基于卷积神经网络的小样本车辆检测与识别

吴玉枝[1],吴志红[2],熊运余[2]
(1. 四川大学 计算机学院,四川 成都 610064;
2. 四川大学 图形图像研究所,四川 成都 610064)

摘 要:设计了一种快速准确的算法,实现了环境复杂、样本缺少情况下实时车辆检测和车型识别,特别是对三轮车的识别。利用一种改进的卷积神经网络(convolutional neural network,CNN)快速学习车辆特征,采用微调、分段训练以及多层特征图结合的策略增强网络特征学习能力,在小样本下尽可能全面地学习目标特征。摒弃繁琐耗时的区域推荐算法和后分类算法,利用单个网络直接预测图片中目标车辆的位置和车型类别,大幅提高了算法性能。实验结果表明,采用 GeForce GTX 1080 GPU 时,该算法对各类车型识别准确度较为平衡,平均检测准确率高达 72.2%,每秒检测帧数 46.57,在雨天、晴天、夜晚、强光和树荫等各种复杂场景下均有较好的适应性,适用于真实视频监控下智能交通系统精确实时的要求。

关键词:卷积神经网络;车辆检测;车型识别;多特征结合;分段训练
中图分类号:TP391.4 **文献标识码**:A **文章编号**:1673-629X(2018)06-0001-06
doi:10.3969/j.issn.1673-629X.2018.06.001

Vehicle Detection and Recognition of a Few Samples Based on Convolutional Neural Network

WU Yu-zhi[1],WU Zhi-hong[2],XIONG Yun-yu[2]
(1. School of Computer Science,Sichuan University,Chengdu 610064,China;
2. Institute of Image and Graphics,Sichuan University,Chengdu 610064,China)

Abstract:We design a quick and accurate algorithm to achieve the detection and recognition of vehicles,especially the tricycles,in the complex environments of lacking of samples. Firstly,an improved convolutional neural network is used to learn vehicle features rapidly,then many methods such as fine-tuning neural network,combining predictions from multiple feature maps and phased training are used to enhance network's learning with a few samples. By eliminating the tedious and time-consuming regional recommendation algorithm and the post-classification algorithm,the position and category of the target vehicle in the image are directly predicted by using a single network,which greatly improves the performance of the algorithm. The experiment shows that when using GeForce GTX 1080 GPU,the vehicle recognition accuracy of the proposed algorithm is relatively balanced,with an average detection accuracy by 72.2%,and the number of frames per second is 46.57. It owns better adaptability in all kinds of complicated scenarios such as rainy day,sunny day,night,light and shade and so on,which is suitable for the precisely real-time requirements of intelligent transportation system under the real video monitoring.

Key words:convolutional neural network;vehicle detection;vehicle recognition;multiple feature maps;phased training

1 概 述

随着城市规模的扩大和道路车辆增多,智能交通系统(intelligent transportation system,ITS)逐渐成为图像视觉领域的一个研究热点。车辆检测和车型识别是 ITS 的重要组成部分,两者在规避交通事故、TC 收费系统和交通量调查等方面应用广泛。

目标检测的任务是确定图像中有无感兴趣的目标,并给出目标的具体坐标。车辆检测是目标检测的一个分支,可分为基于视频的车辆检测和基于图像的车辆检测。基于视频的检测是在图像检测的基础上,利用视频流的帧连续性实现车辆检测,主要采用帧差算法、边缘检测算法和背景差算法实现对车辆数量、类

收稿日期:2017-07-24 **修回日期**:2017-11-29 **网络出版时间**:2018-02-07
基金项目:国家高技术研究发展计划(2015AA016405)
作者简介:吴玉枝(1993-),女,硕士,研究方向为图像处理;吴志红,博士,副教授,通讯作者,研究方向为数字图像处理、智能系统与信息处理、嵌入式系统、可编程逻辑器件开发与应用。
网络出版地址:http://kns.cnki.net/kcms/detail/61.1450.TP.20180307.1034.002.html

型、车流量、车流密度、平均车速以及交通事故等的检测。其中背景差算法使用最多。常用的背景建模法有均值法、高斯平均法、中值法、卡尔曼滤波模型法[1]及混合高斯模型法(Gaussian mixture model, GMM)[2]。传统的基于图像特征的车辆检测采用视觉特征、统计特征、变换系数特征和代数特征等车辆显著特征,机器学习在各个领域取得成功后,HOG(histogram of oriented gradient)[3]和 LBP(local binary pattern)[4]等纹理特征也被相继应用于车辆检测。近年来,深度卷积神经网络被应用于计算机视觉各个领域,在诸如图像分类[5]、人脸识别[6]、行人检测[7]和车辆分类等多个方面取得了成功,文中将卷积神经网络应用到车辆检测。

车型识别技术是以计算机视觉、数字信号处理、模式识别等技术为基础,通过视频监控或高速拍照等方式取样,以工控机或嵌入式处理器为处理平台,完成对不同车型的分类和识别。其识别的结果信息可以全面应用于道路交通视频监控系统,也可有效配合智能交通信息管理中心完成规划路网、管理通流、高效收费等智能化应用,最终有效改善道路拥堵,提高路网通流效率,优化交通运输环境。车型识别一般有身份识别和身份鉴定两种,身份识别即判断车的类型,身份鉴定则判断车辆对象是否属于某已知库,文中所指的车型识别专指身份识别。传统的基于视频图像识别车型的主要方法有基于模板匹配的识别方法、基于统计模式的识别方法、基于仿生模式的识别方法和基于支持向量机的识别方法[8]。文中用一个深度卷积神经网络先检测出图像中的车辆,再识别检测到的目标车型,借鉴文献[9-10]的方法,将车辆检测和车型识别统一于同一个网络。

SSD[11]是将不同分辨率的特征图与先验包围框结合的网络,达到了计算量小、适应性强、检测精度高的效果。文中借鉴 SSD 中的做法改进 VGG-16,在 VGG-16 网络结构的基础上修改全连接层,增加新的卷积层,结合五层不同的卷积特征图用于预测,并抛弃了繁琐的区域推荐和后分类算法,用单个网络实现了多类型车辆检测。由于样本数据集有限,通过微调和分步训练的策略提高模型收敛度和精确度。

2　车辆检测

与一般的深度神经网络的目标检测方法不同,文中利用单个卷积神经网络进行多类目标检测,将检测与分类统一与一个网络,而不是先检测再分类,同时也抛弃了繁琐的区域推荐算法,大大减少了时间消耗,速度更快。YOLO[12]同样通过减少区域推荐算法达到更快的速度,但最后仅仅使用顶层特征图的预测结果作为检测依据,并且在输入图片时将图片转化为同一尺寸,再投入网络进行训练得到最终模型,同样检测时也需要将待检测图片进行相同的处理,导致对不同尺寸和车辆目标的适应性较差。与 YOLO 不同的是,文中算法结合了顶层和低层多层特征图的预测结果,也并不改变输入图片的尺寸大小,以达到增加模型输出层的平移不变性、减少过拟合、改善检测性能的目的。

2.1　训练方法

假设一张图片中总共有 n 个先验,用 $b_i(i \in [0, n))$ 表示,每一个先验对应一个包围框和若干车型置信度,这些置信度标志着该先验位置的目标是属于某个特定车型的概率。c_p^i 表示第 i 个先验是第 p 种车型的置信度,$l_i \in R_4$ 表示第 i 个先验包围框的坐标,$g_p^j \in R_4$ 表示图片中第 p 类车型的第 j 个真实包围框。值得注意的是,最终的预测包围框是根据网络输出的坐标偏移量来调整先验包围框得到的。另外,先验框和真实框都采用基于整张图的单元坐标系。坐标系归一化使得整张图片都可以在一个坐标单元内取得,这样就不必关心输入图片的大小,可以随意地对比坐标。

2.1.1　匹配策略

需要在训练阶段将真实包围框同先验关联起来,其实现方式有两种:直接使用先验包围框坐标或者在网络预测出坐标偏差后调整先验框坐标再使用。为了描述方便,文中将先验包围框称为源框,真实包围框简称为真实框。在训练过程中需要确定正负样本,能够与真实框匹配的源框是正样本,其余的源框是负样本。有两种备选的匹配方式,其一是双向匹配,其二是全预测匹配(perprediction matching)。双向匹配算法中每一个真实框同与其相似度最高的源框匹配,它保证了每一个真实框都一定有一个源框相匹配。文中采用另一种匹配算法:全预测匹配。全预测匹配算法首先执行双向匹配,保证每个真实框同一个源框匹配,然后在剩下的源框中选择与真实框相似度超过特定阈值(经过实验对比,采用阈值 0.5)的几个源框,这几个源框也同真实框相匹配。全预测匹配可以为每个真实框匹配好几个正样本源框—使用全预测匹配算法能够为多个交叉的先验框预测出高置信度,而不会像双向匹配一样非要选出一个最好的匹配源框。

为了检测识别多类车型(几百甚至上千),需要为每个源框预测出一个针对所有目标车型的包围框偏移量,也就是说在匹配阶段不必考虑源框的车型类别,直接将源框和真实框匹配。

2.1.2　训练目标

为了达到检测并识别多种类型车辆的训练目标,如果第 i 个源框和第 p 类车型的第 j 个真实框相匹配,则 $x_{ij}^p = 1$,否则 $x_{ij}^p = 0$。若使用双向匹配算法,则有

$\sum_i x_{ij}^p = 1$。若使用全预测匹配算法,则 $\sum_i x_{ij}^p \geqslant 1$,即对于第 j 个真实框存在至少一个匹配源框。损失函数为所有框的位置损失值(L_{loc})和置信度损失值(L_{conf})的线性加权和:

$$L(x,c,l,g) = L_{conf}(x,c) + 0.06L_{loc}(x,l,g) \quad (1)$$

其中,0.06 是通过交叉验证选取的合适值,位置的损失函数是 L2 范数(计算预测框和真实框的损失值):

$$L_{loc}(x,l,g) = 0.5\sum_{i,j} x_{ij}^p \| l_i - xg_{ij}^p \| \quad (2)$$

而置信度损失函数采用 logistic 函数:

$$L_{conf}(x,c) = \sum_{i,j,p} \log(c_i^p) - \sum_{i,p}(1 - \sum_{j,p=q} x_{i,j}^q)\log(1 - c_i^p) \quad (3)$$

2.1.3　难例最小化

在匹配步骤之后,大部分的源框都会被标记为负样本,导致训练过程中正负样本之间的数量差异巨大。将负样本中的源框按照所有车型分类的置信度由高到低进行排序后,选取置信度最高的几个负样本而不是所有的负样本使用,以平衡正负样本的比例,使得负样本和正样本的比例总是接近 3∶1。

2.1.4　共享先验包围框

借鉴 SSD 的做法,通过所有车型分类共享包围框的方式,在训练之后直接检测多个分类。假设特征图大小为 $m \times m$,特征图中每个位置有 k 个先验框(用先验框的左上角和右下角的坐标表示这个先验框),总共有 c 种车型。这样,对 k 个先验框就有 $4k$ 个坐标输出,检测 c 个类别就会有 ck 个置信度输出,最后每张特征图位置共有 $(4+c)k$ 个输出。累计所有位置的输出参数,共有 $(4+c)km^2$ 个输出,但其中只有 $(4+c)k$ 个参数需要学习。如果不同类别不采用共享位置的方法,则一共有 $5ckm^2$ 个输出,总共有 $5ck$ 个参数。显然,如果有很多种车型需要检测识别($c \geqslant 1$),以上数字会迅速增大。

2.1.5　分段训练

由于样本数据集数量很小,为了达到更高的准确率,在训练过程中采用微调和分步训练的策略。通过 fine-tuning 技术微调 PASCAL VOC 训练得到 VGG-16 模型,修改 VGG-16 的网络结构,增加更多的卷积层用于多特征图预测。将训练过程分为两个阶段,第一阶段将样本数据集随机分配为训练集和验证集,使得训练集和验证集的比例接近于 3∶1,迭代训练至 loss 稳定。第二阶段将验证集和训练集中随机的等量的数据进行替换,仍然保证训练集和测试集的比例接近 3∶1,迭代至 loss 稳定。通过实验验证该策略可以一定程度地降低 loss,提高模型的准确率。

2.2　网络结构

一些算法将样本图片转化为不同的尺寸,对转化后的图片分别进行处理,之后再把处理结果结合起来,以适应不同尺寸目标的检测;而改进的 VGG-16 网络通过结合单个网络中的低层卷积特征图和顶层特征图达到了同样的效果。众所周知,卷积网络低层相较于高层获取了更多输入目标的细节,有助于改善语义分割的质量;而顶层特征图中池化得到的全局特征有助于平滑语义分割结果。改进的 VGG-16 网络使用低层和高层的特征图而不是仅仅使用顶层特征图,可以获得更为全面的目标特征,其网络的主体结构如图 1 所示。

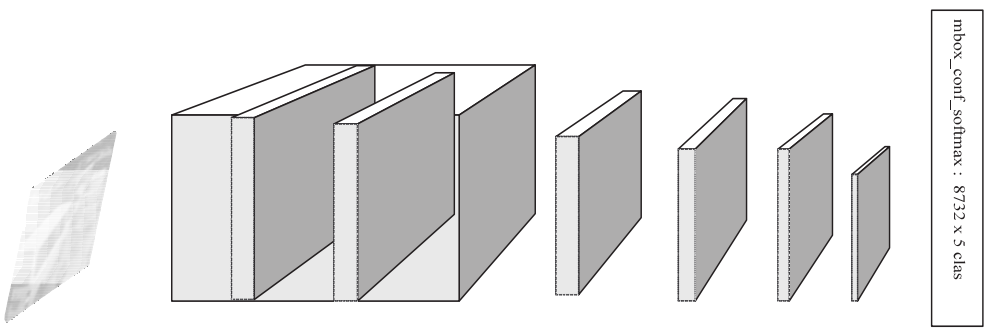


图 1　改进的 VGG-16 网络

2.2.1　单网络检测

借鉴 Faster-RCNN[13] 中 anchor 的概念,摒弃繁琐耗时的区域推荐算法,提高了算法的性能。首先,将图片样本调整到一个固定尺寸,然后将图片划分为 $N \times N$ 个网格,利用卷积特征图提取图像特征,每个格子负责预测中心点落在该格子区域内的目标物体,最终输出多个包围框和每个包围框所属类别的置信度。这样就不需要一个专门的网络或者算法推荐区域,只要训练之前定义一系列不同长宽比和尺寸的先验包围框,然后在训练中由卷积神经网络筛选合适的包围框,并根据网络输出的位置偏移量对包围框的大小和位置进行调整,最终得到预测包围框。网络最终输出车辆目标的包围框及该包围框所属车型的置信度,不需要像 R-CNN[14] 等多种算法一样在卷积神经网络后再增加

一个分类器判断包围框目标所属类别,实现了单个网络下的多分类检测识别。

2.2.2　多层特征图预测

大多数的卷积神经网络都会在深度层通过池化等手段减少特征图的大小,以减少计算量和内存消耗,并增加算法的平移不变性和尺度不变性,但最后只采用顶层特征图用作预测,这造成了一定程度上特征信息的流失。通过改进 VGG-16 网络,在其基础上增加新的卷积层,并结合多层卷积特征图用作预测,可以获得更为全面的特征信息。

一个网络中不同层的特征图通常感受野也不同。如果网络中的卷积先验需要同使用的每一层感受野相关联,计算量将非常大。为了解决以上难题,通过一定的调整使得特征图的某个位置通过学习负责预测图片特定区域和特定大小的检测目标。假设有 m 个特征图用做预测。简单起见,用 $f_k \in [1,m]$ 表示第 k 个特征图(按尺寸非递增排序)的大小。每个特征图的源框尺寸计算如下:

$$s_k = s_{\min} + (k+1)\frac{s_{\max} - s_{\min}}{m-1} \tag{4}$$

其中,$s_{\min} = 0.1$,$s_{\max} = 0.7$,表示最低层的先验尺寸为 0.1,最高层的先验尺寸为 0.7,两者之间所有层的先验尺寸线性递增。

采用不同长宽比的先验框,用 $a_r \in \{1,2,\frac{1}{2},\frac{1}{3}\}$ 表示,然后计算每个先验框的宽($w_{ka} = s_k\sqrt{a_r}$)和高($h_{ka} = s_k/\sqrt{a_r}$)。对于那些长宽比为 1 的先验框,为其添加一个附属先验框,这个附属先验框尺寸为 $s_k' = \sqrt{s_k s_{k+1}}$,这样每个特征图位置都有六个先验框。每个先验框的中心点表示为 $(\frac{i+0.5}{f_k},\frac{j+0.5}{f_k})$,其中 $i,j \in [0,f_k)$,然后再调小先验框的坐标值使其取值范围在 0 到 1 之间。实际上,可以为不同的检测任务设计不同的先验框。

通过结合多张特征的所有位置上不同尺寸和长宽比的先验框预测结果,得到一系列离散的的预测结果,包含了不同大小不同形状的检测目标。

3　实　验

3.1　多车型标签车辆数据集

传统图片数据库中的车辆图片与真实的道路摄像头采集到的车辆图片有很大的区别。前者车辆目标只有小车和公交,车辆目标往往较大较少,多是平视的视角,背景各异,包含车辆的正面、侧面和背面;而后者车辆目标有小车、公交、三轮车和货车四种车型,受高摄像头和密集车流影响,图片中车辆目标更多,多为俯视视角,背景是树荫和道路,除了包含传统的车辆正面、侧面和背面外,还包含车顶图片,使用球机摄像头时目标可能还有轻微变形。使用传统数据集如 PASCAL VOC 训练得到的模型,难以准确检测和识别道路上的三轮车与货车。文中用到的数据集大部分采自于真实道路摄像头,筛选了不同摄像头下 1 100 多张图片,以确保角度和场景的多样性,另外采集了 500 多张网络车辆图片,以增加模型自适应性。在筛选出来的图片中标注货车、三轮车、小车和公交车四个车型的真实包围框,制作样本标签集。

为了使模型更加鲁棒,可以适应不同尺寸和形状的输入目标,随机选择以下的一种方式对每一张训练图片进行取样:

· 使用原始图片;

· 取图片的一部分,这部分图片与目标的最小 jaccard 相似度取值 ∈ {0.1,0.3,0.5,0.7};

· 取图片的一部分,这部分图片与目标的最大 jaccard 相似度为 0.5。

在进行以上取样步骤后,每一个样本以二分之一的概率进行水平翻转,并且随机选择一些图片做图像扭曲变换处理。

3.2　特征提取

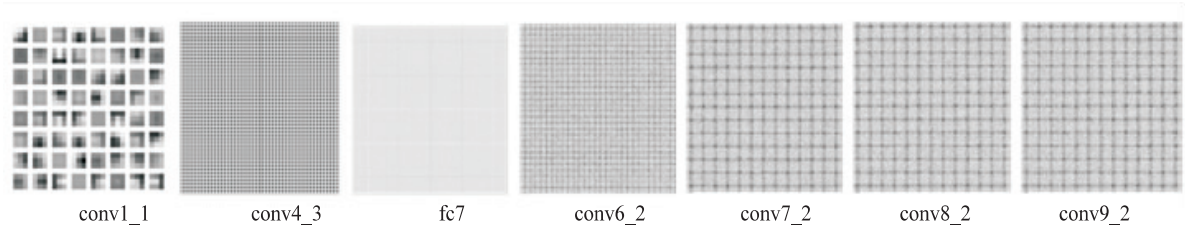
通过 CNN 卷积学习到的特征具有一定的辨别性,针对目标提取的特征识别度较高,背景激活度较低,具有一定程度的位移、尺度、形变不变性。图 2 为特征可视化结果,其中 Conv1_1 是网络中的第一个卷积层,提取的特征较为具象化,图(b)第一张图片能够看到车辆目标的轮廓。算法中用作预测的六层特征图如图 2 中后六张图片所示,可以看到越底层的特征越具象化,提取的多是颜色、轮廓等基本特征;越顶层的特征越抽象化,越具有辨别性。多层特征的结合,可以兼顾到细节和全局两方面的特征信息,使得预测结果更为准确可靠。

根据坐标偏移量调整特征图上所有位置的先验框,输出该特征图的预测框。结合多张特征图的预测输出,可以得到一系列离散的预测结果,包含了不同大小不同形状的检测目标。最后的检测层输出 1 000×6 维向量,即 1 000 个预测框,每个框的信息包括目标左上角和右下角的坐标、何种车型及属于该车型的置信度。预测效果示例如图 3 所示。

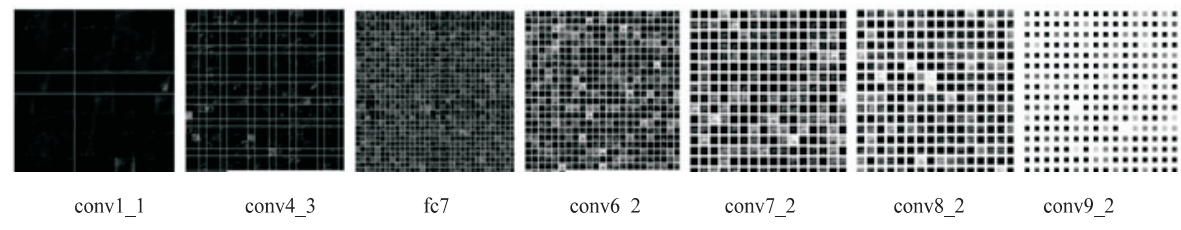
3.3　实验结果及分析

以 PASCAL VOC 数据集下训练好的 VGG-16 模型作为预训练模型,采用 two-phase 策略分成两个阶段进行训练。第一阶段随机分配训练集和验证集,迭代 8 万次至 loss 稳定;第二阶段微调第一阶段生成的模型,按照分段训练策略打乱训练集和验证集,迭代 2

万次,得到最终的训练模型。两阶段中数据集数量分布如表 1 所示。

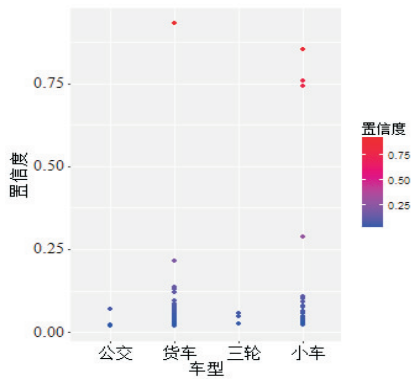


(a)卷积核提取的特征



(b)输入图片的各层特征图

图 2　特征可视化



(a) 四类车辆目标预测图

(b) 输出结果

图 3　网络输出预测效果示例

表 1　四种车型的目标数量分布

	第一阶段		第二阶段	
	训练集	测试集	训练集	测试集
图片数量	1248	416	1 248	416
货车	975	276	918	333
三轮车	396	154	402	148
小车	3 953	1 343	3 926	1 370
公交车	339	116	329	126

表 2　不同模型实验结果

模型	小车	货车	公交车	三轮车	mAP	FPS
改进 VGG-16 分段迭代 10 万次	0.753	0.789	0.781	0.757	0.772	46.57
改进 VGG-16 迭代 10 万次	0.759	0.729	0.730	0.653	0.718	42.14
Faster RCNN 迭代 12 万次	0.833	0.602	0.571	0.446	0.613	19.16
改进 VGG-16 分段迭代 12 万次	0.745	0.768	0.741	0.667	0.718	42.98
改进 VGG-16 迭代 12 万次	0.739	0.730	0.734	0.664	0.718	41.15

训练采用的硬件配置如下:CPU,Intel Core i7-6800k@3.40 GHz×12;GPU,GeForce GTX 1080/PCIe/SSE2。Faster R-CNN 是一种国际前沿的目标检测算法,使用 RPN(region proposal network)代替区域推荐算法,大大提高了检测速度,同时保持了与 Fast R-CNN 持平的检测精度。FasterR-CNN 和改进的 VGG-16 使用相同的样本训练得到的模型检测效果如表 2 所示。

从表 1 可以看出,训练样本中四类车辆目标的数量严重不平衡,基于道路实际情况,小车数目远远多于其他三类车辆数目。而表 2 中 Faster-RCNN 模型识别四种车型的准确度也存倾斜度相同的不平衡,其中小车的 AP(average precision)远高于其他三种车型的 AP,这反映出数据集的匮乏和不平衡影响了 Faster R-CNN 算法的准确度。而使用改进的 VGG-16,除了三轮车的 AP 稍低,其余三种车型 AP 均为 0.7+,比 Fas-

ter R-CNN 有更好的平衡性和适应性。两种网络检测效果如图 4 所示。

图 4　两种网络检测效果

实验对比发现,过多的训练次数容易造成过拟合。从表 2 可以看到,改进的 VGG-16 网络在迭代 10 万次后平均准确率为 0.718,此后继续增加迭代次数至 12 万次,mAP 不再增加。在采用分段训练策略时,迭代 10 万次平均准确率达 0.772,此后继续增加迭代次数反而造成 mAP 的下降。对比迭代 10 万次时改进的 VGG-16 有无采用分段训练策略得到的模型发现,采用分段训练策略可以将 mAP 提高 5%,说明小样本情况下分段训练是一种有效提高检测精度的方法。总体上,与 Faster R-CNN 网络相比,文中采用改进 VGG-16 网络针对四类车型的 AP 较为平衡,mAP(mean average precision)也比前者高约 10%,且速度约为其 2.5 倍,说明文中算法在小样本上有较为理想的检测效果。与此同时,文中算法的网络误检率更低,适应性强,在各种复杂环境诸如夜晚、强光、树荫下等均有稳定的检测和识别效果。

4　结束语

采用改进的 VGG-16 网络,用 fine-tuning 和分段训练的策略实现在小样本下快速准确的车辆目标检测和车型分类。通过实验结果分析与对比发现,该算法的平均准确度可达 77.2%,运行效率上平均每秒处理帧数 46.57 帧,比 Faster-RCNN 快约 1.5 倍,可以应用到道路摄像头实时检测中。但是算法对图片中远处的小目标车辆无法有效识别,漏检率较高;且部分机动三轮车由于背面和侧面和小车极为相似,有时会被误检为小车;同时,因为目前的训练样本数量,算法的最终 mAP 还是不够高。针对这些问题,将继续深入研究,对算法进行改进。

参考文献:

[1] 田鹏辉,隋立春,肖　锋. 用 Kalman 滤波改进的背景建模红外运动目标检测[J]. 四川大学学报:自然科学版,2014,51(2):287-291.

[2] 王宝珠,胡　洋,郭志涛,等. 基于朗斯基函数的混合高斯模型运动目标检测[J]. 计算机应用研究,2016,33(12):3880-3883.

[3] 凌永国,胡维平. 复杂背景下的车辆检测[J]. 计算机工程与设计,2016,37(6):1573-1578.

[4] 董恩增,魏魁祥,于　晓,等. 一种融入 PCA 的 LBP 特征降维车型识别算法[J]. 计算机工程与科学,2017,39(2):359-363.

[5] 朱　威,屈景怡,吴仁彪. 结合批归一化的直通卷积神经网络图像分类算法[J]. 计算机辅助设计与图形学学报,2017,29(9):1650-1657.

[6] TAIGMAN Y, YANG Ming, RANZATO M, et al. DeepFace:closing the gap to human-level performance in face verification[C]//IEEE conference on computer vision and pattern recognition. Columbus,OH,USA:IEEE,2014:1701-1708.

[7] 徐　超,闫胜业. 改进的卷积神经网络行人检测方法[J]. 计算机应用,2017,37(6):1708-1715.

[8] 邓　柳,汪子杰. 基于深度卷积神经网络的车型识别研究[J]. 计算机应用研究,2016,33(3):930-932.

[9] JIANG Changyu,ZHANG Bailing. Weakly-supervised vehicle detection and classification by convolutional neural network[C]//International congress on image and signal processing, biomedical engineering and informatics. Datong, China:IEEE,2016:570-575.

[10] BAUTISTA C M,DY C A,MAÑALAC M I,et al. Convolutional neural network for vehicle detection in low resolution traffic videos[C]//IEEE region 10 symposium. Bali,Indonesia:IEEE,2016:277-281.

[11] LIU Wei,ANGUELOV D,ERHAN D,et al. SSD:single shot multibox detector[C]//European conference on computer vision.[s.l.]:Springer International Publishing,2016:21-37.

[12] REDMON J,DIVVALA S,GIRSHICK R,et al. You only look once:unified, real-time object detection[C]//IEEE conference on computer vision and pattern recognition. Las Vegas,NV,USA:IEEE,2016:779-788.

[13] GIRSHICK R. Fast R-CNN[C]//Proceedings of the IEEE international conference on computer vision.[s.l.]:IEEE,2015:1440-1448.

[14] GIRSHICK R,DONAHUE J,DARRELL T,et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Computer vision and pattern recognition.[s.l.]:IEEE,2014:580-587.