

基于中心差分转换器图卷积的骨架识别模型

何璠¹ 林亮² 张楠³

1 江西工程学院, 江西宁余, 338000;

2 江西邻通科技有限公司, 江西宁余, 338000;

3 中国电信南昌分公司, 河南商丘, 476000;

摘要: 当前骨架动作识别的主流方法多采用时空图卷积技术。然而, 此类方法不仅导致特征提取缺乏足够的灵活性, 且局限于对节点信息的聚焦, 难以充分挖掘数据中的深层关联。因此, 提出了一个 CDGT 的新型模型用于骨架动作识别。该模型首先对骨架数据进行精准建模, 将关节与骨骼的运动信息分别输入模型进行初步处理; 随后, 经过预处理的数据被传送至 CDGC 模块进行聚焦节点自身信息与节点间的关联信息。处理之后, 模型会调用提出的 STT 模块, 有效捕捉被关节标记的空间依赖关系与时间依赖关系。为验证 CDGT 模型的性能, 在大型数据集 NTU-RGB+D60 上进行了系统的训练与验证, 并将其结果与基线方法 ASGC-STT 对比。实验结果显示, 在数据集的两个交叉子集 CS 和 CV 上, 该模型的准确率分别提升了 0.17% 和 0.62%。实验结果充分证明, 提出的 CDGT 网络能够有效增强骨架动作识别的性能, 并且在大型数据集上展现出了显著且稳定的效果。

关键词: 动作识别; 时空图卷积; 中心差分算法; 时空转换器

DOI: 10.69979/3029-2808.25.10.048

引言

目前, 人体动作识别是一个热门且具有挑战性的研究课题, 由于其在视频分析、人机交互等领域的广泛应用^[1]。

传统动作识别模型中, 注意增强网络的应用较广泛, 能对特征表现明显的信息进行针对性增强, 这一机制在一定程度上为识别精度的提升带来了帮助。但动作包含的动态梯度信息以及局部依赖关系, 对于识别精度的优化而言同样重要。

1 相关工作

1.1 图卷积方法

在研究的基于骨架动作识别的方法中, 基于 CNN^[2] 和基于 RNN^[3] 的方法相比, GCN^[4, 5] 更有效。现在大部分基于 GCN 的方法都是直接将关节位置或坐标作为其模型的输入。文献^[5]提出了一种基于 GCN 的多流模型, 旨在从骨架运动序列中学习丰富的鉴别特征, 提高模型的鲁棒性。

1.2 中心差分算法

中心差分图卷积通过融合中心节点的局部梯度特征, 显著提升了模型的表征能力。该方法已成功应用于

人脸活体检测、非接触式心率监测等多个视觉任务, 展现出优异的性能表现。现有研究多基于 GCN 框架, 或通过增强时空感知模块^[5], 或构建多流网络结构^[4]来丰富特征表达。然而, 这些方法普遍未能有效利用中心节点与邻域节点间的局部运动特征, 限制了模型的判别能力。

1.3 时空转换器机制

转换器网络在自然语言处理领域取得了突破性进展, 在计算机视觉领域引发了广泛且浓厚的兴趣。转换器的核心优势在于其全局自注意力机制, 而自注意力操作与卷积操作二者结合有望为相关任务带来新的突破。受此启发, 本工作将转换器引入到 GCN 网络中。尽管此方法对骨架识别方法的研究提供了帮助, 但也带来了计算复杂度和模型参数数量增加的问题。

2 提出的方法

结合中心差分算法和转换器机制, 本文提出 CDGT, 用于捕获局部特征聚合过程中中心节点与其关联的相邻节点之间局部运动的动态梯度信息, 时空转换器机制可捕获全局时空依赖关系。

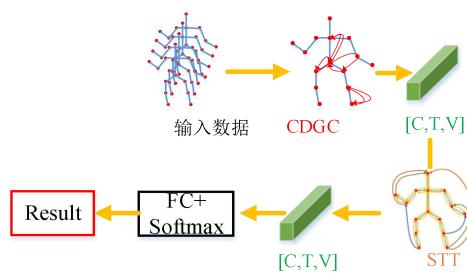


图1 CDGT 总体架构图

CDGT 网络架构如图 1 所示, 包含四个关键处理阶段:

- 1) 骨架数据输入模块, 负责接收关节和骨骼运动信息;
- 2) 中心差分图卷积网络 (CDGC), 通过中心节点与邻域节点的空间梯度特征更新顶点信息;
- 3) 时空转换器注意力网络 (STT), 建模全局时空依赖关系;
- 4) 全连接分类层, 对特征进行加权融合并输出动作预测结果。

该架构首先通过 CDGC 提取局部运动特征, 再利用 STT 捕获全局时空上下文, 最终实现端到端的动作识别。

2.1 数据输入

动作识别的输入数据是提取骨架的关节信息, 本文将骨架数据作为一个无环图, 将骨架的关节看作图中的点, 骨骼为边, 骨架的重心为图的根节点。

2.2 中心差分图卷积训练

本文将 CDGC 作为 CDGT 的子网络。并在数据集中进行实验。

对采样顶点的中心梯度进行聚合, 公式如下所示:

$$y(v_i) = \alpha \sum_{v_j \in R_i} \frac{1}{Z_{ij}} w(l_i(v_j)) \cdot (x(v_j) - x(v_i)) + (1 - \alpha) \sum_{v_j \in R_i} \frac{1}{Z_{ij}} w(l_i(v_j)) \cdot x(v_j) \quad (1)$$

其中 x 和 y 分别代表骨架的输入特征和输出特征。

超参数 $\alpha \in [0,1]$, 用于平衡节点和梯度信息的贡献。 R_i 表示 v_i 一阶相邻顶点的距离, l_i 是划分函数, 其作用是将 R_i 中的顶点划分为固定数量的子集, 每个子集都有独立的卷积权向量, w 为权重函数。 Z_{ij} 是一个标准化向量, 主要用来平衡每个子集对输出的贡献。当 $v_j = v_i$ 时, 相对于中心位置 v_i 本身, 梯度值总是等于 0。

在 CDGC 网络中, 当交换两个相邻节点的中心位置时, 其对应的梯度向量呈现方向相反的特性。具体而言, 若以节点 A 为中心计算节点 B 的梯度向量, 与以节点 B 为中心计算节点 A 的梯度向量, 二者大小相等但方向相反。这种对称性质确保了网络对节点间相对运动特征的

稳定建模。根据公式 1, 当 v_1 是中心, v_2 相当于 v_1 点的梯度方向从 v_1 到 v_2 , 反之类似。其次, 当在自然骨架图中为原本不连通的两个节点建立连接后, 二者之间的差异会表现得更为显著。

由公式可知, 训练参数在 $x(v_i)$ 前的总和可通过对邻接矩阵的第二维求和得出, 所以, CDGC 网络的表达式之和可表示为:

$$Y = \alpha \cdot (AX - \hat{A} \odot X)W + (1 - \alpha) \cdot AXW \quad (2)$$

$$= (AX - \alpha \hat{A} \odot X)W$$

公式中的 A 是归一化邻接矩阵。在本文模型中, 输入特征为 (C, T, V) 的张量, 其中三个元素分别表示为通道, 帧数和顶点的数量。 \hat{A} 是对邻接矩阵 A 的第二维求和, 延伸得到的 $N \times 1$ 向量 \bar{A} 到 $N \times C$ 的矩阵 \hat{A} , 其中 $\hat{A} = \bar{A} \times 1$, 其中 1 是 $1 \times C$ 维的向量。其中权重矩阵 W 是将多个输出通道的权值向量叠加形成的。

2.3 时空转换器网络训练

本文提出的时空转换器网络采用三模块级联架构: 关节标记嵌入层负责特征编码; 空间注意力模块 (STA) 专司关节间空间关系建模; 时间注意力模块 (TTA) 主理动作时序演化分析。关节标记作为特征提取器, 专门用于从中心差分网络中解码时空特征。其空间注意力组件聚焦于关节间的结构关联性建模, 时间注意力组件则专精于动作序列的动态演变分析。

本文提出的转化器如图 2 所示, 每个块有两个子层, 分别是多头自注意层 (MHSA) 和全连接前馈层 (FFN), 输入序列先被投影到三个不同的矩阵中, 分别是查询 $Q \in R^{N \times D_q}$, 键 $K \in R_{N \times D_k}$ 和值 $V \in R_{N \times D_v}$ 。 D_q , D_k 和 D_v 分别表示查询, 键和值的维度。

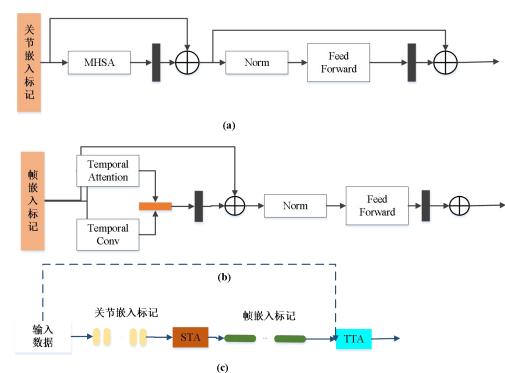


图2 (a) STA 框架; (b) TTA 框架; (c) STT 框架

MHSA 机制将 D 维表示映射到具有 H 个不同的学习投影集的多个子空间中。对于每个投影查询、键和值，根据公式来计算单个注意头。

FFN (前馈网络) 由两层逐点前馈网络构成。转换器在每个模块后，会先采用一个残差连接，随后再进行一层归一化处理，来稳定训练过程并提升模型性能。

2.3.1 时空标记嵌入

本文是在 CDGC 中提取时空特征，并将其重塑为联合标记，给定 CGDC 网络的输出数据 $X_{CD} \in R^{C_{dim}}$ ，通过批处理移动时间维度，将其重塑为关节标记 $X_j \in R^{T \times N \times C_{dim}}$ ，其中 C_{dim} ， T 和 V ，分别表示为输入特征通道，帧和关节的数量。帧标记 $X_f \in R^{N \times T \times C_{dim}}$ 由前一个 STA 块输出重塑而成，如图 2(c)。此外，通过改变时间卷积的步长，可对标记特征维数和帧数进行调整。

2.3.2 空间转换器注意模块

本文提出空间转换器注意力 (STA) 模块，其结构如图 2(a) 所示，基于标准转换器架构。该模块通过 MHSA 机制对关节标记进行特征建模，其中对关节特征施加多个可学习变换，将查询、键和值分别映射至不同子空间。STA 模块的计算流程与标准转换器一致，其自注意力机制可学习关节间的全局依赖关系，有效突破了图卷积神经网络中局部邻域的限制。

2.3.3 时间转换器注意模块

本文提出时间转换器注意力 (TTA) 模块，采用双分支结构设计 (如图 2(b))。与标准自注意力不同，该模块通过并行分支分别建模短期和长期时间依赖关系。其中，一个分支利用时间卷积捕捉局部帧间上下文，另一分支采用标准 MHSA 学习全局时序特征。两个分支的输出特征经拼接后送入 FFN 层，在保持全局建模能力的同时有效降低了计算复杂度。此方法既保留了自注意力机制的全局建模优势，又通过局部卷积增强了对邻近帧关系的捕捉能力。

3 实验效果及分析

为验证本文方法的有效性，在 NTU-RGB+D60^[6] 数据集上并与一些其他同类型的方法进行比较，评估本文方法的先进性。

3.1 数据集

本文在 NTU-RGB+D60 数据集上进行实验，该数据集包含 60 类动作的 56,880 个视频片段。针对 3D 骨架动

作识别任务，主要采用两种标准评估：跨主体 (CS) 协议将不同受试者分配到训练集和测试集；跨视角 (CV) 协议则区分不同摄像机视角下的数据。实验以 top-1 分类准确率作为评估指标，分别在 CS 和 CV 两种设置下验证模型性能。该数据集提供多模态数据，本研究仅使用其中的 3D 关节坐标信息进行动作识别研究。

3.2 训练过程

本文提出的模型采用批处理大小为 64，使用 SGD 优化器进行训练。优化过程中采用交叉熵损失函数计算梯度，并设置 0.0005 的权重衰减系数以防止过拟合。该配置确保了模型训练的稳定性和收敛性。

3.3 实验结果

如表 1 所示，本文在 NTU-RGB+D60 数据集上评估模型性能，分别在跨目标 (CS) 和跨视图 (CV) 两种基准设置下进行测试。

表 1 NTU-RGB+D60 数据集的识别精度比较

Method	CS(%)	CV(%)
JDFG-STGCN ^[2]	90.1	93.2
MSAA-GSRU ^[3]	86.6	94.0
MRFAM-CDAGC ^[7]	89.1	96
SparseShift-GCN ^[8]	91.8	94.3
MS-SlowFastGCN ^[9]	91.5	93.8
ASGC-STT ^[10]	92.70	96.90
DSTANet ^[11]	92.70	96.90
CDGT	92.87	97.52

本文以 ASGC-STT^[10] 为基线模型进行对比实验，本方法在数据集上展现出显著优势。在跨目标 (CS) 和跨视图 (CV) 基准上，识别准确率分别提升 0.17% 和 0.62%，充分验证了该模型在骨架动作识别任务中的优越性能。

3.4 实验结果分析

本文提出的网络模型通过双通道注意力机制，同时聚焦于骨架关键节点特征和节点间梯度信息。该模型采用时空增强策略，通过顺序的空间关系建模和时间动态捕捉，有效提升了骨架动作的识别精度。

4 结论

本文提出中心差分图变换网络 (CDGT)，该架构由中心差分图卷积 (CDGC) 和时空变换器 (STT) 组成。CDGC 模块创新性地融合了关节节点特征和骨骼梯度信息，突破了传统图卷积的局限性，显著提升了特征表达能力。STT 模块通过关节标记嵌入、空间注意力机制和时间注意力机制三个子模块，有效建模了骨架序列的时空动态特

征。该模型以关节和骨骼运动数据为输入，在 NTU-RGB+D60 数据集上的实验验证了其优越的识别性能。

参考文献

[1] 何璠. 基于注意力机制骨架动作识别的算法研究[D]. 江西:江西理工大学, 2022.

[2] 赵俊男、余青山、孟明、陈云. 基于关节点距离特征图的骨架动作识别方法[J]. 计算机学报, 2023, 46(3): 567-576.

[3] 杨世强, 杨江涛, 李卓, 王金华, 李德信. 基于LSTM神经网络的人体动作识别[J]. 图学学报, 2021, 42(2): 174-181.

[4] 刘红, 赵亮等. 语义引导下多流自适应拓扑推理图卷积网络的人体动作识别[J]. 广东工业大学学报, 2022, 39(6): 88-95.

[5] 李扬志, 袁家政, 刘宏哲等. 基于时空注意力图卷积网络的人体骨架动作识别算法[J]. 计算机学报, 2021, 44(5): 937-949.

[6] 金成铭, 应忍冬, 王伟行等. 基于三维双流网络的深度图人体动作识别[J]. 信息技术, 2022(04): 44-48+53.

[7] 白杉, 冯秀芳. 基于注意力增强的中心差分自适应图卷积的骨架行为识别[J]. 国防科技大学学报, 2023, 45(4): 177-184.

[8] 瞿生根, 黄方怡, 孙界平. 稀疏移位图卷积网络的轻量化骨架动作识别[J]. 软件学报, 2022, 33(5): 1617-1629.

[9] 瞿生根, 黄方怡, 孙界平. 基于多尺度慢快图卷积的轻量化骨架动作识别[J]. 软件学报, 2022, 33(12): 4117-4129.

[10] 王磊, 张宇, 李明等. ASGC-STT: 基于自适应空间图卷积和时空 Transformer 的人体行为识别[J]. 计算机应用研究, 2024, 41(12): 1-10.

[11] 王磊, 张宇, 李明等. 解耦时空注意网络 (DSTANet) 的骨架动作识别[J]. 自动化学报, 2024, 50(8): 1895-1906.

基金项目: 江西工程学院科学技术研究课题: 基于骨架动作识别的中心差分转换器图卷积(2024-JGKJ-24).

作者简介: 何璠(1997-), 女, 江西省新余市人, 江西工程学院人工智能与数据科学学院教师, 计算机技术专业硕士, 研究方向为模式识别。