计算机科学
COMPUTER SCIENCE

基于多模态体育教育数据的图空间融合动作识别方法

陈海涛, 梁俊威, 陈晨, 王宇帆, 周宇

引用本文

陈海涛, 梁俊威, 陈晨, 王宇帆, 周宇. 基于多模态体育教育数据的图空间融合动作识别方法[J]. 计算机科学, 2026, 53(2): 89-98.

CHEN Haitao, LIANG Junwei, CHEN Chen, WANG Yufan, ZHOU Yu. Multimodal Physical Education Data Fusion via Graph Alignment for Action Recognition [J]. Computer Science, 2026, 53(2): 89-98.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.250800007

基于多模态体育教育数据的图空间融合动作识别方法

陈海涛[1]　梁俊威[2]　陈　晨[3]　王宇帆[4]　周　宇[1]
1 深圳大学计算机与软件学院　广东 深圳 518060
2 深圳信息职业技术学院　广东 深圳 518172
3 西北大学文学院　西安 710127
4 上海交通大学机械工程学院　上海 200240
(2410105035@mails.szu.edu.cn)

摘　要　在智能体育与教育信息化的背景下,精细化的人体动作识别已成为体育教学与训练评估中的关键技术。针对传统动作识别方法在复杂运动场景中存在的模态信息利用不足、时空结构表达受限等问题,提出了一种融合骨架数据与可穿戴传感器信息的多模态图卷积网络模型。首先,提出了一种基于"虚拟传感器"的融合方法,将可穿戴传感器信号映射至骨架关节构建的时空图结构中并融合,有效提升了对动作细节的建模能力与跨模态语义一致性。其次,构建了针对复杂运动模式的多层图卷积网络,通过对身体进行局部划分,增强了模型在复杂体育场景下的识别能力。此外,面向击剑这一技术动作复杂的竞技项目,自主采集并构建了一套涵盖不同典型技术动作与运动水平层次的多模态数据集,为精细化动作识别与水平评估提供了数据支持。在该数据集与多个标准数据集上进行的实验表明,所提方法在动作识别精度与技术水平判断上优于现有主流方法,为体育教育场景中的智能识别与评估提供了新的建模框架与技术支持,具有良好的应用前景。

关键词:动作识别;多模态数据融合;图卷积网络;体育教育;击剑数据集
中图分类号　TP391

Multimodal Physical Education Data Fusion via Graph Alignment for Action Recognition

CHEN Haitao[1],LIANG Junwei[2],CHEN Chen[3],WANG Yufan[4] and ZHOU Yu[1]
1 College of Computer Science and Software Engineering,Shenzhen University,Shenzhen,Guangdong 518060,China
2 Shenzhen Institute of Information Technology,Shenzhen,Guangdong 518172,China
3 Faculty of Liberal Arts,Northwest University,Xi'an 710127,China
4 School of Mechanical Engineering,Shanghai Jiao Tong University,Shanghai 200240,China

Abstract　In the context of intelligent sports and educational informatization,fine-grained human action recognition has become a key technology in physical education and training assessment. To address the limitations of traditional methods in utilizing multi-modal information and capturing spatio-temporal structures in complex motion scenarios,this paper proposes a multi-modal graph convolutional network model that fuses skeleton data and wearable sensor information. Firstly,it proposes a fusion method based on "virtual sensors," which maps wearable sensor signals onto a spatio-temporal graph constructed from skeletal joints,enabling effective integration of multimodal information and enhancing fine-grained motion modeling and cross-modal semantic consistency. Secondly,it designs a multi-layer graph convolutional network tailored for complex sports movements,incorporating local body part segmentation to improve recognition performance in challenging scenarios. Thirdly,it constructs a high-quality multimodal dataset for fencing,covering various technical actions and skill levels,to support fine-grained action recognition and skill assessment. Experimental results on both this dataset and several public benchmarks demonstrate that the proposed method outperforms existing approaches in both action recognition accuracy and skill level classification. This work provides a novel mode-

到稿日期:2025-08-04　返修日期:2025-11-03
基金项目:国家自然科学基金面上项目(72271168);广东省自然科学基金面上项目(2024A1515012485);广东省重点领域研发计划(2024B0101120003);深圳市科技重大专项(KJZD20230923114111021);深圳市基础研究面上项目(JCYJ20220810112354002);广东省基础与应用基础研究区域联合基金-青年基金项目(2023A1515110070)
This work was supported by the Surface Project of the National Natural Science Foundation of China(72271168),Surface Project of the Natural Science Foundation of Guangdong Province,China(2024A1515012485),Key Field Research and Development Program of Guangdong Province,China(2024B0101120003),Major Science and Technology Special Project of Shenzhen,China(KJZD20230923114111021),Surface Project of the Basic Research of Shenzhen,China(JCYJ20220810112354002) and Joint Funds of the Basic and Applied Basic Research Area of Guangdong Province,China-the Program of the Young Scientists Fund(2023A1515110070).
通信作者:周宇(zhouyu_1022@126.com)

ling framework and technical support for intelligent recognition and evaluation in sports education.

Keywords Action recognition, Multimodal data fusion, Graph convolutional network, Physical education, Fencing dataset

1 引言

随着计算机视觉与人机交互技术的不断进步，人体动作识别(Human Action Recognition，HAR)作为智能感知与行为理解的重要分支，被广泛应用于教育、医疗、安防等多个领域。尤其在体育教育中，动作识别技术为传统教学与训练模式注入了新的智能化元素。通过对学生或运动员动作过程的自动化分析，系统不仅能够客观评估其技术动作的完成质量，还可辅助教师及时发现错误动作并进行针对性纠正，从而显著提升教学效率与训练效果。同时，动作数据的实时采集与处理，也为远程教学、个性化训练方案的制定提供了坚实的数据基础。

尽管动作识别在体育教育中具有显著的应用价值，但其在真实复杂环境中的部署仍面临一系列挑战。一方面，不同个体在动作幅度、运动节奏以及身体结构上的差异，要求识别模型具备较强的泛化与个性化适应能力；另一方面，教学与训练环境中普遍存在遮挡干扰、噪声污染与多视角变换等问题，导致传统基于单一视觉模态的方法在鲁棒性和准确性方面表现受限。近年来，随着深度传感器、惯性测量单元等精密传感技术的发展，越来越多区别于传统视觉模态的数据被挖掘，许多研究开始采用不同类型的数据模式进行输入，如骨架[1]、深度图[2]、点云[3]和红外序列[4]等。其中，Zhu 等[5]提出的 FencerNet架构将 2D 姿态数据作为输入，并使用基于骨架的动作识别方法对动作进行分类，用于自动化击剑中细粒度步法技术的分类。但是，仅依赖骨架模态进行动作识别，在捕捉局部动作细节和复杂动作形态方面仍存在受视角变换和局部数据缺失影响的局限性。Tao 等[6]构建了多 IMU 传感器注意力融合架构，通过对不同身体部位传感器的特征进行加权学习，实现传感器之间的动态互补与信息选择。但此类方法多限于惯性传感器内的融合。随着研究的深入，多模态数据的融合逐渐成为提升动作识别性能的主流方向。Ahmad 等[7]提出将深度图转换为图像、IMU 信号转换为时间信号图后分别提取特征并融合，实验结果表明，该模型的性能优于单模态方法。然而，多模态融合系统往往对设备同步性和通信稳定性要求较高，并且特征融合策略往往会带来较高的计算复杂度。从实验结果来看，现有的多模态数据融合方法在性能上往往优于单模态的动作识别方法，但是也存在以下问题：一方面，多模态数据间缺乏结构化、可泛化的深度融合机制，未能充分挖掘模态间的互补性；另一方面，对复杂人体动作的空间结构表达和高阶语义刻画仍显不足，尤其在不同体育场景下存在动作类内差异大、类间区分度低等问题。

为解决上述问题，本文提出了一种基于骨架与可穿戴传感器融合的多模态动作识别框架，结合视觉感知的骨架数据与惯性信息，以提升动作识别的精度与稳定性。具体而言，设计了一种将传感器信号映射至骨架图空间的融合策略。不同于传统的拼接式融合方法，本文在每个骨架关节点上构造“虚拟传感器”，并借助 LSTM 网络建立有向无环图传播机制，从而使得惯性信息能以图结构的方式与骨架数据进行融合并参与动作建模，以实现更深层次的跨模态特征交互。在此基础上，进一步提出基于身体功能区域划分的图卷积网络结构，将人体骨架划分为头部、双手与双腿 3 个功能子区域，以突出关键部位的动作差异性，并使得不同的局部区域能够相互进行有效动作特征的分享。通过引入注意力机制对不同区域特征进行加权融合，使模型可根据动作特性动态调整对关键部位的关注程度。与此同时，本文利用 Vietoris-Rips 复合体[8]的拓扑建模能力，从融合图中提取高阶空间结构特征，有效增强对复杂运动模式的表征能力。本文的贡献如下：

1)提出了一种融合骨架数据与可穿戴传感器数据的多模态动作识别框架，通过构建虚拟传感器与 LSTM 预测机制，使得惯性数据与骨架数据能够在图空间中进行融合；

2)引入身体局部区域划分策略，增强模型对关键动作区域的表达能力，结合 Vietoris-Rips 拓扑特征进行结构建模，提升对骨架与传感器融合图中高阶关系的建模能力；

3)自采集了一个击剑动作识别的多模态数据集，其中包含骨架数据与可穿戴传感器的同步数据，为教育智能化及技术动作分析提供了数据支持与实验基础；

4)在自采集击剑数据集与多个标准数据集上进行实验验证，结果表明，本文方法在动作识别任务中具备显著优势。

本文第 2 章回顾了与本研究相关的动作识别领域的研究进展；第 3 章详细介绍了融合骨架与惯性传感器的动作识别框架；第 4 章展示了实验设置、评估指标以及在多个数据集上的性能对比结果；最后总结全文并对未来研究方向进行展望。

2 相关工作

2.1 图卷积网络

图卷积网络(Graph Convolutional Network，GCN)因在非欧几里德结构数据上具备建模优势，被广泛应用于基于骨架的人体动作识别任务中[9]。骨架序列具有天然的拓扑结构，可自然构建为时空图，节点表示关节，边表示连接关系，从而使得 GCN 能够有效捕捉骨骼之间的空间依赖性与动作的动态特征。Yan 等[10]提出了时空图卷积网络(Spatial-Temporal GCN)，利用空间图卷积提取骨架帧内的结构信息，结合时间卷积建模帧间的时序动态。Li 等[11]提出了动作-结构图卷积网络(Actional-Structural GCN)，一方面从动作中捕获活动链接，以获得动作间潜在的依赖关系，另一方面扩展原有的骨架图以表示高阶依赖关系，通过两者捕获更多的全局关节信息以及隐式的联系。Shi 等[12]提出了 2s-AGCN———一种双流自适应图卷积网络，将骨骼的一阶信息(关节坐标)与二阶信息(骨骼的长度和方向)进行了有效结合，提升了对复杂动作的表达能力。这些方法有效提升了骨架数据的建模精度和动作分类性能，验证了图结构建模在动作识别中的适用性。但传统基于 GCN 的动作识别方法多数依赖于静态、预定义的骨架拓扑结构，难以自适应不同动作类别下骨架关系的动态变化，缺乏对局部区域间独立运动模式的建模能力。

近些年来，Hu 等[13]提出的 STGAT 通过在图注意力模块中引入时间维度的流动建模来强化动作特征的表达能力。Chen 等[14]提出的 MST-GCN 通过多尺度空间和时间图卷积扩展模型的感受域，从而提升模型对远程依赖关系的捕获能力。这些方法在时空依赖建模和图结构自适应学习方面均展现出良好效果，但普遍仍未充分关注跨模态信息融合与细粒度身体区域建模的问题。本文在设计上继承了众多图卷积网络在时空特征提取上的优势，并借鉴了 AGCN 中可学习邻接矩阵的思想，同时进一步引入基于局部身体区域划分的建模策略，并结合传感器模态信息，以实现跨模态的协同表征。这样不仅延续了现有方法在时空依赖与图结构学习上的成果，而且在细粒度区域建模和多模态融合方面进行了拓展，能更好地应对体育场景下复杂动作识别与运动水平评估的需求。

2.2 多模态数据融合动作识别

多模态融合作为提升动作识别系统性能的重要策略，旨在通过整合不同模态数据的互补优势，实现更准确的特征表达与分类判断。目前的研究表明，使用多模态的数据可以获得更高的精度[15]。目前常见的多模态融合策略主要包括分数融合、特征融合与数据融合 3 类。分数融合通过独立训练的模型在决策层面整合预测结果，虽然具有实现简单、灵活性强的优点，但往往忽略了不同模态间的低层语义协同，融合深度有限[16]。特征融合将各模态提取的中间特征进行拼接或映射，从而增强表示能力；但其融合质量高度依赖于特征对齐方式与网络设计，容易受到不同模态间时序或尺度差异的干扰[17]。相比之下，数据融合在输入阶段就对多模态数据进行融合，使模型能够在特征学习初始阶段便不断挖掘不同模态间的底层互补信息，从而在一定程度上缓解对齐问题，并保留更多原始结构信息[18]。近年来，多模态数据融合在人体动作识别中不断推进，其核心在于输入早期或中间阶段整合互补模态，以缓解单模态对场景变化、遮挡与数据缺失等问题的脆弱性。Choi 等[19]提出的多模态预训练范式利用多模态信号进行对比式学习，显著提升了小样本和跨域设置下的识别性能。Hu 等[20]提出的人体多模态融合网络以骨架为先导，对 RGB、深度与光流等视觉模态进行重构与对齐，从而在复杂视点与动态背景下获得更稳健的融合表征。在训练与监督策略上，Yuan 等[21]采用多尺度知识蒸馏将强模态向弱模态迁移，有效缓解了多模态不同步或缺失时的性能退化。总体来看，这些工作从表示空间对齐、注意力交互、蒸馏迁移等维度进行了多模态融合，但在骨架与其他模态数据融合的早期统一建模能力上仍相对不足，许多方法侧重中期特征拼接或注意力交互，需要额外的时序与尺度对齐模块，且没有利用传感器捕捉到的局部动力学以及与骨架全局拓扑关系的协同。与之相比，本文采用数据级早期融合，利用了骨架数据具备的良好空间结构建模能力，能够清晰刻画全身关节间拓扑关系的特性；在特征提取之前将惯性传感器数据映射在图空间中与骨架数据进行融合，从学习初始阶段就显式地捕获跨模态的底层互补性，将传感器主导的局部动态与骨架主导的全局拓扑进行协同对齐与统一表征，以获得更稳健的细粒度动作辨析，提升系统在动作细节建模与复杂场景适应中的表现。

2.3 体育教育与智能设备

随着智能设备与可穿戴技术的迅速发展，体育教育这一学科正逐步通过引入各类具备数据采集与分析能力的设备，来提升教学效率，并实现个性化指导与科学化评估[22]。Han 等[23]提出了一种轻量级的视频识别方法，该方法通过实施时空修剪方案来减少参数数量，在基于边缘计算的在线体育教学系统中取得了良好的表现。Ding 等[24]设计了一种基于动作识别的体育教学效果评价系统，并验证了该系统的动作识别准确率和体育教学效果，符合体育教学的实际需求。

当前常用的智能设备包括搭载惯性测量单元（Inertial Measurement Unit，IMU）的可穿戴装置、深度摄像头、红外感应系统等。IMU 设备因部署灵活、数据精确、不依赖外部视觉条件，在运动轨迹记录和姿态分析等任务中具有显著优势。通过智能设备采集到的多维度行为数据，不仅丰富了教学中可观测的信息类型，也推动了体育课堂从基于经验的评估方式向量化和可视化的实时分析转变。例如，Fu 等[25]将智能设备应用于篮球专项教学中，为教师提供了客观、可追溯的数据支撑，显著提升了训练评估的科学性。此外，无论是体育课堂中的基础体态训练，还是竞技项目中的复杂技战术动作，传统依赖经验判断与肉眼观测的方式在精度与效率上已难以满足现代体育教育的智能化发展需求。Sri-Iesaranusorn 等[26]基于低成本传感设备，设计了一种判别初学者与专业选手差异的通用分析方法，用于辅助技能提升与动作矫正。Yuan 等[27]将检测到的选手构建成时空图，对选手间的动态交互与潜在动作关系进行建模，所提方法在排球数据集中取得了优异的分类性能。尽管已有研究初步验证了智能设备在体育教学中的应用潜力，但在竞技项目中，由于缺乏高质量、多模态标注数据集，难以全面支撑个性化训练与技术动作分类等任务。为此，本文基于可穿戴传感器，构建了一套多模态动作识别数据集，其中涵盖骨架与传感器数据，涉及多类动作与不同水平层次，为后续精细识别与多目标分类研究提供了数据支持。

3 多模态数据融合图卷积网络

3.1 总体设计

本文模型的系统架构如图 1 所示。

多模态数据融合图卷积网络的完整工作流程大致分为两个部分：1）通过 Sensor to Graph（STG）模块将传感器数据映射至图空间中与骨架数据进行融合；2）对融合好的数据使用图卷积网络进行特征提取，最终输出分类结果。对于一般的公共数据集，骨架数据 S 的结构一般为 (C_s, T_s, N_s)，其中 C_s 代表关节点的通道维度，T_s 代表骨架数据帧的数量，N_s 代表关节点的数目。传感器数据 I 的结构一般为 (C_i, T_i)，其中 C_i 代表传感器数据的通道维度，T_i 代表传感器数据帧的数量。本文算法可以分为以下几步。

1）通常来说，骨架数据与传感器数据两者间的帧数与通道维度往往不一致，所以首先需要进行对齐处理。通过线性插值重采样的方法，将 t 时刻的传感器数据 I_t 的原先帧数 T_i 调整为目标帧数 T_s，以实现与骨架数据的对齐。对于通道维度，通过一个全连接层 FC 降低 C_i 的维度，使得 $C_i = C_s$，最终得到对齐后的传感器数据 $\hat{I}_t$。

$$\hat{I}_t = FC(I_t) \tag{1}$$

2）对于每一个时间戳 t 的传感器数据 $\hat{I}_t$，通过 STG 模块

的函数，由原先的数据结构(C_i, T_i)塑造成与骨架数据 S 相似的结构(C_s, T_s, N_s)。

$$\widetilde{I}_t = STG_\theta(\hat{I}_t, S_t) \tag{2}$$

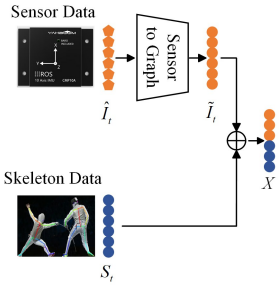
其中，θ 是一个可学习的参数，共享于不同的时间戳。STG 的作用是将传感器数据与骨架数据在图空间中进行对齐，为之后的数据融合做准备。具体来说，STG 假设每个骨架节点上都有一个对应的虚拟传感器，随后使用一个基于有向无环图的 LSTM(DAG-LSTM)模块，通过每一帧的传感器数据$\hat{I}_t$与骨架数据S_t计算出所有虚拟传感器的数据$\widetilde{I}_t$。

3)将得到的虚拟传感器数据$\widetilde{I}_t$与骨架数据S_t在图空间中对应地进行数据层面的融合，得到多模态数据 X 的结构为(C_s+C_i, T_s, N_s)。

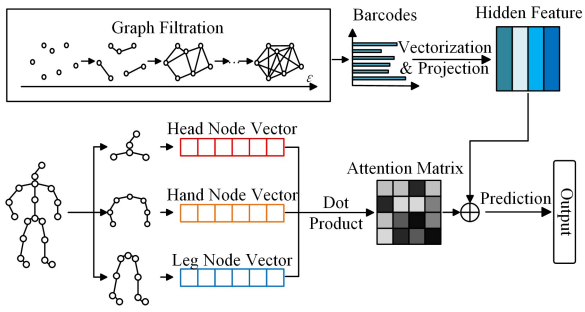
$$X = Concat(\widetilde{I}_t, S_t) \in \mathbb{R}^{(C_s+C_i)\times T_s \times N_s} \tag{3}$$

4)通过图卷积网络提取 X 中的多模态时空特征，对给定样本进行动作分类。

$$\hat{y} = f_{GCN}(X) \tag{4}$$



(a)数据融合



(b)图卷积网络

图 1 本文模型系统架构

Fig. 1 Overall structure of the proposed model

3.2 STG 模块

算法 1 展示了 STG 模块进行数据融合的过程。假定所有关节点中仅有一个关节点存在真实传感器，STG 模块会认为剩余的N_s-1个节点都存在一个虚拟传感器。根据这一个真实传感器的数据与所有的骨架数据，计算出所有虚拟传感器的数据，使得传感器数据可以拥有图结构的特性，从而与骨架数据进行融合，获得更多的特征信息。

算法 1 基于有向无环图的长短时记忆网络

输入：$\hat{I}_t$/* 真实传感器数据 */

S_t/* 骨架数据 */

n_s/* 真实传感器所在位置 */

$\mathcal{G}$/* 有向无环图 */

输出：$\widetilde{I}_t$/* 映射在图空间的传感器数据 */

1. Let $C_i = C_s$
2. Let $\widetilde{I}_t = 0$
3. Let $\widetilde{I}_t^{n_s} = \hat{I}_t^{n_s}$
4. $\mathcal{N} \leftarrow$ Set of all node with no incoming edges
5. insert n_s into $\mathcal{N}$
6. While $\mathcal{N}$ is not empty do
7. remove a node n_s from $\mathcal{N}$
8. for each node with an edge e from n_s to n_v do
9. $n_s n_v = S_t^{n_v} - S_t^{n_s}$
10. $\widetilde{I}_t^{n_v} = f(\widetilde{I}_t^{n_s}, n_s n_v)$
11. remove edge e from the $\mathcal{G}$
12. if n_v has no other incoming edges then
13. insert n_v into $\mathcal{N}$
14. end if
15. end for
16. end while
17. return $\widetilde{I}_t$

当真实传感器位于关节点n_s时，对于n_s的任一邻居节点n_v，可以通过 t 时刻两个关节点的骨架数据$S_t^{n_v}$与$S_t^{n_s}$得到关节点n_s与n_v之间的向量$\boldsymbol{n_s n_v}$。

$$\boldsymbol{n_s n_v} = S_t^{n_v} - S_t^{n_s} \tag{5}$$

通过位于关节点n_s的真实传感器数据$\widetilde{I}_t^{n_s}$与$\boldsymbol{n_s n_v}$，就可以计算出关节点n_v的传感器数据$\widetilde{I}_t^{n_v}$。

$$\widetilde{I}_t^{n_v} = f(\widetilde{I}_t^{n_s}, \boldsymbol{n_s n_v}) \tag{6}$$

其中，f 是一个预测函数，使用了 LSTM 的单元。输入和隐藏状态分别是不同关节点的相对位置和对应虚拟传感器的数据。仿照计算关节点n_v的虚拟传感器数据的过程，可以计算出关节点n_s的所有邻接关节点的虚拟传感器数据。

当关节点n_v的虚拟传感器数据计算完成后，STG 模块将会继续计算n_v的邻居关节点中仍未完成虚拟传感器数据计算的关节点。随着这个过程的不断进行，最终所有关节点的虚拟传感器数据将被计算完成。其传播过程如图 2 所示。以真实传感器在大腿左侧为例，基于骨架图中从真实传感器所在关节点开始的拓扑顺序进行计算。

$$\widetilde{I}_t^{(v)} = f(\widetilde{I}_t^{(u)}, \boldsymbol{n_s n_v}), \forall V \in \mathcal{N}(U) \tag{7}$$

其中，$\widetilde{I}_t^{(v)}$与$\widetilde{I}_t^{(u)}$代表 t 时刻关节点 V 与关节点 U 的虚拟传感器数据；$\mathcal{N}(U)$代表关节点 U 的邻居节点集合，即骨架图中与关节点 U 直接连接的关节点。

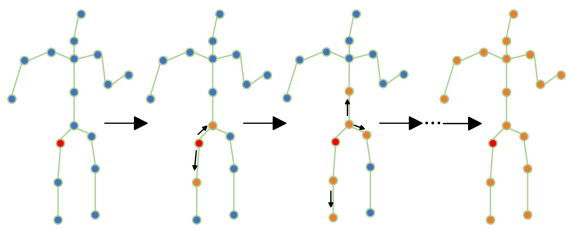


图 2 传感器数据计算顺序

Fig. 2 Calculation sequence of the sensor data

在实际情景中，被试者往往不会仅仅配备一个传感器，可

能会在多个关节点同时配备传感器进行数据的收集。在这种情况下，假设共有 K 个关节配备了传感器，相应地会构建出 K 个 STG 模块，每个 STG 模块根据对应的真实传感器的数据与骨架数据进行计算，得到所有其余虚拟传感器的数据。

如果不同传感器所测量的是相同类型的数据，则直接将位于不同位置的 K 个传感器数据 $\widetilde{I}_t^{(1)}$，$\widetilde{I}_t^{(2)}$，…，$\widetilde{I}_t^{(K)}$ 相加，得到最终的传感器数据 $\widetilde{I}_t$。

$$\widetilde{I}_t=\widetilde{I}_t^{(1)}+\widetilde{I}_t^{(2)}+\cdots+\widetilde{I}_t^{(K)} \tag{8}$$

如果数据类型不同（如其中一个传感器测量的是加速度，另外几个传感器测量的是角速度），则不能直接相加，需要在通道维度上进行拼接。

$$\widetilde{I}_t=Concat(\widetilde{I}_t^{(1)},\widetilde{I}_t^{(2)},\cdots,\widetilde{I}_t^{(K)}) \tag{9}$$

3.3 局部注意力-拓扑图卷积网络

在基于骨架数据的动作识别任务中，图卷积网络可以捕捉图中节点的邻域信息，有效学习节点关系，因此被广泛应用。其中，AGCN[12]通过引入可学习的邻接矩阵，使得模型能够自动学习最优的空间连接关系，提升了模型对不同动作之间骨架动态结构差异的适应能力。

$$f_{\text{out}}=\sum_k^{K_v}W_k(f_{\text{in}}\boldsymbol{A}_k)\odot\boldsymbol{M}_k \tag{10}$$

其中，W_k 为卷积核参数，f_{in} 与 f_{out} 为输入特征与输出特征，$\boldsymbol{A}_k$ 为第 k 个预定义邻接矩阵，$\boldsymbol{M}_k$ 为可学习的自适应邻接矩阵。

然而，AGCN 的全局自适应建图机制往往忽略了不同身体区域在动作执行中的差异性与独立性，可能导致非关键部位对识别结果产生不必要干扰，尤其是在动作表现依赖某些局部肢体的场景中更为明显。

为解决上述问题，本文提出了一种全新的网络架构，如图 1(b)所示。算法 2 展示了网络的运行流程。通过对身体进行局部划分，在 AGCN 基础上引入了局部注意力机制。局部动作识别可以通过检测同一时间戳内多个身体部位间的相关信息而受益。例如，在检测走路这一整体动作时，对抬腿这一动作的识别可以通过识别手臂是否同时进行了摆动得到更精准的判断。如图 3 所示，本文将人体骨架节点依据解剖结构划分为 3 个区域，即上肢、下肢与头部，并在每个区域内部单独构建子图，分别提取图卷积特征。将预定义的整体邻接矩阵 $\boldsymbol{A}$ 对应划分为 3 个子矩阵 $\boldsymbol{A}^{(\text{hand})}$，$\boldsymbol{A}^{(\text{leg})}$ 和 $\boldsymbol{A}^{(\text{head})}$。

$$\boldsymbol{A}=\boldsymbol{A}^{(\text{hand})}\oplus\boldsymbol{A}^{(\text{leg})}\oplus\boldsymbol{A}^{(\text{head})} \tag{11}$$

算法 2 局部注意力-拓扑图卷积网络

输入：X /* 多模态融合数据 */

L /* 网络层数 */

A /* 局部区域邻接矩阵 */

输出：$\hat{y}$ /* 分类结果 */

1. Reshape(X) to shape [N * M,C,T,V]
2. BatchNorm(X)
3. For i=1 to L do
4. If use_attention[i]==True then
5. Z←unit_gcn(X,**A**,attention=True)
6. Else
7. Z←unit_gcn(X,**A**,attention=False)
8. Z←unit_tcn(Z)
9. topo_feature= topotrans_i(X)
10. Z←topo_feature
11. X=ReLU(Z+residual(X))
12. End For
13. Let X← reshape X to [N,M,C_out,T * V]
14. X= Mean(X,dim=3)
15. X= Mean(X,dim=1)
16. $\hat{y}$=FC(X)
17. Return $\hat{y}$

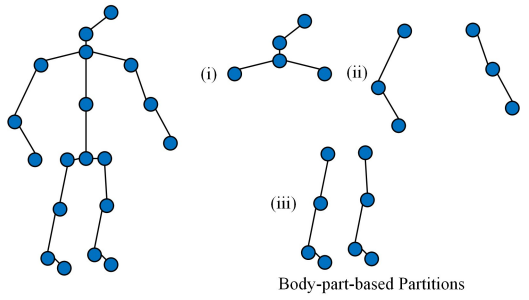


图 3 身体结构划分

Fig. 3 Division of body structure

随后，通过注意力机制对不同区域的特征进行加权融合，使模型能够根据具体动作场景动态调整对各区域的关注程度。F 为每个节点的特征，通过不同节点间的相似度进行打分，得到最终的邻接矩阵 $\boldsymbol{A}^*$。

$$\boldsymbol{A}^*=\boldsymbol{A}\odot\text{Softmax}(\boldsymbol{F}\boldsymbol{F}^{\text{T}}) \tag{12}$$

该方法可以提升模型对关键身体部位动作模式的捕捉能力，并显著增强对复杂动作中局部运动的辨识效果。

此外，为进一步提高模型对骨架结构信息的建模能力，在每一层图卷积模块中并行引入拓扑特征提取通道。该模块从图结构提取基于 Vietoris-Rips 复形的持久性拓扑特征，以描述骨架在不同时间片中的局部连接变化与全局拓扑演化。$\mathcal{VR}_\epsilon(V)$ 代表点集 V 在距离阈值 ϵ 下构造的 Vietoris-Rips 复形，σ 代表点集中任意两点之间的距离。

$$\mathcal{VR}_\epsilon(V)=\{\sigma\subseteq V\mid diam(\sigma)\leqslant\epsilon\} \tag{13}$$

最终，通过特征融合机制将图卷积输出与拓扑特征集成，形成更具结构敏感性的特征表达，从而提升动作识别的精度与泛化能力。

4 实验设计与结果分析

4.1 MFAD 数据集

在传统体育教学模式中，教师往往依赖肉眼观察与经验判断来评估学生动作的完成质量与技术水平，这种方式不仅效率低，准确性有限，且难以实现大规模的客观评测与远程教学场景下的动作监督。尤其是在对动作标准化要求极高的竞技项目，如击剑这类技术主导型的对抗性运动中，如何利用智能感知手段对运动员的动作表现进行定量分析与水平判定，已成为体育教育领域中亟待解决的问题之一。

针对上述需求，本文自采集并构建了一个面向体育教学任务的多模态击剑数据集 MFAD(Multimodal Fencing Action Dataset)，旨在服务于击剑教学中的动作识别与水平分级

的评估任务。数据集采集自 27 名具有不同训练背景的击剑参与者，涵盖业余、次精英与精英 3 个水平层次，显著提升了数据在能力分布上的代表性与广泛性。动作样本总数为 928 条，覆盖了弓步、向前一步弓步与箭步 3 种典型击剑技术动作。具体动作分解如图 4 所示，由上到下分别是弓步、向前一步弓步与箭步。

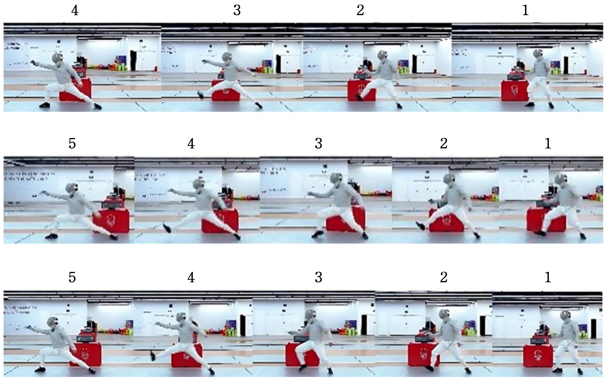


图 4 击剑动作分解

Fig. 4 Decomposition of fencing action

弓步是击剑中最基本、最常用的攻击动作，主要动作包括首先击剑运动员的前腿向前大迈一步，并随之移动手臂和身体。向前一步弓步则比弓步多往前走出一步。向前一步是为了产生更多的动量和惯性，以实现更快、更远的弓步，后续的弓步动作基本保持不变。箭步也称为“跑动攻击”，首先将所有体重转移到前腿，以流线型运动推动击剑运动员向前，随后击剑运动员踮起脚尖，从而实现更快的移动，而不是单纯地向前一步跑动。

每条样本均同步记录骨架与传感器两种模态的数据。骨架数据由微软高速运动深度相机采集，采样帧率为 100 Hz，其能完整捕捉全身运动轨迹；传感器数据则来自于四肢的 IMU 装置，采样率为 50 Hz，加速度计测量范围为±8 g，陀螺仪测量范围为±1000 度/秒。其中提供了高精度的加速度与角速度信息，能有效弥补骨架模态在局部运动细节上的不足。另外，由于传感器的佩戴位置对动作识别的精度与稳定性具有重要影响，为保证数据的可重复性与动作的自然性，本数据集在传感器布局时充分考虑了人体工程学与击剑运动的技术特点：一方面，将 IMU 固定于四肢的远端（如小腿、手腕等位置），使其既能捕捉到肢体的主要动力学信息，又避免对运动员持剑与步伐发力造成干扰；另一方面，传感器采用轻量化与贴合式佩戴方式，确保动作在高速转换过程中不影响动作质量与运动舒适性。这样既保证了局部运动信号的敏感性，又降低了对实际击剑表现的干扰。骨架节点编号与传感器具体分布如图 5 所示，该数据集目前已公开。

MFAD 数据集不仅具有技术层面的多模态结构，还充分考虑了体育教学的实际需求，设计了 3 个子任务，即动作类别识别、技术水平评估与联合识别。该数据集的构建过程严格控制采集标准与动作规范，确保了数据的可复用性与教学适配性。在击剑教育场景中，该数据集可以支持学生动作规范性检测、技术水平分层反馈与训练轨迹量化分析，有助于建立以数据驱动为核心的教学评估机制。

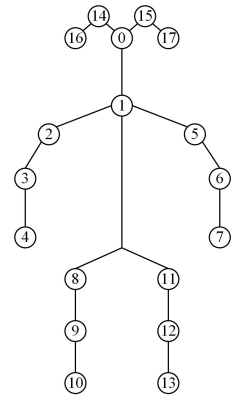


图 5 MFAD 数据集的骨架结构

Fig. 5 Skeleton structure of MFAD dataset

4.2 实验环境与设置

4.2.1 参数设置

所有实验均按照以下配置环境与相应设置进行：RTX A4000 GPU（16 GB），PyTorch 1.11.0，Ubuntu 20.04，CUDA 11.3。对于所有实验，学习率设置为 0.001，迭代次数设置为 77，均使用 Adam 优化器进行优化，并使用余弦拟退火学习率调度器，其中 T0 参数设置为 20，表示学习率调度器完成完整周期并恢复到初始值的 epoch 位置。

4.2.2 数据划分

根据不同数据集的特点，采用相应的数据划分策略，以确保模型评估的公平性与科学性。对于自采集的 MFAD 数据集，采用十折交叉验证方法进行评估。在 MFAD 数据集中，不同动作类别的分布较为均衡，但不同技能水平的样本分布存在差异。为确保取样的代表性，根据技能水平对样本进行分层，并在每个动作类别内进行随机化抽取。具体而言，将数据集划分为十折，每个折中保持技能水平与动作类别比例的一致性。最终，记录 10 个折的平均分类准确率并将其作为主要评估指标，以确保评估结果的稳定性与可靠性。在 MFAD 数据集上，为进一步验证模型性能差异的统计显著性，记录了每一折的详细结果，并采用配对 t 检验将所提模型与所有基线模型进行两两比较。显著性水平（α）设定为 0.05。对于其他公开数据集，严格遵循其原始论文中的划分要求进行模型验证，并以 F1 分数与分类准确率作为评估指标。在 MMACT 数据集的划分中，采用 Cross-Subject 方式，将 80% 的受试者（受试人员 ID 从 1 到 16）样本用于训练模型，其余 20%的受试者样本用于测试。该划分方式能更好地模拟实际应用场景，评估模型在不同受试者之间的适应能力。在 UTD-MHAD 数据集的划分中，将编号为 1，3，5，7 的被试者样本作为训练集，其余被试者样本作为测试集。该划分方式能有效评估模型在未见被试者数据上的泛化能力。对于 CZU-MHAD 数据集，采用 T5 划分策略，将被试者 1 号和 2 号的样本作为训练样本，其余被试者样本作为测试样本。该划分方式与原始研究一致，便于与其他方法进行公平对比。

4.3 MFAD 数据集实验

MFAD 数据集设计了动作类别识别、技术水平评估与联合识别 3 个子任务。为验证本文模型在 MFAD 数据集上的性能，将其与下面几种方法进行了比较。

首先是仅需要骨架数据的双流自适应图卷积网络(AGCN)[12]与基于骨架数据的时空图注意网络(STGAT)[28]。选择这两个模型在MFAD数据集上进行实验，是因为本文采用的GCN的一部分有参考AGCN中的自适应邻接矩阵的想法，并且对时空特征的提取与STGAT有相似之处。

此外，选择了基于骨架和传感器多模态动作识别领域的代表性模型Fusion-GCN[29]和MMT[30]进行比较。这两篇论文专注于骨架和传感器数据的融合，以实现高效的多模态动作识别，与本研究方向一致；并且开源了代码，为对比实验提供了良好的参照。

每种方法均为在同一实验环境下得到，并且均使用准确率作为评估指标。实验结果如表1所列，结果表明，所提方法在动作识别任务与联合识别任务上取得了最佳效果；且对比AGCN与STGAT均有明显的优势，证明了引入传感器数据对于模型精度提升具有有效性。

表1 MFAD数据集上的准确率比较

Table 1 Accuracy comparison of MFAD dataset

(%)

方法	数据类型	动作识别	水平评估	联合识别
AGCN	骨架	99.02	97.84	96.87
Fusion-GCN	骨架、传感器	98.48	95.46	90.51
STGAT	骨架	98.91	95.57	94.49
MMT	骨架、传感器	**99.45**	97.84	95.79
Ours	骨架、传感器	99.35	**98.81**	**97.61**

4.4 公共数据集实验

为验证本文方法的泛用性，本节在MMACT[31]，CZU-HMAD[32]和UTD-MHAD[33]3个公共数据集上进行了实验。

在3个公共数据集上，均进行了本文模型与AGCN模型的性能对比，以验证引入传感器数据进行融合的模型与传统的AGCN模型相比，在公共数据集上同样具有良好的性能。对于公共数据集的对比实验，则选取了在这些多模态数据集上报告了性能的且最先进的方法作为基准。一方面，选取含有骨架数据或传感器数据的多模态数据融合方法，探讨不同模态数据带来的优势互补的效果；另一方面，则选取与本文模型相类似的同时采用骨架数据与传感器数据融合的模型进行对比。为了确保一致性，特定数据集的评估指标(F1分数或准确度)取决于相应比较方法使用的指标。

1)MMACT数据集：选用基于多阶段的多模态数据特征融合(MSBFF)[34]、双流自适应图卷积网络(AGCN)[12]、融合图卷积网络(Fusion-GCN)[29]、多模态时间片段注意力网络(MMTSA)[35]、视频到传感器知识蒸馏方法(VKSD)[36]进行比较。实验结果如表2所列，对比方法的实验结果取自相应的论文，均使用F1分数作为评估指标。

表2 MMACT数据集上的性能比较

Table 2 Performance comparison of MMACT dataset

方法	数据类型	*F*1分数/%
DMSBFF	RGB视频、传感器	84.78
AGCN	骨架	87.74
Fusion-GCN	骨架、传感器	85.50
MMTSA	RGB视频、传感器	87.41
VKSD	骨架、传感器	73.64
Ours	骨架、传感器	**92.50**

2)CZU-MHAD数据集：选用双流自适应图卷积网络(AGCN)[12]、融合图卷积网络(Fusion-GCN)[29]、基于骨架的有向扩散图卷积网络(DD-GCN)[37]、基于骨架的移位图卷积网络(Shift-GCN)[38]、自适应多模态图表示融合方法(AMGRF)[39]进行比较。实验结果如表3所列，对比方法的实验结果取自相应的论文，均使用准确率作为评估指标。

表3 CZU-MHAD数据集上的性能比较

Table 3 Performance comparison of CZU-MHAD dataset

方法	数据类型	准确率/%
AGCN	骨架	93.17
Fusion-GCN	骨架、传感器	94.38
DD-GCN	骨架	94.62
Shift-GCN	骨架	95.64
AMGRF	骨架、传感器	**98.67**
Ours	骨架、传感器	96.97

3)UTD-MHAD数据集：选用基于多阶段的多模态数据特征融合方法(MSBFF)[34]、双流自适应图卷积网络(AGCN)[12]、融合图卷积网络(Fusion-GCN)[29]、3D卷积神经网络(Res3D-101)[40]、多模态深度时空卷积网络(Ds-CNN)[41]、视频到传感器知识蒸馏方法(VKSD)[36]、单一主体多视图关键信息表示与多模态融合方法(Fusion)[42]进行比较。实验结果如表4所列，对比方法的实验结果取自相应的论文，均使用准确率作为评估指标。

表4 UTD-MHAD数据集上的性能比较

Table 4 Performance comparison of UTD-MHAD dataset

方法	数据类型	准确率/%
MSBFF	RGB视频、传感器	96.05
AGCN	骨架	95.81
Fusion-GCN	骨架、传感器	94.42
Res3D-101	骨架、深度图	93.57
DsCNN	骨架、深度图	95.34
VKSD	骨架、传感器	96.97
Fusion	骨架、深度图	93.26
Ours	骨架、传感器	**97.21**

3个公共数据集各有特点，保证了测试环境的多样性。MMACT是一个较大的数据集，涵盖了多种多样的场景，并且由于在采集中有时会存在遮挡等问题，造成了部分数据缺失的情况，这种数据的不完整性为模型处理现实世界中的挑战性场景提供了测试机会；另外两个数据集UTD-MHAD与CZU-MHAD规模相对较小，场景设置相对受限，但也同时确保了骨架数据的完整性，为模型在理想环境下的性能评估提供了基准。从实验结果来看，本文方法除了相比AMGRF方法在CZU-MHAD数据集上略微落后，在其他数据集上均优于其他方法。尤其是在3个公共数据集上的效果均优于骨架AGCN方法，这表明本文方法有效结合了骨架数据与传感器数据模态，使得二者的数据能够进行信息上的互补，并且在多个数据集上均取得了良好的效果，证明了本文方法的实用性与泛用性。

4.5 消融实验

为了深入研究所提方法中每个组成部分的有效性，本节将进行相应的实验评估。

1)多传感器的互补效应：为了验证多传感器的互补效应，在MMACT数据集上进行了消融实验。

具体来说，位于不同关节点的传感器有能力捕捉与其相应关节点的局部运动信息，并且相比于仅依赖视觉信号提取的骨架序列，这类局部传感器信号不受视角、遮挡或关键点识

别误差的影响。此外，各种传感器可以提供互补信息，有助于更全面地理解人体运动过程。

为论证上述内容，设计了以下几组对比实验：不使用传感器数据，仅使用骨架数据的基础情况（w/o sensor）；额外融合右腕的局部运动信息手表传感器数据 W；额外融合右大腿的局部运动信息的手机传感器数据 P；以及同时引入两者的 $W+P$。实验结果如图 6 所示，表明本文方法可以自然地保持和改进多传感器的互补效应这一特性。

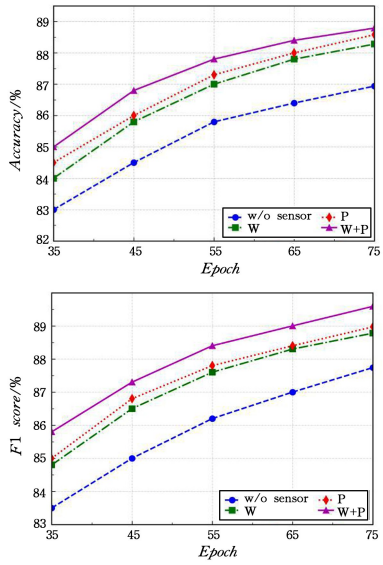


图 6 不同传感器组合的性能比较

Fig. 6 Performance comparison of different sensor combinations

2）图卷积网络的改进效应：为了验证在图卷积网络结构中引入拓扑增强模块（Topo）与局部身体划分注意力机制（Local Attention）对模型性能的提升作用，在 UTD-MHAD 数据集上进行了消融实验。

Topo 模块通过引入拓扑特征编码机制，对原始骨架图进行几何结构的增强建模，有效提升了模型对空间结构变化的敏感性，尤其在涉及非刚性变形或长距离交互的动作中表现更佳。而 Local Attention 模块将骨架图划分为多个区域，以强化模型对局部动态特征的建模能力，使模型能够分别捕捉不同部位的运动模式，并在融合阶段实现区域间的协同建模，从而提升整体动作识别准确率。

为评估这两个模块的独立及协同作用，设计了以下几组对比实验：基础模型 AGCN，加入拓扑模块的 Topo，加入局部划分模块的 Local，以及同时引入两者的 Topo+Local。实验结果如图 7 所示。

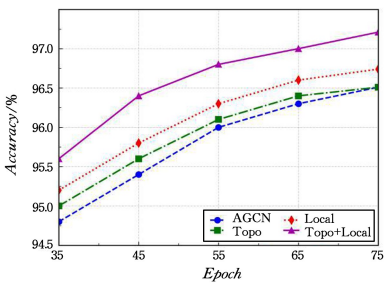


图 7 不同模块组合的 GCN 性能比较

Fig. 7 Performance comparison of GCN with different module combinations

可以看出，两者结合后效果最优，说明二者在建模思路上具有良好的互补性，在保持原空间建模能力的同时，加强了对局部信息的敏感度。

结束语 本文针对传统骨架动作识别在局部细节建模和结构表达能力方面的不足，提出了一种基于骨架与可穿戴传感器数据融合的多模态动作识别框架。在自采集击剑数据集与标准动作识别数据集上的实验结果表明，本文方法在动作类别识别任务中展现出了卓越水平，验证了多模态融合策略的有效性以及模型的泛化能力。特别是在复杂、快速的击剑动作场景中，将传感器数据与骨架信息融合能够更准确地刻画运动轨迹和细节变化。但本文方法仍存在一些不足：对当前局部区域划分的策略为静态设定，未能根据不同动作进行自适应调整；以及传感器信息的建模仍依赖于手动设计的映射关系，存在进一步优化空间。未来的研究将考虑引入其他模态数据，寻找能够更有效融合各模态数据的方法，并探索基于图学习的区域自适应划分方法，以提升模型对多样化动作的适应能力。

参 考 文 献

[1] SONG Y F, ZHANG Z, SHAN C, et al. Constructing stronger and faster baselines for skeleton-based action recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(2): 1474-1488.

[2] RAO D S, RAO L K, BHAGYARAJU V, et al. Enhanced Depth Motion Maps for Improved Human Action Recognition from Depth Action Sequences[J]. Traitement du Signal, 2024, 41(3): 1461-1472.

[3] LAI Y T, LIN C H, CHOU P Y. Real-Time Point Cloud Action Recognition System with Automated Point Cloud Preprocessing [C]// 2024 IEEE International Conference on Consumer Electronics (ICCE). IEEE, 2024: 1-7.

[4] YANG Y, YANG H, LIU Z, et al. Fall detection system based on infrared array sensor and multi-dimensional feature fusion [J]. Measurement, 2022, 192: 110870.

[5] ZHU K, WONG A, MCPHEE J. Fencenet: Fine-grained footwork recognition in fencing[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 3589-3598.

[6] TAO W, CHEN H, MONIRUZZAMAN M, et al. Attention-Based Sensor Fusion for Human Activity Recognition Using IMU Signals[J]. arXiv: 2112. 11224, 2021.

[7] AHMAD Z, KHAN N. Towards improved human action recognition using convolutional neural networks and multimodal fusion of depth and inertial sensor data[C]// 2018 IEEE International Symposium on Multimedia (ISM). IEEE, 2018: 223-230.

[8] AKTAS M E, AKBAS E, FATMAOUI A E. Persistence homology of networks: methods and applications[J]. Applied Network Science, 2019, 4(1): 1-28.

[9] LE V T, TRAN-TRUNG K, HOANG V T. A comprehensive review of recent deep learning techniques for human activity recognition [J]. Computational Intelligence and Neuroscience, 2022, 2022(1): 8323962.

[10] YAN S,XIONG Y,LIN D. Spatial temporal graph convolutional networks for skeleton-based action recognition[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2018.

[11] LI M,CHEN S,CHEN X,et al. Actional-structural graph convolutional networks for skeleton-based action recognition[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019:3595-3603.

[12] SHI L,ZHANG Y,CHENG J,et al. Two-stream adaptive graph convolutional networks for skeleton-based action recognition [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019:12026-12035.

[13] HU L,LIU S,FENG W. Spatial Temporal Graph Attention Network for Skeleton-Based Action Recognition[J]. arXiv: 2208.08599,2022.

[14] CHEN Z,LI S,YANG B,et al. Multi-Scale Spatial Temporal Graph Convolutional Network for Skeleton-Based Action Recognition[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2021:1113-1122.

[15] SUN Z,KE Q,RAHMANI H,et al. Human action recognition from various data modalities: A review[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(3): 3200-3225.

[16] BOULAHIA S Y,AMAMRA A,MADI M R,et al. Early,intermediate and late fusion strategies for robust deep learning-based multimodal action recognition[J]. Machine Vision and Applications,2021,32(6):121.

[17] CHEN T,MO L. Swin-fusion:swin-transformer with feature fusion for human action recognition[J]. Neural Processing Letters,2023,55(8):11109-11130.

[18] QIU S,FAN T,JIANG J,et al. A novel two-level interactive action recognition model based on inertial data fusion[J]. Information Sciences,2023,633:264-279.

[19] CHOI H,BEEDU A,ESSA I. Multimodal Contrastive Learning with Hard Negative Sampling for Human Activity Recognition [J]. arXiv:2309.01262,2023.

[20] HU Z,XIAO J,LI L,et al. Human-centric multimodal fusion network for robust action recognition[J]. Expert Systems with Applications,2024,239:122314.

[21] YUAN Z,YANG Z,NING H,et al. Multiscale knowledge distillation with attention based fusion for robust human activity recognition[J]. Scientific Reports,2024,14(1):12411.

[22] CHEN Z,SONG X,ZHANG Y,et al. Intelligent Recognition of Physical Education Teachers' Behaviors Using Kinect Sensors and Machine Learning[J]. Sensors & Materials,2022,34(3): 1241-1253.

[23] HAN J Z,ZHAO J J,YUE Y,et al. Edge Computing-based Video Action Recognition Method and Its Application in Online Physical Education Teaching[J]. IEEE Access, 2024, 12: 148666-148676.

[24] DING X,PENG W,YI X. Evaluation of physical education teaching effect based on action skill recognition[J]. Computational Intelligence and Neuroscience,2022,2022(1):9489704.

[25] FU D,CHEN L,CHENG Z. Integration of wearable smart devices and internet of things technology into public physical education[J]. Mobile Information Systems, 2021, 2021(1): 6740987.

[26] SRI-IESARANUSORN P,GARCIA F C,TIAUSAS F,et al. Toward the perfect stroke:A multimodal approach for table tennis stroke evaluation[C]// 2021 Thirteenth International Conference on Mobile Computing and Ubiquitous Network (ICMU). IEEE,2021:1-5.

[27] YUAN H,NI D,WANG M. Spatio-temporal dynamic inference network for group activity recognition[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 7476-7485.

[28] HU L,LIU S,FENG W. Spatial temporal graph attention network for skeleton-based action recognition[J]. arXiv: 2208.08599,2022.

[29] DUHME M,MEMMESHEIMER R,PAULUS D. Fusion-gcn: Multimodal action recognition using graph convolutional networks[C]// DAGM German Conference on Pattern Recognition. Cham:Springer,2021:265-281.

[30] IJAZ M,DIAZ R,CHEN C. Multimodal transformer for nursing activity recognition[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022:2065-2074

[31] KONG Q,WU Z,DENG Z,et al. Mmact:A large-scale dataset for cross modal human action understanding[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019:8658-8667.

[32] CHAO X,HOU Z,MO Y. CZU-MHAD:a multimodal dataset for human action recognition utilizing a depth camera and 10 wearable inertial sensors[J]. IEEE Sensors Journal, 2022, 22(7):7034-7042.

[33] CHEN C,JAFARI R,KEHTARNAVAZ N. UTD-MHAD:A multimodal dataset for human action recognition utilizing a depth camera and a wearable inertial sensor[C]// 2015 IEEE International Conference on Image Processing (ICIP). IEEE, 2015:168-172.

[34] CHOI H,BEEDU A,HARESAMUDRAM H,et al. Multi-stage based feature fusion of multi-modal data for human activity recognition[J]. arXiv:2211.04331,2022.

[35] GAO Z,WANG Y,CHEN J,et al. Mmtsa:Multi-modal temporal segment attention network for efficient human activity recognition[J]. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies,2023,7(3):1-26.

[36] NI J,SARBAJNA R,LIU Y,et al. Cross-modal knowledge distillation for vision-to-sensor action recognition[C]// ICASSP 2022-2022 IEEE International Conference on Acoustics,Speech and Signal Processing(ICASSP). IEEE,2022:4448-4452.

[37] LI C,HUANG Q,MAO Y. Dd-gcn:Directed diffusion graph convolutional network for skeleton-based human action recognition[C]// 2023 IEEE International Conference on Multimedia and Expo(ICME). IEEE,2023:786-791

[38] CHENG K,ZHANG Y,HE X,et al. Skeleton-based action recognition with shift graph convolutional network[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020:183-192.

[39] LIU X, YUAN G, BING R, et al. When Skeleton Meets Motion: Adaptive Multimodal Graph Representation Fusion for Action Recognition[C]// 2024 IEEE International Conference on Multimedia and Expo(ICME). IEEE, 2024: 1-6.
[40] WU H, MA X, LI Y. Spatiotemporal multimodal learning with 3D CNNs for video action recognition[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 32(3): 1250-1261.
[41] ZHAO C, CHEN M, ZHAO J, et al. 3d behavior recognition based on multi-modal deep space-time learning[J]. Applied Sciences, 2019, 9(4): 716.
[42] CHAO X, JI G, QI X. Multi-view key information representation and multi-modal fusion for single-subject routine action recognition[J]. Applied Intelligence, 2024, 54(4): 3222-3244.

CHEN Haitao, born in 2001, postgraduate. His main research interests include multimodal data fusion methods in the field of action recognition, and so on.

ZHOU Yu, born in 1987, Ph.D, associate professor, is a member of CCF (No. P7618M). His main research interests include computational intelligence, machine learning and intelligent information processing.

(责任编辑:柯颖)

CCF 上海会员活动中心换届,新一届执监委产生

2026 年 1 月 18 日, CCF 上海会员活动中心(简称 CCF 上海)换届选举会议在杨浦区长阳创谷举行。CCF 秘书长唐卫清、CCF 副秘书长、CCF 业务总部 & 学术交流中心总经理束庆山以及上海部分高校、企事业单位的 29 名 CCF 上海委员和 10 余名会员参加。

CCF 上海换届筹备组组长谷大武教授(CCF 上海主席,2017-06－2019-06)介绍了换届选举规则,并对换届筹备过程及 CCF 上海委员的产生规则作了简要介绍。各候选人上台陈述过去对 CCF 所做工作、未来的工作设想以及自己对 CCF 的理解等。经过限时竞选演讲、现场答辩和无记名投票,产生了 CCF 上海新一届执行委员会和监督委员会成员。(名单后附)。

CCF 会士、上海前主席王晓阳教授(CCF 上海主席,2015-09－2017-06)、白硕博士(CCF 上海主席,2019-06－2023-12)均对分部未来工作做出了建议。CCF 秘书长唐卫清总结了 CCF 上海近年的发展成绩,在吸引优秀人才、国际合作、政府服务等方面寄予厚望,并期待新一届执委会和监委会能够持续打造 CCF 上海的品牌活动,增强会员活跃度和认可度。

未来 CCF 上海将继续奋勇前行。在新一届执委会和监委会成员带领下,打造高品质会员活动,举办更多创新的品牌活动,力求更好地服务会员,助力会员成长,期待更多有志之士加入。

附:CCF 上海新一届执行委员会、监督委员会成员名单(按姓氏拼音排序)

主　席

韩伟力　复旦大学计算与智能创新学院教授

副主席

方志军　东华大学信息与智能科学学院执行院长
丁　炎　上海麒辑教育科技有限公司首席执行官

秘书长

王昊奋　同济大学设计创意学院教授

执委会委员

韩　威　上海启迪创业孵化器有限公司董事长
李建华　华东理工大学计算机系主任
林俊宇　中国科协-复旦大学科技伦理与人类未来研究院高工
吕　俊　上海磐测信息技术有限公司首席执行官
唐心悦　联想集团副总裁
王加溢　蓝忆(上海)智能科技有限公司首席执行官
温　蜜　上海电力大学信息化与数据管理中心主任
张楚炜　上海外事服务中心教授级高工
赵登吉　上海科技大学信息科学与技术学院研究员

监委会主席

孔令和　上海交通大学电子信息与电气工程学院副院长、教授

监委会委员

熊　赟　复旦大学计算与智能创新学院教授
于晓东　上海杉达学院信息科学与技术学院教授

据 CCF 微信公众号