

姚砺, 魏钰菁, 万燕. 基于密集图卷积和注意力的攀爬行为识别技术[J]. 智能计算机与应用, 2024, 14(6): 50-55. DOI: 10.20169/j.issn.2095-2163.240607

# 基于密集图卷积和注意力的攀爬行为识别技术

姚 砺, 魏钰菁, 万 燕

(东华大学 计算机科学与技术学院, 上海 201620)

**摘要:** 为解决大量特征信息在经过特征提取网络层层传递过程中不断被削弱以及针对时空和通道信息提取不充分的问题, 本文基于 MST-GCN 提出了基于注意力和密集图卷积的攀爬行为识别方法。首先, 在 MST-GCN 原先的时空图卷积网络中引入了密集连接, 重构了特征提取网络, 获取更丰富的关节之间的上下文关系; 其次, 在每层卷积块中引入卷积块注意力模块 (CBAM), 沿着通道和空间维度生成注意力特征图, 加强模型对关键信息的特征提取能力。实验结果表明, 本文所提出的方法相对于基线网络对攀爬行为的识别准确率提升了 11.2%, 并且超过当前其他方法。

**关键词:** 骨架攀爬行为识别; 密集连接; CBAM

中图分类号: TP391.4

文献标志码: A

文章编号: 2095-2163(2024)06-0050-06

## Climbing behavior recognition technology based on dense graph convolution and attention

YAO Li, WEI Yujing, WAN Yan

(College of Computer Science and Technology, Donghua University, Shanghai 201620, China)

**Abstract:** To address the problem that a large amount of feature information is continuously weakened during the process of feature extraction network layer-by-layer propagation, as well as the insufficient extraction of spatiotemporal and channel information, this paper proposes an attention-based and dense Graph Convolutional Network (GCN) method for rock climbing behavior recognition based on the MST-GCN. Firstly, dense connections are introduced into the spatial-temporal GCN network of the MST-GCN, reconstructing the feature extraction network to obtain a more comprehensive contextual relationship between joints. Then, the Convolutional Block Attention Module (CBAM) is introduced into each layer of the convolutional block to generate attention feature maps along channel and spatial dimensions, enhancing the model's ability to extract key feature information. Experimental results show that the proposed method in this paper improves the recognition accuracy of climbing behavior by 11.2% compared with the baseline network, and surpassing other current methods.

**Key words:** skeleton-based climbing behavior recognition; dense connection; CBAM

## 0 引言

随着社会的发展和技术的进步, 人们在生活中越来越多地使用监控视频, 以确保公共安全。城市化进程的加速和高层建筑的增多, 使攀爬行为成为一个亟需关注的社会问题。攀爬行为也可能造成公共设施损坏、治安问题以及危及他人生命安全等不容忽视的一系列后果。传统的监控技术通常利用视频监控设备, 以及配备一定的人员, 安全人员需要实时监控画面, 容易疲惫和疏忽, 这样一来将无法

及时发现异常情况, 面对突发事件的实时性则难以保证。因此, 研究如何基于监控视频自动化识别攀爬行为具有重要意义。

人体动作识别是计算机视觉中一项重要的任务, 在诸如人机交互<sup>[1]</sup>、安防与监控<sup>[2]</sup>和医疗保健等领域中得到广泛应用。人体行为识别主要就是在视频中理解人类的行为, 和图像特征不同的是, 视频中的特征不仅包括空间信息, 还包括重要的时间信息, 这也正是人体行为识别的关键。

早期的行为识别研究主要是基于 RGB 视频序

**作者简介:** 魏钰菁 (1997-), 女, 硕士研究生, 主要研究方向: 图像处理; 万 燕 (1970-), 女, 博士, 教授, 硕士生导师, 主要研究方向: 图像处理, 纤维自动识别。

**通讯作者:** 姚 砺 (1967-), 男, 博士, 副教授, 硕士生导师, 主要研究方向: 软件测试技术。Email: weiyujing0226@163.com

收稿日期: 2023-04-28

列,包括传统的方法<sup>[3-5]</sup>,利用机器学习提取人工设计特征,然后通过选择合适的分类算法进行识别分类;还包括基于深度学习<sup>[6-8]</sup>的方法。由于基于RGB数据的行为识别容易受到环境光线、背景干扰等因素的影响,对人体姿势的变化和遮挡比较敏感,而基于骨架信息的行为识别则不受这些因素的影响,因其只关注人体骨架信息,抗干扰能力更强、鲁棒性更高。目前基于骨架信息的行为识别也开始受到研究者的广泛关注。

传统的骨架信息行为识别方法也是手工设计提取特征,选择合适的分类算法进行分类识别<sup>[9-10]</sup>。早期基于深度学习的方法是通过将骨架序列手动构造坐标向量序列或者伪图像,再输入到循环神经网络(RNN)<sup>[11-12]</sup>或者卷积神经网络(CNN)<sup>[13-14]</sup>去学习人体运动特征。而人类的骨骼是典型的非欧数据,CNN和RNN无法在这样的结构上充分发挥其作用。GCN可以将人体骨架序列数据表示为图结构,利用这个图结构进行特征学习和预测,从而实现对人体行为的识别和分类。ST-GCN<sup>[15]</sup>第一个将图卷积应用到人体行为识别中,ST-GCN将骨架数据视为由节点和边组成的图形,其中节点表示人体关节,边表示关节之间的连接。在该图形上,ST-GCN采用基于卷积的方法从局部区域中提取特征,并使用多个ST-GC模块相结合学习全局特征,最终用于分类。Shi等学者<sup>[16]</sup>提出2个独立的流来分别处理时间和空间信息,并且通过自适应图卷积网络学习不同的GCN层和骨架的拓扑结构。MST-GCN<sup>[17]</sup>提出多尺度时空图卷积来捕捉空间和时间域的短期和长期依赖关系,丰富了模型在时空域上的感受野。

上述方法都是经过若干层神经网络提取人体行为特征,没有考虑到特征信息在多层图卷积网络传递过程中会在很大程度上被削弱的问题,并且都忽略了通道和空间的信息交互,只关注自身的信息。

本文针对上述问题,对骨架序列行为识别模型MST-GCN进行改进。在MST-GCN中,针对通过多层GCN特征提取会削弱大量局部特征信息的问题,同时受到DeepGCNs<sup>[18]</sup>中DenseGCN的启发,将多尺度时空图卷积中的普通图卷积改为密集图卷积。同时在每层卷积块中引入CBAM<sup>[19]</sup>模块,通过注意力模块能够聚焦更加有效的空间和通道信息,抑制无用的信息,获得更准确的行为特征,提升行为识别的准确率。随后对攀爬行为数据集Climb-skeleton进行研究。结果表明本文提出的方法识别准确率明显高于其他方法在Climb-skeleton的识别准确率。

## 1 改进的MST-GCN模型结构

MST-GCN是一种时空图卷积网络,可以对空间和时间信息进行建模。其中,多尺度时空图卷积模块不仅能获得近距离关节依赖关系,也能捕获长距离关节依赖关系,并且考虑到了远程时间信息,可以很好地应用于骨架序列建模。每个骨架序列被建模为一个图输入到MST-GCN网络中,先是通过多尺度空间图卷积模块,主要是将空间特征和图卷积细分为一系列子集,构造成分层残差结构,从而分层处理特征,有利于捕获到较远距离的关节依赖关系。接着通过多尺度时间图卷积,同样细分为一系列子集,每个子集中使用固定内核大小以及不同的输入进行卷积操作。通过10层特征提取网络后,将通过特征提取网络提取出来的特征再经过池化层和全连接层实现分类识别。

多层图卷积神经网络扩大了感受野,获得了更大范围的邻居信息,但大量特征信息通过层层GCN的传播在很大程度上会被削弱。同时在骨架结构特征的传播过程中,卷积块会有大量冗余信息。对于空间模块,每一个动作都会有关键的关节点;而对于通道模块,每一个通道对于每一个特征都有特定的语义信息,通道信息和空间信息很难得到充分的交互,导致预测不准确。

本文基于上述问题,针对骨架序列的行为识别提出了基于密集图卷积和注意力的攀爬行为识别方法,整体框架如图1所示。

本文在MST-GCN的基础上对原始的多尺度时空图卷积引入密集连接,以及在每层的卷积块中引入CBAM注意力模块两项改进。图卷积采用密集连接可以解决过多层的图卷积带来的特征信息被削弱的问题。注意力模块能够聚焦更加有效的空间和通道信息,抑制无用的信息,获得更准确的行为特征。对此可展开论述如下。

(1)本文的人体骨架信息利用人体姿态估计算法OpenPose获得的,从视频中获取每一帧的人体骨架数据,每一帧上的所有人体的关节点对应一个坐标,分别是 $x, y$ 和置信度(score)。

(2)改进后的模型首先将骨架序列数据构造为具有时空信息的图结构,为了消除输入之间的差异性和防止梯度消失,对骨架序列数据进行归一化处理。

(3)接下来,输入到10层特征提取网络。首先会通过引入密集连接实现多尺度时空密集图卷积,

能够在多尺度上反复使用上下文特征信息,增加特征之间的连接,可以更好地捕捉行为之间的关系和依赖,让网络可以获得更加全面的时空特征信息,并且保留了更多人体姿势的结构信息。

(4)然而在基于骨骼的行为识别中,通道信息和空间信息都非常重要,共同构成了骨骼序列的特征表示,都可以区分运动的较为明显的特征,综合利

用通道信息和空间信息可以提高基于骨骼的行为识别任务的性能和可靠性。因此在每个卷积块中,引入卷积块注意力模块(CBAM),在特征提取时关注更关键的通道和空间信息,实现信息的交互。通过在添加了CBAM模块的卷积块中进行密集图卷积,捕获到更加丰富的全局和局部信息,提高模型的识别准确率,最终完成攀爬行为识别工作。

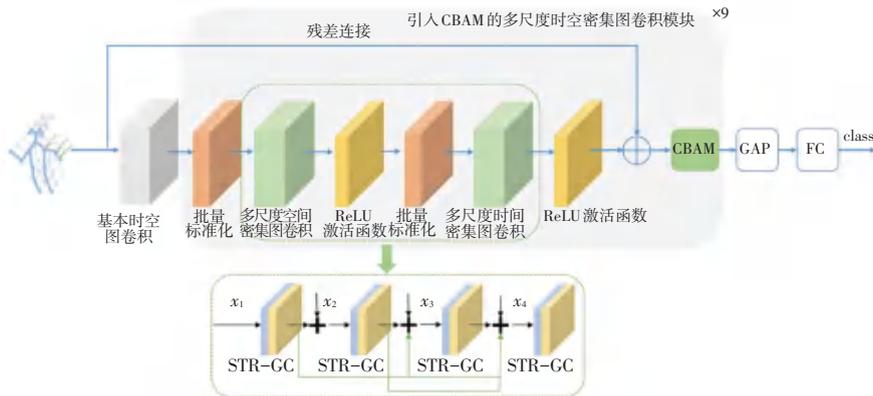


图1 模型整体框架结构图

Fig. 1 Overall framework diagram of the model

### 1.1 多尺度时空密集图卷积

在图卷积网络中,人体骨架图被定义为  $G = \{V, E\}$ , 其中  $V$  是骨架序列中所有骨架的集合,  $E$  是骨架序列中所有骨骼的集合。然后再将相邻帧的相同关节点连接起来作为时间边, 此时就将骨架序列构成了骨架时空图。所以在此网络中, 中间的特征可以表示为  $X \in R^{C \times T \times V}$ , 其中  $C$  表示通道数,  $T$  表示视频帧数,  $V$  表示骨架序列中的关节点数。多尺度时空图卷积主要是将特征和图卷积细分为一系列子集, 同时将这些子集构造成分层残差结构, 可以分层处理特征。

本文利用密集连接可以复用上下文特征信息的优点, 将 MST-GCN 中原始的多尺度时空图卷积模块中引入密集连接, 多尺度时空密集图卷积模块如图 2 所示。

在本模块中, 将输入特征  $X$  根据通道维度分割为  $n$  个片段, 表示为  $x_i$ , 其中  $i \in \{1, 2, \dots, n\}$ 。此时, 将每个时空图卷积细分为  $n$  个子图卷积, 残差和密集连接设置在 2 个相邻片段中, 这重用了上下文特征信息, 增加特征之间的连接, 可以更好地捕捉行为之间的关系和依赖, 每一个多尺度时空密集图卷积可以表示为:

$$y_i = \begin{cases} TC_i(GC_i(x_i)), & i = 1 \\ TC_i(GC_i(x_i + y_{i-1} + y_{i-2} + \dots + y_2 + y_1)), & i > 1 \end{cases} \quad (1)$$

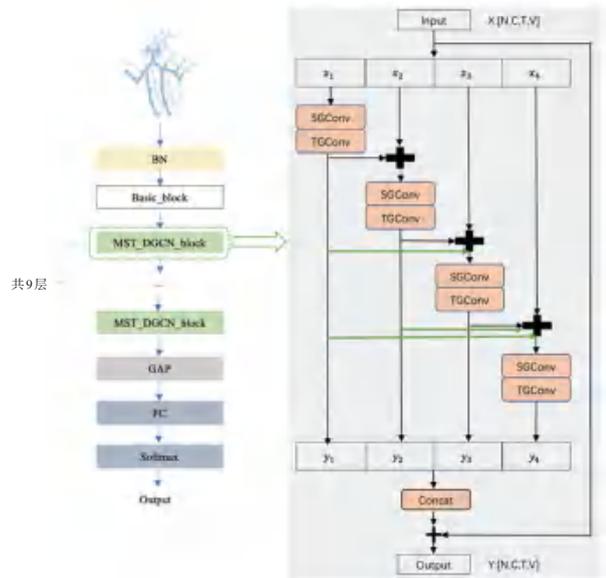


图2 多尺度时空密集图卷积模块结构图

Fig. 2 Multi-scale spatial-temporal dense map convolution module structure diagram

最后, 将所有的片段都拼接起来, 并将额外的残差连接应用到整个模块, 让模型收敛得更好。整个模块输出的计算公式为:

$$Y = \sigma([y_1; y_2; \dots; y_n] + X) \quad (2)$$

其中,  $\sigma$  表示激活函数。通过此模块输出的特征  $Y$  复用了多层特征信息, 充分利用了骨架序列的数据, 增强了特征之间的连接, 避免了特征信息在多层卷积后逐渐消失的问题, 更好地捕捉行为之间的

关系和依赖,可以提取更准确的行为特征。

## 1.2 引入卷积块注意力模块

前面提出的多尺度时空密集图卷积可以有效避免特征提取过程中特征信息被削弱的问题,但同时也会因为密集连接带来很多冗余信息,对特征提取造成干扰。在识别过程中,并不是所有的关节信息都是有用的,一个动作的区别主要在于关键的关节点。并且每一个通道对于每一个特征片段都有不同的语义信息。

卷积块注意力模块(CBAM)包含2个子模块:通道注意力模块和空间注意力模块。其中,通道注意力模块用于学习不同通道的重要性,以增强特征图中的有用信息。空间注意力模块则用于学习不同位置的重要性,以增强特征图中具有重要位置的信息。CBAM设计结构如图3所示。

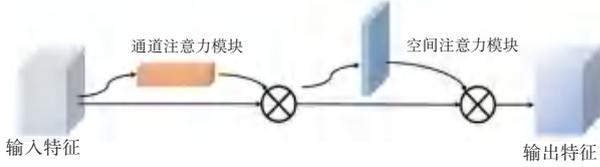


图3 卷积块注意力模块结构图

Fig. 3 Structure diagram of Convolutional Block Attention Module

CBAM中的通道注意力模块和空间注意力模块可以让通道信息和空间信息进行有效交互。因此,本文将CBAM引入到模型内每层的卷积块中,聚焦图中的重要特征,抑制无用的特征信息。具体来说,是将经过多尺度密集图卷积模块产生的中间特征 $X \in R^{C \times T \times V}$ 输入到CBAM模块中,首先通过通道注意力模块,产生一维通道注意力特征图 $A_c \in R^{C \times 1 \times 1}$ ,然后通过空间注意模块生成二维空间注意力特征图 $A_s \in R^{1 \times T \times V}$ 。整个过程可以表示为:

$$\begin{cases} X' = A_c(X) \otimes X \\ X'' = A_s(X') \otimes X' \end{cases} \quad (3)$$

其中,“ $\otimes$ ”表示元素级的乘法。在乘法过程中,中间采用广播机制进行维度变换和匹配, $X''$ 为最终的特征输出。

## 2 实验和结果

### 2.1 实验环境和数据集

本文的实验环境基于Ubuntu18.04 64位LTS操作系统,系统内存为32G,使用2块RTX 2080Ti显卡进行训练,使用pytorch 1.2和python 3.6。所有实验的学习率均设置为0.015,优化方法首先采用SGD随机梯度下降方法,损失函数选取交叉熵损失函数,

每次参加训练的批次(*Batch size*)为10,总共训练110轮。模型在50轮之后,每一轮训练结束都会进行一次保存操作。

本文使用的数据集主要是来自于包含400类动作视频数据集Kinetics-400<sup>[20]</sup>,每段视频的长度在10s左右。相较于其他视频动作识别数据集,这是一个更大规模、更多样化的数据集,使得行为识别算法的研究变得更加准确和普适。Kinetics-skeleton是从Kinetics 400提取出的骨架数据集,本文主要研究监控视频是否存在攀爬行为,所以在Kinetics-skeleton筛选出适用于识别监控视频中是否出现攀爬行为的数据集Climb-skeleton。Climb-skeleton包括climbing a rope、ice climbing、climbing ladder等与攀爬行为相关数据,共有3684条数据。其他的是随机选取3684条数据使之与攀爬类相近的数据量。最后,训练集共有6868条数据,验证集有500条数据。

## 2.2 实验结果比较和分析

### 2.2.1 多尺度时空密集图卷积前后的实验结果和分析

第一部分实验针对特征信息在多层图卷积神经网络长时间传递中不断被削弱导致关节之间依赖关系不充分,将多尺度时空图卷积中普通的图卷积模块改进为密集连接的图卷积模块,称为多尺度时空密集图卷积。为了验证密集图卷积的有效性,在Climb-skeleton数据集进行了实验,先后测试了普通图卷积模块和密集图卷积模块。实验结果见表1。

表1 改进多尺度时空图卷积前后对比实验结果

| Table 1 Experimental results comparison before and after convolution of improved multi-scale spatio-temporal maps | %    |
|---|------|
| Action recognition algorithm  | Acc  |
| (a) MST-GCN baseline  | 40.8 |
| (b) MST-DGCN(Ours)  | 50.4 |

通过表1可以看出,本次实验改进后的模型在Climb-skeleton数据集上取得了较好的结果,准确性提高了9.6%。表明多尺度时空密集图卷积通过获取更丰富的关节间依赖关系和保留更多的人体姿势结构信息,使其能够更好地理解人体的运动模式,并能够更准确地识别和分析不同的姿势,提高网络模型的性能。

### 2.2.2 引入CBAM前后的实验结果和分析

第二部分实验主要是为了验证引入CBAM是

否能提升多尺度时空密集卷积网络对通道和空间的信息的提取能力,因此对第一部分实验改进后包含 CBAM 模块的网络和不含 CBAM 模块的网络在 Climb-skeleton 数据集进行了实验,同时也对 2 个网络在验证集做了验证对比,实验结果见表 2。

表 2 引入 CBAM 前后对比实验结果表

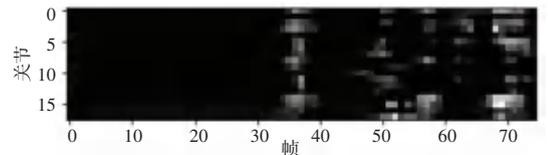
Table 2 Table of experimental results comparison before and after the introduction of CBAM %

| Action recognition algorithm | Acc  |
|------------------------------|------|
| (a) MST-DGCN                 | 50.4 |
| (b) MST-DGCN + CBAM          | 52.0 |

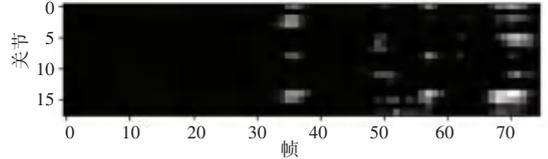
由表 2 可知,在多尺度时空密集卷积图中引入 CBAM 模块,在 Climb-skeleton 数据集上的准确性进一步提高了 1.6%,表明 CBAM 模块聚焦有效的通道信息和空间关节信息可以减少冗余信息的干扰,提升了网络模型的性能。图 4 给出了基于 Climb-skeleton 数据集输入特征在通过 CBAM 前后得到的可视化特征,图 4 中的横轴和纵轴分别表示帧和关节。由图 4(a)可以看出在特征输入到 CBAM 模块前存在大量冗余的时空信息;图 4(b)是通过 CBAM 后的特征,明显看出很多冗余信息被削弱,聚焦到有效信息,进一步证明了引入 CBAM 模块对提取时空特征的有效性。

### 2.3 经典算法对比

为了体现本文提出的模型的优越性,在 Climb-skeleton 数据集上与最先进的算法<sup>[15-16,21-23]</sup>进行了对比。实验的结果见表 3。



(a) 通过注意力模块前的特征



(b) 通过注意力模块后的特征

图 4 注意特征可视化

Fig. 4 Visualization of attention feature

表 3 攀爬行为识别准确率

Table 3 Climbing behavior recognition accuracy %

| Action recognition algorithm | Acc         |
|------------------------------|-------------|
| ST-GCN <sup>[15]</sup>       | 30.0        |
| 2s-AGCN <sup>[16]</sup>      | 40.4        |
| MSAAGCN <sup>[21]</sup>      | 50.4        |
| MS-G3D <sup>[22]</sup>       | 44.0        |
| ST-GAT <sup>[23]</sup>       | 44.4        |
| Improved MST-GCN(ours)       | <b>52.0</b> |

实验证明,本文提出的基于改进 MST-GCN 的算法相比 ST-GCN 和 2s-AGCN 等主流网络在识别攀爬行为时具有较明显的优势。

图 5 展示了本文的模型与 ST-GCN 联合 OpenPose 在一些实例识别的可视化对比结果,并且在类似于攀爬行为的识别中,通过获得更丰富的全局上下文信息即可根据上下文信息判断该行为是否为攀爬行为,以及辅以更准确的时空特征,得到更加准确的识别结果。

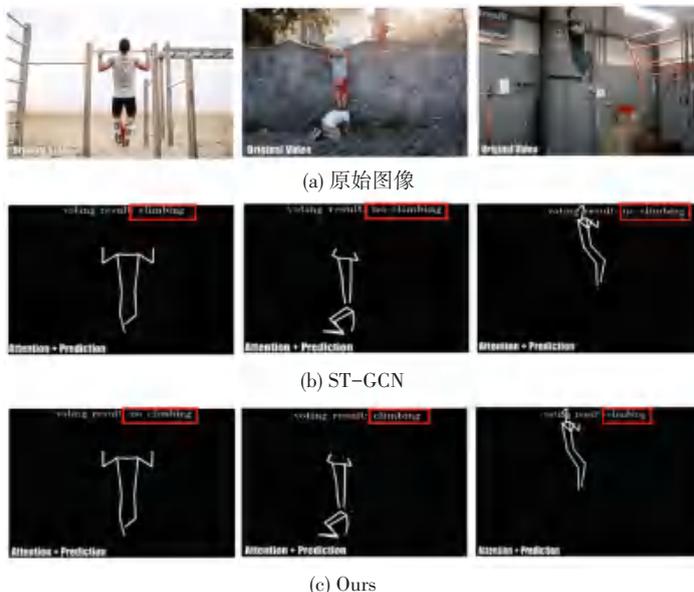


图 5 识别结果对比图

Fig. 5 Comparison chart of recognition results

### 3 结束语

在基于骨架序列的攀爬行为识别方法中,关键就在于提取充分的时空和通道特征以及获得关节间的全局上下文信息。本文基于 MST-GCN 提出了基于注意力和密集图卷积的攀爬行为识别方法。首先,是在 MST-GCN 原始多尺度时空图卷积引入密集连接,以此来解决关节特征信息在多层特征提取网络传递过程中不断被削弱的问题。然后,在每个多尺度时空密集图卷积块中引入 CBAM 模块,聚焦行为识别中的有效信息,增强模型对时空、通道特征的提取能力。对比其他算法,本文提出的方法能够有效地提升攀爬行为识别结果,高于其他算法。

### 参考文献

- [1] WANG Heng, SCHMID C. Action recognition with improved trajectories [C]// The IEEE International Conference on Computer Vision (ICCV). Sydney, Australia: IEEE, 2013: 3551–3558.
- [2] CHEN Y, WANG J, LI Y, et al. A review of human action recognition with depth cameras in surveillance systems[J]. Journal of Visual Communication and Image Representation, 2019, 65: 102649.
- [3] BOBICK A F, DAVIS J W. The recognition of human movement using temporal templates [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23(3): 257–267.
- [4] LAPTEV I. On space-time interest points[J]. International Journal of Computer Vision, 2005, 64(2–3): 107–123.
- [5] WANG Heng, KLSER A, SCHMID C, et al. Dense trajectories and motion boundary descriptors for action recognition [J]. International Journal of Computer Vision, 2013, 103(1): 60–79.
- [6] SIMONYAN K, ZISSERMAN A. Two-stream convolutional networks for action recognition in videos[C]//Proceedings of the 27<sup>th</sup> International Conference on Neural Information Processing Systems. Cambridge, USA: MIT Press, 2014: 568–576.
- [7] TRAN D, BOURDEV L, FERGUS R, et al. Learning spatiotemporal features with 3D convolutional networks [C]//Proc. of IEEE International Conference on Computer Vision. Washington DC, USA: IEEE Computer Society, 2015: 4489–4497.
- [8] DONAHUE J, HENDRICKS L A, GUADARRAMA S, et al. Long-term recurrent convolutional networks for visual recognition and description [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 677–691.
- [9] VEMULAPALLI R, ARRATE F, CHELLAPPA R. Human action recognition by representing 3D skeletons as points in a lie group [C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA: IEEE, 2014: 588–595.
- [10] XIA LU, CHEN C C, AGGARWAL J K. View invariant human action recognition using histograms of 3D joints [C]//Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington DC, USA: IEEE Computer Society, 2012: 20–27.
- [11] LIU Jun, SHAHROUDY A, XU Dong, et al. Spatio-temporal LSTM with trust gates for 3D human action recognition [C]//Proceedings of the 14<sup>th</sup> European Conference on Computer Vision. Cham: Springer, 2016: 816–833.
- [12] ZHANG Pengfei, LAN Cuiling, XING Junliang, et al. View adaptive neural networks for high performance skeleton-based human action recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(8): 1963–1978.
- [13] KE QiuHong, BENNAMOUN M, AN Senjian, et al. A new representation of skeleton sequences for 3D action recognition [C]//Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition. Washington DC, USA: IEEE Computer Society, 2017: 1063–6919.
- [14] LI Chao, ZHONG Qiaoyong, XIE Di, et al. Co-occurrence feature learning from skeleton data for action recognition and detection with hierarchical aggregation [C]//Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence. Stockholm, Sweden: ACM, 2018: 786–792.
- [15] YAN Sijie, XIONG Yuanjun, LIN Dahua. Spatial temporal graph convolutional networks for skeleton-based action recognition [J]. Proceedings of the 32<sup>nd</sup> AAAI Conference on Artificial Intelligence, (AAAI-18), the 30<sup>th</sup> Innovative Applications of Artificial Intelligence (IAAI-18), and the 8<sup>th</sup> AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18). New Orleans, USA: AAAI, 2018: 7444–7452.
- [16] SHI Lei, ZHANG Yifan, CHENG Jian, et al. Two-stream adaptive graph convolutional networks for skeleton-based action recognition [J]. arXiv preprint arXiv:1805.07694, 2019.
- [17] CHEN Zhan, LI Sicheng, YANG Bing, et al. Multi-scale spatial temporal graph convolutional network for skeleton-based action recognition [C]//Proceedings of the AAAI Conference on Artificial Intelligence. AAAI, 2021, 35(2): 1113–1122.
- [18] LI Guohao, MULLER M, THABET A, et al. Deepgcns: Can GCNs go as deep as CNNs? [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, Republic of Korea: IEEE, 2019: 9267–9276.
- [19] WOO S, PARK J, LEE J, et al. Cbam: Convolutional block attention module [C]//Proceedings of the European Conference on Computer Vision (ECCV). Munich, Germany: dblp, 2018: 3–19.
- [20] KAY W, CARREIRA J, SIMONYAN K, et al. The kinetics human action video dataset [J]. arXiv preprint arXiv:1705.06950, 2017.
- [21] SHI Lei, ZHANG Yifan, CHENG Jian, et al. Skeleton-based action recognition with multi-stream adaptive graph convolutional networks [J]. IEEE Transactions on Image Processing, 2020, 29: 9532–9545.
- [22] LIU Ziyu, ZHANG Hongwen, CHEN Zhenghao, et al. Disentangling and unifying graph convolutions for skeleton-based action recognition [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE, 2020: 143–152.
- [23] SONG J, SON J, SEO D, et al. ST-GAT: A spatio-temporal graph attention network for accurate traffic speed prediction [C]//Proceedings of the 31<sup>st</sup> ACM International Conference on Information & Knowledge Management. Atlanta, USA: ACM, 2022: 4500–4504.