

doi:10.11835/j.issn.1000-582X.2017.07.005

Faster-RCNN 的车型识别分析

桑 军^{a,b}, 郭 沛^b, 项志立^b, 罗红玲^b, 陈 欣^{a,b}

(重庆大学 a.信息物理社会可信服务计算教育部重点实验室; b.软件学院, 重庆 400044)

摘 要: 车型识别是目标检测领域在智能交通的重要应用,也是近年来国内外学者的研究热点之一。针对已有车辆检测方法缺乏识别车型能力的问题,提出了基于 Faster-RCNN 目标检测模型与 ZF、VGG-16 以及 ResNet-101 3 种卷积神经网络分别结合的策略,实验对比了该策略中的 3 种结合模型方案在 BIT-Vehicle 和 CompCars2 种大型车型数据库的车型识别能力。在 BIT-Vehicle 数据集上,基于 Faster-RCNN 与 ResNet-101 结合模型方案的车型识别率高与其余 2 种结合模型方案,其车型识别率高达 91.3%;在迁移测试 CompCars 数据集上,3 种结合模型方案均展现了很好的泛化能力。

关键词: 车型识别; 目标检测; Faster-RCNN; 卷积神经网络

中图分类号: TP391

文献标志码: A

文章编号: 1000-582X(2017)07-032-05

Vehicle detection based on faster-RCNN

SANG Jun^{a,b}, GUO Pei^b, XIANG Zhili^b, LUO Hongling^b, CHEN Xin^{a,b}

(a. Key Laboratory of Dependable Service Computing in Cyber Physical Society, Ministry of Education;

b. School of Software Engineering, Chongqing University, Chongqing 400044, P.R.China)

Abstract: As one of the object detection, vehicle detection, which has been a hot research area in recent years, is one of the important application in intelligent transportation system. To figure out the problem that vehicle detection is lack of the ability of vehicle category recognition, we adopted the strategy of integrating the Faster-RCNN (region-based convolutional neural networks) model with 3 different convolutional neural networks (ZF, VGG-16 and ResNet-101) respectively. By comparing the vehicle category recognition results of the 3 integrating strategies on BIT-Vehicle database and CompCars database, the strategy integrating the Faster-RCNN model with ResNet-101 shows the best result among the 3 models and recognition accuracy reaches 91.3% on BIT-Vehicle database. On the migration test CompCars database, 3 strategy models show good generalization ability.

Keywords: vehicle detection; object detection; faster-RCNN; convolutional neural network

随着大众生活水平的提高,汽车的数量逐渐增多,道路交通压力也日益增大。近年来,部分车辆故意使用“套牌”或者虚假车牌等方法来逃避道路交通的监控,给交通管理部门带了诸多麻烦。出现这类现象的主要原因是:目前大多数交通监控缺乏详细的车型信息。即便只套换车牌号的最后一位数,仅从车牌号可能性

收稿日期: 2017-02-10

基金项目: 高等学校博士学科点专项科研基金博导类资助项目(20130191110027)。

Supported by the Research Fund for the Doctoral Program of Higher Education of China under Grant (20130191110027).

作者简介: 桑军(1968-),男,重庆大学教授,博士生导师,主要从事机器学习、图像处理、软件工程等方向研究,
(E-mail) yunzhou@shmtu.edu.cn。

组合来排查,也是极其耗时耗力的工作。因此,为了从根本上解决这类问题,人们开始利用各种新技术来解决这类问题。智能交通系统就在这种背景下应运而生。车型识别是目标检测领域在智能交通的重要应用,也是积极响应“互联网+”模式的一种方式。

目前,应用于车辆检测和识别的技术有很多。在文献[1]中,Lim T R.等使用 Gabor 滤波和支持向量机的方法完成车辆检测;Cui B 等在文献[2]中提出了利用变形模板来研究车辆检测。近年来,也有基于深度学习的方法识别车型。在文献[3]中,提出了一种基于无监督学习的卷积神经网络的方法进行车型分类;在文献[4]中,Heikki Huttunen 等用基于神经网络的方法实现车型识别。

卷积神经网络^[5](convolutional neural networks, CNNs)已经被广泛应用于各个领域,比如目标检测^[6]、语音识别^[7]等。在目标检测领域中,基于区域的卷积神经网络(region-based CNN, R-CNN)的发展非常迅速。其中,Faster-RCNN^[8]模型在2015年COCO检测大赛一举夺魁,并在PASCAL VOC 2007和PASCAL VOC 2012上实现了目前最高的目标检测准确率。

目前,Faster-RCNN模型应用于车型识别研究的并不多。在文献[9]中,Shoaib Azam等使用Faster-RCNN从4个不同方位(前、后、左和右)的进行车辆方向识别(vehicle pose detection);Tianyu Tang等在文献[10]中,通过航拍所获得车辆数据集进行基于Faster-RCNN的车辆检测研究。在实际生活中,不论是道路监控、高速路收费或停车场,车辆图像的信息来源都是单一摄像头所获取的。因此,车辆图像信息也只会来源某一个特定的方向,多方向车辆图片信息的处理通常不适用于实际生活中。

研究将Faster-RCNN应用于道路监控的车辆,仅从一个方向获取更多的车辆信息,实现监控车辆的车型识别。旨在从实际出发,研究解决现有道路交通监控所存在的问题。

1 Faster-RCNN

继R-CNN^[11]和Fast-RCNN^[12]之后,为了进一步减少检测网络的运行时间,微软Shaoqing Ren等提出了最新的目标检测方法Faster-RCNN^[8]。他们设计一种区域建议网络(region proposal network, RPN)来生成建议区域(region proposals)。RPN的出现替代了之前Selective Search^[13]和EdgeBoxes^[14]等方法,它和检测网络共享全图的卷积特征,使得区域建议检测几乎不花时间。

RPN是一个基于全卷积的网络^[15],它能同时预测输入图片的每个位置目标区域框和目标得分(属于真实目标的概率值)。RPN是通过end-to-end的方式进行网络训练,旨在生成高质量的区域建议框,用于Fast-RCNN的分类检测。通过一种简单的交替运行优化方法,RPN和Fast-RCNN可以在训练时共享卷积特征。由此可见,Faster-RCNN的整体结构可认为是“RPN + Fast-RCNN”的集成。二者结合,分工明确。RPN网络主要作用于生成高质量建议区域框,Fast-RCNN则是起着学习高质量建议区域特征以及分类的作用。

Faster-RCNN在生成建议区域的改进,使得检测效率提升,提供了Faster-RCNN应用于监控车辆车型识别的可行性。选取ZF-Net^[16]、VGG-16^[17]以及ResNet-101^[18]3种不同的卷积神经网络与Faster-RCNN相结合,通过实验对比研究Faster-RCNN在车型识别领域的应用。

2 数据集与预处理

目前,Faster-RCNN还未应用于识别道路监控车辆的车型,其主要原因是道路监控数据较难收集且耗时。选择了2种数据集:BIT-Vehicle数据集^[19]和CompCars数据集^[20]。BIT-Vehicle数据集是由北京理工大学所收集,其车辆图片来源于道路监控。此数据集包含9 580张车辆图片,共6类车型:客车、小型客车、小型货车、轿车、城市越野以及卡车(bus, microbus, minivan, sedan, SUV, truck)。各类车型图片的数量是558,883,476,5922,1392和822。图片的尺寸分为2种:1 600 * 1 200和1 600 * 1 080,它们的取样在不同时间地点(包含白天与夜晚)的2个摄像头所获取。CompCars数据集则由香港中文大学所提供,此数据集包含2大类车辆图片:一类是从互联网中所收集的商业汽车模型图片,包含1 687种车型类别;另一大类是通过监控摄像头所获取的车辆图片,拥有40,000多张车辆图片(包含白天与黑夜)。选取了BIT-Vehicle数据集作为训练验证集以及测试集;将CompCars数据集的第二大类作为迁移测试集,以便研究车型识别的泛化能力。

笔者将BIT-Vehicle数据集以及CompCars数据集做了相应的预处理。所选取的CompCars数据集只含2个类别:轿车以及SUV。此外,由于CompCars数据集并没有标签且数量较大,从中随机抽选了1 000

张图片作为迁移测试集,人工进行标注以便于迁移测试。BIT-Vehicle 是数据集和 CompCars 数据集部分数据分别如图 1、图 2 所示。



图 1 BIT-Vehicle 车型数据集部分数据图

Fig.1 The example of images of BIT-Vehicle dataset



图 2 CompCars 车型数据集部分数据图

Fig.2 The example images of CompCars dataset

3 实验结果与分析

通常目标检测会将 ImageNet 分类^[21]预训练的模型用于训练网络的初始化,这也是目标检测方法的标准惯例^[11]。将 ImageNet 分类预训练所得到的 3 种卷积神经网络模型用于 Faster-RCNN 中共享卷积层的初始化。大多数深度学习方法对于数据集的处理是将数据集以一定比例分配生成训练集、验证集和测试集。

在实验中,将 BIT-Vehicle 数据集按 8:2 的比例随机生成了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 BIT-Vehicle 数据集用于 Faster-RCNN 在 3 种不同卷积。

神经网络的训练验证以及测试。实验在 GPU Tesla K80 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结果。在 BIT-Vehicle 数据集上的车型识别结果如表 1 所示。

表 1 识别 BIT-Vehicle 车型数据集的实验数据表

Table 1 Detection results on BIT-Vehicle database

方法	Bus	Microbus	Minivan	Sedan	Suv	Truck	mAP
Dong et al. ^[16]	0.98	0.84	0.83	0.84	0.91	0.90	0.8811
Faster-RCNN + ZF	0.909 1	0.904 9	0.898 2	0.908 5	0.903 2	0.901 5	0.904 2
Faster-RCNN + VGG	0.909 1	0.906 5	0.901 9	0.906 7	0.898 5	0.903 3	0.904 3
Faster-RCNN + ResNet	0.906 2	0.944 2	0.906 7	0.906 3	0.912 5	0.900 7	0.912 8

在实验过程中,发现 BIT-Vehicle 数据集存在一部分车辆图片没有标注完整。为了保证与 Dong et al.研究成果的可对比性,并没有完善未完成标注的车辆图片,一定程度上影响了方法在车型识别精确度上的提高。在表 1 中,Dong et al.中出现了 BUS 类单精确度高其余精确度低的情况,也因此将其平均精确度大幅提高,其原因可能是该类参与训练的样本多而验证样本少而造成的。尽管如此,研究中的 3 种方法的平均精度值(mean average precision, MAP)在 Dong et al.的基础上均有提高,且并未出现单类高精度的特殊情况,各类精确度相对平均,最高识别率达到 91.28%,提高了 3.2% 左右。主要评估检测结果的 mAP,因为 mAP 是对目标检测的实际度量标准。

此外,ZF、VGG-16 以及 ResNet-101 在 BIT-Vehicle 数据集的实验结果相差不多。为了进一步确定 3 种网络的车型识别能力,将 CompCars 数据集的部分车辆图片作为迁移测试的数据集。由于 CompCars 数据集的数量大且没有标注,故只随机抽取了 1 000 张进行人工标注用于测试(Random_Comp 测试集)。在标注图片的过程中,随机抽取的图片中有一部分车型辨识度很低,如图 2 所示。由于此数据集夜间取样时光过暗,以至于部分图片人也无法识别具体车型,已不具备研究车型识别的意义。为了保证迁移测试的随机性,仍然将随机抽取的所有车辆图片都进行了标注,并将识别度相对较高的图片作为对比数据集(Comp 测试集)。按对车型的分类,在 CompCars 数据集中只包含 2 类:SUV 和轿车,故只针对这 2 类进行车型识别的迁移测试。

在表 2 中,3 种卷积神经网络在辨识度较高的测试集上检测精度明显上升。其中,ResNet-101 在 2 个迁移测试集下,车型识别率均明显高于 ZF 和 VGG-16 的迁移测试结果,证明 ResNet-101 在车型识别能力优于

ZF 和 VGG-16。从整体上观察,迁移结果并不太好,其主要原因可能是由于 CompCars 数据集中所获取的道路监控车辆图片,只截取了车的正面部分,仅有“车脸”信息,缺乏车型信息,并且还存在着大部分夜间不清晰的车辆图像,导致识别效果欠佳。

表 2 识别 CompCars 车型数据集的实验结果表

Table 2 Detection results on BIT-Vehicle database

方法	Random_Comp	Comp
Faster-RCNN + ZF	0.681 55	0.790 3
Faster-RCNN + VGG-16	0.724 15	0.825 8
Faster-RCNN + ResNet-101	0.769 05	0.851 1

参与 Faster-RCNN 训练的 BIT-Vehicle 数据集不仅图片辨识度高,而且包含了“车脸”以及车型信息。通过这样信息量大且辨识度高数据集进行训练之后的模型,依然能够识别辨识度相对低且信息量小的车辆数据集,且最高识别率达到 85.11%,证明 Faster-RCNN 在车型识别上具有很好的泛化能力。

实验中,Faster-RCNN+ResNet101 模型的部分检测结果图 3 所示。在图 3 中,上方 6 张子图片来自 BIT-Vehicle 数据集,下方 3 张子图片来自 CompCars 数据集;在子图中,不同颜色的方框代表模型所检测出的不同车型类别;方框左上角代表模型对于方框区域目标的分类结果:车型类别和类别得分(属于此车型类别的概率值)。

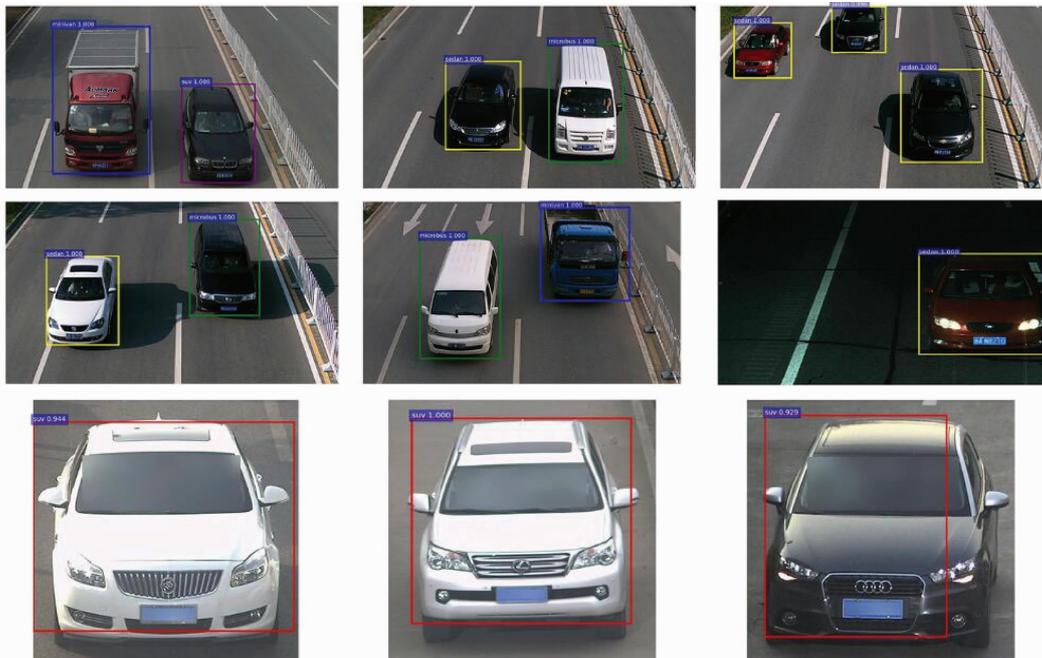


图 3 Faster-RCNN+ResNet101 模型的部分检测结果图

Fig.3 Parts of Faster-RCNN with ResNet101

4 结 论

研究用 Faster-RCNN 模型与不同卷积神经网络相结合的方案研究与实现车型识别。在 BIT-Vehicle 数据集中,基于 Faster-RCNN 的方法的识别率明显优于 Dong^[19]的识别结果。不仅如此,为了进一步确认 3 种卷积神经网络在车型识别的能力,在 CompCars 数据集上做了迁移测试。通过实验结果验证,ResNet-101 相对于 ZF 和 VGG-16 效果更佳。实验表明,基于 Faster-RCNN 的方法能用于道路监控车辆的车型识别,且识别准确率也达到了 91.3%。由于道路监控车辆车型识别的场景过于特殊,其训练数据集收集困难,在已有的监控车型数据集上验证了基于 Faster-RCNN 车型识别的可行性。在未来的工作里,会通过相关部门收集更多且具有实际意义的车型数据集,以便进一步研究车型识别,在研究基础上进一步提升车型检测的效率。从实际出发,解决目前车辆检测获取较少车型信息的问题。

参考文献:

- [1] Lim T R, Guntoro A T. Car recognition using Gabor filter feature extraction[C]// Proceedings of Asia-Pacific Conference on Circuits and Systems, APCCAS 2002, Denpasar, Bali, Indonesia, October 28-31, 2002. Bali: IEEE, 2002(2):451-455.
- [2] Cui B, Xue T, Yang K. Vehicle recognition based on support vector machine[C]// Proceedings of 2nd 2008 International Symposium on Intelligent Information Technology Application Workshop, IITA 2008, Shanghai, China, January 21, 2008-December 22, 2008. Shanghai: IEEE, 2008:443-446.
- [3] Dong Z, Pei M, He Y, et al. Vehicle type classification using unsupervised convolutional neural network[C]// 22nd International Conference on Pattern Recognition, ICPR 2014, Stockholm, Sweden, August, 24-28, 2014. Sweden: IEEE, 2014:172-177.
- [4] Huttunen H, Yancheshmeh F S, Chen K. Car type recognition with deep neural networks[C]// 2016 IEEE Intelligent Vehicles Symposium, IV 2016 IEEE, Gotenburg, Sweden, June 19-22. Sweden: IEEE, 2016:1115-1120.
- [5] LeCun Y, Boser B E, Denker J S, et al. Handwritten digit recognition with a back-propagation network[C]// Advances in neural information processing systems, Denver, Colorado, USA, November 27-30, 1989. USA: NIPS Conference, 1989: 396-404.
- [6] He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[C]// 13th European Conference on Computer Vision, ECCV 2014, Zurich, Switzerland, September 6-12, 2014. Springer: Cham, 2014:346-361.
- [7] Hinton G, Deng L, Yu D, et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups[J]. IEEE Signal Processing Magazine, 2012, 29(6): 82-97.
- [8] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[C]// 29th Annual Conference on Neural Information Processing Systems, NIPS 2015, Montreal, QC, Canada, December 7-12, 2015. Canada: NIPS Conference, 2015:91-99.
- [9] Azam S, Rafique A, Jeon M. Vehicle pose detection using region based convolutional neural network [C] // 5th International Conference on Control, Automation and Information Sciences, ICCAIS 2016, Ansan, Korea, Republic of, October, 27-29, 2016. Korea: IEEE, 2016:194-198.
- [10] Tang T, Zhou S, Deng Z, et al. Vehicle Detection in Aerial Images Based on Region Convolutional Neural Networks and Hard Negative Example Mining[J]. Sensors, 2017, 17(2): 336.
- [11] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]// Proceedings of the IEEE conference on computer vision and pattern recognition, Columbus, OH, United states, June, 23-28, 2014. USA: IEEE, 2014:580-587.
- [12] Girshick R. Fast r-cnn[C]// Proceedings of the IEEE international conference on computer vision, Santiago, Chile, December 11-18, 2015. USA: IEEE, 2015:1440-1448.
- [13] Uijlings J R R, Van De Sande K E A, Gevers T, et al. Selective search for object recognition[J]. International journal of computer vision, 2013, 104(2): 154-171.
- [14] Zitnick C L, Dollár P. Edge boxes: Locating object proposals from edges[C]// 13th European Conference on Computer Vision, ECCV 2014, Zurich, Switzerland, September, 6-12, 2014. Zurich: Springer, Cham, 2014:391-405.
- [15] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, United states, June 7-12, 2015. USA: IEEE, 2015: 3431-3440.
- [16] Zeiler M D, Fergus R. Visualizing and understanding convolutional networks[C]// European conference on computer vision, Zurich, Switzerland, September, 6-12, 2014. Zurich: Springer, Cham, 2014:818-833.
- [17] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. ArXiv preprint arXiv, 2014:1409.
- [18] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]// Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, NV, United states, June 26 - July 1, 2016. USA: IEEE, 2016: 770-778.
- [19] Dong Z, Wu Y, Pei M, et al. Vehicle type classification using a semisupervised convolutional neural network[J]. IEEE transactions on intelligent transportation systems, 2015, 16(4): 2247-2256.
- [20] Yang L, Luo P, Change Loy C, et al. A large-scale car dataset for fine-grained categorization and verification[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, United states, June 7-12, 2015. USA: IEEE, 2015: 3973-3981.
- [21] 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.