arXiv:2304.03164v2 [cs.CV] 5 Nov 2023

合成任何人、在任何地点、以任何姿势呈现的图像

H˚akonHukkel˚as Frank Lindseth 挪威
科技大学hakon.hukkelas@ntnu.no

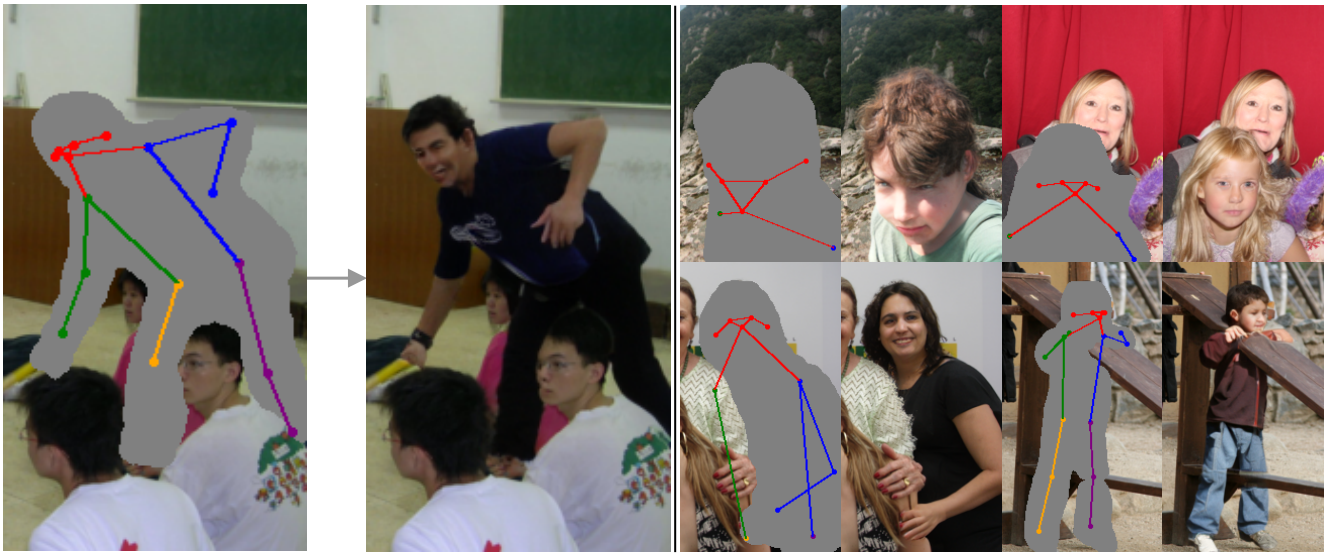


图1。TriA-GAN 可以根据一张带掩码的图像和一组稀疏的关键点合成逼真的人体。

摘要

我们研究野外场景中的人体合成任务，其主要目标是给定任意图像中的任意区域，合成一个完整人体。野外场景中的人体合成长期以来一直是一项具有挑战性且探索不足的任务，当前方法难以处理极端姿势、遮挡物以及复杂背景。

我们的主要贡献是 *TriA-GAN*，这是一种关键点引导*GAN*，能够以任意姿势合成任何人、身处任何地点。我们方法的关键在于将 *Projected GAN* 与精心设计的训练策略相结合，使简单的生成器架构也能成功应对真实场景全身合成的挑战。我们证明，*TriA-GAN* 显著优于以往的真实场景全身合成方法，同时合成所需的条件信息更少（关键点与*DensePose*）。最后，我们还证明，*TriA-GAN* 的潜在空间兼容标准的无条件编辑技术，因此能够通过文本引导编辑生成的人体形象。

1. 引言

给定一张存在缺失区域的图像，你能想象出一个适合填入其中的人体外观吗？如果缺失区域旁边有一个足球，你想象中的人会随之改变吗？这是一个引人入胜且颇具挑战性的问题，因为有无数种可能的解都符合上下文。我们将这一任务称为真实场景人体合成。要解决这个问题，系统必须深入理解人体外观，以及人体外观如何根据不同的环境条件、视角、姿势和缺失区域大小而变化。这样的系统在内容创作、时尚 [37], 乃至匿名化方面 [18]都有广泛的应用。

人体合成是一个成熟的研究领域，拥有许多高层次目标。然而，真实场景中的 人体合成是一项困难且尚未得到充分探索的任务。以往的方法聚焦于更简单的任务，例如将已知外观迁移到给定姿势中 [2, 4], 迁移服装 [14, 52], ，或在纯色背景中进行全身合成 [9]。它们往往忽视了真实场景合成的关键难点，例如物体重叠、部-

分人体、复杂背景和极端姿势。事实上，近期研究会从数据集中剔除这些困难案例，以提高合成质量 [9, 10]。据我们所知，只有少数研究应对过这些挑战，重点关注用于匿名化的全身合-成 [18, 20] [1]。虽然以往的方-法 [18]能够生成视觉效果出色的结果，但它们严重依赖 DensePose 估计，在复杂场景中表现不佳。此外，生成的图像难以编辑 [18]。

当前真实场景人体合成方法的一个关键问题是依赖 DensePose 标注[18,20]。带有此类标注的数据集要么规模有限 [12,20]，要么是通过自动方式标注的 [18]。我们认为，这种依赖会限制这些方法：它们要么在小型数据集上过拟合 [20]，要么受到 DensePose 标注 [18]产生的大量错误影响。

本文探讨以稀疏二维关键点为条件的全身合成，从而无需昂贵的 DensePose 标注。然而，这显著增加了建模复杂度，因为生成模型现在必须同时推断人体的纹理和结构。我们发现，当前的 GAN [18]在缺乏 DensePose 对应关系的情况下，难以合成逼真的人体。

我们的贡献旨在解决如何扩展 GAN、使其能够在缺乏 DensePose 对应关系的情况下处理真实场景全身合成这一挑战。我们方法的关键在于用 Projected GAN [53] 取代传统的 GAN 判别器。通过将 Projected GAN 与精心设计的训练策略相结合，我们的方法能够生成结构连贯、纹理视觉效果出色的人体。

我们的贡献总结如下。首先，我们将 Projected GAN [53]改用于图像修复（第 3.1 节），并提出一种新颖的掩码感知补丁判别器（第3.2 节）。其次，我们研究了判别器所用预训练特征网络的表征能力（第 3.3 节）。实验表明，先前使用的分类网络 [53, 54] 并不适合用于判别人像。因此，我们在判别器中结合使用多个自监督特征网络，显著提升了样本质量。最后，我们提出一种适用于 U-Net [50] 架构的渐进式训练技术（第3.4 节），使我们能够轻松扩展至更高分辨率和更大的模型规模。

我们的贡献最终为野外人体合成带来了新的最先进水平。据我们所知，我们的方法首次实现了无需 DensePose 标注即可生成近乎照片级真实感的人体，同时有效应对极端姿势、复杂背景、不完整的人体以及遮挡。源代码和附录：http://github.com/hukkelas/deep_privacy2/。

[1]请注意，其他研究也探讨了类似任务，[40,60],但它们关注的是更简单的数据集（即 Market1501 [75], DeepFasion [37]），其中重叠或相互遮挡的物体较少。

2. 相关工作

2.1. 全身人体合成

人体合成有广泛的应用，先前研究也有各种各样的高层目标。我们将人体合成分为基于迁移和基于合成模型。基于迁移的方法将源外观（或服装 [14, 52]）迁移到新的姿势 [2, 33, 39, 47, 52, 56],、动作 [4] 或场景 [57]中。虽然其中一些方法适用于野外人物合成 [57, 67],，但它们需要源外观，这使得合成的身份受限于纹理库或外观图像数据集。相比之下，我们的方法可以直接合成新身份。对于后一个目标，基于合成的方法可以合成外观，条件可以是姿势 [40, 60, 68],、场景 [8, 18, 20],，也可以不设条件 [5, 9, 10]。其中有几种方法适用于野外人体合成 [18, 40, 60],，但它们受限于低分辨率 [8, 40],，难以处理复杂背景 [9, 60],，而且只有少数方法能够处理重叠物体 [18, 20]。

无论目标是什么，大多数方法都会使用某种形式的姿态信息，通过 DensePose 标注 [18, 20, 43, 52], 语义分割 [5, 60, 67], 稀疏关键点 [2, 4, 8, 14, 33, 39, 40, 47, 56, 57, 67], 或人体的 3D 姿态 [32, 68]来提升合成质量。

以往研究主要聚焦于基于 GAN 的方法，但近期研究已采用扩散模型 [59]进行人体合成 [22]。我们的工作聚焦于 GAN，因为它能够快速采样高质量图像。

2.2. 生成对抗网络

生成对抗网络 [11]（GAN）长期以来一直是各种全身合成任务中的主流生成模型。GAN 众所周知难以训练，而一个重要的研究重点是实现生成器的稳定训练；为此，人们提出了不同的技术，例如新颖的目标函数 [1], 架构 [24, 26–28], 训练策略 [25], 和正则化 [13, 41]，以提高稳定性和合成质量。近期提出的 Projected GAN [53]在判别器中使用预训练特征网络，以减少训练时间并提升图像质量；随后，该方法又被扩展用于在 ImageNet [6]数据集 [54]上进行高分辨率图像合成。我们延续这一研究方向，将投影式 GAN 适用于条件合成。

2.3. 图像修复

图像修复 [3] 旨在补全自然图像中的缺失区域。与通用图像修复不同，我们补全的是自然图像中随机位置出现的人体所在的缺失区域。长期以来，GAN 一直是自由形状图像修复的主流方法。

图像修复 [46, 70], 而此前的大多数工作都侧重于对生成器进行架构改动。例如，为了处理缺失值 [19, 35, 70], 生成更高分辨率的图像 [69],利用辅助信息 [23, 31, 42], 或通过注意力机制 [71] 或傅里叶卷积来改善感受野 [62]。此前的方法会调整传统的 GAN 判别器，通常是局部块判别器 [21, 36, 49, 65, 70],并结合感知图像相似度损失 [36, 49]和逐像素 l_1 损失 [49, 65]。据我们所知，我们是首个将 Projected GAN [53] 应用于图像修复的方法，并且仅针对对抗目标进行训练。

3. TriA-GAN——一种关键点引导GAN

在本节中，我们逐步引入改动以提升合成质量（表 1）。配置A（第 3.1节）从一种基于StyleGAN的 [27] U-Net [50] 架构开始，该架构与在 [18], 中使用 EfficientNet-Lite0 [63] 通过 ProjectedGAN [54] 训练时所采用的架构相似。配置 B引入了我们的掩码感知判别器目标（第 3.2节），配置 C则将EfficientNet-Lite0 替换为 ViT-L16$_{\text{MAE}}$和 RN50$_{\text{CLIP}}$ （第 3.3 节）。配置 D 引入了我们的渐进式训练技术（第 3.4 节），最后，配置E增大了生成器模型规模。为减少训练时间，我们在低分辨率图像（72×40）上对我们的方法进行了消融实验。最后，第3.5 节将分辨率提升至 288×160。附录 A 包含实验和架构细节。

问题定义 我们将真实场景下的全身合成表述为图像修复任务。我们的目标是补全受损图像中缺失的区域 $I = I \odot M$，其中 I 是真实图像， M 是指示缺失区域的掩码（$M_i = 1$ 已知像素为 1，缺失像素为 0），而 $\odot$ 表示逐元素相乘。为提高合成质量，我们根据 COCO 的 17 个关键点对生成器进行条件约束 [34] 关键点格式

数据集 我们的实验在 FDH数据集[18]上进行。FDH数据集是一个大型未筛选数据集，在 FDH 上训练的模型能够很好地适应真实场景[18]。该数据集包含 187 万张训练图像和 3 万张验证图像。每张图像以单个人物为主体，但同一张图像中可能包含多个人物。每张图像均带有二维关键点标注、用于指示待修复人物的分割掩码，以及像素到表面对应关系（即 T形三维人体的表面）。请注意，TriA-GAN 不使用像素到表面对应关系。

我们发现，FDH数据集中的大量关键点标注存在错误。因此，我们使用 ViTPose 自动重新标注所有图像 [66]（见附录 B）。

表 1。我们方法的迭代开发。每次添加都建立在前一版本之上。配置 A-C 均训练至判别器观察到 50M 个图像。

配置	FID ↓	FID$_{\text{CLIP}}$ ↓	PPL ↓	OKS ↑
A: 基线	1.73	1.74	55.8	0.916
B: + 掩码感知判别器	1.65	1.63	52.8	0.912
C: + 改进的特征网络	1.79	0.47	49.2	0.951
D: + 渐进式生长	1.66	0.40	52.0	**0.954**
E: + 更大的 G（62M → 110M）	**1.62**	**0.30**	**52.0**	0.948

姿态表示 我们将关键点表示为独热编码的空间图，具体而言， $P \in \{0,1\}^{K \times H \times W}$其中 $K = 17$ 且 $P_{k,\ y,\ x} = 1$ 表示位置为（x， y）的关键点 k，其他位置的值均为 0。此外，我们还加入一张绘制人体骨架的空间图（S）。具体而言，空间图 $S \in \{0,1\}^{6 \times H \times W}$ 采用独热编码，分为 6 个类别：连接身体相邻关节的线段分别归为 6 类（左/右臂/腿、躯干、头部）。独热编码的姿态图和骨架图与生成器的输入图像拼接在一起。

评估样本质量 我们使用 Fr 评估样本质量´echet Inception Distance (FID) [16] 和 FID$_{\text{CLIP}}$[2]。此外，我们通过感知路径长度 (PPL) 报告潜在解耦情况 [27], 该指标与形状的一致性和稳定性相关 [28]。

此外，我们还提出了一种用于评估生成的人体形象样本质量的新指标，即物体关键点相似度（*OKS*），它将生成的姿势与真实关键点进行比较。提出这一指标的目的是获得一种不受判别器所用特征网络影响的评估指标。众所周知，Projected GAN [53] 在基于特征的指标 [30], 上会取得人为偏高的分数，因此难以在不同类型的特征网络之间进行定量比较。我们的实验清楚地表明了这一点：配置 B（其判别器使用 ImageNet 特征）生成的受损图像明显多于配置 E，但取得的 ImageNet FID 却相近。

物体关键点相似度（*OKS*）的计算方法是先使用 ViTPose [66], 预测关键点，然后按照 COCO [34]计算预测结果与真实关键点之间的 OKS。与直接计算欧几里得距离相比，OKS 考虑到了预测关键点可能与真实关键点略有偏差，而且不同关键点所允许的偏差程度也不同（例如，肩部关键点的偏差可以大于眼部关键点）。

[2]ImageNet-FID 会给包含 ImageNet 对象的图像更高的评分，而且对人脸不敏感 [30]。使用 FID$_{\text{CLIP}}$ 可以减轻这些问题；我们采用 CLIP [48] 预训练的 ViT-B/32 提取特征。

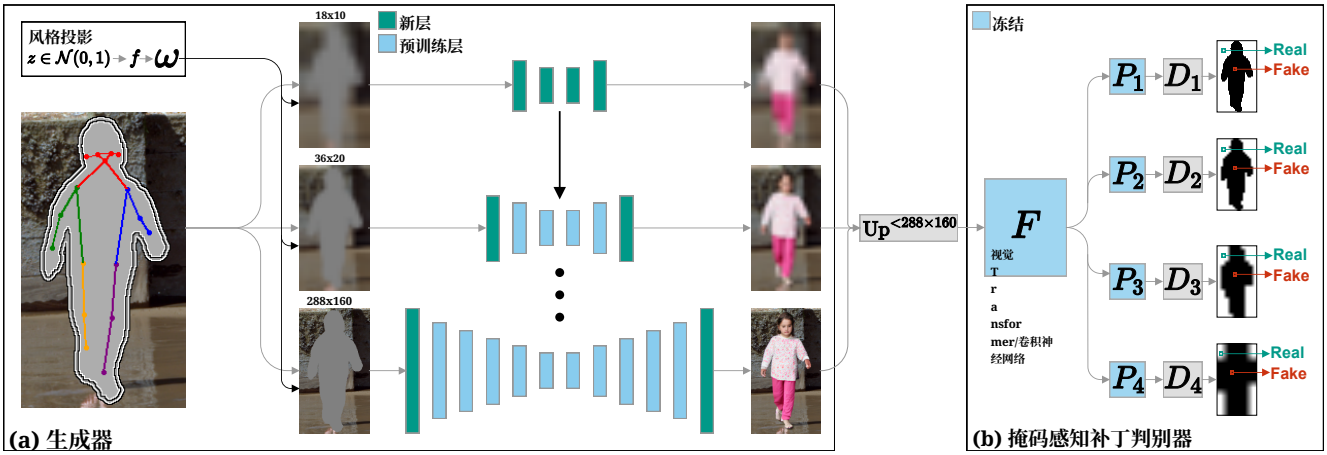


图 2。(a) 给定 17 个关键点后，我们的生成器会填补缺失区域。生成器层采用自适应实例归一化，根据 ω 对生成器进行条件化，其中 ω 是风格映射网络的输出。配置 D&E 从 18×10 分辨率开始渐进式训练，随后通过在编码器/解码器的起始/末尾添加层来提高分辨率。请注意，在整个训练过程中，所有层始终保持可训练状态。(b) 对于每个特征网络 F，我们使用四个浅层局部块判别器来处理其特征（具有不同的空间分辨率），其中每个特征都经过随机可微操作投影（P_1-P_4）。给定投影后的特征，每个判别器都会预测给定图像块是否对应真实或伪造的图像区域。

3.1. 用于图像修复的 Projected GAN

Projected GAN [53]采用预训练特征网络来区分真实图像和伪造图像。给定一幅图像 I，对抗目标函数表示为

$$\min_G \max_{D_\ell} \sum_{\ell \in \mathcal{L}} \mathbb{E}_{I \sim p_{data}} \left[\log \left(D_\ell \left(P_\ell \left(I \right) \right) \right) \right] + \mathbb{E}_{z \sim p_z} \left[\log \left(1 - D_\ell \left(P_\ell \left(G \left(z, \bar{I} \right) \right) \right) \right) \right], \tag{1}$$

其中 $\{D_\ell\}$ 是一组独立的判别器，各自在其特征投影器上运行 P_ℓ。每个投影器在训练期间都被冻结，并由一个预训练的特征网络组成 F，其中来自 F 的特征通过可微操作进行随机投影。对于基线（**配置 A**），我们使用 EfficientNet-Lite0 [63]作为 F，遵循 [53]，，我们将在第 3.3节中重新讨论这一点。对于每个判别器 D_ℓ，我们采用了图块判别器架构，详见第 3.2节。

公式 (1) 并未强制条件 (I与生成的图像保持一致，因此生成器在实际应用中学会完全忽略 I。因此，我们通过对生成器的输出应用掩码来强制条件一致性。具体而言，我们设定 $G(z, I) = I \odot (1 - M) + G(z, I) \odot M$，其中 $G(z, I)$ 是 G 中最后一层的输出。

3.1.1 稳定生成器

直接将 Projected GAN 用于图像修复时，训练过程不稳定，并且容易在训练早期发生模式崩溃。这是因为生成器难以跟上预训练判别器的步伐，导致判别器在训练早期压制生成器。为改善稳定性，我们对对抗训练设置进行了若干修改。首先，在训练开始时对输入判别器的图像进行模糊处理，并在 4M 张图像的训练过程中线性减弱模糊效果。长时间的模糊处理可防止判别器关注由生成器输出的掩码造成的高频边缘。以往的方法会在前 200k 张图像的训练过程中对判别器输入进行模糊处理 [26, 54]，，而我们发现显著延长这一阶段很有益。此外，U-net 架构通过映射网络和 StyleGAN 风格调制注入潜在代码（z）。我们将映射网络设为 2 层，并将 2 [28]的维度降至 64，遵循 z。此外，我们将残差跳跃连接缩放为 [54]（类似于 $1/\sqrt{2}$），并在存在残差 U-net 连接时，将 [28]用于跳跃连接。最后，我们使用实例归一化代替权重解调 $1/\sqrt{3}$，因为我们发现这样训练起来更稳定。

3.2. 掩码感知补丁判别器

Projected GAN [53, 54] 采用四个浅层判别器，它们作用于不同的特征投影（P_ℓ），具有不同的空间分辨率。每个判别器输出的 logits 分辨率相同（4×4）。相比之下，我们发现局部块判别器更适用于图像修复任务，因为每个判别器尝试分类局部图像块，而非整幅图像。具体而言，每个 D_ℓ（输入来自投影 P_ℓ的特征）由三层卷积组成，其中 D_ℓ 的输出空间分辨率是P_ℓ的一半。我们发现，用局部块判别器替换 [53] 中的判别器，能显著提升性能。

局部块判别器广泛应用于图像修



(a) FID ↓ (b) FID_{CLIP} ↓ (c) PPL ↓ (d) OKS ↑

图 3. 使用标准投影 GAN 目标（公式（1））和掩码感知判别器目标（公式（2））时，不同特征网络的比较。所有模型均训练至判别器观察到 5000 万张图像。

复 [62, 70, 73, 74]。通常，每个图像块都被分类为属于原始图像的类别，因此，与真实图像对应的所有图像块都会被分类为真实。然而，这会给图像修复任务带来歧义，因为某些特征（例如，CNN 的浅层特征）可能仅依赖真实像素，但由于感受野有限，整幅图像仍是伪造的。因此，我们提出了一种掩码感知判别器目标，根据判别器的图像块对应图像中的真实区域还是伪造区域，将其归类为真实类或伪造类。新的目标函数如下：

$$\min_G \max_{D_\ell} \sum_{\ell \in \mathcal{L}} \mathbb{E}_{I \sim p_{data}} \left[\log\left(D_\ell\left(P_\ell\left(I\right)\right)\right)\right] + \mathbb{E}_{z \sim p_z} \left[\sum_y^{H_\ell} \sum_x^{W_\ell} M_\ell^{y,x} \cdot \log\left(D_\ell^{y,x}\left(P_\ell\left(G\left(z, \bar{I}\right)\right)\right)\right) + (1 - M_\ell^{y,x}) \cdot \log(1 - D_\ell^{y,x}(P_\ell(G(z, \bar{I})))) \right], \quad (2)$$

其中 $D_\ell \in \mathbf{R}^{H_\ell \times W_\ell}$，$M_\ell$ 通过最小池化从 M 下采样至 $H_\ell \times W_\ell$。

公式（2）消除了全局类别分配导致的图像块分类歧义，从而为生成器提供更详细且空间连贯的响应。此外，该公式为判别器引入了一项辅助任务，已知这有助于提升合成质量 [45]。在我们的设定中，辅助任务是对与生成区域相对应的区域进行空间分割。

图 3 证实，公式 (2) 能提升图像质量（FID/FID_{CLIP}）和 OKS，且适用于多种特征网络，包括采用不同预训练任务和架构（CNN 和 ViT）的特征网络。此前已有研究将类似的分割判别器用于其他任务 [55, 61, 68]。我们的工作进一步验证了这一概念的普适性：无论使用何种特征网络，都可以利用预训练特征网络构建极浅的判别器架构 F。

3.3. 用于人体合成的判别特征网络

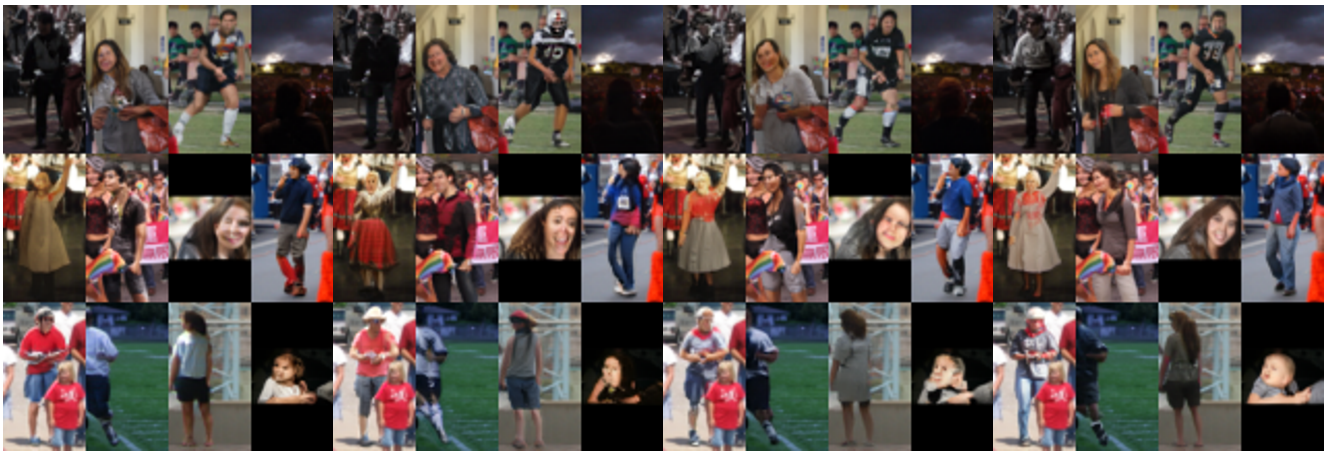
长期以来，GAN 在对齐的人体合成方面取得了令人瞩目的成果，尤其是在 FFHQ [27] 和 CelebA-HQ [25, 38] 数据集上。然而，众所周知，Projected GAN 在 FFHQ 上进行人脸合成时会产生伪影[53]，而且难以生成逼真的非对齐人体图像 [54] [3]。我们发现，人体合成质量不佳源于判别器所用预训练特征空间中的不变性。早期工作 [53, 54] 采用了预训练的 ImageNet [6]分类网络。这些特征网络学习特征表示的唯一目标是进行分类，即将图像映射到排名第一的类别。因此，它们会忽略与分类目标无关的特征。尽管这种不变性有利于图像分类，但我们发现它会损害人体合成所需的判别性表示。

我们为判别器探索了不同的特征网络（包括 CNN/ViT 的各种变体），这些网络采用差异很大的预训练任务。具体而言，图3对采用以下预训练任务的特征网络进行了消融研究：

- **在**：ImageNet 分类：ResNet50（**RN50**）、**ViT-B16**（DeIT 变体）、EfficientNet-Lite0（**EN-L0**）。
- **CLIP**：对比式语言-图像预训练 [48]：**RN50**、**ViT-B16**。
- **MAE**：掩码自编码器 [15]：**ViT-B16**、**ViT-L16**。
- **CSE**：DensePose 姿态估计 [44]：ResNet50（**RN50**）。

我们将每个模型称为架构$_{task}$，例如，$\text{RN50}_{\text{CLIP}}$指采用 CLIP 预训练权重的 ResNet-50。直接从标准生成指标（FID/FID_{CLIP}）中选出最佳特征网络并不明确，因为众所周知，投影 GAN 在基于特征的指标 [30]上会取得不自然的高分。我们发现，ImageNet 模型会因匹配预训练而获得不自然的高 FID。

[3]请参阅 [54]中的附录。



(a) ImageNet (ViT-B/16) (b) CLIP (ViT-B/16) (c) MAE (ViT-B/16) (d) ViT-L/16_{MAE} +$\text{RN50}_{\text{CLIP}}$

图 4。定性比较 用于判别器的各种特征网络的儿子。值得注意的是，这些示例是 n ot 经过整理，但从以下内容中选取 验证集中的前 12 张图像。

任务，而 ViT 在 FID 上得分更高$_{\text{CLIP}}$ 由于架构匹配 [4]。

无论架构如何，我们都观察到所有 ImageNet [6] 模型都会生成严重失真的人脸，如图 4所示。这很可能是由于这些特征网络中的人脸描述符具有不变性所致，这种现象在 [30]中也有报道。请注意，附录 C 中的图 3对所有网络进行了比较。

从图 3的结果可以看出，**配置 C**用 ViT-L 替换了 EfficientNet-Lite0$16_{\text{MAE}}$ 和 $\text{RN50}_{\text{CLIP}}$。将这些网络配对的动机，是利用架构和预训练任务完全不同的特征。此外，这些网络在 OKS、FID_{CLIP}和 PPL 方面均表现出色。最后，$\text{RN50}_{\text{CLIP}}$ 很好地补充了 ViT：RN50 使用原始宽高比（288×160），而 ViT 的输入尺寸固定为 224×224 [5]。

3.4. 渐进式增长

渐进式训练 [25] 已知可以提高 GAN 的训练稳定性，近期又被重新引入，用于使用 Projected GAN 进行无条件合成 [54]。StyleGAN-XL [54] 首先在 16×16 分辨率下进行训练，然后通过在解码器末端添加新层来提高分辨率。请注意，在训练下一阶段时，StyleGAN-XL 会冻结已经训练好的层和样式网络。

我们采用一种直接扩展图像到图像转换的方法，逐步训练 U-net 架构，分别在编码器和解码器的起始端/末端添加层（见图。2）。我们观察到，在编码器起始端添加新模块会导致训练不稳定，因为这会对已训练层的输入。为缓解这一问题，我们为每个残差块引入 LayerScale [64]，其初始值为 10^{-5}，以减小新块的贡献。此外，我们按照 [27] 添加输出跳跃连接。与 StyleGAN-XL 不同，我们在训练期间不冻结任何块，因为计算收益微乎其微：我们仍需计算编码器起始处各层的梯度。引入这些更改后，最终图像质量显著提升（**配置 D**）

我们注意到，我们尝试过更先进的渐进式训练技术，例如级联 U-net[17]，，或对编码器/解码器进行非对称训练（即从全分辨率编码器和低分辨率解码器开始）。然而，我们发现，直接的渐进式训练技术在训练时间和最终图像质量方面都更优。

3.5. 扩展生成器的规模

配置E 在编码器/解码器中，将每个分辨率对应的残差块数量加倍，使生成器中的参数量达到110.4M，相比之前的 62.2M。该模型可以稳定训练至 288×160 分辨率，这是 FDH数据集的最大分辨率。

4. 与表面引导生成对抗网络的比较

表2 将 TriA-GAN 与表面引导生成对抗网络（SG-GAN）进行比较 [20]采用 DeepPrivacy 训练2 [18]，这是目前真实场景全身合成领域的最先进方法。图1 展示了 TriA-GAN 的合成结果，图5 将 TriA-GAN 与 SG-GAN 进行比较。附录D 包含随机选取的样本。

TriA-GAN 与 SG-GAN 的主要区别[18]在于 TriA-GAN 改进了训练策略，并采用了

[4] FID_{CLIP} 根据 ViT-B/32 的特征进行计算，具体遵循 [30]。[5]对于所有模型，ViT 输入分辨率均设置为 224×224，因为与训练分辨率相比，ViT 特征对分辨率变化的鲁棒性较低。

表 2. SG-GAN 的定量比较 [18]与我们的方法。

方法	FID ↓	FID_{CLIP} ↓	PPL ↓	OKS ↑
SG-GAN [18]	1.97	1.25	70.2	0.950
TriA-GAN（我们的方法）	**1.68**	**0.43**	**47.8**	**0.972**

更稀疏的条件信息（关键点而非密集的表面对应关系）。TriA-GAN 在处理重叠物体、局部身体（例如与图像边缘相交）以及纹理合成（例如头发、服装）方面表现更佳。此外，TriA-GAN 的上下文处理能力也有所提升，例如能够推断一位老妇人很可能会坐在桌旁（图5顶行），或推断自行车上有一位骑摩托车的人（图5第 3 行）。

最后，TriA-GAN 更便于用于下游任务，因为我们的方法不依赖 DensePose 检测结果。例如，在交互式编辑应用中，关键点更容易编辑。此外，DensePose 检测难度较大，且在远距离检测时可靠性较低，这限制了它在许多场景中的应用（例如，对街上的行人进行匿名化）。失败案例示例见附录E 。

5. TriA-GAN 的可编辑性

StyleGAN [27] 以其解耦的潜空间而闻名，并广泛用于用户引导的图像编辑，例如通过文本提示修改图像 [29]。然而，大多数图像编辑方法都专注于无条件 GAN（或类别条件 GAN），而将其应用于图像修复的研究较少。StyleMC [29]对于使用图像修复方法编辑面部非常有效 [18],但同一项研究发现，在真实场景中编辑人体要困难得多 [18]。我们认为，这一局限源于 DensePose条件，其中描述性条件可能与特定属性相关联。这会缩小采样概率，使得从随机采样的图像中找到有意义的方向更加困难。

图 6 证明 StyleMC [29] 与 TriA-GAN 结合，能够在 GAN 潜在空间中找到语义上有意义的方向。StyleMC 通过 CLIP 编码器将随机图像朝文本提示的方向进行编辑，从而找到全局方向 [48], , 这些方向是基于 1280 张图像找到的。我们发现，StyleMC 与 TriA-GAN 结合后能够编辑多种属性，甚至是相当具体的属性，例如耳朵的大小。不过，我们也注意到，编辑某些属性会导致其他相关属性发生变化。例如，“金发”编辑会轻微改变肤色。此外，有些属性更难编辑。例如，内向

将“红唇”引入推断为男性的身体可能会导致显著的语义变化（图 6 的顶行）。目前尚不清楚这一局限是否源于编辑技术

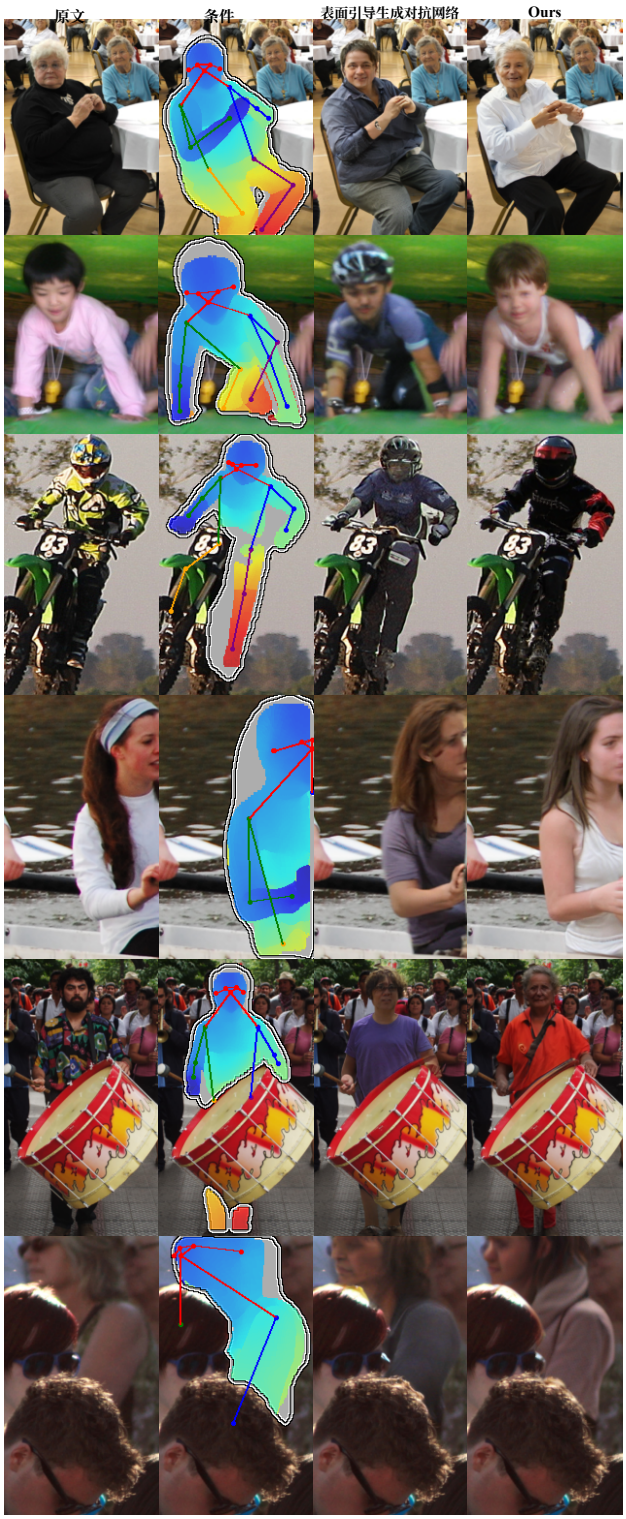


图5。对比表面引导生成对抗网络 [18] 与TriA-GAN。请注意，TriA-GAN 未使用表面信息（以蓝黄色调显示）。



图 6. StyleMC [29] 使用 TriA-GAN 进行编辑，其中将一个全局方向（来自每列上方的文本提示）添加到原始图像（最左侧）的样式代码中。

或 TriA-GAN 本身。我们认为，这些相关性固有地存在于 CLIP 或 TriA-GAN 的训练数据集中。

6. 结论

TriA-GAN 让我们能够根据一组稀疏的关键点，在任意期望的姿势和位置生成人体，并在 FDH数据集上的人物合成方面达到了新的最先进水平。我们方法的关键在于为判别器利用预训练的特征网络。我们证明，精心设计的训练策略与适用于判别人体的特征网络相结合，能够显著提升合成质量。TriA-GAN 首次展示了通过文本提示可靠地编辑人体属性，我们相信这在许多应用中都将具有很高的实用价值。

社会影响合成人体在各个领域都有一系列实用应用，从内容创作到匿名化处理。然而，与所有基于学习的生成模型类似，合成的人体遵循数据集的采样概率。在我们的案例中，数据集源自Flickr，这意味着我们的生成器会遵循其中的偏见，不太可能合成来自该网站上代表性不足群体的人物。此外，我们的工作专注于生成逼真的人类，这也带来了被滥用的可能性（例如深度伪造）。我们注意到，社区已经通过 DeepFake 检测等倡议，齐心协力应对这一问题，



图 7。TriA-GAN 的失败案例。

tion 挑战 [7], 或将水印嵌入生成模型生成的图像中 [72]。

6.1. 局限性

TriA-GAN在野外场景人体合成方面达到了新的当前最佳水平。探索将潜在空间与姿态、身体形状和环-境解耦的方法，是令人兴奋的未来研究方向。目前，TriA-GAN的采样空间高度依赖条件信息，在给定某些条件时，可能会坍缩为单一的合成人体身份。通过将姿态、外观和上下文解耦，解耦的人物图-像生成可以缓解这一问题。然而，当前方法需要包含成对图像 [40, 51], which 的数据集，而这类数据集多样性较低且规模较小。

TriA-GAN 的主要局限在于难以处理与物体之间更复杂的交互（图7）。生成逼真的手和手指时尤其如此，例如弹钢琴。若 DensePose 信息明确描述了交互，SG-GAN [18] 在这类场景中通常比 TriA-GAN 表现更好。但在交互不明确的情况下，它仍然难以应对（例如演奏被掩码遮挡的小号）。

对于 FDH数据集中较少出现的属性，Tri A-GAN 很难进行编辑。例如，许多图像不包含下半身，尝试寻找“穿着红裤子的人”的编辑方向时，也会导致其他属性发生变化。这究竟是编辑方法的局限，还是 TriA-GAN 本身的局限，仍是一个悬而未决的问题。

致谢 计算在 NTNU IDUN/EPIC 计算集群 [58]以及由 Anne C. Elster 教授领导的 Tensor-GPU 项目上完成。此外，我们感谢 Rudolf Mester 提供的一般性支持和富有思考的讨论。

参考文献

[1] MartinArjovsky、Soumith Chintala和 L´eon Bottou。Wasserstein生成对抗网络。载于国际机器学习会议，第214–223页，2017年 2 –。[2] Guha Balakrishnan、Amy Zhao、Adrian V Dalca、Fredo Du-rand 和John Guttag。合成未知姿势下的人体图像。载于2018年IEEE/CVF计算机视觉与模式识别会议，第8340–8348页。IEEE，2018年6月。1，2[3] M. Bertalmio、G. Sapiro、V. Caselles 和 C.Ballester。图像修复。载于ACM SIGGRAPH计算机图形学会议论文集，第417–424页，2000年。2[4] Caroline Chan、Shiry Ginosar、Tinghui Zhou 和 Alexei Efros。现在每个人都跳舞。载于2019年IEEE/CVF国际计算机视觉会议（ICCV），第49卷，第5932–5941页。IEEE，2019年10月。1，2[5] Bindita Chaudhuri、Nikolaos Sarafianos、LindaShapiro和Tony Tung。面向3D人体的高分辨率可编辑纹理半监督合成。载于IEEE计算机学会计算机视觉与模式识别会议论文集，第7987–7996页，2021年。2[6] Jia Deng、Wei Dong、RichardSocher、Li-JiaLi、KaiLi 和Li Fei-Fei。ImageNet：大规模分层图像数据库。载于2009年IEEE计算机视觉与模式识别会议，第248–255页。IEEE，2009年6月。2，5，6[7] Brian Dolhansky、Joanna Bitton、Ben Pflaum、Jikuo Lu、Russ Howes、Menglin Wang 和 Cristian Canton Ferrer。DeepFake检测挑战赛（DFDC）数据集。arXiv预印本arXiv:2006.07397，2020年。8[8] PatrickEsser和Ekaterina Sutter。用于条件外观与形状生成的变分U-Net。载于2018年IEEE/CVF计算机视觉与模式识别会议，第8857–8866页。IEEE，2018年6月。2[9] Anna Fruhstuck、Krishna Kumar Singh、Eli Shechtman、Niloy J Mitra、Peter Wonka和 Jingwan Lu。用于全身图像生成的InsetGAN。载于2022年IEEE/CVF计算机视觉与模式识别会议（CVPR），第7713–7722页。IEEE，2022年6月。1，2[10] Jianglin Fu、Shikai Li、Yuming Jiang、Kwan-Yee Lin、Chen Qian、ChenChange Loy、Wayne Wu 和 ZiweiLiu。StyleGAN-Human：以数据为中心的人体生成奥德赛。载于计算机视觉——ECCV 2022，第13676卷，第1–19页。Springer，Cham，2022年。2[11] Ian Goodfellow、Jean Pouget-Abadie、Mehdi Mirza、Bing Xu、David Warde-Farley、Sherjil Ozair、Aaron Courville 和YoshuaBengio。生成对抗网络。载于神经信息处理系统进展，第2672–2680页，2014年 2 – 。[12] Riza Alp Guler、Natalia Neverova和 IasonasKokkinos。DensePose：野外密集人体姿态估计。载于2018年IEEE/CVF计算机视觉与模式识别会议，第7297–7306页。IEEE，2018年6月。2[13] IshaanGulrajani、Faruk Ahmed、Martin Arjovsky、VincentDumoulin 和 Aaron C Courville。改进的训练Wasserstein生成对抗网络。在《神经信息处理系统进展》中，第5767–5777页，2017年。 2[14] Xintong Han、Zuxuan Wu、Zhe Wu、Ruichi Yu 和 Larry S Davis。VITON：基于图像的虚拟试穿网络。载于《2018年IEEE/CVF计算机视觉与模式识别会议》，第7543–7552页。IEEE，2018年6月。 1, 2[15] Kaiming He、Xinlei Chen、Saining Xie、Yanghao Li、Piotr Dollar 和 Ross Girshick。掩码自编码器是可扩展的视觉学习器。载于《2022年IEEE/CVF计算机视觉与模式识别会议（CVPR）》，第2022年6月卷，第15979–15988页。IEEE，2022年6月。5[16]Martin Heusel、Hubert Ramsauer、Thomas Unterthiner、Bernhard Nessler 和 Sepp Hochreiter。通过双时间尺度更新规则训练的生成对抗网络收敛至局部纳什均衡。载于《神经信息处理系统进展》，第6626–6637页，2017年。 3[17] Jonathan Ho、Chitwan Saharia、William Chan、David J. Fleet、Mohammad Norouzi 和 Tim Salimans。用于高保真图像生成的级联扩散模型。《机器学习研究期刊》，23:1–33，2022年。 6[18]H˚akon Hukkel˚as 和 Frank Lindseth。DeepPrivacy2：迈向逼真的全身匿名化。载于《2023年IEEE/CVF冬季计算机视觉应用会议（WACV）》，第1329–1338页。IEEE，2023年1月。1, 2, 3, 6, 7, 8[19] H˚akon Hukkel˚as、Frank Lindseth 和 Rudolf Mester。利用可学习特征填补进行图像修复。载于《DAGM德国模式识别会议》，第388–403页。施普林格出版社，2021年。 3[20] H˚akon Hukkel˚as、Morten Smebye、Rudolf Mester 和 Frank Lindseth。利用表面引导生成对抗网络实现逼真的全身匿名化。载于《2023年IEEE/CVF冬季计算机视觉应用会议（WACV）》，第1430–1440页。IEEE，2023年1月。 2, 6[21] Phillip Isola、Jun-Yan Zhu、Tinghui Zhou 和 Alexei A Efros。利用条件生成对抗网络进行图像到图像转换。载于《2017年IEEE计算机视觉与模式识别会议（CVPR）》，第5967–5976页。IEEE，2017年7月。 3[22] Yuming Jiang、Shuai Yang、Haonan Qiu、Wayne Wu、Chen Change Loy 和 Ziwei Liu。Text2human：文本驱动的可控人体图像生成。《ACM图形学汇刊》，41(4):1–11，2022年7月。 2[23] Youngjoo Jo 和 Jongyoul Park。SC-FEGAN：利用用户草图和颜色进行人脸编辑的生成对抗网络。载于《2019年IEEE/CVF国际计算机视觉会议（ICCV）》，第1745–1753页。IEEE，2019年10月。 3[24] Animesh Karnewar 和 Oliver Wang。MSG-GAN：用于生成对抗网络的多尺度梯度。载于《2020年IEEE/CVF计算机视觉与模式识别会议（CVPR）》，第7796–7805页。IEEE，2020年6月。 2[25] Tero Karras、Timo Aila、Samuli Laine 和 Jaakko Lehtinen。通过生成对抗网络渐进式增长提升质量、稳定性和多样性。载于《国际学习表征会议》，2018年。 2, 5, 6

[26] 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》。载于《神经信息处理系统进展》，第34卷，863、2021 852，第2、4页 – 。[27] Tero Karras、Samuli Laine 和 Timo Aila。《生成对抗网络的风格化生成器架构》。载于《2019年IEEE/CVF计算机视觉与模式识别会议》（CVPR），第4396–4405页。IEEE，2019年6月。2、3、5、6、7[28] Tero Karras、Samuli Laine、Miika Aittala、Janne Hellsten、Jaakko Lehtinen 和 Timo Aila。《分析并提升StyleGAN的图像质量》。载于《2020年IEEE/CVF计算机视觉与模式识别会议》（CVPR），第8107–8116页。IEEE，2020年6月。2、3、4[29] Umut Kocasari、Alara Dirik、Mert Tiftikci 和 Pinar Ya- nardag。《StyleMC：基于多通道的快速文本引导图像生成与操控》。载于《2022年IEEE/CVF冬季计算机视觉应用会议》（WACV），第3441–3450页。IEEE，2022年1月。7、8[30] Tuomas Kynk¨a¨anniemi、Tero Karras、Miika Aittala、Timo Aila 和 Jaakko Lehtinen。《ImageNet类别在弗雷歇特征距离中的作用》。arXiv预印本arXiv:2203.06026，2022年。3、5、6[31] Avisek Lahiri、Arnav Kumar Jain、Sanskar Agrawal、Pabi- tra Mitra 和 Prabir Kumar Biswas。《先验引导的基于GAN的语义图像修复》。载于《2020年IEEE/CVF计算机视觉与模式识别会议》（CVPR）。IEEE，2019年6月 2020 3。Christoph [32] Lassner、Gerard Pons-Moll 和 Peter V. Gehler。《着装人物生成模型》。载于《2017年IEEE国际计算机视觉会议》（ ICCV），第2017-Octob卷，第853–862页。IEEE，2017年10月。2[33] Yining Li、Chen Huang 和 Chen Change Loy。《用于人体姿态迁移的稠密内在外观流》。载于《2019年IEEE/CVF计算机视觉与模式识别会议》（CVPR），第3688–3697页。IEEE，2019年6月。2[34] Tsung-Yi Lin、Michael Maire、Serge Belongie、James Hays、Pietro Perona、Deva Ramanan、Piotr Doll´ar 和 C. Lawrence Zitnick。《Microsoft COCO：上下文中的常见物体》。载于《欧洲计算机视觉会议》，第8693卷《计算机科学讲义》，第740–755页。Springer，Cham，2014年。3[35] Guilin Liu、Fitsum A Reda、Kevin J Shih、Ting-Chun Wang、Andrew Tao 和 Bryan Catanzaro。《使用部分卷积修复不规则孔洞图像》。载于《欧洲计算机视觉会议论文集》（ECCV），第85–100页，2018年。3[36] Hongyu Liu、Bin Jiang、Yi Xiao 和 Chao Yang。《用于图像修复的连贯语义注意力》。载于《2019年IEEE/CVF国际计算机视觉会议》（ICCV），第4169–4178页。IEEE，2019年10月。3[37] Ziwei Liu、Ping Luo、Shi Qiu、Xiaogang Wang 和 Xiaoou Tang。《DeepFashion：借助丰富标注实现稳健的服装识别与检索》。载于《2016年IEEE计算机视觉与模式识别会议》（CVPR），第 1096 1104 2016 1, 2 –页。IEEE，6月。[38] Ziwei Liu、Ping Luo、Xiaogang Wang 和 Xiaoou Tang。《野外场景中的深度学习人脸属性》。载于《2015年IEEE】【。国际计算机视觉会议（ICCV），第3730–3738页。IEEE，2015年12月。5[39] Liqian Ma、Xu Jia、Qianru Sun、Bernt Schiele、Tinne Tuyte-laars 和 Luc Van Gool。姿态引导的人物图像生成。载于《神经信息处理系统进展》，第2017-Decem卷，第406–416页，2017年。2[40] Liqian Ma、Qianru Sun、Stamatios Georgoulis、Luc Van Gool、Bernt Schiele 和 Mario Fritz。解耦人物图像生成。载于2018年IEEE/CVF计算机视觉与模式识别会议，第99–108页。IEEE，2018年6月。2, 8[41] Takeru Miyato、Toshiki Kataoka、Masanori Koyama 和 Yuichi Yoshida。用于生成对抗网络的谱归一化。国际学习表征会议，2018年。2[42] Kamyar Nazeri、Eric Ng、Tony Joseph、Faisal Qureshi 和 Mehran Ebrahimi。EdgeConnect：使用边缘预测的结构引导图像修复。载于2019年IEEE/CVF国际计算机视觉会议研讨会（ICCVW），第3265–3274页。IEEE，2019年10月。3[43] Natalia Neverova、Riza Alp G¨uler 和 Iasonas Kokkinos。稠密姿态迁移。载于欧洲计算机视觉会议，第11207卷《计算机科学讲义》，第128–143页，2018年。2[44] Natalia Neverova、David Novotny、Vasil Khalidov、Marc Szafraniec、Patrick Labatut 和 Andrea Vedaldi。连续表面嵌入。载于《神经信息处理系统进展》，第33卷，第17258–17270页。Curran Associates, Inc.，2020年11月。5[45] Augustus Odena、Christopher Olah 和 Jonathon Shlens。利用辅助分类器生成对抗网络进行条件图像合成。载于第34届国际机器学习会议（国际机器学习会议），第6卷，第4043–4055页，2017年。5[46] Deepak Pathak、Philipp Krahenbuhl、Jeff Donahue、Trevor Darrell 和 Alexei A Efros。上下文编码器：通过图像修复进行特征学习。载于2016年IEEE计算机视觉与模式识别会议（计算机视觉与模式识别会议），第2536–2544页。IEEE，2016年6月。3[47] Albert Pumarola、Antonio Agudo、Alberto Sanfeliu 和 Francesc Moreno-Noguer。任意姿态下的无监督人物图像合成。载于2018年IEEE/CVF计算机视觉与模式识别会议，第8620–8628页。IEEE，2018年6月。2[48]Alec Radford、Jong Wook Kim、Chris Hallacy、Aditya Ramesh、Gabriel Goh、Sandhini Agarwal、Girish Sastry、Amanda Askell、Pamela Mishkin、Jack Clark、Gretchen Krueger 和 Ilya Sutskever。通过自然语言监督学习可迁移视觉模型。载于国际机器学习会议，第8748 8763页，2021年 3, 5, 7 – 。[49] Yurui Ren、Xiaoming Yu、Ruonan Zhang、Thomas H Li、Shan Liu 和 Ge Li。StructureFlow：通过结构感知外观流进行图像修复。载于2019年IEEE/CVF国际计算机视觉会议（ICCV），第181–190页。IEEE，2019年10月。3[50] Olaf Ronneberger、Philipp Fischer 和 Thomas Brox。U-Net：用于生物医学图像分割的卷积网络。载于国际医学图像

计算与计算机辅助干预，第234–241页。Springer，2015年。2, 3[51] Kripasindhu Sarkar、Vladislav Golyanik、Lingjie Liu 和 Christian Theobalt。从单目视角控制人物图像合成的风格与姿态。arXiv预印本 arXiv:2102.11263，2021年。8[52] Kripasindhu Sarkar、Dushyant Mehta、Weipeng Xu、Vladislav Golyanik 和 Christian Theobalt。从单张图像对人物进行神经重渲染。载于欧洲计算机视觉会议，第12356卷《计算机科学讲义》，第596–613页。Springer Science and Business Media Deutschland GmbH，2020年。1, 2[53] Axel Sauer、Kashyap Chitta、Jens M¨uller 和 Andreas Geiger。Projected GAN收敛速度更快。载于《神经信息处理系统进展》，第17480–17492页，2021年。2, 3, 4, 5[54] Axel Sauer、Katja Schwarz 和 Andreas Geiger。StyleGAN-XL：将StyleGAN扩展至大型多样化数据集。载于ACM SIGGRAPH 2022会议论文集，第1–10页，加拿大不列颠哥伦比亚省温哥华，2022年8月。计算机协会。2, 3, 4, 5, 6[55] Edgar Schonfeld、Bernt Schiele 和 Anna Khoreva。用于生成对抗网络的U-Net判别器。载于2020年IEEE/CVF计算机视觉与模式识别会议（计算机视觉与模式识别会议），第8204–8213页。IEEE，2020年6月。5[56] Chenyang Si、Wei Wang、Liang Wang 和 Tieniu Tan。基于姿态的人物图像合成的多阶段对抗损失。载于2018年IEEE/CVF计算机视觉与模式识别会议，第118–126页。IEEE，2018年6月。2[57] Aliaksandr Siarohin、Enver Sangineto、Stephane Lathuiliere 和 Nicu Sebe。用于基于姿态的人物图像生成的可变形GAN。载于2018年IEEE/CVF计算机视觉与模式识别会议，第3408–3416页。IEEE，2018年6月。2[58] Magnus Sj¨alander、Magnus Jahre、Gunnar Tufte 和 Nico Reissmann。EPIC：节能、高性能GPGPU计算研究基础设施，2019年。8[59] Jascha Sohl-Dickstein、Eric Weiss、Niru Maheswaranathan 和 Surya Ganguli。利用非平衡热力学进行深度无监督学习。载于国际机器学习会议，第2256–2265页。PMLR，2015年。2[60] Sijie Song、Wei Zhang、Jiaying Liu、Zongming Guo 和 Tao Mei。通过语义解析变换实现无配对人物图像生成。《IEEE模式分析与机器智能汇刊》，43(11)：4161–4176，2021年11月。2[61] Vadim Sushko、Edgar Sch¨onfeld、Dan Zhang、Juergen Gall、Bernt Schiele 和 Anna Khoreva。OASIS：仅通过对抗监督实现语义图像合成。《国际计算机视觉杂志》，130(12)：2903–2923，2022年12月。5[62] Roman Suvorov、Elizaveta Logacheva、Anton Mashikhin、Anastasia Remizova、Arsenii Ashukha、Aleksei Silvestrov、Naejin Kong、Harshith Goka、Kiwoong Park 和 Victor Lempitsky。《分辨率鲁棒大型掩码傅里叶卷积图像修复》。载于《2022年IEEE/CVF冬季计算机视觉应用会议》（WACV），第3172–3182 2022 3, 5 页。IEEE，2015年12月。[63] Mingxing Tan 和 Quoc V Le。《EfficientNet：重新思考卷积神经网络的模型缩放》。载于国际机器学习会议论文集，第6105–6114页，2019年 3, 4 – 。[64] Hugo Touvron、Matthieu Cord、Alexandre Sablayrolles、Gabriel Synnaeve 和 Herve Jegou。《深入探索图像Transformer》。载于《2021年IEEE/CVF国际计算机视觉大会》（ICCV），第32–42页。IEEE，2019年10月。6[65] Wei Xiong、Jiahui Yu、Zhe Lin、Jimei Yang、Xin Lu、Connelly Barnes 和 Jiebo Luo。《前景感知图像修复》。载于《2019年IEEE/CVF计算机视觉与模式识别会议》（CVPR），第5833–5841页。IEEE，2018年6月 2019 3 。Yufei Xu、Jing Zhang、Qiming Zhang[66] 和 Dacheng Tao。《ViTPose：人体姿态估计的简易视觉Transformer基线》。arXiv预印本 arXiv:2204.12484，2022年。3[67] Chaojie Yang、Hanhui Li、Shengjie Wu、Shengkai Zhang、Haonan Yan、Nianhong Jiao、Jie Tang、Runnan Zhou、Xi-aodan Liang 和 Tianxiang Zheng。《BodyGAN：通用可控神经人体生成》。载于《2022年IEEE/CVF计算机视觉与模式识别会议》（CVPR），第7723–7732页。IEEE，2022年6月。2[68] Zhuoqian Yang、Shikai Li、Wayne Wu 和 Bo Dai。《3DHumanGAN：迈向照片级真实感的3D感知人体图像生成》。arXiv预印本 arXiv:2212.07378，2022年。2, 5[69] Zili Yi、Qiang Tang、Shekoofeh Azizi、Daesik Jang 和 Zhan Xu。《用于超高分辨率图像修复的上下文残差聚合》。载于《2020年IEEE/CVF计算机视觉与模式识别会议》（CVPR），第7505–7514页。IEEE，2020年6月。3[70] Jiahui Yu、Zhe Lin、Jimei Yang、Xiaohui Shen、Xin Lu 和 Thomas Huang。《使用门控卷积进行自由形状图像修复》。载于《2019年IEEE/CVF国际计算机视觉会议》（ICCV），第4470–4479页。IEEE，2019年10月。3, 5[71] Jiahui Yu、Zhe Lin、Jimei Yang、Xiaohui Shen、Xin Lu 和 Thomas S. Huang。《结合上下文注意力的生成式图像修复》。载于《2018年IEEE/CVF计算机视觉与模式识别会议》，第5505–5514页。IEEE，2018年6月。3[72] Ning Yu、Vladislav Skripniuk、Sahar Abdelnabi 和 Mario Fritz。《生成模型的人工指纹：将深度伪造归因追溯至训练数据》。载于《2021年IEEE/CVF国际计算机视觉大会》（ICCV），第14428–14437页。IEEE，2021年10月。8[73] Tao Yu、Zongyu Guo、Xin Jin、Shilin Wu、Zhibo Chen、Weip-ing Li、Zhizheng Zhang 和 Sen Liu。《用于图像修复的区域归一化》。AAAI人工智能会议论文集，34(07):12733–12740，2020年4月。5[74] Yu Zeng、Zhe Lin、Jimei Yang、Jianming Zhang、Eli Shecht-man 和 Huchuan Lu。《高分辨率图像修复

与迭代置信度反馈和引导式上采样。载于 欧洲计算机视觉会议，第1–17页。施普林格出版社，2020年。 5[75] 郑亮、沈丽月、田路、王生进、Jing-dong Wang 和田奇。《可扩展行人重识别：基准》。载于 *2015*年 *IEEE* 国际计算机视觉会议（*ICCV*），第1116–1124页。IEEE，2015年12月。2