

基于图卷积网络的行为识别方法综述

孔 玮, 刘 云[†], 李 辉, 王传旭

(青岛科技大学 信息科学技术学院, 山东 青岛 266061)

摘 要: 行为识别技术具有巨大的应用前景和潜在的经济价值, 广泛应用于视频监控、视频检索、人机交互、公共安全等领域。卷积神经网络虽被广泛应用, 但对非欧氏空间数据的处理具有局限性。而图卷积网络表现出基于图数据的依赖关系进行建模的强大功能, 成为行为识别领域的研究热点。该文主要概述基于图卷积网络的行为识别方法。图卷积网络主要有两大方法, 基于频谱的方法和基于空间的方法。首先, 该文从不同侧面分析了两种方法的优缺点, 综述了两种方法分别在行为识别领域的应用与发展。其次, 根据行为识别中图网络模型和算法设计的差异, 总结网络构造的关键方面, 对比不同算法对模型性能产生的影响。最后, 针对图卷积网络在行为识别中存在的问题, 对未来图卷积网络的发展进行了展望。

关键词: 图卷积网络; 行为识别; 神经网络; 深度学习; 非欧氏空间; 综述

中图分类号: TP391

文献标志码: A

DOI: 10.13195/j.kzyjc.2020.0514

开放科学(资源服务)标识码(OSID):



A Survey of Action Recognition Methods Based on Graph Convolution Network

KONG Wei, LIU Yun[†], LI Hui, WANG Chuan-xu

(College of Information Science and Technology, Qingdao University of Science and Technology, Qingdao 266061, China)

Abstract: Action recognition technology has great application prospects and potential economic value, and is widely used in video surveillance, video retrieval, human-computer interaction, public security and other fields. Although Convolutional Neural Network is widely used, it has limitations in dealing with data of non-Euclidean space. Graph Convolution Network shows the powerful function of modeling based on graph data dependency. It has become a research hotspot in the field of action recognition. This paper mainly summarizes action recognition methods based on Graph Convolution Network. There are two main methods of Graph Convolution Network: spectral-based method and spacial-based method. Firstly, for the two methods, this paper analyzes advantages and disadvantages from different aspects, summarizes their application and development in the field of action recognition. Secondly, according to differences of graph network model and algorithm design in action recognition, key aspects of network construction are summarized, and the influence of different algorithms on model performance is compared. Finally, according to problems existing in the action recognition based on Graph Convolution Network, future development of Graph Convolution Network is prospected.

Keywords: graph convolution network; action recognition; neural network; deep learning; non-Euclidean space; survey

0 引 言

在深度学习的发展背景下, 行为识别在智能监控、人机交互、虚拟现实和运动分析等领域有着更为广泛的需求, 识别性能得到明显提升。卷积神经网络(CNN)^[1-2]作为深度学习的代表性网络, 具有平移不变性, 可实现参数共享, 比传统的神经网络的识别效果有巨大的提高。尽管 CNN 在基于欧氏

空间的图片序列的特征提取过程中取得了巨大的成功, 但在许多实际的应用场景中的数据是从非欧氏空间生成的。例如, 在场景图^[3]的应用中, 物体与物体或物体与人之间的交互是基于图的学习系统。图结构含有更丰富的语义信息, 但图数据是不规则的, 导致一些重要的操作(例如卷积)在图像上很容易计算, 但不再适合直接用于图域。图数据的处

收稿日期: 2020-05-04; 修回日期: 2020-10-04.

基金项目: 国家自然科学基金项目(61702295, 61672305).

[†]通讯作者. E-mail: lyun-1027@163.com.

理对深度学习算法提出了重大挑战,同样需要进行有效的分析,图神经网络^[4]发展产生,成为深度学习^[5]中另一个新的研究热点。图神经网络主要划分为五类^[6]:图卷积网络(GCN)、图注意力网络、图自编码器、图生成网络和图时空网络。GCN作为图神经网络最活跃的分支,在行为识别领域得到广泛的关注。本文着重介绍图卷积网络(GCN)在行为识别中的应用。

图卷积方法主要有两大流派:基于频谱的方法和基于空间的方法。基于频谱的方法主要有Spectral CNN^[7](第一代GCN)、第二代GCN^[8]、ChebyNet^[9]、CayleyNet^[10]、一阶ChebyNet^[11]等。在行为识别领域,ChebyNet及一阶近似被广泛应用。基于空间的方法主要有消息传递与聚合、图采样与聚合、图结构序列化、关注卷积方式、基于注意力机制、关注感知域的方法,方法的详细介绍见第1.2部分。本文组织结构如图1所示。

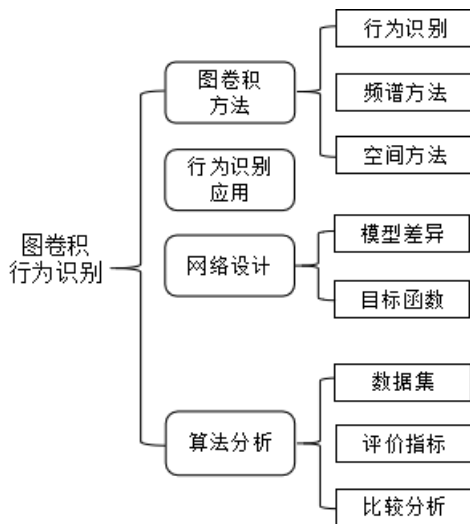


图1 组织结构图

1 相关工作

1.1 行为识别

根据行为复杂程度可把行为识别分为3大类:个体行为(Actions),交互动作(Interactions)以及团体活动(Group Activities)。其中,个体行为是指单个人的基本运动动作,如挥手、跑步等,也可以看作是多个姿势的组合;交互动作一般指的是人与人的交互或者人与物的交互,如握手、弹琴等,也是目前较受关注的行为识别类型;团体活动是指一个场景中包含多人和多物的活动,如排球比赛、团体会议等,也是最复杂的行为识别类型。

行为识别的方法,分为传统的方法和深度学习的方法。传统的方法是基于手工特征和规则提取特征的方法,例如SIFT^[12]方法、HOG^[13]方法、

SURF^[14]方法、iDT^[15-16]方法等,虽然识别性能逐步提升,但仍不能避免计算复杂度高、网络收敛速度慢、特征描述符不灵活等缺点,难以应用到复杂的场景中。基于深度学习的行为识别方法比较主流的有two-stream^[17-18]网络、C3D网络^[19]和LSTM^[20]循环网络等,利用端到端的学习,提取图像域中共有的局部特征,通过深度架构能减少网络中的参数数量。尽管该领域已取得丰硕的研究成果,但卷积架构适合图像序列等规则数据的处理,并且在复杂应用场景中,目标通常会受光照变化、噪声、环境等因素的影响,引起较大外观变化,导致算法性能下降。而基于人体骨架的行为识别方法因对光照和场景变换具有很好的鲁棒性,得到了越来越多的关注。通过人体姿态检测算法或高精度的深度摄像头可以较易获取骨架特征,并且骨骼点连接的人体关节自然形成图结构。给定人体关节位置的时间序列,能应用图卷积网络学习人体行为。在群组行为识别中,个体的行为决定群组的行为,对于“组”这个群体来说,每个个体可视为一个节点,那么图卷积网络也可以推广到群组行为识别中。

1.2 图卷积网络方法

随着GCN方法的发展,其方法主要分为两大类:基于频谱的方法(spectral-based)和基于空间的方法(spatial-based)。基于频谱的图卷积网络是将图信号^[21]与频谱信号进行转换和分析,然后再恢复图信号所在的空域,从而完成图信号的降噪与特征提取。以“降低复杂度”为主线,将常用于图像的卷积神经网络应用到图数据上。其主要方法如表1所示,K为阶数,E为边数。

表1 频谱图卷积方法

频谱方法	优点	缺点	计算复杂度
第一代GCN	泛化到非欧氏空间	扰动大,参数多	$O(n^3)$
第二代GCN	局部卷积,参数少	计算复杂度高	$O(n^2)$
ChebyNet	简便运算,预处理	参数多,不灵活	$O(K E)$
CayleyNet	捕获窄频带,灵活	泛化能力不强	线性
GCN	核小易训练,实用	计算消耗增加	$O(E)$

基于频谱的图卷积通过添加自我连接单位矩阵、归一化邻接矩阵A的方式解决了基于空间的图卷积忽略节点自身特征、邻接矩阵过于庞大的两个问题。但由于基于频谱的图卷积方法存在灵活性不高、普适性不强、运行效率较低等问题,基于空间的图卷积方法利用邻域聚合的思想降低了复杂度,增强了泛化能力,提高了运行效率,是行为识别领域在图结构数据上进行分析的基本思想。行为识别的空间图卷积涉及到的主要方法如表2所示。

表 2 空间图卷积方法

分类	主要方法	优点	缺点	发展方向
消息传递	NN4G ^[22]	特征多, 自适应	数值不稳	泛化性
	MPNN ^[23]	可通用, 易扩展	效率较低	自适应
	DCNN ^[24]	刻画高阶信息	难以拓展	灵活性
	MoNet ^[25]	核共享, 应用广	不利优化	拓展
图采样	GraphSage ^[26]	邻采样, 建难图	不好收敛	深网络
	FastGCN ^[27]	层采样, 速度快	自适应差	拓展
	Huang 等 ^[28]	自适应, 易训练	重复计算	优化
	Cluster-GCN ^[29]	子聚类, 效率高	类联系少	稀疏性
图序列化	Mathias 等 ^[30]	通用, 易并行	易过拟合	简计算
关注卷积	ST-GCN ^[31]	图首用, 新分割	结构固定	自适应
	GAT ^[32]	多注意, 强表达	对齐问题	拓展
	GaAN ^[33]	能力强, 稳训练	算力开销	通用
	h/cGao ^[34]	成本低, 易学习	难以优化	优化
注意力	SPAGAN ^[35]	加深图, 自更新	点关系少	强关联
	EGAT ^[36]	关注边, 增权重	点不关联	拓展
感知域	Non-Local ^[37]	感受野大, 灵活	信息单一	多特征
	StoGCN ^[38]	可大图优化	资源耗多	优化

2 图卷积网络在行为识别中的应用

行为识别的主要任务是对给定的动作信息提取特征, 并根据这些特征进行分类。这些动作信息包括图片上的动作行为、视频动作序列以及骨架信息等。对于图卷积网络, 主要应用在基于骨架图的单人行为识别, 交互网络以及基于个体建立关系图的群组行为识别等方面。在视频动作序列的建模过程中, 主要应用时空双流网络。如何利用图卷积在图结构上提取空域特征并与时域特征结合在一起, 如何使得提取的时空特征在行为识别中更具判断力, 仍是目前行为识别领域的重要挑战。在行为识别中涉及到的图卷积方法见表 3 所示。

2.1 频谱方法的应用

基于谱的主要方法已经在第 1 部分介绍, 而在行为识别邻域, ChebyNet 及一阶 ChebyNet 是主流。2017 年提出的一阶 ChebyNet, 即 GCN 是使用最广泛的图卷积模型, 从 2018 年开始, 许多行为识别方法就围绕上述谱方法展开。

(一) 不同构图方式。GGCN 提出基于骨架的行为识别的广义图卷积网络, 根据权重区分强弱连接, 捕捉关节坐标的变化, 实现高效的动作特征学习与识别。由于 3 个连续帧构成一个单元, 建模过程中忽略了内部帧之间的关联, 未来还需要挖掘广义图的内部关系, 提高其扩展性。与 GGCN 构图方式不同, 给定骨骼点视频序列, DPRL 深度渐进式强化学习利用连续帧的关节空间关系建模, 与利用马尔科夫决策过程提取的关键帧相互结合、相互优化, 实现行为类别预测。此模型由于结构图的邻接矩阵使用手工参数, 灵活性不足, 所以有必要探索一些自动设计权重的方法来提高图的自适应能力。

NAS 打破主流谱 GCN 一阶近似结构, 通过提供多个动态图模块丰富搜索空间, 加强节点的时空相关性, 探索了基于骨架动作识别的最佳 GCN 结构^[41]。

(二) 复杂图应用。为了简化卷积核, 早期图卷积网络只能处理低维数据, 卷积核关注局部特征^[57]而限制了其灵活性能, 不能处理大规模的复杂图。基于此, AGCN^[58]提出了一种新型频谱图卷积网络。网络的输入是多种不同的图结构, 频谱卷积核不再共享, 而是设置不同的拉普拉斯矩阵, 定制不同的频谱滤波器, 富有灵活性。这为行为识别推广到复杂图提供了借鉴和参考。Bayesian GC-LSTM 为了提高模型的泛化能力, 提出了一种对抗先验算法来调整模型参数^[43]。为了更好地捕捉数据的随机性和变异性, 将整个模型在贝叶斯框架下扩展为概率模型, 用一个贝叶斯推理问题来解决分类问题。此类方法由于存在密集矩阵运算, 计算复杂度高, 还需要采取合适的优化策略, 以方便训练和计算。

(三) 深层网络研究。用于行为识别的图卷积网络, 普遍存在网络层次不深, 节点特征容易平滑的缺点, PPNP 和 APPNP 网络的提出对此问题的解决提供了方向。使用 GCN 与 PageRank 之间的关系推导基于 PageRank 的个性化的传播方案^[59], 增加了传回中心节点的机会, 保留了局部性, 避免过渡平滑。与大多数模型相比, 此模型使用更少的参数和训练时间, 并且计算复杂度与边的数量呈线性关系。然而, 特征传播过程不进行参数学习, 因此可用在半监督任务中, 限制了其通用性能, 可作为行为识别领域日后继续研究的内容。

2.2 空间方法的应用

(一) 提升灵活性。ST-GCN 基于骨架建立的时空模型, 弥补了基于 RGB 模型中的不足, 其灵活性能为行为识别开辟了新的方向和可能。但是, ST-GCN 只构造人体的自然连接图, 忽视了非物理连接关节点间的联系, 缺乏对所有层中包含的多级语义信息进一步建模的灵活性和能力。并且, 此模型对数据依赖性不强, 对诸如“跳起来”和“坐下去”这种类似动作不好区分。基于此, 2S-AGCN 提出自适应的双流图卷积网络结构, 它允许加入自然连接以外的新连接来动态调整图结构, 以更好地适应模型的层次结构。AS-GCN 提出自监督的动作和结构连接结构, 分别挖掘潜在关节联系和高阶邻域信息。但此模型结构复杂, 不易计算, 可显示出对未来姿态预测有希望的结果。值得注意的是, ST-GCN 中只利用了节点的二维或三维坐标的一阶特征, 特

征单一, 而骨骼的长度和方向等特征对于行为识别来说更具信息性。为此, PB-GCN 把几何特征和运动特征相结合代替位置坐标, 丰富特征, 分图卷积。但对于不相连的子图, 特征传播比较困难, 不利于优化与收敛。有向无环图神经网络 (DGNN)

被提出。这个模型不仅能提取关节点和骨骼信息, 还能提取它们之间的方向关联信息。上面的研究, 主要应用的是普通图卷积的一阶邻域信息, 限制了网络的感受野。

表 3 基于图卷积的行为识别方法及目标函数

分类	频谱方法	空间方法	空间方法目标函数	符号含义
关注卷积	STGC ^[39]	ST-GCN ^[31] PB-GCN ^[44]	$f_{out} = \sum_j \Lambda_j^{-\frac{1}{2}} A_j \Lambda_j^{-\frac{1}{2}} f_{in} W_j$ $\sum_i W_{agg}(i) Y_i$	j 是子集, f_{in} 是输入, Λ 是度矩阵 子图加权融合, Y_i 是子图
自适应	DPRL ^[40]	2s-AGCN ^[45] AS-GCN ^[46]	$f_{out} = \sum_k^{K_v} W_k f_{in} (A_k + B_k + C_k)$ $X_{act} = \sum_{c=1}^C \hat{A}_{act}^{(c)} X_{in} W_{act}^{(c)}$	K_v 是子集, B 、 C 定义新连接 C 是连接种类, 可加新连接
注意力	—	RAGCN ^[47] BAGCN ^[48]	$f_{out} = \sum_{d=0}^{D_{max}} W_d f_{in} \left(\Lambda_d^{-\frac{1}{2}} A_d \Lambda_d^{-\frac{1}{2}} \otimes M_d \right)$ $X_{out} = \sum_{s \in S} M'_s \otimes \hat{A}_s^d \hat{X}_{mid} W'_s$	D_{max} 、 S 是子集, M 是掩膜 M 定义连接强度, 分配注意力
感受野	NAS ^[41]	AGC-LSTM ^[49] SemGCN ^[50] Wang 等 ^[51] Wen 等 ^[52]	$Y_{out} = \sum_{k=1}^K \Lambda_k^{-\frac{1}{2}} A_k \Lambda_k^{-\frac{1}{2}} X W_k$ $Z = \sum_i G_i X W_i$ $Z_t^M = (D^M)^{-1} \sum_{k=1}^{K_M} A_k^M X_t W_k^M$	K 是子集, 含全局和局部特征 — i 是类型图, G 是亲密度矩阵 D 是度矩阵, K 是子集, M 捕捉高阶信息
权值调整	GGCN ^[42]	SemGCN ^[50]	—	—
信息融合	Bayesian GC-LSTM ^[43]	DGNN ^[53] ARG ^[54] GPNN ^[55] SGN ^[56]	$Z^{(l+1)} = \sum_{i=1}^{N_g} \sigma(G^i Z^{(l)} W^{(l,i)})$ $Y_t = G_t Z_t W_y$	— N_g 是子图, G 计算特征关系 — G 是亲密度矩阵, Z 语义特征

(二) 关注感受野。为了提升感受野, 更好的挖掘图卷积网络的高阶信息, 同时又为了加深网络深度, 避免图中节点信息过于平滑, 有些研究分别从细化卷积方式^[60–61]、改变网络体系结构^[51–52,62]、丰富输入网络的特征信息^[49]等方面展开。在行为识别中, Wang 等^[51]通过构建不同的图结构, 加强物体间的相关关系, 关注感知域的提升。Wen 等提出可编码空间分层结构, 利用可变的时间密集块捕获不同范围骨骼序列的局部时间高阶信息^[52]。Non-local 模块^[37,50]可以提高网络的感受野。但是, 文中构造的图节点只有一种信息。为此, AGC-LSTM^[49]提出时序层次结构, 提高了时间接受域和高层语义表示的学习能力, 显著降低了网络计算量。对于感受野的提升只能应用在小规模图结构上, 对于大图, 易造成网络负担影响性能。所以, 感受野的“度”需要适当把握。

(三) 注意力机制应用。原始的 GCN 无法捕捉哪些节点是关键的邻节点, 不知道节点对中心节点的分类贡献大小。有些研究提出了辨别邻居节点贡献能力的注意力机制^[63–64], 并取得了不错的实验效果。RAGCN 对每一个邻接矩阵, 学习新的矩阵表示边的重要性, 间接获取注意力。双向注意图卷积网络 BAGCN 在聚焦过程中, 引入注意模块学习帧的隐式依赖, 增强节点上下文信息的传递^[48]。虽

然注意力机制使神经网络易于通过反向传播进行有效学习, 同时也导致了二次计算代价。对于位置敏感的模型, 易出现决策困难, 框架不可微, 难以优化等问题。所以, 注意力机制需要在注意力类型选择、计算效率、可微性等方面进行参数权衡。

(四) 权值共享的研究。在之前介绍的图卷积中, 为了使图卷积网络应用在任意的拓扑结构上, 核函数 W 对关系图中的边权值共享, 这样导致邻边及其图内部结构不能充分利用。语义图卷积 SemGCN 解决了这一问题, 通过动态更新邻接矩阵的值间接调整 W 不再共享。此模型为今后的研究提供了思路, 例如, 如何整合视频的时态信息等。针对此类问题, 应合理设置权重分配, 考虑不同标签策略, 也可为节点学习独立权重; 关于邻接矩阵的设计, 既可以让网络自学习也可以根据先验知识调整。

(五) 应用拓展。GCN 除应用在单人行行为识别领域之外, 也被应用到群组行为识别领域和人-物交互领域。群组行为识别中, 根据个体的位置和外貌特征, 以每个个体为节点, 以个体之间的关系为边建立多个关系图。图卷积网络 ARG 对关系图进行关系推理, 将各个关系图的推理结果融合在一起, 生成所有参与者个体的关系表示, 分别进行个体行为识别和群组行为识别。图卷积网络在群组行为识别中的应用, 为群组行为识别开辟了新的研究道

路。图解析神经网络 GPNN, 用于检测和识别图像或视频中人物-物的交互^[55], 是一种端到端可微的结构知识集成框架。实验验证了模型在大型数据集上的可伸缩性和对时空设置的适用性。

3 基于图卷积的行为识别网络设计

基于图卷积的行为识别网络模型的设计, 如图 2 所示。图中“BN”代表正则化, “ATT”是注意力机制, “TCN”代表时域图卷积, “POOL”是池化操作, “FC”是全连接。残差连接下面的块可视为一个图卷积模型块。通过观察图 2, 不同的图卷积行为识别模型设计的区别主要在于构造图拓扑结构和 GCN 两大部分。

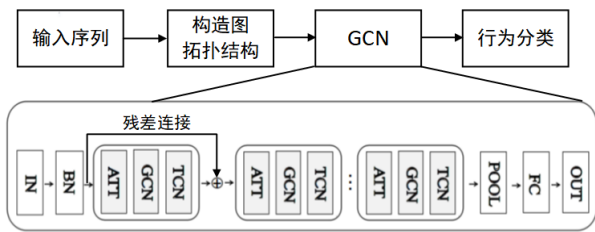


图 2 基于图卷积的行为识别基本模型

3.1 构造图拓扑结构的变化

谱方法的构图方式已在 2.1 节介绍, 本节重点介绍基于空间的图卷积在行为识别中的构图。按照输入特征分, 图构造分为基于点的构图和基于面的构图。按照图的完整性分, 图构造分为部分图、拆解图。按照图的方向性分, 分为有向图和无向图。按照图的动态性分, 分为固定图和自适应图。

3.1.1 按照输入特征区分

基于点的构图主要是骨架图^[31,44-49,52-53]。骨架图的构造建立在空间域和时域上。如图 3 所示。空间域是帧内图构造, 图中的实心圆点是关节点, 特征表示为坐标, 关节点之间的连边代表骨骼。时域是帧间构造, 是相同关节点在不同的时间点的连接。基于面的构图是对于交互^[51,55]和群组行为^[54]识别。在空间域, 一个标定框形成的面构成一点, 构图输入除了点的坐标还有一些外观特征、运动特征等。

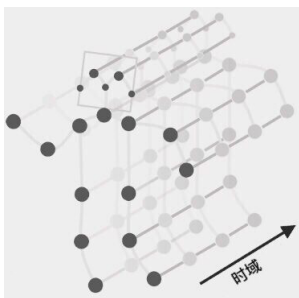


图 3 骨架时空图

3.1.2 按照图的完整性区分

部分图是将人体的完整骨架图划分为多个子图, 每一部分分别卷积再进行特征融合^[44]。这种划分方式能更好的抽取局部关节信息, 同时加快信息传播速度。但是, 部分图划分数量不易控制, 数量过少, 关节之间的差异性容易平滑, 数量过多, 非边缘节点间的信息传播困难。拆解图是将骨架图拆分为点图和边图^[45-46], 将骨骼流独立于关节流进行训练。如图 4(a) 所示。预测时, 两个分支融合, 得到最终行为结果。关节点数据和骨骼数据的拆解增强了数据, 提高了网络对空间特征的抽取能力。但构图方式比较复杂。

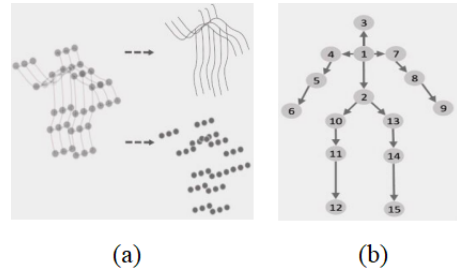


图 4 骨架拆解图和有向无环图

3.1.3 按照图的方向性区分

骨架无向图简单、易操作, 但不能将骨头和关节点之间的依赖关系完全探索出来。骨骼信息包含人体骨头的方向和长度, 它和关节点坐标信息紧密相连。DGNN 以关节点为节点、骨头为边设计了一个有向无环图网络, 增强了关节点和骨头信息间依赖关系的提取, 进一步提高了抽取空间特征的细粒度, 促进了行为识别, 如图 4(b) 所示。但参数数量相对较多。

3.1.4 按照图的动态性区分

仅使用人体的物理连接构造的固定结构图使得模型在行为识别中应用狭窄, 只能识别挥手、跑步等简单的动作, 不能充分利用骨架数据的图形结构, 难以推广到具有任意形式的骨架。对于超越物理连接的动作, 比如拥抱、拍手、下蹲等, 需要根据样本数据自适应的调整图的拓扑结构^[45-46], 这种数据驱动的方法提高了模型的灵活性, 为适应各种数据样本带来了更广泛的通用性。

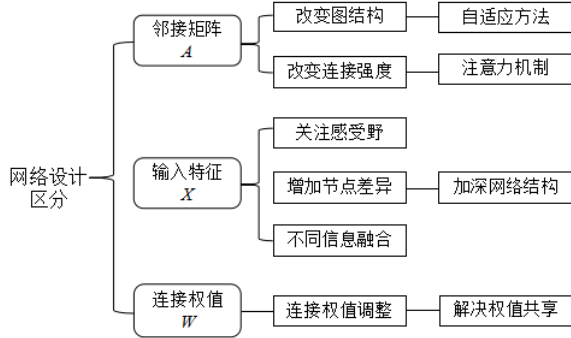
3.2 不同的 GCN 模型

假设图卷积网络只有一层, 则图卷积定义如下:

$$X^{(1)} = \sigma(\hat{A}X^{(0)}W). \quad (1)$$

从 (1) 式可以看到, 等式右侧的激活函数内主要涉及邻接矩阵 A 、输入特征 X 和连接权值 W 三部分, 所以, 图卷积网络的不同设计主要体现在这三个方

面。行为识别中,图卷积网络设计的变化及解决的问题如图5所示。三大部分对应的图卷积行为识别的方法及介绍见表3和第2部分。



3.3 目标函数的设计

在行为识别中,基于空间的图卷积的目标函数主要是在 ST-GCN 的基础上进行设计与改进。目标函数的改变正是针对图5中提到的邻接矩阵 A 、输入特征 X 和连接权值 W 三大方面,以此来改变图卷积的网络结构。空间方法的目标函数见表3第四列,等号左侧是经过图卷积后的输出。基于频谱图卷积的行为识别^[40-43]目标函数主要是切比雪夫多项式,表达式见式(2)。

$$g_{\theta} * x = \sum_{k=0}^{K-1} \theta_k T_k(L)x. \quad (2)$$

式中, $*$ 代表卷积运算, g_{θ} 和 θ_k 是需要学习的卷积核和参数, L 是归一化的拉普拉斯矩阵, $T_k(L)$ 是 k 阶切比雪夫多项式。式(3)是切比雪夫一阶近似, D

是度矩阵表示与图中节点的连接数量, H^l 表示第 l 层的输入。

$$H^{(l+1)} = \sigma \left(\hat{D}^{-1/2} \hat{A} \hat{D}^{-1/2} H^{(l)} W^{(l)} \right). \quad (3)$$

4 基于图卷积的行为识别算法分析

4.1 数据集及评价指标

用于图卷积网络的行为识别的数据集主要有 Kinetics、NTU-RGB+D、HDM05、Northwestern-UCLA、UT-Kinect、MSR Action3D、UTD MHAD、SYSU 及 Florence 3D, 汇总见表4。

在行为识别中,对算法的评价指标一般是行为识别的准确率,定义如式(4)。式中, N_{corr} 和 N_{total} 分别代表分类正确的数量和总分类数量。在 Kinetics 数据集中,用 top-1 和 top-5 准确率进行表示,其它数据集都是用最高准确率表示。各种算法的识别率见表5所示。

4.2 算法比较与分析

ST-GCN 在两个大规模的数据集上减少人工设计,取得了优越的性能,成为其他行为识别方法的基准。由于 ST-GCN 存在的问题(见2.2部分),AS-GCN 提出 Action 和 Structure 两种结构,找到行为的隐藏关节点,利用了高阶特征;2s-AGCN 加入物理连接之外的新连接形成自适应结构;DGNN 提出有向图,较 Kinetics 数据集的 top-1 分别提高 3.1、5.4 和 6.2 个百分点。

$$f = \frac{N_{corr}}{N_{total}}. \quad (4)$$

表4 数据集汇总表

数据集	视频数	行为类	受试人数	关节点数	训练集	测试集
Kinetics	30 万	400	400	18	24 万	2 万
NTU-RGB+D	56880	60	40	25	CS:40320 CV:37920	CS:16560 CV:18960
HDM05	2337	130	5	31	一半数据	一半数据
Northwestern-UCLA	1494	10	10	25	前两台摄像机	后一台摄像机
UT-Kinect	200	10	10	20	9 名受试者	1 名受试者
MSR Action3D	557	20	10	18	奇数受试者	偶数受试者
UTD MHAD	861	27	8	20	奇数受试者	偶数受试者
SYSU	480	12	40	18	奇数受试者(或一半数据)	偶数受试者(或一半数据)
Florence 3D	215	9	10	15	9 名受试者	1 名受试者

NTU-RGB+D 数据集是普遍应用的数据集,ST-GCN 在 CS 和 CV 两个基准上的识别率分别为 81.5% 和 88.3%。后续的算法分别在图结构拆解,例如 PB-GCN; 使用注意力机制,例如 RAGCN、BAGCN; 加入自适应结构,例如 AS-GCN, 2S-AGCN; 提高模型感受野,例如 DPRL; 研究有向图等方面展开深入研究,使得识别性能不断提升。目前

为止,在 NTU-RGB+D 数据集上最好的识别率是 BAGCN, 分别为 90.3% 和 96.3%。其他数据集应用较少,贝叶斯 GC-LSTM 由于贝叶斯的应用,在 MSR Action3D、UTD MHAD、SYSU 上都取得了不错的识别性能。Florence 3D 在 GGCN 网络体系中,由于广义图在时域上体现的是时空图的分割,在空域表达了节点间的强弱连接,识别性能较优。

表 5 基于图卷积的行为识别方法的准确率比较

图卷积方法	Kinetics		NTU-RGB+D		HDM05	Northwestern-UCLA	UT- Kinect	MSR Action3D	UTD HAD	SYSU	Florence 3D
	Top-1	Top-2	CS	CV	准确识别率 (%)						
ST-GCN[31]	30.7	52.8	81.5	88.3	—	—	—	—	—	—	—
DPRL ^[40]	—	—	83.5	89.8	—	—	98.5	—	—	76.9	—
NAS ^[41]	37.1	60.1	89.4	95.7	—	—	—	—	—	—	—
GGCN ^[42]	—	—	87.5	94.3	—	—	98.5	—	—	77.9	98.4
Bayesian	—	—	—	—	—	—	—	—	—	—	—
GC-LSTM ^[43]	—	—	81.8	89.0	—	—	—	94.5	92.1	82.0	—
PB-GCN ^[44]	—	—	87.5	93.2	88.17±0.99	—	—	—	—	—	—
2S-AGCN ^[45]	36.1	58.7	88.5	95.1	—	—	—	—	—	—	—
AS-GCN ^[46]	34.8	56.5	86.8	94.2	—	—	—	—	—	—	—
RAGCN ^[47]	—	—	85.9	93.5	—	—	—	—	—	—	—
BAGCN ^[48]	37.3	60.2	90.3	96.3	—	—	—	—	—	—	—
AGC-LSTM ^[49]	—	—	89.2	95.0	—	93.3	—	—	—	—	—
Wen 等 ^[52]	84.2	96.1	84.2	90.2	—	—	—	—	—	—	—
DGNN ^[53]	36.9	59.6	89.9	96.1	—	—	—	—	—	—	—
SGN ^[56]	—	—	86.6	93.4	—	—	—	—	—	86.9	—

5 行为识别中 GCN 的问题与挑战

图神经网络的应用, 虽然提高了识别性能, 但仍有许多问题需要解决。

(一) 图卷积网络为浅层结构。传统的基于 CNN 的深度学习模型往往是通过堆叠卷积层数来获得精细的信息特征, 从而获得更少的网络参数和更多的非线性变换, 以提高网络的识别性能^[65]。但是, 在图神经网络的卷积层的设置中, 实验结果表明, 如果网络层数大于两层, 网络的性能反而下降^[11]。原因是, 如果堆叠的层数过多过深, 会造成图中的每一个节点过度平滑, 无法区分差异性, 从而不利于识别性能的提高^[66]。因此, 图卷积网络的卷积层数往往只有两到三层, 无法加深网络结构。

(二) 网络结构不稳定。为了使图卷积网络应用在任意的拓扑结构上, 核函数 W 对关系图中的所有边都是权值共享的, 这和传统的卷积神经网络 CNN 相同。这样导致邻边及其整个图的内部结构都没有得到充分利用。有些研究通过增加掩膜与邻接矩阵进行矩阵的乘法来间接调整点与点之间的权重^[31,45,50], 使得网络识别性能进一步提升。但是, 当图结构发生变化时, 如果各个节点分配的权重不同, 这些不同的权重就会失去效用, 导致图网络结构不稳定。

(三) 图的拓扑结构自适应能力不强。在人体运动的过程中, 行为是不断发生变化的, 随着动作的改变, 图的拓扑结构也随之发生改变, 节点与节点之间的关系以及关键节点的选择都会随之发生变化。当新加入节点之间的连接或者节点之间的关联消失的时候, 图卷积网络的自适应能力不强。虽然也做了一些自适应的研究^[45-46], 但都是基于规则, 不能动态生成。每次卷积后需要动态调整边的权

重, 但因参数过多, 容易引起过拟合。

(四) 非结构场景数据利用不充分。行为识别过程中, 在图结构数据融合的过程中, 一般是根据骨骼点的 2D、3D 坐标或者外貌特征进行融合, 融合的数据结构一般是矩阵的结构形式, 而对于外部场景、人体之外的物体、交互等非结构化数据信息没有充分利用, 除了骨架图^[31,44-49,52-53]、群组中的关系图^[54]之外, 对利用非结构化数据生成的图在行为识别应用中研究的少。

6 对 GCN 在行为识别中的展望

(一) 卷积池化相结合, 体现网络层级结构。关于增加图卷积网络深度的问题, 本文认为可以在图卷积网络中加入池化层, 通过池化增大节点间差异性的几率。提升节点之间的稀疏性, 就可以逐渐增加网络深度, 逐渐由中心节点附近的点扩展到离它较远的点, 直到图模型的边缘节点, 体现出图卷积网络的层次性。

(二) 挖掘节点新特征, 建立多样类型图。在将来的工作中, 可以充分挖掘 GCN 的灵活性来建立更多类型的图, 分析非结构数据的图。针对动作识别任务, 在骨架图各个节点坐标、外貌等特征的基础上引入更多的视觉特征, 如交互特征、场景图、RGB 特征等, 并将它们统一在图卷积网络的框架下。以此, 建立一个性能好, 可解释性强的行为识别模型。

(三) 考虑动态图模型的建立。目前的图卷积神经网络一经训练都是固定的模式, 图结构一旦发生变化, 训练出来的参数可能就会失效, 自适应能力以及延展性不强, 本文认为设计动态变化的图卷积神经网络是未来的趋势。

(四) 在构建图神经网络时所有节点共享同样的

模型参数并且拥有同样维度的特征空间,目前构造的图都是同构的。异构图中可以存在不同类型的节点和边,使其拥有不同维度的特征或属性。本文认为可以对异构图模型展开研究,冲破权值共享限制,获得更加鲁棒的图模型。

7 结束语

图卷积网络对非欧氏空间的图数据处理非常有效,受到越来越多的关注。图卷积神经网络的核心思想主要是通过深度学习方法,将图结构中的节点或边映射到向量空间,然后进行聚类、分类等处理。本文主要总结了近年来图卷积网络在行为识别领域的应用与发展。本文在第一部分重点介绍了行为识别、图卷积网络的两种方法,并把两种方法进行了对比。第二部分总结了图卷积网络在行为识别领域的应用与发展,并分析了不同网络模型的优缺点。第三部分总结了基于图卷积的行为识别的网络设计的不同方面,并比较了不同网络模型的目标函数。第四部分介绍了基于图卷积的行为识别的数据集和评价指标,并对多种算法进行比较与分析。第五、六部分分析了图卷积网络在行为识别领域面临的问题与挑战,讨论了图卷积网络在行为识别中未来的发展方向。

图卷积网络在复杂图结构上的建模能力确实令人惊叹,在行为识别领域也取得了很好的识别性能。鉴于其有效性,相信在不久的将来,图卷积网络将在深度学习的发展中发挥重要作用。

参考文献 (References)

- [1] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based Learning Applied to Document Recognition[J]. *Proceedings of the IEEE*, 1998, 86(11): 2278-2324.
- [2] 张顺, 龚怡宏, 王进军. 深度卷积神经网络的发展及其在计算机视觉领域的应用 [J]. *计算机学报*, 2019, 42(3): 453-482.
(Zhang S, Gong Y H, Wang J J. The development of deep convolution neural network and its applications on computer vision[J]. *Chinese Journal of Computers*, 2019, 42(3): 453-482.)
- [3] Xu D F, Zhu Y K, Christopher B C, et al. Scene graph generation by iterative message passing[C]. *Proc of the 30th IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, USA: IEEE, 2017: 3097-3106.
- [4] Du Y, Fu Y, Wang L. Skeleton based action recognition with convolutional neural network[C]. *Proc of the 3rd IAPR Asian Conference on Pattern Recognition*. Kuala Lumpur, Malaysia: IEEE, 2015: 579-583.
- [5] Zhang Z W, Cui P, Zhu W W. Deep Learning on Graphs: A Survey. *arXiv preprint arXiv: 1812.04202*, 2018.
- [6] Wu Z H, Pan S R, Chen F W, et al. A Comprehensive Survey on Graph Neural Networks. *arXiv preprint arXiv: 1901.00596*, 2019.
- [7] Joan B, Wojciech Z, Arthur S, et al. Spectral networks and deep locally connected networks on graphs[C]. *Proc of the 2nd International Conference on Learning Representations*. Banff, Canada: ICLR, 2014: 1-14.
- [8] Michaël D, Xavier B, Pierre V. Convolutional neural networks on graphs with fast localized spectral filtering[C]. *Proc of the 30th Annual Conference on Neural Information Processing Systems*. Barcelona, Spain: NIPS, 2016: 3844-3852.
- [9] David K H, Pierre V, R'emi G. Wavelets on graphs via spectral graph theory[J]. *Applied and Computational Harmonic Analysis*. 2011, 30(2): 129-150.
- [10] Ron L, Federico M, Xavier B, et al. Cayleynets: Graph convolutional neural networks with complex rational spectral filters[J]. *IEEE Transactions on Signal Processing*, 2019, 67(1): 97-109.
- [11] Thomas N, Max W. Semi-supervised classification with graph convolutional networks[C]. *Proc of the 5th International Conference on Learning Representations*. Toulon, France: ICLR, 2017: 1-14.
- [12] David G. Distinctive Image Features from Scale-Invariant Keypoints[J]. *International Journal of Computer Vision*, 2004, 60(2): 91-110.
- [13] Dalal N, Triggs B. Histograms of Oriented Gradients for Human Detection[C]. *Proc of the IEEE Conference on Computer Vision and Pattern Recognition*. San Diego, USA: IEEE, 2005: 886-893.
- [14] Herbert B, Andreas E, Tinne T, et al. Speeded-Up Robust Features (SURF)[J]. *Computer Vision and Image Understanding*, 2008, 110(3): 346-359.
- [15] Wang H and Cordelia S. Action recognition with improved trajectories[C]. *Proc of the 14th IEEE International Conference on Computer Vision*. Sydney, Australia: IEEE, 2013: 3551-3558.
- [16] Heng W, Alexander K, Cordelia S, et al. Dense Trajectories and Motion Boundary Descriptors for Action Recognition[J]. *International Journal of Computer Vision*, 2013, 103(1): 60-79.
- [17] Karen S, Andrew Z. Two-Stream Convolutional Networks for Action Recognition in Videos[C]. *Proc of the 28th Annual Conference on Neural Information Processing Systems*. Montreal, Canada: NIPS, 2014: 568-576.
- [18] Christoph F, Axel P, Andrew Z. Convolutional Two-Stream Network Fusion for Video Action Recognition[C]. *Proc of the 29th IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, USA: IEEE Computer Society, 2016: 1933-1941.
- [19] Du T, Lubomir B, Rob F, et al. Learning Spatiotemporal Features with 3D Convolutional Networks[C]. *Proc of the 15th IEEE International Conference on Computer Vision*. Santiago, Chile: IEEE, 2015: 4489-4497.
- [20] Jeff D, Lisa A H, Marcus R, et al. Long-term Recurrent

- Convolutional Networks for Visual Recognition and Description[J]. *Potential Analysis*, 2019, 49(2): 225-245.
- [21] David I S, Sunil K, Pascal F, et al. Signal Processing on Graphs: Extending High-Dimensional Data Analysis to Networks and Other Irregular Domains. *arXiv preprint arXiv: 1211.0053*, 2012.
- [22] Micheli A. Neural network for graphs: A contextual constructive approach[J]. *IEEE Transactions on Neural Networks*, 2009, 20(3): 498-511.
- [23] Justin G, Samuel S S, Patrick F R, et al. Neural message passing for quantum chemistry[C]. *Proc of the 34th International Conference on Machine Learning*. Sydney, Australia: IMLS, 2017: 2053-2070.
- [24] James A, Don T. Diffusion-Convolutional Neural Networks[C]. *Proc of the 30th Annual Conference on Neural Information Processing Systems*. Barcelona, Spain: NIPS, 2016: 2001-2009.
- [25] Federico M, Davide B, Jonathan M, et al. Geometric deep learning on graphs and manifolds using mixture model cnns[C]. *Proc of the 30th IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, USA: IEEE, 2017: 5425-5434.
- [26] William L, Rex Y, Jure L. Inductive Representation Learning on Large Graphs[C]. *Proc of the 31st Annual Conference on Neural Information Processing Systems*. Long Beach, USA: NIPS, 2017: 1025-1035.
- [27] Chen J, Ma T F, Xiao C. FastGCN: Fast learning with graph convolutional networks via importance sampling[C]. *Proc of the 6th International Conference on Learning Representations*. Vancouver, Canada: ICLR, 2018: 1-15.
- [28] Huang W B, Zhang T, Rong Y, et al. Adaptive Sampling Towards Fast Graph Representation Learning[C]. *Proc of the 32nd Conference on Neural Information Processing Systems*. Montreal, Canada: NIPS, 2018: 4558-4567.
- [29] Chiang W L, Liu X Q, Si S, et al. Cluster-GCN: An Efficient Algorithm for Training Deep and Large Graph Convolutional Networks[C]. *Proc the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Anchorage, USA: ACM, 2019: 257-266.
- [30] Mathias N, Mohamed A, Konstantin K. Learning Convolutional Neural Networks for Graphs[C]. *Proc of the 33rd International Conference on Machine Learning*. New York, USA: IMLS, 2016: 2958-2967.
- [31] Yan S J, Xiong Y J, Lin D H, et al. Spatial temporal graph convolutional networks for skeleton-based action recognition[C]. *Proc of the 32nd AAAI Conference on Artificial Intelligence*. New Orleans, USA: AAAI press, 2018: 7444-7452.
- [32] Petar V, Guillem C, Arantxa C, et al. Graph Attention Networks[J]. *Lecture Notes in Computer Science*, 2019, 11731(2005): 566-577.
- [33] Zhang J N, Shi X J, Xie J Y, et al. GaAN: Gated attention networks for learning on large and spatiotemporal graphs[C]. *Proc of 34th Conference on Uncertainty in Artificial Intelligence*. Monterey, USA: AUAI, 2018: 339-349.
- [34] Gao H Y, Ji S W. Graph Representation Learning via Hard and Channel-Wise Attention Networks[C]. *Proc of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Anchorage, USA: ACM, 2019: 741-749.
- [35] Yang Y D, Wang X C, Song M L, et al. SPAGAN: Shortest Path Graph Attention Network[C]. *Proc of the 28th International Joint Conference on Artificial Intelligence*. Macao, China: IJCAI, 2019: 4099-4015.
- [36] Gong L Y, CHeng Q. Adaptive edge features guided graph attention networks[C]. *Proc of the 32nd IEEE Conference on Computer Vision and Pattern Recognition*. Long Beach, USA: IEEE, 2019: 9203-9211.
- [37] Wang X L, Ross G, Abhinav G, et al. Non-local neural networks. *arXiv preprint arXiv: 1711.07971*, 2017.
- [38] Chen J F, Zhu J, Song L. Stochastic Training of Graph Convolutional Networks with Variance Reduction[C]. *Proc of the 35th International Conference on Machine Learning*. Stockholm, Sweden: IMLS, 2018: 1503-1532.
- [39] Li C L, Cui Z, Zheng W M, et al. Spatio-Temporal Graph Convolution for Skeleton Based Action Recognition[C]. *Proc of the 32nd AAAI Conference on Artificial Intelligence*. New Orleans, USA: AAAI press, 2018: 3482-3489.
- [40] Tang Y S, Tian Y, Lu J W, et al. Deep Progressive Reinforcement Learning for Skeleton-Based Action Recognition[C]. *Proc of the 31st IEEE Conference on Computer Vision and Pattern Recognition*. Salt Lake City, USA: IEEE Computer Society, 2018: 5323-5332.
- [41] Peng W, Hong X P, Chen H Y, et al. Learning Graph Convolutional Network for Skeleton-based Human Action Recognition by Neural Searching. *arXiv preprint arXiv: 1911.04131*, 2019.
- [42] Gao X, Hu W, Tang J X, et al. Generalized Graph Convolutional Networks for Skeleton-based Action Recognition. *arXiv preprint arXiv: 1811.12013*, 2018.
- [43] Zhao R, Wang K, Su H, et al. Bayesian Graph Convolution LSTM for Skeleton Based Action Recognition[C]. *Proc of the 17th IEEE International Conference on Computer Vision*. Seoul, Korea: IEEE, 2019: 6881-6891.
- [44] Kalpit T, P.J. N. Part-based Graph Convolutional Network for Action Recognition[C]. *Proc of the 29th British Machine Vision Conference*. Newcastle, Britain: BMVA Press, 2019.
- [45] Shi L, Zhang Y F, Cheng J, et al. Two-stream adaptive graph convolutional networks for skeleton-based action recognition[C]. *Proc of the 32th IEEE Conference on Computer Vision and Pattern Recognition*. Long Beach, USA: IEEE Computer Society, 2019: 12018-12027.
- [46] Li M S, Chen S H, Xu C, et al. Actional-Structural Graph Convolutional Networks for Skeleton-based Action

- Recognition[C]. Proc of the 32nd IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE Computer Society, 2019: 3590-3598.
- [47] Song Y F, Zhang Z, Wang L. Richly Activated Graph Convolutional Network For Action Recognition With Incomplete Skeletons[C]. Proc of the 26th IEEE International Conference on Image Processing. Taipei, Taiwan: IEEE Computer Society, 2019: 1-5.
- [48] Gao J L, He T, Zhou X, et al. Focusing and Diffusion: Bidirectional Attentive Graph Convolutional Networks for Skeleton-based Action Recognition. arXiv preprint arXiv: 1912.11521, 2019.
- [49] Si C Y, Chen W T, Wang W, et al. An Attention Enhanced Graph Convolutional LSTM Network for Skeleton-Based Action Recognition[C]. Proc of the 32nd IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE, 2019: 1227-1236.
- [50] Zhao L, Peng X, Tian Y, et al. Semantic Graph Convolutional Networks for 3D Human Pose Regression[C]. Proc of the 32nd IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE Computer Society, 2019: 3420-3430.
- [51] Wang X L, Abhinav G. Videos as space-time region graphs[C]. Proc of the 15th European Conference on Computer Vision. Munich, Germany: Springer Verlag, 2018: 413-431.
- [52] Wen Y H, Gao L, Fu H B, et al. Graph CNNs with Motif and Variable Temporal Block for Skeleton-based Action Recognition[C]. Proc of the 9th AAAI Symposium on Educational Advances in Artificial Intelligence. Honolulu, USA: AAAI Press, 2019: 8989-8996.
- [53] Shi L, Zhang Y F, Cheng J, et al. Skeleton-Based Action Recognition with Directed Graph Neural Networks[C]. Proc of the 32nd IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE Computer Society, 2019: 7904-7913.
- [54] Wu J C, Wang L M, Wang L, et al. Learning Actor Relation Graphs for Group Activity Recognition[C]. Proc of the 32nd IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE Computer Society, 2019: 9956-9966.
- [55] Qi S Y, Wang W G, Jia B X, et al. Learning Human-Object Interactions by Graph Parsing Neural Networks[C]. Proc of the 15th European Conference on Computer Vision. Munich, Germany: Springer Verlag, 2018: 407-423.
- [56] Zhang P F, Lan C L, Zeng W J, et al. Semantics-Guided Neural Networks for Efficient Skeleton-Based Human Action Recognition. arXiv preprint arXiv: 1904.01189, 2019.
- [57] Jiang G B, Lin D D. Graph Laplacian Regularized Graph Convolutional Networks for Semi-supervised Learning. arXiv preprint arXiv: 1809.09839, 2018.
- [58] Li R Y, Wang S, Zhu F Y, et al. Adaptive Graph Convolutional Neural Networks[C]. Proc of the 32nd AAAI Conference on Artificial Intelligence. New Orleans, USA: AAAI press, 2018: 3546-3553.
- [59] Johannes K, Aleksandar B, Stephan G. Predict then Propagate: Graph Neural Networks meet Personalized PageRank[C]. Proc of the 7th International Conference on Learning Representations. New Orleans, USA: ICLR, 2019: 1-15.
- [60] Wang Y W, Hu Z N, Ye Y S, et al. Demystifying Graph Neural Network Via Graph Filter Assessment[C]. Proc of the International Conference on Learning Representations. Addis Ababa, ETHIOPIA: ICLR, 2020: 1-15.
- [61] Sami A, Amol K, Bryan P, et al. N-GCN: Multi-scale Graph Convolution for Semi-supervised Node Classification[C]. Proc of the 35th Conference on Uncertainty in Artificial Intelligence. Tel Aviv, Israel: AUAI, 2019: 1-9.
- [62] Hu F Y, Zhu Y Q, Wu S, et al. Hierarchical Graph Convolutional Networks for Semi-supervised Node Classification[C]. Proc of the 28th International Joint Conference on Artificial Intelligence. Macao, China: IJCAI, 2019: 4532-4539.
- [63] Kiran K, Chong W, Sewoong O, et al. Attention-based graph neural network for semi-supervised learning. arXiv preprint arXiv: 1803.03735, 2018.
- [64] Chen F W, Pan S R, Jiang J, et al. DAGCN: Dual Attention Graph Convolutional Networks[C]. Proc of the 2019 International Joint Conference on Neural Networks. Budapest, Hungary: IEEE, 2019: 1-8.
- [65] Li C, Zhong Q Y, Xie D, et al. Skeleton based action recognition with convolutional neural network[C]. Proc of the 3rd IAPR Asian Conference on Pattern Recognition. Kuala Lumpur, Malaysia: IEEE, 2015: 579-583.
- [66] Li Q M, Han Z C, Wu X M. Deeper insights into graph convolutional networks for semi-supervised learning[C]. Proc of the 32nd AAAI Conference on Artificial Intelligence. New Orleans, USA: AAAI press, 2018: 3538-3545.

作者简介

孔玮(1986—), 女, 博士研究生, 从事计算机视觉的研究, E-mail: kongwei1902@126.com;

刘云(1962—), 男, 教授, 博士(博士生导师), 从事计算机视觉等研究, E-mail: lyun-1027@163.com.

李辉(1984—), 男, 副教授, 博士, 从事计算机视觉的研究, E-mail: lihui@qust.edu.cn.

王传旭(1968—), 男, 教授, 博士, 从事计算机视觉的研究, E-mail: wangchuanxu_qd@163.com.