CN 43-1258/TP
ISSN 1007-130X

计算机工程与科学
Computer Engineering & Science

第 42 卷第 8 期 2020 年 8 月
Vol. 42, No. 8, Aug. 2020

文章编号:1007-130X(2020)08-1393-13

# 智能视频异常事件检测方法综述*

王思齐,胡婧韬,余　广,祝　恩,蔡志平
(国防科技大学计算机学院,湖南 长沙 410073)

**摘　要**:视频异常事件检测问题是计算机视觉领域的重要研究课题之一,旨在基于模式识别和计算机视觉方法智能地从监控视频中自动检测出需要关注的异常事件或行为,在实际生活中有广泛的应用和巨大的潜在需求,是人工智能技术落地的重要方向之一。同时,近年来以深度学习为代表的新兴机器学习技术及其在各个领域中取得的巨大成功,极大地启发了各类先进技术在视频异常事件检测问题中的应用。首先回顾了视频异常事件检测问题的定义和面临的主要挑战,随后从视频异常检测包含的 3 个最主要的技术环节(视频事件提取、视频事件表示、视频事件建模与检测)对当前主流视频异常事件检测技术进行了介绍,并对其各自的优缺点进行了分析和总结。最后,介绍视频异常检测领域中常用的基准测试数据集和相应的评价指标,对比当前主流方法的视频异常事件检测性能,对这些方法进行讨论并给出结论和展望。

**关键词**:视频异常检测;机器学习;人工智能;前景提取;特征提取;表示学习;正常事件建模
**中图分类号**:TP391　　**文献标志码**:A
**doi**:10.3969/j.issn.1007-130X.2020.08.009

# A survey of video abnormal event detection

WANG Si-qi, HU Jing-tao, YU Guang, ZHU En, CAI Zhi-ping
(School of Computer, National University of Defense Technology, Changsha 410073, China)

**Abstract**: Video anomaly detection is one of the most significant research tasks in computer vision area. It aims to intelligently identify the events that do not conform to expected behavior based on pattern recognition and computer vision methods. Video anomaly detection is widely applied and there is an enormous potential demand in modern society. Meanwhile, inspired by the successful achievements in various area of emerging deep learning technologies, more and more newly-emerged methods are conducted on video anomaly detection problem. Firstly, we retrospect the definition and main challenges of video anomaly detection. Secondly, we introduce the mainstream video anomaly detection methods from three primary technical steps (video event extraction, video event representation, video event modeling and detection) of video anomaly detection, and conclude their advantages as well as drawbacks respectively. Finally, we introduce the benchmark datasets and evaluation metrics of video anomaly detection, compare the performance of mainstream methods and give conclusions and prospects.

**Key words**: video anomaly detection; machine learning; artificial intelligence; foreground extraction; feature extraction; representation learning; modeling normal events

* 收稿日期:2019-12-31;修回日期:2020-03-23
基金项目:国家重点研发项目(2018YFB1003203);湖南省自然科学基金(2020JJ5673);国防科技大学科研计划(ZK20-10)
通信地址:410073 湖南省长沙市国防科技大学计算机学院
Address: School of Computer, National University of Defense Technology, Changsha 410073, Hunan, P. R. China

## 1 引言

面对现实生活中不断涌现的各类安全威胁和形形色色的突发情况,视频监控在公共安全、交通路况、市政管理等各个领域扮演着不可或缺的重要角色。随着以“天网监控系统”[1]为代表的各类视频监控系统建设的不断完善,各类视频监控设备监控也已经全面深入到现代社会公共场所的各个角落。然而,快速增长的视频监控设备每时每刻产生的海量视频数据对基于人工判读的视频异常事件发现带来了巨大的挑战,使后者在成本、效率甚至准确率上都已经变得越来越难以为继。因此,发展能够不依赖于人工判读、基于机器学习和计算机视觉方法来自动从监控视频中发现异常情况的智能视频异常事件检测技术,对于降低人力物力成本,提高监测效率,增强监控安全可靠性具有极其重要的意义和作用。在现代社会日益错综复杂的安全形势下,智能视频异常事件检测技术在许多现实生活场景中都具有迫切的需求,其应用前景光明,且具备客观的潜在商业价值,得到了来自学术界和工业界越来越高的重视。智能视频异常检测技术作为智能安防领域的核心任务之一,受到了来自商汤、格林深瞳等新兴人工智能公司以及华为、海康、大华等传统巨头的持续关注,也使其成为近年来兴起的人工智能浪潮的重要落地方向之一。

智能视频异常事件检测旨在基于各类机器学习和计算机视觉方法自动地检测和定位监控视频中各类违反常规的事件或行为,比如人群不正常的奔跑或聚集,人行道上车辆的出现等。相较于计算机视觉领域中的一些经典的目标检测任务(如行人检测、文字检测、人脸检测等),视频异常事件检测任务具有以下显著的特殊性:

(1)异常事件的定义具有抽象性,其内涵较为丰富,意义比较模糊(即“违反常规情况的事件”),这使得异常事件检测并不像经典的目标检测任务一样具有语义清晰且无二义性的检测对象,其待检测的异常事件往往并不特指某一种或者几种对象。例如,人行道上行驶的汽车和长时间徘徊逗留的行人都是需要注意的异常事件。

(2)异常事件往往具有很强的不可预测性。视频异常事件检测通常需要遵循一个“开放世界”假设(即所有不符合训练数据中出现的正常事件的情况都视为待检测的异常事件),而不能像一般的经典目标检测任务一样采用“闭合世界”假设(把检测的异常对象局限于训练数据给定的特定种类对象)。

(3)异常事件的稀疏性。由于异常事件本身被定义为是反常的,这就从本质上决定了视频中的异常事件发生的频率要远小于正常事件发生的频率,因而使得收集异常事件的数据要远远难于收集正常事件数据,甚至在一些情况下根本不可能预先收集到异常事件的数据。

(4)异常事件的定义具有相对性。在视频中,不同的对象和事件根据其所处的时空上下文环境的不同,往往具有不同的异常程度判定。例如,高速公路上行驶的汽车是正常事件,而在人行道上穿行的汽车则是需要注意的异常事件。

以上这些特殊性都使得视频异常事件检测相较于经典的目标检测任务更加困难。此外,相对于其他领域的异常检测任务(如网络流量异常检测),视频异常事件检测是一个面向自然图像/视频的计算机视觉任务,这使得它和其他计算机视觉任务一样需要面对真实场景下的各类复杂因素的挑战,比如光照变化、模糊、形变、拥挤场景、镜头抖动等。同时,从图像或者视频帧低层次的像素中提取出具备人类能够理解的语义信息的高层次特征来表示各类视频事件也是一个十分具有挑战性的任务。以上因素使得视频异常事件检测一直是计算机视觉中的一个十分具有挑战性的任务,至今仍有待进一步的探索。

## 2 视频异常事件检测实验设定

由于视频异常事件检测相较于经典的目标检测任务的特殊性,使得收集数量充分、种类齐全的异常事件数据集在现实中通常不具有可操作性,因此视频异常事件检测往往不能采用最常见的、效果最好的监督式分类技术作为解决方案(监督式分类技术要求使用同时包含标注好的正常和异常事件的数据进行训练)。根据使用的实验设定不同,现有的视频异常事件检测技术一般分为以下 3 种类型:

(1)半监督视频异常事件检测。半监督视频异常事件检测使用仅包含正常事件的视频作为训练视频来构建一个正常事件模型,而在测试环节则将所有偏离该正常事件模型的事件判定为异常事件。半监督视频异常事件检测技术假设完全没有任何异常事件作为先验信息,最符合异常检测中的“开放世界”假设。因此,目前文献中的大部分方法都

属于这种类型,其中比较有代表性的工作可参见文献[2-6]。

(2)无监督视频异常事件检测。无监督视频异常事件检测完全不使用任何人为标注的数据(无论是正常事件还是异常事件),仅根据这些视频事件数据自身所展示出的性质和分布特点来找出其中最与众不同的异常事件。无监督视频异常事件检测的设定比半监督视频异常事件检测更加贴近现实需求,对训练数据的要求远低于后者,但是其难度也要高于后者,属于近年来视频异常事件检测中新兴的研究方向。其中比较有代表性的工作可参见文献[7-9]。

(3)弱监督视频异常事件检测。弱监督视频异常事件检测使用少量或者具有弱标注的包含异常事件的视频(弱标注即不精确标注出异常行为的像素或者视频帧位置,仅在粗粒度层面标注出某一个视频是否包含异常行为,例如标注一整个视频是否包含异常而不具体到某一帧)来作为训练数据,旨在克服半监督视频异常事件检测完全忽视已知异常事件的先验信息的问题,更加贴近于现实生活中异常检测的设定。弱监督视频异常事件检测也是视频异常事件检测中新出现的研究方向,目前此类型的方法较少,其具有代表性的工作可参见文献[10]。

## 3 视频异常事件检测技术环节

现有视频异常事件检测技术通常都包含3个共同的技术环节:

(1)视频事件提取:由于监控视频摄像头绝大部分情况下处于静止状态,视频事件提取环节旨在剔除监控视频中每一帧里始终静止不变的背景部分,提取出其中需要关注的正在运动或者具有运动可能的前景对象,并将这些前景组织为待分析的基本视频事件单元。

(2)视频事件表示:在提取出视频中的前景部分并将其构建的基本单元作为待分析的视频事件后,需要进一步从这些视频事件中提取出有判别性的特征来作为这些视频事件的表示,从而方便利用这些特征来进一步判别出哪些事件属于异常事件。

(3)视频事件建模:根据从视频事件中提取出的特征表示构建出一个视频事件的模型,该模型能够对视频事件的特点(纹理、速度、分布等)进行描述,最终利用构建好的模型发现视频中包含的异常事件。

本节将从上面3个技术环节出发,对现有视频异常事件检测解决方案进行介绍。

### 3.1 视频事件提取

监控视频中的视频帧通常含有大量重复的、始终静止的背景。通常情况下,这些背景部分是不需要关注的,因为异常事件往往体现在发生运动或者有潜在发生运动可能的前景对象上。因此,针对背景部分进行运算开销庞大且没有必要。此外,背景部分也可能引入大量噪声和冗余信息,干扰视频事件的建模并降低异常检测的效率和质量。因此,现有许多方法的常用做法是先提取出前景,再按照一定的形式将提取出的前景组织成一定的结构作为视频异常事件检测任务中最基本的视频事件单元。其中,视频事件提取包含2个关键的技术要点。

#### 3.1.1 前景区分

提取视频事件的基础在于区分视频帧上哪些像素属于前景,哪些像素属于背景,比较典型的方法包括:

(1)梯度法(又称帧间差分法)。

该类方法是实现前景区分最简单的方法,旨在通过判断每一个像素位置上是否发生了运动,即相邻两帧之间相同像素位置上的灰度值是否发生剧烈变化,来确定该像素是否属于前景所在的区域。文献[4]利用相邻两帧相减计算每一个视频帧在时间方向上的梯度的模值来作为该像素位置的运动强度的表示,并以此作为区分前景和背景像素的依据。文献[11]将相邻帧相减得到的梯度进行二值化,并进一步通过形态学滤波和连通性分析实现更加完整的前景区分。此类方法虽然简单且容易实现,但是容易受到噪声的影响,且不能区分出短暂静止的前景。

(2)背景减除法。

该类方法先通过对输入若干视频帧中的背景进行建模得到背景模型,再将每帧图像和背景模型图像进行相减提取出其中明显与背景模型不同的部分作为前景。例如,文献[12,13]将视频中的背景和前景部分分别视为矩阵中的低秩和高秩部分,利用鲁棒主成分分析法[14]求解出其中的低秩部分作为背景模型,但需要输入较多的视频帧用于保证建模效果,且鲁棒主成分分析计算开销较高,而文献[5,15]则利用了经典的ViBe(Visual Backgroud extractor)方法[16]来对背景进行建模,能够在保持一定计算效率的前提下得到比简单的梯度法更加准确的前景区分效果。

(3)基于深度神经网络的方法。

近年来,深度神经网络[17]得到了迅猛的发展,其在目标检测、语义分割等一系列计算机视觉任务上都取得了巨大的突破。深度神经网络的优异表现自然启发了研究者将其应用在前景区分上。例如,文献[18,19]使用在大规模数据集上预训练出来的全连接卷积网络来从视频帧上找出具有显著特征相应的像素位置作为前景,而文献[20]则直接使用在大规模目标检测数据集上预训练好的深度目标检测网络来给出包含前景物体的约束框,将约束框内的像素作为待提取的前景。此类在大规模数据集上预训练好的深度神经网络能够取得较好的前景区分效果,并且能够通过 GPU 等硬件的辅助实现非常高效的前景区分。然而,这种策略显著的缺点在于这些预训练的深度模型往往都受限于"闭合世界"的假设——它们只能够检测出在训练时学习过的类别的物体;一旦出现训练数据之外的新类别的前景物体,它们就无法将其成功区分出来。

### 3.1.2 前景提取

前景提取的目的是在区分每一个像素是否属于前景之后,根据这些前景像素的位置将其按照一定的形式提取出来并构建成为基本的视频事件单元。目前较为常见的前景提取方法主要有以下几种:

(1)滑动窗口法。

滑动窗口法是最简单高效的前景提取方法,其通常使用一个滑动窗口按照一定步长在视频帧上进行从左至右、从上到下的扫描,并将其中符合一定条件的窗口内提取出来的前景块作为一个视频事件。例如,文献[4]将窗口内像素时间方向上梯度模值之和大于一定阈值的窗口中的前景块提取出来作为视频事件基本单元,而文献[5]则将前景像素所占比例超过 40%的窗口视作视频事件的基本单元。除此之外,滑动窗口法一般会将视频帧缩放到多个不同尺度,从而提取到包含不同尺度大小前景的前景块,使得能够更加全面地提取出所有可能的视频事件的基本单元。另外,在一些文献中(如文献[3])还会将与当前提取出的前景块在时间或者空间位置上相邻的前景块一并提取出来,组合在一起成为时空立方体,并以此作为视频事件基本单元,从而更好地捕捉视频中的时空上下文信息。

图 1 展示了文献[5]利用滑动窗口法提取视频事件基本单元的全过程。虽然此种策略简单高效,但是滑动窗口法的主要缺陷在于对前景中的对象往往是非常粗略的提取,会将一个独立完整的前景(例如行人、车辆)划分在几个不同的前景块里,这阻碍了对于前景物体的精确建模,在一定程度上影响了异常事件检测效果。同时,多个尺度前景的提取往往会产生过多的前景块,造成较大的计算开销。

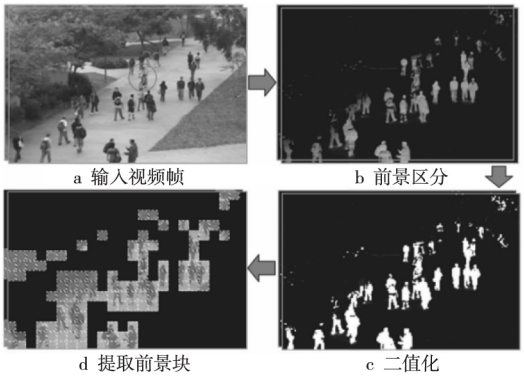


Figure 1 Sliding window method for video event extraction[5]

图 1 滑动窗口法提取视频事件[5]

(2)目标检测法。

基于深度神经网络的目标检测[21]近年来取得了长足的进步,其不仅能够识别出图像中包含的不同对象所属的种类,还能同时给出一系列边界框用于精确地定位出这些前景物体的位置,这使得其能够直接同时实现前景区分与前景提取的功能。自然而然,预训练好的目标检测器能够作为优良的前景提取工具。目前,许多在大规模数据目标检测数据集(例如 Microsoft COCO)上预训练的目标检测器(例如 Faster R-CNN[22], YOLO[23], RetinaNet[24])均已经能够在保证接近实时的情况下以较高的准确率鲁棒地检测出大部分日常生活中常见的物体,甚至在一些比较困难的场景(例如拥挤的场景)中也能够较为准确地定位各类不同尺度的前景对象。因此,近年来一些文献[20,25]已经开始将预训练好的目标检测器用到视频异常事件检测中进行前景提取,并将使用边界框提取出的前景块归一化为统一的大小,作为视频事件的基本单元。如前所述,目标检测法的优点虽然能较为完整准确地得到前景目标,但其无法检测视频帧中出现的预训练数据集中没有的目标,这也是其最为致命的缺陷。图 2 展示了使用目标检测法在视频帧上进行前景提取的结果。

在进行前景区分和前景提取后,视频前景往往会以缩放到同一尺度大小的前景图像块或者时空立方体的形式被提取出来,作为视频事件的基本单

Figure 2 Foreground bounding box obtained by object detection

图 2 使用目标检测法得到的前景边界框

元,方便后续视频事件的表示和建模,也会有助于对检测出的异常事件进行定位。值得一提的是,随着具有较强建模能力的深度神经网络的提出和发展,近年来一些视频异常事件检测方法省略了视频事件提取的步骤,直接对原始视频帧进行处理。例如,文献[6]引入了医疗图像分割领域的 U-Net 用于直接预测整幅视频帧中的异常事件,文献[26]设计了深度自编码器来直接重建整幅视频帧。

### 3.2 视频事件表示

视频事件提取环节得到的视频事件基本单元数据往往具有很高的数据维度。例如,一个较小的、边长为 32 的彩色视频前景块就总共包含3 072 个数据,这会带来很大的计算开销,且其中含有大量对于视频异常事件检测没有帮助的冗余信息。因此,我们需要从这些原始的视频事件基本单元中找出维度较低且具有判别性的特征,用其更好地表示得到的视频事件并服务于随后的视频异常事件检测。目前已有的视频异常事件检测技术所使用的视频事件表示方法一般可以分为 2 类。

#### 3.2.1 基于特征工程的视频事件表示

基于特征工程的视频事件表示方法一般是从人类专家对于视频事件表示的一些领域知识出发,设计相应的特征算子,再利用这些特征算子从提取出的视频事件基本单元计算出其特征表示。早期的视频异常事件检测方法一般都采用这种方法进行视频事件表示,其提取的特征可以分为高层次特征和低层次特征。

对于高层次特征而言,首先基于目标检测或跟踪等方法得到较高语义层次上的前景目标对象(例如行人、车辆等),随后对这些检测或者跟踪得到的目标对象中包含的一些含有明显高层次语义的信息(例如运动速度、轨迹长度、方向、形状)设计相应的特征来表示视频事件[27-30]。高层次特征的优点是在场景简单、目标清晰、分辨率较高的视频中效果较好,可解释性较强;但其缺点是在有较多拥挤的真实场景下会失效,这是因为目标间的相互遮挡、光照变化等因素,会导致目标检测与跟踪器难以对目标进行连续准确的捕捉。

基于提取高层次特征所遇到的困难,研究人员提出了更加鲁棒、计算更加简便的低层次特征,即不再首先获取高层次对象,而是直接在视频帧像素层面上设计基于各类统计信息的特征算子对视频进行表示。文献中常用的低层次特征有三维方向梯度[31]、方向梯度直方图[32]、光流直方图[33]及其变种(如多尺度光流直方图[3]和空间局部化光流直方图 SL-HOF(Spatially Localized Histogram of Optical Flow)[34])、局部二值化模式[35]以及局部梯度模式[36]等。低层次特征往往具有一些高层次特征所不具备的优点,例如光照变化下的不变性。图 3 展示了一个低层次特征算子的计算流程。

基于特征工程的视频事件表示具有较好的可解释性,但是其特征需要根据不同的场景由领域专家进行设计,门槛较高;同时,与诸如图像分类、目标检测等经典的计算机视觉任务一样,这些基于人类认知进行手工特征工程设计的视频表示方法往往并不能得到视频事件的最优表示,判别性较弱。

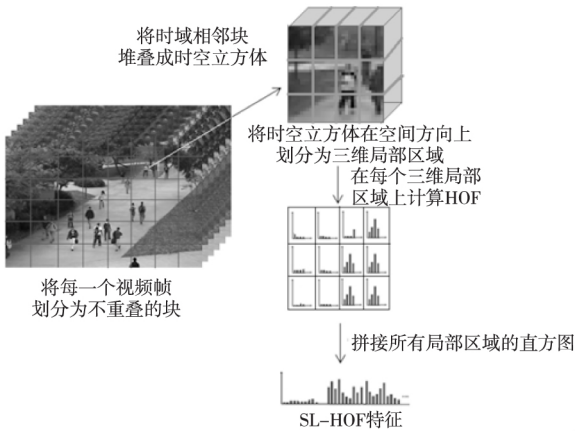


Figure 3 Process of spatially localized histogram of optical flow

图 3 空间局部化光流直方图特征算子的计算流程

#### 3.2.2 基于深度表示学习的视频事件表示

随着近年来深度学习的快速兴起,使用深度神经网络从输入的原始数据中直接学习出对于特定任务最优的特征表示的深度表示学习方法已经成为计算机视觉领域中最明显的发展趋势。这也使得近年来许多的视频异常事件检测方法开始逐步将深度学习作为视频事件表示的新选择[37]。文献

[5]作为这个方面的开创性工作,首次用深度堆叠去噪自编码器网络学习重建像素和光流时空立方体,并将自编码器的中心隐层或末尾全连接层的输出作为学习到的视频事件表示。文献[20,38]则分别使用了卷积深度自编码器和赢家通吃卷积深度自编码器(Convolutional Winner-take-all AUtoencoder)来替代由全连接层组成的深度堆叠去噪自编码器(如图4所示),因为卷积操作能够更好地捕捉二维图像上的空间结构,得到更好的视频特征。文献[39]利用多任务卷积神经网络提取行人面部区域的特征。文献[40]利用预训练的AlexNet模型提取图像特征。值得一提的是,最近更多基于深度神经网络的视频异常事件检测方法已经不再将深度神经网络学习到的视频事件表示单独提取出来,而是直接利用深度神经网络端到端得到视频事件的检测结果。

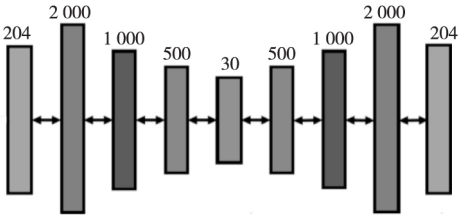


Figure 4 Schematic diagram of deep autoencoder

图4 深度自编码器网络的结构示意图

### 3.3 视频事件建模与检测

视频事件建模和检测是视频异常事件检测的核心技术环节,旨在根据训练数据中的视频事件建立相应的数学模型,从而实现在测试时对视频事件的异常程度进行打分,找出其中的异常事件。文献中提出了丰富的视频事件建模方法,其主要可以分为以下几类:

#### 3.3.1 基于经典机器学习技术的方法

此类视频事件建模方法使用前面提取出的视频事件的特征表示(通常以特征向量的形式进行表示)作为输入,利用经典机器学习算法进行建模,包括经典的稀疏编码、分类、聚类和离群点检测算法等。例如,基于稀疏编码技术,文献[3]将字典学习的方法引入视频异常检测任务中,作为视频异常事件的建模方法,其从给定的训练数据中学习出一组能够最优表示正常事件的字典,而无法用学习到的字典良好表示的事件被判定为异常。该方法随后衍生出一系列同样基于稀疏编码的变种[4,41-45]。基于单类别分类技术,文献[5]则将表示视频事件的特征向量输入单类别支持向量机建立异常检测模型。同时,单类别自编码器以及类似的判别式单分类分类器(例如单类别极限学习机[36]等)也是文献中广泛使用的用于视频异常事件检测的经典机器学习技术之一。除此之外,二分类或者多分类分类器也常用于对正常事件建模。例如,文献[20]将视频事件特征向量利用$k$均值聚类算法进行聚类,并为聚类得到的不同簇打上不同的伪标签,然后利用伪标签将一类作为正样本、其他类作为负样本循环训练$k$个支持向量机二分类器,支持向量机的输出作为检测异常的指标。类似地,文献[46]对所有局部活动的图袋进行$k$中值聚类处理后,使用标准的二分类支持向量机进行分类。基于聚类技术[47],文献[48]利用k均值聚类对视频事件特征进行聚类,并将少数集群作为异常结果。文献[49]利用近邻传播聚类进行字典学习。基于离群点检测技术,文献[25]用核密度估计对分类分数建立密度分布,并将密度的倒数作为异常值得分。

除了这几种最直观的建模方法外,文献中还涌现了非常丰富的视频事件建模方法。例如,文献[2]使用条件随机场来建模视频事件;文献[50]以社会力模型(Social Attribute-aware Force Model)来描述视频事件中前景对象中的相互作用;文献[51]使用层次化高斯混合模型来构建正常的视频事件概率模型;文献[12]使用概率图模型来解释正常视频事件中出现的前景对象;文献[7,8]将局部变化检测技术应用于无监督视频异常事件检测。

基于经典机器学习技术的视频异常事件检测方法往往具有较坚实的理论基础以及良好的可解释性,但是其要求必须根据不同的数据集和不同的场景进行视频事件的提取与表示。另外,部分经典的机器学习方法的计算复杂度较高、可扩展性较差,不利于处理较大规模的视频数据。

#### 3.3.2 基于深度神经网络的方法

此类方法利用深度神经网络对视频事件进行建模和预测,也是近年来视频异常事件检测领域重要的发展趋势。深度神经网络一般具有从诸如图像和视频的复杂数据之中自动学习到优良特征表示的能力,使其可以端到端地同时实现视频事件表示和视频事件建模与预测2个环节,避免了2个环节的衔接性问题,增加了实现的紧凑型和便利性。同时,深度神经网络自动学习特征表示的能力也增强了相应视频异常事件检测解决方案在不同数据集、不同场景之间的通用性。此类方法目前主要分为2种:

(1)基于重构的方法。

此类方法的核心思想是深度生成式神经网络

(如自编码器网络、对抗生成网络),通过训练使其学习到自动重构输入数据的能力。在视频异常事件检测任务中,最普遍的做法是在给定的正常视频事件数据上训练生成式深度神经网络,使其能够在测试阶段输入正常视频事件数据时得到良好的、具有较低重构误差的重构数据,而在输入没有训练过的异常事件数据时将得到较差的重构数据、较大的重构误差,通过区分性的重构误差实现对正常事件的建模和对异常事件的检测。目前,有许多利用或基于自编码器重构的视频异常检测方法[52]。文献[26]是此类方法的先驱者,实现了2种深度自编码器,其中一种利用传统特征算子(梯度方向直方图和光流方向直方图)表示的视频事件输入到全连接深度神经网络,另一种深度卷积自编码器直接以整个原始视频帧作为输入,并发现直接将原始视频帧作为输入的深度卷积自编码器能够获得优良的异常检测性能。文献[53]把只能处理二维图像的自编码器拓展为三维时空自编码器,可以更好地建立视频帧与视频帧之间的联系。文献[18]将提取出的视频事件输入到深度变分自编码器进行重构,这种更加鲁棒和强大的生成式神经网络保证了其效果。文献[54]设计了一种两路输出的自编码器网络,将重构单个视频帧和重构其对应的光流结合起来,并使重构网络和预测网络共享一个编码网络。文献[55]则结合自编码器网络和记忆增强网络使得模型对正常和异常视频帧的重构误差更具有区分性。

(2)基于预测的方法。

此类方法的出发点是将预料之外、无法准确预测的事件作为异常,符合人类对于异常的认知。其实现的核心思想是,将若干时间上相邻的视频帧作为历史信息输入到一个深度生成网络来预测这些视频帧的下一帧(即未来帧),并将预测值和实际值之间的差异作为判定异常的要素。一般而言,用正常数据训练的预测网络能准确地预测正常数据,而对于预测异常数据的表现较差。文献[53]首次在使用时空自编码器对视频帧进行重建时,引入了额外的预测未来分支,以改善异常检测的性能。

除了最常见的用自编码器网络生成未来帧的做法之外,在视频异常检测领域,另一种应用较广的预测网络是U-Net[56]。文献[6]首次引入U-Net作为生成未来帧的深度神经网络,其网络结构如图5所示。其中,U-Net通过降采样网络和上采样网络的跳层连接,融合了低层次特征和高层次特征,解决了梯度消失和每层信息不平衡等问题。文献[6]将前$t$帧输入到U-Net中以预测第$t+1$帧,同时在训练时加入最小化预测帧和实际帧的光流和梯度差异的目标函数作为运动约束,以及鼓励预测帧尽可能接近真实视频帧的对抗损失函数。在此基础上文献[57]则通过使用U-Net对重构历史帧的任务做出了改进。另外,除了直观地用历史帧预测未来帧的做法,文献中还出现了跨模态的预测方法,即使用一个模态的视频数据预测另一个模态的视频数据。文献[58]使用对抗生成网络[59]作为生成模型,在原始视频帧和其对应的光流场之间进行相互预测。文献[60]则将U-Net作为生成网络来从原始视频帧预测光流。

## 4 公开数据集

### 4.1 UCSD数据集

UCSD行人数据集[2]是视频异常事件检测研究中使用最广泛的主流数据集之一,它分为UCSD Pedestrian 1(Ped1)和UCSD Pedestrian 2(Ped2)2个数据集,是从加州大学圣地亚哥分校UCSD(University of California, San Diego)2个不同地点

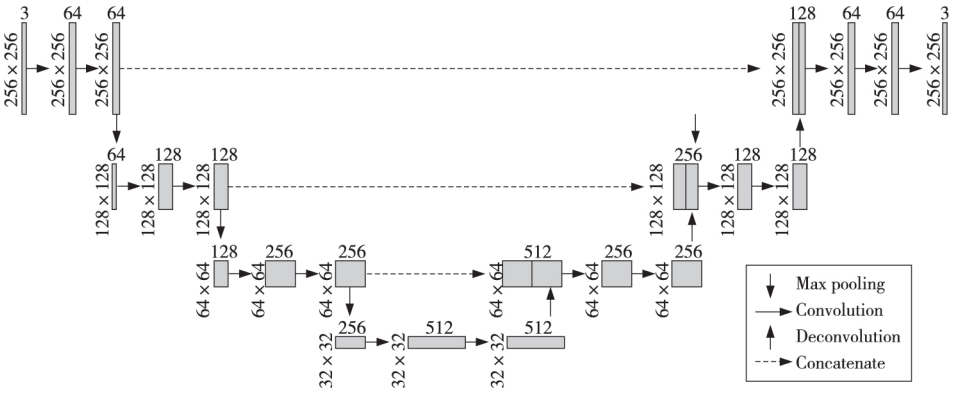


Figure 5 Schematic diagram of U-Net

图5 用于预测未来的U-Net网络结构

的人行道的视频监控中收集而来。在 UCSD Ped1 和 UCSD Ped2 数据集中,走路的行人被视为正常事件,而在人行道中出现的其他移动物体则被视为异常事件,例如在人行道上骑自行车、滑滑板、推手推车、驾驶汽车等。行人在非人行道区域(例如草坪)行走也被视作异常事件。此外,这 2 个数据集包含大量视频前景物体相互遮挡的拥挤场景,尤其是行人的遮挡,增加了检测异常事件的难度。测试集均提供了像素级别的标注,即在每个视频帧上标注出了异常像素。

UCSD Ped1 数据集包含 34 个训练视频样本和 36 个测试视频样本,总共包含 14 000 帧,每帧分辨率为 158×238 像素。UCSD Ped2 数据集也是一个广泛使用的数据集,包含 16 个训练视频序列和 12 个测试视频序列,其视频帧的分辨率为 240×360 像素,总共包含 4 560 帧。与 UCSD Ped1 数据集相比,Ped2 数据集具有更少的视频帧和更大的帧分辨率,这使得它相较于 Ped1 更为常用。在实际实验中,通常的做法是把 2 个数据集分开测试。图 6 展示了 UCSD Ped1 和 Ped2 数据集上具有代表性的正常事件(图 6a 和图 6c)和异常事件(图 6b 和图 6d)的视频帧。


a UCSD Ped1正常事件 b UCSD Ped1异常事件: 骑自行车
c UCSD Ped2正常事件 d UCSD Ped2异常事件: 卡车

Figure 6 Representative frames on UCSD Ped1 and Ped2 Datasets

图 6 UCSD Ped1 和 Ped2 数据集上具有代表性的视频帧

### 4.2 Avenue 数据集

Avenue 数据集[4]采集自香港中文大学 CUHK(The Chinese University of Hong Kong)校内一个走廊的监控摄像机。在此视频数据集中,行人在走廊上按正常方向行走是正常事件,而在走廊上扔书包、洒纸片、奔跑、推自行车等各种类型的行为被当成异常事件。为了接近现实世界中的实际情况,Avenue 数据集的测试视频中会发生一些轻微的相机抖动,从而使我们更加聚焦于异常的移动对象上。

Avenue 数据集包含共 15 328 帧的 16 个训练视频和共 15 324 帧的 21 个测试视频,其中一共有 47 个异常事件。视频帧的分辨率为 360×640 像素。同时,训练集包含少量的异常,测试集提供了像素级别的标注。

图 7 展示了 Avenue 数据集上具有代表性的正常事件(图 7a)和异常事件(图 7b~图 7d)的视频帧。


a Avenue正常事件 b Avenue异常事件: 奔跑
c Avenue异常事件: 推自行车 d Avenue异常事件: 错误方向

Figure 7 Representative frames on Avenue dataset

图 7 Avenue 数据集上具有代表性的视频帧

### 4.3 ShanghaiTech 数据集

ShanghaiTech 数据集[41]是目前为止最大的视频异常检测数据集,与其他数据集只包含一个场景不同(即一个固定的摄像头采集所有视频),ShanghaiTech 数据集一共包含了 13 个有不同光照条件和摄像头角度的场景,这使其成为文献中最具挑战性的数据集之一。训练集包含共 274 515 帧的 330 个视频,测试集包含共 42 883 帧的 107 个视频,包含 130 个异常事件。视频帧的分辨率为 480×856 像素。同时,测试集提供了像素级别的标注。

图 8 展示了 ShanghaiTech 数据集上具有代表性的正常事件(图 8a)和异常事件(图 8b~图 8d)的视频帧。


a ShanghaiTech正常事件 b ShanghaiTech异常事件: 翻越围栏
c ShanghaiTech异常事件: 骑自行车 d ShanghaiTech异常事件: 跑步、滑滑板

Figure 8 Representative frames on ShanghaiTech dataset

图 8 ShanghaiTech 数据集上具有代表性的视频帧

## 5 评价指标

异常检测的任务中,最常用的评估指标分为帧层次标准和像素层次标准[2]。对于帧层次标准,如果算法检测到视频帧中存在至少一个像素点为异常,则将视频帧视为异常;像素层次标准规定,一次对异常事件成功的检测被认定为至少超过40%的像素被检测为异常且真实标签也为异常。

在视频异常检测领域,为了进行定量比较,实验中通常计算相应受试者工作特征曲线ROC(Receiver Operating Characteristic)下的面积*AUC*(Area Under the corresponding ROC Curve)和等错误率*EER*(Equal Error Rate)以评估性能。具体而言,首先,根据分类器的预测结果对数据进行排序。然后,通过逐步改变判定异常的阈值(相当于"截断点"),将样本分为2部分,可计算得到一系列真阳率*TPR*(True Positive Rate)和假阳率*FPR*(False Positive Rate)。真阳率表示分类器预测是正样本且正确的(预测值等于真实值)占总正样本的比率。假阳率表示分类器预测是正样本但错误的(当前被错误分到正样本中实际是负样本)占总正样本的比率。

如图9所示,把得到的一组*FPR*设为横坐标的值,*TPR*设为纵坐标的值,可以绘制出ROC曲线。

ROC曲线下的面积(图9)的值代表正样本排在负样本之前的概率,取值在0~1,值越大表示异常样本越有可能排在负样本之前,即分类性能越好。

等错误率是当分类器的真阳率和假阳率满足$FPR=1-TPR$时,被错分的视频帧数量占所有的视频帧数量的比例,其数值越小表明方法的性能越好。具体来讲,*EER*是ROC曲线与ROC空间中对角线的交点,即图9中虚线与曲线的交点。

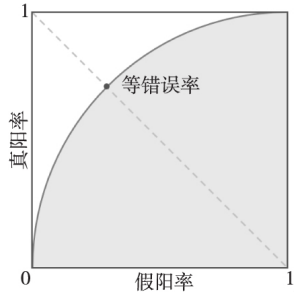


Figure 9 Schematic of ROC curve and *EER*

图9 ROC曲线和*EER*的示意图

## 6 主流方法实验效果对比

### 6.1 UCSD Ped1数据集实验效果对比

表1展示了主流方法在UCSD Ped1数据集上的实验效果,并按照实验设定的不同将方法分为半监督和无监督2类。其中,效果最佳的结果在表中加粗表示。

Table 1 Frame-level and pixel-level *EER* and *AUC* evaluation on UCSD Ped1 dataset

表1 UCSD Ped1数据集上帧层次和像素层次标准的*EER*和*AUC*结果对比表

| | 方法 | 评价指标 | | | |
|---|---|---|---|---|---|
| | | 帧层次 | | 帧层次 | |
| | | *EER* | *AUC* | *EER* | *AUC* |
| 半监督 | Adam等人[60] | 38.0 | 65.0 | 76.0 | 13.3 |
| | MPPCA[61] | 40.0 | 59.0 | 82.0 | 20.5 |
| | SF[62] | 31.0 | 67.5 | 79.0 | 19.7 |
| | SF+MPPCA[2] | 32.0 | 77.0 | 71.0 | 21.3 |
| | MDT[2] | 25.0 | 81.8 | 44.0 | 55.0 |
| | Bertini等人[63] | 31.0 | — | — | — |
| | SRC[3] | 19.0 | 86.0 | 54.0 | 46.1 |
| | Lu等人[4] | 15.0 | 91.8 | 40.9 | 63.8 |
| | HMDT-CRF[64] | 18.0 | — | **35.0** | 66.2 |
| | CAE[26] | 27.9 | 81.0 | — | — |
| | STAE[53] | 15.3 | 92.3 | — | — |
| | WTA-CAE[38] | 14.8 | 91.6 | 35.7 | 68.7 |
| | Liu等人[6] | — | 83.1 | — | — |
| | $S^2$-VAE[18] | 14.3 | — | — | **94.3** |
| | Ravanbakhsh等人[58] | **8.0** | **97.4** | **35.0** | 70.3 |
| 无监督 | Unmasking[8] | — | 68.40 | — | 52.50 |
| | Wang等人[9] | **29.20** | **77.80** | **46.30** | **55.50** |

### 6.2 UCSD Ped2数据集实验效果对比

表2展示了主流方法在UCSD Ped2数据集上的实验效果,并按照实验设定的不同将方法分为半监督和无监督2类。其中,效果最佳的结果在表中加粗表示。

### 6.3 Avenue数据集实验效果对比

表3展示了主流方法在Avenue数据集上的实验效果,并按照实验设定的不同将方法分为半监督和无监督2类。其中,效果最佳的结果在表中加粗表示。

Table 2 Frame-level and pixel-level *EER* and *AUC* evaluation on UCSD Ped2 dataset

表 2 UCSD Ped2 数据集上帧层次和像素层次标准的 *EER* 和 *AUC* 结果对比表

| | 方法 | 评价指标 | | | |
|---|---|---|---|---|---|
| | | 帧层次 | | 帧层次 | |
| | | *EER* | *AUC* | *EER* | *AUC* |
| 半监督 | Adam 等人[60] | 42.00 | 63.00 | — | — |
| | MPPCA[61] | 30.00 | 77.00 | — | — |
| | SF[62] | 42.00 | 63.00 | — | — |
| | SF+MPPCA[2] | 36.00 | 71.00 | — | — |
| | MDT[2] | 25.00 | 85.00 | 55.00 | — |
| | Bertini 等人[63] | 34.00 | — | — | — |
| | HMDT-CRF[64] | 18.50 | — | 29.90 | — |
| | OWC-MTT[65] | 13.90 | 94.00 | 21.10 | 83.60 |
| | CAE[26] | 21.70 | 90.00 | — | — |
| | STAE[53] | 16.70 | 91.20 | — | — |
| | TSC-sRNN[41] | — | 92.20 | — | — |
| | WTA-CAE[38] | **8.90** | 96.60 | **16.90** | **89.30** |
| | Liu 等人[6] | — | 95.40 | — | — |
| | Nguyen 等人[54] | — | 96.20 | — | — |
| | MemAE[55] | — | 94.10 | — | — |
| | Ravanbakhsh 等人[58] | 14.00 | 93.50 | — | — |
| | Ionescu 等人[20] | — | **97.80** | — | — |
| 无监督 | Unmasking[8] | — | 82.20 | — | — |
| | Wang 等人[9] | **8.90** | **96.40** | **19.40** | **85.90** |

Table 3 Frame-level and pixel-level *EER* and *AUC* evaluation on avenue dataset

表 3 Avenue 数据集上帧层次和像素层次标准的 *EER* 和 *AUC* 结果对比表

| | 方法 | 评价指标 | | | |
|---|---|---|---|---|---|
| | | 帧层次 | | 帧层次 | |
| | | *EER* | *AUC* | *EER* | *AUC* |
| 半监督 | Lu 等人[4] | — | 80.9 | — | — |
| | CAE[26] | 25.1 | 70.2 | — | — |
| | STAE[53] | 24.4 | 80.9 | — | — |
| | TSC-sRNN[41] | — | 81.7 | — | — |
| | WTA-CAE[38] | **24.2** | 82.1 | **45.2** | **55.0** |
| | Liu 等人[6] | — | 84.9 | — | — |
| | $S^2$-VAE[18] | — | 87.6 | — | — |
| | Nguyen 等人[54] | — | 86.9 | — | — |
| | Ionescu 等人[20] | — | **90.4** | — | — |
| 无监督 | Del Giorno 等人[7] | — | 78.3 | — | — |
| | Unmasking[8] | — | 80.6 | — | — |
| | Wang 等人[9] | **23.9** | **85.3** | **41.2** | **52.7** |

### 6.4 ShanghaiTech 数据集实验效果对比

表 4 展示了主流方法在 ShanghaiTech 数据集上的实验效果，并按照实验设定的不同将方法分为半监督和弱监督 2 类。其中，效果最佳的结果在表中加粗表示。由于 ShanghaiTech 数据集比前 3 个数据集更具有挑战性，可以看到不同方法的实验效果在此数据集上还存在提升空间。

Table 4 Frame-level *AUC* evaluation on ShanghaiTech dataset

表 4 ShanghaiTech 数据集上帧层次标准的 *AUC* 结果对比表

| | 方法 | *AUC*/% |
|---|---|---|
| 半监督 | MemAE[55] | 71.2 |
| | Ionescu 等人[20] | **84.9** |
| | CAE[26] | 60.9 |
| | TSC-sRNN[41] | 68.0 |
| | Liu 等人[6] | 72.8 |
| | Lu 等人[4] | 65.5 |
| 弱监督 | Sultani 等人[10] | **76.5** |

## 7 结束语

本文主要从 3 个模块对视频异常检测进行了介绍和分析：视频事件前景提取、视频事件表示和视频事件建模与检测。同时介绍了 3 个常用的公开数据集。随着深度学习的发展和兴起，现有视频异常检测技术越来越贴近于深度学习，借助于深度神经网络强大的建模能力，视频异常检测取得了突破性的进展。然而实验表明，在许多真实复杂的场景下，视频异常检测的性能还有待提升，并且现有建模方法更多地采用了基于重构的自编器，而许多深度学习相关的成果并没有迁移到这个领域。因此，结合更多前沿的深度学习技术实现更高效准确的检测，是视频异常检测领域一个非常有前景和巨大需求的研究方向。

### 参考文献：

[1] Zhao Ya-fei, Wang He. The review of "Sky Net" project planning[J]. Chinese Public Security, 2014(9): 181-183. (in Chinese)

[2] Mahadevan V, Li W X, Bhalodia V, et al. Anomaly detection in crowded scenes[C]// Proc of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2010: 1975-1981.

[3] Cong Y, Yuan J, Liu J. Sparse reconstruction cost for abnor-

mal event detection[C]// Proc of 2011 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2011: 3449-3456.

[4] Lu C W, Shi J P, Jia J Y. Abnormal event detection at 150 FPS in Matlab[C]// Proc of 2013 IEEE International Conference on Computer Vision, 2013: 2720-2727.

[5] Xu D, Yan Y, Ricci E, et al. Detecting anomalous events in videos by learning deep representations of appearance and motion[J]. Computer Vision and Image Understanding, 2017, 156: 117-127.

[6] Liu W, Luo W X, Lian D Z, et al. Future frame prediction for anomaly detection—A new baseline[C]// Proc of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 6536-6545.

[7] del Giorno A, Bagnell J A, Hebert M. A discriminative framework for anomaly detection in large videos[C]// Proc of European Conference on Computer Vision, 2016: 334-349.

[8] Ionescu R T, Smeureanu S, Alexe B, et al. Unmasking the abnormal events in video[C]// Proc of 2017 IEEE International Conference on Computer Vision, 2017: 2895-2903.

[9] Wang S Q, Zeng Y J, Liu Q, et al. Detecting abnormality without knowing normality: A two-stage approach for unsupervised video abnormal event detection[C]// Proc of the 26th ACM International Conference on Multimedia, 2018: 636-644.

[10] Sultani W, Chen C, Shah M. Real-world anomaly detection in surveillance videos[C]// Proc of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 6479-6488.

[11] Zhang Xia, He Zheng-ran. Video abnormality judgment based on grayscale and optical flow detection[J]. Chinese Journal of Electron Devices, 2019, 42(3): 718-721. (in Chinese)

[12] Antic B, Ommer B. Video parsing for abnormality detection[C]// Proc of 2011 International Conference on Computer Vision, 2011: 2415-2422.

[13] Cheng Hao-gui, Xu Le-ling, Tang Xu-qing. Hidden Markov model based non parametric Bayesian algorithm for video anomaly detection[J]. Control Engineering of China, 2019, 26(9): 1763-1769. (in Chinese)

[14] Wright J, Ganesh A, Rao S, et al. Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization[C]// Proc of 2009 3rd Annual Conference on Neural Information Processing Systems, 2009: 2080-2088.

[15] Huang Xin, Xiao Shi-de, Song Bo. Detection of vehicle's abnormal behaviors in surveillance video[J]. Computer Systems & Applications, 2018, 27(2): 125-131. (in Chinese)

[16] Barnich O, van Droogenbroeck M. ViBe. A universal background subtraction algorithm for video sequences[J]. IEEE Transactions on Image Processing, 2011, 20(6): 1709-1724.

[17] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015, 521(7553): 436-444.

[18] Wang T, Qiao M, Lin Z, et al. Generative neural networks for anomaly detection in crowded scenes[J]. IEEE Transactions on Information Forensics and Security, 2019, 14(5): 1390-1399.

[19] Sabokrou M, Fayyaz M, Fathy M, et al. Deep-anomaly: Fully convolutional neural network for fast anomaly detection in crowded scenes[J]. Computer Vision and Image Understanding, 2018, 172: 88-97.

[20] Ionescu R T, Khan F S, Georgescu M I, et al. Object-centric auto-encoders and dummy anomalies for abnormal event detection in video[C]// Proc of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 7842-7851.

[21] Liu L, Ouyang W L, Wang X G, et al. Deep learning for generic object detection: A survey[J]. International Journal of Computer Vision, 2020, 128: 261-318.

[22] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.

[23] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]// Proc of 2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.

[24] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]// Proc of 2017 IEEE International Conference on Computer Vision, 2017: 2980-2988.

[25] Hinami R, Mei T, Satoh S. Joint detection and recounting of abnormal events by learning deep generic knowledge[C]// Proc of 2017 IEEE International Conference on Computer Vision, 2017: 3619-3627.

[26] Hasan M, Choi J, Neumann J, et al. Learning temporal regularity in video sequences[C]// Proc of 2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016: 733-742.

[27] Fu Z Y, Hu W M, Tan T N. Similarity based vehicle trajectory clustering and anomaly detection[C]// Proc of IEEE International Conference on Image Processing 2005, 2005: II-602.

[28] Piciarelli C, Micheloni C, Foresti G L, et al. Trajectory-based anomalous event detection[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2008, 18(11): 1544-1554.

[29] Basharat A, Gritai A, Shah M. Learning object motion patterns for anomaly detection and improved object detection[C]// Proc of 2008 IEEE Conference on Computer Vision and Pattern Recognition, 2008: 1-8.

[30] Zhang T Z, Lu H Q, Li S Z. Learning semantic scene models by object classification and trajectory clustering[C]// Proc of 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009: 1940-1947.

[31] Kratz L, Nishino K. Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models[C]

//Proc of 2009 IEEE Conference on Computer Vision and Pattern Recognition,2009: 1446-1453.

[32] Roshtkhari M J,Levine M D. Online dominant and anomalous behavior detection in videos[C]//Proc of 2013 IEEE Conference on Computer Vision and Pattern Recognition, 2013: 2611-2618.

[33] Du Gui-ying,Chen Ming-jin. Research on anomaly detection algorithm of moving objects based on intelligent video analysis[J]. Video Engineering,2018,42(12): 23-26. (in Chinese)

[34] Wang S Q, Zhu E, Yin J P, et al. Anomaly detection in crowded scenes by SL-HOF descriptor and foreground classification[C]//Proc of 2016 23rd International Conference on Pattern Recognition,2016: 3398-3403.

[35] Hu X,Huang Y P,Gao X M,et al. Squirrel-cage local binary pattern and its application in video anomaly detection[J]. IEEE Transactions on Information Forensics and Security, 2019,14(4): 1007-1022.

[36] Wang S Q,Zhu E,Yin J P. Video anomaly detection based on ULGP-OF descriptor and one-class ELM[C]//Proc of 2016 International Joint Conference on Neural Networks, 2016: 2630-2637.

[37] Wu Pei-ji,Mei Xue,He Yi,et al. Method of detecting abnormal behavior in video sequences based on deep network models[J]. Laser & Optoelectronics Progress, 2019, 56(13):131101-1-131101-7. (in Chinese)

[38] Tran H T M,Hogg D. Anomaly detection using a convolutional winner-take-all autoencoder[C]//Proc of the British Machine Vision Conference,2017:1-13.

[39] Li Jun-jie,Liu Cheng-lin,Zhu Ming. Fast abnormal pedestrians detection based on multi-task CNN in surveillance video [J]. Computer Systems & Applications,2018,27(11): 78-83. (in Chinese)

[40] Lei Li-ying,Chen Hua-hua. Video anomaly detection based on AlexNet [J]. Journal of Hangzhou Dianzi University (Natural Science),2018,38(6): 16-21. (in Chinese)

[41] Luo W X,Liu W,Gao S H. A revisit of sparse coding based anomaly detection in stacked RNN framework[C]//Proc of 2017 IEEE International Conference on Computer Vision, 2017: 341-349.

[42] Zhao Y,Qiao Y,Yang J,et al. Abnormal activity detection using spatio-temporal feature and Laplacian sparse representation[M]//Neural Information Processing. Cham:Springer International Publishing,2015: 410-418.

[43] Zhao B, Li F-F,Xing E P. Online detection of unusual events in videos via dynamic sparse coding[C]//Proc of 2011 IEEE Conference on Computer Vision and Pattern Recognition,2011: 3313-3320.

[44] Xia L M,Hu X J, Wang J. Anomaly detection in traffic surveillance with sparse topic model[J]. Journal of Central South University,2018,25(9): 2245-2257.

[45] Zheng Hao,Liu Jian-fang,Liao Meng-yi. Human abnormal behavior detection and recognition based on hybrid algorithm in indoor video surveillance[J]. Computer Applications and Software,2019,36(7): 224-230. (in Chinese)

[46] Pan Zhi-an. Abnormal behavior detection algorithm of surveillance video based on bag of word[J]. Journal of Southwest China Normal University (Natural Science Edition), 2018,43(7): 60-66. (in Chinese)

[47] Liu Li-qi-ming,Xu Xiang-hua,Zhang Ling-jun. Video abnormal event detection based on improved dynamic clustering [J]. Artificial Intelligence and Robotics Research, 2018, 7(2): 78-88. (in Chinese)

[48] Liu Su,Sun Chen. Slow moving sparse target anomaly detection in museum monitoring video[J]. Science Technology and Engineering,2018,18(22): 84-89. (in Chinese)

[49] Hu Zheng-ping,Zhang Le,Yin Yan-hua. Video anomaly detection by AP clustering sparse representation based on spatial-temporal deep feature model[J]. Journal of Signal Processing,2019,35(3): 386-395. (in Chinese)

[50] Zhang Y H,Qin L,Yao H X,et al. Abnormal crowd behavior detection based on social attribute-aware force model [C]//Proc of 2012 19th IEEE International Conference on Image Processing,2012: 2689-2692.

[51] Cheng K W,Chen Y-T,Fang W-H. Video anomaly detection and localization using hierarchical feature representation and Gaussian process regression[C]//Proc of 2015 IEEE Conference on Computer Vision and Pattern Recognition,2015: 2909-2917.

[52] Cai De-xiu, Yang Da-wei. Adaptive anomaly detection method in video based on autoencon-der framework[J]. Microprocessors,2019,40(5): 60-64. (in Chinese)

[53] Zhao Y R,Deng B,Shen C,et al. Spatio-temporal autoencoder for video anomaly detection[C]//Proc of the 2017 ACM International Conference on Multimedia,2017: 1933-1941.

[54] Nguyen T N, Meunier J. Anomaly detection in video sequence with appearance-motion correspondence[C]//Proc of 2019 IEEE/CVF International Conference on Computer Vision,2019: 1273-1283.

[55] Gong D,Liu L Q,Le V,et al. Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection[C]//Proc of 2019 IEEE/CVF International Conference on Computer Vision, 2019: 1273-1283.

[56] Ronneberger O,Fischer P,Brox T. U-Net: Convolutional networks for biomedical image segmentation[C]//Proc of International Conference on Medical Image Computing and Computer-Assisted Intervention,2015: 234-241.

[57] Tang Y,Zhao L,Zhang S S,et al. Integrating prediction and reconstruction for anomaly detection[J]. Pattern Recognition Letters,2019,129:123-130.

[58] Ravanbakhsh M,Nabi M,Sangineto E,et al. Abnormal event detection in videos using generative adversarial nets[C]//Proc of 2017 IEEE International Conference on Image Processing,2017: 1577-1581.

[59] Goodfellow I J,Pouget-Abadie J,Mirza M,et al. Generative

adversarial nets[C]// Proc of the 27th International Conference on Neural Information Processing Systems, 2014: 2672-2680.

[60] Adam A, Rivlin E, Shimshoni I, et al. Robust real-time unusual event detection using multiple fixed-location monitors [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(3): 555-560.

[61] Kim J, Grauman K. Observe locally, infer globally: A space-time MRF for detecting abnormal activities with incremental updates[C]// Proc of 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009: 2921-2928.

[62] Mehran R, Oyama A, Shah M. Abnormal crowd behavior detection using social force model[C]// Proc of 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009: 935-942.

[63] Bertini M, Del Bimbo A, Seidenari L. Multi-scale and real-time nonparametric approach for anomaly detection and localization [J]. Computer Vision and Image Understanding, 2012, 116 (3): 320-329.

[64] Li W X, 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.

[65] Lin H H, Deng J D, Woodford B J, et al. Online weighted clustering for real-time abnormal event detection in video surveillance[C]// Proc of 2016 ACM International Conference on MultiMedia, 2016: 536-540.

## 附中文参考文献:

[1] 赵亚飞,王鹤.“天网”工程成环成网规划综述[J].中国公共完全(综合版),2014(9):181-183.

[11] 张霞,贺正然.基于灰度与关键帧光流检测的视频异常判断[J].电子器件,2019,42(3):718-721.

[13] 陈皓圭,许乐灵,唐旭清.隐式马尔科夫 HDP 非参数贝叶斯视频异常检测[J].控制工程,2019,26(9):1763-1769.

[15] 黄鑫,肖世德,宋波.监控视频中的车辆异常行为检测[J].计算机系统应用,2018,27(2):125-131.

[33] 都桂英,陈铭进.基于智能视频分析的运动目标异常行为检测算法研究[J].电视技术,2018,42(12):23-26.

[37] 吴沛佶,梅雪,何毅,等.基于深度网络模型的视频序列中异常行为的检测方法[J].激光与光电子学进展,2019,56(13):131101-1-131101-7.

[39] 李俊杰,刘成林,朱明.基于多任务 CNN 的监控视频中异常行人快速检测[J].计算机系统应用,2018,27(11):78-83.

[40] 雷丽莹,陈华华.基于 AlexNet 的视频异常检测技术[J].杭州电子科技大学学报(自然科学版),2018,38(6):16-21.

[45] 郑浩,刘建芳,廖梦怡.室内视频监控下基于混合算法的人体异常行为检测和识别方法[J].计算机应用与软件,2019,36(7):224-230.

[46] 潘志安.基于词袋模型的监控视频异常活动检测算法[J].西南师范大学学报(自然科学版),2018,43(7):60-66.

[47] 刘李启明,徐向华,张灵均.基于改进的动态聚类的视频异常事件检测[J].人工智能与机器人研究,2018,7(2):78-88.

[48] 刘速,孙晨.博物馆监控视频中慢速移动稀疏目标异常轨迹检测[J].科学技术与工程,2018,18(22):84-89.

[49] 胡正平,张乐,尹艳华.时空深度特征 AP 聚类的稀疏表示视频异常检测算法[J].信号处理,2019,35(3):386-395.

[52] 蔡德秀,杨大为.基于自编码器框架的自适应视频异常检测方法[J].微处理机,2019,40(5):60-64.

## 作者简介:

**王思齐**(1992 -),男,四川犍为人,博士,助理研究员,研究方向为模式识别和异常检测。**E-mail**:wangsiqi10c@gmail.com

**WANG Si-qi**, born in 1992, PhD, assistant research fellow, his research interests include pattern recognition, and outlier detection.

**胡婧韬**(1995 -),女,北京人,博士生,研究方向为计算机视觉和模式识别。**E-mail**:hujingtao17@nudt.edu.cn

**HU Jing-tao**, born in 1995, PhD candidate, her research interests include computer vesion, and pattern recognition.

**余广**(1996 -),男,四川广元人,硕士生,研究方向为异常检测。**E-mail**:179342757@qq.com

**YU Guang**, born in 1996, MS candidate, his research interest includes outlier detection.

**祝恩**(1976 -),男,湖南益阳人,博士,教授,CCF 会员(16689S),研究方向为机器学习和模式识别。**E-mail**:enzhu@nudt.edu.cn

**ZHU En**, born in 1976, PhD, professor, CCF member (16689S), his research interests include machine learning, and pattern recognition.

**蔡志平**(1975 -),男,湖南益阳人,博士,教授,CCF 会员(E200011640S),研究方向为计算机网络和人工智能。**E-mail**:zpcai@nudt.edu.cn

**CAI Zhi-ping**, born in 1975, PhD, professor, CCF member (E200011640S), his research interests include computer network, and artificial intelligence.