复杂视频场景人群行为分析研究

胡同花[1]，胡紫英[2]
1. 永州职业技术学院 信息网络中心，湖南 永州 425000
2. 湖南科技学院 理学院，湖南 永州 425000
DOI:10.61369/ASDS.2025050016

摘　　要： 随着人群聚集场所中异常事件频发，研究复杂视频场景下的人群行为分析在公共安全维护、智能监控系统搭建等关键领域中愈发重要。本文提出复杂视频场景的人群行为分析方法、人群行为检测常见手段和检测方法、主要存在的问题及人群行为检测和定位模型，并从人群计数及密度估计、人群全局异常检测及人群局部异常行为检测和定位三个角度提出解决方案及应用场景。

关 键 词： 复杂视频场景；人群行为分析；异常检测

Research on the Analysis of Crowd Behavior in Complex Video Scenes

Hu Tonghua[1], Hu Ziying[2]
1.Yongzhou Vocational and Technical College Information Network Center, Yongzhou, Hunan 425000
2.Faculty of Science, Hunan University of Science and Engineering, Yongzhou, Hunan 425000

Abstract: With the frequent occurrence of abnormal events in crowded places, the analysis of crowd behavior in complex video scenes has become increasingly important in key areas such as public safety maintenance and the construction of intelligent monitoring systems. This article proposes a method for analyzing crowd behavior in complex video scenes, common methods and detection techniques for crowd behavior detection, main existing problems, and crowd behavior detection and localization models. Solutions and application scenarios are proposed from three perspectives: crowd counting and density estimation, global anomaly detection of crowds, and local anomaly behavior detection and localization of crowds.

Keywords: complex video scenes; analysis of crowd behavior; outlier detection

引言

随着人群聚集的情况频繁在各种公共场所中出现，因人群拥挤引发的灾难性事件逐年增多。近年来因人群拥挤引发的典型人群灾难性事件造成巨大财产、生命损失，主要包括：2021年3月21日，坦桑尼亚悼念活动踩踏事件、2022年10月 29日，韩国首尔梨泰院踩踏事件、2024年4月19日，也门萨那慈善物资发放踩踏事件、2024年12月1日，加沙代尔拜莱赫抢面饼踩踏事件、2025年2月15日，印度新德里火车站踩踏事件等。人群灾难性事件的频繁出现，给公共安全管理带来了新的挑战，成为各级安全管理部门迫切需要重视和解决的现实问题。新时期下视频监控技术主要基于目标检测、跟踪、识别和分析等方法，对公共环境中人群监控场景下个体间存在大量的遮挡和不规则运动很难适用。因此，本文重点研究复杂环境下人群监控场景的人群行为特性，监测并预警人群动态，避免人群灾难性事件的发生。

一、人群行为分析方法国内外现状

现有的人群行为分析方法大致分为基于传统特征表示的方法和基于深度特征表示的方法两大类。本文主要从人群计数及密度估计、人群异常检测和人群运动建模三个方面进行国内外现状分析。

（一）人群计数及密度估计研究

基于传统特征表示的人群计数及密度估计方法通过设计不同的时空特征表示（如：梯度直方图、形状描述子、局部纹理特征、运动轨迹等），采用全局/局部检测或回归模型完成行人检测、定位及计数。中国科学院操晓春研究员团队[1]首次提出端到端的CNN回归模型对高密度人群图像进行密度估计。上海交通

基金项目：2024年度湖南省自然科学基金科教联合项目基金（复杂视频场景人群异常行为检测研究 2024JJ8086）。
作者简介：胡同花（1982.01-），女，主要研究领域为网络技术、智能应用。 E-mail: yzzyhutonghua@163.com。
第二作者：胡紫英（1974.12-），女，主要研究领域为嵌入式研究、生物涂层材料。E-mail:409467800@qq.com。

大学杨小康教授团队[2]针对已有模型应用到新的人群场景时密度估计精度急剧下降的问题，首次提出基于跨场景 CNN 模型的人群计数方法。针对人群场景的尺度变化问题，上海科技大学高盛华博士团队[3]基于三种具有不同大小感受野的卷积核，构造出多列 CNN 网络结构估计人群密度。上海交通大学杨华博士团队[4]采用条件 GAN 将人群图像转换为对应的人群密度图。Shen 等[5]采用 U 型 GAN 分别对人群整体场景和人群块进行训练。为避免尺度变化引起的密度估计不一致性，引入交叉尺度一致性损失进行训练。目前急需解决具有复杂性、尺度变化、运动模式多样性等特点人群监控场景，如何提取人群场景的深度密度特征表示等问题。

（二）人群异常事件检测研究

基于传统特征表示的人群异常事件检测研究通常基于人群视频的可视特征，采用不同的统计模型（HMM 模型、贝叶斯模型等）或物理模型（社会力模型、能量模型等）完成人群异常事件检测。Yuan 等[6]提出上下文结构描述子建模行人关系，根据结构描述子的时空变化检测人群事件是否发生。杭州电子科技大学的张旭光教授团队[7]等根据人群场景能量分布的变化检测人群全局异常事件。

基于深度特征表示的人群异常事件检测研究通常采用深度网络学习人群常规事件的深度特征表示，再基于不同分类器判断是否出现异常事件。广东工业大学蔡瑞初博士团队[8]对人群整体动态的时间序列建模，提出了基于多尺度时间递归神经网络的人群异常事件检测和定位方法。人群场景异常检测通常包括全局异常检测和局部异常检测，从以上各团队成功改进中可以看到，人群场景的可视运动特征的深度表示，是增强人群异常检测性能的关键。

（三）人群运动建模研究

基于传统特征表示的人群运动建模方法通常基于人群运动特征（如：光流、轨迹、时空特征等），采用聚类或流模型或概率图模型建模人群运动模式。杨华博士团队基于运动向量的局部特征及全局运动结构，提出基于卷曲和发散度的人群运动轨迹描述子。该描述子具有尺度和旋转不变性，在人群运动模式分析应用中表现出优异的性能。香港中文大学王晓刚教授团队[9]使用 KLT 轨迹特征，基于马尔可夫链构建人群集体性转换先验器，提取人群场景的集体性、冲突性、一致性、统一性描述子建模人群运动模式。上海交通大学赵旭博士团队[10]根据人群轨迹相似性定义人群的全局运动一致性及局部运动一致性，提出聚类算法检测人群组运动模式。山东师范大学刘弘教授团队[10]基于社会力模型构建疏散路径集合，考虑影响行人路径选择的四种因素，提出基于疏散路径集合的路径选择和人群疏散模型。虽然这些方法在具有规则运动的人群场景中检测精度高、速度快；但在密集的复杂人群场景中，由于人群遮挡、相机运动、光照问题等引起的低级特征提取的不稳定性，造成算法检测精度不稳定。

二、人群行为检测常见手段和检测方法

复杂视频场景中人群行为检测常用手段有视频监控设备和传感器，如：分布于不同场景的监控摄像头、全景摄像头、高速摄像头等，红外传感器可检测人体的存在和移动，与视频数据结合能更准确地判断人群行为；压力传感器布置在地面等区域，可感知人群的分布和活动强度等信息。检测方法主要分为传统计算机视觉方法、机器学习方法和深度学习方法三大类，各类方法基于不同的技术原理，在检测精度、实时性和场景适应性上呈现差异化特征，对比分析如表1所示。

表1 人群行为检测手段和方法对比与应用场景

方法类型	代表算法	优点	缺点	典型应用场景
传统计算机视觉	光流法（Optical Flow）	像素级运动捕捉，快速动作敏感	计算量大，抗干扰能力弱	简单场景个体动作分析
	背景建模（Background Modeling）	实时性强，异常检测效率高	动态背景适应性差	固定监控场景异常预警
机器学习	支持向量机（Support Vector Machine，SVM）	小样本泛化能力强，分类边界清晰	手工特征表征能力有限	工业动作规范性检测
	隐马尔可夫模型（Hidden Markov Model，HMM）	时序建模能力强，适合短序列动作	长程依赖建模不足	手语识别、简单动作序列分析
深度学习	卷积神经网络（Convolutional Neural Network，CNN）	自动提取时空特征，复杂场景适应性强	时序信息捕捉不足	暴力行为检测、动作分类
	循环神经网络（Recurrent Neural Network，RNN）及其变种（LSTM/GRU）	长时序动态建模，行为过程表征好	计算效率低，空间特征提取较弱	人群异常行为预测
	时空图神经网络（Spatio-Temporal Graph Neural Network，ST-GNN）	显式建模个体交互与群体动态	依赖姿态估计，密集场景建模困难	团队协作分析、群体异常检测

复杂视频场景下的人群行为检测已从传统手工特征方法逐步发展为深度学习主导的端到端模型。传统方法在简单场景中仍具实时性优势，而深度学习方法通过时空特征联合建模，在复杂场景（如拥挤、遮挡、多目标交互）中表现更优。

三、人群行为分析主要存在的问题

1. 标注成本与数据多样性矛盾。由于人群监控视频的数据标注需耗费大量人力物力，导致现有研究多针对某类数据集，缺乏针对复杂环境下多场景多类别人群场景的研究。

2. 跨域分布差异显著。不同场景（如地铁站 vs 校园）、不同

摄像头视角（俯视 vs 平视）、不同时段（白天 vs 黑夜）的数据分布差异大，模型泛化易受 “域偏移” 影响。

3. 隐私保护与数据采集冲突。高清视频采集涉及个体隐私（如面部、行为轨迹），公开数据集多经过脱敏处理（如模糊化），导致特征完整性受损。现有模型多为监督学习模型，在应用到新场景时，检测精度常明显下降，模型适用性受限。

四、人群行为分析检测和定位模型

（一）跨场景多尺度的复杂人群密度估计研究

基于人群场景的尺度不变性，本文设计一种多尺度的人群密度特征表示，并基于深度特征域适应模型完成跨场景人群密度估计。该模型方法如图1所示。

1. 基于 CNN 的人群场景密度先验器

针对不同监控环境下人群场景密度存在较大差异的问题，本文设计人群密度先验器为人群场景密度估计提供先验知识。将数据集中的人群场景密度分为10类，采用卷积神经网络对人群场景进行密度分类，采用5层带有 Relu 激活函数的卷积层，每个卷积层后跟一个池化层。

2. 基于 GAN 的多尺度密度估计预训练模型

对已有密度标注的源域训练集采用深度网络训练，得到源域数据的特征分布。针对 CNN 平均卷积核引发的生成密度图模糊问题，基于人群场景的尺度不变性特征，提出基于 GAN 的多尺度场景密度估计预训练模型。模型训练过程如图1(a) 所示。

学习三种尺度下的人群密度生成图，通过归一化不同尺度下的生成图得到场景密度生成图。模型中的判别器与生成器的网络结构如下：生成器编码器采用五个卷积层，解码器对应采用五个反卷积层；判别器则采用三个卷积层和一个全卷积层。在训练目标函数方面，除生成对抗网络的重构损失之外，为减少生成密度图与标注密度图的差异，提高密度估计精度，引入密度一致性损失。通过多轮训练求解网络参数，得到训练集中人群场景的密度特征表示。

3. 基于权值共享的域适应人群密度估计算法

针对测试集中通常只有少量标注样本或没有标注样本的实际情况，通过微调训练阶段得出的密度估计模型，基于权值共享的实现域适应的人群密度估计算法。测试模型如下：将标注样本集与未标注样本集分别输入特征编码器 E，且同时使用预训练模型中的参数初始化编码器。对标注样本，计算样本场景分类损失；对所有样本，采用最小化平均偏差 (MMD) 定义源域特征分布与目标域特征分布的距离。测试模型如图1(b) 所示，通过重新训练编码器，调整网络参数，增强编码器生成的深度特征表示的域不变性，从而实现跨场景多尺度的人群密度估计。（图1(b) 仅图示了一种尺度下的测试过程）

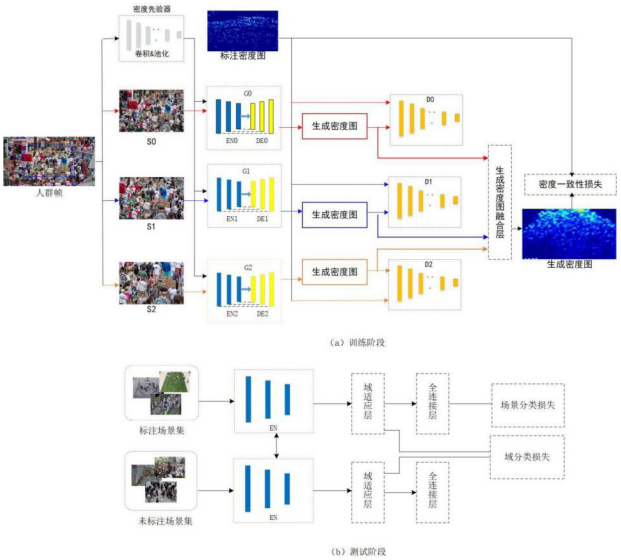

图1 跨场景多尺度人群密度估计模型

（二）无监督域适应人群全局异常检测模型

鉴于人群监控视频中场景特征的空间结构性及时序相关性，设计基于时空能量特征的域适应人群深度特征表示模型，并研究无监督的人群全局异常检测方法。

1. 基于时空能量特征的人群运动表示

由于在复杂人群监控场景中存在大量遮挡且运动变化微小，传统的光流特征不能及时捕捉局部运动变化。引入时空能量特征模型，基于高斯滤波器的三阶导数设计人群场景的时空能量特征表示，对频域内通过原点的所有平面的能量响应和进行能量归一化，获得人群时空能量特征表示。

2. 域适应的人群场景深度特征表示模型

考虑到人群视频的运动特征变化同时受到场景空间结构及时序相关性的影响，提出域适应的人群深度特征表示方法，建模人群场景的运动特征。模型方法如图2所示。对源域视频和目标域视频采用 GAN 网络学习深度特征表示，图中、分别表示源域特征生成器和判别器；、分别表示目标域特征生成器和判别器。为利用视频特征间的时序相关性，引入源域特征与目标域特征间的循环对抗损失。为减小源域特征分别与目标域特征分别的距离，引入特征匹配损失。此外，针对源域人群场景的类别标签，引入场景分类损失训练网络。为保持深度特征的域不变性，引入域判别器，基于域对抗损失训练网络以确保深度特征的域不变性。

模型采用的网络结构如下：与采用相同的网络结构，采用10个卷积层，每个卷积层后紧随批归一化层及 ReLU 激活层；且与采用相同参数进行初始化；与也采用相同的网络结构，采用6个卷积层及一个全卷积层；域判别器采用具有3个卷积层的网络结构。模型目标函数包括图2中所示的循环对抗损失、生成对抗损失、场景分类损失、特征匹配损失及域对抗损失。模型训练问题最终转化为最小最大问题求解，通过多轮训练求解网络参数，得到人群场景密度生成模型。

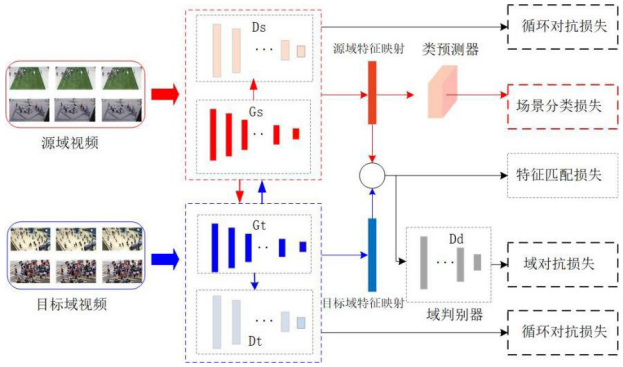

图2 基于时空相关性的域适应人群深度特征表示模型

3. 无监督人群全局异常检测算法

鉴于在一段时间内人群场景的运动变化具有连续性特点，提取滑动窗口内生成的深度运动特征表示，计算其在给定时间段内的累计变化。当特征累计变化超出给定阈值时，认为发生了全局异常事件。

（三）基于时空特征的人群局部异常行为检测和定位模型

针对现有人群数据集中只有少量像素级异常行为标注的问题，本文设计基于时空特征的人群局部异常行为检测模型，模型方法如图3所示。

1. 基于动态颗粒流的人群块分割算法

首先，基于人群序列的时空能量特征，分析并讨论人群颗粒间的时空相关性及社会交互影响力，建立人群颗粒之间的时空交互模型，提出基于拉格朗日动态颗粒流人群运动分割模型，得到人群视频的可视特征块及能量特征块。

2. 基于时空特征的人群异常行为检测模型

基于人群视频块的时空特征建立域适应的异常行为检测模型，实现无监督 / 半监督的人群异常行为检测和定位。首先采用两个 CNN 网络建模源域人群块和目标域人群块的深度特征表示，并引入域判别器最小化源域特征分布与目标域特征分布的距离，判别器的输入为源域及目标域人群块。为更好的利用源域及目标域中的已有标注信息，引入异常判别器判别当前人群块的真伪，异常域判别器的输入仅为源域中的标注人群块。模型目标函数包括图3中定义域对抗损失及异常预测损失。模型采用的网络结构类似于图2中的网络结构，其中源域生成器与目标域生成器采用参数共享的同一网络结构，异常判别器与域判别器则都采用多层 CNN 结构。该网络模型适用于无监督 / 弱监督的人群局部异常检测与定位。

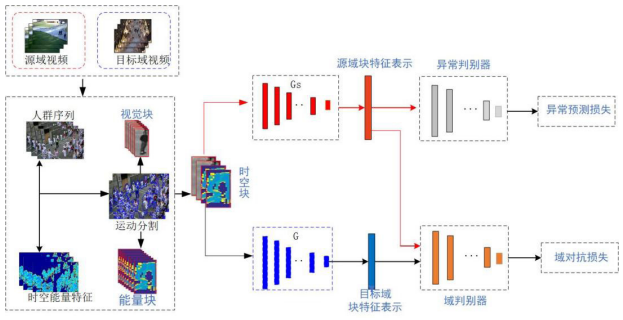

图3 基于时空特征的人局部异常行为检测模型

跨场景多尺度人群密度估计模型优势：通过多尺度特征融合与跨场景迁移，实现不同环境下的人群密度精准统计，通用性强。局限：聚焦宏观数量估算，缺乏个体行为细节分析，计算成本较高。基于时空相关性的域适应人群深度特征表示模型优势：融合时空动态特征与域适应技术，擅长捕捉群体行为模式并跨场景迁移应用。局限：个体动作分辨率不足，对长程时序依赖和极端场景变化的适应性较弱。基于时空特征的人局部异常行为检测模型优势：利用时空注意力聚焦个体局部特征，对摔倒、暴力等异常动作的实时定位精度高。局限：缺乏全局场景理解，对新型异常模式泛化能力差，多目标交互时易误检。

五、人群行为定位模型的定位精度与误差分析

（一）跨场景多尺度的复杂人群密度估计模型

定位精度：在主流数据集（如 ShanghaiTech、UCF-QNRF）上，平均绝对误差（MAE）为82.4 - 98.7，均方误差（MSE）为159.7-334.7。引入多尺度注意力机制后，小目标检测精度提升30%，误检率降低15 - 20%。主要误差源于目标粘连（密集区域像素重叠导致计数偏差）和尺度敏感性（极端尺度变化贡献约35% 误差）。

（二）基于时空相关性的域适应人群深度特征表示模型

定位精度：跨场景域适应后定位精度提升12 - 18%（如地铁站到机场场景），多传感器融合下定位误差达亚米级（0.8m，σ =0.32m）。域偏移（场景差异导致特征分布偏移）和时空失配（动态场景时序异步引发15 - 20% 轨迹断点）是核心误差源。小样本过拟合使定位波动增加30%，需通过流形学习和黎曼几何分析抑制分布偏移。

（三）基于时空特征的人局部异常行为检测模型

定位精度：在 UCSD 等数据集上异常检测 F1值达85.7 - 92.3%，基于骨架的 GCN 模型关节级定位准确率94.2%。行为边界模糊（快速动作或遮挡导致30% 边界偏差）和环境干扰（光照变化与遮挡引发12% 关键点丢失）是主要挑战。个体行为差异使模型置信度波动 ±18%，需结合滑动窗口和多模态融合优化。

平均绝对误差 (Mean Absolute Error，MAE)：预测总人数与实际总人数之差的绝对值的平均值。均方误差 (Mean Squared Error，MSE)：预测总人数与实际总人数之差的平方的平均值。F1分数 (F1 Score)：精确率和召回率的调和平均数。基于图结构的卷积网络（Graph Convolutional Network，GCN）：处理关节拓扑关系（如人体骨架）。σ（Sigma，标准差）：描述误差离散程度。

表2 检测模型综合对比与对比与优化方向

模型类型	优势	典型精度 / 误差范围	误差改进方向
多尺度密度估计	高密度场景适应性	MAE 82.4 - 98.7；F1 71.2%	抑制粘连、增强尺度鲁棒性
时空域适应特征表示	跨场景泛化能力	定位误差0.8m；泛化误差↓23.6%	优化域偏移补偿、时序同步
局部异常行为检测	细粒度关节级定位	F1 85.7 - 92.3%；误报率5.1 - 8.3%	多模态融合、边界精细化

六、总结

本文围绕复杂视频场景的人群行为分析方法、常见检测手段和方法、主要存在的问题，研究基于深度特征的人群行为检测和定位模型，研究方法从人群计数及密度估计、人群全局异常检测及人群局部异常行为检测和定位三个角度提出解决方案，总结分析了跨场景多尺度人群密度估计模型、基于时空相关性的域适应人群深度特征表示模型、基于时空特征的人局部异常行为检测模型这3种人群行为定位模型的定位精度和误差，综合对比后给出应用场景建议，为进一步探索人群行为分析提供了重要研究基础。

参考文献

[1] Wang, C., Zhang, H., Yang, L., Liu, S., Cao, X, Deep people counting in extremely dense crowds[J]. in Proceedings of the 23rd ACM international conference on Multimedia, 2015:1299-1302.

[2] 卢博文 . 基于深度学习的监控视频中的异常行为的检测算法研究 [D]. 南京邮电大学 ,2020.DOI:10.27251/d.cnki.gnjdc.2020.000865.

[3] 徐涛 , 田崇阳 , 刘才华 . 基于深度学习的人群异常行为检测综述 [J]. 计算机科学 , 2021, 48(09): 125-134.

[4] 亢洁，田野，杨刚 . 基于改进 SSD 的人群异常行为检测算法研究 [J]. 红外技术 ,2022,44(12):1316-1323.

[5] 葛文超 , 魏超 , 王玉涛，鲁迎春 , 易茂祥 . 基于潜在空间矩阵的半监督异常检测 [J]. 计算机应用研究 .2020.37(S2):318-320.

[6] Yuan, Y., J. Fang, and Q. Wang, Online anomaly detection in crowd scenes via structure analysis[J]. IEEE Transactions on Cybernetics, 2015. 45(3): 548-561.

[7] Zhang X, Zhang Q, Hu S, et al. Energy level-based abnormal crowd behavior detection[J]. Sensors, 2018, 18(2): 423.

[8] 蔡瑞初 , 谢伟浩 , 郝志峰等 . 基于多尺度时间递归神经网络的人群异常检测 [J]. 软件学报 , 2015,26(11):2884-2896.

[9] 罗朝阳 . 基于深度学习的人体异常行为识别算法研究 [D]. 陕西理工大学 ,2021.DOI:10.27733/d.cnki.gsxlg.2021.000168.

[10] 鱼春燕 , 徐岩 , 缑丽莎 , 等 . 基于单列深度时空卷积神经网络的人群计数 [J]. 激光与光电子学进展 ,2021,58(8):143-151.

