Figure1：使用用户随身携带的各种移动设备，IMUPoser可以估计全身姿势。在最佳情况下，用户可以携带智能手机、智能手表和耳机（从3个设备获取姿势信息）。当然，设备数量会随时间变化，例如，耳机使用是间断的，并非每个人都佩戴智能手表。这意味着IMUPoser必须跟踪当前存在的设备、它们的位置，并使用可用的IMU数据。缩写键：L-左，R-右，H-手，P-口袋。

摘要

在行动中跟踪身体姿势在健身、移动游戏、上下文感知虚拟助手和康复等领域具有强大的用途。然而，用户不太可能购买并穿戴特殊的服装或传感器阵列来实现这一目标。相反，在这项工作中，我们探讨了使用许多用户已经拥有的设备（即智能手机、智能手表和耳机）中已经存在的IMUs来估计身体姿势的可行性。这种方法面临着多个挑战，包括来自低成本通用IMUs的嘈杂数据，以及用户身体上的测量点既稀疏又不断变化的事实。我们的流程接收可用的IMU数据的任意子集，甚至可能只来自单一设备，并生成最佳猜测的姿势。

为了评估我们的模型，我们创建了IMUPoser数据集，该数据集由10名参与者穿戴或携带现成的消费者设备，并跨足多种活动场景进行了收集。我们对我们的系统进行了全面的评估，在我们自己的IMU数据集和现有的IMU数据集上进行了基准测试。

1 引言

全身动作捕捉在电影视觉效果中很常见，正在逐渐进入消费者领域，尤其是虚拟现实领域。全身姿势跟踪在游戏[38]、健身[28]、康复[39]、生活记录[24]和上下文感知界面[1, 5]等领域具有明显的应用。例如，了解姿势的数字助手可以帮助足球运动员改进动作，或者帮助手术后康复患者监测步态变化。然而，目前大多数消费者没有工具来跟踪他们的姿势，也不想在家中添加传感器或穿戴特殊的服装或附件设备。然而，如果可以从用户已经拥有的设备中生成有用的姿势信息，将产生重大影响。

我们随身携带的大多数计算设备都包含IMUs，尤其是智能手机、智能手表和无线耳机。在这项工作中，我们研究如何利用这一戴戴式移动设备生态系统来实时估计用户的身体姿势，而无需外部基础设施。这种方法引入了新的挑战，先前的稀疏IMU姿势模型（例如[21, 60]）未曾面对。独特的是，跟踪的身体位置的位置和数量可以在移动过程中发生变化。例如，用户可以将手机从左口袋拿到右手，或者用户可以通过佩戴耳机增加感测点的数量。我们的模型必须接受各种不完整输入的组合，并在活动设备数量减少时（可能降至一个）能够优雅降级。其次，我们的系统必须使用来自消费者设备的IMU数据，这些数据比专业级动作捕捉服（例如XSens [63]）的数据要嘈杂，这些数据在高度相关的先前工作中使用，例如Sparse Internal Poser [60]，Deep Inertial Poser [21]和TransPose [65]。

表1提供了最相关的先前工作的概述。

为了评估我们的方法，我们创建了一个新的数据集：专业级的Vicon光学跟踪与普通佩戴/持有位置的设备IMU数据相配对。毫不奇怪，考虑到我们最多需要估计数百自由度的三个身体位置，我们的姿势输出是一种近似。然而，它很少会大幅出错，大多数情况下，用户的主要姿势和运动模式得以捕捉。这种“低保真度”的姿势输出不适用于高保真度应用，例如特效动作捕捉或虚拟现实头像，用户期望得到大致忠实的身体表示。尽管如此，IMUPoser的自适应和移动特性使其能够对用户进行被动和长期感测（甚至可能仅来自一个设备），因此特别适用于健康和健康应用。例如，这种低保真度的身体跟踪对于提高卡路里计数的准确性，跟踪康复疗程的进展，以及监测运动形式和次数都具有价值。我们在图2中突出显示了一些示例用途。

2 相关工作

现在，我们回顾全身数字化领域的相关工作。我们将研究外部和佩戴式捕捉系统，然后审查与我们的工作最相关的基于IMU的姿势感测方法。关于过去和现在的姿势估计方法的深入审查，我们建议读者参考Desmarais等人[15]和Nguyen等人[43]的调查报告。

2.1 使用外部传感器进行身体捕捉

存在一系列使用外部传感器来估计用户姿势的方法和解决方案。商业系统，如Vicon [58]和OptiTrack [41]，使用专门的硬件，包括高速红外摄像机，可以跟踪附着在用户的整个身体或个别部位（如脸部或手部）的反光标记。经过校准程序后，这些系统可以以毫米级的准确度跟踪大型空间。这些方法通常用于电影、游戏和角色动画制作。然而，其成本和设置要求通常不适合大多数消费者应用。

Figure2：由消费者移动设备提供动力的实时姿势估计（插图中的照片），可在许多领域中应用，包括体育（A）、康复（B）、健身（C）和交通（D）。还请注意，IMUPoser对遮挡（E）和光照条件（F）具有鲁棒性。缩写键：L-左，R-右，H-手，P-口袋。

目前，用于消费者应用的全身姿势估计通常依赖于成本更低且需要较少校准的传感器。深度摄像头，如微软Kinect [45, 53, 61]和英特尔RealSense [22]，使用中等成本传感器提供足够的姿势准确性，适用于虚拟现实和游戏应用。例如，Zimmermann等人[71]和Michael等人[37]扩展了这些方法，将深度图像与RGB数据结合，以提高准确性。这些商业可用的传感器在成本、可用性和准确性之间提供了良好的平衡，但它们通常是静态设置。计算机视觉和深度学习领域的最新成功使得从单目RGB摄像头中提取姿势数据成为可能。这包括从单一图像[11, 12, 46]、多个摄像头[14]或从2D图像估算3D姿势[34]的方法。

还存在用于虚拟现实和增强现实中的姿势跟踪的专门外部硬件[33]。例如，HTC Vive [19, 20]、Oculus Rift [36]和PlayStation VR [54]使用外部传感器基站跟踪头部、手柄和其他肢体附着物。任何未感知的关节可以使用逆运动学来粗略估计[47, 50]。用于姿势估计的其他非佩戴式外部方法包括电容感应[68]、磁场[44, 49]、射频[69]和机械联动[55]。

2.2 佩戴式非IMU传感器的身体捕捉

使用佩戴式传感器进行姿势估计比需要外部组件的系统更加便携和灵活。许多研究已经集中在捕捉特定身体部位上。例如，在虚拟现实/增强现实应用中，手部姿势非常重要，可以使用腕戴式摄像头[10, 29, 62]、电阻抗成像[67]、肌电图[30]、深度传感器[16]、磁性跟踪器[13]和专用手套[17]等方式来跟踪手部姿势。面部姿势也很重要，通常使用摄像头[4, 57]捕捉，但也探索了其他方法，如超声波[23]和肌电图[18]。

我们的研究更关注全身姿势估计。存在许多佩戴式身体感测方法，包括外骨骼[35]、超声波信标[59]、压力传感器[66]和射频识别（RFID）[27]。佩戴式摄像头方法特别受欢迎，例如Shiratori等人[52]、Ng等人[42]和Ahuja等人[2, 3]的工作。所有后者的方法都需要专门的额外硬件，而大多数用户并不拥有这些设备。我们的工作目标是将佩戴式姿势估计的灵活性带给用户，而不需要他们购买任何新设备。与类似理念的先前工作之一是“Pose-On-The-Go” [6]，该工作仅使用智能手机中的传感器估计用户的全身姿势，并且仅在手中持有时有效。然而，该方法通过测量绝对运动并使用动画化的（IK）虚拟角色来猜测大部分全身姿势（而非跟踪）。

2.3 使用佩戴式IMUs进行身体姿势估计

在这一部分，我们关注依赖于佩戴式IMUs的姿势系统，这与我们的技术方法最为匹配（表1提供了先前系统的概述）。虽然单个IMUs曾被用来跟踪个别肢体（例如ArmTrack [51]中的手臂姿势），但更常见的是在全身分布“舰队”式的IMUs（例如流行的XSens [63]套装）用于全身姿势估计。重要的是，这些设置在IMUs方面（因此在性能和噪声方面）是均匀的，并且倾向于使用高质量传感器以高帧率运行，这是消费者移动设备通常不具备的。正如我们发现的那样，在苹果自己的生态系统中使用的IMUs因设备而异，因此数据在质量、噪声和帧率方面也不同。

在以往的工作中，使用许多佩戴式IMUs，我们看到Tautges等人[56]能够使用四个XSens IMUs生成外观合理的动作流。稀疏惯性姿势[60]和深度惯性姿势[21]使用基于优化和深度学习的方法进行全身姿势估计，使用了6到17个佩戴式IMUs。这两个系统都使用SMPL [31, 48]，即统计身体模型，作为其姿势输出。TransPose [65]或Physical Inertial Poser [64]等方法基于这些努力，提供更准确的表示和更好的模型。所有这些工作都利用了使用的IMUs具有已知和校准的位置以及相同的噪声特性。

3 可能的设备组合

智能手机、智能手表和耳机具有不同的可能身体位置。例如，智能手机可以放在左口袋或右口袋，或者手持在左手或右手，也可以举到头部（接电话），或者用户根本不携带手机（6种可能状态）。对于智能手表，它们可以佩戴在左手腕或右手腕，或者用户根本不佩戴手表（3种可能状态）。对于类似耳机的设备，它们可以佩戴在头部，放在充电盒中并存放在左口袋或右口袋，或者用户根本不携带耳机（4种可能状态）。尽管目前将耳机放入充电盒通常会让它们休眠，但我们假设将来的固件更新可能会允许连续的IMU数据流传输，特别是考虑到充电盒中的较大电池。

完全列举，这会产生72种可能的设备位置组合。然而，我们排除了三种组合，即用户既佩戴耳机又将手机举到头部（接电话），因为这不是典型的使用情况。另一个无效的组合是没有手机、智能手表或耳机，因此我们的系统根本无法运行。这留下了68种可能的排列组合，其中有14种组合具有1个活动设备，36种组合具有2个活动设备，18种组合具有3个活动设备。

接下来，重要的是要考虑，某些设备组合对我们的目的来说并不提供大不相同的身体数据。例如，用户可以在左手腕上佩戴智能手表，并在左手持有智能手机 - IMU数据将高度相关，因此我们将其视为一个身体点。另一个例子是将智能手机存放在右口袋中，同时将耳机放在充电盒中 - 再次，IMU数据将相似。因此，我们的系统真正关心的是由68种可能的设备位置组合启用的身体位置组合 - 这24种组合在图3中有示例。

Figure3：我们支持并调查的24种可能的设备位置组合。

我们注意到我们的系统对身体位置做了一些简化的假设。例如，要使手的位置被视为活动位置，需要手持智能手机或佩戴智能手表在该手腕上。即使信号不完全相同，它们高度相关，因此信息功率是相似的。同样，将智能手机放在任一耳朵旁边被视为头部位置（而不是左耳或右耳）。我们之所以做出后一种简化，是因为苹果的AirPods（我们在实时实施中使用）融合了它们的IMUs以提供单一头部的6自由度估计，而不是从每个AirPod单独提供IMU数据。

4 实施

图4提供了我们流程的概述。我们关注三种流行的消费者设备：智能手机、智能手表和无线耳机/头戴耳机。每个这些设备都包含一个IMU、无线传输数据的能力以及一些本地计算能力。我们设想我们的模型在用户携带的性能最强的设备上执行，而其他性能较弱的设备通过蓝牙等方式流式传输其IMU数据。

4.1 模型

在学习架构方面，我们使用了一个受以前的工作启发的双向LSTM，尤其是[21, 65]。虽然我们尝试了较新的架构，比如transformers，但我们发现这些模型在实践中表现不佳。与我们测试过的其他模型相比，LSTMs产生了更平滑的输出预测。对于每个可用的IMU，我们的系统使用方向（表示为3×3的旋转矩阵）以及加速度作为输入，都是在全局坐标参考系中。与先前的工作不同，我们不会将这些输入归一化为相对于根IMU传感器位置（如骨盆）的位置，因为我们可用的设备各不相同。我们将这些输入展平并连接起来，形成一个大小为60的输入向量：5个可能的IMU位置×（3个加速度轴+3×3的方向旋转矩阵），我们将其输入到模型中。值得注意的是，我们的模型可以接受任何可用IMU数据的子集，不存在的设备会被屏蔽（即值设为零）。 输入向量首先通过一个ReLU激活的线性层转换为维度为512的嵌入。接下来，这些嵌入按顺序馈送到一个隐藏维度为512的双向LSTM中。最后，一个线性层输出144个SMPL [31]姿势参数，表示为24个关节，每个关节用6D旋转表示（这是一个更平滑的表示空间[70]）。SMPL还提供了一个身体网格（6890个顶点），可以在图1、图4和图7中看到。总之，我们的神经网络模型有1070万个可训练参数。 在训练过程中，前向运动学模块从这些姿势参数中计算关节位置，并相对于地面真实值进行进一步最小化。

4.2 IMU数据集合成

为了训练我们的姿势模型，我们需要大量的数据。为此，我们可以利用现有的运动捕捉数据库来生成大规模的合成语料库。具体来说，我们使用了AMASS [32]，这是包含了24个运动捕捉数据集（ACCAD、BioMotion、CMU MoCap、MIXAMO、Human Eva、Human 3.6M等）的汇编，总共包含了近63小时的高质量、高分辨率的SMPL [31]格式的运动捕捉数据。其中包括各种各样的动作和活动环境，如行走、体育、跳舞、锻炼、烹饪和自由互动等。有关AMASS数据集的构成的更多细节，请参考[32]。 我们注意到，AMASS已经被用于许多以前的工作[21, 64, 65]作为生成合成数据集的基础。 我们用于研究和实时演示（在第5.2.1和第8节中描述）的消费者设备以共同的帧速率为25 FPS运行。因此，我们将AMASS的60∼120 FPS的数据重新采样为25 FPS。然后，我们遵循了TransPose [65]和DIP [21]中使用的合成数据生成过程。简而言之，我们在SMPL网格的特定顶点（左右手腕、左右前裤袋和头皮）上"附加"虚拟IMUs，并使用全局坐标参考系中相邻帧计算合成加速度数据。为了生成合成方向数据，我们计算关节旋转相对于全局坐标的旋转，从关节开始依次叠加本地旋转，直到骨盆（根）为止，按照SMPL运动链的规则。我们将加速度数据（m/s²）缩放30倍，以适用于神经网络[65]。最后，我们不是向我们的数据集添加合成的高频噪音，而是使用长度为5帧（200毫秒）的平均窗口对合成和真实世界数据进行平滑处理，类似于[26]。 我们使用这个流程创建了24组数据，每组对应我们的24种设备位置组合（图3），然后将它们组合成一个单一的数据集。我们通过屏蔽（即将IMU数据设为零）模拟缺失的设备。例如，即使在我们最好的情况下有三个设备存在，这意味着输入向量的2/5是空的。AMASS数据63小时 × 25 FPS × 24种设备位置组合产生了134.8百万个用于训练的合成IMU实例，与配对的SMPL姿势数据。

Figure4：我们实时系统架构的概述。

4.3 训练

模型使用PyTorch和PyTorch Lightning深度学习框架进行端到端训练。我们使用批量大小为256，并使用Adam优化器来更新权重，学习率为3e-4。在训练过程中，我们使用非重叠的窗口，每个窗口包含5秒（125个样本）的配对IMU和姿势数据。如前所述，我们训练我们的模型使用均方误差（MSE）损失来回归全身姿势和全身关节位置。我们的总损失是这两个单独损失的总和。我们在NVIDIA Titan X GPU上进行了80个时代（22小时）的模型训练。

4.4 关节旋转精细调整

作为我们推断流程的最后一步，我们采用了[25]中提出的逆运动学细化方法，对我们的输出姿势进行最后的精细调整。尽管我们的模型预测了腿、手和头部的旋转，但即使在我们的损失项中给予了重要权重，它不一定会完全遵守IMUs提供的绝对方向。然而，对于带有设备的肢体来说，充分利用IMU的方向是合乎逻辑的，因为它既是绝对值，又比加速度计数据要嘈杂得多。更具体地说，由于我们从IMUs中获得了绝对方向，我们为每个设备的关节优化了某些骨骼方向。特别是对于手腕关节（智能手表/手机），我们优化了肘部和肩部的方向，同样适用于头部（耳机/手机）和臀部（手机/耳机盒）关节。我们使用PyTorch框架实现了这一点，并使用MSE损失和Adam优化器来优化这个误差。我们允许这个优化在每一帧上运行10次，我们发现这不会影响实时性能。

5 评估

我们系统地分析和评估了IMUPoser在不同数据集和条件下的有效性。

5.1 DIP-IMU 数据集

为了测试我们的模型在真实的（而不是合成的）IMU数据上的性能，我们使用了DIP-IMU [21]，这是一个基于IMU的MoCap数据集。虽然比我们用于训练的AMASS数据集小，但它提供了五个类别中各种不同的姿势和活动：上半身（举手、伸展、摆动等），下半身（抬腿、深蹲、弯腿等），互动（与日常物品互动的手势），自由式（仰卧起坐、出拳、踢腿等）和行走（步行、侧步等）。使用DIP-IMU的另一个好处是它已经在其他类似的工作中用于评估[21, 26, 64, 65]，可以进行直接比较。DIP-IMU使用了商用Xsens [63]基于IMU的系统来捕捉了10名参与者的数据。数据以60 Hz的采样率进行采样，总数据集大小约为90分钟。

5.2 IMUPoser 数据集

如上所述，DIP-IMU使用了专业级的XSens系统进行数据收集，其成本约为4000美元。所有的IMUs都是匹配的，提供了类似的噪声和跟踪性能。为了补充这个数据集，我们收集了我们自己的数据集，使用了与消费者设备等价的设备。

5.2.1 数据收集装置。我们的数据收集装置由两部智能手机（苹果iPhone 11 Pro）放置在左右前口袋，两块智能手表（苹果Watch Series 6）放置在左右手腕，以及一副戴在耳朵里的苹果AirPods Pro 组成（见图5）。我们的系统采样率配置为25赫兹，与AirPods的最大采样率相符。苹果手表和AirPods通过蓝牙与iPhone通信，两部iPhone将所有IMU数据传输到一台笔记本电脑进行数据处理和记录。尽管用户在数据捕捉期间携带了所有五个设备，但我们只使用其中的一部分设备进行姿势估计，如第3节所述。作为地面真实姿势，我们使用了Vicon Motion Capture System系统 [58]，配备了十二台MX40摄像机和四台T160摄像机，以120FPS的速度捕捉。我们使用Vicon Blade 3.2进行捕捉和数据导出，使用Vicon IQ 2.5进行数据清理。我们对Vicon数据进行降采样，并将其与我们收集的IMU数据流同步。在分析过程中，我们使用Mosh++ [32]将SMPL网格拟合到Vicon数据上。

5.2.2 设备校准。与商用基于IMU的运动捕捉系统（如XSens）不同，智能手机、智能手表和耳机无法提供IMU方向的常用（全局）参考框架。如果设备内含有磁力计，制造商通常会提供一种方法来获取设备相对于地球引力和磁场定向的全局方向。尽管我们在这项研究中使用的iPhone和苹果手表都含有磁力计，但我们发现它们的全局方向数据相当嘈杂。此外，苹果AirPods不含有磁力计，因此只提供相对于头部初始参考框架的方向。因此，我们选择使用Swift CoreMotion API [8]提供的XArbirtraryCorrectedZVertical参考框架。

在研究开始之前，我们将所有设备对齐到一个共同的参考框架，并记录它们在三秒窗口内的方向数值。这充当了校准数据，将所有设备带入相同的全局参考框架。在实际操作中，由于AirPods仅在参与者戴在耳朵上时才对IMU数据进行采样，因此常用的参考框架设置为该参考框架。与以前的研究一致[21, 65]，我们要求参与者做一个T形姿势三秒钟，以计算设备和其连接的骨骼关节之间的方向偏差。T形姿势充当了一个模板姿势，其中旋转是单位矩阵，因此每个关节的旋转都是已知的。这有助于校准穿戴设备在不同方向上的用户，例如，手机握在手中与手腕上戴表的用户。

5.2.3 数据收集过程。在我们的数据收集中，我们招募了10名参与者（其中5名为女性，5名为男性），平均年龄为22岁。研究大约持续了45分钟，参与者获得了20美元的补偿。我们要求参与者以他们感到最自然的方式佩戴和存放这五个设备。除了要求参与者穿带口袋的裤子之外，我们没有控制服装、口袋款式或手表佩戴位置的差异，以获取现实世界中的变化。为了获得基于Vicon的真实数据，我们在参与者身上放置了41个光学标记。为了保持标记的牢固，我们要求参与者把衬衫塞进裤子里，并在需要的地方提供了钩和环带。

受到之前的研究启发[6, 21]，我们采用了一种“障碍赛道”式的程序来收集数据。我们扩展了DIP-IMU数据集中的类别，并包括以下动作： • 上半身：右臂上举、左臂上举、双臂上举、右臂摆动、左臂摆动、双臂摆动、双臂横过胸前、双臂横过头后。 • 下半身：右腿抬高、左腿抬高、深蹲、左腿前跨、右腿前跨。 • 步行：直线行走、八字行走、圆圈行走、交叉腿的侧步、双腿触地的侧步。 • 自由式：开合跳、网球挥拍、交替挥拍、用主要腿踢、俯卧撑、篮球运球。 • 头部动作：上下晃动头部、左右晃动头部、从一侧的肩膀倾向另一侧的肩膀、绕圈移动头部。 • 互动：坐在椅子上滚动智能手机。 • 杂项：挥动右臂、挥动左臂、鼓掌、单腿跳跃（右腿）、单腿跳跃（左腿）、直线慢跑、圆圈慢跑。

这些上半身、下半身、步行、自由式、头部动作、互动和杂项的场景的平均持续时间分别为69.7、43.4、95.3、76.2、36.8、19.2和74.3秒，从而每位参与者的数据约为7分钟。所有动作都是连续的，数据也在参与者从一个类别过渡到另一个类别时进行捕获。

5.3 评估协议 为了与以前的研究进行比较，我们按照[21, 26, 64, 65]中详细描述的方法进行操作。具体来说，我们使用DIP-IMU的前八名参与者的数据作为训练数据，使用后两名参与者的数据进行测试。我们对我们的AMASS训练模型进行微调，使用将采样率降低到25 Hz的训练数据，以匹配我们的AMASS训练数据和我们的实时系统的性能。我们进一步在我们的IMUPoser数据集上测试这个模型，以帮助评估真实世界的准确性和性能。我们以在线方式评估这个模型。具体来说，我们提供一个滚动窗口，包括125个样本（5秒历史数据），重叠一个样本，模拟真实世界的使用。这些数据使用平均滤波器进行平滑，如第4.2节所述。我们使用不同的评估指标来分析不同的设备位置组合下的结果。此外，按照以前的做法[21, 64, 65]，我们使用以下评估指标来量化我们的全身姿势估计流程的性能： （1）关节旋转平均误差（Mean Per Joint Rotation Error，MPJRE）：度量所有关节的平均全局角度误差，以度（°）为单位。 （2）关节位置平均误差（Mean Per Joint Position Error，MPJPE）：度量所有估计关节与根关节（骨盆）对齐的平均欧几里得距离误差，以厘米（cm）为单位。 （3）关节顶点平均误差（Mean Per Joint Vertex Error，MPJVE）：度量估计的SMPL网格的所有顶点与根关节（骨盆）对齐的平均欧几里得距离误差，以厘米（cm）为单位。 （4）关节抖动平均误差（Jitter）：抖动度量预测运动的平均急变。较低的抖动值表示更平滑和更自然的运动。 对于大多数任务，我们使用网格误差（MPJVE）作为主要评估指标，因为它易于理解，并且作为与以前研究的比较的基准的实用工具。

6 结果

首先，我们将描述IMUPoser在不同设备位置组合下的准确性，然后将焦点转移到按身体部位的结果。最后，我们将通过与其他相关系统的比较来结束这一部分。

6.1 不同设备位置组合的准确性 为了简化结果的呈现，我们根据设备数量和它们的身体位置将24种可能的设备位置组合（如图3所示）分为12个超集。图6呈现了IMUPoser和DIP-IMU数据集的结果。请注意，我们的模型没有在IMUPoser数据集上进行微调。在所有设备组合中，我们发现在IMUPoser数据集上的MPJVE为14.1厘米，而在DIP-IMU数据集上为12.1厘米。当将两个数据集的结果进行平均时，用户只有一个设备的MPJVE为16.27厘米（标准差=9.93厘米），当第二个设备存在时，MPJVE为13.9厘米（标准差=8.36厘米）。不出所料，当有三个设备时，误差最低，即在所有可能的三个设备组合中，MPJVE为11.1厘米（标准差=6.51厘米）。

图7提供了不同设备位置组合下的示例网格预测。正如预期的那样，只有当耳机存在时才能准确估计头部方向。其他时候，头部会根据身体面对的方向回归到最自然的方向。全局身体方向在至少有两个设备存在时效果最好。最后，具有特定节奏的运动，如行走，在所有组合中都效果很好。类似地，具有对称肢体运动的活动，如跳跃，即使重要肢体没有传感器数据，也能够相当好地完成。另一方面，具有不相关肢体运动的活动除非肢体被连接，否则会失败。

Figure7：不同设备放置和组合的示例SMPL网格预测。红色表示每个顶点的误差，单位为米（范围从0到1米）。

6.2 不同身体区域的准确性 图8提供了IMUPoser和DIP-IMU数据集在不同身体区域的系统准确性分布。我们注意到，具有传感点的肢体的准确性总是高于没有传感点的肢体。例如，在两个数据集和右手有IMU的情况下，右臂的MPJVE为14.65厘米（右手=17.2厘米），而左臂的MPJVE为21.6厘米（左手=26.9厘米）。毫不奇怪，当特定身体区域的肢体都没有IMU数据时，误差最高。 同样不足为奇的是，当左右肢体都有IMU时，可以实现最低的误差。例如，只有一只手臂有IMU时，两只手臂的MPJVE为18厘米（右手=22.17厘米；左手=20.4厘米）。而当两只手臂都有IMU时，MPJVE为14.5厘米（右手=17.35厘米；左手=16.9厘米）。与这一趋势有些不同的是腿部。与手臂不同，腿部倾向于一起运动（行走时相位相反，跳跃等活动时相位相同）。这意味着即使腿部只有一个IMU，仍然可以非常有效地预测两只腿，而两个IMUs只会带来一些小的增益。从我们的结果来看，当IMU位于左口袋时，左腿的MPJVE为10.3厘米（左脚=15.65厘米），右腿的误差为10.4厘米（右脚=16厘米）。当两个IMU都存在时（即左口袋和右口袋），左腿和右腿的误差都只下降得很少，分别为10.05厘米和9.75厘米。 我们注意到误差会在运动链上积累（参见图7）。在所有条件下，末端效应器（左手、右手、左脚、右脚、头部）的平均误差在IMUPoser数据集和DIP-IMU数据集上分别为20.28厘米和17.29厘米（相对于不是末端效应器的关节的12.92厘米和11.23厘米）。

6.3 与以往研究的比较

根据我们的了解，以往的研究尚未调查过如何从装备IMU的稀疏的消费者级设备中获取全身姿势。表2提供了与以往关键研究的定量比较，所有这些研究都在相同的DIP-IMU数据集上进行评估[21]。

毫不奇怪，对于使用1到3个IMU的系统，我们的模型比使用6个传感器的模型（即每个肢体上都有IMU的模型）精度更低。然而，与DIP[21]和Transpose[65]相比，我们的MPJVE仅分别差3.2厘米和5.0厘米。有趣的是，我们的系统的抖动与以前的研究相符（1.9与Transpose的1.4相符）。总体而言，即使采用了贫瘠的传感配置，我们仍能够生成自然、逼真且平滑的姿势估计序列。在未来，我们希望结合基于物理的模型（如PIP）来进一步改进我们系统的姿势估计。

图 8：在DIP-IMU和IMUPoser数据集上对不同身体部位的总体准确性结果概述。

7 主动设备跟踪  
我们的姿势模型在运行之前需要的关键信息有：1) 用户身上有哪些设备？2) 这些设备位于身体的哪个位置？为此，我们创建了一个独立的软件，与我们的姿势模型并行运行。

7.1 实施 要确定设备位于身体的哪个位置，我们需要用户提供三方面的信息，我们设想这些信息可以在用户首次购买设备时进行收集。这三方面信息包括：1) 他们通常把手机放在哪个口袋？2) 他们通常用哪只手拿手机？3) 如果他们戴智能手表，它会戴在哪只手腕上？在进行了这些基本初始化之后，我们使用一系列自动启发式算法。

我们假设如果智能手机的屏幕是开启的并且IMU传感器报告了微小的运动，那么用户正在用手拿着智能手机。如果用户戴着智能手表，我们可以使用手表和手机之间的距离（由Apple的NINearbyObject API提供，该API使用UWB技术）来自动猜测用户的拿手机的手（参见图10）。如果用户没有戴智能手表，我们的系统会回退到用户在设置时指定的拿手机的手。如果智能手机的屏幕处于关闭状态并且红外线接近传感器被触发，我们假设手机在口袋中。如果用户佩戴智能手表，我们可以类似地使用UWB技术得出口袋的猜测。如果没有佩戴智能手表，我们会默认用户指定的口袋。

由于大多数用户会在一致的位置佩戴手表，智能手表的逻辑更加简单。如果智能手表与iPhone连接并且在移动，我们假设它戴在用户指定的手腕上。类似地，对于AirPods，如果它们与iPhone连接，我们知道它们戴在耳朵里。当AirPods放在充电盒中时，它们会进入休眠状态并停止传输IMU数据。但我们相信Apple未来可以修改AirPods的固件，使其能够在放在口袋里时继续传输IMU数据。

7.2 评估 为了初步评估我们的主动设备跟踪预测，我们进行了一项用户研究，共有7名参与者（其中5名男性，2名女性；平均年龄27.8岁；都是右撇子，偏好佩戴手表在左手腕）。该研究持续大约15分钟，参与者获得了5美元的报酬。为了初始化我们的系统，我们记录了参与者对前面部分提到的三个偏好问题的答案。然后，我们要求用户在15种不同的设备组合之间进行切换，顺序是随机的，这些组合见于图9。当某个设备不在请求的组合中时，它会被放在附近的桌子上。对于每个请求的设备-位置组合，我们要求参与者四处走动约10秒，然后短暂坐下，再次站起，最后返回到起始位置。在下一次试验开始之前，所需的设备会被分发给或收回参与者。在整个研究过程中，我们的主动设备跟踪过程一直在运行，实时预测哪些设备处于活动状态以及它们的位置。经过培训的实验员进行了这项研究，标记了每个设备组合试验的开始和结束，同时记录了地面真实标签。

在所有参与者和所有数据实例中，由于耳机和智能手表的位置是已知的，佩戴与否也非常容易检测，所以耳机和智能手表的跟踪精度为100%。智能手机的跟踪是最具挑战性的，因为有五种可能的状态（不在场、左口袋、右口袋、左手、右手）。我们发现智能手机跟踪的逐实例准确率为90.8%。

Figure9：作为我们主动设备跟踪研究的一部分测试的设备组合。蓝色表示存在于集合中。

8 实时实现

为了帮助展示我们方法的即将可行性，我们创建了IMUPoser的实时实现，可以在我们的视频图中看到。它由两个主要的进程共同工作组成。首先是主动设备跟踪，它监视哪些设备可用于提供IMU数据，并预测它们在身体上的位置。其次是我们的姿势模型，它接收位置推断和IMU数据。

8.1 概念验证设备生态系统

作为一个概念验证的实现，我们使用了一台Apple iPhone 11 Pro、一块Apple Watch Series 6和一对AirPods Pro。Apple提供了一个成熟的跨设备API，允许这些设备交换数据。每个设备以不同的速率报告6自由度（6DOF）的IMU数据，其中AirPods的速度最慢，大约为25帧每秒。我们还注意到，尽管AirPods是成对出售的，但它们将各自的IMU数据融合成一个单独的6DOF头部估计。

8.2 输出

作为一个概念验证，我们使用了一部iPhone，可选择连接到Apple Watch和AirPods。iPhone将所有可用的IMU数据传输回一台运行我们的主动设备跟踪和姿势估计流程的MacBook Air（2021），平均推理时间为26.8毫秒。我们相信通过额外的工程努力，我们的模型可以在移动电话上运行。无论模型运行在哪里，它的帧率被限制为25帧每秒，即我们最慢的IMU（AirPods）的报告频率。在运行我们的系统之前，我们必须执行与数据收集中相同的校准过程（请参见第5.2.2节）。对于实时输出，我们可视化SMPL网格。

Figure10：在不同设备组合中的主动设备跟踪。活动设备以白色圆圈进行突出显示以进行说明，前景的笔记本电脑显示了实时跟踪结果。

10 ****限制与未来工作****

尽管IMUPoser允许通过最少的用户设备仪器来进行全身姿势估计，但它像任何其他技术方法一样有其优点和缺点。虽然IMUPoser可以获取有关它没有直接传感器数据的肢体姿势的见解，但需要注意，这种姿势只是一个近似结果。在那些仪器连接的关节的运动完全独立于未连接的关节的情况下，IMUPoser往往趋向于回归到平均姿势。IMUPoser可以通过使用相应的SMPL网格顶点进行训练来支持新关节位置的加入。因此，在未来，IMUPoser有可能支持和跟踪新的设备放置位置，例如将手机放在后口袋、外套口袋、臂带等等。通过集成额外的消费者设备（例如智能鞋、眼镜、戒指）可以提高系统的保真度。这将有助于扩展IMUPoser支持的姿势范围，使其能够跟踪自行车、皮划艇、滑雪等动态活动。

与Transpose [65]和PIP [64]不同，当前的IMUPoser实现不会预测全局根部平移。在未来，使用更好的学习方法和在可用时使用多模态线索（例如来自智能手机的视觉里程计 [6]）可以帮助预测平移。类似的，通过包括上下文线索，如正在执行的活动 [7]或用户位置，可以提高系统的总体准确性。

我们系统的另一个限制是主动设备跟踪。目前，这是一个基本的概念验证实现，需要进一步完善才能用于消费者使用。此外，所有设备需要在同一生态系统中（例如，Apple）才能有效工作。在未来，使用行业共同标准来连接和网络化不同的消费者设备之间可以帮助缓解这个问题。

最后，我们设想IMUPoser在用户携带的性能最强大的设备上执行。在大多数情况下，这将是智能手机，但未来情况可能会有所不同。例如，今天用户可以只带着智能手表和无线耳机去跑步，而不带手机。在不久的将来，似乎有可能会有像AirPod一样的设备，可以独立运行（例如，可听设备）。

****结论****

本文介绍了IMUPoser - 一种使用消费者设备（如手机、智能手表和无线耳机）中的IMU进行实时全身姿势估计的系统。我们的系统必须自动跟踪可用的设备以及它们当前在身体上的位置，并使用流媒体IMU数据来估计姿势。我们的评估表明，IMUPoser能够应对消费者IMUs的嘈杂信号，并生成自然且时间连贯的姿势估计，即使只有一个设备。这为全身姿势应用提供了新颖而有趣的可能性，而无需额外的用户仪器。

IMUPoser是一项创新的工作，它提供了以前不可思议的全身姿势估计方法，通过充分利用普及的消费者设备中的传感器。这项研究不仅为虚拟现实、增强现实等领域提供了潜在的应用，还为将来的研究提供了广阔的可能性，包括更多设备的支持、姿势精度的提高和多样性活动的跟踪。此外，研究人员还提供了数据集、体系结构、训练模型和可视化工具的开源，以鼓励其他研究者和从业者构建和改进IMUPoser系统。

当然，以下是关于这篇论文主要内容的分点概括：

****目标和动机****：

* + 论文的目标是实现全身姿势估计。
  + 动机在于探索使用消费者设备中的IMUs来实现这一目标，而不需要额外的传感器或装备。

****使用的设备****：

* + 论文讨论了使用智能手机、智能手表和无线耳机等设备中的IMU传感器进行姿势估计。
  + 不同设备有不同的潜在身体位置，如手机可以放在口袋里或手中，智能手表可以佩戴在左手或右手，等等。

****挑战****：

* + 论文提到了使用低成本IMU传感器产生的嘈杂数据。
  + 用户身体上的传感点分布稀疏，可能随时间变化。

****系统概述****：

* + IMUPoser系统的目标是处理这些挑战，通过处理IMU数据估计用户的全身姿势。
  + 该系统的核心思想是接收可用的IMU数据，并产生最佳猜测的用户姿势。

****应用前景****：

* + 论文提到了这一技术的广泛应用前景，包括在健身、移动游戏、虚拟助手和康复等领域中的潜在用途。

这些要点简要总结了论文的主要内容和目标。