

基于卷积自编码器分块学习的视频异常事件检测与定位

李欣璐, 吉根林, 赵斌

(南京师范大学计算机与电子信息学院/人工智能学院, 南京 210023)

摘要: 视频异常事件检测与定位旨在检测视频中发生的异常事件,并锁定其在视频中发生的位置。但是视频场景复杂多样,并且异常发生的位置随机多变,导致发生的异常事件难以被精准定位。本文提出了一种基于卷积自编码器分块学习的视频异常事件检测与定位方法,首先将视频帧进行均匀划分,提取视频帧中每一块的光流和方向梯度直方图(Histogram of oriented gradient, HOG)特征,然后为视频中的不同图块分别设计卷积自编码器以学习正常运动模式特征,最后在异常事件检测过程中利用卷积自编码器的重构误差大小进行异常判断。该方法可以有效地针对视频不同区域进行特征学习,提升了异常事件定位的准确度。所提方法在UCSD Ped1、UCSD Ped2、CUHK Avenue三个公开数据集上进行实验,结果表明该方法能够准确定位异常事件,并且帧级别AUC(Area under the curve)平均提升了5.61%。

关键词: 视频异常事件检测;异常事件定位;分块学习;卷积自编码器;深度学习

中图分类号: TP391 **文献标志码:** A

Convolutional Auto-Encoder Patch Learning Based Video Anomaly Event Detection and Localization

LI Xinlu, JI Genlin, ZHAO Bin

(School of Computer and Electronic Information/School of Artificial Intelligence, Nanjing Normal University, Nanjing 210023, China)

Abstract: Video anomaly event detection and localization aim to detect abnormal events and lock its localization in the video. However, the video scenes are complex and diverse, and the localizations where anomaly events occur are random and changeable, which makes it difficult to accurately locate the occurred abnormal events. This paper proposes a video anomaly event detection and localization method based on convolutional auto-encoder patch learning. Firstly, we divide the video frames evenly into patches and extract the optical flow and the histogram of oriented gradient (HOG) feature of each patch. Then, at the different patches in the video, we individually design a convolutional auto-encoder to learn the feature in the normal motion mode. During the anomaly event detection process, the reconstruction loss of the convolutional auto-encoder is used for anomaly detection. The proposed method can effectively perform feature learning for different regions of the video and improve the accuracy of anomaly event localization. Experimental results on three public datasets, UCSD Ped1, UCSD Ped2, and CUHK Avenue, demonstrate that the frame level AUC (area under the curve) of this method is increased by 5.61% on average and can accurately locate anomaly events.

Key words: video anomaly event detection; anomaly event localization; patch learning; convolutional auto-encoder; deep learning

引 言

视频异常事件检测是计算机视觉的重要应用之一,它可以将潜在的危險事件检测出来,并发出警报,从而提高有关部门和机构的响应效率。目前视频异常事件检测得到了广泛的研究,人们提出了针对各种场景和检测要求的异常事件检测方法,其中的关键问题就是视频前景中运动物体的特征表示。由于视频场景具有多样性,且异常的定义具有主观不确定性,异常事件检测任务存在一定的挑战。例如,为了发现跌倒或打架斗殴等指定异常动作,研究者通常以视频中的行人交互问题为中心,重点提取行人的姿态特征和运动特征^[1-2],采用姿态估计^[3]和动作识别^[4]的方式检测视频中的行人是否存在异常动作。但是,在现实场景中发生的异常往往是不可预计的,而该类异常检测方法仅能检测出指定的人体异常动作,通用性不强。并且在人群密集的场景下,存在行人遮挡和视频分辨率低的问题,姿态估计和动作识别的准确度不高,直接影响了异常检测的准确性。本文针对基于姿态和动作的异常检测方法存在的不足,采用表观特征和运动特征共同表示前景物体的运动特征,将方向梯度直方图(Histogram of oriented gradient, HOG)作为表观特征,并选取深度神经网络提取出的光流作为运动特征。HOG特征的提取较为简单并且可以清晰表示视频画面中的纹理信息,能够检测出可疑物体。深度神经网络提取出的光流比传统光流法提取的光流更加准确且能够充分表示移动物体的运动状态,能够检测出运动异常。采用光流和HOG特征相结合的方式,不仅能检测出视频中的异常行为,还可以检测出非行人实体闯入视频区域的异常情况,增强了异常检测模型的普适性。

异常事件在现实场景中不常发生,因此正常样本的数量远大于异常样本的数量。针对正负样本数量不均的问题,本文采用卷积自编码器进行特征学习,仅用正常视频样本训练模型,无需对训练集和测试集进行重新划分。通过学习正常的行为模式,将与正常运动模式不匹配的情况设定为异常,可检测所有的非正常事件。异常检测的常用方法是将视频帧的特征提取出来之后,直接放入异常检测模型中进行特征学习。而本文考虑到在视频的不同区域,视觉信息和移动物体的运动状态差异较大,首先将视频帧均匀划分成互不重叠的图块,再对每一个图块进行特征提取。让卷积自编码器针对视频不同区域中的特征进行学习,在检测时判断视频某一位置上的图块是否存在异常事件,不仅使模型的学习内容更有针对性,同时能实现异常事件定位。

1 相关研究工作

视频异常事件检测可分为两个部分:视频特征的提取和异常事件检测模型的建立。视频特征提取对于异常事件检测的精确度起到关键作用,常用的特征主要分为手工设计的特征和深度模型提取的特征。Mahadevan等^[5]利用混合动态纹理(Mixtures of dynamic textures, MDT)对人群中的局部异常进行检测。张俊阳等^[6]考虑到异常事件通常发生在前景区域,利用自适应混合高斯模型进行背景分割,提取视频前景进行人群异常事件检测,减少计算量。光流可以较好地表示移动物体的运动特征,Cong等^[7]改进光流直方图,利用多尺度的光流直方图(Multi-scale histograms of optical flow, MHOF)捕捉视频中物体的运动信息,Colque等^[8]提出了基于光流场矢量的方向、模长、像素熵值的运动特征HOFME(Histogram of optical flow orientation and magnitude and entropy)。Mehran等^[9]提取光流信息中更高层次的语义特征,利用社会力模型(Social force, SF)进行人群异常检测。另一种具有高层语义的特征是移动物体的轨迹,常用于交通异常事件检测^[10],但是在存在遮挡时轨迹提取效果不佳。多种特征融合的方式比单一特征具有更强的表达能力,Xu等^[11]对视频前景图像块与光流场块进行像素级别的融合,形成新的图块,再将其输入去噪自编码器,以学习融合图块的深度特征。近年来深度学习快速发展,研究者常用深度模型对视频进行特征提取。异常检测任务中常用的深度模型有:全卷积神经网络^[12]、稀疏去噪自编码器(Sparse denoising auto-encoders, SDAE)^[13]等。此外,双流卷积网络^[14]可提取出人体的动

作特征来对视频中行人的动作进行识别,可用于异常动作检测。人体骨架特征可以较好地表示行人姿态,Morais等^[15]利用深度神经网络提取出视频中行人的骨架特征,通过单个行人的骨架序列进行异常行为检测,但是在人群密集的场景下异常检测效果不佳。

根据不同的建模角度,视频异常事件检测模型可以分为基于传统概率推断的模型和基于深度学习的模型。混合概率主元分析(Mixture of probabilistic principal component analysis, MPPCA)是基于概率推断的一种经典模型, Kim等^[16]将MPPCA和时空马尔可夫随机场(Markov random field, MRF)相结合以检测人群异常行为。Sabokrou等^[17]在将每个视频划分为立方块之后,计算相邻块的相似度值,并利用高斯分布对正常的视频块进行建模,在测试时计算提取的特征和正常特征的距离大于阈值的判定为异常。基于传统概率推断的建模方法常用高斯混合模型(Gaussian mixture model, GMM)和其他的模型进行组合,以提高异常检测的准确度, Feng等^[18]将多个GMM堆叠起来对正常行为模式建模, Li等^[19]首先假设正常事件的表现及运动特征的深度表示都服从多变量高斯分布,结合自编码器和生成对抗网络进行联合训练,对多变量高斯分布进行拟合。基于深度学习的模型中,常用的两种方法是重构和分类。Hasan等^[20]利用传统的自编码器对HOG特征和HOF特征进行学习,并用卷积自编码器(Convolutional auto-encoders, Conv-AE)来学习视频序列中的正常模式。Chong等^[21]更关注视频帧在时间维度上的变化,利用时空自编码器(SpatioTemporal-AE)学习视频片段中的时空特征,在异常检测时利用重构误差计算异常得分,得分低的判为异常视频帧。袁静等^[22]在SDAE中增加梯度差约束,以提升自编码器的异常检测效果。利用分类进行异常检测通常需要在训练时就对正常样本和异常样本进行标注,因此会面临正负样本不均的问题。Xu等^[11]使用单类支持向量机(One-class support vector machine, OCSVM)以生成决策边界,对深度融合表示后的特征进行分类。Sultani等^[23]提出了弱监督算法框架,利用深度多实例排序方法训练分类网络,减少了异常样本少带来的问题。2020年, Nawaratne等^[24]将模糊聚合与时空自编码器的异常检测方法结合起来,提出时空增量学习(Incremental spatio-temporal learner, ISTL)方法实现异常检测模型的更新,不断提高模型的性能。

2 视频异常事件检测与定位方法

2.1 处理流程

本文提出的视频异常事件检测与定位方法处理流程如图1所示。首先将视频样本拆分成视频帧序列 $\{\text{frame}_1, \text{frame}_2, \dots, \text{frame}_t, \dots, \text{frame}_n\}$,然后将视频帧均匀划分成 $M \times N$ 个互不重叠的图块,每个图块的大小均为 $S \times S$ 。由于运动特征可以反映出视频前景中的运动物体的运动状态,可检测出奔跑、抛掷物品等运动异常,而表现特征可以体现视频画面中是否存在异常物体,例如卡车、自行车等非行人对象,因此需要对视频图块中的运动特征和表现特征进行提取。

在提取运动特征时,将第 t 个视频帧中某一位置的图块 $\text{patch}_{(t,i)}$ 和第 $t+1$ 个视频帧中对应位置的图块 $\text{patch}_{(t+1,i)}$ 共同输入到FlowNet2.0^[25],可得到第 t 个视频帧中该图块的光流。FlowNet2.0是一个基于卷积神经网络的深度模型,对于两个视频帧的输入可以计算出前一帧相对于后一帧的光流。对光流的幅值和方向进行颜色编码,移动物体的运动方向用不同的颜色进行表示,而速度大小则用颜色深浅进行表示,最终可得到1个RGB三通道的彩色光流图。对视频帧中每1个图块计算HOG,可提取视频帧图块的表现特征。

对于训练样本和测试样本先进行视频帧的划分和视频帧图块的特征提取;对于某一图块的光流和HOG特征,分别设置一个异常检测卷积自编码器(Anomaly detection convolutional auto-encoders, AD-ConvAE)进行训练和测试。一个位置图块上的AD-ConvAE仅关注该视频位置区域里的人群运动情况,利用分块学习的方式可以更有效地学习局部特征。在训练过程中视频仅包含正常样本,AD-Con-

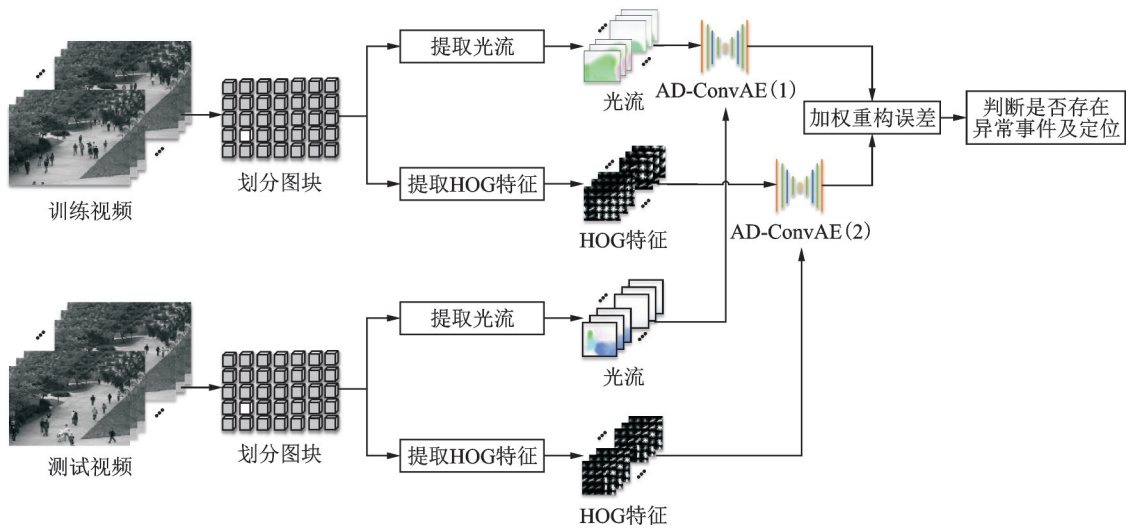


图1 视频异常事件检测与定位方法处理流程

Fig.1 Pipeline of video anomaly event detection and localization method

vAE通过视频帧图块的光流和HOG特征学习某一区域的正常运动模式。在测试时,将测试视频中该区域图块的光流和HOG特征放入AD-ConvAE中进行重构,根据光流的重构误差和HOG特征的重构误差计算加权重构误差,若重构误差大,则说明该图块内存在异常事件。

2.2 AD-ConvAE网络结构

在特征学习的过程中使用卷积自编码器作为特征学习模型,设计的异常检测卷积自编码器AD-ConvAE网络结构如图2所示。卷积自编码器和传统自编码器两者的区别在于卷积自编码器采用卷积方法对输入内容进行变换。传统自编码器一般使用全连接层作为网络的基本结构,但对于二维图像来说会丢失一定的空间信息,而卷积层可以更好地保留空间信息。不同于文献[20]中将HOG特征和HOF特征输入到传统全连接自编码器的方法,本文将光流和HOG特征分别输入到卷积自编码器中进行重构,模型在训练过程中学习正常运动模式,在测试时计算重构误差进行异常事件检测。AD-ConvAE的网络结构由编码和解码部分组成,在编码部分,对于1个 $64 \times 64 \times 3$ 大小的输入,经过3个卷积层和3个池化层,最终得到特征的深度表示。在解码部分,对特征的深度表示进行重构,采用卷积操作和上采样操作,输出一个和输入图像同样大小的图像。整个网络中使用到的池化核大小均为 2×2 ,卷积核大小均为 3×3 ,图中 c 表示通道数,在卷积层中使用 n 个卷积核就会得到 n 个通道的特征图。AD-

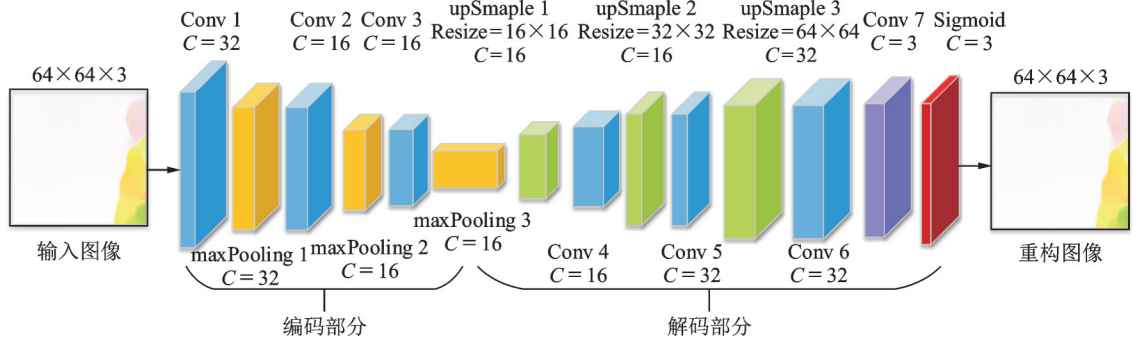


图2 AD-ConvAE的网络结构图

Fig.2 Overview of AD-ConvAE structure

ConvAE中使用交叉熵损失函数计算 Loss 值,其计算公式为

$$\text{Loss}(x) = \max(x, 0) - x \times y + \log(1 + e^{-|x|}) \quad (1)$$

式中: x 表示重构图像中某一点的像素值, y 表示输入图像中该点像素值。经式(1)计算,可以得到单个像素点的 Loss 值。对每个像素点都计算 Loss 值后取平均,可得到图像的平均重构误差。在训练时重构出来的图像和输入图像越像越好,即平均重构误差越小越好。

2.3 异常事件检测与定位方法

在测试阶段,将第 t 帧的第 (m, n) 位置上的视频图块提取出来的光流和 HOG 特征分别放入已训练完成的 AD-ConvAE 中计算重构误差。重构误差大表示测试样本存在异常,重构误差小表示测试样本正常。根据得到的 HOG 特征重构误差 $\text{Loss_hog}_{(m, n, t)}$ 和光流重构误差 $\text{Loss_optical}_{(m, n, t)}$, 计算该位置上的总重构误差 $\text{sumLoss}_{(m, n, t)}$ 。第 t 帧的第 (m, n) 位置上的总重构误差计算公式为

$$\text{sumLoss}_{(m, n, t)} = \alpha \times \text{Loss_hog}_{(m, n, t)} + \beta \times \text{Loss_optical}_{(m, n, t)} \quad (2)$$

式中: α 为 HOG 特征重构误差的权重, β 为光流重构误差的权重,且 $\alpha + \beta = 1$ 。

通过设置重构误差阈值 θ 判断该图像块是否存在异常事件。 $F(m, n, t) = 0$ 表示不存在异常事件, $F(m, n, t) = 1$ 表示存在异常事件,判定规则为

$$F(m, n, t) = \begin{cases} 0 & \text{sumLoss}_{(m, n, t)} \leq \theta \\ 1 & \text{sumLoss}_{(m, n, t)} > \theta \end{cases} \quad (3)$$

由于采用视频帧分块的方式,让异常事件检测模型针对视频帧中的某一区域判断是否存在异常,因此在异常事件检测的同时完成了异常事件定位。视频异常事件检测模型的训练过程如算法 1 所示。

算法 1 视频异常事件检测模型的训练算法

输入: 训练视频帧集合 TrainFrames, 训练视频帧的个数 frameCount, 分块个数 patchNumber, 迭代次数 epochs

输出: 异常事件检测模型 AD-ConvAE(1), AD-ConvAE(2)

(1) for each train $_i$ in TrainFrames do:

(2) 将 train $_i$ 均匀划分成互不重叠的图块 patches

(3) for $i=0$ to frameCount do:

(4) for $j=0$ to patchNumber do:

(5) { Opt $_{(i, j)}$ = patch $_{(i, j)}$ 的光流

(6) Hog $_{(i, j)}$ = patch $_{(i, j)}$ 的 HOG 特征 }

(7) for $k=0$ to patchNumber do:

(8) { epoch = 0

(9) while(epoch < epochs) do:

(10) { reconOpt $_{(k)}$ = forward(AD-ConvAE(1) $_{(k)}$, Opt $_{(k)}$)

// 重构第 k 个图块的光流

(11) bp(AD-ConvAE(1) $_{(k)}$, reconOpt $_{(k)}$) // 反向传播更新 AD-ConvAE(1) $_{(k)}$ 的参数

(12) reconHog $_{(k)}$ = forward(AD-ConvAE(2) $_{(k)}$, Hog $_{(k)}$)

// 重构第 k 个图块的 HOG

(13) bp(AD-ConvAE(2) $_{(k)}$, reconHog $_{(k)}$) // 反向传播更新 AD-ConvAE(2) $_{(k)}$ 的参数

(14) epoch = epoch + 1 }

(15) return AD-ConvAE(1), AD-ConvAE(2)

3 实验与结果

3.1 实验设置

本文在3个公开数据集(UCSD Ped1、UCSD Ped2和CUHK Avenue)上验证算法的有效性。UCSD数据集中包含的异常情况主要为:非行人实体闯入人行道,例如卡车、轮椅、自行车等以及异常的行人运动模式,例如奔跑、玩滑板车、推车等。CUHK Avenue数据集中包含的异常情况有:行人奔跑、行人抛掷物品、行人行走方向错误、出现非行人物体,例如自行车等。本文实验部分使用的数据集中,训练集部分仅包含正常模式,而测试集部分包含正常模式和异常模式。

由于UCSD Ped1和UCSD Ped2两个数据集中的视频帧大小不统一,并且为了方便使用FlowNet2.0进行光流提取,在实验开始时首先对视频帧大小进行调整。在实验中,将UCSD Ped1和UCSD Ped2的视频帧大小统一调整为256像素 \times 192像素,将CUHK Avenue的视频帧大小调整为640像素 \times 384像素。在对视频帧进行均匀分块的步骤中,将UCSD Ped1和UCSD Ped2中的视频帧划分成互不重叠的48个图块,每个图块的大小为32像素 \times 32像素。在CUHK Avenue数据集上,将视频帧划分成60个互不重叠的图块,每个图块的大小为64像素 \times 64像素。

实验中使用的GPU型号为NVIDIA GeForce GTX 1060,在Tensorflow框架下实现。在使用卷积自编码器AD-ConvAE进行训练时,根据观察Loss值的下降速度,将迭代次数设置为50次。实验结果使用异常检测任务常用性能评价标准:帧级别AUC(Area under the curve)和等错误率EER(Equal error rate)。

3.2 实验结果

首先对HOG特征的重构误差权重 α 和光流的重构误差权重 β 这两个超参数进行调整,图3给出了在CUHK Avenue数据集上绘制的ROC曲线。其中横轴表示假正例率(False positive rate, FPR),纵轴表示真正例率(True positive rate, TPR)。计算ROC曲线下面积可得到AUC,AUC在不同 α 和 β 下的值如表1所示。实验结果表明,当 $\alpha=0.6, \beta=0.4$ 时,AUC达到最大值,即表观特征的重要性要略高于运动特征,但是仅考虑移动物体的表观纹理信息或仅考虑移动物体的运动信息时,异常事件检测的效果欠佳。因此结合移动物体的表观特征和运动特征进行异常事件检测合理有效。

本文所提方法在UCSD Ped1、UCSD Ped2和CUHK Avenue三个公开数据集上,与基于传统手工特征的异常检测方法^[5,8-9,16]和其他基于深度学习的异常检测方法^[20-21,24]进行对比。由实验结果可知,基

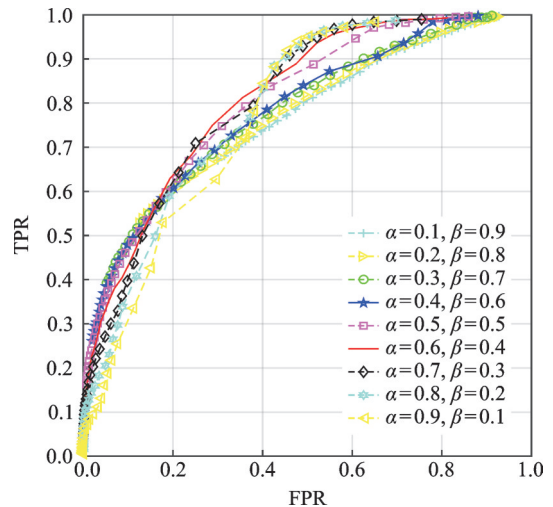


图3 α 和 β 取不同值时的ROC曲线图

Fig.3 ROC curves when α and β take different values

表1 α 和 β 取不同值时AUC值的比较

Table 1 Comparison of AUC with different values of α and β

α	β	AUC/%
0.1	0.9	75.92
0.2	0.8	77.28
0.3	0.7	78.13
0.4	0.6	78.95
0.5	0.5	78.98
0.6	0.4	81.65
0.7	0.3	81.39
0.8	0.2	75.87
0.9	0.1	76.19

于深度学习的异常检测方法整体优于传统的异常检测方法。其中文献[20]中的异常检测方法没有进行分块学习与检测,由此可见针对视频的不同位置进行特征学习,可以有效地提升异常检测性能。本文所提方法在帧级别的 AUC 和 EER 与其他方法的对比如表 2 所示。

表 2 本文所提方法与其他异常检测方法的比较

Table 2 Comparison of our method with other anomaly detection methods

方 法	UCSD Ped1		UCSD Ped2		CUHK Avenue	
	AUC	EER	AUC	EER	AUC	EER
MPPCA ^[16]	59.00	40.00	69.30	30.0		
SF ^[9]	67.50	31.00	55.60	42.0		
SF+MPPCA ^[5]	66.80	36.00	61.30	36.0		
Mohammad et al. ^[8]			67.50	23.8		
ConvAE ^[20]	81.00	27.90	90.00	21.7	70.20	25.1
SpatioTemporal-AE ^[21]	81.00	27.90	87.40	12.0	80.30	20.7
ISTL ^[24]	75.20	29.80	91.10	8.9	76.80	29.2
Ours	85.92	23.35	90.45	15.6	81.65	27.7

3.3 视频异常事件检测与定位结果可视化

实验最后将本文提出的异常事件检测与定位方法在 3 个实验数据集上进行检测结果可视化呈现,在 UCSD ped1 和 UCSD Ped2 数据集上的可视化异常事件检测结果如图 4 所示,图 5 给出了在 CUHK Avenue 数据集上的可视化异常事件检测结果。其中,红色的色块标识出视频帧中包含的异常。实验结果表明,本文提出的异常事件检测与定位方法可以有效地检测出视频中存在的异常事件,并给出准确定位。

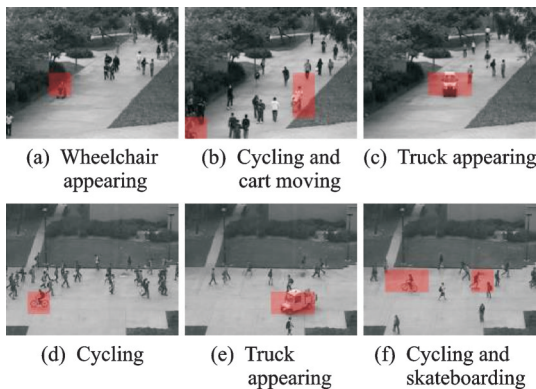


图 4 UCSD Ped1 和 UCSD Ped2 数据集上的可视化异常事件检测结果示例

Fig.4 Visualization of abnormal event detection results on the UCSD Ped1 and UCSD Ped2 datasets

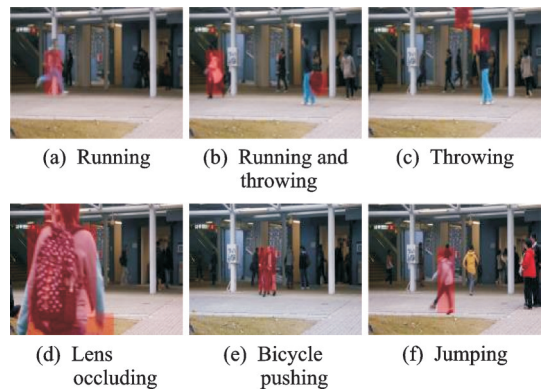


图 5 CUHK Avenue 数据集上的可视化异常事件检测结果示例

Fig.5 Visualization of visual abnormal event detection results on the CUHK Avenue dataset

4 结束语

本文提出了一种基于卷积自编码器分块学习的视频异常事件检测与定位方法,将视频帧均匀划分

成互不重叠的图块,并提取各个图块的光流和HOG特征。利用分块学习的方式,为视频不同位置上的图块分别设计卷积自编码器进行正常模式的特征学习。在异常事件检测时,根据图块中光流和HOG特征的重构误差大小进行异常判断,并实现异常事件定位功能。实验结果表明本文提出的异常事件检测与定位方法优于其他方法,并且可以准确地检测和定位异常事件。

参考文献:

- [1] ESEN E, ARABACI M A, SOYSAL M. Fight detection in surveillance videos[C]//Proceedings of the 11th International Workshop on Content-Based Multimedia Indexing. Veszprém, Hungary: Institute of Electrical and Electronics Engineers, 2013: 131-135.
- [2] CARNEIRO S A, SILVA G P, GUIMARÃES S J F, et al. Fight detection in video sequences based on multi-stream convolutional neural networks[C]//Proceedings of the 32nd SIBGRAPI Conference on Graphics, Patterns and Images. Rio de Janeiro, Brazil: Institute of Electrical and Electronics Engineers, 2019: 8-15.
- [3] PAN Huadong, YIN Jun, KU Haohua, et al. Fighting detection based on pedestrian pose estimation[C]//Proceedings of the 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics. Beijing, China: Institute of Electrical and Electronics Engineers, 2018: 1-5.
- [4] AKTI S, TATAROGLU G A, EKENEL H K. Vision-based fight detection from surveillance cameras[C]//Proceedings of the Ninth International Conference on Image Processing Theory, Tools and Applications. Istanbul, Turkey: Institute of Electrical and Electronics Engineers, 2019: 1-6.
- [5] MAHADEVAN V, LI Weixin, BHALODIA V, et al. Anomaly detection in crowded scenes[C]//Proceedings of the Conference on Computer Vision and Pattern Recognition. San Francisco, CA, USA: Institute of Electrical and Electronics Engineers, 2010: 1975-1981.
- [6] 张俊阳, 谢维信, 植柯霖. 基于运动前景效应图特征的人群异常行为检测[J]. 信号处理, 2018, 34 (3): 296-304.
ZHANG Junyang, XIE Weixin, ZHI Kelin. Abnormal crowd behavior detection based on motion effect map features of moving foregrounds[J]. Journal of Signal Processing, 2018, 34 (3): 296-304.
- [7] CONG Yang, YUAN Junsong, LIU Ji. Sparse reconstruction cost for abnormal event detection[C]//Proceedings of the Conference on Computer Vision and Pattern Recognition. Providence, RI, USA: Institute of Electrical and Electronics Engineers, 2011: 3449-3456.
- [8] COLQUE R V H M, CAETANO C, ANDRADE M T L, et al. Histograms of optical flow orientation and magnitude and entropy to detect anomalous events in videos[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2017, 27 (3): 673-682.
- [9] MEHRAN R, OYAMA A, SHAH M. Abnormal crowd behavior detection using social force model[C]//Proceedings of the Conference on Computer Vision and Pattern Recognition. Miami, Florida, USA: Institute of Electrical and Electronics Engineers, 2009: 935-942.
- [10] 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.
HUANG Xin, XIAO Shide, SONG Bo. Detection of vehicle's abnormal behaviors in surveillance video[J]. Computer Systems & Application, 2018, 27 (2): 125-131.
- [11] XU Dan, YAN Yan, 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.
- [12] SABOKKROU 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.
- [13] NARASIMHAN M G, SOWMYA K S. Dynamic video anomaly detection and localization using sparse denoising autoencoders [J]. Multimedia Tools and Applications, 2018, 77 (11): 13173-13195.
- [14] SIMONYAN K, ZISSERMAN A. Two-stream convolutional networks for action recognition in videos[C]//Proceedings of the Annual Conference on Neural Information Processing Systems. Montreal, Quebec, Canada: Institute of Electrical and Electronics Engineers, 2014: 568-576.

- [15] MORAIS R, LE V, TRAN T, et al. Learning regularity in skeleton trajectories for anomaly detection in videos[C]// Proceedings of the Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA: Institute of Electrical and Electronics Engineers, 2019: 11996-12004.
- [16] KIM J, GRAUMAN K. Observe locally, infer globally: A space-time MRF for detecting abnormal activities with incremental updates[C]// Proceedings of the Conference on Computer Vision and Pattern Recognition. Miami, Florida, USA: Institute of Electrical and Electronics Engineers, 2009: 2921-2928.
- [17] SABOKROU M, FATHY M, HOSSEINI M, et al. Real-time anomaly detection and localization in crowded scenes[C]// Proceedings of the Conference on Computer Vision and Pattern Recognition. Boston, MA, USA: Institute of Electrical and Electronics Engineers, 2015: 56-62.
- [18] FENG Yachuang, YUAN Yuan, LU Xiaoqiang. Learning deep event models for crowd anomaly detection[J]. Neurocomputing, 2017, 219: 548-556.
- [19] LI Nanjun, CHANG Faliang. Video anomaly detection and localization via multivariate gaussian fully convolution adversarial autoencoder[J]. Neurocomputing, 2019, 369: 92-105.
- [20] HASAN M, CHOI J, NEUMANN J, et al. Learning temporal regularity in video sequences[C]// Proceedings of the Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: Institute of Electrical and Electronics Engineers, 2016: 733-742.
- [21] CHONG Y S, TAY Y H. Abnormal event detection in videos using spatiotemporal autoencoder[C]// Proceedings of the International Symposium on Neural Networks. Hokkaido, Japan: Springer, Cham, 2017: 189-196.
- [22] 袁静, 章毓晋. 融合梯度差信息的稀疏去噪自编码网络在异常行为检测中的应用[J]. 自动化学报, 2017, 43 (4): 604-610. YUAN Jing, ZHANG Yujin. Application of sparse denoising auto encoder network with gradient difference information for abnormal action detection[J]. Acta Automatica Sinica, 2017, 43 (4): 604-610.
- [23] SULTANI W, CHEN Chen, SHAH M. Real-world anomaly detection in surveillance videos[C]// Proceedings of the Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: Institute of Electrical and Electronics Engineers, 2018: 6479-6488.
- [24] NAWARATNE R, ALAHAKOON D, SILVA D S, et al. Spatiotemporal anomaly detection using deep learning for real-time video surveillance[J]. IEEE Transactions on Industrial Informatics, 2020, 16 (1): 393-402.
- [25] ILG E, MAYER N, SAIKIA T, et al. FlowNet 2.0: Evolution of optical flow estimation with deep networks[C]// Proceedings of the Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: Institute of Electrical and Electronics Engineers, 2017: 1647-1655.

作者简介:



李欣璐(1996-),女,硕士研究生,研究方向:大数据分析
与挖掘技术,E-mail:
182202013@njnu.edu.cn。



吉根林(1964-),通信作者,
男,博士,教授,博士生导师,
研究方向:大数据分析
与挖掘技术,E-mail:
glji@njnu.edu.cn。



赵斌(1978-),男,博士,副教授,
研究方向:大数据分析
与挖掘技术,E-mail:
zhaobin@njnu.edu.cn。

(编辑:刘彦东)