doi:10.3969/j.issn.1001-4616.2025.03.013

基于多流自适应时空图卷积网络的人体行为识别

古学茹[1],张备伟[1],丁　雯[2]

(1.南京财经大学信息工程学院,江苏 南京 210023)
(2.沙洲职业工学院数字化与微电子学院,河北 张家港 215699)

[**摘要**]　针对当前人体骨骼动作识别算法全局关系描述不够详尽、时空特征挖掘不够充分等问题,本文提出了一种基于多流自适应时空图卷积网络的人体骨骼动作识别模型(MsAST-GCN). 首先使用注意力机制和神经张量网络(NTN)算法求解每对关节点之间的连接强度,构建全局邻接矩阵;接着利用 topK 策略,根据连接强度动态选择前 K 个邻居节点,更新全局邻接矩阵;然后采用混合池化模型提取全局上下文信息及时间关键帧特征. 通过对关节信息、骨骼信息、关节运动信息、骨骼运动信息同时进行建模,加强模型提取的特征对动作的表征能力;最后在数据集 NTU-RGB+D 上开展实验. 结果表明,该模型在人体骨骼动作识别任务中取得了良好的性能,有效提高了动作识别的准确率.

[**关键词**]　动作识别,注意力机制,图卷积神经网络,多流网络
[**中图分类号**]TP391.41;TP183　[**文献标志码**]A　[**文章编号**]1001-4616(2025)03-0112-08

Human Behavior Recognition Based on Multi-stream Adaptive Spatio-temporal Graph Convolutional Networks

Gu Xueru[1],Zhang Beiwei[1],Ding Wen[2]

(1.School of Information Engineering,Nanjing University of Finance and Economics,Nanjing 210023 China)
(2.Shazhou Polytechnic College,Zhangjiagang 215699,China)

Abstract:Aiming the problems of insufficient mining of temporal and spatial features in current human action recognition algorithms,a human skeleton action recognition model based on multi-stream adaptive spatio-temporal graph convolutional networks is proposed in this paper. Firstly,the attention mechanism and neural tensor network(NTN)algorithm are used to solve the connection strength between each pair of nodes,and the global adjacency matrix is constructed. Then,topK strategy is used to dynamically select the first K neighbor nodes according to the connection strength and update the global adjacency matrix. Next a hybrid pooling model is used to extract the global context information and time key frame features. By modeling joint information,bone information,joint movement information and bone movement information,the ability of the features extracted from the model to represent the movement is strengthened. Finally,experiments are carried out on the data set NTU-RGB+D. The results show that the model has a good performance in the task of human skeleton action recognition,and effectively improves the accuracy of action recognition.

Key words:action recognition,attention mechanism,graph convolutional neural network,multistream network

基于骨骼的动作识别[1-2]通过对人体骨骼运动信息的分析与识别,实现对各种动作的理解与分类. 相较于基于图像的动作识别,基于骨骼的动作识别不易受外界因素的干扰,具有较高的精确性和鲁棒性. 这一技术在安全监控、健康医疗、智能家居等多个领域展现出了极其关键的社会应用价值[3-4]. 例如,骨骼动作识别可以识别异常行为(如打架斗殴),提升公共安全;在健康医疗中,实时监测和分析患者动作有助于制定个性化康复治疗和健康管理;在智能家居中,通过捕捉用户动作信息自动调节家居环境,为日常生活

收稿日期:2024-10-25.
基金项目:国家自然科学基金项目(61877061).
通讯作者:张备伟,博士,副教授,研究方向:大数据处理、图像处理、模式识别. E-mail:zhangbeiwei@nufe.edu.cn

带来便利.

基于深度学习的人体动作识别方法主要有卷积神经网络(CNN)、循环神经网络(RNN)、Transformer网络和图神经网络(GCN). 前三种方法通常将数据处理成格状结构,而骨骼序列更符合图序列的定义,图卷积网络应运而生. Yan 等[5]引入了时空图的概念并构建了能体现人体结构拓扑的邻接矩阵. Zhao 等[6]提出自适应空间图卷积层来自适应地学习关节之间的相关性. Bai 等[7]提出了具有显著图像特征的中心连通图卷积网络(SIFE-CGCN),捕捉相似动作间的微小差异.

尽管图卷积神经网络在许多领域实现了显著进步,但仍然面临一些挑战:(1)大部分图卷积网络采用人体拓扑结构来构建对应的邻接矩阵,不能充分挖掘人体结构在空间维度上的特征;(2)人体运动中每个肢体的交互范围是有限的,直接使用全连接矩阵定义全局关系不可靠;(3)时间卷积仅采用 $K_t\times1$ 的卷积核提取特征;(4)大部分网络只对关节流进行了建模,没有充分考虑到其他数据流,缺少一定的数据支撑.

为此,本文设计了一种多流自适应时空图卷积网络的人体骨骼动作识别模型,并开展了以下研究:(1)将全局上下文感知注意力和 NTN 网络结合,创新性地提出了一种自适应全局邻接矩阵的构建方法,在提高节点表示准确性的基础上,增强对人体空间维度结构特征的提取能力;(2)提出 topK 全局邻接关系计算模型,消除一些不必要的节点间的协同关系;(3)在时间维度上引入混合池化模型,并结合时间维度的卷积提取方法,以增强时间特征提取的精确度;(4)同时对关节流、肢体流、关节运动流、肢体运动流进行建模,进行动作识别.

1 相关原理技术

1.1 图卷积神经网络

由于传统 CNN 模型只能用于处理规则的一维序列或二维矩阵数据,但实际中的许多数据并不具有这样的规则结构,于是研究人员将传统卷积扩展到图结构,提出了图卷积网络. 图卷积可以简单描述为将一个节点的所有邻居按照不同的权重叠加. 图卷积网络的计算公式为:

$$H^{(l+1)}=\sigma(\tilde{\boldsymbol{D}}^{-\frac{1}{2}}(\tilde{A}+M)\tilde{\boldsymbol{D}}^{-\frac{1}{2}}H^{l}W^{l}), \tag{1}$$

式中,H^l 表示第 l 层的节点特征表示,$\tilde{A}=A+I$,$\boldsymbol{I}$ 为单位矩阵,$\tilde{\boldsymbol{D}}$ 为 $\tilde{A}$ 的度矩阵,M 是一个可学习的掩码,$\boldsymbol{W}$ 是网络学习的权重参数矩阵,$\sigma(\cdot)$是激活函数.

1.2 相关网络架构

Shi 等[8]设计的双流自适应图卷积网络(2s-AGCN)采用双流结构,同时对关节流和骨骼流进行建模. 该模型通过 BP 算法以端到端的方式学习图的拓扑. 架构由 9 层自适应图卷积块堆叠而成,每一模块包含空间图卷积层、批归一化层、ReLU 非线性激活、Dropout 正则化、时间图卷积层和残差连接. 其中,空间图卷积层将代表人体物理结构的邻接矩阵和由高斯函数确定的依赖于数据的图相结合,形成新的邻接矩阵进行图卷积,提取空间特征信息. 尽管该模型在捕捉时空特征方面表现优异,但其计算复杂度高且训练时间较长.

Wang 等[1]提出的基于多模态图自注意力网络(MGSAN)利用自注意力机制聚合局部和全局特征,并通过多模态学习增强对不同骨骼关节间依赖关系的建模能力. 然而,模型仍可能面临计算复杂度较高的问题,尤其是在处理大规模数据时,图自注意力的计算开销可能导致效率瓶颈. 此外,模型在数据依赖、跨模态融合、训练效率等方面仍存在一定的不足.

Bai 等[7]提出显著图像特征增强中心连通图卷积网络(SIFE-CGCN)用于识别相似动作. 该模型引入中心连接策略捕捉关节间的连接关系,并融合显著性图像特征以改善动作区分. 尽管该模型在识别相似动作方面取得了优异成绩,但它面临较高的计算开销,图像特征提取对识别精度的依赖较大,且骨骼数据和图像特征的融合仍需进一步优化.

2 模型结构

2.1 构建骨骼时空图

实际情况中,关节点之间的关系在运动过程中是不断变化的,故 ST-GCN 提出的固定拓扑图 A 并不能

表示关节点之间不断变化的连接模式. 因此,本文设计了一个自适应的时空图卷积层.

$$f_{\text{out}}=\sum_{k}^{K_v}\boldsymbol{W}_k f_{\text{in}}(A_k+B_k)\,, \tag{2}$$

式中,f_{in}和f_{out}分别为卷积层的输入和输出,$\boldsymbol{W}_k$ 为权重矩阵. 具体来说,图的邻接矩阵由 A_k 和 B_k 两个部分构成. A_k 代表了预定义的拓扑图,用以描绘人体的物理连接. B_k 是基于样本数据动态生成的图,反映了在特定时间关节点之间的连接强度. 为了确定关节点之间的连接关系及其强度,本文应用神经张量网络[9]来计算关节点间的相似性,并通过 topK 策略筛选出比较重要的连接. 通过这种自适应设计,邻接矩阵可以根据动作的时域特征动态调整,确保在不同的运动阶段捕捉到关键的动态变化.

2.1.1　注意力机制

在人体行为识别任务中,不同帧中同一关节点的运动呈现出一定的相关性,但各帧的动作对整个运动序列的重要性存在差异. 这要求在不同帧间为各个关节点赋予差异化的权重分配. 图 1 展示了全局上下文感知注意力机制的具体架构,其详细的计算公式如下:

$$h=\sum_{t=1}^{T}f(v_t^{\mathrm{T}}c)v_t=\sum_{t=1}^{T}f\left(v_t^{\mathrm{T}}\tanh\left(\left(\frac{1}{T}\sum_{n=1}^{T}v_t\right)\boldsymbol{W}\right)\right)v_t \tag{3}$$

式中,v_t 为关节点在 t 帧上的特征,c 表示关节点的全局上下文特征,f 表示 sigmoid 激活操作,T 是该动作样本的帧数,$\boldsymbol{W}$ 是可学习的权重矩阵.

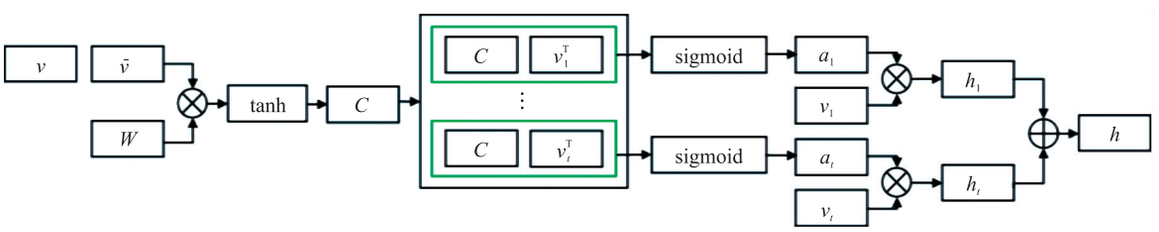


图 1　注意力机制架构图

Fig. 1　The framework diagram of attention mechanism

具体地,c 的计算公式为:$c=\tanh\left(\left(\frac{1}{T}\sum_{t=1}^{T}v_t\right)W\right)$,它是经过非线性变化的简单平均值. 通过权重矩阵 W,c 提供了该关节点的时间全局结构和特征信息. 根据 c,可以为该关节点的每一帧计算注意力权重 a_t. 对于关节点 v_t,通过计算 c 和关节点的内积来获取该关节点的注意力权重,这可以确保类似全局上下文的关节点获得更高的权重. 然后,应用 sigmoid 函数,确保值始终在(0,1)之间. 最后,关节点特征 $h\subseteq R^C$ 是帧的加权,$h=\sum_{t=1}^{T}a_t v_t$.

2.1.2　NTN 网络

针对前面生成的关节特征,对两个关节间的关系进行建模的最简单的办法就是内积,但内积会导致两者的交互不足. 因此,本文采用 NTN 来模拟两个关节点之间的关系:

$$g(h_i,h_j)=f\left(h_i^{\mathrm{T}}\boldsymbol{W}^{[1:K]}h_j+V\begin{bmatrix}h_i\\h_j\end{bmatrix}+b\right), \tag{4}$$

式中,h_i、h_j 是注意力层生成的关节点特征,$\boldsymbol{W}^{[1:K]}$ 是权重张量,[ ]代表连接操作,V 也是权重向量,b 是长度为 K 的偏移向量,$f(\cdot)$代表激活函数,K 是一个超参数,控制模型为每对关节生成的相关性分数数量. 神经张量网络示意图如图 2 所示.

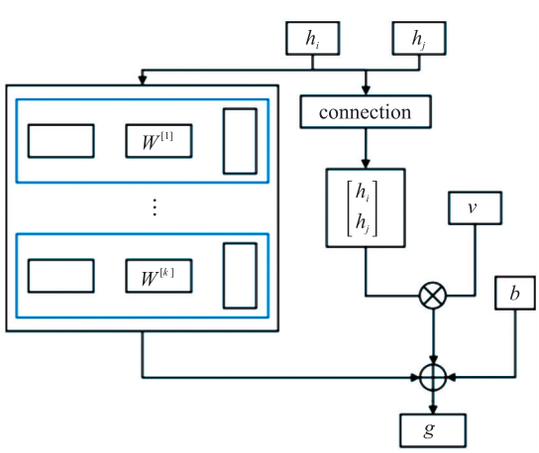


图 2　NTN 网络

Fig. 2　NTN network

2.1.3　topK 全局邻接关系自适应策略

实际运动中,单个肢体无法与人体的全部肢体同

时互动,且在不同的动作阶段,同一肢体的互动对象可能会随着时间变化. 为解决这一问题,提出了基于topK 的全局邻接关系更新策略(简称 topK 策略).

topK 策略通过选择每个肢体在特定时间节点上与其最相关的 K 个邻接肢体,自适应调整邻接关系. 作为排序算法,$f_{topK}(\cdot)$的主要功能是筛选出排序后向量中前 K 个元素的位置索引,并返回. K 值基于特定实验结果设定.

在动态邻接矩阵的生成过程中,首先通过 NTN 网络计算每对关节点间的相似度,这些相似度反映了关节点间的连接强度. 然后使用 topK 策略从相似度值中选择前 K 个较大值,构建稀疏的动态邻接矩阵. 对于每一帧,动态邻接矩阵只保留了与目标关节点交互强度最高的 K 个连接,这样可以有效减小计算复杂度,集中关注重要的关节点交互. 图 3 展示了自适应图卷积层的架构,清晰地标出了如何从 NTN 网络生成的相似度矩阵出发,结合 topK 策略动态构建邻接矩阵.

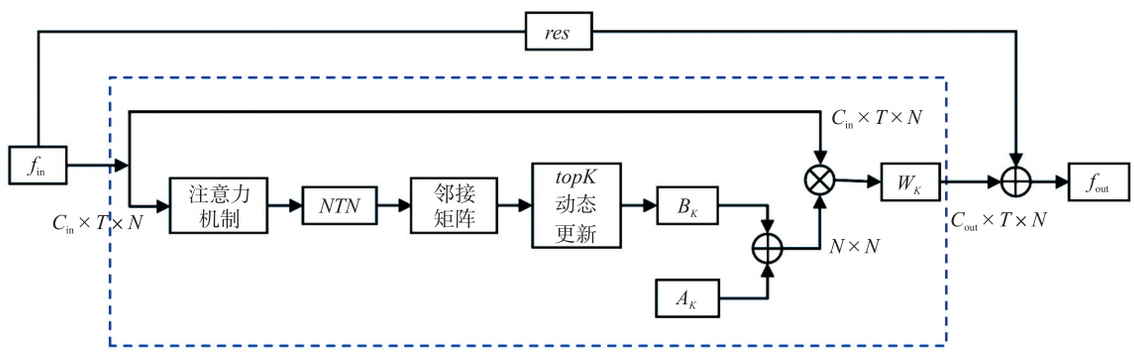


图 3　自适应图卷积层的架构示意图

Fig. 3　Schematic diagram of the architecture of the adaptive graph convolutional layer

2.2　时间维度上的混合池化模型

传统的 $K_t\times1$ 卷积核在处理复杂动作模式时,难以全面捕捉不同时间尺度上的动态变化,可能导致对关键时间维度特征的忽视. 为了有效提高时间特征的提取能力,本文引入了一种混合池化模型,结合不同类型的池化操作,通过多层次的特征融合增强了对时间特征的捕捉能力,如图 4 所示.

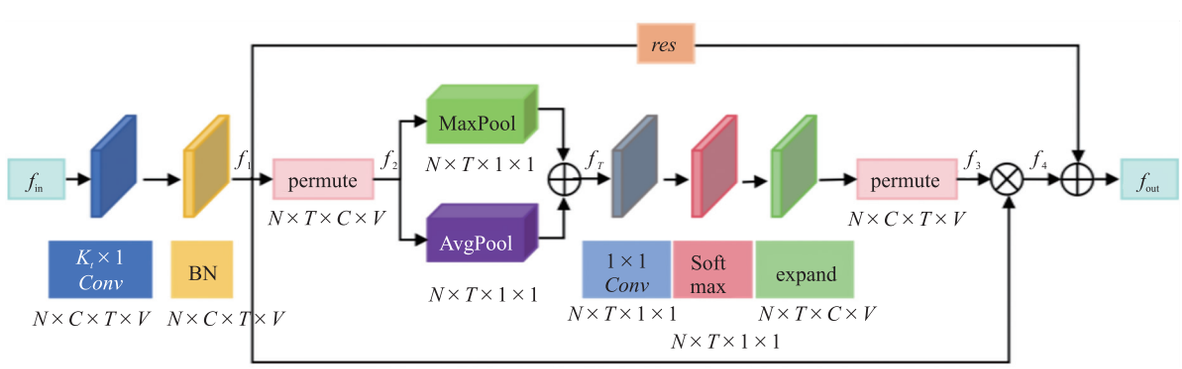


图 4　混合池化模型

Fig. 4　Hybrid pooling model

该模型的实现过程包括以下几个步骤:首先输入特征,尺寸为 $N\times C\times T\times V$;第 2 步为时间卷积,使用 $K_t\times1$ 的卷积核对特征进行卷积,随后进行批归一化处理;第 3 步为混合池化,在卷积层输出后,维度置换操作随之进行,然后分别进行全局平均池化和最大池化,如式(5)和(6)所示,f_2 为归一化后的特征,f_T 为混合池化后的特征;第 4 步为特征增强,利用 1×1 的卷积对特征 f_T 进行处理,应用 softmax 实施归一化,并通过 expand 将特征尺寸恢复至池化前的状态,然后再次进行置换操作. 过程如式(7)所示,f_3 为再次置换操作后的特征;第 5 步为输出,引入残差 Res,将 f_4 与 f_1 相加,得到特征输出 f_{out},如式(8)所示.

$$f_T=Concat(MaxPool(f_2),AvgPool(f_2)) \tag{5}$$

$$f_T=\frac{2}{C\times V}\sum_{k=1}^{C}\sum_{j=1}^{V}f_2(k,j) \tag{6}$$

$$f_4 = f_1 \otimes f_3 = \sum_{k=1}^{C}\sum_{i=1}^{T}\sum_{j=1}^{V} f_1^k(i,j) f_3^k(i,j) \tag{7}$$

$$f_{\text{out}} = Res(f_1, f_4) = f_1 + f_4 \tag{8}$$

第 3 步中通过将通道维度和时间维度进行交换,增强了模型对时间序列动态变化的敏感性. 上述内容揭示了混合池化模型的优势,它能同时提取时间维度上的关键帧动作特征和全局上下文特征. 当这些特征与通过时间卷积获取的特征相结合时,能显著提升对动作信息时间维度特性的捕捉效能.

2.3　整体结构

本文使用多流结构框架,对关节点信息、骨骼信息、关节点运动信息、骨骼运动信息分别进行建模,通过提取更多特征来增强动作识别效果. 图 5 描述了多流自适应时空图卷积网络的整体结构. 首先对骨骼数据进行预处理,得到原始数据;然后提取特征并输入至模型中. 模型通过 L 层的时空图卷积模块实现从输入到输出的全程端到端训练. 假设骨骼的两个关节点坐标为 $v_1=(x_1,y_1,z_1)$,$v_2=(x_2,y_2,z_2)$,则骨骼的坐标为 $bone_{(v1,v2)}=(x_2-x_1,y_2-y_1,z_2-z_1)$. 运动信息指的是同一关节点在不同帧间的变化. 假设第 t 帧的第 i 个关节点坐标为 $v_{i,t}$,则运动信息为 $motion_{i,t}=v_{i,t}-0.5v_{i,t-1}+v_{i,t+1}$. 骨骼运动信息的计算同理. 最后对多流网络的分类结果进行加权求和. 具体的权重比例由实验确定.

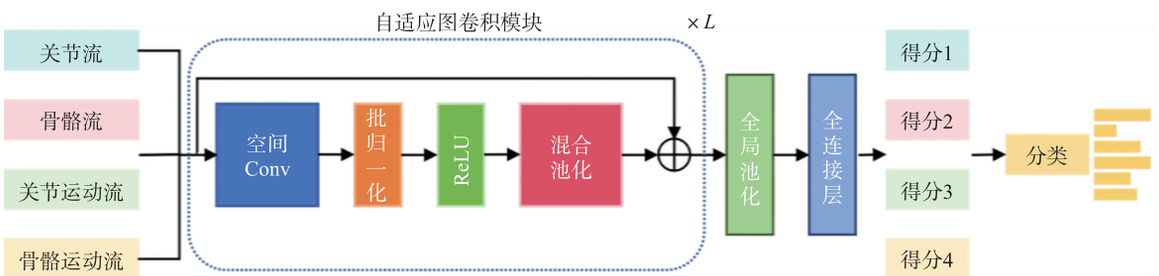


图 5　模型整体结构框架

Fig. 5　The overall structural framework of the model

3　实验结果与讨论

3.1　数据集

NTU-RGB+D 数据集是 Shahroudy 等[10]在 2016 年提出的,由 56680 个动作样本组成,包含 60 个动作类别. 该数据集分为三类:40 个日常动作(如“挥手”、“跳跃”等)、9 个健康相关动作(如“跌倒”、“头痛”等)和 11 个交互动作(如“拥抱”、“握手”等). 图 6 展示了部分采样帧图片. 该数据集提出交叉实验者评估设置(X-Sub)和交叉视角评估设置(X-View). X-Sub 根据人物 ID 将 40 个实验者分为两类,训练集和测试集各 20 个实验者;X-View 根据相机 ID 划分,相机 ID 为 2,3 的为训练集,相机 ID 为 1 的为测试集.

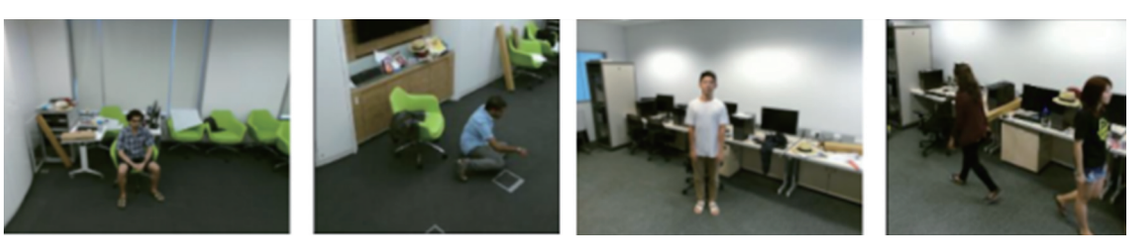


图 6　NTU-RGB+D 数据集样本示例

Fig. 6　Sample example of the NTU-RGB+D dataset

3.2　参数配置

该实验基于 Pytorch 框架,在系统为 Ubuntu20.04,GPU 为 NVIDIA GeForce RTX 4090 D,显存为 24 GB 的服务器上运行. 训练轮数设置为 40,确保模型有足够的时间优化和收敛. 对于 NTU-RGB+D 数据集,训练批次大小为 64,测试批次大小为 256,较大的批次有助于提高训练效率并保证足够的内存处理大规模数据. 设置初始学习率为 0.1,以加快训练过程并保证网络能够快速找到合适的参数范围. 在优化器的选择上,该实验采用的是 SGD(Stochastic Gradient Descent,随机梯度下降),并配置了 Nesterov 动量参数为 0.9,

以及权重衰减设定为 0.0001. 衰减值过大会导致模型的表达能力下降,衰减值过小可能达不到有效正则化的效果,而 0.000 1 的设定值提供了良好的平衡. 与梯度下降(GD)相比,SGD 具有更快的收敛速度,增加了避免陷入局部最优解的可能. 选择交叉熵(Cross-Entropy)作为损失函数,反向传递误差.

3.3 实验结果与分析

模型的层数不同,其识别准确率大概率不同. 如果模型的层数过少,可能会导致提取的特征不足;反之,则可能会导致过拟合问题. 为了研究该模型的最佳层次结构,本文依据骨骼流数据构建了 3 至 9 层的 MsAST-GCN 模型,并在 X-Sub 上进行了对比实验,结果见表 1. 当层数为 3 时,其性能达到了最佳水平.

表 1 MsAST-GCN 模型不同层数识别精确度对比

Table 1 Comparison of recognition accuracy of different layers in the MsAST-GCN model %

层数	3 层	4 层	5 层	6 层	7 层	8 层	9 层
骨骼流	88.19	86.00	85.35	83.73	81.22	70.70	83.77

为了验证注意力机制和 NTN 的有效性,以传统内积和无注意力机制的模型为基线,结合 NTN、注意力机制构建多流网络,并在 X-Sub 上进行消融实验,实验结果见表 2. 从表中可以看出,引入 NTN、注意力机制的模型相较于基线模型,准确率有所提升,且 MsAST-GCN 的识别精度优于其他模型. 消融实验充分验证了 NTN 和注意力机制的有效性.

表 2 X-Sub 上的消融实验识别精确度对比

Table 2 Comparison of recognition accuracy of ablation experiments on X-Sub %

模型	关节流	骨骼流	关节运动流	骨骼运动流	双流	多流
基线	87.43	87.64	88.61	84.83	88.78	89.20
引入 NTN	87.60	87.93	88.72	85.27	88.92	89.35
引入注意力机制	87.71	87.78	88.85	85.04	88.94	89.42
MsAST-GCN	87.88	88.19	89.06	85.39	89.06	89.50

为了研究不同 K 值对识别准确率的影响,在 X-Sub 上基于骨骼流进行实验验证. 不同 K 值下该模型的识别准确率如表 3 所示. 根据表 3 可知,K 的取值不同,其最终识别准确率也不相同,即不同的全局邻接矩阵确实会对识别准确率造成影响,加入该模块后准确率有不小的提高. 由数据可知,K 值在 5~15 之间准确率上升,15~25 之间下降,$K=15$ 时准确率较高,表明此时的全局关系比较合理. 故 K 的最佳取值为 15.

表 3 模型不同 K 值识别精确度对比

Table 3 Comparison of recognition accuracy of different K values in the model %

K	5	10	15	20	25
准确率	85.35	86.51	88.19	85.54	84.05

为了研究多流、双流、单流网络对识别准确率的影响,在 NTU 数据集上进行对比实验,不同流模型的最终准确率见表 4. 根据表 4 可知,双流网络 MsAST-GCN 模型的准确率分别为 89.06%、95.77%,而多流网络 MsAST-GCN 模型的识别准确率分别为 89.50%、96.81%,相较于双流模型,多流网络模型分别提升了 0.44%、1.04%,进一步证明了多流网络在性能上的优势.

表 4 该模型各分支的识别精确度

Table 4 The recognition accuracy of each branch of this model %

数据集	关节流	骨骼流	关节运动流	骨骼运动流	双流	多流
X-Sub	87.88	88.19	89.06	85.39	89.06	89.50
X-View	94.80	94.08	92.19	92.52	95.77	96.81

为了验证该模型的优越性能,在 NTU-RGB+D 数据集上,将该模型与其他主流的骨骼动作识别模型进行对比实验. 实验的具体结果参见表 5. 本文的 MsAST-GCN 模型取得了较为优异的成绩,其识别准确率高于大部分的算法,这充分说明了该模型优越的识别性能. 与 GCN 领域的经典方法 ST-GCN 和 2s-AGCN 相比,本文模型的识别准确率分别提升了 8.00%,8.51%和 1.00%,1.71%,这显现出相较于传统 GCN 方法的

显著改进. 同时,与最新的 GCN 模型 3D-GCN 和 CGCN 相比,模型的识别准确率分别提升了 3.10%,4.71%和 0.82%,1.26%,这说明该模型在与当前最先进的 GCN 方法竞争时也展现出了一定的优势.

表 5　NTU-RGB+D 数据集上多种模型的识别精确度

Table 5　Recognition accuracy of multiple models on the NTU-RGB+D dataset　%

模型	ST-GCN[5]	PeGCN[11]	3D-GCN[12]	Co-ConvT[13]	2s-AGCN[8]	CGCN[7]	KA-AGTN[14]	IA-ASGCN[6]	STFE-GCN[15-16]	**MsAST-GCN**
X-Sub	81.50	85.60	86.40	88.10	88.50	88.68	88.80	90.50	89.80	**89.50**
X-View	88.30	93.40	92.10	94.30	95.10	95.55	94.60	95.80	96.00	**96.81**

尽管该模型在动作识别任务中表现良好,但仍存在一些识别错误,如对于动作“阅读”和“撕纸”的混淆. 这可能源于它们在视觉上具有一定的相似性,尤其是在动作初期(如双手接近纸张并开始移动),而模型可能过多关注了这些前期的相似特征. 在“跳跃”与“跳起来”动作的识别中,模型混淆率较高. 分析认为,这与两者关节运动差异较小且过渡较快有关,导致模型难以准确区分. 通过混淆矩阵分析模型在不同动作类别中的准确率(表 6),可见某些动作的错误率较高,表明模型在处理具有相似运动模式的动作时存在一定困难.

表 6　部分动作的识别准确率与错误率

Table 6　Recognition accuracy and error rate of some actions　%

动作类别	阅读	撕纸	跳跃	跳起来
识别准确率	88.5	85.2	92.3	90.1
错误率	11.5	14.8	7.7	9.9

4　结论

本文提出了一种多流自适应时空图卷积网络的人体骨骼动作识别模型. 使用注意力机制和 NTN 算法计算关节点间的连接强度,构建全局邻接矩阵,并采用 topK 策略动态选择前 K 个邻居节点,更新全局邻接矩阵,进一步详尽地描述全局关系. 采用混合池化模型提取全局上下文特征和时间关键帧特征,进一步充分提取时间特征. 对关节、骨骼、关节运动和骨骼运动信息同时进行建模,进一步加强了模型提取的特征对动作的表征能力.

实验结果表明,MsAST-GCN 模型在人体骨骼动作识别任务中取得了良好的性能,有效提高了动作识别的准确率. 实验结果不仅证明了多流输入在捕捉动作特征中的优势,还为后续研究提供了参考. 未来的工作可以进一步探索其他信息流的融合方式及其对模型性能的影响,同时扩大数据集和场景的多样性,以进一步提升模型的泛化能力和应用范围.

[参考文献]

[1] WANG J, LI Z, LIU B, et al. MGSAN: multimodal graph self-attention network for skeleton-based action recognition[J]. Multimedia systems, 2024, 30(6): 1-12.

[2] XIANG X, LI X, LIU X, et al. A GCN and transformer complementary network for skeleton-based action recognition[J]. Computer vision and image understanding, 2024, 249: 104213.

[3] WANG Z, CHEN T. Human-UAV interactive perception: skeleton-based iterative perspective optimization algorithm for UAV patrol tracking of large-scale pedestrian abnormal behavior[J]. Applied soft computing, 2024, 167: 112467.

[4] ZHANG Y, ZHAO C, YAO Y, et al. Human posture estimation and action recognition on fitness behavior and fitness[J]. Alexandria engineering journal, 2024, 107: 434-442.

[5] YAN S, XIONG Y, LIN D. Spatial temporal graph convolutional networks for skeleton-based action recognition[C]// Proceedings of the 32nd AAAI Conference on Artificial Intelligence. New Orleans, USA: AAAI Press, 2018: 7444-7452.

[6] ZHAO Y, WANG J, WANG H, et al. Adaptive spatiotemporal graph convolutional nerwork with intermediate aggregation of multi-stream skeleton features for action recognition[J]. Neurocomputing, 2022, 505: 116-124.

[7] BAI Z, DING Q, XU H, et al. Skeleton-based similar action recognition through integrating the salient image feature into a center-connected graph convolutional network[J]. Neurocomputing, 2022, 507: 40-53.

[8] SHI L, ZHANG Y, CHENG J, et al. Two-stream adaptive graph convolutional networks for skeleton-based action recognition[J]. The IEEE conference on computer vision and pattern recognition (CVPR). Long Beach, USA: IEEE Computer Society, 2019: 12026-12035.

[9] BAI Y, DING H, BIAN S, et al. SimGNN: a neural network approach to fast similarity computation[C]//The 20th Internationsl Conference on Web Search and Data Mining. Melbourne, Australia: ACM, 2019: 384-392.

[10] SHAHROUDY A, LIU J, NG T, et al. NTU RGB+D: a large scale dataset for 3D human activity analysis[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE Computer Society, 2016: 1010-1019.

[11] YOON Y, YU J, JEON M. Predictively encoded graph convolutional network for noise-robust skeleton-based action recognition[J]. Applied Intelligence, 2022, 52(3): 2317-2331.

[12] ZHANG G, WEN S, LI S, et al. Fast 3D-graph convolutional networks for skeleton-based action recognition[J]. Applied soft computing, 2023, 145: 110575.

[13] 石跃祥,朱茂清. 基于骨架动作识别的协作卷积 Transformer 网络[J]. 电子与信息学报,2023,45(4):1485-1493.

[14] LIU Y, ZHANG H, XU D, et al. Graph transformer network with temporal kernel attention for skeleton-based action recognition[J]. Knowledge-based systems, 2022, 240: 108146.

[15] 曹毅,吴伟官,李平,等. 基于时空特征增强图卷积网络的骨架行为识别[J]. 电子与信息学报,2023,45(8):3022-3031.

[16] 王小军,王兴,林羽,等. 基于时间感知注意力与拥塞驱动图卷积的交通流量预测[J]. 福建师范大学学报(自然科学版),2025,41(1):1-10.

[责任编辑:杜忆忱]