软件学报 ISSN 1000-9825, CODEN RUXUEW
*Journal of Software*,2015,26(11):2884−2896 [doi: 10.13328/j.cnki.jos.004893]
©中国科学院软件研究所版权所有.
E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

# 基于多尺度时间递归神经网络的人群异常检测*

蔡瑞初[1,2], 谢伟浩[1], 郝志峰[1], 王丽娟[1], 温 雯[1]

[1](广东工业大学 计算机学院,广东 广州 510006)

[2](计算机软件新技术国家重点实验室(南京大学),江苏 南京 210023)

通讯作者: 蔡瑞初, E-mail: cairuichu@gmail.com, https://sites.google.com/site/cairuichu/

**摘 要**: 如何在人群密度大、变化快、存在大量遮挡的密集场景中实现可靠的人群事件检测,是领域研究的难点和热点.在密集场景时空建模的基础上提出了一种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.首先对人群场景进行网格化划分,并利用多尺度光流直方图对每个网格的人群动态进行刻画;然后,连接各个局部的人群动态获得整体的人群动态,实现整体人群动态的时间序列建模;最后,利用多尺度时间递归神经网络实现异常事件的检测和定位.其中,多尺度隐含层实现了密集场景中不同规模相邻网格之间的空间联系,节点间的反馈关系则为时间维度上的关系表达提供了有效方案.与多种代表性算法的对比实验,验证了本方法的有效性.

**关键词**: 视频监控;人群异常事件检测;时间递归神经网络;多尺度

**中图法分类号**: TP183

中文引用格式: 蔡瑞初,谢伟浩,郝志峰,王丽娟,温雯.基于多尺度时间递归神经网络的人群异常检测.软件学报,2015,26(11):2884−2896. http://www.jos.org.cn/1000-9825/4893.htm

英文引用格式: Cai RC, Xie WH, Hao ZF, Wang LJ, Wen W. Abnormal crowd detection based on multi-scale recurrent neural network. Ruan Jian Xue Bao/Journal of Software, 2015,26(11):2884−2896 (in Chinese). http://www.jos.org.cn/1000-9825/4893.htm

## Abnormal Crowd Detection Based on Multi-Scale Recurrent Neural Network

CAI Rui-Chu[1,2], XIE Wei-Hao[1], HAO Zhi-Feng[1], WANG Li-Juan[1], WEN Wen[1]

[1](School of Computer, Guangdong University of Technology, GuangZhou 510006, China)

[2](State Key Laboratory for Novel Software Technology (Nanjing University), Nanjing 210023, China)

**Abstract**: Because of the great variations of crowd density and crowd dynamics, as well as the existence of many shelters in scenes, the abnormal crowd event detection and localization are still challenging problems and hot topics of the crowd scene analysis. Based on the spatial-temporal modeling of the crowd scene, this paper proposes an abnormal crowd event detection and localization approach based on multi-scale recurrent neural network. Firstly, the crowd scenes are split into grids and presented using multi-scale histogram of optical flow (MHOF). Then, different grids are connected to obtain a global time series model of the crowd scene. Finally, a multi-scale recurrent neural network is devised to detect and locate the abnormal event on the time series model of the crowd scene. In the multi-scale recurrent neural network, the multi-scale hidden layers are used to model the spatial relation among different scale neighbors, and the feedback loops are used to catch the temporal relation. Extensive experiments demonstrate the effectiveness of the presented approach.

**Key words**: video surveillance; abnormal crowd event detection; recurrent neural network; multi-scale

随着人口和活动多样性的增加,密集人群现象越来越普遍,给公共空间管理和社会安全带来了巨大的挑战.如果能够利用广泛分布的视频监控设施进行事件检测,则可方便公共空间管理,提升社会管理能力.由于人群场

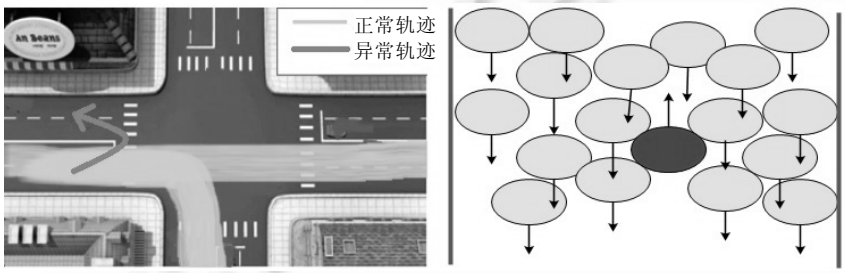
* 基金项目: 国家自然科学基金(61202269, 61472089); 广东省自然科学基金(2014A030306004, 2014A030308008); 广东省科技计划(2013B051000076); 广东省高校学科专业建设与质量工程专项(PT2011JSJ); 广州市科技计划(201200000031, 2013Y2-00034, 2014Y2-00027); 计算机软件新技术国家重点实验室开发课题(KFKT2014B03, KFKT2014B23)

收稿时间: 2015-05-01; 修改时间: 2015-07-14, 2015-08-11; 定稿时间: 2015-08-26

© 中国科学院软件研究所 http://www.jos.org.cn

景分析的巨大应用价值,人群场景分析已经成为近几年视频监控领域的研究热点,吸引了大量研究者的关注[1−4].人群场景分析是指对由大量目标所构成的场景进行分析,主要包括人群密度估计[5]、人群中的目标追踪[6]、移动模式分割[7]、人群行为识别[8]、人群异常事件检测[9−31]等内容.人群密度高、模式变化快、场景中存在着巨大的遮挡等挑战,使得传统视频监控技术不能直接应用于人群场景[1,2],这使得人群场景分析仍然是一个有待解决的问题,涌现了大量相关研究[1−4].但是,仍然没有被普遍接受的用于人群场景分析任务的解决方案[4].在人群场景分析的众多问题中,人群异常事件检测是研究焦点之一[9−31].

人群异常事件检测指的是对人群场景中不符合规则的事件进行检测.这里的不符合规则的事件即异常的定义往往带有主观性[3],比如,可以把人群恐慌当作异常事件,也可以把在场景中打架斗殴当作异常事件,或者是在人行道骑车等等.人群异常检测可被建模为“正常-异常”二分类问题[32].基于异常类型,可以进一步区分为空间异常事件检测和时间异常事件检测[9].图 1 给出了这两种事件的例子:在图 1(a)中,浅色的轨迹为随时间变化形成的符合规则的车辆运行轨迹,深色的 U 型弯为随时间变化形成的异常的轨迹,这种情况被视为时间异常事件;而在图 1(b)中,每一个椭圆代表了行进中的个体,而深色那个个体是不同于周围的个体的,这种情况被视为空间异常事件.




(a) 时间异常事件 (b) 空间异常事件

Fig.1 Two examples of abnormal situation

图 1 两种异常情况的例子

空间建模和时间建模分别为上述两种异常检测提供了有效手段,是异常检测建模的主要方法.例如,文献[10−12]以时空块的形式来表示时间异常事件和空间异常事件,但是这些方法只考虑了局部信息,而忽略了各个局部之间的关系.但是在空间异常事件中,每个个体与周围个体之间存在着一定的关系,而只以时空块的形式来表示异常事件则只是通过对局部的区域进行观察来发现是否是异常事件,这样往往会忽略掉一些有助于空间异常事件检测的信息;而对于时间异常事件,不同时间异常事件持续的时间是不一样的,而只以时空块的形式来表示异常事件则只是通过对较短的一段时间的观察来发现是否是异常事件,这样往往会忽略掉一些有助于时间异常事件检测的信息.虽然也有一些方法考虑了各局部空间维度之间的相互关系[13−15],或者考虑了各局部时间维度之间的相互关系[16,17],但是它们只是考虑了其中一方面,而没有同时考虑各局部时间维度的相互关系和各局部空间维度的相互关系.

目前,也有极少部分方法同时考虑到了各局部时间维度的相互关系和各局部空间维度的相互关系,但是基本上都是对它们进行独立建模[18−21],然后再整合它们的检测结果.例如,Mahadevan 等人[18]首先基于混合动态纹理为时间正常事件建立模型,时间异常等同于低概率事件,而空间异常则使用了基于混合动态纹理的判别空间显著性的方法;然后,在分别得到空间异常检测结果和时间异常检测结果之后,将它们的结果合并起来作为异常检测和定位的结果用于人群异常检测和定位;最后,通过实验证明了该方法要优于大部分的人群异常检测和定位方法.之后,Li 等人[21]进一步对 Mahadevan 等人的工作进行完善,在原有的工作上加入了多尺度的空间异常情况检测,并提出了使用判别模型来整合空间和时间异常情况来进行人群异常检测和定位的方法、使用适合在线检测的在线条件随机场(conditional random fields,简称 CRF)以整合空间和时间异常情况以进行人群异常的检

© 中国科学院软件研究所 http://www.jos.org.cn

测和定位的方法.Li 等人[21]也对时空关系对于人群异常事件检测的影响进行了研究,证明了同时考虑空间维度的相互关系以及时间维度的相互关系的确有助于人群异常事件的检测.使用混合动态纹理并结合时间异常和空间异常的方法虽然比许多人群异常检测方法效果要好,但是这个方法需要大量的计算,检测大小为 240×160 的视频图像需要花费 25s,很难保证实时性[10].

为了实现全局时空建模,本文提出一种基于多尺度时间递归神经网络(multi-scale recurrent neural network,简称 MRNN)的人群异常事件检测和定位方法.首先,本方法将人群场景进行网格化划分,计算每一帧的光流,并从各个局部网格中提取多尺度光流直方图(multi-scale histogram of optical flow,简称 MHOF)特征;然后,链接各个局部网格特征作为整体的人群动态,并将整体人群动态当作时间序列;最后,利用多尺度时间递归神经网络模型同时对时间维度的关系和空间维度的关系进行建模.利用多尺度时间递归神经网络的隐含层来发现人群场景中的不同尺度的相邻网格之间的关系,利用反馈节点来发现时间维度的关系,利用递归神经网络的输出层对不同网格进行判别,以此来发现和定位异常事件.本文的贡献在于:基于时间递归神经网络,对视频信息中的空间维度相互关系和时间维度相互关系进行了统一建模;根据时空关系的局部性,构造了多尺度时间递归神经网络模型,提高了检测效果和检测效率.

## 1 相关工作

根据引发异常的个体数目,可以将人群异常检测分成两类:局部异常事件检测和整体异常事件检测[22].其中,局部异常事件检测指的是个体行为与其周围邻居的行为存在差异,如人行道上骑车;而整体异常事件检测指的是整个场景中的一群人的行为是异常的,如人群恐慌[4].整体异常事件检测的任务是检测异常事件,并确定异常事件的起始和终止位置以及它们之间的过渡;局部异常事件检测的任务是检测异常事件,并定位异常发生的位置[4].

整体异常事件检测往往是通过将人群看作一个整体,然后从中提取主要的移动模式来检测异常事件[4],如文献[14,23−26].其中,Chen 等人[14]将人群表示成图的形式,各个独立的区域被当作是一个节点,为了有效地建模拓扑变化,他们综合考虑局部特征和全局特征,并结合它们作为指示器去发现在场景中是否出现异常情况;最近, Wu 等人[26]提出了一个用于逃跑恐慌行为检测的贝叶斯框架,他们使用光流场来表示人群移动,然后构造对应的光流的类条件概率密度函数,最后,通过贝叶斯框架进行逃跑恐慌行为检测.局部异常事件检测除了要检测异常以外,还需要进一步定位异常发生的位置.为了实现局部异常事件检测,已提出了不同的异常检测技术[9−13,16−22,27−31].主流技术主要有两类:一类是基于视觉的方法,这种方法只使用来自计算机视觉邻域的技术来进行模型的学习和预测,如隐含马尔可夫模型(hidden Markov model,简称 HMM)[20]、动态纹理[18,21]、词袋模型[28]、稀疏表示[22]、流型学习[29]等;另一类是受物理启发的方法,这种方法结合了用于人群动态表示的相关物理模型,并使用学到的模型来进行异常的检测,如流场模型[30]、社会力模型[13]、人群能量模型[31]等.更加具体的信息可以参见综述文献[1−4].

## 2 多尺度时间递归神经网络模型

人群异常检测主要的目标在于寻找人群状态的变化与人群异常之间存在的关系,其关键在于:(1) 构建能刻画人群状态变化的特征;(2) 构建人群状态的变化与人群异常之间存在的关系的模型.针对上述两个关键问题,本文设计了如图 2 所示的两阶段异常事件检测框架:在特征表示阶段,首先需要对整个场景进行网格化划分,并从每个网格中提取相应的人群动态特征,最后连接各个网格的特征一起作为当前时刻的人群状态;在建模阶段,拟利用多尺度时间递归神经网络模型对时间维度的异常和不同尺度的空间维度的异常情况进行同时建模,进而检测每个网格的异常情况.

© 中国科学院软件研究所 http://www.jos.org.cn

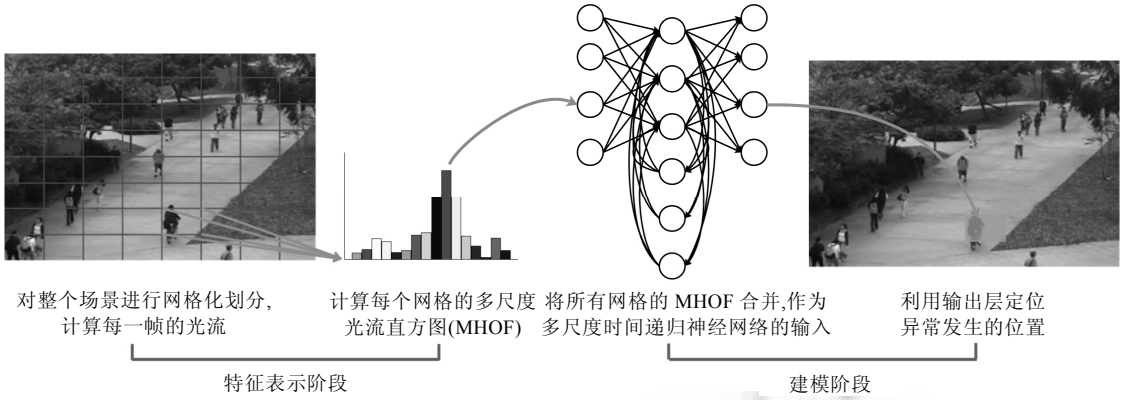



Fig.2 Whole frame of the method of crowd abnormal events detection and location based on multi-scale recurrent neural network

图 2 基于多尺度的时间递归神经网络的人群异常事件检测和定位方法的整体框架

### 2.1 特征表示

人群动态特征是人群异常事件检测的基础.由于 MHOF 同时保存了时间和空间上下文信息[22],本文拟使用 MHOF 作为底层特征表示方法.如图 3 所示,在基于光流算法获得光流场后,我们将每个网格中的像素量化成 MHOF.具体来说,使用如下方式对网格中的每个像素进行量化,计算 MHOF:

$$h(x,y)=\begin{cases} round\left(\dfrac{p\theta(x,y)}{2\pi}\right) \bmod P, & r(x,y)<\tau \\ round\left(\dfrac{p\theta(x,y)}{2\pi}\right) \bmod P+P, & \tau\leqslant r(x,y)<\varsigma \\ round\left(\dfrac{p\theta(x,y)}{2\pi}\right) \bmod P+P+P, & r(x,y)\geqslant\varsigma \end{cases} \tag{1}$$

其中,$p$ 表示方向数,$r(x,y)$和$\theta(x,y)$分别表示光流在像素点$(x,y)$上的大小和方向.对于直方图的层数,我们测试了 2 层和 3 层 HOF 的划分方法,实验结果表明,3 层的 HOF 比 2 层的 HOF 要好.实验结果将会在实验部分给出.由于更高层的 HOF 将导致模型维度过高,训练复杂度大,因此我们没有对 3 层以上的 HOF 模型进行测试.在之后的实验中,我们都采用 3 层的光流直方图,并将 $p$ 设置为 8,$\tau$设置为 1,$\varsigma$设置为 3,总共有 24 个区间.

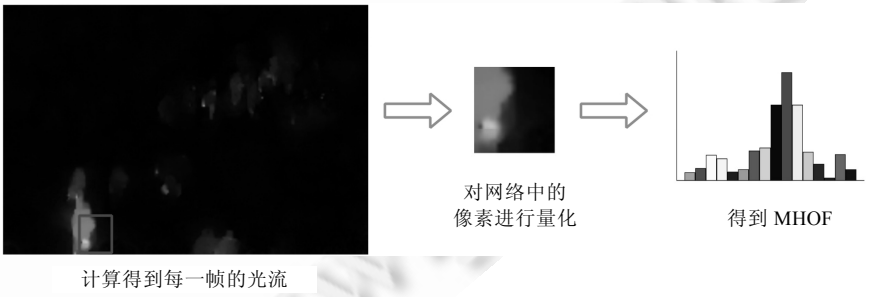



Fig.3 A diagram of feature extraction

图 3 特征提取示意图

在得到每个网格的 MHOF 之后,我们连接各个局部的人群动态获得整体的人群动态,构建人群动态的时间序列表示.具体来说,令 $N$ 表示网格的数目,$G_n(G_n\in\Re^{24})$表示第 $n$ 个网格的 MHOF 特征,我们将 $N$ 个网格链接成

© 中国科学院软件研究所 http://www.jos.org.cn

为 $x_t=(G_1,G_2,G_3,\ldots,G_N)^T$,总共 $N\times 24$ 维的一个向量来表示当前时刻的人群动态;由多个连续时刻的人群动态 $x_1,x_2,\ldots,x_T$,我们获得了人群动态的时间序列表示.

### 2.2 多尺度时间递归神经网络模型

对于人群异常事件检测,另一个关键点在于构建人群状态的变化与人群异常之间存在的关系的模型.为了更好地发现时间异常事件和空间异常事件,我们基于时间递归神经网络对视频信息中的空间维度相互关系和时间维度相互关系进行了统一建模,并根据时空关系的局部性构造了多尺度时间递归神经网络模型.下面给出详细的论述.

对于视频序列来说,它是天然的时间序列,并且在人群异常事件检测问题中存在着时间维度的异常,而时间递归神经网络是一个随时间运作的神经网络,它能利用输入序列中的时间信息,如图 4 所示.在时间递归神经网络(recurrent neural network,简称 RNN)的每个时间段中,当前输入与上一个时刻的中间状态共同作用产生新的中间状态,这个中间状态就表示了当前时刻与过去时刻之间的相互关系,当前的输出由过去时刻与当前时刻之间的相互关系决定.为此,在本问题中,我们利用 RNN 的上述特性来发现时间异常事件;通过网格化划分,利用 RNN 的隐含节点来发现各个局部之间的相互关系以及空间异常事件.在上面工作的基础上,将网格是否异常作为目标,以此来实现空间以及时间维度的整体建模.

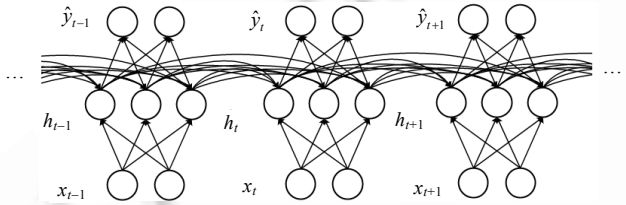



Fig.4 Architecture of recurrent neural network

图 4 时间递归神经网络体系结构

具体来说,如图 5 所示,$G_n$ 表示每个网格的 MHOF 特征,$\hat{o}_n(\hat{o}_n\in[0,1])$ 表示对应的网格是否异常.于是,当前时刻的输入可以表示为 $x_t=(G_1,G_2,G_3,\ldots,G_N)^T(X\in\Re^{N\times 24},N$ 为网格数);当前时刻的输出可以表示为 $\hat{y}_t=(\hat{o}_1,\hat{o}_2,\hat{o}_3,\ldots,\hat{o}_N)^T(y\in\Re^N)$.对于给定的一个序列 $x_1,x_2,\ldots,x_T(X\in\Re^{N\times 24})$,通过以下公式来计算隐含层状态 $h_1,h_2,\ldots,h_T(h\in\Re^m)$和预测值 $\hat{y}_1,\hat{y}_2,\ldots,\hat{y}_T(y\in\Re^N)$:

$$\left.\begin{aligned} t_i &= W_{hx}x_i + W_{hh}h_{i-1} + b_h \\ h_i &= e(t_i) \\ s_i &= W_{yh}h_i + b_y \\ \hat{y}_i &= g(s_i) \end{aligned}\right\} \tag{2}$$

其中,$W_{hx},W_{hh},W_{yh}$ 分别为输入层到隐含层、隐含层到隐含层和隐含层到输出层的权重矩阵,而 $b_h,b_y$ 分别为隐含层和输出层的偏置项,$t_1,t_2,\ldots,t_T(t\in\Re^k)$和 $s_1,s_2,\ldots,s_t(s\in\Re^k)$分别为隐含层节点的输入和输出层节点的输入,$e$ 和 $g$ 为激活函数,一般情况下为非线性函数.对于隐含层的激活函数 $e$,本文使用反正切函数 tanh,对于输出层的激活函数 $g$,本文使用 sigmoid 函数.

令 $\theta=[W_{hx},W_{hh},W_{yh},b_h,b_y]$表示所有参数组成的一个大向量,则目标函数可以形式化定义为

$$f(\theta)=L(\hat{y};y) \tag{3}$$

其中,$L$ 为距离函数用来计算真实值与预测值之间的误差.一般情况下,$L$ 可以取平方误差 $\sum_i\|\hat{y}_i-y_i\|^2/2$ 或者是交叉熵误差 $-\sum_i\sum_j y_{ij}\log(\hat{y}_{ij})+(1-y_{ij})\log(1-\hat{y}_{ij})$.在本文中,我们使用了平方误差.为了得到更好的泛化效果,在目标函数中加入了正则项.所以,最终的目标函数为

$$f(\theta)=L(\hat{y};y)+\lambda reg(\theta) \tag{4}$$

© 中国科学院软件研究所 http://www.jos.org.cn

其中,$\lambda$为正则项所占权重,*reg* 为 L1 范式.整理可以得到如下模型:

$$\min f(\theta)=L(\hat{y};y)+\lambda reg(\theta)$$
$$\text{s.t.}\begin{cases}t_i=W_{hx}x_i+W_{hh}h_{i-1}+b_h\\ h_i=e(t_i)\\ s_i=W_{yh}h_i+b_y\\ \hat{y}_i=g(s_i)\end{cases} \tag{5}$$

最后,我们使用 BPTT[33]来训练上述的模型,并使用随机梯度下降来更新权重.

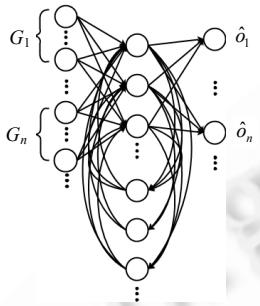
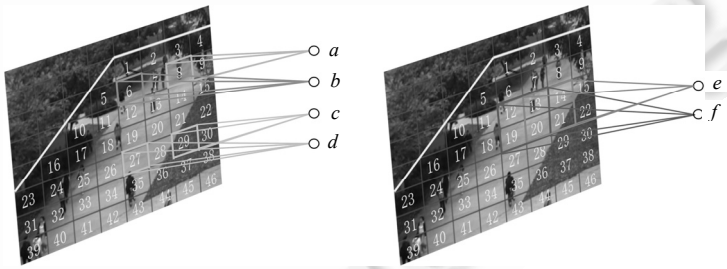



Fig.5 A diagram of inputs and outputs of every moment

图 5 每个时刻的输入、输出示意图

由于在局部人群异常事件检测中,每一个触发异常的个体,只与它周围的个体存在着一定的关系,而与离这个个体较远的个体基本上没有任何关系,而且对于触发异常的个体的大小或者是不同类型的异常事件来说,影响范围是不一样的,为此,我们可以使用不同的隐含节点来发现不同尺度的空间关系,构造一个多尺度的时间递归神经网络模型来更好地发现空间异常事件.

本文拟采用不同的隐含节点来发现两种尺度的空间关系:第 1 种尺度为 1 个网格大小,第 2 种尺度为相邻 4 个网格大小.为了发现第 1 种尺度的空间关系,我们使用一部分隐含节点来发现网格本身各维度之间的关系,即,每个隐含节点对应于一个网格大小的滑动窗口,滑动窗口的滑动距离为 1.如图 6(a)所示,图中给出了 4 个隐含节点 *a*,*b*,*c*,*d*,分别用于发现第 8 个、第 6 个、第 27 个和第 29 个网格本身各特征维度之间的关系.




(a) 用于发现网格本身各特征维度关系的隐含层节点示意图

(b) 用于发现相邻 4 网格之间相互关系的隐含层节点示意图

Fig.6 A diagram of the mode of connection from input layer to hidden layer

图 6 输入层到隐含层的连接方式示意图

为了发现第 2 种尺度的空间关系,我们使用了一部分隐含节点来发现相邻的 4 个网格之间的相互关系,即,每个隐含节点对应于 4 网格大小的滑动窗口,滑动窗口的滑动距离为 1.如图 6(b)所示,图中给出了 2 个隐含节点 *e*,*f*,其中,隐含节点 *e* 用于发现 13,14,20,21 这 4 个网格之间的关系;隐含节点 *f* 用于发现 12,13,19,20 这 4 个网格之间的关系.我们把它们称为 1 倍多尺度隐含节点.之后,我们使用多倍的多尺度隐含节点来更好地发现不同网

© 中国科学院软件研究所 http://www.jos.org.cn

格之间的关系.

本文拟采用屏蔽矩阵方式来刻画多尺度网络间的局部链接关系.令 $m_{hx}$ 表示从输入层到隐含层参数的一个屏蔽矩阵,用于屏蔽掉与相应的隐含节点无关的参数,并令 $CEM(M_1,M_2)$来表示两个相同维度矩阵对应元素相乘,最后,我们得到了如下的多尺度时间递归神经网络模型:

$$\begin{aligned}&\min f(\theta)=L(\hat{y};y)+\lambda reg(\theta)\\&\text{s.t.}\ \begin{cases}t_i=CEM(m_{hx}W_{hx})x_i+W_{hh}h_{i-1}+b_h\\h_i=e(t_i)\\s_i=W_{yh}h_i+b_y\\\hat{y}_i=g(s_i)\end{cases}\end{aligned}\tag{6}$$

## 3 实验与分析

为了对本文提出的算法进行验证和分析,我们选取了 5 种代表性的方法进行比较,包括 SF[13]、MPPCA[19]、SF-MPPCA、Adam 等人的方法[27]和 MDT[18].在本文中,我们使用 RNN 来表示时间递归神经网络模型,使用 MRNN 来表示多尺度的时间递归神经网络模型.我们在圣地亚哥加州大学的异常检测数据集上对上述方法进行了测试及比较.该数据集包含了两个不同场景的视频子集,分别为 Ped1 数据集和 Ped2 数据集.在 Ped1 数据集上,RNN 和 MRNN 的默认参数为:学习率为 0.003,学习率衰减比例为 0.999,迭代次数为 2 000,L1 范式的系数为 0.005.对于 RNN,我们使用了 500 个隐含节点;对于 MRNN 的隐含节点的设置,我们分别用 46 个隐含节点来发现网格本身各维度特征之间的关系,用 32 个隐含节点来发现相邻 4 个网格之间的关系,总共 5 倍规模.序列长度设置为 39 个时刻.在 Ped2 数据集上,RNN 和 MRNN 的默认参数为:学习率为 0.01,学习率衰减比例为 0.999,迭代次数为 1 000,L1 范式的系数为 0.01.对于 RNN,我们使用了 400 个隐含节点;对于 MRNN 的隐含节点的设置,我们分别用 50 个隐含节点来发现网格本身各维度特征之间的关系,用 36 个隐含节点来发现相邻 4 个网格之间的关系,总共 5 倍规模.序列长度设置为 29 个时刻.

### 3.1 实验数据

数据集使用圣地亚哥加州大学的异常检测数据集[18,34],数据主要通过安装在高地、俯瞰人行道的静态摄像机获得.在人行道上的人群密度是可变的,人群密度从稀疏到非常拥挤.其中,正常的情况为场景中只有正常行走的行人在人行道上.异常情况主要有以下两点:1) 非行人实体在人行道上流动;2) 异常的人群行为模式.其中,异常包括有:较多的情况为在人行道上骑自行车、溜冰、开小汽车或者是行人横跨人行道或者在草地上践踏;较少的情况为坐着轮椅的人.所有的异常情况都是自然出现的.数据集被分成了两个子集,分别为 Ped1 数据集和 Ped2 数据集,每个子集包含一个场景.在 Ped1 子集中,包含了由 200 帧组成的多个视频片段;在 Ped2 数据集中,包含了分别有 180,150,120 帧组成的多个视频片段.

- 视频场景 1(Ped1 数据集):一群人朝着视像头或者是远离摄像头移动,有少许的视角的扭曲.其中包含了 34 个训练视频样本和 36 个测试视频样本.
- 视频场景 2(Ped2 数据集):行人的移动平行于摄像平面.其中包含了 16 个训练视频样本和 12 个测试的视频样本.

由于我们的方法使用了有监督的方式,所以对于 Ped1 数据集,我们使用了其中的 36 个测试视频样本(其中,正常样本帧数约为 3 200,异常样本帧数约为 4 000),并使用了 5 折交叉验证来验证模型的有效性.对于 Ped2 数据集,我们在原有的 12 个测试视频样本的基础上加入了 8 个训练视频样本,总共 20 个视频序列(其中,正常样本帧数约为 1 652,异常样本帧数约为 1 648),并在这 20 个视频序列上做了 5 折交叉验证,以此来验证模型的有效性.本文分别给出了在 Ped1 数据集上的异常检测以及定位结果和在 Ped2 数据集上的异常检测结果.

如图 7 所示,在实验中:

- 本文首先去除了一些无关的背景,对于其他方法,我们并没有对背景进行手工剪除.但是对于传统方法来说,不剪除背景对于检测结果不会有太大的影响.因为背景部分在不同帧之间是固定不变的,不会产

© 中国科学院软件研究所 http://www.jos.org.cn

生大的光流点.

- 然后,对于 Ped1 数据集,将场景划分为 30×20 的 46 个网格;对于 Ped2 数据集,我们剔除了上部分背景,并将场景划分为 36×30 的 50 个网格.
- 最后,对每个网格进行标记,获得训练及测试样本.


Fig.7 Diagram of the gridding partition of crowd scene

图 7 人群场景网格化划分示意图

### 3.2 评估方式

局部人群异常事件检测任务主要由两部分组成:异常检测和异常定位.在这里,我们沿用了 Mahadevan 等人[18]提出的评估方式,即,对于异常检测,当有某个像素被认为是异常的时候就判定该帧为异常;对于异常定位,如果引起异常的对象有 40%的像素被检测到,就认为异常定位正确.该评估方式被后续的许多文献(如文献[9−12,21,28,29])所采用,作为传统方法的评估方式.所以,将其用于传统方法来评估异常检测以及异常定位是否准确是合适的.

对于本文提出的方法,我们使用受试者工作特征曲线(receiver operating characteristic curve,简称 ROC)来对其进行评估.ROC 曲线的横坐标表示本身为负类被检测为正类的比例,即,假阳性率(false positve rate,简称 FPR);纵坐标表示本身为正类被检测为正类的比例,即,真阳性率(true positive rate,简称 TPR).我们在两个数据集上做了 5 折交叉验证,并使用阈值平均[35]方法求取 5 折交叉验证的平均 ROC 曲线.该方法在给定的阈值下得到每条 ROC 曲线对应的点,然后对这些点求均值,得到在该阈值下的平均值.变换阈值得到不同阈值下的平均值,最终得到 5 折交叉验证的平均 ROC 曲线.

本文分别给出了异常检测结果的等错误率(equal error rate,简称 EER)(即,在漏检率(miss rate,简称 MR)等于 FPR 情况下的错误率);在等错误率情况下,异常定位结果的检测率;异常检测结果和异常定位结果的 ROC 曲线.

### 3.3 实验结果

图 8 给出了在 Ped1 数据集上使用特征分别为 2 层 HOF 和 3 层 HOF 的 MRNN 进行异常检测的结果.从图 8 中我们可以看出,3 层的 HOF 比 2 层的 HOF 检测效果要好.

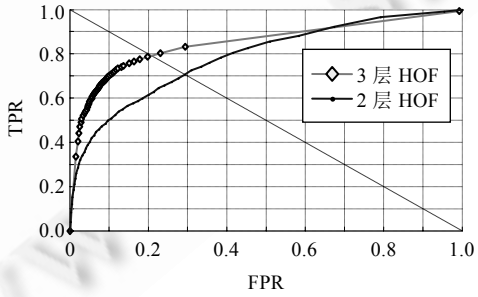



Fig.8 ROC curves of 2 layers HOF and 3 layers HOF MRNN on Ped1 data set for abnormal detection

图 8 2 层 HOF 和 3 层 HOF 的 MRNN 在 Ped1 数据集上的异常检测结果的 ROC 曲线

© 中国科学院软件研究所 http://www.jos.org.cn

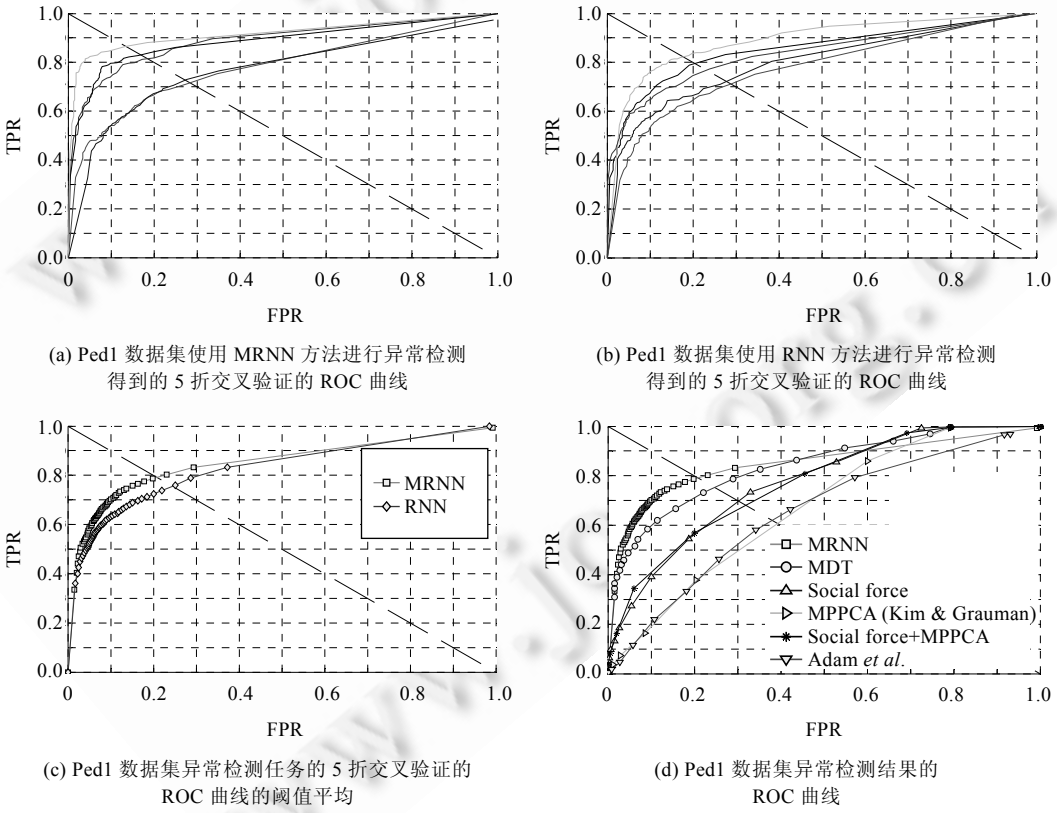
表 1 给出了在 Ped1 和 Ped2 数据集上,RNN 和 MRNN 与其他经典算法异常检测结果的等错误率.从表 1 中可以看到,MRNN 比其他经典算法有更低的等错误率.

**Table 1** Experimental result of abnormal detection: equal error rate (EER) (%)

**表 1** 异常检测实验结果:等错误率(EER)(%)

| | SF[13] | MPPCA[19] | SF-MPPCA | Adam, *et al.*[27] | MDT[18] | RNN | MRNN |
|---|---|---|---|---|---|---|---|
| Ped1 | 31 | 40 | 32 | 38 | 25 | 24.6 | 20.8 |
| Ped2 | 42 | 30 | 36 | 42 | 25 | 15 | 12 |

图 9 给出了异常检测结果的 ROC 曲线,其中,图 9(a)、图 9(e)分别为 Ped1 和 Ped2 数据集上 MRNN 方法在异常检测方面的 5 折交叉验证的 ROC 曲线,从中可以看出,我们的方法具有一定的稳定性;图 9(b)、图 9(f)分别为 Ped1 和 Ped2 数据集上 RNN 方法在异常检测方面的 5 折交叉验证的 ROC 曲线,从中可以看出,RNN 方法也具有一定的稳定性;图 9(c)、图 9(g)分别为 Ped1 和 Ped2 数据集上 MRNN 和 RNN 方法在异常检测方面的 5 折交叉验证的 ROC 曲线的阈值平均,从中可以看出,MRNN 比普通的 RNN 效果要好;图 9(d)、图 9(h)分别为 Ped1 和 Ped2 数据集上 MRNN 方法和其他经典算法在异常检测任务方面的 ROC 曲线,从中可以看出,MRNN 比经典的方法要好.从以上实验结果可以发现,MRNN 比 RNN 在异常检测方面更好的原因在于:在异常事件检测中,目标个体的大小、异常事件的时空范围均存在较大差异.与 RNN 相比,MRNN 通过多个尺度领域间的关系可以更好地发现不同粒度的异常事件,从而获得了比 RNN 更加准确的异常检测效果.MRNN 比其他经典的方法在异常检测方面更好的原因在于:MRNN 通过同时建模空间维度的关系和时间维度的关系,可以更好地发现时间维度的相互依赖关系和多尺度空间维度的相互关系,从而能获得比经典方法更好的效果.




(a) Ped1 数据集使用 MRNN 方法进行异常检测得到的 5 折交叉验证的 ROC 曲线

(b) Ped1 数据集使用 RNN 方法进行异常检测得到的 5 折交叉验证的 ROC 曲线

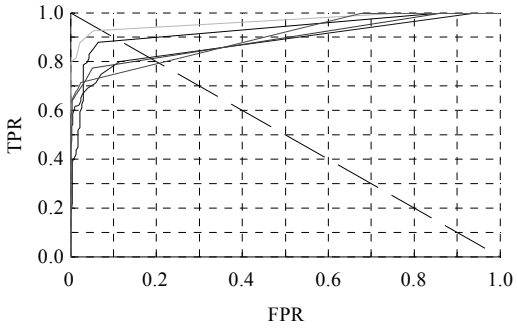
(c) Ped1 数据集异常检测任务的 5 折交叉验证的 ROC 曲线的阈值平均

(d) Ped1 数据集异常检测结果的 ROC 曲线

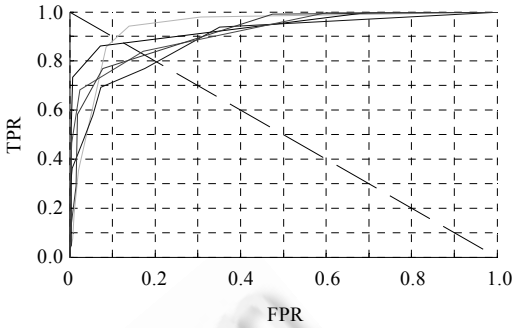
Fig.9 ROC curves for abnormal detection

图 9 异常检测结果的 ROC 曲线

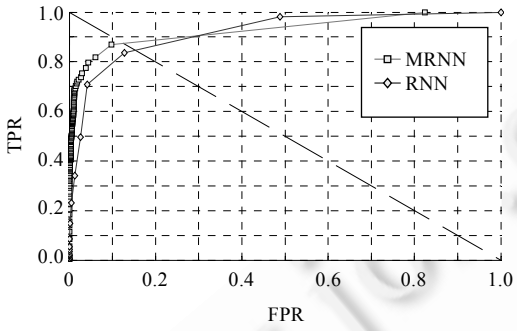
© 中国科学院软件研究所 http://www.jos.org.cn



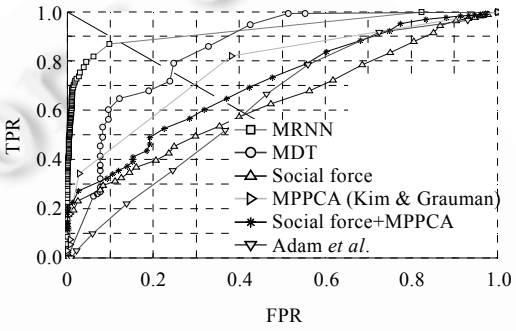

(e) Ped2 数据集使用 MRNN 方法进行异常检测得到的 5 折交叉验证的 ROC 曲线




(f) Ped2 数据集使用 RNN 方法进行异常检测得到的 5 折交叉验证的 ROC 曲线




(g) Ped2 数据集上异常检测任务的 5 折交叉验证的 ROC 曲线的阈值平均




(h) Ped2 数据集异常检测结果的 ROC 曲线

Fig.9 ROC curves for abnormal detection (Continued)

图 9 异常检测结果的 ROC 曲线(续)

表 2 给出了在等错误率的情况下,Ped1 数据集上,RNN 和 MRNN 与其他经典算法异常定位结果的检测率.从表 2 中可以看到,MRNN 和 RNN 比其他经典算法在定位异常方面具有更大的优势.

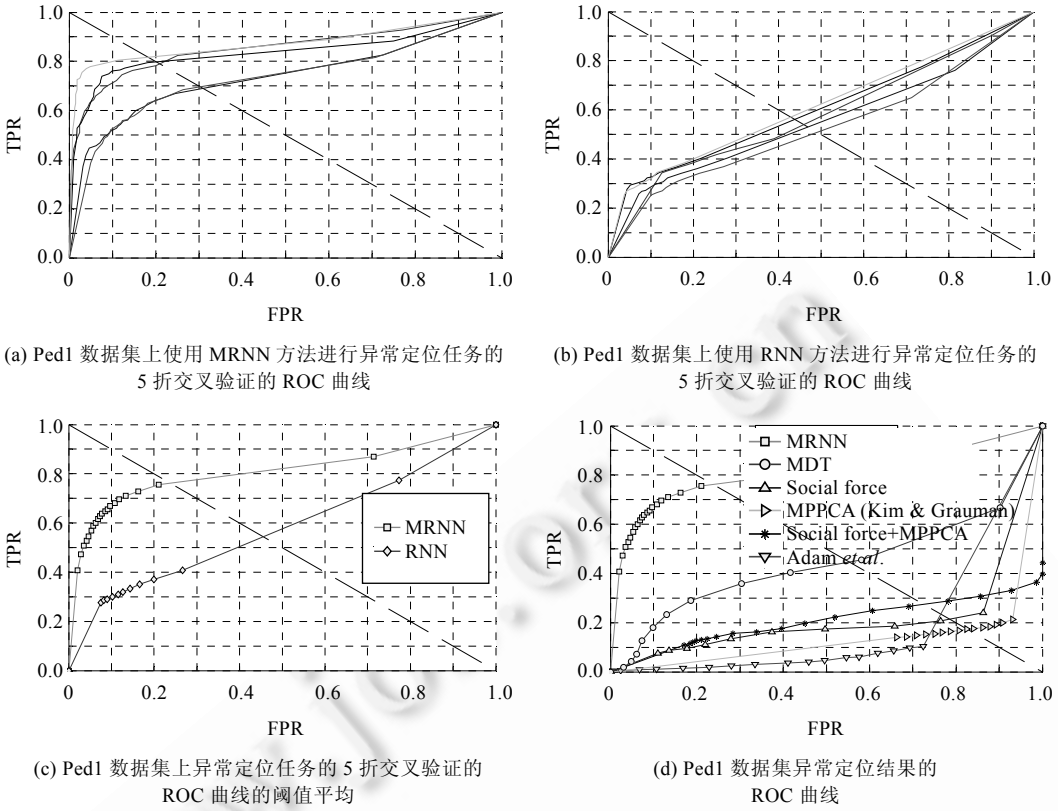
**Table 2** Experimental result of abnormal location: detection rate (DR) (%)

**表 2** 异常定位实验结果:检测率(DR)(%)

| | SF[13] | MPPCA[19] | SF-MPPCA | Adam *et al*.[27] | MDT[18] | RNN | MRNN |
|---|---|---|---|---|---|---|---|
| 定位 | 21 | 18 | 28 | 24 | 45 | 53 | 76 |

图 10 给出了异常定位结果的 ROC 曲线,其中,图 10(a)为 Ped1 数据集上 MRNN 方法在异常定位方面的 5 折交叉验证的 ROC 曲线;图 10(b)为 Ped1 数据集上 RNN 方法异常定位方面的 5 折交叉验证的 ROC 曲线;图 10(c)为 Ped1 数据集上 MRNN 和 RNN 在异常定位方面的 5 折交叉验证的 ROC 曲线的阈值平均,从图 10(c)中可以看出,MRNN 在异常定位方面比 RNN 要好;图 10(d)为 Ped1 数据集上 MRNN 和经典算法在异常定位任务方面的 ROC 曲线,从图 10(d)中可以看出,我们的方法在异常定位方面要远远优于经典方法.从以上实验结果可以发现,MRNN 比 RNN 在异常检测方面更好的原因在于:在异常定位中,目标个体的大小、异常事件的时空范围均存在较大差异.与 RNN 相比,MRNN 通过多个尺度领域间的关系可以更好地发现不同粒度的异常事件,从而获得了比 RNN 更加准确的异常定位效果.MRNN 比其他经典的方法在异常定位方面更好的原因在于:MRNN 通过同时建模空间维度的关系和时间维度的关系,可以更好地发现时间维度的相互依赖关系和多尺度空间维度的相互关系,从而获得比经典方法更好的定位效果.

© 中国科学院软件研究所 http://www.jos.org.cn




(a) Ped1 数据集上使用 MRNN 方法进行异常定位任务的 5 折交叉验证的 ROC 曲线

(b) Ped1 数据集上使用 RNN 方法进行异常定位任务的 5 折交叉验证的 ROC 曲线

(c) Ped1 数据集上异常定位任务的 5 折交叉验证的 ROC 曲线的阈值平均

(d) Ped1 数据集异常定位结果的 ROC 曲线

Fig.10 ROC curves for abnormal localization

图 10 异常定位结果的 ROC 曲线

## 4 结束语

在人群异常事件检测和定位任务中存在着两种类型的异常情况——空间异常事件和时间异常事件.通过对时间维度的关系和空间维度的关系进行建模,有助于异常事件的检测和定位.为此,在密集场景时空建模的基础上,本文提出了一种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.与经典算法的对比实验,证明了本方法的有效性.同时,由于多尺度时间递归神经网络模型的普适性,我们的方法不仅可以用于异常事件检测,而且可以用于具体事件检测或者辅助人群中的追踪和定位任务.虽然本方法的检测速度受限于光流算法,但是基于 GPU 加速等策略,仍然可以达到实时检测异常的目的.另外,鉴于 MRNN 是基于有监督的方法,需要大量的人工标注,在实际应用中可以使用众包的方式来进行人工标注,并且每隔一段时间对检测错误的视频帧进行标注,将其纳入到训练集中,并更新模型,以此来不断提升异常检测和异常定位效果.

**References**:

[1] Zhan B, Monekosso DN, Remagnino P, Velastin SA, Xu LQ. Crowd analysis: A survey. Machine Vision and Applications, 2008, 19(5-6):345−357. [doi: 10.1007/s00138-008-0132-4]

[2] Junior JCSJ, Musse SR, Jung CR. Crowd analysis using computer vision techniques. IEEE Signal Processing Magazine, 2010,27(5): 66−77. [doi: 10.1109/MSP.2010.937394]

[3] Thida M, Yong YL, Climent-Pérez P, Eng HL, Remagnino P. A literature review on video analytics of crowded scenes. In: Proc. of the Intelligent Multimedia Surveillance. Berlin, Heidelberg: Springer-Verlag, 2013. 17−36. [doi: 10.1007/978-3-642-41512-8_2]

© 中国科学院软件研究所 http://www.jos.org.cn

[4] Li T, Chang H, Wang M, Ni BB, Hong RC, Yan SC. Crowded scene analysis: A survey. IEEE Trans. on Circuits and Systems for Video Technology, 2014,25(3):367−386. [doi: 10.1109/TCSVT.2014.2358029]

[5] Kong D, Gray D, Tao H. A viewpoint invariant approach for crowd counting. In: Proc. of the 18th Int'l Conf. on Pattern Recognition. 2006. 1187−1190. [doi: 10.1109/TCSVT.2014.2358029]

[6] Kratz L, Nishino K. Tracking with local spatio-temporal motion patterns in extremely crowded scenes. In: Proc. of the Computer Vision and Pattern Recognition (CVPR). 2010. 693−700. [doi: 10.1109/CVPR.2010.5540149]

[7] Benabbas Y, Ihaddadene N, Djeraba C. Motion pattern extraction and event detection for automatic visual surveillance. Journal on Image and Video Processing, 2011,2011(1):413−447. [doi: 10.1155/2011/163682]

[8] Solmaz B, Moore BE, Shah M. Identifying behaviors in crowd scenes using stability analysis for dynamical systems. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2012,34(10):2064−2070. [doi: 10.1109/TPAMI.2012.123]

[9] Cong Y, Yuan J, Tang Y. Video anomaly search in crowded scenes via spatio-temporal motion context. IEEE Trans. on Information Forensics and Security, 2013,8(10):1590−1599. [doi: 10.1109/TIFS.2013.2272243]

[10] Reddy V, Sanderson C, Lovell BC. Improved anomaly detection in crowded scenes via cell-based analysis of foreground speed, size and texture. In: Proc. of the Computer Vision and Pattern Recognition Workshops (CVPRW). 2011. 55−61. [doi: 10.1109/CVPRW.2011.5981799]

[11] Cong Y, Yuan J, Liu J. Sparse reconstruction cost for abnormal event detection. In: Proc. of the Computer Vision and Pattern Recognition (CVPR). 2011. 3449−3456. [doi: 10.1109/CVPR.2011.5995434]

[12] Cheng KW, Chen YT, Fang WH. Abnormal crowd behavior detection and localization using maximum sub-sequence search. In: Proc. of the 4th ACM/IEEE Int'l Workshop on Analysis and Retrieval of Tracked Events and Motion in Imagery Stream. ACM Press, 2013. 49−58. [doi: 10.1145/2510650.2510655]

[13] Mehran R, Oyama A, Shah M. Abnormal crowd behavior detection using social force model. In: Proc. of the Computer Vision and Pattern Recognition (CVPR). 2009. 935−942. [doi: 10.1109/CVPR.2009.5206641]

[14] Chen DY, Huang PC. Visual-Based human crowds behavior analysis based on graph modeling and matching. Sensors Journal, IEEE, 2013,13(6):2129−2138. [doi: 10.1109/JSEN.2013.2245889]

[15] Antic B, Ommer B. Video parsing for abnormality detection. In: Proc. of the Computer Vision (ICCV). 2011. 2415−2422. [doi: 10.1109/ICCV.2011.6126525]

[16] Wu S, Moore BE, Shah M. Chaotic invariants of lagrangian particle trajectories for anomaly detection in crowded scene. In: Proc. of the Computer Vision and Pattern Recognition (CVPR). 2010. 2054−2060. [doi: 10.1109/CVPR.2010.5539882]

[17] Zhang D, Gatica-Perez D, Bengio S, Mccowan L. Semi-Supervised adapted hmms for unusual event detection. In: Proc. of the Computer Vision and Pattern Recognition (CVPR). 2005. 611−618. [doi: 10.1109/CVPR.2005.316]

[18] Mahadevan V, Li W, Bhalodia V, Vasconcelos N. Anomaly detection in crowded scenes. In: Proc. of the Computer Vision and Pattern Recognition (CVPR). 2010. 1975−1981. [doi: 10.1109/CVPR.2010.5539872]

[19] Kim J, Grauman K. Observe locally, infer globally: A space-time MRF for detecting abnormal activities with incremental updates. In: Proc. of the Computer Vision and Pattern Recognition (CVPR). 2009. 2921−2928. [doi: 10.1109/CVPR.2009.5206569]

[20] Kratz L, Nishino K. Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models. In: Proc. of the Computer Vision and Pattern Recognition (CVPR). 2009. 1446−1453. [doi: 10.1109/CVPR.2009.5206771]

[21] Li W, Mahadevan V, Vasconcelos N. Anomaly detection and localization in crowded scene. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2014,36(1):18−32. [doi: 10.1109/TPAMI.2013.111]

[22] Cong Y, Yuan J, Liu J. Abnormal event detection in crowded scenes using sparse representation. Pattern Recognition, 2013,46(7):1851−1864. [doi: 10.1016/j.patcog.2012.11.021]

[23] Thida M, Eng HL, Dorothy M, Remagnino P. Learning video manifold for segmenting crowd events and abnormality detection. In: Proc. of the Computer Vision (ACCV 2010). Berlin, Heidelberg: Springer-Verlag, 2011. 439−449. [doi: 10.1007/978-3-642-19315-6_34]

[24] Andrade EL, Blunsden S, Fisher RB. Modelling crowd scenes for event detection. In: Proc. of the 18th Int'l Conf. on Pattern Recognition. 2006. 175−178. [doi: 10.1109/ICPR.2006.806]

© 中国科学院软件研究所 http://www.jos.org.cn

[25] Benabbas Y, Ihaddadene N, Djeraba C. Motion pattern extraction and event detection for automatic visual surveillance. Journal on Image and Video Processing, 2011,2011:Article ID 163682. [doi: 10.1155/2011/163682]

[26] Wu S, Wong HS, Yu Z. A Bayesian model for crowd escape behavior detection. IEEE Trans. on Circuits and Systems for Video Technology, 2014,24(1):85−98. [doi: 10.1109/TCSVT.2013.2276151]

[27] Adam A, Rivlin E, Shimshoni I, Reinitz D. Robust real-time unusual event detection using multiple fixed-location monitors. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2008,30(3):555−560. [doi: 10.1109/TPAMI.2007.70825]

[28] Roshtkhari MJ, Levine MD. An on-line, real-time learning method for detecting anomalies in videos using spatio-temporal compositions. Computer Vision and Image Understanding, 2013,117(10):1436−1452. [doi: 10.1016/j.cviu.2013.06.007]

[29] Thida M, Eng HL, Remagnino P. Laplacian eigenmap with temporal constraints for local abnormality detection in crowded scenes. IEEE Trans. on Cybernetics, 2013,43(6):2147−2156. [doi: 10.1109/TCYB.2013.2242059]

[30] Loy CC, Xiang T, Gong S. Salient motion detection in crowded scenes. In: Proc. of the 5th IEEE Int'l Symp. on Communications Control and Signal Processing (ISCCSP). 2012. 1−4. [doi: 10.1109/ISCCSP.2012.6217836]

[31] Xiong G, Wu X, Chen Y, Ou Y. Abnormal crowd behavior detection based on the energy model. In: Proc. of the IEEE Int'l Conf. on Information and Automation (ICIA). 2011. 495−500. [doi: 10.1109/ICINFA.2011.5949043]

[32] Sodemann AA, Ross MP, Borghetti BJ. A review of anomaly detection in automated surveillance. IEEE Trans. on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 2012,42(6):1257−1272. [doi: 10.1109/TSMCC.2012.2215319]

[33] Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature, 1986,323(6088):533−536.

[34] http://www.svcl.ucsd.edu/projects/anomaly

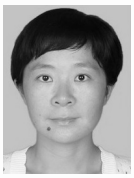
[35] Fawcett T. ROC graphs: Notes and practical considerations for researchers. Technical Report, HPL-2003-4, Palo Alto: HP Laboratories, 2003.

**蔡瑞初**(1983－),男,浙江温州人,博士,副教授,CCF 高级会员,主要研究领域为数据挖掘,机器学习,信息检索.

**谢伟浩**(1989－),男,硕士生,CCF 学生会员,主要研究领域为计算机视觉,机器学习.

**郝志峰**(1968－),男,博士,教授,博士生导师,主要研究领域为机器学习,人工智能.

**王丽娟**(1978－),女,博士,讲师,主要研究领域为高维数据聚类分析,多分类器融合.



**温雯**(1981－),女,博士,副教授,CCF 会员,主要研究领域为机器学习,模式识别,信息检索.

© 中国科学院软件研究所 http://www.jos.org.cn