

基于一致性模型的 3D 人体运动姿势预测

刘浩杰

指导教师：高随祥教授

中国科学院大学

2024 年 12 月

- 1 研究背景与意义
- 2 国内外研究现状
- 3 研究目标与创新点
- 4 研究方法与技术路线
- 5 实验设计与可行性
- 6 已有科研基础与所需的科研条件
- 7 研究进展与计划

研究背景与意义

- 虚拟现实、增强现实、运动捕捉等领域快速发展，对实时高效的 3D 人体姿势预测提出更高要求。人体运动预测 (HMP) 旨在利用计算机视觉和机器学习技术，从一系列观察到的人体运动数据中推断出下一个可能的运动状态。
- 近年来，得益于深度学习等人工智能技术的进步，特别是卷积神经网络 (CNN) 和循环神经网络 (RNN) 在图像识别与序列数据分析上的应用，使得 3D 人体姿态估计达到了前所未有的准确度。
- 然而，在实际应用场景中，如何确保长时间内姿态预测的一致性与流畅性仍是一大挑战。因此，探索基于一致性模型的方法对于提高预测精度、减少误差累积具有重要意义。
- 研究意义：
 - 在娱乐、医疗、体育等领域，精确预测人体动作可提升用户体验。
 - 有助于开发更加智能化的人机交互系统，提高效率和安全性。

国内外研究现状

人体运动预测领域：

● 确定性模型：

- 基于 RNN 的序列预测方法，存在累积误差问题。
- 引入图卷积网络（GCN）挖掘人体关节间关系。

● 随机预测模型：

- 基于生成模型（如 VAE、GAN）的多样性预测。
- 基于扩散模型：BelFusion, MotionDiff, HumanMac¹

¹Chen, Ling-Hao, et al. "Humanmac: Masked motion completion for human motion prediction." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023.

扩散模型

- **扩散模型**：强调从噪声到数据的反向采样过程，生成高质量样本。它首先构建了一个从数据样本到随机噪声的过程，然后考虑其反向变换，通过反复执行反向变换来完成数据样本的生成。扩散模型通过一个随机微分方程（SDE）开始对数据分布 $p_{data}(\mathbf{x})$ 进行扩散：

$$d\mathbf{x}_t = \boldsymbol{\mu}(\mathbf{x}_t, t)dt + \sigma(t)d\mathbf{w}_t \quad (1)$$

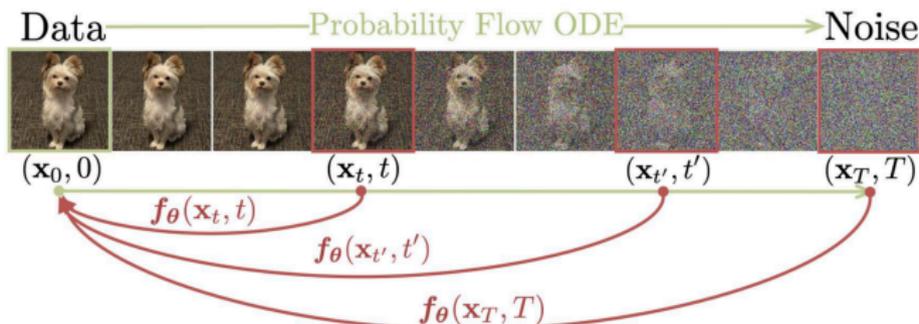
其中 $t \in [0, T]$, $T > 0$ 是一个固定的常数，而 $\{\mathbf{w}_t\}_{t \in [0, T]}$ 表示标准布朗运动。这个 SDE 的一个显著特性是存在一个常微分方程（ODE），被 Song 等人称为概率流（PF）ODE，其在时间 t 采样的解轨迹分布为 $p_t(\mathbf{x})$ ：

$$d\mathbf{x}_t = \left[\boldsymbol{\mu}(\mathbf{x}_t, t) - \frac{1}{2}\sigma(t)^2 \nabla \log p_t(\mathbf{x}_t) \right] dt. \quad (2)$$

这里 $\nabla \log p_t(\mathbf{x})$ 是 $p_t(\mathbf{x})$ 的得分函数；因此扩散模型也被称为得分生成模型。训练目标是通过得分匹配训练一个得分模型 $s_\phi(\mathbf{x}, t) \sim \nabla \log p_t(\mathbf{x})$ 。

一致性模型

一致性模型在连续时间扩散模型中的概率流 (PF) 常微分方程 (ODE) 之上构建，其轨迹将数据分布平滑地过渡到可处理的噪声分布。模型目的是学习一个将任何时间步的任意点映射到轨迹起点的模型。一致性模型的一个显著特性是自一致性，即相同轨迹上的点映射到相同的初始点。一致性模型允许通过只需一步转换随机噪声向量 (ODE 轨迹的端点，图中的 x_T) 生成数据样本 (ODE 轨迹的初始点，图中的 x_0)。



研究问题

- 人体运动预测 (HMP) 旨在利用计算机视觉和机器学习技术, 从一系列观察到的人体运动数据中推断出下一个可能的运动状态。将观察到的历史序列表示为 $\mathcal{D} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T\}$ 由 T 帧组成, 预测序列 $\mathcal{X} = \{\mathbf{x}_{T+1}, \mathbf{x}_{T+2}, \dots, \mathbf{x}_{T+f}\}$ 由 f 帧组成
- 现有 RNN, GCN 等模型预测效果差, 基于扩散模型的 HumanMac, MotionDiff, HumanMac 等模型效果较好, 但计算成本高, 无法满足实时需求, 特别是在需要快速响应的应用场景中
- 保持扩散模型的高精准度的同时, 减少计算成本, 实现单步生成。

创新点

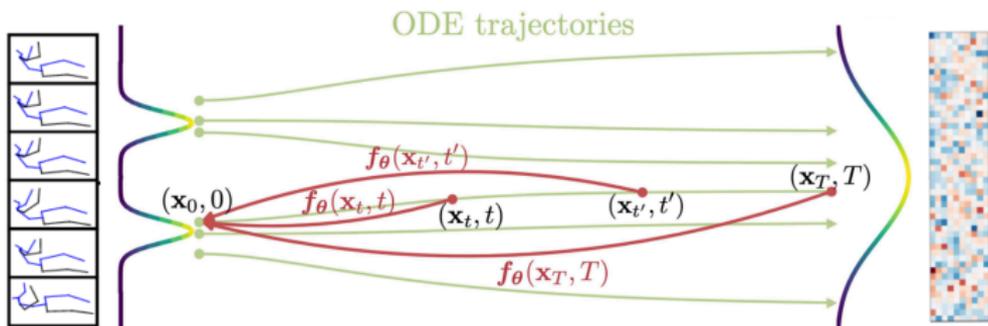
研究创新点

- **模型设计**：引入一致性约束，优化扩散生成流程，实现单步生成。
- **技术突破**：解决当前扩散模型在人体姿势预测生成速度上的瓶颈。
- **综合优化**：生成质量与速度的动态平衡。

研究方法与技术路线

- 初始条件：历史序列与噪声作为模型输入。
- 将观察到的历史序列表示为 $\mathcal{D} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T\}$ 由 T 帧组成，预测序列 $\mathcal{X} = \{\mathbf{x}_{T+1}, \mathbf{x}_{T+2}, \dots, \mathbf{x}_{T+f}\}$ 由 f 帧组成，其中 $\mathbf{x}_t \in \mathbb{R}^{3 \times J}$ 是时间戳 t 的 3D 坐标， J 是身体关节的数量。
- 训练：通过一致性约束优化生成过程。
- 然后将训练好的得分模型 $s_\phi(\mathcal{X}, t)$ 代入式子中利用求解器得到解轨迹 $\{\mathcal{X}_t\}_{t \in [\epsilon, T]}$ ，我定义一致性函数 $f: (\mathcal{X}_t, t) \mapsto \mathcal{X}_\epsilon$ 。如图所示，一致性模型的目标（符号为 f_θ ）是通过学习强制自一致性属性来估计数据中的一致性函数 f 。

模型基础架构

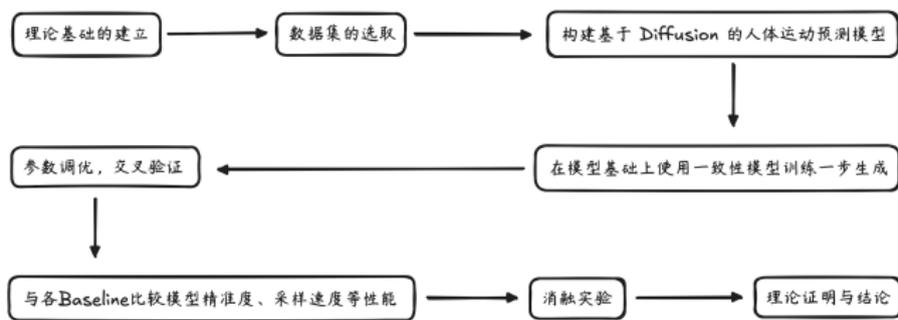


- **生成**：单步生成未来 3D 姿势序列。通过一个训练良好的一致性模型 $f_\theta(\cdot, \cdot)$ ，我可以从初始分布 $\hat{\mathbf{x}}_T \sim \mathcal{N}(\mathbf{0}, T^2 \mathbf{I})$ 采样生成样本，然后评估一致性模型以得到 $\hat{\mathbf{x}}_\epsilon = f_\theta(\hat{\mathbf{x}}_T, T)$ 。这仅涉及通过一致性模型的一次前向传递，因此在单步中生成样本。重要的是，也可以通过交替去噪和噪声注入步骤多次评估一致性模型，以提高样本质量。

实验设计与可行性

● 实验设计

- 基准比较：与传统模型（如 RNN、GCN）、扩散模型进行性能对比。
- 消融实验：验证一致性约束对生成性能的影响。
- 指标评估：Average Pairwise Distance (APD)、Average Displacement Error (ADE)。



● 可行性分析

- 一致性模型已在多种生成任务中证明其优越性。
- 实验所需的数据集、计算资源均已具备。

已有科研基础与所需的科研条件

- **已有科研基础：**本人对人体运动预测相关技术有一定的了解与使用经验，也曾选修过图论与网络算法、机器学习、深度学习等相关课程，有一定的机器学习基础。对一致性模型有一定的了解，此外还阅读了大量的一致性模型与人体运动预测相关的论文，并且对其中部分论文进行了代码复现。
- **所需科研条件：**所需的研究条件主要集中在计算资源上，Diffusion 和 3D 数据的训练和实验通常需要较高的计算资源，包括但不限于 GPU 或 CPU 等高性能计算设备或服务器。

研究进展

- 已完成：理论框架初步搭建，确定技术路线。
- 初步实验验证一致性模型在生成任务中的效果。

研究计划

- **2024.12–2025.3**: 对已有的人体运动预测进行归纳总结，并指出不同方法的优劣。同时提出基于一致性模型的人体运动预测加速的想法，完成论文初稿。
- **2025.3–2025.6**: 撰写小论文并投稿。
- **2025.6–2025.10**: 扩展应用场景，新增研究内容：
 - 基于一致性模型与文本条件的多模态 3D 人体运动预测。
 - 基于 ControlNet 搭建可控的 3D 人体运动预测模型。
- **2025.10–2025.11**: 完成学位论文。