

基于 Yolo 和 GOTURN 的景区游客翻越行为识别

周巧瑜^{1,2}, 曹 扬^{2,3}, 詹瑾瑜^{1,2}, 江 维¹, 李 响^{2,3}, 杨 瑞^{2,3}

(1. 电子科技大学 信息与软件工程学院, 四川 成都 610054;

2. 中电科大数据研究院有限公司, 贵州 贵阳 550022;

3. 提升政府治理能力大数据应用技术国家工程实验室, 贵州 贵阳 550022)

摘 要:近年来,随着旅游市场的快速发展,在旅游景区出现的一些违规行为,不仅危害了人身安全,而且也给社会造成了许多负面影响。由于出现该类行为的频率不高,通过人工观察耗费大量人力资源且效率不高,使用深度学习算法对具体行为进行识别,帮助景区监管人员快速预警违规行为,已成为必然趋势。针对这一问题,结合目标检测与目标跟踪任务,该文提出了一种基于 Yolo 和 GOTURN 的景区游客翻越行为识别方法。首先将视频转为视频帧,再经过 Yolo 目标检测和 GOTURN 目标跟踪得到人员边界框坐标和视频帧轨迹点集合,再进入轨迹分析得出最终结果标签(是否为翻越行为),形成一个完整的翻越行为识别方法。实验数据表明,基于 Yolo 和 GOTURN 的景区游客翻越行为识别方法相对于其他方法具有较高的准确率,应用在实际的景区游客翻越行为识别系统中得到了 93.7% 的准确率。

关键词:深度学习;目标检测;目标跟踪;翻越行为识别;Yolo;GOTURN

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2022)01-0134-07

doi:10.3969/j.issn.1673-629X.2022.01.023

A Fence Climbing Behavior Recognition of Scenic Area Tourist Based on Yolo and GOTURN

ZHOU Qiao-yu^{1,2}, CAO Yang^{2,3}, ZHAN Jin-yu^{1,2}, JIANG Wei¹,

LI Xiang^{2,3}, YANG Rui^{2,3}

(1. School of Information and Software Engineering, University of Electronic Science and
Technology of China, Chengdu 610054, China;

2. CETC Big Data Research Institute Co., Ltd., Guiyang 550022, China;

3. Big Data Application on Improving Government Governance Capabilities National
Engineering Laboratory, Guiyang 550022, China)

Abstract: In recent years, with the rapid development of tourism market, there are some tourism violation behaviors, which not only endanger personal safety, but also cause many negative effects on the society. Due to the infrequent occurrence of such behaviors, manual observation costs a lot of human resources and is inefficient. It has become an inevitable trend to use deep learning algorithms to identify specific behaviors and help scenic area supervisors to quickly warn violation behaviors. For this, combining target detection and target tracking tasks, we introduce a fence climbing behavior recognition method for the scenic area tourists based on Yolo and GOTURN. Firstly, the video is converted to video frame, and then the boundary frame coordinates and the video frame track point set are obtained by Yolo target detection and GOTURN target tracking. Finally, through trajectory analysis, the final result label (whether it is a fence climbing behavior or not) is obtained to form a fence climbing behavior recognition method. The experiment shows that the proposed fence climbing behavior recognition method has higher accuracy compared with the other methods, and the accuracy of 93.7% is obtained in the actual scenic spot tourist jump behavior recognition system.

Key words: deep learning; object detection; object tracking; fence climbing behavior recognition; Yolo; GOTURN

收稿日期:2021-02-03

修回日期:2021-06-03

基金项目:提升政府治理能力大数据应用技术国家工程实验室开放基金项目(W-2019007);四川省科技项目(2018CC0136);中科院计算机体系结构国家重点实验室开放课题(CARCH201811);中央高校基本科研业务费(ZYGX2018J077, ZYGX2019J078)

作者简介:周巧瑜(1996-),女,硕士研究生,CCF会员(B3078G),研究方向为软件工程与视频行为识别;通信作者:詹瑾瑜(1978-),女,博士,副教授,CCF会员(B3099M),研究方向为深度学习、大数据处理。

0 引言

近年来随着经济的快速发展,人民生活水平逐渐提高,旅游市场逐渐兴起,景区监管制度逐渐完善,在旅游景区中经常会有很多危险行为和违规行为发生,也有一些关于游客违规行为事件的相关新闻报道。2020年11月19日,四川黄龙景区游客翻越栏杆踩踏万年五彩池,违反了景区规定,破坏了自然景观。2020年12月12日,国外一名女子在景区翻越栏杆摆姿势拍照,从80米高的观景台跌落身亡。这些报道显示了翻越行为对社会治安造成的不良影响,暴露出景区监管制度的不完善。

人工通过实时监控视频进行特定行为识别,不仅耗费了大量人力资源,而且效率也偏低,同时,随着应用场景的多样性变化,特定行为识别技术受到了限制。为使特定行为识别技术发挥最大的作用,人们开始把目标转向机器学习,应用机器学习方法对人体进行特定行为识别实现快速识别并且识别率较高的效果。然而,由于传统机器学习方法对于翻越行为有一定的局限性,且实时性不够高,因此,如何对特定的行为进行识别且能够达到较高的实时性和准确率,是一个亟需解决的问题。

在视频图像领域中运用深度学习方法进行行为识别的技术已经较为成熟。Karen等人^[1]提出一个时空双流网络结构(two-stream CNN),它运用基于卷积网络的时间和空间识别流得到运动信息。最后经过Softmax后,做分类分数的融合得到行为分类。Wang Limin等人^[2]提出了一个基于视频的动作识别框架的时间段网络(TSN),它结合了一个稀疏时间采样策略和视频级监督,使用整个动作视频的高效学习提高了动作识别的准确率。Song Sijie等人^[3]提出了一个端到端空间和节奏注意模型的人体动作识别骨架数据和一种正则化的交叉熵损失来驱动模型学习过程,并相应地制定联合训练策略。Du等人^[4]利用在大规模监督视频数据集上训练的深度三维卷积网络(3D ConvNets)进行时空特征学习,同时证明了使用线性分类器的C3D特征可以在不同的视频分析基准上优于或接近之前最好的方法。Diba等人^[5]提出了一种视频卷积网络名为时域3D ConvNet(T3D)及其新的时域层时域过渡层(TTL),引入了一种新的时间层来模拟可变时间卷积核深度。Yang等人^[6]提出了一种时间保持卷积(TPC)网络,在每帧动作预测和分段级时间动作定位上取得了显著改善。Mabrouk等人^[7]综述了行为表示的特征提取和描述相关技术,对视频监控系统的行为表示和行为建模方面进行了研究,给出了行为建模的分类方法和框架。

上述使用深度学习方法进行的行为识别主要是针

对多种常规行为进行识别,缺少对于特定且不常见行为的识别,下面是针对不同场景对于翻越行为进行检测并识别的方法。Yu等人^[8]设计了一种基于块的隐马尔可夫训练方法组成翻越行为识别系统,通过2SS算法计算人体星形骨架特征,并通过该特征训练一个HMM模型,将视频中人体的动作分为行走、攀爬、跨越、下降四种状态。当攀爬、跨越、下降三种状态连续出现的时候,就判定发生了翻越行为,该方法是一种新型的翻越行为识别方法。类似地,Yu等人^[9]利用时间序列表示的隐马尔可夫模型(HMM)技术进行识别,在包含步行和攀爬等混合动作的图像序列上的实验结果证明了所提方法的有效性。之后,Yu等人^[10]又提出了效果更好的VSS算法。

基于以上提到的翻越行为识别方法,张泰等人^[11]提出了一种视频监控中人员翻越行为检测算法。该算法通过训练前景判断分类器与头部检测分类器Adaboost来实现目标检测,将混合高斯模型法得到的运动前景区域与KLT算法得到的特征点的运动信息结合起来,得到了一个仅使用灰度图像作为输入,能够一定程度上适应目标形变及遮挡的,鲁棒性强、实时性强的跟踪算法,最后基于先验知识对跟踪轨迹进行分析,得到最终是否发生翻越行为的结果。

由于此方法针对性较强,准确率较高,因此文中采用该过程进行翻越行为的判定,但该方法使用的Adaboost头肩检测方法在实际应用中标注的人物边界框大小固定不变,无法适应人物大小,并且检测速率过慢,有时人物过小则无法检测到目标人物,既无法达到实时性,也不能更准确地得到人物框坐标,存在极大的缺陷。

该文提出了一种基于Yolo和GOTURN的景区游客翻越行为识别方法,通过绘制与人物大小相同的边界框,克服了传统目标检测方法中实时性不高以及边界框大小固定的缺点;采用Yolo目标检测网络^[12]进行图像特征类别预测,采用GOTURN网络^[13]进行目标跟踪;最后通过先验知识的方法快速运用栏杆与轨迹点集合的相对位置关系来判定是否为翻越行为,若是翻越行为则输出翻越标签并发起警告,最终达到93.7%的准确率。

1 系统模型架构

该文提出了一种基于Yolo和GOTURN的景区游客翻越行为识别方法,解决了旅游景区场景下的翻越行为识别问题,系统模块如图1所示。

由图1可知,系统主要分为输入层、视频分割层、模型处理层和输出层,在模型处理层中分为Yolo模块、GOTURN模块、轨迹分析模块。

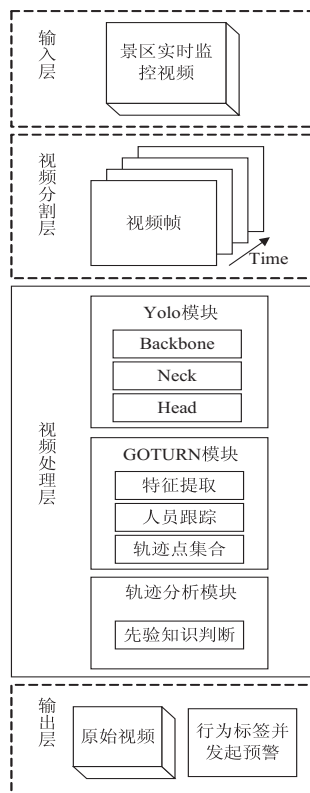


图 1 系统模块

2 基于 Yolo 和 GOTURN 的景区游客翻越行为识别方法

该文采用视频分割、目标检测、目标跟踪、轨迹分析的步骤对翻越行为进行分析与识别。其中目标检测部分通过 Yolo 方法进行人员目标检测,再通过 GOTURN 方法实现目标跟踪,最后根据轨迹集合与栏杆线的相对位置关系运用先验知识判定是否为翻越行为。

2.1 视频分割方法

该步骤主要是对视频数据进行处理:筛选和分割,筛选出有翻越栏杆的行为及在栏杆附近其他行为的视频,并将视频分割为视频帧的图片。

一般来说,视频分割方法主要分为基于时域的视频对象分割方法、基于运动的视频对象分割方法、交互式视频对象分割方法。该文主要应用交互式分割方法,该方法主要是通过图形界面对视频图像进行初始分割,然后对后续帧利用基于运动和空间的信息进行分割。

2.2 Yolo 目标检测方法

Yolo 方法是 One-stage 方法之一,它基于一个单独的 End-to-end 网络,将物体检测作为回归问题求解,完成从原始图像的输入到物体位置和类别的输出。该方法比一般 Two-stage 方法在速度上快很多,整个检测网络管道简单。测试证明,Yolo 对于背景图像的误检率低于 Two-stage 中 Fast R-CNN 方法误检率的一半。因此,该文采用此方法进行目标检测。

Yolo 检测网络包括 24 个卷积层和 2 个全连接层,Yolo 网络借鉴了 GoogLeNet 分类网络结构。不同的是,Yolo 使用 1×1 卷积层 + 3×3 卷积层的简单替代。Yolo 全连接层将输入图像分成 $S \times S$ 个格子,每个格子负责检测‘落入’该格子的物体, S 表示单元格数量,如 $S = 7$ 时, $S \times S$ 表示把图像划分成 7×7 个单元格。若某个物体的中心位置坐标落入到某个格子,那么这个格子就负责检测出这个物体。每个格子输出 B 个边界框信息,以及 C 个物体属于某种类别的概率信息。边界框信息包含 5 个数据值,分别是 x, y, w, h 和 confidence 。其中 x 和 y 是指当前格子预测得到的物体边界框的中心位置的坐标, w 和 h 是边界框的宽度和高度。注意:在实际训练过程中, w 和 h 的值使用图像的宽度和高度进行归一化到 $[0, 1]$ 区间; x 和 y 是边界框中心位置相对于当前格子位置的偏移值,并且被归一化到 $[0, 1]$ 。 confidence 反映当前边界框是否包含物体以及物体位置的准确性,计算方式如下:

$$\text{confidence} = P(\text{object}) \times \text{IOU} \quad (1)$$

其中,若边界框包含物体,则 $P(\text{object}) = 1$;否则 $P(\text{object}) = 0$ 。IOU (intersection over union) 为预测边界框。

Yolo 使用均方和误差作为 loss 函数来优化模型参数,即 Yolo 检测网络输出的 $S \times S \times (5B + C)$ 维向量与真实图像对应该向量的均方和误差:

$$\text{loss} = \sum_{i=0}^{S^2} (\text{coordError}_i + \text{iouError}_i + \text{classError}_i) \quad (2)$$

其中, coordError_i 、 iouError_i 、 classError_i 分别表示预测值与真实值之间的坐标误差、IOU 误差和分类误差。

该文通过 Yolo 目标检测网络对视频分割得到的视频帧进行检测。具体而言,Yolo 网络包括三个部分:Backbone 部分、Neck 部分、Head 部分,如图 2 所示。

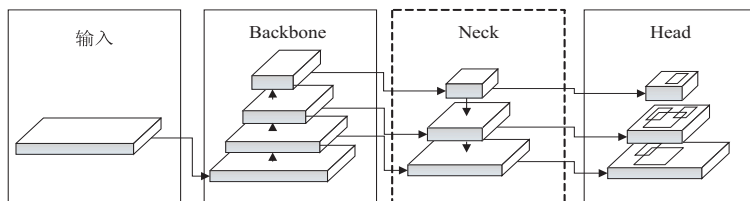


图 2 Yolo 模型结构

通过 Backbone 将输入视频帧通过 CSPResNext50 神经网络聚合并形成图像特征,以实现图像特征提取;通过 Neck 部分将图像特征运用 SPP-block 和 PANet 组合并传递图像特征到预测层;通过 Head 部分对图像特征进行预测,生成边界框并预测类别,若预测类别为‘person’则进入下一步的目标跟踪,否则继续检测下一个视频帧,直到找到该预测类别。

2.3 GOTURN 目标跟踪方法

GOTURN 方法是一种离线学习神经网络的方法,该方法训练一组带标签的训练视频和图像,但不需要任何类别级别的标签或关于被跟踪对象类型的信息。同时,该方法建立了一个新的跟踪框架,在这个框架中,外观和动作之间的关系以一种通用的方式离线学习。

GOTURN 网络的卷积层采用的是 5 层结构,该结构参照了 CaffeNet 里面的结构,其中激励函数都采用了 Relu,部分卷积层后面添加了池化层,而全连接层是由 3 层组成,每层 4 096 个节点,各层之间采用 Dropout 和 Relu 激励函数,以防过拟合和梯度消失。将上一帧的目标和当前帧的搜索区域同时经过 CNN 的卷积层,然后将卷积层的输出通过全连接层,用于回归当前帧目标的位置。

为具体化运动平滑的思想,该文将当前帧 (c'_x, c'_y) 中的边界框中心相对于前一帧 (c_x, c_y) 中的边界框中心建模为:

$$c'_x = c_x + \omega \cdot \Delta x \quad (3)$$

$$c'_y = c_y + h \cdot \Delta y \quad (4)$$

其中, Δx 和 Δy 都可以用均值为 0 的拉普拉斯分布建模。同样,跟踪器模型的大小也会发生变化。

$$\omega' = \omega \cdot \gamma_\omega \quad (5)$$

$$h' = h \cdot \gamma_h \quad (6)$$

其中, γ_ω 和 γ_h 由均值为 1 的拉普拉斯分布模拟。通过交叉验证,本实验最终选定的分布参数为:

$$\Delta \sim f(0, \frac{1}{5}), \gamma \sim f(1, \frac{1}{15}) \quad (7)$$

另外,轨迹点由每一帧边界框组成,其公式如下:

$$\text{tra}_{x,y} = (\frac{c'_x}{2}, \frac{c'_y}{2}) \quad (8)$$

然后,每个视频帧中的轨迹由 $\text{tra}_{x,y}$ 组成,公式如下:

$$\text{trajectory} = \{ \text{tra}_{x_0,y_0}, \text{tra}_{x_1,y_1}, \dots, \text{tra}_{x_i,y_i} \} \quad (9)$$

文中给出了轨迹坐标识别算法,如算法 1 所示。

算法 1: 轨迹坐标识别。

输入: 视频地址 v , 初始边界框坐标 (c_x, c_y)

$$\begin{cases} H \times (0.4 - 0.05 \times (|45 - |\text{angle}|/15)) < D & |\text{angle}| > 5^\circ \\ H \times (0.3 - 0.1 \times (|10 - |\text{angle}|/2)) < D & \text{else} \end{cases} \quad (13)$$

1. 设置参数: Δ, γ, b_x, b_y

2. 初始化 trajectory = list

3. Do each frame

4. for frame in v

5. $(c'_x, c'_y) \leftarrow (c_x, c_y)$

6. $\text{tra}_{x,y} \leftarrow (\frac{c'_x}{2}, \frac{c'_y}{2})$

7. trajectory $\leftarrow \{ \text{tra}_{x,y}, \dots \}$

8. end for

9. end Do

输出: 轨迹坐标 trajectory

在该方法中,将边界框、预测类别为人、当前视频帧传递到 GOTURN 网络进行目标跟踪。首先,从当前帧向 GOTURN 网络输入边界框坐标,将当前视频帧和边界框进行裁剪得到带目标的中心区域,然后,将得到的上一帧目标和当前帧的搜索区域同时经过 CNN 的卷积层,然后回归当前帧目标的边界框位置,并绘制当前帧坐标框中心点作为轨迹点,输出轨迹坐标。

2.4 轨迹分析方法

在上一步的目标跟踪过程中,对处理的每一帧视频都会得到一个跟踪轨迹。首先,去除太短或太长的轨迹。然后,确定轨迹是否与标记的轨线位置的线段相交。另外,轨迹的最高点应在轨迹的中间,也就是说,轨迹的中间点高于坐标系中的起始点和终点。轨迹的最高点和最低点满足以下条件:

$$A_{(x,y)} \begin{cases} A_x = \frac{\max(c_x) + \min(c_x)}{2} \\ A_y = \frac{\max(c_y) + \min(c_y)}{2} \end{cases} \quad (10)$$

其中, c_x, c_y 分别表示该轨迹点的横坐标和纵坐标, A_x, A_y 分别表示横纵坐标平均值。轨迹点满足以下条件:

$$A_{(x,y)} < 50\% \times \text{trajectory} \quad (11)$$

其中, trajectory 表示轨迹点集合。轨迹的高度 H 和宽度 L 满足以下条件:

$$\begin{cases} \frac{L}{H} < 5 & |\text{angle}| > 5^\circ \\ \frac{L}{H} < (5 - 0.6 \times (10 - |\text{angle}|/2)) & \text{else} \end{cases} \quad (12)$$

其中, angle 表示轨迹线与栏杆线之间的夹角。

以上条件作为条件 1。

另外,距离 D 的起始点和终点之间的轨迹满足以下条件:

$$\begin{cases} H \times 4.5 > D & | \text{angle} | > 5^\circ \\ H \times (0.4 - 0.05 \times (| 45 - | \text{angle} || 15)) > D & \text{else} \end{cases} \quad (14)$$

上面的条件分别表示为条件 2 和条件 3, 如果满足上述 3 个条件则判定该轨迹是翻越行为。

该文给出了翻越行为的判定算法, 如算法 2 所示。

算法 2: 翻越行为判定。

输入: 轨迹坐标 trajectory, 栏杆坐标 line

1. 设置参数: 阈值 M , 最大检测数量 N

2. 初始化 str(label)

3. Do list \leftarrow trajectory

4. for tra_{x,y} in trajectory

5. if 条件 1, then

6. if 条件 2 or 条件 3, then

7. label \leftarrow 'crossing'

8. end if

9. end if

10. end for

11. end Do

输出: 标签 label

3 实验分析

3.1 实验设置

文中实验的硬件环境: CPU 为 Intel i7 9700K, 内存为 16G RDD4, GPU 为两块 Nvidia RTX 2080ti, 运行环境为 Linux 操作系统 (Ubuntu 16.04.6)。编程语言为 Python3.7。神经网络由 Pycaffe 框架搭建, 并使用 Django 作为后端框架。

3.2 数据集构建

3.2.1 目标跟踪模型数据集构建

由于在 GOTURN 模型中的分类标签较多, 为使模型更适用于本实验, 增加了部分有翻越行为的视频数据并删除了部分不符合监控场景的视频数据。另外, 由于原始数据集中的视频序列过少, 该文采用数据增强方法 (随机裁剪、翻转、镜像等) 增加视频序列。

实验的目标跟踪任务采用整理后的 ALOV300 数据集^[14]的 1 575 个视频序列与 ILSVRC2014 数据集^[15]的 134 821 张静态图片数据进行训练。其中在 ALOV300 数据集中每个视频大约每五帧都标记了被跟踪对象的位置, 这些视频通常很短, 从几秒钟到几分钟不等。本实验将这些视频中的 1 281 个作为训练集, 294 个作为验证集, 在选择超参数之后, 使用整个训练集 (训练+验证) 来训练该模型, 另外, 测试集由来自 VOT 2014 年跟踪挑战^[16]的 175 个视频组成。

3.2.2 轨迹分析方法数据构建

由于在公开数据集中缺少翻越行为相关视频, 本实验使用自行模拟监控场景拍摄的翻越行为视频数据。

在整个翻越行为识别过程中, 为检测在不同监控场景下对该行为识别的准确性, 运用 4 个场景对该行为进行检测, 如图 3 所示, 由上到下分别为场景 1、场景 2、场景 3、场景 4。场景 1 有 7 个视频数据, 场景 2、3、4 分别有 6 个视频数据, 合计有 25 个视频数据, 其中每个场景有 1 个走路行为, 其余都为翻越行为。同样地, 该文采用数据增强方法增加视频序列, 合计 175 个视频数据, 如表 1 所示。



图 3 场景展示图

表 1 各视频场景下视频数量

视频场景	视频类别	类别名	数量/个
场景 1	翻越	Crossing	42
	其他	Others	7
场景 2	翻越	Crossing	35
	其他	Others	7
场景 3	翻越	Crossing	35
	其他	Others	7
场景 4	翻越	Crossing	35
	其他	Others	7
总计			175

3.3 性能评估

3.3.1 目标检测模型评估与分析

将文中方法与文献[11]提出的 Adaboost 目标检测方法进行比较, 选取表 1 中的 175 个视频数据, 当检测方法能够完整描绘人物边界框则视为检测准确, 以平均帧率 FPS 和准确率 Accuracy (表示描绘出完整人物边界框) 作为评价指标。

表 2 Yolo 模型性能评估

方法	FPS	视频数 (正确数)	Accuracy/%
Adaboost	25	175 (91)	52
Yolo	65	175 (165)	95.42

由表 2 可知,Yolo 方法相较于 Adaboost 方法在 FPS 上更高,实时性更好,准确率上也高出近 43%。在实际测试中,由图 4 可知,Adaboost 方法描绘的人物边界框大小不可改变,实用性较差,而 Yolo 方法描绘的边界框贴近人物实际大小,能更好地适用于后续的跟踪部分。



图 4 Adaboost 方法(上)与 Yolo 方法(下)实测图

相较于文献[11]提出的方法,该文在目标检测中使用实时性和准确率更高的 Yolo 方法,而不使用前景判断分类器与头部检测分类器 Adaboost 相结合的方式,简化了目标检测的步骤,提高了实时性与人物边界框的准确率。

3.3.2 目标跟踪模型评估与分析

图 5 为该模型在 Iteration = 200 000 次时的 loss 函数的变化。由图 5 可知,当迭代到 40 000 到 60 000 之间,loss 值有显著波动趋势,之后,随着迭代次数的增多,呈现梯度下降趋势,损失值稳定在 30 左右。

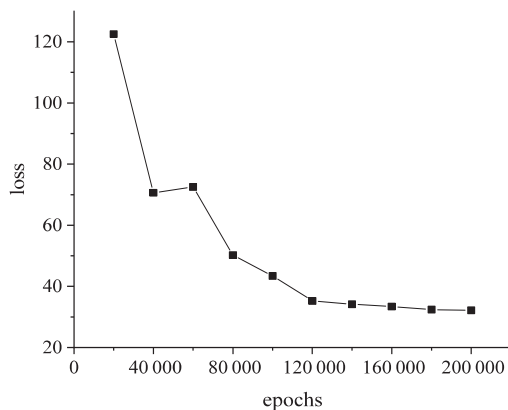


图 5 GOTURN 目标跟踪模型的损失变化

表 3 GOTURN 模型性能评估 %

模型训练方式	Overall errors	Accuracy errors	Robustness errors
仅训练图像	33.33	50.06	18.22
仅训练视频	30.46	46.80	15.08
同时训练	22.38	37.28	10.2

该模型采用 Overall errors、Accuracy errors 以及 Robustness errors 来评估模型性能。由表 3 可见,“仅训练视频”比“仅训练图像”时模型的误差率更低,并且,通过同时训练图像和视频的方式,该跟踪器学会了在不同条件下跟踪各种移动对象,实现了性能最大化。

3.3.3 轨迹分析方法评估与分析

由于实验标签为二分类任务,第一种分类表示预测值与真实值不同(识别失败),第二种分类表示预测值与真实值相同(识别成功),因此该方法采用 Accuracy 来度量轨迹方法的性能。

表 4 轨迹分析方法性能评估

视频场景	视频数(识别数)	Accuracy/%
场景 1	49(42)	85.71
场景 2	42(41)	97.62
场景 3	42(40)	95.24
场景 4	42(41)	97.62
总计	175	93.71

由表 4 可知,在 4 种场景下,该方法基本都能够实现对翻越行为的精确判定,在场景 1 和场景 3 各有 1 次错误识别,出现该错误识别可能是由于拍摄的视频中镜头有抖动,导致轨迹分析中轨迹点不准确。总的来说,该方法达到了 93.7% 的准确率,能够实现对各种旅游景区场景中的翻越行为识别。

将文中方法与现有方法相比较,使用准确率(Accuracy)指标评估模型性能,如图 6 所示。

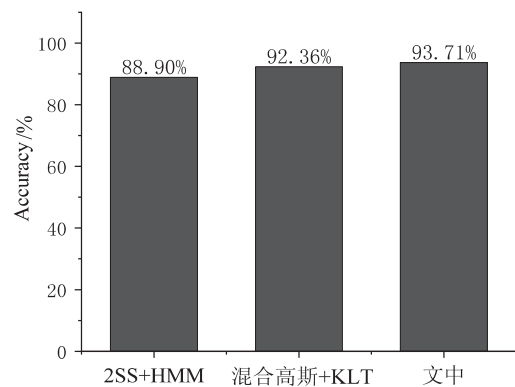


图 6 翻越行为识别不同模型比较

在使用轨迹分析方法数据的情况下,提出的基于 Yolo 和 GOTURN 方法的准确率达到 93.71%。将 2SS+HMM 方法作为基线,则混合高斯+KLT 方法比基线高出近 3%,这表明该方法比基线更有效,而文中方法比混合高斯+KLT 方法高出近 1%,能更快速、更准确地定位人物框,从而更有效地完成翻越行为识别任务。

4 结束语

随着旅游市场的快速发展,景区监管制度不断完善,对于各种违规行为和危险行为的识别变得越来越

重要,通过对这类行为进行识别并提出预警已成为社会关注点之一。该文提出了基于 Yolo 和 GOTURN 的景区游客翻越行为识别方法,主要通过视频分割方法得到每一帧视频,然后通过 Yolo 目标检测方法得到视频帧中的人员坐标,再通过 GOTURN 目标跟踪方法得到轨迹点集合,最后通过轨迹分析得到“crossing”或“no”标签,以形成完整的翻越行为识别过程。实验表明,提出的基于 Yolo 和 GOTURN 的景区游客翻越行为识别方法应用到监控场景下可以达到 93.7% 的准确率。另外,在旅游景区实时监控场景下的单一行为识别任务中,还可以通过人员与边界线之间的相对位置关系判定其他行为,例如湖边的落水行为,这些还有待进一步研究。

参考文献:

- [1] SIMONYAN K, ZISSERMAN A. Two-stream convolutional networks for action recognition in videos[C]//Annual conference on neural information processing systems (NIPS). Montreal, Quebec, Canada; [s. n.], 2014:568-576.
- [2] WANG Limin, XIONG Yuanjun, WANG Zhe, et al. Temporal segment networks: towards good practices for deep action recognition[C]//European conference on computer vision. Amsterdam, The Netherlands; Springer, 2016:20-36.
- [3] SONG Sijie, LAN Cuiling, XING Junliang, et al. An end-to-end spatio-temporal attention model for human action recognition from skeleton data[C]//Proceedings of the thirty-first AAAI conference on artificial intelligence (AAAI). San Francisco, California, USA; AAAI, 2017:4263-4270.
- [4] TRAN D, BOURDEV L, FERGUS R, et al. Learning spatio-temporal features with 3D convolutional networks[C]//IEEE international conference on computer vision (ICCV). Santiago, Chile; IEEE, 2015:4489-4497.
- [5] DIBA A, FAYYAZ M, SHARMA V, et al. Temporal 3D ConvNets: new architecture and transfer learning for video classification[J]. arXiv:1711.08200, 2017.
- [6] YANG Ke, QIAO Peng, LI Dongsheng, et al. Exploring temporal preservation networks for precise temporal action localization[C]//Proceedings of the thirty-second AAAI conference on artificial intelligence (AAAI). New Orleans, Louisiana, USA; AAAI, 2018:7477-7484.
- [7] MABROUK A B, ZAGROUBA E. Abnormal behavior recognition for intelligent video surveillance systems: a review[J]. Expert Systems with Applications, 2018, 91(1):480-491.
- [8] YU E, AGGARWAL J K. Detection of fence climbing from monocular video[C]//8th international conference on pattern recognition (ICPR). Hong Kong, China; [s. n.], 2006:375-378.
- [9] YU E, AGGARWAL J K. Recognizing persons climbing fences[J]. International Journal of Pattern Recognition and Artificial Intelligence, 2009, 23(7):1309-1332.
- [10] YU E, AGGARWAL J K. Human action recognition with extremities as semantic posture representation[C]//IEEE conference on computer vision and pattern recognition (CVPR). Miami, FL, USA; IEEE, 2009:1-8.
- [11] 张泰, 张为, 刘艳艳. 周界视频监控中人员翻越行为检测算法[J]. 西安交通大学学报, 2016, 50(6):47-53.
- [12] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//IEEE conference on computer vision and pattern recognition (CVPR). Las Vegas, NV, USA; IEEE, 2016:779-788.
- [13] HELD D, THRUN S, SAVARESE S. Learning to track at 100 FPS with deep regression networks[C]//Computer vision - ECCV 2016. Amsterdam, The Netherlands; Springer, 2016:749-765.
- [14] SMEULDERS A W M, CHU D M, CUCCHIARA R, et al. Visual tracking: an experimental survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(7):1442-1468.
- [15] RUSSAKOVSKY O, DENG J, SU H, et al. Imagenet large scale visual recognition challenge[J]. International Journal of Computer Vision, 2015, 115(3):211-252.
- [16] KRISTAN M, PFLUGFELDER R P, LEONARDIS A, et al. The visual object tracking VOT2014 challenge results[C]//Computer vision - ECCV 2014. Zurich, Switzerland; Springer, 2014:191-217.