

路面交通标志检测调研：数据集及算法[⊖]

王子恒，齐勇刚，刘军
北京邮电大学 宽带网络监控教研中心

【摘要】 交通标志检测是自动驾驶系统中的重要组成部分，本文从数据集和检测算法两方面入手，通过对照实验对交通标志检测现状进行了调研。此外本文根据真实情况提出了新的数据集构建原则及评价标准，并构建了小规模的新数据集与其他数据集进行对比。实验结果展示了当前交通标志识别算法在数据集之间的差异，也验证了本文所提出的数据集有着更强的针对性，更能反映待测方法在真实路面情况下的识别效果，同时也能为算法模型的改进提供合理参考。

【关键词】 交通标志识别，数据集，深度学习，调研

A Survey on Road Traffic Sign Recognition: Datasets and Algorithms

WANG Ziheng, QI Yonggang, LIU Jun

Research Center of Network Monitoring and Analysis, Beijing University of Posts and Telecommunications

Abstract: Traffic sign recognition is an important part of the automatic driving system. In this paper, we investigate the status of traffic sign recognition through contrast experiments in two aspects: dataset and recognition algorithm. In addition, A new dataset construction principles and evaluation criteria are proposed in this paper, and we set up a new small scale dataset to compared with other datasets. The experiment results show the difference between the current traffic sign recognition algorithms and datasets. It also proves that the new dataset proposed in our paper is more pertinent, which can reflect the recognition effect of different algorithms in the real road and provide reasonable suggestion for the improvement of algorithms.

Key words: traffic sign recognition, datasets, deep learning, survey

引 言

在近年的汽车智能化的研究过程中，自动驾驶技术取得了重大突破，尤其是其中的视觉系统，短短几年时间便在识别效果上实现了质的飞跃，这也让越来越多的智能车原型能够走出实验室，在真实道路环境中进行测试，逐渐向实际应用迈进。为了与真实的道路交通环境中的复杂场景和驾驶人生物视觉习惯相匹配，自动驾驶视觉系统在相关研究中通常被分为一系列子任务，例如识别动态目标（车辆、行人等）或静态目标（交通信号灯、道路标志线等），每一类子任务的识别方法和表现都有差异。交通标志是保障道路交通安全顺畅的重要措施，据统计，乡村道路上事故频发的主要原因正是交通标志的缺失或者不规范，这也表明：准确迅速地对标志进行识别将是实现可靠自动驾驶的关键环节，因此本文选取路面交通标志作为研究对象，并从数据集和检测算法两个方面对当前研究及应用状况进行调研。

任何一个模型在投入应用前都需要经历一系列的测试评估工作，以保障系统的准确性及可靠性，那么由相关数据构成数据集进行测试评估将扮演重要的角色。自动驾驶系统更是如此，因道路实测时所存在的风险，需要采取更加严谨、合理测试评估方案。作为自动驾驶中的重要部分，交通标志识别在实际应用中同样对于准确率有着严格的要求，并且要全面考虑真实条件下可能出现的各种复杂情况，因此我们在构建相应数据集及设置评测准则时应该更加具有针对性，例如按照出现交通标志的像素大小分为不同的类，使较小目标在测

试时对漏判、误判拥有更高的容忍度，符合实际情况中驾驶人眼睛的识别优先级。本文调查了已有交通标志数据集的构成，针对其中不合理处进行分析；同时下载了网络上行车记录仪数据并进行整理，尝试构建符合期望的全新数据集。

作为综述性文章，本文主要完成了以下几部分工作：

- 列举了当前交通标志识别中广泛使用的 5 个数据集，从规模、类别设置等方面对比分析它们的特点，同时提出更加合理的数据集的构建原则。
- 将目标识别算法分为传统方法和深度学习方法两类，选取其中具有代表性的 5 个算法，分别在上述数据集中进行对照测试，横向比较算法之间的差异。
- 结合以上实验结果对数据集和算法进行评价，并且讨论新数据集的可行性。

在文章第 1 部分中，本文将介绍交通标志识别领域中已有数据集及按照更科学的方法搜集的新数据集；第 2 部分是列举当前各种识别算法的先进之处及发展趋势，并从中挑选一系列合适的算法用于实验；本文第 3 部分记录了实验的各方面工作，从数据准备到结果讨论；最后本文在第 4 部分做出总结，并对未来的研究进行了展望。

1 交通标志数据集调研

1.1 German Traffic Sign Detection Benchmark (GTSDB)

GTSDB 是在 2010 年德国波鸿鲁尔大学 INI 研究所提出的数据集基础上整理而来的，并作为 IJCNN 2013 目标检测

⊖ 经“第五届国际智能网联汽车技术年会”组委会推荐，本论文发表在《2018 中国汽车工程学会年会论文集》。

竞赛任务的基准。它由 900 张 1360×800 大小的彩色图片构成,所有出现的交通标志均以矩形框标注,根据种类分为 43 类,交通标志大小从 15×15 到 250×250 不等。GTSDB 在深度学习模型 YOLOV2 中最佳识别效果已经能达到 96% 的准确率和 86% 的召回率 ($t=0.5$),但由于数据集的规模小,其已不适合作为衡量自动驾驶中交通标志检测的标准。

1.2 Belgian Traffic Sign Dataset (BTSD)

类似于 German Traffic Sign Dataset, Belgian Traffic Sign Dataset (BTSD) 最初是作为交通标志分类、检测、跟踪等任务的基准而提出的。来自比利时鲁汶大学研究人员通过驾驶一台搭载 8 个高清摄像头,并记录下拍摄视频中交通标志的位置、摄像头姿势等信息构建了该数据集,因此与其他数据集相比, BTSD 还包含了交通标志的 3D 信息,这些数据同样可以用来帮助自动驾驶视觉系统对目标进行更准确的检测,但是在一般的交通标志识别任务中,研究者更多地会使用其中的 2D 信息进行测试。BTSD 数据集中有超过 10000 张图片,其中 8000 余张包含标注信息,图片的像素为 1628×1232 ,包含交通标志共 62 类,在文献 [4] 的交通标志分类中研究用到了该数据集并且取得了不错的效果。

1.3 LISA Traffic Sign Dataset (LTSD)

LTSD 是美国智能安全汽车实验室提出的交通标志数据集,包含视频序列及图片,共有 6610 张图片,质量有高低,其中交通标志种类数为 47,像素大小从 6×6 到 167×168 不等,标注形式同样是矩形框。

1.4 Tsinghua-Tencent 100K Dataset (TT100K)

TT100K 在清华大学 Zhe Zhu 等人的研究中被提出,是清华大学联合腾讯公司共同建立的,从中国 5 个不同城市 10 个区域的道路街景图片中挑选了 10 万张图片,有着 2048×2048 的高像素,其中包含交通标志的有 10000 张,并且采用了多种形式的标注:如矩形、多边形、椭圆等,有超过 100 类的交通标志实例,然而考虑到样本数量带来的不平衡因素,在不做数据增强的情况下选择其中样本数超过 100 个的 45 类标志来进行评估是比较科学的做法。使用 Zhe Zhu 等人的改进 Overfeat-CNN 模型能在 TT100K 上达到 91% 的召回率和 88% 的准确率。

1.5 Russian Traffic Sign Dataset (RTSD)

RTSD 由来自俄罗斯 NRU 高等经济学院的 V. I. Shakhuro 小组在研究中提出,总共有 179138 张图片,156 类交通标志,是几个数据集中规模最大的,图片质量从 1280×720 到 1920×1280 不等。

以上 5 个交通标志数据集按照大致的出现时间排序,可以发现:伴随着检测技术的迅速发展,研究人员对数据集的要求也逐渐提高,无论是在图片质量、数据规模还是类别数量上都有所体现。虽然后来出现的数据集在尽可能降低识别误差方面做出了许多尝试,例如优化标注格式、提供额外辅助信息等,但是从根本上来说这些数据集都没有对路面复杂多变的情况进行更加细致的考虑,无法体现待测模型的全面性指标,势必导致在道路实测时更大的潜在风险及更长的测

试周期,因此提出新的数据集及评价准则是非常有必要的,它能够让研究人员直观地掌握设计模型的实际效果,并能准确地定位到算法的不足之处,进而更加有效地进行模型优化。结合已有交通标志数据集及检测方法的先进之处,本文提出了如下一些交通标志数据集的构建及评价标准原则:

1) 在原始数据的采集阶段,主要考虑两方面的因素:①素材所在的地区:以建立我国通用交通标志数据集为例,由于我国幅员辽阔,各地区间的道路状况在视觉上也有着差异,因此搜集时应将多个不同地区的城市作为来源,并挑选具有地区特点的数据进入数据集,如江南地区的梅雨条件下和华北地区雾霾条件下的照片,保证数据集的多态性;此外在单个城市中的数据也应该尽量覆盖各个区域,包括中心区道路、郊区道路等。②素材的提取方式:对于图片为主体的数据集而言,连续的图片比抽样得到的离散图片包含更多的信息,可以利用物体在连续图片中的连贯性来保证检测的效果,同时连续图片还能用于交通标志跟踪等相关领域的工作。接下来在对数据标注的时候,最初的策略是采用矩形框框出交通标志位置,而在 TT100K 数据集中则增加以像素为单位的两种标注,这样能获得比矩形框更加精准的标注,在一些新兴模型,如 Mask-RCNN 等支持像素标记的模型中能很容易的得到更好的识别效果。

2) 通过查看网络中行车记录仪数据以及对驾驶人实际行车体验进行调查,本文总结了一系列在真实路面条件下存在的、对交通标志识别有干扰的因素,并且为相关的图片打上该因素下的标签,如图 1 所示。首先是交通标志的大小,驾驶人往往会抱怨远处的目标看不清楚,这些小目标对于大部分视觉检测系统来说也是极大的挑战,因此在数据集中需要标记出这些交通标志所属大小范围,这有助于衡量检测算法的检测效果,在本文中,像素值将作为标志大小的划分标准:小于 32×32 的交通标志为小目标,在 $[32 \times 32, 96 \times 96]$ 范围内的为中目标,而大于 96×96 的为大大目标。光照也是在行车过程中影响视觉效果的重要因素,尤其是交通标志牌,在强日光、夜间灯光的直射下或是背光时都会明显区别于正常状态,图 1 中 b 就展示了暗光和强光条件下的交通标志区别。此外,树叶、电线杆(图 1c)等遮挡物以及行车过程中的视角变换(图 1d)也会或多或少地对识别结果造成干扰,记录下这些信息能完善数据集的健壮性。



图 1 真实环境中交通标志的各种不同识别形态

3) 有了按照上述原则生成的交通标志数据集后,研究人员在使用此数据集进行模型评估时不仅能像往常一样测试整体的准确率和召回率,更可以在按照标志大小(小/大)、光照(强光/弱光)、遮挡(有/无)和视角(侧面/正面)等进一步测试模型在这些影响下的检测识别情况,同样按照准确率/召回率评价。

本文按照设想的数据集构建原则，整理了网络上驾驶人上传的行车记录仪视频，并以此为素材构建了小规模 of 国内交通标志数据集，图片规模为 1045 张，其中 808 张含有交通标志，分辨率在 200 × 130 到 2040 × 1530 范围之间，包括 113 个交通标志类别，标注形式为矩形框，设置的场景标签

为天气 {晴，雨，阴，雪，雾}、光线 {好，坏}、遮挡 {有，无}、距离 {远，近}、污损 {有，无} 以及是否正对镜头 {是，否}，具体数据分布列举在表 1 中。在接下来的实验中，该数据集将与以上 5 个数据集一起进行实验并按照提出的更加细化的测试方法分析结果。

表 1 新数据集详情

数据集构成（张）														
天气					光线		遮挡		距离		污损		是否正对	
晴	阴	雨	雪	雾	好	差	有	无	远	近	有	无	是	否
507	241	47	4	9	717	91	77	731	343	365	21	787	603	25

2 交通标志识别方法调研

纵观计算机视觉领域的发展历程，图像识别研究在两个时间段最为活跃，期间均涌现出了许多经典的算法和模型。在 2012 年以前，识别算法的基本流程是：提出不同尺度的候选框，对候选框内图像区域使用 LBP、LTP、HOG 等算法进行特征提取，最后使用分类算法，如决策树、支持向量机（SVM），判断是否是待测目标。这些方法在各自的领域中都取得良好的效果，但受限于算法性能，许多检测任务，如识别图片中交通标志，通常被分为两步进行：首先检测出标志所在位置，再判断所属标志种类。而在 2012 年后，随着深度学习研究的迅速发展，基于 CNN 的目标检测模型无论是在识别效果上还是检测速度上都较之前有着明显提升，能够将传统方法中的两步任务通过一个模型高效的完成，成功地实现了图片的端到端检测。本文据此将目标检测算法分为传统方法和深度学习方法，并选择其中具有代表性的算法在数据集上进行测试。虽然传统方法在整体效果和性能上不如深度学习方法，但在某些领域也值得深度学习借鉴，且在小数据集中具有易调参、训练快等优势，因此也在将其列入待测列表。

2.1 Haar/AdaBoost

AdaBoost 是 Boosting 分类算法中的一种，其核心思想是训练多个弱分类器最终组合成更强的分类器，具有自适应、不容易过拟合等优点。在 2001 年 Viola 和 Jones 等人提出利用 Haar-like 小波特征结合 AdaBoost 分类器对物体进行快速检测，并且在人脸检测上取得了不错的效果。Haar 使用特征模板（图 2），通过区域间矩形像素相减得到该区域的特征。和人脸区域类似，交通标志同样具有固定的形状，并且存在一定的对称性，因此 Haar/AdaBoost 同样能用于交通标志识别。

2.2 HOG/SVM

HOG 特征的全称是 Histogram of Oriented Gradients，它是表征图像区域内梯度方向和梯度强度分布特性的描述，在 CVPR2005 上由 Dalal 提出。与其他方法相比，HOG 特征对于几何、光学形变具有很好的不变性。在实际应用中，HOG 配合经典的机器学习分类算法 SVM 在许多检测领域如行人检测、车辆检测等方面都取得了成效，表现了其良好的适应

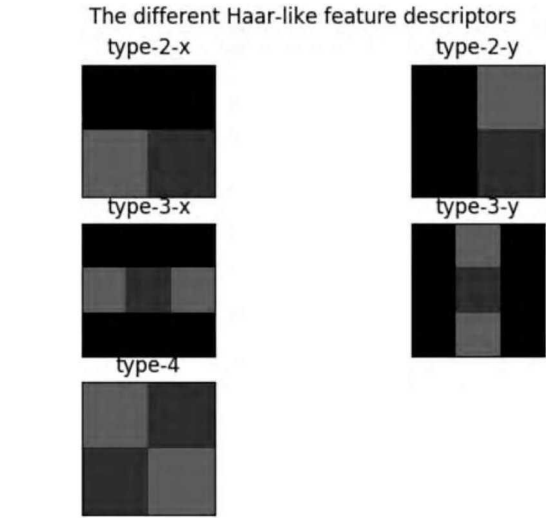


图 2 Haar 特征模板

能力，其中 HOG/SVM 的改进版本在 IJCNN 中关于德国交通标志分类的任务中有着优异表现。

2.3 DPM/LSM

Deformable Parts Model（DPM）特征提取算法可以看作是 HOG 特征算法的扩展，主要是在提取时使用了多模型和各自的子模型，充分适配目标在不同视角、不同动作下的姿态，其本质是弹簧形变模型。而 Latent-SVM 在 SVM 的基础上加入了混合模型中的隐变量，在特征中对应部件位置的配置变量，可以进一步发掘各部件间的关系。利用 DPM 特征的检测算法连续获得三年的 VOC 目标检测算法的冠军，但其在概念上更适合如行人、自行车等具有不同部件的目标，对于交通标志检测而言其相对于其他方法并没有明显的优势。

2.4 RCNN/Fast-RCNN

Region CNN（RCNN）是利用深度学习进行目标检测的开山之作，作者 Ross Girshick 在 VOC2012 数据集上达到了 53.3% 的平均准确率，大幅度超越了之前基于 DPM 的检测算法。RCNN 检测步骤（图 3）为：通过候选框提取算法（如 Selective Search，EdgeBox 等）滑动提出可能的候选框，接着将候选区域分别送入 CNN 网络中进行特征提取，最后

对每一类目标使用 SVM 二元分类,同时使用回归器对目标位置进行修正。RCNN 虽然在检测效率上较传统方法提升不少,但由于存在特征提取部分存在冗余和单独训练分类器回归器所占空间较大等问题,其仍无法满足实际应用的需求。改进版本 Fast-RCNN 有效地解决了上述问题,极大加快了检测速度,但由于候选框算法的存在,始终不能完成端到端的训练和检测过程。

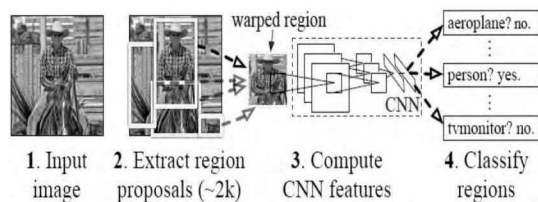


图3 RCNN 检测流程

2.5 Faster-RCNN

Faster-RCNN 是 Ross Girshick 团队在 RCNN 基础上的改进模型,在 VOC2007 数据集中识别效果上提升不大,准确率为 73.2%,但是在时间效率上快了近 15 倍,并且实现了端到端的检测。Faster-RCNN 主要是通过添加一个与共享 Fast-RCNN 结构共享权值的 RPN 网络来提出候选框(图4),代替了效率不高的 Selective Search 等传统方法,其巧妙的设计思路也给后来的深度学习模型提供了很多参考,已经在一系列任务中取得了丰富的成果。

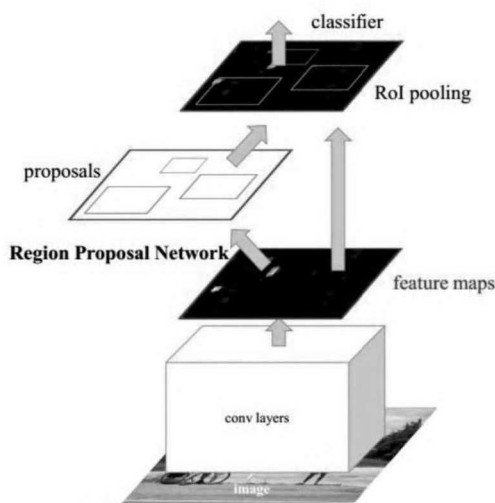


图4 Faster-RCNN 结构

2.6 YOLO

2016 年出现 YOLO 模型,将识别效率进一步提高,达到了惊人的 45FPS,是 Faster-RCNN 的 7 倍,然而识别准确率只有不到 60%。YOLO 同样实现了端到端检测,但并没有使用 RPN 类似的候选框提取步骤,而是将整个图片划分为多个网格区域,并且将候选框检测当作回归问题处理。在改进版本 YOLOv2 中,识别准确率在 67FPS 的效率下已经提升到 Faster-RCNN 水平,同时还具有很强的泛化能力。

2.7 Single Shot MultiBox Detector

Single Shot MultiBox Detector (SSD) 借鉴了 Faster-RCNN 和 YOLO 各自的优点,在 VOC2007 数据集上有着 73.9% 的准确率和 58 的 FPS。它通过将 Faster-RCNN 中的 RPN 和 YOLO 中的回归设计结合,并加入了多尺度检测,在保持高准确率的前提下加快了检测效率。

综上所述,本文在传统方法中挑选 Haar/AdaBoost 算法和 HOG/SVM 两个经典目标分类算法,而在深度学习中选择了 RCNN 中能实现端到端检测的 Faster-RCNN、YOLO、SSD 进行对照实验。

3 实验

本文的实验内容主要是使用 2 中所提到的 5 个算法对 GTSDb、BTSD、LTSD、TT100K、RTSD 和本文所构建的小规模新数据集 (Ours) 测试,记录实验结果并进行对比,同时将对包含干扰因素、信息标记的 Ours 数据集进行有针对性的评估。

3.1 实验准备

实验的前期准备包括标注清理和训练集/测试集划分。在许多数据集中都存在着标注误差,例如标注的范围超过了图片的宽高,这将会导致在读取交通标志区域时发生异常,从而影响实验效果,因此在实验前要将这些异常标注修改,使之在合理范围内,同时不影响标注准确性。训练集(验证集)/测试集分别对应模型训练和测试两个阶段,对于已有划分合理的训练/测试数据的数据集(如 TT100K),本文直接使用它们进行实验;而对于没有训练/测试划分的数据集(LTSD、RTSD 等),本文按照训练:验证:测试 = 6:3:1 的比例划分(若无验证步骤则按训练:测试 = 9:1 划分),并保证所有的类别样本都等比例地出现在这些集合中,此外,GTSDb 虽然划分了训练/测试数据,但其测试数据中并没有包括所有的类,这影响了实验效果,本文同样按照上述划分为其重新生成训练/测试集。

3.2 训练过程

本文实验所有的训练和测试过程在搭载了 12 核 CPU、128G 内存的服务器上完成,并使用了 Nvidia Geforce 1080Ti 11G GPU 加速深度学习方法。对于传统方法中的两类算法,由于其只能检测区域内是否某类交通标志,因此本文使用 Selective Search 提取的候选框作为它们的前置条件,通过这样的组合方式来模拟深度学习方法中才能实现的端到端检测。与此同时,对于数据量较小、图片质量不高的数据集 GTSDb、BTSD、LTSD 和 Ours,迭代总次数为 80000 次;对于图片质量高或者规模大的数据集,如 TT100K 和 RTSD,迭代次数设置为 160000。

1) 传统方法:在传统方法的实验中,本文主要通过 Python 来完成训练,使用 scikit-image 模块中的 haar_like_feature_coord 函数和 hog 函数分别提取像素区域的 Haar 特征和 HOG 特征,以及 sklearn 模块中的 AdaBoostClassifier 函数和 SVC 函数来实现 AdaBoost 和 SVM 分类工作。传统方法分类器要求保持特征向量长度固定,因此训练和测试输入的目标

区域将通过 OpenCV 模块统一变换为 50×50 。在 Haar/AdaBoost 中, 本文使用了 10×10 的 type-4 特征窗口提取 Haar 的特征, 并设置了 1000 个决策树作为级联弱分类器, 采用样本集分类效果作为 AdaBoost 中弱分类器的权重, 同时超参数学习率设为 1.5。在 HOG/SVM 实验里, 提取 HOG 特征的 cell 以及 block 包含的 cell 数分别设为 (8, 8) 和 (3, 3), 在 8 个方向上进行滑动, SVM 分类器选用的核函数为 RBF 核, 其中的超参数 Gamma 和 Cost 通过 sklearn 模块里的 GridSearch 函数拟合训练集中的数据分布并在局部寻找得到最优的值。而在作为前置任务的 Selective Search 算法中, 设置候选框的最大规模为 5000, sigma 值为 0.9, 同时最小的候选框大小为 20。

2) 深度学习方法: 对于深度学习方法中的三个模型, 为了控制变量进而得到客观的比较, 本文在相关训练中均使用 0.001 的初始学习率, 同时选择 SGD 作为学习率优化策略, 每 20000 步衰减 0.1, 同时 batch_size 统一设置为 16。由于深度学习中这些算法的参数均是对 VOC 数据集适配的, 因此本文并未涉及模型内部细节的修改及优化, 而是直接测试以观察其原始版本对于交通标志的识别情况。在 Faster-RCNN 中, 本文选择 ZF 网络结构用于训练; 在 YOLO 算法中, 本文采用 YOLOv2 发布的高速版本 tiny-yolo 模型进行训练; 而在 SSD 的训练则在 SSD_VGG300 模型上完成。

3.3 实验结果

本文所有的测试均在 iou 阈值为 0.5 的条件下完成。在识别效率方面中, 传统方法结合候选框算法一般需要 1min 以上才能完成单张高像素图片的识别, 而深度学习方法中, 三者的平均耗时分别为: 0.447s, 0.21s, 0.05s。考虑到传统方法通常将识别任务分为检测 (指出目标位置) 和分类 (识别区域内目标的类别) 两步, 因此本文使用两类任务中对应的两种指标来作为评价, 记为 R' 和 P' , 如 (1) (2) 分别所示:

$$R' = \frac{tb}{npos} \quad (1)$$

$$P' = \frac{tp}{tb} \quad (2)$$

式中, tb 代表所有与目标 iou > 0.5 的候选框数; tp 代表中满足与目标 iou > 0.5 的且分类预测正确的候选框数; $npos$ 代表测试集中目标的个数。而深度学习中的 P 、 R 如式 (3)、式 (4) 所描述:

$$R = \frac{tp}{npos} \quad (3)$$

$$P = \frac{tp}{nob} \quad (4)$$

式中, tp 代表中满足与目标 iou > 0.5 的且分类预测正确的候选框数目; $npos$ 代表测试集中目标的个数; nob 代表所有检测出的候选框个数。

本文依据以上指标将实验结果总结成表 1、表 2 和表 3 三个表格。通过在实验结果中对识别方法进行横向比较可以发现, 传统方法中 HOG/SVM 的组合在大部分数据集的分类正确率上要比 Haar/AdaBoost 高出不少, 但在 RTSD 中效果很差, 原因在于 SVM 超参数的设置可能未设置合适, 在这方面 AdaBoost 则要方便许多, 并且可以发现前置检测算法 Selective Search 通常只能找准 20% 左右的候选框, 与深度学习方法有差距, 因此传统方法离实际应用还有很长一段距离; 深度学习方法中的 Faster-RCNN 和 YOLO 在大规模数据集中效果较好, 但在检测效率中 YOLO 的速度大约是 Faster-RCNN 的两倍, 效果较差的 SSD 又比 YOLO 快了 4 倍, 综合来看在本次实验中 YOLO 的表现会更加好一点, 而 SSD 算法表现不佳的原因可能是本文所使用的 SSD300 模型过于追求检测速度, 而舍弃了检测质量。此外, 高分辨率下小目标的识别是几个深度学习方法中的通病, 原因在于其不采用候选框提取的核心思想天生就不太适合小目标, 相比之下反而是传统方法的候选框提取如 Selective Search 或 EdgeBox 能更好地适应小目标检测。但考虑到深度学习方法的高速特点, 在实际应用中通过使用将大图片切割成多张较小图片进行检测等技巧可以有效地改善这一问题。另一方面, 纵向比较相同算法在不同数据集上的表现可以发现, 数据越接近真实情况 (类别多、样本多、分辨率高), 识别的难度也就越大, 无论是从各种算法的准确率还是召回率都可以看出, 尤其是 TT100k 数据集, 2048×2048 的高分辨率对于每个算法而言都是极大的挑战, 但也有例外, 算法对数据规模最大同时也是类别最多的 RTSD 数据集识别准确率很高, 通过观察数据构成发现: 其可能原因是 RTSD 中图片均为连续图片, 对于同一目标有足够的、由远及近的样本去训练, 此外个别类出现的频率较高也导致了结果的异常。

在本文新提出的小规模数据集的实验中, 由于样本数量少, 在 YOLO 和 SSD 上总体识别效果还不错, 在 Faster-RCNN 上训练时可能出现了过拟合现象, 导致结果召回率不错, 而正确率急剧下降。结合实际因素标签及每一项测试结果中可以发现很多有意思的地方: 多数算法对于雪天目标的识别效果很差, 同时在大雨天气中召回率显著降低, 而在其他气候下识别效果并未受很大干扰; 而光照、遮挡、距离和角度的影响只会减少召回率, 准确率未受显著干扰; 但污损的目标则会导致识别效果急剧下降, 几个模型都很难识别出其中的目标。因此可以看出现有算法对于这些特定场景下的检测仍有改进的空间。

表 2 传统方法实验结果

单位: %

数据集	GTSDb		BTSD		LTSD		TT100K		RTSD		Ours	
	R'	P'	R'	P'	R'	P'	R'	P'	R'	P'	R'	P'
Haar/AdaBoost	17.9	13.9	33.0	21.1	38.2	11.9	7.8	22.7	37.1	21.6	7.8	19.8
HOG/SVM	75	13.9	59.3	21.1	63.4	11.9	74.9	22.7	1.3	21.6	51.0	19.8

表 3 深度学习实验结果

单位:%

数据集	GTSD		BTSD		LTSD		TT100K		RTSD		Ours	
	P	R	P	R	P	R	P	R	P	R	P	R
Faster- RCNN	81.7	72.8	90.3	78.1	70.2	25.6	63.4	20.6	96.3	53.7	35.2	62.0
YOLOv2	98.8	39.6	92.7	65.8	99.0	57.7	67.5	27.7	98.9	12.0	80.1	53.1
SSD	100	19.0	76.6	47.1	92.2	31.9	75.5	5.2	93.1	10.6	85.5	22.9

表 4 针对场景标签测试结果

单位:%

检测数据集	天气										光线				遮挡				距离				污损				是否正对			
	晴		阴		雨		雪		雾		好		差		有		无		远		近		有		无		是		否	
	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R
Faster- RCNN	33	61	42	69	34	57	0	0	0	0	35	63	39	57	29	50	36	64	41	55	32	68	18	60	35	62	34	61	43	65
YOLOv2	79	50	81	40	50	5	0	0	100	50	80	47	70	25	77	42	92	64	82	50	76	39	0	0	79	45	80	48	78	29
SSD	87	26	100	16	33	5	0	0	100	50	87	23	80	14	86	18	87	23	93	11	85	33	0	0	87	23	88	23	82	19

4 结论

本文对当前交通标志识别领域的算法及数据集进行调研及实验,为相关领域的研究者们选择合适的算法和测试数据集提供了参考依据和建议。需要说明的是,本文所用的算法对不同数据集的测试,大多从开源代码发布版本修改而来,涉及检测类别和标注形式的匹配、修正,并未进行网络结构的优化,因此与其他一些经过精细调参使用相同数据集和算法的相关研究相比,本文实验得到的准确率和召回率有一定程度的下降,不过对于本文的目的而言并不造成影响;本文实验旨在客观地比较不同交通标志识别算法及数据集的优缺点,检测各种算法对于不同数据集的性能表现,放弃了可能会带来不平衡影响的一些优化操作。通过比较实验效果的差异,我们能够对业界后续深入研究及实际应用上提供考量

标准。

此外,本文提出一种全新、包含更多复杂真实环境的交通标志识别数据集。虽然所提出的新数据集规模不大,但它涵盖了更多能够表征现实复杂环境下的图片,比如不同天气、光线、遮挡等,相比于市面上已有的公开数据集,我们提出的数据集为交通标志的检测提出了更加全面的考量,其多个评测指标可以有效地衡量待测模型的不足之处,同时也能正确反映实际情况下路面交通标志识别主要干扰因素,为模型的改进指出方向,且其数据构成也能准确体现国家或地区的实际情况。在日后的研究工作中,我们将联合其他研究机构,采集更加全面的数据对数据集进行扩充,并且丰富场景标签,为将来的自动驾驶标识识别系统的标准化测试奠定基础。

参 考 文 献

[1] Zhang, J., Huang, M., Jin, X., & Li, X. (2017). A Real-Time Chinese Traffic Sign Detection Algorithm Based on Modified YOLOv2 [J]. Algorithms, 10 (4), 127.

[2] Prisacariu, V. A., Timofte, R., Zimmermann, K., Reid, I., & Van Gool, L. (2010, August). Integrating object detection with 3d tracking towards a better driver assistance system [C]. In Pattern Recognition (ICPR), 2010 20th International Conference on (pp. 3344-3347). IEEE.

[3] Timofte, R., Zimmermann, K., & Van Gool, L. (2014). Multi-view traffic sign detection, recognition, and 3d localisation [J]. Machine vision and applications, 25 (3), 633-647.

[4] Huang, Z., Yu, Y., Gu, J., & Liu, H. (2017). An efficient method for traffic sign recognition based on extreme learning machine [J]. IEEE transactions on cybernetics, 47 (4), 920-933.

[5] Zhu, Z., Liang, D., Zhang, S., Huang, X., Li, B., & Hu, S. (2016). Traffic-sign detection and classification in the wild [C]. In Proceedings of the IEEE Conference on

Computer Vision and Pattern Recognition (pp. 2110-2118).

[6] Shakhuro, V. I., & Konouchine, A. S. (2016). Russian traffic sign images dataset [J]. Computer Optics, 40 (2), 294-300.

[7] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks [C]. In Advances in neural information processing systems (pp. 1097-1105).

[8] Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting [J]. Journal of computer and system sciences, 55 (1), 119-139.

[9] Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features [C]. In Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on (Vol. 1, pp. 1-1). IEEE.

[10] Dalal, N., & Triggs, B. (2005, June). Histograms of oriented gradients for human detection [C]. In Com-

- puter Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on (Vol. 1, pp. 886-893). IEEE.
- [11] Cortes, C., & Vapnik, V. (1995). Support-vector networks [J]. *Machine learning*, 20 (3), 273-297.
- [12] Benenson, R., Omran, M., Hosang, J., & Schiele, B. (2014, September). Ten years of pedestrian detection, what have we learned? [C]. In *European Conference on Computer Vision* (pp. 613-627). Springer, Cham.
- [13] Felzenszwalb, P., McAllester, D., & Ramanan, D. (2008, June). A discriminatively trained, multiscale, deformable part model [C]. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on* (pp. 1-8). IEEE.
- [14] Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation [C]. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 580-587).
- [15] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks [C]. In *Advances in neural information processing systems* (pp. 91-99).
- [16] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016, October). Ssd: Single shot multibox detector [C]. In *European conference on computer vision* (pp. 21-37). Springer, Cham.