E-mail: jig@aircas.ac.cn
Website: www.cjig.cn
Tel: 010-58887035
中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS

中图法分类号:TP391　　**文献标识码**: A　　**文章编号**: 1006-8961(2025)11-3604-13

论文引用格式:Wu Z Z, Wan L, Hong F H, Tang Z D, Sun F, Zou L and Wang X F. 2025. Graph convolutional network integrating skeleton large kernel operators and global context information. Journal of Image and Graphics, 30(11):3604-3616(吴志泽, 万龙, 洪芳华, 汤正道, 孙斐, 邹乐, 王晓峰. 2025. 融合骨架大核算子和全局上下文信息的图卷积网络. 中国图象图形学报, 30(11):3604-3616)[DOI:10.11834/jig.240353]

融合骨架大核算子和全局上下文信息的图卷积网络

吴志泽[1],万龙[1],洪芳华[2],汤正道[3],孙斐[1],邹乐[1],王晓峰[1*]

1. 合肥大学人工智能与大数据学院,合肥　230601; 2. 合肥市公安局网安支队,合肥　230001;
3. 安徽省产品质量监督检验研究院,合肥　230051

摘　要:**目的**　骨架数据不仅体量轻巧,而且其内在的拓扑结构与图卷积网络(graph convolution network, GCN)高度契合,基于图卷积网络的骨架人体行为识别技术在行为识别领域得到广泛关注。然而,传统图卷积难以有效建模远距离节点关系,从而限制了其在复杂动作识别中的表现,针对这一问题,提出一种融合骨架大核算子和上下文信息的骨架图卷积网络(skeleton large-kernel and contextual GCN,SLK-GCN)。**方法**　该方法从两种不同的角度实现空间特征的增强。首先设计一种新颖的骨架大核卷积算子(skeleton-large kernel convolution, SLKC),通过扩大感受野并增强通道适应性,以增强空间特征提取能力。具体而言,SLKC通过引入大核卷积网络,模拟节点之间的远程依赖关系,从而提升模型在处理空间复杂性时的表现。同时,SLKC利用扩展的感受野捕捉更多的全局信息,增强特征提取的深度和广度。此外,引入轻量级的全局上下文建模模块(global context modeling, GCM),该模块能够自动学习和适应骨架拓扑结构,并从全局视角整合上下文特征。GCM通过捕捉不同节点之间的全局关系,进一步提升了模型的表征能力和鲁棒性。**结果**　所提出的SLK-GCN在NTU RGB+D、NTU RGB + D 120和Northwestern-UCLA数据集上的准确率分别为96.8%(最高)、91.0%和96.8%(最高),实验结果表明,SLK-GCN在人体行为识别任务中表现出了显著的优势。**结论**　SLKC与GCM的引入和结合,使得SLK-GCN在处理复杂骨架数据时能够更加有效地提取和利用空间特征。

关键词:人体骨架;行为识别;图卷积网络(GCN);上下文建模;大核卷积

Graph convolutional network integrating skeleton large kernel operators and global context information

Wu Zhize[1], Wan Long[1], Hong Fanghua[2], Tang Zhengdao[3], Sun Fei[1], Zou Le[1], Wang Xiaofeng[1*]

1. School of Artificial Intelligence and Big Data, Hefei University, Hefei　230601, China;
2. Cybersecurity Division, Hefei Public Security Bureau, Hefei　230001, China;
3. Anhui Provincial Institute of Product Quality Supervision and Inspection, Hefei　230051, China

Abstract: **Objective** Skeleton data have become a prime candidate for use with graph convolutional network (GCN) because of their lightweight nature and intrinsic topological structure. The alignment between skeleton data and GCN has

收稿日期:2024-06-24;**修回日期**:2024-11-01;**预印本日期**:2024-11-08

*通信作者:王晓峰　xfwang@hfuu.edu.cn

基金项目:国家自然科学基金项目(62406095);安徽省自然科学基金项目(2308085MF213);安徽省重点研发计划资助(2022K07020011);安徽省高校科学研究创新团队项目(2022AH010095);合肥市揭榜挂帅项目(2023SGJ011)

Supported by: National Natural Science Foundation of China (62406095); Natural Science Foundation of Anhui Province, China (2308085MF213); Key R&D Program of Anhui Province, China (2022K07020011); Innovation Team of the Anhui Higher Education Institutions of China (2022AH010095); Hefei Key Technology R&D "Champion-Based Selection" (2023SGJ011)

attracted considerable attention in developing human action recognition techniques based on these data. These techniques leverage the strengths of GCN to interpret the skeletal structures and movements inherent in human actions. However, traditional graph convolution methods encounter challenges in effectively modeling long-range node relationships, which are crucial for accurately recognizing complex actions. This limitation arises from the inherent design of conventional graph convolutions, which typically focus on local neighborhood information and struggle with capturing dependencies between distant nodes in the graph. **Method** To overcome this challenge, we propose a novel approach called skeleton large-kernel and contextual GCN (SLK-GCN). This innovative network aims to enhance spatial features from two distinct perspectives, which improves the capability to model long-range dependencies and capture the complexity of human actions effectively. The first key component of our SLK-GCN is the skeleton-large kernel convolution (SLKC) operator. This operator is designed to expand the receptive field and enhance channel adaptability, which leads to improved spatial feature extraction. Traditional convolutional kernels have limited capability in capturing extensive spatial relationships due to their relatively small receptive fields. By contrast, SLKC employs large kernel convolution networks, which significantly broaden the receptive field. This broader receptive field allows the model to simulate long-range dependencies between nodes effectively. In this way, SLKC enhances the capacity of the model to handle the spatial complexities inherent in human action recognition tasks. The large kernel approach not only captures a wide array of spatial information but also ensures that the extracted features are comprehensive and nuanced, which contributes to enhanced overall model performance. In addition to SLKC, we introduce a lightweight global context modeling (GCM) module as the second key component of SLK-GCN. The GCM module is designed to automatically learn and adapt to the topological structure of the skeleton while integrating contextual features from a global perspective. Traditional models often fail to consider the global context; instead, they focus on local node interactions. However, capturing global relationships between nodes is essential for understanding the full scope of human actions, especially those involving complex movements that span multiple joints and limbs. The GCM module addresses this gap by capturing global relationships and contextual information across the entire skeleton. This integration of global context enhances the representational capacity and robustness of the model, which allows it to accurately interpret and classify various human actions. **Result** To validate the effectiveness of our proposed SLK-GCN, we conducted extensive experiments on several widely used datasets, including NTU RGB+D, NTU RGB+D 120, and Northwestern-UCLA. These datasets are well regarded in the field of human action recognition and provide a diverse set of scenarios and action types for comprehensive evaluation. Experimental results demonstrate that SLK-GCN exhibits significant advantages and effectiveness in human action recognition tasks. Specifically, the incorporation and combination of SLKC and GCM enable SLK-GCN to effectively extract and utilize spatial features when processing complex skeleton data. This enhanced feature extraction capability translates to improved accuracy and robustness in recognizing various human actions. The success of SLK-GCN can be attributed to several factors. First, the capability of the SLKC operator to simulate long-range dependencies ensures that the model captures the intricate spatial relationships between different parts of the skeleton. This capability is particularly important for recognizing actions that involve coordinated movements across multiple joints. Second, the integration of global context by the GCM module provides a holistic view of the skeleton, which enables the model to consider the broader context in which individual movements occur. This holistic perspective is crucial for accurately interpreting complex actions that cannot be understood solely through local interactions. Furthermore, the combination of SLKC and GCM in SLK-GCN represents a synergistic approach to feature enhancement. While SLKC focuses on expanding the receptive field and capturing long-range dependencies, GCM complements this feature by integrating global contextual information. Overall, these components ensure that SLK-GCN has a comprehensive understanding of local and global spatial features, which leads to superior performance in human action recognition tasks. The implications of our work extend beyond the immediate scope of skeleton-based human action recognition. The principles underlying SLK-GCN, particularly the emphasis on large receptive fields and GCM, can be applied to other domains where capturing complex spatial relationships is essential. For example, similar approaches could be adapted for use in gesture recognition, sign language interpretation, and even broader applications in computer vision and robotics where understanding spatial dependencies is critical. **Conclusion** The spatial feature enhanced graph convolutional network (SLK-GCN) represents an important advancement in the field of human action recognition. By addressing the limitations of traditional graph convolutions in modeling long-

range node relationships, SLK-GCN offers a robust solution for capturing the complexity of human actions. The innovative combination of SLKC and GCM enables SLK-GCN to effectively extract and utilize spatial features, which results in improved accuracy and robustness. Our extensive experimental validation on multiple datasets underscores the effectiveness of this approach, which highlights its potential for broader applications in understanding and interpreting complex spatial data.

Key words: human skeleton; action recognition; graph convolution network(GCN); context modeling; large kernel convolution

0 引 言

行为识别是计算机视觉领域备受关注的分类任务,涉及多种数据模态,包括RGB视频、深度图像、光流和骨架等(王帅琛 等,2022)。其中骨架数据是用少量关节坐标描述人体结构,传达人类行为信息(卢健 等,2023)。它计算高效,对复杂背景和不同条件更稳健,对于动作识别十分重要。近年来,鉴于图卷积(graph convolution network, GCN)在处理非欧几里得数据上的优势与人体骨架的非欧几里得特点,基于图卷积网络的骨架行为识别引起了广泛关注。

虽然图卷积网络在挖掘顶点之间的相互作用方面具有显著优势,但随着网络的加深,顶点之间的依赖关系逐渐趋于一致。具体而言,动作识别取决于时空图中节点之间的依赖关系,它们可以分为直接依赖关系和间接依赖关系。前者是指两个相邻节点之间的依赖关系,作为先验知识输入到网络。后者表示拓扑上不相邻关节之间的依赖关系,即远程依赖关系。例如,在“拍手”动作中,直接依赖由骨架显式表示。实际上,表示左手和右手的关节不是相邻的,但仍存在长期的间接依赖关系,特别在考虑节点之间的远程依赖关系时,传统图卷积由于难以有效建模远距离节点关系,导致捕捉到这些关系的难度较大,这导致在人体行为识别任务中的性能下降。此外,现有方法(Yan等,2018;Shi等,2019)局限于远距离建模等因素,对全局信息的表征能力也有待提高。

针对远程建模问题,Ye等人(2020)和Liu等人(2023)分别提出上下文建模和大核卷积,但是大核卷积局限于多流数据的使用,并且上下文建模难以有效地提取全局的空间拓扑结构,仍然存在提升的空间。因此,本文从空间通道拓扑方面,结合全局的上下文建模和大核卷积,进一步探索图卷积在空间维度上的远程建模方法,提出一种以通道拓扑优化图卷积网络(channel-wise topology refinement graph convolution network, CTR-GCN)(Chen等,2021b)为基线的基于空间特征增强方法,该方法以两种不同的角度实现空间特征的增强,首先是骨架大核卷积算子(skeleton-large kernel convolution, SLKC),它聚合了隐藏在大核注意间接依赖中的关节特征,即可以在使用多流数据的情况下建模联合相关性,捕获远程依赖关系,并学习通道自适应。利用大核算子的优势,建模骨架节点之间因距离而产生的间接依赖。其次是全局上下文建模(global context modeling, GCM)模块捕获顶点之间的微妙关系,根据节点之间的邻接矩阵,丰富全局上下文信息。

将GCM模块和SLKC模块相结合,进一步提出SLK-GCN(sheleton large-kernel and contextual GCN)用于人体骨架行为识别。在NTU RGB+D 60(Shahroudy等,2016)、NTU RGB+D 120(Liu等,2020a)上的大量实验表明提出模型的有效性。贡献总结如下:1)提出通道拓扑下的全局上下文建模模块,通过对输入数据维度的交叉变换并重塑,在极小计算开销下实现了全局上下文的高效建模;2)针对传统图卷积难以有效建模远距离节点关系这一问题,提出骨架大核算子,通过建模局部联合相关性,利用大核优势捕获远程依赖关系,同时学习通道自适应;3)探究全局上下文建模模块和骨架大核算子的最佳融合方案,寻求最佳权重配比,对不同的实验参数进行比较,寻找最适合本文方法的效果,提出SLK-GCN,同时在基于骨架的行为识别基准测试上取得了显著的效果。

1 相关工作

1.1 图卷积网络

为了有效处理图这类非欧几里得数据,学者们

逐步开发图卷积网络(GCN)。在不同的GCN变体中,Kipf和Welling(2017)提出的GCN因其简单性而广泛适用于多种任务中,特别是基于骨架的行为识别(Yan等,2018;Shi等,2019;Chen等,2021c;姜权晏等,2022)。但是该方法在时空连续性上表现欠佳,针对这一问题,Yan等人(2018)提出时空图卷积网络(spatial temporal graph convolutional network, ST-GCN),即从时空图卷积出发,有效解决人体骨架在时空连续性上存在的问题,为后续的工作奠定了基础。但对于骨架动作识别任务而言,这种方式难以有效建模远距离节点的关系,不利于得到精确的空间特征。本文提出GCM来建模远距离节点的关系,得到精确的空间特征。在此基础上,再进一步研究远距离节点关系建模与空间特征增强的关系。

1.2 基于GCN的骨架动作识别

GCN能够有效地处理像骨架数据(Liu等,2017)这样的不规则结构图。给定具有N个节点的骨架数据,图拓扑可以用$N\times N$邻接矩阵$\boldsymbol{A}$很好地表示。基于GCN的方法的关键在于图拓扑的设计,即邻接矩阵$\boldsymbol{A}$。最直接的方法是根据人体的物理连接定义一个固定的图,ST-GCN采用了这种方法。为了关注边,Cheng等人(2020a)还创建了一个可学习的掩码,将其与物理邻接矩阵相乘或相加。在上述方法中,邻接矩阵仅仅是预定义的,为了使图拓扑更加灵活,Li等人(2019)尝试针对不同的样本构建不同的图,但是在探究两个关节之间的依赖关系时,只考虑了两个关节的特征,而忽略了上下文关节的影响。Wu等人(2024)利用自注意力机制建模远距离节点关系,并利用GCN建模局部节点关系。在本文中,所有上下文连接的特征都与引入的GCM模块完全结合。通过这种方式学习的图可以更加鲁棒和富有表现力。

1.3 大核网络

在早期,代表性的Inception(Huang等,2020)认为更大的感觉野有利于优化模型,并且已经有一些相关的杰出工作,如Ding等人(2022)提出的大核网络,总结设计了大核模型的5条准则,并进一步提出重参数化卷积神经网络(convolutional neural network, CNN)模型,名为RepLKNet,内核大小为31×31。

虽然这些工作证明了大核模型的有效性,但它们通常用于处理二维图像数据。由于受特征维数和内部结构的限制,如何设计针对骨架数据的大核识别模型仍然是一个挑战。为了解决大核卷积给人体骨架行为识别任务带来的困难,Liu等(2023)将大核卷积引入骨架行为识别任务,用于建模远程依赖关系。然而其也面临着一些限制,如网络轻量化,以及无法精准识别特定动作。针对以上问题,本文将大核算子与人体骨架相结合,提出SLKC,不仅增大了模型的感受野,而且在轻量化的同时进一步改善了建模远距离节点关系。

2 方 法

模型主要框架如图1所示,模型由L个空间特征增强模块堆叠而成。在本节中,首先简要介绍基于骨架的动作识别所需要的预备知识。然后再详细说明空间增强图卷积网络的细节。

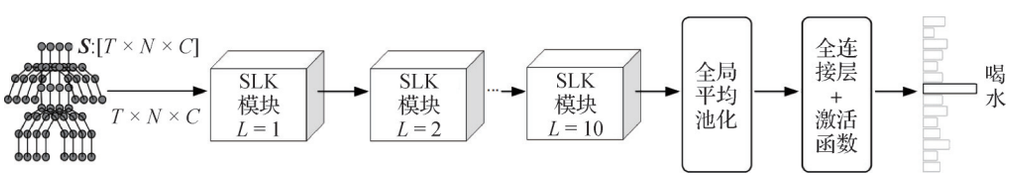

图1 SLK-GCN网络的整体框架

Fig. 1 Overall framework of SLK-GCN network

2.1 预备知识

2.1.1 符号

人体骨架表示为以关节为顶点、以骨架为边的图。该图表示为$\boldsymbol{G}=(\boldsymbol{V},\boldsymbol{\varepsilon})$,$\boldsymbol{V}=\{v_1,v_2,\cdots,v_N\}$表示人体骨架$N$个关节节点的集合,$\boldsymbol{\varepsilon}$表示为邻接矩阵$\boldsymbol{A}\in\mathbf{R}^{N\times N}$捕获的骨架边集,其中元素$a_{ij}$反映了节点$v_i$与节点$v_j$之间的连接强度,关节节点$v_i$的相邻区域统一表示为$\boldsymbol{N}(v_i)=\{v_j\,|\,a_{ij}\neq 0\}$。为了捕获精确的位置信息,根据ST-GCN(Yan等,2018)的研究,将邻接矩阵$\boldsymbol{A}$分为3个子集:根节点本身、向心子集和离心子集。$\boldsymbol{S}\in\mathbf{R}^{T\times N\times C}$代表一个人体行为的骨架序列,其中$T$表示一个人体行为的时间帧长度,$N$表示

节点数，C表示特征的维度，即$\boldsymbol{S}=\{S_1, S_2, \cdots, S_T\}$，表示为每一帧节点特征的集合，$S_i=\mathbf{R}^{N\times C}$表示某一帧的骨架节点特征。

2.1.2 图卷积

对于输入的一帧骨架数据S_i，拓扑共享的图卷积利用权重$\boldsymbol{W}$进行特征变换，通过邻接矩阵$\boldsymbol{A}$聚合该帧中每个节点的邻居节点特征，表示为

$$\boldsymbol{Z}_t=\sigma\left(\sum_{i=1}^{3}\hat{\boldsymbol{A}}_i S_i W_i\right) \tag{1}$$

式中，$\hat{\boldsymbol{A}}_i=\boldsymbol{D}_i^{-\frac{1}{2}}\boldsymbol{A}_i\boldsymbol{D}_i^{\frac{1}{2}}$是第$i$个子集的归一化邻接矩阵，$\boldsymbol{D}_i$为第$i$个子集的度矩阵，$\sigma(\cdot)$表示激活函数。对于静态方法，$\boldsymbol{A}$可以是手动定义或是可学习的参数，对于动态方法，$\boldsymbol{A}$通常由模型根据输入样本生成。$\boldsymbol{A}$定义为可学习的参数。

2.1.3 通道拓扑空间建模

并行使用3个通道拓扑模块以捕捉人体关节之间的相关性，然后将它们的结果汇总为输出，如图2所示。CTR-GC的工作流程如图3所示，目标是从输入骨架张量$\boldsymbol{S}\in\mathbf{R}^{T\times N\times C}$中提取图的特征，在时间维度上进行最大池化和平均池化，将池化后的结果按通道维度结合并成对相减来推断通道拓扑。

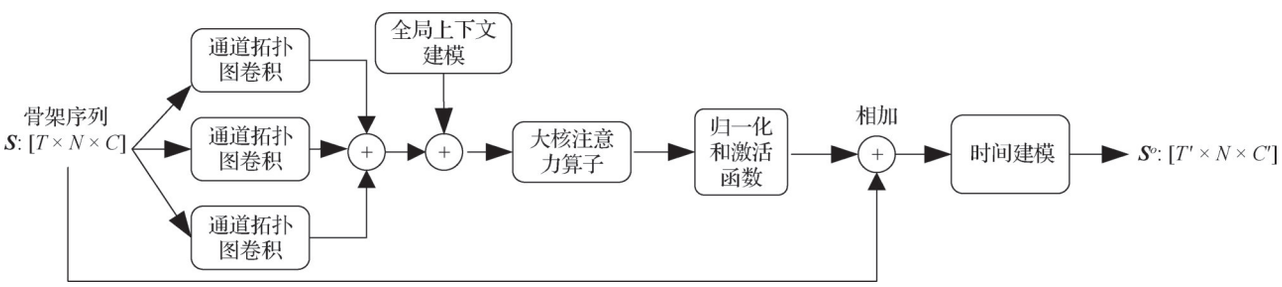

图2 SLK-GC模块

Fig. 2 SLK-GC module

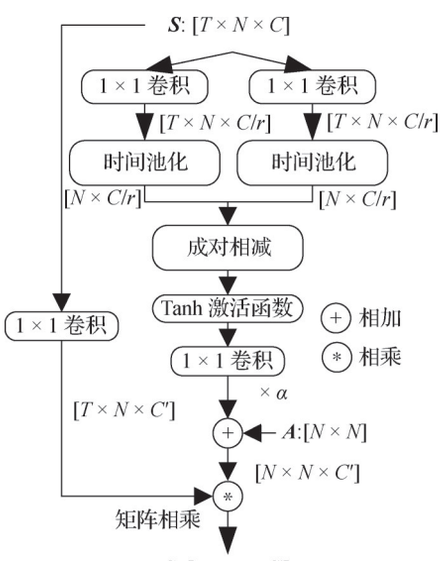

图3 通道拓扑空间建模

Fig. 3 Channel topology spatial modeling

具体而言，CTR-GC模块用于提取具有输入特征图的特征，首先由两个并行的双分支1×1卷积提取紧凑的表示，然后利用时间池化对时间特征进行聚合。之后，将特征进行成对减法和激活。随后用卷积操作提高激活的通道维度以获得特定于通道的相关性，这些相关性用于改进共享拓扑$\boldsymbol{A}$以获得通道拓扑。最后，在每个骨架图中进行逐通道聚合（通过批处理矩阵乘法实现）以获得输出表示。

$$M_1(\Psi(S_i), \Phi(S_j))=\sigma(\Psi(S_i), \Phi(S_j)) \tag{2}$$

式中，Ψ和Φ分别为图3双分支中的成对相减操作，$\sigma(\cdot)$表示激活函数。

2.2 全局上下文建模

本文尝试利用全局上下文建模来弥补预定义的邻接矩阵在对人体骨架数据进行空间建模时的全局感受野问题。受Dynamic-GCN（Ye等，2020）的启发，本文利用全局上下文进行建模，提出一种将图卷积与上下文建模结合的空间建模方法。

在图卷积网络中，邻接矩阵$\boldsymbol{A}$充分反映了图的拓扑结构，对应于不同骨架节点之间的依赖关系。然而，当$\boldsymbol{A}$被先验知识预定义时，拓扑信息是静态且有限的。目前的学习方法（Zhang等，2020；Duan等，2022）通常独立地预测两个关节之间的依赖关系，并手动设计函数（例如内积）将输入特征映射到这些依赖关系。与这些方法不同，本文利用全局上下文建模预测节点之间的依赖关系，它将整个特征映射作为输入，并直接预测完整的邻接矩阵$\boldsymbol{A}$。值得注意的是，本文探索了沿关节、时间和特征维度的上下文

信息，从而产生了更灵活和更具表现力的图拓扑。

全局上下文建模的体系结构如图4所示。给定的骨架序列映射 $\boldsymbol{S}\in\mathbf{R}^{C\times T\times N}$，首先通过两个1×1卷积层对每个关节的特征和时间维度进行压缩。接着，将联合维作为卷积通道，利用单个1×1卷积层将 N 维向量映射到 $N\times N$ 邻接矩阵中。从而在测量每对关节之间的依赖关系时，能够充分考虑所有其他关节的影响。随后，将拓扑表示为形状为(batch，N，N)的矩阵。此外，对邻接矩阵的每一行进行归一化，以简化优化过程。

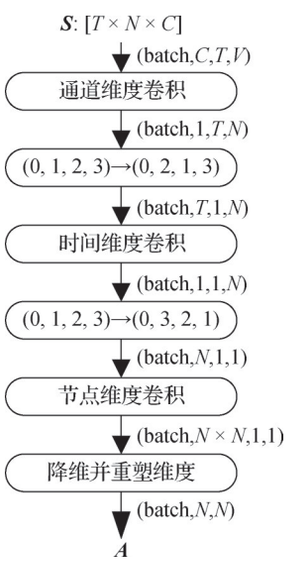

图4　全局上下文建模模块

Fig. 4　Global context modeling module

值得注意的是，全局上下文建模模块在学习每个样本和每个邻接矩阵时展现了其独特的特性。与传统的神经网络结构不同，其在处理图数据时不能简单地将图拓扑在不同样本之间共享。这样处理数据的好处在于，即使这些样本属于相同的操作类，图的拓扑结构也必须针对每个样本重新学习。这意味着全局上下文建模中的参数并非像传统网络中那样由手工制定的函数，而是以数据驱动的方式进行学习，不受任何预先设定的假设约束。

此外，该模块还通过将关节维度作为通道的方式处理全局上下文信息，使可训练的卷积核能够对所有关节的全局上下文信息进行编码。这种做法使其能够更好地捕捉图数据中的全局特征，并在学习过程中充分利用整个图的结构信息，从而提高其对复杂关系的建模能力和泛化能力。最后，将全局上下文建模模块预测的邻接矩阵作为其图拓扑表示馈入SLKC模块中。

$$\boldsymbol{A}=F\&R(Conv\_J(Dim\_trans_2(Conv\_T(Dim\_trans_1(Conv\_C(\mathrm{S})))))) \tag{3}$$

式中，$F\&R$ 代表对张量的降维和重塑，$trans_1$ 和 $trans_2$ 代表维度变换，而 $Conv\_J$、$Conv\_T$、$Conv\_C$ 分别代表对关节、时间以及通道部分进行卷积操作。

2.3　骨架大核算子

SLKC的体系结构如图5所示。首先，利用二维卷积改变通道数，这一步骤有助于调整特征图的深度，从而更好地捕捉和表征输入数据的抽象特征。接着，通过采用GELU激活函数对特征进行非线性变换，以提高模块的表达能力和拟合能力，从而使整个模型能够更有效地发挥性能。

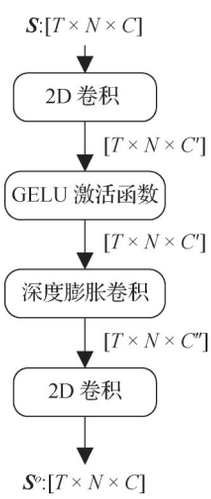

图5　骨架大核算子

Fig. 5　Skeleton large kernel operator

随后，利用深度膨胀卷积对模型前期全局上下文建模模块优化后的空间特征进行聚合。通过引入膨胀卷积，可以扩大特征图的感受野，这意味着能够更好地捕获全局信息，并在更大范围内进行特征信息的聚合。

$$\boldsymbol{S}^o=Conv(DWDConv(GELU(\mathrm{Conv}(\boldsymbol{S})))) \tag{4}$$

式中，$\boldsymbol{S}$ 为输入，$GELU$ 为激活函数，$DWDConv$ 为深度膨胀卷积，$Conv$ 为2D卷积，$\boldsymbol{S}^o$ 为输出。

由于使用了膨胀卷积，聚合的特征的空间感受野将会更大，这使得能够在更广泛的范围内捕获和整合特征信息，进而增强空间特征的表达能力和鲁棒性。因此，大核膨胀卷积的引入对于聚合空间特征和处理复杂任务具有重要意义。

2.4　多尺度时间建模模块

为了对不同持续时间的动作进行建模，本文采

用多尺度时间建模模块。如图6所示，该模块包含4个分支，每个分支包含一个1×1的卷积来减少通道维度。前3个分支分别包含两个时间卷积层，每个层具有不同的扩张率，并且一个最大池化层，然后经过一个1×1的卷积。最后，将4个分支的结果连接起来得到输出。

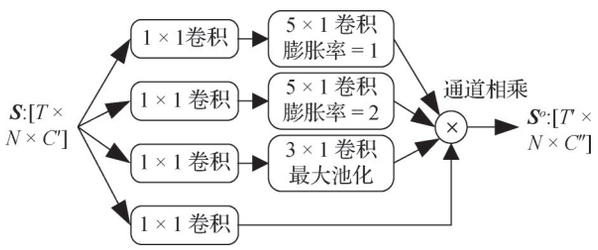

图6　多尺度时间建模模块

Fig. 6　Multi scale temporal modeling module

2.5　空间特征增强图卷积神经网络

将通道拓扑下的GCM和SLKC与其他的基础模块进行结合，形成SLK基本模块，通过堆叠L个SLK基本模块（如图1所示），进一步提出SLK-GCN用于骨架的人体行为识别。

对于输入的T帧人体骨架序列，分别将其传入GCM模块与SLKC模块，在GCM模块中提取出具有全局感受野的空间上下文建模后的邻接矩阵，将其与空间建模模块相加，随后，将输出的结果输入到SLKC当中，利用大核卷积聚合全局空间特征以实现空间特征增强，具体表示为

$$\boldsymbol{O} = SLKA(GCM(\boldsymbol{S}) + \boldsymbol{S}) \tag{5}$$

式中，GCM为全局上下文建模模块，$SLKC$为骨架大核卷积算子，$\boldsymbol{S}$为输入，$\boldsymbol{O}$为输出。

之后，将空间建模的结果传入时间卷积模块，用于建模不同帧之间的时间相关性。对于时间卷积，采用与CTR-GCN一致的结构。

在经过L层连续的空间与时间交替建模之后，最终的输出特征会依次经过全局平均池化层与全连接层来获得行为分类的分数。

3　实　验

为了验证本文方法在基于人体骨架的行为识别任务中的有效性，在NTU RGB+D（Shahroudy等，2016），NTU RGB+D 120（Liu等，2020a）以及Northwestern-UCLA（Wang等，2014）数据集上进行了大量的实验，识别结果采用Top-1准确度。

3.1　数据集

NTU RGB+D数据集包含60类动作，共56 880个样本，在划分训练集和测试集时使用两个标准：1）cross-subject（X-Sub）：指定的20个人作为训练集，其余的作为测试集。2）cross-view（X-View）：将1个摄像头采集的数据作为训练集，其余2个摄像头采集的数据作为测试集。

NTU RGB+D 120数据集是最大的可用人体运动3D联合数据集之一。该数据集通过添加57 367个骨架序列和60个额外的动作类别扩展了NTU RGB+D数据集。该数据集由106名志愿者用3个摄像头拍摄，共113 945个样本，并且沿用NTU RGB+D的划分标准。

Northwestern-UCLA数据集是用3个摄像头从多个角度捕获的，共1 494个视频片段，包含10个动作类别，由10个不同的对象执行。使用的训练集来自两个摄像头，测试集来自另一个Kinect摄像头。

3.2　执行细节

所有实验均在PyTorch深度学习框架的3090GPU上执行。整体模型通过连续堆叠10个SLKBlock以及一个全连接层组成，前4个块的输出通道为64，在第5与第8个块时，通道数分别是64的2倍与4倍，时间帧长度也会以2的倍数下采样。

损失函数选择交叉熵方法，模型使用动量（0.9）的随机梯度下降进行训练。在2个数据集上训练模型时，先对前5个epoch使用了预热策略。学习率设置为0.1，在第35个和第55个epoch时使用衰减，衰减率为0.1，在第65个epoch时结束训练。

对于NTU RGB+D 60&120数据集，使用与CTR-GCN相同的数据预处理，将batchsize设置为64，在UCLA数据集上则设置为32。

3.3　与最先进的方法的比较

本小节将提出的方法（SLK-GCN）与最先进的方法在3个公开数据集上进行比较。为了公平比较，本文方法同样采用多流融合策略，包含关节流、骨架流、关节运动流和骨架运动流这4个流的数据。其中关节流以原始骨架坐标为输入，骨架流以空间坐标的微分（骨架关节的二阶信息）为输入，关节运动流和骨架运动流使用对应数据的相邻帧之间的时间差作为输入。

在NTU RGB+D与NTU RGB+D 120数据集上，本文方法实验分别汇报关节流（Js）、骨架流（Bs）、关

节与骨架流(2s),以及4流融合(4s)的结果。为了验证方法的通用性,在NTU RGB+D、NTU RGB+D 120以及UCLA数据集上与最先进方法进行比较,结果如表1和表2所示。在3个大规模数据集上,除了NTU RGB+D数据集的X-View评估标准,本文方法在其余基准测试上均取得了最先进的性能。

表1　NTU RGB+D 60和120数据集上Top-1精度与最先进方法的比较

Table 1　Comparison of top-1 accuracy and state-of-the-art methods on NTU RGB+D 60 and 120 datasets

方法	出版者	NTU 60/%		NTU 120/%		参数量/M	浮点运算数/G
		Sub	View	Sub	Set		
2s-AGCN(Shi等,2019)	CVPR19	88.5	95.1	82.9	84.9	6.90	37.30
Dynamic-GCN(Ye等,2020)	ACMMM20	91.5	96.0	87.3	88.6	14.40	–
MS-G3D Net(Liu等,2020b)	CVPR20	91.5	96.2	86.9	88.4	2.80	48.80
Shift-GCN(Cheng等,2020b)	CVPR20	90.7	96.5	85.9	87.6	2.80	10.00
MST-GCN(Chen等,2021c)	AAAI21	91.5	96.6	87.5	88.8	12.00	–
DualHead-Net(Chen等,2021a)	ACMMM21	92.0	96.6	88.2	89.3	12.00	–
CTR-GCN(Chen等,2021b)	ICCV21	92.4	96.8	88.9	90.6	5.80	7.90
PSUMNet(Trivedi和Sarvadevabhatla,2022)	ECCV22	92.9	96.7	**89.4**	90.6	2.80	**2.70**
InfoGCN(Chi等,2022)	CVPR22	92.7	**96.9**	**89.4**	90.7	6.40	7.40
SMotif-GCN+TBs(Wen等,2023)	TPAMI23	90.5	96.1	87.1	87.7	–	–
ML-STGNet(Zhu等,2023)	TIP23	91.9	96.2	88.6	90.0	2.90	–
EfficientGCN-B4(Song等,2023)	IEEE23	92.1	96.1	88.7	88.9	**2.00**	15.20
FRF-GCN(Yun等,2024)	AAAI24	91.3	96.5	87.1	88.4	2.42	–
SLK-GCN(本文)		**93.0**	96.8	**89.4**	**91.0**	7.40	8.20

注:加粗字体表示各列最优结果。“–”表示未提供该项指标相关信息。

表2　Northwestern-UCLA数据集上Top-1精度与最先进方法的比较

Table 2　Comparison of top-1 accuracy and state-of-the-art methods on the Northwestern-UCLA

方法	出版者	精确度/%
AGC-LSTM(Si等,2019)	CVPR19	93.3
Shift-GCN(Cheng等,2020b)	CVPR20	94.6
DC-GCN+ADG(Cheng等,2020a)	ECCV20	95.3
CTR-GCN(Chen等,2021b)	ICCV21	96.5
FGCN(Yang等,2022)	TIP22	95.3
InfoGCN(4s)(Chi等,2022)	CVPR22	96.6
MV-IGNet(Wang等,2023)	TPAMI23	93.1
SLK-GCN(本文)		**96.8**

注:加粗字体表示最优结果。

具体而言,如表1所示,本文方法在NTU RGB+D数据集的corss-subject评估标准上,当采用4流融合时,相较于当前先进的方法InfoGCN(Chi等,2022)提升了0.3%,并且InfoGCN使用了两个额外的最大均值差异(maximum mean discrepancy,MMD)损失,而本文方法仅使用了交叉熵损失。需要注意的是,为了比较的公平性,仅与InfoGCN的4流融合结果进行比较。

对于NTU RGB+D 120数据集,如表1所示,在X-Sub评估标准上的性能超过基线模型CTR-GCN 0.5%,并与InfoGCN持平;在Cross-Set评估标准上,本文方法取得了最先进的性能,超过基线模型CTR-GCN 0.4%。表2中,在Northwestern-UCLA数据集的4流融合结果同样优于InfoGCN方法0.2%。通过在3个大规模数据集上与先进方法的比较,证明了本文方法的优越性。

3.4 消融实验

本节进行消融研究，验证提出的GCM、SLKC模块及其组合的有效性。所有消融实验都使用CTR-GCN作为基线，只展示在NTU RGB+D0Cross-Subject评估标准的4流融合结果，实验设置与3.2节描述的一致。

由表3得知，在单独加入SLKC模块后，参数量增加了约0.1 M，但精度提升了0.3%；单独加入GCM模块，参数量虽然提升了近0.4 M，但是精度提升了0.5%。将提出的两个模块融合后，模型的性能进一步提升，相较于基线提升0.6%。这同时也表明，两个模块之间可以共存且相互协作，共同增强模型的性能。

表3 NTU RGB+D数据集上的消融实验

Table 3 Ablation experiments on the NTU RGB+D dataset

方法	参数量/M	精确度/%
Baseline（CTR-GCN）	1.46	92.4
Baseline+SLKC	1.59	92.7
Baseline+GCM	1.83	92.9
Baseline+GCM+SLKC	1.85	**93.0**

注：加粗字体表示最优结果。

3.5 GCM模块与GCN融合方式对实验的影响

为了验证GCM模块融合到模型中的方式的有效性，对其融合方式进行了实验。具体如表4所示。表中的融合方式1为采取将GCM模块与GCN模块直接相加的融合方式。融合方式2为将GCM模块输出后的结果按通道维度与GCN模块进行拼接。由表4看出，方法1的融合方式无论是在精度还是参数量上效果都是最佳。产生这样的结果可能是因为这种融合方式还原了GCM模块原始的处理数据的方式，即当GCM模块提取出全局上下文的特征信息后，将原始数据直接与GCN的原始输出数据一同汇入时间卷积网络（temporal convolution network，TCN）当中，起到上下文特征信息补充的作用。而按照通道维度拼接可能一定程度上压缩了GCM处理信息的能力，限制了GCM模块的效果。所以方法1由于保留了原始提取的特征数据，虽然简单，但也是科学的处理初级信息的方式。

3.6 不同卷积核大小对实验结果的影响

为了验证提出的SLKC模块的有效性，分别在基线上使用了不同大小的卷积核，并以此为唯一变量分别进行实验，实验结果如表5所示。在基线中引入不同大小核的SLKC模块后，性能最好的卷积核大小为21时，效果最佳，其性能从92.4%改善到92.7%，相较于基线模型得到0.3%的提升，证明了提出的SLKC模块的有效性。结果表明，使用到的深度膨胀卷积对模型前期空间特征进行聚合的方法是可行的。同时表明通过引入膨胀卷积，可以扩大特征图的感受野，这意味着能够更好地捕获全局信息，并在更大范围内进行特征信息的聚合。由于使用了膨胀卷积，聚合的特征的空间感受野将会更大，这样就能够在更广泛的范围内捕获和整合特征信息，进而增强空间特征的表达能力和鲁棒性。因此，大核膨胀卷积的引入对于聚合空间特征和处理复杂任务具有重要意义。

表4 NTU RGB+D数据集上不同方法的X-Sub精度比较

Table 4 Comparison of X-Sub accuracy of different methods on the NTU RGB+D dataset

方法	X-Sub/%	参数量/M
Baseline	92.4	1.42
Baseline+GCM（融合方式1）	**92.9**	**1.83**
Baseline+GCM（融合方式2）	92.5	2.29

注：加粗字体表示各列最优结果。

表5 NTU RGB+D数据集上不同核大小的X-Sub精度的比较

Table 5 Comparison of X-Sub precision with different kernel sizes on the NTU RGB+D dataset

卷积核大小	参数量/M	精确度/%
$k=7$	1.52	92.5
$k=14$	1.55	92.6
$k=21$	1.59	**92.7**
$k=28$	1.65	92.6
$k=35$	1.73	92.6

注：加粗字体表示最优结果。

3.7 模块不同放置位置对实验结果的影响

表6中，方法1为SLKC放在3个CTR相加之后，在与GCM相加之前；方法2为LKA放在与GCM相加之后。由表6可知：方法2比方法1有着更少的参数量以及更高的精度表现。因为选择方法1的放置位置，根据流程会在GCM模块聚合空间特征之前进行

LKA操作，将LKA聚合过后的信息再次送入GCM模块中，会导致信息聚合的过度泛化，从而无法充分有效地聚合空间中的有效信息。而方法2的优势在于利用深度膨胀卷积对模型前期全局上下文建模模块(GCM)优化后的空间特征进行聚合，能够充分提取空间特征，由于将LKA放置在空间卷积的最后一层进行收尾，实现了空间信息的最大程度提取。

表6　NTU RGB+D数据集上不同位置模块放置方式下X-Sub精度的比较

Table 6　Comparison of X-Sub precision of module placement at different positions on the NTU RGB+D dataset

方法	参数量/M	精确度/%
方法1	2.29	92.8
方法2	1.91	**93.0**

注：加粗字体表示最优结果。

3.8　远距离节点建模

挑选多个全局动作进行动作识别，将基线模型和所提方法进行对比，如图7所示，在所给出的动作中，可以看出所提出模型的精确度比基线模型更高。

本文提出的全局上下文建模模块有效捕捉了远距离节点的时空关系，尤其是多节点之间的长程依赖建模；并且提出的骨架大核算子也捕捉了这种全局性的信息，增强了模型建模远距离节点的能力。因此，由图7看出，在动作识别任务中，尤其是全局动作的识别，动作可能会涉及到身体不同部位的协调，甚至是身体与环境的交互。这意味着动作中不只是局部的节点关系(如邻近关节之间的关系)重要，远距离节点之间的关联性(如手与脚之间的配合)也起到了至关重要的作用。

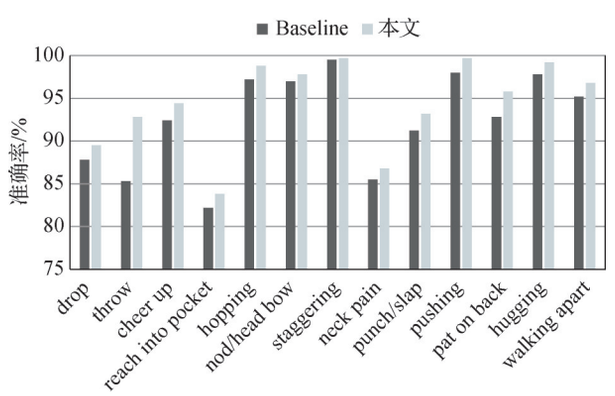

图7　不同全局动作精确度对比

Fig. 7　Comparison of accuracy of different global actions

基线模型在局部关系建模上可能表现良好，能够识别一些依赖于邻近关节的动作(如“敬礼”这种简单的手臂动作)，但在需要远距离节点配合的复杂动作中(如hugging或walking apart)，其表现就不如本文方法。本文方法通过更好的远距离建模策略，使得复杂动作中的多个关节互动关系能够得到充分利用，进而提升了动作识别的精确度。

通过以上分析，远距离节点关系建模有助于提升动作识别性能。所提方法更全面地捕捉到了远距离节点之间的互动关系，从而提高复杂动作的识别精度。这种全局建模的能力是动作识别任务中非常关键的要素，尤其是在涉及多个关节的协调动作时，远距离节点的依赖关系更加重要。

3.9　可视化分析

为了更清晰地展现提出的模块在动作识别中的作用和效果，对特定动作进行了动作可视化，并展示可视化权重图。图8展示了人体“跳跃(jump)”动作的骨架图与相应的邻接矩阵，颜色的深浅表示节点之间的关联性，颜色越深，表示关联性越强；颜色越浅，表示关联性较弱。通过分析邻接矩阵权重图，可以看出如何建模远距离关系，以下是详细分析。

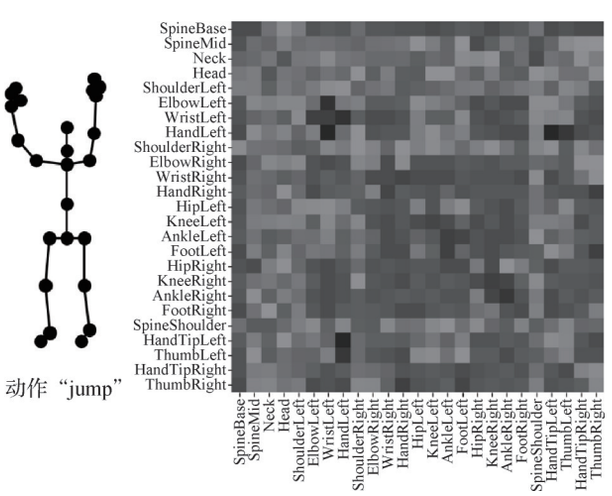

图8　邻接矩阵可视化

Fig. 8　Adjacency matrix visualization

3.9.1　节点关联性分析

1)下肢的强关联性。在这一动作中，腿部和髋部是主导的运动区域。权重图中，HipLeft、KneeLeft、AnkleLeft以及对应的右腿节点HipRight、KneeRight、AnkleRight显示出较深的颜色，表明这些关节在跳跃过程中有高度的协作关系。跳跃动作的

核心在于腿部发力和髋关节的带动，因此这些节点的强关联性反映了腿部与髋部的紧密联系。

2）躯干与下肢的协调。在邻接矩阵中，躯干节点（如SpineBase、SpineMid）与下肢（如HipLeft、HipRight）之间也表现出较强的关联性。跳跃动作需要躯干提供支撑和平衡，以保持身体的稳定。这些节点之间的强关联性表明，模型在处理跳跃动作时需要重点建模躯干和下肢的协同作用。

3）上肢的平衡作用。跳跃动作中手臂的作用主要是保持身体的平衡，因此ShoulderLeft、ElbowLeft、HandLeft以及右手对应的节点（如ShoulderRight、ElbowRight、HandRight）的关联性相对较弱，颜色较浅。这表明手臂在跳跃动作中并非主要运动部位，但仍起到平衡作用。

3.9.2　远距离关系的建模

1）腿部与上肢的远距离关联。在“jump”动作中，尽管下肢是主要的发力部位，但上肢（如HandLeft、HandRight）也通过躯干与下肢存在一定的远距离关联。权重图中，上肢与下肢的颜色虽然较浅，但仍有一定的关联性。这表明在跳跃的过程中，模型能够捕捉到手臂在维持平衡、调整动作姿态时与下肢之间的协同作用。

2）上下肢与躯干的整体协调。通过权重图的颜色变化可以看到，尽管“jump”这一动作主要依赖腿部发力，但模型仍然捕捉到了腿部、躯干和手臂之间的远距离关联。上下肢的协调对动作的完成有重要作用，尤其是在跳跃时的空中姿态调整和落地时的平衡控制。

因此，本文提出的方法通过捕获远距离节点之间的联系进行了有效建模，从而更进一步提升了动作的识别精度。

4　结　论

针对建模远距离节点关系等问题提出一种空间特征增强方法，包括一种骨架大核算子SLKC和一种全局上下文建模模块GCM，分别通过大核网络建立大接受域来模拟远程依赖以增强对空间特征的提取以及自动学习骨架拓扑结构。通过将两者结合，提出空间特征增强图卷积网络SLK-GCN。在3个标准数据集的实验，证明了所提出的SLK-GCN的有效性。当前，时间维度上的建模优化的局限性会干扰模型准确性，仍然是有待解决的难点问题。下一步将从如何更好地以不同时间尺度的建模关系着手，探索更优的表示学习方法，使得任务之间的交互更加合理和充分，提高模型的分类效果。

参考文献（References）

Chen T L, Zhou D S, Wang J, Wang S D, Guan Y, He X M and Ding E R. 2021a. Learning multi-granular spatio-temporal graph network for skeleton-based action recognition//Proceedings of the 29th ACM International Conference on Multimedia. [s. l.]: Association for Computing Machinery: 4334-4342 [DOI: 10.1145/3474085.3475574]

Chen Y X, Zhang Z Q, Yuan C F, Li B, Deng Y and Hu W M. 2021b. Channel-wise topology refinement graph convolution for skeleton-based action recognition//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 13339-13348 [DOI: 10.1109/ICCV48922.2021.01311]

Chen Z, Li S C, Yang B, Li Q H and Liu H. 2021c. Multi-scale spatial temporal graph convolutional network for skeleton-based action recognition//Proceedings of the 35th AAAI Conference on Artificial Intelligence. [s.l.]: AAAI: 1113-1122 [DOI: 10.1609/aaai.v35i2.16197]

Cheng K, Zhang Y F, Cao C Q, Shi L, Cheng J and Lu H Q. 2020a. Decoupling GCN with DropGraph module for skeleton-based action recognition//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 536-553 [DOI: 10.1007/978-3-030-58586-0_32]

Cheng K, Zhang Y F, He X Y, Chen W H, Cheng J and Lu H Q. 2020b. Skeleton-based action recognition with shift graph convolutional network//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 180-189 [DOI: 10.1109/CVPR42600.2020.00026]

Chi H G, Ha M H, Chi S, Lee S W, Huang Q X and Ramani K. 2022. InfoGCN: representation learning for human skeleton-based action recognition//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 20154-20164 [DOI: 10.1109/CVPR52688.2022.01955]

Ding X H, Zhang X Y, Han J G and Ding G G. 2022. Scaling up your kernels to 31×31: revisiting large kernel design in CNNs//Proceedings of 2022 IEEE/CVF conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 11953-11965 [DOI: 10.1109/CVPR52688.2022.01166]

Duan H D, Wang J Q, Chen K and Lin D H. 2022. DG-STGCN: dynamic spatial-temporal modeling for skeleton-based action recognition [EB/OL]. [2024-06-24]. https://arxiv.org/pdf/2210.05895.pdf

Huang Z, Shen X, Tian X M, Li H Q, Huang J Q and Hua X S. 2020.

Spatio-temporal inception graph convolutional networks for skeleton-based action recognition//Proceedings of the 28th ACM International Conference on Multimedia. Seattle, USA: ACM: 2122-2130 [DOI: 10.1145/3394171.3413666]

Jiang Q Y, Wu X J and Xu T Y. 2022. M2FA: multi-dimensional feature fusion attention mechanism for skeleton-based action recognition. Journal of Image and Graphics, 27(8): 2391-2403 (姜权晏, 吴小俊, 徐天阳. 2022. 用于骨架行为识别的多维特征嵌合注意力机制. 中国图象图形学报, 27(8): 2391-2403) [DOI: 10.11834/jig.210091]

Kipf T N and Welling M. 2017. Semi-supervised classification with graph convolutional networks//Proceedings of the 5th International Conference on Learning Representations. Toulon, France: OpenReview.net:

Li B, Li X, Zhang Z F and Wu F. 2019. Spatio-temporal graph routing for skeleton-based action recognition//Proceedings of the 33rd AAAI Conference on Artificial Intelligence. Honolulu, USA: AAAI: 8561-8568 [DOI: 10.1609/aaai.v33i01.33018561]

Liu J, Shahroudy A, Perez M, Wang G, Duan L Y and Kot A C. 2020a. NTU RGB+D 120: a large-scale benchmark for 3D human activity understanding. IEEE Transactions on Pattern Analysis and Machine Intelligence, 42(10): 2684-2701 [DOI: 10.1109/TPAMI.2019.2916873]

Liu M Y, Liu H and Chen C. 2017. Enhanced skeleton visualization for view invariant human action recognition. Pattern Recognition, 68: 346-362 [DOI: 10.1016/j.patcog.2017.02.030]

Liu Y N, Zhang H, Li Y Q, He K J and Xu D. 2023. Skeleton-based human action recognition via large-kernel attention graph convolutional network. IEEE Transactions on Visualization and Computer Graphics, 29(5): 2575-2585 [DOI: 10.1109/TVCG.2023.3247075]

Liu Z Y, Zhang H W, Chen Z H, Wang Z Y and Ouyang W L. 2020b. Disentangling and unifying graph convolutions for skeleton-based action recognition//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE: 140-149 [DOI: 10.1109/CVPR42600.2020.00022]

Lu J, Li X F, Zhao B and Zhou J. 2023. A review of skeleton-based human action recognition. Journal of Image and Graphics, 28(12): 3651-3669 (卢健, 李萱峰, 赵博, 周健. 2023. 骨骼信息的人体行为识别综述. 中国图象图形学报, 28(12): 3651-3669) [DOI: 10.11834/jig.230046]

Shahroudy A, Liu J, Ng T T and Wang G. 2016. NTU RGB+D: a large scale dataset for 3D human activity analysis//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 1010-1019 [DOI: 10.1109/CVPR.2016.115]

Shi L, Zhang Y F, Cheng J and Lu H Q. 2019. Two-stream adaptive graph convolutional networks for skeleton-based action recognition//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 12018-12027 [DOI: 10.1109/CVPR.2019.01230]

Si C Y, Chen W T, Wang W, Wang L and Tan T N. 2019. An attention enhanced graph convolutional LSTM network for skeleton-based action recognition//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE: 1227-1236 [DOI: 10.1109/CVPR.2019.00132]

Song Y F, Zhang Z, Shan C F and Wang L. 2023. Constructing stronger and faster baselines for skeleton-based action recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(2): 1474-1488 [DOI: 10.1109/TPAMI.2022.3157033]

Trivedi N and Sarvadevabhatla R K. 2022. PSUMNet: unified modality part streams are all you need for efficient pose-based action recognition//Proceedings of 2022 European Conference on Computer Vision. Tel Aviv, Israel: Springer: 211-227 [DOI: 10.1007/978-3-031-25072-9_14]

Wang J, Nie X H, Xia Y, Wu Y and Zhu S C. 2014. Cross-view action modeling, learning, and recognition//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA: IEEE: 2649-2656 [DOI: 10.1109/CVPR.2014.339]

Wang M S, Ni B B and Yang X K. 2023. Learning multi-view interactional skeleton graph for action recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(6): 6940-6954 [DOI: 10.1109/TPAMI.2020.3032738]

Wang S C, Huang Q, Zhang Y F, Li X, Nie Y Q and Luo G C. 2022. Review of action recognition based on multimodal data. Journal of Image and Graphics, 27(11): 3139-3159 (王帅琛, 黄倩, 张云飞, 李兴, 聂云清, 雒国萃. 2022. 多模态数据的行为识别综述. 中国图象图形学报, 27(11): 3139-3159) [DOI: 10.11834/jig.210786]

Wen Y H, Gao L, Fu H B, Zhang F L, Xia S H and Liu Y J. 2023. Motif-GCNs with local and non-local temporal blocks for skeleton-based action recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(2): 2009-2023 [DOI: 10.1109/TPAMI.2022.3170511]

Wu Z Z, Sun P P, Chen X, Tang K K, Xu T, Zou L, Wang X F, Tan M, Cheng F and Weise T. 2024. SelfGCN: graph convolution network with self-attention for skeleton-based action recognition. IEEE Transactions on Image Processing, 33: 4391-4403 [DOI: 10.1109/TIP.2024.3433581]

Yan S J, Xiong Y J and Lin D H. 2018. Spatial temporal graph convolutional networks for skeleton-based action recognition//Proceedings of the 32nd AAAI Conference on Artificial Intelligence. New Orleans, USA: AAAI: 7444-7452 [DOI: 10.1609/aaai.v32i1.12328]

Yang H, Yan D, Zhang L, Sun Y D, Li D and Maybank S J. 2022. Feedback graph convolutional network for skeleton-based action recognition. IEEE Transactions on Image Processing, 31: 164-175 [DOI: 10.1109/TIP.2021.3129117]

Ye F F, Pu S L, Zhong Q Y, Li C, Xie D and Tang H M. 2020. Dynamic GCN: context-enriched topology learning for skeleton-

based action recognition//Proceedings of the 28th ACM International Conference on Multimedia. Seattle, USA: ACM: 55-63 [DOI: 10.1145/3394171.3413941]

Yun X, Xu C L, Riou K, Dong K W, Sun Y J, Li S, Subrin K and Le Callet P. 2024. Behavioral recognition of skeletal data based on targeted dual fusion strategy//Proceedings of the 38th AAAI Conference on Artificial Intelligence. Vancouver, Canada: AAAI: 6917-6925 [DOI: 10.1609/aaai.v38i7.28517]

Zhang P F, Lan C L, Zeng W J, Xing J L, Xue J R and Zheng N N. 2020. Semantics-guided neural networks for efficient skeleton-based human action recognition//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 1109-1118 [DOI: 10.1109/CVPR42600.2020.00119]

Zhu Y S, Shuai H, Liu G C and Liu Q S. 2023. Multilevel spatial-temporal excited graph network for skeleton-based action recognition. IEEE Transactions on Image Processing, 32: 496-508 [DOI: 10.1109/TIP.2022.3230249]

作者简介

吴志泽，男，教授，主要研究方向为深度学习驱动的视频、图像处理与理解。E-mail: wuzz@hfuu.edu.cn

王晓峰，通信作者，男，教授，主要研究方向为人工智能与计算机视觉。E-mail：xfwang@hfuu.edu.cn

万龙，男，硕士研究生，主要研究方向为人体行为识别。E-mail：22085400608@stu.hfuu.edu.cn

洪芳华，男，博士研究生，主要研究方向为人工智能安全。E-mail：fanghuahong@stu.ahu.edu.cn

汤正道，男，高级工程师，主要研究方向为人工智能。E-mail：zhengdaotang@ustc.edu.cn

孙斐，男，教授，主要研究方向为计算机视觉。E-mail：sunfei@hfuu.edu.cn

邹乐，男，教授，主要研究方向为人工智能与计算机视觉。E-mail：zoul@hfuu.edu.cn