arXiv:2302.00556v3 [cs.CV] 4 Mar 2024

无对应关系的在线人体运动重定向

Rim Rekik*[1] Mathieu Marsot*[1] Anne-H´el`eneOlivier[2] Jean-S´ebastien Franco[1] StefanieWuhrer[II] 格勒诺布尔-阿尔卑斯大学、Inria、法国国家科学研究中心、格勒诺布尔国立综合理工学院 †、LJK，法国格勒诺布尔，邮编38000[2] 雷恩大学、Inria、法国国家科学研究中心、IRISA、M2S，法国雷恩，邮编35000*共同第一作者：rim.rekik-dit-nekhili@inria.fr, mathieu.marsot@gmail.com firstname.lastname@inria.fr

摘要

我们提出了一种数据驱动框架，用于无监督人体动作重定向，使目标主体随着源主体的动作进行动画呈现。我们的方法无需对应关系，既不需要源形状与目标形状之间的空间对应关系，也不需要源运动不同帧之间的时间对应关系。因此，该方法可以使用任意人体运动序列为目标形状生成动画，这些序列可能由*4D*采集平台或消费级设备捕获。我们的方法融合了两类现有方法的优点：骨架动作重定向利用长期时间上下文，而基于表面的重定向则保留表面细节。为此，我们将几何感知变形模型与骨架感知动作迁移方法相结合，从而在考虑表面细节的同时利用长期时间上下文。在推理阶段，我们的方法可在线运行，即输入可以按序列逐帧处理，且每帧仅需一次前向传播即可完成重定向。实验表明，在训练期间纳入长期时间上下文，可提高该方法在骨骼运动准确度和细节保留方面的表现。此外，我们的方法能够泛化到未观测过的动作和身体形状。我们证明，该方法在两个测试数据集上取得了最先进的结果，并且可以利用多视图采集平台的输出为人体模型生成动画。代码可在以下网址获取：
https://gitlabb.inria.fr/rrekikdi/human-motion-retargeting2023。

1. 引言

人体动作重定向是指使用源角色的动作来驱动目标角色的动画。我们针对密集采样的三维表面和任意时长的动作研究这一问题。这项技术可应用于视频游戏、电影制作、虚拟形象动画，以及扩充现有的4D人体动作数据集例如。 [6, 7, 29, 40]。

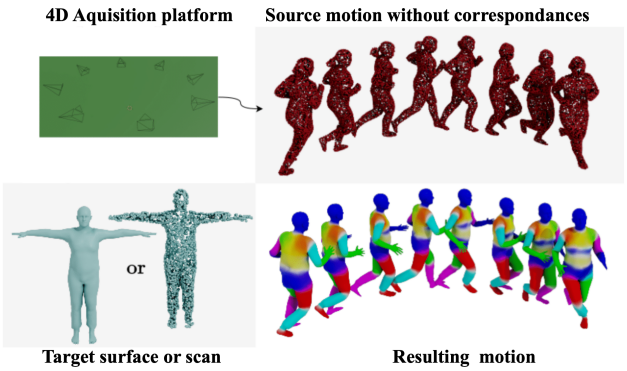

图 1. 给定由源主体执行的未跟踪运动（上方）和目标身体形状（左下），我们的方法使用源动作驱动目标，同时保留输出运动的时间对应关系（右下）。

近年来，运动重定向取得了显著进展。基于骨架的方法能够考虑约 2 秒的长期时间上下文例如。 [15, 24,38]。基于表面变形的方法可以在重定向几何细节的同时捕捉短时动态例如。 [5, 35, 36, 41]。

这些工作大多以结构化数据作为输入，例如骨架或模板对齐表面，且这些数据具有已知的对应关系。

本文旨在实现无对应关系的重定向，其输入为非结构化四维数据，其中既没有源形状与目标形状之间的对应关系，也没有源动作不同帧之间的对应关系。这样便可直接使用任意人体运动序列为人体目标形状制作动画，这些序列也可能由 4D采集平台采集。对于非结构化三维数据，解决重定向问题颇具挑战，因为计算对应关系是一个组合问题。图1 展示了该问题的输入和输出。

†格勒诺布尔-阿尔卑斯大学工程学院

近期，有人提出了首个无对应关系的动作重定向方案 [18]，该方案引入通用动作先验，用于描述身体形状如何在低维空间中运动，并将长期时间上下文纳入考量。尽管这为新方案开辟了道路，但在推理过程中，该方法仅限于重定向固定长度的序列。受这项工作的启发，我们的框架也学习长期时间上下文。

在这项工作中，我们提出了首个能够学习长期时间上下文的在线动作重定向无对应关系方法。我们的方法能够学习时长为 $1s$ 的时间上下文，我们通过具有挑战性的样本证明，这能够提高动作重定向准确度。

我们结合基于骨架的动作重定向和基于表面的动作重定向的优势，同时学习时间上下文和几何细节，从而实现这一目标。具体而言，我们将基于骨架的动作重定向与形状变形建模相结合，从而能够处理高维时空数据。这是因为，骨架重定向方法通过编码长期时间上下文 [38] 取得了良好效果，而近期基于骨架变形的形状变形方法 [26, 32, 42, 43] 能够准确地建模几何细节。我们的假设是，源序列的运动信息可在骨架层级上得到解释，而稠密表面细节则是目标形状固有的。因此，我们的思路是最小化动作保持损失，将源序列的骨骼运动传递至目标，并最小化形状保持损失，以保留目标的形状细节。

我们的主要贡献如下：

- 首个无需对应关系、能够学习时间上下文的在线稠密人体运动重定向方法。
- 证明长期时间上下文能够提升运动重定向准确度。
- 提出一种方法，在几何细节保持和骨架层面的运动重定向方面均优于无需对应关系的方法。

2. 相关工作

针对运动重定向的研究可分为三类：在骨架层面进行处理的研究、在稠密几何层面传递形状变形的研究，以及将运动先验应用于运动重定向的研究。

2.1. 骨骼动作迁移

早期的运动重定向工作基于骨架参数化，即用一组稀疏的关节位置来描述人体姿态。Gleicher 等。 [15]提出了这项任务，并将其视为一个针对整个动作序列、受运动学约束的优化问题。随后，可使用人工计算的蒙皮权重 [25, 44] 和混合技术 [19, 22]对生成的骨架参数化进行动画处理。

后续工作 [12, 21, 23, 37] 同样需要通过迭代优化，为特定动作手动设计运动学约束。随着采集的动作数据日益易于获取，且深度学习技术效率不断提升，新的数据驱动方法 [2, 14, 16, 38] 无需大量手工设计的能量项，便展现出卓越效果。其中一些数据驱动方法 [14, 16] 需要配对训练数据，而设计这些数据需要投入人力。因此，另一些研究工作 [2, 38] 提出了无监督重定向策略。Villegas 等。 [38]提出了一种基于对抗式循环一致性的无监督运动重定向框架，以确保重定向后动作的合理性。该方法能够为未见角色生成自然的动作，但仅在骨架层级上运行，不包含几何细节。

较新的研究 [39] 提出纳入几何信息，并研究结合数据驱动网络与推理后优化的混合骨架式运动重定向，以保留自接触并防止相互穿透。该方法展现出卓越的效果，但需要将手工制作的蒙皮权重作为额外输入。

与我们的方法类似，近期有一种方法结合骨架层面和几何信息进行运动重定向 [45]。不同于我们的工作，该方法侧重于保持接触，尤其针对使用计算机辅助设计工具创建、并以干净骨架模型作为输入的角色。相比之下，我们专注于直接重定向采集平台的原始输出，无需进行成本高昂且容易出错的预处理来寻找对应信息或拟合骨架。

2.2. 形状变形迁移

形状变形迁移方法直接作用于表面：一种方式是使用能够解耦形状与姿态的三维人体形状表示 [13, 49]，另一种方式是优化网格变形 [4, 8, 36, 48]。一些研究 [4, 36] 将运动重定向视为姿态变形迁移：对源角色的姿态进行编码，并将关联网格的变形迁移至目标角色。另一些研究 [5]则将运动重定向视为形状变形迁移：对目标角色的形状身份进行编码，并将其迁移至源角色。近期一项研究通过考虑连续 4 帧 [35]引入了短期时间上下文。这些研究在具有对应关系的网格上进行操作，取得了令人瞩目的效果。由于建立源网格与目标网格之间的时空对应关系十分困难，Wang 等人。 [41] 提出了一种新的风格迁移方法，该方法能够学习将一个角色的形状风格迁移到另一个具有姿态的角色上，同时学习结构不同的点云之间的对应关系。Chen 等人。 [9]在此基础上进一步纳入了 3 帧的短期时间信息。这样虽然提升了效果，但代价是需要输入目标序列，而非单个

目标帧进行训练。由于计算复杂度较高，基于变形的方法无法扩展到学习超出少数连续帧范围的长期时间上下文。我们通过实验表明，我们的方法优于神经姿态迁移（NPT）[41], ——这是唯一一种无需序列级监督即可训练的现有无对应关系方法。

2.3. 通用运动先验

近期，能够在潜在空间中表示稠密人体运动，并将运动与稠密身体形状信息解耦的通用运动先验，已被应用于运动重定向 [18, 30]。它们能够在考虑长期时间上下文的同时，将身体运动重定向到不同的身体形状。Marsot 等人 [30] 使用结构化数据，并学习稠密身体形状与基于骨架的运动之间的相关性。与我们的方法更为接近的是，H4D [18] 处理非结构化点云，并保留目标形状的稠密表面细节。尽管运动先验能够利用长期时间上下文，但由于它们以归一化时序片段为基础，因此难以在推理时应用于时长任意的序列。我们通过实验表明，我们的方法优于 H4D [18]。

2.4. 定位

表 1 总结了我们的工作相对于最先进方法的定位。我们根据四项标准对方法进行分类：训练时是否使用长期时间上下文（$0.5s$ 或更多）、是否支持在线推理、是否对几何细节建模，以及是否处理无已知对应关系的非结构化数据。结合基于骨骼的重定向和形状变形模型的优势，我们提出了首个无对应关系重定向方法，该方法能够学习长期时间上下文，支持推理时处理任意时长，并对几何细节建模。

3. 运动重定向方法

我们的运动重定向方法以主体 A 的源运动作为输入，该运动由一系列 n 点云$\{\mathcal{S}_i^A\}_{i=1}^n$ 构成，且没有时间对应关系；同时输入的还有处于 T 姿势的目标形状 B，以点云形式表示，并且可能包含连接信息，$\mathcal{S}_{tpose}^B$。我们的目标是生成一系列重定向扫描 $\{\mathcal{S}_i^B\}_{i=1}^n$，其中 $\mathcal{S}_i^B$模仿 $\mathcal{S}_i^A$ 的姿势，同时保留 $\mathcal{S}_{tpose}^B$的身体形状，并与 $\mathcal{S}_{tpose}^B$相对应。

运动重定向的一种常见策略是将高层运动与身体形状引起的形状变形解耦，并将源动作与目标身体形状相结合。若考虑每帧包含数千个顶点的密集采样输入表面，这种方法的计算开销很大；尤其当表面之间不存在对应关系时，这会阻碍现有方法

方法	Tem- 时间 上下文	在线 推断 ence	几何 tric 细节	Un- 结构 化的
基于骨架的				
[1, 2, 24, 38]	✓	✓	✗	✗
[39, 45]	✓	✓	✓	✗
形状变形迁移				
[5, 10, 35]	✗	✓	✓	✗
[9, 41]	✗	✓	✓	✓
运动先验				
[30]	✓	✗	✓	✗
[18]	✓	✗	✓	✓
Ours	✓	✓	✓	✓

表1. 与最先进的动作重定向方法的对比。我们提出了首个无对应关系的动作重定向方法，该方法能够学习长期时间上下文，在推理时支持任意时长，同时还能在表面层级对几何细节进行建模。

学习长期时间上下文。

我们通过提出以下假设来克服这一挑战：源动作信息可在骨架层级得到充分解释，而几何细节信息则编码在目标中。基于这一假设，动作重定向目标由训练期间需要优化的两个损失构成。第一个损失称为源动作保持损失 $\mathcal{L}_{motion}$，用于促使重定向后的动作在骨架层级上与 $\{\mathcal{S}_i^A\}_{i=1}^n$ 相似。在骨架层级考虑 $\mathcal{L}_{motion}$ 有三大优势：它提供了输入帧之间的对应信息，大幅降低了数据的维度，并充分编码了形状引起的骨骼长度变化。这使得训练时能够利用长期上下文，而不会遇到复杂度问题。第二个损失称为目标几何保持损失 $\mathcal{L}_{geom}$，用于促使输出序列的每一帧都具有与 $\mathcal{S}_{tpose}^B$相似的几何细节。逐帧独立考虑 $\mathcal{L}_{geom}$，便可处理包含大量几何细节的高维点云。我们的训练过程旨在优化

$$\mathcal{L} = \mathcal{L}_{motion} + \mathcal{L}_{geom}. \tag{1}$$

在推理过程中，我们的方法会将 A 的运动重定向到 B 的骨架层级，并利用由此得到的 B 的骨骼运动，在一次前向传播中将几何细节从 $\mathcal{S}_{tpose}^B$ 迁移过来。这样便能实现高效的在线动作重定向。图2 直观展示了我们的方法。

3.1. 保持源动作

我们的方法旨在保留源角色在骨架层级上的运动。在骨架层级上进行处理是我们工作的关键所在，因为这种简化表示使我们能够利用长期时间上下文进行训练，而不会

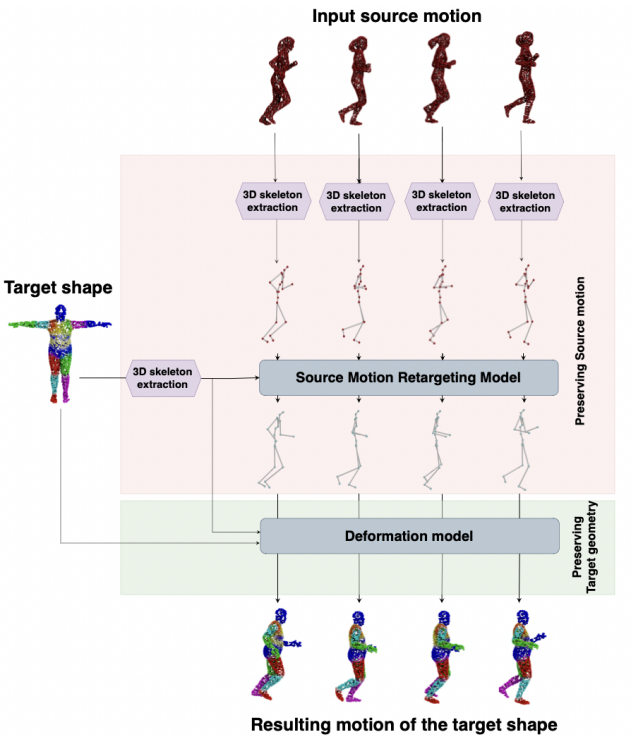

图 2. 我们的方法以一段由无结构点云构成的源序列和一个目标点云作为输入，并输出执行输入动作的目标角色。该方法首先在骨架层级将源动作重定向到目标角色（红色部分），然后使用变形模型为生成的动作添加目标角色的表面细节（绿色部分）。

遇到复杂度问题。我们首先详细介绍网络的架构，并定义源动作保留损失 $\mathcal{L}_{motion}$，然后概述如何根据方法的输入 $\{\mathcal{S}_i^A\}_{i=1}^n$ 和 $\mathcal{S}_{tpose}^B$ 计算骨架表示。下文中，$\{\mathcal{J}_i^A\}_{i=1}^n$ 和 $\mathcal{J}_{tpose}^B$ 分别表示 $\{\mathcal{S}_i^A\}_{i=1}^n$ 和 $\mathcal{S}_{tpose}^B$的骨架表示。

架构 我们考虑将骨骼运动从 $\{\mathcal{J}^A\}_{i=1}^n$ 转移到 $\mathcal{J}_{tpose}^B$ 这一任务。由于缺乏包含两个不同主体执行相同动作的动作捕捉数据，我们采用无监督方式进行训练，不使用动作与其重定向版本之间的真实配对。为解决这一问题，我们受到一种骨架动作重定向模型（SMRM）的启发 [38]该模型以 $\{\mathcal{J}_i^A\}_{i=1}^n$ 和 $\mathcal{J}_{tpose}^B$ 为输入，输出相对关节旋转 $\{\theta_i^B\}_{i=1}^n$。

SMRM 首先使用第一个循环网络从 $\{\mathcal{J}_i^A\}_{i=1}^n$ 中提取高层运动特征。随后利用这些特征以及 $\mathcal{J}_{tpose}^B$，通过第二个循环单元生成关节旋转 $\{\theta_i^B\}_{i=1}^n$。再使用可微正向运动学层将生成的旋转应用于 $\mathcal{J}_{tpose}^B$，从而生成重定向后的骨骼运动 $\{\mathcal{J}_i^B\}_{i=1}^n$。

的定义 $\mathcal{L}_{motion}$ 为了训练我们的网络，我们使用循环一致性损失、对抗损失，以及鼓励时间平滑动作的损失；这些损失均在骨架层级进行评估。循环一致性损失将动作从主体 A 重定向到 B，再将生成的动作重定向回 A，并将主体 A 的生成动作与其初始动作进行比较。我们的损失为

$$\begin{aligned}\mathcal{L}_{motion} &= \mathcal{L}_{cycon} + \mathcal{L}_{adv} \\ &+ \lambda_{rot}\mathcal{L}_{rot} + \lambda_{smooth}\mathcal{L}_{smooth} \end{aligned} \quad (2)$$

权重为 $\lambda_{rot} = 0.01$ 和 $\lambda_{smooth} = 0.1$。

该损失的灵感来自 SMRM，后者使用循环一致性损失 $\mathcal{L}_{cycon} = \text{MSE}(\hat{\mathcal{J}}_i^A, \mathcal{J}_i^A)$，其中 MSE 是均方误差，作用于关节位置，并使用对抗损失$\mathcal{L}_{adv}$。 $\mathcal{L}_{adv}$ 利用判别器网络，该网络与重定向网络通过极大极小博弈进行训练，以区分真实动作和重定向后的动作。为了提升性能，我们加入了两个额外的损失函数来正则化生成的动作。第一个损失是针对关节旋转的循环一致性损失，这有助于避免不自然的动作，可表示为 $\mathcal{L}_{rot} = \text{MSE}(\hat{\theta}_i^A, \theta_i^A)$。与使用四元数表示关节旋转不同，我们采用 6D 表示，因为研究表明，这种表示对深度学习应用更有利 [50]。第二个损失通过重建速度来实现动作的时间平滑，可表示为 $\mathcal{L}_{smooth} = \text{MSE}(\Delta\hat{\mathcal{J}}_i^A, \Delta\mathcal{J}_i^A)$，其中 $\Delta\mathcal{J}_i = \mathcal{J}_{i+1} - \mathcal{J}_i$。

三维骨架提取 下面介绍如何从 $\{\mathcal{S}_i^A\}_{i=1}^n$ 和$\mathcal{S}_{tpose}^B$中提取骨架 $\{\mathcal{J}^A\}_{i=1}^n$ 和 $\mathcal{J}_{tpose}^B$。为支持处理任意长度的输入序列，骨架提取按帧进行。为了处理无结构点云，骨架提取必须对观测点的数量和顺序具有鲁棒性。为此，我们采用 PointFormer [47] 架构从点云中提取顺序不变特征，然后通过多层感知机根据这些特征回归关节位置。我们选择 PointFormer 架构，是因为它能够提取点云的局部特征，从而实现精准的关节预测，并能很好地泛化到未见姿势。我们将这一用于恢复参数化骨架的函数记为 $\mathcal{J} = \text{skeleton}(\mathcal{S})$。骨架提取通过监督方式进行训练，同时最小化真实关节 $\mathcal{J}$ 与预测关节之间的均方误差，即$\mathcal{L}_{\text{skeleton}} = \text{MSE}(\text{skeleton}(\mathcal{S}) \mathcal{J})$。我们的方法独立于所选的骨架参数化方式。在我们的实现中，骨架由 22 个关节位置进行参数化，对应于 SMPL 人体模型 [27]中除根关节以外的所有关节。

3.2. 保持目标几何

我们的方法旨在从表面层面保留目标角色的精细几何。为此，我们学习一种具备空间隐式几何感知能力的变形模型，使目标角色 $\mathcal{S}^B_{tpose}$ 能够摆出在骨架 $\mathcal{J}^B_{tpose}$上定义的任意姿势 θ^B 。隐式模型的关键优势在于，推理时可以对输入进行任意采样。我们将介绍该变形模型，随后详细说明网络架构，并定义目标几何保持损失 $\mathcal{L}_{geom}$。

变形模型 给定骨骼运动来驱动三维网格的一种常见策略，是通过一组蒙皮权重将表面顶点关联到骨架关节，并使用蒙皮技术驱动形状。尽管该领域近期的研究取得了令人瞩目的成果 [11, 32,33, 42, 43], 但这些方法通常适用于配准数据，或在训练过程中使用蒙皮权重进行监督，而我们的情况并不具备这些条件。

为支持任意采样的输入点云，我们学习了一种几何感知变形模型，该模型可用于评估处于 T 姿势的人体表面上的任意点。受构建人体形状空间的相关研究启发 [3, 27], ，我们将变形建模为两个部分：蒙皮权重预测器和姿势校正偏移预测器。为减少伪影，在应用线性混合蒙皮之前，将姿势校正偏移添加到 T 姿势下的顶点上。我们选择这种模型，是因为它简单，而且在建模人体变形时表现出色。

给定目标关节 $\mathcal{J}^B_{tpose}$、旋转集合$\{\theta^B_i\}^n_{i=1}$、蒙皮权重 W^B_j以及点 p_j 在 $\mathcal{S}^B_{tpose}$上的姿势校正偏移 $\{o^B_{i,j}\}^n_{i=1}$ ，我们将生成帧 i 中第 j个顶点的函数记为

$$\mathcal{S}^B_{i,j} = \text{retarget}\left(\theta^B_i, \mathcal{J}^B_{tpose}, W^B_j, p_j + o^B_{i,j}\right). \quad (3)$$

架构 我们的架构由两个网络组成。首先，蒙皮权重预测网络 W_{net} 以目标关节 $\mathcal{J}^B_{tpose}$ 和 $\mathcal{S}^B_{tpose}$ 上的一个三维点 p 为输入，并为每个关节输出一个蒙皮权重 $W = W_{net}(\mathcal{J}^B_{tpose}, p)$。受 [11], 的启发，我们使用 Softmax 激活函数将预测的蒙皮权重约束为总和等于 1。其次，姿势校正偏移网络 O_{net} 以相对关节旋转 θ^B 和 p 为输入，并输出一个位于 $\mathbb{R}^3$ 中的偏移向量，即 $o^B = O_{net}(\theta^B, p)$。这两个网络均由三个隐藏层的多层感知机构成。第一个多层感知机用于建模W_{net} ，后接 Softmax 激活函数。第二个多层感知机用于建模 O_{net}。

定义 $\mathcal{L}_{geom}$ 我们的网络在具有已知姿势 $\mathcal{S}_{tpose}$ 和$\mathcal{S}_\theta$ 关节旋转 θ的同一身份的帧对上进行训练。这使得我们能够使用点对点重建损失。考虑 $\mathcal{S}_{tpose}$ 上的点 p 及其在 $\mathcal{S}_\theta$上的对应点 p' 。我们的损失可写为

$$\mathcal{L}_{geom} = \sum_{(p,p')} ||p' - \text{retarget}(\theta, \mathcal{J}_{tpose}, W, p + o)||^2, \quad (4)$$

其中 $W = W_{net}(\mathcal{J}_{tpose}, p)$ 和 $o = O_{net}(\theta, p)$。

3.3. 训练

受近期研究 [17, 18, 31, 46], 的启发，我们选择分阶段训练策略，以提高稳定性并降低计算复杂度。具体而言，我们通过两个独立阶段分别优化$\mathcal{L}_{motion}$ 和 $\mathcal{L}_{geom}$ ：使用骨架运动来优化 $\mathcal{L}_{motion}$ ，并使用密集采样的静态点云来优化 $\mathcal{L}_{geom}$。

训练时，我们利用 AMASS数据集 [29],，其中包含一系列经过参数化人体模型 [27]拟合的动作捕捉数据集，从而获得密集的逐帧表示。对于每一帧，都提供了对齐的表面信息和骨架信息。所有动作序列均以每秒30帧（FPS）的速度进行时间对齐。作为训练数据，我们选取了120种身体形状，执行2536种不同动作，总计65000秒的运动。作为验证数据，我们选取了3种身体形状，执行24种不同动作，总计147秒的运动；作为测试集，我们选取了7种身体形状，执行44种动作，总计1715秒的运动。我们将此测试集称为*AMASS*测试集。

所有处理几何细节数据的网络，均通过每秒运动采样一帧进行训练。为避免学习模板网格的拓扑偏差，我们在每个三维网格表面随机均匀采样 N 个点，并添加高斯噪声以生成输入扫描模型。

为在骨架层级训练运动重定向，我们从 AMASS 中随机采样固定时长的序列，并且仅考虑其骨骼运动。3D骨架提取器使用从 AMASS 采样的静态三维点云及其对应骨架进行训练。

几何变形模型使用静态 AMASS 数据进行训练，数据中包含同一人的、具有对应关系且已知旋转的帧对 $\{\mathcal{S}_{tpose}, \mathcal{S}_\theta\}$。为在消除 SMPL 拓扑结构所致偏差的同时保留对应关系，我们对模型进行均匀重采样，并保留对应信息。

4. 实验

我们现在评估本文的主要贡献：利用长期时间上下文进行学习能够改善结果；我们的方法优于最先进方法，并且能够泛化到此前未见过的目标形状和源运动。首先，我们通过定量结果表明，考虑长期时间上下文能够提高动作重定向准确度。其次，我们将采用标准误差指标的定量比较，以及通过用户研究进行的评估，与最先进方法进行比较，同时考察几何细节保留和骨架级动作重定向。

我们研究了具有挑战性的形状迁移，其中目标形状既包括裸露形态，也包括着装形态；这些形状和动作在训练期间均未出现。最后，我们展示了将4D多视角采集平台的原始输出重定向到新目标角色的结果。更多结果见补充材料。

数据 为了公平评估比较中考虑的不同方法，我们在两个测试集上评估所有方法。首先，测试集包含执行同一组长时动作的典型裸露人体形状。为构建该数据集，我们使用SMPL 模型创建了 4 种身体形状 [27],，这也是评估人体形变迁移方法时的常见做法 例如。 [35]。我们沿前 2 个主成分以 ± 2 个标准差为间隔采样身体形状，以涵盖人体形状的主要变化，即女性、男性、纤瘦和肥胖。基于骨架的重定向方法例如。 [39] 通常在 Mixamo数据集 ‡上进行评估。受此启发，我们使用 Mixamo 创建一组包含 4 种动作的数据，并将其重定向到所有身体形状上，以生成相应的真实值动作。我们将其称为 *SMPL*测试集。第二个测试集包含穿着服装并执行长时动作的角色。该测试集可用于评估不同方法对目标形状中几何细节的泛化能力。为生成该测试集，我们从 CAPE 中选取 4 个穿着紧身服装的角色 [28],，具体而言，使用 CAPE 中名为“short long”、“short short”、“long short”和“long long”的服装，并将 Mixamo数据集中的 4 种动作分别重定向到这些模型上。我们将其称为 *CAPE* 测试集。请注意，对于 SMPL 和 CAPE 两个测试集，4 种动作——“Walking”、“Jogging”、“Throw”和“Hip Hop Dancing”——在全局轨迹和局部动作方面各不相同，并且这些方法在训练期间均未见过任何一种身体形状或动作。此外，使用 Mixamo 将相同动作重定向到不同身体形状后，即可获得真实重定向结果，用于定量测试。我们还使用通过多视角相机设置采集的未跟踪数据测试集 [30]。对于该数据集，没有可用的真实重定向结果，因此我们提供定性结果。我们将其称为多视图测试集。

评估指标 我们的目标是从整体运动保持情况和目标几何的细节保持情况两个方面评估重定向结果。

为了评估运动，我们采用两个仅在骨架层级进行评估的互补指标。首先，我们考察真实值与预测结果之间的平均关节误差（MPJPE），它用于评估关节位置的准确度，以及 Procrustes 对齐的 MPJPE（PA-MPJPE），后者消除了全局位移误差。其次，为了评估运动平滑性，我们考察真实值与预测运动之间的平均加速度差（Acc），以及其 Procrustes 对齐版本（PA-Acc）。

为了评估目标几何的细节保持情况，我们使用两种互补的表面指标。第一种是计算真实值与重定向结果之间的平均顶点距离（MPVD）和 Procrustes 对齐的 MPVD（PA-MPVD），用于评估预测表面的全局外在准确性。第二种是计算真实值与重定向结果之间的平均边长差（MDEL），以评估内在几何的保持情况。由于我们处理的是点云，因此通过连接真实值中每个点的 6 个最近邻来构造边。

4.1. 利用长期时间上下文进行学习

我们的第一个实验表明，训练时考虑超出少数几帧的时间上下文，有利于运动重定向。我们使用包含不同帧数的动作序列训练模型，即每个模型的所有训练序列帧数固定，范围从5 帧（类似于形状变形迁移方法 [9, 35]）到60 帧（类似于基于骨架的方法 [1, 24, 38]）。表2 展示了结果。纳入长期上下文可将几乎所有指标提升至多达30 帧。在接下来的所有实验中，我们使用以30 帧序列训练的模型。

4.2. 与最先进方法的定量比较

现在，我们将对最先进的运动重定向方法进行比较分析。

对比方法如表1所总结，现有方法可分为三类。基于骨架的方法需要将手工制作的蒙皮权重作为输入，而我们的测试集不具备这些权重，因此无法与我们的方法进行比较。因此，我们将本方法与无需对应关系的形变传递方法和动作先验进行比较。对于形变传递方法，我们将其与

	骨骼运动				细节保留		
上下文时长	MPJPE (m)↓	PA-MPJPE (m)↓	Acc↓	PA-Acc↓	MPVD (m)↓	PA-MPVD (m)↓	MEDL (毫米)↓
0.16秒（5帧）	0.203	0.073	0.021	0.009	0.158	0.061	**0.505**
0.33秒（10帧）	0.167	0.050	0.021	0.008	0.137	0.045	0.507
0.5秒（15帧）	0.161	0.046	**0.020**	0.008	0.136	0.044	0.519
1秒（30帧）	**0.158**	**0.043**	0.021	**0.008**	0.134	**0.040**	0.524
2秒（60帧）	0.160	0.044	0.023	0.008	**0.131**	0.041	0.509

表2. 在SMPL测试集上使用不同时间上下文进行学习。使用 $1s$ 的长期上下文进行训练可提升结果。

‡https://www.mixamo.com

	骨骼运动				细节保留		
	MPJPE (米) ↓	PA-MPJPE (米) ↓	准确度↓	PA-准确度↓	MPVD (m) ↓	PA-MPVD (m) ↓	MDEL (mm) ↓
SMPL测试集中的裸露目标形状							
H4D [18]	0.238	0.096	**0.019**	0.014	0.152	0.078	402.238
NPT [41]	0.388	0.165	0.024	0.014	0.227	0.132	2.739
我们的方法	**0.158**	**0.043**	0.021	**0.008**	**0.134**	**0.040**	**0.524**
来自 CAPE 测试集的着装目标形状							
H4D [18]	0.161	0.091	**0.019**	0.014	0.096	0.074	395.683
NPT [41]	0.373	0.168	0.022	0.013	0.173	0.135	2.845
Ours	**0.105**	**0.040**	0.023	**0.008**	**0.075**	**0.038**	**0.539**

表3. 裸露（上）和着装（下）目标形状与最先进方法的比较。最佳表现分数以粗体显示。

NPT [41]。Aniformers [9] 在训练期间需要不同个体执行相同运动的真实值配对，而我们的情况不具备这些配对，因此无法进行比较。对于动作先验，我们将其与 H4D [18]进行比较。我们向该方法提供完整的身份序列，以提取身体形状参数。

定量结果表3 展示了针对 SMPL 和 CAPE 测试集形状的定量结果。在两个数据集上的几乎所有指标中，我们的方法均显著优于 NPT。这是因为 NPT 逐帧运行，不具备任何时间上下文，因此有时会产生时间上不一致的结果。在两个数据集上的几乎所有评估指标中，我们的方法也优于 H4D。具体而言，经过普罗克鲁斯特对齐后，两个测试集的骨骼关节位置准确度均显著更高；而在未经普罗克鲁斯特对齐的情况下，均值在 $4.9cm$ 裸露目标形状上更准确，而对于着装目标形状则在 $2mm$ 范围内。使用我们的模型时，关节加速度更准确。在两个数据集上，采用普罗克鲁斯特对齐时，我们的模型能显著更好地保留几何细节；而不进行普罗克鲁斯特对齐时，两个模型的误差相近。这表明我们的模型能更好地保留几何细节。

与使用对应关系的重定向方法进行比较 由于我们的表现远胜于无对应关系的重定向方法，我们还将我们的方法与 TST 进行比较 [35], 。TST 是一种考虑短期时间动态的最先进变形迁移方法，在训练和推理时都需要点对点对应关系；我们为该方法提供了对应关系。在评估指标上，我们的方法仅略逊于 TST，所有误差均在 $2cm$ 的 TST 误差范围内。这一接近于利用对应关系的方法的表现，凸显了我们的无对应关系方法的潜力。评估指标的详细数值结果见补充材料。

4.3 用户研究

我们还通过开展用户研究，进一步评估人们对我们重定向结果的感知，考察其在运动保留和几何细节方面的表现。

设计我们设计了一项被试内感知实验，以重定向方法为主要因素。由于 NPT 在定量表现上明显逊于其他方法，为尽量缩短每位用户的实验时长，我们未将其纳入感知评估。我们将本方法的结果与采用对应关系的 H4D 和 TST 进行了比较。30 名无偿志愿者参与了本研究。每个试次中，参与者会看到展示源动作、目标形状和重定向结果的刺激材料，并被要求分别以7点李克特量表评价动作与源动作的相似度，以及形状与目标形状的相似度。针对每种方法，使用 3 种运动类型（行走、嘻哈舞、投掷）和 5 种不同的形状迁移进行了 15 次重复。参与者按随机顺序完成了 45 个试次。统计分析采用 ART方差分析和事后Tukey检验。

结果结果表明，该方法对运动和形状分量均有显著影响（$p<0.0001$）。事后分析显示，用户认为与 H4D 相比，我们的方法在保持运动和身体形状方面显著更好（$p<0.0001$）。TST 在运动保持方面优于我们的方法（$p<0.001$），而我们的方法更能保持目标的身体形状（$p<0.0001$）。有关实验设计和结果的更多详情见补充材料。

4.4. 消融实验

我们针对优化$\mathcal{L}_{motion}$ 和 $\mathcal{L}_{geom}$的架构进行了消融实验。

源运动保持我们首先在 SMPL 测试集密集采样的几何数据上，从骨架层级评估运动迁移。表 4 显示，采用 6D旋转表示并进行旋转监督$\mathcal{L}_{rot}$ 可改善重定向。从四元数转换为 6D 旋转后，几乎所有指标都有显著提升。使用 $\mathcal{L}_{rot}$ 进行监督也能提升脚部、手腕和头部关节周围的几何细节保持。这是因为这些关节是层次化骨骼的叶节点，因此它们不会影响 $\mathcal{L}_{cycon}$，但会影响 $\mathcal{L}_{rot}$。

为评估三维骨架提取效果，我们使用 AMASS测试集。我们比较了 PointNet [34],（考虑全局特征）与 PointFormer [47],（引入局部特征）。使用 PointFormer 可将 MPJPE 从 **46 毫米** 降低到**21 毫米** 。

目标几何保持为消融用于优化 $\mathcal{L}_{geom}$ 的变形模型，我们使用 AMASS测试集。使用姿势校正偏移可通过将平均重建误差从**8.4mm**降低至**5mm**，从而提高精度。更多消融实验见补充材料。

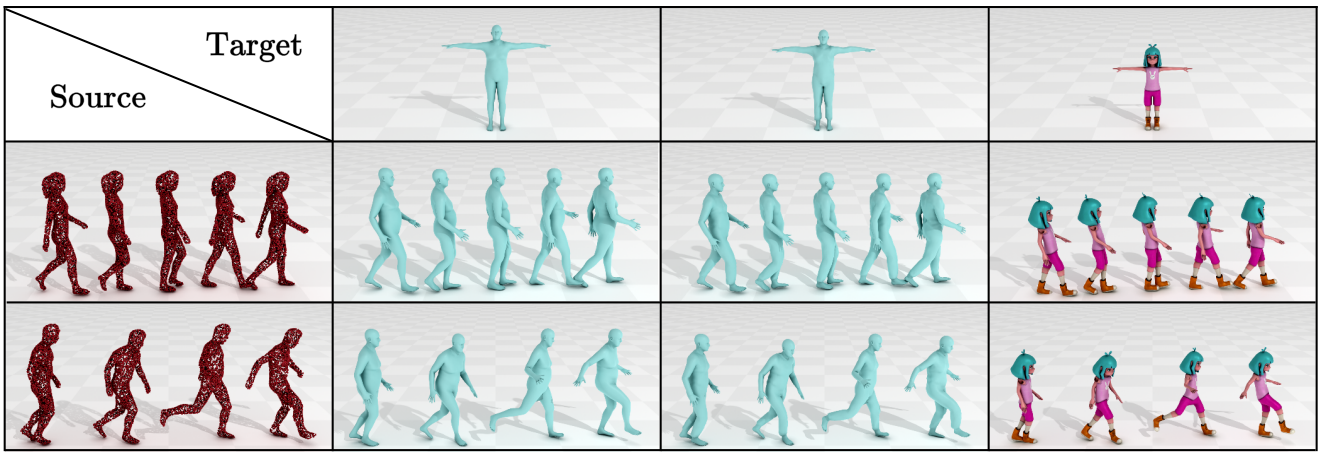

图3。直接使用未跟踪捕获的四维数据为目标形状制作动画。我们考虑了一个行走动作（上）和一个踢腿动作（下），并将其重定向至裸露（左）、着装（中）和CAD生成（右）的目标形状。

	骨骼运动				细节保留		
	MPJPE（米）↓	PA-MPJPE（米）↓	Acc ↓	PA-Acc ↓	MPVD（米）↓	PA-MPVD（米）↓	MEDL（米）↓
[38]	0.163	0.056	0.023	0.011	0.165	0.089	1.597
[38]+6D	0.159	0.043	0.024	0.008	0.137	0.048	0.684
[38]+$\mathcal{L}_{rot}$	0.161	0.044	0.021	0.008	0.156	0.076	1.445
[38]+$\mathcal{L}_{rot}$+6D	**0.158**	**0.043**	**0.021**	**0.008**	**0.134**	**0.040**	**0.524**

表 4. 运动重定向的消融实验。对采用不同旋转表示的 SMPL形状进行定量评估（用于 θ），并对旋转循环一致性损失进行消融 $\mathcal{L}_{rot}$。

4.5. 使用采集的四维数据为目标形状制作动画

我们展示了在使用多视图采集平台的原始 4D 输出为目标形状制作动画时，本方法的性能。我们将多视图测试集中的序列直接重定向到使用 SMPL、CAPE 生成的角色，以及使用计算机辅助设计 (CAD) 工具设计的 Mixamo 角色。请注意，输入序列受到采集噪声的影响，且没有可用的对应信息，即我们输入的是未跟踪的 4D 序列。

图 3 展示了使用我们的方法获得的结果。请注意，我们的方法既保留了源序列的运动，也保留了不同目标形状的几何细节。据我们所知，我们的方法是首个能够在线重定向未经跟踪的 4D 采集数据的方法。

这些样本展示了我们的方法对未见过形状的鲁棒性。第一个源动作呈现了一种包含头发、训练期间未曾见过的身体形状，证明了我们的方法对于未见过的源形状具有鲁棒性。对 CAPE 和 CAD 生成的目标形状的几何保留效果，也表明我们的模型在处理未见过的目标形状时具有良好的泛化能力。

定量和定性结果表明，我们的方法能够很好地泛化到未见过的动作例如来自 Mixamo 的动作，以及未见过的形状例如CAPE 中的穿衣形状和多视角采集平台未经跟踪的 4D 输出。

4.6. 局限性

尽管我们的方法在无对应关系动作重定向问题上取得了最先进方法的效果，但仍存在局限性。由于变形模型本身的限制，该方法无法泛化到穿着宽松服装的形状。该方法还仅适用于可由固定拓扑骨架参数化的形状，也无法捕捉精细的手部或面部动作。潜在的负面社会影响将在补充材料中讨论。

5. 结论

我们提出了首个在线重定向方法，使目标形状能够通过无对应关系的源动作进行驱动。我们证明，在对密集动作进行重定向时，纳入 $1s$ 的长期时间上下文是有益的。我们将低维中间骨骼表示与蒙皮先验相结合，使其能够很好地泛化到未见过的形状和动作。特别是，我们证明了仅在裸身形状上训练的模型能够泛化到包含头发和服装的输入。

未来的工作包括扩展到复杂服装，例如宽松或多层服装。一种方案是在模型中显式纳入服装。将该解决方案扩展到手部和表情也很有意义。

6. 致谢

我们感谢 Jo˜ao Regateiro 和 Edmond Boyer 提供的宝贵讨论。本研究由法国国家研究署（ANR）项目 3DMOVE - 19-CE23-0013 资助。

参考文献

[1] Kfir Aberman、Rundi Wu、Dani Lischinski、Baoquan Chen 和 Daniel Cohen-Or。学习用于二维运动重定向的与角色无关的运动。《图形学汇刊》，38 (4):75:1–14，2019。3, 6[2] Kfir Aberman、Peizhuo Li、Dani Lischinski、Olga Sorkine- Hornung、Daniel Cohen-Or 和 Baoquan Chen。面向深度运动重定向的骨架感知网络。《图形学汇刊》，39(4):62–1，2020。2, 3[3] Brett Allen、Brian Curless、Zoran Popovi´c 和 Aaron Hertz- mann。学习身份与姿势相关的身体形状变化模型，以实现实时合成。载于《2006年ACM SIGGRAPH/Eurographics计算机动画研讨会论文集》，第147–156页。Citeseer，2006。5[4] Ilya Baran、Daniel Vlasic、Eitan Grinspun 和 Jovan Popovi´c。语义形变迁移。载于 SIGGRAPH，第1–6页。2009。2[5] Jean Basset、Adnane Boukhayma、Stefanie Wuhrer、Franck Multon 和 Edmond Boyer。神经人体形变迁移。载于国际3D视觉会议，第545–554页，2021。1, 2, 3[6] Federica Bogo、Javier Romero、Matthew Loper 和 Michael J Black。FAUST：用于三维网格配准的数据集与评估。载于计算机视觉与模式识别会议，第3794–3801页，2014。1[7] Federica Bogo、Javier Romero、Gerard Pons-Moll 和 Michael J Black。Dynamic FAUST：配准运动中的人体。载于计算机视觉与模式识别会议，第6233–6242页，2017。1[8] Adnane Boukhayma、Jean-S´ebastien Franco 和 Edmond Boyer。基于高斯过程回归的表面运动捕捉迁移。载于计算机视觉与模式识别会议，第184–192页，2017。2[9] Haoyu Chen、Hao Tang、Nicu Sebe 和 Guoying Zhao。AniFormer：使用 Transformer 的数据驱动三维动画。载于英国机器视觉会议，2021。2, 3, 6, 7[10] Haoyu Chen、Hao Tang、Henglin Shi、Wei Peng、Nicu Sebe 和 Guoying Zhao。用于无监督三维姿势迁移的内在—外在属性保持生成对抗网络。载于国际计算机视觉会议，第8630–8639页，2021。3[11] Xu Chen、Yufeng Zheng、Michael J Black、Otmar Hilliges 和 Andreas Geiger。SNARF：用于动画化非刚性神经隐式形状的可微前向蒙皮。载于 IEEE/CVF 国际计算机视觉会议论文集，第11594–11604页，2021。5[12] Kwang-Jin Choi 和 Hyeong-Seok Ko。在线运动重定向。《可视化与计算机动画期刊》，11(5):223–235，2000。2[13] Luca Cosmo、Antonio Norelli、Oshri Halimi、Ron Kimmel 和 Emanuele Rodola。LIMP：利用度量保持先验学习潜在形状表征。载于《欧洲计算机视觉会议》，第19–35页。施普林格，2020。2[14] Brian Delhaisse、Domingo Esteban、Leonel Rozo 和 Darwin Caldwell。在运动学结构相似的机器人之间迁移共享潜在空间的迁移学习。载于国际神经网络国际联合会议，页码 4142–4149，2017年。2[15] Michael Gleicher。将运动重定向到新角色。在 SIGGRAPH，页码 33–42，1998年。1, 2[16] Hanyoung Jang、Byungjun Kwon、Moonwon Yu、Seong Uk Kim 和 Jongmin Kim。用于运动重定向的变分 U-Net。在 SIGGRAPH 亚洲海报，页码 1–2。2018年。2[17] Boyi Jiang、Juyong Zhang、Yang Hong、Jinhao Luo、Ligang Liu 和 Hujun Bao。BCNet：从单张图像中学习身体和服装形状。在欧洲计算机视觉会议，页码 18–35。施普林格，2020年。5[18] Boyan Jiang、Yinda Zhang、Xingkui Wei、Xiangyang Xue 和 Yanwei Fu。H4D：通过学习神经组合表示进行人体四维建模。在计算机视觉与模式识别会议，页码 19355–19365，2022年。2, 3, 5, 7[19] Pushkar Joshi、Wen C Tien、Mathieu Desbrun 和 Fr´ed´eric Pighin。学习基于混合变形的逼真面部动画控制。在计算机动画研讨会，页码 17–20。2003年。2[20] Diederik Kingma 和 Jimmy Ba。Adam：一种随机优化方法。在国际学习表征会议，2015年。11[21] Hyeong-Seok Ko、Kwang-Jin Choi、Min Gyu Choi、Seyoon Tak、Byoungwon Choe 和 Oh-Young Song。创建数字演员的研究问题。在 Eurographics 前沿技术报告，2003年。2[22] Paul G Kry、Doug L James 和 Dinesh K Pai。Eigen-skin：用于硬件的实时大变形角色蒙皮。在计算机动画研讨会，页码 153–159，2002年。2[23] Jehee Lee 和 Sung Yong Shin。面向类人形体的交互式运动编辑分层方法。在 SIG-GRAPH，页码 39 48, 1999 2 –。[24] Jongin Lim、Hyung Jin Chang 和 Jin Young Choi。PMNet：学习解耦姿态与运动，实现无监督运动重定向。在英国机器视觉会议，编号 6:1–7，2019年。1, 3, 6[25] Yaron Lipman、Olga Sorkine、David Levin 和 Daniel Cohen-Or。网格的线性旋转不变坐标。《图形学汇刊》，24(3):479–487，2005年。2[26] Lijuan Liu、Youyi Zheng、Di Tang、Yi Yuan、Changjie Fan 和 Kun Zhou。Neuroskinning：利用深度图网络为生产级角色自动进行蒙皮绑定。《图形学汇刊》，38(4):1–12，2019年。2[27] Matthew Loper、Naureen Mahmood、Javier Romero、Gerard Pons-Moll 和 Michael J Black。SMPL：一种蒙皮多人线性模型。《图形学汇刊》，34 16，2015 6 1 4, 5, 6 ( ): –。[28] Qianli Ma 、 Jinlong Yang 、 Anurag Ranjan 、 Sergi Pujades、Gerard Pons-Moll、Siyu Tang 和 Michael J Black。学习为三维人物穿上生成式服装。在 IEEE/CVF 计算机视觉与模式识别会议论文集，页码 6469–6478，2020年。6[29] Naureen Mahmood、Nima Ghorbani、Nikolaus F Troje、Gerard Pons-Moll 和 Michael J Black。AMASS：以表面形状形式呈现的动作捕捉档案。在国际计算机视觉会议，页码 5442–5451，2019年。1, 5

[30] Mathieu Marsot、Stefanie Wuhrer、Jean-Sebastien Franco 和 Stephane Durocher。用于生成人体运动的结构化潜在空间。在3D视觉会议，2022年。3, 6[31]Dushyant Mehta、Oleksandr Sotnychenko、Franziska Mueller、Weipeng Xu、Mohamed Elgharib、Pascal Fua、Hans-Peter Seidel、Helge Rhodin、Gerard Pons-Moll 和 Christian Theobalt。Xnect：使用单个 RGB 相机进行实时多人三维动作捕捉。《图形学汇刊》，39(4):82–1，2020年。5[32] Albert Mosella-Montoro 和 Javier Ruiz-Hidalgo。Skinningnet：用于合成角色蒙皮预测的双流图卷积神经网络。在计算机视觉与模式识别会议，页码 18593–18602，2022年。2, 5[33] Xuming Ouyang 和 Cunguang Feng。Autoskin：利用深度神经网络进行基于骨架的人体蒙皮。《物理学杂志：会议系列》，页码 032163。IOP 出版社，2020年。5[34] Charles R Qi、Hao Su、Kaichun Mo 和 Leonidas J Guibas。PointNet：用于三维分类和分割的点集深度学习。在 IEEE 计算机视觉与模式识别会议论文集，页码 652 660，2017年 7 –。[35] Jo˜ao Regateiro 和 Edmond Boyer。用于三维人体运动的时序形状迁移网络。在国际3D视觉会议，2022年。1, 2, 3, 6, 7, 11, 12[36] Robert W Sumner 和 Jovan Popovi´c。三角网格的变形迁移。《图形学汇刊》，23(3):399–405，2004年。1, 2[37] Seyoon Tak 和 Hyeong-Seok Ko。基于物理的运动重定向滤波器。《图形学汇刊》，24(1):98–117，2005年。2[38] Ruben Villegas、Jimei Yang、Duygu Ceylan 和 Honglak Lee。用于无监督运动重定向的神经运动学网络。在计算机视觉与模式识别会议，页码 8639–8648，2018年。1, 2, 3, 4, 6, 8, 11[39] Ruben Villegas、Duygu Ceylan、Aaron Hertzmann、Jimei Yang 和 Jun Saito。蒙皮动作的接触感知重定向。在国际计算机视觉会议，2021年。2, 3, 6[40] Timo Von Marcard、Roberto Henschel、Michael J Black、Bodo Rosenhahn 和 Gerard Pons-Moll。利用惯性测量单元和移动相机，在野外恢复准确的三维人体姿态。在欧洲计算机视觉会议，页码 601–617，2018年。1[41] Jiashun Wang、Chao Wen、Yanwei Fu、Haitao Lin、Tianyun Zou、Xiangyang Xue 和 Yinda Zhang。通过空间自适应实例归一化实现神经姿态迁移。在计算机视觉与模式识别会议，页码 5831–5839，2020年。1, 2, 3, 7[42] Zhan Xu、Yang Zhou、Evangelos Kalogerakis、Chris Landreth 和 Karan Singh。Rignet：面向关节角色的神经绑定。《图形学汇刊》，39 14，2020 4 58 1 2, 5 ( ): : –。[43] Ze Yang、Shenlong Wang、Sivabalan Manivasagam、Zeng Huang、Wei-Chiu Ma、Xinchen Yan、Ersin Yumer 和 Raquel Urtasun。S3：用于三维人体建模的神经形状、骨架和蒙皮场。载于 IEEE/CVF 计算机视觉与模式识别会议论文集，第 13284–13293 页，2021年。2, 5[44] Yizhou Yu、Kun Zhou、Dong Xu、Xiaohan Shi、Hujun Bao、Baining Guo 和 Heung-Yeung Shum。通过基于泊松的梯度场操控进行网格编辑。载于 SIGGRAPH，第 644–651 页，2004年。2[45] Jiaxu Zhang、Junwu Weng、Di Kang、Fang Zhao、Shaoli Huang、Xuefei Zhe、Linchao Bao、Ying Shan、Jue Wang 和 Zhigang Tu。通过对运动语义与几何进行残差感知的蒙皮运动重定向。载于 IEEE/CVF 计算机视觉与模式识别会议论文集，第 13864–13872 页，2023年。2, 3[46] Jason Y Zhang、Panna Felsen、Angjoo Kanazawa 和 Jitendra Malik。从视频预测三维人体动态。载于国际计算机视觉会议，第 7114–7123 页，2019年。5[47] Hengshuang Zhao、Li Jiang、Jiaya Jia、Philip HS Torr 和 Vladlen Koltun。点变换器。载于 IEEE/CVF 国际计算机视觉会议论文集，第 16259–16268 页，2021年。4, 7, 11[48] Kun Zhou、Weiwei Xu、Yiying Tong 和 Mathieu Desbrun。将变形迁移至多组件对象。载于《计算机图形论坛》，第 319–325 页，2010年。2[49] Keyang Zhou、Bharat Lal Bhatnagar 和 Gerard Pons-Moll。三维网格的无监督形状与姿态解耦。载于欧洲计算机视觉会议，第 341–357 页，2020年。2[50] Yi Zhou、Connelly Barnes、Jingwan Lu、Jimei Yang 和 Hao Li。论神经网络中旋转表示的连续性。载于 IEEE/CVF 计算机视觉与模式识别会议论文集，第 5745–5753 页，2019年。4

无对应关系的在线人体运动重定向

补充材料

本补充材料介绍了我们提出的框架的网络架构与实现细节，并进一步说明了消融实验。我们还展示了本方法与一种使用对应关系的先进重定向方法之间的比较，并补充解释了我们开展的用户研究。

A. 网络架构与实现细节

我们的方法旨在通过源运动重定向模型，在骨架层级保留源角色的运动；同时利用变形模型保留目标形状的详细几何信息。以下将分别介绍这两部分的细节。

下文中， N 表示批量大小， V 表示扫描模型 $\mathcal{S}$ 的顶点数， J 表示骨架 $\mathcal{J}$ 的关节数。

所有部分均使用 PyTorch 实现，并采用 Adam 进行优化 [20]。整个框架的训练在 GeForce RTX 2080 TI 上约需一天。整个模型包含 1460 万个可训练参数。源运动重定向模型包含 5793927 个参数，3D骨架提取模型包含 9 585 730 个参数，变形模型包含 257 564 个参数。

A.1. 源运动保持

源动作重定向模型是对 [38] 的重新实现。我们在表 5 中总结了其架构。FK 表示前向运动学层。

索引	输入	操作	输出形状
(1)	输入	关节位置 $\mathcal{J}$	$N \times 3 \times J$
(1')	输入	目标骨架 $\mathcal{J}^B_{tpose}$	$N \times 3 \times J$
(2)	输入	平移	$N \times 3$
(3)	(1) + (2)	门控循环单元 $(3 \times (J+1) \rightarrow 512,\ 2)$ Dropout $(p = 0.2)$	$N \times 512 \times J$
(4)	(3) + (1')	门控循环单元 $(3 \times J + 512 \rightarrow 512,\ 2)$ Dropout $(p = 0.2)$	$N \times 512$
(5)	(4)	线性层 $(512 \rightarrow 6 \times J)$	$N \times 6 \times J$
(6)	(4)	线性层 $(512 \rightarrow 3)$	$N \times 3$
(7)	(5) + (1')	FK	$N \times 3 \times J$

Tabl图 5. 用于源动作重定向的网络架构 向。

该网络训练 450 个训练轮次，每个训练轮次包含 2 316 个样本。学习率初始为 1e-4，每 150 个训练轮次除以 10。

三维骨架提取表6 介绍了三维骨架提取模块的架构细节。

索引	输入	操作	输出形状	激活函数
(1)	输入	源扫描 $\mathcal{S}^A$	$N \times 3 \times V$	-
(1')	(1)	骨干层 [47]	$N \times 4 \times 512$	-
(2)	(1')	Mean	$N \times 512$	-
(3)	(2)	线性层 $(512 \rightarrow 256)$	$N \times 256$	Relu
(4)	(3)	线性层 $(256 \rightarrow 64)$	$N \times 64$	Relu
(5)	(4)	线性层 $(64 \rightarrow J \times 3)$	$N \times 3 \times J$	-

表6. 用于三维骨架提取的网络架构。

网络训练34个训练轮次，每个训练轮次使用65 000个样本。前30个训练轮次的学习率为1e-4，接着2个训练轮次为1e-5，最后2个训练轮次为1e-6。

A.2. 目标几何保持

变形模块的架构由表 7 中详述的蒙皮权重预测网络和表 8 中详述的姿态校正偏移预测网络组成。

索引	输入	操作	输出形状	激活函数
(1)	$N \times 3 \times (J+1)$	线性层 $(3 \times (J+1)) \rightarrow 256)$ Dropout $(p = 0.2)$	$N \times 256$	Relu
(2)	(1)	线性层 $(256 \rightarrow 256)$ Dropout $(p = 0.2)$	$N \times 256$	Relu
(3)	(2)	线性层 $(256 \rightarrow J)$	$N \times J$	Softmax函数

表7。蒙皮权重预测器的网络架构。

索引	输入	操作	输出形状	激活函数
(1)	$N \times 3 \times (J+1)$	线性层 $(3 \times (J+1) \rightarrow 256)$ Dropout $(p = 0.2)$	$N \times 256$	Relu
(2)	(1)	线性层 $(256 \rightarrow 256)$ Dropout $(p = 0.2)$	$N \times 256$	Relu
(3)	(2)	线性层 $(256 \rightarrow 3)$	$N \times 3$	-

表8. 姿态校正偏移预测器的网络架构。

蒙皮权重预测器和姿态校正偏移预测器网络联合训练28个训练轮次，每个训练轮次包含65 000个样本。第一个训练轮次仅优化蒙皮权重预测网络。学习率为1e-3。

B. 与使用对应关系的重定向方法 TST 的比较

表9 定量比较了我们的方法与 TST（一种 [35],最先进的变形迁移方法，需要点对点对应关系）。请注意，尽管 TST 的表现优于我们的方法，但提升幅度很小。这一结果有力地表明，我们的无对应关系方法仅略逊于一种最先进的

依赖额外对应信息的方法。

	骨骼运动				细节保留		
	MPJPE（m）↓	PA-MPJPE（米）↓	准确度↓	PA-准确度↓	MPVD（米）↓	PA-MPVD（米）↓	MDEL（毫米）↓
来自 SMPL测试集的裸露目标形状							
TST [35]	**0.152**	**0.028**	**0.005**	**0.004**	**0.130**	**0.028**	0.866
Ours	0.158	0.043	0.021	0.008	0.134	0.040	**0.524**
来自 CAPE 测试集的着装目标形状							
TST [35]	**0.096**	**0.027**	**0.005**	**0.004**	**0.064**	**0.028**	0.953
Ours	0.105	0.040	0.023	0.008	0.075	0.038	**0.539**

表 9. 将我们的无对应关系方法与一种需要对应关系的最先进方法进行比较。比较对象为裸露（上）和着装（下）的目标形状。

C. 关于源运动保持的进一步消融研究

现在，我们进一步展示针对源动作重定向模型的消融研究。

首先，正文中的表 4 显示，使用 $\mathcal{L}_{rot}$和六维表示可显著改善所有定量指标。图4进一步展示了采用不同旋转表示以及使用和不使用循环一致性损失时的一个重定向示例。请注意，我们的改动显著改善了重定向结果。

其次，我们在表 10 中用定量结果表明，添加 $\mathcal{L}_{smooth}$ 可改善源动作重定向。我们评估骨架运动，因为 $\mathcal{L}_{smooth}$ 是在骨架上运行的。

	骨骼运动			
	MPJPE（米）↓	PA-MPJPE（米）↓	准确度↓	PA-Acc↓
$\mathcal{L}_{motion}$ 不使用 $\mathcal{L}_{smooth}$	0.160	0.044	0.023	0.008
$\mathcal{L}_{motion}$ 使用 $\mathcal{L}_{smooth}$	**0.158**	**0.043**	**0.021**	**0.008**

表 10. 对 SMPL形状进行 $\mathcal{L}_{smooth}$ 的定量消融实验。

三维骨架提取 在本段中，我们研究了不同点云编码对3D骨架提取模型的影响。正文通过定量结果表明，在此任务中使用 PointFormer 优于使用 PointNet。图5 展示了一个具有挑战性的示例对比。使用 PointNet 时，躯干预测与真实值相差甚远；而使用 PointFormer 时，明显的误差仅出现在右脚和腕关节上。

D. 用户研究

为了进一步评估我们的方法，我们设计了一项用户研究，以评估人们感知到的生成运动相对于源动作的效果，以及感知到的生成形状相对于目标形状的效果。

D.1. 参与者

30名未领取报酬的参与者自愿参加了实验（14名女性，16名男性；年龄：平均=32±11，范围为21–64岁）。他们在动画（中位数 =2，四分位距（IQR） = 3，采用李克特量表评分，范围为1–7，其中1表示新手，7表示专家）和人体运动（中位数 = 3，IQR = 4，采用李克特量表评分，范围为1–7，其中1表示新手，7表示专家）方面的专业水平各不相同。他们的视力均正常或经矫正后正常，并签署了书面知情同意书。本研究遵循《赫尔辛基宣言》，并经机构伦理委员会批准。

D.2. 设计

每个试次中，参与者坐在一台24英寸的计算机屏幕前，观看展示源动作、目标形状和重定向结果的刺激材料，并分别对运动相似度和形状相似度进行评分；源动作和目标形状的评分采用7点李克特量表，如图6所示。

我们使用 PsychoPy 软件构建了实验。实验采用被试内设计，并以重定向方法作为主要因素。我们考察了 3 种重定向方法：我们的方法和 H4D（不使用对应关系），以及 TST（使用对应关系）。每种方法针对不同的动作和形状迁移进行了 15 次重复。我们选取了 3 种动作类型，如图 7所示：
1. 行走，包含全局位移和周期性身体动作；2. 投掷，是一种局部性较强且离散的上半身动作；3. 嘻哈舞，是一种复杂动作。

图 8展示了选定的形状迁移，包括性别、身体人体测量数据和服装变化，即：
1. 女性到男性，2. 男性到女性，3. 肥胖到纤瘦，4. 纤瘦到肥胖，5. 裸露到穿衣。
经过 4 个训练试次后，参与者总共进行了 45 个试次。试次按形状迁移分块呈现。每个区块开始时，参与者会看到源身体形状和目标身体形状的 T 姿势。我们采用拉丁方设计随机排列区块顺序。随后，在每个区块中，参与者按随机顺序完成 9 个试次（3 种方法 × 3 动作）。对于每位参与者和每个试次，因变量为

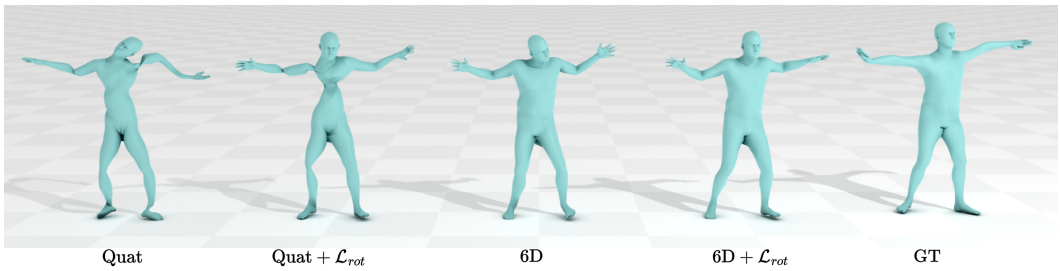

图 4. SMRM 消融实验。从女性身体形状到男性身体形状的一个具有挑战性的嘻哈动作的重定向结果。四元数容易发生扭转；加入 $\mathcal{L}_{rot}$ 可改善头部和脚部的重定向。

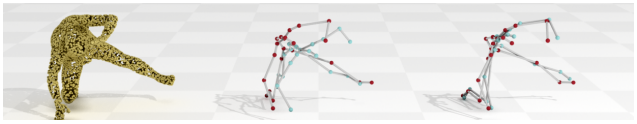

图5. 三维骨架提取消融实验：在 AMASS测试集中的一个具有挑战性的姿势上，使用 PointNet（中）和 PointFormer（右）进行比较（右）。真实值以红色显示，回归结果以蓝色显示。

运动和形状相似度评分。统计分析使用 R 完成，采用重复测量ART方差分析和事后Tukey检验。显著性水平设为 0.05。

D.3. 结果

结果表明，重定向模型对感知到的运动（$F(2, 58) = 297.9, p < 0.0001$）和形状相似度（$F(2, 58) = 85.3, p < 0.0001$）均有显著主效应。该效应具有重要意义，因为效应量分别为 0.91 和 0.74。正如图 9 所示，事后检验表明，与源动作和目标形状相比，我们的方法在运动（$t(58) = 19.01, p < 0.0001$）和形状相似度（$t(58) = 12.74, p < 0.0001$）方面的评分均优于 H4D。此外，尽管我们的方法在感知运动方面的评分低于 TST（$t(58) = 3.75, p < 0.001$），但在感知形状方面优于 TST（$t(58) = 8.87, p < 0.0001$）。

E. 潜在的负面社会影响

我们提出了一种方法，可在不同数字化人体模型之间进行长期且几何细节丰富的运动重定向。未经用户同意，该方法可能被用于为静态三维扫描模型，甚至是由二维图像生成的三维重建结果制作动画，例如用于生成虚假信息。

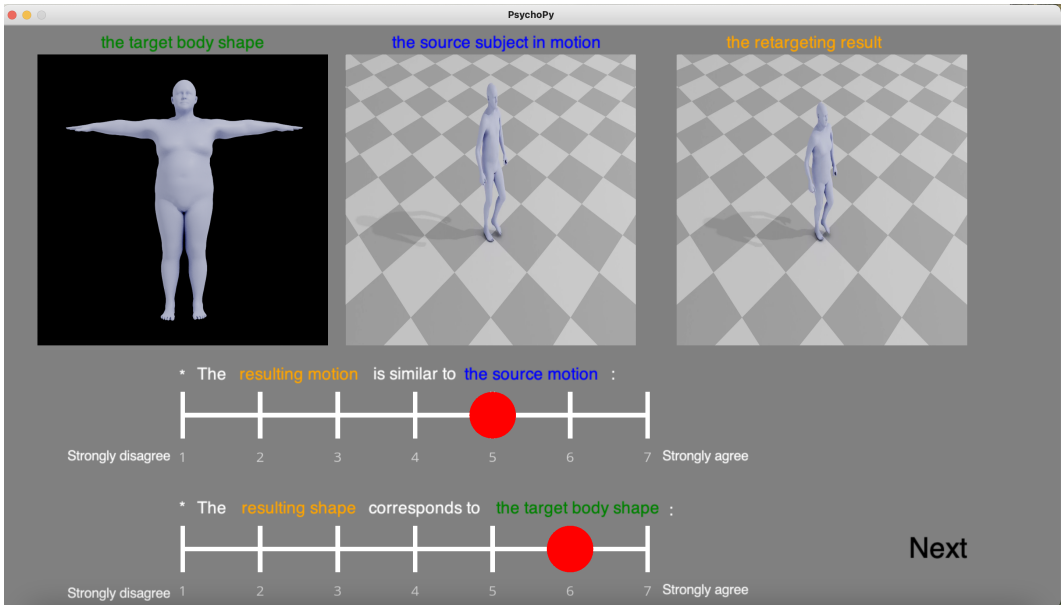

图6. 用户研究中向参与者展示的刺激材料和问题截图。

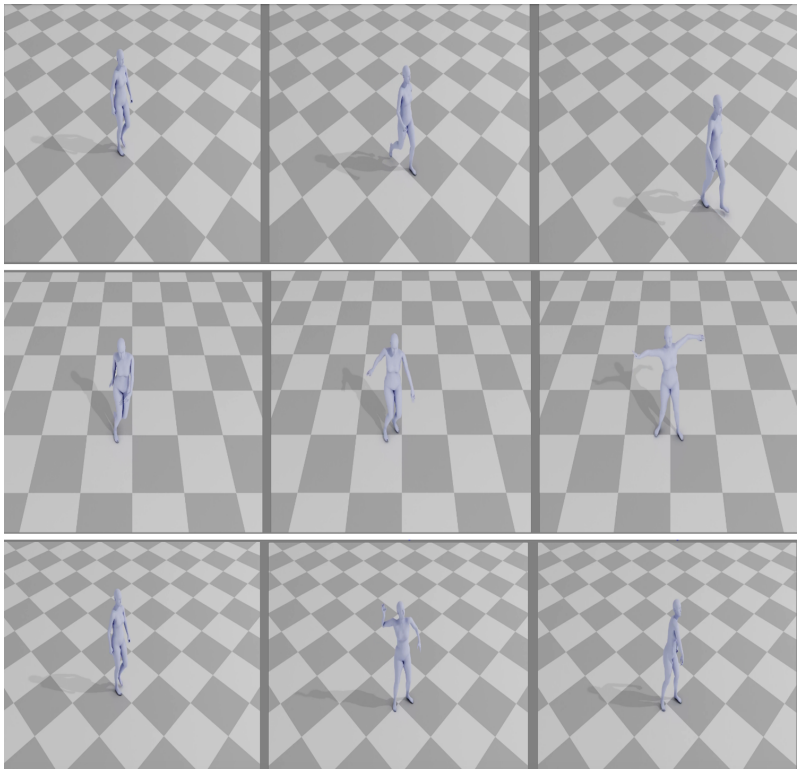

图7. 用户研究中选取的3种动作示意图。上：行走动作；中：嘻哈动作；下：投掷动作。这些样本被用作女性形体的源动作。

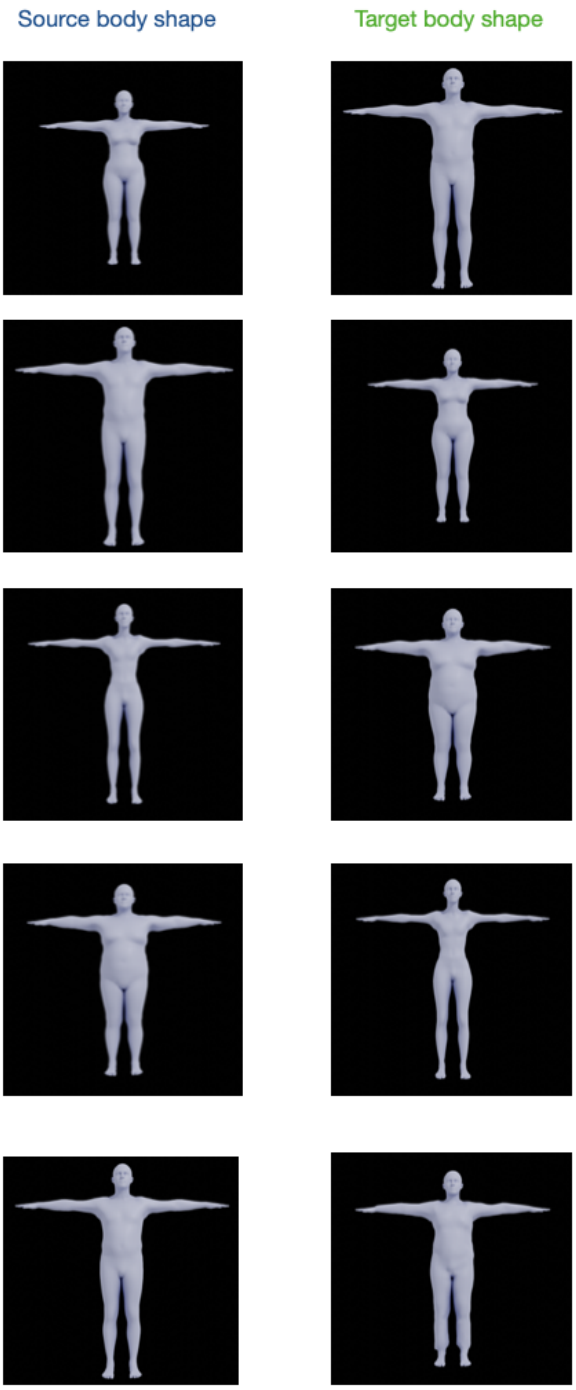

图 8. 用户研究中选取的 5 种形状迁移示意图。左侧为源形状，右侧为目标形状。

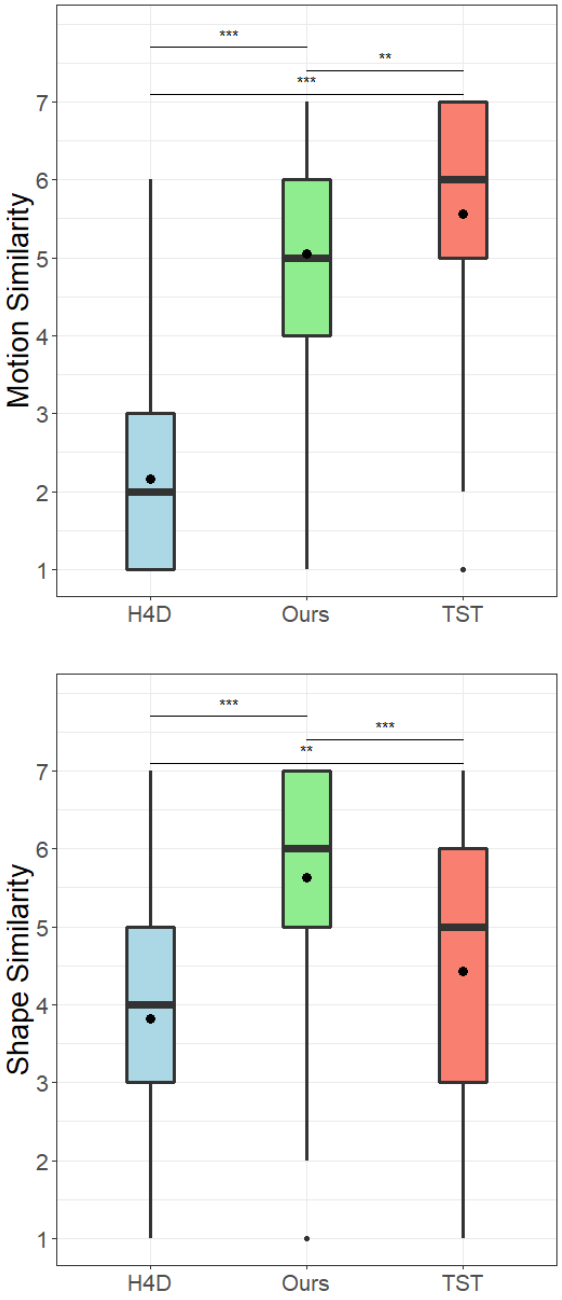

图 9. 用户研究结果。上：箱线图展示了参与者对源动作与重定向结果之间运动相似度的评分。下：箱线图展示了参与者对目标形状与重定向结果之间形状相似度的评分。星号标示事后检验中模型之间的显著差异（$**$：$p < 0.001$，$***$：$p < 0.0001$）。