结果见表#列中指定的表格。

	数据集	表 #	Gain
图像生成	C-10 [22]	表 4	-20.0% 弗雷歇起始距离 ↓
	CelebA [30]	表4	-54.0% 弗雷歇起始距离 ↓
	FFHQ [19]	表5	-4.9% 弗雷歇起始距离 ↓
	AFHQ-CAT [2]	表5	-43.5% 弗雷歇起始距离 ↓
	FFHQ-5k [19]	表。6	-47.9% 弗雷歇起始距离 ↓
图像分类	C-10 [22]	表 11	**+0.7% 准确率** ↑
	C-100 [22]	表 11	**+1.5% 准确率** ↑
	T-ImageNet [4]	表 11	**+5.87% 准确率** ↑
	ImageNet [38]	表 11	**+1.9% 准确率** ↑
鲁棒分类	C-10 [22]	表 7	**+4.6×准确率** ↑
	C-100 [22]	表 8	**+5.2×准确率** ↑
	T-ImageNet [4]	表 8	**+1.1 × 准确率** ↑
OOD 检测	基准测试 [51]	表 12	**+3.5% F1** ↑

增强后的分类学习目标中使用基于混合的交叉熵损失 [53]，，因为我们使用 Mixup 生成困难样本（有关分类的更多详细信息，请参阅补充材料）。

3.5. 基于 AdaptiveMix 的分布外检测

AdaptiveMix 可以轻松集成到 [51] 的最先进分布外检测模型中。给定所有训练样本 $x \in \mathcal{X}$ 作为输入，我们可以通过训练好的 $\mathcal{F}(\cdot)$获得相应的表征 $v \in \mathcal{V}$。然后，通过计算 $\{\mathcal{F}(x_i)|x_i \in \mathcal{X} \arg\max \tilde{y}_i = k\}$ 的第一个奇异向量，可以得到第 k 类的代表性表征 v_k^*。请注意，v_k^* 是通过奇异值分解而非 $\mathcal{F}(x_i)$ 计算得出的。给定一个测试样本 x_t，x_t 属于分布外的概率ϕ_t 计算如下：

$$\phi_t = \min_k \arccos\left(\frac{|\mathcal{F}^T(x_t)v_k^*|}{||\mathcal{F}(x_t)||}\right), \tag{10}$$

当 ϕ_t 大于预定义阈值 ϕ*时，x_t 被归类为分布外样本。

4. 实验

为了评估我们方法的性能，我们在图像生成、图像识别、鲁棒图像分类和分布外检测等多种任务上进行了广泛实验。表 1显示，所提出的方法显著提升了这些任务上的基线方法。下面，我们首先评估所提出方法在图像生成上的性能，然后测试该方法在图像分类和分布外检测等视觉识别任务上的表现。请注意，实验中使用的数据集、其他实验以及实现细节详见补充材料。

表 2 10 。
[22] ±
所提出方法在 CIFAR 上的消融研究。平均FID标准差表示模型分别训练100、400和800轮时FID分数的均值和标准差。最小FID是训练期间的最佳结果。收敛曲线见补充材料

	平均FID ±标准差 ↓	最小FID ↓
基线	106.81 ± 21.79	55.96
基线 + AdaptiveMix	**39.52 ± 7.85**	**30.85**

表3. 在 FFHQ- 5k [19] 上使用 StyleGAN-V2 [20]评估利普希茨连续性。该指标表示给定样本对的嵌入距离图像距离 的平均值，即$\mathbb{D}_v(\mathcal{F}(x_i)\ \mathcal{F}(^{x_j}))\ \mathbb{D}_x(x_i,^{x_j})$ 。数值越小越好。计算详情见补充材料。

	真实样本	生成样本	Both
StyleGAN-V2	3.167	0.263	0.735
StyleGAN-V2 +我们的方法	**1.391**	**0.166**	**0.291**

4.1. 图像生成性能

消融研究。 为了量化 AdaptiveMix 的贡献，我们测试了使用和不使用 AdaptiveMix 时的图像生成性能。如表2所示，所提出的方法在所有情况下都显著优于基线 [1] 。这一提升表明，所提出的方法增强了生成对抗网络的训练。

除了定量分析外，我们还进行了可视化，以展示所提出方法的有效性。如图 3所示，我们使用一个玩具数据集训练模型，其输入是遵循图 3中分布的大量二维点。可以看出，与另外两种基线方法相比，AdaptiveMix 能更好地模拟这种分布。为进一步研究，我们在图 3的第二行中可视化了判别器在各个位置的输出。如果判别器像 Std-GAN [5], 那样增大真实样本与生成样本之间的距离，生成器就难以从判别器获得有效指导，从而导致生成结果不佳或模式崩溃。相比之下，所提出的方法会缩小输入样本之间的距离，因此置信度分数图趋于平坦，从而能够为生成器提供更有利的梯度，带来更好的生成性能。我们还在实际场景中对这一现象进行了量化。如表 3所示，我们计算了特征空间与图像空间中距离的平均比值。采用 AdaptiveMix 后，所有场景中的相应比值均达到最小，可视为利普希茨连续性的保证。

与现有方法比较。 为进一步展示所提出方法的优越性，我们在玩具数据集、CIFAR10 [22], 和 CelebA[31]上，将所提出方法的表现与知名损失函数进行比较。如表 4所示，所提出的方法在两个数据集上的表现均大幅优于其他方法。与近期方法 Realness GAN 相比，

表4. DCGAN 的 FID分数 [37] 在 CelebA 上采用各种学习目标时 [31] 以及 CIFAR10 [22]。

学习目标	CIFAR-10	CelebA
WGAN [1] (ICML'17)	55.96	-
HingeGAN [57] (ICLR'17)	42.40	25.57
LSGAN [33] (ICCV'17)	42.01	30.76
DCGAN [37] (ICLR'16)	38.56	27.02
WGAN-GP [6] (NIPS'17)	41.86	70.28
重新实现的 WGAN-GP	38.63	70.16
Realness GAN-目标函数1 [46] (ICLR 2020)	36.73	-
Realness GAN-目标函数2 [46] (ICLR 2020)	34.59	23.51
Realness GAN-目标函数3 [46] (ICLR 2020)	36.21	-
AdaptiveMix（我们的方法）	**30.85**	**12.43**

表5. 所提出方法在 AFHQ [2] 和 FFHQ [19] 上用于 StyleGAN-V2 [20] 的弗雷歇起始距离和初始分数，与其他最先进的生成对抗网络训练解决方案的比较

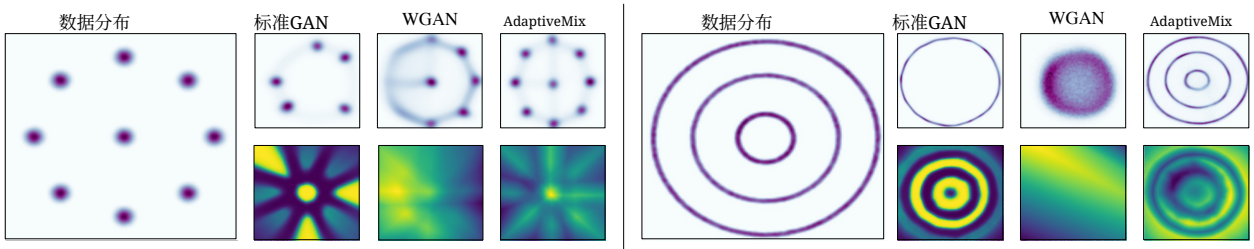
方法	AFHQ-Cat-5k		FFHQ（完整集）	
	FID	IS	FID	IS
StyleGAN-V2 [20]（CVPR 2020）	7.737	1.825	3.862	5.243
StyleGAN-V2（重新实现）	7.924	1.890	3.810	5.185
LC-Reg [43]（CVPR'21）	6.699	1.943	3.933	**5.312**
Style 生成对抗网络-V2 + 我们的方法	**4.477**	**1.972**	**3.623**	5.222
ADA [17]（NIPS 2020）	6.053	2.119	4.018	5.329
ADA（重新实现）	5.582	2.059	3.713	5.200
ADA + 我们的方法	**4.680**	**2.069**	**3.681**	**5.335**
APA [15] (NIPS 2021)	4.876	2.156	3.678	5.336
APA（重新实现）	4.645	2.093	3.752	5.281
APA+我们的方法	**4.148**	**2.096**	**3.609**	**5.296**

表6. 我们的方法与此前在 FFHQ- 5 [19] 2 k 上对生成对抗网络进行 [20] 正则化的方法在弗雷歇起始距离和初始分数上的比较。StyleGAN-V 以基线为基准。

正则化	FID	IS
基线 [20] (CVPR 2020)	37.830	4.018
基线（重新实现）	36.053	4.097
实例噪声 [42] (ICLR'17)	40.981	4.231
单边最小二乘 [39] (NIPS'16)	33.978	4.029
LC-Reg [43] (CVPR'21)	35.148	3.926
Ours	**18.769**	**4.332**
APA [15] (NIPS 2021)	13.249	4.487
APA（重新实现）	14.368	4.855
APA+我们的方法	**11.498**	**4.866**

所提出的方法在 CIFAR10 上使 FID 提升了 10.81%。同样，在 CelebA 上，Realness GAN 和所提出方法的 FID 分数分别为 23.51 和 12.43，这有力地展示了 AdaptiveMix 的优势。请注意，表 4中的结果取自 [46] ，其采用相同的架构，即 DC-GAN。相应的可视化结果见补充材料。

为了全面论证 AdaptiveMix，我们还将所提出的方法与近期针对生成对抗网络的正则化方法进行比较。表 5 展示了所提出的方法



(a) 混合高斯分布　　(b) 混合圆线

图3. 合成数据集上的实验结果：将来自 (a) 九个高斯分布和 (b) 三个圆的二维点用作生成对抗网络的训练数据。每种情况下，从左到右依次为 Std-GAN、WGAN 和 AdaptiveMix 生成的样本。第一行为生成结果，第二行为判别器对应的置信度图。

能够帮助生成对抗网络在不同数据集上收敛，并取得显著成果。作为一种即插即用模块，所提出的方法还可以与最先进的基于增强的方法 ADA [17] 和 APA[15], 结合，进一步提升生成对抗网络的生成性能（弗雷歇起始距离平均提升 13.5%）。最后，我们使用有限的训练数据评估所提出方法的性能。如表 6所示，仅给定 5k 个样本时，所提出的方法便能将基线的弗雷歇起始距离从 37.830 显著降低至 18.769。结合 APA 后，AdaptiveMix 可取得最佳的弗雷歇起始距离（11.498）和初始分数（4.866）。

4.2. 视觉识别性能

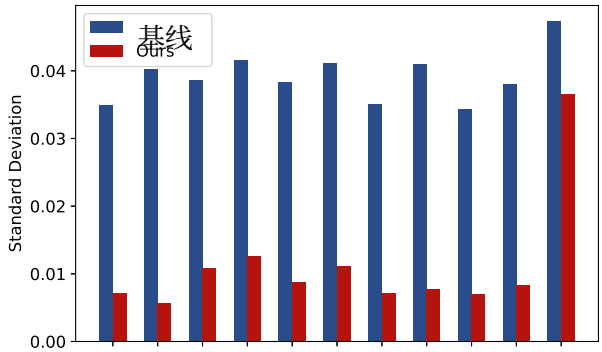


图4. CIFAR-10 上嵌入簇的紧凑度（即标准差）。标准差根据同一标注下的嵌入编码计算。“总计”表示测试集中所有嵌入编码的整体紧凑度。

消融研究。 为验证我们的方法能够缩小特征空间中样本的区域，我们还分析了 CIFAR-10 上特征空间中每个类别的聚类紧致度（即，聚类的标准差），结果如图 4 所示。可以观察到，我们的 AdaptiveMix 的类别级标准差远低于

表7. 使用各种方法训练的 WRN-28-10 搭配正交分类器（正交）在 CIFAR-10 上的准确率（%）。

CIFAR10	FGSM (8/255)	PGD-8 (4/255)	PGD-16 (4/255)	CW-100 (c=0.01)	CW-100 (c=0.05)
基线	38.03	0.92	0.28	11.1	0.39
Mixup [53]	60.17	3.97	1.16	30.32	2.36
正交 +Mixup	44.80	3.99	2.66	71.12	49.47
M.-Mixup [44]	59.32	7.97	2.97	51.47	11.12
正交 +M.-Mixup	38.76	5.77	4.38	69.08	53.98
Ours	**74.18**	**32.12**	**22.12**	**81.39**	**74.72**

表8. 基于 WRN-28- 的各种对抗攻击在 CIFAR-100 和 Tiny-ImageNet 上的准确率（%）10 [52] 以及 PreActResNet-18 [9]分别。

数据集	方法	FGSM (8/255)	PGD-8 (4/255)	PGD-16 (4/255)	CW-100 (c=0.01)	CW-100 (c=0.05)
C-100	基线	11.71	0.79	0.42	4.42	0.23
	Mixup [53]	27.34	0.28	0.11	4.83	0.28
	M.-Mixup [44]	**29.73**	1.19	0.49	10.75	0.77
	Ours	24.28	**8.22**	**7.40**	**42.02**	**26.18**
T-ImageNet	基线	4.26	0.81	0.60	27.92	7.52
	Mixup [53]	4.23	0.98	0.77	29.13	15.41
	M.-Mixup [44]	3.04	0.82	0.59	29.69	16.86
	Ours	**7.10**	**4.66**	**4.98**	**35.93**	**34.22**

基线方法的相应值。“总计”一项衡量特征空间中所有样本区域的紧致度。图4 表明，与基线方法（即，不使用 AdaptiveMix）相比，我们的方法缩小了特征空间中样本的区域。

表9. 在不同噪声项 σ 下，使用所提出的方法训练的 WRN-28-10 在 CIFAR-10 上的准确率（%）。

噪声	FGSM (8/255)	PGD-8 (4/255)	PGD-16 (4/255)	CW-100 (c=0.01)	CW-100 (c=0.05)
σ=0.1	71.90	32.54	**23.31**	79.58	71.96
σ=0.01	71.56	34.04	25.96	80.68	72.00
σ=0.005	71.31	27.79	20.64	80.46	71.58
σ=0.001	70.51	27.42	17.98	79.47	67.00
σ=0.05	**74.18**	**32.12**	22.12	**81.39**	**74.72**

鲁棒图像识别。 为评估所提方法的对抗鲁棒性，我们将其与 Mixup 进行比较

表10. 使用所提出的方法训练的 WRN-28-10 在 CIFAR-10 上的准确率（%）；该方法使用贝塔分布中的多种 $alpha$ 来生成混合系数 λ。

阿尔法	FGSM (8/255)	PGD-8 (4/255)	PGD-16 (4/255)	CW-100 (c=0.01)	CW-100 (c=0.05)
α=2.0	71.27	31.85	**22.39**	79.18	70.50
α=1.0	**74.18**	**32.12**	22.12	**81.39**	**74.72**

表11. 所提出的 AdaptiveMix 在不同基线方法和数据集上的准确率（%）。Res. 表示输入分辨率。

数据集	架构	Res.	基线	Ours
CIFAR-10	WRN-28-10 [52]	32^2	96.11	**96.80**
CIFAR-100	WRN-28-10 [52]	32^2	80.82	**82.02**
T-ImageNet	PreActResNet-18 [9]	64^2	57.23	**60.59**
ImageNet	ResNet-50 [8]	128^2	67.38	**68.69**

[53] 和流形混合 [44]（记为 M.-Mixup），并在表 7 和表 8中给出结果。如表 7所示，所提方法的平均分类准确率可达 56。91% 在 CIFAR-10 上，分别比 Mixup 和流形混合高出 37。31% 和30。34%。此外，我们还在大规模数据集 即 Tiny-ImageNet 上测试了所提模型。如表 8所示，与其他基于插值的方法相比，所提方法在不同类型的对抗攻击下均取得了更优的性能。具体而言，使用我们的 AdaptiveMix 后，模型的平均准确率可达 17。38%，比流形混合提高了 ∼7%。

在此，我们分析了不同超参数取值的影响，包括 σ 噪声项的取值，以及 αAdaptiveMix 中贝塔分布的取值。如表9所示，网格搜索考虑了多个噪声水平 σ=[0.1、0.05、0.01、0.005、0.001]。当 σ=0.05 时，性能最佳。对于 α，我们设置了两种情况进行比较，包括表 α=110 中的 .0 和 2.010。观察发现，使用 α=1.0 训练的模型优于使用 α=2.0 训练的模型（即56.91% 对比55.04%）。

干净图像识别。为验证所提出方法在图像识别中的有效性，我们在多个标准数据集上测试该方法，并将结果与基线 [8, 9, 52]进行比较。表11显示，我们的方法在所有情况下都提升了基线。尤其是在 Tiny-ImageNet 上，所提出的方法实现了 3% 的绝对提升。实验结果表明，AdaptiveMix 不仅提升了鲁棒性，也有助于泛化。

分布外检测。为了验证我们的方法在分布外检测中的有效性，我们在多个数据集上将其与最先进的分布外检测方法进行比较 [32, 36, 48, 51] 。我们将 1DS 的准确率视为其他方法的上限，如表 [51] 12所示，因为 1DS 采用蒙特卡洛（MC）采样，以牺牲计算效率为代价来获得较高的准确率。因此——

表 12。各类分布外集合上的分布外检测，其中，TIN-C、TIN-R、LSUN-C 和 LSUN-R 分别指 Tiny ImageNet-Crop、Tiny ImageNet-Resize、LSUN-Crop 和 LSUN-Resize 的域外集合。所有数值均为 F1分数（↑），† 表示通过开源代码复现的结果。

域内数据集	CIFAR10			
域外数据集	TIN-C	TIN-R	LSUN-C	LSUN-R
使用 MC 采样的方法				
1DS [51] (CVPR'21)	0.930	0.936	0.962	0.961
采用分布外样本进行验证和微调的方法				
ODIN [26] (ICLR'18)	0.902	0.926	0.894	0.937
马氏距离 [24](NIPS'18)	0.985	0.969	0.985	0.975
软预测 [10] (ICLR'17)	0.803	0.807	0.794	0.815
反事实 [36](ECCV'18)	0.636	0.635	0.650	0.648
CROSR [48](CVPR'19)	0.733	0.763	0.714	0.731
OLTR [32] (CVPR'19)	0.860	0.852	0.877	0.877
1DS 无 MC † [51]	0.890	0.886	0.897	0.907
1DS 无 MC † +我们的方法	**0.922**	(	**0.934**	**0.937**

，1DS 的耗时远高于其他方法。我们通过移除 1DS 的 MC 采样，构建了一个名为"1DS w/o MC"的基线 [51]，而我们的方法则将"1DS w/o MC"与 AdaptiveMix 相结合。表12 显示，由于缺少 MC 采样，"1DS w/o MC"的性能有所下降。不过，我们的 AdaptiveMix 无需付出高昂的计算成本，就能有效提升"1DS w/o MC"的性能。

5. 结论

本文提出了一种名为 AdaptiveMix 的新模块，该模块简单却能有效改进生成对抗网络的训练。通过动态缩小训练样本与其线性组合之间的距离，AdaptiveMix 可以收缩特征空间中的训练数据区域，从而实现生成对抗网络的稳定训练，并提升生成样本的图像质量。我们还证明，AdaptiveMix 是一种合理的方法，可确保对利普希茨连续性的近似估计。除了图像生成，我们还展示了 AdaptiveMix 凭借其即插即用的特性，也可应用于图像分类和分布外检测等其他任务。实验结果表明，在多种任务上，我们的方法能够有效提升基线模型在七个公开数据集上的性能。

致谢

本研究获得阿卜杜拉国王科技大学（KAUST）赞助研究办公室通过视觉计算中心（VCC）提供的资助、广东省重点领域研发计划、中国（编号：2018B010111001）、国家重点研发计划（2018YFC2000702）以及科技创新2030——"新一代人工智能"项目（编号：2020AAA0104100）的支持。

参考文献

[1] Martin Arjovsky、Soumith Chintala 和 L´eon Bottou。Wasserstein生成对抗网络。载于《国际机器学习会议论文集》，第214–223页。PMLR，2017年。1、2、4、6[2] Yunjey Choi、Youngjung Uh、Jaejun Yoo 和 Jung-Woo Ha。StarGAN v2：面向多个域的多样化图像合成。载于《IEEE/CVF计算机视觉与模式识别会议论文集》，第8188–8197页，2020年。5、6[3] Moustapha Cisse、Piotr Bojanowski、Edouard Grave、Yann Dauphin 和 Nicolas Usunier。Parseval网络：提升对抗样本的鲁棒性。载于《国际机器学习会议论文集》，第854–863页。PMLR，2017年。4[4] Jia Deng、Wei Dong、Richard Socher、Li-Jia Li、Kai Li 和 Li Fei-Fei。ImageNet：大规模分层图像数据库。载于《IEEE计算机视觉与模式识别会议论文集》，第248–255页。IEEE，2009年。3、5[5] Ian Goodfellow、Jean Pouget-Abadie、Mehdi Mirza、Bing Xu、David Warde-Farley、Sherjil Ozair、Aaron Courville 和 Yoshua Bengio。生成对抗网络。《神经信息处理系统进展》，第27卷，第2680页，2014年，第2672页。1、6：–。[6] Ishaan Gulrajani、Faruk Ahmed、Martin Arjovsky、Vincent Dumoulin 和 Aaron Courville。改进Wasserstein生成对抗网络的训练。arXiv预印本 arXiv:1704.00028，2017年。2、6[7] Hongyu Guo、Yongyi Mao 和 Richong Zhang。Mixup作为局部线性流形外正则化。载于《AAAI人工智能会议论文集》，第33卷，第3714–3722页，2019年。2、3[8] Kaiming He、Xiangyu Zhang、Shaoqing Ren 和 Jian Sun。用于图像识别的深度残差学习。载于《IEEE计算机视觉与模式识别会议论文集》，第770–778页，2016年。8[9] Kaiming He、Xiangyu Zhang、Shaoqing Ren 和 Jian Sun。深度残差网络中的恒等映射。载于《欧洲计算机视觉会议论文集》，第630–645页。施普林格，2016年。2、7、8[10] Dan Hendrycks 和 Kevin Gimpel。用于检测神经网络中误分类样本和分布外样本的基线。《国际学习表征会议》，2017年。8[11] Dan Hendrycks、Mantas Mazeika 和 Thomas Dietterich。通过异常值暴露进行深度异常检测。载于《国际学习表征会议》，2018年。2[12]Martin Heusel、Hubert Ramsauer、Thomas Unterthiner、Bernhard Nessler 和 Sepp Hochreiter。采用双时间尺度更新规则训练的生成对抗网络收敛到局部纳什均衡。《神经信息处理系统进展》，第30卷，2017年。1[13] Minui Hong、Jinwoo Choi 和 Gunhee Kim。StyleMix：分离内容与风格以增强数据增强。载于 *IEEE/CVF* 计算机视觉大会论文集— 觉与 *Pattern R*识别，页码 14 2–14 87 21 0 3[14] Rui Huang 和 Yixuan Li。《MOS：面向大规模语义空间扩展分布外检测》。载于 IEEE/CVF计算机视觉与模式识别会议论文集，第 8710–8719 页，2021年。2[15] Liming Jiang、Bo Dai、Wayne Wu 和 Chen Change Loy。《Deceive d：利用自适应伪增强，在有限数据条件下训练生成对抗网络》。神经信息处理系统进展，第 34 卷：21655–21667，2021年。2, 3, 6, 7[16] Tero Karras、Timo Aila、Samuli Laine 和 Jaakko Lehtinen。《通过 GAN 渐进式增长提升质量、稳定性和多样性》。国际学习表征会议，2018年。2[17] Tero Karras、Miika Aittala、Janne Hellsten、Samuli Laine、Jaakko Lehtinen 和 Timo Aila。《在有限数据条件下训练生成对抗网络》。神经信息处理系统进展，第 33 卷：12104–12114，2020年。2, 3, 6, 7[18] Tero Karras、Miika Aittala、Samuli Laine、Erik H¨ark¨onen、Janne Hellsten、Jaakko Lehtinen 和 Timo Aila。《无混叠生成对抗网络》。神经信息处理系统进展，第 34 卷：852–863，2021年。1[19] Tero Karras、Samuli Laine 和 Timo Aila。《用于生成对抗网络的基于风格的生成器架构》。载于 IEEE计算机视觉与模式识别会议论文集，第 4401–4410 页，2019年。5, 6[20] Tero Karras、Samuli Laine、Miika Aittala、Janne Hellsten、Jaakko Lehtinen 和 Timo Aila。《分析并改进 StyleGAN 的图像质量》。载于 IEEE计算机视觉与模式识别会议论文集，第 8110–8119 页，2020年。1, 2, 4, 6[21] Jang-Hyun Kim、Wonho Choo 和 Hyun Oh Song。《Puzzle mix：利用显著性和局部统计量实现最优混合增强》。载于国际机器学习会议论文集，第 5275–5285 页。PMLR，2020年。2[22] Alex Krizhevsky。《从微小图像中学习多层特征》。技术报告TR-2009，多伦多大学，多伦多，2009年。3, 5, 6[23] Karol Kurach、Mario Luˇci´c、Xiaohua Zhai、Marcin Michal-ski 和 Sylvain Gelly。《关于生成对抗网络中正则化与归一化的大规模研究》。载于国际机器学习会议论文集，第 3581–3590 页。PMLR，2019年。1[24] Kimin Lee、Kibok Lee、Honglak Lee 和 Jinwoo Shin。《用于检测分布外样本和对抗攻击的简单统一框架》。神经信息处理系统进展，第 31 卷，2018年。3, 8[25] Bing Li、Yuanlue Zhu、Yitong Wang、Chia-Wen Lin、Bernard Ghanem 和 Linlin Shen。《AniGAN：用于无监督动漫人脸生成的风格引导生成对抗网络》。IEEE多媒体汇刊，2021年。1[26] Shiyu Liang、Yixuan Li 和 Rayadurgam Srikant。《提升神经网络中分布外图像检测的可靠性》。国际学习表征会议，2018年。3, 8[27] Haozhe Liu、Bing Li、Haoqian Wu、Hanbang Liang、Yawen Huang、Yuexiang Li、Bernard Ghanem 和 Yefeng Zheng。

通过流形熵估计应对生成对抗网络中的模式崩溃。arXiv预印本 arXiv:2208.12055，2022年。1[28] Haozhe Liu、Haoqian Wu、Weicheng Xie、Feng Liu 和 Lin-lin Shen。《基于组级抑制的特征正则化，用于鲁棒分类》。载于 IEEE/CVF国际计算机视觉会议论文集，第 478–486 页，2021年。2 – [29] Haozhe Liu、Wentian Zhang、Jinheng Xie、Haoqian Wu、Bing Li、Ziqi Zhang、Yuexiang Li、Yawen Huang、Bernard Ghanem 和 Yefeng Zheng。《用于分布外视觉识别的解耦混合增强》。载于欧洲计算机视觉会议论文集，第 451–464 页。施普林格，2023年。2[30] Ziwei Liu、Ping Luo、Xiaogang Wang 和 Xiaoou Tang。《野外人脸属性的深度学习》。载于 IEEE国际计算机视觉会议论文集，第 3730–3738 页，2015年。5 [31] Ziwei Liu、Ping Luo、Xiaogang Wang 和 Xiaoou Tang。《野外人脸属性的深度学习》。载于国际计算机视觉会议（ICCV）论文集，2015年12月。6[32] Ziwei Liu、Zhongqi Miao、Xiaohang Zhan、Jiayun Wang、Boqing Gong 和 Stella X Yu。《开放世界中的大规模长尾识别》。载于 IEEE计算机视觉与模式识别会议论文集，第 2537–2546 页，2019年。3, 8[33] Xudong Mao、Qing Li、Haoran Xie、Raymond YK Lau、Zhen Wang 和 Stephen Paul Smolley。《最小二乘生成对抗网络》。载于 IEEE计算机视觉与模式识别会议论文集，第 2794–2802 页，2017年。6[34] Takeru Miyato、Toshiki Kataoka、Masanori Koyama 和 Yuichi Yoshida。《生成对抗网络的谱归一化》。国际学习表征会议，2018年。1, 2[35]Aamir Mustafa 和 Rafał K Mantiuk。《变换一致性正则化——一种用于图像到图像转换的半监督范式》。载于《计算机视觉—ECCV 2020：第 16 届欧洲计算机视觉会议》，英国格拉斯哥，2020年8月23日至28日，论文集，第 XVIII 卷第 16 部分，第 599–615 页。施普林格，2020年。3[36] Lawrence Neal、Matthew Olson、Xiaoli Fern、Weng-Keen Wong 和 Fuxin Li。《利用反事实图像进行开放集学习》。载于欧洲计算机视觉会议论文集，第 613–628 页，2018年。8[37] Alec Radford、Luke Metz 和 Soumith Chintala。《利用深度卷积生成对抗网络进行无监督表征学习》。arXiv预印本 arXiv:1511.06434，2015年。1, 2, 6[38] Olga Russakovsky、Jia Deng、Hao Su、Jonathan Krause、Sanjeev Satheesh、Sean Ma、Zhiheng Huang、Andrej Karpathy、Aditya Khosla、Michael Bernstein 等。《ImageNet大规模视觉识别挑战赛》。国际计算机视觉杂志，115(3)：211–252，2015年。5[39] Tim Salimans、Ian Goodfellow、Wojciech Zaremba、Vicki Cheung、Alec Radford 和 Xi Chen。《改进生成对抗网络训练的技术》。神经信息处理系统进展，第 29 卷，2016年。1, 2, 6 [40] J¨urgen Schmidhuber。关于在模型构建神经控制器中实现好奇心和厌倦的一种可能性。载于《自适应行为模拟国际会议论文集：从动物到人工生命》，第222–227页，1991年 1 – 。[41] J¨urgen Schmidhuber。生成对抗网络是人工好奇心（1990）的特例，也与可预测性最小化（1991）密切相关。《神经网络》，127：58–66，2020年。1[42] Casper Kaae Sønderby、Jose Caballero、Lucas Theis、Wen-zhe Shi和Ferenc Husz´ar。用于图像超分辨率的摊销最大后验推断。国际机器学习会议，2017年。6[43] Hung-Yu Tseng、Lu Jiang、Ce Liu、Ming-Hsuan Yang和Weilong Yang。在有限数据下对生成对抗网络进行正则化。载于IEEE/CVF计算机视觉与模式识别会议论文集，第7921–7931页，2021年。3, 6[44] Vikas Verma、Alex Lamb、Christopher Beckham、Amir Na-jafi、Ioannis Mitliagkas、David Lopez-Paz和Yoshua Ben-gio。流形混合：通过插值隐藏状态获得更好的表征。载于国际机器学习会议，第6438–6447页。PMLR，2019年。2, 3, 4, 7, 8[45] Xintao Wang、Ke Yu、Shixiang Wu、Jinjin Gu、Yihao Liu、Chao Dong、Yu Qiao和Chen Change Loy。ESRGAN：增强型超分辨率生成对抗网络。载于欧洲计算机视觉会议（ECCV）研讨会论文集，第0–0页，2018年。1 [46] Yuanbo Xiangli、Yubin Deng、Bo Dai、Chen Change Loy和Dahua Lin。真实还是虚假，这是个问题。国际机器学习会议，2020年。6[47] Jingkang Yang、Haoqi Wang、Litong Feng、Xiaopeng Yan、Huabin Zheng、Wayne Zhang和Ziwei Liu。语义一致的分布外检测。载于IEEE/CVF国际计算机视觉会议论文集，第8301–8309页，2021年。3[48] Ryota Yoshihashi、Wen Shao、Rei Kawakami、Shaodi You、Makoto Iida和Takeshi Naemura。用于开放集识别的分类-重构学习。载于IEEE计算机视觉与模式识别会议论文集，第4016–4025页，2019年。8[49] Qing Yu和Kiyoharu Aizawa。通过最大分类器差异进行无监督分布外检测。载于IEEE/CVF国际计算机视觉会议论文集，第9518–9526页，2019年。2[50] Sangdoo Yun、Dongyoon Han、Seong Joon Oh、Sanghyuk Chun、Junsuk Choe和Youngjoon Yoo。CutMix：利用可定位特征训练强分类器的正则化策略。载于IEEE/CVF国际计算机视觉会议论文集，第6023–6032页，2019年。2[51] Alireza Zaeemzadeh、Niccol`o Bisagno、Zeno Sambugaro、Nicola Conci、Nazanin Rahnavard和Mubarak Shah。使用一维子空间并集进行分布外检测。载于IEEE计算机视觉与模式识别会议论文集，第9452–9461页，2021年。2, 3, 5, 8

[52] Sergey Zagoruyko和Nikos Komodakis。宽残差网络。载于英国机器视觉会议。英国机器视觉协会，2016年。2, 7, 8[53] Hongyi Zhang、Moustapha Cisse、Yann N Dauphin和David Lopez-Paz。Mixup：超越经验风险最小化。arXiv预印本arXiv:1710.09412，2017年。2, 3, 5, 7, 8 [54] Han Zhang、Ian Goodfellow、Dimitris Metaxas和Augustus Odena。自注意力生成对抗网络。载于国际机器学习会议，第7354–7363页。PMLR，2019年。1[55] Han Zhang、Zizhao Zhang、Augustus Odena和Honglak Lee。生成对抗网络的一致性正则化。国际学习表征会议，2020年。3[56] Wenlong Zhang、Yihao Liu、Chao Dong和Yu Qiao。RankSRGAN：带排序器的图像超分辨率生成对抗网络。载于IEEE/CVF国际计算机视觉会议论文集，第3096–3105页，2019年。1[57] Junbo Zhao、Michael Mathieu和Yann LeCun。基于能量的生成对抗网络。载于国际学习表征会议，2017年。6[58] Jiachen Zhong、Xuanqing Liu和Cho-Jui Hsieh。通过对抗训练提升GAN的速度和质量。arXiv预印本arXiv:2008.03364，2020年。3 [59] Mingchen Zhuge、Dehong Gao、Deng-Ping Fan、Linbo Jin、Ben Chen、Haoming Zhou、Minghui Qiu和Ling Shao。Kaleido-BERT：时尚领域的视觉-语言预训练。载于IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，第12647–12657页，2021年6月。3