arXiv:2303.05076v2 [cs.CV] 6 Mar 2024

用于步态识别与匿名化的行人属性编辑

Jingzhe Ma†、Dingqiang Ye†、Chao Fan†和Shiqi Yu

摘要—**作为一种生物特征识别方式，行人的步态信息因可在无需目标配合的情况下远距离采集，自工业界和学术界以来便受到广泛关注。近年来，这一研究方向带来了令人振奋的机遇，同时也面临着令人担忧的挑战：积极的一面是，用于嫌疑人检索和安全检查等安防应用的步态识别正展现出越来越大的潜力；消极的一面是，滥用步态信息可能引发隐私问题，因为即使在佩戴口罩、改变衣着的情况下，违法者仍可利用步态特征追踪关注对象。为应对这一把双刃剑，我们提出了一种名为 GaitEditor 的步态属性编辑框架。该框架能够对真实步态序列进行不同程度的属性编辑，同时保持视觉真实性，分别用于步态数据增强和去身份化，从而根据用户意图自适应地提升或降低步态识别性能。在实验方面，我们在三个广泛使用的步态基准数据集上，依据步态识别和匿名化协议进行了全面评估。大量实验结果表明，灵活运用 GaitEditor 能有效提升步态识别性能，并生成生动的去身份化可视化结果，以保护个人隐私。据我们所知，GaitEditor 是首个能够编辑多种步态属性，同时有益于步态识别和步态匿名化的框架。GaitEditor 的源代码将在 https://github.com/ShiqiYu/OpenGait 上发布。**

索引词—**步态识别、步态匿名化、行人属性编辑、基于风格的生成对抗网络**

一、引言

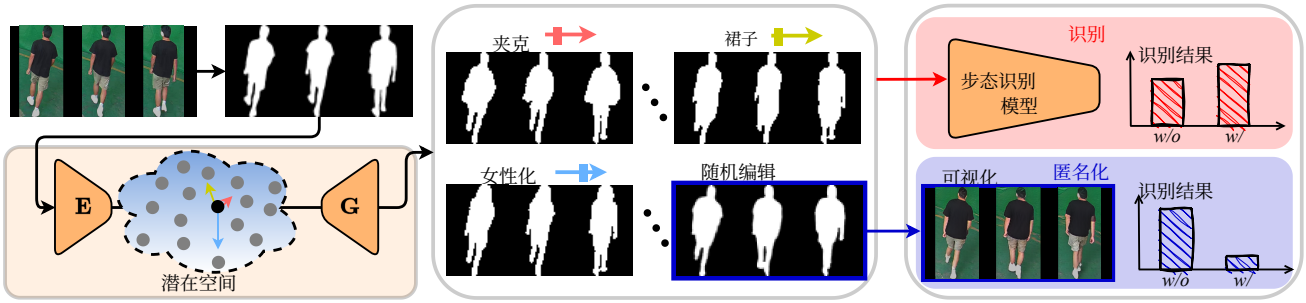
NOTHING 很完美。行人步态以身体重心的动态位移为特征，涵盖下肢和躯干的协调运动，以实现运动 [1]。值得注意的是，步态信息无需目标配合，且可轻松从远距离获取，因此其应用具有双刃剑效应。从积极方面来看，步态是一种用于个体识别的生物特征模态，尤其应用于步态识别领域 [2]。这一能力在犯罪调查、监控系统和社会安全中发挥着关键作用 [3]–[6]。但另一方面，利用步态信息可能会暴露行人的个人信息，从而引发隐私问题。即使行走时面部被遮挡或替换、服装有所改变，行人步态所固有的独特特征仍然可以被识别和追踪。

近年来，随着深度生成方法 [7]–[10]，特别是基于风格的生成对抗网络（生成对抗网络） [9]，[11]–[13],的兴起，物体属性编辑持续受到越来越多的关注。在这一更广泛的背景下，行人步态属性编辑（为方便起见，简称步态属性编辑）成为一个独特的子领域，专注于操控行人在行走时的身体外观（空间维度）、运动（时间维度）或这两个方面。已有多项研究表明，步态属性编辑方法能够有利于识别 [14]–[17]或去身份化 [18]，[19]，即，**仅关注其中积极或消极的一面**。本文旨在设计一种统一的步态属性编辑方法，以期灵活驾驭这一把双刃剑。

在步态识别领域，一项主要挑战是应对可能显著影响步态外观的各种因素 [6],，例如跨服装、跨视角和跨携带物挑战。因此，许多研究 [14]–[17] 尝试通过使用生成式方法编辑或生成步态属性来应对这些挑战，从而提升步态识别的性能。尽管这些方法使这些因素变得多样化，并生成了生动的轮廓图，但它们要么在训练过程中严重依赖成对数据样本 [14]–[16], ，要么创建规模固定的虚拟步态数据集 [17]。在步态匿名化领域，已有多项研究提出，其主要目标是在尽量减少对整体纹理外观修改的同时，防止步态识别模型识别出个体 [19]。通常，这些方法会操控个体特有的因素，例如静态体型特征 [20]–[22],、动态运动特征，或两者兼有 [18], [19]。然而，这些方法往往会以同质化的外观编辑不同个体的体型和步行阶段，这可能导致匿名化步态序列被深度学习模型逆向还原。为可靠地保护步态身份，还必须考虑步态属性变化的多样性。

上述关于步态识别和匿名化的讨论反映了这些方法的主要问题：它们只能编辑有限的属性。本文通过利用基于风格的生成对抗网络的潜在空间，并控制不同语义方向上的操控程度，发现步态属性的多样性可以得到提升。这一进展对于扩充步态数据集和实现行人匿名性都至关重要，最终能够实现两个目标：提高识别准确率并保护

Jingzhe Ma、Dingqiang Ye、Chao Fan 和 Shiqi Yu 均来自南方科技大学计算机科学与工程系，中国深圳 518055。电子邮箱：{12031127, 12232422、 12131100}@mail.sustech.edu.cn，以及 yusq@sustech.edu.cn†表示这些作者对本研究贡献相同。通讯作者：Shiqi Yu。



(a) 方法：GaitEditor **(b) 编辑结果** **(c) 应用**

图1：我们提出了**GaitEditor** （子图(a)），用于对真实步态序列（子图(b)）进行属性条件化的语义编辑，从而提升步态识别性能，同时缓解隐私问题（子图(c)）。对于**GaitEditor**，我们通过精心设计的编码器**E**将真实步态数据投影到构建的潜在空间中，沿多个语义方向操纵潜在编码，并通过生成器**G**将修改后的编码投影到图像域。

个人隐私。值得注意的是，我们主要关注对行人身体外观的编辑。

与以往使用基于风格的生成对抗网络进行人脸编辑任务的方法 [23]–[27], 不同，**步态本身固有的四项关键挑战**需要加以考虑：a) 二值步态轮廓仅包含稀疏的结构信息，例如边缘和形状，这可能使语义可控的步态属性难以解耦。b) 步态数据是序列，而非单幅图像，因此编辑步态序列必须保持时间连续性和身份一致性。c) 在分别进行人脸编辑 [28] 和物体编辑 [29],时，姿态变化通常会损害人脸和物体的保真度。因此，跨视角编辑步态可能是一项颇具挑战性的任务。d) 步态识别与匿名化的目标不同，这带来了额外挑战：如何将设计好的统一框架有效应用于这两项不同的任务。

为克服这些挑战，我们提出了一种开创性的步态属性编辑方法，称为**GaitEditor**，如图1（a）所示。首先，训练一个基于风格的生成对抗网络，将随机噪声编码转换为逼真轮廓图，从而构建语义丰富的潜在空间。在该空间中，可以发现各种语义可控的方向。随后，我们通过编码器应用生成对抗网络反演 [30] ，将真实步态序列投影到该潜在空间，同时保持时间和身份完整性。具体而言，我们集成了一个时空混合模块，并精心设计了身份损失，以确保这些特性不受破坏。沿这些方向操控潜在编码，即可编辑输入步态序列中对应的属性。此外，我们在训练中引入了视角转换分支，以应对跨视角编辑的挑战。

在实验中，我们开展了一项用户研究，系统地分析编辑后属性的真实性。结果表明，GaitEditor能够编辑多种步态属性，例如衬衫、裤子、视角、体型、年龄特征和性别特征，如图1(b)所示。此外，GaitEditor的设计简单明了，可作为即插即用模块轻松集成到下游任务中，如图1(c)所示。得益于这种便捷的集成方式，我们发现，巧妙控制编辑的强度和频率（即编辑程度）会影响编辑前后的身份一致性。对于输入步态序列，轻微、细微的编辑造成的身份一致性损失很小，适用于步态数据增强。相反，强烈、大幅度的编辑可能会削弱身份一致性，但无论以轮廓图还是RGB格式呈现，仍能带来生动的可视化效果。这些发现显著促进了GaitEditor在步态识别和匿名化中的应用。总而言之，本文作出以下贡献。

- 据我们所知，GaitEditor 是首个尝试使用单一、简洁的生成器，以无监督方式编辑多种步态属性的方法，显著提升了步态编辑研究的实用性。
- 我们成功地将 GaitEditor 用作在线数据增强模块，以增强步态识别算法。在具有挑战性的跨服装数据集 CCPG 上，多个广受欢迎的步态模型均取得了显著提升。具体而言，在各种行走条件下，GaitSet、GaitPart、GaitBase 和 DeepGaitV2 的平均 Rank-1准确率分别提高了 +1.00%、 +2.12%、 +1.60% 和+1.29%。
- 在步态匿名化方面，GaitEditor 可以对步态序列中的身份进行匿名化，同时保留自然的纹理外观信息。在 GaitBase 框架下，GaitEditor 在三个具有挑战性的数据集 CCPG、OU-MVLP 和 Gait3D 上，分别使 Rank-1准确率大幅下降 89.24%、57.34% 和 56.00%，展现了其保护步态数据中个人身份的强大能力。

II. 相关工作

A. 步态识别

现有的步态识别方法通常可分为两类：基于模型的方法和基于外观的方法。其中，前者以人体的二维/三维底层结构（如骨架和SMPL模型）为输入，用于推断身份信息 [31]–[37]。这些人体模型通常会明确排除携物和穿着物品等与步态无关的视觉因素。

在此过程中，人体形状这一最具区分性的步态特征之一通常会同时受到破坏，甚至被移除。这被认为是基于模型的方法相比更关注步态外观的基于外观的方法，识别性能不尽如人意的主要原因 [38], [39]。

基于外观的步态识别方法 [38], [40]–[44]主要从轮廓图或 RGB 图像中提取步态特征，因此易于捕捉形状特征。这些方法通常侧重于空间特征提取和步态时序建模。特别是，GaitSet [40] 是最知名的方法之一，其创新之处在于将步态序列视为无序集合。GaitPart [41]细致地探索输入轮廓图的局部细节，并利用微运动捕获模块 [41]对时序依赖关系进行建模。GaitGL [42]指出，全局空间步态表征往往会忽略重要细节，而局部的基于区域的描述符则无法捕捉相邻部位之间的关系。因此，GaitGL引入了全局和局部卷积层 [42]。OpenGait [38] 从实验角度重新审视了几项具有代表性的工作，并构建了稳健的基线模型 GaitBase。近期，DeepGaitV2 [43]提出了一种统一视角，用于探索如何构建先进的步态识别深度模型，并在多个具有挑战性的基准上取得了突破性进展。

B. 步态匿名化

步态匿名化方法分为视觉抽象和替换两类，通过对相关人体区域进行像素化或模糊处理，或微妙地变形轮廓图，以防止步态识别模型成功识别身份。基于视觉抽象的步态匿名化研究有限。具体而言，Agrawal 等人 [20]提出了一种技术，通过对视频中的整个人体区域应用模糊滤波器来实现步态匿名化。Mitsugami 等人 [21] 建议在监控视频中用棒状符号覆盖单个人体区域，以此保护个人隐私。然而，基于视觉抽象的方法往往会造成不自然的外观。为了解决这一问题，Tieu 等人在一系列研究中 [22], [45], [46]提出了基于替换的步态匿名化方法。他们具体采用生成式方法，将目标轮廓图与噪声轮廓图结合，合成一幅与目标轮廓图略有不同的新轮廓图。此外，Hirose 等人 [19] 从体型和步行阶段两个方面对步态序列中每一帧的轮廓图进行变形，以防止准确识别输入视频中的人物。然而，这些方法往往以同质化的外观编辑不同人物的体型和步行阶段，这可能导致匿名化步态序列被深度学习模型还原。相比之下，GaitEditor能够自主、反复地编辑各种属性，从而确保步态属性的多样性。

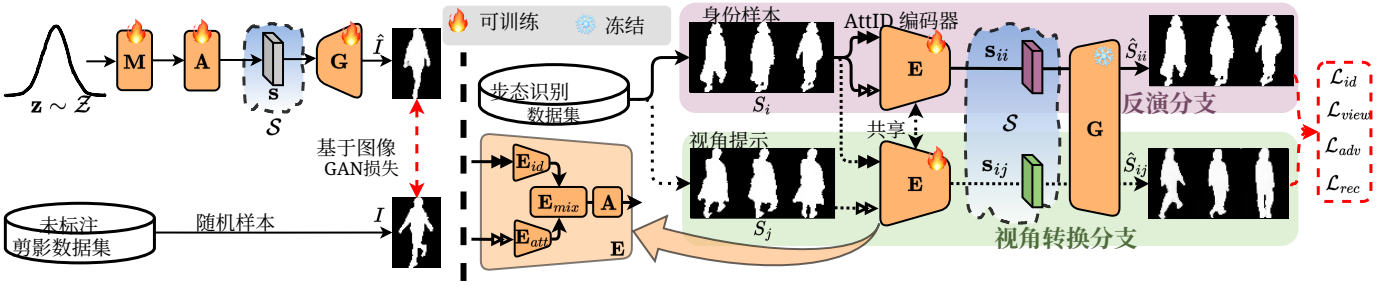
C. 步态数据合成

许多方法利用广受欢迎的生成模型来生成逼真的步态数据。例如，Yu 等人在 [14] 通过生成对抗网络（GAN）生成了视角不变的步态能量图（GEI） [47] 。他们使用身份判别器来确保生成的步态能量图具有身份一致性。Li 等人 [15] 通过对比学习在潜在空间中解耦特定协变量属性，以合成新的步态能量图，进而将协变量特征从一个受试者转移到目标受试者。Chen等人 [16] 提出了多视角步态生成对抗网络（MvG-GAN），用于生成具有不同视角的步态序列，以弥补现有步态数据集视角不足的问题。Xiang 等人 [48]将行人重识别和步态识别的先验知识引入训练损失，以便在转移受试者视角时保持身份一致性。近期，Zhang等人 [17] 利用 Unity3D 工具包构建了大规模虚拟步态数据集，从而丰富步态数据集中与身份无关的因素。尽管这些方法能够获得逼真的步态数据，但它们需要带标签的数据对进行训练，或者旨在通过操作单一的特定步态属性（例如摄像机视角 [16], [48] 或携带物品的变化 [15]）来构建固定规模的虚拟步态数据集 [17] 。相比之下，GaitEditor可以作为在线模块，为输入步态序列增添各种步态属性变化。

D. 基于风格的生成对抗网络与生成对抗网络反演

基于风格的生成对抗网络，即 StyleGAN [9] 指通过隐式学习到的层次化潜在风格来生成图像。具体而言，该模型利用风格潜在编码（由映射网络获得）和随机变化（通过噪声层实现）来进行图像合成。为提高生成图像的感知质量，研究者提出了 StyleGAN2 模型 [11]。为了在数据有限的情况下确保 StyleGAN2 的训练过程稳定，StyleGAN2- ADA [12]引入了自适应判别器增强机制。这些模型的潜在空间包含丰富且相互解耦的语义方向，可用于操控生成图像的各种属性，例如面部表情、妆容、物体颜色、图像风格等。

为了将这一特性应用于真实图像，研究人员提出了生成对抗网络反演方法。生成对抗网络反演方法通常可分为基于优化的方法 [23], [25], [49],[50]和基于编码器的方法 [27], [51]–[56]。虽然基于优化的模型具有较高的反演质量，但针对每张输入图像都需要进行大量优化步骤 [57], ，导致耗时显著。为克服这一局限，基于编码器的方法应运而生，成为一种有前景的解决方案，可实现实时运行 [27], [51]–[56]。这类方法旨在训练编码器，将真实世界图像映射到潜在空间。尽管生成对抗网络反演在基于图像的人脸编辑方面取得了显著成效，但这一技术在步态属性编辑领域尚未受到关注。本文考虑到潜在空间蕴含丰富语义信息的特性，并将其用于促进步态属性编辑。然而，与基于 RGB 的人脸图像不同，包含二值轮廓的步态序列在空间维度上的结构信息较为稀疏



（a）阶段1： 构建潜在空间　　（b）阶段2： 训练用于反演和视角转换的编码器

图 2：GaitEditor 的整体训练架构。训练过程分为两个阶段。(a) 在无标签剪影数据集上训练基于图像的生成对抗网络模型，以构建潜在空间。(b) GaitEditor 包含两个分支：反演和视角转换。第一个分支将源步态序列投影到已构建的潜在空间中；另一个分支则着重将目标样本的视角转换为源样本的视角。为确保保留步态序列中的所有相关信息，我们设计了四种损失函数：身份损失（$\mathcal{L}_{id}$）、视角损失（$\mathcal{L}_{view}$）、视频对抗损失（$\mathcal{L}_{adv}$）和重建损失（$\mathcal{L}_{rec}$）。

，而在时间维度上则包含更多个体运动特征 [18]。这一差异使步态反演尤其具有挑战性。

III. 方法

如图 2 所示，我们的 GaitEditor 训练方案包含两个阶段：构建潜在空间（阶段1），以及将真实步态序列投影到该潜在空间中，同时解决视角转换相关的挑战（阶段2）。阶段1涉及构建一个融入语义信息、可供编辑的潜在空间，具体内容将在 III-A 小节中阐述。阶段2侧重于训练一个统一的编码器，将真实步态序列投影到该潜在空间中，并处理视角转换带来的挑战，如 III-B 小节所示。随后，III-C 小节将介绍训练 GaitEditor 的训练方案和精心设计的损失函数；III-D 小节则详细说明如何利用我们构建的潜在空间编辑输入步态序列的步态属性。最后，我们将在 III-E 小节中阐述 GaitEditor 在步态识别和匿名化方面的应用。

A. 潜在空间构建

编辑各种步态属性的能力，在很大程度上取决于所学习到的潜在空间的语义可分性。StyleGAN-ADA [12]，是一种强大的基于风格的生成对抗网络模型，以其在潜在空间中出色的解耦特性而闻名，尤其是在面部属性编辑场景中表现突出 [24], [26], [58]。为了将这些潜在空间的能力应用于步态轮廓图像，我们在 OU-MVLP数据集 [59] 上重新训练了 StyleGAN-ADA，遵循官方网络设计 [12]，如图 2 (a) 所示。形式上，给定输入潜在空间 $\mathcal{Z}$中的一个噪声向量 $\mathbf{z}$ ，其服从高斯分布ˆ，合成图像 I 可按如下方式生成：

$$\hat{I} = \mathbf{G}(\mathbf{A}(\mathbf{M}(\mathbf{z}))), \quad \mathbf{z} \in \mathcal{Z}, \tag{1}$$

其中$\mathbf{G}$ 表示合成网络。符号$\mathbf{M}$表示一个非线性映射网络，用于将噪声向量$\mathbf{z}$ 编码为中间潜在向量，$\mathbf{A}$ 表示线性层，用于将中间潜在向量投影到 $\mathcal{S}$ 空间 [26], 中的$\mathbf{s}$ ，而$\mathbf{G}$ 表示生成器网络。值得一提的是，这一训练过程不依赖身份级标签。考虑到真实轮廓图像通常会调整至 64×64，且像素值为二值，因此我们采用该尺寸作为生成器$\mathbf{G}$的默认输出分辨率。此外，我们使用无标签轮廓图像数据集和图像判别器进行对抗学习。经过这一阶段的训练后， $\mathcal{S}$ 潜在空间得以构建，生成器$\mathbf{G}$ 可以接收潜在向量$\mathbf{s}$ 作为输入，生成轮廓图像。

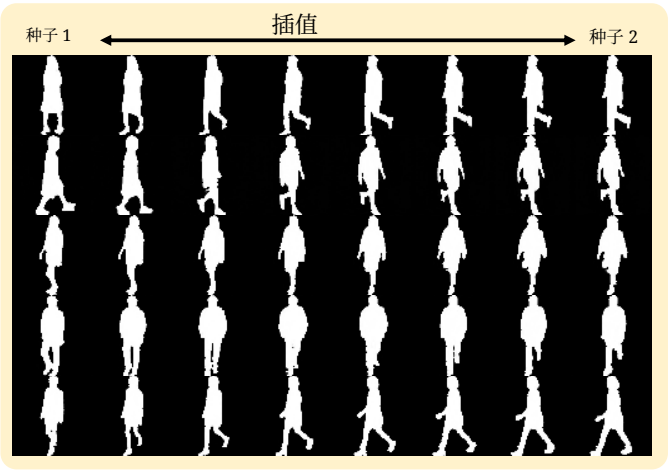


图 3：我们构建的潜在空间的插值结果。每行序列展示了两个独特随机噪声之间的各个阶段。每行的第一张和最后一张图像分别是根据这两个随机噪声生成的合成可视化图像。

为展示在由二值步态轮廓构建的潜在空间中解耦步态属性的有效性，我们对两个输入噪声进行插值，并利用中间结果生成图像。如图 3 所示，在对两幅合成轮廓图进行插值时，生成图像的属性呈现出

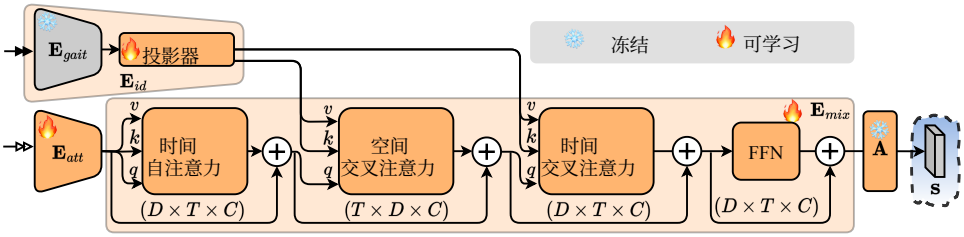


图4：属性-身份编码器的架构。

平滑过渡，而其他属性保持不变。这一现象表明，由二值步态轮廓构建的潜在空间同样包含丰富的语义信息。

在人脸编辑中，有许多基于StyleGAN的编辑方法可用于在其潜在空间中寻找语义方向，例如StyleSpace [26], SeFa [60], InterFaceGAN [24]。在我们的研究中，我们利用StyleSpace [26] 寻找语义方向，以编辑轮廓图像中相应的图像区域。经过详细的人工筛选，我们确定了十个与十种特定行人属性相关的方向，包括性别特征、体型、年龄特征和服装类型。值得注意的是，这些属性的内涵本质上具有主观性，并以人类感知为基础。例如，当观察到行人的头发较长时，我们将其解释为女性气质；这一趋势在OU-MVLP数据集中很常见。我们通过用户研究严格验证了这些属性的有效性和感知一致性。研究结果见第IV-C小节。

此外，在筛选过程中，我们发现沿特定方向操纵潜在编码，只会导致拍摄视角发生轻微变化。其主要原因是，视角变化会显著影响轮廓图的外观，即 仅操纵单一方向无法完成视角转换任务。操纵合成轮廓并非我们的最终目标，我们关注的是在潜在空间中编辑真实步态序列。基于上述发现，我们提出了一种名为属性-身份（AttID）编码器**E** 的双流编码器架构，将真实步态序列投影到既定的潜在空间中，并解决视角转换这一难题。

B. 属性-身份编码器

在阶段2中，GaitEditor 包含两个分支：反演分支（紫色背景）和视角转换分支（绿色背景），如图 2 (b) 所示。前者用于将真实步态序列投影到构建的潜在空间中，后者用于处理视角转换这一挑战。在此阶段，预训练线性层 $\mathbf{A}$ 和生成器 $\mathbf{G}$ 始终处于冻结状态。对于特定的步态识别数据集，随机采样两个不同的真实步态序列 S_i和 S_j ，分别记为身份样本和视角提示。对于反演任务，仅将 S_i 输入 AttID 编码器 $\mathbf{E}$，再输入线性层 $\mathbf{A}$ ，以获得潜在编码 $\mathbf{s}_{ii}$。与此同时，对于视角转换任务，AttID 编码器 $\mathbf{E}$ 接收 将 S_i 和 S_j 都作为输入，然后将 $\mathbf{E}$的输出输入线性层 $\mathbf{A}$ ，生成视角交换潜在编码 $\mathbf{s}^{ij}$。具体而言，AttID 编码器 $\mathbf{E}$ 由三个精心设计的模块组成：身份提取器 $\mathbf{E}_{id}$、属性提取器 $\mathbf{E}_{att}$和特征混合模块 $\mathbf{E}_{mix}$，分别负责提取身份、属性信息并将二者融合。形式上，获取中间潜在向量的过程可表示如下：

$$\begin{aligned}\mathbf{s}_{ii} &= \mathbf{A}(\mathbf{E}(S_i, S_i)) = \mathbf{A}(\mathbf{E}_{mix}(\mathbf{E}_{id}(S_i), \mathbf{E}_{att}(S_i))),\\ \mathbf{s}_{ij} &= \mathbf{A}(\mathbf{E}(S_i, S_j)) = \mathbf{A}(\mathbf{E}_{mix}(\mathbf{E}_{id}(S_i), \mathbf{E}_{att}(S_j))).\end{aligned} \tag{2}$$

最后，冻结的$\mathbf{G}$ 随后将$\mathbf{s}_{ii}$ 和$\mathbf{s}^{ij}$ 解码为虚假的步态ˆ ˆ 序列 S_{ii}，以及 S_{ij} 逐帧结果。形式上，投影与生成过程可表示如下：

$$\begin{aligned}\hat{S}_{ii} &= \mathbf{G}(\mathbf{s}_{ii}),\\ \hat{S}_{ij} &= \mathbf{G}(\mathbf{s}_{ij}).\end{aligned} \tag{3}$$

ˆ 值得注意的是， S_{ij} 包含与视角提示 S_j 相对应的合成视角，同时保留了身份样本 S_i的身份信息。

在模型架构方面，对于 $\mathbf{E}_{id}$，为减少步态分布的类内差异，我们采用预训练的基于外观的步态识别模型作为骨干网络（$\mathbf{E}_{gait}$），并在 GaitEditor 训练期间将其冻结。因此，为完成阶段（ii）中的步态序列重建任务，可学习的 $\mathbf{E}_{att}$ 将被迫作为 $\mathbf{E}_{id}$的补充，即 尽可能充分地挖掘输入步态序列的多样性。同时，对于 $\mathbf{E}_{att}$，按照 [51]，我们使用以 ResNet18 [61] 为骨干网络的类 pSp 网络。具体而言，$\mathbf{E}_{att}$ 根据输入的视角提示逐帧处理轮廓图，得到维度为 $T \times D \times C$的特征，其中 T、 D和 C分别表示序列长度、嵌入维度和通道维度。同时，$\mathbf{E}_{gait}$ 将输入的身份样本逐帧压缩为形状为 $T \times P_{gait} \times C_{gait}$的身份嵌入，其中 P_{gait}、 C_{gait} 分别表示部位维度和通道 [38]。为使身份与属性特征的维度对齐，我们引入了一个可学习投影模块，该模块为非线性模块。经过该投影模块后，身份特征的形状变为 $T \times D \times C$。

在图4所示的架构中，我们的$\mathbf{E}_{mix}$ 包含四个关键组件：时间自注意力、空间交叉注意力、时间交叉注意力和前馈模块。这些组件协同工作，以融合身份与属性嵌入。首先，将$\mathbf{E}_{att}$ 的输出输入时间自注意力模块，以捕获帧间

关系，从而提升合成步态序列的行走真实性和帧间一致性。随后，E_{id}提取的身份嵌入同时作为键（k）和值（v），而时间自注意力与属性嵌入相加的结果则用作查询（q）。接着，分别将这些元素输入空间和时间交叉注意力模块。此过程有助于将身份嵌入与从视角提示中提取的嵌入进行帧内和帧间融合。在这一框架中，视角提示提供较粗粒度的全局信息，而身份样本则提供更细粒度的局部细节。通过采用交叉注意力模块，我们能够有效融合不同粒度的信息。

C. 损失函数与训练方案

在训练阶段，我们采用四种损失函数，即重建损失、身份损失、视频对抗损失和视角分类损失。

重建损失。 $\mathcal{L}_{rec}$ 可以直接保持生成序列的视觉真实性。因此，我们引入像素级和感知级重建项，其公式如下：

$$\begin{aligned}\mathcal{L}_{rec} &= \mathcal{L}_{pix} + \mathcal{L}_{per},\\ \mathcal{L}_{pix} &= \left\|\hat{S}_{ii} - S_i\right\|_2,\\ \mathcal{L}_{per} &= \left\|V(\hat{S}_{ii}) - V(S_i)\right\|_2,\end{aligned} \tag{4}$$

其中，$\|\cdot\|_2$ 和 V 分别表示 L_2 距离函数和在ImageNet [62] 上训练的VGG-16模型。 $\mathcal{L}_{per}$ 表示LPIPS损失 [63] ，该损失广泛用于保持视觉属性的相似性。需要注意的是，只有当两个输入步态序列完全相同时，重建过程才会起作用，即反演分支的情况。

视频对抗损失。为确保合成步态序列的真实性，我们设计了一种基于视频的对抗损失，该损失采用最小二乘GAN损失 [64] ，其公式如下：

$$\begin{aligned}\mathcal{L}_{adv}^{\mathbf{D}_{vid}} = \frac{1}{2}\sum_{y\in\{i,j\}}&\left(\mathbb{E}_{S_i\sim\mathcal{D}}\left[(\mathbf{D}_{vid}(S_i)-1)^2\right]\right.\\ &\left.+\mathbb{E}_{\hat{S}_{iy}\sim\hat{\mathcal{D}}}\left[(\mathbf{D}_{vid}(\hat{S}_{iy}))^2\right]\right),\\ \mathcal{L}_{adv}^{\mathbf{E}} = \frac{1}{2}\sum_{y\in\{i,j\}}&\mathbb{E}_{\hat{S}_{iy}\sim\hat{\mathcal{D}}}\left[(\mathbf{D}_{vid}(\hat{S}_{iy})-1)^2\right],\end{aligned} \tag{5}$$

其中，$\mathbf{D}_{vid}$ 表示视频判别器，它是一个采用3D卷积块的块判别器 [65]。 $\mathcal{D}$ 和 $\hat{\mathcal{D}}$分别表示真实步态数据集和合成步态数据集。在对抗训练过程中， $\mathcal{L}_{adv}^{\mathbf{D}_{vid}}$和 $\mathcal{L}_{adv}^{\mathbf{E}}$ 交替进行优化，分别更新视频判别器$\mathbf{D}_{vid}$ 和AttID 编码器$\mathbf{E}$的训练参数。对抗过程作用于所有分支，即反演分支和视角转换分支。

身份损失。 为了保持合成步态序列的身份一致性，我们再次利用现成的步态识别模型 $\mathbf{E}^{gait}$ ，该模型已在图4中提及，以构成身份约束。具体而言，对于反演分支，我们利用均方误差损失和余弦相似度损失ˆ来缩小 S_i 和 S_{ii} 在身份嵌入空间中的距离。

$$\begin{aligned}\mathcal{L}_{id}^{inv} = &(1-\left\langle\mathbf{E}_{gait}(\hat{S}_{ii}),\mathbf{E}_{gait}(S_i)\right\rangle)\\ &+\left\|\mathbf{E}_{gait}(\hat{S}_{ii})-\mathbf{E}_{gait}(S_i)\right\|_2,\end{aligned} \tag{6}$$

其中，$\langle\cdot,\ \cdot\rangle$ 表示余弦相似度函数。

对于视角转换分支，我们在身份嵌入空间中采用InfoNCE损失 [66]。具体而言，考虑一个批量大小为 N的小批量，并将 $(S_{ij}^n,\ S_i^n)$ 定义为正样本对，将 $(S_{ij}^n,\ S_i^m)$ 与 $n\ \neq\ m$ 以及 $(S_{ij}^n,\ S_j^l)$ 定义为负样本对，其中$n,\ m,\ l\in\{1,\ 2,\ \cdots,\ N\}$，我们将身份InfoNCE损失表述如下：

$$\begin{aligned}\mathcal{L}_{id}^{vt} &= -\frac{1}{N}\sum_{n=1}^{N}[\log(M_{id}^n)],\\ M_{id}^n &= \frac{Q(\hat{S}_{ij}^n,S_i^n)}{\sum_{m=1}^{N}Q(\hat{S}_{ij}^n,S_i^m)+\sum_{l=1}^{N}Q(\hat{S}_{ij}^n,S_j^l)},\\ Q(A,B) &= \exp(\mathrm{sim}(\mathbf{E}_{gait}(A),\mathbf{E}_{gait}(B))/\tau),\end{aligned} \tag{7}$$

其中

$$\mathrm{sim}(\mathbf{E}_{gait}(A),\mathbf{E}_{gait}(B)) = \frac{\mathbf{E}_{gait}(A)^\top\cdot\mathbf{E}_{gait}(B)}{\|\mathbf{E}_{gait}(A)\|_2\|\mathbf{E}_{gait}(B)\|_2}$$

表示余弦相似度，而 $\tau\ >\ 0$ 表示温度超参数，设为 16。

身份损失的总和为这两项之和，即：

$$\mathcal{L}_{id} = \mathcal{L}_{id}^{inv} + \mathcal{L}_{id}^{vt}. \tag{8}$$

视角损失。为保持生成的步态序列与视角提示之间的视角一致性，应引入属性损失。具体而言，我们使用在OU-MVLP [59] 上训练的视角分类器$\mathbf{E}_{view}$ 来构建损失：

$$\begin{aligned}\mathcal{L}_{view} = &D_{KL}(\mathbf{E}_{view}(\hat{S}_{ii})||\mathbf{E}_{view}(S_i))\\ &+D_{KL}(\mathbf{E}_{view}(\hat{S}_{ij})||\mathbf{E}_{view}(S_j)),\end{aligned} \tag{9}$$

其中 $D_{KL}(\cdot||\cdot)$ 表示 KL散度。

训练方案与总损失。在我们的研究中，我们首先训练反演分支，然后同时训练反演分支和视角转换分支。总损失定义为不同损失的加权组合：

$$\begin{aligned}\mathcal{L}_{total} = &\lambda_{rec}\mathcal{L}_{rec}+\lambda_{id}\mathcal{L}_{id}+\lambda_{view}\mathcal{L}_{view}\\ &+\lambda_{adv}(\mathcal{L}_{adv}^{\mathbf{D}_{vid}}+\mathcal{L}_{adv}^{\mathbf{E}}).\end{aligned} \tag{10}$$

我们将 $\lambda_{rec}\ =\ 1$, $\lambda_{id}\ =\ 0.6$, $\lambda_{view}\ =\ 0.1$, $\lambda_{adv}\ =\ 0.4$ 设为仅用于反演分支训练，并将 $\lambda_{rec}=0.6$, $\lambda_{id}=1.0$,$\lambda_{view}\ =\ 1.2$, $\lambda_{adv}\ =\ 0.1$ 设为同时用于反演分支和视角转换分支训练。

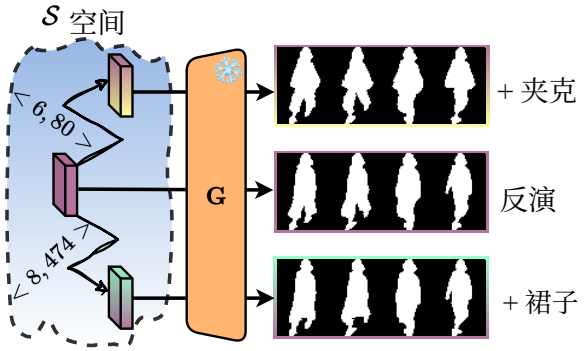


图5：非视角属性编辑示意图。

D. 步态属性编辑

通过上述训练过程，GaitEditor 展现出能够有效地将真实步态序列投影到构建的 $\mathcal{S}$ 空间，并应对视角属性编辑挑战的能力。

编辑非视角属性。在这一阶段中，我们希望使用阶段1中选出的十个方向来编辑真实步态序列。因此，我们采用反演分支的流程。该阶段将接收两个相同的真实步态序列作为输入，通过 $\mathbf{E}$ 将其投影到 $\mathcal{S}$ 空间中，然后在 $\mathcal{S}$空间中操纵语义方向，以编辑非视角属性。 $\mathcal{S}$ 空间中的语义方向通常可表示为 $< l, c >$，其中 l和 c 分别表示层和通道维度。对于潜在编码 s，我们可以操纵其第 l层和第c个通道，以编辑相应的步态属性。图5展示了如何使用特定方向 $< l, c >$操纵潜在编码 s_{ii} 。操纵过程可表示为：

$$s_{ii}[l][c] \mathrel{+}= \sigma(s_{ii}[l])[c] * (\alpha * \beta), \tag{11}$$

其中， $\sigma(s_{ii}[l])$ 表示 l层每个通道的标准差， α 表示编辑强度的数量，β 表示编辑强度的系数。在我们的实验中，方向 $< 8, 474 >$和 $< 6, 80 >$分别表示穿着衬衫和夹克的操纵。

编辑视角属性。在这一过程中，使用视角转换分支的流程，并且$\mathbf{E}$ 将接收两个不同的真实步态序列，即 S_i 和 S_j作为输入。随后，将 S_i 视为身份样本，主要提供身份信息；将 S_j 视为视角提示，主要提供视角信息。最后，通过我们的$\mathbf{E}$ 和$\mathbf{G}$生成合成步态序列 S_{ij} ，其视角为 S_j ，身份为 S_i。

E. 应用

如图 6 所示，GaitEditor 可以作为即插即用模块，轻松集成到步态识别和匿名化任务的工作流程中。在我们的实验中，我们将利用 $\mathcal{S}$ 空间中的全部语义方向。虽然阶段1识别出了十个可用于编辑十种不同步态属性的语义方向，但 $\mathcal{S}$ 空间还包含范围更广的其他语义方向。这些向量能够以不同程度编辑轮廓图像中的特定区域，从而为进一步探索步态识别和匿名化任务提供丰富的多样性。

为了进一步利用这些丰富的语义方向，我们引入了一种随机编辑迭代策略。具体而言，我们对一个步态序列执行多次编辑，每次随机选择一个语义方向，并调整不同的强度系数 β，如公式 11所示。为便于说明，我们将编辑次数命名为编辑频率ξ，并将强度系数命名为编辑强度 β。实验观察表明，编辑强度和编辑频率都会显著影响生成的步态序列中因素的多样性以及身份保持程度，如图 7所示。为了定量验证这一点，我们在CCPG 数据集 [67]上使用三个指标：Rank-1准确率、已知因素特征角度（KFFA） [68],和Fr´échet Inception Distance（FID） [69], ，在不同的编辑频率和编辑强度下分别评估身份保持程度、因素多样性以及生成的步态序列质量。

具体而言，为评估 Rank-1准确率，我们使用 GaitEditor 编辑 CCPG 数据集的整个测试探针集 [67]，并设置不同条件，从而评估 GaitBase 模型 [38] 的识别准确率。对于 KFFA，我们使用 GaitSSB [70]将生成的 CCPG 数据集样本映射到其潜在空间。类似地，我们采用 192 通道版本的 InceptionV3 [71] 计算 FID。如图 7 (a) 所示，结果表明识别准确率与编辑程度呈负相关，这说明编辑频率或编辑强度的增加会因引入变化而降低准确率。相反，如图 7 (b) 所示，KFFA指标呈现上升趋势，表明更高的编辑频率或编辑强度能够提升步态特征的多样性。编辑程度越高，FID分数也越高；但如图 7 (c) 所示，当编辑频率低于 32 且编辑强度低于 4 时，FID分数仍处于可接受范围内。因此，我们谨慎控制编辑强度和编辑频率，以分别提升步态识别和匿名化任务的性能。

1）步态识别：在此应用场景中，GaitEditor 作为步态模型训练阶段的在线数据增强工具，旨在提升模型在各种挑战基准上的识别准确率。步态识别面临的一项重大挑战是训练集中的因素多样性有限，这会对识别性能造成不利影响。GaitEditor 通过调节编辑频率和编辑强度来解决这一问题，如图5（上方分支）所示。对于每个步态序列，都会进行多次编辑，以编辑强度系数 β 随机修改视角和非视角属性。对于非视角属性编辑，GaitEditor 首先将真实或已编辑的步态序列投影到 $\mathcal{S}$ 空间，得到一个潜在编码。随后，多次沿不同方向调整该编码，生成最终的步态序列。对于视角属性，GaitEditor 使用 AttID 编码器，自适应地将视角从提示中迁移

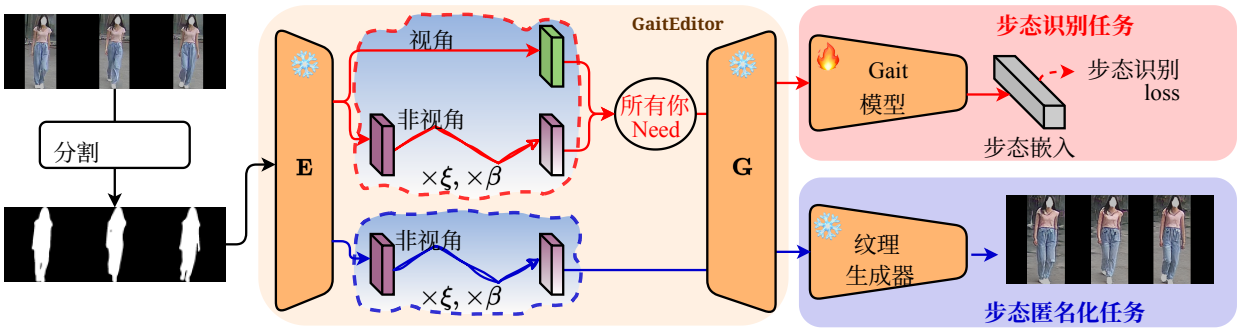
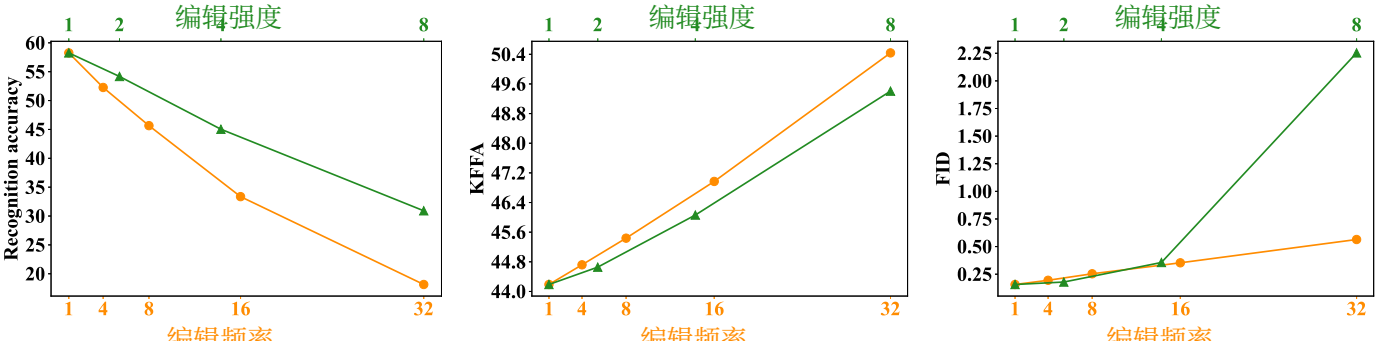


图6：GaitEditor应用概览。其中， ξ 和 β 分别表示编辑频率和编辑强度。



(a) 识别准确率指标。 (b) KFFA指标。 (c) FID指标。

图 7：在不同编辑强度和频率下，三项指标的表现：(a) 识别（Rank-1）准确率，(b) KFFA，(c) FID。

表I：展示了通过编辑各种属性实现的身份一致性结果。术语“身份相似度”指编辑后的步态序列与真实步态序列之间的余弦相似度。“反演”指使用我们的 GaitEditor 重建输入样本，代表身份相似度的上界。相反，“负样本对”表示计算两个不同个体之间的身份相似度，作为身份相似度的下界。

编辑属性	反演	肥胖	瘦弱	女性化	年长	年轻
身份相似度 (CCPG |OU-MVLP)	0.9785 | 0.9336	0.9399 | 0.7985	0.9327 | 0.8026	0.9322 | 0.8319	0.9457 | 0.8500	0.9329 | 0.8610
编辑属性	夹克	裙子	裤子	连帽衫	视角	负样本对
身份相似度 (CCPG |OU-MVLP)	0.9576 | 0.8039	0.9513 | 0.7872	0.9295 | 0.7612	0.9202 | 0.7167	0.9380 | 0.7813	0.8806 | 0.5153

到身份样本。为在引入因素多样性的同时保持身份一致性，我们将 ξ 设为 1 到 8，将β 设为 0 到 2。这些参数设置的具体取值取决于训练数据集，如表 III 所示。

2）步态匿名化：在这一特定应用场景中，GaitEditor 主要用于匿名化步态序列中的身份，同时尽可能减少视觉失真。根据图 7 的结果，我们观察到，随着 ξ 或β增加，识别准确率会下降，但对图像质量的影响可以忽略不计。利用这一发现，GaitEditor 能够在不显著改变个体外观的情况下，对步态序列中的身份进行匿名化。具体而言，该过程首先检测并分割输入的步态序列，以提取其轮廓图。随后，GaitEditor 在 $\mathcal{S}$ 空间中以编辑强度 β 多次编辑该轮廓图序列的属性。在本研究中，我们将 ξ 设为 32，将 β 设为 1，用于步态匿名化任务。随后，预训练的图像到图像生成网络 VQGAN [72]，会将编辑后的轮廓图序列转换为逼真的 RGB 视频格式。该流程详见图6（下方分支）。

IV. 实验

A. 数据集和评估指标

在本研究中，我们采用了四个公开步态数据集，分别是应用最广泛的数据集 OU-MVLP [59]，、规模最大且包含换装情况的数据集 CCPG [67]，、因素最全面的数据集 CASIA-E [73]，，以及最新且广受欢迎的野外环境数据集 Gait3D [39]。除非另有说明，我们的实现均遵循官方协议，包括训练/测试划分以及图库/探测集划分策略。

OU-MVLP [59]是目前最具规模的室内步态数据集之一。该数据集涵盖10,307名个体的数据，每位参与者的数据均反映单一的

表 II：不同属性的属性正确性。我们开展了一项用户研究进行评估。有Exp. 表示受试者具有步态经验，无Exp. 表示受试者没有步态经验。

	准确率（%）										
	肥胖	瘦弱	男性化	女性化	年长	年轻	夹克	裙子	裤子	连帽衫	Mean
有 实验。	72.41	88.37	74.42	82.61	69.35	76.47	90.57	94.55	88.24	77.59	81.46
无 实验。	73.33	82.69	45.00	68.52	58.33	55.74	75.00	85.42	86.27	82.00	71.23
Mean	72.87	85.53	59.71	75.57	63.84	66.11	82.79	89.99	87.26	79.80	76.34

步行场景，即正常步行。该数据集通过从14个固定拍摄角度拍摄的视频记录每次步行，角度范围为 $[0°, 90°]$ 至 $[180°, 270°]$。因此，每位受试者包含28个序列。根据该数据集的官方协议，我们将受试者分为两组：5,153人用于训练，5,154人用于测试。在测试阶段，每位受试者的第一个序列（NM#0）用作探测集，而后续序列（NM# 1）则用于图库。

CCPG [67] 是规模最大的换装步态数据集。该数据集包含 200 名受试者，每名受试者穿着七种不同的服装风格，包括整套服装变换、上衣变换、裤装变换以及携带包袋等情形；每名受试者包含 10 个不同视角的序列。该数据集包含超过 16,000 个序列，涵盖 RGB 和轮廓模态。这些受试者和序列被分为训练集和测试集两组，其中前 100 名受试者用于训练，其余 100 名用于测试。测试子集中，我们按照官方指南设置了四种换装设置。1）换装（**携物**）：探测时使用 U0D0 和 U0D0BG 序列，图库集则使用 U1D1、U2D2 和 U3D3 序列。U0D0BG 表示不同的服装组合，其中“U”表示上衣，“D”表示裤子，“BG”表示包袋，“0”表示服装编号。2）上衣变换（**上坡**）：探测集包含 U3D3 序列，图库集则使用 U0D3 序列。3）裤装变换（**下坡**）：此设置中，探测时使用 U1D0 序列，图库则使用 U1D1 序列。4）包袋变换（**背包**）：探测时使用 U0D0BG 序列，图库则使用 U0D0 序列。

Gait3D [39]采集于一家大型超市，包含 1,090 小时的高分辨率视频，每段视频的分辨率为 1920×1080 像素，帧率为 25 FPS。该数据集由 4,000 名受试者的记录组成，包含超过 25,000 个步行序列。这些受试者被分为两个不同的组，分别用于训练和测试：3,000 名受试者用于训练，其余 1,000 名用于测试。在测试子集中，每个身份标识仅选取一个序列，构成包含 1,000 个序列的探测集；其余序列则构成图库集，共计 5,369 个序列。

对于 GaitEditor 的评估，我们采用峰值信噪比（PSNR）、结构相似性（SSIM） [74],FID [69],学习感知图像块相似度（LPIPS） [75],和均方误差（MSE）。前两项指标和 MSE 用于评估重建性能，FID 用于评估轮廓图像质量，以及使用 LPIPS 评估轮廓图像的感知质量。对于步态识别评估，采用 Rank-1准确率作为主要评估指标。对于步态匿名化评估，采用 Rank-1准确率和 FID。

表 III：不同实验的调参详情。LR 为初始学习率。如果 LR=1e-4，则前两个卷积块、后两个卷积块和头部层的初始学习率分别设为 LR、LR×10 和 LR×100。Exp_1 和 Exp_2 分别表示对不同数据集和步态模型的评估。

实验	数据集	步态模型	里程碑 / 步骤	LR	ξ	β
Exp_1	CCPG	GaitBase	(10K、15K) / 20K	1e-4	4	1
	OU-MVLP		(10K、15K) / 20K	1e-4	4	1
	Gait3D		(10K、15K) / 20K	1e-4	4	0.8
Exp_2	CCPG	GaitSet	(10K、20K、30K) / 40K	1e-1	4	1
		GaitPart	(10K、15K) / 20K	1e-4	4	1
		GaitBase	(10K、15K) / 20K	1e-4	4	1
		DeepGaitv2	(20K、40K、50K) / 60K	1e-1	8	2

B. 实现细节

在阶段1中，GaitEditor 在 OU-MVLP 训练集上进行训练，无需身份级别的标注。StyleGAN2-ADA [12] 模型 [1] 采用 16500 (∼3d) tricks 进行训练。我们使用多种 ADA 选项，例如水平翻转、图像空间滤波、图像损坏、各向同性缩放、整数平移、各向异性缩放、亚像素平移，以及正负十度旋转。其他超参数设置与官方协议 [12]相同。使用 2048 通道版本的 InceptionV3 得到的 FID 分数 [69], 为 2.93，证明了生成轮廓图像的逼真度。

在阶段2中，我们冻结预训练的 StyleGAN2-ADA 模型，并训练 AttID 编码器 $\mathbf{E}$ 和视频判别器$\mathbf{D}_{vid}$。对于不同的数据集，我们分别利用相应数据集的训练集来训练不同的阶段2。我们使用官方提供的 GaitBase 网络 [38] 作为$\mathbf{E}_{id}$，该网络已在相应数据集上完成预训练。我们首先在第一个 N_{epoch} epoch 训练反演分支，然后同时训练反演分支和视角转换分支。对于 OU-MVLP数据集，我们设置 $N_{epoch} = 8$，总训练轮数为 50（∼25h）。对于 CCPG 和 Gait3D 数据集，我们设置 $N_{epoch} = 32$，总训练轮数为 100（∼18h）。AttID 编码器$\mathbf{E}$ 和视频判别器$\mathbf{D}_{vid}$ 使用 Adam [76], 进行优化，学习率分别为 1e-4 和 2e-4。

[1]公开发布于 https://github.com/NVlabs/stylegan2-ada-pytorch.git

表IV：以 GaitBase 为基线时，不同数据集的识别准确率。各条件下列出的 Rank-1准确率百分比展示了步态模型处理这些特定变化的性能。

数据集	GaitEditor	Rank-1准确率（%）	
		不同条件	Mean
OU-MVLP	✘	正常行走：89.85	89.85
	✔	正常行走：90.27	**90.27**
CCPG	✘	携物：71.56，上坡：75.00，下坡：76.85，背包：78.52	75.48
	✔	**携物：73.21，上坡：76.98，下坡：77.81，背包：80.32**	**77.08**
Gait3D	✘	64.60	64.60
	✔	**65.40**	**65.40**

表 V：不同步态模型在 CCPG 数据集上的识别准确率。各条件下列出的 Rank-1 准确率百分比展示了步态模型处理这些特定变化的性能。

步态模型	GaitEditor	Rank-1 准确率 (%)				
		CL	UP	DN	BG	Mean
GaitSet	✘	60.18	65.22	65.11	68.53	64.76
	✔	**61.30**	**66.09**	**66.21**	**69.45**	**65.76**
GaitPart	✘	64.25	67.76	68.58	71.68	68.07
	✔	**65.28**	**70.15**	**71.36**	**73.98**	**70.19**
GaitBase	✘	71.56	75.00	76.85	78.52	75.48
	✔	**73.21**	**76.98**	**77.81**	**80.32**	**77.08**
DeepGaitv2	✘	78.37	84.67	80.95	89.45	83.36
	✔	**79.76**	**85.70**	**82.30**	**90.83**	**84.65**

C. 属性编辑评估

在本节中，我们分别通过余弦相似度和用户研究，评估编辑不同属性时的身份一致性和属性正确性。

身份一致性评估。在 CCPG [67]和 OU-MVLP 数据集 [59] 上使用 GaitBase 模型 [38] 进行的身份一致性实验评估表明，编辑不同属性时，身份相似度结果各不相同，如表 I 所示。值得一提的是，每种属性恰好编辑一次，编辑强度设为 1，即 $\xi = \beta = 1$。不同属性的身份一致性大多高于 CCPG 数据集的身份相似度中位基准 0.9296 ($(0.8806 + 0.9785)/2$) 和 OU-MVLP 数据集的身份相似度中位基准 0.7245 ($(0.9336 + 0.5153)/2$)。这表明，GaitEditor 在编辑这些属性时能够有效保持身份一致性。属性“视角”的身份一致性得分高得出乎意料，在 CCPG 数据集和 OU-MVLP 数据集上分别达到 0.9380 和 0.7813。考虑到视角转换会显著改变人体外观，这一结果尤其值得注意，表明我们提出的 GaitEditor 能够成功应对视角转换这一挑战。在 CCPG 数据集上，“裤子”和“连帽衫”属性的身份一致性得分略低。这可能是因为这些属性分别会导致腿部和颈部区域发生显著的形态变化。此类编辑可能会扰乱 GaitBase 模型捕捉到的独特步态模式；该模型在 CCPG 数据集上训练时，对这些特定区域的编辑尤为敏感。

表 VI：以 GaitBase 为基线，在不同数据集上的匿名化评估。

数据集	匿名化方法	Rank-1 (%)	FID (↓)
OU-MVLP	Raw	正常行走：89.85	-
	Blur	正常行走：11.79	7.92
	GaitEditor	正常行走: 0.61	**0.60**
CCPG	Raw	携物: 71.56，上坡: 75.00，下坡: 76.85，背包: 78.52	-
	Blur	携物: 31.84，上坡: 27.43，下坡: 30.51，背包: 31.65	8.36
	GaitEditor	**携物：13.79，上坡：22.09，下坡：23.93，背包：12.76**	**0.56**
Gait3D	Raw	64.60	-
	Blur	20.71	10.46
	GaitEditor	**8.60**	**1.05**

属性正确性评估。由于缺乏专门针对步态轮廓的细粒度属性分类器，我们通过用户研究评估编辑后步态序列的属性正确性。本研究共有 101 名参与者，分为两组：52 名具有步态研究经验的人员和 49 名普通参与者。每位参与者需回答十道随机生成的问题，题型为单项选择或判断题。问题按加权比例分配，其中 70% 为单项选择题，用于评估参与者对已编辑属性的识别能力；其余 30% 为判断题，用于验证特定属性编辑的正确性。单项选择题的题干为：“请指出紫色边框视频相较于蓝色边框参考视频中发生改变的属性。”判断题的题干为：“与蓝色边框的参考视频相比，紫色边框的视频是否呈现了经过编辑的‘X’属性？”，其中“X”代表某一特定步态属性。紫色边框标示编辑后的序列，蓝色边框则标示未经编辑的原始步态序列。

表 II 所示的研究结果全面展示了不同专业水平的受试者对各种属性判断的准确性。具有步态经验的参与者总体准确率为 81.46%，而非专家组的总体准确率略低，为 71.23%。在识别性别特征（“男性化”和“女性化”）及年龄特征（“年轻”和“年长”）时，准确率差异尤为明显。对于“男性化”“女性化”“年轻”和“年长”这几项，缺乏步态经验的受试者与具有步态经验的受试者相比，准确率分别降低了 29.42%、14.09%、11.02% 和 20.73%。这些结果表明，要准确识别步态属性在特定区域的细微变化，需要具备专业知识。

D. 步态识别评估

评估包含两个方面：使用 GaitEditor 在各种步态数据集上增强不同的步态模型。

在不同步态数据集上的评估。为进行对比分析，我们采用了三个不同的步态数据集：跨服装数据集 CCPG、跨视角数据集 OU-MVLP，以及室外数据集 Gait3D，并遵循其官方提供的数据集划分方案。作为基线模型，我们

采用了在上述各数据集（CCPG、OU-MVLP 和 Gait3D）上进行官方预训练的 GaitBase 模型。为提升 GaitBase 模型在不同数据集上的性能，我们以 0.2 的概率，使用由 GaitEditor 生成的编辑后输入序列对 GaitBase 进行微调，其中编辑视角和非视角的概率分别为 0.1 和 0.9。该微调过程的具体细节见表 III。值得一提的是，在引入 GaitEditor 对 GaitBase 进行调优时，采用了其标准数据增强策略，并遵循官方训练方案。

通过应用 GaitEditor 获得的性能提升见表 IV。具体而言，与 GaitBase 基线相比，在 OU-MVLP、CCPG 和 Gait3D 数据集上的平均 Rank-1 准确率分别提高了 +0.42%、 +1.60% 和 +0.80%。这些提升表明，GaitEditor 有助于在跨服装、跨视角和室外条件下更有效地学习步态表征。

不同步态模型上的评估。为验证 GaitEditor 能够提升各种步态方法的性能，我们采用了四种流行的步态识别模型：GaitSet、GaitPart、GaitBase 和 DeepGaitv2。这些模型均遵循其官方设置，并已在 CCPG 数据集上预训练。随后，我们利用 GaitEditor，并将其作为在线数据增强模块接入不同模型。接下来，我们以 0.2 的概率使用 GaitEditor 生成的编辑后输入序列对这些模型进行微调，其中视角和非视角编辑的概率分别为 0.1 和 0.9。详细的微调设置见表 III。

使用 GaitEditor 时，不同步态模型在 CCPG 数据集上的表现如表 IV 所示。深度学习步态模型的先驱 GaitSet 在借助 GaitEditor 增强后，在所有条件下的表现均有所提升：CL 条件下 Rank-1准确率提高了 +1.12%，UP 条件下提高了 +0.87%，DN 条件下提高了 +1.10%，BG 条件下提高了 +0.92%，平均提升 +1.00%。同样，考虑时间依赖关系的 GaitPart 也受益于 GaitEditor，表现出明显提升，尤其是在 UP 和 BG 条件下。包含 3D 卷积层、深度更深的模型 DeepGaitv2，在使用 GaitEditor 后准确率也有所提高。这些结果共同表明，将 GaitEditor 应用于不同的步态模型，能够有效提升步态表征的学习效果。

E. 步态匿名化评估

为评估步态匿名化的有效性，我们使用 GaitEditor 编辑整个测试探针集中的步态序列，然后计算 Rank-1准确率。我们通过计算 FID分数来评估编辑后的图像质量。对于不同数据集上的评估，我们使用 GaitBase 模型评估 Rank-1准确率；该模型在三个数据集（CCPG、OU-MVLP 和 Gait3D）各自的训练集上进行了预训练。对于不同步态模型的评估，我们采用四种常用的步态模型（GaitSet、GaitPart、GaitBase 和 DeepGaitv2），这些模型均在 CCPG 数据集上进行了预训练。

表 VI 展示了在不同数据集上进行评估的结果。模糊处理是通过应用高斯核来模糊

表VII：CCPG 数据集上不同步态模型的匿名化评估。术语“GaitBase*”表示经过 GaitEditor 调优的 GaitBase 模型，调优时采用了表III中列出的相同设置。

模型	GaitEditor	匿名化Rank-1准确率（↓）				
		CL	UP	DN	BG	Mean
GaitSet	✘	60.18	65.22	65.11	68.53	64.76
	✔	**14.53**	**19.66**	**21.31**	**11.10**	**16.65**
GaitPart	✘	64.25	67.76	68.58	71.68	68.07
	✔	**16.96**	**21.84**	**23.23**	**13.20**	**18.81**
GaitBase	✘	71.56	75.00	76.85	78.52	75.48
	✔	**13.79**	**22.09**	**23.93**	**12.76**	**18.14**
DeepGaitv2	✘	71.22	76.17	75.66	80.09	75.79
	✔	**13.17**	**23.79**	**25.48**	**14.35**	**14.20**
GaitBase*	✘	73.21	76.98	77.81	80.32	77.08
	✔	**16.09**	**26.23**	**27.62**	**13.10**	**20.76**

步态序列中的身份，从而确保匿名性。值得注意的是，对于跨视角数据集 OU-MVLP，应用 GaitEditor 后，Rank-1准确率从 89.848% 骤降至仅 0.609%，与使用模糊处理相比，展现出强大的匿名化效果。轮廓质量的 FID分数也表明效果良好，较低的 0.6038 分数意味着编辑后生成的图像质量较高。同样，对于跨服装数据集 CCPG，应用 GaitEditor 后，在各种条件（携物、上坡、下坡、背包）下的 Rank-1准确率均显著下降。对于室外数据集 Gait3D，应用 GaitEditor 后，Rank-1准确率从未使用 GaitEditor 时的 64.60% 大幅降至 8.60%。此外，无论在身份匿名化的效果还是匿名化后的图像质量方面，GaitEditor 都显著优于模糊技术。

表VII 展示了不同步态模型在 CCPG 数据集上结合 GaitEditor 进行测试的评估结果。GaitSet、GaitPart、GaitBase 和 DeepGaitv2 模型的平均 Rank-1准确率分别大幅下降了 48.11%、49.26%、57.34% 和 58.88%。这些结果表明，GaitEditor 能够有效提升不同模型架构下步态数据的匿名性，从而解决步态隐私问题。此外，实验还评估了调优后的 GaitBase 模型，称为 GaitBase*。尽管该模型经过了 GaitEditor 的优化，匿名化过程中编辑的多样性仍能确保去身份化的有效性，从而显著有助于保护个人隐私。

F. 可视化

图 8 展示了我们的 GaitEditor 使用多种属性编辑后，忠实保留的步态轮廓图。示例步态序列来自 OU-MVLP 数据集。如紫色框最后一行所示，通过分别调整强度 β和编辑频率 ξ，GaitEditor 可以连续、叠加地修改步态属性。对于视角属性，GaitEditor 还能够根据给定的两段序列进行视角转换，结果如绿色

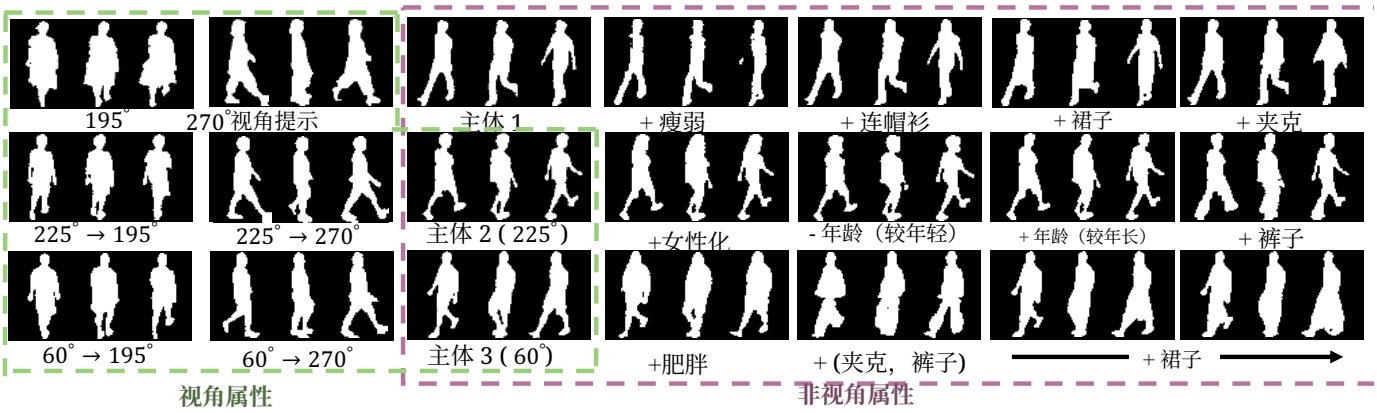


图 8：带有非视角属性（紫色框）和视角属性（绿色框）的步态属性编辑可视化。 $+/-$表示添加/移除该属性， $a^{\circ} \rightarrow b^{\circ}$ 表示视角从 a° 变为 b°。

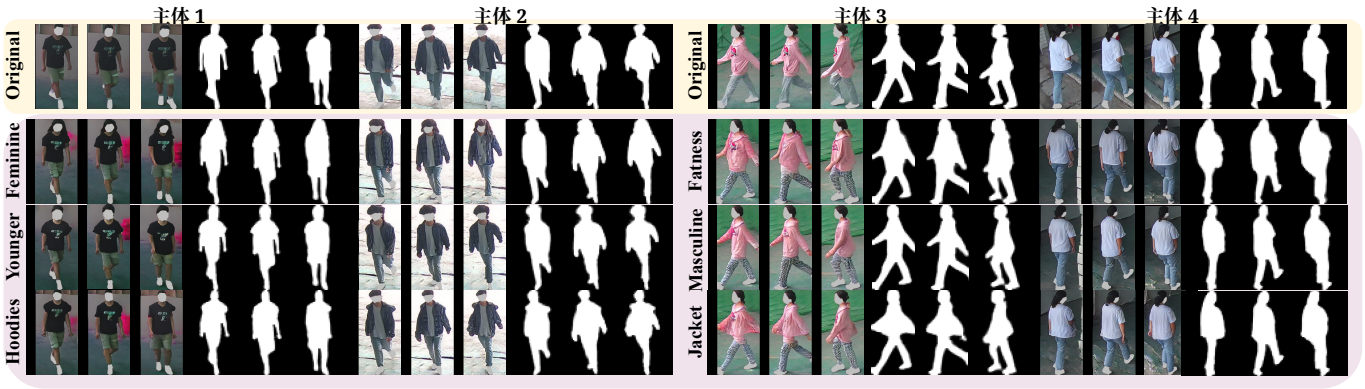


图 9：非视角属性可视化的 RGB 结果。第一行为原始序列（黄色），最后三行为编辑结果（紫色）。

表 VIII：针对 GaitEditor 中不同模块或损失函数的消融研究。我们使用一系列 GAN指标对这些组件的效能进行量化，这些指标衡量反演质量、轮廓质量和身份相似度。“↑”表示数值越高，性能越好；而“↓”表示数值越低，性能越好。

设置	峰值信噪比（↑）	结构相似性指数（↑）	FID（↓）	LPIPS（↓）	MSE（↓）	身份相似度（↑）
不含 $\mathcal{L}_{adv}$	20.904	0.865	0.082	0.046	0.045	0.966
不含 $\mathcal{L}_{id}$	20.776	0.874	0.099	0.047	0.046	0.956
不含 $\mathbf{E}_{id}$	20.514	**0.901**	**0.059**	0.045	0.039	0.951
不含 $\mathbf{E}_{mix}$	19.884	0.848	0.154	0.054	0.060	0.969
我们的完整模型	**21.523**	0.902	0.069	**0.034**	**0.030**	**0.978**

框。图 9 展示了我们使用 GaitEditor 编辑非视角属性后，再由 VQGAN 渲染为 RGB格式的步态图像。示例步态序列来自 CCPG 数据集的测试集。可以看到，GaitEditor 能够操控特定特征，同时保持步态的自然性。与先前依赖使用 RGB图像进行训练的方法 [77] 相比（该方法通常着重于编辑服装长度和款式等纹理特征），GaitEditor 能够准确捕捉并编辑与行人体型相关的属性，例如女性气质和体型大小。这一区别凸显了 GaitEditor 在保护行人隐私方面的独特能力。它能够有效地匿名化个人的可识别特征，同时不损害自然外观信息。

G. 消融研究

在本节中，我们通过消融研究评估 GaitEditor 在 CCPG 数据集上的表现。我们通过修改模块和损失函数设计了四个变体：1）不含 $\mathcal{L}_{adv}$，在训练阶段移除视频对抗损失。2）不含 $\mathcal{L}_{id}$，在训练阶段移除身份损失函数。3）不含 $\mathbf{E}_{id}$，移除预训练步态模型，并从头开始训练。4）不含 $\mathbf{E}_{mix}$，移除特征混合模块 $\mathbf{E}_{mix}$，然后利用

加法操作在潜在空间中融合不同特征。与各种变体的定量比较见表VIII。如表VIII所示，移除身份损失项（$\mathcal{L}_{id}$）或作为先验知识使用的预训练步态模型（$\mathbf{E}_{id}$）都会导致身份保持能力下降，这一点可由身份相似度指标体现。值得注意的是，变体$\mathbf{E}_{id}$ 在结构相似性指数（SSIM）和 FID 指标上表现最佳，但在身份相似度指标（ID-Simility）上表现最差。这是因为预训练步态模型确保了反演帧中身份的一致性。然而，这也会在一定程度上损害帧内成像的质量。省略对抗损失会导致反演质量下降，具体表现为峰值信噪比（PSNR）降低以及 FID 分数升高。值得注意的是，缺少特征混合模块 $\mathbf{E}_{mix}$ 会产生显著影响，导致图像质量大幅下降。我们的完整模型集成了所有组件，在所有指标上均取得了最高分。

V. 局限性

我们承认，我们的方法存在一些局限性，例如：1）难以准确转换视角以实现大幅度变化，例如实现 180° 旋转；2）难以准确编辑年龄特征和性别特征，尤其是在编辑结果与传统观念相矛盾时，例如男性留长发；3）在时间维度上编辑属性也面临挑战，例如步行阶段。

VI. 结论与未来工作

我们提出了 GaitEditor，这是一种开创性的步态属性编辑方法，可用于编辑多种步态属性，并应用于步态识别和匿名化。GaitEditor 的研究动机源于构建统一框架所面临的挑战：该框架既要提升个体识别能力，又要缓解隐私问题。我们在三个广泛使用的步态基准数据集上，采用四种著名的步态识别算法，对 GaitEditor 的效能进行了定量评估，结果表明，该方法在识别和匿名化任务中均具有多功能性和有效性。此外，它能够生成生动的轮廓图和 RGB 序列，进一步彰显了其在步态识别和匿名化中的应用价值。

展望未来，从无标注行走视频中学习通用的步态表征是一项颇具挑战、但极具吸引力的任务，因为收集大量带标注的步态数据成本高昂，且需求永无止境。GaitEditor 可作为数据增强模块，为此提供一种潜在解决方案，从而助益**无监督步态识别**。与此同时，GaitEditor 也有望助益**跨X步态识别**。跨视角和跨服装是当前步态数据集最突出的两种测试设置，因此现有大多数方法主要针对这些问题。然而，在实际应用中，识别模型常会遇到许多其他复杂情况，例如跨帽子、跨发型、跨体型、跨年龄，甚至更加特殊的跨X情形。由于缺乏细粒度标注数据对，难以准确评估面向跨X情形的步态识别。幸运的是，GaitEditor 能够模拟包含多种步态属性变化的虚拟数据集，从而提供一种潜在解决方案。最后但同样重要的是，尽管 GaitEditor 可以通过扰动身体外观来匿名化步态序列中的身份，但这种身体外观也会呈现不自然的效果。例如，家庭成员更愿意看到真实的面容和体型，而非匿名化后的样子；人们也可能希望分享仅限特定授权方解读的数据。因此，也需要**步态匿名化的可逆性**。

总而言之，GaitEditor 标志着我们通过统一框架同时编辑多种步态属性的能力取得了重要进展。它展现了提升真实数据集规模和多样性的潜力，有望激发宏大的构想，使步态应用摆脱真实数据集在规模和多样性方面的限制。

参考文献

[1] A. S. Alharthi、S. U. Yunas 和 K. B. Ozanyan，“用于人体步态监测的深度学习：综述”，*IEEE*传感器期刊，第19卷，第21期，第9575–9591页，2019年。[2] C. Shen、S. Yu、J. Wang、G. Q. Huang 和 L. Wang，“深度步态识别综合综述：算法、数据集与挑战”，*arXiv*预印本 *arXiv:2206.13732*，2022年。[3] I. Bouchrika、M. Goffredo、J. Carter 和 M. Nixon，“步态在法医生物特征识别中的应用”，法医学杂志，第56卷，第4期，第882–889页，2011年。[4] H. Iwama、D. Muramatsu、Y. Makihara 和 Y. Yagi，“用于刑事调查的步态验证系统”，信息与媒体技术，第8卷，第4期，第1187–1199页，2013年。[5] N. Lynnerup 和 P. K. Larsen，“步态作为证据”，*IET*生物特征识别，第3卷，第2期，第47–54页，2014年。[6] Z. Wu、Y. Huang、L. Wang、X. Wang 和 T. Tan，“基于深度卷积神经网络的跨视角步态人体识别综合研究”，*IEEE*模式分析与机器智能汇刊，第39卷，第2期，第209–226页，2016年。[7] D. P. Kingma 和 M. Welling，“自编码变分贝叶斯”，*arXiv*预印本 *arXiv:1312.6114*，2013年。[8] I. Goodfellow、J. Pouget-Abadie、M. Mirza、B. Xu、D. Warde-Farley、S. Ozair、A. Courville 和 Y. Bengio，“生成对抗网络”，神经信息处理系统进展，第27卷，2014年。[9] T. Karras、S. Laine 和 T. Aila，“用于生成对抗网络的基于风格的生成器架构”，载于*IEEE/CVF*计算机视觉与模式识别会议论文集，2019年，第4401–4410页。[10] J. Ho、A. Jain 和 P. Abbeel，“去噪扩散概率模型”，神经信息处理系统进展，第33卷，第6840–6851页，2020年。[11] T. Karras、S. Laine、M. Aittala、J. Hellsten、J. Lehtinen 和 T. Aila，“分析并提升StyleGAN的图像质量”，载于*IEEE/CVF*计算机视觉与模式识别会议论文集，2020年，第8110–8119页。[12] 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”，神经信息处理系统进展，第33卷，12 114，2020年12 104页–。[13] T. Karras、M. Aittala、S. Laine、E. H¨ark¨onen、J. Hellsten、J. Lehtinen 和 T. Aila，“无混叠生成对抗网络”，神经信息处理系统进展，第34卷，第852–863页，2021年。[14] S. Yu、H. Chen、E. B. Garcia Reyes 和 N. Poh，“GaitGAN：使用生成对抗网络提取不变步态特征”，载于*IEEE*计算机视觉与模式识别会议研讨会论文集，2017年，第30–37页。[15] X. Li、Y. Makihara、C. Xu、Y. Yagi 和 M. Ren，“通过半监督解耦表示学习身份与协变量特征进行步态识别”，载于*IEEE/CVF*计算机视觉与模式识别会议论文集，2020年，第13 309–13 319页。

[16] X. Chen、X. Luo、J. Weng、W. Luo、H. Li 和 Q. Tian，“用于跨视角步态识别的多视角步态图像生成”，IEEE图像处理汇刊，第30卷，第3041–3055页，2021年。[17] P. Zhang、H. Dou、W. Zhang、Y. Zhao、Z. Qin、D. Hu、Y. Fang 和 X. Li，“面向自然场景模拟的大规模合成步态数据集及比较研究”，ACM多媒体计算、通信与应用汇刊，第19卷，23，2023年1，1号，第–页。[18] J. Ma、X. Zhang 和 S. Yu，“用于人体动作迁移的身份保持框架”，arXiv预印本 arXiv:2204.06862，2022年。[19] Y. Hirose、K. Nakamura、N. Nitta 和 N. Babaguchi，“基于轮廓变形与纹理迁移的视频人体步态匿名化”，IEEE信息取证与安全汇刊，第17卷，第3375–3390页，2022年。[20] P. Agrawal 和 P. Narayanan，“视频中的人物去身份化”，IEEE视频技术电路与系统汇刊，第21卷，第3期，第299–310页，2011年。[21] I. Mitsugami、M. Mukunoki、Y. Kawanishi、H. Hattori 和 M. Minoh，“用于感知网络的隐私保护摄像机”，载于《基于知识系统的不确定性信息处理与管理：第13届国际会议，IPMU 2010，德国多特蒙德，2010年6月28日至7月2日，论文集，第II部分13》。施普林格，2010年，第622–631页。[22] N.-D. T. Tieu、H. H. Nguyen、H.-Q. Nguyen-Son、J. Yamagishi 和 I. Echizen，“一种使用深度学习进行步态匿名化的方法”，载于《2017年IEEE信息取证与安全研讨会（WIFS）》。IEEE，2017年，第1–6页。[23] R. Abdal、Y. Qin 和 P. Wonka，“Image2StyleGAN：如何将图像嵌入StyleGAN潜空间？”，载于《IEEE/CVF国际计算机视觉会议论文集》，2019年，第4432–4441页。[24] Y. Shen、J. Gu、X. Tang 和 B. Zhou，“解读生成对抗网络的潜空间以进行人脸语义编辑”，载于《IEEE/CVF计算机视觉与模式识别会议论文集》，2020年，第9243–9252页。[25] R. Abdal、Y. Qin 和 P. Wonka，“Image2StyleGAN++：如何编辑嵌入的图像？”，载于《IEEE/CVF计算机视觉与模式识别会议论文集》，2020年，第8296–8305页。[26] Z. Wu、D. Lischinski 和 E. Shechtman，“StyleSpace分析：用于StyleGAN图像生成的解耦控制”，载于《IEEE/CVF计算机视觉与模式识别会议论文集》，2021年，第12 863–12 872页。[27] S.-J. Moon 和 G.-M. Park，“IntereStyle：为稳健的StyleGAN反演编码感兴趣区域”，载于《计算机视觉—ECCV 2022：第17届欧洲计算机视觉会议，以色列特拉维夫，2022年10月23日至27日，论文集，第XV部分》。施普林格，2022年，第460–476页。[28] A. Fr¨uhst¨uck、N. Sarafianos、Y. Xu、P. Wonka 和 T. Tung，“Vive3D：使用3D感知生成对抗网络进行视点无关视频编辑”，载于《IEEE/CVF计算机视觉与模式识别会议论文集》，2023年，第4446–4455页。[29] H. Liu、Y. Song 和 Q. Chen，“深入探索用于图像编辑的StyleGAN反演：基础潜空间视角”，载于《IEEE/CVF计算机视觉与模式识别会议论文集》，2023年，第10 072–10 082页。[30] W. Xia、Y. Zhang、Y. Yang、J.-H. Xue、B. Zhou 和 M.-H. Yang，“生成对抗网络反演：综述”，IEEE模式分析与机器智能汇刊，2022年。[31] G. Ariyanto 和 M. S. Nixon，“基于模型的三维步态生物特征识别”，载于《2011年国际生物特征联合会议（IJCB）》。IEEE，2011年，第1–7页。[32] R. Liao、S. Yu、W. An 和 Y. Huang，“一种结合人体姿态与人体先验知识的基于模型步态识别方法”，模式识别，第98卷，文章号107069，2020年。[33] T. Teepe、A. Khan、J. Gilg、F. Herzog、S. H¨ormann 和 G. Rigoll，“GaitGraph：用于基于骨架的步态识别的图卷积网络”，载于《2021年IEEE国际图像处理会议（ICIP）》。IEEE，2021年，第2314–2318页。[34] T. Teepe、J. Gilg、F. Herzog、S. H¨ormann 和 G. Rigoll，“深入理解基于骨架的步态识别”，载于《IEEE/CVF计算机视觉与模式识别会议论文集》，2022年，第1569–1577页。[35] X. Li、Y. Makihara、C. Xu 和 Y. Yagi，“包含人体网格的多视角大规模人群步态数据库及其性能评估”，IEEE生物特征、行为与身份科学汇刊，第4卷，第2期，第234–248页，2022年。[36] Y. Fu、S. Meng、S. Hou、X. Hu 和 Y. Huang，“GPGait：广义姿态步态识别”，arXiv预印本 arXiv:2303.05234，2023年。[37] X. Li、Y. Makihara、C. Xu 和 Y. Yagi，“利用同步多视角姿态约束实现端到端基于模型的步态识别”，载于 IEEE/CVF国际计算机视觉会议论文集，2021年，第4106–4115页。[38] C. Fan、J. Liang、C. Shen、S. Hou、Y. Huang 和 S. Yu，“OpenGait：重新审视步态识别，以提升实用性”，载于 IEEE/CVF计算机视觉与模式识别会议论文集，2023年，第9707–9716页。[39] J. Zheng、X. Liu、W. Liu、L. He、C. Yan 和 T. Mei，“利用稠密三维表示和基准实现野外环境中的步态识别”，载于IEEE/CVF计算机视觉与模式识别会议论文集，2022年，第20 228–20 237页。[40] H. Chao、K. Wang、Y. He、J. Zhang 和 J. Feng，“GaitSet：将步态视为深度集合，实现跨视角步态识别”，IEEE模式分析与机器智能汇刊，2021年。[41] C. Fan、Y. Peng、C. Cao、X. Liu、S. Hou、J. Chi、Y. Huang、Q. Li 和 Z. He，“GaitPart：用于步态识别的时序部件模型”，载于 IEEE/CVF计算机视觉与模式识别会议论文集，2020年，第14 225–14 233页。[42] B. Lin、S. Zhang 和 X. Yu，“通过有效的全局-局部特征表示和局部时序聚合实现步态识别”，载于 IEEE/CVF国际计算机视觉会议论文集，2021年，第14 648–14 656页。[43] C. Fan、S. Hou、Y. Huang 和 S. Yu，“探索用于实用步态识别的深度模型”，arXiv预印本 arXiv:2303.03301，2023年。[44] J. Liang、C. Fan、S. Hou、C. Shen、Y. Huang 和 S. Yu，“Gaitedge：超越普通端到端步态识别，提升实用性”，载于《计算机视觉—ECCV 2022：第17届欧洲会议，以色列特拉维夫，2022年10月23日至27日，论文集，第V部分》。施普林格，2022年，第375–390页。[45] N.-D. T. Tieu、H. H. Nguyen、H.-Q. Nguyen-Son、J. Yamagishi 和 I. Echizen，“用于步态匿名化的时空生成对抗网络”，《信息安全与应用》期刊，第46卷，第307–319页，2019年。[46] N.-D. T. Tieu、H. H. Nguyen、F. Fang、J. Yamagishi 和 I. Echizen，“用于低质量轮廓图的RGB步态匿名化模型”，载于2019年亚太信号与信息处理协会年度峰会与会议（APSIPA ASC）。IEEE，2019年，第1686–1693页。[47] J. Han 和 B. Bhanu，“利用步态能量图进行个体识别”，IEEE模式分析与机器智能汇刊，第28卷，第2期，第316–322页，2005年。[48] W. Xiang、H. Yang、D. Huang 和 Y. Wang，“多视角步态视频合成”，载于第30届ACM国际多媒体会议论文集，2022年，第6783–6791页。[49] A. Creswell 和 A. A. Bharath，“反演生成对抗网络的生成器”，IEEE神经网络与学习系统汇刊，第30卷，第7期，第1967–1974页，2018年。[50] X. Pan、X. Zhan、B. Dai、D. Lin、C. C. Loy 和 P. Luo，“利用深度生成先验实现多功能图像复原与操控”，IEEE模式分析与机器智能汇刊，第44卷，第11期，第7474–7489页，2021年。[51] E. Richardson、Y. Alaluf、O. Patashnik、Y. Nitzan、Y. Azar、S. Shapiro 和 D. Cohen-Or，“Encoding in Style：用于图像到图像转换的StyleGAN编码器”，载于 IEEE/CVF计算机视觉与模式识别会议论文集，2021年，第2287–2296页。[52] O. Tov、Y. Alaluf、Y. Nitzan、O. Patashnik 和 D. Cohen-Or，“为StyleGAN图像操控设计编码器”，ACM图形学汇刊（TOG），第40卷，第4期，第1–14页，2021年。[53] Y. Alaluf、O. Patashnik 和 D. Cohen-Or，“Restyle：通过迭代细化实现的残差式StyleGAN编码器”，载于 IEEE/CVF国际计算机视觉会议论文集，2021年，第6711–6720页。[54] Y. Xu、Y. Shen、J. Zhu、C. Yang 和 B. Zhou，“通过合成图像获得生成式层次特征”，载于IEEE/CVF计算机视觉与模式识别会议论文集，2021年，第4432–4442页。[55] Q. Bai、Y. Xu、J. Zhu、W. Xia、Y. Yang 和 Y. Shen，“利用填充空间实现高保真GAN反演”，载于《计算机视觉—ECCV 2022：第17届欧洲会议，以色列特拉维夫，2022年10月23日至27日，论文集，第XV部分》。施普林格，2022年，第36–53页。[56] X. Yao、A. Newson、Y. Gousseau 和 P. Hellier，“用于基于风格的GAN反演的特征风格编码器”，arXiv电子版，第arXiv–2202页，2022年。[57] H. Kim、Y. Choi、J. Kim、S. Yoo 和 Y. Uh，“利用GAN潜变量的空间维度实现实时图像编辑”，载于 IEEE/CVF计算机视觉与模式识别会议论文集，2021年，第852–861页。

[58] E. Collins、R. Bala、B. Price 和 S. Susstrunk，“风格编辑：揭示生成对抗网络的局部语义”，载于《IEEE/CVF计算机视觉与模式识别会议论文集》，2020年，第5771–5780页。[59] N. Takemura、Y. Makihara、D. Muramatsu、T. Echigo 和 Y. Yagi，“多视角大规模人群步态数据集及其跨视角步态识别性能评估”，《IPSJ计算机视觉与应用汇刊》，第10卷，第1期，第1–14页，2018年。[60] Y. Shen 和 B. Zhou，“生成对抗网络中潜在语义的闭式因子分解”，载于《IEEE/CVF计算机视觉与模式识别会议论文集》，2021年，第1532–1540页。[61] K. He、X. Zhang、S. Ren 和 J. 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》，2016年，第770–778页。[62] 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《2009年IEEE计算机视觉与模式识别会议论文集》。IEEE，2009年，第248–255页。[63] R. Zhang、P. Isola、A. A. Efros、E. Shechtman 和 O. Wang，“深度特征作为感知度量的惊人效能”，载于CVPR，2018年。[64] X. Mao、Q. Li、H. Xie、R. Y. Lau、Z. Wang 和 S. Paul Smolley，“最小二乘生成对抗网络”，载于《IEEE国际计算机视觉会议论文集》，2017年，第2794–2802页。[65] S. Tulyakov、M.-Y. Liu、X. Yang 和 J. Kautz，“MoCoGAN：分解运动与内容以实现视频生成”，载于《IEEE计算机视觉与模式识别会议论文集》，2018年，第1526–1535页。[66] M. Gutmann 和 A. Hyv¨arinen，“噪声对比估计：一种用于未归一化统计模型的新估计原理”，载于《第十三届国际人工智能与统计会议论文集》。JMLR研讨会与会议论文集，2010年，第297–304页。[67] W. Li、S. Hou、C. Zhang、C. Cao、X. Liu、Y. Huang 和 Y. Zhao，“换装条件下行人重识别与步态识别的深入探索”，载于《IEEE/CVF计算机视觉与模式识别会议论文集》，2023年，第13 824–13 833页。[68] Y. Li、Y. Li、J. Lu、E. Shechtman、Y. J. Lee 和 K. K. Singh，“用于多样化解耦前景生成的对比学习”，载于《欧洲计算机视觉会议论文集》。施普林格，2022年，第334–351页。[69] M. Heusel、H. Ramsauer、T. Unterthiner、B. Nessler 和 S. Hochreiter，“采用双时间尺度更新规则训练的生成对抗网络会收敛至局部纳什均衡”，《神经信息处理系统进展》，第30卷，2017年。[70] C. Fan、S. Hou、J. Wang、Y. Huang 和 S. Yu，“从大量无标注行走视频中学习步态表征：基准”，《IEEE模式分析与机器智能汇刊》，2023年。[71] C. Szegedy、V. Vanhoucke、S. Ioffe、J. Shlens 和 Z. Wojna，“重新思考用于计算机视觉的Inception架构”，载于《IEEE计算机视觉与模式识别会议论文集》，2016年，第2818–2826页。[72] P. Esser、R. Rombach 和 B. Ommer，“驯服变换器以实现高分辨率图像合成”，载于《IEEE/CVF计算机视觉与模式识别会议论文集》，2021年，第12 873–12 883页。[73] C. Song、Y. Huang、W. Wang 和 L. Wang，“CASIA-E：用于步态识别的大型综合数据集”，《IEEE模式分析与机器智能汇刊》，第45卷，第3期，第2801–2815页，2022年。[74] Z. Wang、A. C. Bovik、H. R. Sheikh 和 E. P. Simoncelli，“图像质量评估：从误差可见性到结构相似性”，《IEEE图像处理汇刊》，第13卷，第4期，第600–612页，2004年。[75] R. Zhang、P. Isola、A. A. Efros、E. Shechtman 和 O. Wang，“深度特征作为感知度量的惊人效能”，载于《IEEE计算机视觉与模式识别会议论文集》，2018年，第586–595页。[76] D. P. Kingma 和 J. Ba，“Adam：一种随机优化方法”，arXiv预印本 arXiv:1412.6980，2014年。[77] J. Fu、S. Li、Y. Jiang、K.-Y. Lin、C. Qian、C. C. Loy、W. Wu 和 Z. Liu，“StyleGAN-Human：以数据为中心的人体生成探索之旅”，载于《欧洲计算机视觉会议论文集》。施普林格，2022年，第1–19页。