GAN 反演:一项调查

Weihao Xia, Yulun Zhang, Yujiu Yang*, Jing-Hao Xue, Bolei Zhou*, Ming-Hsuan Yang*

摘要:GAN反演旨在将给定图像反演回预训练 GAN 模型的潜在空间,以便生成器可以从反演代码忠实地重建图像。作为一种连接真实和虚假图像域的新兴技术,GAN 反演在启用预训练 GAN 模型（例如 StyleGAN 和 BigGAN)以应用于真实图像编辑方面发挥着至关重要的作用。此外,GAN 反演解释了 GAN 的潜在空间并检查了如何生成逼真的图像。在本文中,我们对 GAN 反演进行了调查,重点关注其代表性算法及其在图像恢复和图像处理中的应用。我们进一步讨论了未来研究的趋势和挑战。可以在此github上找到 GAN 反演方法、数据集和其他相关信息的精选列表地点。

索引词 生成对抗网络、可解释机器学习、图像重建、图像处理

1简介

生成对抗网络（GAN）是一种深度学习生成新数据的生成模型[1]。它由两个神经网络组成:一个生成器G 和一个鉴别器D,它们通过对抗过程联合训练。 G的目标是合成与真实数据相似的假数据,而D的目标是区分真假数据。通过对抗训练过程,生成器G尝试生成与真实数据分布相匹配的假数据来欺骗鉴别器。近年来,GAN 已应用于计算机视觉任务,从图像翻译[2]、 [3]、 [4]、图像处理[5]、 [6]、 [7],到图像恢复[8]、 [9]， [10]。

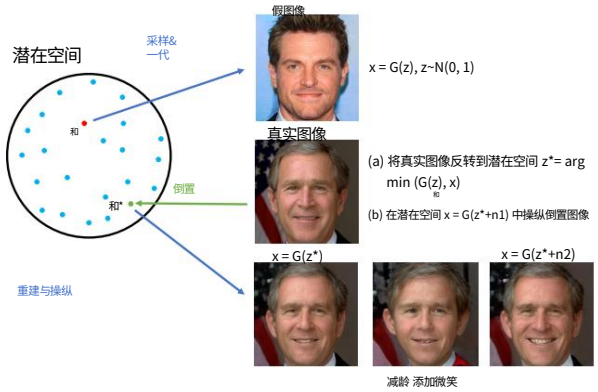


图 1. GAN 反演示意图。与使用训练生成器 G 的传统采样和生成过程不同,GAN 反演将给定的真实图像x映射到潜在空间并获得潜在代码z *。然后通过x = G(z)获得重建图像x 。

通过在不同的可解释方向上改变潜在代码z * ,例如+ n2 ,其中n1和n2分别模拟潜在空间中z * + n1和z * 的年龄和微笑,我们可以编辑真实图像的相应属性。重构结果来自[19]。

已经开发了许多 GAN 模型,例如 PGGAN [11]、 BigGAN [12]和 StyleGAN [13]、 [14],以从随机潜在代码中合成具有高质量和多样性的图像。最近的研究表明,GAN 在图像生成的监督下有效地在中间特征[15]和潜在空间[16]、 [17]、 [18]中编码丰富的语义信息。这些方法可以合成具有多种属性的图像,例如具有不同年龄和表情的面部,以及具有不同光照条件的场景。通过改变潜在代码,我们可以操纵某些属性,同时保留生成图像的其他属性。然而,由于 GAN 缺乏推理能力,潜在空间中的这种操作仅适用于 GAN 生成器生成的图像,而不适用于任何给定的真实图像。

GAN 反转旨在将给定图像反转回预训练 GAN 模型的潜在空间。然后可以通过生成器从反码中忠实地重建图像。由于 GAN 反演在桥接真实和虚假图像域中起着至关重要的作用,因此已经取得了重大进展[14]、 [17]、 [18]、 [20]、 [21]、 [22]、 [23]、 [24]， [25]。 GAN 反转使得在现有训练的 GAN 的潜在空间中发现的可控方向适用于编辑真实图像,而不需要任何临时监督或昂贵的优化。如图1所示,将真实图像倒置到潜在空间后,我们可以沿着一个特定的方向改变其代码来编辑图像的相应属性。作为结合 GAN 和可解释机器学习技术的快速发展方向,GAN 反演不仅是一个灵活的图像编辑框架,而且有助于揭示深度生成模型的内部工作原理。

*通讯作者 W. Xia 和 Y. Yang 就读于中国清华大学清华深圳国际研究生院。电子邮箱:xiawh3@outlook.com, yang.yujiu@sz.tsinghua.edu.cn Y. Zhang 在计算机视觉实验室工作,ETH Z ¨urich, Z ¨urich 8092, Switzerland。邮箱:yulun100@gmail.com J.-H.薛就读于英国伦敦大学学院统计科学系。电子邮件:jinghao.xue@ucl.ac.uk B. Zhou 就职于加州大学洛杉矶分校计算机科学系。电子邮件:bolei@cs.ucla.edu M.-H。杨在加州大学默塞德分校、延世大学和谷歌任职。电子邮件： mhyang@ucmerced.edu

在本文中,我们对

GAN 反演方法,重点是算法和应用。据我们所知,这项工作是对快速增长的 GAN 反演的第一次调查,具有以下贡献。我们对 GAN 反演方法进行了全面回顾,并比较了它们的不同属性和性能。我们进一步讨论了未来研究的挑战、未解决的问题和趋势。

本调查报告的其余部分组织如下。
我们首先在第 2 节中给出了 GAN 反演的问题公式。获得的给定图像的潜在代码应该具有两个属性:1)完全和逼真地重建输入图像的真实性和 2)促进下游任务。实现这两个属性也是 GAN 反演的目标。3.1节介绍了许多不同的预训练 GAN 模型G(z)。随后的部分介绍了不同 GAN 反演方法为达到目标所付出的努力。为了评估 GAN 反演方法的性能,我们在第3.2 节中考虑了两个重要方面,即重建图像的逼真度(感知质量)和忠实度(反演精度)。第一个方面取决于公式是如何解决的。由于G(z)的非凸性,它通常是一个非凸优化问题,很难找到准确的解决方案。第二个方面主要取决于使用哪个潜在空间。4.1节介绍、分析和比较了不同潜在空间的特征。在 4.2、4.3和4.4节中,我们介绍了现有方法如何尝试提供解决方案,并讨论这些 GAN 反演方法的一些重要特征。

第5节和第 6节介绍了 GAN 反演的应用和未来方向。

2问题定义与概述

众所周知,GAN [1]、 [11]、 [13]可以生成高分辨率和逼真的假图像。然而,由于缺乏推理能力,将这些无条件 GAN 应用于真实图像的编辑仍然具有挑战性。给定一张图像,GAN 反演旨在恢复预训练无条件 GAN 模型的潜在空间中的潜在代码,从而通过操纵潜在代码来启用众多图像编辑应用程序。在这种情况下,可以在不修改架构的情况下使用预训练的无条件 GAN 模型。理想情况下,给定图像的找到的潜在代码应该实现两个目标:1)忠实地和逼真地重建输入图像和 2)促进下游任务。

我们首先在统一的数学公式下定义 GAN 反演问题。无条件 GAN 的生成器学习映射G : Z → X 。当z1， z2 ∈ Z在Z空间中很接近时,对应的图像x1， x2 ∈ X在视觉上是相似的。 GAN 反演旨在找回潜在表示z ,或者等效地,找到可以完全由训练有素的生成器G合成并保持接近真实图像 x 的图像x * 。形式上,将要反转的信号表示为x ∈ R n,将训练有素的生成器表示为G : R m → R n,将潜在向量以下反转问题:

,

$$和^{*} = \arg\min_{和} \quad `(G(z), x), \qquad (1)$$

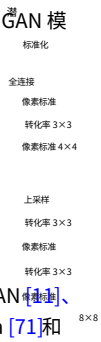
其中`(·)是图像或特征空间中的距离度量， G被假定为前馈神经网络。通常， `(·)可以基于`1、 `2、感知[26]或 LPIPS [27]度量。在实践中也可以包括对潜在代码[19]或面部身份[28]的一些其他限制。从得到的z我们可以得到原始图像;我们可以改变z以进一步获得操纵后的图像。

*,
*

促进下游任务的第二个目标主要取决于使用哪个潜在空间(参见第4.1 节)。
第一个目标取决于如何准确求解方程(1) ,由于G(z)的非凸性,这通常是一个非凸优化问题。因此,很难找到准确的解决方案。已经开发了许多方法[20]、 [21]、 [28]来使用基于学习、优化或两者的公式来求解方程(1) 。基于学习的反演方法旨在学习编码器网络将图像映射到潜在空间中,从而使基于潜在代码的重建图像看起来尽可能与原始图像相似。基于优化的反演方法通过反向传播直接求解目标函数,以找到最小化像素级重建损失的潜在代码。混合方法首先使用编码器生成初始潜在代码,然后使用优化算法对其进行细化。通常,基于学习的 GAN 反演方法不能忠实地重建图像内容。例如,已知基于学习的反演方法在重建人脸图像时有时无法保留身份以及其他一些细节[19]， [28]。虽然基于优化的技术已经实现了卓越的图像重建质量,但它们不可避免的缺点是计算成本明显更高[21]、 [22]。因此,最近基于学习的 GAN 反演方法的改进主要集中在如何忠实地重建图像,例如,在训练期间集成额外的面部身份损失[28]、 [29]或提出迭代反馈机制[30]。最近对基于优化的方法的改进强调如何更快地找到所需的潜在代码,因此提出了几种初始化策略[21]、 [22]和优化器[20]、 [24]。现有的反演方法无法同时实现重建质量和推理时间,导致“质量时间权衡”。尽管还提出了一些混合方法来平衡这种权衡,但快速找到准确的潜在代码仍然是一个挑战。

与 GAN 反演类似,一些任务也旨在学习 GAN 模型的逆映射。一些方法[31]、 [32]、 [33]、 [34]使用额外的编码器网络来学习 GAN 的逆映射,但它们的目标是与生成器和鉴别器共同训练编码器,而不是使用训练有素的 GAN 模型。其他一些方法,例如 PULSE [35]、 ILO [36]或 PICGM [37],也依赖于预训练的生成器来解决逆问题,例如修复、超分辨率或去噪。他们设计了不同的优化机制来搜索满足给定退化观察的潜在代码。由于他们的目标是从退化的观察(例如,噪声图像)中搜索准确和可靠的估计(例如,去噪图像) ,而不是对给定图像的忠实重建,因此我们在本次调查中不将它们归类为 GAN 反演方法

纸。但是关注这些工作是有益的,因为它们具有相同的想法,即在预训练的 GAN 模型的潜在空间中找到所需的潜在代码。



3预赛

3.1 GAN 模型和数据集

诸如 GAN [1]之类的深度生成模型已经
用于对自然图像分布进行建模和合成
逼真的图像。 GAN 的最新进展,如 DCGAN [38]、 WGAN [44]、 PGGAN [11]、 BigGAN [12]、 StyleGAN [13]、 StyleGAN2 [14]、 StyleGAN2 -Ada [71]和 StyleGAN3 [72]开发了更好的架构、损失和训练方案。这些模型在不同的数据集上进行训练,包括人脸(CelebA-HQ [11]、 FFHQ [13]、 [14]、 AnimeFaces [73]和 AnimalFace [74])、场景(LSUN [41])和对象(LSUN [41]和 ImageNet [53])。具体来说,在 ImageNet 上预训练的 BigGAN、CelebA-HQ 上的 PGGAN 以及 FFHQ 或 LSUN 上的基于样式的 GAN 被广泛用于 GAN 反演方法。与上述 2D GAN 相比,最近开发的 3D 感知 GAN [75]、 [76]弥合了 2D 图像和 3D 物理世界之间的差距。基于这些 3D 感知 GAN 的反演方法目前研究较少,但在图像、视频和 3D 应用中具有巨大潜力。

3.1.1 GAN 模型

DCGAN [38]在鉴别器中使用卷积,在生成器中使用分数步长卷积。

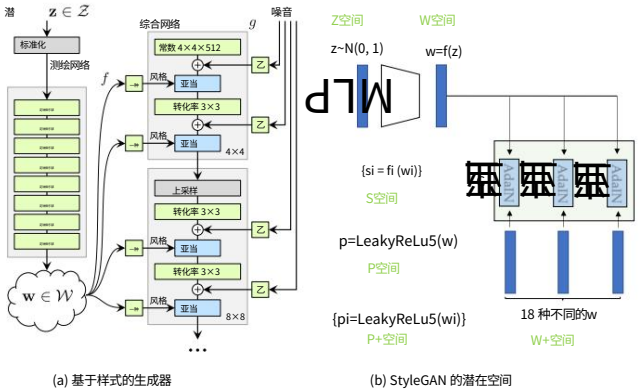
WGAN [44]最小化了生成的数据分布和真实数据分布之间的 Wasserstein 距离,这提供了更高的模型稳定性并使训练过程更容易。

BigGAN [12]生成高分辨率和高质量的图像,并通过放大、架构更改和正交正则化进行修改,以提高大规模 GAN 的可扩展性、鲁棒性和稳定性。 BigGAN 可以在 ImageNet [53]上以256×256和512×512 进行训练。

PGGAN [11],也称为 ProGAN 或渐进式 GAN,在训练过程中使用增长策略。关键思想是从生成器和判别器的低分辨率开始,然后添加新层,随着训练的进行,对越来越细粒度的细节进行建模。

这种方法提高了训练速度和稳定性,从而促进了更高分辨率的图像合成,例如 1024×1024像素的 CelebA 图像。

基于样式的 GAN,例如 StyleGAN [13],隐式学习分层潜在样式以生成图像。该模型通过操纵每个通道的均值和方差来有效地控制图像的样式[77] 。如图2(a) 所示,StyleGAN 生成器将样式向量(由映射网络f 定义)和随机变化(由噪声层提供)作为图像合成的输入。这提供了对不同细节级别的生成图像样式的控制。 StyleGAN2 模型[14]通过提出权重解调、路径长度正则化、生成器重新设计和去除渐进式增长,进一步提高了感知质量。 StyleGAN2-Ada [71]提出了一种自适应鉴别器增强机制来稳定有限数据的训练。 StyleGAN3 [72]观察到 GAN 中的“纹理粘连”问题(混叠)并提出



(a) 基于样式的生成器 (b) StyleGAN 的潜在空间

图 2. (a) 基于样式的生成器的架构。 (b) 构建反演方法的潜在空间。 (b) 中的合成网络 g 和 AdaIN 与 (a) 中的相同。

一种新的架构,通过考虑连续域中的混叠效应并对结果进行适当的低通滤波,更适合视频和动画。

对于 StyleGAN 和 StyleGAN2,它们的层数L由输出图像大小R 决定: L= 2 log2 R−2;它还具有18层的最大分辨率为1024×1024 。

对于 StyleGAN3,层数是一个自由参数,与输出分辨率没有直接关系。

3.1.2 数据集

ImageNet [53]是用于视觉对象识别研究的大规模手动注释数据集,包含超过 1400 万张图像,超过 20,000 个类别。

CelebA [45]是一个大规模的人脸属性数据集,由 20 万张名人图像组成,每张图像有 40 个属性注释。
CelebA 及其后续的 CelebA-HQ [11]和 CelebAMask-HQ [78]被广泛用于人脸图像的生成和处理。

Flickr-Faces-HQ (FFHQ) [13]是从 Flickr 爬取的高质量人脸图像数据集,由 70,000 张1024 × 1024像素的高质量人脸图像组成,在年龄、种族方面包含相当大的差异和图像背景。

LSUN [41]为 10 个场景类别(例如,卧室、教堂或塔楼)和 20 个对象类别(例如,鸟、猫或公共汽车)中的每一个包含大约一百万个标记图像。教堂和卧室场景图像以及汽车和鸟类物体图像通常用于 GAN 反演方法。

一些 GAN 反演研究也使用其他数据集
在他们的实验中,例如DeepFashion [79]、 Anime Faces [73]和 StreetScapes [80]。

3.2 评价指标

评估 GAN 反演方法有不同的维度,例如照片写实度、重建图像的忠实度和反演潜码的可编辑性。

3.2.1 写实主义

IS、FID 和 LPIPS 指标被广泛用于评估 GAN 生成图像的逼真质量。其他指标,例如 Fre´chet 分割距离 (FSD) [23]

表格1

GAN 反演方法的属性。类型包括基于学习的 (L.)、基于优化的 (O.) 和混合 (H.) GAN 反演。 S.-A.,L.-W.,
和 S.-R 分别表示语义感知、分层和支持的分辨率。 GAN 模型和 Dataset 表示哪些 GAN 模型
在方法正在反转的数据集上进行训练,可以在第3.1 节中找到。

方法	出版物类型 S.-AL-W。				S.-R 空间 GAN 模型 64 Z [38] [38]、 [44]		数据集	关键词
朱等人。 [20]	ECCV’ 16	H。			[38] [11]		[39], [40], [41]	GAN 的反演
克雷斯韦尔等人。 [42]、 [43]	NeurIPS 16	这。			第128章		[39], [45]	首先使用术语反转
佩拉瑙等人。 [46]	NeurIPS 16	L.			64 Z		[45], [47]	条件GAN的反演
GANPaint [48]	TOG 19	H.X		X 256 Z			[41]	学习特定于图像的生成器
GANSeeing [23]	ICCV’ 19	H.X		X 256	Z、 W [11]、 [13]、 [44]		[41]	模式崩溃的可视化
Image2StyleGAN [21]	ICCV’ 19	O.X			1024瓦[13]		[13]	StyleGAN 的第一次反转
Image2StyleGAN++ [22]	CVPR’ 20	O.X		X 1024 W+ [11]、 [13]	[11]、 [13]		[11], [13]	
mGANPrior [49]	CVPR’ 20	O.X		X 256 Z			[11], [13], [41]	多码 GAN 先验
风格编辑[50]	CVPR’ 20	O.X			1024瓦[11]、 [13]、 [14]		[13], [41]	
YLG [51]	CVPR’ 20	这。			第128章	[52]	[53]	注意力
胡等人。 [24]	ECCV’ 20	O.X			第1024章	[12], [14]	[13], [41], [53]	类条件
IDInvert [19]	ECCV’ 20	H.X		X 256瓦+		[13]	[13]、 [41]	域内
SG-蒸馏[54]	ECCV’ 20	O.X		X 1024瓦+		[14]	[13]	
MimicGAN [55]	IJCV 20	这。			64 Z	[45]	[38]	对于损坏的图像
柴等人。 [56]	ICLR 21	L.	X	X 1024 Z, W+ [11], [14]			[11], [13], [41]	数据增强
pSP [28]	CVPR’ 21	L.	X	X 1024瓦+		[14]	[11]	map2style 模块
风格空间[57]	CVPR’ 21	O.X		X 1024	小号	[14]	[13], [41]	S-空间
GH-壮举[58]	CVPR’ 21	L.	X	X 256	小号	[13]	[13], [41], [47]	生成层次特征
GANEnsembling [59]	CVPR’ 21	H.X		X 1024瓦+		[14]	[13], [41]	
e4e [60]	TOG 21	L.	X	X 1024瓦+		[14]	[11], [13], [41]	编辑用编码器
徐等人。 [61]	ICCV’ 21	O.X		X 1024瓦+		[13]	[13], [62]	对于连续图像
重新设计[30]	ICCV’ 21	L.	X	X 1024瓦+		[14]	[11], [13], [41], [63]迭代细化	
BD反转[64]	ICCV’ 21	O.X		X 1024 F/W+ [11], [14] [13], [14] [14] [14] [14]			[11], [13], [41]	超出范围， F/W+-空间
朱等人。 [65]	arxiv 21	O.X		X 1024 P		[13], [14]	[13] [11],	P和P+空间
魏等人。 [29]	arxiv 21	L.	X	X 1024瓦+			[13] [11],	高效的编码器架构
PTI [66]	arxiv 21	H.X			1024瓦		[13] [11],	围绕枢轴潜在代码调整G
超风格[67]	CVPR’ 22	H.X			1024瓦		[13], [68] [11],	学习优化生成器
HFGI [69]	CVPR’ 22	L.	X	X 1024瓦+			[13], [68]	
超级逆变器[70]	CVPR’ 22	L.	X		1024瓦	[14]	[11]、 [13]、 [41]	两相反转

和切片 Wasserstein 差异 (SWD) [81]也
用于图像感知质量评估。在[82]中，
徐等人。对评估进行实证研究
GAN 模型的指标。

初始分数(IS) [83]是一种广泛使用的衡量指标
GAN 生成的图像的质量和多样性
楷模。它计算合成图像的统计信息
使用在_
图像网[85]。分数越高越好。

Fre´chet 起始距离[86] (FID) 由
特征向量与真实和
基于 Inception-v3 [84] pool3 层生成的图像。
较低的 FID 表示更好的感知质量。

学习感知图像块相似度(LPIPS) [27]
使用 VGG 模型测量图像感知质量[87]
在 ImageNet 上预训练。较低的值意味着较高
图像块之间的相似性。

3.2.2 忠实

忠实度衡量真实图像之间的相似度
和生成的。它可以近似为
图像相似度。最广泛使用的指标是 PSNR
和 SSIM。一些方法使用逐像素重建
距离,例如,平均绝对误差 (MAE),均方
误差 (MSE) 或均方根误差 (RMSE)。

峰值信噪比(PSNR) 是最广泛使用的信噪比之一
使用标准来衡量重建的质量。这
地面实况图像和重建之间的 PSNR 由图像的最大可能像素值定义

图像和图像之间的均方误差。

结构相似性（SSIM） [88]测量结构相似性
基于亮度、对比度和结构方面的独立比较的图像之间的相似性。这

这些术语的详细信息可以在[88]中找到。

3.2.3 可编辑性

可编辑性衡量倒置的可编辑灵活性
关于输出的某些属性的潜在代码
来自生成器的图像。直接评估可编辑性
隐藏代码是困难的。现有方法使用
余弦或欧几里得距离[89]或分类精度[90]以评估输入x和

输出x [0] （即修改目标属性的同时保持
其他不变）。现有方法侧重于评估
面部数据和面部属性的可编辑性。例如，
尼赞等人。 [89]使用余弦相似度来比较
面部表情保存的准确性,由x的 2D 地标之间的欧几里得距离计算

和x [0]. 相反,姿势保存计算为
x和x的欧拉角之间的欧几里得距离
达尔等人。 [91]开发编辑一致性分数（回归
由一个属性分类器)来衡量一致性

[0].抗体

编辑的人脸图像基于这样的假设,即当使用属性分类器进行分类时,不同的编辑排列应该具有相同的属性分数。这些方法测量人脸身份的保存情况以评估编辑图像的质量。我们注意到上述方法可能不适用于面部以外的所有图像域。

3.2.4 主观指标除了上述指

标外,一些研究[65]、 [91]还包括用于绩效评估的人类评估者或用户研究。例如,对于主观图像质量评估,人类评价者被要求为图像分配感知质量分数,例如,从1 (差)到5 (好) 。最终分数,通常称为平均意见分数 (MOS) 或差异平均意见分数 (DMOS),计算为所有评分的算术平均值。典型的用户研究要求参与者从给定的三组图像(来源、基线结果和建议的方法)中选择最能满足问题的一个。问题可以是“从给定的两个编辑图像中选择一个更好地保留源图像中人的身份的图像”或“哪个编辑图像更真实?”最终的响应百分比表明所提出的方法相对于基线的偏好率。这些指标的缺点包括人类判断的非线性尺度、潜在的偏差和方差以及高昂的人力成本。

4 GAN反演方法

本节介绍 GAN 模型的不同潜在空间、代表性 GAN 反演方法及其属性。随着 StyleGAN 模型实现了最先进的图像合成,已经开发了许多基于 StyleGAN 的各种潜在空间[13]、 [14]、 [72]的 GAN 反演方法。除了通用 GAN 的Z空间外,还专门为 StyleGAN 设计了几个潜在空间,包括W、 W+、 S 和P空间。

4.1 嵌入哪个空间　从Z空间到P空间无论 GAN 反演方法如何,一个重要的设计选择是嵌入图像的潜在空间。

一个好的潜在空间应该是解开的并且易于嵌入。这种潜在空间中的潜在代码具有以下两个属性:它忠实地和逼真地重建输入图像,并促进下游图像编辑任务。本节介绍潜在空间分析和正则化从原始Z空间到最近的P空间的潜在空间的工作。 Z空间适用于所有 GAN ,一些潜在空间是专门为 StyleGAN [21]、 [57]、 [65]、 [92] 设计的。潜在空间的选择取决于预训练的模型和任务。例如,使用 StyleGAN 进行图像编辑主要在W+空间中执行。

Z空间。 GAN 架构中的生成模型学习将从简单分布(例如正态或均匀分布)采样的值映射到生成的图像。这些直接从分布中采样的值通常称为潜在代码或潜在表示(由z ∈ Z 表示) ,如图 2 所示。它们形成的结构通常称为潜在Z空间。 Z空间适用于所有无条件 GAN 模型,例如 DCGAN [38]、 PGGAN [11]、 BigGAN [12]和 StyleGANs [13]、 [14]、 [71]。然而, Z空间服从正态分布的约束限制了它的表示能力和对语义属性的区分。

W和W+空间。最近的 GAN 反演方法大多采用 StyleGAN 中使用的潜在空间。这些潜在空间具有更高的自由度,因此比Z空间更具表现力。图2说明了构建反演方法的潜在空间。各种潜在空间都是从原始Z空间衍生而来的。 StyleGAN [13]通过使用8 层多层感知器 (MLP)实现的非线性映射网络将原生z转换为映射样式向量w 。这个中间的潜在空间被命名为W空间。由于映射网络和仿射变换, StyleGAN的W空间比Z空间包含更多的分离特征。一些研究[18]、 [21]分析了W和Z空间的可分离性和语义。然而, W空间的表现力仍然有限,限制了可以忠实重建的图像范围。因此,一些工作[21]、 [22]使用了另一个逐层潜在空间W+,其中不同的中间潜在向量w 通过 AdaIN [77]被馈送到生成器的每个层。

然而,将图像反转到W+空间会以牺牲可编辑性为代价来减轻失真。最近的方法[60]、 [67]旨在通过预测W+中靠近W 的潜在代码来平衡重建可编辑性权衡。对于具有 18 层的 StyleGAN, w ∈ W有 512 维, w ∈ W+有18× 512 个维度。

S空间。样式空间S [57]由通道样式参数 s 跨越,其中 s通过对生成器的每一层使用不同的学习仿射变换从w ∈ W变换。在具有 18 层的 1024×1024 StyleGAN2 中, W、 W+和S分别具有 512、9216 和 9088 维。提出这个S空间是为了在超出语义级别的空间维度上实现更好的空间解缠结。空间纠缠主要是由基于样式的生成器[13]的内在复杂性和 AdaIN 归一化[77] 的空间不变性引起的。

徐等人。 [93]用编码器学习的解开的多级视觉特征替换原始样式代码。他们将这些样式参数所跨越的空间称为Y空间,但实际上可以将其视为S空间的一种。通过直接干预样式代码s ∈ S,基于S空间的方法[57]、 [94]实现了对局部翻译的细粒度控制。

P空间。最近的一种方法 PULSE [35]在搜索生成模型的潜在空间以找到所需点时观察到“肥皂泡”效应。顾名思义,“肥皂泡”效应是高维高斯的大部分密度位于超球面附近。上述作者提出将图像嵌入到Z空间中的超球体表面上。基于观察,朱等人。 [65]提出了一个P空间。由于最后一个leaky ReLU 使用了0.2 的斜率,所以从W空间到P空间的变换是x = LeakyReLU5.0(w),其中w和x分别是W和P空间中的潜在代码。他们做了最简单的假设,即潜在代码的联合分布近似为多元高斯分布,并进一步提出PN空间来消除

依赖性并消除冗余。 P空间到PN空间的变换是通过PCA白化得到的： v^= Λ−1 · CT (x−µ),其中Λ−1是缩放矩阵， C是正交矩阵， µ是均值向量。参数C、 Λ 和µ是从PCA(X) 获得的,由P空间中的 100 万个潜在样本组成。这种变换将分布归一化为零均值和单位方差,从而导致P空间在所有方向上都是各向同性的。 P N空间从PN空间扩展而来： v = Λ−1CT (xi − µ) 其中X ∈ R 106×512

+ 18 i=1。潜在代 每个 码的一部分用于解调不同层的相应 StyleGAN 特征图。

4.2 GAN 反演方法

图3显示了 GAN 反演的三种主要技术,即基于学习、优化或混合公式将图像投影到潜在空间中。倒置代码具有其他属性,即具有支持分辨率、语义感知、分层和分布外通用性。表1列出了现有 GAN 反演方法的一些重要属性。

4.2.1 基于学习的 GAN 反演 基于学习的 GAN 反演

[20]、 [46]、 [95]通常涉及训练编码神经网络E(x; θE)以将图像x 映射到潜在代码z经过

$$\theta^*_{和} = \arg 最小_{\theta E} \sum_n L(G(E(xn; \theta E)), xn), \quad (2)$$

其中xn表示数据集中的第 n 个图像。 (2)中的目标让人想起自动编码器管道,具有编码器E和解码器G。解码器G在整个训练过程中是固定的。除了准确的重建,一个好的 GAN 反演编码器应该具有以下特点:1)轻量级； 2) 数据效率； 3) 支持高分辨率图像（见第4.3.1 节）； 4) 对任意图像的泛化性（见第4.3.4 节）。

Perarnau 等人提出了一种较早的基于学习的 GAN 反演方法。 [46]。给定条件 GAN (cGAN) 模型,真实图像x由潜在代码编码， z通过改变y 合成。这种方法包括使用经过训练的条件 GAN (cGAN)训练编码器E。不同于朱等人。 [20]， 和一个属性向量y,一个修改后的图像x 0 该编码器E由两个模块组成： Ez,将图像编码为z,以及Ey,将图像编码为y。为了训练Ez,该方法使用生成器创建生成图像x和潜在向量z 的数据集,最小化z和Ez(G(z, y0))之间的平方重建损失Lez ,并通过直接用ky - Ey训练来改进Ey (x)k

0

2 2 . Ey最初是通过使用生成的及其条件信息y来训练的。由于图像x 0 StyleGAN [13]、 [14]、 [71]、 [72] 的流行,最近基于学习的方法为 StyleGAN 设计了一个编码器。理查森等人。 [28]提出MAP2STYLE模块从相应的特征图中学习风格,其中 18 个单层潜在代码分别预测。

Wei 等人并没有使用 18 个模块来学习 StyleGAN 的风格。 [29]提出了一个简单有效的头部,它

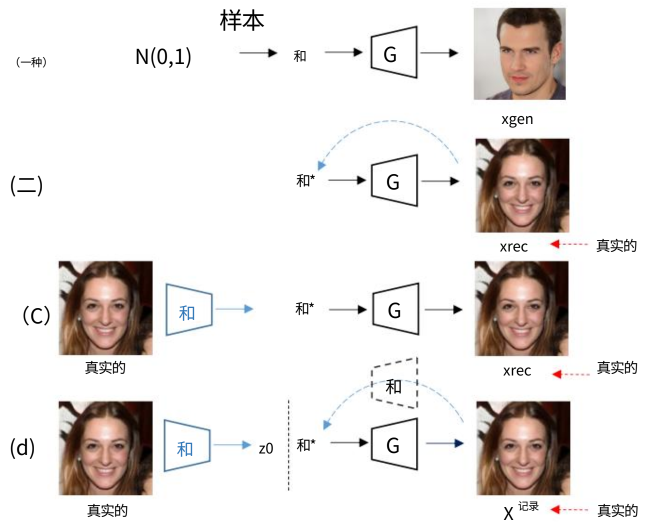


图 3. GAN 反演方法的说明。 (a) 给定一个训练有素的 GAN 模型G,照片般逼真的图像x gen可以从随机采样的潜在向量z 中生成。 (b) 基于学习的反演方法旨在学习编码器网络将图像映射到潜在空间,从而使基于潜在代码的重建图像看起来尽可能与原始图像相似。基于优化的反演方法通过反向传播直接求解目标函数,以找到最小化像素级重建损失的潜在代码。混合方法首先使用编码器生成初始潜在代码,然后使用优化算法对其进行细化。由虚线E 描绘,训练有素的编码器包含在[19]中,作为优化的正则化器。

蓝色块代表可训练或迭代模块,红色虚线箭头表示监督。

仅由一个平均池化层和一个全连接层组成。给定由特征金字塔网络（FPN） [96]获得的三个不同语义级别的特征,这三个头从浅层、中层和深层特征产生w15、 · · · w14和w18 、 w10 、 · · · w9 ， 分别。在[60] 中， Tov 等人分析了 StyleGAN 潜在空间内失真、感知质量和可编辑性之间的权衡 w1, · · · ，精度,Alaluf 等人。 [30]为编码器引入了一种迭代细化机制。在步骤t 中,编码器不是在前向传播中直接预测给定真实图像的潜在代码,而是对通过将给定图像x与预测图像连接获得的扩展输入进行操作： ∆t = E(x, yt),其中yt = G(wt)。然后将步骤t+1的潜在代码更新为wt+1 = ∆t + wt。

w0和y0的初始化值分别设置为平均潜码及其对应的图像。

虽然一些方法[31]、 [32]、 [33]、 [97]使用加法编码器网络来学习 GAN 的逆映射,但我们没有将它们归类为 GAN 逆,因为它们的目标是与两者共同训练编码器生成器和鉴别器,而不是确定经过训练的 GAN 模型的潜在空间。

4.2.2 基于优化的GAN反演

现有的基于优化的 GAN 反演方法通常通过优化潜在的

向量

$$\text{和}^{*} = \arg\min_{\text{和}} \; `(x, G(z; \theta)), \tag{3}$$

其中x是目标图像， G是由θ 参数化的 GAN 生成器。

选择优化器至关重要,因为好的优化器有助于缓解局部最小值问题。有两种类型的优化器:基于梯度的（ADAM [98]、 L BFGS [99]、哈密顿蒙特卡罗（HMC） [100])和无梯度（协方差矩阵自适应（CMA） [101])方法。基于优化的 GAN 反演方法使用不同的优化器。例如,ADAM [98]用于 Image2StyleGAN [21]， L-BFGS 被 Zhu 等人使用。 [20]。

胡等人。 [24]系统地试验了基于梯度和无梯度优化器的不同选择,发现 CMA 及其变体 BasinCMA 在将具有挑战性的数据集（例如 LSUN Cars [41])中的图像反转到StyleGAN2 [14] 的潜在空间。

基于优化的 GAN 反演的另一个重要问题是潜在代码的初始化。由于等式(1)是高度非凸的,因此重建质量强烈依赖于z的良好初始化（对于 StyleGAN [13]有时是w ）。实验表明,不同的初始值会导致生成的图像存在显着的感知差异[11]、 [12]、 [13]、 [38]。一个直观的解决方案是从几个随机初始值开始,以最小的成本获得最好的结果。Image2StyleGAN [21]研究了两种初始化选择,一种基于随机选择,另一种基于平均潜在代码w。然而,在获得稳定的重建[20]之前,可能会测试大量的随机初始值,这使得实时处理变得不可能。因此,一些[20]、 [102]改为训练深度神经网络以直接最小化(1) ,如第4.2.1 节所述。一些[20]、 [95]提出使用编码器为优化提供更好的初始化,这将在第4.2.3 节中讨论。

我们注意到,基于优化的方法[21]、 [22]、 [43]通常在内存和运行时间方面都需要昂贵的迭代过程,因为它们必须独立地应用于每个潜在代码。

4.2.3 Hybrid GAN Inversion 混

合方法[19]、 [20]、 [23]、 [95]利用了上述两种方法的优点。作为该领域的开创性工作之一,朱等人。 [20]提出了一个框架,该框架首先通过训练一个单独的编码器E(x; θE)来预测给定真实照片x的z ,然后使用获得的z作为优化的初始化。学习到的预测模型可作为非凸优化问题(1)的快速自下而上初始化。

随后的研究遵循这个框架并提出了几种变体。例如,为了反转G,Bau 等人。 [95]首先训练网络E以获得潜在代码z0 = E(x)及其中间表示r0 = gn(· · · (g1(z0))) 的合适初始化,其中gn(· · · (g1(·))在G(·)的分层表示中。然后,该方法使用r0初始化对r的搜索,以获得接近目标x的= G(r ）（有关详细信息,请参阅第4.3.3节的重构x ）。朱等人。 [19]表明,在大多数现有方法中,生成器G不提供

$*$

$$_{0}$$

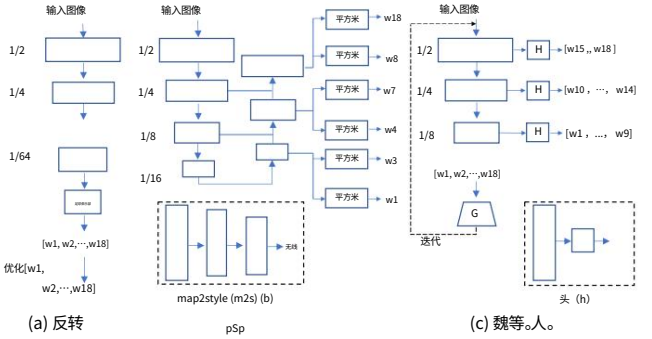


图 4. 三种基于学习的方法的编码器结构:(a) IDin vert [19]、 (b) pSp [28]和 (c) Wei 等人。 [29]。

它的领域知识来指导编码器E的训练,因为G(·)的梯度根本没有被考虑在内。为了解决这个问题,开发了一种特定领域的 GAN 反演方法,它既可以重建输入图像,又可以确保反演代码对语义编辑有意义（有关此方法的更多详细信息,请参见第4.3.2节）。

与以前的方法相比,Roich 等人。 [66]开发了一种发电机调谐技术。使用初始潜在代码作为枢轴,他们轻轻调整预训练生成器,以便可以忠实地重建输入图像。这个过程被称为关键调整,它有助于将域外图像忠实地映射到域内潜在代码[19] 。

阿拉鲁夫等人。 [67]进一步介绍了一个超网络[103] ,它学习针对给定的输入图像细化生成器的权重。超网络由轻量级特征提取器和一组细化块组成。

4.3 GAN 反演方法的性质在本节中,我们讨论 GAN 反

演方法的重要性质,即支持分辨率、语义感知、分层和分布外泛化性。

4.3.1 支持的分辨率 GAN 反演方法

可以支持的图像分辨率主要取决于生成器的容量和反演机制。朱等人。 [20]使用在多个数据集上训练的 GCGAN,图像为64 × 64像素,Bau 等人。 [48]、 [104]采用了 PGGAN [11] ,这些 PGGANs [11] 使用来自 Lsun [41]的256 × 256像素大小的图像进行训练。然而,一些方法不能充分利用预训练的 GAN 模型。朱等人。 [19]提出了一种编码器,将给定的图像映射到 StyleGAN 的潜在空间。该方法（图4 （a））对于256×256像素的图像表现良好,但由于计算成本高（图中 1/n 表示语义特征图） ,不能很好地扩展到1024×1024像素的图像的 1/n 原始输入分辨率）。相反,Richardson 等人提出的 pSp 方法。 [28] （图4 （b））可以合成1024 × 1024像素的图像,无论输入图像大小如何,因为他们提出的 18 个 map2style 模块用于分别预测 18 个单层潜在代码。魏等人。 [29]提出了一个类似的模型,但带有一个轻量级的编码器。与[28] 类似,来自三个语义级别的特征用于预测图像的不同部分

潜码。尽管如此,该模型从每个语义级别预测了 9、5 和 4 层潜在代码,如图4 (c) 所示。最近的应用,如百万像素上的人脸交换[105]、 [106]和无限分辨率图像合成[107]被开发为可以支持高分辨率图像编辑的图像反转方法。

4.3.2 语义感知 具有语义感知特性的

GAN 反演方法可以在像素级别执行图像重建,并将反演代码与潜在空间中出现的知识对齐。通过重用 GAN 模型中编码的丰富知识,语义感知的潜在代码可以更好地支持图像编辑。现有方法通常随机采样一组潜在代码z并将它们输入G(·)以获得相应的合成x E(·)然后由

.编码器

$$\min_{\theta_E} LE = kz - E(G(z))k2, \tag{4}$$

其中k· k2表示l2距离， ΘE表示编码器E(·)的参数。柯林斯等人。 [50]使用潜在对象表示来合成具有不同风格的图像并减少伪影。然而,仅通过重建z （或等效地，合成数据)的监督不足以训练准确的编码器。

为了缓解这个问题,朱等人。 [19]提出了一种特定领域的 GAN 反演方法来在像素和语义级别恢复输入的真实图像。该方法首先训练域引导编码器E将图像空间映射到潜在空间,使得编码器生成的所有代码都是域内潜在代码。编码器E被训练来恢复真实图像,而不是用合成数据训练来恢复潜在代码。然后,他们通过将这个训练有素的E作为正则化项来执行实例级域正则化优化,以在z优化期间微调语义域中的潜在代码。这种优化有助于更好地重构像素值,而不影响倒排码的语义特性。训练过程被制定为

$$LE = kx - G(E(x))k2 + \lambda 1kF(x) - F(G(E(x)))k2 \min_{\theta E} - \lambda 2E[D(G(E(x)))], \tag{5}$$

其中F(·)表示VGG特征提取， E[D(·)]是判别器损失， λ1和λ2分别是感知和判别器损失权重。来自所提出的域引导编码器的反码可以很好地基于预训练的生成器重建输入图像,并确保代码本身在语义上是有意义的。但是,代码仍然需要改进以更好地适应像素值处的单个目标图像。基于域引导编码器,他们设计了具有两个模块的域正则化优化：（i)以域引导编码器的输出为起点,以避免局部最小值并缩短优化过程,以及（ii）) 域引导编码器用于规范生成器语义域内的潜在代码。目标函数是

$$z^* = \arg\min_{和} kx - G(z)k2 + \lambda = {}^{\text{①}}1kF(x) - F(G(z))k2 + \lambda {}^{\text{②}}2kz - E(G(z))k2, \tag{6}$$

其中x是要反转的目标图像， λ 2是分别对应于感知损失和编码器正则化器的损失权重。

4.3.3 逐层当层数很大

时,确定方程（1)定义的完全反演问题的生成器是不可行的。最近开发了一些方法[23]、 [108] ,通过将生成器G分解为层来解决易处理的子问题：

$$G = Gf(gn(\cdots((g1(z))),\text{其中}g1,\ldots,gn\text{是}G \tag{7}$$

的早期层， Gf构造G的所有后面的层。

最简单的逐层 GAN 反演基于一层。雷等人。 [108]考虑G = g(z) = ReLU(Wz + b)形式的单层模型。当问题可以实现时,要找到一个可行的z使得x = G(z),可以通过解决线性规划问题来反转函数：

$$\text{无线} >z + bi = xi \ \forall i \text{ s.t. } xi > 0, >z + bi \leqslant \text{无线} 0, \forall i \text{ s.t. } xi = 0. \tag{8}$$

(8)的解集是凸的并形成一个多面体。然而,它可能包括不可数的可行点[108],这使得如何进行分层反演不清楚。

几种方法做出了额外的假设,以将上述结果推广到更深层次的神经网络。雷等人。[108]假设输入信号被`1或`∞方面的有界噪声破坏,并提出了一种使用线性程序逐层生成模型的反演方案。

确保稳定反演的分析仅限于以下情况：(1)网络的权重应该是高斯独立同分布变量； (2)每一层都应该扩大一个常数因子； (3) 最后一个激活函数应该是 ReLU [109]或leaky ReLU [110]。

然而,这些假设在实践中往往不成立。

为了反转复杂的最先进的 GAN,Bau 等人。 [23]提出解决反转最终层Gf的更简单问题：

$$\text{×}_0 = Gf(r^*), \tag{9}$$

其中r * sentation， `是图像特征空间中的潜在向量,他们在两步混合 GAN 反演框架中解决了反演问题(1) :首先构建一个神经网络E ,它近似地反转整个G并计算估计z0 = E(x) ,然后解决一个优化问题以识别rgn (· · · (g1(z0)))生成重建图像Gf (r *)以密切恢复x。对于每一层gi ∈ {g1, ..., gn, Gf },首先训练一个小型网络ei来反转gi 。也就是说,当定义ri = gi(ri-1) 时,目标是学习一个网络ei ,它近似于计算ri-1 ≈ ei(ri)并确保网络ei的预测能够很好地保留层gi ,即ri ≈ gi(ei(ri))。因此,训练ei以最小化左右反转损失： LL = Ez[||ri-1 - e(gi(ri-1)||1], LR = Ez[||ri - gi (e(ri))||1], ei = arg min LL + λR LR,

$$\text{和} \tag{10}$$

在哪里|| · ||1表示L1损失， λR设置为0.01以强调ri-1的重构。为了专注于在生成器生成的表示流形附近进行训练，该方法使用样本z和层gi来计算ri-1和ri的样本，使得ri-1 = gi-1(· · · g1(z))。一旦所有层都被反转，所有G的反转网络可以组成如下：

$$和^{*} = e1 (e2 (\cdots (en (ef (x))))) 。 \quad (11)$$

可以通过微调组合网络E*来进一步改进结果，将G联合为一个整体，得到最终结果E。

对于 StyleGAN [13]、 [14]、 [71]，中间潜在向量w ∈ W+或s ∈ S在各层之间是不同的，并通过 AdaIN [77]或仿射变换[57]馈送到生成器的相应层.因此，将图像反转到W+或S空间可以看作是分层的。

4.3.4 分布外的泛化性 GAN 反演方法可以支持图

像的反演，尤其是任何给定的真实图像，这些图像不是由训练数据的相同过程生成的。我们将这种能力称为分布外泛化性[111]、 [112]、 [113]。

具体来说，给定一个在 FFHQ 数据集上预训练的 StyleGAN，该属性与以下两个方面密切相关：1)生成具有所有面部属性组合的面部图像，即使某些组合在训练数据集中不存在； 2)处理与训练集样本不同的图像，例如损坏的图像、漫画或黑白照片。此属性是 GAN 反演方法编辑更广泛图像的先决条件。分布外的泛化性已在许多 GAN 反演方法中得到证明。

朱等人。 [19]提出了一种特定领域的 GAN 反演方法来恢复像素和语义级别的输入图像。尽管仅使用 FFHQ 数据集进行训练，但他们的模型不仅可以推广到来自多个人脸数据集[114]、 [115]、 [116]的真实人脸图像，还可以推广到从互联网收集的绘画、漫画和黑白照片。康等人。 [64]提出了一种反转超出范围图像的方法。以面部图像为例，超出范围的图像可能是具有极端姿势的图像或损坏的图像，以前的方法通常无法处理。能够反转超出范围的图像允许 GAN 反转方法应用于更广泛的领域，而不是有限的设置。一些方法[22]、 [56]探索了将图像反转为所需潜在代码的潜力，只是给出了降级或部分观察。除了图像之外，最近的方法还显示了对其他模态的分布外泛化能力，即草图[28]、 [29]和文本[94]、[117]。

当与基于潜在代码的编辑方法（参见第 4.4 节） [90]、 [118]、[119]、 [120]结合时，GAN 反演的分布外泛化性有助于开放世界的图像处理。一个显着的缺点是，包含看不见的属性的反转图像很容易导致意外结果，因为它们位于预训练图像生成器的域之外。这限制了将 GAN 反演扩展到更广泛的应用，例如由不常见的文本描述引导的图像合成[117]。

最近的一些方法旨在通过将在一个图像域上预训练的 GAN 转移到一个新的域来缓解这个问题，由来自一个或几个目标图像[121] （few-shot 和 one-shot)的某些参考或语义引导，预训练语言-图像模型[122]（零镜头），或两者[123]。

4.4 潜在空间导航GAN 反演不是最终

目标。我们将真实图像反转到经过训练的 GAN 模型的潜在空间的原因是，它允许我们通过改变潜在空间中某个属性的反转代码来操纵图像。

这种技术通常被称为潜在空间导航或遍历[124]、 [125]、 GAN 可操纵性[17]、 [119]或潜在代码操作[18]。虽然通常被视为一个独立的研究领域，但它成为 GAN 反演[94]、 [126]不可或缺的应用。许多反演方法[30]、 [60]还探索了对所需潜在代码的有效发现。 4.1节介绍了不同的潜在空间。本节介绍在 GAN 的潜在空间中发现可预测和解开的方向。

4.4.1 发现可解释方向 一些方法支持发现潜在空间

中的可解释方向，即通过在所需方向n上以步长α 改变潜在代码z来控制生成过程，这被认为是向量算术= z + αn。这样的方向可以通过有监督、无监督或自我监督的方式来识别。最近的方法也被提出来直接从预训练模型计算封闭形式的可解释方向，而无需任何类型的训练或优化。[0]

监督设置。现有的基于监督学习的方法通常随机采样大量潜在代码，合成一组相应的图像，并通过引入预训练的分类器（例如，预测人脸属性或光线方向)用一些预定义的标签对它们进行注释[16]、 [17]、 [18]、[91]或提取统计图像信息（例如，颜色变化） [127]。例如，为了解释 GAN 学习的人脸表示，Shen 等人。 [18]使用一些现成的分类器来学习作为分离边界的潜在空间中的超平面并预测合成图像的语义分数。阿卜杜勒等人。 [91]通过使用连续归一化流 (CNF) 来学习Z空间和W空间之间的语义映射。

这两种方法都依赖于属性的可用性（通常通过人脸分类器网络获得），这对于新数据集可能难以获得，并且可能需要手动标记工作。

无监督设置。监督设置会在实验中引入偏差，因为用作监督的采样代码和合成图像在每次采样中都不同，并且可能导致可解释方向的不同发现[120]。它还严重限制了现有方法可以发现的一系列方向，尤其是在缺少标签的情况下。此外，通过这些方法发现的单个控件通常是纠缠的，影响多个属性，并且是十个非局部的。因此，一些方法[90]、 [125]、 [128]、[129]旨在发现潜在空间中的可解释方向

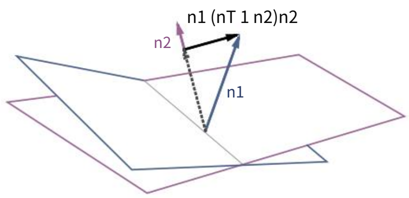


图 5. 发现多个属性的解开方向的图示。从n1中减去n1在n2上的投影,得到一个新方向n1 - (n> n2)n2。该图来自[18]。

1

以无监督的方式,即不需要配对数据。例如,哈克¨等人。[129]通过基于潜在或特征空间中应用的 PCA 识别重要的潜在方向,为图像合成创建可解释的控制。得到的应,主成分的选择性应用允许对许多图像属性进行控制。这种方法被认为是"无监督的",因为 PCA 可以在不使用任何标签的情况下发现方向。需要人工干预和监督才能将这些方向注释到目标操作以及它们应该应用于哪些层。相比之下,Jahanian 等人。[17]以自我监督的方式优化轨迹(线性和非线性)。以线性游走w为例,给定一个倒置的源图像G(z),他们学习w为

$$w^* = \arg\min_{在} E_{z,\alpha} [L (G (z + \alpha w), 编辑 (G (z), \alpha))], \quad (12)$$

其中L测量在潜在方向上采取α 步后生成的图像G(z + αw)与目标图像编辑 (G(z), α) 之间的距离。这种方法被认为是"自我监督的",因为目标图像(G(z), α) 可以从源图像G(z) 导出。

封闭式解决方案。最近的一些方法[119]、[120]、[130]、[131]表明,图像合成的可解释方向可以直接以封闭形式获得,无需训练或优化。沉等人。[120]提出了一种基于预训练GAN第一层权重的奇异值分解的语义分解方法。他们观察到图像的语义转换,通常表示为将潜在代码向某个方向n = z + αn 移动,实际上是由潜在方向n 决定的,而潜在方向 n 与采样代码z 无关。开发了语义分解 (SeFa) 方法来发现可能导致输出图像Δy发生显着变化的方向n ,即Δy = y - y = (A(z + αn) + b) - (Az + b) = αAn,其中A和b分别是G中某些层的权重和偏差。得到的公式Δy = αAn表明,可以通过将项αAn添加到投影代码上来实现所需的方向为 n的编辑,并表明权重参数A应该包含图像变化的基本知识。因此,探索潜在语义的问题可以通过解决以下优化问题来分解:

0

$$n^* = \arg\max_{\{n \in R^d: n^T n=1\}} ||安||_2^2 . \quad (13)$$

期望的方向n^*,即 GAN 中潜在语义的闭合形式分解,应该是矩阵AT A 的特征向量。与 SeFa [120] 相比,一种基于正交雅可比正则化的方法应用于生成器的多个层以确定可解释的图像合成方向[130]。

4.4.2 发现分离的方向 当涉及多个属性时,编辑一个可能会影响另一个,因为某些语义没有分离。一些方法旨在解决多属性图像处理而不受干扰。这一特性在文献中也被称为多维[89]或条件编辑[18] 。目标是发现所需属性的解开方向。例如,要编辑多个属性,Shen 等人。[18]制定基于反演的图像 ma = G(z + αn),其中n是单位法线 nipulation 作为x向量,表示由两个潜在代码z1和z1 定义的超平面。在该方法中, k个属性{z1, · · · , zk}可以形成m个(其中m≤k(k-1)/2)个超平面{n1, · · · , nm}。

0

为了在不相互干扰的情况下编辑多个属性,这些解开的方向{n1, · · · , nm}应该是正交的。如果这个条件不成立,那么一些语义就会相互关联,可用于测量两个语义之间的纠缠度。特别是,该方法使用投影来正交化不同的向量。如图5 所示,给定两个具有法线向量n1和n2的超平面,目标是找到投影方向n1 - (n 1 n2) n2 ,使得沿这个新方向移动的样本可以改变"朝贡方向"而不影响"属性二"。对于涉及多个属性的情况,它们从原始方向减去投影到由所有条件方向构成的平面上。基于预训练 StyleGAN [13]或StyleGAN2 [14]模型的其他 GAN 反演方法[54]也可以操纵多个属性,因为W空间的可分离性比Z空间的强。然而,正如最近的方法[57]、[132] 所观察到的,一些属性仍然纠缠在W空间中,当我们操纵给定图像时会导致一些不需要的语义变化。因此,需要在W空间中进行操作。[57]提出了S空间(风格空间)。

样式代码是通过连接 StyleGAN2 [14]生成器的所有仿射层的输出形成的。实验表明, S空间可以缓解空间纠缠变化并施加精确的局部修改。通过直接干预样式代码s ∈ S ,他们的方法可以在不影响其他人的情况下操纵不同的面部属性以及各种语义方向,并且可以实现对局部翻译的细粒度控制。

5应用

找到反演问题的准确解决方案使我们能够匹配目标图像,而不会影响下游任务中的编辑能力。版本中的 GAN 不需要特定任务的密集标记数据集,可以应用于许多任务,例如图像处理、图像插值、图像恢复、样式迁移、新视图合成,甚至对抗性防御。

除了常见的图像编辑应用程序外,在

过去几个月,GAN 反演技术已被广泛应用于许多其他任务,例如 3D 重建[133]、 [134]、图像理解[135]、 [136]、多模态学习[94]、 [117]、[132]、 [137]和医学成像[138]、 [139]、 [140],这表明它对不同任务的多功能性和有利于更大的研究社区的力量。

5.1 图像处理给定图像x,我们

希望通过改变其潜在代码z来编辑某些区域,并通过对训练过的 GAN 模型 G 的潜在表示进行线性变换来获得目标图像[illegible]。

$$x^{0} = G(z^{*} + \alpha n), \tag{14}$$

其中n是对应于潜在空间中特定语义的法线方向， α是操作的步骤。换句话说,如果一个潜在代码在某个方向上移动,那么输出图像中包含的语义应该会发生相应的变化。例如,Voynov 等人。 [90]在不改变前景的情况下逐渐确定对应于背景去除或背景模糊的方向。沉等人。 [18]通过投影和正交化不同的向量来实现单个和多个面部属性操作。最近,朱等人。 [19]通过降低或增加语义程度来执行语义操作。 [18]、 [19]两种方法都使用投影策略来搜索语义方向n。

一些方法可以执行感兴趣区域编辑,这允许通过用户操作来编辑给定图像中的一些所需区域。此类操作通常涉及选择所需区域的附加工具。

例如,阿卜杜勒等人。 [21]、 [22]分析了在 FFHQ [13]上训练的 StyleGAN 的缺陷图像嵌入,即带有遮罩区域的图像嵌入。实验表明,StyleGAN 嵌入对图像中的缺陷非常稳健,并且不同面部特征的嵌入相互独立[21]。

根据他们的观察,他们开发了一种基于掩模的局部操作方法。他们为掩码之外的区域找到了一个合理的嵌入,并在掩码像素中填充了合理的语义内容。朱等人。 [19]使用他们的域内反转方法进行语义扩散。此任务是将目标人脸插入到上下文中并使它们兼容。他们的方法可以在保留目标图像的显着特征（例如人脸身份）的同时适应上下文信息。

一些方法还可以处理除语义之外的图像,例如几何、纹理和颜色。

例如， [21]、 [91]改变姿势旋转以进行面部操作,而[90]可以操作几何（例如,缩放/移位/旋转）、纹理（例如,背景模糊/添加草/锐度)和颜色（例如,照明/饱和度）。

5.2 图像生成针对图像生成

任务提出了几种基于 GAN 反演的方法,例如发型转移[141]、少镜头语义图像合成[142]和无限分辨率图像合成[107]。萨哈等人。 [141]通过优化 StyleGAN2 [14] 的扩展潜在空间和噪声空间,开发了一种逼真的发型转移方法。远藤等人。 [142]假设共享相同语义的像素具有相似的 StyleGAN 特征,可以从潜在空间中的随机噪声生成图像和相应的伪语义掩码,并使用最近邻搜索进行合成。该方法将编码器与固定的 StyleGAN 生成器集成在一起,并以有监督的方式使用伪标记数据训练编码器以控制生成器。程等人。 [143]提出了一种基于 GAN 反转的图像修复和外画方法。一个以坐标为条件的生成器被设计用来合成要拼接成一个完整图像的补丁。潜在代码根据联合潜在代码及其坐标,合成与输入图像重叠的图像。可用输入补丁的最佳潜在代码在输出绘制阶段在经过训练的基于补丁的生成器的潜在空间中确定。 GAN 反演方法可以应用于交互式生成,即从用户绘制的笔画开始,生成最能满足用户约束的自然图像。朱等人。 [20]表明,用户可以使用画笔工具从头开始生成图像,然后不断添加更多涂鸦来改进结果。阿卜杜勒等人。 [22]反转 StyleGAN 以根据用户涂鸦执行语义本地编辑。使用这种方法,通过将简单的涂鸦嵌入到 StyleGAN 的某些层中,可以将其转换为逼真的编辑。该应用程序有助于现有的交互式图像处理任务,例如草图到图像生成[144]、 [145]、 [146]和基于草图的图像检索[147]、 [148],这些任务通常需要密集标记的数据集。

5.3 图像恢复假设在采集过

程中通过x ^ = ϕ(x)得到x^,其中x是无失真图像， ϕ是退化变换。许多图像恢复任务可以看作是在给定x^ 的情况下恢复x 。一种常见的做法是学习从x^到x 的映射,这通常需要针对不同的ϕ进行特定于任务的训练。或者,GAN 反演可以使用存储在某个先验中的x的统计数据,并通过将x视为x的部分观察值,在x 的空间中搜索与x最匹配的最优x 。例如,阿卜杜勒等人。 [21]、 [22]观察到 StyleGAN 嵌入对图像中的缺陷非常稳健,例如,被遮罩的区域。基于这一观察,他们提出了一种基于反转的图像修复方法,将源缺陷图像嵌入到W+空间的早期层中以预测丢失的内容,并嵌入到后面的层中以保持颜色一致性。潘等人。 [25]声称固定的 GAN 生成器不可避免地受到训练数据分布的限制,并且其反演不能忠实地重建看不见的复杂图像。因此,他们提出了一种轻松且更实用的重建公式,用于在训练的 GAN 模型中捕获自然图像的统计数据,就像之前的方法一样,即深度生成先验 (DGP)。

具体来说,他们重新制定（3） ,以便它允许在目标图像上动态微调生成器参数：

$$\theta^{*}, \text{和}^{*} = \arg\min \theta, \dot{z}(\hat{X}, \phi(G(z; \theta))。 \tag{15}$$

他们的方法在着色[149]、修复[150]和超分辨率[151]方面的性能与最先进的方法相当。虽然 GAN 模型[11]、 [13]合成的人脸图像中有时会出现伪影,但Shen 等人。 [18]表明在潜在空间中编码的质量信息可用于恢复。 PGGAN [11]产生的伪影可以通过使用线性 SVM [152]将潜在代码移向由分离超平面定义的正质量方向来纠正。

5.4 图像插值使用 GAN 反演,可以

通过在给定图像的相应潜在向量之间变形来插值新结果。给定一个训练有素的 GAN 生成器G和两个目标图像xA和xB,它们之间的变形自然可以通过在它们的潜在向量zA和zB 之间进行插值来实现。通常， xA和xB之间的变形可以通过应用线性插值来获得[6]、 [25]： $z = \lambda zA + (1 - \lambda)zB, \lambda \in (0, 1)$。

(16)

这种操作可以在[21]、 [89]中找到。此外,在 DGP [25] 中,重建两个目标图像xA和xB将分别产生两个生成器GθA和GθB ,以及相应的潜在向量zA和zB ,因为它们也微调G。在这种情况下， xA和xB可以通过潜在向量和生成器参数的线性插值来实现： $z = \lambda zA + (1 - \lambda)zB, \theta = \lambda\theta A + (1 - \lambda)\theta B, \lambda \in (0, 1)$,

(17)

并且可以使用新的z和θ 生成图像。

5.5 3D 重建

对于 3D 数据,Pan 等人。 [133]和张等人。 [134]提出了基于 GAN 反演的单张图像的 3D 形状重建和点云补全。给定由 GAN 生成的图像,从初始椭圆体 3D 对象形状开始,Pan 等人。 [133]首先使用各种随机采样的视点和光照条件（称为伪样本）渲染许多不自然的图像。通过使用 GAN 重建它们,这些伪样本可以将原始图像引导到 GAN 流形中的采样视点和照明条件,从而生成许多看起来自然的图像（称为投影样本）。这些投影样本可以用作可微渲染过程的基本事实,以细化先前的 3D 形状。 Zhang 等人没有使用现有的在图像上训练的 2D GAN。 [134]首先以点云的形式在 3D 形状上训练生成器G。预训练的生成器使用潜在代码来生成完整的形状。给定一个部分形状,他们寻找一个目标潜在向量z并微调G的参数θ ,以通过梯度下降最好地重建完整的形状。

5.6 图像理解一些方法利用训练过的

GAN 模型的表示并将这些表示用于语义分割和 alpha matting [135]、 [136]。特里特隆等人。 [135]首先将图像嵌入到潜在空间中 z并将其输入到具有多个激活图的生成器中。这些映射被上采样并沿通道维度连接以形成所需的表示。

使用一些手动注释的图像和提取的表示来训练分割模块。在推理过程中,从测试图像中提取表示并输入分割器以获得分割图。在[136] 中,两个预训练生成器、一个 alpha 网络和一个鉴别器用于抠图任务。一个生成器G(z)负责生成前景图像，另一个生成器Gbg (z)处理背景。 alpha 网络用于预测图像抠图的掩码A(z) G(z) 。合成图像可以通过使用A(z) G(z) + (1 − A(z)) Gbg (z)混合背景和前景来获得,判别器D无法将其与真实图像区分开来。在训练过程中,两个生成器被冻结，只有 alpha 网络和判别器通过对抗学习进行训练。

0

0

5.7 多模态学习对于多模态学习,最

近的几项研究集中在语言驱动的图像生成和使用 StyleGAN 的操作上。夏等人。 [132]通过训练编码器将文本映射到 StyleGAN 的潜在空间并执行样式混合以产生不同的结果,为文本到图像的生成和文本引导的图像处理任务提出了一种新颖的统一框架。在[137] 中， Wang 等人。提出了类似的想法,但在反演期间引入了循环一致性训练,以学习更稳健和一致的反演潜码。另一方面,一些方法[94]， [117]首先获得给定图像的潜在代码,并在一些强大的预训练语言模型的指导下找到所需属性的目标潜在代码,例如,CLIP [153]或对齐[154]。 Logacheva 等人。 [155]提出了一种基于 StyleGAN 反转的风景动画视频生成模型。李等人。 [156]提出了一种声音引导的图像编辑框架。他们训练音频编码器将声音编码到多模态潜在空间中,其中音频表示与文本图像表示对齐以指导图像处理。

5.8 医学成像GAN 反演技术最

近已被引入医学应用[157]。这些方法[138]、 [139]用于数据增强,其中公开可用的医学数据集通常已过时、有限或注释不充分。通常,这些方法在特定领域的医学图像数据集上训练 GAN 模型,例如计算机断层扫描 (CT) 或磁共振 (MR),并使用现有的 GAN 反演方法进行反演和操作。费蒂等人。 [139]提出了一种基于 Style GAN 模型[21]的方法,其中具有所需属性的 CT 或 MR 图像可以通过遍历潜在空间中的点（参见第4.4 节）或风格混合[13] 来合成。为了合成具有所需属性的大小医学图像,Ren 等人。 [138]使用特定领域的 GAN 反演技术[19]为心理物理实验生成具有所需形状和纹理的乳房 X 线照片。总的来说,这些基于GAN 反演的方法在医学图像合成中实现了更好的可解释性和可控性。

6挑战和未来方向理论理解。尽管在将 GAN 反演应用于图像编

辑应用程序方面已经做出了巨大努力,但很少关注对潜在空间的更好的理论理解。数据中的非线性结构可以紧凑地表示,并且诱导几何尝试需要使用非线性统计工具[158]、黎曼流形和局部线性方法。相关领域的成熟理论可以促进从不同角度进行理论理解。最近的一些方法[131]、 [159]将潜在空间视为流形结构,其中涉及不同的概念和度量。

反转类型。除了 GAN 反演之外,还开发了一些基于编码器-解码器架构的生成模型的反演方法。 IIN 方法[160]学习变分自动编码器 (VAE) [161] 的可逆解纠缠解释。朱等人。 [34]开发了潜在的可逆自动编码器来学习面部图像的解耦表示,从中可以根据属性编辑内容。 LaDDer 方法[162]使用基于生成先验(包括加性 VAE 和超先验混合)的元嵌入将训练有素的 VAE 的潜在空间投影到低维潜在空间,其中多个 VAE 模型用于形成分层表示。探索结合 GAN 反转和编码器-解码器反转是有益的,这样我们就可以利用两全其美。

领域泛化。如第5 节所述,GAN 反转被证明在跨域应用中是有效的,例如样式迁移和图像恢复,这表明预训练模型已经学习了与域无关的特征。来自不同领域的图像可以倒置到相同的潜在空间中,从中可以得出有效的度量。已经开发出多任务方法来协同利用视觉线索,例如在 GAN 框架内的图像恢复和图像分割[163]或语义分割和深度估计[164]、 [165] 。开发有效且一致的方法来反转中间共享表示是具有挑战性但值得的,这样我们就可以在统一的框架下处理不同的视觉任务。

隐式表示。一些基于预训练 GAN 的方法[90]、 [91]可以操纵几何(例如,缩放、移位和旋转)、纹理(例如,背景模糊和锐度)和颜色(例如,照明和饱和度)。这种能力表明在大规模数据集上预训练的 GAN 模型已经从现实世界场景中学习了一些物理信息。隐式神经表示学习[166]、[167]、 [168]是 3D 计算机视觉的最新趋势,它学习 3D 形状或场景的隐式函数,并能够控制场景属性,例如照明、相机参数、姿势、几何、外观和语义结构。它已被用于体积性能捕获[169]、 [170]、 [171]、新视图合成[172]、 [172]、面部形状生成[173]、对象建模[174]和人体重建[175]、[176]、 [177]、 [178]。最近的 StyleRig 方法[102]被训练以将 3D 可变形模型 (3DMM) [179]的参数与 Style GAN [13] 的输入对齐。它打开了一个有趣的研究方向来反转预训练 GAN 的隐式表示以进行 3D

重建,例如,使用 StyleGAN [13]进行人脸建模或延时视频生成。

精确控制。 GAN 反演可用于查找图像处理的方向,同时保留身份和其他属性[18]、 [91]。但是,需要进行一些调整才能在细粒度级别上实现所需的精确控制粒度,例如,凝视重定向[6]、 [180]、重新照明[181]、[182]、 [183] 和连续视图控制[184]。这些任务需要精确控制,即摄像机视图或注视方向的1 。当前的 GAN 反演方法无法处理这些任务。因此需要更多的努力,例如创造更多解开的潜在空间和发现更多可解释的方向。 。

多模式倒置。现有的 GAN 反演方法主要集中在图像上。然而,生成模型的最新进展超出了图像领域,例如用于音频合成的 GPT-3 语言模型[185]和 WaveNet [186] 。这些复杂的深度神经网络经过各种大规模数据集的训练,被证明能够代表广泛的不同内容、风格、情感和主题。在这些不同的模态上应用 GAN 反演技术可以为语言风格迁移等任务提供新的视角。

此外,许多 GAN 模型是为多模态生成或翻译而开发的[187]、 [188]、[189]。将这种 GAN 模型反转为多模态表示以创建新颖的内容、行为和交互是一个很有前途的方向。

评估指标。新的感知质量指标,可以更好地评估逼真和多样化的图像或与原始图像一致的身份,仍有待探索。当前的评估主要集中在测量真实感,或者生成图像的分布是否与使用真实图像训练的模型在分类[23]或分割[90]准确性方面与真实图像一致。然而,仍然缺乏有效的评估工具来评估预测结果与预期结果之间的差异或更直接地测量倒置潜码。

7结论

GAN 等深度生成模型通过对图像生成的弱监督来学习训练数据的潜在变化因素。在图像生成中发现和引导可解释的潜在表示有助于广泛的图像编辑应用程序。本文全面介绍了 GAN 反演方法,重点是算法和应用。我们总结了 GAN 潜在空间和模型的重要性质,然后介绍了四种 GAN 反演方法及其关键性质。然后,我们介绍了 GAN 反演的几个引人入胜的应用,包括图像处理、图像生成、图像恢复以及图像处理之外的最新应用。我们最后讨论了 GAN 反演的挑战和未来方向。

参考

[redacted]

[55] R. Anirudh、JJ Thiagarajan、B. Kailkhura 和 P. Bremer，“Mim icGAN：具有腐败模仿的图像流形上的鲁棒投影”，IJCV，第一卷。 128，没有。 10，pp. 2459–2477，2020。4 [56] L. Chai、J. Wulff 和 P. Isola，“使用潜在空间回归分析和利用 GAN 中的组合性”，ICLR，2021年。4、9

[57] Z. Wu、D. Lischinski 和 E. Shechtman，“StyleSpace 分析：StyleGAN 图像生成的分离控制”，CVPR，2021。 4、5、9、10 [58] Y. Xu、Y. Shen 、J. Zhu、C. Yang 和 B. Zhou，“通过合成图像生成层次特征”，CVPR，2021 年。 4

[59] L. Chai，J.-Y。 Zhu、E. Shechtman、P. Isola 和 R. Zhang，“具有深刻生成观点的集合”。在 CVPR，2021年。4 [60] O. Tov、Y. Alaluf、Y. Nitzan、O. Patashnik 和 D. Cohen-Or，“为 StyleGAN 图像处理设计编码器”，TOG，2021 年。4、5 、6、9 [61] Y. Xu、Y. Du、W. Xiao、X. Xu 和 S. He，“从连续性到可编辑性：用连续图像反转 GAN”，ICCV，2021。 4

[62] SR Livingstone 和 FA Russo，“情感语音和歌曲 (ravdess) 的 ryerson 视听数据库：北美英语中面部和声音表达的动态、多模态集”，PloS one，vol。 13，没有。第 5 页。 e0196391，2018。 4 [63] Y. Choi、Y. Uh、J. Yoo 和 J.-W。 Ha，“StarGAN v2：多域的多样化图像合成”，CVPR，2020 年。 4 [64] K. Kang、S. Kim 和 S. Cho，“具有几何变换的超范围图像的 GAN 反演，” in ICCV，2021。 4、9 [65] P. Zhu、R. Abdal、Y. Qin、J. Femiani 和 P. Wonka，“我证明了 StyleGAN 嵌入：好的潜伏者在哪里？” arXiv 预印本 arXiv:2012.09036，2020。 4、5 [66] D. Roich, R. Mokady, AH Bermano 和 D. Cohen-Or，“基于潜在的真实图像编辑的关键调整”，arXiv 预印本 arXiv:2106.05744 , 2021。 4、7

[67] Y. Alaluf、O. Tov、R. Mokady、R. Gal 和 AH Bermano，“Hy perstyle：Stylegan inversion with hypernetworks for real image editing”，CVPR，2022。 4、5、7

[68] J. Krause、M. Stark、J. Deng 和 L. Fei-Fei，“用于细粒度分类的 3d 对象表示”，ICCV 研讨会，2013 年。4

[69] T. Wang、Y. Zhang、Y. Fan、J. Wang 和 Q. Chen，“用于图像属性编辑的高保真 gan 反演”，CVPR，2022年。4 [70] TM Dinh，AT Tran , R. Nguyen 和 B.-S。 Hua，“Hyperinverter：通过超网络改进 stylegan 反转”，CVPR，2022 年。 4

[71] T. Karras、M. Aittala、J. Hellsten、S. Laine、J. Lehtinen 和 T. Aila，“使用有限数据训练生成对抗网络”，NeurIPS，2020。 3、5、6、9 [72] T. Karras、M. Aittala、S. Laine、E. Hark ¨ onen、J. Hellsten、J. Lehtinen 和 T. Aila，“无别名生成对抗网络”，NeurIPS，2021 年。 3、5、6 [73] Yan 和 H. Zhang，“使用生成对抗网络实现高质量动漫角色生成”，NeurIPS 研讨会，2017 年。3 [74] M.-Y. Liu、X. Huang、A. Mallya、T. Karras、T. Aila、J. Lehtinen 和 J. Kautz.，“Few-shot unsueprvised image-to-image translation”，ICCV，2019。 3 [75] J. Gu、L. Liu、P. Wang 和 C. Theobalt，“Stylenerf：用于高分辨率图像合成的基于样式的 3d 感知生成器”，ICLR，2022年。3 [76] ER Chan、M. Monteiro 、P. Kellnhofer、J. Wu 和 G. Wetzstein，“pi-GAN：用于 3d 感知图像合成的周期性隐式生成对抗网络”，CVPR，2021年。3 [77] X. Huang 和 S. Belongie，“具有自适应实例规范化的实时任意风格迁移”，ICCV，2017。 3、5、9 [78] C.-H。 Lee、Z. Liu、L. Wu 和 P. Luo，“MaskGAN：迈向多样化和交互式面部图像处理”，CVPR，2020。3

[79] Z. Liu、P. Luo、S. Qiu、X. Wang 和 X. Tang，“深度时尚：用丰富的符号支持强大的衣服识别和检索”，CVPR，2016 年，第 1096-1104 页。 3 [80] N. Naik、J. Philipoom、R. Raskar 和 C. Hidalgo，“Streetcore 预测一百万街景的感知安全性”，CVPR 研讨会，2014 年，第 779-785 页。 3 [81] J. Rabin、G. Peyre、J. Delon 和 M. Bernot，“Wasserstein 重心及其在纹理混合中的应用”，SSVM， 2011。4

[82] Q. Xu、G. Huang、Y. Yuan、C. Guo、Y. Sun、F. Wu 和 K. Wein berger，“关于生成对抗网络评估指标的实证研究”，arXiv 预印本 arXiv： 1806.07755，2018。 4 [83] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford 和 X. Chen，“训练 GAN 的改进技术”，NeurIPS，2016。 4 [84] C. Szegedy、V. Vanhoucke、S. Ioffe、J. Shlens 和 Z. Wojna，“重新思考计算机视觉的初始架构”，CVPR，2016 年，第 2818-2826 页。 4 [85] J. Deng, W. Dong, R. Socher, L.-J. Li、K. Li 和 L. Fei-Fei，“Imagenet：大型分层图像数据库”，载于 CVPR，2009 年，第 248-255 页。 4 [86] M. Heusel、H. Ramsauer、T. Unterthiner、B. Nessler 和 S. Hochreiter，“通过两个时间尺度更新规则训练的 GAN 收敛到局部纳什均衡”，NeurIPS，2017 年。4 [87] K. Simonyan 和 A. Zisserman，“用于大规模图像识别的非常深的卷积网络”，ICLR，2015 年。 4 [88] Z. Wang、AC Bovik、HR Sheikh 和 EP Simoncelli，“图像质量评估：从错误可见性到结构相似性，”

提示，卷。 2004 年 13 月4 日

[89] Y. Nitzan、A. Bermano、Y. Li 和 D. Cohen-Or，“通过潜在空间映射解开人脸身份”，TOG，卷。 39，第 1-14 页， 2020年。4、10、12

[90] A. Voynov 和 A. Babenko，“GAN 潜在空间中可预测方向的无监督发现”， ICML， 2020年。4、9、11、13

[91] R. Abdal、P. Zhu、N. Mitra 和 P. Wonka，“StyleFlow：使用条件连续归一化流对 StyleGAN 生成的图像进行属性条件探索”，TOG， 2021年。4、5、9、11 、 13

[92] Q. Bai、Y. Xu、J. Zhu、W. Xia、Y. Yang 和 Y. Shen，“具有填充空间的高保真 gan 反转”，arXiv 预印本 arXiv:2203.11105，2022。 5 [93] J. Xu、H. Xu、B. Ni、X. Yang、X. Wang 和 T. Darrell，“基于层次风格的运动合成网络”，ECCV，2020 年。5 [94] O. Patashnik，Z . Wu、E. Shechtman、D. Cohen-Or 和 D. Lischin ski，“StyleCLIP： StyleGAN图像的文本驱动操作”，ICCV， 2021年5、9、11、12

[95] D. Bau，J.-Y。 Zhu、J. Wulff、W. Peebles、H. Strobelt、B. Zhou 和 A. Torralba，“大型发电机的反相层”，ICLR 研讨会，第一卷。 2，没有。 2019 年 3 月，第 3 页。 4. 6, 7 [96] T.-Y. Lin、P. Dollar、R. Girshick、K. He、B. Hariharan 和 S. Belongie，“用于对象检测的特征金字塔网络”，CVPR，2017，第 2117-2125 页。 6 [97] S. Pidhorskyi、DA Adjeroh 和 G. Doretto，“对抗性潜在

自动编码器”，CVPR，2020 年。6

[98] D. Kingma 和 J. Ba，“Adam：一种随机优化方法”，ICLR，2015 年。 7 [99] DC Liu 和 J. Nocedal，“关于大规模优化的有限内存 BFGS 方法，” 数学编程，1989 年。7 [100] S. Duane、AD Kennedy、BJ Pendleton 和 D. Roweth，“混合蒙特卡罗”，物理字母 B，1987 年。7 [101] N. Hansen 和 A. Ostermeier，“进化策略中完全去随机化的自我适应。进化计算，”

进化计算，2001。 7

[102] A. Tewari、M. Elgharib、G. Bharaj、F. Bernard、H.-P。 Seidel、olhofer 和 GAN 用于对肖像图像进行 3D 控制”，CVPR，2020。 3 Style-P. Perez, M. Z

13

[103] D. Ha、A. Dai 和 QV Le，“超网络”，ICLR，2016 年。7 [104] D. Bau，J.-Y。朱，H. Strobelt，B. Zhou，JB Tenenbaum，WT Freeman 和 A. Torralba，“GAN 解剖：可视化和理解生成对抗网络”，ICLR，2019 年。 7

[105] Y. Zhu、Q. Li、J. Wang、C. Xu 和 Z. Sun，“One shot face swapping on megapixels”，CVPR，2021年。8 [106] Q. Bai、W. Xia， F. Yin 和 Y. Yang，“具有多模态轮廓条件的身份引导人脸生成”，arXiv 预印本 arXiv:2110.04854，2021。 8 [107] CH Lin, Y.-C.程，H.-Y。 Lee、S. Tulyakov 和 M.-H. Yang，“InfinityGAN：迈向无限分辨率图像合成”，ICLR，2022 年。8、11 [108] Q. Lei、A. Jalal、IS Dhillon 和 AG Dimakis，“一次一层反转深度生成模型，” in NeurIPS, 2019. 8 [109] V. Nair 和 GE Hinton，“整流线性单元改进了受限的 Boltzmann 机”，ICML，2010。 8

[110] AL Maas、AY Hannun 和 AY Ng，“整流器非线性改进神经网络声学模型”，ICML，2013 年。8 [111] J. Ren，PJ Liu，E. Fertig，J. Snoek，R. Poplin，M. Depristo，J. Dillon 和 B. Lakshminarayanan，“分布外检测的似然比”，NeurIPS，2019 年，第 14 707–14 718 页。9 [112] D. Hendrycks 和 K. Gimpel，“在神经网络中检测错误分类和分布外示例的基线”，ICLR，2017年。 9

[113] K. Lee、K. Lee、H. Lee 和 J. Shin，“用于检测分布外样本和对抗性攻击的简单统一框架”，NeurIPS，2018 年，第 7167-7177 页。 9 [114] O. Chelnokova、B. Laeng、M. Eikemo、J. Riegels、G. Løseth、H. Maurud、F. Willoch 和 S. Leknes，“美丽的回报：阿片类药物系统介导人类的社会动机，”分子精神病学，第一卷。 19，没有。 7，pp. 746–747，2014. 9 [115] R. Courset，M. Rougier，R. Palluel-Germain，A. Smeding，JM

Jonte、A. Chauvin 和 D. Muller，“高加索和北非法国面孔 (CaNAFF)：面孔数据库”，《国际社会心理学评论》，卷。 31，没有。 1，2018. 9 [116] R. Yi，Y.-J.刘，Y.-K。 Lai 和 PL Rosin，“ApdrawingGAN：使用分层 GAN 从面部照片生成艺术肖像画”，CVPR，2019 年，第 10 743-10 752 页。9

[117] W. Xia，Y. Yang，J.-H. Xue 和 B. Wu，“走向开放世界的文本引导人脸图像生成和操作”，arxiv 预印本 arxiv：2104.08910，2021. 9，11，12

[118] Y. Han，J. Yang 和 Y. Fu，“通过实例感知潜在空间搜索进行分离的人脸属性编辑”，IJCAI，2021 . 9 [119] S. Nurit、B. Ron 和 M. Tomer，“没有

优化”，ICLR，2021. 9，10

[120] Y. Shen 和 B. Zhou，“潜在语义的闭式分解 GAN 中的抽动”，CVPR，2021. 9，10

[121] C. Yang、Y. Shen、Z. Zhang、Y. Xu、J. Zhu、Z. Wu 和 B. Zhou，“一次性生成域适应”，arXiv 预印本 arXiv:2111.09876，2021. 9

[122] R. Gal、O. Patashnik、H. Maron、G. Chechik 和 D. Cohen-Or，“StyleGAN-NADA：图像生成器的 CLIP 引导域自适应”，arXiv 预印本 arXiv:2108.00946，2021. 9 [123] P. Zhu、R. Abdal、J. Femiani 和 P. Wonka，“注意差距：用于生成对抗网络的单次域自适应的域间隙控制”，ICLR，2022年。9 [124] P . Zhuang、O. Koyejo 和 AG Schwing，“享受您的编辑：通过潜在空间导航进行图像编辑的可控 GAN”，ICLR，2021年。9 [125] A. Cherepkov、A. Voynov 和 A. Babenko，“导航 GAN 参数空间进行语义图像编辑”，载于 CVPR，2021年。9 [126] Y. Alaluf、O. Patashnik 和 D. Cohen-Or，“只是风格问题：使用基于风格的年龄转换回归模型，”TOG， 2021。9

[127] A. Plumerault、HL Borgne 和 C. Hudelot，“控制具有连续变化因素的生成模型”，ICLR，2020 年。9

[128] Y.-D。卢，H.-Y。李，H.-Y。曾和 M.-H.杨，“GAN 中解开流形的无监督发现”，arXiv 预印本 arXiv:2011.11842，2020. 9 [129] H. Erik、H. Aaron、L. Jaakko 和 P. Sylvain，“GANSpace：Discoverable GAN 控制，” in NeurIPS，2020. 9，10 [130] Y. Wei，Y. Shi，X. Liu，Z. Ji，Y. Gao，Z. Wu，and W. Zuo，“Orthogon nal Jacobian regularization for unsupervised disentanglement in图像生成，”在 ICCV，2021年。10 [131] J. Zhu、R. Feng、Y. Shen、D. Zhao、Z. Zha、J. Zhou 和 Q. Chen，“GAN 中的低秩子空间，” in NeurIPS，2021. 10，13 [132] W. Xia，Y. Yang，J.-H. Xue 和 B. Wu，“TediGAN：文本引导的多样化图像生成和操作”，CVPR，2021，第 2256-2265 页。 10，11，12 [133] X. Pan，B. Dai，Z. Liu，CC Loy 和 P. Luo，“2D GAN 知道 3D 形状吗？来自 2D 图像 GAN 的无监督 3D 形状重建”，ICLR，2021. 11，12 [134] J. Zhang，X. Chen，Z. Cai，L. Pan，H. Zhao，S. Yi，CK Yeo，B. Dai 和 CC Loy，“通过 GAN 反转完成无监督 3D 形状”，CVPR，2021年。11、12 [135] N. Tritrong、P. Rewatbowornwong 和 S. Suwajanakorn，“将 GAN 重新用于一次性语义部分细分”，CVPR，2021. 11，12

[136] R. Abdal、P. Zhu、N. Mitra 和 P. Wonka，“Labels4Free：使用 StyleGAN 的无监督分割”，ICCV，2021. 11，12

[137] H. Wang、G. Lin、SCH Hoi 和 C. Miao，“用于文本到图像合成的循环一致逆 GAN”，ACM MM，2021. 11，12

[138] Z. Ren、SX Yu 和 D. Whitney，“通过生成对抗网络生成可控医学图像”，人类视觉和电子成像，2021 . 11，12

[139] L. Fetty、M. Bylund、P. Kuess、G. Heilemann、T. Nyholm、D. Georg 和 T. Lofstedt，“通过 StyleGAN 进行高分辨率医学图像合成的潜在空间操作”，Zeitschrift f ¨ur Medizinische Physik，第一卷。 30，没有。 4，pp. 305–314，2020. 11，12 [140] GB Daroach、JA Yoder、KA Iczkowski 和 PS LaViolette，“使用 StyleGAN 进行高分辨率可控前列腺组织学合成”，生物成像，2021. 11 [141] R. Saha、B. Duke、F. Shkurti、G. Taylor 和 P. Aarabi，“Loho：通过正交化优化发型”，CVPR，2021. 11

[142] Y. Endo 和 Y. Kanamori，“使用 StyleGAN 先验进行少量语义图像合成”，arXiv 预印本 arXiv：2103.14877，2021. 11 [143] Y.-C. Cheng，CH Lin，H.-Y. Lee，J. Ren，S. Tulyakov 和 M.-H。
杨，“In&out：通过 GAN 反转进行多样化图像外绘”，CVPR，2022年 11 月

[144] W. Xia、Y. Yang 和 J.-H。薛，“Cali-sketch：笔画校准和完成，用于从画得不好的草图中生成高质量人脸图像”，神经计算，2021 . 11 [145] A. Ghosh，R. Zhang，PK Dokania，O. Wang，AA Efros，小灵通

Torr 和 E. Shechtman，“交互式草图和填充：多类草图到图像的转换”，ICCV，2019年。11 [146] S.-Y。 Chen、W. Su、L. Gao、S. Xia 和 H. Fu，“DeepFaceDrawing：从草图中深度生成人脸图像”，TOG，第一卷。 39，没有。 4，pp. 72–1，2020. 11 [147] M. Eitz、K. Hildebrand、T. Boubekeur 和 M. Alexa，“基于草图的图像检索：基准和特征袋描述器”，TVCG，卷。 17，没有。 11，pp. 1624–1636，2010. 11 [148] S. Dey、P. Riba、A. Dutta、J. Llados 和 Y.-Z。 Song，“Doodle to search：基于零样本草图的实用图像检索”，载于 CVPR，2019 年，第 2179-2188 页。 11 [149] G. Larsson、M. Maire 和 G. Shakhnarovich，“自动着色的学习表示”，ECCV，2016 年。 12 [150] D. Ulyanov、A. Vedaldi 和 V. Lempitsky，“Deep图像优先，”

在 CVPR，2018 年，第 9446-9454 页。 12

[151] TR Shaham、T. Dekel 和 T. Michaeli，“SinGAN：从单个自然图像中学习生成模型”，ICCV，2019 年，第 4570-4580 页。 12 [152] C. Cortes 和 V. Vapnik，“支持向量网络”，机器学习，第一卷。 20，没有。 3，pp. 273–297，1995. 12 [153] A. Radford，JW Kim，C. Hallacy，A. Ramesh，G. Goh，S. Agar wal，G. Sastry，A. Askell，P. Mishkin，J. Clark 等人，“从自然语言监督中学习可迁移的视觉模型”，ICML，2021年。12

[154] C.贾，Y.杨，Y.夏，Y.-T。 Chen，Z. Parekh，H. Pham，Q.V. Le，Y. Sung，Z. Li 和 T. Duerig，“通过嘈杂的文本监督扩大视觉和视觉语言表示学习”，ICML，2021年。12 [155] E. Logacheva，R. Suvorov，O. Khomenko，A. Mashikhin 和 V. Lempitsky，“Deeplandscape：景观视频的对抗性建模”，ECCV，2020 年，第 256-272 页。 12 [156] SH Lee、W. Roh、W. Byeon、SH Yoon、CY Kim、J. Kim 和 S. Kim，“声音引导语义图像处理”，CVPR，2022年。12

[157] X. Yi、E. Walia 和 P. Babyn，“医学成像中的生成对抗网络：综述”，医学图像分析，2019 . 12 [158] L. Kuhnel、T. Fletcher、S. Joshi，和 S. Sommer，“潜在空间非线性统计”，arXiv 预印本 arXiv:1805.07632，2018 年。13 [159] J. Choi、C. Yoon、J. Lee、JH Park、G. Hwang 和 M. Kang，“不要从流形中逃脱：发现 GAN 潜在空间上的局部坐标”，ICLR，2022年。13 [160] P. Esser、R. Rombach 和 B. Ommer，“用于解释的解开可逆解释网络潜在表示”，CVPR，2020 年。13 [161] DP Kingma 和 M. Welling，“自动编码变分贝叶斯”，ICLR，2013 年。13 [162] S. Lin 和 R. Clark，“LaDDer：潜在数据具有生成先验的分布建模”，BMVC，2020 年。13 [163] W. Xia、Z. Cheng、Y. Yang 和 J.-H。薛，“不利环境条件下的协同语义分割和图像恢复”，arXiv预印本arXiv :1911.00679，2019.13

[164] V. Nekrasov、T. Dharmasiri、A. Spek、T. Drummond、C. Shen 和 ID Reid，“使用不对称注释的实时联合语义分割和深度估计”，ICRA，2019. 13 [165] W. Zhan、X. Ou、Y. Yang 和 L. Chen，“Dsnet：用于场景分割和视差估计的联合学习”，ICRA，2019。13 [166] Z. Chen 和 H. Zhang，“学习生成形状建模的隐式场”，CVPR，2019 年，第 5939-5948 页。 13 [167] R. Tucker 和 N. Snavely，“多平面图像的单视图合成”，CVPR，2020，第 551-560 页。 13 [168] S. Rajeswar、F. Mannan、F. Golemo、J. Parent-Levesque、D. Vazquez、D. Nowrouzezahrai 和 A. Courville，“Pix2shape：使用视图从图像中实现 3d 场景的无监督学习-基于表示，”IJCV，第 1-16 页，2020 年。13 [169] A. Chen、R. Liu、L. Xie 和 J. Yu，“SofGAN：具有动态样式的肖像图像生成器”，TOG，2021 . 13 [170] L. Liu、W. Xu、M. Habermann、M. Zollhoefer、F. Bernard、H. Kim、W. Wang 和 C. Theobalt，“通过学习动态纹理和渲染的神经人类视频渲染-到视频翻译，”

TVCG，2020 年。
13 [171] S. Lombardi、T. Simon、J. Saragih、G. Schwartz、A. Lehrmann 和 Y. Sheikh，“神经体积：从图像中学习动态可渲染体积”，TOG，2019年。13 [172] R. Martin-Brualla、N. Radwan、MS Sajjadi、JT Barron、A. Doso vitskiy 和 D. Duckworth，“野外 Nerf：无限制照片集的神经辐射场”，CVPR，2021 年。13 [173] S. Wu、C. Rupprecht 和 A. Vedaldi，“从野外图像中对可能对称的可变形 3d 对象进行无监督学习”，CVPR，2020 年，第 1-10 页。 13

[174] T. Nguyen-Phuoc、C. Richardt、L. Mai、Y.-L。 Yang 和 N. Mi tra，“BlockGAN：从未标记的图像中学习 3D 对象感知场景表示”，NeurIPS，2020 年。13 [175] Z. Zheng、T. Yu、Y. Liu 和 Q. Dai，“ Pamir：基于图像的人体重建的参数模型条件隐式表示，”TPAMI，2021年。13

[176] BL Bhatnagar、C. Sminchisescu、C. Theobalt 和 G. Pons-Moll，“结合隐式函数学习和参数模型进行 3d 人体重建”，ECCV，2020 年。13 [177] T. He, J. Collomosse、H. Jin 和 S. Soatto，“Geo-pifu：用于单视图人体重建的几何尝试和像素对齐隐式函数”，NeurIPS，2020 年。13

[178] S. Saito、T. Simon、J. Saragih 和 H. Joo，“Pifuhd：高分辨率 3D 人体数字化的多级像素对齐隐式函数”，CVPR，2020 年，第 84-93 页. 13 [179] B. Egger、WA Smith、A. Tewari、S. Wuhrer、M. Zollhoefer、T. Beeler、F. Bernard、T. Bolkart、A. Kortylewski、S. Romdhani 等人，“3D 变形面模型 过去、现在和未来，”

TOG，卷。 39，没有。 5, pp. 1–38, 2020.
13 [180] Z. He, A. Spurr, X. Zhang 和 O. Hilliges，“使用生成对抗网络的逼真的单目注视重定向”，ICCV，2019. 13 [181] H. Zhou、S. Hadap、K. Sunkavalli 和 DW Jacobs，“深度单图像人像重新照明”，ICCV，2019年。13 [182] X. Zhang，JT Barron， Y.-T。 Tsai、R. Pandey、X. Zhang、R. Ng 和 DE Jacobs，“人像阴影操纵”，TOG，2020 年。13 [183] T. Sun，JT Barron，Y.-T。 Tsai、Z. Xu、X. Yu、G. Fyffe、C. Rhemann、J. Busch、PE Debevec 和 R. Ramamoorthi，“单幅肖像重新照明”，TOG，2019. 13

[184] X. Chen、J. Song 和 O. Hilliges，“基于单目神经图像的连续视图控制渲染”，ICCV，2019年。13 [185] TB Brown，B. Mann，N. Ryder，M. Subbiah, J. Kaplan, P. Dhari wal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell 等人，“语言模型是少数学习者”，NeurIPS，2020. 13 [186] A. vd Oord、S. Dieleman、H. Zen、K. Simonyan、O. Vinyals、A. Graves、N. Kalchbrenner、A. Senior 和 K. Kavukcuoglu，“WaveNet：原始音频的生成模型”，arXiv预印本 arXiv:1609.03499, 2016. 13 [187] B. Li, X. Qi, T. Lukasiewicz 和 PHS Torr，“可控文本到图像生成”，NeurIPS，2019. 13 [188] Y. Jia, Y . Zhang、RJ Weiss、Q. Wang、J. Shen、F. Ren、Z. Chen、P. Nguyen、R. Pang、I. Lopez-Moreno 和 Y. Wu，“将学习从说话人验证转移到多说话人文本语音合成”，NeurIPS，2018 年，第 4485-4495 页。 13 [189] KR Prajwal、R. Mukhopadhyay、VP Namboodiri 和 C. Jawa har，“学习个人说话风格以实现准确的唇语合成”，CVPR，2020 年。13