

文章编号: 2095-2163(2023)06-0189-04

中图分类号: TP391.4

文献标志码: A

基于深度学习的社区安防异常行为检测研究

余胤翔¹, 黄胤政², 李 宁¹, 徐 威¹, 储茂祥¹

(1 辽宁科技大学 电子与信息工程学院, 辽宁 鞍山 114051; 2 西南科技大学 经济管理学院, 四川 绵阳 621010)

摘要: 在人工智能快速发展的浪潮下, 社区监控安防体系得到逐步加强和完善, 安防监控行业迎来了智能化升级阶段。在安防监控图像处理技术领域, 传统的机器视觉算法很难做到高精度和快速识别监控中的异常行为。基于此, 本文提出了一种基于 Faster-RCNN+SlowFast 网络的社区安防异常行为识别方法。该方法通过提取视频关键帧中时间和空间信息进行目标定位与行为分析, 提高了识别精度和速率。通过实时分析视频监控中的人体动作, 识别出社区中发生的打架、盗窃、破坏公物等行为, 并将信息上报给社区安保系统。通过采取及时的预警措施, 有效地维护了社区治安稳定。

关键词: SlowFast; Faster-RCNN; 目标检测; 时空分析; 深度学习

Research on community security abnormal behavior detection based on deep learning

YU Yinxiang¹, HUANG Yinzheng², LI Ning¹, XU Wei¹, CHU Maoxiang¹

(1 School of Electronics and Information Engineering, Liaoning University of Science and Technology, Anshan Liaoning 114051, China;

2 School of Economics and Management, Southwest University of Science and Technology, Mianyang Sichuan 621010, China)

[Abstract] With the rapid development of artificial intelligence, the community monitoring and security system has been gradually strengthened and improved, and the security monitoring industry has stepped into an intelligent upgrade phase. In the field of security monitoring image processing technology, traditional machine vision algorithms are difficult to achieve high-precision and rapid recognition on abnormal behavior in monitoring. To solve this problem, this paper proposes a community security abnormal behavior recognition method based on Faster RCNN + SlowFast network, which improves recognition accuracy and speed by extracting temporal and spatial information from video key frames for target location and behavior analysis. By real-time analysis of human movements in video surveillance, identification of fights, theft, vandalism, and other behaviors occurring in the community, this system can report relevant information to the community security system, which effectively helps to maintain community security and stability.

[Key words] SlowFast; Faster-RCNN; target detection; spatiotemporal analysis; deep learning

0 引言

智慧社区是新形势下社会管理创新的一种新模式, 随着物联网、云计算、人工智能等新一代信息技术的快速发展, 居民的生活环境变得越来越安全和便利^[1]。社会对公共安全重视程度的逐渐加深, 使基于视频监控的智能感知技术成为安全监控发展的重要方向^[2]。由于传统视频监控系统仅具备拍摄和存储功能, 一旦发现盗窃、抢劫、打架等暴力事件,

公安部门在后期收集证据和案件审查过程中会消耗大量的人力、物力和时间。随着国内监控范围基本实现全覆盖, 其海量的视频数据所带来的工作量及人工甄别的方式已不能适应居民对社区安全的需求。利用人工智能的高计算能力和高效能特点, 通过对视频监控数据进行实时处理分析, 及时发现异常行为和意外事件, 做到提前预警, 从而减少意外导致的经济损失和居民人身危害。

针对异常行为基于时空信息进行定义可将其分

基金项目: 大学生创新创业训练计划项目。

作者简介: 余胤翔(2002-), 男, 本科生, 主要研究方向: 嵌入式软件和机器视觉; 黄胤政(2002-), 男, 本科生, 主要研究方向: 市场细分和目标市场选择; 李 宁(2002-), 男, 本科生, 主要研究方向: 数学建模与 PLC 控制; 徐 威(2002-), 男, 本科生, 主要研究方向: 机器视觉; 储茂祥(1978-), 男, 博士, 教授, 博士生导师, 主要研究方向: 模式识别、机器学习。

通讯作者: 储茂祥 Email: chu52_2004@163.com

收稿日期: 2023-03-24

为3类:

- (1)在某一场景中不寻常位置发送的动作;
- (2)在某一场景中不寻常时刻发生的动作;
- (3)在某一场景中正常位置、正常时刻发生的不寻常动作^[3]。

由于异常行为的种类很多,在社区中行为自由度较高的地方,不合规的动作很难穷举。因此,异常行为检测存在以下技术难点:

(1)相比于正常行为,社区内一般异常行为发生的几率很低,因此很难获取到足够的数量,由于正负样本不均衡,计算机很难学习到足够多的异常行为特征^[4]。

(2)通过社区监控获取到的画面质量参差不齐,外界阳光和室内灯光的变化导致视频画面亮度值不同;监控摄像头拍摄的角度和远近不同。

(3)视频画面中人员运动速度较快,视频画面会产生一定的模糊,且人员通过周围环境或者衣物进行遮挡,也会影响到动作判定的精确度。

(4)视频流处理数据非常耗时,如果将视频中每个帧数的数据都提取出来处理,任务量非常庞大,使整个系统运行的实时性效果变差。因此需要对视频数据抽取部分关键帧,从而减少无效数据量,提升运行速率。

早期对人体异常行为的检测研究多利用手工特征描述符来表示行人的外观和运动特征^[5],其中包括轨迹^[6]、梯度方向直方图(HOG)^[7]和光流直方图(HOF)^[8]等等。但是,这类视觉特征提取方法在捕获行为特征时性能较低。因此,本文结合目前主流的目标检测算法,提出一种基于Faster-RCNN+SlowFast的异常行为识别方法,提升识别的精度,对社区公共场所内人员的异常行为进行监测预警。

1 异常行为检测原理和实现方法

1.1 SlowFast网络原理

SlowFast网络^[9]分为两个分支,慢速通道(Slow pathway)和快速通道(Fast pathway)。Slow pathway分支运行低帧率,在低时序分辨率下,提取空间中的有效信息。Fast pathway负责运行高帧率,在高时序分辨率下,精准提取有效的运动信息,同时减少通道数量,使其变得轻量化,更能够适应一些边缘AI的应用场景。尽管Fast pathway能在高速的环境下做出反应,刺激反应快,但是对空间信息不敏感,而Slow pathway则能够提供空间中的详细信息,但是刺激反应速度较慢,因此需要对两条通路进行横

向连接进行互补融合。通过改变两条通道的时间维度,使得网络在快慢通道的每个阶段进行匹配。最后,分别获取Fast pathway和Slow pathway两条通道的特征向量。将两个的特征向量串联起来作为全连通分类器层的输入^[10]。SlowFast网络结构如图1所示。

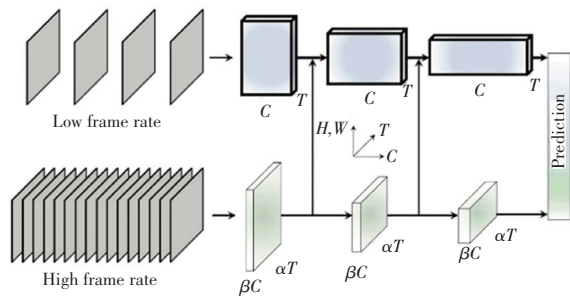


图1 SlowFast网络结构

Fig. 1 SlowFast network structure

1.2 SlowFast模型应用

首先需要建立时空动作识别模型,视频通过密集抽帧获取时空信息。由于相邻帧之间信息冗余较大,所以只获取关键帧,之后对其进行数据处理,使关键帧匹配其上下文的数据,最后加载检测该模型。

整个运行过程中,系统在关键帧中利用检测模型检测图像中的人,获得人的位置、置信度等信息,将提前加载好的模型输入,构建的时空动作识别模型,之后进行时空动作识别模型推理,得到动作识别结果。最后将整个过程进行可视化操作输出视频,用来观测动作识别结果。模型应用流程如图2所示。

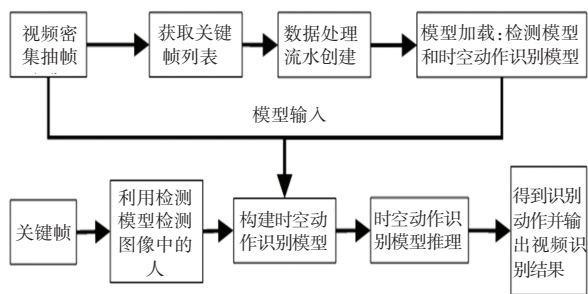


图2 模型应用流程

Fig. 2 Specific flow chart of model application

1.3 SlowFast+FasterRCNN实现

如图3所示,SlowFast+FasterRCNN的实现分为两个阶段检测。在准备完数据集后,先利用SlowFast模型提取时序特征,作为整个网络的主干网络Backbone;其次通过FasterRCNN中的Roi_head获得目标检测中框的特征,并将该操作扩展到3维,最后

将检测结果输入分类器, 判别人体动作并输出当前人的动作。总结来讲, 系统先定位到视频帧的每个人动作, 提取其位置等特征, 得到每个人的特征检测框, 再放到分类模型中, 对每个人的动作进行分类。

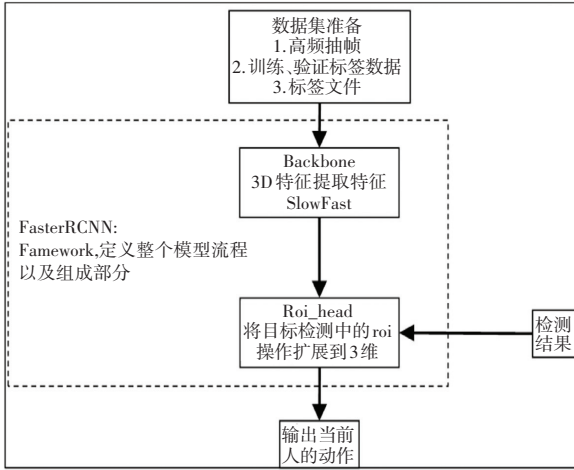


图 3 SlowFast_FasterRCNN 实现过程

Fig. 3 SlowFast_FasterRCNN implementation

2 数据集

2.1 数据集建立

本文实验针对社区环境进行人员识别, 视频数据主要来源于监控画面, 数据集主要来源于谷歌开源的 AVA 数据集。考虑到部分异常行为的发生频率较高, 数据集相对较少, 因此自建数据集进行补充。如: 打架的数据集中又加入了 fight-detection-surv-dataset、A-Dataset-for-Automatic-Violence-Detection-in-Videos 和 UBI_FIGHTS 等开源数据集。最后, 基于 AVA 公开数据集^[11] 格式, 建立异常行为检测数据库。

2.2 数据处理

建立数据集后, 需要对数据集进行帧标注, 在电脑上导入已经标注好的数据集进行本地模型训练。

首先从所有视频数据中提取视频帧, 以每帧 3s 的帧率 (fast 分支帧率) 进行视频抽帧采样, 之后进行帧标注。帧数标注使用的标签数据 (CSV) 格式为: 视频 ID、视频帧时间戳、识别的框图位置 (x1, y1, x2, y2)、动作 ID、人的 ID。

通过 PaddleDetection 提供的 FasterRCNN 模型得到检测结果后, 将检测结果转成 SlowFast_FasterRCNN 模型需要的输入格式, 制作标签数据 list 文件, ID 从 1 开始, 见表 1。

在 slow 分支帧的数据集的基础上重复上述操

作, 以较大帧率 (实验使用为 40 fps) 抽取视频帧, 制作 fast 分支数据集。

表 1 标签 ID
Tab. 1 Label ID

ID	Name
1	踢腿
2	打架
3	争吵
4	摔倒
5	破坏公共设施
6	追赶
7	挥棒

3 实验与结果分析

3.1 实验环境和参数设定

实验在 Windows10 Version 21H2 操作系统下进行, 采用飞桨 AI Studio 社区提供的 PaddleDetection 2.0 框架, 软件环境为 CUDA 11.2, Python3.7。训练时的初始学习率为 0.003。

3.2 模型优化

3.2.1 调节学习率

在训练模型的过程中, 需要对模型进行优化。如果一开始模型精度设置过大, 可能导致训练时跳过最优解, 随着训练次数越多, 模型精度就会越低。因此, 采用学习率预热 (Warmup) 和分段常数衰减 (Piecewise_constant) 的方法, 分别设置两个常数值: 开始的学习率 $warmup_start_lr$ 和基础学习率 $step_base_lr$ 。初始训练时, 选择较小的学习率训练, 再将学习率调节为设置好的学习率继续训练, 从而提高模型精度。实验最后确定的 $warmup_start_lr$ 参数为 0.002、 $step_base_lr$ 为 0.04。

3.2.2 α 超参数优化

α 参数是 slow 和 fast 两个分支帧的比例。如公式 (1):

$$fast = slow * \alpha \quad (1)$$

通过调小 backbone 和数据处理的 PackOutput 中 α 参数, 再结合学习率参数的变化趋势进行修改, 从而进一步提升模型的识别精度。

3.2.3 数据集优化

数据集的质量会对系统异常行为检测精度产生影响, 为此进行了对比实验。训练选择两组数据集, 第一组只用 AVA 数据集, 第二组使用自建数据集。观察模型部署后, 系统对异常行为的平均正确率 (AP)。

考虑到不同异常行为的数据量不同,训练出的识别精度也不同,因此需要挑选出识别精度相对较高的异常行为,选择以官方的 AVA 数据进行集中测试。实验结果见表 2。

表 2 AVA 数据测试结果
Tab. 2 AVA data test results

动作	训练样本	测试样本	AP
摔倒	303	70	0.335
吸烟	2 563	646	0.538
打架	2 342	420	0.566
踢腿	46	11	0.005 7

从表 2 中数据可知,由于踢腿动作的数据集较少,AP 值极低,不具备参考价值。因此,选用打架、吸烟两个识别精度较高的动作作为实验对照选取的动作进行实验,结果见表 3、表 4。

表 3 吸烟数据集对比
Tab. 3 Comparison of smoking data sets

吸烟	样本数量	AP
AVA	3 209	0.538
自建数据	5 584	0.234

表 4 打架数据集对比
Tab. 4 Comparison of fighting datasets

打架	样本数量	AP
AVA	2 762	0.566
自建数据	3 285	0.434

根据数据得出,在相同动作下,自建数据质量低于 AVA 数据集,即便在数据量大于 AVA 的情况下,自建数据集的 AP 值依旧低于 AVA 的 AP 值。由此可见,数据质量对模型精度存在一定的影响,因此可以通过进一步优化数据集样本,提高模型的识别精度。

3.3 实验效果

使用 Faster-RCNN+SlowFast 网络对异常行为识别数据集的测试集中视频进行测试,通过可视化操作输出视频,用来观测动作识别结果。这里导入了两段打架视频进行测试。

识别效果如图 4 所示,两组图片都进行了输入图像和输出图像的对照,系统对打架双方的打架动作进行目标检测后在视频图像上进行了框选,并在方框上自动显示检测动作的名称 fight。由实验结果可以看出,该检测方法可让用户直接根据视频的框选标记,判断社区中是否发生异常行为及其种类,能够有效实现异常行为的检测,具有较好的实用性。



图 4 实验效果

Fig. 4 Experimental effect

4 结束语

本文提出了一种基于 Faster-RCNN+SlowFast 网络的社区安防异常行为识别方法,系统通过训练和部署检测模型和时空动作识别模型,可以实时检测视频监控中出现的异常行为。本文对该网络的框架和原理进行了简单阐述,提出了数据集的优化和参数的调节方法,并通过实验进行验证,给出了相关实验数据进行分析,一定程度上提高了模型的识别精度。人体异常行为识别的研究,对于未来社区安防事业的发展具有重要意义。

参考文献

- [1] 陈唯石. 社区银行小额贷款业务发展路径分析[D]. 福建: 福建师范大学, 2020.
- [2] 钟婷, 彭略. 视频监控中异常行为检测在安防领域的研究进展[J]. 智能城市, 2022, 8(9): 11-14.
- [3] 周义. 行人异常行为检测方法的研究与系统实现[D]. 南京: 东南大学, 2016.
- [4] 彭嘉丽, 赵英亮, 王黎明. 基于深度学习的视频异常行为检测研究[J]. 激光与光电子学进展, 2021, 58(6): 51-61.
- [5] 徐涛, 田崇阳, 刘才华. 基于深度学习的人群异常行为检测综述[J]. 计算机科学, 2021, 48(9): 125-134.
- [6] MARCO B, ALBERTO D B, LORENZO S. Multi-scale and real-time non-parametric approach for anomaly detection and localization[J]. Computer Vision and Image Understanding, 2012, 116(3): 320-329.
- [7] Tomoki Watanabe, Satoshi Ito, Kentaro Yokoi. Co-occurrence histograms of oriented gradients for human detection[J]. IPSJ Transactions on Computer Vision and Applications, 2010, 2.
- [8] DALAL N, TRIGGS B, SCHMID C. Human detection using oriented histograms of flow and appearance [C]//Computer Vision - ECCV 2006, 9th European Conference on Computer Vision, 2006: 428-441.
- [9] SIMONYAN K, ZISSERMAN A. Two-stream convolutional networks for action recognition in videos [J]. Advances in neural information processing systems, 2014, 27(1): 568-576.
- [10] 汪威, 胡旭晓, 吴跃成, 等. 基于深度学习的自动扶梯视频人体动作识别[J]. 软件工程, 2021, 24(9): 24-27.
- [11] GU C, SUN C, ROSS D A, et al. Ava: A video dataset of spatio-temporally localized atomic visual actions [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018.