

文章编号: 2095-2163(2022)05-0095-04

中图分类号: TP391

文献标志码: A

基于增量时空学习的实时异常行为识别算法研究

胡六四, 尹汪宏, 陈俊生, 王 林, 夏红霞

(安徽电子信息职业技术学院 软件工程学院, 安徽 蚌埠 233030)

摘要: 为了实现诸如监控视频等时空数据流的异常检测定位, 设计了实时异常行为识别算法(RAID)。该算法使用无监督深度学习深度学习方法, 进行带有模糊聚合的主动学习, 不断更新和学习随时间演变的异常行为。使用通用的真实数据集对 RAID 进行性能评估, 实验结果表明, 与现有方法相比 RAID 在准确性和鲁棒性方面均有较好的表现。

关键词: 异常行为识别; 时空数据; 深度学习

Research on real-time abnormal behaviors recognition algorithm based on incremental spatio-temporal learning

HU Liusi, YIN Wanghong, CHEN Junsheng, WANG Lin, XIA Hongxia

(School of Software Engineering, Anhui Vocational College of Electronics & Information Technology, Bengbu Anhui 233030, China)

[Abstract] In order to realize anomaly detection and positioning of spatio-temporal data streams such as the surveillance video, a real-time abnormal behaviors recognition algorithm (RAID) is designed. The algorithm uses an unsupervised deep learning method to perform active learning with fuzzy aggregation, and constantly updates and learns abnormal behaviors that evolve over time. Common real data sets are used to evaluate RAID performance. The experimental results show that compared with the existing methods, RAID has better performance in terms of accuracy and robustness.

[Key words] abnormal behaviors recognition; spatio-temporal data; deep learning

0 引言

智慧城市的营建推动了智能视频监控系统的展, 智能视频监控系统是现代工业和城市环境的开发、运营以及实现可持续发展的关键因素。视频监控不仅能够有效保障人们日常生活生产的安全性, 还有利于提高城市管理和运营的效率^[1]。随着智能摄像头的数量和部署范围的增加, 海量的视频数据流为处理分析带来了巨大的压力, 而使用人工对视频数据流进行准确分析也是不切实际的。

人工智能(AI)技术的发展为自动化监控系统提供支撑, AI技术适用于视频摘要、对象检测和重识别、行为检测和异常识别等领域。其中, 异常行为识别是当下重要的研究方向之一^[2-3], 也是自动化视频分析中的一项基本任务。异常检测的目的是识别不符合预期、即异常的行为, 异常行为和事件通常是未知的, 这也使得异常识别成为一个复杂研究课题。

高维的时空视频数据以及跨视频帧数据处理增

加了视频处理和计算的复杂性和成本^[4-6]。而且, 异常行为的定义并不明确, 正常行为和异常行为之间的界限往往是模糊的, 并且具有上下文关系。另外一般情况下, 正常行为很可能会随着时间的推移而演变, 从而导致识别检测过程出现误差。现有的异常行为技术主要围绕着优化计算复杂性和识别上下文异常来展开, 迄今为止仍然缺乏对异常行为动态性、即异常行为随时间演变的研究与探讨。

本文基于增量时空学习, 提出了 RAID 算法以实现实时异常行为识别。RAID 从时空视频数据流中主动学习正常行为的时空模式。主动学习可以随着环境的变化和新信息的出现不断完善学习过程, 并结合观察者来改善迭代的学习结果。

1 RAID 算法概述

1.1 算法概述

RAID 的架构如图 1 所示。首先, 将从时间 t_0 到 t_u 的实时监控视频作为时空模型训练的输入。然

基金项目: 2019 年安徽省高等学校自然科学研究项目(KJ2019A1067); 2019 年度安徽省技术技能型大师工作室项目(2019dsgzs34); 2020 年高校优秀青年人才支持计划(gxyq2020142)。

作者简介: 胡六四(1981-), 男, 硕士, 实验师, 主要研究方向: 人工智能与软件技术研究。

通讯作者: 胡六四 Email: huliusi@163.com

收稿日期: 2021-12-19

后,将训练后的模型用于进行异常行为识别。接下来,人们对检测到的异常进行验证,将验证输入用于构建更新后的正常行为。这种更新后的正常行为会被反馈到 RAID 模型中以进行持续的学习。

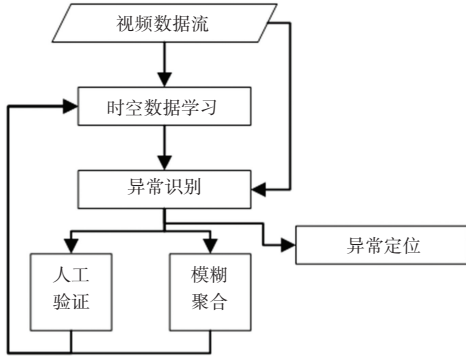


图1 系统架构

Fig. 1 System architecture

用于训练的视频流 $X_{train} \subset R$ 由一系列高度为 h 、宽度为 w 的帧组成, R 是包含正常行为的视频帧集合。用于测试的视频流 X_{test} 包含正常和异常行为的视频帧。模型的目标是从 X_{train} 中学习正常行为的表示 Ω , 随后使用 X_{test} 对其进行验证以识别异常行为。现有的工作需要完整的正常行为训练数据集,而 RAID 是基于实时接收的连续视频的时空信息主动更新先前学到的知识 Ω , 同时 RAID 考察了使用者对异常检测结果的反馈。

1.2 时空学习

RAID 模型由时空自动编码器组成,从 X_{train} 中学习正常行为的时空表示。自编码器是一种无监督学习算法,采用反向传播通过最小化重构误差来将目标值设置为等于输入。在 RAID 模型中,时空自动编码器由一系列用于学习空间表示的 CNN 层和一系列用于学习时间表示的 ConvLSTM 层组成。

研究中,要对原始视频数据进行预处理,以增强时空自动编码器模型的学习能力。首先,将视频数据提取为连续帧,转换为灰度以减少维度,调整为 224×224 像素并通过缩放来标准化像素值。时空自动编码器模型的输入是视频帧的时间长方体,使用长度为 T 的滑动窗口进行提取,不进行任何特征变换。将长度为 T 的连续帧堆叠在一起以构建时间长方体。

在视频分析中,卷积层可以通过使用过滤器学习特征表示来保留输入帧内的空间关系,过滤器的值是在训练过程中学习的。RAID 模型由 2 个卷积层和 2 个反卷积层组成。

LSTM 在处理空间数据方面存在缺陷,这是因为空间信息没有在其状态转换中进行编码。在保留

空间结构的同时,从监控视频流中学习时间规律是很重要的。因此,利用卷积 LSTM (ConvLSTM),即通过将其输入、隐藏状态、门和单元输出设计为 3D 张量来克服这个缺点。此外,在输入和门中的矩阵运算被卷积运算取代。ConvLSTM 能够从输入帧序列中捕获时空特征,ConvLSTM 模型可以用式(1)~(5)进行表示:

$$i_t = \sigma(W_{xi} * X_t + W_{hi} * H_{t-1} + W_{ci} \circ C_{t-1} + b_i) \quad (1)$$

$$f_t = \sigma(W_{xf} * X_t + W_{hf} * H_{t-1} + W_{cf} \circ C_{t-1} + b_f) \quad (2)$$

$$c_t = f_t \circ C_{t-1} + i_t \circ \tanh(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c) \quad (3)$$

$$o_t = \sigma(W_{xo} * X_t + W_{ho} * H_{t-1} + W_{co} \circ C_{t-1} + b_o) \quad (4)$$

$$H_t = o_t \circ \tanh(C_t) \quad (5)$$

其中,符号“ $*$ ”是指卷积操作;符号“ \circ ”是指哈达玛积; X_t 表示输入; C_t 表示单元状态; H_t 表示隐藏状态;门 i_t 、 f_t 和 o_t 都是三维的张量; $\sigma(\cdot)$ 是激活函数; W_x 和 W_h 是二维卷积核。RAID 模型由 3 个 ConvLSTM 层组成。时空自动编码器的结构如图 2 所示。

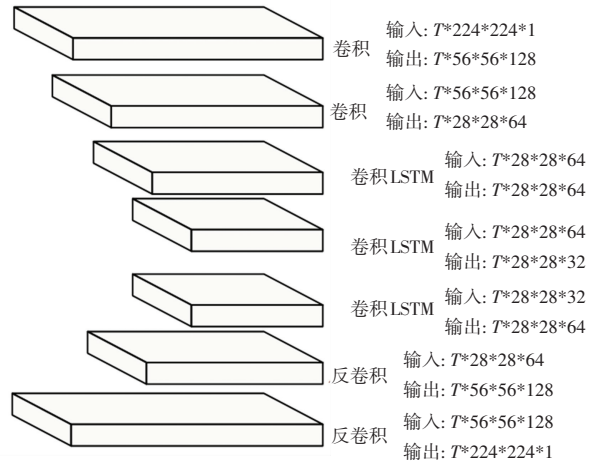


图2 时空自动编码器的结构

Fig. 2 The structure of spatio-temporal automatic encoder

1.3 异常检测和定位

RAID 模型可用于以像素级精度重建输入视频的正态性。然而,经过训练的自动编码器不具备准确重建异常行为的能力。这是因为训练数据中并不包含异常行为。重建误差 E 的计算方式如式(6)所示:

$$E = \sqrt{\sum_{i=0}^T \sum_{j=0}^w \sum_{k=0}^h |X_{i,j,k} - \bar{X}_{i,j,k}|^2} \quad (6)$$

其中, X 是输入, \bar{X} 是重建的输入。

使用重建误差阈值 μ 来区分正常和异常行为。在实际应用中,人们可以根据监控应用所需的灵敏度选择 μ 的值。较低的 μ 值会导致模型具有更高敏感性。

此外,引入了时间阈值 λ , 该阈值用于减少由于遮挡、运动模糊和高强度照明条件导致的监控视频流突然变化而导致的误报。

模型能够在视频帧定位出异常行为发生的特定区域。在检测到一段视频中有异常行为后,通过计算非重叠时空局部长方体窗口的重建误差来定位异常的具体位置。

在实际视频监控环境中主动学习的目的是实现动态变化环境中的异常识别。为了提高模型的泛化能力,RAID 模型通过采用模糊聚合的主动学习方法,不断地训练和学习以适应具有新行为的模型。

RAID 模型首先使用预先确定的正常行为进行训练,并用于异常检测。如果视频帧被检测为异常,即输入长方体的重构误差高于异常阈值,则输入长方体被归类为异常。然后将该结果发送给使用者进行验证。如果检测到的视频帧是错误检测(误报),那么该视频帧会被手动地标记为正常,并将该结果用于连续学习阶段。RAID 模型的连续学习更新的输入是来自连续接收的监控视频流的时空信息以及人们对监测结果的反馈。

1.4 主动学习

RAID 模型通过视频帧的模糊聚合来保持学习迭代的稳定性。在检测阶段,模型根据重构误差用模糊度量 g_λ 标记所有被评估的视频帧,并基于 g_λ 将视频帧分为 n 个组。随后,在连续学习阶段,算法将从每组选择包含最高 g_λ 的 k 个视频帧来训练 RAID 模型。使用式(7)来选择连续训练的场景,具体数学公式可写为:

$$d = \sum_{i=0}^n \max_j(s_i) \quad (7)$$

2 实验评估

2.1 数据集与实验设置

实验部分使用 2 个常用的数据集对 RAID 模型进行评估。这 2 个数据集分别是 CUHK 街道数据集^[7]和 UCSD 行人数据集^[8]。实验采用的服务器配置有 Intel Xeon Gold 32 核 3.20 GHz 处理器、128 GB 内存和双 NVIDIA Quadro GPU。

CUHK 街道数据集是由 16 个包含正常人类行

为的训练视频样本和 21 个包含异常事件及行为的测试视频样本组面。其中,异常行为是指乱扔垃圾、在草地上行走和遗弃物品等。UCSD 人行道数据集有 34 个训练视频样本和 36 个测试视频样本。正常行为包含行人在路径上行走的场景,而测试视频样本包含异常行为和意外行为。

在异常识别的评估中,将 RAID 模型与 MPCCA、ConvLSTM-AE 和 Unmasking 进行对比。其中,MPCCA 模型利用时空马尔可夫随机场和视频光流进行异常检测。ConvLSTM-AE 集成了 CNN 和 ConvLSTM 自动编码器,用于编码正常行为中的空间和时间模式。Unmasking 是一种基于 unmasking 技术的异常检测方法,该方法采用从三维梯度捕获的运动特征和来自预训练 CNN 的外观特征来进行异常检测。

2.2 实验结果与分析

使用 ROC 曲线的曲线下面积 (AUC) 和等错误率 (EER) 作为指标来评估各个模型的结果,结果见表 1。分析表 1 可知,RAID 模型的性能要优于基于人工设计特征的 MPCCA。与基于深度学习的方法,即 ConvLSTM-AE 和 Unmasking 相比,RAID 模型在 UCSD 行人数据集上具有最好的性能。

表 1 各模型的性能对比

Tab. 1 Performance comparison of models

模型	CUHK 街道数据集		UCSD 人行道数据集	
	AUC	EER	AUC	EER
MPCCA	60.2	35.1	61.3	36.0
ConvLSTM-AE	77.1	27.2	88.1	21.7
Unmasking	80.2	24.8	82.2	16.5
RAID	79.8	25.1	92.2	8.7

接下来,对模型的异常定位性能进行定性分析,部分结果如图 3 所示。在 UCSD 人行道数据集中,RAID 能定位视频中的诸如滑板、骑自行车和小车等异常行为。在 CUHK 街道数据集中,RAID 模型可以定位出诸如丢包、走错方向、跑步等异常。

为了评估 RAID 模型的主动学习能力,将 UCSD 人行道数据集(CUHK 街道数据集)中骑自行车(踩草地)行为的测试样本的所有异常检测标记为正常,以此评估 RAID 模型的连续学习能力。结果见表 2。在人为干预后,RAID 模型对异常行为的识别率大大下降了,这也表明了 RAID 模型能够主动地进行学习。

(下转第 101 页)