

基于高速公路场景的车辆目标跟踪^①



宋焕生¹, 李莹¹, 杨瑾², 云旭¹, 张韞², 解熠²

¹(长安大学 信息工程学院, 西安 710064)

²(陕西省交通建设集团公司, 西安 710064)

通讯作者: 李莹, E-mail: 734778674@qq.com

摘要: 车辆目标检测与跟踪是高速公路视频监控实时获取交通参数的关键步骤. 本文提出了一种面向高速公路场景的目标轨迹时序信息结合核相关滤波 KCF 算法的车辆目标跟踪方法, 实现了车辆目标的高精度持续跟踪. 该方法首先采用基于深度学习的单目标检测 SSD 算法, 通过建立车辆数据集, 实现了适用于高速公路场景的车辆目标的分类与检测. 然后, 基于目标轨迹时序信息实现目标车辆与轨迹的匹配, 并且采用 KCF 跟踪算法对丢失目标进行预测重定位, 从而实现车辆目标轨迹的持续跟踪. 实验表明, 该跟踪方法精度高, 且适应多种不同场景, 具有较高的应用价值.

关键词: 目标跟踪; SSD; 目标检测; KCF

引用格式: 宋焕生, 李莹, 杨瑾, 云旭, 张韞, 解熠. 基于高速公路场景的车辆目标跟踪. 计算机系统应用, 2019, 28(6): 82-88. <http://www.c-s-a.org.cn/1003-3254/6853.html>

Vehicle Object Tracking Method Based on Highway Scenario

SONG Huan-Sheng¹, LI Ying¹, YANG Jin², YUN Xu¹, ZHANG Yun², XIE Yi²

¹(School of Information Engineering, Chang'an University, Xi'an 710064, China)

²(Shaanxi Provincial Communications Construction Group Corporation, Xi'an 710064, China)

Abstract: Vehicle object detection and tracking is a key step for the real-time monitoring and acquisition of traffic parameters in the video monitoring system of expressway. A vehicle tracking method based on trajectory temporal information with KCF algorithm is proposed to realize high precision continuous tracking. Firstly, data sets is established, the classification and detection of vehicle applicable to highway scenario are realized by using SSD algorithm based on deep learning. Then, based on the trajectory of temporal information, matching between object and trajectory is realized, and KCF algorithm is applied to forecast the missing object positioning, so as to realize the vehicle trajectory tracking. The experiment result shows that this tracking method has high precision, can adapt to many different scenarios, thus has high application value.

Key words: object tracking; SSD; object detection; KCF

随着社会经济飞速发展, 交通运输对我国经济和社会发展有着重要的促进作用. 智能交通系统 (Intelligent Transportation Systems, ITS) 作为一个综合的交通监控管理系统, 道路交通信息则是必不可少的前提和内容^[1].

因此, 基于视频的交通信息获取成为智能交通系统的核心. 高速公路作为当前道路的主要交通方式, 基于当前大量的高速公路视频监控设备, 通过对视频图像进行处理与分析, 实现车辆目标的检测与跟踪, 从而获取

① 基金项目: 国家自然科学基金 (61572083); 陕西省重点研发计划重点项目 (2018ZDXM-GY-047); 中央高校团队培育项目 (300102248402)

Foundation item: National Natural Science Foundation of China (61572083); Key Research and Development Program of Shaanxi Province (2018ZDXM-GY-047); Team Cultivation Project for the Central Universities of China (300102248402)

收稿时间: 2018-10-16; 修改时间: 2018-11-06; 采用时间: 2018-11-19; csa 在线出版时间: 2019-05-25

道路交通信息,自动地对高速公路场景进行监控监测,实现交通管理的智能化。但是,高精度的车辆目标检测与跟踪仍是智能交通系统的一个难点。

传统的目标检测算法包括:帧差法^[2]、背景差分法^[3]、光流法^[4]等。但是,当目标运动缓慢时,帧差法易将运动目标当做背景图像造成目标检测失败;背景差分法是对背景进行建模得到前景的运动目标,受外界光照的变化,背景运动等因素的影响;光流法采用光流信息描述目标, Meye 等提出了基于轮廓的目标检测跟踪算法,实现了较好的检测跟踪效果,但该方法存在抗噪能力差,计算复杂,对硬件要求高等缺点。基于深度学习的目标检测算法与传统算法相比拥有更高的检测精度,其主要运用卷积神经网络 CNN^[5]。目前基于深度学习的目标检测主流算法有,基于区域建议的检测算法 Fast R-CNN^[6]和基于回归的检测算法 YOLO^[7]、SSD^[8]。

针对目标跟踪,跟踪预测算法本质上是通过提取已检测目标的特征信息,在后续图像序列中搜索定位与目标模板相似的区域作为候选位置。目标特征的提取的一种方式是通过提取特征点,然后利用特征点进行目标匹配的重定位。常用的特征点提取方式有 Harris^[9]角点检测、SIFT^[10]算子特征点检测,基于特征点的匹配能够适应目标的旋转和光照变化,但是特征点提取过多导致匹配困难,特征点过少又容易造成误检,而且特征点提取过程计算复杂,耗时严重。另一种方式是对目标整体描述,常用到的目标特征包括图像边缘、形状、纹理、颜色直方图统计特征等,通过融合多个特征增强目标特征的可靠性。对目标进行特征提取之后,采用相似性度量方式进行目标重定位,实现目标跟踪。基于特征的跟踪算法对运动目标的尺度、形变和亮度等变化不敏感,即使目标部分被遮挡,还能够根据部分特征信息完成跟踪任务。但是其对于图像模糊、噪声等较敏感,特征的提取效果也依赖于各种提取算子及其参数的设置。

因此本文采用基于深度学习的 SSD 目标检测算法,通过数据集训练得到适应高速公路场景的模型,实现高精确度的车辆目标检测。在目标检测结果的基础上,本文提出了一种目标轨迹时序信息结合 KCF^[11]算法的车辆目标跟踪方法,实现车辆目标的持续跟踪,具有较好的应用价值。

1 车辆目标跟踪总体框架

目前,多目标跟踪有两种解决方案。方案一:对视

频序列的每一帧进行运动目标检测,基于相邻帧间的检测结果完成运动目标轨迹连接,获取运动目标信息。方案二:在初始帧对目标进行检测得到有效描述,然后在后续图像序列中查找与已检测目标相匹配的目标区域,最终跟踪得到运动目标轨迹。

为了实现高速公路场景下的高精准检测和车辆持续跟踪,本文将上述两种方案相结合,提出了一种面向高速公路场景车辆目标检测及跟踪方法,实现了车辆目标的持续跟踪。高速公路场景下车辆目标的跟踪,分两大模块,一是目标车辆检测,二是目标跟踪。目标车辆检测采用基于深度学习的 SSD (Single Shot MultiBox Detector) 目标检测算法,目标的持续跟踪首先利用轨迹时序信息进行数据关联,然后结合 KCF 算法校正断开轨迹。如图 1 是面向高速公路场景下的车辆目标跟踪过程的总体框架。

2 车辆目标跟踪方法

2.1 目标检测方法

本文采用基于深度学习的 SSD 目标检测算法,其网络结构如图 2 所示。其中使用 VGG16^[12]作为基础网络,该网络包括 16 层,网络模型如图 3 所示,SSD 在 VGG16 的基础上添加了特征提取层。具体的 SSD 算法请参考文献[8]中刘伟等人的工作。

为了更适用高速公路场景下的车辆目标检测,本文将 SSD 的检测类别由原来的 20 种改为 4 种,并参考国内标准的车型分类手册^[13],将车辆目标分为三类:小车 (Car)、客车 (Bus)、卡车 (Truck),如图 4 为车型类别的样本示例图。本文以杭金衢高速公路的监控视频作为样本,针对 4 种不同场景下的视频每 75 帧提取一帧作为本文的训练样本集,总计提取 6288 张图片,并通过 LamImage 工具对目标车辆进行位置标注,对类别用标签进行区分,得到 6288 个标注文件。其标注工具如图 5 所示,生成的标签文件如图 6 所示。

本文的输入网络图像大小为 512×512,训练迭代次数设置为 12 000 次,根据本实验的 GPU 的性能将 BatchSize 设置为 4,训练学习率 LearningRate 设置为 0.004×0.05,然后将制作好的数据集输入 SSD 网络进行训练,其中网络训练过程的 Loss 曲线图如图 7 所示, Loss 随着迭代次数的增加逐渐降低并趋于稳定,说明达到了训练的收敛要求,即得到适合高速公路场景的网络模型。

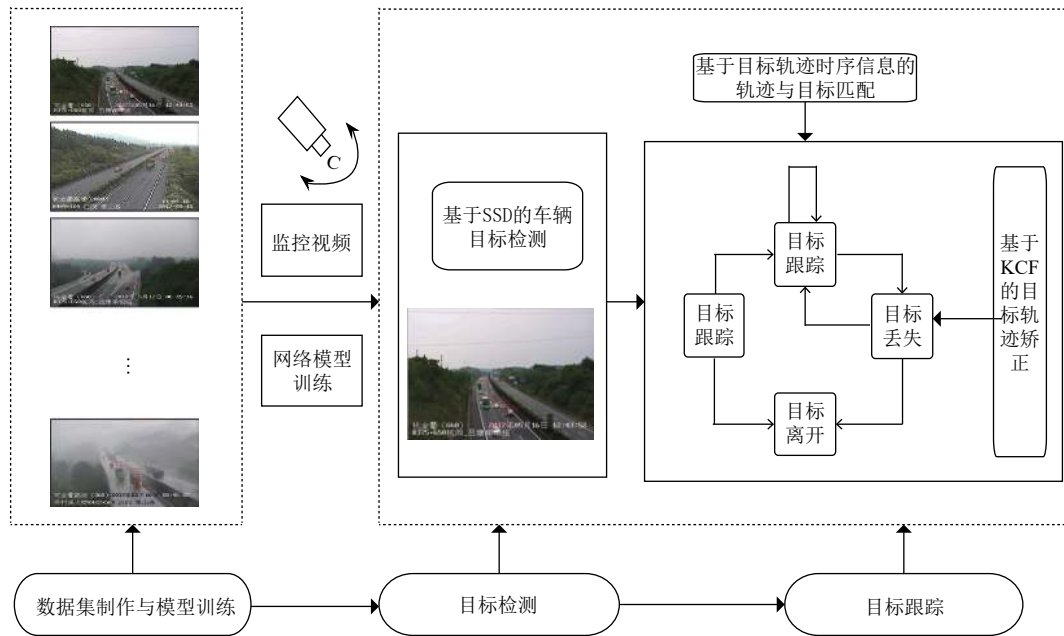


图1 车辆目标跟踪过程的总体框架

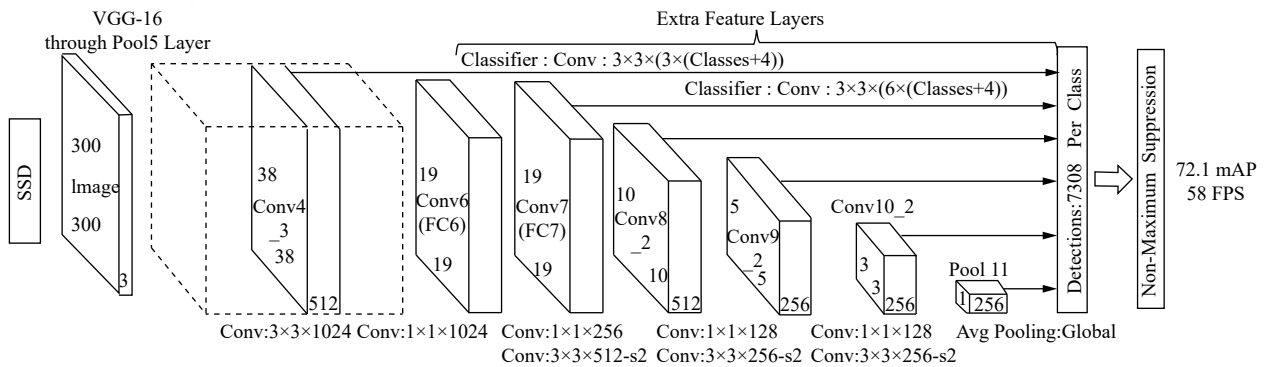


图2 SSD网络结构

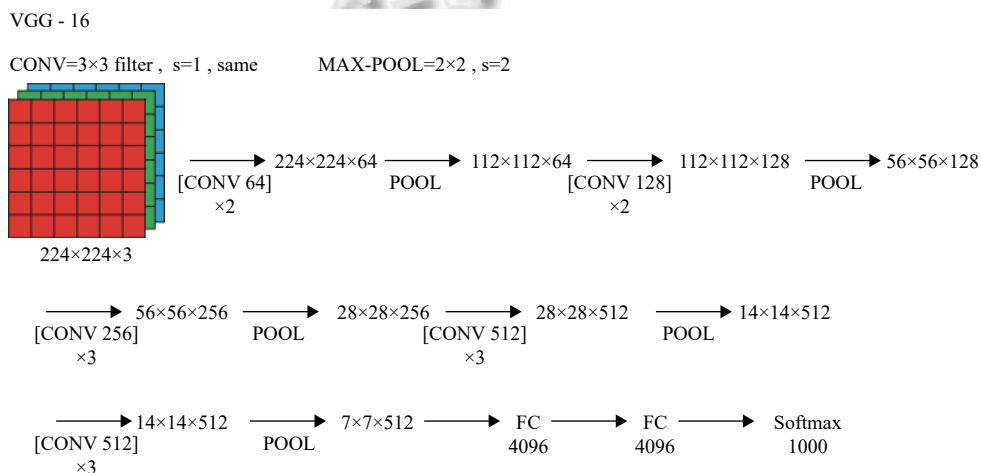


图3 VGG16网络模型及每层的输出



图4 车型类别的样本示例



图5 LabelImg 图像标注工具

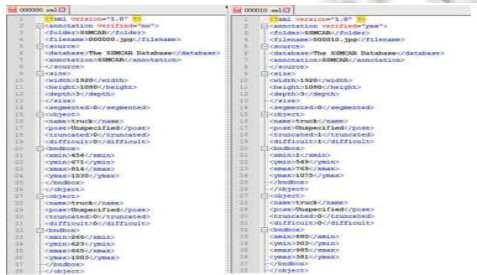


图6 LabImage 生成的标签文件

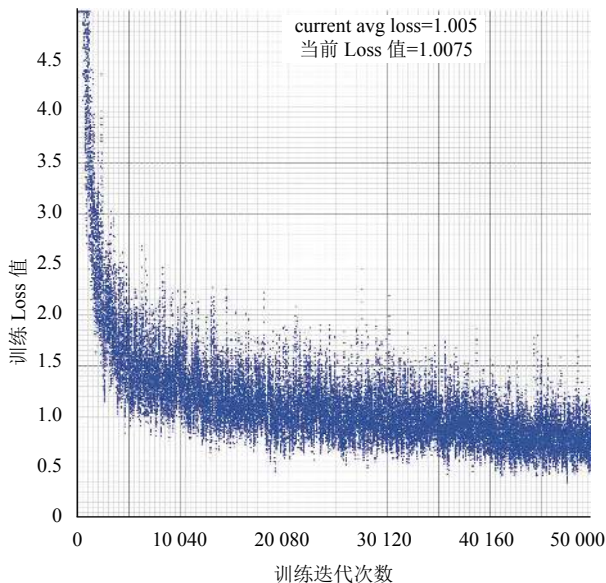


图7 网络训练过程的 Loss 曲线

2.2 基于轨迹时序信息的轨迹与目标匹配方法

在检测结果的基础上, 需要通过数据关联方式将同一目标车辆连接起来形成轨迹. 高速公路场景下的

高帧率(一般为 25 FPS) 监控视频, 在相邻帧中同一个车辆目标具有明显的高重叠面积 IOU (Intersection-Over-Union)^[14]. 如图 8 是实际场景中连续帧间的目标位置.



图8 高帧率下目标高重叠面积图

本文提出了一个简单的 IOU 跟踪器, 计算当前帧的检测目标框与现有轨迹的 IOU 匹配值 $Overlap(box1, box2)$. 由于高速公路场景下车辆目标之间存在相互遮挡, 帧间相互遮挡的目标 IOU 匹配值只有在一定范围有效, 本文将有效范围设置为 $[0.6, 1]$. 其中, IOU 匹配值的定义如式 (1) 所示.

$$Overlap(box, box2) = \frac{Area(box1) \cap Area(box2)}{Area(box1) \cup Area(box2)} \quad (1)$$

其中, $box1$ 和 $box2$ 是两个目标检测框, $Area(box1)$, $Area(box2)$ 是其对应的图像面积大小.

2.3 基于 KCF 的核相关滤波算法

在高速公路场景下, 由于光照、遮挡、相机架设位置等因素的影响, 即使是 SSD 算法, 也不能达到 100% 的检测精度, 为了实现目标轨迹的持续跟踪, 当目标检测丢失时, 需要通过跟踪预测算法完成轨迹信息的校正, 从而得到完整的轨迹信息.

本文利用 KCF 跟踪算法对目标车辆进行位置预测, 完成对目标轨迹的校正. KCF 跟踪算法是通过提取目标的 HOG^[15]特征在目标跟踪过程中训练一个目标检测器, 利用目标检测器在下一帧图像中判别预测位置是否是目标, 然后再用新检测结果更新目标检测器. 具体的 KCF 原理过程参考文献[12]. KCF 目标跟踪的基本过程如下:

- 1) 在第 t 帧中, 在当前目标位置 p 附近进行采样, 并且采用循环矩阵增加正负训练样本, 利用脊回归训练目标检测器.
- 2) 在第 $t+1$ 帧中, 在前一帧位置 p 附近进行采样, 用训练的目标分类器计算每一个采样对应的响应值.
- 3) 响应最强的采样作为第 $t+1$ 帧的目标位置.

2.4 目标轨迹时序信息结合 KCF 的 track-by-detection 方法

基于 SSD 的车辆目标检测结果, 一个车辆目标从首次检测到跟踪形成一条有效轨迹的过程中, 目标车辆的状态转换过程如图 9 所示的目标跟踪模块。

1) 目标激活: 当一个目标首次被检测到, 即该目标被激活. 新目标的初始位置框 $box_{i,t}$ (i 表示当前帧的检测序号, t 表示当前处理的帧号) 作为一条新轨迹形成的初始位置, 新轨迹表示为 $trace_{I,t}$ (I 表示轨迹的唯一标识 ID). 与此同时, 为该轨迹建立一个 KCF 跟踪器 $trackerBox_{I,t+1}$, 目的是为了后续防止目标丢失时, 对目标能够进行位置预测。

一个处于激活状态的目标可以经 s1 转换成跟踪和经 s7 转换到离开两种状态. 实际过程中不可避免的会出现错误的目标检测结果, 但是由于误检目标并不能持续地被检测跟踪, 所以本文设定阈值, 如果一个目标处于激活状态, 连续 5 帧被检测且匹配跟踪时, 目标转换到跟踪状态, 否则, 该目标就被认为是错误的检测目标, 状态转换为离开, 与之对应的轨迹也被删除。

2) 目标跟踪: 在第 $t+1$ 帧检测得到 $box_{i,t+1}$, 轨迹 $trace_{I,t}$ 基于其时序信息与当前检测框计算得到匹配代价表. 如果某条轨迹 $trace_{I,t}$ 与某目标框 $box_{i,t}$ 满足式 (1) 的范围取值时, 则表示两者匹配成功, 目标经时转换到跟踪状态, 得到更新轨迹 $trace_{I,t+1}$. 如果轨迹未匹配到目标框, 则该轨迹目标经 s3 进入丢失状态. 如果

该目标未匹配到轨迹, 则该目标是一个新的目标, 经 s0 进入目标激活状态。

3) 目标丢失: 理想情况下, 一个目标不离开场景时会一直保持跟踪状态时, 但是由于光照, 遮挡等原因目标可能未被检测到, 此时轨迹 $trace_{I,t}$ 找不到与之匹配到目标 $box_{i,t+1}$, 则该轨迹目标转换到丢失状态. 当该目标再次被检测到时, 目标可以从丢失状态经 s4 过程重新回到跟踪状态. 当目标处于丢失状态时, 本文采用 KCF 目标跟踪算法, 在当前帧为轨迹目标预测新位置 $trackerBox_{I,t+1}$, 并将预测位置添加到轨迹中。

在本文中, 当一个新目标被激活时, 为该目标初始化一个跟踪器, 在跟踪过程中训练一个目标检测器, 并且通过不断地更新提高目标检测器的准确性, 但是当目标运动过程中形态尺度发生变化时, 检测器的更新过程会造成不断地错误累积, 导致目标漂移, 跟踪失败. 为了适应目标的变化, 每 25 帧重新初始化一次模板, 适应目标的变化。

4) 目标离开: 目标转换到离开状态的情况分三种:

① 目标离开运动场景时, 目标既不会被检测到也不再 KCF 的预测操作, 目标将从跟踪状态先经 s3 转到丢失状态, 再经 s5 转换到离开状态;

② 目标被激活但是不满足连续 5 帧被检测跟踪, 被认定为错误目标, 目标从激活状态转换到离开状态;

③ 当一条轨迹目标连续 40 帧一直处于丢失状态, 即轨迹目标一直处于 s6 的转换过程, 则认为该目标离开。

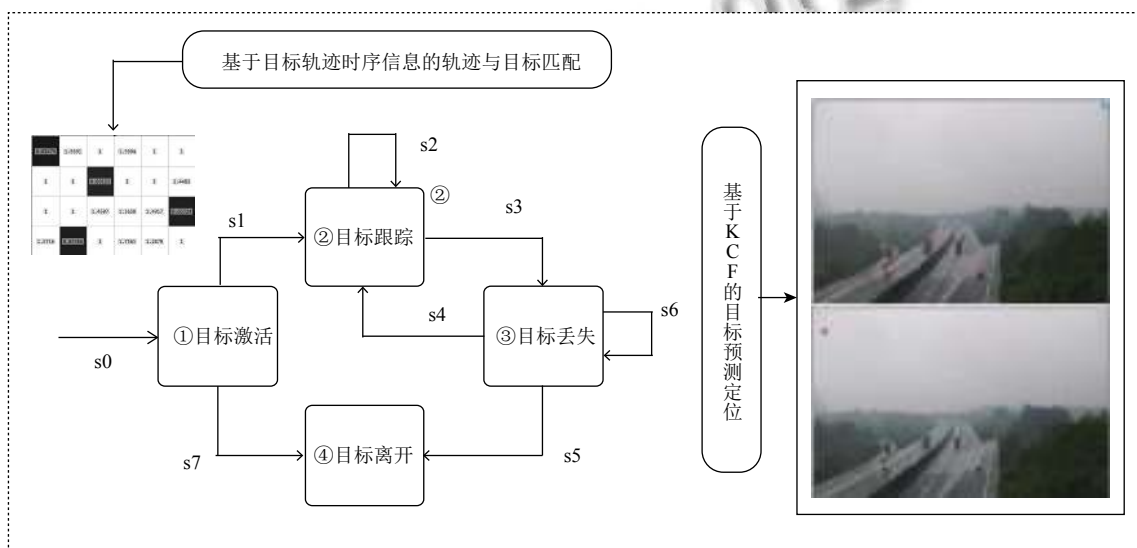


图 9 目标跟踪模块状态转换

2.5 算法步骤

步骤一: 车辆目标检测. 将视频帧序列送入训练好的 SSD 网络模型进行目标检测.

步骤二: 对目标轨迹进行跟踪, 建立匹配代价表.

- 1) 若检测目标与轨迹匹配成功, 更新轨迹;
- 2) 若跟踪态轨迹未匹配到检测目标, KCF 预测目标位置, 更新轨迹;
- 3) 若检测目标未匹配到轨迹, 记为激活态轨迹;
- 4) 激活态轨迹连续 5 帧跟踪成功, 确认更新为跟踪态轨迹;
- 5) 若轨迹满足 2.4 小节中目标离开的 3 个条件之一, 则删除轨迹.

步骤三: 重复步骤一、二. 对车辆目标进行持续跟踪.

3 实验结果分析

本文实验所采用的视频是杭金衢高速 (G60 国

道) 的不同时间和不同路段的监控视频. 表 1 是本文实验的四种不同场景.

表 1 四种不同实验场景

场景	路段	相机安装位置	监测时间	视频分辨率
1	杭向-吕塘角枢纽	K375+650	2017-5-16 12:00-14:00	1920×1080
2	江向-常山县	K449+100	2017-5-16 11:00-12:00	1280×720
3	杭向-吕塘角枢纽	K375+650	2017-5-16 6:00-7:00	1920×1080
4	江向-常山县	K422+560	2017-5-12 15:00-16:00	1920×1080

为了验证 SSD 检测算法的精确度, 本实验针对场景 1 和场景 2 将其与传统的背景建模检测算法相比较, 得到表 2. 通过对比发现, 背景建模检测算法的精度均低于 90%, 而基于深度学习的 SSD 检测算法相较于具有更高的精度, 实验结果显示平均检测精度在 95% 左右.

表 2 SSD 检测算法与传统的背景建模检测算法的精度分析

场景	测试总帧	车辆类型	实际总车数/辆	背景建模方法		SSD 目标检测方法	
				测试车数 (辆)	精度	测试车数 (辆)	精度
1	30 913	Car	692	548	0.792	684	0.945
		Bus	116	82	0.707	113	0.940
		Truck	429	331	0.771	427	0.953
2	4481	Car	95	81	0.853	93	0.959
		Bus	15	11	0.733	14	0.938
		Truck	9	8	0.889	9	1.00

本实验将 KCF 跟踪算法与传统卡尔曼滤波 Kalman 跟踪算法进行比较, 为了公平起见, 我们针对

同一场景同一时间段的车辆目标进行跟踪分析. 如图 10(a) 和 10(b) 是两种跟踪算法的实验结果.



图 10 (a) 卡尔曼滤波 Kalman 跟踪 (b) 核相关滤波 KCF 跟踪

从结果可以看出, 虽然两种跟踪算法都能将目标位置锁定, 但是 Kalman 算法得到的跟踪框位置明显存在较大误差, 而 KCF 算法能够更加精确地锁定目标位

置, 同时也能从上图发现, 即使存在部分遮挡的情况下, KCF 算法也可以实现跟踪.

以上本文单独针对目标检测和目标跟踪做了实验

对比分析,为了验证本文算法的整体精度,针对场景1、2、3本文通过对轨迹条数进行统计分析,并且将其与实际人工检测的数据进行对比,得到表3。本文利用简单的切面检测线方式进行统计计数,精度平均可以达到96.3%。其中小车(Car)类型的平均检测精度

95.6%,卡车(Truck)类型的平均检测精度97.1%,客车(Bus)类型的平均检测精度95.9%。小车精度较低原因在于高速公路场景中,小车(Car)目标小,容易被遮挡,当小车目标是黑色时容易漏检。从整体的检测跟踪精度来分析,本文的方法满足实际高速公路场景的应用需求。

表3 轨迹计数精度分析

场景	测试总帧	车辆类型	测试总车数(辆)	实际总车数(辆)	漏计车数(辆)	多计车数(辆)	每类精度	总精度
1	30 913	Car	684	692	23	15	0.945	0.946
		Bus	113	116	5	2	0.940	
		Truck	427	429	11	9	0.953	
2	4481	Car	93	95	3	1	0.959	0.966
		Bus	14	15	1	0	0.938	
		Truck	9	9	0	0	1.000	
3	25 086	Car	52	54	2	0	0.963	0.975
		Bus	1	1	0	0	1.000	
		Truck	50	51	0	1	0.961	

4 结论与展望

通过实验发现利用基于深度学习的SSD目标检测算法,可以获得了高精度的目标检测结果。其次,采用检测与跟踪相结合的思想,利用轨迹时序信息进行车辆目标的数据关联,采用KCF跟踪算法对丢失检测结果的轨迹进行校正,能够有效完成轨迹的持续跟踪。实验表明,该车辆目标跟踪方法具有较高精度,并且能够适应各种实际高速公路场景,接下来为了达到实时化的检测跟踪效果,需要针对深度学习检测方法进行优化使其速度更快,此外,可以通过已得到的轨迹信息完成交通参数的获取,交通事件的实时监测,对于智能交通领域具有广泛的应用前景。

参考文献

- 杨东凯,吴今培,张其善. 智能交通系统及其信息化模型. 北京航空航天大学学报, 2006, 26(3): 270-273.
- Lv YS, Duan YJ, Kang WW, *et al.* Traffic flow prediction with big data: A deep learning approach. *IEEE Transactions on Intelligent Transportation Systems*, 2015, 16(2): 865-873.
- Ohta N. A statistical approach to background subtraction for surveillance systems. *Proceedings of IEEE International Conference on Computer Vision*. Vancouver, BC, Canada. 2001. 481-486.
- Barron JL, Fleet DJ, Beauchemin SS. Performance of optical flow techniques. *International Journal of Computer Vision*, 1994, 12(1): 43-77. [doi: 10.1007/BF01420984]
- Bouvier J. Notes on convolutional neural networks. *Neural Nets*, 2006.
- 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*. Montreal, Canada. 2015. 91-99.
- Redmon J, Divvala S, Girshick R, *et al.* You only look once: Unified, real-time object detection. *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, NV, USA. 2016. 779-788.
- Liu W, Anguelov D, Erhan D, *et al.* SSD: Single shot multibox detector. *Proceedings of the 14th European Conference on Computer Vision*. Amsterdam, The Netherlands. 2016. 21-37.
- Harris C, Stephens M. A combined corner and edge detector. *Proceedings of the Fourth Alvey Vision Conference*. Manchester, UK. 1988.147-152.
- Luo J, Gwon O. A comparison of SIFT, PCA-SIFT and SURF. *International Journal of Image Processing*, 2017, 3(4): 143-152.
- Henriques JF, Caseiro RC, Martins P, *et al.* High-speed tracking with kernelized correlation filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(3): 583-596. [doi: 10.1109/TPAMI.2014.2345390]
- Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556, 2014.
- 朱殿举. 我国汽车分类标准. *交通世界*, 1997, (8): 13.
- Bochinski E, Eiselein V, Sikora T. High-speed tracking-by-detection without using image information. 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance. Lecce, Italy. 2017. 1-6.
- Dalal N, Triggs B. Histograms of oriented gradients for human detection. *Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. San Diego, CA, USA. 2005. 886-893.