



优化 EfficientDet 深度学习的车辆检测

摘要

针对深度学习 EfficientDet 模型的车辆检测性能进行分析,基于训练过程中容易陷入局部最优进行优化改进,构建分阶段自适应的训练模型,利用该训练模型对短距离和远距离车辆进行检测,并将检测结果与基于 Cascade R-CNN 和 CenterNet 方法进行比较,从计算复杂度、耗时及检测精度三方面分析显示本文方法优于其他两种方法.同时,对不同角度和不同距离车辆检测结果进行分析,确定了检测的最优距离和角度.最后,通过实例验证了本文方法可以用于大范围车辆的检测.

关键词

深度学习;EfficientDet;车辆检测;学习率

中图分类号 TP391.41

文献标志码 A

收稿日期 2021-09-28

资助项目 国家自然科学基金(42171428);湖北省安全生产专项资金科技项目(SJHX 20211006);重庆市技术创新与应用发展专项(cstc2019jscx-msxmX0051)

作者简介

陈西江,男,博士,副教授,从事激光点云及目标识别方向研究.cxj_0421@163.com

- 1 武昌理工学院 人工智能学院,武汉,430223
- 2 武汉理工大学 安全科学与应急管理学院,武汉,430070
- 3 重庆市计量质量检测研究院,重庆,401120

0 引言

随着居民生活水平的不断提高,越来越多的人拥有车辆,而车辆的不断增加也加剧了城市对车辆监控的难度.为了有效地打击违法、违规行为,保障城市安全,城市主要路段都有监控系统^[1-2],实现对机动车和非机动车等目标的有效监管.车辆检测成为城市车辆智能监控系统中的重要功能.在实际应用中,由于受光照条件、部分遮挡及背景杂乱等环境因素的影响,车辆检测的准确度会受到极大影响.为了提高车辆检测的可靠性及准确度,学者们对其进行了大量的研究.

车辆检测方法主要包含基于图像的检测方法和基于深度学习的检测方法.基于图像的检测方法主要是通过车辆图像特征和方向梯度直方图特征来对车辆目标进行检测.例如:文献[3]使用图像的特征变换进行车辆的分类检测;文献[4]通过分析车辆图像的边缘特征结合主成分分析法来实现对车辆的检测.基于图像特征的车辆检测方法虽然计算复杂度较低,能够快速检测车辆,但受环境影响较大,特别是对于车辆有部分遮挡或者光照变化的区域很难对其进行准确检测.因此,文献[5]通过融合 Sobel 和 Sift 特征来对车辆图像特征进行增强处理,并由此实现对车辆的检测;文献[6]利用 Sobel 边缘检测对车位中的车辆进行检测,并实现了不同种类车辆的检测得分;文献[7]使用基于模糊 c 均值聚类 and 超参数调优的分类器来实现对车辆的检测,并基于车辆检测结果完成了车辆在某个车位的定位.由于车辆本身的图像具有独特的纹理特征,所以可以根据车辆纹理特征来实现对车辆的检测^[8].基于车辆图像纹理特征及车辆边缘特征的车辆检测方法的主要缺点是受光照的影响较大.随着深度学习的不断发展,越来越多的学者开始研究利用深度学习来实现对车辆的检测.文献[9]利用卷积神经网络对车辆图像进行训练,并将训练后的模型用于车辆检测;文献[10]提出一种基于 Haar-AdaBoosting 算法和卷积神经网络的车辆检测方法;文献[11]提出超区域候选网络,该深度学习检测模型可以实现对较远摄像头拍摄的小车辆进行检测.在检测车辆的过程中,很难保证车辆不被遮挡,为了解决车辆在遮挡情况下能够实现对其检测,文献[12]将对抗学习引入到 RCNN 目标检测过程中,提高了车辆检测精度.RCNN 算法的优点是检测精度较高,但其检测速度较慢,很难适用于实时视频流车辆检测.为了提高检测效率,文献[13]引入 YOLO 系列算法来实现对车辆的检测,该检测网络主要

是基于 SSD 网络,对车辆的检测速度较快,但其精度不是特别高.为了提高检测精度,文献[14]对 SSD 网络进行了改进,通过融合级联模块和元素模块来实现高精度的车辆检测,但由于融合了较多的模块,使得检测速度有所下降.目前,深度学习在车辆检测领域得到了学术界的广泛关注^[15].因此,本文通过对深度学习 EfficientDet 模型训练阶段进行优化,构建分阶段训练模型,实现快速准确的车辆检测.

1 车辆检测模型的构建

首先利用 EfficientDet 深度学习模型对车辆进行训练,并用训练好的模型对车辆进行检测.基于深度学习的目标识别网络主要包括 Cascade R-CNN^[16]、SpineNet^[17]和 CenterNet^[18]等.为了验证不同深度学习网络的性能,表 1 给出了不同识别网络模型的检测性能.表 1 中的 COCO mAP^[^1]是用于衡量检测性能的平均精度,其值越大说明检测性能越好.

由表 1 可以明显看出, EfficientDet 系列的深度学习网络模型检测性能最好,其次是 Cascade R-CNN_ResNet 深度学习网络模型.而 EfficientDet 系列中检测性能最好的为 EfficientDet-D7x 网络,该网络的识别精度达到了 55.1,其次是 EfficientDet-D3. Efficient-Det-D7x 网络的缺点是复杂度太高,无法实现车辆实时检测.综合考虑,本文选用 EfficientDet-D3 作为车辆检测网络.

表 1 不同识别网络模型的性能

网络模型	速度/ms	COCO mAP[^1]
Cascade R-CNN_ResNet-101	410	42.8
CenterNet_DLA-34	31	41.6
RetinaNet_ResNet-101	32	39.9
EfficientDet-D1	16	40.5
EfficientDet-D3	37	45.6
EfficientDet-D7x	285	55.1

为了构建基于 EfficientDet-D3 优化的车辆检测模型,采集了不同时间段不同场景下的车辆图像数据集,如图 1 所示.该数据集分为训练数据集和验证数据集.总共 7 533 张图像,其中,6 387 张为训练数据集,1 146 张为验证数据集,均通过 LabelImg 工具进行注释.每张图像生成的注释文件通过 Python 脚本文件转换为 Tensorflow 统一的 TFRecord 数据格式.

根据数据集的特点,训练前调整模型对应的配置文件中的参数,包括类别数量、批大小、初始学习率以及相关的数据读取路径.车辆检测模型不受车型的影响,在迭代开始时,选取一部分数据,通过深度学习算法得到预测值.根据数据集大小和计算机配置,首先选择初始训练的批大小,然后根据损失函数值的变化和检测效果进行调整.利用标准梯度下降算法训练车辆检测模型,在模型训练阶段, Tensor-

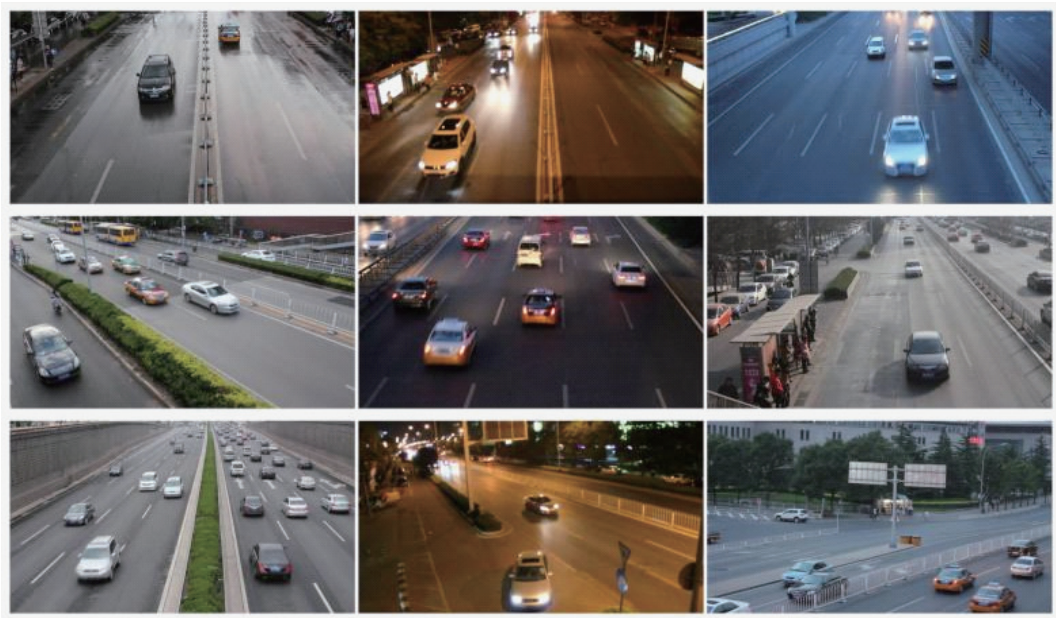


图 1 车辆数据集
Fig. 1 Vehicle data set

Flow_Slim 模块提供了简单但很强大的训练模型.在模型训练过程中,损失函数的返回值即为每次迭代过程中产生的目标函数值.另外,检测模型对车辆定位损失和置信度损失之和是衡量预测模型性能的指标.为了调试和优化模型训练过程,TensorFlow 提供了一个可视化工具 TensorBoard,通过读取记录的数据文件来监控和显示训练过程.批处理的大小越大,内存利用率就越高,处理数据的速度就越快.在训练中,当使用单个 NVIDIA GeForce RTX 2080Ti GPU 时,本文将批大小设置为 16.学习率是影响模型性能最大的参数之一,学习率太大会导致检测不稳定,学习率太小会导致过拟合和收敛速度慢等现象.为了加快模型的收敛速度,本文利用 AdamW 算法^[19]来优化学习率.在训练阶段,如果连续 3 个周期损失不下降,学习率就会降低.本文依据总损失值变化情况对车辆检测模型训练阶段学习率参数的变化情况调整,结果如图 2 所示.

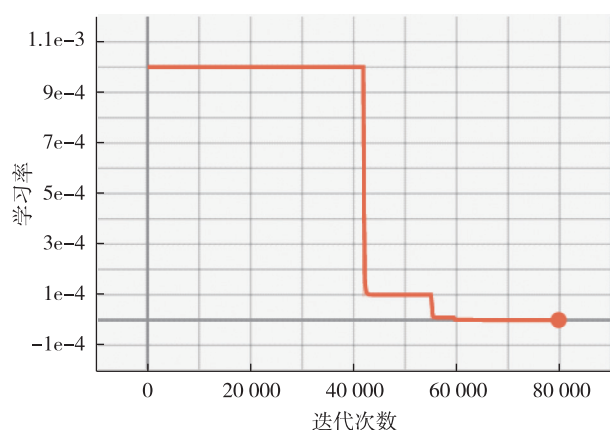


图2 学习率参数的变化

Fig. 2 Variation of learning rate

根据图 2,第 1 阶段为迭代从 0 到 41 870 阶段,初始学习率被设定为 $1e-3$,此时,由图 3 可明显看到,总损失值较为迅速地递减到 0.26,意味着学习率需要调整.第 2 阶段迭代从 41 871 到 54 980,学习率设置为 $1e-4$,由图 3 可知总损失值降低为 0.225,且在其附近趋于平滑.为了确定学习率是否仍然需要优化,第 3 阶段从 59 431 次迭代到结束,学习率被降低到 $1e-6$,总损失值仍然没有变化,其值为 0.225,并趋于平滑.整个模型训练过程结束,总的训练收敛时间约为 37 h,训练得分大于 99.5%,总的学习率参数经过 3 次调整,最终完成深度学习车辆检测模型的构建.

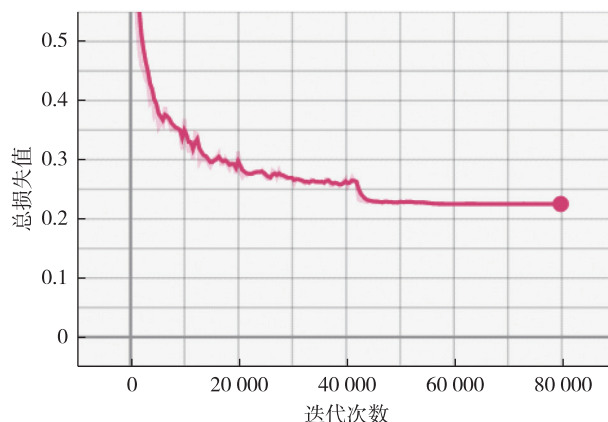


图3 总损失值变化趋势

Fig. 3 Variation trend of total loss value

2 实验分析

1) 不同方法比较

为了验证本文方法的性能,利用 Cascade R-CNN 和 CenterNet 方法对远距离车辆进行检测,并将检测结果与本文方法进行比较.表 2 给出了 3 种方法对车辆检测时需要使用的参数数量.

表2 3种不同方法的参数数量

Table 2 Number of parameters for the three detection methods

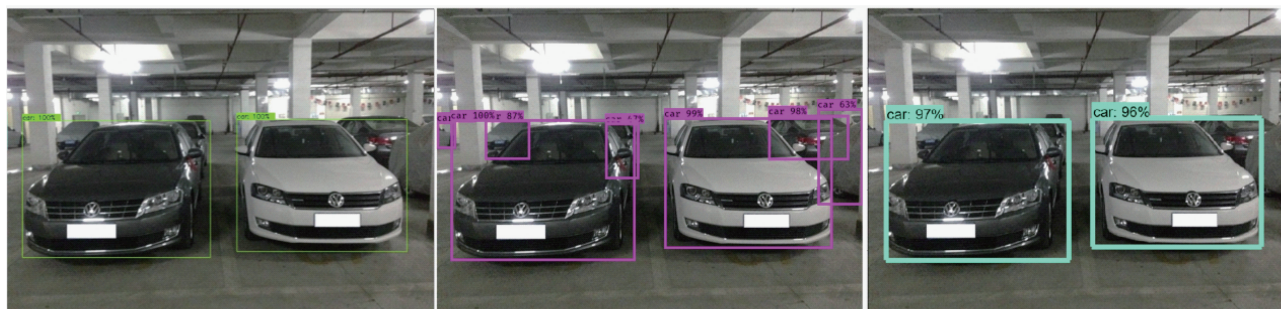
方法	参数/MB
Cascade R-CNN	345
CenterNet	185
本文方法	12

由表 2 可以明显看出,本文方法对车辆检测使用的参数数量最少,Cascade R-CNN 需要使用的参数数量最多,即该方法的计算复杂度最高,其次是 CenterNet 模型.利用这 3 种方法对近距离车辆进行检测,结果如图 4 所示.

由图 4 可明显看出,3 种方法对近距离车辆的检测得分都大于 80%,表明都适合该情况下的车辆检测.其中,本文方法对 2 辆车的检测得分均为 100%,高于其他 2 种方法.

利用 3 种方法对远距离车辆进行检测,检测得分如图 5 所示.

图 5 显示:本文方法对 3 辆车的检测得分均为 100%,基于 Cascade R-CNN 方法对 3 辆车的检测得分分别为 99%、100%和 100%,基于 CenterNet 方法对 3 辆车的检测得分分别为 99%、98%和 99%.本文方法的检测得分最高且最稳定.



a. 本文方法

b. 基于Cascade R-CNN方法

c. 基于CenterNet方法

图4 不同方法对近距离车辆检测效果

Fig. 4 Performance of different methods on short-range vehicle detection



a. 本文方法

b. 基于Cascade R-CNN方法

c. 基于CenterNet方法

图5 不同方法对远距离车辆检测效果

Fig. 5 Performance of different methods on long-distance vehicle detection

为了评估本文方法和其他2种方法在检测车辆上面的检测效率,得到对车辆检测的时间成本,结果如图6所示.由图6可明显看出,基于 Cascade R-CNN 检测时间最长,需要 140 ms 左右,而本文和基于 CenterNet 方法的检测时间相似,且2种方法只需要耗费 32 ms 左右,低于基于 Cascade R-CNN 方法.从检测时间成本上可以看出,本文和基于 CenterNet 方法较好,但是,基于 CenterNet 方法需要的参数较多,计算复杂度较高.因此,从检测时间成本、检测精度及复杂度3个方面综合分析,本文方法要优于基于 Cascade R-CNN 和 CenterNet 方法.

2) 不同场景检测比较

为了评价本文所构建的深度学习车辆检测模型适用于不同场景下的车辆检测情况,对不同场景下的车辆进行检测,结果如图7所示.车辆的检测得分越高,检测效果越好.

由图7可明显看出,3种不同背景下的车辆检测得分均为100%,特别是只有部分车身在图像中显示的情况下,车辆检测得分仍然是100%,表明本文方法对车辆检测效果较好,适用于不同背景下车辆的

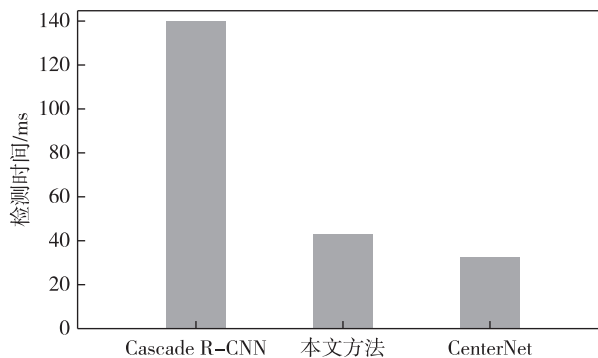


图6 3种不同方法检测时间

Fig. 6 Detection time of three different methods

检测,且能够实现对非完整车辆图像进行检测.

3) 不同距离车辆检测分析

为了验证不同距离对本文方法的影响,选取4个不同距离:8、15、22和30 m,使用训练好的深度学习车辆检测模型对不同距离的车辆进行检测,如图8所示.由图8a可明显看出,在8 m的情况下,4辆车都能被检测到,且检测分数都为100%.摄像头距离车辆15 m的情况下,7辆车检测得分都为



图7 深度学习车辆检测模型对不同场景下车辆的检测效果

Fig. 7 Performance of the proposed model for vehicle detection in different scenarios

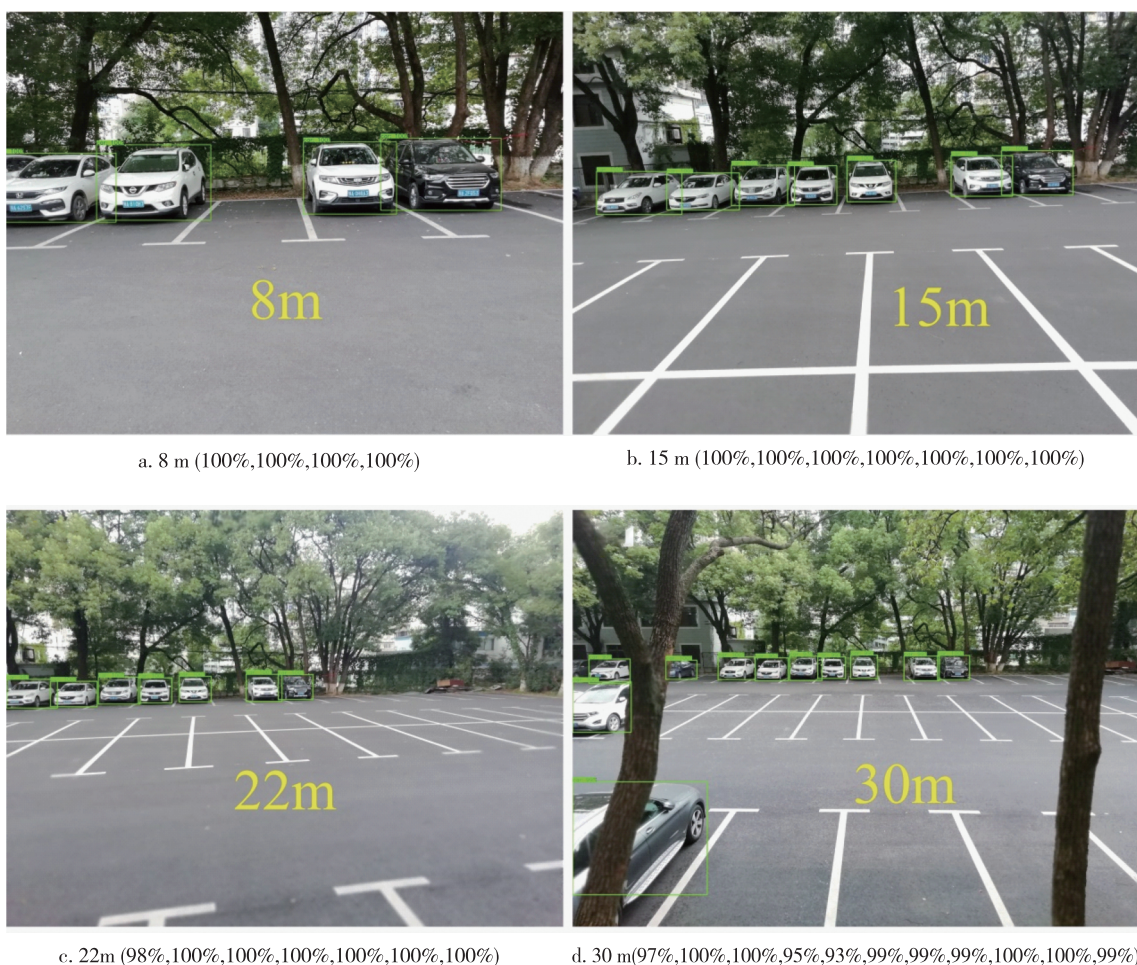


图8 不同距离车辆检测得分

Fig. 8 Vehicle detection scores from different distances

100%,如图8b所示.当距离增加到22 m时,检测得分稍微有所降低,其中有1辆车检测得分为98%,剩余6辆为100%,如图8c所示.从以上分析可知,当距离为22 m以内时,本文方法对每辆车的检测得分几乎不受距离的影响.主要原因是本文方法训练模型所使用的车辆距离都超过了22 m.当检

测距离增加到30 m时,大部分车辆的检测得分大于93%,如图8d所示.即随着距离的增加,检测效果略有降低.因此,在训练模型时,所使用的车辆数据集的距离最好要大于检测车辆的距离,这样可以保证检测车辆的精度.

4) 不同角度车辆检测分析

为了验证不同角度车辆对检测的影响,本文设置6个不同角度:0°、20°、40°、60°、80°和90°,结果如图9所示。

图9a为正对着车辆拍摄,即0°方向,本文方法对所拍摄4辆车均可检测,且检测得分均为100%。图9b显示的是20°的情况,本文方法对5辆车的检测得分分别为100%、100%、100%、100%及89%。当角度变为40°时,相邻车辆图像之间存在重叠区域,由图9c覆盖的5辆车中仅有3辆车能够被检测到。图9d显示的是60°的情况,仅有4辆车被检测到,其得分分别为92%、86%、100%和100%。而从图9e可以看出,12辆车中只有2辆车的检测分数大于90%,其余都无法被检测到。如图9f所示,当角度增加到90°时,几乎所有车辆都无法被检测到。所以,车辆检测效果受角度的影响较大,造成这种现象的可

能原因是训练数据集中不包含不同角度的车辆数据集;同时,角度较大时,图像中的相邻车辆间存在着重叠。因此,为了对车辆进行准确检测,应尽量使用正对着拍摄的车辆。

3 实例分析

利用本文训练好的深度学习车辆检测模型对停车场中的车辆进行检测,其检测后的得分如图10所示。

由图10检测结果可知,实现了对拍摄的12个车辆的检测。其中,最高检测得分为100%,最低检测得分为74%,且只有1辆车检测得分低于90%,其余车辆检测得分均高于90%。高于70%的检测得分被认为是有效检测,即本文检测的结果认定这12个检测均为车辆,与实际情况相同,表明本文方法的检测



图9 6个不同角度车辆检测结果

Fig.9 Vehicle detection scores from six angles

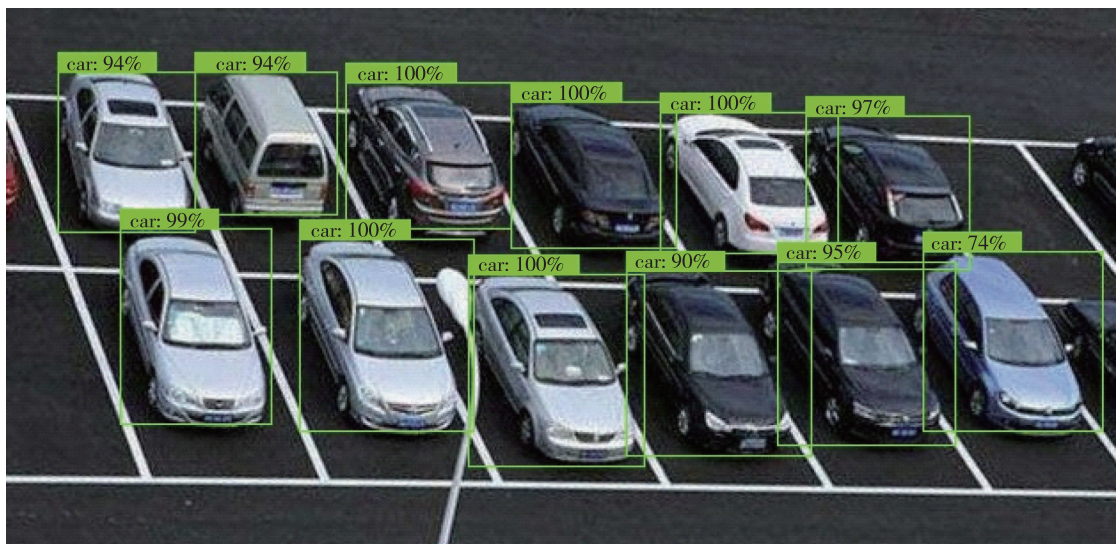


图 10 较大停车位中每个车辆的检测得分情况

Fig. 10 Detection scores of each vehicle in a large parking lot

为可靠检测。

4 结论

本文提出一种基于深度学习的车辆检测方法,建立了基于 EfficientDet-D3 模型的车辆检测模型。在 EfficientDet-D3 模型训练阶段,为了提高检测模型的训练效率,依据总损失值的变化情况,给出分阶段训练学习参数的设定,建立了自适应的模型训练的学习机制。

将本文检测模型与基于 Cascade R-CNN 和 CenterNet 方法进行比较,本文方法使用的参数最少,其计算复杂度最低,本文方法对车辆的检测耗时与 CenterNet 方法相当,低于 Cascade R-CNN 方法,同时 3 种方法的检测精度基本相同。因此,从计算复杂度、耗时及检测精度 3 方面分析确定本文方法优于其他 2 种方法。

利用该检测模型对不同距离和角度的车辆进行判断。结果表明,当检测的车辆距离小于训练的车辆距离时,检测效果不受距离的影响,当检测的车辆距离大于训练的车辆距离时,检测效果随着距离的增加而降低。因此,为了使得摄像头覆盖的所有车辆获得良好的检测效果,本文使用车辆数据集距离较远。角度对车辆检测效果影响很大,特别是当摄像头方向与车辆方向的夹角大于 20° 时,有些车辆无法被成功检测到。因此,摄像头的角度要尽可能小,以保证能够准确地检测到每辆车。

参考文献

References

- [1] 张家旭,杨雄,施正堂,等.汽车紧急换道避障的路径规划与跟踪控制[J].华南理工大学学报(自然科学版),2020,48(9):86-93,106
ZHANG Jiayu, YANG Xiong, SHI Zhengtang, et al. Path planning and tracking control for emergency lane change and obstacle avoidance of vehicles[J]. Journal of South China University of Technology (Natural Science Edition), 2020, 48(9): 86-93, 106
- [2] 赵娜,袁家斌,徐晗.智能交通系统综述[J].计算机科学,2014,41(11):7-11,45
ZHAO Na, YUAN Jiabin, XU Han. Survey on intelligent transportation system[J]. Computer Science, 2014, 41(11): 7-11, 45
- [3] Aarthi R, Padmavathi S, Amudha J. Vehicle detection in static images using color and corner map[C]//2010 International Conference on Recent Trends in Information, Telecommunication and Computing. March 12-13, 2010, Kerala, India. IEEE, 2010: 244-246
- [4] Matos F M D S, Souza R M C R D. An image vehicle classification method based on edge and PCA applied to blocks[C]//2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC). October 14-17, 2012, Seoul, Korea (South). IEEE, 2012: 1688-1693
- [5] Iqbal U, Zamir S W, Shahid M H, et al. Image based vehicle type identification[C]//2010 International Conference on Information and Emerging Technologies. June 14-16, 2010, Karachi, Pakistan. IEEE, 2010: 1-5
- [6] 裴明涛,沈家峻,杨敏,等.复杂光照环境下的车辆检测方法[J].北京理工大学学报,2016,36(4):393-398
PEI Mingtao, SHEN Jiajun, YANG Min, et al. Vehicle detection method in complex illumination environment[J]. Transactions of Beijing Institute of Technology, 2016, 36(4): 393-398

- [7] Ghaffarian S, Gökaşar I. Automatic vehicle detection based on automatic histogram-based fuzzy C-means algorithm and perceptual grouping using very high-resolution aerial imagery and road vector data [J]. Journal of Applied Remote Sensing, 2016, 10(1) : 015011
- [8] Li Y, Li B, Tian B, et al. Vehicle detection based on the AND-OR graph for congested traffic conditions [J]. IEEE Transactions on Intelligent Transportation Systems, 2013, 14(2) : 984-993
- [9] 申铨京, 沈哲, 黄永平, 等. 基于非局部操作的深度卷积神经网络车位占用检测算法 [J]. 电子与信息学报, 2020, 42(9) : 2269-2276
SHEN Xuanjing, SHEN Zhe, HUANG Yongping, et al. Deep convolutional neural network for parking space occupancy detection based on non-local operation [J]. Journal of Electronics & Information Technology, 2020, 42(9) : 2269-2276
- [10] Xiang X Z, Lv N, Zhai M L, et al. Real-time parking occupancy detection for gas stations based on Haar-AdaBoosting and CNN [J]. IEEE Sensors Journal, 2017, 17(19) : 6360-6367
- [11] Tang T, Zhou S L, Deng Z P, et al. Vehicle detection in aerial images based on region convolutional neural networks and hard negative example mining [J]. Sensors, 2017, 17(2) : 336
- [12] Wang X L, Shrivastava A, Gupta A. A-fast-RCNN: hard positive generation via adversary for object detection [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). July 21 - 26, 2017, Honolulu, HI, USA. IEEE, 2017 : 3039-3048
- [13] Lu J Y, Ma C, Li L, et al. A vehicle detection method for aerial image based on YOLO [J]. Journal of Computer and Communications, 2018, 6(11) : 98-107
- [14] Cao G M, Xie X M, Yang W Z, et al. Feature-fused SSD: fast detection for small objects [C] // Proc SPIE 10615, Ninth International Conference on Graphic and Image Processing (ICGIP 2017), 2018, 1061 : 381-388
- [15] Liang S D. Smart and fast data processing for deep learning in internet of things: less is more [J]. IEEE Internet of Things Journal, 2019, 6(4) : 5981-5989
- [16] Cai Z W, Vasconcelos N. Cascade R-CNN: delving into high quality object detection [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. June 18-23, 2018, Salt Lake City, UT, USA. IEEE, 2018 : 6154-6162
- [17] Du X Z, Lin T Y, Jin P C, et al. SpineNet: learning scale-permuted backbone for recognition and localization [C] // 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 13 - 19, 2020, Seattle, WA, USA. IEEE, 2020 : 11589-11598
- [18] Zhou X, Wang D, Krähenbühl P. Objects as points [J]. arXiv e-print, 2019, arXiv : 1904. 07850
- [19] Loshchilov I, Hutter F. Decoupled weight decay regularization [J]. arXiv e-print, 2017, arXiv : 1711. 05101

Optimized EfficientDet deep learning model for vehicle detection

CHEN Xijiang^{1,2} AN Qing¹ BAN Yan³

¹ School of Artificial Intelligence, Wuchang University of Technology, Wuhan 430223

² School of Safety Science and Emergency Management, Wuhan University of Technology, Wuhan 430070

³ Chongqing Academy of Metrology and Quality Inspection, Chongqing 401120

Abstract At present, deep learning has been widely applied in object detection, such as vehicle detection. In this paper, the deep learning EfficientDet model was analyzed, and its advantages in vehicle detection were confirmed. A phased adaptive training model was constructed to avoid local optimum in training process, then it was used to detect vehicles from both short and long distance. The detection results showed that compared with detection methods based on Cascade R-CNN and CenterNet, the proposed model was superior in terms of computational complexity, time consumption and detection accuracy. Meanwhile, further analysis figured out the optimal detection distance and angle. Finally, an example is given to verify that the proposed method can be applied to a large range of vehicle detection.

Key words deep learning; EfficientDet; vehicle detection; learning rate