

基于神经网络的密集人群视频异常检测

黄 荷 俞亚萍 张之江

(上海大学通信与信息工程学院 上海 200072)

摘要: 人群场景分析是当今智能监控技术研究领域关注的一个热点,尤其在拥挤的环境中,行人间严重的遮挡、光照分布不均以及人群分布的多样性,使其具有很大的挑战性。本文针对视频的特点,提出了两种描述人群运动信息的、适用于卷积神经网络的输入特征图,用于检测密集人群场景中的异常事件。首先,以提取的光流场为基础,分别构建出堆叠光流图和脉线图。其次,将不同类型的输入特征图作为训练样本分别送入卷积神经网络进行训练。接着,将训练出的模型作为异常检测器来判断相应类型特征图的测试样本的正异常情况。最后,通过实验比较分析不同类型特征图的优劣,并且证明了该方法可以有效的检测异常事件。

关键词: 异常事件检测; 密集场景; 堆叠光流; 脉线

中图分类号: TP751 **文献标识码:** A **国家标准学科分类代码:** 510.4050

Anomaly detection in dense crowd video based on neural network

Huang He Yu Yaping Zhang Zhijiang

(School of Communication and Information Engineering, Shanghai University, Shanghai 200072, China)

Abstract: Crowd scene analysis is a highly focused research area of the intelligent surveillance. Due to severe occlusions between individuals, different illumination and diverse crowd distributions, it's a challenging task in crowded environments. In this paper, two kinds of input feature maps of convolutional neural network (CNN) are proposed to characterize crowd motion for abnormal event detection in crowd videos. First, stacked optical flow map and streaklines map are constructed based on optical flow field. Next, different kinds of feature maps are fed into CNN to train models which are then treated as a classifier for anomaly detection. Finally, the experiments, conducted on publicly available datasets, evaluate the results of different feature maps and show the effectiveness of the proposed method.

Keywords: abnormal event detection; crowded scene; stacked optical flow; streaklines

0 引言

随着人口的增长和人们活动的日益丰富,现实生活中的密集场景越来越常见。视频监控系統被运用在很多公共场所,比如地铁站、飞机场、广场、街道等,作为城市安全保障的重要依据,并由此产生了大量的视频。由人力来对这些视频进行监控处理是非常耗时且容易发生漏报误报的现象。因此,对监控视频的自动分析和检测技术的研究得到越来越多的关注。

近年来,对场景的理解或分析,已经成为计算机视觉领域的一个热门的研究方向,尤其是针对密集人群场景的异常检测研究。异常检测可以分为全局异常检测和局部异常检测。全局异常检测^[1-5]旨在找出视频中出现异常事件的帧,而局部异常检测^[6-9]不仅仅要找到出现异常的帧,更要在图像帧中定位出出现异常的区域。本文研究的人群异常事件检测属于全局异常检测问题。

目前,根据提取特征的方式的不同,密集人群场景的异常研究主要分为以下几类:

1) 基于时空立方体的研究。由于密集场景中人群遮挡严重,对每一个行人单独进行分析是不现实的,因此很多研究将人群作为一个整体进行分析。例如,Kratz 等人^[6]将视频分为许多时空立方体,在每个立方体上提取时空梯度特征,并以该特征为基础对局部时空运动模式聚类,接着结合隐马尔科夫模型实现密集场景中多目标的稳定预测与异常检测。Li 等人^[10]在 Kratz 的基础上在时空立方体中加入了上下文信息,实现异常检测。

2) 基于直方图的研究。Chaudhry 等人^[11]提出了方向光流直方图特征。Kaltsa 等人^[12]串联梯度方向直方图和群体方向直方图对场景建模,实现异常检测。

3) 基于运动学的研究。Mehran 等人^[13]将人群运动看作粒子流运动,结合社会力模型,将人群的个人运动和相互

运动用牛顿力学来表征,构建词袋模型进行人群异常检测。Wu 等人^[14]利用拉格朗日粒子轨迹混沌不变原理来探测人群异常行为。

上述方法都是在底层特征的基础上,结合运动学模型或统计模型来手工地构造高维特征。近年来,深度学习的方法在图像领域得到广泛的应用,如物体分类^[15]、物体骨架提取^[16]、人脸识别^[17]、文字识别^[18-19]等。本文研究的对象为视频,视频比图像多了时间轴上的信息,所以这些模型不能直接用于视频场景的异常检测。Sermanet 等人^[20]证明了深度模型提取的特征比手工特征具有更好的性能。针对以上特点,本文设计了包含运动信息的特征图,结合卷积神经网络(convolutional neural network, CNN),能够提高密集人群场景下异常事件检测的精度。

1 基于运动特征的密集人群 CNN 模型

本文系统框架如图 1 所示。对于输入的视频序列,首先提取视频的运动特征,本文用光流场表示视频的运动特征。然后利用计算出的光流构建堆叠光流图和脉线图。最后将堆叠光流图和脉线图分别作为 CNN 的输入数据,从而实现视频场景的异常检测。

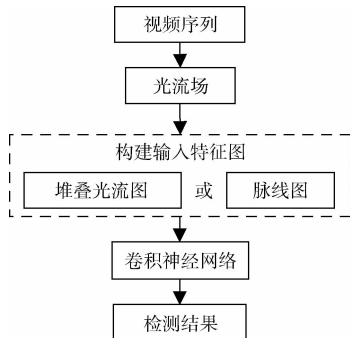


图 1 基于神经网络的人群异常检测系统

1.1 堆叠光流图

光流场是图像中所有像素的二维瞬时速度场,是空间运动物体的三维速度矢量在二维图像上的投影,十分适用于人群密集的场景视频分析以及对前景运动对象的特征检测。因为本文的特征图是以光流为基础构建的,所以光流估计的准确度对系统性能有很大的影响。本文采用 Brox 提出的光流估计方法计算光流场。

用 $O(t)$ 表示视频序列在第 t 帧求得的光流场,则光流的水平方向和垂直方向的分量可分别表示为 $O_u(t)$ 和 $O_v(t)$ 。这两个分量可以近似看作是光流图的两个通道。堆叠光流图的构建方法为在视频中选取连续 L 帧光流图,每个光流图都含有两个通道,将这 $2L$ 个通道堆叠起来,则 L 帧的堆叠光流图 $I_\tau(n)$ 数学表示为:

$$I_\tau(2k-1) = O_u(\tau+k-1) \quad (1)$$

$$I_\tau(2k) = O_v(\tau+k-1) \quad (2)$$

式中: $k \in [1, L]$, τ 为堆叠光流图的起始帧号, n 为堆叠光流图的通道编号。

1.2 脉线图

密集人群的群体运动也可以看做是流体的运动。在流体力学中,运用流场法来研究流体质点的运动,这种方法引入了脉线^[1]。脉线是在一定时间范围内相继通过一固定空间位置的流体质点依次串联而成的曲线,其可以在时间和空间上描述流体运动情况。

在图像中,将每一个像素点看做是脉线的初始位置,将第 i 帧图像中像素点 p 处出现的质点在 t 时刻的位置记为 $(x_i^p(t), y_i^p(t))$, 其中 $i, t = 0, 1, 2, \dots, L$ 。质点的移动路径表示为:

$$x_i^p(t+1) = x_i^p(t) + O_i^x(x_i^p(t), y_i^p(t)) \quad (3)$$

$$y_i^p(t+1) = y_i^p(t) + O_i^y(x_i^p(t), y_i^p(t)) \quad (4)$$

式中: O_i^x 和 O_i^y 代表 t 时刻位置 $(x_i^p(t), y_i^p(t))$ 处光流的水平分量和垂直分量。

图像中,长度为 L 的初始位置为 p 的脉线 S 可以表示为一系列点的集合,即:

$$S^p(L) = \{(x_L^p(L), y_L^p(L)), (x_{L-1}^p(L), y_{L-1}^p(L)), \dots, (x_0^p(L), y_0^p(L))\} \quad (5)$$

脉线图 $SL_\tau(p, n)$ 的数学表示为:

$$SL_\tau(p, 2k-1) = x_{L-k}^p(L) - x_{L-k+1}^p(L) \quad (6)$$

$$SL_\tau(p, 2k) = y_{L-k}^p(L) - y_{L-k+1}^p(L) \quad (7)$$

式中: $k \in [1, L]$, n 表示脉线图的通道编号。

1.3 密集人群 CNN 模型

本文参考文献[21]设计了密集人群的 CNN 结构,具体描述如表 1 所示。另外,在每一个卷积层之后都分别接了一个修正线性单元(rectified linear unit, ReLU),在全连接层 Fc1 和 Fc2 后分别接了 dropout 层避免训练过程中会出现的过拟合现象。

除此之外,本文还采取了数据扩增的方法来平衡正负样本数的不均匀,同时也达到避免过拟合的效果。具体操作是:对所有异常样本采用水平翻转、垂直翻转和旋转操作;对随机选取的部分正常样本水平翻转、垂直翻转或旋转。

表 1 卷积神经网络层结构

层类型	层名称	核大小	层输出数
卷积层	Conv1	7×7	48
采样层	Pool1	3×3	--
卷积层	Conv2	5×5	128
采样层	Pool2	3×3	--
卷积层	Conv3	3×3	256
卷积层	Conv4	3×3	256
卷积层	Conv5	3×3	256
全连接层	Fc1	--	2 048
全连接层	Fc2	--	512
全连接层	Fc3	--	2

2 实验与结果分析

为了使实验具有一定的代表性,本文分别在UMN数据集和PET2009数据集上进行实验,评估不同输入特征图的性能,并且将文本提出的方法与现有的方法进行比较。

2.1 UMN数据集

UMN数据集是由明尼苏达大学提供的,包含了3个不同的场景,共11段视频序列,其分辨率为 240×320 ,如图2(a)~(c)所示。每个序列以正常事件开始,以异常事件结束。在UMN数据集中,把行人在场景中随意行走作为正常事件,而行人在同一时间突然四散逃跑作为异常事件。在UMN数据集中,3个场景分别包含2个、6个和3个视频序列,从中分别选取1个、3个和2个序列作为训练样本,其余的作为测试样本。为了评价异常检测的结果,本文采用精度作为评价指标,即判断正确的样本数与总样本数的比值。

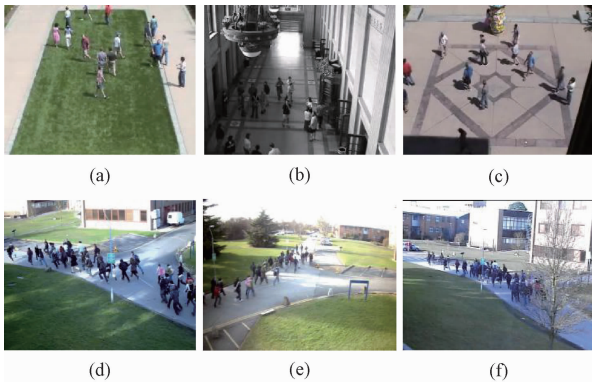


图2 UMN数据集和PETS2009数据集

实验将光流图、8帧堆叠光流图、16帧堆叠光流图以及长度为8和16的脉线图,共5种特征图分别作为CNN的输入,将输出结果做对比,检测结果如表2所示。

表2 不同输入特征图的结果比较

特征图	精度/%
光流图	97.47
8帧堆叠光流图	98.23
16帧堆叠光流图	98.88
脉线图(长度8)	98.42
脉线图(长度16)	97.73

实验结果显示,本文构建的堆叠光流图与脉线图比一般的光流图具有更好的检测结果,因为一般光流图只能捕获瞬时的运动信息,无法捕获一段时间内的运动信息。比较相同长度的堆叠光流图与脉线图,可得堆叠光流图具有更好的检测效果。同时,长度较长的脉线图的检测精度较低。分析可得,UNM数据集中行人分布比较稀疏,存在许

多光流为0的区域间隙,这导致脉线图较难捕获到完整的运动信息。

为了评估本文方法的有效性,将其与现有方法的比较结果列于如表3所示。从表3可以看出本文方法具有更好的检测结果。

表3 算法结果对比

方法	精度/%
力场模型	81.04
混沌不变量	87.91
社会力模型	85.09
稀疏重构	84.70
贝叶斯模型	96.40
本文方法	98.45

2.2 PETS2009数据集

PETS2009数据集包含不同类型的人群活动,每种人群活动都由多个摄像头从不同视角进行拍摄。该数据集比UMN数据集更加复杂,因为其场景中人群的行为是由正常到异常逐渐过渡的,没有很明显的区分度,且场景的光照分布不均匀,对光流的准确提取造成了极大的困难,如图2(d)~(f)所示。在PETS2009数据集中,正常事件是一群人沿着道路从一头走向另一头,而某时刻人群发生的快速奔跑行为作为异常事件。本文选取了从3个不同视角拍摄的9段视频序列作为实验样本,每次从中选取其中两个视角拍摄的序列作训练样本,另外一个作为测试样本,重复3次。检测结果如表4所示。

表4 不同输入特征图的结果比较

特征图	视角1	视角2	视角3
光流图	76.21	77.17	87.46
8帧堆叠光流图	80.06	93.31	96.28
16帧堆叠光流图	94.12	85.97	99.85
脉线图(长度8)	91.45	88.85	99.63
脉线图(长度16)	94.57	81.45	99.89

实验结果显示,本文构建的堆叠光流图与脉线图比一般的光流图具有更好的检测结果。对于视角1,较长的脉线图表现出较好的检测结果;对于视角2,由于光照影响了光流的提取,脉线图的检测结果劣于堆叠光流图;对于视角3,所有特征图都能很好地捕获人群连续运动信息。对3个视角的检测结果进行比较,可以得出行人间的遮挡、光照影响和人群密度不均对检测结果有一定的影响。在人群运动连续且人群密度分布均匀的场景下,脉线图能获得更好的结果。

与现有方法的比较结果如表5所示,从表5可以看出本文方法具有较好的检测结果。

表5 不同输入特征图的结果比较

方法	视角1	视角2	视角3
力场模型	37.74	37.74	37.74
混沌不变量	56.60	83.02	81.13
社会力模型	63.21	70.76	52.83
贝叶斯模型	92.45	83.02	89.62
本文方法	90.05	87.40	98.91

3 结 论

异常检测是密集人群场景分析研究中的热点及难点。目前,针对该问题的研究多侧重于提取底层特征并转化为高维特征进行分类。近年来,深度学习在计算机视觉领域得到广泛的应用,且深度特征表现出优于手工特征的性能。传统的神经网络多应用于二维图像,针对视频的特点,本文提出了两种含有人群运动信息的神经网络输入特征图,即堆叠光流图和脉线图,并将其分别送入卷积神经网络,实现异常事件的检测。通过实验评估了两种特征图的性能,并证明了该方法比现有的方法具有更好的识别结果。

另外,对于场景尺度不同、人群密度分布不均、光照不均等问题,本方法还有一定的局限性。克服这些缺点,将是课题组的下一步研究方向。

参考文献

- [1] MEHRAN R, MOORE B E, SHAH M. A streakline representation of flow in crowded scenes [C]. European Conference on Computer Vision, 2010.
- [2] SOLMAZ B, MOORE B E, SHAH M. Identifying behaviors in crowd scenes using stability analysis for dynamical systems[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(10): 2064-2070.
- [3] SU U, YANG H, ZHENG S, et al. The large-scale crowd behavior perception based on spatio-temporal viscous fluid field [J]. IEEE Transactions on Information Forensics and Security, 2013, 8(10): 1575-1589.
- [4] BENABBAS Y, IHADDADENE N, DJERABA C. Motion pattern extraction and event detection for automatic visual surveillance[J]. Eurasip Journal on Image and Video Processing, 2011, 2011(1):1-15.
- [5] 殷梦娇,俞亚萍,张之江. 基于四元数模型的密集人群视频特征提取[J]. 电子测量技术, 2016, 39(7):72-75.
- [6] KRATZ L, NISHINO K. Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models [C]. IEEE Conference on

- Computer Vision and Pattern Recognition, 2009.
- [7] LI W, MAHADEVAN V, VASCONCELOS N. Anomaly detection and localization in crowded scenes[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(1):18-32.
- [8] ROSHRKHARI M J, LEVINE M D. An on-line, real-time learning method for detecting anomalies in videos using spatio-temporal compositions [J]. Computer Vision and Image Understanding, 2013, 117(10): 1436-1452.
- [9] 张旭光,王梦伟,左佳倩,等. 介观小团体运动聚类的人群异常检测[J]. 仪器仪表学报, 2015, 36(5): 1106-1114.
- [10] LI N, WU X, XU D, et al. Spatio-temporal context analysis within video volumes for anomalous-event detection and localization[J]. Neurocomputing, 2015(155): 309-319.
- [11] CHAUDHRY R, RAVICHANDRAN A, HAGER G, et al. Histograms of oriented optical flow and binet-cauchy kernels on nonlinear dynamical systems for the recognition of human actions [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2009:1932-1939.
- [12] KAL TSA V, BRIASSOULI A, KOMPATSIARIS I, et al. Swarm intelligence for detecting interesting events in crowded environment [J]. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, 2015, 24(7): 2153-2166.
- [13] MEHRAN R, OYAMA A, SHAH M. Abnormal crowd behavior detection using social force model[C]. Computer Vision and Pattern Recognition, 2009.
- [14] WU S, MOORE B E, SHAH M. Chaotic invariants of Lagrangian particle trajectories for anomaly detection in crowded scenes[C]. Computer Vision and Pattern Recognition, 2010:2054-2060.
- [15] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Image net classification with deep convolutional neural networks [C]. International Conference on Neural Information Processing Systems, 2012:1097-1105.
- [16] SHEN W, ZHAO K, JIANG Y, et al. Object skeleton extraction in natural images by fusing scale-associated deep side outputs[C]. Computer Vision and Pattern Recognition, 2016:222-230.
- [17] TAIGMAN Y, YANG M, RANZATO M, et al. Deep face: Closing the gap to human-level performance in face verification[C]. Computer Vision and Pattern Recognition, 2014:1701-1708.

- [18] 吴国新,丁春艳,徐小力,等. 东巴经典古籍象形文字智能识别研究[J]. 电子测量与仪器学报, 2016, 30(11):1774-1779.
- [19] 曲景影,孙显,高鑫. 基于CNN模型的高分辨率遥感图像目标识别[J]. 国外电子测量技术, 2016, 35(8): 45-50.
- [20] SERMANET P, EIGEN D, ZHANG X, et al. Overfeat: Integrated recognition, localization and detection using convolutional networks [J]. Eprint ArXiv, 2013.
- [21] CHATFIELD K, SIMONYAN K, VEDALDI A, et al. Return of the devil in the details: Delving deep into convolutional nets[J]. Computer Science, 2014, arXiv:1405.3531.

作者简介

黄荷(通讯作者),1991年出生,硕士研究生,主要研究方向为密集人群场景分析、模式识别。

E-mail:wihw@t.shu.edu.cn

俞亚萍,1992年出生,硕士研究生,主要研究方向为人群异常检测、计算机视觉。

张之江,1968年出生,工学博士,教授,主要研究方向为计算机视觉、模式识别。