

基于 SSD 的不平衡样本车辆检测与识别

马浩良, 谢林柏

(江南大学 物联网工程学院, 江苏 无锡 214122)

摘要:为了实现在复杂环境, 车辆样本不平衡情况下的实时车辆检测与识别, 基于 SSD 算法搭建了车辆检测与识别的框架。针对车辆数据存在车型难易样本不均衡以及 SSD 方法存在的正负样本不平衡问题, 在 SSD 引入改进的损失函数来挖掘难易样本, 通过提高难样本的学习比例来更好地识别样本较少的车辆类型。引入 SSD 级联的网络结构, 在第一级 SSD 挖掘正负样本, 在第二级 SSD 根据第一级 SSD 的指导过滤掉大量的负样本。构建了拥有 7 480 幅图像, 包含 4 种车辆类型的数据集对该方法进行验证。实验结果表明, 基于改进 SSD 的方法提高了少样本车辆类型的准确率, 使整体检测精度取得了 90.0% 的准确率。针对不均衡样本的车辆数据集有较好的通用性, 适用于车辆检测与识别任务。

关键词: 车辆检测与识别; SSD; 样本不平衡; 难易样本挖掘; 正负样本挖掘

中图分类号: TP301

文献标识码: A

文章编号: 1673-629X(2019)12-0135-06

doi: 10.3969/j.issn.1673-629X.2019.12.024

Vehicle Detection and Recognition in Unbalanced Samples Based on SSD

MA Hao-liang, XIE Lin-bo

(School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, China)

Abstract: In order to realize real-time vehicle detection and recognition in a complex environment with unbalanced vehicle samples, a framework of vehicle detection and recognition is built based on SSD algorithm. Aiming at the imbalance of difficult and easy samples in vehicle data and the imbalance of positive and negative samples in SSD method, an improved loss function to mine difficult and easy samples in SSD is introduced to identify vehicle types with fewer samples better by increasing the learning proportion of difficult samples. The cascade SSD network structure is introduced to mine positive and negative samples in the first-level SSD and filter out a large number of negative samples in the second-level SSD according to the guidance of the first-level SSD. A data set with 7 480 images and 4 vehicle types is constructed to verify this method. The experiment shows that the improved method based on SSD improves the accuracy of vehicle types with fewer samples and the overall detection accuracy achieves 90.0%. This method has excellent generality for vehicle data sets with unbalanced samples and is suitable for vehicle detection and recognition.

Key words: vehicle detection and recognition; SSD; sample imbalance; hard sample mining; positive and negative sample mining

0 引言

随着城市规模的扩大和道路车辆的增多, 交通事故发生的风险逐年增加, 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。车辆检测与识别^[1]作为 ITS^[2]的重要组成部分, 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。

车辆检测与识别是目标检测^[3]的一个分支, 传统的基于视频的车辆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。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、ViBe 算法、卡尔曼滤波模型法^[4]及混合高斯模型法 (Gaussian mixture model, GMM)^[5]。文献[6]提出一种自适应阈值与双特征的 ViBe 运动车辆检测算法抑制鬼影和空洞现象。传统的基于静态图像的车辆检测算法首先提取图像特征, 文献[7]采用 Harris 特征与 SIFT 特征结合的方式实现车型识别, 文献[8]采用空间金字塔结合 Adaboost 分类器的方式对车辆进行检

收稿日期: 2019-01-08

修回日期: 2019-05-09

网络出版时间: 2019-06-27

基金项目: 国家自然科学基金(61374047, 60973095); 江苏省博士后科研资助计划(1601085C)

作者简介: 马浩良(1994-), 男, 硕士研究生, CCF 会员(92423G), 研究方向为深度学习、目标检测; 谢林柏, 博士, 教授, 研究方向为智能检测、故障诊断、过程建模与控制。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20190627.1105.058.html>

测。传统方法具有效率高、算法复杂度低的特点,但是需要针对特定场景设计不同特征,而在数据中心处理的图像数据来自不同路口、场景复杂多变,传统方法无法满足处理所有数据的要求。

随着深度学习的迅速发展,深度学习在车辆检测与识别中地应用愈加广泛^[9-10]。现有的物体检测算法主要有基于区域的检测算法(two-stage detector),如 Fast R-CNN^[11]、Faster R-CNN^[12],以及基于回归的物体检测算法(one-stage detector),如 SSD^[13]、YOLO^[14]。针对车辆检测与识别,在目前算法地基础上,研究人员主要在以下两个方面进行改进。一方面,结合车辆的尺寸特征来优化车辆检测的性能。在文献[9]中,在 SSD 算法中添加更符合车辆尺寸的先验候选框来提高车辆的检测效果。另一方面,充分利用卷积神经网络提取车辆特征来提升车辆检测的性能。文献[15]在 Faster R-CNN 中嵌入一个轻量级的建议网络,通过应用不同的特征融合技术,可以对建议区域进行精确的定位和识别。文献[16]中结合 Faster R-CNN 车辆检测模型与 3 种不同大小的卷积神经网络相结合进行迭代训练,择优选择最佳的模型。

以上车辆检测算法通常将车辆作为一类物体,没有考虑车辆类型内部差异,车辆数据集的车型数据不均,实际环境中背景的干扰,使得模型对前景的区分不

足,导致误检漏检的问题。文中基于 SSD 搭建了车辆检测与识别的框架,挖掘车辆类型之间的差异,提升不同车辆类型的检测效果,挖掘前景和背景来抑制背景以提升前景的检测效果。

1 原始 SSD 方法

1.1 SSD 模型结构

SSD 算法是端到端的图像目标检测算法,如图 1 所示。SSD 在训练过程中,首先将不同尺寸的图片变换为 300 * 300 的图像作为输入。为了获取小尺度的特征信息,SSD 选取 VGG 网络模型中两个尺度的特征层作为目标检测的依据,分别为尺度 38 * 38 的 conv4_3 层和尺度为 19 * 19 的 fc7 层。为了获取大尺度的特征信息,SSD 在基础网络 VGG16 后依次增加了 4 个不同尺度的特征提取层 conv6_2、conv7_2、conv8_2 和 conv9_2,其输出 feature map 的尺度分别为 10 * 10、5 * 5、3 * 3 和 1 * 1。SSD 将这 6 个不同尺度的特征层作为目标检测的依据,经过 mbox_priobox 网络、mbox_loc 网络和 mbox_conf_flattern 网络,得到目标所在位置,目标所属类别及其置信度等信息。根据预测结果与真值数据通过 LOSS 层来计算损失,通过批量随机梯度下降算法(SGD)实现模型的权值更新。

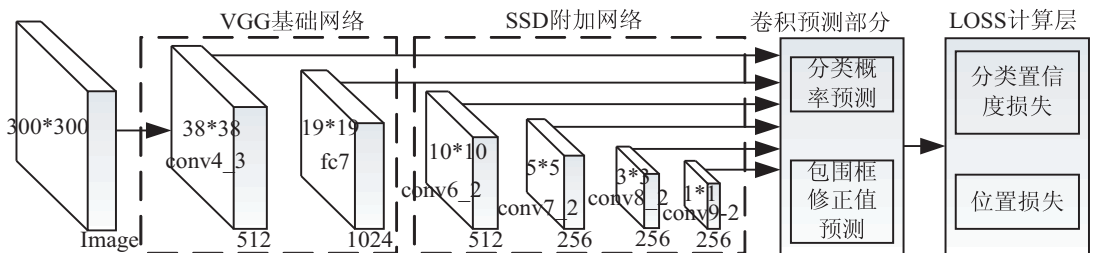


图 1 SSD 的网络框架

1.2 SSD 作为车辆检测与识别方法存在的问题及改进思路

1.2.1 车辆类型样本不平衡问题

在不同的道路场景,不同类型的车辆存在数量上的差异,比如在市區中以公交车与私家车为主,而在高速路上货车的比例较高。

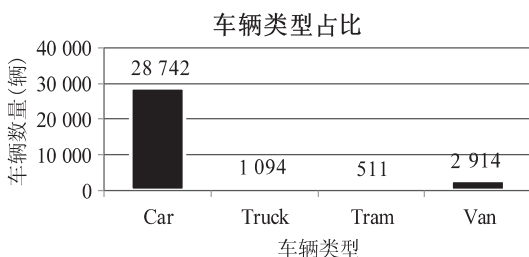


图 2 车辆类型在数据集中的占比

图 2 为文中构建的道路图像数据集中不同车辆类

型的数量分布。该数据集总共有四种车辆类型,其中 Car 有 28 742 辆,Truck 有 1 094 辆,Tram 有 511 辆, Van 有 2 914 辆。可以看出,按照车辆类型划分的数据集呈现出长尾现象。这种现象会导致在训练过程中样本多的车辆类型占据主导,样本少的车辆类型难以学习到更好的参数模型。

1.2.2 先验候选框生成部分的正负样本不平衡问题

在一幅图像中大部分都是背景图像,车辆通常占比很小,而 SSD 算法在六个不同尺度的特征图的候选框上进行回归分类,在训练过程中会引入大量的负样本,从而对模型的优化增加了难度。假设特征图的尺寸为 $m * m$,特征图中每个位置有 k 个先验候选框(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 种类型。这样, k 个先验框就有 $4k$ 个坐标输出,检测 c 个类别就会有 $c * k$ 个置信度输出,最后每张

特征图位置一共有 $(4 + c) \times k$ 个输出。累计所有位置的输出参数,一共有 $(4 + c) \times km^2$ 个输出。

2 SSD 的改进

2.1 车辆类型难易样本挖掘

在训练阶段,SSD 的总损失项是由位置损失和分类置信度损失的加权和定义的。

$$L(x, c, l, g) = \frac{1}{N} [L_{\text{conf}}(x, c) + \partial L_{\text{loc}}(x, l, g)] \quad (1)$$

其中, N 为匹配的默认框的数量; $L_{\text{conf}}(x, c)$ 为分类置信度损失; $L_{\text{loc}}(x, l, g)$ 为位置损失; ∂ 为平衡因子。

如图 2 所示,在车辆数据集中 Car 的数量远大于 Truck, Tram, Van 三种车辆的数量。SSD 在训练过程中,Car 车辆类型的置信度的损失和会远大于 Truck, Tram, Van 三种车辆的置信度的损失和,因此样本较少的车辆类型难以学习到更好的模型。为了改善这个问题,可以通过减少易分类样本在计算损失函数时的权重的方式,提高难学习样本在损失总和中中的占比,从而避免难分类样本淹没在易分类样本当中。在 SSD 中,置信度的损失采用了标准交叉熵的计算方式:

$$\text{CE}(p_i) = -\log(p_i) \quad (2)$$

其中, p_i 为分类置信度; $\text{CE}(p_i)$ 为标准的交叉熵。

$\text{FL}(p_i)$ 表示针对难易样本不平衡问题引入的改进的交叉熵:

$$\text{FL}(p_i) = -(1 - p_i)^\gamma \log(p_i) \quad (3)$$

其中, p_i 为分类置信度; γ 为调制参数。

比较式 2 和式 3, $\text{FL}(p_i)$ 比 $\text{CE}(p_i)$ 多了一个调制系数 $(1 - p_i)^\gamma$, $\text{CE}(p_i)$ 相当于 $\text{FL}(p_i)$ 中 $\gamma = 0$ 时的特殊情况。在训练过程中,置信度越高,与之对应的车辆类型越容易分类,它的损失会越小。然而通常来讲,样本多的车辆类型是属于更容易分类的,因此其损失和是远远大于样本少的难分类车辆类型的损失和,这会导致在训练过程中易分类样本的损失占据主导作用,难分类样本的损失难以得到体现。如图 3 所示,为了减少易分类样本的占比,可以通过调制参数 γ 来进行调节,随着 γ 的增大,易分类车辆类型的损失越接近于零,其损失和在整体损失中的占比会越少,从而使得模型在训练时更专注于难分类的车辆类型。

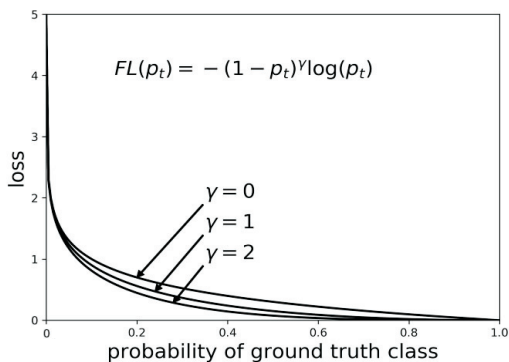


图 3 挖掘难分类样本的交叉熵损失

为了方便比较,将其命名为 SSD-1,其网络结构如图 4 所示。与 SSD 结构相比,SSD-1 只修改了在训练阶段的损失函数,以挖掘难以分类的车辆类型。

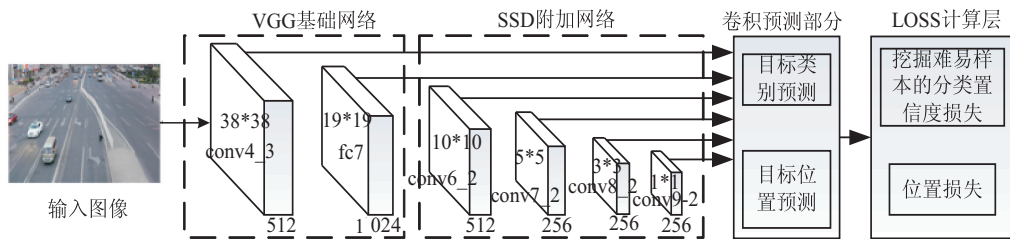


图 4 SSD-1 的网络框架

2.2 在线正负样本挖掘

与基于候选区域的方法不同,SSD 使用卷积神经网络对图像进行卷积后,在不同尺度的特征图上生成一系列不同尺寸和长宽比的候选框,通过评估特征图的每个候选框中是否包含车辆目标并进行车辆类型分类。假设特征图的大小为 $m \times m$,特征图中每个位置有 k 个候选框。根据式 1 可得,每个特征图累计输出 $2km^2$ 个损失。但是在一张图片中,背景通常占据绝大部分,只有少数位置是包含车辆目标的,在匹配步骤之后,大部分的候选框都会被标记为负样本,导致训练过程中正负样本的数量差异巨大,因此在 $m \times m$ 的特征图上求得的 $2km^2$ 个损失中绝大部分是负样本的损失,

不利于车辆检测模型的优化。

为了解决正负样本不平衡的问题,SSD 算法把负样本中的候选框按照所有车型分类的置信度由高到低进行排序,选取置信度最高的几个负样本而不是所有的负样本,使得负样本和正样本的比例总是接近 3 : 1,以平衡正负样本的比例,然而要取得合适比例需要进行大量的实验。为了在训练阶段更好地解决正负样本不平衡问题,通过借鉴基于候选区域的方法在提取候选区域阶段过滤负样本的思想,文中引入了两个 SSD 级联的网络框架^[17]。首先在第一级 SSD(ARM)中通过二分类挖掘正负样本并进行粗略的定位,然后在第二级 SSD(ODM)中以第一级 SSD 的分类结果为指导

筛除掉大量的负样本,最后再进一步检测和识别物体,这种做法会取得更加精确的检测结果。其中第二级 SSD 的损失函数定义为:

$$L(p_i, x_i, c_i, t_i) = \frac{1}{N_{\text{arm}}} \left\{ \sum_i L_b(p_i, [l_i^* \geq 1]) + \sum_i [l_i^* \geq 1] L_r(x_i, g_i) \right\} + \frac{1}{N_{\text{odm}}} \left\{ \sum_i L_m(c_i, l_i^*) + \sum_i [l_i^* \geq 1] L_r(t_i, g_i^*) \right\} \quad (4)$$

其中, i 表示一个 batch 中的第 i 个候选框, l_i^* 表示第 i 个候选框对应的 ground truth box 的类别, g_i^* 表示第 i 个候选框对应的 ground truth box 的位置, p_i 表示在 ARM 中第 i 个候选框的二分类置信度; x_i 表示在

ARM 中第 i 个候选框下的车辆预测坐标, c_i 表示在 ODM 中第 i 个候选框下的多类别置信度, t_i 表示 ODM 中第 i 个候选框下的车辆预测坐标; N_{arm} 和 N_{odm} 分别表示 ARM 和 ODM 中的 positive anchor 数量; L_b 表示二分类损失(有目标\没有目标), L_m 表示多类别损失, L_r 表示回归损失。 $l_i^* \geq 1$ 表示第 i 个候选框对应的 ground truth box 的类别是车辆目标。

为了提高车辆检测与识别的精度,文中将上述两种改进进行融合,使其充分挖掘车辆数据集中的难易样本和正负样本。为了方便比较,将其命名为 SSD-2,其网络架构如图 5 所示。其中第一级 SSD 和第二级 SSD 之间的连接中加入了反卷积操作,这样有利于增加上下文信息,对于模型提取特征有增强的作用。

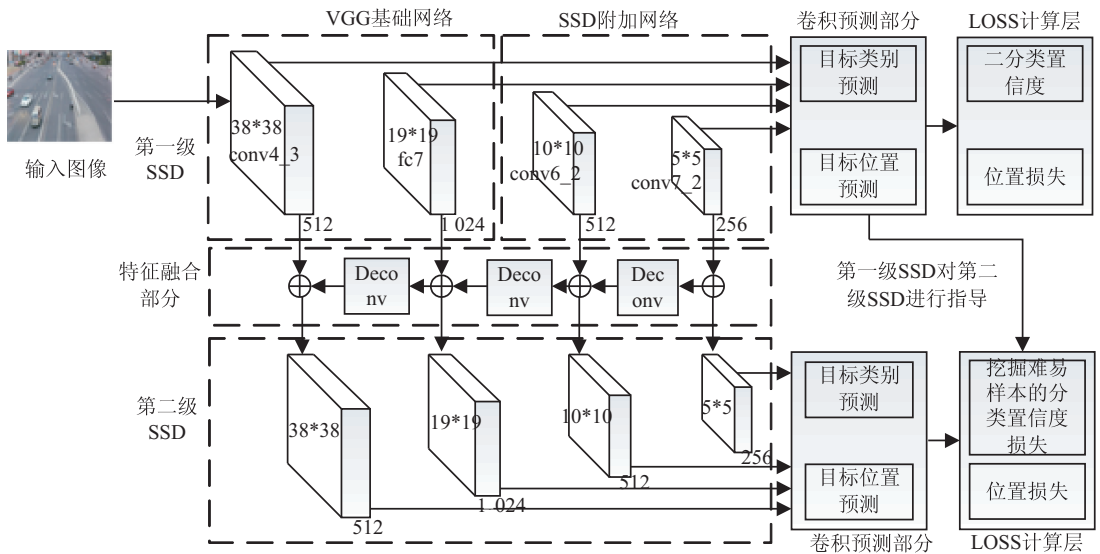


图 5 SSD-2 的网络框架

3 实验与分析

3.1 数据集和实验环境

为了验证车辆检测与识别算法的性能,构建了一个拥有 7 480 幅图像的车辆数据集,该数据集总共包括四种车辆类型,其中 Car 有 28 742 辆,Truck 有 1 094 辆,Tram 有 511 辆, Van 有 2 914 辆。从数据集中随机选取 5 980 幅图像作为训练集,其余 1 500 幅图像作为测试集。各种车型标注数量如表 1 所示。

表 1 数据集车型类别数量

类别	训练集	测试集	合计
Car	28 151	591	28 742
Truck	896	198	1 094
Van	2 347	567	2 914

实验采用的硬件配置如下:CPU, Intel Xeon E5-1650@3.20GHz×12;GPU, NVIDIA GeForce GTX 1070/PCIE/SSE2。软件环境基于 64 位的 ubuntu16.04 操作

系统,采用 Caffe^[18] 框架。以在 PASCAL VOC 数据集下训练好的 VGG-16 模型作为预训练模型,以 0.001 作为初始学习率,采用 multistep 学习策略进行训练。第一阶段模型训练迭代 8 万次至 loss 稳定,第二阶段降低学习率,微调第一阶段生成的模型,迭代至 12 万次,得到最终的训练模型。

3.2 车辆检测与识别结果分析

在评估检测性能时,文中采用检测精度(AP)作为每一类车辆检测结果的评估指标,使用 mAP(mean AP)作为数据集检测结果的评估指标。AP 是反映模型误差程度的参数,检测精度越高,说明模型对不同类型的车辆判别越准确。车辆检测与识别的实验结果如表 2 所示。

P-R 曲线是精确率(precision)和召回率(recall)的曲线,以 recall 为横坐标,precision 为纵坐标。图 6 是四种车辆类型分别在 SSD,SSD-1 和 SSD-2 下的 P-R 曲线。

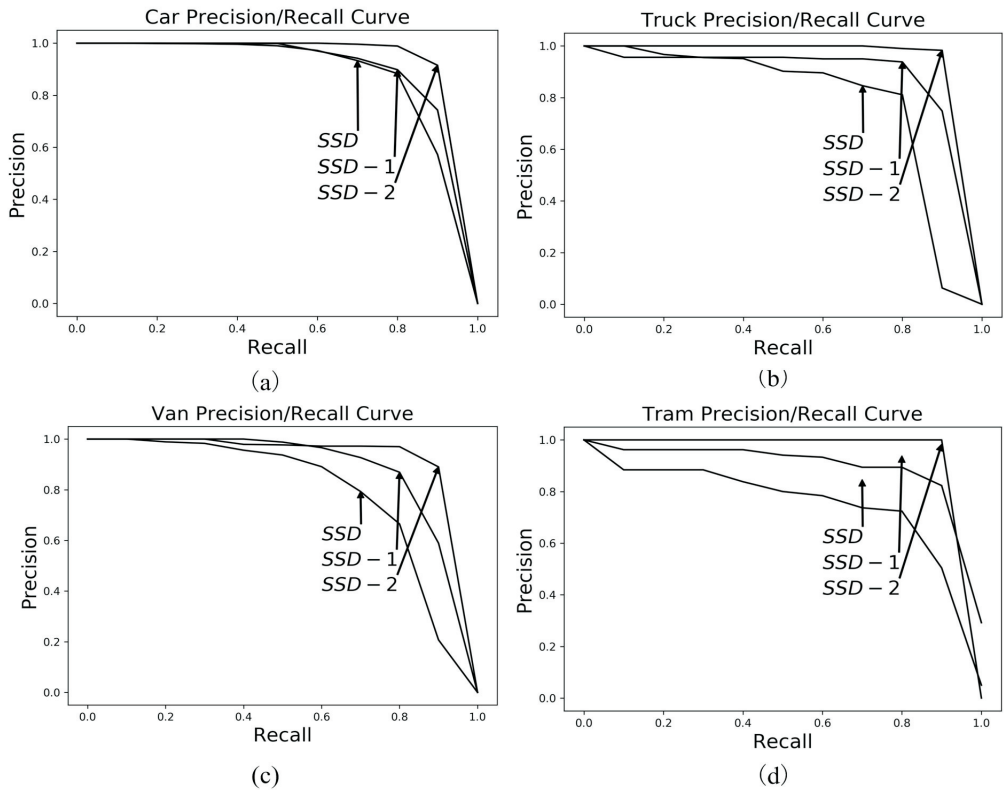


图 6 各类车型在三种算法中的 P-R 曲线

表 2 车辆检测准确度对比情况 %

方法	Car	Truck	Van	Tram	mAP	FPS
SSD	84.9	76.3	76.5	73.6	77.8	57
SSD-1	86.7	85.2	84.9	87.5	86.1	57
SSD-2	90.0	90.6	88.7	90.9	90.0	51

对比 SSD-1 和 SSD 算法的检测精度可以看出,对于 Car 的检测,SSD-1 比 SSD 提高了 1.8%,提升效果不大。但是对于 Truck, Van, Tram 三种车辆类型的检测,SSD-1 比 SSD 分别提高了 8.9%, 8.4%, 13.9%。由图 6 也可以看出,在同样的 recall 下,Truck, Van, Tram 三种车辆类型的 precision 提升明显。分析 SSD-1 相比于 SSD 算法做出的改动。文中通过在 SSD-1 中引入了难易样本挖掘来提升难学习样本的学习比例,使得模型更加专注于难学习的车辆类型,改善了 Truck, Van, Tram 三种车辆类型样本较少的问题。实验结果表明,挖掘难易样本对于解决样本不平衡问题是一个很好的方法。

对比 SSD-2 和 SSD-1 算法,对于 Car, Truck, Van, Tram 四种车辆类型的检测精度,SSD-2 比 SSD-1 分别提高了 3.3%, 5.4%, 3.8%, 3.9%。由图 6 可看出,SSD-2 算法对四种车辆类型的精确率和召回率都有所提升。文中借鉴了基于区域提取的检测算法的思想,在 SSD-1 的基础上引入了两级 SSD 串联的网络结构挖掘正负样本,改进了基于回归的检测算法采用大量候选框的缺点。根据图 7 所示,改进后的 SSD-2

网络在车辆误检、漏检等方面都比 SSD 有所提升。

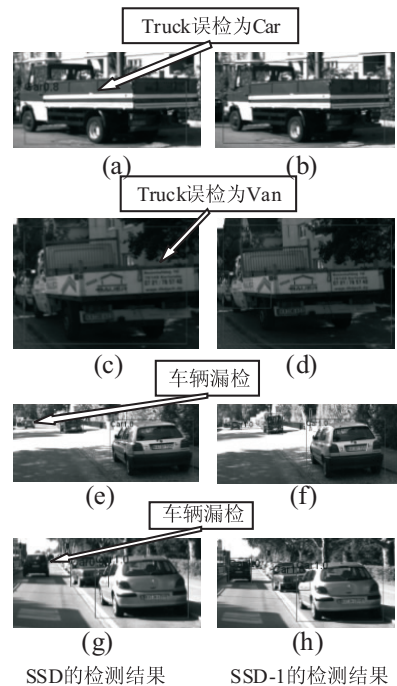


图 7 实验结果对比

3.3 计算性能分析

在测试阶段,SSD 只需一次前向计算便可完成检测。在文中实验平台上分别对三种算法进行了 2 000 次前向计算,各算法的运行速度如表 2 所示。SSD-1 相比 SSD 算法仅改进了计算损失的函数,因此在测试阶段计算速度相同。SSD-2 相比 SSD 增加了网络结

构,计算速度稍慢。可以看出,针对车辆检测与识别任务,文中方法取得了更高的检测精度,同时计算速度仍与 SSD 处于同一量级。

4 结束语

以 SSD 作为车辆检测与识别的基本框架,针对车辆数据集中车辆类型存在样本不平衡的问题,引入了专注于难样本的交叉熵损失函数,同时引入了两级 SSD 串联的网络结构来挖掘正负样本。实验结果证明,改善样本不平衡问题可以提高车辆检测与识别的性能。文中方法有待完善的地方还有很多,比如两级 SSD 串联的网络结构导致计算量增大,还有进一步精简网络结构的空间,对于车辆识别,仅实现了车辆类型的分类,车辆的其他属性如品牌、姿态、颜色等信息的识别需进一步研究。

参考文献:

- [1] SIVARAMAN S, TRIVEDI M M. Looking at vehicles on the road: a survey of vision-based vehicle detection, tracking, and behavior analysis[J]. IEEE Transactions on Intelligent Transportation Systems, 2013, 14(4): 1773-1795.
- [2] TIAN Bin, MORRIS B T, TANG Ming, et al. Hierarchical and networked vehicle surveillance in ITS: a survey [J]. IEEE Transactions on Intelligent Transportation Systems, 2017, 18(1): 25-48.
- [3] 张 慧,王坤峰,王飞跃. 深度学习在目标视觉检测中的应用进展与展望[J]. 自动化学报, 2017, 43(8): 1289-1305.
- [4] 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.
- [5] 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.
- [6] 杨先凤,吴姝泓. 基于自适应阈值与双特征的 ViBe 运动车辆检测算法[J]. 计算机工程, 2018, 44(10): 241-245.
- [7] 康维新,曹宇亭,盛 卓,等. 车辆的 Harris 与 SIFT 特征及车型识别[J]. 哈尔滨理工大学学报, 2012, 17(3): 69-73.
- [8] 曹晓明. 基于多图像特征金字塔的车辆检测[D]. 北京: 北京交通大学, 2016.
- [9] KIM H, LEE Y, YIM B, et al. On-road object detection using deep neural network [C]//Proceedings of the IEEE international conference on consumer electronics-Asia. Seoul, South Korea: IEEE, 2016: 1-4.
- [10] ZHOU Yi, LIU Li, SHAO Ling, et al. DAVE: a unified framework for fast vehicle detection and annotation [C]//Proceeding of 14th European conference on computer vision. Amsterdam, Netherlands: Springer, 2016: 278-293.
- [11] GIRSHICK R. Fast R-CNN [C]//Proceedings of IEEE international conference on computer vision. Washington D. C., USA: IEEE, 2015: 1440-1448.
- [12] REN Shaoqing, HE Kaiming, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [13] LIU Wei, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector [C]//Proceeding of 14th European conference on computer vision. Amsterdam, Netherlands: Springer, 2016: 21-37.
- [14] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C]//Proceedings of the 2016 IEEE conference on computer vision and pattern recognition. Piscataway, NJ: IEEE, 2016: 779-788.
- [15] WANG Li, LU Yao, WANG Hong, et al. Evolving boxes for fast vehicle detection [C]//2017 IEEE international conference on multimedia and expo. Hong Kong: IEEE, 2017: 1135-1140.
- [16] 史凯静,鲍 泓,徐冰心,等. 基于 Faster RCNN 的智能车道路前方车辆检测方法[J]. 计算机工程, 2018, 44(7): 36-41.
- [17] ZHANG Shifeng, WEN Longyin, BIAN Xiao, et al. Single-shot refinement neural network for object detection [C]//IEEE conference on computer vision and pattern recognition. [s. l.]: [s. n.], 2018: 4203-4212.
- [18] JIA Yangqing, SHELHAMER E, DONAHUE J, et al. Caffe: convolutional architecture for fast feature embedding [C]//Proceedings of the 22nd ACM international conference on multimedia. Orlando, Florida: ACM, 2014: 675-678.