计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2019,28(3):51−58 [doi: 10.15888/j.cnki.csa.006830]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于 SSD 卷积神经网络的公交车下车人数统计①

李继秀, 李啸天, 刘子仪

(西南交通大学 唐山研究生院, 唐山 063016)
通讯作者: 李继秀, E-mail: 804138731@qq.com

摘　要: 传统典型的公交车人数统计方法在准确率和速度方面存在一些不足, 且提取目标特征的效果较差. 本文提出了基于深度卷积神经网络的公交车人数统计系统解决人群计数问题. 首先制作数据集, 难点在于所有用于训练的数据集均是手工标注. 并且公交车摄像头角度比以往文献覆盖更广区域. 本文首先比较了多种不同的深度卷积神经网络模型对乘客进行全身检测的效果. 综合考虑检测速率、准确率等方面, 最终采用单次检测器深度卷积神经网络模型对乘客进行人头目标检测, 在线实时目标追踪算法实现人头的多目标追踪, 跨区域人群计数方法统计公交车下车人数. 系统准确率达到 78.38%, 运行速率约为每秒识别 19.79 帧. 实现了人群计数.

关键词: SSD 目标检测; 卷积神经网络; SORT 目标追踪; 跨区域人数统计

引用格式: 李继秀,李啸天,刘子仪.基于 SSD 卷积神经网络的公交车下车人数统计.计算机系统应用,2019,28(3):51–58. http://www.c-s-a.org.cn/1003-3254/6830.html

Statistics on Number of People Getting off Bus Based on SSD Convolutional Neural Network

LI Ji-Xiu, LI Xiao-Tian, LIU Zi-Yi

(Graduate School at Tangshan, Southwest Jiaotong University, Tangshan 063016, China)

Abstract: The statistics of traditional and typical bus passengers have some shortcomings in accuracy and speed, and the effect of extracting target features is poor. This study proposes a bus counting system based on deep convolutional neural network to solve the crowd counting problem. The first thing to make a dataset is that all the datasets used for training are hand-labeled. And the bus camera angle is wider than the previous literature. This study first compares the effects of various deep convolutional neural network models on the whole body detection of passengers. Considering the detection rate and accuracy, the single-detector deep convolutional neural network model is used to detect passengers' heads. The simple online and real-time target tracking algorithm implements multi-target tracking of human heads, and the cross-region crowd counting method is used to count the number of passenger getting off the bus. The system accuracy rate reaches 78.38% and the operating rate is approximately 19.79 frames per second. the passenger count is achieved.

Key words: SSD target detection; convolutional neural network; SORT target tracking; cross-region population statistics

在各大城市公共交通系统中, 随着人口密度的不竭增长, 城市公共交通系统经常出现满载现象. 乘客进出公交车的人数信息是控制和管理公交车交通系统非常重要的一环. 早期计数的方法大部分为人工计数, 不仅耗时长, 效率低下, 而且消耗大量的人力物力, 加重交通系统运营成本. 深度卷积神经网络的发展在特征提取上给视频监控下的交通控制带来了惊人的便捷, 可以端对端训练公交车视频监控场景下人群计数算法

① 基金项目: 四川省国土资源厅项目 (KJ-2018-16)
Foundation item: Fund of Land and Resources Bureau of Sichuan Province (KJ-2018-16)
收稿时间: 2018-09-29; 修改时间: 2018-10-23, 2018-10-29; 采用时间: 2018-10-31; csa 在线出版时间: 2019-02-22

模型, 省去了前景分割和人为地设计和提取特征等步骤, 通过神经网络多层卷积之后得到的高层特征, 能使得人群计数算法的功能愈加卓越, 增强了人数统计数目的可信度. 因此, 本文提出基于深度卷积神经网络技术的公交车视频监控场景下的人数统计系统, 给公交车管理人员合理地调度车辆提供参考.

在许多基于视频图像处理的人流量统计技术中, 根据检测模块的不同, 可以分为人头检测[1,2]、头肩检测[3,4]和全身检测[5]. 近些年, 又发展了运用卷积神经网络[2,6]实现人流量统计. 公交车的视频监控场景下进行人数统计主要有三种: 基于差分统计的公交车人流量统计[7,8]、基于计数线法和阈值法的公交车人流量统计[9]和基于使用块运动功能的公交车人流量统计[1].

本课题的难点有三: 首先, 所针对的公交车视频数据分辨率较低 (352×288). 且由视频获取的视频帧图像往往存在运动模糊的缺点. 其次, 摄像头角度覆盖范围广. 最后, 本课题摄像头角度与已知文献中的公交车摄像头角度有较大区别, 传统过线计数统计的方法无法适用本课题. 本论文利用 SSD (Single Shot Detector) 深度卷积神经网络模型进行人头目标检测, SORT (Simple Online and Real-time Tracking) 目标追踪算法实现人头的多目标追踪, 跨区域人群计数方法统计公交车下车人数. 系统准确率达到 78.38%, 实现了基于深度卷积神经网络模型实现公交车监控场景下的人数统计.

1 系统框架

基于深度卷积神经网络算法实现公交车视频监控场景下的人群数量统计, 并实现一个监控系统, 对人群数量和人群密度实施实时监控、数据剖析. 公交车视频监控场景下的人数统计的系统框架如图 1 所示. 主要是三大部分: 公交车人头目标检测、对人头运动轨迹进行目标跟踪、跨区域人数统计.

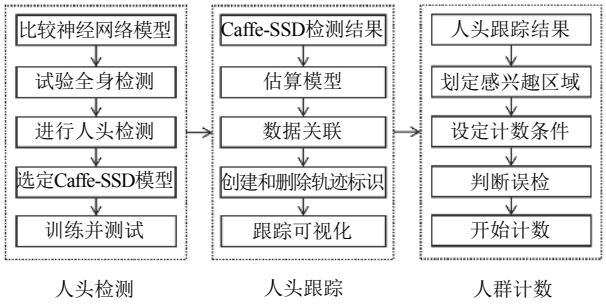


图 1 系统框架

2 本文算法

本章首先阐述数据集的分析和处理, 然后运用 YOLO (You Only Look Once) 两个不同的版本和 Faster-RCNN (Faster Region based Convolutional Neural Network) 以及 SSD 目标跟踪算法对公交车乘客进行全身检测. 然而在公交车这一特定视频监控场景空间里, 由于摄像头涉及空间狭小, 行人之间遮挡严重, 效果并不理想, 漏检率和误检率较高. 故综合考虑, 本论文目标检测采取运用 Caffe-SSD 网络模型进行人头检测的方法. 再利用 SORT 目标追踪算法实现人头的多目标追踪, 最后结合跨线人群计数方法统计公交车下车人数.

2.1 数据集的分析及处理

输入视频源, 使用 OpenCV 计算机视觉库的相关 API 库, 来对输入的 200 个原始视频帧率为 15FPS 的视频文件按每秒抽取一帧的方式切帧, 得到公交车后门摄像头所能覆盖区域的 306 712 张图片, 图片的分辨率是 352×288. 图片涉及多种场景, 包括白天、黑夜; 涉及多种人群密度, 包括拥挤、疏散等.

为了验证对乘客进行全身检测的可行性, 本系统首先分别使用不同深度神经网络模型, 如 YOLO[10]模型, Faster-RCNN[11]模型, SSD[12]模型, 设置置信度均为 50%, 对本论文使用的公交车上下车数据进行目标检测. 检测结果如图 2 所示, 图 2(a) 是 YOLOv1-darknet 网络模型对行人进行全身检测的检测结果, 由图可知, 检测结果较准确, 但系统识别框内为“person”类的概率较低, 置信度太低. 图 2(b) 是 YOLOv3-darknet 模型检测结果, 由图可知, 识别框内为“person”类的概率较高, 但框中把两个人识别为同一人. 图 2(c) 是 VGG+Faster-RCNN 模型检测结果. 识别框内识别框内为“person”类的概率较高, 但把背景中的不是“person”的类别识别成了“person”类, 误检率较高. 图 2(d) 是使用 SSD 模型对人数较少情况进行全身检测的检测结果, 图 2(e) 是 SSD 模型对人数较多情况进行全身检测的检测结果. 由检测结果可知, SSD 模型无论在哪种人群密度下, 都能较准确识别出车中所有乘客, 但全身检测面积太大, 不利于后续的目标跟踪和人群计数. 这是本论文选择使用 SSD 模型作为目标检测的原因之一.

由图 2 所知, 图片中的矩形框是模型识别出的乘客, 虽然这些深度学习网络模型都能在一定程度上识别出乘客, 但由于公交车摄像头覆盖区域空间狭小, 乘

客之间存在严重遮挡, 还有视频抽帧存在的运动模糊等问题. 对乘客进行全身检测, 漏检率和误检率都很高. 考虑摄像头角度问题和乘客头部面积较小且遮挡不严重等特征, 同时为了便于实现目标帧之间的乘客追踪和人群计数. 因此本论文目标检测采用 SSD 深度神经网络模型对乘客进行头部检测. 经过手工标注数据集, 本论文最终选取了 10 654 张图像用于标注, 其中 70% 的数据集用于训练, 30% 的数据集用于测试.

(a) YOLOv1 (b) YOLOv3 (c) VGG+Faster-RCNN

(d) SSD人数较少 (e) SSD人数较多

图 2 不同网络模型对乘客进行全身检测结果

2.2 运用 Caffe-SSD 模型进行公交车人头检测

SSD[12]深度学习目标检测算法是在 YOLO[10]算法上改进而来, 基于端对端方法, 无区域提名, 使用 VGG-16-Atrous 作为基础网络, 沿用了 YOLO 中直接回归 bbox 和分类概率的方法, SSD 与 YOLO 差异之处是除了在选取的 5 个特征图上进行预测, 还有在最终特征图上做目标检测. SSD 还参考了 Faster R-CNN 目标检测算法, 大批使用 anchor 来提升识别精确度, 应用全图全部位置的多尺度区域特征并进行回归. 由于结合这两种结构, SSD 综合了 Faster R-CNN 的 anchor box 和 YOLO 端对端的单个神经网络检测思路. 所以 SSD 能保持较高的识别准确度和识别速度.

相比 YOLO 模型, SSD 模型不仅在检测速度上有了很大的提升, 而且检测识别精确度也有所加强. SSD 是 YOLO 的多尺度版本, 由于 YOLO 对小目标检测效果不好, 所以 SSD 在不同的 feature map 上分割成 grid 然后采用类似 RPN 的方式做回归, 较好地解决了这一问题, 因此本论文选择的目标检测算法为 SSD 算法.

本论文首先创建 VOC 格式的数据集, 所有人头数据集均为手工标注, 耗费了大量的人力. 难点在于将自己的数据集整合到 SSD 卷积神经网络中, 以期提高实验性能. 数据集有多种种类, 如 JPG 图像、带有数据标注的 XML 文件、带有图像信息和标注图像坐标等信息的 TXT 文件. 文件归类、训练集和测试集分别处理, 并将数据集转换为 LMDB 格式. 接着需要修改卷积神经网络结构代码中相关路径; 修改 Max_iter, 分别为 4000 和 8000, 即训练和测试的迭代次数; 修改 Batch_size, 本文设为 16, 既能较好提高内存利用率, 又能使修正方向得到较好的收敛; 修改学习率, 使得学习率随着训练帧数的变化而变化, 防止过拟化; 设置置信度为 50%, 当目标检测结果在置信区间内, 就能检测出来; 最后是设置了类的数量, 本文只需用到一个类, 故 num_class 设置为一加上类别数等于二. 在 VGG16+SSD 网络结构的基础上, 结合本论文使用的公交车人头检测数据集, 进行了训练和测试. 参考 VGG16+SSD 原文的网络框架, 配置网络框架环境, 读取数据集中验证集和测试集的数据. 训练过程可视化结果如下:

(1) 训练过程中迭代 40 000 次, 准确率呈曲线增长, 后逐渐平缓,并逐渐稳定在 84.18% 左右. 训练迭代 40 000 次准确率过程变化图如图 3(a) 所示; 训练过程中迭代 80 000 次, 准确率在 84.20% 附近波动, 最终准确率为 84.22%, 比迭代 40 000 次的准确率高了 0.04%. 训练迭代 80 000 次, 准确率过程变化图如图 3(b) 所示.

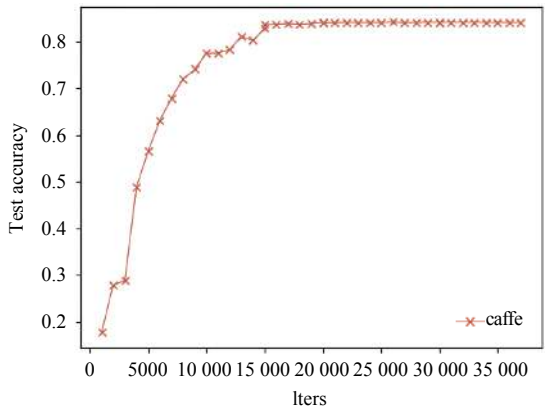
(2) 为了获得更好的训练效果, 迭代 40 000 次训练过程中, 学习率呈梯状逐渐减小, 迭代 40 000 次的学习率变化如图 4(a) 所示; 迭代 40 000 次到 80 000 次的学习率都是 0.000 000 01, 如图 4(b) 所示.

(3) 迭代 40 000 次训练过程中, loss 函数的值呈曲线逐渐减小, 并渐渐稳定. 在训练过程中 loss 丢失变化图如图 5(a) 所示; 迭代 40 000 次到 80 000 次训练过程中, loss 丢失值在 3 附近波动, 和迭代 40 000 次效果差不多. 在训练过程中 loss 丢失变化图如图 5(b) 所示.

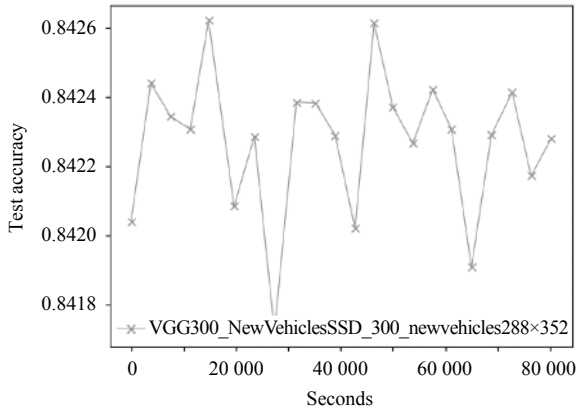
2.3 公交车人头跟踪和人群计数

本论文使用的目标追踪算法: SORT 目标追踪算法[13]. SORT 目标追踪算法是一个通过对 MOT 基准进行测试来实现的基于深度学习的多目标追踪算法. 尽管 SORT 目标追踪算法只是简单的结合了卡尔曼滤波器追踪算法和匈牙利跟踪算法, 但是却能有效将对象与联机和实时应用程序关联起来. 同时又由于本论文使用的 SORT 算法方法简单, 复杂度低, 所以速度性能不错, 较好地协调了精度和速度, 算法实时性好, 能够实现在线更新. 比其他在线追踪器要快 20 倍左右. 在

本章中, 由于基于 CNN 的检测器的灵活性, 它自然可以推广到公交车头部对象类. 因而该方法适用于在公交车视频监控场景下对乘客进行头部跟踪.



(a) 迭代40 000次



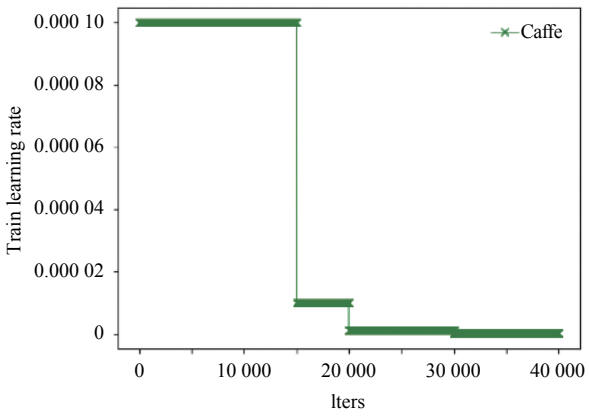
(b) 迭代40 000到80 000次

图 3 训练过程准确率变化

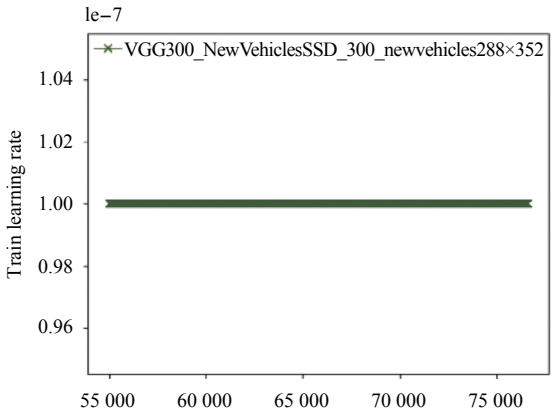
本论文运用了跨区域人群计数方法来做人流量统计. 算法的核心思想是通过划定一个矩形区域, 也就是划线. 如果框内有人, 结合目标跟踪, 如果目标运动 N 个像素点后消失, 或人本不在框内, 等待进入框内后, 运动 M 个像素点后消失, 上下车人数就加一. 本文的人群计数算法增加了判断功能, 也就是误检的情况. 误检的判断条件为: 是否小于 60 个像素点. 最终确定的计数算法的一些关键约束: 行人在 Y 方向运动的距离阈值设置为 25, 划线检测的位置设置为 85. 本论文跨线人群计数技术的流程图如图 6 所示.

3 实验结果及分析

本文基于卷积神经网络的人头检测训练过程中, 所使用的数据集包含多种人流量情况和不同的公交车行车环境, 如图 7 所示. 保留置信度在 50%, 模型每训练 10 张, 记录一次输出结果. 训练总时长达 50 小时, 记录了模型迭代 40 000 次和迭代 80 000 次的结果. 设置训练过程中图像的高为 352, 宽为 288, 每次训练的图像数为 16 张, 为了防止模型过拟化, 所以学习率随着训练图像的增多, 要逐渐减少. 本系统目标检测无需分类, 只有一个类. 用于训练的数据集包括测量环境变化和乘客数量变化等多种情况, 如图 7(a)–(d) 所示, SSD 网络模型训练本论文使用的数据集的检测结果如图 8 所示, 人头目标检测准确率约为 84.22%. 图 8(a) 是乘客比较稀疏的检测结果, 虽然大部分人头都检测到了, 但仍然存在一定的漏检. 图 8(b) 是乘客比较密集的检测结果, 几乎图片中所有的人头都检测到了. 图 8(c) 是比较稀疏且光照环境为黑夜下的检测, 检测结果比较准确,只有一个人头没有检测到. 图 8(d) 是光照环境为黑夜, 但漏检率较高的情况. 综上所知, 本系统公交车人头检测适用于各种光照环境.

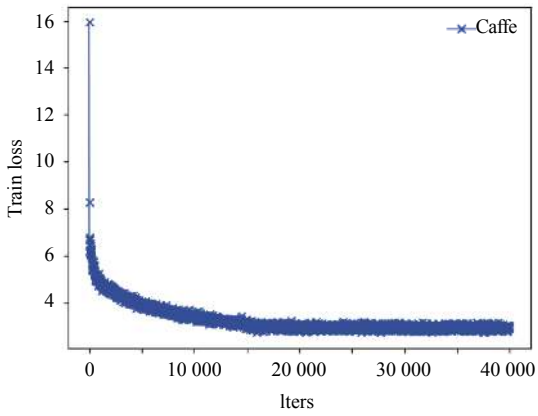


(a) 迭代40 000次

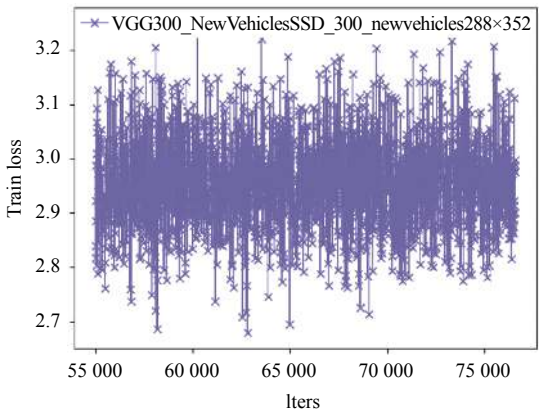


(b) 迭代40 000到80 000次

图 4 训练过程中学习率的设定



(a) 迭代40 000次



(b) 迭代40 000到80 000次

图 5 训练过程中 loss 丢失变化图

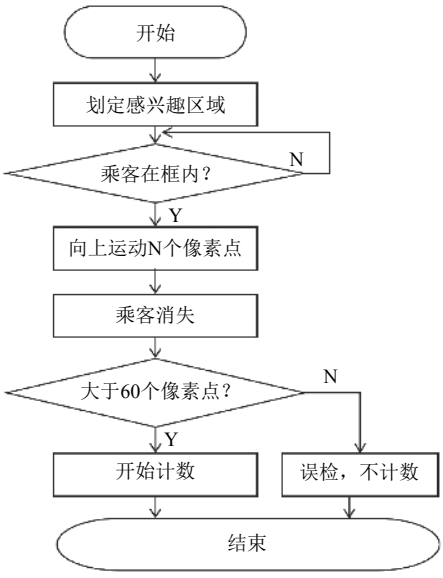


图 6 跨线人群计数流程图

本文对数据进行了训练集、测试集进行了公交车人头检测. 使用的检测架构是 Caffe-SSD 深度神经网络模型, 人头检测结果准确率约为 84.22%. 再由 SORT 目标跟踪算法集中处理帧到帧之间的关联, 实现人头检测目标的跟踪. 整合公交车人头检测、SORT 目标跟踪和本论文使用的感兴趣区域计数方法后, 就完成了基于深度卷积神经网络的公交车人数统计系统. 由于每个视频都包括很多下车片段, 故随机抽取了三个视频样本, 运行速率等于运行总帧数和运行时间的比值, 视频准确率的计算方法等于时间段内检测人数和实际下车人数的比值. 三段视频运行速率和准确率信息表如表 1 所示, 得出实验结果: 本论文的公交车人数统计的系统准确率约为 78.38%, 运行速率约为每秒识别 19.79 帧.

3.1 正常情况

人数稀疏情况下的目标跟踪和人数统计的追踪效果如图 9(a)–(f) 所示, 这是比较有代表性的一组照片. 是实现目标跟踪, 以及乘客一旦下车就进行的计数的效果连帧中能表现出下车过程的抽帧. 图中由点连成的曲线是乘客的运动轨迹, 图像左上角数字是当前视频的下车人数, 图像中的两条横线是越区域人数统计划定的矩形框, 人头上的数字是视频中系统创建的第 N 条轨迹. 可以看到乘客人头形心中点离开矩形框后几帧内, 下车人数加一.

人数拥挤情况下的目标跟踪和人数统计的追踪效果如图 10(a)–(f) 所示, 是检测效果连帧中能表现出下车过程的抽帧. 该视频片段中, 上一次下车的人数共有 4 人, 故下车人数统计之前, 左上角的数字即为 4. 图 10(a) 和图 10(b) 是乘客开始下车, 图 10(c) 是第一个人下车后, 左上角下车人数加一; 图 10(d) 是第二个人下车后, 下车人数再加一. 接下来的检测效果帧都是乘客下车, 人数就自动加一. 图 10(e) 是该下车片段的第 40 帧, 可以看到左上角的数字由刚开始的 4 变成了 11, 图 10(f) 中公交车开始关门, 本次下车人数统计结束, 共统计到 7 名乘客下车, 实际下车人数为 8 名, 可见系统有较高准确率.

黑暗行车环境情况下的目标跟踪和人数统计的追踪效果如图 11(a)–(f) 所示. 由于本论文使用的所有视频数据中, 黑暗行车环境数据本不多, 而黑暗行车环境下乘客下车的数据为零, 故只能在图 11 系列图片中看到系统检测到乘客人头, 并对乘客人头进行追踪的检测效果.

(a) 白天行车环境 (b) 黑色行车环境 (c) 人数拥挤情况 (d) 人数稀疏情况

图 7 数据集的多样性

(a) 白天人数稀疏 (b) 白天人数密集 (c) 黑夜检测较准确情况 (d) 黑夜检测漏检率较高情况

图 8 训练 SSD 网络模型后的人头检测结果

表 1 三段视频运行速率和准确率信息表

视频序列号	1	2	3	合计
视频总帧数	14 486	9428	8360	32 274
运行时间 (s)	729	554	348	1631
实际下车人数	33	29	12	74
系统统计人数	26	22	10	58
运行速率 (帧/s)	19.87	17.02	24.02	19.79
准确率 (%)	78.79	72.41	75.00	78.38

3.2 误检情况

特殊情况下, 计数系统会存在小概率的误检漏检. 如图 12(a) 到图 12(b) 所示误检情况, 如圆圈圈中的人头所示, 乘客未下车, 却进行计数. 这种情况是乘客一直在划定的矩形框内, 若视频连续几帧没有检测到该人头, 计数系统则认为是目标运动了 N 个像素点, 并消失, 并且大于 60 个像素点, 所以图片左上角的计数加一.

(a) 下车第1帧 (b) 下车第4帧 (c) 下车第8帧 (d) 下车第12帧 (e)下车第16帧 (f)下车第20帧

图 9 稀疏情况下的目标跟踪和人数统计效果图

(a) 下车第1帧 (b) 下车第6帧 (c) 下车第10帧 (d) 下车第15帧 (e)下车第40帧 (f)下车第42帧

图 10 拥挤情况下的目标跟踪和人数统计效果图

(a) 行车第1帧 (b) 行车第4帧 (c) 行车第8帧 (d) 行车第12帧 (e)行车第16帧 (f)行车第20帧

图 11 黑夜情况下的目标跟踪和人数统计效果图

(a) 误检情况之未计数 (b) 误检情况之已计数

图 12 误检情况

3.3 漏检情况

计数系统也存在小概率的漏检情况. 检测过程中可能出现如图 13(a)–(d) 所示情况, 圆圈内的乘客就是漏检情况, 乘客已经下车, 计数系统却没有加一. 这是因为乘客的身高达不到我们划定的矩形框感兴趣区域, 或者仅能达到矩形框的下框线, 这样就不能再框内判断乘客运动轨迹是否是向上运动. 同样, 图 13 中左上角显示的是此段视频当前已下车人数的总和.

最后, 选取了和本文实现人群计数方法比较相近的文献[2], 并从目标检测方法、目标跟踪方法和人数统计方法和人数统计系统的准确率这四个方面和本文所引用的文献[2]做一个比较, 比较结果如表 2 所示. 本文运用了较为先进的 SSD 深度卷积神经网络实现了公交车人头目标检测, 获得了更高的人数统计系统准确率.

(a) 运动第一帧 (b) 运动第二帧 (c) 运动第三帧 (d) 运动第四帧

图 13 漏检情况

表 2 实验结果和引用文献的比较

论文名称	本文	文献[2]
目标检测方法	SSD 实现人头检测	Adaboost 人头检测器模型
目标跟踪方法	SORT 人头跟踪	在线人头检测
人群计数方法	感兴趣区域计数	越线人群计数
系统准确率 (%)	78.38	72.00

4 结论与展望

本系统提出了一种基于深度卷积神经网络实现公交车视频监控场景中的人数统计. 该系统分三步走, 首先是用 Caffe-SSD 神经网络框架进行目标检测, 准确率约为 84.22%, 运行速率约为每秒识别 19.79 帧. 然后是运用 SORT 目标跟踪算法实现了公交车乘客多目标追踪, 最后依靠越线人数统计技术对人群进行人数统计. 兼顾检测质量与速度, 实现了一个前后端均较稳定的公交车人数统计系统, 且系统准确率达 78.38%. 能够较准确对公交车下车人数进行统计, 但也存在部分漏检和误检的情况. 本系统可初步试验用于城市公交系统, 帮助公交车运营管理人员分析客流量结果, 从而合理地调度车辆.

参考文献

1 Chen CH, Chang YC, Chen TY, *et al*. People counting system for getting in/out of a bus based on video processing. 2008 Eighth International Conference on Intelligent Systems Design and Applications. Kaohsiung, China. 2008. 565–569.

2 张雅俊, 高陈强, 李佩, 等. 基于卷积神经网络的人流量统计. 重庆邮电大学学报 (自然科学版), 2017, 29(2): 265–271.

3 Xu HZ, Lv P, Meng L. People counting system based on head-shoulder detection and tracking in surveillance video. 2010 International Conference on Computer Design and Applications. Qinhuangdao, China. 2010. 394–398.

4 Zeng CB, Ma HD. Robust head-shoulder detection by PCA-based multilevel HOG-LBP detector for people counting. 2010 20th Conference on Pattern Recognition. Istanbul, Turkey. 1995. 2069–2072.

5 Li JW, Huang L, Liu CP. An efficient self-learning people counting system. The First Asian Conference on Pattern Recognition. Beijing, China. 2011. 125–129.

6 付敏. 基于卷积神经网络的人群密度估计[硕士学位论文]. 成都: 电子科技大学, 2014.

7 李衡宇, 何小海, 吴炜, 等. 基于计算机视觉的公交车人流量统计系统. 四川大学学报 (自然科学版), 2007, 44(4): 825–830. [doi: 10.3969/j.issn.0490-6756.2007.04.022]

8 Rigoll G, Eickeler S, Müller S. Person tracking in real-world scenarios using statistical methods. Proceedings of Fourth IEEE International Conference on Automatic Face and Gesture Recognition. Grenoble, France. 2000. 342–347.

9 鲜晓东, 石亚麋, 唐云建, 等. 基于乘客多运动行为的公交客流计数判定方法. 计算机工程, 2015, 41(4): 176–180, 186. [doi: 10.3969/j.issn.1000-3428.2015.04.033]

10 Redmon J, Divvala S, Girshick R, *et al*. You only look once: Unified, real-time object detection. 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. 2016. 779–788.

11 Ren SQ, He KM, Girshick R, *et al*. Faster R-CNN: Towards real-time object detection with region proposal networks. Proceedings of the 28th International Conference on Neural Information Processing Systems. Cambridge, MA, USA. 2015. 91–95.

12 Liu W, Anguelov D, Erhan D, *et al*. SSD: Single shot multibox detector. In: Leibe B, Matas J, Sebe N, eds. Lecture Notes in Computer Science. Cham: Springer, 2016. 21–37.

13 Bewley A, Ge ZY, Ott L, *et al*. Simple online and realtime tracking. 2016 IEEE International Conference on Image Processing. Phoenix, AZ, USA. 2016. 3464–3468.