

基于双维度拓扑优化超图卷积网络的骨架动作识别^①



黄国辉, 罗会兰

(江西理工大学 信息工程学院, 赣州 341099)

通信作者: 罗会兰, E-mail: luohuilan@sina.com

摘要: 图卷积网络 (GCN) 在基于骨架的动作识别中表现出色, 但由于关节间距离较大和现有方法特征聚合能力有限, 识别精度受到限制. 为了解决这一问题, 本文提出了一种时序与关节的双维度拓扑优化的超图卷积网络 (DDTO-HGCN), 用于骨架动作识别. 该方法通过超图理论扩展传统图结构, 利用超边提取局部与全局信息, 捕捉高阶依赖关系. 设计了时序拓扑优化超图卷积 (TTO-HGC) 和关节拓扑优化超图卷积 (JTO-HGC), 分别在时间维度和关节维度优化拓扑结构, 增强特征表示能力. 此外, 结合多尺度时序卷积网络 (MS-TCN) 丰富了时序特征表达, 并通过四流集成方法将骨架中心的相对向量作为补充数据流, 提升识别性能. 在 NTU RGB+D 和 NTU RGB+D 120 数据集上的实验结果表明, 所提出的方法优于现有一些先进方法.

关键词: 骨架动作识别; 超图卷积网络; 图拓扑优化

引用格式: 黄国辉, 罗会兰. 基于双维度拓扑优化超图卷积网络的骨架动作识别. 计算机系统应用, 2025, 34(9): 92-103. <http://www.c-s-a.org.cn/1003-3254/9952.html>

Dual-dimension Topology-optimized Hypergraph Convolution Network for Skeletal Action Recognition

HUANG Guo-Hui, LUO Hui-Lan

(School of Information Engineering, Jiangxi University of Science and Technology, Ganzhou 341099, China)

Abstract: Graph convolutional network (GCN) excels in skeleton-based action recognition, but the recognition accuracy is limited by large inter-joint distances and the restricted feature aggregation capability of existing methods. To tackle this issue, this study proposes a dual-dimension topology-optimized hypergraph convolution network (DDTO-HGCN) for skeleton action recognition. The method extends traditional graph structures using hypergraph theory, leveraging hyperedges to extract local and global information while capturing high-order dependencies. A temporal topology-optimized hypergraph convolution (TTO-HGC) and a joint topology-optimized hypergraph convolution (JTO-HGC) are proposed to optimize the topological structure along the temporal and joint dimensions respectively, thereby enhancing feature representation capabilities. Additionally, a multi-scale temporal convolution network (MS-TCN) is incorporated to enrich temporal feature expression, and a four-stream ensemble method which utilizes relative vectors of the skeleton center as supplementary data streams is introduced, improving recognition performance. Experimental results on the NTU RGB+D and NTU RGB+D 120 datasets demonstrate that the proposed method significantly outperforms some state-of-the-art methods.

Key words: skeletal action recognition; hypergraph convolution network; graph topology optimization

^① 基金项目: 国家自然科学基金 (62361032); 江西省主要学科学术和技术带头人领军人才项目 (20213BCJ22004); 江西省研究生创新专项资金 (YC2023-S619); 江西省多维智能感知与控制重点实验室 (2024SSY03161)

收稿时间: 2025-01-13; 修改时间: 2025-02-12, 2025-03-18; 采用时间: 2025-03-24; csa 在线出版时间: 2025-07-25

CNKI 网络首发时间: 2025-07-28

人类动作识别,即识别和理解人类行为,对于许多现实世界的应用至关重要.它可用于视觉监视系统识别危险的人类活动,以及自主导航系统感知人类行为以实现安全操作.此外,它对于许多其他应用也很重要,例如视频检索、人机交互、娱乐和医疗康复等领域^[1-4].通常可以从RGB视频和骨架序列数据中识别人体行为,其中基于RGB视频的人体动作识别已被广泛研究.但利用RGB视频学习到的动作表示对于身体尺度、相机视角和背景干扰的变化并不稳健.近年来,由于深度传感器的发展及其对复杂背景的鲁棒性,基于骨架的人体动作识别备受关注.与RGB图像相比,骨骼数据仅包含多个人体关节的2D或3D位置的时间序列,提供了极其抽象和结构良好的信息.作为高级信息的骨架序列传达了关于人体运动的紧凑信息,可以为描述人类行为提供有效且稳健的表示.此外,骨架信息可通过多种方式获取,例如人体姿势估计方法^[5]和Kinect^[6]等移动传感器,成功地满足了实时动作识别的要求.目前骨架动作识别任务的总体工作流程和方法:将处理好的数据输入到基于递归神经网络(RNN)、卷积神经网络(CNN)和图卷积网络(GCN)中进行空间和时域建模,得到输入数据的运动特征表达信息,从而进行动作分类识别.

当前骨架动作识别的研究主要集中于图卷积网络(GCN)方法^[7-18],同时基于其他理论框架的时空建模方法也取得了显著进展^[19-21].MotionBERT^[19]提出了基于大规模多样化数据的预训练通用运动表征学习框架,其核心在于采用双流时空Transformer(DSTformer)网络构建运动编码器.该编码器通过多头自注意力机制实现了骨骼关节间远程时空关系的自适应建模,并创新性地将2D姿态序列提升至3D空间进行表示学习.MotionBERT的统一框架可同时应用于3D姿态估计、动作识别和人体网格恢复等多个下游任务,突破了传统方法单一任务建模的局限性.文献[20]提出了一种自监督学习范式,通过构建速度变化的正样本和运动破坏的负样本,结合骨架插值技术提升识别性能,为无监督学习提供了新思路.PoseC3D^[21]则采用关键点热图的三维堆叠表示骨架序列,利用3D-CNN进行时空特征提取,在识别精度和泛化能力等方面表现优异.与上述方法不同,本文聚焦全监督学习框架下的单一骨架动作识别任务,重点研究图卷积网络的性能优化.

在GCN中,图拓扑主导特征聚合,是提取代表性特征的关键.它们识别人类骨骼中的物理连接与其他

方法相比,通过捕捉关节之间更多的语义关系,表现出显著的性能提升.这些基于GCN的方法明显优于使用手工特征的方法.由于现实生活中的动作通常需要多个关节的协同运动,而这些关节在人体图结构中可能相距较远,以往的研究^[7,9,11,16]忽略非身体连接关节之间的相关联,仅关注人体物理结构的骨架连接,导致卷积核感受野受限,无法有效捕捉远距离关节的协同关系,影响了信息交换效率.Yan等人^[8]提出了时空图卷积网络(ST-GCN),将GCN应用于基于骨架的动作识别,实现了从骨架数据中自动学习时空模式.2s-AGCN^[9]结合人体的物理图与自适应学习的骨架拓扑,优化适用于不同动作样本的拓扑,从而聚合骨架特征信息可以为不同的GCN层和骨架样本以端到端方式自适应地学习图拓扑.Wen等人^[18]提出了基于主题的GCN,采用可变时间密集块架构.Li等人^[10]将动作链接和结构链接结合到广义骨架图中,以学习动作识别的时空特征信息.CTR-GCN^[11]提出在通道维度中获取特定图拓扑以执行骨架图卷积.SGN^[15]在网络输入中提供高层语义(关节类型和帧索引),以增强特征学习能力.MS-G3D^[14]提出统一的多尺度时空模块,结合多邻接GCN和MS-TCN,以利用时空维度中的直接信息流.ML-GCNet^[12]中提出的多级图卷积(MLGCN)能够全面建模人体的层次关系.AdaSGN^[22]针对消除无信息关节的问题,首先通过预训练多组数据驱动的变换矩阵,对空间维度进行不同数量关节的降采样,将相应的变换特征传递至空间建模模块,并输入到策略网络中以选择最佳关节数量.尽管这些方法取得了进展,但在某些场景下可能无法灵活高效地学习和优化骨架图.

人的动作通常与头、腕、踝关节等骨骼关节的相互作用和组合有关,同时这些关节对每个动作的影响可能不同.例如,“手指”和“头”之间的关系对于区分“挥手”和“摸头”很重要.然而,物理连接的关节图很难探索这两部分之间的依赖关系.因此,对于基于骨架的图卷积网络的人体动作识别,如何充分利用骨架数据的图结构获取高阶的语义信息,使网络可以得到更好的训练,任务性能可以得到进一步的提高.对于人类动作识别来说,结构上距离较远的关节之间以及相邻关节之间的关系是密切相关的.为了解决上述问题,受到超图神经网络^[23]的启发,本文提出了基于超图理论的骨架图拓扑结构,进行超图卷积.本文方法利用定义的超边关联多个关节,破除传统图拓扑限制,更好地捕捉不相邻关节联系,为动作识别提供精确描述.

ST-GCN^[8]通过利用时空图卷积,基于图结构建模关节间的关系来提取运动特征.但其定义的身体物理连接图难以捕捉非自然连接关节间的关系,限制了模型表达能力.为提升GCN骨架动作识别的性能,近期研究方法^[9,11,13,15]实现了骨架拓扑结构的自适应注意力机制学习.尽管这些方法可以突破身体物理图的不相关的关节间的局限性,但它们在图卷积过程中,对不同通道和时序维度上共享生成的固定拓扑结构,导致不同通道和时序帧中特征聚合的灵活性受限.考虑到不同时序帧反映动作执行的不同阶段,而不同通道表征多样化的运动特征,关节间的相关性会随特征类型和时间变化而动态变化,因此单一共享拓扑结构并非最优方案,限制了GCN的表达能力优化.为了能够提高GCN对骨架动作数据的表达能力和鲁棒性,本文提出利用数据本身从节点维度(JTO-HGC)和时序(TTO-HGC)维度并行生成自适应的细化拓扑结构来丰富网络整体的自适应骨架超图卷积表达能力,解决共享图拓扑结构限制了深度学习模型的学习拟合能力,从而实现模型的整体性能提升.为了获取关节和时序维度的骨架拓扑无法聚合的骨架节点之间相近的高级语义特征,本文借用文献^[24]的EdgeConv组件进一步优化模型聚合特征的能力,该结构在3D点云中被广泛使用.根据样本数据本身来学习拓扑结构,可以很好地得到样本间的相关共性和信息以及样本自身特定类别动作所需关联的骨架拓扑,利用深度模型去学习上述的特征信息可以较好地提高模型泛化能力.

本文在两个广泛使用的基于骨架的动作识别数据集NTU RGB+D 120^[25]和NTU RGB+D^[26]上进行了实验.实验结果表明,DDTO-HGCN在识别准确率方面优于部分最新的先进方法.本文的主要贡献如下.

(1) 提出了一种基于超图构建的人体非物理连接的图拓扑结构.

(2) 基于超图拓扑设计了超图卷积网络DDTO-HGCN,实现骨架动作识别.

(3) 实验结果表明,所提出的DDTO-HGCN在主流数据集NTU RGB+D和NTU RGB+D 120上显示出优秀的性能.

1 方法

本节详细说明由提出的DDTO-HGC Block组成的DDTO-HGCN,如何去解决传统ST-GCN仅包含人体物理连接的边集骨架图,以及介绍相关的经典图卷

积知识.我们提出利用超图理论HNN^[23]来提升基于骨架数据的动作识别任务的性能,同时动态建模具有时序和关节维度特异性的拓扑结构,根据输入样本数据学习获得不同帧之间的骨架拓扑结构和全局骨架数据共用的关节维度拓扑结构,使模型拥有更大的灵活度和更强的学习拟合能力.

1.1 骨架图卷积和超图理论

ST-GCN^[8]是一种用于学习图结构数据表示的通用且有效骨架数据的时空图卷积模型.首先,将骨架序列数据表示成骨架图 $G=\{V, E\}$,它是一个普通的图结构,其中 V 是关节的集合, E 是只关联两个关节的骨架边集合. $A \in \{0,1\}^{N \times N}$ 为骨架图的邻接矩阵,其中 $A_{ij}=1$ 表示节点 i 和 j 之间存在边.ST-GCN设计了一种分区策略,将每个节点 V_i 的邻域 N_i 分为3个子集:(1)节点 V_i 本身;(2)向心群,比根节点更靠近骨架重心的相邻节点;(3)否则为离心群.骨架中各关节的平均坐标被视为人体重心.ST-GCN的关键部分是空间图卷积运算,令 $X_{in} \in R^{C_{in} \times T \times N}$ 为输入特征,其中 C_{in} 表示输入通道数, T 表示时间长度, N 表示关节数.令 $X_{out} \in R^{C_{out} \times T \times N}$ 为输出特征,其中 C_{out} 表示输出通道数.空间图卷积如式(1)所示:

$$X_{out} = \sum_k^K M_k \odot \tilde{A}_k X_{in} W_k \quad (1)$$

其中, K 表示空间图卷积的内核大小,根据ST-GCN的分区策略,将 K 设置为3. $\tilde{A}_k = \Lambda_k^{-\frac{1}{2}} A_k \Lambda_k^{-\frac{1}{2}}$, A_k 表示第 k 个子邻域对应的邻接矩阵,其中的 A_k^{ij} 表示顶点 v_j 是否在顶点 v_i 的对应子集中, $\Lambda_k^{ii} = \sum_j (A_k^{ij})$ 是归一化对角矩阵表示关节的度; $W_k \in R^{C_{out} \times C_{in}}$ 是卷积运算的权重向量; M_k 是一个 $N \times N$ 的掩码,用来学习每个顶点的重要性和相关性; \odot 表示点积.ST-GCN在利用多层空间图卷积学习到骨架序列的空间信息后,进行时间维度的卷积,在时间维度上使用一维卷积,以学习各关节的时间信息.

传统的GCN方法中,节点之间的连接通常以成对形式存在.实际骨架的数据结构往往超越了这种成对连接的限制,呈现出更加复杂的关系模式.与传统图结构中所有边的度数被限制为2不同,超图可以利用度数不受限制的超边来编码更高阶的数据相关性(超越成对连接).相比于普通图结构,超图通过其灵活的超边能够同时包含多个顶点,从而表示更加复杂的关系^[23].在许多计算机视觉任务中,超图结构被用于建模数据

间的高阶相关性. 因此, 超图已被广泛应用于诸如分类和检索等计算机视觉任务中. 本文所提出的超图基于对人类动作的高语义认知, 能够获取丰富的关节连接信息, 从而增强特征聚合的能力.

1.2 骨架超图构建和超图卷积

对于骨架动作识别而言, 成对连接的普通图结构无法表达在动作过程中身体之间不存在物理连接的部分关节间的配合关系. 与所有边的度数强制为 2 的普

通图结构相比, 超图可以使用度数没有限制的超边对高阶数据相关性 (超越成对连接) 进行编码. 超图定义为 $g = (v, \mathcal{E})$, 其中包括顶点集 v , 超边集 \mathcal{E} . 超图可以用关联矩阵 $H \in |v| \times |\mathcal{E}|$ 表示:

$$h(v, e) = \begin{cases} 1, & \text{if } v \in e \\ 0, & \text{if } v \notin e \end{cases} \quad (2)$$

图 1 为基于人体运动时各关节 (25 个关节点) 之间的关联性的人为认知以及大量实验的择优结果.

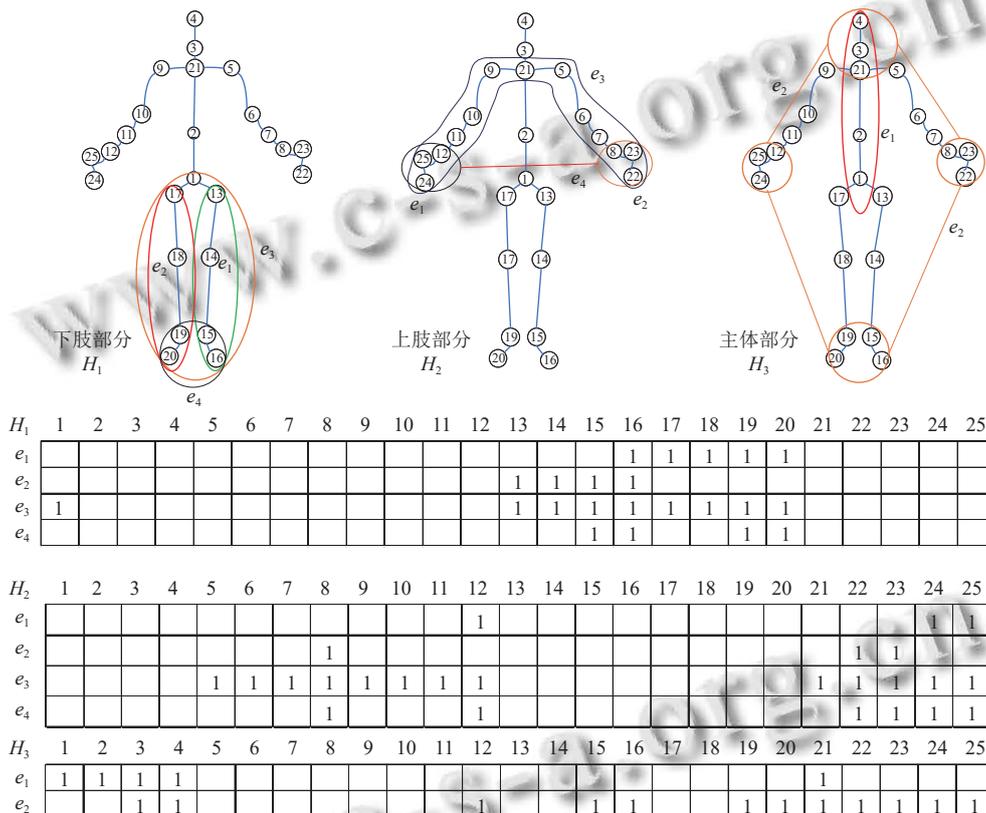


图 1 人体骨架 3 部分超图构建依据人体动作语义设置了 3 个部分的骨架超图, 构建关联矩阵 H 及可视化骨架的超边集

本文定义的超图拓扑基于人为认知, 具有较强的可解释性, 能解释每个超图节点和边所代表的具体意义, 有助于进一步优化网络结构和提升模型的可控性. 设计了 3 个部分的骨架超图, 分别对应上肢部分、下肢部分和主体部分, 利用超边对相应部位的多个关节进行关联. 同时实现了超边, 既包含了同一骨骼串连的关节对, 也关联了不存在骨骼连接的关节. 这种超图拓扑结构的设计打破了普通图拓扑中两个关节之间只有一条边相连的限制, 因此能够更好地表达非身体物理结构中各类部位关节的联系, 这对于表述动作任务中不相邻的关节响应有实质性的帮助. 上肢部分的

动作主要涉及手臂和手部末端关节的位置变化 (例如, “打电话”“刷牙”). 因此在构建超边时, 着重考虑了手臂和手部关节之间的联系. 这种超图的构建使得本文能够表达出普通身体结构图无法捕捉到的左右手关节之间的关系. 对于下肢部分也是同理, 下肢部分的动作空间位置改变的明显的地方在于下端的双脚关节以及下肢的两端腿部关节集 (例如, “踢东西”“侧踢”) 根据人的动作, 它们存在很强的关联性. 主体部分起到关注全局的作用, 两条超边分别包含主干以及联系到下肢和上肢部分关节的关联. 基于对人体执行动作的理解, 我们对该超图进行拓扑搭建, 这对于模型学习内部的

高级语义信息有显著的支持,可以很好地提高模型泛化能力.基于超图关联矩阵 H 搭建超图最终骨架数据的拓扑邻接矩阵 $A^H \in R^{N \times N}$, N 个关节的邻接矩阵,如式(3):

$$A^H = D_v^{-1} H D_e^{-1} H^T D_v^{-1} \quad (3)$$

该拓扑矩阵基于超边和关节的关联,有效利用了关节数据之间的高阶相关性特性.其中 $H \in R^{N \times E}$, N 个关节与 E 条超边的关联. $D_v \in R^{N \times N}$ 和 $D_e \in R^{E \times E}$ 分别表示超边度和顶点度的对角矩阵,在式(3)中起到归一化的作用.该邻接矩阵通过 H^T 和 H 来实现将超边

和关节关系最后聚合成关节之间的高阶关系.相较于式(1),超图卷积本质区别在式(4)所示,其与普通图卷积不同是图拓扑邻接矩阵的建模理论,超图卷积网络层的基本公式:

$$X^{(l+1)} = \beta(X^l A^H W^{(l)}) \quad (4)$$

1.3 DDTO-HGCN 整体模型结构

整体网络结构如图2所示,本文提出了双维度拓扑优化超图卷积块 DDTO-HGC Block,并利用它来构建深度超图卷积网络,捕获时空信息和高阶依赖关系,提高动作识别的性能.

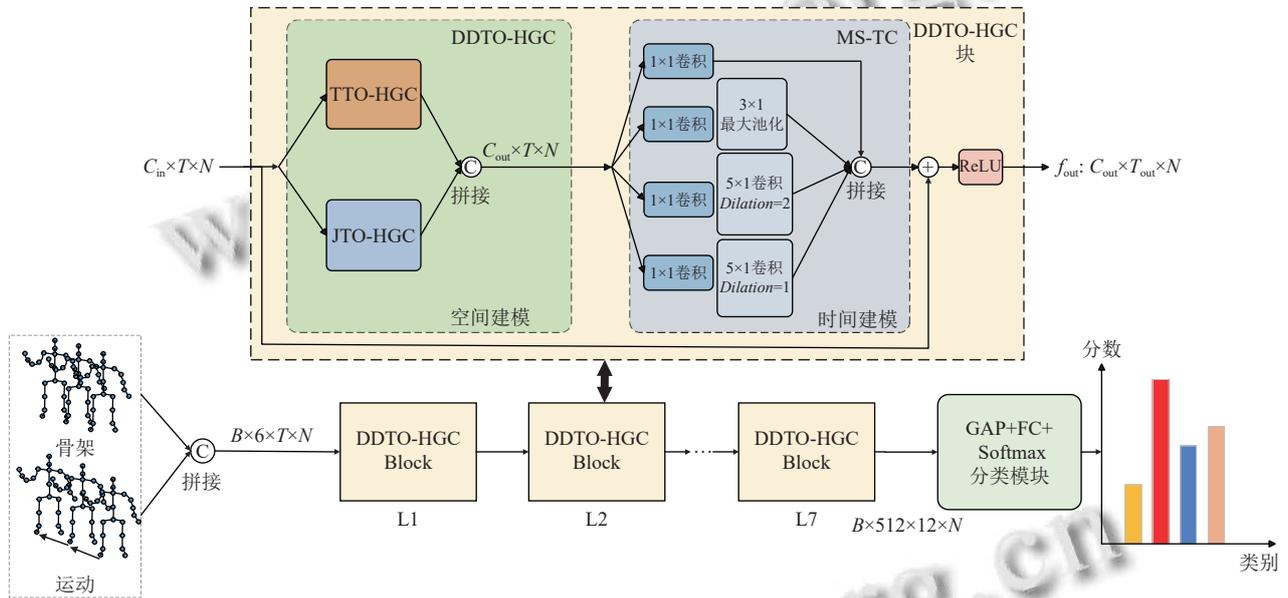


图2 DDTO-HGCN 网络结构

由图2可知,模型由7个DDTO-HGC块、1个全局平均池化层、1个全连接层和Softmax分类器组成. B 为样本批量大小,6为通道数, T 为时序帧数, N 为关节数.在经过7个DDTO-HGC块提取骨架和骨架运动(motion)序列数据(两者通过通道拼接整合)特征,获得具有判别力的高级特征表示后,在全局平均池化和全连接层的作用下,最后由一个Softmax处理得到各动作类别的预测分数,确定输出的动作类别,分类训练损失为二元交叉熵损失. DDTO-HGC Block分成空间建模和时间建模.在空间建模部分,首先并行使用时序拓扑细化超图卷积TTO-HGC和关节拓扑细化超图卷积JTO-HGC,这样能够使网络在时序和关节维度上学习到高级的语义拓扑结构来优化本文定义的骨架的超图拓扑,并且超图拓扑和网络一起进行训练.进一步提高超图卷积网络学习聚合特征的能力,从而提高网

络性能和高级语义的表达.然后将二者的结果在通道维度拼接汇总.

骨架特征在空间建模后输入到时间建模结构中,为了对具有不同持续时间的动作进行建模,使用了CTR-GCN^[11]中的多尺度时间卷积.如图1所示,该部分包含4个分支,每个分支都设有1个1x1卷积以减少通道维度.第1个分支在时间维度上使用了扩张率为1的卷积核大小为5的一维卷积,以获得局部运动信息,第2个分支在时间维度上使用了扩张率为2的卷积核大小为5的一维卷积,以获得更大感受野的局部运动信息,第3个分支在时间维度上使用了全局最大池化(MaxPool),以捕获全局运动信息,第4个分支用于保留输入信息.将4个分支的结果在通道上进行串接融合后,为了使得网络训练稳定高效,还使用了残差结构来保留输入特征,经过残差连接和激活后得到DDTO-

HGC 块的最终输出. DDTO-HGC 的总体计算过程如式 (5) 所示:

$$f_{out} = MT(Concat(G_J(f_{in}), G_T(f_{in})) \quad (5)$$

其中, G_J 、 G_T 、 MT 分别表示 TTO-HGC、JTO-HGC 和 MS-TC, $Concat$ 表示在通道维度上对并行的输出结果进行拼接. 在该整体网络中, 7 个模块的输出通道维度依次是 64-64-128-128-256-512-512, 分别第 3、5 和 6 块中的时间建模部分, 通过对时间建模部分的步长 stride 设置为 2 来对时序维度进行减半输出.

1.4 双维度拓扑优化超图卷积 (DDTO-HGC)

尽管超图定义了比传统图更有意义的节点关系, 但它可能仍然无法提取反映特征空间中所有节点之间相似性的样本智能键关系. 为了使得 DDTO-HGC 能够学习到泛化能力优秀的超图卷积网络, 采用 EdgeConv 作为 TTO-HGC 和 JTO-HGC 输出的拼接操作, 通过特征空间中的局部邻域图提取图特征. 使用 EdgeConv 网络可以提取骨架节点之间的相关性, 有利于获取除骨架拓扑学习到的特征外的高语义特征. 对于该方法, 如图 3 所示, EdgeConv 模块先平均池化时间维度提升计算效率, 获得融合整体时序的骨架信息, 然后通过基于欧几里得距离的 KNN 形成具有局部边缘的局部图, 并且基于图的局部边缘和恒等边缘通过可训练参数 $W_{edge} \in R^{C' \times 2C'}$ 进行聚合, 并将其与原始特征 X 进行通道维度的拼接. 最终的输出复制回到原有的时间维度得到 $F_{out} \in R^{C \times T \times N}$. 对于深度神经网络, 物理上接近的边缘会反映到最初的浅层, 但随着深度的加深, 特征空间中语义上相似的节点边缘之间的关系会被识别和学习.

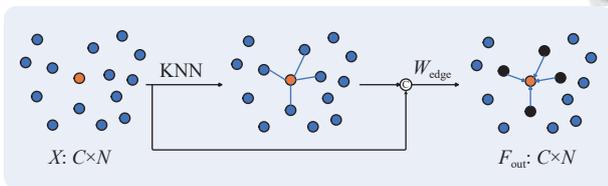


图3 EdgeConv 模块, K 设置为 5

1.4.1 时序拓扑优化超图卷积

基于超图卷积的理论框架, 生成与骨架数据相关的时序特定拓扑结构, 而非依赖单一的预定义拓扑. 该方法利用大型数据集的内在相关性和深度学习模型的强大学习能力, 有效地推广并解决任务, 从而提升超图卷积的性能. 鉴于人体关节运动的感知特性, 不同关节在不同时间点表现活跃, 同时相关联的关节也可能发生变化. 因此, 基于时间维度生成不同的骨架拓扑能够

帮助模型动态聚合骨架间的空间特征信息, 从而增强捕获高级语义特征的能力. TTO-HGC 模块的计算过程如式 (6) 所示:

$$f_{out} = Concat\left(\sum_{i=1}^K ((A_i^H + \alpha G_i^T T_s) X W), EdgeConv(X)\right) + X \quad (6)$$

该模块的输入和输出维度和实现细节如图 4 所示. 在虚线框中, $T_s \in R^{T \times 1 \times 1}$ 为第 3 分支的输出, 其中时序帧的关节空间特征通过卷积进行平均, 从而聚合跨时序帧的关节特征. 通过卷积和 Softmax 激活实现的维度变换, 为每个帧分配不同权重, 自适应地突出某些帧在动作识别中的重要性. 对于特定动作类别, 某些关键动作位置能够直接表明正在执行的行为 (从人类感知的角度来看, 可以从一张照片中判断某人正在做什么). 这一过程应用于 $G_i \in R^{T \times N \times N}$ 聚焦于不同时序帧的拓扑, 从而影响骨架超图卷积特征信息. 这是本文提出的时序骨架拓扑优化的独特细节. 将第 1.2 节定义的 3 部分骨架超图拓扑 A_i^H , K 是对应 3 个超图拓扑, 再和第 2 个分支得到的样本生成的时序特异性拓扑 G_i 进行一个自适应的权重比例 α 后相加, 让模型自身选择二者的重要性. 最后二者输出, 得到时序维度上的特异性拓扑 $R^{T \times N \times N}$ 和第 1 个分支输入骨架信息, 进行超图空间特征卷积后 X 经过 $W \in R^{C_{out}/2 \times C_{in}}$ 卷积得到 $C_{out}/2$ 维度的骨架特征, 在每个骨架图中进行逐时序帧聚合骨架空间信息 (矩阵批量乘法) 以获得输出.

在该模块中, 并行地实现了上述第 1.2 节 3 部分骨架超图拓扑的细化超图卷积. 其分别对输入的骨架数据样本进行超图拓扑细化, 在细化的最后, 对每个超图卷积后的特征进行相加聚合, 以更全面地捕捉输入数据的拓扑结构和聚合卷积到的高语义信息后, 再与输入数据 X 通过 EdgeConv 获得的特征张量进行通道维度的拼接归一化激活后, 通过残差连接保留输入特征获得最终的 f_{out} .

1.4.2 关节拓扑优化超图卷积

JTO-HGC 模块是为了学习细化关节之间的拓扑结构后进行骨架超图卷积, 通过骨架数据本身来生成在关节维度上具有动作关联意义的拓扑结构. 图 4 所示的 JTO-HGC 模块, 同理上述的 TTO-HGC 模块, 主要区别在于其利用骨架数据生成的动态拓扑的方式和应用的维度不同, 其通过并行的 2 个分支, 经过卷积和维度整合后将通道和时序维度整合进行自注意力机制的矩阵乘法生成关节细化拓扑 $G_j \in R^{N \times N}$, 经过 \tanh 激

活函数其相较于 Softmax 可以灵活的表达拓扑矩阵内部的关系, 和定义的超图拓扑 A_i^H 进行相加优化关节之间的拓扑结构信息, 后续为第 3 分支输入的骨架特征提供超图卷积, 优化聚合特征. 整体模块的核心过程如下:

$$f_{out} = \text{Concat} \left(\sum_{i=1}^K ((A_i^H + \alpha G_j^i) X W), \text{EdgeConv}(X) \right) + X \quad (7)$$

相应维度和实现细节都在图 4(a)、(b) 中标示, 最后整体的输出和 TTO-HGC 同理不再赘述.

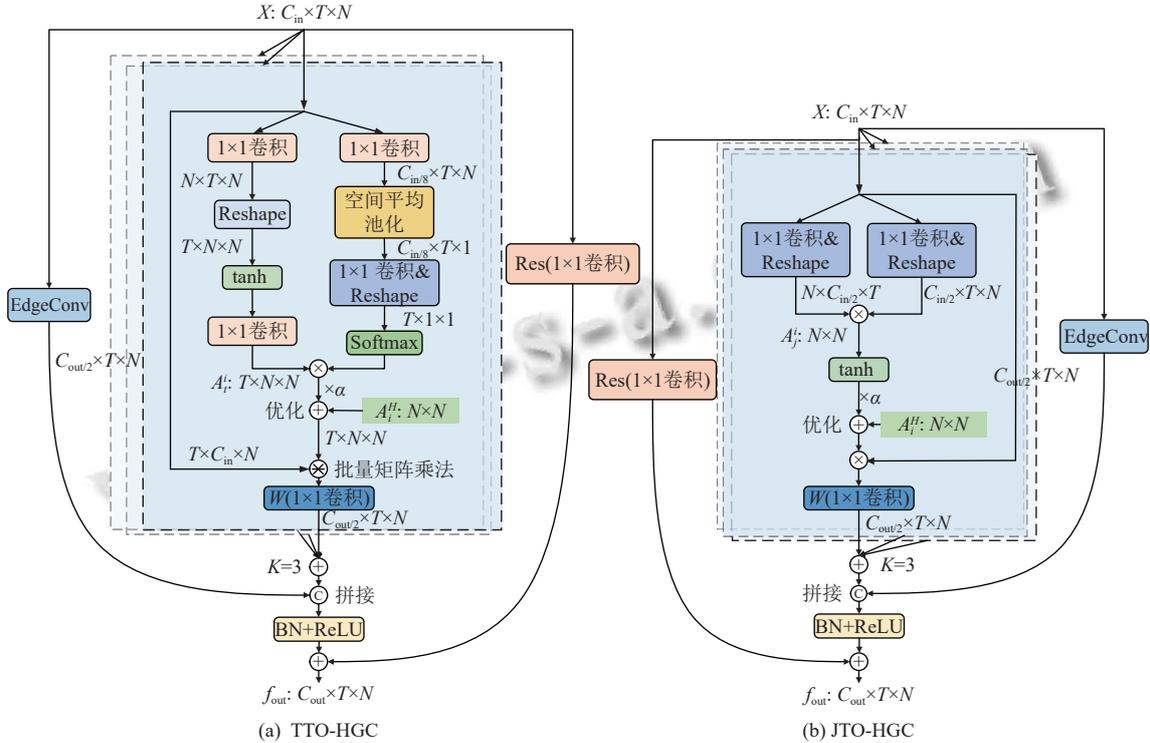


图 4 TTO-HGC 和 JTO-HGC 模块结构

1.5 利用相对中心流的四流融合

2s-AGCN^[9,13]使用了关节、骨骼、关节运动和骨骼运动流的四流集成方法中每一种模态数据独立网络训练. 因此, 为了减少模型参数量, 并且在一定程度上增加骨架和运动数据之间的交互, 本文提出了一种高效的多流数据融合策略, 通过在通道维度进行节点和对应的运动数据流、骨骼和对应的运动数据流的拼接, 形成单一流输入至 DDTO-HGCN, 有效降低了模型复杂度并增强了模态间交互. 在特征表示方面, 骨骼特征由相邻关节的三维坐标差值计算获得, 节点运动信息则通过相邻帧间关节的坐标变化量提取.

骨架相对于人体中心位置的变化, 对于动作识别可以起到补充原有骨骼和关节所忽略的一些动作信息, 提高模型的性能. 在关节、骨骼和运动流的基础上, 本文引入了相对节点数据、相对骨骼数据、相对运动数据. 如图 5 所示, 通过计算各节点与中心节点的相对坐标向量 $R(\text{joint+motion})$ 、各骨骼与中心骨骼的相对坐

标向量 $R(\text{bone+motion})$, 以及对应运动骨架与中心运动的相对向量, 有效补充了全局位置信息. 本文的模型 (DDTO-HGCN) 采用四流并行架构, 包括 (joint+motion)、(bone+motion)、 $R(\text{joint+motion})$ 和 $R(\text{bone+motion})$ 数据流. 各流独立训练测试后, 通过 Softmax 得分融合得到最终分类结果. 该设计补充了原有模态数据流, 丰富了特征提取提高模型性能.

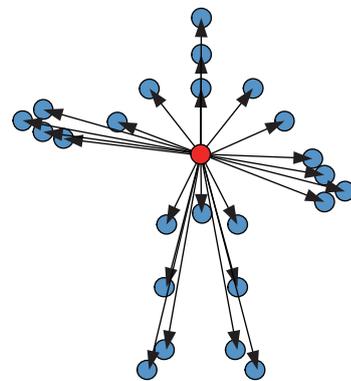


图 5 骨架相对中心的相对向量示意图

2 实验

2.1 骨架数据集

• NTU RGB+D: 是一个广泛用于骨架动作识别的大型数据集, 包含 56 880 个骨架动作样本, 由 40 位参与者完成, 分为 60 个类别. 该数据集的作者推荐了两种基准测试方法.

(1) 跨主体 (X-Sub): 40 位参与者中的 20 位用于训练, 其余 20 位用于验证.

(2) 跨视角 (X-View): 3 个摄像机视角中的两个用于训练, 另一个用于验证.

• NTU RGB+D 120: 是 NTU RGB+D 数据集的扩展版本, 增加了 60 类新动作, 包含 114 480 个视频. 相比 NTU RGB+D 数据集, 新增加了 57 367 个动作样本, 总计 114 480 个骨架动作样本, 涵盖 120 个动作类别, 由 106 位不同主体完成. 该数据集的作者推荐了两种基准测试方法.

(1) 跨主体 (X-Sub): 106 位主体中, 53 位用于训练, 剩余 53 位用于验证.

(2) 跨设置 (X-Set): 32 个设置中, 偶数设置的样本用于训练, 奇数设置的样本用于验证.

2.2 实验设置

在所有实验中, 本文将训练轮数设置为 90, 并在前 5 个轮次应用 warm-up 策略, 以实现更稳定的学习. 本文采用带有 0.9 Nesterov 动量的 SGD 优化器进行网络训练, 并设置权重衰减为 0.000 4, 初始学习率为 0.1, 在第 40、65 和 80 轮时以 0.1 的因子逐步降低学习率 0.000 1. 本文使用 CTR-GCN^[11]的数据预处理方法, 并将批量大小设为 64, 考虑到更多的帧数 T 能为样本提供更丰富的时空信息, 从而有助于提升模型性能, 本文将骨架视频帧的大小设置为 96. 所有实验均在 3 块 Tesla P100 GPUs 上进行.

2.3 与先进方法的对比

本文使用第 1.5 节描述的四流融合方法来提升模型的识别性能, 并将本文的方法与两个数据集 (NTU RGB+D 和 NTU RGB+D 120) 上的先进模型进行对比. 比较结果如表 1 所示.

表 1 DDTO-HGCN 在骨架动作识别数据集 NTU RGB+D 和 NTU RGB+D 120 上与其他先进方法的比较

方法	会议/期刊	年份	NTU RGB+D		NTU RGB+D 120	
			X-Sub (%)	X-View (%)	X-Sub (%)	X-Set (%)
ST-GCN ^[8]	AAAI	2018	81.5	88.3	70.7	73.2
AS-GCN ^[10]	CVPR	2019	86.8	94.2	78.3	79.8
2S-AGCN ^[9]	CVPR	2019	88.5	95.1	82.5	84.2
SGN ^[15]	CVPR	2020	89.0	94.5	79.2	81.5
MS-AAGCN ^[13]	TIP	2020	90.0	96.2	—	—
MS-G3D ^[14]	CVPR	2020	91.5	96.2	86.9	88.4
Shift-GCN ^[16]	CVPR	2020	90.7	96.5	85.9	87.6
Hyper-GNN ^[27]	TIP	2021	89.5	95.7	—	—
SD-HGCN ^[28]	ICONIP	2021	90.9	96.7	87.0	88.2
CTR-GCN ^[11]	ICCV	2021	92.4	96.8	88.9	90.6
AdaSGN ^[22]	ICCV	2021	90.5	95.3	85.9	86.8
MCC ^[20] (finetune)	ICCV	2021	89.7	96.3	81.3	83.3
SAMotif-GCN ^[18]	TPAMI	2022	91.7	96.7	88.4	88.9
FGCN ^[17]	TIP	2023	90.2	96.3	85.4	87.4
ML-STGNet ^[12]	TIP	2023	91.85	96.18	88.6	90.0
MotionBERT ^[19] (scratch)	ICCV	2023	87.7	94.1	—	—
MD ² TL-GCN ^[7]	电子学报	2024	92.64	96.47	89.29	90.49
DDTO-HGCN (joint+motion)	—	—	90.7	95.5	85.5	87.4
DDTO-HGCN (bone+motion)	—	—	90.8	95.3	86.8	88.0
2s-DDTO-HGCN	—	—	92.1	96.5	88.8	90.3
DDTO-HGCN R (joint+motion)	—	—	90.3	95.4	85.4	87.2
DDTO-HGCN R (bone+motion)	—	—	90.2	95.0	86.5	88.0
4s-DDTO-HGCN	—	—	92.61	96.74	89.44	90.83

表 1 中, 2s-DDTO-HGCN 指的是 (joint+motion) 和 (bone+motion) 这两个流的融合结果, 而 4s-DDTO-

HGCN 则是通过第 1.5 节中提到的相对向量方法整合获得模型四流融合的结果. DDTO-HGCN 几乎所有指

标上优于许多先进方法. 尤其是在 NTU RGB+D 120 数据集上, 与先进的 CTR-GCN^[11]相比, 本文的模型在 X-Sub 基准上的识别精度提高了 0.6%, 在 X-Set 基准上的识别精度提高了 0.2%. 与单一时空图卷积 ST-GCN^[8]相比, 分别提升了 18.7% 和 17.6%. 此外, 本文进一步比较了复现的 ST-GCN 和 DDTO-HGCN 在 NTU RGB+D 数据集中 7 个具有代表性 (涉及精细手部动作和需要捕捉全身协调运动的类别) 的“穿鞋”“阅读”“打字”“吃饭”“鼓掌”“搓手”和“踢”动作类中使用 (joint+motion) 流数据作为输入的 X-Sub 准确度, 如图 6 所示.

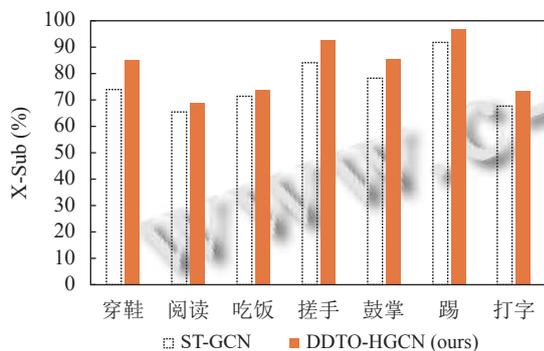


图 6 基于 (joint+motion) 流的 7 类动作识别精度对比

在所有 7 个动作类别的性能都显著优于 ST-GCN 模型. 以上的对比说明本文所提出的 DDTO-HGCN 可以更好地匹配不同关节响应的运动模式的动作类别, 并对骨架信息图卷积提取到更加全面的判别力特征, 侧面说明了本文超图卷积能够更好地建模关节间的复杂依赖关系.

2.4 四流融合

本文利用相对于人体中心的相对向量骨架信息, 并将相应的运动流沿通道维度进行拼接, 以形成本文的四流模型集成的数据流. 每个流的具体性能及其在本文模型中的集成融合结果如表 1 所示. 与表 1 中的其他方法进行对比, 本文方法即使是单流方法的性能, 也超越了某些多流集成的模型的整体表现. 表 1 中数据流 DDTO-HGCN $R(\text{joint}+\text{motion})$ 和 DDTO-HGCN $R(\text{bone}+\text{motion})$ 的性能相对弱于对应的 (joint+motion) 和 (bone+motion), 这可能与相对向量数据流的表达能力有关, 在某些动作的关节位移较小或相对位置变化不显著时, 模型可能未能充分捕捉到关键的时空特征. 通过补充该数据流, 帮助模型更全面地理解动作, 提高识别精度.

2.5 消融实验

为了验证本文模块的有效性, 本文重现了 7 层 ST-GCN, 记作 ST-GCN-L7, 并将其深度和通道维度对应 DDTO-HGCN 作为基准模型. 本文进一步在 ST-GCN 中添加了残差连接作为基本模块, 并将其时间卷积替换为第 1.3 节中描述的时间建模模块, 以确保公平比较. 输入数据遵循第 2.2 节中描述的预处理步骤, 帧大小设置为 96 帧. 本文的模型与基准模型 (baseline) 的性能基于第 1.5 节中描述的数据输入 (joint+motion). 性能指标为 NTU RGB+D 120 数据集上的跨主体 (X-Sub) 和跨设置 (X-Set) 分类准确率.

(1) 3 部分超图拓扑有效性. 本文复现了原始的 ST-GCN 和 CTR-GCN (两者基于 joint 数据流). 随后将原始物理身体连接图替换为本文提出的超图拓扑, 并将 ST-GCN 的图拓扑设置为可训练参数, 以确保对比实验的公平性. 实验在 NTU RGB+D 120 数据集的关节流数据上进行, 采用跨主体 (Cross-Subject) 和跨设置 (Cross-Setup) 分类准确率作为评估指标.

如表 2 所示, 相较于基准模型, 采用超图拓扑的 DDTO-HGCN 模型在性能指标上取得显著提升. 实验结果表明: 与传统物理连接图相比, 本文提出的超图拓扑能够更有效地提取骨架数据的特征信息; 基于人体动作认知定义的超图结构增强了模型对关节关系的理解能力, 为动作识别任务提供了更优的特征表示; 超图拓扑突破了传统物理连接图的局限性, 实现了更精细的关节信息交互, 从而提升了模型的训练效果和识别性能. 这证实了超图拓扑在骨架动作识别任务中的优越性, 其不仅具有更强的语义表达能力, 还能更有效地聚合信息特征, 为模型优化提供了有力支撑.

表 2 添加本方法提出超图拓扑的模型对比 (%)

方法	X-Sub	X-Set
ST-GCN	83.6	85.7
ST-GCN + hypergraph	84.0 (0.4 \uparrow)	86.2 (0.5 \uparrow)
ST-GCN_L7 (baseline)	85.0	86.9
ST-GCN_L7 + hypergraph	85.2 (0.2 \uparrow)	87.0 (0.1 \uparrow)
DDTO-HGCN (joint+motion) (ours)	85.5 (0.2 \uparrow)	87.4 (0.2 \uparrow)
DDTO-HGCN w/o hypergraph	85.3	87.2
CTR-GCN	84.5	86.2
CTR-GCN + hypergraph	84.8 (0.3 \uparrow)	86.6 (0.4 \uparrow)

(2) 验证 DDTO-HGC 模块的各个组件的有效性. 本文基于 baseline 用本文的模块替换掉其原有的空间图卷积部分, 进行了详细的消融对比实验. 其结果如表 3 所示.

表3 DDTO-HGC 各个模块组件的性能对比 (%)

方法	DDTO-HGC			X-Sub	X-Set
	JTO-HGC	TTO-HGC	EdgeConv		
ST-GCN-L7 (baseline)	—	—	—	85.2	87.0
A	√	—	√	85.30	87.14
B	—	√	√	85.42	87.21
C	√	√	—	85.35	87.34
DDTO-HGCN (joint+motion)	√	√	√	85.5	87.4

A 和 B 与 baseline 比较下, 性能都有提升, 这说明 JTO 和 TTO-HGC 模块在单独使用都能有效提高模型的性能, 验证了这两个模块的有效性. DDTO-HGCN 与 A、B 和 baseline 比较, 在 2 个指标下其性能都高于其他三者, 这说明本文模块 DDTO-HGC 从时序和关节维度出发动态生成的拓扑是能够综合学习到泛化能力更好的骨架拓扑信息, 进一步提升了模型的超图卷积能力和性能. C 和 DDTO-HGCN 进行比较, 说明加入的 EdgeConv 模块能够辅助 JTO 和 TTO-HGC 学习到高语义相邻的骨架特征让模型性能得到提升.

(3) 骨架时序帧数 T 的设置. 为了说明本文模型对输入帧数的选择, 本文对比了 baseline 模型和自己模型

在不同帧数设置下的表现, 结果见表 4. 随着帧数增加, 两个模型的准确性有所提升, 表明更多帧数提供了更丰富的时空信息, 有助于提高准确性. 然而, 当帧数增加到 $T=128$ 时, 精度提升趋于平缓, 这个结果可能是数据处理或模型能力的限制. 最后本文综合选择了 $T=96$ 作为骨架数据帧数. 在相同帧数下, 本文模型优于基准模型, 表明 DDTO-HGCN 能更有效地利用更多帧数的骨架数据, 优化时序帧拓扑和全局关节拓扑, 从而提升超图卷积的聚合能力和动作识别性能.

表4 在 NTU RGB+D 120 X-Sub 上不同骨架帧数 (T) 的效果对比 (%)

方法	64	80	96	128
Acc-baseline	84.90	84.84	85.20	85.22
Acc-DDTO-HGCN	84.88	85.32	85.50	85.56

(4) DDTO-HGC 模块的内部结构设置. 本文对比 4 种实例化模块结构的方法: DDTO-HGC-E、DDTO-HGC-F、DDTO-HGC-G 和确定的结构. 在这些方法中, JTO 和 TTO-HGC 分别以串联、并行、输出相加和通道维度拼接以及输入通道维度拆分的方式形成内部结构, 如图 7 所示, 精度对比结果如表 5 所示.

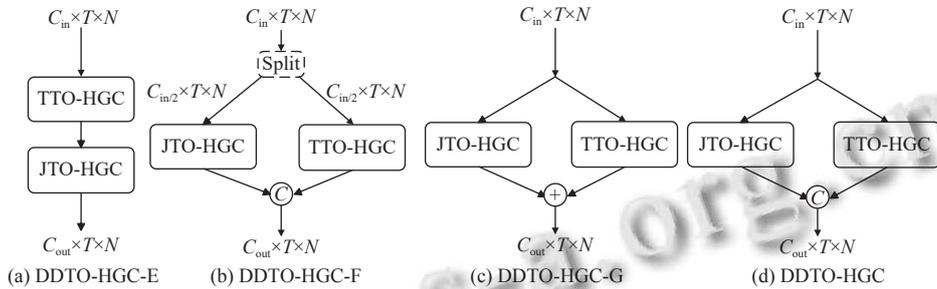


图7 DDTO-HGC 中 4 个不同内部结构实例

表5 DDTO-HGC 内部模块不同结构的对比 (%)

方法	X-Sub	X-Set
图7(a)	85.42	87.33
图7(b)	85.00	87.28
图7(c)	85.40	86.73
DDTO-HGCN (joint+motion)	85.50	87.40

本文的 DDTO-HGC 在两个指标上实现了相当的综合性能, 其余 3 个结构的性能相对较差. 这表明, 本文的结构设置较为完美地整合空间和时序拓扑信息所提取的特征, 验证了 JTO-HGC 和 TTO-HGC 两者联合时空建模的信息的重要性和有效性.

2.6 可视化学习的超图拓扑

本研究提出的超图网络旨在建模超越物理连接的

高级语义关系. 基于 X-Set 指标的 joint+motion 流数据, 图 8 通过彩色矩阵可视化了训练得到的人体 3 部分超图拓扑, 其中颜色亮度与关节连接强度成正比. 可视化结果表明: 图 8(a)、图 8(b) 矩阵呈现多个局部高亮区域, 而图 8(c) 矩阵则表现出更强的全局关键关节连接, 证实了本文定义的 3 部分超图能够同时捕获局部和全局关节关系. 与传统骨架图相比, 该超图显著提升了信息聚合能力和识别准确率, 验证了第 1.2 节基于人体动作认知定义超图的合理性.

通过对比第 1 和 5 层的图 8(a)、图 8(b) 和图 8(c) 矩阵发现, 随着网络层数加深, 关节关系得到更精细的区分和优化: 局部和全局高亮关节关系呈现更细腻的

区分;建立了更有效的关节连接;捕获了更高层次的语义信息.这表明本文定义的超图拓扑能够引导模型在

逐层优化中深化对关节关系的理解,促进信息流动,从而提升识别性能.

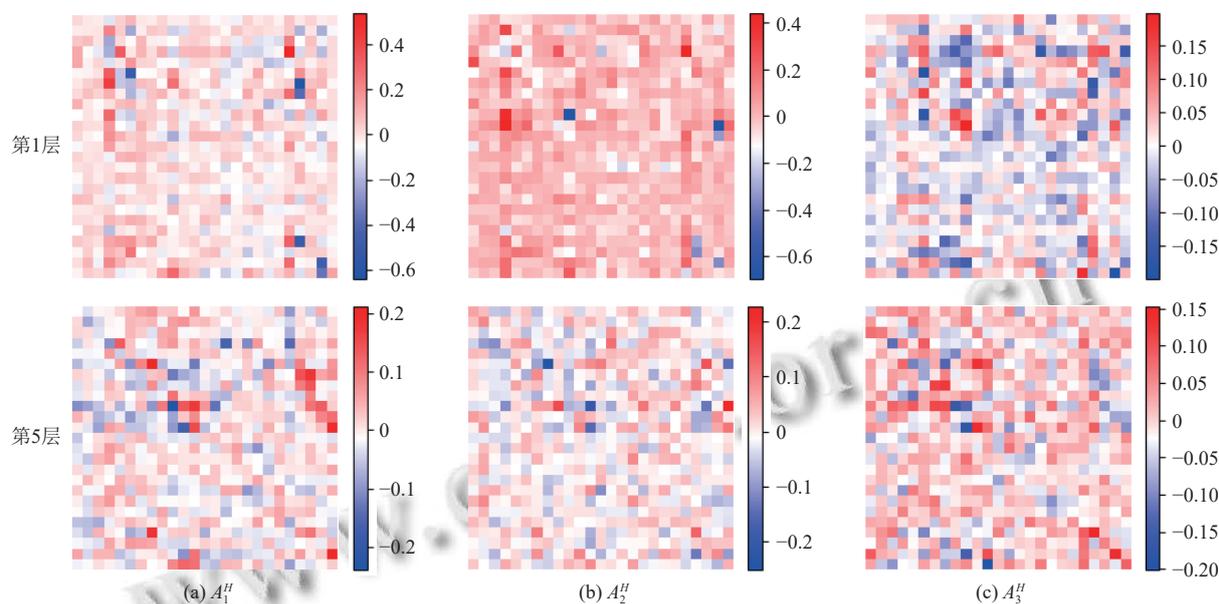


图8 DDTO-HGCN第1、5层学习到的3个超图 A_i^H

3 结论

在本研究中,提出了一种超图卷积网络 (DDTO-HGCN) 用于基于骨架的动作识别.本文引入了一种创新的骨架超图拓扑结构,突破了传统的物理连接限制,并利用超图理论对骨架进行建模.所提出的 DDTO-HGC 模块通过在时间和关节维度生成动态拓扑结构优化超图拓扑,聚合高层语义特征以提升模型性能.在多流融合中,引入了一种相对向量数据流,实现了四流融合.与以往的骨架数据流融合方式不同,该相对数据流方法提升了整体性能.本文方法在两个基准数据集上超越了多种先进方法.

参考文献

- 1 Li C, Zhong QY, Xie D, *et al.* Co-occurrence feature learning from skeleton data for action recognition and detection with hierarchical aggregation. Proceedings of the 27th International Joint Conference on Artificial Intelligence. Stockholm: AAAI Press, 2018. 786–792.
- 2 Jiang YG, Dai Q, Liu W, *et al.* Human action recognition in unconstrained videos by explicit motion modeling. IEEE Transactions on Image Processing, 2015, 24(11): 3781–3795. [doi: 10.1109/TIP.2015.2456412]
- 3 Gaur U, Zhu Y, Song B, *et al.* A “string of feature graphs” model for recognition of complex activities in natural videos.

Proceedings of the 2011 International Conference on Computer Vision. Barcelona: IEEE, 2011. 2595–2602.

- 4 罗会兰, 童康, 孔繁胜. 基于深度学习的视频中人体动作识别进展综述. 电子学报, 2019, 47(5): 1162–1173. [doi: 10.3969/j.issn.0372-2112.2019.05.025]
- 5 Cao Z, Simon T, Wei SE, *et al.* Realtime multi-person 2D pose estimation using part affinity fields. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 1302–1310.
- 6 Zhang ZY. Microsoft kinect sensor and its effect. IEEE Multimedia, 2012, 19(2): 4–10. [doi: 10.1109/MMUL.2012.24]
- 7 罗会兰, 曹立京. 基于多维动态拓扑学习图卷积的骨架动作识别. 电子学报, 2024, 52(3): 991–1001.
- 8 Yan SJ, Xiong YJ, Lin DH. Spatial temporal graph convolutional networks for skeleton-based action recognition. Proceedings of the 32nd AAAI Conference on Artificial Intelligence and the 30th Innovative Applications of Artificial Intelligence Conference and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence. New Orleans: AAAI, 2018. 912.
- 9 Shi L, Zhang YF, Cheng J, *et al.* Two-stream adaptive graph convolutional networks for skeleton-based action recognition. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2019. 12018–12027.
- 10 Li MS, Chen SH, Chen X, *et al.* Actional-structural graph

- convolutional networks for skeleton-based action recognition. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2019. 3590–3598.
- 11 Chen YX, Zhang ZQ, Yuan CF, *et al.* Channel-wise topology refinement graph convolution for skeleton-based action recognition. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE, 2021. 13339–13348.
- 12 Zhu YS, Shuai H, Liu GC, *et al.* Multilevel spatial-temporal excited graph network for skeleton-based action recognition. IEEE Transactions on Image Processing, 2023, 32: 496–508. [doi: [10.1109/TIP.2022.3230249](https://doi.org/10.1109/TIP.2022.3230249)]
- 13 Shi L, Zhang YF, Cheng J, *et al.* Skeleton-based action recognition with multi-stream adaptive graph convolutional networks. IEEE Transactions on Image Processing, 2020, 29: 9532–9545. [doi: [10.1109/TIP.2020.3028207](https://doi.org/10.1109/TIP.2020.3028207)]
- 14 Liu ZY, Zhang HW, Chen ZH, *et al.* Disentangling and unifying graph convolutions for skeleton-based action recognition. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 140–149.
- 15 Zhang PF, Lan CL, Zeng WJ, *et al.* Semantics-guided neural networks for efficient skeleton-based human action recognition. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2020. 1109–1118.
- 16 Cheng K, Zhang YF, He XY, *et al.* Skeleton-based action recognition with shift graph convolutional network. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 180–189.
- 17 Yang H, Yan D, Zhang L, *et al.* Feedback graph convolutional network for skeleton-based action recognition. IEEE Transactions on Image Processing, 2022, 31: 164–175. [doi: [10.1109/TIP.2021.3129117](https://doi.org/10.1109/TIP.2021.3129117)]
- 18 Wen YH, Gao L, Fu HB, *et al.* Motif-GCNs with local and non-local temporal blocks for skeleton-based action recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(2): 2009–2023. [doi: [10.1109/TPAMI.2022.3170511](https://doi.org/10.1109/TPAMI.2022.3170511)]
- 19 Zhu WT, Ma XX, Liu ZY, *et al.* MotionBERT: A unified perspective on learning human motion representations. Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision. Paris: IEEE, 2023. 15039–15053.
- 20 Su YK, Lin GS, Wu QY. Self-supervised 3D skeleton action representation learning with motion consistency and continuity. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021. 13308–13318.
- 21 Duan HD, Zhao Y, Chen K, *et al.* Revisiting skeleton-based action recognition. Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022. 2959–2968.
- 22 Shi L, Zhang YF, Cheng J, *et al.* AdaSGN: Adapting joint number and model size for efficient skeleton-based action recognition. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021. 13393–13402.
- 23 Feng YF, You HX, Zhang ZZ, *et al.* Hypergraph neural networks. Proceedings of the 33rd AAAI Conference on Artificial Intelligence. Honolulu: AAAI, 2019. 3558–3565.
- 24 Wang Y, Sun YB, Liu ZW, *et al.* Dynamic graph CNN for learning on point clouds. ACM Transactions on Graphics (TOG), 2019, 38(5): 146.
- 25 Liu J, Shahroudy A, Perez M, *et al.* NTU RGB+D 120: A large-scale benchmark for 3D human activity understanding. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(10): 2684–2701. [doi: [10.1109/TPAMI.2019.2916873](https://doi.org/10.1109/TPAMI.2019.2916873)]
- 26 Shahroudy A, Liu J, Ng TT, *et al.* NTU RGB+D: A large scale dataset for 3D human activity analysis. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 1010–1019.
- 27 Hao XK, Li J, Guo YC, *et al.* Hypergraph neural network for skeleton-based action recognition. IEEE Transactions on Image Processing, 2021, 30: 2263–2275. [doi: [10.1109/TIP.2021.3051495](https://doi.org/10.1109/TIP.2021.3051495)]
- 28 He CX, Xiao C, Liu ST, *et al.* Single-skeleton and dual-skeleton hypergraph convolution neural networks for skeleton-based action recognition. Proceedings of the 28th International Conference on Neural Information Processing. Sanur: Springer, 2021. 15–27.

(校对责编: 张重毅)