

基于图神经网络与时序融合的人体动作预测方法

刘 瑞¹ 黎润霖¹ 阴明旭¹
LIU Rui LI Runlin YIN Mingxu

摘 要

针对人体动作预测中时空建模能力不足的问题,文章提出了一种融合图神经网络(graph neural network, GNN)与 Transformer 架构的人体动作时空预测方法。该方法首先利用图卷积网络从每帧骨架图中提取空间结构特征,再通过 Transformer 对时间序列建模,捕捉动作随时间演化的动态变化,从而实现对人体动作的全局建模。并在 CMU Mocap 数据集上开展了系统实验,选取平均关节位置误差(mean per joint position error, MPJPE)与角度误差(angular error, AE)为主要评估指标,并与多种主流模型进行了对比。实验结果表明,所提模型在短期与长期预测任务中均取得最优性能,具备更强的时空表示能力与泛化性能。此外,消融实验进一步验证了空间建模、时序建模与特征融合模块在整体性能中的关键作用。该研究为提升人体动作预测的准确性与连续性提供了一种有效思路。

关键词

人体动作预测; 图神经网络; 时序融合

doi: 10.3969/j.issn.1672-9528.2025.10.034

0 引言

人体动作预测是计算机图形学、机器人、游戏和虚拟现实等领域的关键技术,旨在根据过去或当前的动作序列预测未来的人体姿态变化。准确、自然的人体动作预测对于提升数字角色的逼真度、实现智能人机交互以及进行运动分析具有重要意义。然而,人体动作数据固有的高维度、复杂的关节间空间关联以及跨越长时期的时间依赖性,使得精准可靠的动作预测成为一个具有挑战性的难题^[1]。

人体动作预测作为一种典型的时空建模任务,主要是基于历史的骨架序列预测未来一段时间内的动作演化。早期研究多采用基于循环神经网络(recurrent neural network, RNN)及其变体(如 LSTM、GRU)进行建模,通过学习时间序列中的动态特征完成预测任务^[2]。然而,受限于 RNN 类模型在长序列建模中的梯度传播问题,这类方法往往在长期预测任务中表现欠佳。随后,一些工作尝试引入卷积网络(convolutional neural network, CNN)和时序卷积网络(temporal convolutional network, TCN)等结构,在一定程度上缓解了长时依赖问题,但空间结构建模能力仍然有限。

近年来,图神经网络(graph neural network, GNN)因其在结构化数据建模中的优势被引入到骨架建模任务中,能较

好地捕捉关节之间的空间关系;同时,Transformer 模型凭借自注意力机制在长序列建模中的优势,被广泛用于时序建模任务。

另一方面,Transformer 架构因其基于自注意力机制的设计,在自然语言处理和时间序列建模中表现卓越。近期也有研究将其引入动作预测任务中,如将 Transformer 直接应用于关节序列,捕捉动作随时间的非线性演化过程,这类方法具有良好的全局建模能力。

为此,本文提出一种基于图神经网络与 Transformer 融合的人体动作时空预测方法。该方法首先利用 GNN 提取骨架序列中每一帧的空间结构特征,然后将其输入 Transformer 进行时序建模,从而实现对人体动作的全局动态理解。在 CMU Mocap 数据集上进行了大量实验,结果表明,本文方法在多个预测指标上均优于现有主流方法,具备良好的精度与泛化能力。

1 人体动作预测方法概述

1.1 图神经网络概述

图神经网络是一类用于处理图结构数据的神经网络模型,广泛应用于社交网络、生物信息、推荐系统以及动作识别与预测等领域。传统神经网络在处理序列或图像数据方面表现良好,但无法直接应用于非欧几里得结构的数据,如人体骨架这种具有明确拓扑结构的骨架图。GNN 通过邻接矩阵和归一化操作,使得节点能够聚合其邻居节点的信息,从而实现结构感知的特征学习^[3]。

1. 吉利学院智能科技学院 四川成都 641423

[基金项目] 吉利学院“一院一品”教学改革项目(2024JG30048);吉利学院“一院一品”教学改革项目(2024JG30035)

1.2 Transformer 概述

Transformer 结构在自然语言处理任务中取得了突破性进展,近年来也被广泛引入到动作建模等时序预测场景。与传统的循环神经网络(如 RNN、LSTM)不同,Transformer 完全基于自注意力机制,可对序列中任意位置间的关系进行建模,具备出色的全局依赖建模能力和高度的并行计算效率^[4]。

在人体动作建模中,Transformer 可以处理整段骨架序列的空间嵌入,捕捉关节随时间变化的复杂动态规律。相较于传统时序模型,Transformer 在建模长距离依赖、保留全局动态信息方面具有显著优势,特别适用于多帧预测、复杂动作过渡等任务。

1.3 时空建模方法概述

人体动作具有明显的时空特性,其动态演化过程既依赖于骨架结构中各关节之间的空间拓扑关系,也依赖于动作随时间发展的连续性与变化规律^[5]。因此,高质量的人体动作预测模型需要同时具备空间建模能力与时间建模能力,实现对动作序列中“结构+动态”的联合建模^[6]。

但大多数 GNN 模型主要聚焦于每一帧骨架的空间关系建模,对于帧与帧之间的时间动态建模通常依赖于简单的时间卷积或滑动窗口机制。因此,将 GNN 与更强的时序建模方法结合,是当前动作预测任务中的重要发展趋势。

尽管 Transformer 在时序建模中表现优异,但其通常将骨架序列视为平坦的一维向量序列,忽略了人体动作中内在的空间结构。为此,近年来有研究尝试将 Transformer 与图神经网络结合,实现对空间结构与时间动态的统一建模,以提升预测精度和物理合理性^[7]。

因此,如何设计一种更高效、更具表达能力的时空融合模型,仍是当前人体动作预测研究中的重要方向。

2 研究方法

2.1 模型整体框架

为了高效建模人体动作的空间拓扑结构与时序动态特征,本文提出一种融合图神经网络(GNN)与 Transformer 的时空预测框架。

整体架构如图 1 所示,主要由三部分组成:(1)空间建模模块(GNN):针对每一帧骨架图,利用图卷积网络(如 GCN)提取空间结构特征,挖掘关节间的相互依赖。(2)时序建模模块(Transformer):将所有帧的空间特征作为序列输入,采用多层 Transformer 编码器建模长时依赖,捕捉动作演化规律。(3)预测解码模块:将融合后的特征输入线性变换层,输出未来 H 帧的骨架姿态。该模型在结构上实现了空间和时序的解耦建模,又通过端到端训练实现整体时空特征融合,兼具精度与泛化能力,适用于多种骨架预测场景。

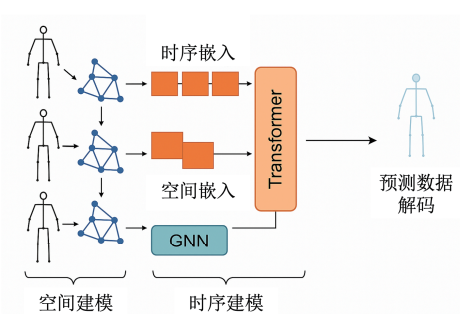


图 1 模型结构示意图

2.2 问题定义

本文关注的人体动作预测任务,目的是根据一个已知长度的骨架序列,预测未来若干帧的人体姿态。设历史输入序列为:

$$\mathbf{X} = \{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_t\} \quad (1)$$

式中: $\mathbf{X}_t \in \mathbb{R}^{J \times D}$, 表示第 t 帧的骨架姿态,包含 J 个关节点,每个关节点有 D 维特征(例如 3D 坐标或关节角度)。预测目标为输出未来 H 帧的人体动作序列:

$$\mathbf{Y} = \{\mathbf{X}_{t+1}, \mathbf{X}_{t+2}, \dots, \mathbf{X}_{t+H}\} \quad (2)$$

通过一个可学习函数 F_θ 实现从输入序列到未来帧的映射,即:

$$\hat{\mathbf{Y}} = F_\theta(\mathbf{X}) \quad (3)$$

式中: $\mathbf{X}$ 为输入的历史骨架序列, $\hat{\mathbf{Y}}$ 为模型输出的预测序列。通过最小化 $\hat{\mathbf{Y}}$ 与真实未来序列 $\mathbf{Y}$ 之间的欧氏距离,优化模型参数,即:

$$L = \frac{1}{H} \sum_{i=1}^H \|\mathbf{x}_{t+i} - \hat{\mathbf{y}}_{t+i}\|^2 \quad (4)$$

2.3 空间建模模块

图神经网络模块主要用于学习单帧人体骨骼姿态中的空间依赖关系。人体骨架具有天然的结构特性,每一帧骨架数据可形式化为一个无向图 $G = (V, E)$, 其中 V 表示关节点集合; E 表示关节之间的骨骼连接关系。

为有效建模关节之间的空间依赖关系,本文采用图卷积网络(graph convolutional network, GCN)对每一帧骨架图进行空间特征提取。在 GCN 中,节点(关节点)特征的更新过程可表示为:

$$\mathbf{H}^{(l+1)} = \sigma(\tilde{\mathbf{D}}^{-1/2} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-1/2} \mathbf{H}^{(l)} \mathbf{W}^{(l)}) \quad (5)$$

式中: $\tilde{\mathbf{A}}$ 为骨架图的原始邻接矩阵; $\mathbf{I}$ 为单位矩阵,则 $\tilde{\mathbf{D}} = \mathbf{A} + \mathbf{I}$ 为添加了自环的邻接矩阵; $\tilde{\mathbf{D}}$ 为对角线元素为 $\tilde{\mathbf{A}}$ 每行元素之和的度矩阵; $\mathbf{H}^{(l)}$ 为第 l 层 GCN 的输入节点特征矩阵; $\mathbf{W}^{(l)}$ 为第 l 层的可学习权重参数矩阵; σ 为非线性激活函数。该公式通过对邻接节点的特征进行加权聚合,实现了关节点间信息的有效传播与交互。

通过堆叠多层图卷积,模型能够逐步扩展每个节点的感受野,从而捕捉人体骨架中局部关节连接与整体空间结构的

关系。最终，每一帧骨架的节点特征被编码为一个结构感知的空间嵌入表示 f_i ，这些在不同时间步提取的空间嵌入向量序列构成了 $f_1, f_2, f_3, \dots, f_n$ ，供后续的时序融合模块使用。

2.4 时间建模模块

空间特征序列作为时间序列输入 Transformer 模型。Transformer 通过多头自注意力机制建模动作随时间的动态变化，核心操作为：

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}(\mathbf{Q}\mathbf{K}^T/\sqrt{d})\mathbf{V} \quad (6)$$

式中： $\mathbf{Q}$ 、 $\mathbf{K}$ 、 $\mathbf{V}$ 分别为查询、键、值向量； d 为缩放因子。

每个空间嵌入 f_i 加入位置编码 p_i 后构成序列 $z_i = f_i + p_i$ ，再输入多层 Transformer 编码器得到时间建模后的特征 $Z' = \{z'_1, z'_2, \dots, z'_N\}$ 。

2.5 特征融合与预测模块

在空间与时间建模完成后，将 Transformer 输出的特征通过融合机制进行整合，使用平均池化或残差连接等方式形成统一表示。该表示输入至线性预测层，输出未来 H 帧的姿态向量：

$$\hat{\mathbf{Y}} = \mathbf{W}_{\text{out}} \cdot \text{Fusion}(\mathbf{Z}) + \mathbf{b} \quad (7)$$

式中： $\mathbf{W}_{\text{out}}$ 与 $\mathbf{b}$ 表示可学习参数； $\text{Fusion}(\mathbf{Z})$ 表示融合后的特征。最终输出尺寸为 $H \times J \times D$ ，即预测未来每帧所有关节的位置。

2.6 损失函数与训练策略

采用均方误差作为主要优化目标，定义为：

$$\mathcal{L}_{\text{MSE}} = \frac{1}{H} \sum_{i=1}^H \|x_{T+i} - \hat{x}_{T+i}\|^2 \quad (8)$$

为提升预测的自然性与连续性，引入关节速度平滑正则项，抑制突变并增强动作连贯性。

模型使用 Adam 优化器进行训练，初始学习率为 0.001，批量大小为 32，采用早停机制防止过拟合。

3 实验

3.1 实验配置

本研究使用 CMU Motion Capture 数据集进行验证。该数据集包含多个被试者执行的丰富动作序列，包括行走、跑步、跳跃、挥手等，覆盖多种姿态变化情形^[8]。选取常用的子集（如 walking、eating、smoking、discussion 等）作为实验对象，并按照标准流程将每帧表示具有 22 个关节的三维坐标序列。

预处理步骤包括：

（1）对每段动作序列进行标准化，使每个关节坐标减去根节点（hip）位置并归一化；

（2）使用滑动窗口划分样本，输入序列长度为 $T=50$ ，预测未来 $H=10、25、50$ 帧；

（3）数据按 8:1:1 划分为训练、验证和测试集。

实验训练在以下硬件配置下进行：CPU：Intel i7 14700K；内存：128 GB；GPU：1 张 NVIDIA RTX4080。

3.2 评估指标

为了全面评估预测模型的性能，本文采用以下两种常用的动作预测指标：

（1）平均关节点位置误差（mean per joint position error, MPJPE），是衡量预测姿态与真实姿态在欧氏空间中偏差的主指标。计算方式为预测骨架与真实骨架每个对应关节的三维坐标差值的均方根，单位通常为毫米（mm）。该指标反映整体预测精度，值越小表示预测越接近真实。

（2）平均关节角度误差（angle error, AE）用于评估预测动作的自然性与连贯性，表示预测与真实角度的平均差值，单位为度（°）。该指标适用于需要考虑物理合理性的动作序列，如行走与跑步等高动态动作。

本文在不同预测步长（10 帧、25 帧、50 帧）下分别计算上述指标，以评估模型在短期、中期及长期预测中的性能表现。

3.3 对比实验

为了验证所提出模型的有效性，本文在 CMU Mocap 数据集上与当前主流的人体动作预测方法进行了系统对比。所选对比方法涵盖了传统的时序建模模型、图结构建模方法以及最新的基于 Transformer 的模型，具体包括：

（1）ERD：早期基于 RNN 的动作预测模型，采用递归结构进行短期姿态回归；

（2）Residual Sup. RNN (RSRNN)：引入残差连接机制的深层循环网络，提升了动作连续性；

（3）ST-GCN：采用空间图卷积与时间卷积结合建模人体骨架的时空关系；

（4）DCT-GCN：利用离散余弦变换处理时间维度信号，提升 GNN 在频率域的建模能力；

（5）ST-Transformer：纯基于 Transformer 架构进行序列建模，具有良好的长期依赖捕捉能力。

平均关节点位置误差实验结果对比如表 1 所示，本文方法在 10 帧、25 帧以及 50 帧三个预测步长上的平均关节点位置误差分别为 11.7、20.5 和 31.8，在所有预测步长上均取得最优性能，尤其在长期预测任务中，显著优于其他方法，说明所提模型在建模时空依赖关系方面具备更强的表示能力与泛化性能。

表 1 平均关节点位置误差对比实验结果

方法	不同预测步长的平均关节点位置误差		
	10 帧	25 帧	50 帧
ERD	18.7	31.6	48.3
RSRNN	16.2	27.4	42.7
ST-GCN	14.8	25.5	39.9
DCT-GCN	13.6	23.7	35.1
ST-Transformer	12.9	22.8	36.4
本文方法	11.7	20.5	31.8

平均关节角度误差对比实验结果对比如表 2 所示。

表 2 关节角度误差对比实验结果

方法	不同预测步长的平均关节角度误差		
	10 帧	25 帧	50 帧
ERD	13.2°	19.4°	26.7°
RSRNN	11.7°	16.5°	24.1°
ST-GCN	10.5°	16.8°	21.5°
DCT-GCN	9.8°	13.1°	19.6°
ST-Transformer	9.3°	12.6°	18.4°
本文方法	8.1°	10.9°	16.2°

整体来看，本文方法在各预测步长下均取得最优表现，其中在短期预测（10 帧）中角度误差为 8.1°，相比 ST-Transformer（9.3°）与 DCT-GCN（9.8°）进一步降低；在中期（25 帧）与长期（50 帧）预测中，误差分别为 10.9° 和 16.2°，显著优于其他方法，表明本文模型在保持动作连贯性与物理合理性方面具有更强的建模能力。相比传统时序方法，本文方法的优势更加明显，说明图结构与时间建模的有效融合对于提升预测角度准确性具有重要意义。

3.4 消融实验

为进一步验证模型中关键组件的有效性，本文设计了若干消融实验，逐步移除或替换模型中的核心模块，分析其对整体预测性能的影响。关注以下 3 个变体：

- （1）无图结构建模（w/o GNN）：移除空间建模模块，直接将原始骨架展平成向量输入 Transformer；
- （2）无时序建模（w/o Transformer）：仅使用图神经网络对每帧建模，再通过线性映射预测未来帧；
- （3）无时空融合（w/o Fusion）：空间特征与时间特征分别建模后直接串联，不进行融合策略优化。

在 CMU Mocap 数据集上，使用 MPJPE 作为主要指标，评估不同变体在 10 帧、25 帧与 50 帧 3 个预测步长下的性能表现，结果如表 3 所示。可以看出，完整模型在所有预测步长下均表现最佳。移除 GNN 后，模型无法有效捕捉关节之间的空间拓扑结构，导致短期与长期误差均大幅上升；而移除 Transformer 则影响了动作演化的动态建模能力，尤其在长期预测下 MPJPE 增加明显；未使用时空融合机制虽然仍可提取特征，但特征之间缺乏交互，导致预测精度下降。

表 3 消融实验结果

模型	不同预测步长的 MPJPE		
	10 帧	25 帧	50 帧
w/o GNN	15.4	26.4	42.1
w/o Transformer	14.1	24.9	38.5
w/o Fusion	12.6	22.3	34.7
完整模型	11.7	20.5	31.8

以上分析表明，本文提出的空间建模 + 时间建模 + 融合策略三者协同作用，共同构建出具有良好结构感知性与动态理解能力的时空预测模型。

3.5 定性分析

为进一步验证所提出模型在空间结构保持性与时间动态一致性方面的表现，对多个典型动作序列进行了可视化分析，并与真实骨架及现有方法的预测结果进行了对比。此处选取了一次 50 帧预测步长的行走动画作为演示。效果如图 2 所示，图 2（a）为本文方法，图 2（b）为对比方法中效果最好的 ST-Transformer 方法。其中实线表示真实骨架，虚线表示预测骨架。

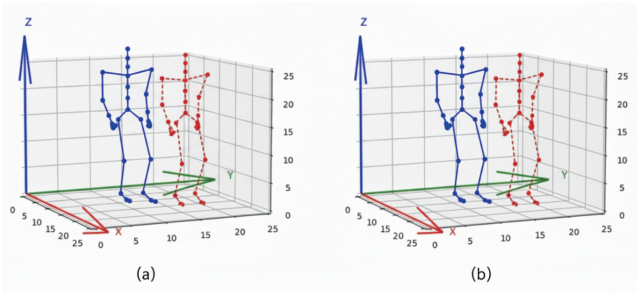


图 2 预测结果可视化

从可视化结果可以看出，对比方法 ST-Transformer 在预测过程中出现不同程度的动作僵化现象，说明其在结构保持方面存在一定不足，而本文方法能够较好地保持关节结构的一致性动作趋势。

4 结语

研究针对人体动作预测任务中时空特征建模不足的问题，提出了一种基于图神经网络与 Transformer 融合的人体动作时空预测方法。该方法以人体骨架的图结构为基础，利用图卷积网络有效建模关节间的空间依赖关系，同时借助 Transformer 架构强大的时序建模能力，捕捉长时依赖的动作演化过程。通过引入结构与时间特征的融合策略，模型能够在保持骨架拓扑合理性的同时，实现对未来动作序列的高精度预测。

在 CMU Mocap 数据集上进行了大量实证研究，结果表明，本文方法在多个预测步长下均显著优于主流基线方法，在位置误差与角度误差两个维度上均取得最优性能。消融实验进一步验证了空间建模、时序建模与融合机制对整体性能的关键作用。

尽管本文方法在准确性与稳定性方面均表现出良好性能，但仍存在一定局限性，如对极端或异常动作的适应能力尚有限，模型结构在效率与参数量上仍有优化空间。未来工作可考虑引入人体动力学约束、多模态感知信息或轻量级 Transformer 结构，以进一步提升模型的泛化能力与实用性。

基于图注意力网络的电力通信网络动态路由优化算法

柳 茜¹

LIU Qian

摘 要

在对电力通信网络进行动态路由优化时,使用单头注意力机制难以准确计算网络节点的权重值,导致优化效果不佳。对此,文章提出一种基于图注意力网络的电力通信网络动态路由优化算法。从电力通信网络数据中提取节点特征和边特征,并构建通信网络特征向量。将特征向量作为输入变量输入到图注意力网络中,计算网络节点相对于邻近节点的注意力系数。在多头注意力机制的作用下,计算注意力系数的平均值,并据此分配节点的权重。在优化网络动态路由时,优先选择权重值最高的路由节点作为下一节点,生成网络路由方案,并对其进行评估。选择评估值最高的方案作为最优路由方案输出。实验结果表明,应用设计的优化方法后,网络动态路由的吞吐量达到 5 225.26 bit/s,带宽利用率为 93.52%,显示出较高的优化效果。

关键词

图注意力网络;电力通信网络;网络路由;动态路由;路由优化

doi: 10.3969/j.issn.1672-9528.2025.10.035

0 引言

电力通信网络作为支撑电网安全稳定运行的基础,其对于推进新型电力系统建设具有重要意义。由于风电、光伏等新能源具有不确定性,其会导致电力通信网络的拓扑结构呈现较高的波动性,进而引发一系列的网络问题。现有的网络

动态路由无法适应通信网络的动态变化,且容易受到链路故障的影响,出现较长时间的网络波动,影响方法的应用效果。

在该研究背景下,不少研究学者针对路由优化展开了研究,并给出了自己的观点。赵星珠等人^[1]根据计算机网络路由运行特性,设定最小化平均网络时延作为优化目标,设定路由唯一性等多个约束条件,根据人工鱼群算法,设定初始参数,根据鱼群特点,设计人工鱼的行为规则,引入公告牌机制,记录全局最优解,通过多次迭代优化,输出最优的路

1. 甘肃中网通讯技术有限公司 甘肃兰州 730000

参考文献:

- [1] 孟祥璞,李硕,苑明哲,等.基于人体骨架的动作识别:综述与展望[J].信息与控制,2025,54(1):1-27.
- [2] DUBEY S, DIXIT M. A comprehensive survey on human pose estimation approaches[J]. Multimedia systems, 2023, 29: 167-195.
- [3] KHEMANI B, PATIL S, KOTTECHA K, et al. A review of graph neural networks: concepts, architectures, techniques, challenges, datasets, applications, and future directions[J]. Journal of big data, 2024, 11: 18.
- [4] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing System. New York: ACM, 2017: 6000-6010.
- [5] 刘宝龙,周森,董建锋,等.基于骨架的人体动作识别技术研究进展[J].计算机辅助设计与图形学学报,2023,35(9): 1299-1322.

- [6] 赵晨,彭舰,黄军豪.基于时空关节映射的骨架动作识别方法[J/OL].计算机科学,1-14[2025-05-26].<https://link.cnki.net/urlid/50.1075.tp.20241209.1518.010>.
- [7] 游浩,丁苍峰,马乐荣,等.Graph Transformer 技术与研究进展:从基础理论到前沿应用[J].计算机应用研究,2025,42(4): 975-986.
- [8] DE LA TORRE F, HODGINS J K, BARGTEIL A W, et al. Guide to the carnegie mellon university multimodal activity (CMU-MMAC) database[EB/OL].(2008-08-22)[2025-06-12].<https://www.ri.cmu.edu/publications/guide-to-the-carnegie-mellon-university-multimodal-activity-cmu-mmact-database/>.

【作者简介】

刘瑞(1997—),男,硕士研究生,四川成都人,研究方向:复杂网络、虚拟现实技术, email: 836287308@qq.com。

(收稿日期: 2025-05-28 修回日期: 2025-10-15)