

基于改进的 FAST R-CNN 的前方车辆检测研究

史凯静 鲍 泓

(北京联合大学北京市信息服务工程重点实验室 北京 100101)

摘 要 目前,前方车辆检测的研究主要通过机器学习的方法,然而其难以解决遮挡和误检的问题。在这种背景下,使用深度学习的方法检测前方车辆更为有效。首先采用了选择性搜索方法获得样本图像的候选区域,然后使用改进的 FAST R-CNN 训练网络模型,检测道路前方车辆。已在 KITTI 车辆公共数据集上对该方法进行了测试,实验结果表明,所提方法的检测率高于 CNN 直接检测的结果,很大程度上解决了遮挡和误检的问题。而且,与先提取 Harr-Like 特征然后利用 Adaptive Boosting 分类器的算法相比,该方法在 TSD-MAX 交通场景数据库测试中实现了较高的性能。结果表明,该方法提高了车辆检测的准确性和鲁棒性。

关键词 前方车辆检测,样本图像,卷积神经网络,准确率

中图分类号 TP311 文献标识码 A

Forward Vehicle Detection Research Based on Improved FAST R-CNN Network

SHI Kai-jing BAO Hong

(Beijing Key Laboratory of Information Service Engineering, Beijing Union University, Beijing 100101, China)

Abstract The current research on vehicle detection is mainly about machine learning, but it is still difficult to deal with occlusion and false detection. In this paper, using deep learning methods to detect forward vehicles is more effective. This paper firstly adopts the selective search method to obtain the candidate area of the sample image, and then uses the improved FAST R-CNN training network to detect the forward vehicles on the road. The method has been tested in the KITTI vehicle public dataset. The experimental results show that the detection rate of this method is higher than that of the direct test based on CNN. The problem of occlusion and error detection is largely solved. Moreover, the widely used method extracts the circulated Harr-Like features, and then uses the Adaptive Boosting classifier algorithm. Compared in TSD-MAX traffic scene dataset, the proposed method provides a higher performance. The results show that this method improves the accuracy and robustness of vehicle detection.

Keywords Forward vehicle detection, Sample image, Convolutional neural network(CNN), Accurate rate

1 引言

前方车辆检测是计算机视觉的研究内容之一,目的是识别和定位图像或视频中的车辆。在国内外 ITS 以及 IV 研究领域,针对前方车辆检测已提出了许多算法和实施手段。研究人员将机器学习应用到车辆检测^[1]的研究中取得了一定的进展,但目前研究前方车辆检测仍面临以下主要问题:实际的道路情况复杂多样,车辆类型、颜色和尺寸有所不同,部分车辆存在大面积遮挡问题。这在一定程度上增加了基于车辆检测模型的匹配难度。而且智能驾驶对于实时 SVM 或神经网络尤其苛刻,基于机器的学习方法在车辆检测耗时方面不能满足智能驾驶的安全要求^[2]。深度学习理论的提出在目标检测领域取得了惊人的效果。通过深度学习的方法构建深度网络来提取车辆目标特征是目前前方车辆检测技术中很受关注的研究方向^[3]。

为了解决这些问题,本文提出了一种基于改进的 FAST R-CNN 的方法,利用选择性搜索(Selective Search)方法^[4]获

得样本图像的候选区域,将候选区域坐标与样本图像一起输入到网络进行学习。图像首先经过深度卷积神经网络计算得到深度卷积特征,然后基于改进的 Fast R-CNN 网络对提取的感兴趣区域通过池化层规格化卷积特征,最后将特征输入到不同的全连接分支,计算特征的分类以及回归计算检测框坐标值。经过多次迭代训练,最后得到具有训练好的权重参数的车辆检测模型。利用该模型在 KITTI 数据集和 TSD-MAX 交通场景数据库上进行测试。实验结果表明,与已有的车辆检测方法相比,所提方法的性能得到提升,为前方车辆的检测问题提供了更泛化和简洁的解决思路。

2 前期工作

车辆检测的性能由特征提取和分类器训练两部分决定。车辆特征提取的方法有: Haar-like 特征、HOG 特征、SIFT 特征、直方图等。常用的分类算法有 Adaptive Boosting、SVM、神经网络等。基于机器学习的方法,文献^[3]将 Haar-Like 特征和 AdaBoost 算法应用于特征提取和分类器

本文受国家自然科学基金重大研究计划(91420202, NSFC61271370)资助。

史凯静(1992—),女,硕士生,主要研究方向为图像处理, E-mail: 1247678671@qq.com; 鲍 泓(1958—),男,教授,主要研究方向为图像处理, E-mail: xxtbaohong@buu.edu.cn。

构造来检测车辆。文献[6]提出了一种基于 LBP 特征的 Haar-Like 特性的鲁棒检测算法。这种方法按照局部二元模式(LBP)计算图像的局部纹理特征;然后,当利用 Adaboost 分类器进行训练时,小部分的特征从大数据集局部纹理特征中被选择出来;最后利用 Adaboost 分类器进行分类并选择特征。实验结果表明,分类器的鲁棒性得到了很大的提高,使得分类器能够准确地检测车辆。张全发等[7]提出的算法利用梯度方向直方图特征(HOG)作为图像描述,并结合支持向量机(SVM)训练分类器,实现了图像中工程车辆的检测。为了提高检测效率和准确度,采用 Hough 变换直线检测的方法提取包含目标的区域。实验表明,该算法能够在较大分辨率的复杂图像中识别出检测目标,有较高的准确度,对于实验中不同的环境具有较强的鲁棒性。其在正常光照下的识别率为 96.52%,误识别率为 3.59%,但在车辆左右边界的定位还不够精准。文献[8]改进了多维 Haar-like 特征以及学习 Adaboost 算法,来实现级联分类器的训练,从而实现可靠的车辆检测。车道间和其他区域的分离使用的是 lane-edge 检测算。实验结果表明,与传统演算法方法相比,该算法能有效地提高车辆检测的准确性和效率。基于深度学习的方法,文献[1]根据卷积神经网络某些层对于目标透视变形具有不敏感性,得到适合车辆检测的卷积神经网络模型;针对未标定的图片,提出了 EdgeCNN 检测框架。首先通过 Edge box 算法获取图片中的预选区域,并根据阈值划分成正、负样本,之后训练得到二分类器卷积神经网络模型。测试过程中,将测试样本输入到 Edge box 中获取预选区域,再将预选区域输入到深度学习网络中进行判别。实验表明,该方法的平均准确度高于其他方法,并且具有一定的鲁棒性。文献[10]先得到候选区域生成网络(RPN),再共享全图像卷积特征与检测网络,来进行前方车辆检测。文献[11]提出了适应的混合神经网络,通过将最后一层分为可变大小的几个网络来提取不同尺度特征。实验结果表明,提出的车辆检测方法超过其他传统方法的检测结果。文献[12]改进了卷积神经网络提出了统一多尺度的深度卷积神经网络 MS-CNN,其通过分成建议网络和检测网络两个子网络,进行车辆检测。结果表明其准确率有明显提高,并且在内存和计算方面有了很大的改进。MS-CNN 检测器的检测率高达 15 帧每秒。文献[13]利用了 Fast R-CNN,在训练 VGG16 网络时运行速度比 RCNN 提高了 9 倍。

本文第 2 节介绍了车辆检测的相关工作;第 3 节详细介绍了前方车辆检测的方法;第 4 节给出了实验结果和分析;最后,总结全文并进行了简要总结和展望。

3 提出的方法

本文提出一种改进的基于可行驶区域的 FAST R-CNN 前方车辆检测方法,该方法利用可行驶区域可以快速准确地确定车辆的位置区域。该算法首先选用了选择性搜索方法得到前方车辆检测图像样本的若干候选区域,深度卷积神经网络的训练输入为候选区域坐标与图像样本,通过深度卷积神经网络中的卷积层和池化层计算提取出卷积特征;然后接入改进的 Fast RCNN 网络,通过 ROI 池化层规格化特征;最后将特征输入到全连接层,通过计算将特征进行分类并回归出检测框坐标值。经过多次迭代训练,最后得到具有训练好的

权重参数的车辆检测模型。该方法在 KITTI 车辆公共数据集上进行了测试,并与基于 Fast R-CNN 网络直接检测的结果[9]进行对比实验。而且,与被广泛应用的提取 Harr-Like 特征并且使用 Adaptive Boosting 分类器的算法[5]相比,该方法在 TSD-MAX 交通场景数据库测试中的性能有所提高。结果表明,该方法提高了车辆检测的准确性和鲁棒性。

3.1 前方车辆检测的整体流程

本文方法分为两个阶段:训练阶段和测试阶段。训练阶段中,对预训练后具有初始参数的网络模型进行再次迭代训练,得到前方车辆检测模型;测试阶段中,将测试样本输入车辆检测模型得到检测结果。如图 1 所示,整个计算流程包含如下 6 个步骤:1)获取训练样本;2)构造训练数据,使用 selective search 方法抽取若干个候选区域,可将前方车辆检测者看作训练一个二分类(车辆类与背景类)的图像分类网络;3)利用改进的基于区域的卷积神经网络训练得到每个区域的深度特征;4)使用 ImageNet 图像分类网络预训练卷积神经网络模型。其参数权重作为本文训练的初始参数权重;5)将样本图像和训练集中手工标注的标定值以及 Selective Search 方法产生的候选区域作为输入,通过训练的模型进行多次训练,从数次结果中选择最优的得到最终的车辆目标检测模型;6)利用车辆测试样本对车辆目标检测模型进行测试,得到检测结果。

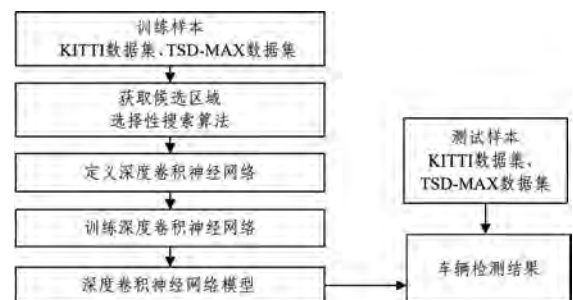


图 1 前方车辆检测的技术框架流程图

3.2 训练数据构造

对 KITTI 和 TSD-MAX 数据集的训练样本进行标定,标定时记录车辆左上和右下的坐标值,构成矩形框的坐标值 $P = \{x_1, y_1, x_2, y_2\}$ 。

3.3 Selective Search 方法

假设某一帧图像上有 n 个预分割的区域,表示为 $R = \{R_1, R_2, \dots, R_n\}$,计算每个区域与它相邻区域(注意是相邻的区域)的相似度,得到一个 $n \times n$ 的相似度矩阵(同一个区域之间和一个区域与不相邻区域之间的相似度可设为 S),从矩阵中找出最大相似度值对应的两个区域,将这两个区域合二为一,此时图像中还剩下 $n-1$ 个区域;重复上述过程(只需要计算新的区域与它相邻区域的新相似度,其他的不用重复计算),重复一次,区域的总数目就少 1,直到最后所有的区域都合并成为同一个区域,即此过程进行了 $n-1$ 次,区域总数目最后变成了 1。

3.4 利用预训练模型初始化参数权重

此次实验使用深度学习 Caffe 框架预训练模型参数权重来自于 ImageNet 数据集下的训练迭代 40000 次的结果。

3.5 改进的 Fast R-CNN 的结构和训练过程

本文首先对图像进行预处理,获得感兴趣区域后,采用训练 ImageNet 数据集时得到的网络模型初始化本文网络模型

中 RoI 层之前的所有层,对训练样本进行迭代训练,得到车辆检测模型。其中边界框 Bounding Box 回归用于检测车辆目标的修正后的位置坐标信息,Softmax 回归用于计算两类目标的(车辆类和背景类)概率估计值。本文方法的网络结构如图 2 所示。

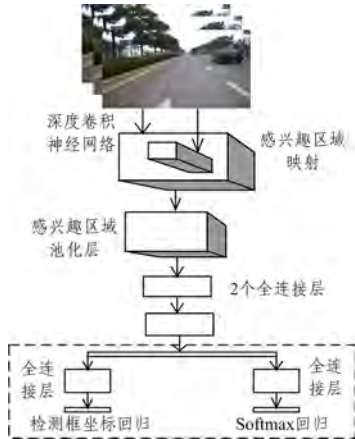


图 2 卷积神经网络的结构图

3.6 ROI 池化层

Selective Search 方法提取到的候选区域的大小都是不同的,但是 RoI 池化层可以针对输入的每一个 RoI 区域提取出固定维度的特征图,以保证后续对每个区域的训练能够正常进行。

本文通过计算每个样本图像的标定框与候选区域的覆盖率,可以得到每个样本图像对应的一组感兴趣区域(RoIs)。其中每帧图片的 R 个 RoIs 的构成方式如表 1 所列。感兴趣区域是一个矩形窗口,每个感兴趣区域由指定其左上角坐标 $(u_1; v_1)$ 及其高度和宽度 $(h; w)$ 的四元组 $(u_1; v_1; h; w)$ 定义。RoI 池化层使用最大池化将任何有效的感兴趣区域内的特征转换成具有 $H \times W$ 的固定空间范围的小特征地图,其中 H 和 W 是层超参数,独立于任何特定的 RoI 层。

表 1 R 个感兴趣区域的构成方式

类别	比例/%	方式
车辆类	25	与某个真值重叠在 $[0.5, 1]$ 的候选框
背景类	75	与真值重叠的最大值在 $[0.1, 0.5)$ 的候选框

在训练过程中,定义一个随机梯度下降(SGD)小批量块(mini batches),它由 N 个图像样本和 R 个 RoIs 构成。本文使用 $R=128, N=2$,即每组批量块有 2 个图像同时每个图像有 64 个 RoIs。在反向传播过程中经过 RoI 池化层时,进行反向传播的计算,如式(1)所示:

$$\frac{\partial L}{\partial x_i} = \sum \sum [i=i^*(r, j)] \frac{\partial L}{\partial y_{rj}} \quad (1)$$

其中, $[\cdot]$ 表示当残差在 ROI 层反向传播时,需判断该残差所在节点 i 是否连接到该 ROI 输入值的最大值。如果是则累计残差,中括号值为 1,否则为 0。

3.7 输出层 Bounding Box 回归

在网络中有两个输出层:1)针对每个 ROI 区域的分类概率预测, $P=(p_0, p_1)$;2)针对每个 ROI 区域坐标的偏移优化, $t^k=(t_u^k, t_v^k, t_h^k, t_w^k)$ ($0 \leq k \leq 1$) 是多类检测的类别序号。

假设对于类别 k^* ,在图片中标注了一个 groundtruth 坐标 $t^*=(t_u^*, t_v^*, t_h^*, t_w^*)$,而预测值 $t=(t_u, t_v, t_h, t_w)$ 。二者理论上越接近越好,这里定义损失函数为:

$$L_{Loc}(t, t^*) = \sum_{i \in \{u, v, h, w\}} smooth_{L_i}(t_i - t_i^*) \quad (2)$$

$$smooth_{L_i}(X) = \begin{cases} 0.5X^2, & |X| \leq 1 \\ |X| - 0.5, & otherwise \end{cases}$$

3.8 车辆检测过程

本文改进的 Fast R-CNN 网络通过 Softmax 分类和 Bounding Box 回归联合训练的方式更新所有参数。用于 Softmax 分类任务和 Bounding Box 回归的 fc 层参数用标准差介于 0.001~0.01 之间的高斯分布初始化。

改进的 Fast R-CNN 网络被微调完毕,检测相当于运行前向传递(假设候选框是预先计算的)。网络将图像和待计算概率的 R 个候选框的列表作为输入。在测试时, R 通常在 2000 左右。

4 实验结果和分析

本文实验采用的硬件平台是 DELL precision 深度学习工作站,操作系统为 Ubuntu16.04 平台,编程环境为 Python,内存为 32GB, GPU 为 NVIDIA GeForce 1080Ti(11G)。任务测试数据来源于西安交通大学人工智能与机器人研究所提供的 2016 TSD-MAX 交通场景数据库。实验中,将数据集按 8:2 的比例随机生成训练验证集和测试集;在训练验证集中,将其 80% 作为训练数据集,剩余的 20% 作为验证数据集。将已分配的数据集用于改进的 Fast R-CNN 网络模型的训练验证以及测试。实验在 DELL 深度学习工作站上完成,整个实验均基于深度学习框架 Caffe 进行,通过数次实验择优原则获取最终结果。本次前方车辆检测包括:轻型客车、轿车(乘用车)、越野车、轻型载货汽车、大型客车、重型载货汽车等常见车型,如图 3 所示。



图 3 来自 TSD-MAX 的部分车辆车型

样本图像来自西安交通大学的 TSD-MAX 交通场景数据库,图 4 给出了几种常见类型。



图 4 TSD-MAX 的车辆图像样本

TSD-MAX 交通场景数据库的训练集和测试集的数量如表 2 所列。

表2 样本数量

数据集	正样本	负样本
训练集	20000	48000
测试集	2000	5000

评价标准如式(3)、式(4)所示:

$$DR = \frac{TP}{TP + FP + FN} \quad (3)$$

$$FPR = \frac{FP}{TP + FP} \quad (4)$$

其中, TP 为所有前方车辆的正检总数, FP 为所有前方车辆的误检总数, FN 为所有视频帧中前方车辆的漏检总数。误检总数 FP 为所有检测到的前方车辆数目与正检车辆总数 TP 之差; 漏检总数 FN 为所有标注车辆数目与正检车辆总数 TP 之差。实验对比结果如表3和表4所列。

表3 KITTI测试集的检测结果(数据结果来自算法对应文献)

方法	时间/s	车辆检测率 DR/%
LSVM-MDPM-SVM ^[14]	10	44.18
Faster-RCNN ^[10]	2	71.12
SDP+RPN ^[15]	0.4	78.38
MS-CNN ^[12]	0.4	76.71
本文方法	2	87.90

表4 TSD-MAX交通场景数据库的检测结果

(来自2016年的中国智能车未来挑战赛)

(单位: %)

方法	检测率 DR	误检率 FPR
文献[5]方法	86.41	11.20
文献[17]方法	84.70	40.83
本文方法	84.10	1.60

通过表4可以发现, 文献[5]是先提取 Harr-Like 特征, 然后使用 Adaptive Boosting 分类器进行前方车辆检测; 文献[17]通过帧差法和改进的 Adaboost 算法进行车辆检测。虽然这两种方法的检测率略高于本文方法, 但是误检率远高于本文方法。本文方法解决图像中车辆目标误检的效果较好, 具有理论意义。

本文方法的实验结果如图5所示。



图5 本文方法在 TSD-MAX 交通场景数据库上的测试结果

结束语 本文通过对前方车辆图片提取感兴趣区域后, 利用改进的 Fast R-CNN 模型进行训练, 最终得到车辆目标检测模型。该方法很大程度上解决了传统的利用机器学习的方法在车辆目标被大面积遮挡时效果差和误检率高的问题。经过实验, 利用海量数据训练的 Fast R-CNN 深度学习方式可以提高算法准确率并且该检测模型能够实现较好的检测效果。然而, 这需要大量的有效样本进行提取特征, 前期样本候选区域的提取过程和标注工作较为费时。在保证检测效果的前提下, 解决上述问题是未来需要研究的重点问题。

参考文献

- [1] 林晓翠. 基于深度学习的车辆检测研究[D]. 厦门: 厦门大学, 2016.
- [2] TANG Y, ZHANG C Z, GU R S, et al. Vehicle detection and recognition for intelligent traffic surveillance system[J]. *Multimedia Tools & Applications*, 2017, 76(4): 5817-5832.
- [3] 郑胤, 陈权崎, 章毓晋. 深度学习及其在目标和行为识别中的新进展[J]. *中国图象图形学报*, 2014, 19(2): 175-184.
- [4] UIJLINGS J R R, SANDE K E A, GEVERS T, et al. Selective Search for Object Recognition[J]. *International Journal of Computer Vision*, 2013, 104(2): 154-171.
- [5] ZHANG F F, BAO H, WANG D. A Revised Multi-lane Vehicle Detection Model with Adaptive Boosting Algorithm[J]. *Journal of Residuals Science & Technology*, 2016, 13(8): 1-6.
- [6] QIU Q J, LIU Y, CAI D W. Vehicle detection based on LBP features of the Haar-like Characteristics[C]// *Intelligent Control and Automation*. IEEE, 2015: 1225-1227.
- [7] 张全发, 蒲宝明, 李天然, 等. 基于 HoG 特征和机器学习的工程车辆检测[J]. *计算机系统应用*, 2013, 22(7): 104-107.
- [8] ZHANG Y Z, SUN P F, LI J F, et al. Real-time vehicle detection in highway based on improved Adaboost and image segmentation[C]// *IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems*. IEEE, 2015: 2006-2011.
- [9] 曹诗雨, 刘跃虎, 李辛昭. 基于 Fast R-CNN 的车辆目标检测[J]. *中国图象图形学报*, 2017, 22(5): 671-677.
- [10] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2017, 39(6): 1137.
- [11] KONOPLICH G V, PUTIN E, FILCHENKOV A A. Application of deep learning to the problem of vehicle detection in UAV images[C]// *IEEE International Conference on Soft Computing and Measurements*. 2016: 4-6.
- [12] CAI Z W, FAN Q F, FERIS R S, et al. A Unified Multi-scale Deep Convolutional Neural Network for Fast Object Detection [M]// *Computer Vision-ECCV 2016*. Springer International Publishing, 2016.
- [13] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]// *CVPR*. 2014.
- [14] GEIGER A, WOJEK C, SAARBRUCKEN M, et al. Joint 3d estimation of objects and scene layout[C]// *NIPS*, 2011. 2011: 1467-1475.
- [15] FAN Y, CHOI W G, LIN Y Q. Exploit all the layers: Fast and accurate cnn object detector with scale dependent pooling and cascaded rejection classifiers[C]// *Computer Vision and Pattern Recognition*. IEEE, 2016: 2129-2137.
- [16] DOLL'AR P, WOJEK C, SCHIELE B, et al. Pedestrian Detection: An Evaluation of the State of the Art[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2012, 34(4): 743-761.
- [17] 刘洋, 王海晖, 向云露, 等. 基于改进的 Adaboost 算法和帧差法的车辆检测方法[J]. *华中科技大学学报(自然科学版)*, 2013, 41(s1): 379-382.