

中图法分类号: TP399 文献标识码: A 文章编号: 1006-8961(2025)06-2275-29

论文引用格式: Li S, Wang H Y and Hu R M. 2025. A review of public social behavior understanding and hidden groups discovery. Journal of Image and Graphics, 30(6):2275-2303(李姝, 王怀远, 胡瑞敏. 2025. 公共场所社交行为理解研究综述与隐密社团发现. 中国图象图形学报, 30(6):2275-2303)[DOI:10.11834/jig.240784]

# 公共场所社交行为理解研究综述与隐密社团发现

李姝, 王怀远, 胡瑞敏\*

西安电子科技大学网络与信息安全学院, 西安 710000

**摘要:** 公共场所中的社交群体检测旨在分析监控视频数据, 采用社交互动、时空位置关系或计算机视觉等技术手段去表达人类的社交行为特征, 进而识别交互群体。挖掘人类的社交行为模式并识别正在交互的行人群组对于轨迹预测、群体异常活动识别以及人机交互等研究领域意义重大, 已成为计算机视觉领域的热门研究课题。尽管取得了一些进展, 群体交互现象仍然缺乏正式的规则和精确的社会解释。且由于数据采集困难, 社交群体检测面临着轻量化网络设计与小样本学习困难。本文全面回顾现有的社交行为理解与群体检测工作: 首先依据不同的社交行为建模角度, 将公共场所中的群体检测方法分为基于启发式规则与基于学习的方法两大类, 重点介绍基于空间、视觉内容和行为模式的方法与主流深度学习框架; 其次归纳总结评价指标、数据集与检测性能; 最后讨论当前研究面临的挑战和局限性以及该领域未来可能的研究方向。通过对主要期刊/会议论文的调查分析, 发现图神经网络与Transformer等的融合模型仍然是主要发展趋势, 但最近的基于行为模式的方法取得了最优的检测性能。因此, 如何挖掘人类的本质社交模式将极具发展潜力, 同时还应探索此前未被关注的隐式交互模式挖掘: 隐秘犯罪团体发现。

**关键词:** 社交群体检测; 社交行为理解; 交互检测; F-formation; 行为模式

## A review of public social behavior understanding and hidden groups discovery

Li Shu, Wang Huaiyuan, Hu Ruimin\*

School of Cyber Engineering, Xidian University, Xi'an 710000, China

**Abstract:** Social group detection in public spaces has become a popular research topic in computer vision and social behavior analysis in recent years. This research aims to analyze surveillance video data, utilizing technologies such as social interaction, spatiotemporal relationships, and computer vision to identify and understand human social behavior patterns, thereby detecting pedestrian groups that are actively interacting. With the advancement of urbanization, the demand for public safety and social behavior monitoring has increased significantly. Social group detection is crucial in various research fields, including trajectory prediction, group anomaly detection, human-computer interaction, and intelligent surveillance. Exploring and identifying human social behavior patterns, especially recognizing pedestrian groups in interaction, provides strong support for applications such as predicting pedestrian behavior, detecting abnormal group activities, and enhancing public safety. This, in turn, propels further developments in computer vision and artificial intelligence. Despite some progress in this field, accurately modeling group interaction phenomena remains challenging. First, social

收稿日期: 2024-12-29; 修回日期: 2025-02-18; 预印本日期: 2025-02-25

\* 通信作者: 胡瑞敏 rmhu@xidian.edu.cn

**基金项目:** 国家自然科学基金项目(U22A2035); 广西自然科学基金项目(2021GXNSFDA075011); 广西壮族自治区产业技术研究院产研计划项目(CYY-HT2023-JSJJ-0021)

**Supported by:** National Natural Science Foundation of China (U22A2035); Guangxi Natural Science Foundation Program, China (2021GXNSFDA075011); Projects of the Industrial Research Plan of Guangxi Institute of Industrial Technology (CYY-HT2023-JSJJ-0021)

group behaviors are diverse and complex, lacking formal rules and clear social interpretations, making social group detection a difficult task. Second, most existing methods rely on high-quality annotated data, which poses significant costs and technical challenges, especially in real-world scenarios where the number and behavior patterns of social groups are constantly changing. As a result, existing datasets often fail to cover all possible scenarios. Furthermore, social group detection faces challenges related to lightweight network design and few-shot learning. Particularly in the absence of large-scale annotated data, how to effectively train models and improve detection accuracy remains a key issue in current research. This paper reviews the existing research on social behavior understanding and social group detection. It categorizes methods for social group detection in public spaces into two main categories: heuristic-rule-based methods and learning-based methods. In heuristic-rule-based methods, researchers define a series of rules to determine the interaction features of pedestrian groups, such as using distance, speed change, relative direction, and visual descriptors to define the occurrence of social behavior, or extracting high-level semantic expressions of human interaction patterns, such as social force models or social interaction fields. However, the applicability of these methods is often limited by the accuracy and complexity of the rules. Conversely, learning-based methods leverage machine learning or deep learning techniques to automatically learn the patterns of social behaviors from large datasets, thereby achieving higher detection accuracy. In recent years, deep learning methods, particularly convolutional, recurrent, and graph neural networks, have been widely applied in social group detection and have achieved significant results. In terms of evaluation, this paper summarizes common evaluation metrics, datasets, and detection performance. Common evaluation metrics include accuracy, recall, and F1 score, which effectively reflect the performance of social group detection algorithms. For example, when using mainstream deep learning frameworks for group detection, the average precision and recall are typically above 85%, whereas state-of-the-art behavior pattern-based methods achieve F1 scores of up to 89%. This highlights the potential of behavior patterns in recognizing social behavior and detecting interacting groups. Significant performance differences exist across various datasets, such as detection accuracy reaching 87% on the Crowd-Tracking dataset and slightly lower performance (around 83%) on the ETH dataset. These differences suggest that the diversity and complexity of datasets affect detection results to some extent, and current datasets cannot sufficiently cover all possible social behavior patterns. Lastly, the paper discusses the challenges and limitations of social group detection and outlines future research directions. Current challenges mainly focus on how to effectively process large-scale video data, achieve accurate group detection in data-scarce environments, and address algorithms' real-time performance and computational efficiency. Furthermore, social group detection research faces challenges in integrating cross-disciplinary knowledge, such as incorporating behavioral patterns and social psychology theories to enhance detection accuracy. Future research could expand in the following areas. First, researchers could explore implicit social patterns, particularly covert criminal group detection. Second, interpretable fundamental social patterns could be extracted to enhance the accuracy and robustness of social group detection. Third, unsupervised and self-supervised learning methods could be used to address the issue of few-shot learning, thereby improving the applicability of algorithms in real-world scenarios. In conclusion, social group detection is not only a technical challenge but also an interdisciplinary research issue with vast application potential. With continuous advancements in algorithms and hardware, future social group detection technologies will play an increasingly important role in public safety, smart cities, and human-computer interaction.

**Key words:** social group detection; understanding interaction behavior; interaction detection; F-formation; behavioral pattern

## 0 引言

公共场所中的社交群体检测,通常是指在监控视频中利用计算机视觉、传感器和人工智能等技术手段,在公共场所中识别出人群中哪些个体在进行

互动或形成了社交群体。近年来,它已被认为是分析和预测人群行为的必要步骤,研究数量稳步增长,并且促进了多个研究领域的发展,如事件识别(Yuan等,2015)、语义场景分割、人群跟踪(Liu, 2016)、异常检测(Palanisamy 和 Manikandan, 2017)、监控视频分析(Wang等,2020;Liu等,2023;Raman等,2023)、

人机交互(human-computer interaction, HCI)(Robinson等, 2023; Li等, 2021; Montesdeoca等, 2022)和社会心理学(Sheng和Li, 2021; Zitouni等, 2019)。

Kendon(1990)最早提出社会群体和互动的概念。Nicolas和Hassan(2023)、Zanlungo等人(2014)认为群体成员与一个或多个行人建立社会关系,并一起朝着共同的目标前进,而他们被Moussaïd等人(2010)和Huang等人(2018b)视为具有社会关系并有意一起行走的个体,例如朋友或家人。社会心理学家Stangor认为两人组成的群体在社会群体中也普遍被接受。著名的F-formation(Zhang和Hung, 2021; Ferrera-Cedeño等, 2019)被定义为自由站立且进行社交互动的一群人,是由两个或两个以上的人在社交交往中形成的空间模式。Barua等人(2024)将社交群体定义为“由两个或两个以上具有相似特征和共同的团结意识的个体之间的互动和联系形成的新兴实体”。因此,从广泛意义上讲,Barua等人(2024)的定义更符合社交群体的概念。

在社会公共场所中,研究人类的社交行为并进行社交群体划分对群体行为分析领域意义重大。本文在如下主流数据库中进行了详尽的检索:IEEE Xplore数据库、ACM数据库、Scopus数据库、Science-Direct数据库以及Google Scholar。所用检索关键词为“群体检测”、“F-formation”、“群体互动”、“视频社交互动”、“行人分组”、“会话组检测”、“交互组检测”等,共检索出与调研主题相关的170余篇文献。其中,中国学者占比约为36%,且99%的中国学者选择在国外刊物发表所调研主题的论文。

如图1所示,以每3年为单位绘制了文献数量柱状图。可以看出,自2013年以来,“监控视频中的社交行为理解与群体检测”相关研究取得了长足进展,论文数量基本处于增长趋势,到2024年底,预计会有更多的论文发表。据调查,目前的综述性文章有Barua等人发表的两篇主题为“社交机器人如何使用F-formation感知和加入社会互动群体”的一项调查综述(Barua等, 2020, 2024),其关注的是社会公共场所中社交群体检测的人机交互技术,缺少对其他一般性方法的概述。此外还有北京邮电大学智能通信软件与多媒体北京市重点实验室(王正等, 2021)与四川大学电子信息学院(高建军和卿粼波, 2019)发表的社交关系识别研究综述,其关注具体的社交关系识别方法。因此,在社交行为理解与社交群体检测领域,并没有

综述性文章能够提供对研究、技术状况以及未来发展的总体概述。本文全面涵盖了几几乎所有的技术分类和研究领域,目的是全面地讨论该领域,方便科学家、研究人员和感兴趣者对该领域有一个快速、全面的认识,能够帮助他们在未来开展更深入的研究。

近年来,社会群体的检测得到大量研究,但其检测性能仍有很大提升空间。由于人类社会行为具有复杂性和不可预测性,不同的空间群体结构和多个活动群体中的人类行为会持续发生变化(Lu等, 2023),导致社会群体检测面临两个重大挑战:如何表达具有紧凑和有意义的社会交互特征以及如何利用这些特征形成良好的群体聚类/检测。研究者针对这两个问题开展了一系列研究,本文制作了一个研究年表,如表1所示。表内显示了各年份研究使用/提出的交互建模方法或群体检测技术,便于读者观察主要研究方法及其演变趋势、研究热点与主要团队。

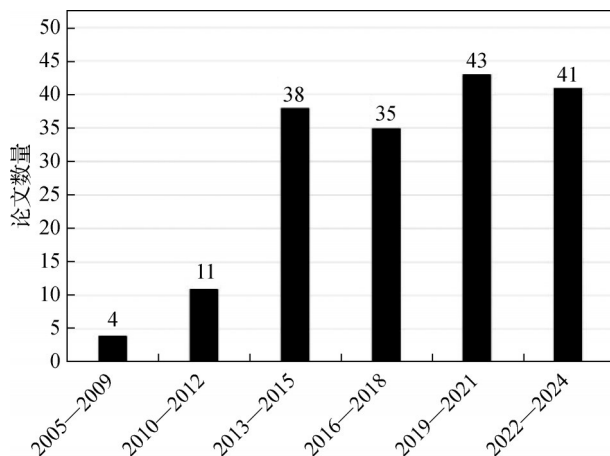


图1 2005—2024年论文发表数量统计

Fig. 1 Statistics on the number of published papers from 2005 to 2024

从2009年到2015年间,基于规则的固定启发式方法盛行,这一时期的研究主要考虑人们在交互时产生的距离范围、轨迹相似度和速度一致性等空间几何特征,利用轨迹亲和力矩阵或运动相似性判定人与人之间的交互强度,进而利用DBSCAN(density-based spatial clustering of applications with noise)聚类、Hough投票机制、基于图论的优势集法(dominant Sets)以及支持向量机(support vector machine, SVM)等聚类手段划分群体。

突出理论包括描述互动者空间排列模式的F-formation(Kendon, 2010)与霍尔提出的人类接近

距离理论(human proxemics),这些考虑轨迹、位置、身体方向等因素的方法也称为“社会信号处理(social signal processing, SSP)”(Dim和Kuflik, 2010)。事实上,这一时期的关注焦点在于“空间社会信号处理”。2014年后,研究者逐步转向以视觉信号和时空关系为核心的交互建模,他们更加关注身体姿态与注视注意力等身体交互信号,结合空间几何特征形成多种视觉交互描述符来构建交互图,以此表达人类的交互特征,人机交互技术也得到了快速发展。这些加入视觉信号的SSP也称为全域社会信号处理(Cristani等, 2013)。研究者或利用显著点、ViBe背景减除法等其他视觉处理技术去提取和表达群体聚

集特征。聚类方法多使用高斯混合模型(Gaussian mixture module, GMM)、马尔可夫随机场、贝叶斯网络、图聚类等机器学习模型。此类研究在2017年达到顶峰。同时,尤其是2020年之前,研究者趋向于设计人类交互模式的高级语义表达,如社会力模型(social force model, SFM)、社会互动场或轨迹群模式等进行社交群体检测,这些模型通常借助“力”或“场”等物理学原理去表达人类的相互作用力,从本质上揭示社交行为模式,或关注人与环境之间的交互,如“事件”、“因果”等社会属性去手动定义行为模式。以上所述研究本质上都是利用一定的规则去定义交互强度。

表1 主要研究方法年表(2005—2024)  
Table 1 Statistical analysis of major research methods over 2005—2024

| 年份   | 主要理论/方法/技术                                                                                                                                                                                                                                                                                                                                                                                       |
|------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 2005 | DBSCAN轨迹相似性聚类(Kalnis等, 2005)                                                                                                                                                                                                                                                                                                                                                                     |
| 2008 | 马尔可夫随机场和排斥力建模相互作用力(Pang等, 2008),交互事件驱动的多层次动态贝叶斯网络(Dai等, 2008)                                                                                                                                                                                                                                                                                                                                    |
| 2009 | 基于Hausdorff距离和速度的层次聚类(Ge等, 2009)                                                                                                                                                                                                                                                                                                                                                                 |
| 2010 | F-formation(Kendon, 2010),基于位置接近度和语音的社会信号处理(Dim和Kuflik, 2010),基于距离和方向的高斯混合模型(GMM)(Groh等, 2010)                                                                                                                                                                                                                                                                                                   |
| 2011 | F-formation(Hung和Kröse, 2011; Cristani等, 2011),基于轨迹的社会力模型(social force model, SFM)(Šochman和Hogg, 2011),SFM(Leal-Taixé等, 2011),从视觉概念推断社会关系(Ding和Yilmaz, 2011)                                                                                                                                                                                                                                     |
| 2012 | 多模态传感器(Kjærgaard等, 2012),基于位置和速度的在线贝叶斯(Zanotto等, 2012),基于Hausdorff距离和速度的聚类方法(Ge等, 2012),动作交互式短语(Kong等, 2012)                                                                                                                                                                                                                                                                                     |
| 2013 | 行为/交互描述符(Li等, 2013),霍尔人类亲近学理论(Solera和Calderara, 2013),轨迹/位置与方向(Yücel等, 2013),语义驱动的时空描述符(Nabi等, 2013),霍尔社会距离度量的结构化SVM(Solera和Calderara, 2013),F-formation(Setti等, 2013a, b),SFM(Mazzon等, 2013),基于认知语义的事件表示(Yin等, 2013),跨摄像机生成模型(Qin等, 2013),基于谈话/手势/对视等社交行为(Gorriani等, 2013),社会信号处理:F-formation与SFM(Cristani等, 2013)                                                                              |
| 2014 | 基于空间和几何关系识别结构化群体(Choi等, 2014),跨摄像头轨迹的社会亲和力图(Alahi等, 2014),基于位置、速度相似性的投票机制(Ai Masum等, 2014),运动特征与视觉注意力特征(Leach等, 2014)                                                                                                                                                                                                                                                                            |
| 2015 | 基于稀疏运动信息的GMM(Poiesi和Cavallaro, 2015),F-formation(Setti等, 2015; Ricci等, 2015),基于轨迹起止点的关联矩阵(Khan等, 2015),距离/方向相似性(Bozorgtabar等, 2015; Chandran等, 2015; Li等, 2015),基于互动形成期的子图搜索(Zhang等, 2015b),显著点轨迹聚类(Shao等, 2015),Panoptic Studio:多视点社交互动感知(Joo等, 2015),基于演变的轨迹交互网络(Feng和Bhanu, 2015),基于区域运动线索的暴力互动检测(Rota等, 2015),集合层次聚类(Setia和Mittal, 2015),社交属性感知力模型(Zhang等, 2015a),基于轨迹空间分布的静止人群检测(Yi等, 2015) |
| 2016 | 四层结构生成模型(Zhou等, 2016),基于社交身份轨迹的支持向量机(SVM)(Solera等, 2016),基于轨迹关键点的多级滤波GMM(Zhao等, 2016),基于轨迹/频率/速度的期望最大化聚类(expectation maximization clustering, EM)(Voon等, 2016),学习和遗忘策略建模亲密关系(Islas等, 2016),互动社会图模型(Nunes等, 2016),位置/速度/角度(Favaretto等, 2016),基于位置/方向的视觉注意力场(Inaba和Aoki, 2016),群体活动状态变化模型(Nakamura等, 2016),基于像素编码的静止人群检测(Yi等, 2016)                                                                |
| 2017 | 上下文描述符描述运动特征(Li等, 2017),引入手势信息(Habe等, 2017),基于锚点的流形排序(Chen等, 2017),基于时空/速度相关性的集体转换算法(Shuaibu等, 2017),引入高级特征:视觉关注锥体(Vascon和Bazzani, 2017),基于几何关系、运动和相对姿态的交互描述符(Lathuilière等, 2017),对话和牵手行为(Tan等, 2017),在线学习的条件随机场(Chen等, 2017),正交亲和力编码的跨摄像机SVM(Solera等, 2017),基于音频/视频特征的互动SVM(Bano等, 2017)                                                                                                        |

续表 1 主要研究方法年表(2005—2024)

Continued Table 1 Statistical analysis of major research methods over 2005—2024

| 年份   | 主要理论/方法/技术                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 2018 | 头部/身体姿态弱监督学习(Varadarajan 等, 2018), 时空能量流(Huang 等, 2018a, b), 轨迹群模式语义(Sawas 等, 2018), 多模态 SVM(Bano 等, 2018), 运动相似性(Zaki 和 Sayed, 2018), 基于轨迹的固定宽度聚类(Yang 等, 2018), 距离/方向/同步移动(Yucel 等, 2018)                                                                                                                                                                                                                                                                                                                                             |
| 2019 | 基于显著点的群体描述符(Chen 和 Lai, 2019), 轨迹群模式语义(Sawas 等, 2019), F-formation(Ferrera-Cedeño 等, 2019), 基于共存时间和距离的 DBSCAN 聚类(Cheng 等, 2019), 基于生成对抗网络(GAN)的无监督学习(Fernando 等, 2019), 几何线索与情感相关性(Yoo 等, 2019), 基于轨迹因果关系的分层贝叶斯模型(Xie 等, 2019), 基于距离、方向和姿势的社交互动场模型(Zhou 等, 2019), 社交力图学习人群交互信息(Lin 等, 2019), 霍尔人类亲近学理论(Favaretto 等, 2019), 运动相似性(Bisagno 等, 2019)                                                                                                                                                                                       |
| 2020 | F-formation(Livramento 等, 2020; Hedayati 等, 2020; Barua 等, 2020), 图注意力网络(Ehsanpour 等, 2020), 基于位置、速度的 CNN(Zitouni 等, 2020), 基于多源特征描述符的 SVM(Basalamah 和 Khan, 2020), 基于位置/距离的稀疏图分析(Pouw 等, 2020), 基于位置/朝向的图亲和网络(Swofford 等, 2020), 非递归运动相似性聚类算法(Zhao 等, 2020a), ViBe 背景去除算法(Pandey 等, 2020), 基于结构化上下文的多视图聚类(Wang 等, 2020), 基于速度/方向的结构化 SVM(Zhao 等, 2020b), 核相关性语义分析(Fernández-Ramírez 等, 2020), 基于密集光流速度/距离的层次聚类(Taylor 等, 2020), 递归社会行为图(Sun 等, 2020), 线积分卷积(line integral convolution, LIC)的流场特征(Zhang 等, 2020), 运动信息(Zitouni 等, 2020) |
| 2021 | F-formation 工具包(Rivas 等, 2021), 图神经网络(GNN)(Thompson 等, 2021), 图形嵌入式 CNN(Tzelepi 和 Tefas, 2021), 基于链接预测的 GNN(Schmuck 和 Celiktutan, 2021), 配对表示学习(Huang 等, 2021), 距离/方向/姿势(Wright 等, 2021), 基于自监督社会关系学习的 GNN(Li 等, 2021), F-formation(Zhang 和 Hung, 2021), 图形嵌入式卷积神经网络(CNN)(Tzelepi 和 Tefas, 2021)                                                                                                                                                                                                                                        |
| 2022 | F-formation(Pathi 等, 2022), Transformer(Tamura 等, 2022), 基于距离的无监督图聚类(Su 等, 2022), 距离(Yang 等, 2022; Su 等, 2022), 基于动态 Epsilon 的改进型 DBSCAN(Chen 等, 2022), 基于时空上下文亲和度的长短期记忆网络(long-short term memory, LSTM)(Tan 等, 2022), 基于身体视锥体的图卷积网络(graph convolutional network, GCN)(Atıcı 等, 2022), 基于人群指示特征和卡尔曼滤波器的 CNN(Taylor 和 Riek, 2022), 多尺度时空图卷积网络(Zhou 等, 2022), 距离(Quiroz 等, 2022), 基于距离亲和矩阵的自调谐谱聚类(Han 等, 2022)                                                                                                                            |
| 2023 | 空间/姿态约束的双分支深度学习(Lu 等, 2023), 基于轨迹与视觉上下文的图聚类(Li 等, 2023b), 基于轨迹的图 Transformer(Jahangard 等, 2023), 轨迹(Küpper 和 Seyfried, 2023), 基于关系存在判别的 ResNet(Li 等, 2023a), 基于距离的交并比(intersection over union, IoU)聚类(Kang 等, 2023), 基于区域划分的 K-means 和 DBSCAN 聚类(Chen 等, 2023), 基于 GNN 的多注意力上下文图谱模型(Yan 等, 2023), 端到端 CNN(Bhaskara 等, 2023)                                                                                                                                                                                                             |
| 2024 | F-formation(Chen 和 Liang, 2024), 基于多模态交互线索的 GCN 模型(Lyu 等, 2024b), 基于注意力机制的多尺度 GCN 模型(Zhang 等, 2024), 基于 LSTM/RNN 的轨迹深度亲和网络(Nasri 等, 2024), 基于视觉环境感知的 Transformer(Yu 等, 2024), 基于群体互动场的图注意力网络(Wang 等, 2024), 基于时空信息和视觉特征的循环条件随机场(Zhang 等, 2024), 监控视角与鸟瞰视角社交距离(Ghasemi 等, 2024), 运动相似性(Lyu 等, 2024b), 基于光流的 SFM 模型(Wang 等, 2024), 基于空间接近性与时间连续性的行人流网络(Wu 等, 2024)                                                                                                                                                                        |

2019 年之后, 基于学习或数据驱动的方法, 如深度学习和强化学习取得了突出的进展, 并且每年都在加快步伐。学者们广泛应用卷积神经网络(convolutional neural network, CNN)、图神经网络(graph neural network, GNN)、Transformer、图注意力网络和生成对抗网络(generative adversarial network, GAN)等深度学习技术进行自动交互模式学习与群体分类, 尤其对注意力机制的运用, 进一步推动了该领域的发展。值得注意的是, F-formation 技术成为贯穿多个年份的热点。各种社会信号也通常作为深度学习网络的基本数据输入, 因此在近 3 年, 基于规

则的社交建模方法往往辅助深度学习方法进行监督、半监督或无监督学习。本文由此总结: 时空信号和视觉信息是理解、研究人类社交行为时不可或缺的因素。此外, 代表性研究者如 Kalnis、Zanotto、Cristani、Solera、Li 等团队, 在不同阶段提出关键的理论与模型, 推动了群体交互理解从静态到动态、多视角的转变。

综上所述, 本文根据现有研究对人类互动行为建模方法的不同, 将群体检测方法分为基于启发式规则的方法与基于学习的方法, 如图 2 所示。其中基于规则的方法又分为基于空间的方法、基于视觉

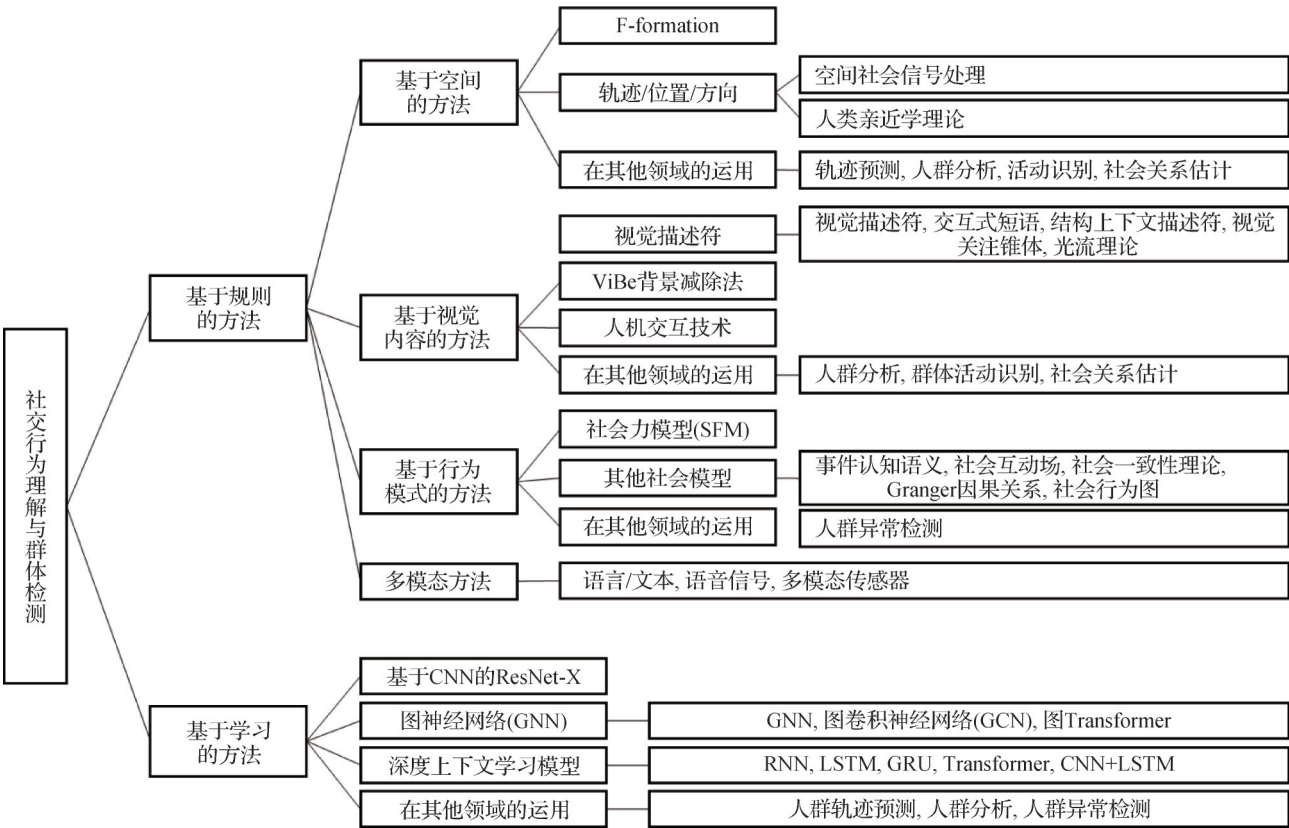


图2 文献分类导图

Fig. 2 Literature classification map

内容的方法、基于行为模式的方法和多模态方法。在每种分类中,着重介绍其中的研究热点,并介绍它们在其他研究领域的运用,最后总结它们的优缺点。这自然地形成了本文的组织结构,分别为基于空间的方法:F-formation技术、空间社会信号处理与人类亲近学理论;基于视觉内容的方法:视觉描述符、ViBe背景减除法与人机交互技术;基于行为模式的方法:社会力模型与其他社会模型、多模态方法;基于学习的方法:基于CNN的ResNet-X网络、图神经网络(GNN)与深度上下文学习模型。

1 基于规则的群体检测算法

基于规则的方法通常基于手工制定的启发式规则。大多数研究者试图用轨迹相似性与空间位置关系来分析社会互动。在过去的10年中已经提出许多成熟的方法。

1.1 基于空间的方法

基于空间的方法根据个体的相对物理位置和身体或头部的方向来检测个体之间的社会互动,主要

分为基于F-formation技术和基于轨迹/位置/方向的方法。

1.1.1 F-formation技术

参与者群体互动的F-formation理论最早由社会心理学家Adam Kendon提出,用于描述人们在面对面交流时常见的空间排列模式(Kendon,2010)。如图3所示,F-formation用来表示人与人之间的距离和互动关系的社会空间有3个:O空间、P空间和R空间。O空间是最核心的区域,参与者面向彼此,在这个空间内进行直接的互动和沟通,身体通常围绕这个空地呈现面对面的姿势,没有其他人能够进入这个区域,除非也成为互动的一部分;P空间位于O空间的外圈,是参与者站立区域,这个位置接近O空间,参与者能够间接参与互动;R空间是最外层的区域,用于旁观者或不直接参与互动的人,这个区域的人不会打扰到O空间内的互动,但能够观察或等待加入。根据社会科学,Kendon(1990)提出6种标准队形排列,如图4所示,它们是面对面形、并排形、L形、列形、方形和圆形。除此之外,还有许多其他类型的队形,如半圆形、矩形、三角形、V形和勺形。

Kendon 解释说,有些听众可能是 F-formation 的一部分,尽管他们不直接交流,只是倾听其他互动的人。在这种情况下,他们可以位于  $P$  空间。

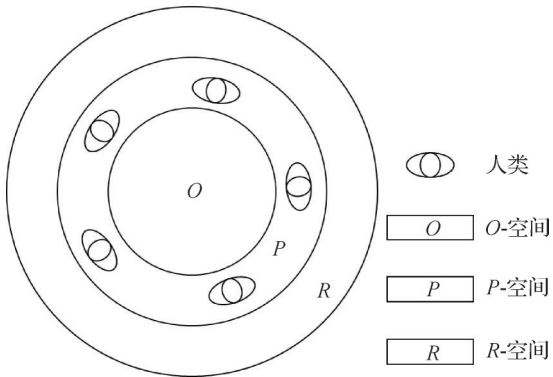


图3 F-formation的3种社会空间

Fig. 3 Three social spaces corresponding to F-formation

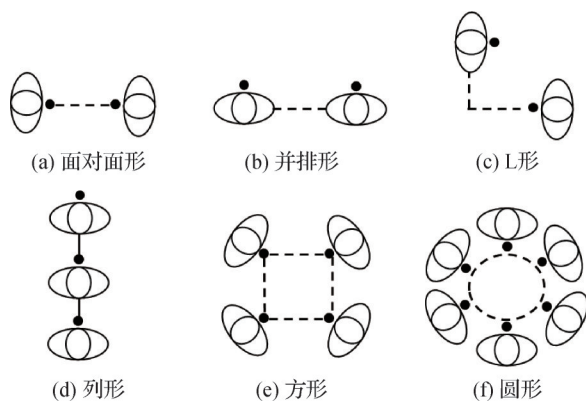


图4 常见的F-formation队形排列

Fig. 4 Common F-formation formations ((a) via-a-vis; (b) side-by-side; (c) L-shape; (d) column; (e) square; (f) circle)

1) 群体检测技术。F-formation 最具代表性算法是 Hough 投票机制与基于图论的优势集法。Hough 投票法通过推断交互空间的中心位置来识别 F-formation (Cristani 等, 2011), 而优势集法 (dominant sets) 则利用图论中的最大团概念, 将人与人的位置和方向数据输入聚类算法。其中又以 Setti 团队为主。研究证明, 当仅有位置数据时, 优势集法表现更佳 (Hung 和 Kröse, 2011); 而在有位置和方向数据时, Hough 投票法更为有效 (Setti 等, 2013a)。

Setti 团队通过多个投票会话的竞争来解决传统的组检测方法往往局限于单一尺度, 无法适应不同规模的群体这一问题, 每个会话专注于特定的组大小, 最终通过信息论标准评估所有投票, 生成最终

的组集合 (Setti 等, 2013b)。2015 年, 他们又提出一种基于图切割的算法, 该方法根据人的位置和方向信息构建图模型, 利用图切割技术来更新 F-formation 的中心, 找到社交群体的  $O$  空间, 并检测这些交互群体。Hung 和 Kröse (2011) 在优势集理论的基础上, 使用了人与人之间的距离以及身体方向信息, 改进了现有的基于模块性切割 (modularity cut) 的方法, 提供一种更精确的方式来检测这些交互群体。还有一些学者引入了位置和头部朝向的模糊关系理论, 生成模糊矩阵来检测独立对话组 (Ferrera-Cedeño 等, 2019), 该方法提升了在拥挤场景中检测群体时的准确率和稳定性。

在检测不像亲密成员那样紧密参与对话的群组成员时, 有学者提出一种新的特征表示方法, 引入了注意力视锥模型, 考虑空间环境的影响, 如家具布置和人群密度不均匀等因素 (Zhang 和 Hung (2021))。Rivas 等人 (2021) 结合人类接近距离理论, 开发了一个用于 F-formation 模式检测的 Proxemics 工具包, 结合 proxemics 度量和 F-formation 模式的概念预定义各种互动类型, 并通过人与人之间的距离、身体方向以及相对位置来估计必要值。

2) 人机交互技术。调查发现, F-formation 往往与人机交互技术息息相关, 主要研究目标是基于自然数据驱动的方法来解决人形移动机器人如何接近一群人的问题。机器人会首先处理来自机器人自带传感器 (RGB 相机和激光) 的数据 (Pathi 等, 2022), 根据周围环境和空间布局来检测个人与交互群体 (Livramento 等, 2020), 或通过跟踪人的位置、头部方向、附近感兴趣物体的位置及其群组情况来估计下半身方向, 以 F-Formation 为导向来识别互动组 (Ricci 等, 2015), 这种办法解决了当前多数互动型机器人只能被动等待人类靠近的局限性。

近几年, 数据驱动的机器学习方法成为 F-formation 技术的主要关注点 (Paay 等, 2013)。Hedayati 等人 (2020) 提出一种数据驱动机器学习算法 REFORM。该算法分为 3 个步骤: 数据解构、成对分类和重构。首先, 通过将每个场景的帧分解成成对的数据点 (即两个人的相对位置和方向), 接着利用欧几里得距离和身体旋转角度等特征进行分类, 判断交互对。最后通过贪婪重构算法, 将成对数据重新组装为完整的 F-formation。基于深度学习的实时 F-formation 检测系统首先通过 PoseNet 模型检测

人群中的关键点,并利用条件随机场(conditional random field, CRF)模型预测场景中的组成员及离群者。随后,使用多分类支持向量机(SVM)对 F-formation 进行分类,判断人群的社交模式,如面对面、并排站立、L型或三角形结构(Barua等,2020)。

最近,学者探索了能够帮助机器人在复杂社交场景中理解并参与群体互动的一种基于拓扑数据分析(topological data analysis, TDA)的创新方法(Chen和Liang,2024))。通过利用持续同源性(persistent homology)定义了一种六维几何结构,量化和评估群体的空间稳定性。该方法首先通过几何结构(如视觉交点)捕捉个体间的空间关系,随后利用 Alpha 复杂结构(alpha complexes)生成滤波树,分析不同群体内外的连接性和紧密性。

### 1.1.2 基于轨迹/位置/方向的方法

在过去的20年中,基于轨迹、位置和方向的社交行为理解和群体检测已经成为一个迅速发展并且热度不减的关注点。研究者们利用移动设备或计算机视觉工具收集大量的轨迹数据与身体方向,探讨人类行为的模式、社交互动以及群体动态。他们通常根据位置的接近程度、轨迹相似性或运动相似性来确定一个连贯的群体。通过分析这些空间社会信号,研究人员能够识别社交活动的热点区域、互动频率和参与者之间的关系(Pang等,2008)。此外,群体检测技术可以揭示群体形成的机制和动态变化(Li等,2023a)。许多研究对社交行为的理解提供了新的视角。

1)空间社会信号处理。其中一个突出的理念是社会信号处理(SSP)。Cristani等人(2013)将 SSP 定义为:通过利用非语言信号,如面部表情、身体姿势和空间距离,提供一种新的视角来理解人类互动,而 Dim 和 Kuflik(2013)的 SSP 中包含了语音交流信息。在这些定义中包含了视觉、空间和语言信息。从调研情况来看,视觉内容特征的研究已经自成一派,且语音相关研究较少。因此,将社会信号处理分为空间社会信号处理与考虑视觉内容以及其他因素的全域社会信号处理(见1.2节)。Dim 和 Kuflik(2010)通过社交信号处理(SSP)检测小组互动,通过监测小组成员的物理位置接近度和语音交流来推断其社交情境。实验观察了58个随机选择的小组,记录了每分钟的成员位置接近度和语音互动时间。通过 K-means 聚类分析将小组分为3种类型:始终接近、

始终分离和偶尔变化。

与 F-formation 或静态群体的固定交互模式挖掘不同,事实上,动态群体的位置和内容是不断随时间变化的,但群体的核心特性仍保持不变。基于轨迹/位置/方向的方法通常假设一组对象在较长时间段内以相似路径移动,即轨迹相似性(Yücel等,2013)。且成员趋向相同的运动方向,瞬时速度或阶段性位置聚集等,即运动相似性(Yücel等,2013;Li等,2015)。这种方法首先利用先进的行人检测和多目标跟踪算提取行人轨迹,接着通过计算行人 ID(identity document)之间的相对距离将距离接近的行人聚类到同一群体中(Yi等,2015),体现了轨迹的空间一致性;或结合位置和速度信息的对称 Hausdorff 距离进行聚类,以识别低到中密度人群中共同移动的小团体(Ge等,2009),还有学者使用主成分分析(principal components analysis, PCA)提取位置和速度的相似性亲和特征(Al Masum等,2014)。常用聚类方法有 DBSCAN(density-based spatial clustering of applications with noise)(Kalnis等,2005;Chen等,2022,2023)、优势集聚类(Pavan和Pelillo,2007)、K-means(Chen等,2022,2023)、交并比(IoU)聚类(Yi等,2015)、SVM(Solera和Calderara,2013;Zhao等,2020b)、图聚类算法(Tran等,2013;Inaba和Aoki,2016;Su等,2022)或卷积神经网络等深度网络(Chen等,2022)。

密度聚类算法(DBSCAN)对每个时间片进行空间聚类,并结合时间序列形成移动群体(Kalnis等,2005)。高斯混合模型(GMM)能够模拟这些交点的空间概率密度函数,以准确预测人们的互动位置(Poiesi和Cavallaro,2015)。还有学者引入了区域划分技术(Chen等,2023),利用区域内的密集度和距离,再结合 K-means 和 DBSCAN 等聚类算法进行群体检测。Groh等人(2010)提出一种新方法,通过交互几何(主要为人际距离和相对身体方向)定量测量小时间和空间尺度上的社交互动。该方法使用 D-Track 软件和 ARTrack2 IR 摄像头的红外信标跟踪系统,以高精度记录参与者的空间位置和方向,建立多种概率分类器,包括高斯混合模型(GMM),对数据进行训练和评估,通过一系列实验测量两个人之间的这两个参数,以识别现有的社交场景。

利用基于距离、速度和运动方向的非递归运动

相似性聚类算法(non recursive motion similarity clustering, NMSC)自动识别社交群体也是研究热点(Chandran 等, 2015; Li 等, 2015; Favaretto 等, 2016; Zhao 等, 2020a)。这种方法假设群体内部的行人表现出相似的运动特征,如空间位置的接近性和运动方向的一致性。如 Yücel 等人(2013)提出两种运动模型:位置模型和方向模型,利用行人之间的距离和方向一致性,在不同环境和群体配置下识别行人群体。Zhao 等人(2020b)提出基于运动相似性的结构化支持向量机(structured SVM)的在线群检测方法。Küpper 和 Seyfried(2023)监测行人之间的距离及其随时间的稳定性,以此识别出同一社交群体中的成员。有些研究还能通过运动相似性自动分析群体成员的速度偏好和行走策略(Zaki 和 Sayed, 2018)。Khan 等人(2015)提出一种不同于以往轨迹相似度计算的方式,该方法只考虑行人的起点和终点(即“源点”和“汇点”),通过构建“关联矩阵”来表示所有行人的起点和终点的联合概率分布。并通过分层聚类来发现潜在的社会群体。为了克服遮挡问题,研究者们基于距离、出现频率和速度相关性的权重计算公式,来衡量和推断轨迹簇中个体之间的连接度,最终通过 EM 算法推断出最终的群体结构(Voon 等, 2016)。

另一个研究热点是图聚类算法,通过将个体的位置和身体方向信息表示为完整图,使用亲和函数来定义个体之间的紧密度,进而检测 F-formation(Inaba 和 Aoki, 2016)。基于演变的轨迹交互网络将行人轨迹表示为网络中的节点,并使用边及其权重描述分组行为,创新性地探索了社会群体的动态特性(Feng 和 Bhanu, 2015)。Cheng 等人(2019)还考虑了共存时间和欧几里得距离,能够输出具有高 IoU 值的可靠结果。

以上监督式的群体检测方法在处理复杂的拥挤运动场景时容易产生鲁棒性不足的问题,一些学者研究了自适应性的无监督判别方法。Choi 等人(2014)假设人与人之间的空间和几何关系是发现群体的重要线索。提出“结构化群体”的概念,这些群体根据人物之间的空间交互关系来分类,例如相对面坐着、并排站立等。这些模式通过学习无监督的判别互动模式来实现。作者还提出一种迭代增强算法,用于在不明确群体数量的前提下,推断图像中的群体结构,并得出群体的三维布局。自适应聚

类算法通过一种自动前景检测策略,首先在前景区域追踪显著点生成密集的轨迹片段。然后,利用自适应初始化聚类技术,构建一个分层聚类模型,根据不同特征逐层将人群划分为群体(Shao 等, 2015)。Zhao 等人(2016)结合高斯混合模型(GMM)和 KLT (Kanade-Lucas-Tomasi)跟踪器,用于提取前景关键点,并通过多级滤波策略实现群体检测。这些滤波器包括时空距离相邻滤波器、速度角滤波器和速度相关性滤波器,分别用于计算邻近关键点之间的空间距离、速度角度以及速度相关性,从而过滤掉不一致的运动行为。

2) 人类亲近学理论(human proxemic theory)。霍尔在近 60 年前首次研究了人类亲近学理论,也称“人类亲密距离理论”,认为人与人之间的距离能够反映社交关系的亲密程度,常被用于社交互动的关键特征,如图 5 所示,依据人与人之间的距离将其划分为亲密距离、个人距离、社交距离和公共距离(Solera 和 Calderara, 2013; Su 等, 2022)。

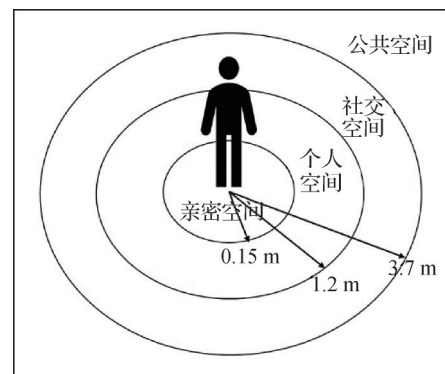


图5 人类亲近学理论距离空间

Fig. 5 Distance Space in human proxemic theory

在这类方法中,首先通过时空轨迹数据测量个体间的欧几里得距离和 Fréchet 距离,来量化人际距离。结合这些距离数据,设计社会互动识别方案(Su 等, 2022),并通过聚类技术划分社交群体。通过引入了基于“霍尔人类亲近学距离”理论的社会约束,群体检测能够更加符合人类社会互动的特点。Zanotto 等人(2012)提出一种在线贝叶斯非参数方法,该方法基于 Dirichlet 过程混合模型(Dirichlet process mixture model, DPMM),通过将群体表示为无穷混合模型的组成部分,个体则表示为这些组成部分生成的观测数据。该模型利用位置和速度特征进行无监督学习,并通过一种顺序变分推理框架,实

现实时的群体检测。结构化支持向量机(SVM)能够学习如何根据这些距离特征将行人的轨迹分组为社交群体(Solera和Calderara,2013)。此外便是图聚类方法,Su等人(2022)利用霍尔琴技理论,使用图聚类算法将行人划分为不同的社会群体,生成互动矩阵和社交图。

人类亲近学理论在疫情期间得到了广泛使用,用于监测公共场所中的“距离违规者”。如Pouw等人(2020)基于稀疏图的高效分析方法,实时处理行人的位置和相互距离数据,在疫情期间识别群体中的“距离违规者”和“家庭组”。Yang等人(2022)使用SORT(simple online and realtime tracking)算法进行多目标跟踪。通过鸟瞰式社交距离识别“安全”社交群体与“距离违规者”。Ghasemi等人(2024)同时关注视频社交距离分析器(auto-SDA)和鸟瞰视角社交距离分析器(B-SDA)。

#### 1.1.3 其他应用领域

1)轨迹预测。研究表明,群体互动在行人运动轨迹的预测中具有重要作用。利用一致性过滤器检测具有共同运动趋势的行人群体能够实现更好的轨迹预测(Bisagno等,2019)。Lyu等人(2024a)还引入行人—群体互动模块(pedestrian-group interaction module, PGIM)来捕捉行人之间及其与群体的关系计算行人之间的相似性。

2)人群分析。使用运动一致性信息将个体聚合成群体,还能够用于从视觉监控数据中理解人群的社会认知行为(Zitouni等,2020)。通过分析这些集群的运动模式(如数量、群体质心之间的距离和质心位置的变化速度),可以检测出不同的人群事件变化,如合并和分裂事件(Yang等,2018)。Favaretto等人(2021)利用跟踪行人的速度、位置和角度变化特征检测交互群,并用于分析文化行为。

3)活动识别。Han等人(2022)构建了一个图网络来表示场景中的个体,节点之间的边表示个体之间的关系。利用亲和矩阵和空间距离矩阵识别社交群体,并用于群体活动识别。

4)社会关系估计。Yucel等人(2018)基于跟踪轨迹,根据人际距离,共同方向和同步移动来识别群体,并用于推断群体成员之间的社会关系。

#### 1.1.4 总结

尽管具有潜力,但F-formation技术也有一定的局限性。在群体检测中,F-formation检测中的一个

主要问题是遮挡(Dim和Kuflik,2013)。一个群体中的人可能站在这样的位置,其中一些人可能会遮挡其他人,或观看摄像机的角度被放置在这样一个位置,使整个群体不可见。此外,它只能模拟空间结构的形成,不太适合动态群分析(Setti等,2015)。基于位置结构(如F-formation)或行为检测的方法无法充分捕捉社会群体的本质,往往限制于静态情境或行为预定义(Li等,2023a)。这表明需要更先进的技术来模拟社会群体的时间演变,同时考虑到群体成员在空间和行为方面的变化。

由于对群体密度和相互作用行为变化的敏感性,轨迹信息仍然远远不足以准确度量成对相似度,需要新的领域信息。此外,现有的作品通常测量所有个体的平均位置和速度来表示整个人群的社会关系,这既不准确也不稳健。尽管他们做出了努力,但仅凭轨迹与方向信息不足以正确区分和识别不同密度人群中的群体。

#### 1.2 基于视觉内容/全域社会信号处理的方法

传统的群体检测方法主要依赖于运动特征(如方向、速度和距离)等低级特征,但这些特征在拥挤环境中容易受到限制或干扰。视觉语境为视频内容提供了丰富的信息。近年来,研究者试图对视觉图像所包含的群体相关信息进行聚合。与运动路径相比,可以从图像中发现更多的细节。此外,图像中的视觉语义线索(高级特征)无需从轨迹中观察到(Vascon和Bazzani,2017),例如,如果两个人正在进行聊天(唇部动作)、牵手或拥抱等身体姿态交互,可以直接识别他们来自同一组,而无需分析运动路径,也避免了使用复杂的多目标跟踪系统跟踪行人轨迹。2019年,Nature的一项研究用线性函数拟合了低层特征(距离和身体方向)与高层特征,探讨姿态开放程度对互动强度判别的影响,如图6所示。姿态开放程度对交互强度的相关性高达94.2%,约等于低层与高层特征的联合相关性(Zhou等,2019)。

从调查结果看,研究者通常将视觉内容与空间信息结合使用(Li等,2017;Fernández-Ramírez等,2020;Zhou等,2019),这也称为全域社会信号处理方法。而动作捕捉方法在处理社交互动时通常面临3大挑战:频繁的遮挡,需要在足够大的空间内测量细微运动,以及人类外观和配置的巨大变化(Joo等,2015;Varadarajan等,2018)。

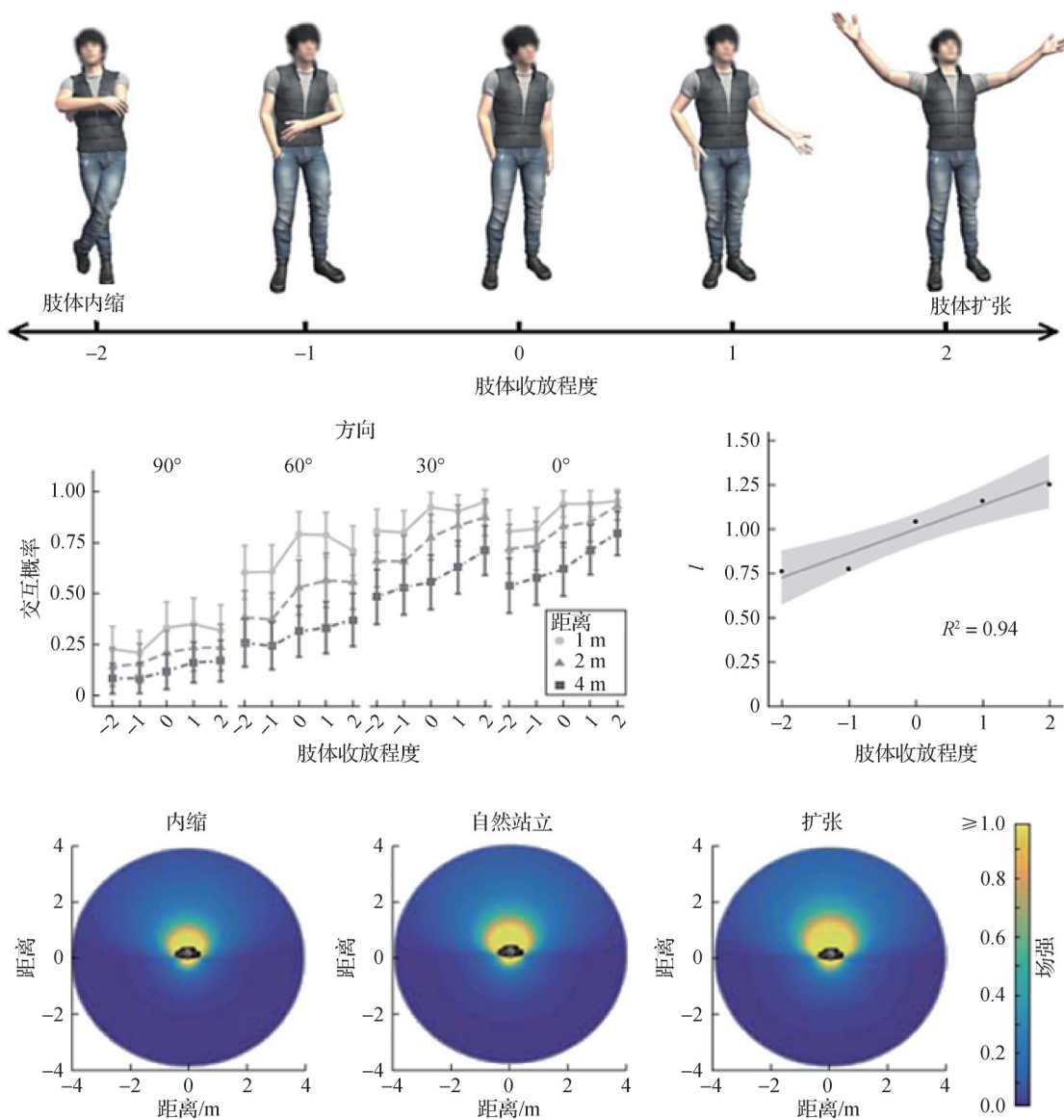


图6 *Nature* 研究中姿态开放程度对互动判断的影响(Zhou 等,2019)

Fig. 6 The influence of openness on interaction judgment in *Nature* research(Zhou et al. , 2019)

1. 2. 1 视觉描述符

在人与环境交互层面,通过高维特征空间到低维字典的映射来编码单人和两人之间的相对姿态和运动特征(Lathuilière 等,2017)。通过检测 5 种典型的个人行为(过马路、等待、排队、说话和走路)来确定视频中的群体(Zhou 等,2016)。这些描述符可用于定义能量函数的单项和成对势能。首先提取高维的时空运动特征和预先提供的边界框信息;其次分析两人之间的几何关系、运动和相对姿态,生成高层次的交互描述符。

此外,研究者通常编码高维视觉语义信息提取“视觉描述符”、“交互式短语”、“视觉关注锥体”、“结构上下文描述符”等特征,用来表征行人的运动交互

信息(尤其针对姿态交互)以及上下文信息(Lathuilière 等,2017; Tan 等,2017),本文将其统一称为“视觉描述符”。Kong 等人(2012)首次引入“交互式短语”这一高层次的描述来表达交互过程中的二元语义运动关系。交互式短语描述的是交互中两个人的运动关系,检测 6 种预定义行为,包括握手、拥抱、踢、指、打和推。该方法采用分层模型,使用潜在支持向量机(SVM)框架将交互式短语视为潜在变量,并明确捕捉它们之间的相互依赖关系,以处理互动中的运动歧义和部分遮挡问题。这些交互式短语实现了对复杂互动的细致描述。

视觉描述符往往关注身体姿态交互,研究假设同一小组成员之间会有频繁的身体互动行为,而不

同小组的个体之间则没有互动。这种方法通常采用计算机视觉模型捕捉视频序列中身体交互行为,并与轨迹信息结合以提高小组检测的准确度。计算成对的相似性权重。然后,使用聚类算法将相似性较高的行人轨迹分组(Tan等,2017)。如将手势互动与传统的距离和视线方向特征结合(Habe等,2017; Tan等,2017),或研究骨骼运动(Joo等,2015),他们提出一种基于生成骨架轨迹的分层方法和轨迹重关联的优化框架。该系统利用了一个由480个摄像头组成的模块化设计,分布在一个5.49 m的测地球体表面,能够捕捉社交群体的细微互动。论文中还介绍了一种新的3D运动重建算法,可以自动捕捉多达6人的互动,无需个体校准或标记。Varadarajan等人(2018)通过弱监督学习算法,基于身体姿势为主要线索推断对话群体,同时结合头部姿势进行优化。框架采用多个遮挡自适应回归函数,通过迭代优化策略,逐步提高对话群体和姿势估计的精度。

众多学者关注上下文视觉信息,通过设计新的结构上下文描述符来表达特征点的结构属性,它能够从宏观角度描述行人运动特征并结合运动与上下文相似性(Li等,2017)。他们往往通过自加权多视图聚类方法,按照其方向和上下文相似度进行聚类,无需人为设置参数或阈值即可自动确定群体数目(Wang等,2020)。还有一些研究假设社会关联的个体往往会相互注视,注视方向能反映其社交关系,结合运动特征和视觉注意特征,提出一种新的社交群体估计模型(Leach等,2014),这种方法尤其适用于人群密集、运动相似的场景,如机场、火车站等。但在拥挤环境下或群体成员分散时,基于这些特征的检测效果会受到限制。基于多层次一致行为特征的社会群体检测新方法结合了运动一致性与和视觉上下文信息。通过光流特征点和邻域信息,提取个体及其周围环境的结构特征描述符,以表达群体内的相互关系(Li等,2023b)。Ramírez等人(2016)提出一种链接方法(link method),根据从视频中提取的低级特征,采用学习和遗忘策略来建模个体之间的亲密关系动态。

Shuaibu等人(2017)提出一种创新的集体转换算法,该算法通过检测和分割人群场景,提取视觉描述符来量化群体属性。Varadarajan等人(2018)通过弱监督学习算法和交替优化策略,实现了头部和身体姿态以及对话群体的联合推断。Vascon和

Bazzani(2017)结合了低级和高级特征。低级特征包括个体的位置、速度和头部/身体方向,而高级特征则结合了社会学的概念,如视觉关注锥体(frustum of visual attention)。本文详细介绍了群体检测模型:鲁棒博弈论会话群体检测器(robust game theory session group detector, R-GTCG)。R-GTCG使用博弈论框架来识别群体,通过进化博弈的稳定策略(evolutionarily stable strategy, ESS)来确定群体的存在。

此外,还有基于数字图像处理的视觉内容方法,如根据多源图像特征(如局部二进制模式、傅立叶分析和灰度共生矩阵)的新方法,通过集成不同的外观特征来提高检测准确度。这种方法可以不依赖运动信息或ViBe背景减除技术,仅使用低级别的外观特征实现了高效的群体检测和分割(Basalamah和Khan,2020)。

### 1.2.2 ViBe背景减除法

ViBe背景减除算法用于从视频帧中分离出运动目标以获得清晰的轮廓信息,并提取视频中的运动特征(Pandey等,2020)。其用于群体检测的步骤为:1)使用ViBe算法进行背景减除,以提取视频中的前景对象,从而获得清晰的轮廓信息;2)通过阈值化轮廓面积和轮廓之间的距离,有效区分不同的群体和个体。

Vibe背景减除法能够有效地识别出前景中的静止区域,利用累积的静止时间信息,应用阈值分割方法来识别静止人群区域。随后,通过聚类算法,将这些静止像素聚合成不同的静止社交群体(Yi等,2016)。Chen和Lai(2019)使用Harris检测器提取群体图像中的显著点,以这些点为中心进行局部采样,生成群体的外观表示集合,利用Vibe背景减除法提取运动信息,并通过聚类分析形成群体整体描述符。最后,使用双向图匹配方法计算两个群体的最优匹配距离,作为群体匹配的相似度度量。

### 1.2.3 人机交互

本节总结了基于视觉内容的人机交互群体检测技术。以机器人为中心的群体感知模型RoboGEM通过自我中心的RGB-D数据(颜色和深度信息)进行无监督学习,并结合密集光流(optical flow)的行人运动和距离信息进行群体检测。进而通过层次聚类算法识别群体(Taylor等,2020)。REGROUP利用了机器人自身的RGB相机,通过追踪一检测方法实时检测并跟踪人群。系统引入了“人群指示特征”

(crowd indication feature, CIF),用于处理高密度人群中的遮挡问题,并利用卡尔曼滤波器(Kalman filter)对群体状态进行预测和更新(Taylor和Riek,2022)。

#### 1.2.4 其他应用领域

1)社会关系估计。Ding和Yilmaz(2011)提出一种利用视觉概念作为行为者社会互动指示器的新方法。利用大规模概念检测器生成概念得分向量来表示视频内容。这些得分进一步映射为分组线索,用于检测社会结构。作者采用了概率图模型结合时间平滑技术来分析行为者之间的社会关系和检测社区。

2)群体活动识别。Nabi等人(2013)提出一种语义驱动的时空描述符,能够在视频序列中有效地检测多人活动,并定位这些活动发生的时空位置。与传统的基于低层次特征的检测方法不同,它能够捕捉集体活动中人体姿势的时间动态特征。Zhou等人(2016)设计了一个4层结构生成模型,能够同时处理多个群体之间的交互和群体内个体之间的互动。该模型包含场景/事件、群体活动、标准姿势和可见姿势4个层次。该模型能够更稳健地应对姿势的遮挡和重叠问题,从而准确识别图像中的混合群体活动。

3)人群分析。通过观察参与者之间的社交互动来进行行人分组,具体包括言语和非言语的交流,如谈话、手势、目光接触、身体方向和空间排列等,用于分析这些群体的动态行为对人群流动和行走速度的影响(Gorrini等,2013)。

#### 1.2.5 总结

虽然收集各种视觉线索是描述群体属性的有效策略,但它总是存在缺陷聚类问题。与标准聚类任务不同,分组的定义在一定程度上是主观的。在大多数情况下,获取人群场景的先验信息是不切实际的,例如聚类数和阈值。这个问题阻碍了将许多经典聚类模型转移到组检测场景中。此外,考虑到标记数据在群体检测中的稀缺性,以无监督的方式联合利用各种特征是至关重要的。很多研究都倾向于通过不同的努力来整合特征,但都没有关注如何协商不同特征的一致部分和不一致部分。最优聚类模型应该能够以不同的方式区分和处理这两部分。在聚类过程中,要去除不一致的部分,保留一致的部分。

### 1.3 基于行为模式的方法

基于行为模式的方法是指通过嵌入合理的人类行为模式来预测人群中的互动,进而提高群体检测的准确性。这种行为模式有两种含义,一种是社交互动的本质行为模式。Rota等人(2015)认为个体行为与群体互动之间存在显著的模式差异,后者涉及到行动/反应原则,基于视频序列或图像中个人或人物动作之间的预定义行为来检测群体。另一种往往涉及人与环境的交互,以物理层面的社交环境信号为基础,利用与地点、事件、原因和因果等高维社会属性(Yin等,2013)相关的“社会力”进行行为模式表达。其中的代表理论为社会力模型(SFM),其他还有诸如认知语义学、社会一致性和时空能量等理论。

#### 1.3.1 社会力模型(SFM)

SFM广泛用于人群模拟,通过指定影响个体的吸引力和排斥力来模拟人群行为,捕捉行人之间的交互力(Leal-Taixé等,2011;Sochman和Hogg,2011;Mazzon等,2013)。SFM包含了一种吸引力,最初由Helbing和Molnár(1995)提出,是一种基于气体颗粒或流体运动与人群运动类比的行人模拟模型。它是一个数学模型,描述了每个个体的运动,考虑与其他个体和环境的相互作用力。假设每个个体*i*都受到(社会)力量 $F_i$ 的驱动。 $F_i$ 不是物理意义上的力(例如,没有作用和反作用力,即力可以是不对称的),但它被用做一个真实的力来更新个体*i*的位置 $x_i^t$ 和速度 $v_i^t$ 。社会力 $F_i$ 本身由4种力量组成,具体为

$$F_i(X^t, G^t) = F_i^0(X^t) + F_i^w(X^t) + F_i^r(X^t, G^t) + F_i^s(X^t, G^t) \quad (1)$$

式中, $F_i^0$ 是驱使个体走向预期目标的力, $F_i^w$ 是防止个体发生碰撞的排斥力, $F_i^r$ 是来自不同群体中其他个体的排斥力, $F_i^s$ 是使个体与同一群体中的其他人在一起的力。参数 $G^t$ 代表群聚类, $F_i(X^t, G^t)$ 表示群体聚类 $G^t$ 是未知变量, $X^t$ 表示位置数据。

SFM模型将个体视为粒子,通过计算这些粒子之间的交互力,来推断人群的整体行为模式(Wang等,2024)。Wang等人(2024)还通过光流技术获取视频中人群的运动信息,生成每个像素的运动向量。利用最后在视频帧上放置粒子网格,模拟人群的流动。通过光流和SFM模型分析人群的整体运动模式。Nunes等人(2016)还运用图论将SFM表示成社会图模型。Pang等人(2011)结合马尔可夫随机场

(Markov random field, MRF)和排斥力来模拟紧密目标之间的相互作用模式。这些模型通过马尔可夫链蒙特卡洛(Markov chain Monte Carlo, MCMC)粒子算法进行顺序推断,能够有效地检测和跟踪群体内的目标,并推断出正确的群体结构。

SFM模型能够很好地用于异常人群检测,其变体:社交属性感知力模型(social attribute-aware force model, SAFM)通过引入社交行为的社会属性,如社会混乱和拥挤程度,来增强对人群交互行为的描述。SAFM通过无监督的方式高效地估计场景规模,结合SFM在线融合社交属性,能够在全局和局部尺度上进行异常检测(Zhang等,2015a)。而反转SFM通过分析个体轨迹推断出导致此行为的社会力量及其关系(Šochman和Hogg,2011)。该方法特别适用于复杂场景中其他方法失效的情况下,尤其在人群密集的环境中表现出色。

另一种描述交互力的模型是2019年在*Nature*提出的“社交互动场模型(social interaction field model, SIFM)”(Zhou等,2019)。SIFM借鉴了物理学“场”的概念,利用虚拟现实技术描述虚拟场景中的人物空间特征(距离和方向)与视觉特征(姿态开放程度),用场强量化了社交场景中一对行人之间的交互强度。最后通过最大连通图识别静态/动态群体。研究发现,更近的人际距离、更直接的人际角度和更开放的身体姿势,会导致一个群体被判断为互动的可能性更高。

### 1.3.2 其他社会模型

1)事件认知语义。事件认知语义通过认知语言学的基本概念,例如“路径”、“地点”、“事物”、“动作”和“原因”来描述视频内容(Yin等,2013)。该方法通过计算同一“地点”中“事物”(即个体)的结构和语义距离,将相似的“事物”合并,减少语义社会熵,从而识别出人群。Dai等人(2008)提出一种基于事件的动态上下文模型,结合事件驱动的多层次动态贝叶斯网络,用于多层次事件的在线检测。该模型通过自底向上和自顶向下的推理机制,实现了上下文的动态感知和组交互分析。Yoo等人(2019)又结合情感相关性(如面部表情、性别和年龄等)来计算群体中个体的互动程度。

2)Granger因果关系。个体间的相互影响可以通过运动轨迹的因果关系来分析。Xie等人(2019)基于Granger因果分析,通过提取个体之间的运动模

式,并通过贝叶斯网络模型结合层次化的Dirichlet分配进行行为模式的表示和识别。Solera等人(2016)通过结构化支持向量机(structural SVM)在线学习人群成员之间的亲和性,灵感来源于亲近理论、Granger因果关系、动态时间规整(dynamic time warping, DTW)和热图,特定设计的特征结合了物理和社交身份。

3)社会一致性。传统的群体检测方法通常只依赖于人们的距离和运动方向,而忽略了社会群体形成的时间一致性特征。社会一致性特征描述了社会群体在空间和时间上的持续互动,在某一时间段内一群人会长时间保持一致的活动目标或运动状态(Zhang等,2015b;Huang等,2018a)。每种群体的活动状态是不同的,而群体内保持一致性(Nakamura等,2016)。为了高效地发现随时间演变的行人群体,Sawas等人(2018,2019)发现社交群体在空间和时间上具有共同的行为模式,并提出一种基于张量的群模式挖掘方法,通过发现随时间变化的行人对,利用这些对作为构建群体的基础,提出3种张量方法:locTgroups、globTgroups和timeWgroups。Li等人(2013)通过比较输入视频中的行为模式与预先定义的示例视频,设计了一个一致性匹配算法,从中识别出具有相似行为的个体小组。

2024年,清华大学学者Wu等人(2024)利用人类交互过程中的空间接近性和时间连续性计算行人之间的相互作用强度,构建具有时变特征的行人流网络。基于加权划分密度的目标函数,利用最优阈值确定帧图像中与人类子群体相对应的群体结构,达到了极高的检测精度。

4)核相关性。基于核相关性分析的方法通过匹配社会语义特征与输出标签,识别出最相关的特征(Fernández-Ramírez等,2020)。

5)社会行为图。传统方法常依赖于行人之间的距离来衡量互动强度,然而这种基于距离的方法无法有效处理复杂的社会行为。通过递归机制提取社会关系,并将其表示为社会行为图。由具有社会学背景的专家进行关系标注。专家根据视频数据判断行人之间的关系,不仅考虑物理特征(如速度、加速度、方向和相对距离),还结合了社会学知识(Sun等,2020)。

### 1.3.3 其他应用

近期,社会力模型多用于人群异常检测(Wang,

2024)。Zhang等人(2020)提出一种基于场景感知的异常人群检测方法,结合了流体力学模型和心理学理论。具体而言,该方法引入流体力学特征,如质量力和表面力来描述人群的运动状态,进而提高了异常检测的准确性。作者首先提出一种基于线积分卷积(LIC)的流场可视化技术来分割运动中的行人群体,接着通过场景感知引导的聚类策略将具有一致运动模式的行人群体进行聚类。该方法模拟了人类认知中的场景感知机制,合理利用了行人的各种属性信息,如位置、速度和运动方向等。

#### 1.3.4 总结

监督交互表示的方法(Choi等,2014;Solera等,2017)、预定义骨架节点序列(Joo等,2019)与情感属性(Yoo等,2019)等监督学习用于检测社会互动。为了在野外检测群体,这些方法大多可以较好地处理不同场景下的群体检测,但难以选择不同潜在变量的距离度量进行聚类,并且依赖于长期运动信息。然而,由于预先定义的群体社会行为,这些方法在检测多个活动群体方面表现出局限性。

#### 1.4 多模态方法

多模态方法不仅利用常规的空间与视觉特征,还辅助以音频数据流或传感器感知数据流进行更精确的互动感知。但语音识别技术与传感器技术往往针对室内场景,在大型公共场所中并不适用。因此此类研究也较少。

Kjærgaard等人(2012)提出一种利用多模态传感器融合技术,通过智能手机检测室内行人群体的系统。传统方法通常依赖单一传感器(如WiFi或加速度计),但这可能会导致准确性不足或对特定传感器的依赖。为此,他们通过融合WiFi、加速度计和指南针等多种传感器,提出一种基于层次聚类的行人群体检测方法。该方法的核心包括:首先从智能手机中的多个传感器收集数据(WiFi信号、加速度和指南针方向),然后通过分层聚类技术,基于这些传感器的特征值来检测群体。WiFi信号用于估算位置和距离,加速度用于捕捉相似的移动模式,指南针则用于比较相对的方向变化。还提出基于加权多数投票法的融合策略,将各个传感器的聚类结果进行融合,以提高检测的鲁棒性和准确性。

Bano等人(2018)利用可穿戴设备捕获的第一人称视角数据,包括视频、音频、GPS(global positioning system)和惯性传感器数据,以自动检测和时间

分割聚焦交互。聚焦交互被定义为两人或多人在共同关注下进行的面对面交流。其创新在于使用支持向量机(SVM)和长短时记忆网络处理音视频流,自动检测这些互动。

## 2 基于学习的群体检测方法

### 2.1 深度学习框架

传统的对话群体检测方法通常依赖于手工制定的启发式规则和数学模型,这些手工交互特征极具复杂性,手工标注也极其耗费人力物力(Thompson等,2021)。且在复杂多变的公共场景中,其泛化能力有限。随着深度学习的发展,基于深度学习的社会群体检测方法已成为近几年的研究热点,其根据模型架构大致可以分为4类:GNN、CNN、RNN/LSTM和Transformer。这些方法部分重叠了基于空间和行为的方法。早在2017年,就出现了利用结构化支持向量机(SVM)分类器学习适合群体实体的相似性权重度量(Solera等,2017)。Tzelepi和Tefas(2021)、Li等人(2023b)、Yoo等人(2019)和Kang等人(2023)都是利用基于CNN的ResNet-X架构从单个图像中学习高级特征,并将其聚合以获得用于社会群体检测的全局特征。考虑到同一群体中的人通常表现出相似的运动模式和空间布局,Hoang等人(2023)和Li等人(2023a)引入RNN或LSTM来测量人的轨迹相似性。同时,由于相互作用群具有图拓扑等空间结构,Tan等人(2022)、Thompson等人(2021)和Schmuck和Celiktutan(2021)使用GNN来探索群的空间关系。

#### 2.1.1 图神经网络(GNN)

图神经网络(GNN)通过学习节点(个体)和边(个体之间的关系)上的特征,可以更有效地检测复杂社交场景中的对话组。该方法首先将场景建模为一个交互图,图中的节点表示个体,边则表示成对关系。随后,GNN通过消息传递机制学习每对个体的亲和度形成亲和矩阵,亲和度表示两人互动的可能性(Thompson等,2021),最后再通过多层感知机(multilayer perceptron, MLP)或其他聚类算法完成群体聚类。与传统的基于距离的启发式方法相比,GNN在对话组检测方面表现更优。尤其是在多种输入特征(如位置、加速度和图像特征等)的条件下,GNN通过结合更多的个体特征,显著提高了群体检测的准确性和鲁棒性。特别是在复杂场景中,该方

法能够捕捉到多种特征之间的关系,从而更准确地划分对话组。

其中,著名的深度图聚类框架有 DANTE (deep affinity network), DANTE 通过空间特征(如位置和朝向)来预测两个人是否属于同一个对话群体。该网络利用图聚类框架,将成对的亲和度分数应用于图模型中的节点,然后通过图聚类算法识别对话群体 (Swofford 等, 2020)。DANTE 的创新之处在于它引入了上下文变换,能够处理包含多个个体的复杂场景,并通过多层感知器 (MLP) 网络计算个体之间的亲和度。此外, DANTE 还利用社会交互图中的其他个体信息,进一步提升了群体检测的准确性。在人一机器人互动的情境下, DANTE 提升了多达 9% 的准确率,能有效地识别对话中的成员及附近的旁观者。在此基础上, Thompson 等人利用 GNN 结合优势集算法对亲和度进行聚类,识别对话组。其 F1 score 显著高于传统方法和 DANTE 方法,尤其在融合了位置信息、加速度和图像特征时, GNN 表现出较强的优势 (Thompson 等, 2021)。

其次是基于 GNN 的深度学习框架 GROWL, 它利用群体内的空间配置而非亲和矩阵来捕捉个体间的互动关系 (Schmuck 和 Celiktutan, 2021)。该方法通过生成基于邻域特征的嵌入,进行节点之间的链接预测,以确定哪些个体属于同一社交群体。GROWL 方法包含两个主要步骤: 1) 基于个体的位置和方向数据构建社交图; 2) 使用图卷积网络生成个体的嵌入特征。这些嵌入被传递给一个浅层的二元分类器(如多层感知机 MLP),用于预测两个个体是否属于同一个群体,或结合卷积神经网络与 GNN 进行集成学习。

### 2.1.2 卷积神经网络(CNN)

一些研究直接将视频图像作为输入,他们利用 CNN 强大的图像学习能力,对图像的不同区域进行特征学习,直接在图像中实现人群划分。常见的用于社交建模的 CNN 学习模型有 I3D (inflated 3D) 网络与 C3D (convolutional 3D) 网络,如 Tzelepi 和 Tefas (2021) 针对无人机开发了一种轻量化的 CNN 模型。该方法能够使用全卷积神经网络 (fully convolutional networks, FCN) 生成概率热力图进行人群划分。并提出一种基于图形嵌入框架的正则化技术,以提升模型的泛化能力,防止过拟合。该方法不仅在无人机图像中的人群检测任务中表现优异,在其他分类

问题中也同样有效。

而 GNN 与 CNN 的组合模型图卷积神经网络 (GCN) 能够更好地学习群模式的空间结构信息 (Atici 等, 2022)。

Zhou 等人 (2022) 利用不同尺度的时空图卷积网络 (spatio-temporal graph convolutional network, STGCN) 来捕捉个体、群体和场景层次的信息,同时结合 LSTM 记忆历史信息。Grouptron 模型通过对行人进行时空聚类,构建了个体层、群体层和场景层的时空图。首先,在个体层中,使用 LSTM 网络对行人历史轨迹进行编码;其次,群体层通过 Hausdorff 距离将行人聚类为群体,进一步使用 STGCN 对群体内的动态进行编码;最后,场景层则对不同群体之间的交互进行建模。其他如改进的图卷积网络 (Lyu 等, 2024b) 与 Grouptron 动态多尺度 GCN 网络 (Zhang 等, 2024) 等学习节点表示以捕捉短期交互模式。

Lu 等人 (2023) 提出一种具有空间和姿态约束的双分支深度学习模型 (visual graph-driven temporal network, VGDTN)。在空间约束分支中,首先基于输入图像的场景俯视图构建交互图; VGDTN 预测交互图的亲和矩阵。通过对亲和矩阵应用优势集聚类 (Pavan 和 Pelillo, 2007) 进行初始群组检测。在姿态约束分支中,估计输入图像中人的三维姿态(如站立、坐在物体上或地面),并利用姿态分类器识别个体姿态。最后,根据人的姿态和初始群体划分对群体检测进行了改进。

### 2.1.3 深度上下文学习模型

最近的研究通过循环神经网络,尤其是长短期记忆神经网络 (LSTM) 来捕捉历史图像帧的群体结构信息和交互信息 (Tan 等, 2022)。随着自注意力机制 (self-attention) 的兴起,更多的学者将 Transformer 架构用于建模社交互动的时空连续性 (Li 等, 2022)。这些网络结构更契合行为特征的时序性与场景上下文。动态长短期记忆网络模型能够学习连续的成对亲和度预测,捕捉人与人之间的社交动态。通过使用位置、头部和身体朝向等时空特征,生成连续的亲和度值,并将这些亲和度用于后续的图聚类算法,从而识别会话组 (Tan 等, 2022)。该方法在具有较高时间分辨率的数据集(如 Confab)上表现尤为出色, F1 score 达到 0.73, 显著优于现有基线模型。

为了克服传统群体检测方法中对手工特征依赖和标注数据需求高的局限性, Li 等人 (2022) 在

ECCV(European Conference on Computer Vision)会议发文提出一种基于自监督学习的多头注意力框架。研究的核心问题是如何表示和划分人类的社会关系。本文设计了一个两阶段的框架:第1阶段使用行为模拟器头(behavioral simulation head, BS-Head)学习社会关系特征表示,利用自监督训练从多人的行为关系中提取社会特征;第2阶段基于所学的社会关系嵌入,使用自注意力网络进行人群群体检测。Tamura 等人(2022)认为社交群组活动识别需要识别多个子群组的活动并确定群组成员。并在 ECCV 发文提出一种基于 Transformer 的框架,其中注意力模块自动识别并聚合与社交群组活动相关的特征,生成每个社交群组的有效特征。通过嵌入群组成员信息,能够通过简单的 Hungarian 匹配算法精确识别群组成员。Yu 等人(2024)设计了一种基于 Transformer 的环境感知机制,整合视觉线索特征,并构建了互动强度矩阵,表示个体之间的互动强度。

调查发现,在 2020—2024 年间,许多学者将 GNN、CNN 与注意力机制结合使用,同时捕捉空间模式与时序上下文信息(李欢和陈念年,2023),自注意力模块用于捕捉个体动作的细节,而图注意力模型(graph attention networks, GAT)则用于建模个体之间的相互关系,从而检测出社交群体及其活动。有研究者同时结合 I3D 骨干网络进行视觉内容的提取(Ehsanpour 等,2020),但与这种大型网络的集成严重阻碍了它们的实时性。因此,后续研究普遍至多集成两种深度学习网络。如社交群组长短期记忆网络(social group long short-term memory network, SG-LSTM)(Bhaskara 等,2023),SG-LSTM 使用 RGB 和深度图像作为输入,通过卷积神经网络(CNN)进行群体检测,生成群体的空间坐标。通过引入社交池化层,SG-LSTM 能够捕捉到群体内部以及群体之间的互动关系。该模型特别适用于密集的场景,能够有效处理群体中的遮挡和复杂的互动;以及 Li 等人(2023a)在 Neurocomputing 发文提出的基于关系存在识别的社会群体检测框架。基于视觉线索的组件利用“CNN + LSTM”组合和多模态特征提取(multimodal feature extraction, MFE)框架 MFE 制挖掘时空信息。在基于非视觉线索的组件中,利用 ResNet-101(residual network-101)将每个轨迹对的相似性作为额外信息进行测量。最后,将这些信息进行互补融合,对社会关系的存在进行分类。

针对大型网络产生的时间/计算复杂度问题,Jahangard 等人(2023)在 ICCV(International Conference on Computer Vision)会议上提出基于 LSTM 和图 Transformer 的轻量级框架,仅使用个体的运动轨迹作为输入数据,并通过图聚类实现社交群体的检测。该模型通过长短期记忆网络(LSTM)编码个体轨迹,将这些编码特征作为图的节点,并利用各个轨迹之间的几何距离作为边缘特征。该框架的核心创新在于使用图 Transformer 更新节点和边缘特征,进而通过图聚类检测出不同的社交群体。该模型在计算速度和推理效率方面表现突出,平均精度均值(mean average precision, mAP)比现有方法提升了 2%~11%。此外,在相同的计算资源下,该模型的推理速度是现有方法的 12 倍,达到实时检测的要求。

最近的研究在现有的 DANTE(Swofford 等,2020)的基础上进行了改进(Nasri 等,2024),提出一种名为 T-DANTE 的深度学习模型。该方法通过引入 RNN、LSTM 和门控循环单元(gated recurrent unit, GRU)来捕捉轨迹数据中的时间动态特征。T-DANTE 结合了个体的时空轨迹与周围环境的上下文信息,构建了一个基于图的社会亲和网络,通过学习个体之间的相互影响和空间关系来识别社会群体。

## 2.2 其他应用领域

1)无人机飞行安全。由于无人机的计算能力限制,传统的深度学习模型可能不适用。因此,Tzelepi 和 Tefas(2021)设计了轻量级的卷积神经网络(CNN)架构,能够有效区分拥挤与非拥挤场景。并提出一种基于图嵌入框架的正则化技术,以防止过拟合并增强模型的泛化能力。

2)群体重识别。Huang 等人(2021)提出“配对表示学习”,即通过捕捉组内成员的邻域关系来增强群体重识别的判别力。该模型不仅学习单个成员的特征,还学习组内成员之间的关系特征,并通过在线加权融合这些特征,最终生成跨摄像机的更精确的群体表达。Yan 等人(2023)提出一种基于图神经网络(GNN)的多注意力上下文图谱(multi-attention contextual graph, MACG)模型,MACG 模型通过构建上下文图谱,将群体成员表示为图节点,并采用多层次的注意力机制来捕捉组内和组间的上下文信息。此外,使用自注意力模块进一步聚合节点级别特征,从而生成稳健的图级别表示。

3)人群轨迹预测。Zhang 等人(2024)提出一种

名为 Grouptron 的动态多尺度图卷积网络。Grouptron 结合动态图建模和时空特征,能够捕捉到群体内个体的细微移动,并根据这些变化预测未来的轨迹。它通过引入注意力机制动态调整各层之间的信息传递,确保重要的时空关系得到突出显示。模型利用轨迹数据,包括位置、速度和加速度等参数,通过多尺度特征融合实现群体感知。

4) 人群分析。Wang 等人(2024)提出一种名为群体互动场(group interaction field, GIF)的新模型,用于学习和解释行人在群体中的预期行为。GIF 通过捕捉多维观察(包括轨迹、视觉方向以及群体互动状态等)来量化群体行为的影响,并提出一个名为 GIFNet 的神经网络模型用于估算 GIF。GIFNet 的核心在于通过图注意力网络(graph attention network, GAT)动态传播群体交互信息,并根据群体规模和动态互动状态自适应地调整模型。

5) 人群异常检测。Lin 等人(2019)提出一种名为社交多实例学习(social mil)的新框架,用于在人群异常检测中引入交互感知。利用异常人群行为的先验知识,通过双分支网络结构来检测语义层面的异常:一个是用于提取全局空间和时间动态信息的空间—时间动态网络分支;另一个是通过社交力图提取人群交互信息的交互动态网络分支。此外,引入了一维依赖捕获注意力模块,增强 C3D 网络在特征提取上的表现。

### 2.3 总结

近几年,深度学习网络在学习交互特性和群体构特性上取得巨大的成效。虽然这些基于深度学习(deep learning, DL)的方法可以在检测交互组方面取得良好性能,但对于广泛的监控视频分析,基于深度学习的方法是计算密集型的,并且需要强大的硬件进行训练和推理(Lu 等, 2023)。因此,有必要开发既高效又有效的轻量级模型,用于现实世界中的监视分析的实际应用,未来需要探索更多轻量型的网络架构。

## 3 数据集

社交数据集提供了人们在现实生活中实际如何相互交往的丰富而真实的表示。这些微妙的协调模式受到个人信念、目标和愿望的影响,这些愿望与个

人在日常生活中进行的活动中可能失去或获得的东西有关。这些条件在实验室环境中不容易复制,需要彻底重新考虑如何收集和收集什么。Hung 等人(2019)提供了在全面考虑整个实验设计和数据收集过程时如何创建此类多模态多传感器数据集的指南。表 2 综合列出了所有文献用于群体检测的数据集,共涉及 48 个数据集,包括数据集名称、摄像机视角、室内/室外拍摄以及公开信息。Barua 等人(2024)将摄像机定位视角分为两种不同类型:1)用于机器人的自我视觉/自我视角(以自我为中心)摄像机,称为“Ego”。自我中心指第一人称视角;例如,可穿戴相机或机器人相机拍摄的图像或视频;2)用于室内或室外环境的外视觉/外视角摄像机,通常指由监视/监控摄像机捕获的图像或视频,称为“Exo”。表 2 沿用这一分类命名。其中 35 个是公开的,12 个是作者/研究人员的自建私有数据集,7 个是未知的。37 个数据集从室外公共场景采集,16 个数据集从室内场景采集,只有 5 个数据集有两种场景。较为常用的数据集有 SALSA(social affective and leadership spatio-temporal analysis)、BIWI(biwi kinect head pose database)、CoffeeBreak、CUHK Crowd(chinese university of Hong Kong crowd dataset)、Social-CAD(social context-aware anomaly detection dataset)、CAC(collective activity classification)、ETH(ETH pedestrian dataset)、CBE(crowds-by-examples)以及 Poster Data 等。

数据收集本身面临巨大的挑战,特别是在采集过程中涉及到他人个人信息和肖像权等伦理问题和影响。隐私和视觉数据的保护显然受到威胁。此外,数据集还可能存在潜在的偏见和歧视问题,即未包含代表性不足的社区的数据。在当今世界,伦理问题主要体现在年龄、种族和性别等因素的多样性和包容性方面。目前,大多数数据集源于西方国家的场景。理想情况下,数据集应包含来自全球不同社区的数据,包括亚洲和非洲国家,并在视觉数据中同时纳入黑人和白人肤色人群。代表性不足会导致模型预测错误的类别。考虑到社会环境中的互动通常不包括婴儿,应在数据采集时关注青少年、成年人与中老年人的恰当比例。年龄方面的多样性同样重要,青少年和成年人的组合相当普遍。最后,性别的平衡也是一个值得关注的问题,现有数据集中男性比例远高于女性,这可能会影响公平性。

表 2 用于群体检测的数据集汇总

Table 2 Summary of datasets for social group detection

| 数据集                | 使用该数据集的检测方法                                                                                                                                                                                                                                    | 相机<br>视角 | 室外/室内 | 公开/<br>私有 |
|--------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------|-------|-----------|
| Meeting room       | Pang 等人(2008)                                                                                                                                                                                                                                  | Exo      | 室内    | 私有        |
| CoffeeBreak        | Setti 等人(2013a, 2013b); Vascon 和 Bazzani(2017); Varadarajan 等人(2018); Lu 等人(2023)                                                                                                                                                              | Exo      | 室外    | 公开        |
| Cocktail Party     | Setti 等人(2013b, 2015); Vascon 和 Bazzani(2017); Varadarajan 等人(2018); Swofford 等人(2020); Tan 等人(2022); Thompson 等人(2021)                                                                                                                        | Exo      | 室内    | 公开        |
| SALSA              | Islas 等人(2016); Inaba 和 Aoki(2016); Varadarajan 等人(2018); Zhang 和 Hung(2021); Zhou 等人(2019); Swofford 等人(2020); Schmuck 和 Celiktutan(2021); Tan 等人(2022); Atıcı 等人(2022); Lu 等人(2023); Chen 和 Liang(2024); Ramírez 等人(2016); Hedayati 等人(2020) | Exo      | 室内    | 公开        |
| ETH                | Mazzon 等人(2013); Nakamura 等人(2016); Nasri 等人(2024); Šochman 和 Hogg(2011); Yücel 等人(2013); Zhang 等人(2024)                                                                                                                                       | Exo      | 室外    | 公开        |
| HOTEL              | Mazzon 等人(2013); Nakamura 等人(2016); Nasri 等人(2024); Šochman 和 Hogg(2011); Yücel 等人(2013); Chandran 等人(2015); Zhang 等人(2024)                                                                                                                    | Exo      | 室外    | 公开        |
| Student003         | Mazzon 等人(2013); Nakamura 等人(2016); Nasri 等人(2024); Šochman 和 Hogg(2011)                                                                                                                                                                       | Exo      | 室外    | 公开        |
| zara01             | Nasri 等人(2024); Zhang 等人(2024)                                                                                                                                                                                                                 | Exo      | 室外    | 公开        |
| VideoWeb           | Qin 等人(2013)                                                                                                                                                                                                                                   | Exo      | 室外    | 公开        |
| Caviar             | Yücel 等人(2013); Chandran 等人(2015); Feng 和 Bhanu(2015)                                                                                                                                                                                          | -        | 室内    | 公开        |
| BIWI               | Yücel 等人(2013); Solera 和 Calderara(2013); Zhang 等人(2015b); Chandran 等人(2015); Nakamura 等人(2016); Vascon 和 Bazzani(2017); Cheng 等人(2019); Fernando 等人(2019); Zhao 等人(2020b)                                                                     | Exo      | 室外    | 私有        |
| PETS2009           | Feng 和 Bhanu(2015)                                                                                                                                                                                                                             | Exo      | 室外    | 私有        |
| UNIV               | Feng 和 Bhanu(2015); Zhang 等人(2024)                                                                                                                                                                                                             | Exo      | 室外    | 公开        |
| APT                | Yücel 等人(2013)                                                                                                                                                                                                                                 | -        | 室内    | 公开        |
| BIT-Interaction    | Kong 等人(2012); Xie 等人(2019)                                                                                                                                                                                                                    | Exo      | 室外    | 私有        |
| UT-Interaction     | Kong 等人(2012); Xie 等人(2019)                                                                                                                                                                                                                    | Exo      | 室外    | 私有        |
| CUHK Crowd Dataset | Kjærgaard 等人(2012); Shao 等人(2015); Zhao 等人(2016); Voon 等人(2016); Chen 等人(2017); Li 等人(2017)                                                                                                                                                    | Exo      | 室外    | 公开        |
| PETS 2007          | Leach 等人(2014)                                                                                                                                                                                                                                 | Exo      | 室外    | 私有        |
| Oxford             | Leach 等人(2014)                                                                                                                                                                                                                                 |          | 室外    | 公开        |
| CBE                | Solera 和 Calderara(2013); Zhang 等人(2015b); Nakamura 等人(2016); Zhao 等人(2020b)                                                                                                                                                                   | Exo      | 室外    | 公开        |
| GVEII              | Zanotto 等人(2012); Solera 等人(2016); Zhao 等人(2020a)                                                                                                                                                                                              | -        | 室外    | 公开        |
| MPT-20x100         | Solera 等人(2016)                                                                                                                                                                                                                                | -        | 室外    | 公开        |
| Poster Data        | Vascon 和 Bazzani(2017); Setti 等人(2013a, 2015); Zhang 和 Hung(2021)                                                                                                                                                                              | Exo      | 室内    | 私有        |
| Social-CAD         | Li 等人(2023a); Yu 等人(2024); Nabi 等人(2013); Yu 等人(2024)                                                                                                                                                                                          | Exo      | 室外    | 公开        |
| PLPS               | Li 等人(2023b)                                                                                                                                                                                                                                   | Exo      | 室外    | 公开        |
| DukeMTMC           | Huang 等人(2021)                                                                                                                                                                                                                                 | Exo      | 室外    | 公开        |
| Road Group         | Huang 等人(2021)                                                                                                                                                                                                                                 | Exo      | 室外    | 公开        |
| JRDB-Act           | Ehsanpour 等人(2022), Jahangard 等人(2023)                                                                                                                                                                                                         | Exo      | 室内和室外 | 部分公开      |
| VSID               | Yu 等人(2024)                                                                                                                                                                                                                                    | -        | 室内    | 公开        |
| CSG                | Yan 等人(2023)                                                                                                                                                                                                                                   | Exo      | 室外    | 公开        |
| BEHAVE dataset     | Yin 等人(2013)                                                                                                                                                                                                                                   | Exo      | 室外    | 公开        |

注:“Exo”指用于室内或室外环境的外视觉/外视角摄像机,“-”表示此项信息未知。

续表 2 用于群体检测的数据集汇总  
Continued Table 2 Summary of datasets for social group detection

| 数据集                        | 使用该数据集的检测方法                                                   | 相机<br>视角 | 室外/室内 | 公开/<br>私有 |
|----------------------------|---------------------------------------------------------------|----------|-------|-----------|
| CAC                        | Zhou 等人(2016); Yin 等人(2013); Xie 等人(2019); Ehsanpour 等人(2020) | Ego      | 室内和室外 | 公开        |
| Volleyball                 | Ehsanpour 等人(2020)                                            | Exo      | 室内    | 公开        |
| PETS                       | Zitouni 等人(2020)                                              | Exo      | 室外    | 私有        |
| UCF-Crime                  | Zitouni 等人(2020); Basalamah 和 Khan(2020)                      | Exo      | 室外    | 公开        |
| MatchNMingle               | Thompson 等人(2021)                                             | Exo      | 室内    | 公开        |
| RICA                       | Schmuck 和 Celiktutan(2021)                                    | -        | 室内    | 公开        |
| Social Interaction Dataset | Pathi 等人(2022); Tan 等人(2022)                                  | -        | 室内和室外 | 私有        |
| PANADA                     | Li 等人(2022)                                                   | Exo      | 室外    | 公开        |
| JRDB-Group                 | Li 等人(2022)                                                   | Exo      | 室内和室外 | 公开        |
| Confab                     | Tan 等人(2022)                                                  | Exo      | 室内    | 部分公开      |
| SCU-VSD-Social             | Su 等人(2022)                                                   | Exo      | 室外    | 私有        |
| Visdrone2019               | Yang 等人(2022)                                                 | Exo      | 室外    | 私有        |
| B-SDA                      | Yang 等人(2022)                                                 | Exo      | 室外    | 私有        |
| SGD                        | Lu 等人(2023)                                                   | Exo      | 室内和室外 | 公开        |
| IGD                        | Lu 等人(2023)                                                   | Exo      | 室外    | 公开        |
| MOT17                      | Chen 等人(2023)                                                 | Exo      | 室外    | 公开        |
| KITTI                      | Chen 等人(2023)                                                 | Exo      | 室外    | 公开        |

注:“Ego”指用于机器人的自我视觉/自我视觉摄像机,“-”表示此项信息未知。

4 评价指标与性能

表 3 列出了公共场所群体的常用评价指标以及 2023—2024 年部分研究所能达到的 SOTA (state of the arts) 效果。常用指标有精度 (Precision, P), 召回率 (Recall, R), F1 score (F1), mAP 等。可以看出,在简单数据集中,2024 年最新研究 (Wu 等, 2024) 报告的群体检测 Recall 能达到 93.9%, 这是一种基于行为模式的行人流动态网络,在相同数据集 (ETH 与 Hotel) 下,超过了深度学习算法 (Nasri 等, 2024)。其他大部分模型/框架仅能达到 80% 左右的性能,在复杂情景数据集,如频繁拥挤、遮挡情况下,情况还会更糟。因此,群体检测相关研究至今依然面临社交行为表达不精确与检测精度不足的困境。

5 限制、挑战与未来方向

目前的方法分为启发式规则和基于学习的两大类。但群体现象是复杂的,仍然缺乏正式的规则和

精确的社会解释。研究者需更多地关注深度学习和强化学习范式,以利用数据驱动的方法应对复杂情况。但在数据量有限的条件下,这一目标难以实现。

1) 小样本学习。现有数据集的规模和质量尚不足以满足这种数据需求。研究领域亟需持续性地投入精力,创建并发布更多数据集,同时全面考虑伦理和生态问题,这将为今后的数据集建议和改进奠定基础。对于当前小数据集环境,需探索对数据依赖性较低的机器学习技术,使少量数据的利用效率最大化。此类情况适合采用统计模型,因其在数据有限时表现最佳。同时,可以尝试少样本学习、一次学习和增量学习等范式,以克服数据不足的挑战。

2) 图神经网络。采用图结构表示数据可以显著提升模型的准确性和效率。近年来,图神经网络 (GNN) 已在该领域取得显著进展,全球多家研究机构已发表相关成果。将 GNN 与其他深度学习方法结合,如通过图卷积网络 (GCN) 探索 GNN 和卷积神经网络 (CNN) 的混合模型,有望创建更适合的检测模型。进一步的组合如递归神经网络 (recurrent neural network, RNN)、卷积神经网络 (CNN) 和

表 3 不同算法在多种评价指标下的性能  
Table 3 Performance of different algorithms under various evaluation metrics

| 算法                 | 数据集               | P           | R           | F1          | mAP         |
|--------------------|-------------------|-------------|-------------|-------------|-------------|
| Zhang 等人(2024)     | Person Trajectory | -           | -           | 83.2        | <b>86.4</b> |
| Wu 等人(2024)        | ETH               | <b>88.6</b> | 89.6        | <b>89.1</b> | -           |
|                    | Hotel             | 82.3        | <b>93.9</b> | 87.7        | -           |
| Nasri 等人(2024)     | ETH               | 79.3        | 81.5        | -           | -           |
|                    | Hotel             | 74.8        | 79.0        | -           | -           |
| Jahangard 等人(2023) | JRDBAct           | -           | -           | -           | 65.9        |
| Li 等人(2023a)       | CoffeeBreak       | 77          | -           | 78          | -           |
| Lu 等人(2023)        | CoffeeBreak       | 68          | 69          | 89          | -           |

注:加粗字体表示各列最优结果,“-”表示此项信息未知。

图 Transformer 网络等也带来了更高的检测精度。因此,图神经网络会得到持续发展。在此基础上,元学习有望在大规模复合数据集上展开应用,并通过大数据和可视化分析简化复杂情境的检测任务。

3)轻量级深度学习网络。基于深度学习的方法是计算密集型的,在处理海量的监控视频数据时需要强大的硬件进行训练和推理。因此,有必要开发既高效又有效的轻量级模型,用于现实世界中的监视分析的实际应用。未来需要探索更多轻量型的网络架构。

4)行为模式挖掘。基于行为模式的分析方法不容小觑。事实证明,若能正确建模社交行为的本质互动模式,这种方法可以在小样本条件下达到超越深度学习方法的最优性能(Wu 等,2024),同时能够避免烦琐的训练任务。因此,未来应持续探索人类的社交行为模式。

5)隐秘社团发现。对于异常群体检测,不仅要关注打架、斗殴等已经构成犯罪的犯罪团伙,还应关注如何在事件发生前就能根据异常模式识别异常人群。在犯罪率高涨的今天,这项研究对于提升公共安全意义重大。

在异常行为尚未发生的隐秘交互桥段,犯罪团伙的行为往往与正常人无异,他们之间存在的是无明显身体交互的隐式交互。目前公共场所社交行为理解及群体检测仅关注显式交互,在隐秘的异常人群检测与人群异常检测中,这些基于显式空间特征与视觉特征(包括轨迹相似性、位置、身体方向和身

体姿态)的方法会完全失效。而深度学习方法也很难直接学习隐性模式。目前针对隐秘尾随行为检测已经成功提出一系列隐秘尾随行为模式的建模方法(Xu 等,2021,2022; Li 等,2024)。

因此,未来研究趋势应聚焦于挖掘多种场景下的全域社交行为模式,尤其是隐秘社交行为与正常社交行为的模式差异,探究隐秘社交的决定性属性特征,进而探索隐秘社团检测方法,填补视频人类行为理解与群体检测在隐匿场景下的理论空缺。

6 结 语

随着计算机视觉、机器人技术和多媒体分析的发展,群体互动的检测与预测逐渐引起广泛关注。然而,该领域面临着两个主要挑战:1)如何有效地表示紧凑且有意义的社交互动特征;2)如何利用这些特征实现高效的群体聚类 and 检测。尽管在过去的 20 年中已经取得了一些进展,但比起如“群体行为识别”方向的研究来说依然很少。本文旨在建立一个系统框架来概述群体和互动检测的问题,这也构成了本文的核心主题。本文全面探讨了这一框架涉及各个方面,包括群体的定义、相关数据集、算法、群体检测应用领域、评估标准和检测能力。同时,还分析了该领域的局限性、面临的挑战及未来的研究方向。研究者们可以利用启发式规则方法和深度学习等先进技术,探索一些此前未被关注的问题,例如隐秘犯罪团体的识别。此外,结合不同类型的神经

网络,也可以在特定应用场景中开发出有效的轻量级模型。

**致谢:** 本文由中国图象图形学学会图像视频通信专业委员会组织撰写,该专业委员会链接为 <https://www.csig.org.cn/16/201612/49312.html>。

## 参考文献(References)

- Alahi A, Ramanathan V and Li F F. 2014. Socially-aware Large-scale Crowd Forecasting. Proceedings 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA: IEEE: 1-9 [DOI: 10.1109/CVPR.2014.76]
- Al Masum A, Rafy M H and Rahman S M M. 2014. Video-based affinity group detection using trajectories of multiple subjects//Proceedings of the 8th International Conference on Electrical and Computer Engineering. Dhaka, Bangladesh: IEEE: 120-123 [DOI: 10.1109/ICECE.2014.7026834]
- Atıcı E, Gökberk B and Akarun L. 2022. Detection of free-standing conversational groups with graph convolutional networks//Proceedings of the 26th International Conference on Pattern Recognition. Montreal, Canada: IEEE: 1041-1047 [DOI: 10.1109/ICPR56361.2022.9956536]
- Bano S, Suveges T, Zhang J and McKenna S J. 2018. Multimodal ego-centric analysis of focused interactions. IEEE Access, 6: 37493-37505 [DOI: 10.1109/ACCESS.2018.2850284]
- Bano S, Zhang J and McKenna S J. 2017. Finding time together: detection and classification of focused interaction in egocentric video//Proceedings of 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), Venice, Italy, 2322-2330 [DOI: 10.1109/ICCVW.2017.274]
- Barua H B, Pramanick P, Sarkar C and Mg T H. 2020. Let me join you! Real-time F-formation recognition by a socially aware robot//Proceedings of the 29th IEEE International Conference on Robot and Human Interactive Communication. Naples, Italy: IEEE: 371-377 [DOI: 10.1109/RO-MAN47096.2020.9223469]
- Barua H B, Mg T H, Pramanick P and Sarkar C. 2024. Enabling social robots to perceive and join socially interacting groups using F-formation: a comprehensive overview. ACM Transactions on Human-Robot Interaction, 13(4): #48 [DOI: 10.1145/3682072]
- Basalamah S and Khan S D. 2020. Pedestrian crowd detection and segmentation using multi-source feature descriptors. International Journal of Advanced Computer Science and Applications, 11(1): 707-713 [DOI: 10.14569/IJACSA.2020.0110187]
- Bhaskara R, Chiu M and Bera A. 2023. SG-LSTM: social group LSTM for robot navigation through dense crowds//Proceedings of 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems. Detroit, USA: IEEE: 3835-3840 [DOI: 10.1109/IROS55552.2023.10341954]
- Bisagno N, Zhang B and Conci N. 2019. Group LSTM: group trajectory prediction in crowded scenarios//Proceedings of the Computer Vision—ECCV 2018 Workshops. Munich, Germany: Springer: 213-225 [DOI: 10.1007/978-3-030-11015-4\_18]
- Bozorgtabar B and Goecke R. 2015. Dominant interaction group detection in team sports//Proceedings of the 21st ACM SIGKDD Conference on Knowledge Discovery and Data Mining. [s.l.]: [s.n.]: 1-4
- Chandran A K, Poh L A and Vadakkepat P. 2015. Identifying social groups in pedestrian crowd videos//Proceedings of the 8th International Conference on Advances in Pattern Recognition. Kolkata, India: IEEE: 1-6 [DOI: 10.1109/ICAPR.2015.7050677]
- Chen M L, Wang Q and Li X L. 2017. Anchor-based group detection in crowd scenes//Proceedings of 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). New Orleans, USA: IEEE: 1378-1382 [DOI: 10.1109/ICASSP.2017.7952382]
- Chen M Z Y, Banitaan S and Maleki M. 2023. Enhancing pedestrian group detection and tracking through zone-based clustering. IEEE Access, 11: 132162-132179 [DOI: 10.1109/ACCESS.2023.3336592]
- Chen M Z Y, Banitaan S, Maleki M and Li Y C. 2022. GDSCAN: pedestrian group detection using dynamic epsilon//Proceedings of the 21st IEEE International Conference on Machine Learning and Applications. Nassau, Bahamas: IEEE: 1748-1753 [DOI: 10.1109/ICMLA55696.2022.00267]
- Chen V K and Liang C. 2024. Persistent homology for capturing social structure and cohesion of F-formation groups//Proceedings of 2024 IEEE International Conference on Robotics and Automation. Yokohama, Japan: IEEE [DOI: 10.1109/ICRA.2024.9376589&#820]
- Chen X H and Lai J H. 2019. Salient points driven pedestrian group retrieval//Proceedings of the 10th International Conference on Image and Graphics. Beijing, China: Springer: 644-656 [DOI: 10.1007/978-3-030-34120-6\_52]
- Cheng H, Li Y and Sester M. 2019. Pedestrian group detection in shared space//Proceedings of 2019 IEEE Intelligent Vehicles Symposium (IV). Paris, France: IEEE: 1707-1714 [DOI: 10.1109/IVS.2019.8813849]
- Choi W, Chao Y W, Pantofaru C and Savarese S. 2014. Discovering groups of people in images//Proceedings of the 13th European Conference on Computer Vision — ECCV 2014. Zurich, Switzerland: Springer: 417-433 [DOI: 10.1007/978-3-319-10593-2\_28]
- Cristani M, Bazzani L, Paggetti G, Fossati A, Tosato D, Del Bue A, Menegaz G and Murino V. 2011. Social interaction discovery by statistical analysis of F-formations//Proceedings of 2011 British Machine Vision Conference. Dundee, UK: BMVC Press: 23.1-23.12 [DOI: 10.5244/C.25.23]
- Cristani M, Raghavendra R, Del Bue A and Murino V. 2013. Human behavior analysis in video surveillance: a social signal processing perspective. Neurocomputing, 100: 86-97 [DOI: 10.1016/j.neu-

- com.2011.12.038]
- Dai P, Di H J, Dong L G, Tao L M and Xu G Y. 2008. Group interaction analysis in dynamic context. *IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics*, 38 (1) : 275-282 [DOI: 10.1109/TSMCB.2007.909939]
- Dim E and Kuflik T. 2010. Social signal processing: detecting small group interaction in leisure activity//*Proceedings of the 15th International Conference on Intelligent User Interfaces*. Hong Kong China: ACM: 309-312 [DOI: 10.1145/1719970.1720018]
- Dim E and Kuflik T. 2013. Social F-formation in blended reality//*Proceedings of 2013 International Conference on Intelligent User Interfaces*. Santa Monica, USA: 25-28
- Ding L and Yilmaz A. 2011. Inferring social relations from visual concepts//*Proceedings of 2011 International Conference on Computer Vision*. Barcelona, Spain: IEEE: 699-706 [DOI: 10.1109/ICCV.2011.6126306]
- Ehsanpour M, Abedin A, Saleh F, Shi J, Reid I and Rezatofighi H. 2020. Joint learning of social groups, individuals action and subgroup activities in videos//*Proceedings of the 16th European Conference on Computer Vision — ECCV 2020*. Glasgow, UK: Springer: 177-195 [DOI: 10.1007/978-3-030-58545-7\_11]
- Ehsanpour M, Saleh F, Savarese S, Reid I and Rezatofighi H. 2022. JRDB-Act: a large-scale dataset for spatio-temporal action, social group and activity detection//*Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, USA: IEEE: 20951-20960 [DOI: 10.1109/CVPR52688.2022.02031]
- Favaretto R M, de Andrade Araujo V F, Vilanova F, Costa A B and Musse S R. 2021. Cultural behaviors analysis in video sequences. *Machine Vision and Applications*, 32 (4) : #103 [DOI: 10.1007/s00138-021-01225-2]
- Favaretto R M, Dohl L L and Musse S R. 2016. Detecting crowd features in video sequences//*Proceedings of the 29th SIBGRAPI Conference on Graphics, Patterns and Images*. Sao Paulo, Brazil: IEEE: 201-208 [DOI: 10.1109/SIBGRAPI.2016.036]
- Favaretto R M, Knob P, Musse S R, Vilanova F and Costa Â B. 2019. Detecting personality and emotion traits in crowds from video sequences. *Machine Vision and Applications* 30: 999-1012 [DOI: 10.1007/s00138-018-0979-y]
- Feng L N and Bhanu B. 2015. Understanding dynamic social grouping behaviors of pedestrians. *IEEE Journal of Selected Topics in Signal Processing*, 9(2) : 317-329 [DOI: 10.1109/JSTSP.2014.2365765]
- Fernández-Ramírez J, Álvarez-Meza A, Pereira E M, Orozco-Gutiérrez A and Castellanos-Dominguez G. 2020. Video-based social behavior recognition based on kernel relevance analysis. *The Visual Computer*, 36(8) : 1535-1547 [DOI: 10.1007/s00371-019-01754-y]
- Fernando T, Denman S, Sridharan S and Fookes C. 2019. GD-GAN: generative adversarial networks for trajectory prediction and group detection in crowds//*Proceedings of the 14th Asian Conference on Computer Vision — ACCV 2018*. Perth, Australia: Springer: 314-330 [DOI: 10.1007/978-3-030-20887-5\_20]
- Ferrera-Cedeño E, Acosta-Mendoza N, Gago-Alonso A and García-Reyes E. 2019. Detecting free standing conversational group in video using fuzzy relations. *Informatica*, 30(1) : 21-32 [DOI: 10.15388/Informatica.2019.195]
- Gao J J and Qing L B. 2019. Survey on social relationship recognition based on images. *Computer Engineering and Applications*, 55(21) : 36-45 (高建军, 卿粼波. 2019. 基于图像的社会关系识别研究综述. *计算机工程与应用*, 55(21) : 36-45) [DOI: 10.3778/j.issn.1002-8331.1908-0308]
- Ge W N, Collins R T and Ruback B. 2009. Automatically detecting the small group structure of a crowd//*2009 Workshop on Applications of Computer Vision*. Snowbird, USA: IEEE: 1-8 [DOI: 10.1109/WACV.2009.5403123]
- Ge W N, Collins R T and Ruback B. 2012. Vision-based analysis of small groups in pedestrian crowds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(5) : 1003-1016 [DOI: 10.1109/TPAMI.2011.176]
- Ghasemi M, Yang Z Y, Sun M F, Ye H Z, Xiong Z H, Ghaderi J, Kostic Z and Zussman G. 2024. Video-based social distancing: evaluation in the COSMOS testbed. *IEEE Internet of Things Journal*, 11(3) : 4987-4997 [DOI: 10.1109/JIOT.2023.3305587]
- Gorrini, A, Bandini, S, Sarvi, M, Dias, C and Shiwakoti, N. 2013. An empirical study of crowd and pedestrian dynamics: the impact of different angle paths and grouping. *Transportation Research Record*. 41:42-47 [DOI: dx.doi.org/]
- Groh G, Lehmann A, Reimers J, Frieß M R and Schwarz L. 2010. Detecting social situations from interaction geometry//*Proceedings of 2010 IEEE Second International Conference on Social Computing*. Minneapolis, USA: IEEE: 1-8 [DOI: 10.1109/SocialCom.2010.11]
- Habe H, Hashimoto T, Mitsugami I, Sumi K and Yagi Y. 2017. Pedestrian group detection with gesture features. *Journal of Japan Society for Fuzzy Theory and Intelligent Informatics*, 29 (3) : 605-610 [DOI: 10.3156/jsoft.29.3\_605]
- Han R Z, Yan H M, Li J C, Wang S M, Feng W and Wang S. 2022. Panoramic human activity recognition//*Proceedings of the 17th European Conference on Computer Vision*. Tel Aviv, Israel: Springer: 244-261 [DOI: 10.1007/978-3-031-19772-7\_15]
- Hedayati H, Muehlbradt A, Szafir D J and Andrist S. 2020. REFORM: recognizing F-formations for Social Robots//*Proceedings of 2020 IEEE/RSJ International Conference on Intelligent Robots and System*. Las Vegas, USA: IEEE: 11181-11188 [DOI: 10.1109/IROS45743.2020.9340708]
- Helbing D and Molnár P. 1995. Social force model for pedestrian dynamics. *Physical Review E*, 51(5) : 4282-4286 [DOI: 10.1103/PhysRevE.51.4282]
- Hoang G, Nguyen T D, Cao T N, Le S D, Hihara K, Utada Y, Torii

- A, Izumi N, Nguyen T and Tran L Q. 2023. F2SD: a dataset for end-to-end group detection algorithms//Proceedings of the 15th International Conference on Machine Vision. Rome, Italy: SPIE: 54-63 [DOI: 10.1117/12.2679967]
- Huang L, Gong J H, Li W H, Xu T, Shen S, Liang J M, Feng Q L, Zhang D and Sun J. 2018a. Social force model-based group behavior simulation in virtual geographic environments. *ISPRS International Journal of Geo-Information*, 7 (2) : #79 [DOI: 10.3390/ijgi7020079]
- Huang S N, Huang D J and Khuhro M A. 2018b. Social pedestrian group detection based on spatiotemporal-oriented energy for crowd video understanding. *KSII Transactions on Internet and Information Systems*, 12(8): 3769-3789 [DOI: 10.3837/tiis.2018.08.012]
- Huang Z L, Wang Z, Tsai C C, Satoh S and Lin C W. 2021. DotSCN: group re-identification via domain-transferred single and couple representation learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 31 (7) : 2739-2750 [DOI: 10.1109/TCSVT.2020.3031303]
- Hung H and Kröse B. 2011. Detecting F-formations as dominant sets//Proceedings of the 13th International Conference on Multimodal Interfaces. Alicante, Spain: ACM: 231-238 [DOI: 10.1145/2070481.2070525]
- Hung H, Raman C, Gedik E, Tan S and Quiros J V. 2019. Multimodal data collection for social interaction analysis in-the-wild//Proceedings of the 27th ACM International Conference on Multimedia. Nice, France: ACM: 2714-2715 [DOI: 10.1145/3343031.3351320]
- Inaba S and Aoki Y. 2016. Conversational group detection based on social context using graph clustering algorithm//Proceedings of the 12th International Conference on Signal-Image Technology and Internet-Based Systems. Naples, Italy: IEEE: 526-531 [DOI: 10.1109/SITIS.2016.89]
- Islas R O A, Varni G, Andries M, Chetouani M and Chatila R. 2016. Modeling the dynamics of individual behaviors for group detection in crowds using low-level features//Proceedings of the 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)\*. [s.l.] : IEEE: 1104-1110 [DOI: 10.1109/ROMAN.2016.7745186]
- Jahangard S, Hayat M and Rezatofighi H. 2023. Real-time trajectory-based social group detection//Proceedings of 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems. Detroit, USA: IEEE: 1901-1908 [DOI: 10.1109/IROS55552.2023.10342121]
- Joo H, Liu H, Tan L, Gui L, Nabbe B, Matthews I, Kanade T, Nobuhara S and Sheikh Y. 2015. Panoptic studio: a massively multiview system for social motion capture//Proceedings of 2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE: 3334-3342 [DOI: 10.1109/ICCV.2015.381]
- Joo H, Simon T, Li X L, Liu H, Tan L, Gui L, Banerjee S, Godisart T, Nabbe B, Matthews I, Kanade T, Nobuhara S and Sheikh Y. 2019. Panoptic studio: a massively multiview system for social interaction capture. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41 (1) : 190-204 [DOI: 10.1109/TPAMI.2017.2782743]
- Kalnis P, Mamoulis N and Bakiras S. 2005. On discovering moving clusters in spatio-temporal data//Proceedings of the 9th International Symposium on Advances in Spatial and Temporal Databases. Angra dos Reis, Brazil: Springer: 364-381 [DOI: 10.1007/11535331\_21]
- Kang D, Mok Y, Kim Y, Kwon S and Paik J. 2023. Human group clustering in a crowded public place using multiple object detection and tracking//Proceedings of 2023 International Conference on Electronics, Information, and Communication. Singapore, Singapore: IEEE: 1-3 [DOI: 10.1109/ICEIC57457.2023.10049978]
- Kendon A. 1990. Conducting Interaction: Patterns of Behavior in Focused Encounters. Cambridge: Cambridge University Press
- Kendon A. 2010. Spacing and orientation in co-present interaction//Proceedings of the 2nd COST 2102 International Training School. Dublin, Ireland: Springer: 1-15 [DOI: 10.1007/978-3-642-12397-9\_1]
- Khan S D, Vizzari G, Bandini S and Basalamah S. 2015. Detection of social groups in pedestrian crowds using computer vision//Proceedings of the 16th International Conference on Advanced Concepts for Intelligent Vision Systems. Catania, Italy: Springer: 249-260 [DOI: 10.1007/978-3-319-25903-1\_22]
- Kjærgaard M B, Wirz M, Roggen D and Tröster G. 2012. Detecting pedestrian flocks by fusion of multi-modal sensors in mobile phones//Proceedings of 2012 ACM Conference on Ubiquitous Computing. Pittsburgh, USA: ACM: 240-249 [DOI: 10.1145/2370216.2370256]
- Kong Y, Jia Y D and Fu Y. 2012. Learning human interaction by interactive phrases//Proceedings of the 12th European Conference on Computer Vision. Florence, Italy: Springer: 300-313 [DOI: 10.1007/978-3-642-33718-5\_22]
- Küttper M and Seyfried A. 2023. Identification of social groups and waiting pedestrians at railway platforms using trajectory data. *PLoS ONE*, 18(3): #e0282526 [DOI: 10.1371/journal.pone.0282526]
- Lathuilière S, Evangelidis G and Horaud R. 2017. Recognition of group activities in videos based on single- and two-person descriptors//Proceedings of 2017 IEEE Winter Conference on Applications of Computer Vision. Santa Rosa, USA: IEEE: 217-225 [DOI: 10.1109/WACV.2017.31]
- Leach M, Baxter R, Robertson N and Sparks E. 2014. Detecting social groups in crowded surveillance videos using visual attention//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Columbus, USA: IEEE: 467-473 [DOI: 10.1109/CVPRW.2014.75]
- Leal-Taixé L, Pons-Moll G and Rosenhahn B. 2011. Everybody needs somebody: modeling social and grouping behavior on a linear pro-

- gramming multiple people tracker//Proceedings of 2011 IEEE International Conference on Computer Vision Workshops. Barcelona, Spain: IEEE: 120-127 [DOI: 10.1109/ICCVW.2011.6130233]
- Li H and Chen N N. 2023. Graph attention-based dual-branch method for social relationship recognition. *Application Research of Computers*, 40(11): 3315-3320 (李欢, 陈念年. 2023. 一种基于图注意力的双分支社会关系识别方法. *计算机应用研究*, 40(11): 3315-3320) [DOI: 10.19734/j.issn.1001-3695.2023.03.0102]
- Li J C, Han R Z, Yan H M, Qian Z K, Feng W and Wang S. 2022. Self-supervised social relation representation for human group detection//Proceedings of the 17th European Conference on Computer Vision. Tel Aviv, Israel: Springer: 142-159 [DOI: 10.1007/978-3-031-19833-5\_9]
- Li J J, Li Z J, Li X D, Feng Y, Hu Y B and Xu B G. 2021. Skill learning strategy based on dynamic motion primitives for human-robot cooperative manipulation. *IEEE Transactions on Cognitive and Developmental Systems*, 13(1): 105-117 [DOI: 10.1109/TCDS.2020.3021762]
- Li L D, Qing L, Guo L and Peng Y H. 2023a. Relationship existence recognition-based social group detection in urban public spaces. *Neurocomputing*, 516: 92-105 [DOI: 10.1016/j.neucom.2022.10.042]
- Li M, Chen T, Du H, Ma N and Xi X W. 2023b. Social group detection based on multi-level consistent behaviour characteristics. *Transportmetrica A: Transport Science*, 19(1): #1976877 [DOI: 10.1080/23249935.2021.1976877]
- Li R N, Porfilio P and Zickler T. 2013. Finding group interactions in social clutter//Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Portland, USA: IEEE: 2722-2729 [DOI: 10.1109/CVPR.2013.351]
- Li S, Hu R M, Li S H and Liao L. 2024. Hidden follower detection: how is the gaze-spacing pattern embodied in frequency domain? //Proceedings of the 38th AAAI Conference on Artificial Intelligence. Vancouver, Canada: AAAI Press: 538-546 [DOI: 10.1609/aaai.v38i1.27809]
- Li X H, Xiong S W, Duan P F, Zheng S W, Li B X and Liu M F. 2015. A study on the dynamic spatial-temporal trajectory features of pedestrian small group//Proceedings of the 2nd International Symposium on Dependable Computing and Internet of Things. Wuhan, China: IEEE: 112-116 [DOI: 10.1109/DCIT.2015.9]
- Li X L, Chen M L, Nie F P and Wang Q. 2017. A multiview-based parameter free framework for group detection//Proceedings of the 31st AAAI Conference on Artificial Intelligence. San Francisco, USA: AAAI Press: 4147-4153 [DOI: 10.1609/aaai.v31i1.11208]
- Lin S H, Yang H, Tang X C, Shi T Q and Chen L. 2019. Social MIL: interaction-aware for crowd anomaly detection//Proceedings of the 16th IEEE International Conference on Advanced Video and Signal Based Surveillance. Taipei, China: IEEE: 1-8 [DOI: 10.1109/AVSS.2019.8909882]
- Liu J Y, Li H and Kong W H. 2023. Multi-level learning counting via pyramid vision transformer and CNN. *Engineering Applications of Artificial Intelligence*, 123: #106184 [DOI: 10.1016/j.engappai.2023.106184]
- Liu X B. 2016. Multi-view 3D human tracking in crowded scenes//Proceeding of the 30th AAAI Conference on Artificial Intelligence. Phoenix, USA: AAAI Press: 3553-3559 [DOI: 10.1609/aaai.v30i1.10463]
- Livramento R, Avelino J and Moreno P. 2020. Natural data-driven approaching behaviors of humanoid mobile robots for F-formations//Proceedings of 2020 IEEE International Conference on Autonomous Robot Systems and Competitions. Ponta Delgada, Portugal: IEEE: 338-344 [DOI: 10.1109/ICARSC49921.2020.9096160]
- Lu X Y, Li X D, Hu C F, Deng J, Sheng W J and Zhu L L. 2023. A two-branch deep learning with spatial and pose constraints for social group detection. *Engineering Applications of Artificial Intelligence*, 124: #106583 [DOI: 10.1016/j.engappai.2023.106583]
- Lyu K, Yuan L and Ni X Y. 2024a. Learning autoencoder diffusion models of pedestrian group relationships for multimodal trajectory prediction. *IEEE Transactions on Instrumentation and Measurement*, 73: #2511412 [DOI: 10.1109/TIM.2024.3375973]
- Lyu Y J, Qin P G, Xu T, Zhu C and Chen E H. 2024b. InteractNet: social interaction recognition for semantic-rich videos. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 20(8): #240 [DOI: 10.1145/3663668]
- Mazzon R, Poesi F and Cavallaro A. 2013. Detection and tracking of groups in crowd//Proceedings of the 10th IEEE International Conference on Advanced Video and Signal Based Surveillance. Krakow, Poland: IEEE: 202-207 [DOI: 10.1109/AVSS.2013.6636640]
- Montesdeoca J, Toibero J M, Jordan J, Zell A and Carelli R. 2022. Person-following controller with socially acceptable robot motion. *Robotics and Autonomous Systems*, 153: #104075 [DOI: 10.1016/j.robot.2022.104075]
- Moussaïd M, Perozo N, Garnier S, Helbing D and Theraulaz G. 2010. The walking behaviour of pedestrian social groups and its impact on crowd dynamics. *PLoS ONE*, 5(4): #e10047 [DOI: 10.1371/journal.pone.0010047]
- Nabi M, Del Bue A and Murino V. 2013. Temporal poselets for collective activity detection and recognition//Proceedings of 2013 IEEE International Conference on Computer Vision Workshops. Sydney, Australia: IEEE: 500-507 [DOI: 10.1109/ICCVW.2013.71]
- Nakamura K, Ono T and Babaguchi N. 2016. Detection of groups in crowd considering their activity state//Proceedings of the 23rd International Conference on Pattern Recognition (ICPR). Cancun, Mexico: IEEE: 277-282 [DOI: 10.1109/ICPR.2016.7899646]
- Nasri M, Maliappis T, Rieffe C and Baratchi M. 2024. T-DANTE: detecting group behaviour in spatio-temporal trajectories using context information//Proceedings of the 22nd International Symposium on Intelligent Data Analysis. Stockholm, Sweden: Springer: 28-39

- [DOI: 10.1007/978-3-031-58553-1\_3]
- Nicolas A and Hassan F H. 2023. Social groups in pedestrian crowds: review of their influence on the dynamics and their modelling. *Transportmetrica A: Transport Science*, 19(1): #1970651 [DOI: 10.1080/23249935.2021.1970651]
- Nunes I O, Vaz de Melo P O S and Loureiro A A F. 2016. Group mobility: detection, tracking and characterization//*Proceedings of 2016 IEEE International Conference on Communications*. Kuala Lumpur, Malaysia: IEEE: 1-6 [DOI: 10.1109/ICC.2016.7511488]
- Paay J, Kjeldskov J, Skov M B and O'Hara K. 2013. F-formations in cooking together: a digital ethnography using youtube//*Proceedings of the 14th IFIP TC 13 International Conference on Human-Computer Interaction*. Cape Town, South Africa: Springer: 37-54 [DOI: 10.1007/978-3-642-40498-6\_3]
- Palanisamy G and Manikandan T T. 2017. Group behaviour profiling for detection of anomaly in crowd//*Proceedings of 2017 International Conference on Technical Advancements in Computers and Communications (ICTACC)*. Melmauravathur, India: IEEE: 11-15 [DOI: 10.1109/ICTACC.2017.14]
- Pandey M, Singhal S and Tripathi V. 2020. An efficient vision-based group detection framework in crowded scene//*Proceedings of the 7th International Conference on FICTA*. Singapore, Singapore: Springer: 201-209 [DOI: 10.1007/978-981-13-9920-6\_21]
- Pang S K, Li J and Godsill S J. 2008. Models and algorithms for detection and tracking of coordinated groups//*Proceedings of 2008 IEEE Aerospace Conference*. Big Sky, USA: IEEE: 1-17 [DOI: 10.1109/AERO.2008.4526445]
- Pang S K, Li J and Godsill S J. 2011. Detection and tracking of coordinated groups. *IEEE Transactions on Aerospace and Electronic Systems*, 47(1): 472-502 [DOI: 10.1109/TAES.2011.5705687]
- Pathi S K, Kiselev A and Loutfi A. 2022. Detecting groups and estimating F-formations for social human-robot interactions. *Multimodal Technologies and Interaction*, 6(3): #18 [DOI: 10.3390/mti6030018]
- Pavan M and Pelillo M. 2007. Dominant sets and pairwise clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(1): 167-172 [DOI: 10.1109/TPAMI.2007.250608]
- Poiesi F and Cavallaro A. 2015. Predicting and recognizing human interactions in public spaces. *Journal of Real-Time Image Processing*, 10(4): 785-803 [DOI: 10.1007/s11554-014-0428-8]
- Pouw C A S, Toschi F, van Schadowijk F and Corbetta A. 2020. Monitoring physical distancing for crowd management: real-time trajectory and group analysis. *PLoS ONE*, 15(10): #e0240963 [DOI: 10.1371/journal.pone.0240963]
- Qin Z, Shelton C R, and Chai L. 2013. Social grouping for target handover in multi-view video. *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, 1-6 [DOI: 10.1109/ICME.2013.6607433]
- Quiroz M, Patiño R, Diaz-Amado J and Cardinale Y. 2022. Group emotion detection based on social robot perception. *Sensors*, 22(10), #3749 [DOI: 10.3390/s22103749]
- Ramírez O A I, Varni G, Andries M, Chetouani M and Chatila R. 2016. Modeling the dynamics of individual behaviors for group detection in crowds using low-level features//*Proceedings of the 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. New York, USA: IEEE: 1104-1111 [DOI: 10.1109/ROMAN.2016.7745246]
- Raman C, Hung H and Loog M. 2023. Social processes: self-supervised meta-learning over conversational groups for forecasting nonverbal social cues//*Computer Vision — ECCV 2022 Workshops*. Tel Aviv, Israel: Springer: 639-659 [DOI: 10.1007/978-3-031-25066-8\_37]
- Ricci E, Varadarajan J, Subramanian R, Bulò, S R, Ahuja N and Lanz O. 2015. Uncovering interactions and interactors: joint estimation of head, body orientation and F-formations from surveillance videos//*Proceedings of 2015 IEEE International Conference on Computer Vision (ICCV)*. Santiago, Chile: IEEE: 4660-4668 [DOI: 10.1109/ICCV.2015.529]
- Rivas M, Alvarez P, Barrientos A and Cuadros M. 2021. Proxemics toolkit for F-formation patterns detection//*Proceedings of the 30th Conference of Open Innovations Association FRUCT*. Oulu, Finland: IEEE: 216-222 [DOI: 10.23919/FRUCT53335.2021.9599996]
- Robinson N, Tidd B, Campbell D, Kulić D and Corke P. 2023. Robotic vision for human-robot interaction and collaboration: a survey and systematic review. *ACM Transactions on Human-Robot Interaction*, 12(1): #12 [DOI: 10.1145/3570731]
- Rota P, Conci N, Sebe N and Rehag J M. 2015. Real-life violent social interaction detection//*Proceedings of 2015 IEEE International Conference on Image Processing (ICIP)*. Quebec, Canada: IEEE: 3456-3459 [DOI: 10.1109/ICIP.2015.7351446]
- Sawas A, Abuolaim A, Afifi M and Papagelis M. 2018. Tensor methods for group pattern discovery of pedestrian trajectories//*Proceedings of the 19th IEEE International Conference on Mobile Data Management (MDM)*. Hong Kong, China: IEEE: 76-85 [DOI: 10.1109/MDM.2018.00024]
- Sawas A, Abuolaim A, Afifi M and Papagelis M. 2019. A versatile computational framework for group pattern mining of pedestrian trajectories. *Geoinformatica*, 23(4): 501-531 [DOI: 10.1007/s10707-019-00353-2]
- Schmuck V and Celiktutan O. 2021. GROWL: group detection with link prediction//*Proceedings of the 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG)*. Jodhpur, India: IEEE: 1-8 [DOI: 10.1109/FG52635.2021.9667061]
- Setia A and Mittal A. 2015. Co-operative pedestrians group tracking in crowded scenes using an MST approach//*Proceedings of 2015 IEEE Winter Conference on Applications of Computer Vision*. Waikoloa, USA: IEEE: 102-108 [DOI: 10.1109/WACV.2015.21]
- Setti F, Hung H and Cristani M. 2013a. Group detection in still images by F-formation modeling: a comparative study//*Proceedings of the*

- 14th International Workshop on Image Analysis for Multimedia Interactive Services. Paris, France; IEEE: 1-4 [DOI: 10.1109/WIAMIS.2013.6616147]
- Setti F, Lanz O, Ferrario R, Murino V and Cristani M. 2013b. Multi-scale F-formation discovery for group detection//Proceedings of 2013 IEEE International Conference on Image Processing. Melbourne, Australia: IEEE: 3547-3551 [DOI: 10.1109/ICIP.2013.6738732]
- Setti F, Russell C, Bassetti C and Cristani M. 2015. F-formation detection: individuating free-standing conversational groups in images. *PLoS ONE*, 10 (5) : #e0123783 [DOI: 10.1371/journal.pone.0123783]
- Shao J, Dong N and Zhao Q. 2015. An adaptive clustering approach for group detection in the crowd//Proceedings of 2015 International Conference on Systems, Signals and Image Processing. London, UK: IEEE: 77-80 [DOI: 10.1109/IWSSIP.2015.7314181]
- Sheng W J and Li X D. 2021. Multi-task learning for gait-based identity recognition and emotion recognition using attention enhanced temporal graph convolutional network. *Pattern Recognition*, 114: #107868 [DOI: 10.1016/j.patcog.2021.107868]
- Shuaibu A N, Malik A S, Faye I and Ali Y S. 2017. Pedestrian group attributes detection in crowded scenes//Proceedings of 2017 International Conference on Advanced Technologies for Signal and Image Processing (ATSIP). Fes, Morocco: IEEE: 1-5 [DOI: 10.1109/ATSIP.2017.8075584]
- Šochman J and Hogg D C. 2011. Who knows who — inverting the social force model for finding groups//Proceedings of 2011 IEEE International Conference on Computer Vision Workshops. Barcelona, Spain: IEEE: 830-837 [DOI: 10.1109/ICCVW.2011.6130338]
- Solera F and Calderara S. 2013. Social groups detection in crowd through shape-augmented structured learning//Proceedings of the Progress in Image Analysis and Processing, ICIAP 2013. Naples, Italy: Springer: 542-551 [DOI: 10.1007/978-3-642-41181-6\_55]
- Solera F, Calderara S and Cucchiara R. 2016. Socially constrained structural learning for groups detection in crowd. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38 (5) : 995-1008 [DOI: 10.1109/TPAMI.2015.2470658]
- Solera F, Calderara S, Ristani E, Tomasi C and Cucchiara R. 2017. Tracking social groups within and across cameras. *IEEE Transactions on Circuits and Systems for Video Technology*, 27(3) : 441-453 [DOI: 10.1109/TCSVT.2016.2607378]
- Su J, Huang J L, Qing L, He X H and Chen H G. 2022. A new approach for social group detection based on spatio-temporal interpersonal distance measurement. *Heliyon*, 8(10) : #e11038 [DOI: 10.1016/j.heliyon.2022.e11038]
- Sun J H, Jiang Q H and Lu C W. 2020. Recursive social behavior graph for trajectory prediction//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 657-666 [https://doi.org/10.1109/CVPR42600.2020.00074]
- Swofford M, Peruzzi J, Tsoi N, Thompson S, Martín-Martín R, Savarese S and Vázquez M. 2020. Improving social awareness through DANTE: deep affinity network for clustering conversational interactants. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW1) : #20 [DOI: 10.1145/3392824]
- Tamura M, Vishwakarma R and Vennelakanti R. 2022. Hunting group clues with transformers for social group activity recognition//Proceedings of the 17th European Conference on Computer Vision (ECCV 2022). Tel Aviv, Israel: Springer: 19-35 [DOI: 10.1007/978-3-031-19772-7\_2]
- Tan K, Xu L F, Liu Y N and Luo B. 2017. Small group detection in crowds using interaction information. *IEICE Transactions on Information and Systems*, E100.D(7) : 1542-1545 [DOI: 10.1587/transinf.2016EDL8192]
- Tan S, Tax D M J and Hung H. 2022. Conversation group detection with spatio-temporal context//Proceedings of 2022 International Conference on Multimodal Interaction. Bengaluru, India: ACM: 170-180 [DOI: 10.1145/3536221.3556611]
- Taylor A, Chan D M and Riek L D. 2020. Robot-centric perception of human groups. *ACM Transactions on Human-Robot Interaction*, 9(3) : #15 [DOI: 10.1145/3375798]
- Taylor A and Riek L D. 2022. REGROUP: a robot-centric group detection and tracking system//Proceedings of the 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI). Sapporo, Japan: IEEE: 412-421 [DOI: 10.1109/HRI53351.2022.9889634]
- Thompson S, Gupta A, Gupta A W, Chen A and Vázquez M. 2021. Conversational group detection with graph neural networks//Proceedings of 2021 International Conference on Multimodal Interaction. Montréal, Canada: ACM: 248-252 [DOI: 10.1145/3462244.3479963]
- Tran K N, Bedagkar-Gala A, Kakadiaris I A and Shah S K. 2013. Social cues in group formation and local interactions for collective activity analysis//Proceedings of 2013 International Conference on Computer Vision Theory and Applications. Barcelona, Spain: SciTePress: 539-548 [DOI: 10.5220/0004256505390548]
- Tzelepi M and Tefas A. 2021. Graph embedded convolutional neural networks in human crowd detection for drone flight safety. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 5(2) : 191-204 [DOI: 10.1109/TETCI.2019.2897815]
- Varadarajan J, Subramanian R, Bulò S R, Ahuja N, Lanz O and Ricci E. 2018. Joint estimation of human pose and conversational groups from social scenes. *International Journal of Computer Vision*, 126 (2) : 410-429 [DOI: 10.1007/s11263-017-1026-6]
- Vascon S and Bazzani L. 2017. Group detection and tracking using sociological features//Murino V, Cristani M, Shah S and Savarese S, eds. *Group and Crowd Behavior for Computer Vision*. Cambridge: Academic Press: 29-66 [DOI: 10.1016/B978-0-12-809276-7.00004-7]
- Voon W P, Mustapha N, Affendey L S and Khalid F. 2016. A new clus-

- tering approach for group detection in scene-independent dense crowds//Proceedings of the 3rd International Conference on Computer and Information Sciences (ICCOINS). Kuala Lumpur, Malaysia: IEEE: 414-417 [DOI: 10.1109/ICCOINS.2016.7783251]
- Wang Q, Chen M L, Nie F P and Li X L. 2020. Detecting coherent groups in crowd scenes by multiview clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42 (1): 46-58 [DOI: 10.1109/TPAMI.2018.2875002]
- Wang X C. 2024. Social force model-based video anomaly detection approaches//Anomaly Detection in Video Surveillance. Singapore, Singapore: Springer: 273-286 [DOI: 10.1007/978-981-97-3023-0\_11]
- Wang X Y, Chen X C, Jiang P H, Yuan X Y, Ji M Q, Huang R Q and Fang L. 2024. Group interaction field for learning and explaining pedestrian anticipation. *Engineering*, 34: 70-82 [DOI: 10.1016/j.eng.2023.05.020]
- Wang Z, Wu B, Wang W Z, Teng Y Y, Shuai J, Xiao Y P and Bai T. 2021. A survey of social relation understanding based on image and video information. *Chinese Journal of Computers*, 44 (6): 1168-1199 (王正, 吴斌, 王文哲, 滕一阳, 帅杰, 肖云鹏, 白婷. 2021. 基于图像和视频信息的社交关系理解研究综述. *计算机学报*, 44 (6): 1168-1199) [DOI: 10.11897/SP.J.1016.2021.01168]
- Wright B, McCurry J M, Lawson W and Trafton J G. 2021. Autonomous group detection, delineation, and selection for human-agent interaction. *Proceedings of ICSR, LNAI 13086*: 329 - 338 [DOI: 10.1007/978-3-030-90525-5\_28]
- Wu W H, Yi W F, Li J H, Chen M Y and Zheng X P. 2024. Automatic identification of human subgroups in time-dependent pedestrian flow networks. *IEEE Transactions on Multimedia*, 26: 166-177 [DOI: 10.1109/TMM.2023.3262975]
- Xie Z, Wu T F, Yang X M, Zhang L M and Wu K W. 2019. Jointly social grouping and identification in visual dynamics with causality-induced hierarchical Bayesian model. *Journal of Visual Communication and Image Representation*, 59: 62-75 [DOI: 10.1016/j.jvcir.2019.01.006]
- Xu D N, Hu R M, Wang Z, Luo L B, Li D S and Zeng W J. 2022. Gaze- and spacing-flow unveil intentions: hidden follower discovery//Proceedings of the 30th ACM International Conference on Multimedia. Lisboa, Portugal: ACM: 2115-2123 [DOI: 10.1145/3503161.3548207]
- Xu D N, Hu R M, Xiong Z X, Wang Z, Luo L B and Li D S. 2021. Trajectory is not enough: hidden following detection//Proceedings of the 29th ACM International Conference on Multimedia (MM'21), Association for Computing Machinery. New York, USA: [s.n.]: 5373-5381 [DOI: 10.1145/3474085.3475664]
- Yücel Z, Zanlungo F, Ikeda T, Miyashita T and Hagita N. 2013. Deciphering the crowd: modeling and identification of pedestrian group motion. *Sensors*, 13(1): 875-897 [DOI: 10.3390/s130100875]
- Yan Y C, Qin J, Ni B B, Chen J X, Liu L, Zhu F, Zheng W S, Yang X K and Shao L. 2023. Learning multi-attention context graph for group-based re-identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(6): 7001-7018 [DOI: 10.1109/TPAMI.2020.3032542]
- Yang M, Rashidi L, Rao A S, Rajasegarar S, Ganji M, Palaniswami M and Leckie C. 2018. Cluster-based crowd movement behavior detection//Proceedings of 2018 Digital Image Computing: Techniques and Applications. Canberra, Australia: IEEE: 1-8 [DOI: 10.1109/DICTA.2018.8615809]
- Yang Z Y, Sun M F, Ye H Z, Xiong Z H, Zussman G and Kostic Z. 2022. Birds eye view social distancing analysis system [EB/OL]. [2025-02-03]. <https://arxiv.org/pdf/2112.07159.pdf>
- Yi S, Li H S and Wang X G. 2015. Understanding pedestrian behaviors from stationary crowd groups//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, USA: IEEE: 3488-3496 [DOI: 10.1109/CVPR.2015.7298971]
- Yi S, Li H S and Wang X G. 2016. Pedestrian behavior modeling from stationary crowds with applications to intelligent surveillance. *IEEE Transactions on Image Processing*, 25(9): 4354-4368 [DOI: 10.1109/TIP.2016.2590322]
- Yin Y F, Yang G and Man H. 2013. Small human group detection and event representation based on cognitive semantics//Proceedings of the 7th IEEE International Conference on Semantic Computing. Irvine, USA: IEEE: 64-69 [DOI: 10.1109/ICSC.2013.20]
- Yoo H, Eom T, Seo J and Choi S I. 2019. Detection of interacting groups based on geometric and social relations between individuals in an image. *Pattern Recognition*, 93: 498-506 [DOI: 10.1016/j.patcog.2019.05.001]
- Yu J Q, Yang J H, Yang H, Pan R J, Lai P R and Zhai G T. 2024. Psychology-guided environment aware network for discovering social interaction groups from videos. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 20(8): #233 [DOI: 10.1145/3657295]
- Yuan Y, Fang J W and Wang Q. 2015. Online anomaly detection in crowd scenes via structure analysis. *IEEE Transactions on Cybernetics*, 45(3): 548-561 [DOI: 10.1109/TCYB.2014.2330853]
- Yucel Z, Zanlungo F, Feliciani C, Gregorj A and Kanda T. 2018. Estimating social relation from trajectories//Proceedings of the 9th International Conference on Pedestrian and Evacuation Dynamics (PED2018). Lund, Sweden: Linköping University Electronic Press: 222-229
- Zaki M H and Sayed T. 2018. Automated analysis of pedestrian group behavior in urban settings. *IEEE Transactions on Intelligent Transportation Systems*, 19(6): 1880-1889 [DOI: 10.1109/ITITS.2017.2747516]
- Zanlungo F, Ikeda T and Kanda T. 2014. Potential for the dynamics of pedestrians in a socially interacting group. *Physical Review E*,

- 89(1): #012811 [DOI: 10.1103/PhysRevE.89.012811]
- Zanotto M, Bazzani L, Cristani M and Murino V. 2012. Online Bayesian nonparametrics for group detection//Proceedings of BMVC 2012. Surrey, UK: BMVA Press: 1-12 [DOI: 10.5244/C.26.111]
- Zhang L and Hung H. 2021. On social involvement in mingling scenarios: detecting associates of F-formations in still images. IEEE Transactions on Affective Computing, 12(1): 165-176 [DOI: 10.1109/TAFFC.2018.2855750]
- Zhang X, Xie X H, Wen L and Lai J H. 2024. Person group detection with global trajectory extraction in a disjoint camera network. Neurocomputing, 574: #127281 [DOI: 10.1016/j.neucom.2024.127281]
- Zhang X G, Ma D X, Yu H, Huang Y, Howell P and Stevens B. 2020. Scene perception guided crowd anomaly detection. Neurocomputing, 414: 291-302 [DOI: 10.1016/j.neucom.2020.07.019]
- Zhang Y H, Qin L, Ji R R, Yao H X and Huang Q M. 2015a. Social attribute-aware force model: exploiting richness of interaction for abnormal crowd detection. IEEE Transactions on Circuits and Systems for Video Technology, 25(7): 1231-1245 [DOI: 10.1109/TCSVT.2014.2355711]
- Zhang Y H, Qin L, Zhang S P, Yao H X and Huang Q M. 2015b. Formation period matters: towards socially consistent group detection via dense subgraph seeking//Proceedings of the 5th ACM on International Conference on Multimedia Retrieval. Shanghai, China: ACM: 475-478 [DOI: 10.1145/2671188.2749305]
- Zhao Q, Shao J and Zhao Y. 2016. A multistage filtering for detecting group in the crowd//Proceedings of 2016 International Conference on Audio, Language and Image Processing. Shanghai, China: IEEE: 771-774 [DOI: 10.1109/ICALIP.2016.7846671]
- Zhao R Y, Li M Y, Wang Y, Liu Q, Dong D H and Li C L. 2020a. An improved dynamic pedestrian grouping model in public transport space//Proceedings of 2020 International Conference on Computing, Networks and Internet of Things (CNIOT). Sanya, China: IEEE: 111-116 [DOI: 10.1145/3398329.3398363]
- Zhao Y L, Hu Z X, Zhou L, Liu M and Liu J T. 2020b. Data-driven online group detection based on structured prediction//Proceedings of the 16th International Conference on Automation Science and Engineering (CASE). Hong Kong, China: IEEE: 1220-1225 [DOI: 10.1109/CASE48305.2020.9216765]
- Zhou C, Han M, Liang Q, Hu Y F and Kuai S G. 2019. A social interaction field model accurately identifies static and dynamic social groupings. Nature Human Behaviour, 3(8): 847-855 [DOI: 10.1038/s41562-019-0618-2]
- Zhou R, Zhou H, Gao H, Tomizuka M, Li J and Xu Z. 2022. Group-tron: dynamic multi-scale graph convolutional networks for group-aware dense crowd trajectory forecasting//Proceedings of 2022 IEEE International Conference on Robotics and Automation (ICRA), 805-811 [DOI: 10.1109/ICRA46639.2022.9811585]
- Zhou Z, Li K, He X J and Li M M. 2016. A generative model for recognizing mixed group activities in still images//Proceedings of the 25th International Joint Conference on Artificial Intelligence. New York, USA: AAAI Press: 3654-3660
- Zitouni M S, Sluzek A and Bhaskar H. 2019. Visual analysis of socio-cognitive crowd behaviors for surveillance: a survey and categorization of trends and methods. Engineering Applications of Artificial Intelligence, 82: 294-312 [DOI: 10.1016/j.engappai.2019.04.012]
- Zitouni M S, Sluzek A and Bhaskar H. 2020. Towards understanding socio-cognitive behaviors of crowds from visual surveillance data. Multimedia Tools and Applications, 79(3): 1781-1799 [DOI: 10.1007/s11042-019-08201-z]

## 作者简介

李姝,女,博士研究生,主要研究方向为隐秘尾随检测与社交行为理解。E-mail: shuli@stu.xidian.edu.cn

胡瑞敏,通信作者,男,教授,主要研究方向为网络空间社会安全感知计算(可信身份计算、隐秘组织发现、复杂关系分析、行为模式挖掘)等大数据自然社会理解的理论与方法。E-mail: rmhu@xidian.edu.cn

王怀远,男,博士研究生,主要研究方向为视频行为模式分析和意图识别。E-mail: wang\_hy96@163.com