Mocap Anywhere：迈向野外基于成对距离的动作捕捉（面向野外）

OFIR ABRAMOVICH，Reichman University，以色列；CYENS卓越中心，塞浦路斯
ARIEL SHAMIR，Reichman University，以色列 ANDREAS ARISTIDOU，塞浦路斯大学，塞浦路斯；CYENS卓越中心，塞浦路斯



图 1. 我们提出了一种动作捕捉方法，该方法基于随时间变化的关节间成对距离测量。我们的方法能够在野外重建不同关节化对象的姿态与运动。

arXiv:2601.19519v1 [cs.CV] 27 Jan 2026

我们提出了一种新颖的动作捕捉系统，仅利用身体佩戴式传感器提供的稀疏成对距离（PWD）测量来重建全身三维运动。我们的方法通过无线节点之间的飞行时间测距，无需外部摄像机，因此能够在不受控的户外环境中稳健运行。与传统光学系统或惯性系统不同，我们的方法具有形状不变性，并能抵御光照和磁干扰等环境限制。我们系统的核心是 Wild-Poser（简称 WiP），这是一种紧凑的实时 Transformer 架构，可直接根据含噪或受损的 PWD 测量预测三维关节位置，之后还可通过学习方法重建关节旋转。WiP 能够适应形态各异的研究对象，包括非人类物种，且无需个体身体测量数据或形状拟合。WiP 实时运行，关节位置误差较低，并展示了在野外对人类和动物主体进行精准三维运动重建的能力。我们的实证分析凸显了该方法在真实环境中实现可扩展、低成本、通用型动作捕捉的潜力。

CCS概念： • **计算方法** → **动画**；**动作捕捉**；**运动处理**；**学习范式**。

附加关键词和短语：动作捕捉、成对距离感测、超宽带、3D姿态重建、稀疏传感器

ACM参考格式：
Ofir Abramovich、Ariel Shamir 和 Andreas Aristidou。2025。《随处动作捕捉：迈向野外（为野外）基于成对距离的动作捕捉》。*ACM*图形学汇刊 1，1（2025年1月），14页。https://doi.org/ 0000001.0000001

1 引言

动作捕捉（mocap）是动画与角色控制、生物力学分析及沉浸式体验的核心技术。它将受试者的运动转换为三维位置数据和/或旋转数据，从而实现对运动的高度逼真呈现。

尽管动作捕捉已广泛应用于电影、游戏、体育科学和虚拟制作等领域，但动作捕捉系统仍受成本、可扩展性及环境要求的限制。

目前，最精确的系统是光学系统，它们依靠多台相机对佩戴在受试者身上的标记进行三角测量。虽然这些系统（例如 Vicon、OptiTrack、PhaseSpace）可达到毫米级精度，但价格昂贵，需要进行多相机标定，而且通常仅限于室内、受控摄影棚环境，并要求照明条件固定。为解决便携性和成本方面的限制，基于惯性测量单元的系统（例如 Xsens、Rokoko）逐渐普及。这些系统利用身体佩戴式传感器的局部加速度和角速度测量值来估算姿势，无需外部设备。然而，惯性测量单元容易产生漂移，需要频繁重新标定，并且难以维持长期稳定性，误差会不断累积，全局定位也较为困难。近年来，一种日益明显的趋势是减少佩戴的传感器数量；稀疏惯性方法尝试利用数量有限的惯性测量单元推断全身运动 [Huang 等，2018；Jiang 等，2022b；Ponton 等，2023；Yi 等， 2022, 2021],但这些方法仍依赖相对运动和预先校准的运动学模型，而且通常针对人体形态进行定制。

在这项工作中，我们提出了一种本质上不同的稀疏动作捕捉方法，该方法具有内在性、可扩展性和形状不变性。我们不依赖视觉标记或惯性线索，而是直接根据稀疏布设于身体上的无线传感器之间的成对距离测量（PWD）重建关节运动。借助超宽带（UWB）飞行时间（ToF）测距 [Fontana2004]，我们的方法不受光照条件和环境复杂度的影响，在不受限制的户外环境及非实验室环境中也能可靠运行。该方法安装简便，无需身体测量数据，并可对任何大型关节运动主体进行连续、准确的跟踪。这使得捕捉对象的范围得以拓展，甚至包括动物（见图 1）。

采用超宽带技术进行动作捕捉是一种相对新颖的方法。它的潜力在于高帧率和低延迟，但仍面临若干挑战，例如噪声、非视距

作者联系信息：Ofir Abramovich，Reichman University，以色列赫兹利亚；CYENS卓越中心，塞浦路斯尼科西亚，ofir1080@gmail.com；Ariel Shamir，Reichman University，以色列赫兹利亚，arik@runi.ac.il；Andreas Aristidou，塞浦路斯大学，塞浦路斯尼科西亚；CYENS卓越中心，塞浦路斯尼科西亚，a.aristidou@ ieee.org。

2025。ACM 1557-7368/2025/1-ART
https://doi.org/0000001.0000001

（NLoS）遮挡和信号多径干扰。近期，Ar- mani 等人 [2024] 通过将超宽带与成对距离估计相结合来增强运动跟踪，并采用扩展卡尔曼滤波器（EKF）处理测量噪声和传感器融合，从而探索了这一方向。他们的研究结果表明，超宽带和成对距离为动作捕捉提供了一条颇具前景的路径，尤其是在与惯性系统相结合时。然而，他们的方法仍主要依赖惯性测量单元信号重建动作，并利用基于超宽带的定位来提升全局精度。

为此，我们提出了**WildPoser**（简称 WiP），这是一种实时的基于 Transformer 的架构，仅依赖 PWD 测量即可预测全身三维关节位置和全局运动，从而无需摄像机或预先了解研究对象的模型，就能在户外环境中进行动作捕捉。该模型采用生成式自回归形式，其输出是对输入的精细化变换，而非时间上的延续。它能够为任何关节角色生成末端执行器位置，支持对各种研究对象进行全身关节化重建，并可基于逆运动学（IK）创建动画机制，以推断相应的关节旋转（例如动物）。当已知角色的形态信息时（例如人类），可应用学习方法来估计关节旋转（例如，[Zuo等人2021]）；而当已获得推断出的旋转信息时，则可以直接重建参数化的三维人体模型，例如 SMPL（例如通过 [Ponton 等人 2025]的方法）。通过在不受限制的环境中对任意关节化对象进行无标记动作捕捉，我们的系统朝着在野外捕捉人类乃至动物迈出了重要一步。它能够以 ~50 FPS 的速度生成平滑的高保真运动，关节误差低于 60 毫米，性能媲美光学系统且优于惯性基线方法。我们的贡献包括：

- 一种便携式动作捕捉方法，仅通过成对距离测量来重建三维姿态和全局平移。我们的方法无需摄像头、外部基础设施或预先校准的运动学模型。
- 一种稳健且形状不变的系统，可适用于不同体型和物种，并采用专门设计的去噪机制，以缓解超宽带信号中的噪声和视距问题，从而在充满挑战的环境中实现精准动作捕捉。当骨骼形态先验已知时，可集成学习模型，生成适用于角色动画的精准关节旋转。
- 一种用于动作捕捉的新型精炼生成模型，可对受损测量数据进行处理，并结合过去的预测结果加以精炼。

为支持我们的论点，我们开展了大量实验，包括对人类和动物进行野外录制，以展现我们方法的鲁棒性和实际应用潜力。全面的评估和消融研究验证了我们的假设，并凸显了我们方法的优越性；与以往基于惯性的方法相比，我们的方法性能具有竞争力，甚至更胜一筹。我们的方法在户外和非受限环境中的动作捕捉方面展现出巨大潜力，为该领域的未来研究开辟了新方向。此外，它还能稳健地泛化至不同的身体形态和物种，取得了与当前光学系统相当的结果。

2 相关工作

2.1 动作捕捉系统

广泛应用于娱乐业、医疗保健、体育和机器人学的动作捕捉技术，大致可分为由外向内 和由内向外 系统。在由外向内 系统中，外部跟踪设备从外部观测目标，通常使用布置在捕捉空间周围的多台摄像机。这类系统包括使用回射标记的被动系统（例如 Vicon、OptiTrack），以及采用可唯一识别的发光二极管的主动系统 （例如 PhaseSpace、Qualisys）。这类系统通过多角度三角测量实现毫米级精度，并具有较高的采样率。然而，它们受限于受控的室内环境，对光照条件变化敏感，需要精确校准，且成本高昂。遮挡、标记交换和数据清理仍是难题 [Aristidou 等人，2018年；Peng 等人，2015年；Yu 等人，2007], 不过，近期的深度学习方法提升了数据去噪和重建效果 [Chen 等人，2021；Holden，2018年；Pan 等人， 2024]。

相比之下，由内向外 系统使用穿戴式传感器估计相对于世界坐标系的运动。基于惯性测量单元的系统 （例如 Xsens、Rokoko）通过陀螺仪、加速度计和磁力计测量旋转数据，并结合传感器融合与生物力学建模来推断关节姿态。这些系统便于携带、独立运行、成本低廉，也适用于户外环境。然而，这类系统因针对特定体型而需要根据受试者的身体进行校准，并依赖正向运动学来应用计算出的旋转，因此位置精度较低，方向漂移会随时间累积，且缺少全局参考（除非辅以全球定位系统或外部信标） [Ami-Williams 等 2024]。近期研究致力于使用更少的惯性测量单元（例如 Vive Tracker、Sony Mocapi）重建全身姿态，并采用逆运动学、学习到的运动先验和机器学习模型。稀疏传感器配置通过深度学习 [Huang 等，2018年；Jiang 等，2022b；Winkler 等，2022年；Yi 等 2021], 以及基于物理的方法 [Yi 等 2022]不断改进。将惯性测量单元与单目摄像机和同步定位与建图相结合的混合方案可减轻漂移 [Guzov 等，2021；Yi 等 2023], ；而六自由度虚拟现实设备则借助自编码器 [Ponton 等 2023], 、Transformer模型 [Jiang 等，2022a；Zheng 等 2023], 和生成式方法 [Castillo 等，2023年；Du 等 2023]增强全局姿态估计。尽管取得了进展，这些系统仍面临稀疏输入变化和约束灵活性方面的挑战 [Lee 等 2023]。包括变分自编码器和运动先验在内的潜在空间模型 [Ling 等，2020年；Rempe 等 2021], 能够实现高效合成与强化学习 [Peng 等 2022], , 而直接潜在优化和学习式逆运动学 [Ponton 等 2025] 则提升了帧级精度。然而，精细控制、接触点保真度和实时细节复现方面的问题仍未解决 [Shi 等 2023]。

一种新兴的内外 跟踪类别采用安装在身体或头戴设备（例如 Meta Quest）上的摄像头，捕捉周围环境并推断用户运动和位置 [Hollidt 等，2024年；Jiang 等，2024年；Millerdurai 等，2024年；Yin 等。 2024]一些其他方法还会从可穿戴设备 [推断惯性测量单元数据，Lee 和 Joo 2024] 以提高准确性。虽然这些第一视角 系统提供了更高的

便携性且无需基础设施，但也面临诸多严重限制，包括有限的视场、妨碍全身姿态估计的自遮挡、运动模糊，以及对摄像头安装位置的敏感性。

总而言之，尽管取得了进展，内外 系统仍难以实现全局定位，存在位置漂移，且在没有外部辅助的情况下缺乏可靠的全局参考。

2.2 超宽带测距与位置重建

超宽带（UWB）凭借低功耗、短时脉冲及出色的抗干扰能力，非常适合高精度定位 [Delamare 等，2020年；Siwiak 和 McKeown，2004年；Zafari 等 2019]。它通过到达时间差（TDoA）估算距离 [Gustafsson 和 Gunnarsson 2005]，并已应用于跟踪、机器人学和增强现实领域 [Alarifi 等，2016年；Lindhé 等，2006年；Lymberopoulos 等 2015]。然而，动态环境和非视距条件会引入显著误差 [Mok 等 2012],，通常通过脉冲响应分析 [Barral 等 2019],、基于卡尔曼滤波的传感器融合 [Piltaver，2015年；Tsaregorodtsev 等 2019], 或统计建模 [Hamie 等，2015 年；Kulikov 等 2020] 来解决。

以往研究也探索了在身体上安装超宽带锚点，以实现点对点定位，而无需依赖固定基础设施 [Di Renzo 等人 2007年；Guizar 等人 2016年；Hamie 等人 2013]。为应对信号因身体遮挡而衰减等挑战，研究人员采用了粒子滤波或卡尔曼滤波等概率方法 [Otim 等人 2020年；Tian 等人 2019]。此外，研究表明，将超宽带测量与惯性测量单元结合可提升姿态估计效果 [Corrales 等人 2008年；Zihajehzadeh 和 Park2017],；近来，超宽带也被整合进基于惯性测量单元的动作捕捉系统，以提高其稳定性 [Armani 等人 2024]。

我们的方法弥补了以往研究的一项关键不足：缺少一种便携式全身动作捕捉系统，能够利用超宽带技术和机器学习领域的最新进展进行姿态重建与去噪。与我们的方法最相关的系统是 Ultra Inertial Poser（UIP） [Armani 等人 2024] 以及 UMotion [Liu 等人 2025],，它们利用惯性测量单元测量数据和超宽带数据构建成对距离矩阵。UIP 使用惯性测量单元估计三维姿态和全局平移，仅将超宽带成对距离用于稳定和优化，而去噪则通过基于扩展卡尔曼滤波的融合来完成。UMotion 是一种状态估计框架，采用无迹卡尔曼滤波和长短期记忆网络进行姿态及逐关节估计，但不估计全局平移。相比之下，我们的方法仅依赖传感器间成对距离，利用 Transformer 模型推断传感器位置，并重建全身姿态和全局平移。我们提出了一种新颖的去噪模型，以应对噪声和非视距挑战，并可选择融合惯性测量单元数据，进一步提高精度。与传统的基于惯性测量单元的系统不同，我们的方法无需基础设施、具有形状不变性，且适用于不同体型和物种，提供了一种便捷、便携的动作捕捉系统。

3 方法

3.1 预备知识

稀疏动作捕捉的常见做法是使用最多六个放置在关键点关节上的惯性测量单元，以估计姿态和全局平移 [Ponton 等，2023年；Yi 等，2022]。这些方法通常基于 SMPL [Loper 等， 2015]生成关节层级结构。然而，我们采用了不同的方法来处理这一任务。我们不再根据惯性数据估算部分姿势，而是提出 *WiP*，一种基于 Transformer 的架构，可根据六个超宽带传感器测得的稀疏含噪成对距离重建三维姿态。本节介绍我们的模型组，随后介绍架构、训练流程和损失函数。

3.2 问题形式化

定义*3.1*（精炼生成模型）。A精炼生成模型$\mathcal{M}_\theta$定义为一个映射：

$$\mathcal{M}_\theta : \mathcal{X} \times \tilde{\mathcal{Y}} \to \mathcal{Y}, \tag{1}$$

其中：X 表示输入特征（例如序列模型的过去预测），Y 表示目标输出的含噪或受损版本（例如辅助信号或中间信号），而 Y 是待重建的干净目标输出。

该模型被称为精炼生成式，因为它利用$\tilde{\mathrm{Y}}$（真实目标 Y 的退化形式）作为辅助信号。从概念上讲，M_θ可被视为一个生成式噪声滤波器，其中先前上下文与噪声引导信号共同用于生成 Y 的去噪或校正版本，并以先验信息为条件。

定义 *3.2*（动作捕捉细化生成模型）。令$P^{[t]} \in \mathbb{R}^{J\times 3}$表示由 J个关节组成的有序三维点集，其时间戳为 t，代表骨架姿态的空间坐标。此外，令 $D^{[t]} \in \mathbb{R}^{J\times J}$表示在相同时间戳下预先计算的成对欧几里得距离矩阵，定义如下：

$$D_{ij}^{[t]} = \|P_i^{[t]} - P_j^{[t]}\|_2, \tag{2}$$

对于所有 i、$j \in \{1、2、\ldots\ldots、J\}$。

动作捕捉细化生成模型 是细化生成模型 M_θ 的一个特定实例，其输入为：

- A temporal window w of prior poses $\mathcal{X} = \left\{P^{[t-w]}, \ldots, P^{[t-1]}\right\}$,
- 当前的含噪观测 $\tilde{\mathrm{Y}} = D^{[t]}$,

并输出当前三维姿态的估计值：

$$\mathcal{M}_\theta(\mathcal{X}, \tilde{\mathcal{Y}}) = \hat{\mathcal{Y}} = \hat{P}^{[t]} \in \mathbb{R}^{J\times 3}. \tag{3}$$

其主要目标是基于成对距离矩阵 $\tilde{\mathrm{Y}} = D^{[t]}$以及先前姿态序列 X，重建时间戳$t$处的姿态 $\hat{P}^{[t]}$。在此表述中：

- $P_j^{[t]}$ denotes the 3D coordinates of joint j at time t,
- $D_{ij}^{[t]}$ encodes the Euclidean distance between joints i and j.

3.3 WiP 架构

WiP 是一种动作捕捉精炼生成模型（记作 RG_θ），结合了编码器-解码器 [Raffel 等人，2020；Vaswani 等人 2017] 和仅解码器 Transformer [Rad- ford 等人，2019；Touvron 等人 2023]的理念。仅解码器模型由 n 个堆叠模块组成，每个模块包含一个掩码多头自注意力操作，后接一个前馈网络，并辅以残差连接和层归一化。与传统方法不同，我们为每个模块加入交叉注意力机制，其中，键 和查询 由成对距离矩阵 $D^{[t]}$得出，而值 则取自过去姿态序列 P_{prev}。这种交互使模型能够动态平衡时间一致性和含噪输入信号，并通过

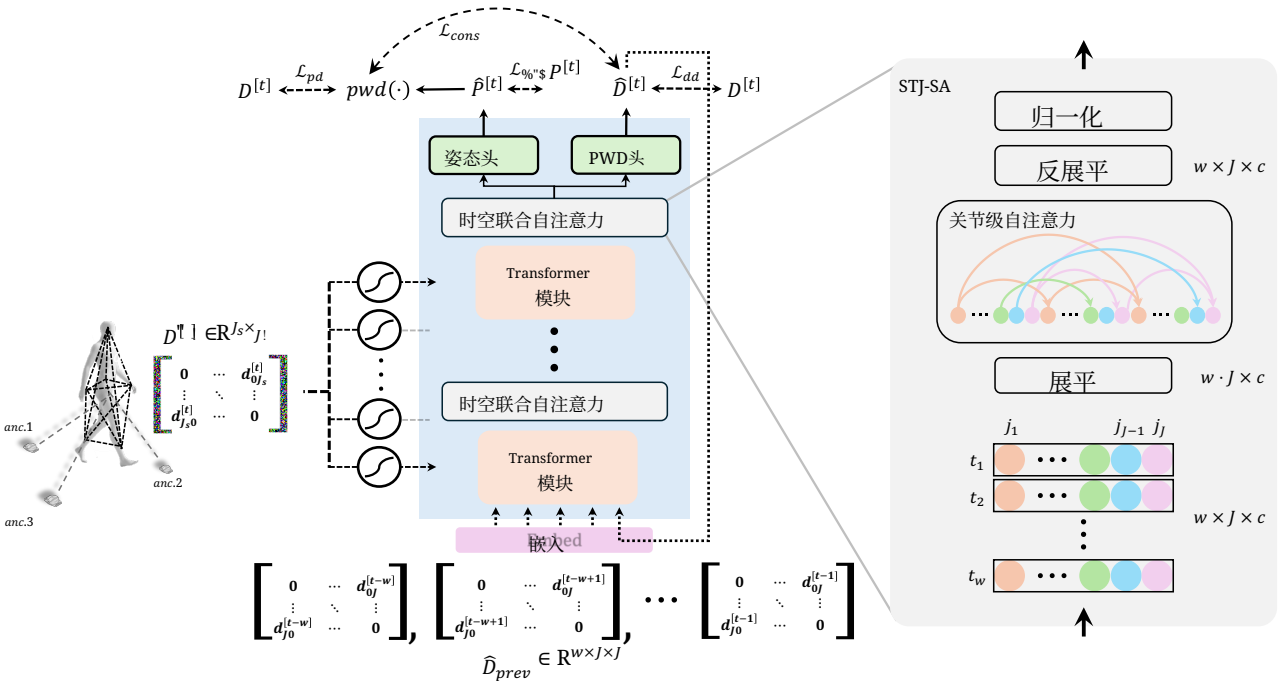

图 2。左：WiP 是一种动作捕捉细化生成模型（见定理 3.2）。给定截至时间戳 $t-1$ 的一段干净成对距离序列 $\hat{}D_{prev} \in \mathbb{R}^{w \times J \times J}$，以及一条稀疏含噪测量 $\tilde{D}^{[t]} \in \mathbb{R}^{J_s \times J_s}$，模型将预测全局坐标中的下一个骨架 $\hat{}P^{[t]} \in \mathbb{R}^{J \times 3}$。与此同时，模型还会预测时间戳 t 对应的去噪成对距离矩阵 $\hat{}D^{[t]} \in \mathbb{R}^{J \times J}$，该矩阵随后将用作预测姿态 $t+1$ 的输入。为确保两种输出模态之间的一致性，我们在分别针对每种模态的重建损失 L^{pd} 和 L_{dd} 之外，还采用一致性损失 L_{cons}。右：插入每个 Transformer 块末尾的 STJ-SA 层，使模型能够学习跨时间的关节对应关系，对齐逐关节轨迹，并减轻含噪测量造成的抖动。对于稀疏配置，我们使用 $J_s = 6$；对于完整的 SMPL 骨架表示，我们使用 $J = 24$。

两个训练阶段实现。图 2 展示了 WiP 架构的整体示意图。

3.3.1 成对距离空间中的生成。我们发现，以 PWD 表示来预测下一姿态更有效。因此，WiP 经过训练，可在成对距离空间中生成运动。具体而言：

$$h^{[t]} = \mathcal{RG}_\theta \left(\hat{D}_{\text{prev}}, \tilde{D}^{[t]} \right) \tag{4}$$

其中，$\hat{}D_{\text{prev}} \in \mathbb{R}^{w \times J \times J}$表示截至时间戳 $t-1$之前预测的 w，$\tilde{D}^{[t]} \in \mathbb{R}^{J \times J}$ 是当前含噪输入。随后，将隐藏状态 $h^{[t]} \in \mathbb{R}^d$ 传入姿态与 PWD 头，分别得到下一姿态$\hat{}$

$P^{[t]} \in \mathbb{R}^{J \times 3}$ $D^{[t]} \in$和 PWD $\hat{}$。

$\mathbb{R}^{J \times J}$

$P^{[t+1]}$ $D^{[t]}$

接下来，为了预测$\hat{}$，将$\hat{}$嵌入并与先前的预测拼接。

3.3.2 刚体生成。虽然 WiP 并非只能用于人体运动，但大多数实际应用都涉及人体捕捉。为此，我们引入一种扩展变体，称为*WiP-H*，并对其进行训练，以重建 SMPL 身体模型的稠密骨骼结构（其中 $J = 24$）。我们分别用 J_s 和 J 表示稀疏关节集和稠密关节集的大小。

给定稀疏输入 $\tilde{D}^{[t]} \in \mathbb{R}^{J_s \times J_s}$，模型会预测稠密点集 $\hat{}P^{[t]} \in \mathbb{R}^{J \times 3}$。为了鼓励模型学习刚体约束和关系，我们引入了刚性损失项（见第 3.4.5 节）。由于先前的研究侧重于人体运动，我们采用这一变体进行所有比较（与先前工作一致，在稀疏设置下我们设定 $J_s = 6$）。

3.3.3 全局位移与旋转。我们的动作捕捉系统无需身体校准或预定义的初始姿势。虽然我们的姿态重建完全依赖身体佩戴式传感器之间的成对距离测量，但我们额外引入三个固定的参考传感器（锚点），并将其置于地面上以构成正交二维基（见图 3）。这些锚点用作空间参考，使模型能够通过监督学习全局方向和位置。为简化设置并提高灵活性，我们不使用第三个轴。这些锚点被整合到分布于受试者身体上的去中心化传感器网络中。为消除 XY 平面中的反射歧义，我们在损失函数中引入重力项，约束预测骨架保持在地表上方。通过纳入参考传感器位置，模型能够在各帧中稳定地恢复全局位移和方向。更多详情见第 3.4.2 和 3.4.3 节。需要强调的是，该组件独立运行，不参与姿态重建评估。类似方法通过使用 Vive追踪器和基站 [Ponton 等人 2023] 或采用近似策略来应对这一挑战。

3.3.4 门控交叉注意力（门控 *CA*）。在标准的编码器-解码器 Transformer架构中 [Raffel 等人，2020；Vaswani 等人， 2017],

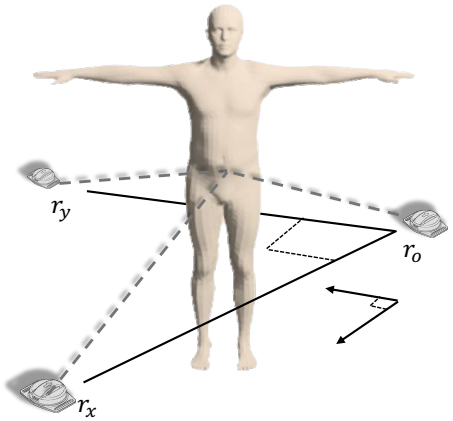

图 3. **参考锚点布置**。为解决记录的研究对象的全局方向问题，我们通过向大小为 $|J_s|$ 的集合添加三个传感器，模拟一个二维全局坐标系。将它们映射到固定的空间位置（$r_o \to (0,\ 0,\ 0)$ $r_x \to (1,\ 0,\ 0)$, $r_y \to (0,\ 1,\ 0)$），并优化公式（7）、（8）和（11）后，WiP 可预测精确的全局位置。

交叉注意力允许上下文引导预测。在我们的细化生成设置中，我们必须平衡对过去姿态预测和当前含噪输入测量的注意力。为此，受 [Alayrac 等人 2022], 的启发，我们用可学习的门控交叉注意力取代传统的交叉注意力。对于一层 l，我们应用：

$$\begin{aligned} A_{\text{ca}} &= \text{CA}_l\left(\text{Emb}\left(\tilde{D}^{[t]}\right), h_{l-1}^{[t]}\right), \\ A_{\text{gate}} &= A_{\text{ca}} \odot \sigma\left(\text{W}_g h_{l-1}^{[t]}\right), \\ h_l^{[t]} &= A_{\text{gate}} + \text{dropout}\left(h_{l-1}^{[t]}\right), \end{aligned} \tag{5}$$

其中，CA_l 是第 l 层中的交叉注意力，$h_{l-1}^{[t]}$ 是前一层 $l-1$ 的输出，而 Emb (·) 是一种嵌入操作。W_g 是一个可学习的门控矩阵，σ 是 Sigmoid函数，$\odot$ 表示 Hadamard乘积运算。鉴于输入测量值含噪，门控机制使模型能够控制信息流，并在每个时间戳动态地关注这些测量值。

3.3.5 WiP-SI（形状不变）。由于 WiP 主要基于人体运动数据进行训练，因此它会隐式地学习关于人体关节构型的强先验。为了提升泛化能力，我们引入了 *WiP- SI*，这是一个仅针对稀疏关节进行重建训练的变体（其中 $J = J_s$）。为了进一步支持对非人类形态（例如四足动物）的泛化，我们在每个训练步骤中应用随机关节置换。设 $\pi\colon 1,\ldots,J \to 1,\ldots,J$ 为采样得到的置换，$D^{[t]}$, 为置换后的矩阵：

$$D_{ij}^{[t]\prime} = D_{ji}^{[t]\prime} = D_{\pi(i)\pi(j)}^{[t]}. \tag{6}$$

这种置换起到了正则化器的作用，在保留全局空间关系的同时，降低模型对特定关节身份的过度依赖。我们将这一变体称为 WiP-SI，因为它无需预先了解研究对象的形状，即可支持任意关节集合。在实验中，我们仅将这一变体用于非人类研究对象。

*3.3.6 PWD*头。WiP 旨在直接预测全局坐标空间中的关节位置。然而，专注于全局预测可能会导致训练过程中丢失重要的局部结构信息。为此，我们增加了一个辅助的 *PWD*头，以促进对骨骼几何结构的保留。该头会预测输出骨架中所有关节（包括锚点）之间的成对距离（PWD），帮助模型维持局部一致性，更好地学习关节之间的空间关系，从而补充全局位置预测。随后，将预测的成对距离（PWD）$\hat{}D^{[t]}$ 与$\hat{}$ D_{prev}拼接，以预测下一姿态（见图 2 左侧）。

3.3.7 时空关节联合自注意力（*STJ-SA*）。从干净的成对距离转向含噪成对距离并非易事，而且每个训练阶段都会带来不同的挑战。为了保持动作平滑且准确，模型必须着重关注关节层面的时间连贯性。为此，我们在每个 Transformer 块的末尾加入一个 STJ-SA 层。该层的核心思想是使模型能够学习跨时间的关节间关系。具体做法是，在执行 SA 操作前，将隐藏状态 $h_l^{[t]} \in \mathbb{R}^{T\times J.c}$展平为形状 $T\cdot\ J \times c$，随后进行后归一化并还原展平状态（图 2 右侧）。这一改动能够在保持原始骨骼结构一致性的同时，对各关节进行动态平滑，从而生成时间上稳定且符合解剖学规律的重建结果。

3.4 损失

为训练模型，我们定义了一个由多个损失项组成的统一目标函数。每个损失项都致力于实现一种平衡，在保持时间连贯性、解剖学合理性以及对噪声的鲁棒性的同时，尽可能降低 PWD 空间中的重建误差。在本节中，我们用 MSE 表示均方误差。

3.4.1 PWD 损失。模型的一个关键目标是保留骨骼结构，为此我们定义以下损失项：

$$\mathcal{L}_{\text{pd}} = \text{MSE}\left(\text{pwd}\left(\hat{P}^{[t]}\right), D^{[t]}\right), \tag{7}$$

$$\mathcal{L}_{\text{dd}} = \text{MSE}\left(\hat{D}^{[t]}, D^{[t]}\right), \tag{8}$$

$$\mathcal{L}_{\text{cons}} = \text{MSE}\left(\text{pwd}\left(\hat{P}^{[t]}\right), \hat{D}^{[t]}\right). \tag{9}$$

其中，pwd(·) 用于计算任意 n 维空间中一个点集的完整成对欧几里得距离矩阵。L^{pd}（式（7））用于惩罚预测三维关节的骨骼距离误差，而 L_{dd}（式（8））则直接监督预测的成对距离。一致性损失 L_{cons}（式（9））用于对齐两个输出头，确保预测姿势与其距离表示之间的内部一致性。这些损失共同确保了解剖结构的一致性和几何保真度。

3.4.2 方向损失。如第 3.3.3 节所述，我们使用地面上的三个固定参考传感器（记为 $r_{\{1,2,3\}}$）来定义全局方向基。通过将它们映射到固定的 3D 位置，我们在训练过程中隐式地锚定模型的全局旋转。此外，我们通过以下方式监督它们的预测位置：

$$\mathcal{L}_{\text{refs}} = \sum_t \left\| r_i^{[t]} - \hat{r}_i^{[t]} \right\|_2, \tag{10}$$

其中，$r_i^{[t]}$和$\hat{r}_i^{[t]}$分别表示参考传感器 i在时间 t的真实位置和预测位置。

3.4.3 重力正则化。 虽然方向锚点提供了全局旋转，但无法消除关于 XY 平面的反射歧义。为此，我们引入一个重力项，约束关节位置位于地面之上：

$$\mathcal{R}_{\text{gravity}} = \mathbb{E}_{\hat{P}[t]}\left[\max\left(0, -\hat{p}_{iz}^{[t]}\right)\right], \tag{11}$$

其中，$\hat{}p_{iz}^{[t]}$是预测姿态$\hat{}P^{[t]}$中关节 i的 z坐标。该正则化器会惩罚任何被预测到地面以下的关节，从而确保相对于重力而言具有正确的全局方向。

3.4.4 速度损失。 受先前工作 [Tevet 等人 2023], 的启发，我们对关节速度施加额外监督，以确保连续时间戳之间的转换平滑且可靠：

$$\mathcal{L}_{\text{velo}} = \sum_t \sum_j \left\| \left(\hat{P}_j^{[t]} - \hat{P}_j^{[t-1]}\right) - \left(P_j^{[t]} - P_j^{[t-1]}\right) \right\|_2. \tag{12}$$

该损失同时跟踪速度大小和方向速度，以避免抖动。此损失在第二训练阶段加入。

3.4.5 刚性损失。 在训练模型预测完整 SMPL 骨架（24 个关节）时，我们引入一项刚性约束，以确保相连关节之间的骨骼长度保持恒定。这通过以下方式维护关节式人体结构：

$$\mathcal{L}_{\text{rigidity}} = \sum_{(i,j)\in\mathcal{B}} \left\| b_{ij} - \hat{b}_{ij} \right\|_2, \tag{13}$$

其中，B 是 SMPL 模型中刚性骨骼对的集合。对于每根骨骼 (i, j) ∈B，其真实骨骼长度和预测骨骼长度分别可轻易地表示为 $b_{ij} = D_{ij}^{[t]}$ 和 $\hat{}\, b_{ij} = \hat{}D_{ij}^{[t]}$。

为此，我们训练过程的总体目标为：

$$\underset{\theta\in WiP}{\operatorname{argmin}} \sum_i \lambda_i \mathcal{L}_i, \tag{14}$$

其中，λ_i是为每个损失项预先定义的权重 i。

3.5 噪声处理

UWB 信号具有较高的时间分辨率和抗干扰能力，因此非常适用于定位任务。然而，它们仍不可避免地容易受到噪声、遮挡和非视距条件的影响，这些因素可能导致严重的运动重建误差（制造商报告的漂移最高可达 15 cm）。为应对这些挑战，我们引入了一个称为去噪阶段的次级训练阶段。在此阶段，模型在加入模拟噪声的原始数据集上进行微调，这些噪声反映了真实世界中的测量失真。我们定义了一个用于模拟原始噪声、同时保持简单通用的噪声函数：

$$\phi\left(D^{[t]}\right) = \frac{1}{w} \sum_{i=t-\lfloor w/2\rfloor}^{t+\lfloor w/2\rfloor} \left(D^{[i]} + \epsilon_i\right), \quad \epsilon_i \sim \mathcal{N}(0,\sigma^2) \in \mathbb{R}^{J_s\times J_s}. \tag{15}$$

我们发现，设置 $\sigma = 15$ cm 和 $w = 5$ 可以较好地近似测得的噪声。这种数据增强策略在提升泛化能力的同时，避免了对特定采集实例过拟合。

4 实验

4.1 实验设置

4.1.1 实现细节。 我们使用 PyTorch 中的 DDP 训练，在 4 块 NVIDIA GeForce RTX 3090 上对模型进行 20 个训练周期的训练（两个训练阶段均如此）。对于基于超宽带的成对距离记录，我们使用 DWTAG100传感器 [Ciholas 2025b], , 并将最高频率设为 20 Hz。更多详情请参见 C 节。

4.1.2 训练流程。 为了训练 WiP 处理 PWD 的原始测量数据，我们将训练分为两个阶段：i）距离到运动阶段，在此阶段，模型接触干净的距离数据，并接受训练以生成连续且准确的运动；ii）去噪阶段，在此阶段，我们在距离空间中合成添加噪声（公式（15）），以破坏输入数据。在这一阶段，模型接受训练以处理噪声输入，同时仍生成合理的动作序列。在第二阶段，我们冻结模型 95% 的参数，仅训练门控注意力和新添加的 STJ-SA 层。采用这种方法的核心动机是，在提升模型对退化和含噪输入数据的鲁棒性的同时，保持其运动重建精度。

4.1.3 数据预处理。 在训练中，我们采用公开可用的动作数据集 AMASS [Mahmood 等人 2019]。不同于以往方法，我们训练 WiP 重建一个预定义的三维点集，该点集由全局世界坐标系中的一组关节构成。因此，我们首先将 SMPL 表示转换为由 24 个关节组成的集合，并创建动作序列。在所有实验中，我们将窗口长度设为 16（~250 毫秒）。随后，我们将所有会话中的头部与骨盆间距归一化为 1，并将地面与 XY 平面对齐。

4.1.4 评估数据采集。为了采集用于评估系统的数据，我们在一名人体受试者身上安装了六个超宽带传感器，分别位于骨盆、头部、手部和脚部。不同于以往将传感器放置在膝盖上的方法 [Armani 等人，2024年；Yi 等人 2022], ，我们将传感器安装在脚部，与 Ponton 等人 [2023]的方法类似。脚部是自然的末端执行器，也是与环境接触的主要部位，因而能够实现更精准的末端执行器定位，提供更长的传感器间距，并减少非视距问题。这种配置顺应了动作捕捉领域采用稀疏式身体传感器的趋势。尽管增加传感器可能提升精度，但传感器间距缩短可能会因超宽带技术在近距离下的局限而加剧噪声敏感性。为获得精确的真实运动数据，我们使用同步光学动作捕捉系统（PhaseSpace Impulse X2E）同时记录超宽带传感器的三维位置，从而确保配对数据集精准对齐。我们总共采集了时长为 ~20 分钟的 10 种不同动作。

4.2 结果

我们在多种场景中测试了系统，包括根据稀疏的 PWD 测量进行全身人体姿态重建、采用 PWD+惯性测量单元混合输入，以及在非受控环境中进行非人体动作捕捉。结果凸显了我们的方法在不同输入模态和研究对象上的多功能性与鲁棒性。动画示例请参阅补充视频。

4.2.1 全身关节位置重建。我们的方法仅使用来自 $J_s = 6$ 身体佩戴式超宽带传感器的 PWD 测量，即可重建完整的 $J = 24$ 关节位置。尽管空间覆盖范围有限，但该方法利用学习先验有效捕捉底层运动动力学，从而生成符合解剖学一致性的全身关节位置。图 4 展示了与真实值对比的代表性重建结果。

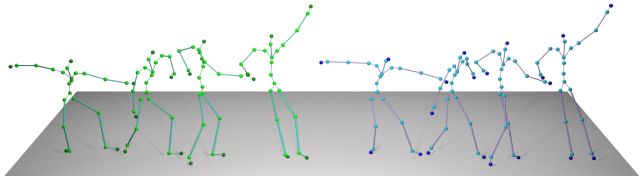

图 4. **全身重建**（来自 DanceDB）。我们的方法（蓝色）仅依靠 PWD 测量，即可重建随时间变化且与真实值（绿色）一致的全身关节位置。

4.2.2 用于角色动画的 *WiP*。 WiP 的设计目标是预测三维空间中的关键点位置。尽管这种表示方式有效，但在捕捉局部扭转关节运动方面能力有限。为解决这一局限，我们扩展了处理流程，通过从三维关节估计 SMPL 姿态参数，以支持全身虚拟角色动画。具体而言，我们采用了 Zuo等人 [2021]（第 3.1 节）使用的初始拟合方法 *SMPLify3D*。该方法通过针对学习到的姿态先验优化基于旋转的层级表示，从三维关节位置估计 SMPL 姿态参数。通过整合上述方法，我们能够恢复关节旋转和全局运动（见图 5）。

4.2.3 基于成对距离的*SMPL*姿态估计 + 旋转测量值。 在可以获取旋转数据的场景中，我们扩展了处理流程，直接预测 SMPL 模型的全身姿态参数。通过将 DragPoser 与基于 Transformer的模型相结合，我们能够准确恢复关节旋转和全局运动。图6展示了我们的重建结果，与真实姿势高度吻合。

4.2.4 野外捕捉。 为展示我们的方法在不同物种间的通用性，我们在一头骆驼和一头驴身上安装了多个超宽带传感器（见图 7）。由于用于训练学习型姿态模型的数据有限，我们采用了 *WiP-SI* 变体来重建传感器的全局三维位置。为了制作动画，我们构建了带约束的绑定骨骼三维模型，并应用逆运动学，将重建位置视为驱动四肢和躯干运动的末端执行器。为提高运动的逼真度，我们对动物的周期性步态模式进行了分段，并计算各步态周期的时间平均值，以减少原始数据中的噪声、异常和离群值。此外，我们施加了足部与地面的接触约束，以增强运动过程中的物理合理性和稳定性。头部、尾巴和耳朵等次要身体部位保持静止，因为这些区域没有安装传感器。图1展示了动画序列中的部分画面，完整视频收录于补充材料中。

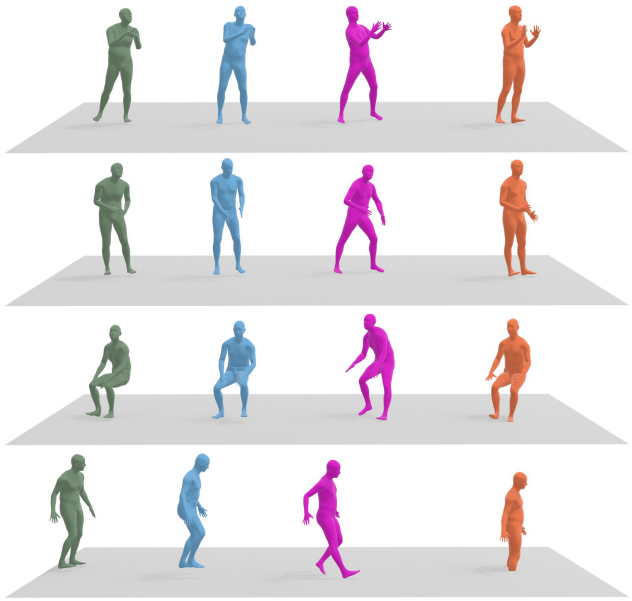

图 5. **UIP-DB 上的定性比较**。 通过集成 SMPLify3D，我们取得了出色的性能，即使与直接预测关节旋转的方法相比也是如此。 蓝色 表示我们的方法，洋红色 表示 UIP，橙色 表示 UMotion，而真实值以 绿色表示。

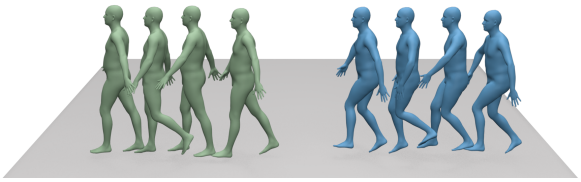

图 6. **SMPL 姿态估计（来自 TotalCapture）**。当成对距离和惯性数据均可用时，我们的方法（蓝色）生成的 SMPL 姿态与真实值（绿色）高度吻合。

5 评估与讨论

本节将分析我们的方法与现有基线方法及最先进方法相比的性能。我们还进行了消融研究，以评估 WiP 架构组件的作用，并展示各个模块如何提升重建运动的鲁棒性。更多实现细节以及扩展的定性和定量结果，请参阅附录 A 和 B。

5.1 实验设置

5.1.1 指标。我们使用一组特定于任务的指标（以厘米为单位）评估了我们方法的性能，包括：

• 位置误差（*PE*）。衡量所有帧中预测的关节位置与真实值之间的平均欧几里得距离，以反映整体重建精度。• 末端执行器位置误差（*EEE*）。专门计算末端执行器的平均位置误差。

表 1。**与以往稀疏传感器动作捕捉方法的比较**。在合成生成的数据上进行评估时，WiP 的表现优于以往方法，尽管后者使用了多种测量模态，包括**O**方向、**A**加速度、角**V**速度和**D**距离。值得注意的是，我们的方法在评估时不需要人体专属校准或初始姿态。粗体表示每项指标中的最佳性能，下划线表示次佳性能。

方法	输入测量 Type	人体校准 Free	TotalCapture					DIP-IMU				DanceDB				
			PE	EEE	GTE	AJE	GSE	PE	EEE	AJE	GSE	PE	EEE	GTE	AJE	GSE
TransPose	O+A+V	✗	5.57	9.07	8.99	0.58	2.35	7.62	7.43	1.60	2.06	8.50	11.41	6.68	1.35	3.61
TIP	O+A+V	✗	5.14	-	-	0.35	-	5.71	7.63	1.31	2.22	8.46	11.66	12.48	2.61	3.52
PIP	O+A+V	✗	5.61	9.17	8.55	0.25	2.39	5.04	7.67	0.40	2.06	5.69	8.55	9.44	1.81	2.50
UIP	O+A+V+D	✗	5.49	10.25	24.70	0.32	**1.52**	5.05	5.59	0.46	1.58	7.61	13.19	20.94	1.05	2.53
UMotion	O+A+V+D	✗	4.46	6.74	-	0.20	1.72	**2.81**	5.15	0.39	**1.31**			-		
WiP-Geo	D	✓	5.61	6.30	4.82	0.12	3.04	3.50	2.71	0.24	1.84	4.95	4.30	2.70	0.64	2.61
WiP-knees	D	✓	**3.07**	6.91	4.98	0.10	2.11	3.40	5.86	**0.23**	2.24	4.79	8.43	**2.28**	0.61	3.21
WiP（我们的方法）	D	✓	4.17	**4.18**	**1.66**	**0.09**	2.71	3.80	**2.14**	0.24	1.73	**4.70**	**3.32**	2.59	**0.58**	**2.50**



图7. **野生动物动作捕捉**。 我们的方法适用于任何类型的运动主体和环境。超宽带传感器可以轻松安装在不同的动物（例如骆驼、驴）身上；得益于我们的免身体校准方法，我们能够捕捉它们的动作。

• 全局平移误差（*GTE*）。衡量预测骨架与真实值骨架之间的全局根节点平移误差。• 绝对抖动误差（*AJE*）。通过比较预测动作的平均关节加加速度与评估数据集的固有抖动之间的绝对差值，反映时间不稳定性。AJE 越低，表示越接近真实动作。单位为 km/s^2。• 几何结构误差（*GSE*）。通过比较所有关节之间的平均成对距离与真实值，衡量姿势合理性。该指标量化预测骨架的关节关系与真实值的吻合程度。• *SIP*误差。量化全局空间中上臂和腿部的平均方向误差。单位为度。• 网格误差。衡量预测网格顶点与真实值网格顶点之间的平均欧几里得距离。

5.1.2 基线方法。 为了进行更全面、深入的比较，我们设计了以下三种基线方法：

多维尺度分析 + Procrustes。 多维尺度分析（MDS）算法是一种经典技术，旨在将高维数据投影到低维空间，同时保留成对距离。Procrustes [Goodall 1991] 旨在通过对一个形状进行平移、旋转和缩放，使其与另一个形状最佳匹配，从而对齐这两个形状，并最小化对应点之间的距离。首先，我们对输入 $\tilde{D}^{[t+1]}$ 运行 MDS，以近似 $\hat{}P^{[t+1]}$。为对齐方向，我们然后在 $\hat{}P^{[t]}$ 和 $\hat{}P^{[t+1]}$ 之间应用 Procrustes。

WiP-Geo。 该基线不在成对距离空间中进行生成，而是直接在三维几何空间中生成。实际应用中，预测结果 $\hat{}P^{[t]}$ 会直接嵌入，并输入 Transformer 以预测下一个姿态。

WiP-feet。 由于以往基于惯性传感器的稀疏动作捕捉工作通常使用膝盖而非脚 [（Armani 等人，2024；Rempe 等，2021；Yi 等人， 2022], ），我们训练了一个额外变体以匹配这一确切设置，称为 *WiP-feet*。不过，我们认为我们采用的设置更稳健，尤其能提升脚与地面接触的准确性（见表4）。

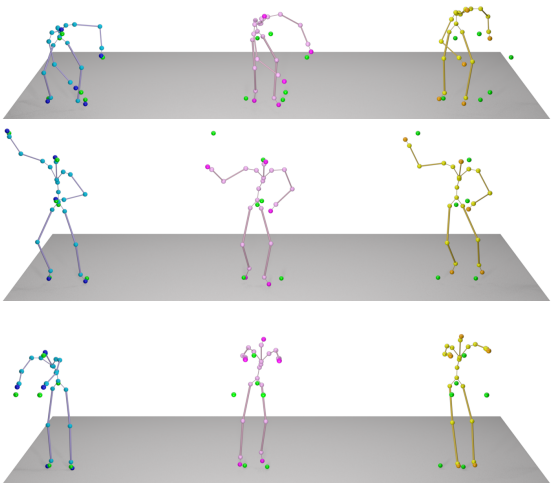

图8。 **末端执行器定位的定性比较**（基于 DanceDB数据集）。请注意，与 UIP（蓝色）和 PIP（洋红色）相比，我们的方法（黄色）生成的姿势中，末端执行器（较深色调）与原始传感器目标位置（绿色球体）的对齐更为准确。

5.2 与先前方法的比较

我们对 WiP 的性能进行了实证分析，并将其与以往工作进行比较：UIP [Armani 等人 2024], PIP [Yi

等人 2022], TIP [Jiang 等，2022b], TransPose [Yi 等人 2021], [Liu 等人 2025]，以及我们提出的基线方法。表1中，我们使用从三个动作捕捉数据集合成的 PWD 对方法进行评估：TotalCapture [Trumble 等人 2017], DIP-IMU [Huang 等人 2018], 和 DanceDB [Aristidou 等人 2019],，这些数据集均未纳入我们的 AMASS [Mahmood 等人 2019] 训练集。值得注意的是，仅使用 PWD，我们的方法便足以达到最先进的性能，优于现有的基于惯性的方法。与以往根据单独测量值估计人体姿态的工作不同，我们的方法构建了一个去中心化网络，能够在每个时间戳同时重建几何结构和全局位置。评估时无需进行人体专属校准，也无需提供初始姿态。如图8所示，WiP 显著提高了末端执行器的位置精度，误差最多降低70%。这种设计大幅减轻了累积漂移，这是以往方法的常见局限（图9）。WiP 还将 GTE 降低了 ~66%（表1）；图10展示了其在全局位置跟踪方面的卓越精度。

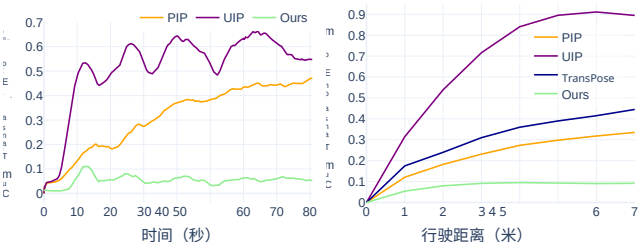

图 9. **累积漂移误差分析**。与基于惯性的方法不同，我们的方法能够有效抵御累积误差，无论是在时间维度（左）还是距离维度（右）。通过利用距离测量并将额外的参考传感器作为锚点，我们可以在每个时间戳估计受试者的全局位置，而几乎不依赖先前的预测。

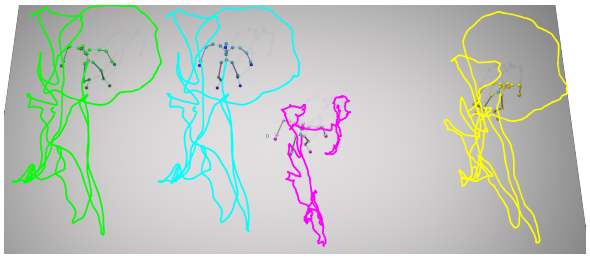

图 10. **全局根位置重建精度对比**（基于 TotalCapture 数据集）。与 UIP（蓝色）和 PIP（洋红色）相比，我们的方法（蓝色）重建的全局根位置更接近原始位置（黄色）。

5.3 原始超宽带测量评估 s

表 2 展示了我们在真实世界超宽带数据上的结果。我们的评估涵盖两个数据集：MA-DB（我们的评估集）和由 UIP 发布的 UIP-DB。为弥合这两个数据集采集条件的巨大差异，我们过滤掉无关指标（MA-DB 中的位置误差，因为该数据集仅使用六个传感器采集；UIP-DB 中的 EEE，因为该数据集

表 2. **真实超宽带数据的定量结果**。我们在两个基准数据库上评估并比较模型：1）MA-DB，即我们的评估数据集；2）UIP-DB，即 UIP 发布数据集的测试集。为公平比较 UIP-DB 上的结果，我们报告了 WiP-knees 变体得出的数值。

方法	MA-DB（脚）			UIP-DB（膝盖）		
	EEE	GTE	AJE	PE	GTE	AJE
UIP		-		10.55	14.19	0.24
UMotion 多维尺度分析 + 过程				10.33	-	0.33
	60.48	57.96	**0.21**	-	-	-
WiP-Geo	28.12	34.55	0.34	7.43	3.52	0.22
WiP	**22.80**	**34.36**	0.35	**6.88**	**6.84**	**0.22**

表 3. **含噪 TotalCapture 和 UIP-DB 上与旋转和网格相关的评估**。我们使用 SMPLify3D，根据 WiP 预测的全局关节位置估计关节局部旋转。因此，WiP 的表现与以往方法相比具有竞争力。

方法	输入测量值 Type	TotalCapture		UIP-DB（测试集）	
		SIP 误差	网格误差	SIP 误差	网格误差
TransPose	O+A+V	16.58	-	-	-
TIP	O+A+V	13.22	7.42	33.01	-
PIP	O+A+V	15.93	6.80	30.47	14.57
UIP	O+A+V+D	11.32	5.09	23.85	14.32
UMotion	O+A+V+D	**10.76**	**4.94**	25.69	**11.68**
WiP +SMPLify3D	D	11.63	6.53	**22.37**	12.01

表4. **足地接触分析**。我们通过在三个已知基准测试中评估足地接触准确度，展示了所提出的传感器布置的潜力。

方法	TotalCapture	DIP-IMU	UIP-DB（测试）
UIP（膝盖）	0.79	0.73	0.81
UMotion（膝盖）	0.84	0.90	0.71
WiP（膝盖）	0.80	0.86	0.77
WiP（脚）	**0.97**	**0.96**	-

是在膝盖上放置传感器采集的）。表中显示，WiP 的表现优于 UIP 和我们的基线方法。我们认为有必要强调，MA-DB 上误差增大主要是由于我们仅进行了最低限度的预处理和噪声过滤（不同于采用扩展卡尔曼滤波器及其他传感器融合技术的 UIP-DB）。我们还要指出，这是有意为之，旨在保留实时性能、避免外部偏差，并直接基于原始测量值评估 WiP。

5.4 基于旋转/网格的指标评估

为了使 WiP 能够支持用于角色动画和网格恢复的人体骨骼层级表示（例如 SMPL 参数模型），可以使用现成的逆运动学求解器来估计关节局部旋转和网格顶点位置。SMPLify3D 是此类技术之一，由 Zuo等人 [2021] 提出。SMPLify3D 利用人体运动先验，估计与 SMPL 模型相一致的关节层级旋转。关于基于旋转和网格的比较，请参见表 3。如表中所示，尽管 WiP 仅依赖 PWD数据，而以往方法同时使用 PWD 和惯性测量单元输入，但它在 TotalCapture 和 UIP-DB 数据集上均取得了有竞争力的性能。

表 5. **对 WiP 各组件的消融研究**。我们报告了含噪 TotalCapture 和 UIP-DB 上的位置相关指标。

方法	TotalCapture（+噪声）				UIP-DB（真实）			
	PE	EEE	AJE	SGE	PE	EEE	AJE	SGE
WiP	4.62	**6.21**	2.56	**1.86**	**6.88**	**15.99**	0.22	**6.32**
使用去噪微调	5.12	6.80	1.42	2.11		-		
不含 STJ-SA	5.17	8.75	**0.71**	2.36	7.42	19.85	0.22	8.16
不含 PWD-Aux	4.66	10.89	2.53	3.03	8.10	23.81	**0.21**	9.83
不含门控	**4.27**	9.35	2.34	2.39	7.14	19.72	0.22	8.71

以往方法同时使用 PWD 和惯性测量单元输入。值得注意的是，尽管我们的模型实现了更低的关节位置误差，但 SMPLify3D 的姿态优化过程也引入了额外误差。

5.5 下半身传感器放置（膝盖与脚）

以往大多数研究都将传感器（通常为惯性测量单元）放置在膝盖上。对于基于惯性的方法而言，这种选择是合理的，因为它能够有效捕捉腿部弯曲情况。然而，我们假设，这种配置会影响脚与地面的接触表现，而这对自然运动建模至关重要。为此，我们将下半身传感器放在脚上，而非膝盖上。为了评估所提方法的有效性，我们测量了脚与地面的接触准确率。按照以往研究的做法，当脚趾和脚跟关节同时满足以下条件时，我们将脚定义为与地面接触：（i）速度低于 0.2 m/s；（ii）垂直方向上距离地面不超过 10 cm。如表4所示，基于膝盖的方案（UIP、UMotion 和 WiP-knees）在估计脚与地面的接触情况时误差较大，而 WiP-feet 的准确率显著更高，接触误差最低，脚与地面的交互也最稳定、最自然。

5.6 消融研究

为了评估 WiP 各个组成模块的贡献，我们在两个数据集上进行了广泛的消融研究：（1）添加合成噪声的 TotalCapture，用于考察对抖动的鲁棒性；（2）UIP-DB，由真实但经过预处理的运动数据组成，用于评估在真实条件下的性能。由于两个数据集的传感器布置不同，我们分别在 TotalCapture 和 UIP-DB 上评估 WiP-feet 和 WiP-knees。如表 5 所示，每个组件都有其独特且互补的贡献。例如，移除 STJ-SA 层可以减少抖动，但会损害运动灵活性，导致末端执行器误差增加。另一方面，去除去噪阶段会使模型更加“受输入引导”而非受运动引导，从而加剧抖动。还可以观察到，用标准交叉注意力层替换门控交叉注意力模块，会使 PWD 测量保持恒定平衡，从而降低关节位置误差，但增加抖动，导致动作更加不稳定且不自然。由于 UIP-DB 经过预处理，我们发现，在 WiP-knees 变体中省略去噪阶段，并在干净数据上进行端到端训练（包括 STJ-SA 层），效果更好。需要注意的是，尽管 WiP 并非在每项指标上都达到最佳表现，但其总体结果最为出色。

6 结论

我们提出了一种新型动作捕捉系统，可根据稀疏的成对距离（具体而言是超宽带测量）重建全身三维姿态，无需依赖外部摄像机或预先校准的身体模型。我们方法的核心是一种紧凑的 Transformer 解码器架构，经过训练，可对时序 PWD 序列进行去噪和解析，从而实现精准的姿态估计。通过大量实验，包括野外录制以及对人类和动物的跟踪，我们证明该系统能够以实时帧率实现平滑、高保真的重建，性能优于现有的惯性基线方法。我们的解决方案便于携带且与体型无关，在动画和机器人学领域具有潜在应用价值。

局限性。 Ciholas DWTAG 设备目前存在显著噪声，影响了测量的整体精度。不过，我们预计超宽带技术将在未来几年持续发展，从而降低信号噪声并提升硬件稳定性和性能。这些进步将进一步为系统的开发与优化带来新的机遇。我们的工作提供了稳健的算法基础，旨在充分利用这些技术改进。随着超宽带技术的发展，我们预计本方法的准确性和可靠性也将相应提高。

未来工作。 为应对噪声敏感性和非视距条件等挑战，未来的工作可以探索先进的分布式网络策略。例如，将传感器组织成具有直接视距的团、采用成对增量拓扑，或使用固定锚站进行多边定位。另一个有前景的方向是在多受试者场景中评估系统性能：多个个体佩戴超宽带传感器，从而产生更密集且可能包含更多信息的 PWD 测量。

参考文献

Abdulrahman Alarifi、AbdulMalik Al-Salman、Mansour Alsaleh、Ahmad Alnafessah、Suheer Al-Hadhrami、Mai A. Al-Ammar 和 Hend S. Al-Khalifa。2016年。超宽带室内定位技术：分析与最新进展。传感器 16，5（2016年5月）。https://doi.org/10.3390/s16050707 Jean-Baptiste Alayrac、Jeff Donahue 等。2022年。Flamingo：用于少样本学习的视觉语言模型。载于第*36*届神经信息处理系统国际会议论文集 （美国路易斯安那州新奥尔良）*(NIPS'22)*。Curran Associates Inc.，美国纽约州 Red Hook，论文1723，21页。Temi Ami-Williams、Christina-Georgia Serghides 和 Andreas Aristidou。2024年。极端服饰下传统舞蹈的数字化：Eyo案例研究。《文化遗产杂志》67（2024年3月），145–157。Andreas Aristidou、Daniel Cohen-Or、Jessica K. Hodgins 和 Ariel Shamir。2018年。动作捕捉清理的自相似性分析。《计算机图形论坛》37，2（2018年5月），297–309。https://doi.org/10.1111/cgf.13362 Andreas Aristidou、Ariel Shamir 和 Yiorgos Chrysanthou。2019年。数字舞蹈民族志：整理大型舞蹈合集。《计算与文化遗产杂志》12，4，论文29（2019年11月），27页。https://doi.org/10.1145/3344383 Rayan Armani、Changlin Qian、Jiaxi Jiang 和 Christian Holz。2024年。Ultra Inertial Poser：利用稀疏惯性传感器和超宽带测距实现可扩展的动作捕捉与跟踪。载于*ACM SIGGRAPH 2024*会议论文 （美国科罗拉多州丹佛）。计算机协会，美国纽约州纽约，论文51，11页。https://doi.org/10.1145/3641519.3657465 Valentín Barral、Carlos J. Escudero、José A. García-Naya 和 Roberto Maneiro-Catoira。2019年。使用低成本超宽带设备进行非视距识别与缓解。传感器 19，16（2019年）。https://doi.org/10.3390/s19163464 Angela Castillo、Maria Escobar、Guillaume Jeanneret、Albert Pumarola、Pablo Arbeláez、Ali Thabet 和 Artsiom Sanakoyeu。2023年。BoDiffusion：通过扩散稀疏观测合成全身人体运动。载于*IEEE/CVF*国际会议

计算机视觉研讨会（*ICCVW '23'*）论文集。4223–4233。https://doi.org/10.1109/ICCVW60793.2023.00456 Kang Chen、Yupan Wang、Song-Hai Zhang、Sen-Zhe Xu、Weidong Zhang 和 Shi-Min Hu。2021。MoCap-solver：一种用于光学动作捕捉数据的神经求解器。*ACM* 图形学汇刊 40，4，论文84（2021年7月），11页。https://doi.org/10.1145/3450626. 3459681 Ciholas。2025a。DWETH111。https://cuwb.io/products/dweth111/。Ciholas。2025b。DWTAG100。https://cuwb.io/products/dwtag100/。J. A. Corrales、F. A. Candelas 和 F. Torres。2008年。通过卡尔曼滤波器融合IMU/UWB数据，对人类操作员进行混合跟踪。载于第三届*ACM/IEEE*人机交互国际会议论文集：与机器人共处（*HRI 2008*）。193–200。Mickaël Delamare、Remi Boutteau、Xavier Savatier 和 Nicolas Iriart。2020年。面向工业应用的UWB定位系统静态与动态评估。《科学》*2020*年，第*2*卷，第*23*页 2，2（2020年4月），23。https://doi.org/10.3390/SCI2020023 Marco Di Renzo、R. Michael Buehrer 和 Jaime Torres。2007年。基于超宽带的身体区域网络中，脉冲形状失真对全身动作捕捉和步态分析测距精度的影响。载于*IEEE*全球电信会议（*GLOBECOM*）。3775–3780。https://doi.org/10.1109/GLOCOM.2007.717 Yuming Du、Robin Kips、Albert Pumarola、Sebastian Starke、Ali Thabet 和 Artsiom Sanakoyeu。2023年。Avatars Grow Legs：利用扩散模型从稀疏跟踪输入生成平滑的人体运动。载于*IEEE/CVF*计算机视觉与模式识别会议（*CVPR '23*）。481–490。https: //doi.org/10.1109/CVPR52729.2023.00054 Robert J. Fontana。2004年。短脉冲超宽带（UWB）技术的近期系统应用。*IEEE*微波理论与技术汇刊 52，9（2004年），2087–2104。Colin Goodall。1991年。形状统计分析中的普罗克鲁斯特斯方法。英国皇家统计学会杂志：*B*辑（方法论） 53，2（1991年12月），285–321。https://doi.org/10.1111/j.2517-6161.1991.tb01825.x Arturo Guizar、Claire Goursaud 和 Jean Marie Gorce。2016年。身体区域网络的体表信道模型下IR-UWB跨层测距协议的性能。《电信年鉴/电信年鉴》 71，9-10（2016年10月），453–463。https://doi.org/10.1007/s12243-016-0500-4 Fredrik Gustafsson 和 Fredrik Gunnarsson。2005年。利用无线网络进行移动定位：基于可用无线网络测量的可能性与根本局限。*IEEE*信号处理杂志 22，4（2005年），41–53。Vladimir Guzov、Aymen Mir、Torsten Sattler 和 Gerard Pons-Moll。2021。人体姿态定位系统（HPS）：利用身体佩戴式传感器，在大型场景中进行三维人体姿态估计与自定位。载于*IEEE/CVF*计算机视觉与模式识别会议（*CVPR '21*）。4316–4327。https://doi.org/10.1109/CVPR46437.2021.00430 Jihad Hamie、Benoit Denis、Raffaele D'Errico 和 Cedric Richard。2015年。面向可穿戴UWB网络动作捕捉应用的基于体表TOA的测距误差模型。环境智能与人性化计算杂志 6，5（2015年10月），603–612。https://doi.org/10.1007/s12652-013-0215-6 Jihad Hamie、Benoit Denis 和 Cedric Richard。2013年。异构身体区域网络中的联合动作捕捉与导航：基于邻域图的距离估计。载于第*10*届定位、导航与通信研讨会论文集（*WPNC '13*）。https://doi.org/10.1109/WPNC.2013.6533282 Daniel Holden。2018年。通过去噪鲁棒求解光学动作捕捉数据。*ACM* 图形学汇刊 37，4，论文165（2018年7月），12页。https://doi.org/10.1145/ 3197517.3201302 Dominik Hollidt、Paul Streli、Jiaxi Jiang、Yasaman Haghighi、Changlin Qian、Xintong Liu 和 Christian Holz。2024年。EgoSim：一种用于运动与活动期间身体佩戴式摄像机的自我中心多视角模拟器及真实数据集。载于神经信息处理系统进展，A. Globerson、L. Mackey、D. Belgrave、A. Fan、U. Paquet、J. Tomczak 和 C. Zhang（编），第37卷。Curran Associates, Inc.，106607–106627。Yinghao Huang、Manuel Kaufmann、Emre Aksan、Michael J. Black、Otmar Hilliges 和 Gerard Pons-Moll。2018年。Deep Inertial Poser：学习从稀疏惯性测量中实时重建人体姿态。*ACM* 图形学汇刊 37，6（2018年12月），185:1–185:15。https://doi.org/10.1145/3272127.3275108 Jiaxi Jiang、Paul Streli、Manuel Meier 和 Christian Holz。2024年。EgoPoser:: Robust real-time egocentric pose estimation from sparse and intermittent observations everywhere. 载于欧洲计算机视觉会议论文集（*ECCV '24*）。Springer。Jiaxi Jiang、Paul Streli、Huajian Qiu、Andreas Fender、Larissa Laich、Patrick Snape 和 Christian Holz。2022a。AvatarPoser：基于稀疏运动感知的关节式全身姿态跟踪。载于欧洲计算机视觉会议论文集（*ECCV '22*），Shai Avidan、Gabriel Brostow、Moustapha Cissé、Giovanni Maria Farinella 和 Tal Hassner（编）。Springer Nature Switzerland，Cham，443–460。https://doi.org/ 10.1007/978-3-031-20065-6 26_

Yifeng Jiang、Yuting Ye、Deepak Gopinath、Jungdam Won、Alexander W. Winkler 和 C. Karen Liu。2022b。Transformer Inertial Poser：利用稀疏惯性测量单元实时重建人体运动并同步生成地形。载于*SIGGRAPH Asia 2022 Conference Papers (SA '22)*。ACM，纽约州纽约，美国，1–9。https://doi.org/10.1145/3550469.3555428 Roman S. Kulikov、Dmitry V. Tsaregorodcev、Nikita I. Petukhov、Alexander A. Chugunov 和 Ana Mitic。2020。使用扩展卡尔曼滤波器进行相关点定位的算法。载于*Proceedings of the 2nd International Youth Conference on Radio Electronics, Electrical and Power Engineering (REEPE '20)*。IEEE。https://doi.org/10. 1109/REEPE49198.2020.9059135 Jiye Lee 和 Hanbyul Joo。2024。Mocap Everyone Everywhere：利用智能手表和头戴式摄像机实现轻量级动作捕捉。载于*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR '24)*。1091–1100。https://doi.org/10.1109/CVPR52733.2024.00110 Sunmin Lee、Sebastian Starke、Yuting Ye、Jungdam Won 和 Alexander Winkler。2023。QuestEnvSim：基于稀疏传感器、具备环境感知能力的模拟运动跟踪。载于*ACM SIGGRAPH 2023 Conference Proceedings*。ACM，纽约州纽约，美国，1–9。https://doi.org/10.1145/3588432.3591504 Magnus Lindhé、Kin Cheong 和 Henrik I. Christensen。2006。融合超宽带距离与角度数据以实现移动机器人定位。载于*Proceedings ofthe IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '06)*。IEEE，4034–4039。Hung Yu Ling、Fabio Zinno、George Cheng 和 Michiel Van De Panne。2020。使用运动变分自编码器的角色控制器。*ACM* 图形学汇刊 39，4（2020年8月），40:40:1–40:40:12。https://doi.org/10.1145/3386569.3392422 Huakun Liu、Hiroki Ota、Xin Wei、Yutaro Hirao、Monica Perusquia-Hernandez、Hideaki Uchiyama 和 Kiyoshi Kiyokawa。2025。UMotion：利用惯性单元和超宽带单元进行不确定性驱动的人体运动估计。载于*IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR'25)*。7085–7094。https://doi.org/ 10.1109/CVPR52734.2025.00664 Matthew Loper、Naureen Mahmood、Javier Romero、Gerard Pons-Moll 和 Michael J. Black。2015。SMPL：一种蒙皮多人线性模型。*ACM* 图形学汇刊 34，6，文章248（2015年10月），16页。https://doi.org/10.1145/2816795.2818013 Dimitrios Lymberopoulos、Jie Liu、Xiaofan Yang、Kamin Whitehouse、Yunxin Liu 和 Harry Husted。2015。微软室内定位竞赛：经验与教训。载于*Proceedings of the ACM International Workshop on Mobile Entity Localization and Tracking in GPS-less Environments (MELT '15)*。ACM，3–8。Naureen Mahmood、Nima Ghorbani、Nikolaus F. Troje、Gerard Pons-Moll 和 Michael J. Black。2019。AMASS：表面形状动作捕捉档案。载于*Proceedings of the International Conference on Computer Vision (ICCV '19)*。5442–5451。Christen Millerdurai、Hiroyasu Akada、Jian Wang、Diogo Luvizon、Christian Theobalt 和 Vladislav Golyanik。2024。EventEgo3D：基于第一人称事件流的三维人体动作捕捉。载于*IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR '24)*。E. Mok、F. Lau、L. Xia、G. Retscher 和 H. Tian。2012。使用超宽带技术进行分米级定位的影响因素。*Survey Review* 44，324（2012），37–44。https://doi.org/10.1179/1752270611Y.0000000009 Timothy Otim、Alfonso Bahillo、Luis Enrique Díez、Peio Lopez-Iturri 和 Francisco Fal- cone。2020。利用身体可穿戴传感器实现亚米级超宽带室内定位。*IEEE Access* 8（2020）。https://doi.org/10.1109/ACCESS.2020.3027669 Xiaoyu Pan、Bowen Zheng、Xinwei Jiang、Zijiao Zeng、Qilong Kou、He Wang 和 Xiaogang Jin。2024。RoMo：一种用于全身无标记光学动作捕捉的鲁棒求解器。载于*SIGGRAPH Asia 2024 Conference Papers*（日本东京）*(SA '24)*。计算机协会，纽约州纽约，美国，文章76，共11页。https://doi.org/10.1145/3680528.3687615 Shu-Juan Peng、Gao-Feng He、Xin Liu 和 Hua-Zhen Wang。2015。利用自适应非负矩阵分解，基于分层模块恢复不完整的人体动作捕捉数据。*Computer & Graphics* 49，C（2015年6月），10–23。Xue Bin Peng、Yunrong Guo、Lina Halper、Sergey Levine 和 Sanja Fidler。2022。ASE：面向物理模拟角色的大规模可复用对抗技能嵌入。*ACM* 图形学汇刊 41，4，文章94（2022年7月），17页。https://doi.org/10.1145/ 3528223.3530110 Rok Piltaver。2015。对使用超宽带实时定位系统采集的人体运动轨迹进行去噪。*Informatica* 39，3（2015）。Jose Luis Ponton、Eduard Pujol、Andreas Aristidou、Carlos Andujar 和 Nuria Pelechano。2025。DragPoser：通过潜在空间优化，基于可变稀疏跟踪信号进行运动重建。《计算机图形论坛》44，2（2025），e70026。https: //doi.org/10.1111/cgf.70026 Jose Luis Ponton、Haoran Yun、Andreas Aristidou、Carlos Andujar 和 Nuria Pelechano。2023。SparsePoser：基于稀疏数据的实时全身运动重建。*ACM* 图形学汇刊 43，1，文章5（2023年10月），14页。https://doi.org/10.1145/ 3625264 Alec Radford、Jeffrey Wu、Rewon Child、David Luan、Dario Amodei 和 Ilya Sutskever。2019。语言模型是无监督多任务学习者。技术报告。OpenAI。技术报告。Colin Raffel、Noam Shazeer、Adam Roberts、Katherine Lee、Sharan Narang、Michael Matena、Yanqi Zhou、Wei Li 和 Peter J. Liu。2020。探索迁移学习的极限

与统一文本到文本转换器。*J. Mach. Learn. Res.*21，1，文章140（2020年1月），67页。Davis Rempe、Tolga Birdal、Aaron Hertzmann、Jimei Yang、Srinath Sridhar 和 Leonidas J. Guibas。2021。HuMoR：用于鲁棒姿态估计的三维人体运动模型。载于 *IEEE/CVF International Conference on Computer Vision (ICCV '21)*。11468–11479。https://doi.org/10.1109/ICCV48922.2021.01129 Mingyi Shi、Sebastian Starke、Yuting Ye、Taku Komura 和 Jungdam Won。2023。PhaseMP：通过相位条件人体运动先验实现鲁棒三维姿态估计。载于*IEEE/CVF International Conference on Computer Vision (ICCV '23)*。14679–14691。https://doi.org/10.1109/ICCV51070.2023.01353 Kazimierz Siwiak 和 Debra McKeown。2004年。超宽带无线电技术。Wiley。https://doi.org/10.1002/0470859334 Guy Tevet、Sigal Raab、Brian Gordon、Yoni Shafir、Daniel Cohen-or 和 Amit Haim Bermano。2023。人体运动扩散模型。载于*The Eleventh InternationalConference on Learning Representations (ICLR '23)*。Qinglin Tian、Kevin I.Kai Wang 和 Zoran Salcic。2019。人体遮挡效应对用于行人跟踪的基于超宽带测距系统的影响。*IEEE*仪器与测量汇刊 68，10（2019年10月），4028–4037。https://doi.org/10.1109/TIM.2018.2884605 Hugo Touvron、Louis Martin 等。2023。Llama 2：开放基础与微调聊天模型。arXiv:2307.09288 [cs.CL] https://arxiv.org/abs/2307.09288 Matt Trumble、Andrew Gilbert、Charles Malleson、Adrian Hilton 和 John Collomosse。2017。Total Capture：融合视频与惯性传感器进行三维人体姿态估计。载于*British Machine Vision Conference (BMVC '17)*。Dmitry Tsaregorodtsev、Alexander Chugunov 和 Nikita Petukhov。2019。超宽带动作捕捉无线电系统。载于*2019 International Multi-Conference on Industrial Engineering and Modern Technologies, FarEastCon 2019*。IEEE。https://doi.org/10.1109/FarEastCon.2019.8934111 Ashish Vaswani、Noam Shazeer、Niki Parmar、Jakob Uszkoreit、Llion Jones、Aidan N. Gomez、Łukasz Kaiser 和 Illia Polosukhin。2017。Attention is all you need。载于*Proceedings of the 31st International Conference on Neural Information Processing Systems*（美国加利福尼亚州长滩）*(NIPS'17)*。Curran Associates Inc.，美国纽约州雷德胡克，6000–6010。Alexander Winkler、Jungdam Won 和 Yuting Ye。2022。QuestSim：利用模拟化身和稀疏传感器进行人体运动跟踪。载于*SIGGRAPH Asia 2022 Conference Papers*。ACM，韩国大邱，1–8。https://doi.org/10.1145/ 3550469.3555411 Xinyu Yi、Yuxiao Zhou、Marc Habermann、Vladislav Golyanik、Shaohua Pan、Christian Theobalt 和 Feng Xu。2023。EgoLocate：利用稀疏身体佩戴式传感器进行实时动作捕捉、定位与稠密建图。*ACM* 图形学汇刊 42，4（2023年7月），76:1–76:17。https://doi.org/10.1145/3592099 Xinyu Yi、Yuxiao Zhou、Marc Habermann、Soshi Shimada、Vladislav Golyanik、Christian Theobalt 和 Feng Xu。2022。Physical Inertial Poser（PIP）：基于稀疏惯性传感器的物理感知实时人体运动跟踪。载于*IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR '22)*。IEEE，美国路易斯安那州新奥尔良，13157–13168。https://doi.org/10.1109/CVPR52688.2022.01282 Xinyu Yi、Yuxiao Zhou 和 Feng Xu。2021。TransPose：利用六个惯性传感器进行实时三维人体平移与姿态估计。*ACM* 图形学汇刊 40，4（2021年7月），86:1–86:13。https://doi.org/10.1145/3450626.3459786 Handi Yin、Bonan Liu、Manuel Kaufmann、Jinhao He、Sammy Christen、Jie Song 和 Pan Hui。2024。EgoHDM：实时第一人称惯性人体动作捕捉、定位与稠密建图系统。*ACM* 图形学汇刊 43，6，文章236（2024年11月），12页。https://doi.org/10.1145/3687907 Qian Yu、Qing Li 和 Zhigang Deng。2007年。针对多个交互式铰接目标的在线动作捕捉标记点标注。《计算机图形论坛》26，3（2007年），477–483。https://doi.org/10.1111/j.1467-8659.2007.01070.x Faheem Zafari、Athanasios Gkelias 和 Kin K. Leung。2019。室内定位系统与技术综述。*IEEE Communications Surveys and Tutorials* 21，3（2019），2568–2599。https://doi.org/10.1109/COMST.2019.2911558 Xiaozheng Zheng、Zhuo Su、Chao Wen、Zhou Xue 和 Xiaojie Jin。2023。通过关节级建模，基于稀疏观测实现逼真的全身跟踪。载于*Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV '23)*。14678–14688。https://doi.org/10.1109/ICCV51070.2023.01349 Shaghayegh Zihajehzadeh 和 Edward J. Park。2017。一种新颖的生物力学模型辅助IMU/UWB融合方法，用于无磁力计下肢动作捕捉。*IEEE*系统、人和控制论汇刊：系统 47，6（2017年6月），927–938。https://doi.org/10.1109/TSMC.2016.2521823

Xinxin Zuo、Sen Wang、Jiangbin Zheng、Weiwei Yu、Minglun Gong、Ruigang Yang 和 LiCheng。2021。SparseFusion：基于稀疏 RGBD 图像的动态人体化身建模。多媒体汇刊23（2021年1月），1617–1629。https://doi.org/10.1109/ TMM.2020.3001506

A 附录

我们扩展了对 WiP 的分析。A.1 节评估了我们的模型生成的成对距离（PWD）。A.2 节考察了 WiP 对噪声输入的鲁棒性，A.3 节则展示了新引入的时空关节联合自注意力（STJ-SA）层的定性结果。

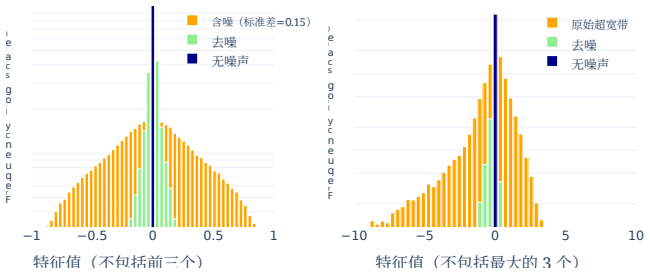

图 11. **WiP 预测的成对距离（PWD）的定性分析**。我们计算成对距离辅助任务头生成的成对距离（PWD）的特征值，排除每个矩阵中最大的三个特征值，并将剩余特征值绘制成直方图。为确保完整性和公平比较，我们仅提取稀疏关节（骨盆、头部、手部和脚）。

A.1 成对距离去噪分析

由三维点构成的定义良好的欧几里得 PWD 矩阵恰好产生三个正特征值，其余特征值则趋近于零。负特征值表明其具有非欧几里得特性，例如违反三角不等式；而额外的正特征值则表明该 PWD 无法完全嵌入三维空间。为评估我们对 PWD 辅助任务的预测质量，我们进行了特征值分析，如图 11 所示。可以观察到，WiP 产生了密集分布在零附近的特征值，这表明它隐式地学习了定义良好的欧几里得 PWD 的结构。表 6 报告了两个指标，展示了我们的模型预测的有效性：

累积解释方差比：累积解释方差比最初源自主成分分析，是应用主成分分析时选择成分数量的一种可靠方法，因为它能够在最大限度保留原始数据与其低维投影之间方差的同时实现降维。CEV 定义如下：

$$CEV\ (k) = \sum_{i=1}^{k} \frac{\lambda_i}{\sum_{j=1}^{n} \lambda_j} \tag{16}$$

在定义良好的欧几里得成对距离中，保留最大的 $k = 3$ 分量通常足以捕获数据中近乎 100% 的方差，反映其底层的三维结构。我们采用这一技术来评估三个最大特征值分量所捕获的方差。

三角不等式得分（*TI* 得分）：定义良好的成对距离（PWD）具有一项重要性质，即任意三个索引都满足三角不等式。因此，我们定义了一种用于评估估计成对距离中

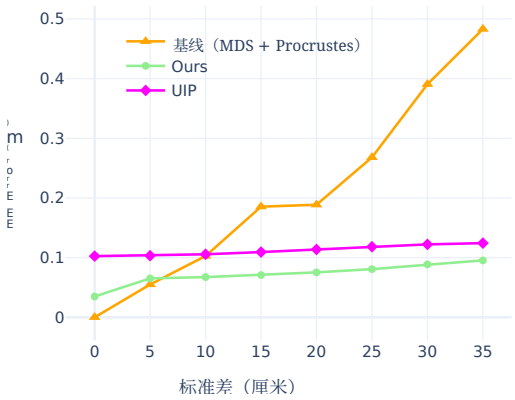

图 12。**随添加噪声增加的末端执行器（EE）误差。**随着输入噪声逐渐增加，EE 误差也随之上升。然而，WiP 仍具有很强的鲁棒性，即使在高噪声水平下，也能始终将 EE 误差保持在较低水平。

这一性质的新指标：

$$\mathrm{TIS}\left(\hat{D}^{[t]}\right) = \frac{1}{J}\sum_{i\neq j\neq k}\begin{cases}1 & \text{if } \hat{D}_{ij}^{[t]} + \hat{D}_{jk}^{[t]} - \hat{D}_{ik}^{[t]} >= 0\\ 0 & \text{otherwise}\end{cases} \tag{17}$$

其中，$\hat{D}^{[t]} \in \mathbb{R}^{J\times J}$ 是 J 关节的估计成对距离。我们计算所有样本的 TI 分数并取平均值。分数最高为 1，表示违反率为 0%。

表 6。**成对距离预测的定性分析。**WiP 经过训练，可估计定义明确的欧几里得成对距离。我们定义定量指标，并与不同输入进行比较。“合成真实值”指根据光学数据计算得到的完美成对距离，“原始输入”指含噪测量值。

方法	TotalCapture 含噪		MA-DB	
	CEV 分数（↑）	TI 分数（↑）	CEV 分数（↑）	TI 分数 (↑)
合成真实值	1.00	1.000	1.000	1.000
原始输入	0.913	0.909	0.500	0.945
WiP（不进行去噪）	0.984	0.989	0.990	0.974
WiP	**0.991**	**0.993**	**0.990**	**0.983**

A.2 WiP 作为 PWD 去噪器

由于遮挡、距离变化和运动速度等因素，原始测量值中的噪声本质上难以预测。因此，具备抗噪能力，并能在过去的预测与当前输入之间自适应地分配注意力至关重要。在图 12 中，我们比较了模型、UIP [Armani 等人 2024], 和经典基线的噪声容忍度。即使噪声水平高达 35 厘米，我们的模型依然稳健，并将 EEE（末端执行器误差）保持在 10 厘米以下，优于另外两种方法。虽然基线在距离测量完全准确时表现出色，但在密集场景下会受到可用传感器数量的限制，且对噪声高度敏感。

A.3 STJ-SA 可视化

为进一步说明我们添加的 STJ-SA 层的有效性，我们在图 13 中通过可视化最终层中关键关节随时间变化的注意力图进行定性分析。对于每个关节，该图显示其在所有 T 时间戳和 J 关节（共 $T\cdot J$ 个）中分配给其他关节的注意力权重。结果表明，某个关节 $j^{[t]}$ 主要关注其自身过去的实例 $j^{[t-i]}$（$0 \le i < t$），凸显了对时间轨迹的关注。此外可以看出，除了关注自身关节外，它们还会关注解剖学相关关节（骨盆 → 右/左锁骨、右手 →右肩、头部 → 右锁骨）。

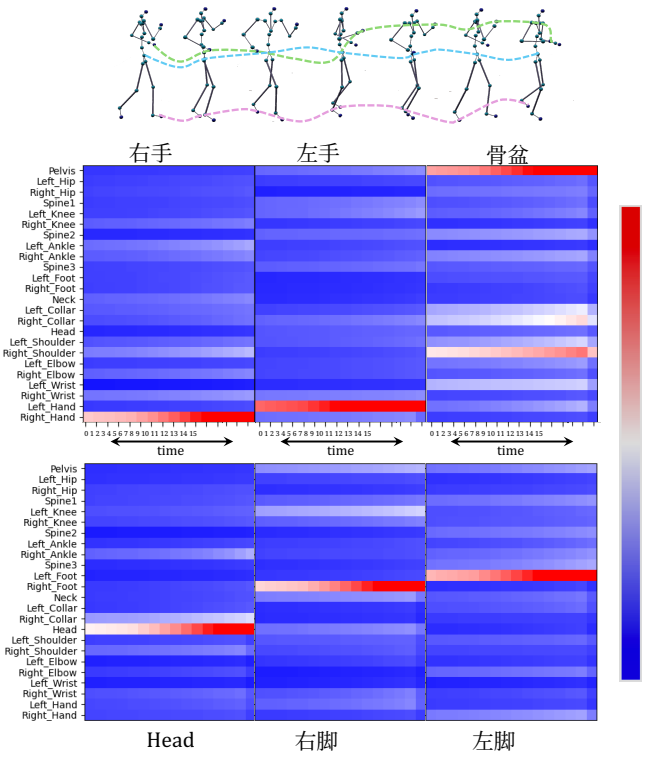

图 13。**末端执行器关节的 STJ-SA 可视化。**我们从 WiP 的最终 STJ-SA 层中提取注意力权重图，并在各个头之间进行平均。红色像素表示较高的注意力值（每张图的总和为 1）。

B 实现细节

B.1 成对距离嵌入

将成对距离（PWDs）编码为 Transformer 模型的输入并非易事。为此，我们提出了一种新颖的 PWD 嵌入策略。在 WiP 中，每个词元表示整个骨架的空间结构化嵌入。为了保留关节特定的空间关系，我们根据 PWD 矩阵中对应的第 $\tilde{D}_j$ 行，为关节 j 推导其嵌入。具体而言，我们使用逐行一维卷积层实现这一过程，其卷积核大小和步幅分别等于 J，输出通道为 d。该操作将每个关节转换为 d 维嵌入。随后，将所得激活值重塑为空间连贯的姿态嵌入 $\mathrm{Emb}\left(\tilde{D}^{[t]}\right) \in \mathbb{R}^{J\times d}$，同时编码全局骨架上下文和关节特定的空间关系（示意图见图 14）。

表7. WiP 的训练方案。

训练阶段	#预热步数	#训练步数	优化器	学习率 Rate	权重衰减	有效批次大小	#可训练 / 总计参数（M）
（I）距离到运动	500	$100K$	AdamW	$1e-4$	$1e-3$	3072	137。1/137。1
（II）去噪						256	10.1/138.3

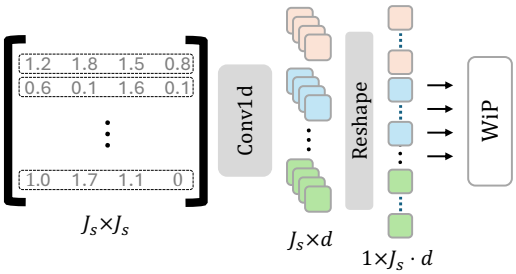

图 14. **PWD嵌入过程**。我们通过应用核大小和步幅均设为 J 的一维卷积，在嵌入中保留空间结构。



图 15. Ciholas DWTAG100 传感器。

B.2 训练

在两个训练阶段中，我们使用 AMASS 运动数据集 [Mahmood 等人 2019],，并排除 TotalCapture [Trumble 等人 2017] 进行评估。在 DanceDB [Aristidou 等人 2019], 上进行评估时，我们将其从训练集中移除，并纳入 TotalCapture。遵循相关文献，我们采用教师强制策略训练模型，以确保稳定性并改善收敛性。更多训练细节见表7。对于损失分量权重 λ_i，我们设定为：$\lambda_{dd} = 1.0$、$\lambda_{pd} = 1.0$、$\lambda_{refs} = 0.5$、$\lambda_{cons} = 0.5$、$\lambda_{velo} = 0.1$、$\lambda_{rigidity} = 1.0$、$\lambda_{gravity} = 0.05$。

B.3 评估

对于生成过程，为了预测时间戳 $t+1$ 处的下一个姿态，我们通过计算最后一个预测姿态 $\hat{P}^{[t]} \in \mathbb{R}^{J \times J}$ 的成对距离来计算 $\hat{D}^{[t]} \in \mathbb{R}^{J \times 3}$。然后，我们将其与上一次迭代得到的预测序列 $\hat{D}^{[t-w]}, \ldots, \hat{D}^{[t-1]}$ 拼接，供下一次迭代使用。为了进一步减少抖动，我们发现，对过去 $\lfloor \frac{w}{2} \rfloor$ 个姿势使用半正态高斯核进行加权平均（其标准差较小）效果良好。为公平比较，我们在所有实验中均采用这种方法，并报告表现最佳的选项。

C 硬件详情

我们使用来自 Ciholas 公司的 DWTAG100 传感器 [Ciholas 2025b], 该传感器支持高频 PWD 测量，并满足以 30 赫兹进行关节运动捕捉的基线要求。定位采用到达时间差和 Ciholas 数据协议，以实现低延迟通信。DWTAG100 标签使用 Ciholas 数据协议进行通信；这是一种低延迟二进制协议，可传输实时计时信息和传感器数据，而定位则通过到达时间差测量计算得出。这些传感器重量轻（17 克），由电池供电，配备超宽带收发器，尺寸紧凑（直径 37.2 毫米，高度 13.4 毫米），并搭载用于惯性运动捕捉的惯性测量单元传感器

。通过与固定的 DWETH111 锚点 [Ciholas2025a],通信，可以实现全局三维定位，从而在非结构化环境中进行高精度跟踪。