PresSim：利用基于物理的三维仿真，从单目视频生成动态地面压力轮廓的端到端框架

Lala Shakti Swarup Ray*†、Bo Zhou*†、Sungho Suh*† 和 Paul Lukowicz*†
*德国人工智能研究中心（DFKI），德国凯泽斯劳滕 67663
†计算机科学系，凯泽斯劳滕-兰道莱茵兰-普法尔茨技术大学，德国凯泽斯劳滕 67663 电子邮箱：
{lala shakti swarup.ray, bo.zhou, sungho.suh, paul.lukowicz}@dfki.de

arXiv:2302.00391v1 [cs.CV] 1 Feb 2023

摘要—人体施加的地面压力是无感普适感知中进行人类活动识别（HAR）的重要信息来源。尽管利用压力传感器收集数据以开发 HAR 解决方案需要投入大量资源和精力，我们提出了一种新颖的端到端框架 PresSim，可根据人体活动视频合成传感器数据，从而大幅减少这类投入。PresSim 采用三阶段流程：首先，利用计算机视觉架构从视频中提取 3D 活动信息；然后，根据 3D 活动信息和包含重力的物理仿真，模拟地面网格形变分布；最后，利用深度学习模型生成模拟的压力传感器数据。我们探索了两种获取 3D 活动信息的方法：结合网格重定向的逆运动学，以及体积姿态与形状估计。我们通过一套实验装置验证了 PresSim：使用单目相机提供输入，并使用压力感应健身垫（空间分辨率为 80×28）提供传感器真实值；九名参与者完成了一组预先设定的瑜伽序列。基于压力形状，将合成压力图与压力传感器真实值进行比较，并通过二值化压力图的 R平方值 0.948，以及地面接触区域内激活传感节点压力值的精确度（校正R平方值为 0.811）进行量化。我们公开了时长九小时的数据集和源代码，以推动更广泛的研究。

索引词—**压力传感器、姿态估计、基于物理的三维仿真、时序 CNN、活动识别**

I. 引言

人体动力学和运动学对于正确理解人类活动场景至关重要。由于人体质量分布与骨骼运动的共同作用，人体施加于周围环境的压力是人体动力学的重要组成部分之一。它具有多种临床应用，包括检测人体的不同类型疾病（如偏瘫）、诊断损伤 [1], 术后康复 [2],分析人体运动[3], 个人健身 [4], [5],以及生物力学研究 [6]。

使用可穿戴设备和外部压力传感器等物理触觉传感设备测量身体施加的地面压力，可能耗时、昂贵且容易发生故障。一种解决方案是利用神经网络，从视频等相对容易采集的信息来源中提取人体压力。这种方法具有一项独特优势，尤其适用于长期监测以及需要大面积作业的任务，相较于使用物理压力传感器更具优势。然而，结合人体姿势与从视频中提取的形状来估算地面压力数据颇具挑战。以往的一些研究利用深度神经网络，根据从RGB视频中提取的数据估算人体施加的地面压力 [7] [8] ，但仅适用于特定姿势和身体部位。

我们通过结合身体形状与姿势，并使用基于物理的三维仿真，提出了一种解决这一复杂问题的新方法，从而创建一种同时包含身体形状和位置信息的中间数据类型。在第一种方法中，我们采用基于运动学的姿态估计模型提取三维运动学信息，然后使用逆运动学创建动画序列。随后，将其重新定位到具有等效质量和相同骨骼结构的人形网格上，并利用基于物理的三维仿真模拟变形轮廓。之后，我们使用仿真数据和提取的三维运动学信息训练时序神经网络回归器，以生成合成压力图。第二种方法使用基于 SMPL+H [9]模型的体积模型进行姿态估计，以获得更准确的变形轮廓。

我们的工作贡献如下：

- 提出一种新方法，结合基于物理的仿真与三维卷积神经网络，从野外单目视频中推断人体施加的地面压力图，如图 1 所示。
- 构建了一套流程，结合逆运动学、网格重定向或体积姿势与形状估计，以及三维基于物理的仿真，从视频中模拟中间变形轮廓。

II. 相关工作

A. 姿态估计

在计算机视觉中，姿态估计用于根据图像或视频预测一个人的关节位置。使用多相机动作捕捉系统预测三维姿态 [10], [11] 存在问题，因为所有摄像机都必须预先校准，而且摄像机之间必须同步，才能准确预测姿态数据。另一种方法是使用单目姿态估计模型，该模型利用 Mediapipe [12], 、Detectron [13],和 Openpose 等二维姿态估计模型的输出

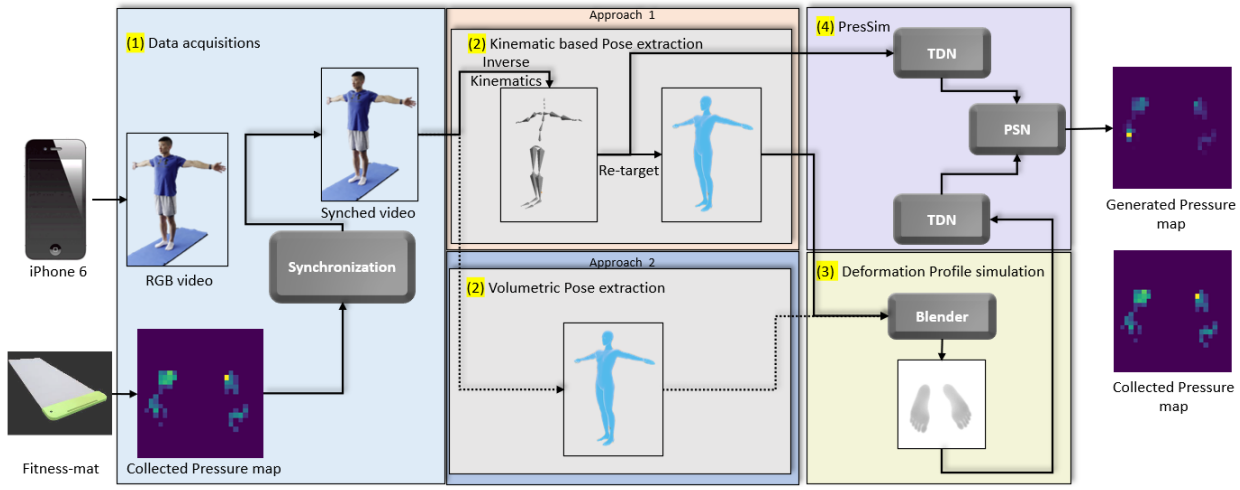

图 1. 所提出的端到端流程示意图，包括：1. 数据采集；2. 姿态提取；3. 变形轮廓模拟（通过网格重定向实现逆运动学，或进行体积姿态与形状估计）；以及 4. 压力图生成。

[14] 来重建三维空间中的关节位置，VideoPose3D [15] 和 Blazepose [16]采用的就是这种方法。我们使用 Detectron2 和 VideoPose3D 进行深度估计，以便在三维空间中重建人体运动学关节，这与 Scott 等人 [7]的方法相同。对于体积姿态估计，我们使用 easymocap [17]，该方法可根据单目视频估计 SMPL+H 的身体形状和姿势。

B. 物理仿真

基于物理的仿真旨在创建自然界中的事件或物理现象，并通过定义仿真开始前自然界的状态，预测现实世界中的物体会对此作出怎样的反应 [18]。随后观察此类事件产生的影响，以打造更稳健的设计。它被用于航空和汽车工业中的原型测试、布料仿真 [19],、物体变形 [20],以及人体生物力学研究 [21]。我们使用 Blender 中的 3D 物理游戏引擎 Eevee [22],，通过刚体仿真和动态绘制变形，渲染人体模型与地面之间的接触点及压力；这一做法与 Clever 等人 [8] 的方法类似。

III. 数据收集与变形仿真

A. 数据采集

为了训练模型，我们使用 Sensing.Tex 的 Fitness-mat 收集了一个大规模双模态数据集，其中包含同步的全身单目视频和压力图序列。该设备配有一个 80×28 传感器网格，感测区域为$560\times1680\ mm$，独立传感器区域为 $12\times16\ mm$，测量范围为 0-5000 毫米汞柱（毫米汞柱）。硬件包括一张由热塑性聚烯烃（TPE）聚合物制成的非弹性、防滑瑜伽垫，其工作湿度范围为 5%–70%，工作范围为 $15°C$– $45\ °C$。视频使用 iPhone 6 的 800 万像素（3840×2160 像素）摄像头拍摄，分辨率为全高清（$1920\times1080p$），帧率为每秒 30 帧。我们考虑了

主题	体重(kg)	身高(cm)	性别	帧
1	74.3	178	Male	40788
2	76.3	183	Male	40428
3	72.2	178	Male	43895
4	94.7	172	Male	43630
5	60.2	157.5	女性	27932
6	57.7	162	女性	30359
7	57.2	160	女性	31600
8	52.3	162	Male	31608
9	57.3	159	女性	31023
总计	-	-	-	321263

表 I 数据统计，包括受试者体重（$kilogram$）、身高（$centimeter$）和性别。

在构建数据集时，我们纳入了九名年龄、性别和身体特征各异的参与者。我们设计了一套连续的瑜伽姿势序列，包含 27 个姿势，涉及任何人都能完成的复杂身体动作，不受性别或年龄限制。每套序列持续约 20 分钟，参与者尝试通过视觉和音频输入跟随练习流程，因此数据集总时长约为 9 小时。表 I 详细列出了每个数据采集环节的信息。

压力数据被转换为大小为 $f\times80\times28$ 的数组，其中 f 表示帧的总数，$80\times28\in N^2$ 表示传感器网格大小。输出传感器数据的范围为 0 至 5000 毫米汞柱。采集的视频数据帧率为 30 fps，而压力图序列的帧率约为 10 fps，两类数据均通过时间戳进行手动同步。对于我们的第一种方法，我们使用了 Detectron2 中的 Keypoint R-CNN，这是一种在 COCO 数据集上训练的二维姿态估计模型 [23]，用于估计每一帧中的关键点，共输出 17 个关键点。为了根据关键点位置序列重建三维运动学，我们使用了 VideoPose3D。该模型在预测二维关键点的深度时考虑了时序数据，并且在从活动视频中预测姿势方面优于其他模型 [24]。VideoPose3D 在 Human3.6M 数据集上进行训练 [25]。输出

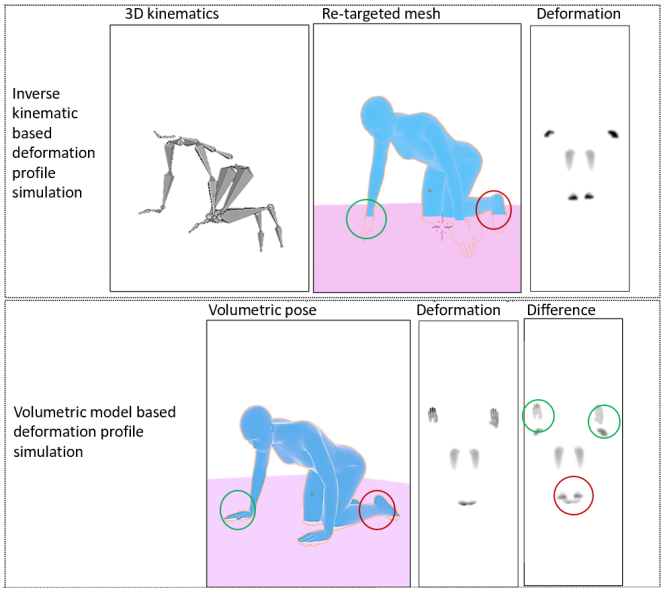

图2. 基于逆运动学方法和体积方法的变形轮廓模拟：提取的三维姿态、重新定向的三维网格，以及三维模拟中的相应变形地面平面；并展示基于这两种方法模拟的变形之间的差异。

身体运动学是一个大小为 $f \times 17 \times 3 \in \mathbb{R}^3$ 的数组，其中 f 是帧的总数，即 $Seconds \times fps$（10）。每一帧包含 17 个关节及其三维（X、Y、Z）坐标。

对于第二种方法，我们使用 EasyMocap，这是一种体积姿态估计模型，可同时估计三维姿态和身体形状 [26]。与前一种方法不同，该模型通过考虑 $f \times 25 \times 3 \in \mathbb{R}^3$ 个关键点而非$f \times 17 \times 3 \in \mathbb{R}^3$来跟踪和估计手部和腿部。与第一种方法相比，模拟生成的中间数据与采集的压力图更为相似。

B. 变形轮廓模拟

我们提出了一种利用物理仿真的变形轮廓模拟流水线，可使用采集的三维姿态数据和估计得到的三维人形网格，自动生成包含大量变形轮廓的数据集。采集的压力图 $P_{80\times28} \in \{0-5000\}$ 与模拟得到的变形轮廓 $P_{80\times28} \in \{0-255\}$ 之间的相关性可用以下数学表达式表示：

$$\begin{bmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,28} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,228} \\ \vdots & \vdots & \ddots & \vdots \\ p_{80,1} & p_{80,2} & \cdots & p_{80,28n} \end{bmatrix} = \alpha \begin{bmatrix} \bar{p}_{1,1} & \bar{p}_{1,2} & \cdots & \bar{p}_{1,28} \\ \bar{p}_{2,1} & \bar{p}_{2,2} & \cdots & \bar{p}_{2,228} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{p}_{80,1} & \bar{p}_{80,2} & \cdots & \bar{p}_{80,28n} \end{bmatrix} \quad (1)$$

其中，α 是某一帧中所有压力点共用的乘法因子常数，但该因子会随帧变化，需要通过神经网络进行估计。

在第一种方法中，我们使用逆运动学将提取的三维运动学数组转换为 Bio Vision Hierarchy（BVH） [27] 动作捕捉格式。我们使用 Blender[22]，（一款开源三维软件）完成了大部分三维网格生成和物理仿真工作。我们创建了一个与受试者身体相似的三维人形网格，并为其绑定了一个包含 17 个关节的 CMU骨架，该骨架与我们的动作捕捉文件中的骨架相同；随后，我们使用人形网格对动作捕捉文件进行重新定向。该人形网格还保留了骨骼信息，例如各关节之间的相对距离。使用生成的动作捕捉文件对其重新定向时，这些信息也会保留在绑定网格中。接着，我们将人形网格转换为刚体，并为其赋予与数据收集所用参与者体重相当的质量。我们在人形网格下方创建了一个可变形平面，并将其转换为动态绘制画布，同时让人形网格充当动态画笔。根据人形网格与平面之间的接触点，平面会发生形变，直至达到一定阈值；接触点消失后，平面则会恢复原状。为确保仿真的变形轮廓不会因意外运动而产生噪声，我们限制了刚体在 X、Y 方向上的移动以及各个方向上的旋转，使刚体的质量成为影响平面对人体反应的唯一因素。我们使用合成节点，通过 Blender 内置摄像机创建了一个尺寸为 (80×28) 像素的自定义渲染视点，其设置与 Fitness-mat 的传感器阵列相同。

渲染视点捕获以变形轮廓形式呈现的渲染结果。根据平面各顶点与渲染视点之间的距离，为其赋予一个介于 255 到 0 之间的数值，以模拟不同的压力点。随后，使用 Eevee 渲染引擎将每一帧渲染为灰度 PNG 格式。之后，使用 OpenCV [28]将图像序列转换为大小为 $f \times 80 \times 28$ 的数组。

对于第二种方法，我们使用 SMPLblender 插件将提取出的体积 SMPL+H 姿势导入 Blender。随后，我们按照之前完全相同的流程模拟变形轮廓。

图2展示了基于我们两种方法，从三维姿态输出生成动作捕捉文件，再到模拟可变形压力轮廓的各个中间步骤。与逆运动学方法相比，基于体积模型的方法模拟变形轮廓的主要优势在于，动画网格考虑了肢体末端的形状和位置，以及人体的整体形状。

IV. PresSim：压力图合成

A. 问题表述

我们的模型旨在利用一种中间数据类型，直接从视频合成压力传感器数据；这种数据类型可通过基于物理的仿真，由从视频中提取的姿势进行仿真生成。使用深度学习技术从三维运动学合成压力图颇具挑战，因为压力图很大程度上取决于身体形状。三维运动学不包含身体形状信息，而身体形状与三维姿态相结合，对于生成准确的地面压力图至关重要。我们使用基于物理的仿真系统，根据从视频中提取的 3D 活动信息生成这种中间数据类型。我们训练网络学习运动学和仿真得到的变形轮廓与相应身体压力图分布之间的关联。

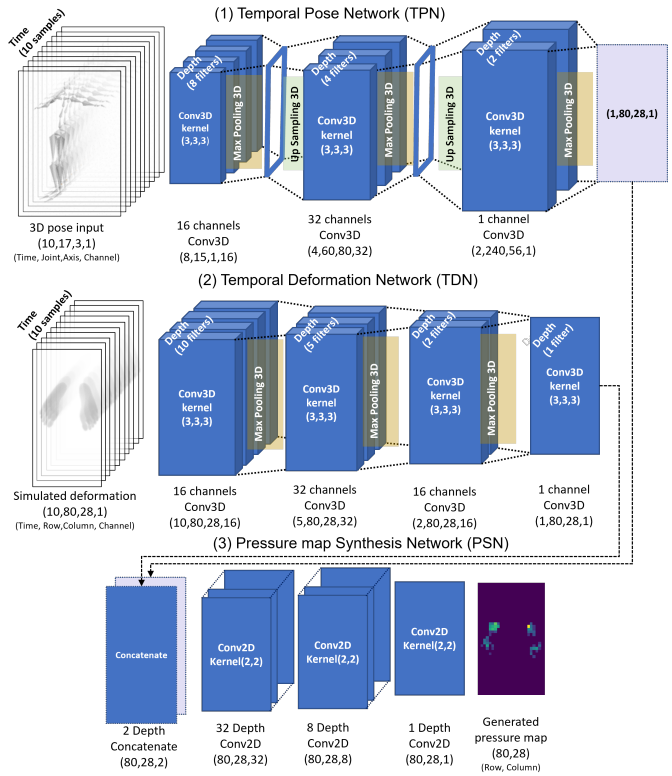

图 3. PresSim总体架构。PresSim由三个网络组成：时序姿态网络（TPN）、时序形变网络（TDN）和压力图合成网络（PSN）

我们采用时序三维卷积神经网络回归器，利用滑动窗口，综合考虑时序姿态序列中的十帧（约 1 秒）及其对应的变形轮廓，生成单帧压力轮廓。深度学习模型的输入是从 $f \times 17 \times 3 \in \mathbb{R}^3$ array 中取出的连续十帧姿态数据，以及从 $f \times 80 \times 28 \in \mathbb{N}^2$ array 中取出的十帧仿真变形轮廓，这些数据由两个同步的滑动窗口选取。真实值是从 $(f-10) \times 80 \times 28 \in \mathbb{N}^2$ array 采集的压力图中取出的一帧。

*B. PresSim*架构

PresSim由三个神经网络组成。两个三维时序卷积神经网络TPN和TDN分别根据三维运动学和二维模拟形变轮廓估计二维特征图。我们使用一个称为PSN的二维卷积神经网络，将前两个网络的输出拼接起来，以合成二维压力图。图3展示了这三个网络及其输入和输出。

a) 时序姿态网络（*TPN*）：时序姿态网络（TPN）是一种三维时序CNN，以十帧大小为 $10 \times 17 \times 3$ 的提取姿势作为输入，并输出大小为 80×28 的二维特征图。TPN的主要用途是预测沿水平轴的平均压力值，随后用于生成合成压力图。我们使用ADAM优化器和L2损失训练网络。该网络估计标记为二维传感器位置 $\{\bar{p}_1, \bar{p}_2 ..., \bar{p}_N\}$ 的压力值，其中每个 $\bar{p}_j \in \mathbb{N}^2$ 表示传感器j的二维压力估计值。我们根据每个关节上欧几里得误差的平方值计算损失：

$$L_{TPN} = \sum_{j=1}^{n} (p_n - \bar{p}_n)^2 \tag{2}$$

我们使用上采样层增大输入维度，使其与采集的压力图尺寸相匹配。我们实现了一个滑动窗口，将三维数组中的连续十帧作为一个数据点输入，用于训练模型。TPN包含十二层，共有15,361个参数，其中15,265个可训练，96个不可训练。

b) 时序形变网络 *(TDN)*：时序形变网络 (TDN) 是另一种三维时序神经网络回归器，它以大小为 $10 \times 80 \times 28$ 的模拟变形轮廓的十帧作为输入，并输出一个大小为 80×28 的二维数组。TDN 的主要用途是在二维特征图中准确预测形状以及预测形状上的相对压力值。与 TPN 一样，我们使用 ADAM优化器和 L2损失训练网络。网络输出带有标签“二维传感器位置” $\{\bar{q}_1, \bar{q}_2 ..., \bar{q}_N\}$ 的压力值估计，其中每个 $\bar{q}_j \in \mathbb{N}^2$ 表示传感器 j 的二维压力估计值。我们根据每个关节上欧几里得误差的平方值计算损失：

$$L_{TDN} = \sum_{j=1}^{n} (q_n - \bar{q}_n)^2 \tag{3}$$

我们创建了一个与 TPN 中使用的滑动窗口时钟相同的滑动窗口，从三维数组中提取连续十帧作为单个数据点，以训练模型。TDN 共包含十一层，有 40,563 个参数，其中 40,371 个可训练，192 个不可训练。

c) 压力图合成网络（*PSN*）：TPN 的输出（尺寸为 80×28 的二维特征图）与 TDN 的输出（尺寸相同的二维特征图），通过压力图合成网络（PSN）——一种二维卷积神经网络——进行拼接。与之前的网络类似，我们使用 ADAM优化器和 L2损失来训练网络。网络输出标注在二维传感器位置 $\{(\alpha\bar{p}_1 + \beta\bar{q}_1), (\alpha\bar{p}_2 + \beta\bar{q}_2) ..... (\alpha\bar{p}_N + \beta\bar{q}_N)\}$ 处的压力值估计，其中每个 $\bar{p}_j, \bar{q}_j \in \mathbb{N}^2$ 分别表示根据公式（2）和公式（3）得出的传感器 j 的二维压力估计值。α 和 β 是为使总分最小化而优化的可变权重。我们根据每个关节上欧几里得误差的平方计算损失：

$$L_{PSN} = \sum_{j=1}^{n} \{(\alpha|p_n - \bar{p}_n| + \beta|q_n - \bar{q}_n|\}^2 \tag{4}$$

此外，还使用三个二维卷积层，将拼接后的数组拟合为所采集压力图的维度。PSN 共包含 23,873 个参数（其中 23,777 个可训练）。

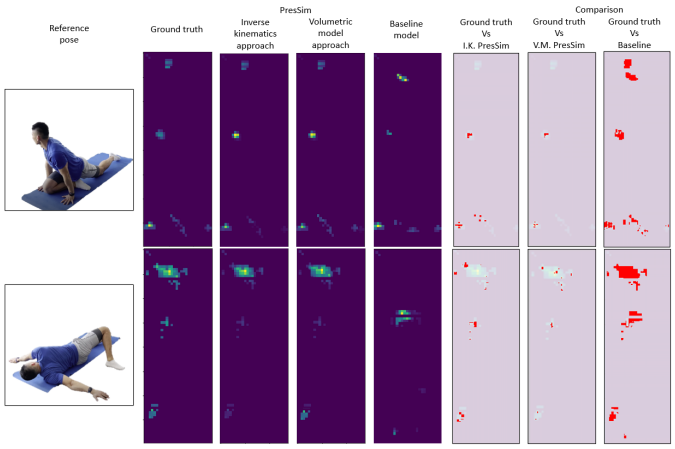

图4. 参考视频帧、使用 Fitness-mat 采集的真实值、基于逆运动学和体积模型方法使用 PresSim 合成的压力图、使用基线模型合成的压力图，以及忽略 ±10 毫米汞柱的不同压力值的所有点。

结合这两个模型的结果，比仅使用一种数据类型生成的合成压力图更加准确。

V. 实验结果

A. 实现细节

我们的网络使用 TensorFlow Keras 2.9.1 实现。我们通过最小化均方误差损失来训练模型，学习率为 $1e^{-4}$，批量大小为 128，并以平均绝对误差作为评估指标。TPN 和 TDN 中均使用了比率为 30% 的 Dropout 层。我们使用神经网络通过 $relu$ 激活函数回归置信度。对于 PSN，Concat 层将前两个模型的输出合并，随后使用三个 Connv2D 层来拟合压力图数组的维度。我们的模型在一个配备 128GB 内存、四个处理器和两块 Nvidia RTX A6000 图形内存的集群上训练。

B. 评估指标

我们使用平均绝对误差作为评估 PresSim 的指标。对于用于训练 PresSim 的数据集，数据范围为 0 到 5000，其中平均绝对误差表示与实际值的偏差。为了了解合成压力图与采集的传感器数据之间的一致性，我们使用决定系数（R^2）这一误差指标来评估模型性能，其取值范围为 0 到 1。

为了比较 PresSim $P_{80\times28} \in \{0-5000\}$ 合成的压力分布形状与地面真实值$P_{80\times28} \in \{0-5000\}$，我们将这两个数组二值化为 $P_{80\times28} \in [0,1]$ 和 $P_{80\times28} \in [0,1]$，其中所有大于零的值均为 1。

为了真实地比较每个合成压力节点与真实值之间的精度，我们仅在接触点处计算R^2。我们通过索引数组中所有压力值非零的点来创建列表，并将其用作掩码，从合成压力图和真实值中获取对应的压力数据$P_{80\times28}\forall\bar{p}_{i,\ j} \geq 1 \cap P_{80\times28}\forall p_{i,\ j} \geq 1$。随后，我们计算均方根

	平均分		
模型	基线模型	I.K. PresSim	体积 PresSim
MAE 掩膜均方根偏差	19.7 7.44	10.6 3.54	10.4 3.52
校正后的 R^2	0.407	0.800	0.811
二值化 R^2	0.659	0.909	0.948

表 II 基线模型与 PresSim 的平均估计分数对比表；两者分别使用通过逆运动学和体积模型模拟的变形轮廓进行训练。

差，并使用该均方根偏差值计算修正后的 R^2 分数。图4展示了我们的基线模型与 PresSim 在使用两种方法模拟的数据上训练后，真实值与合成压力图之间的差异，采用的是掩码均方根偏差。

为了进行比较，我们训练了一个基线模型，其结构与时序姿态网络相同，并额外添加了若干全连接层；该模型仅使用从视频中提取的三维运动学数据。

C. 定性评估

在平均绝对误差方面，我们的基线模型平均得分为 19.7，而使用基于逆运动学的变形轮廓训练的 PresSim 平均绝对误差得分为 10.6，基于体积模型的 PresSim 平均得分为 10.4。使用二值化压力图计算的 R^2，基线模型的平均值为 0.659，而基于逆运动学的 PresSim 方法和基于体积模型的 PresSim 方法的平均值分别为 0.909 和 0.948。通过掩码 - 均方根偏差计算的校正后 R^2，在整个数据集上，我们的基线模型平均值为 0.407。基于逆运动学的 PresSim 方法平均值为 0.8，而基于体积模型的 PresSim 方法平均值为 0.811。与基线模型相比，这两种模型均有显著改进。基于体积模型的 PresSim 略有提升，这是因为变形轮廓生成有所改进，且仿真流程中缺少逆运动学，因此仿真误差并不显著。表 II 列出了各模型的所有估计平均分数。

D. 错误案例

我们遇到的一个错误案例是，四肢与地面之间的接触点完全错误。这种情况并非偶发，而是在整个预测序列中持续存在；在不同的数据采集会话中，受试者执行一组指定的瑜伽姿势时，都能观察到这一现象。任何包含身体部位大幅扭转或弯曲的瑜伽序列，都会导致人形绑定序列不准确，无法在正确模拟形变的同时预测地面接触。我们推测，最可能的原因是使用逆运动学将姿势转换为动画序列，因为我们通过体积方法训练的 PresSim 模型中并未采用这种方法。

另一种失败案例是，PresSim 能够正确预测身体与地面接触点形状的大部分部分。然而，大多数压力

点的数值有所不同，同时合成压力图中缺少一些肢体印迹。由于这种情况在预测序列中零星出现，我们推测，某一帧中某些关节的姿态估计不准确，会导致这类错误；在采用我们任一方法训练的 PresSim 模型中都能观察到这种错误。可以使用更稳健的姿态估计模型来解决，例如多摄像头姿态估计，或从两个角度进行单目姿态估计，以合成形变。

VI. 结论

总之，我们提出了一种框架，结合视频到三维活动、物理仿真和深度学习数据转换，能够根据人体活动的单目视频合成动态压力分布的传感器数据。我们通过比较 PresSim 生成的地面压力数据、基线模型以及使用智能健身垫采集的真实传感器数据，对该框架进行了评估。为比较合成数据与真实值的形状，在整个数据集的二值化压力图上计算的 R^2 分数为 0.948，表明两者高度相似。我们进一步证明，在单个敏感节点层面，使用仅考虑身体与垫子之间接触点的掩码后，真实值与合成压力图之间修正后的 R^2 分数平均为 0.811。本研究提出了一种不同于传统压力传感设备的感知方式，为不受当代压力传感设备局限影响的人体动力学估计以及活动识别开辟了新机遇，并有望应用于智能家居、医疗保健和康复领域。

致谢

本文所报告的研究得到了德国联邦教育与研究部在 VidGenSense 项目（01IW21003）中的资助。

参考文献

[1] R. B. Brund、S. Rasmussen、R. O. Nielsen、U. G. Kersting、U. Laessoe 和 M. Voigt，《足内侧鞋地压力与特定跑步损伤：一项为期一年的前瞻性队列研究》，《体育科学与医学杂志》，第 20 卷，第 9 期，第 830–834 页，2017 年。[2] B. Hwang 和 D. Jeon，《利用力矩和鞋垫压力传感器估算地面步态康复机器人的用户肌肉力矩》，《国际控制、自动化与系统杂志》，第 16 卷，第 1 期，第 275–283 页，2018 年。[3] Z. Taha、M. S. Norman、S. F. S. Omar 和 E. Suwarganda，《用于模拟在地面自然站立的人足模型有限元分析》，《工程学会议论文集》，第 147 卷，第 240–245 页，2016 年。[4] J. Galarza 和 D. I. Caruntu，《额外负重对人体深蹲运动稳定性的影响：地面反作用力与压力中心》，载于动态系统与控制会议，第 84270 卷。美国机械工程师学会，2020 年，第 V001T06A001 页。[5] J. WASIK、W. J. CYNARSKI、D. SZYMCZYK、A. M. VENCESBRITO、G. KOROBEYNIKOV 和 T. ZWIERKO，《空手道运动员进行逆突（出拳）时足部对地压力的变化：案例研究》，《体育科学趋势》，第 26 卷，第 4 期，第 153–156 页，2019 年。[6] M. Lorenzini、W. Kim、E. De Momi 和 A. Ajoudani，《一种协同方法，用于实时估算人体足部地面反作用力和压力中心，并应用于人机协作》，《*IEEE*机器人与自动化快报》，第 3 卷，第 4 期，第 3654–3661 页，2018 年。

[7] J. Scott、B. Ravichandran、C. Funk、R. T. Collins 和 Y. Liu，《从图像到稳定性：从人体姿势中学习动力学》，载于欧洲计算机视觉会议。Springer，2020 年，第 536–554 页。[8] H. M. Clever、P. Grady、G. Turk 和 C. C. Kemp，《Bodypressure——从深度图像推断人体姿势和接触压力》，arXiv 预印本 arXiv:2105.09936，2021 年。[9] G. Pavlakos、V. Choutas、N. Ghorbani、T. Bolkart、A. A. Osman、D. Tzionas 和 M. J. Black，《富有表现力的人体捕捉：从单张图像中获取三维手部、面部和身体》，载于 IEEE/CVF 计算机视觉与模式识别会议论文集，2019 年，第 10 975–10 985 页。[10] N. Nakano、T. Sakura、K. Ueda、L. Omura、A. Kimura、Y. Iino、S. Fukashiro 和 S. Yoshioka，《使用多台摄像机评估基于 Openpose 的三维无标记动作捕捉精度》，《Frontiers in Sports and Active Living》，第 2 卷，第 50 页，2020 年。[11] J. S. Matthis 和 A. Cherian，《FreeMoCap：免费、开源的无标记动作捕捉系统》，2022 年。[在线]。网址：https: //github.com/freemocap/freemocap/ [12] A. K. Singh、V. A. Kumbhare 和 K. Arthi，《使用 Mediapipe 实时检测和识别人体姿势》，载于国际软计算与信号处理会议。Springer，2021 年，第 145–154 页。[13] Y. Wu、A. Kirillov、F. Massa、W.-Y. Lo 和 R. Girshick，《Detectron2》，https://github.com/facebookresearch/detectron2，2019 年。[14] D. Osokin，《CPU 上的实时二维多人姿势估计：轻量级 OpenPose》，arXiv 预印本 arXiv:1811.12004，2018 年。[15] D. Pavllo、C. Feichtenhofer、D. Grangier 和 M. Auli，《通过时间卷积和半监督训练进行视频三维人体姿势估计》，载于计算机视觉与模式识别会议（CVPR），2019 年。[16] I. Grishchenko、V. Bazarevsky、A. Zanfir、E. G. Bazavan、M. Zanfir、R. Yee、K. Raveendran、M. Zhdanovich、M. Grundmann 和 C. Smin- chisescu，《Blazepose ghum holistic：实时三维人体关键点与姿势估计》，arXiv 预印本 arXiv:2206.11678，2022 年。[17] J. Dong、Q. Fang、W. Jiang、Y. Yang、H. Bao 和 X. Zhou，《基于多视角的快速稳健三维多人姿势估计与跟踪》，载于 T-PAMI，2021 年。[18] S. Sangaraju、R. Paolucci、C. Smerzini 和 K. Tarbali，《2016年熊本地震的三维物理地面运动模拟》，载于第六届 IASPEI/IAEE 国际研讨会：地表地质对地震动的影响（ESG6）论文集，2021 年 8 月，日本京都，2021 年。[19] Y. Li、M. Habermann、B. Thomaszewski、S. Coros、T. Beeler 和 C. Theobalt，《面向单目人体动作捕捉的布料形变深度物理感知推断》，载于 2021 年国际三维视觉会议（3DV）。IEEE，2021 年，第 373–384 页。[20] Y. Fang、M. Li、C. Jiang 和 D. M. Kaufman，《具有全局单射保证的三维形变处理》，ACM Trans. Graph.，第 40 卷，第 4 期，第 75–1 页，2021 年。[21] B. Michaud 和 M. Begon，《biorbd：用于分析和模拟人体生物力学的 C++、Python 和 MATLAB 库》，Journal of Open Source Software，第 6 卷，第 57 期，第 2562 页，2021 年。[22] B. O. Community，《Blender——三维建模与渲染软件包》，Blender Foundation、Stichting Blender Foundation，阿姆斯特丹，2022 年。[在线]。网址：http://www.blender.org [23] T.-Y. Lin、M. Maire、S. Belongie、J. Hays、P. Perona、D. Ramanan、P. Doll´ar 和 C. L. Zitnick，《Microsoft COCO：常见场景中的物体》，载于欧洲计算机视觉会议。Springer，2014 年，第 740–755 页。[24] J. Wang、S. Tan、X. Zhen、S. Xu、F. Zheng、Z. He 和 L. Shao，《深度三维人体姿势估计：综述》，Computer Vision and Image Understanding，第 210 卷，第 103225 页，2021 年。[25] C. Ionescu、D. Papava、V. Olaru 和 C. Sminchisescu，《Human3. 6m：面向自然环境中三维人体感知的大规模数据集与预测方法》，IEEE Transactions on Pattern Analysis and Machine Intelligence，第 36 卷，第 7 期，第 1325–1339 页，2013 年。[26] J. Dong、Q. Shuai、Y. Zhang、X. Liu、X. Zhou 和 H. Bao，《从互联网视频中进行动作捕捉》，载于欧洲计算机视觉会议。Springer，2020 年，第 210–227 页。[27] H.-S. Chung 和 Y. Lee，《Mcml：用于整合异构动作捕捉数据的动作捕捉标记语言》，Computer Standards & Interfaces，第 26 卷，第 2 期，第 113–130 页，2004 年。[28] G. Bradski，《OpenCV 库》，Dr. Dobb's Journal of Software Tools，2000 年。