

# 基于区域卷积网络的行驶车辆检测算法

曹长玉<sup>1</sup>, 郑佳春<sup>2</sup>, 黄一琦<sup>1</sup>

(1. 集美大学航海学院, 福建 厦门 361021; 2. 集美大学信息工程学院, 福建 厦门 361021)

**[摘要]** 为解决多种天气与多种场景下主干道路行驶车辆检测存在的实时性、泛化能力差、漏检、定位不准确等问题, 研究了基于 TensorFlow 深度学习框架的区域卷积神经网络 (Faster R-CNN) 算法, 通过引入 VGG16 神经网络模型, 优化 ROI Pooling Layer, 并采用联合训练方法, 得到改进的算法模型。采用 UA\_CAR 数据集进行模型训练, 实现行驶中的车辆检测, 测试结果与优化前 Faster R-CNN 比较, MAP 提高了 7.3 个百分点, 准确率提高了 7.4 个百分点, 检测用时 0.085 s, 提高了对多种环境与场景的适应性。

**[关键词]** 行驶车辆检测; 卷积神经网络; 联合训练

**[中图分类号]** U 495

## Research on Running Vehicle Detection Algorithm Based on Regional Convolution Network

CAO Changyu<sup>1</sup>, ZHENG Jiachun<sup>2</sup>, HUANG Yiqi<sup>1</sup>

(1. Navigation College, Jimei University, Xiamen 361021, China;

2. Information Engineering College, Jimei University, Xiamen 361021, China )

**Abstract:** In order to solve the issue of poor real-time performance, poverty generalization ability, missed detection and inaccurate location of running vehicle detection on currently main roads in various weather and various scenarios, the regional convolutional neural network (Faster R-CNN) algorithm based on TensorFlow framework of deep learning is studied. By introducing the VGG16 neural network model, optimizing the ROI Pooling Layer, and adopting method of joint training, an improved algorithm model is obtained. The UA\_CAR database is used for model training to carry out vehicle detection in the course of driving. Compared with the Faster R-CNN before optimization, the test results show that the MAP is increased by 7.3 percentage points, the accuracy rate is increased by 7.4 percentage points, and the detection time is 0.085s. The network improved adaptability to multiple environments and scenarios.

**Keywords:** running vehicle detection; convolutional neural network; joint training

## 0 引言

在智能交通管理系统<sup>[1]</sup>、智能视频监控系统<sup>[2]</sup>、高级驾驶辅助系统 (ADAS)<sup>[3]</sup>等系统中车辆检测是必不可少的高新技术<sup>[4-5]</sup>。目前国内外关于车辆检测的研究主要有: 文献 [6-7] 提出背景建模法, 检测速度较快且具有一定的适应性, 但检测时易产生黏连现象, 一般只适用于固定场景;

**[收稿日期]** 2019-01-04

**[基金项目]** 福建省科技计划重点项目 (2017H0028); 福建省自然科学基金项目 (2013J01203、2015J01265)

**[作者简介]** 曹长玉 (1994—), 男, 硕士生, 从事交通通信及物联网技术研究。通信作者: 郑佳春 (1965—), 男, 教授, 从事通信工程研究。E-mail: jchzheng@jmu.edu.cn

Lienhart 等<sup>[8]</sup>提出帧间差分法和光流法，可用于行驶车辆检测，前者检测速度较快但准确率较低，而后者反之；文献 [9-12] 提出滑动窗口分类法，使用分类器对感兴趣区域的特征进行分类，其稳定性与准确率均较高，但仍存在提取特征难度大、区域选择易冗余、计算复杂、速度较慢等不足；Girshick 等<sup>[13]</sup>提出 R-CNN 算法，采用选择性搜索法（selective search，SS）<sup>[14]</sup>选取感兴趣区域，检测识别率有所提高，但重复计算导致耗时较长。采取 SPPnet<sup>[15]</sup>思路对 R-CNN 进行改进，进一步提出 Fast R-CNN<sup>[16]</sup>算法，缩短重复计算时间，但感兴趣区域选择耗时问题仍未解决；Ren 等<sup>[17]</sup>通过引入 region proposal networks（RPN）网络，进而提出 Faster R-CNN 算法，极大缩短检测时间；Redmon 等<sup>[18]</sup>提出基于回归思想的 YOLO 算法，检测速度快，但易出现定位问题；后来 Redmon 等<sup>[19]</sup>又提出 YOLOv2 算法，虽速度快、准确率高，但定位准确率低于 Faster R-CNN。

针对多种天气与多种场景下主干道行驶车辆的实时、准确、全面检测问题，本文在对深度卷积神经网络 Faster R-CNN 算法研究基础之上，提出一种新的解决算法。

# 1 Faster R-CNN 算法及优化

## 1.1 Faster R-CNN 算法

Faster R-CNN（faster region-based convolutional neural networks）算法引入 RPN 网络提取目标候选区域框，该网络与检测器共享图片卷积特征。Faster R-CNN 通过将特征提取、候选区域框选择、边界框回归与分类，综合在单个网络中并采用交替训练方式，提升网络的准确率与速度。

该算法通过 ZF 网络模型提取任意输入尺寸图片的特征，而后用于 RPN 网络和 ROI pooling 网络。RPN 网络使用 3×3 的滑动窗口在图片特征上进行滑动，每个位置可产生由 3 种面积 3 种比例（1:1，1:2，2:1）组成的 9 个候选区域框（anchor 框）。其 1×1 的卷积层预测每个 anchor 框的偏移缩放量及目标类别概率。ROI pooling 层综合利用特征图与候选区域框，产生固定大小为 7×7 的 ROI 特征，将该特征送入全连接网络进行分类与回归，输出候选区域框中的目标类别概率及对应目标精确的偏移与缩放量。

## 1.2 算法优化

为了提高算法的性能，本研究对算法做了以下三方面的优化（流程图见图 1）：

1) 提取更深层次的特征。Shared Convolutional layer 层引入 VGG16 网络模型，将卷积层由 5 层扩展至 13 层，激活函数 relu 扩展至 13 个。

2) 缩短候选区域选择时间。计算特征图尺寸  $(P,Q)$ ，利用网络总池化倍数将原图划分为  $P \times Q$  网格，网格点用左上角及右下角坐标形式表示。根据总池化倍数，初始一组 3 种面积 3 种比例组成的 anchor 框，与每个网格点坐标相加，得  $P \times Q \times 9$  个 anchor 框。训练过程中，为简化采样与缩短时间，从单张图片中采样 256 个 anchor 框，用于训练网络。由 3×3 卷积核与并列的 1×1 卷积核组成区域平移网络，预测 anchor 框的偏移、缩放、类别概率。对 anchor 框进行回归修正的计算公式为： $r_x = x_a + p_x w_a, r_w = w_a e^{p_w}, r_y = y_a + p_y h_a, r_h = h_a e^{p_h}$ ，其中， $(p_x, p_y, p_w, p_h)$  为预测 anchor 框的偏移与缩放量， $(x_a, y_a, w_a, h_a)$  为 anchor 框的中心点、宽高坐标。限定超越图像边界的回归修正建议框，使用非极大值抑制（non-maximum suppression，NMS）算法<sup>[20]</sup>提取得分最高的 2000 个回归修正建议框。

3) 固定回归修正建议框特征。用 crop\_and\_resize 方法代替 ROI Pooling Layer 层，根据图像大小对选取的回归修正建议框做归一化处理，通过裁剪调整特征大小使其为 14×14。采用池化层进行降

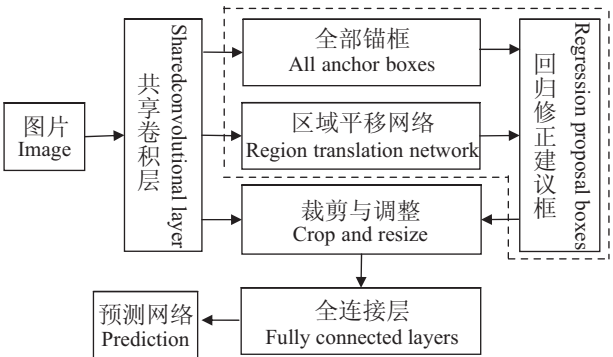


图 1 本文算法流程图  
Fig.1 Flowchart of algorithm

维处理，输出 [128,7,7,512] 形式，用于后续全连接层的输入。为防止过拟合，fully connected layer 使用 dropout 函数。本文研究真实场景行驶车辆检测问题，自制 UA\_CAR 数据集，避开公开数据集，优化后记为 vFaster R - CNN。

2 网络训练

本文使用联合训练方式，对网络进行端到端的训练，可更为准确全面地提取图片特征，训练较为简单有效。损失函数直接影响着网络性能。本文损失函数由分类损失与回归损失组成，分类损失来自区域平移网络的二分类损失和预测网络的多分类损失，计算式<sup>[16]</sup>为： $L(\{c_i\}) = \sum_i L_{cla}(c_i, c_i^*)/M_{cla}$ ， $L_{cla}(c_i, c_i^*) = -c_i^* \lg(c_i)$ ，其中： $i$  为第  $i$  个 anchor 框或回归框； $c_i$  为网络类别预测； $c_i^*$  为真实目标标签； $M_{cla}$  为训练框的个数，在区域平移网络中为 256，在预测网络为 128。

此外，回归损失由区域平移网络回归损失和预测网络回归损失两部分组成。文献 [17] 使用  $\lambda/N_{reg}$  对函数值做归一化处理，本文则使用  $1/M_{cla}$  方式，如：

$$L(\{s_i\}) = \sum_i p_i^* L_{reg}(s_i, s_i^*)/M_{cla}, L_{reg}(s_i, s_i^*) = \text{smooth}_{L_1} \begin{cases} 0.5t^2, \text{if } |t| < 1 \\ |t| - 0.5, \text{otherwise} \end{cases}$$

其中： $s_i$  为预测的偏移与缩放量， $s_i^*$  为与真实框的偏移与缩放量。参考文献 [17]，将分类损失与回归损失相加，得网络总损失函数  $L(\{p_i, s_i\}) = \sum_i p_i^* L_{reg}(s_i, s_i^*)/M_{cla} + \sum_i L_{cla}(p_i, p_i^*)/M_{cla}$ 。

3 仿真实验及结果分析

3.1 实验平台

本实验均在 ubuntu1 6.04 系统进行，处理器为 Intel (R) Core i7 - 7700 CPU @ 3.60GHz × 8，显卡为英伟达 GeForce GTX1080，软件编程语言为 python 3.5，使用 TensorFlow 深度学习框架进行训练。

3.2 数据集处理

本文图片采集于 DETRAC 数据集，该数据集场景丰富且背景复杂，图片分辨率为 960 × 540 像素。从 DETRAC 数据集中采集白天、黄昏、晚上和雨天图片，将采集的 6203 张图片与 VOC\_2007 数据集制成 VOC\_UA 数据集，26 820 张图片制成 UA\_CAR 数据集，使用 LabelImg 对图片中主干道车辆进行标注。UA\_CAR 数据集分为测试集、训练验证集、训练集和验证集四个子集，训练验证集占 UA\_CAR 数据集的 70%，训练集占训练验证集的 70%。VOC\_UA 数据集的处理同 UA\_CAR 数据集。为增强网络的鲁棒性，使用水平翻转实现数据增强，通过召回率和准确率对网络性能进行评价。

回归修正建议框的标签初始化为 0，若与真实目标框的最大交并比 (IOU) 大于等于 0.5，则将回归修正建议框定为目标框，对应类别标签置 1；若 IOU 介于 0.1 与 0.5 之间，则将回归修正建议框定为背景框，其标签保持不变。目标框与背景框之和为 128，目标框个数不高于 32，余下则填充背景框。

3.3 实验结果分析

3.3.1 最优阈值

网络使用 VOC\_UA 数据集训练模型，对 11 张图片进行测试，平均每张图片中主干道 6.54 个“car”。对 NMS 阈值与目标阈值进行控制变量，寻找最优参数，计算每张图平均预测个数、平均检测时间、最快检测时间（如表 1 所示）。

通过观察测试图片：若 NMS 阈值设置过大，无法过滤重叠区域较小的框，同辆车易出现复检问题；若目标阈值设置较大，则易遗漏类别概率较小且正确的预测框，引发漏检问题。经观察类别概率推理验证，当 NMS 阈值大于等于 0.5 且目标阈值小于 0.7 时，复检问题较多；当 NMS 阈值小于 0.4 且目标阈值大于 0.7 时，漏检问题突出；当 NMS 阈值等于 0.4 且目标阈值等于 0.7 时，检测效果较优。故在后续实验中，将 NMS 阈值设为 0.4，目标阈值设为 0.7。

表 1 阈值对检测的影响

Tab.1 The effect of threshold on detection

NMS 阈值 NMS threshold	目标阈值 Object threshold	平均预测个数 Average number of predictions	平均检测时间/s Average detection time	最快检测时间/s Fastest detection time
0.3	0.6	5.545	0.186	0.087
0.3	0.7	5.545	0.187	0.085
0.3	0.8	5.455	0.239	0.085
0.3	0.9	5.455	0.193	0.086
0.4	0.7	5.545	0.186	0.085
0.5	0.6	5.545	0.187	0.086
0.6	0.6	5.545	0.186	0.084

3.3.2 不同数据集测试结果比较

目标框与背景框统称为回归训练框。使用 VOC\_2007 数据集、VOC\_UA 数据集、UA\_CAR 数据集训练模型，计算预测准确率。优化后，使用 UA\_CAR 数据集训练模型预测准确率高至 98.7%，MAP 最高可达 90.8，且用时较少。实验结果如表 2 所示。

表 2 不同数据集训练模型对测试影响

Tab.2 Impact of different dataset training models on testing

数据集 Dataset	回归训练框个数 Number of regression training boxes	平均精度均值(MAP) Mean average precision	预测准确率/% Average prediction accuracy	平均检测时间/s Average detection time
VOC_2007	128	66.9	84.8	0.186
VOC_UA	128	67.3	84.8	0.186
VOC_UA	256	64.9	83.4	0.419
UA_CAR	128	90.8	95.9	0.192
UA_CAR	256	90.8	98.7	0.186

3.3.3 优化前后比较

为证实优化后 vFaster R – CNN 的有效性，使用测试集对模型进行评价。优化前 Faster R – CNN 平均精度均值为 83.5，用时 69 ms，优化后 vFaster R – CNN 平均精度均值达 90.8，耗时 85 ms，即优化后平均精度较优化前提高了 7.3 个百分点，证实了优化的有效性（见图 2、图 3）。主干道行驶车辆检测时，优化前出现定位不准确与漏检现象，而使用本文优化后的方法 vFaster R – CNN，提取更为深入抽象的特征，抗干扰性强，在一定程度上解决了错检、漏检、定位不准确等问题，但随着网络层数的增加，必导致计算量的增大，增加运行时间负担。

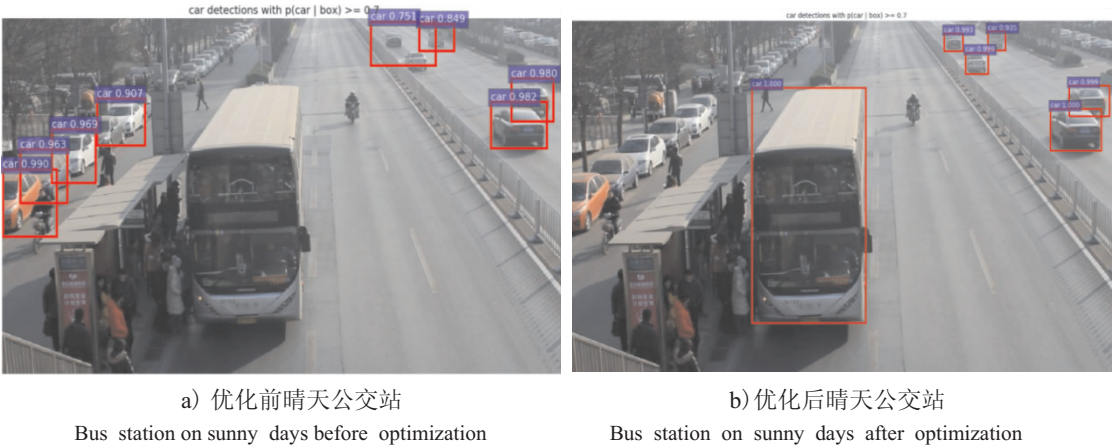


图 2 优化前后晴天效果测试

Fig.2 Performance test before and later optimization on sunny days



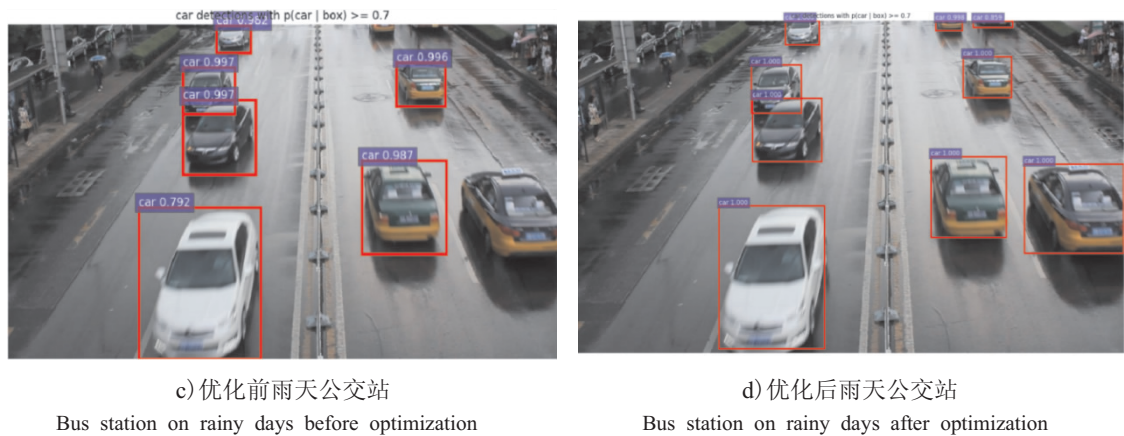


图3 优化前后两天效果测试

Fig.3 Performance test before and after optimization on rainy days

3.3.4 不同算法比较

进一步与 R - CNN, SPPnet, Fast R - CNN 等算法进行车辆检测准确率、召回率的对比, 结果如图 4 所示。随着 R - CNN 算法的不断改进, 提取区域建议框用时逐渐减少, 检测速度、准确率与召回率得以逐步提升, 从而实现对多种场景与环境下主干道行驶车辆的实时检测。本文将优化后算法与其他算法<sup>[21]</sup>的车辆检测平均精度均值进行比较, 结果如表 3 所示。

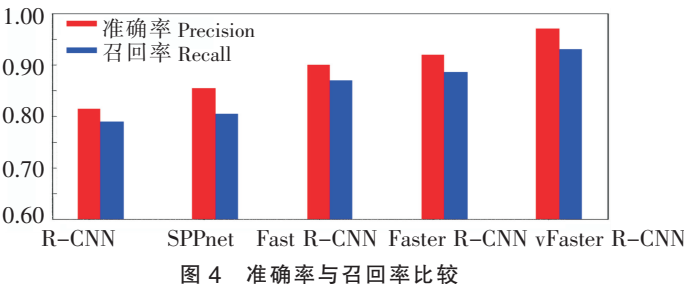


图4 准确率与召回率比较

Fig.4 Comparison of accuracy and recall rate

表3 不同算法的 MAP

Tab.3 MAP of different algorithms

算法 Algorithm	平均精度均值 MAP
YOLO	83.2
SSD	85.8
vFaster R - CNN	90.8

4 结束语

采用本文优化方法可提取更为抽象的特征, 改善了多场景与多天气条件下车辆定位、漏检、错位等问题, 提高了网络的泛化性及对环境的适应性。通过对比其他目标检测算法, 本 vFaster R - CNN 有着较高的平均精度均值 (MAP)。优化后基于本研究实验平台最快检测时间为 0.085s, 准确率达 98.7%。网络训练过程中由于目标尺寸不统一, 为获得更佳效果需训练更长时间。因此, 在接下来的研究中, 将以图片目标尺寸为着入点, 对区域平移网络再度优化处理。

[ 参考文献 ]

[1] TAYARA H, SOO K G, CHONG K T. Vehicle detection and counting in high-resolution aerial images using convolutional regression neural network [J]. IEEE Access, 2018, 6: 2220-2230. DOI:10.1109/ACCESS.2017.2782260.

[2] KOLAROW A, SCHENK S, EISENBACH M, et al. APFeI: the intelligent video analysis and surveillance system for assisting human operators [C] //2013 10th IEEE International Conference on Advanced Video and Signal Based Surveillance. Krakow: IEEE, 2013: 195-201. DOI:10.1109/AVSS.2013.6636639.

[3] PANDEY P S K, KULKARNI R. Traffic sign detection for advanced driver assistance system [C] //2018 International Conference on Advances in Communication and Computing Technology (ICACCT). Sangamner: IEEE, 2018: 182-185. DOI:10.1109/ICACCT.2018.8529455.

- [4] CHEN Z, WANG C, LUO H, et al. Vehicle detection in high-resolution aerial images based on fast sparse representation classification and multiorder feature [J]. IEEE Transactions on Intelligent Transportation Systems, 2016, 17(8): 2296-2309. DOI:10.1109/TITS.2016.2517826.
- [5] LIU K A, MATTYUS G. Fast multiclass vehicle detection on aerial images [J]. IEEE Geoscience and Remote Sensing Letters, 2015, 12(9): 1938-1942. DOI:10.1109/LGRS.2015.2439517.
- [6] ZIVKOVIC Z. Improved adaptive Gaussian mixture model for background subtraction [C] //Proceedings of the 17th International Conference on Pattern Recognition, 2004. Cambridge: IEEE, 2004, 2: 28-31. DOI:10.1109/ICPR.2004.1333992.
- [7] ZANG X, LI G, YANG G, et al. Adaptive difference modelling for background subtraction [C] //2017 IEEE Visual Communications and Image Processing (VCIP). St Petersburg, FL: IEEE, 2017: 1-4. DOI:10.1109/VCIP.2017.8305025.
- [8] LIENHART R, MAYDT J. An extended set of Haar-like features for rapid object detection [C] //International Conference on Image Processing. Rochester, NY, USA: IEEE, 2002: 1. DOI:10.1109/ICIP.2002.1038171.
- [9] LEE D. Effective Gaussian mixture learning for video background subtraction [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(5): 27-32. DOI:10.1109/TPAMI.2005.102.
- [10] ROSER M, MOOSMANN F. Classification of weather situations on single color images [C] //2008 IEEE Intelligent Vehicles Symposium. Eindhoven: IEEE, 2008: 798-803. DOI:10.1109/IVS.2008.4621205.
- [11] FELZENSZWALB P F, GIRSHICK R B, MCALLESTER D, et al. Object detection with discriminatively trained part-based models [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(9): 1627-1645. DOI:10.1109/TPAMI.2009.167.
- [12] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection [C] //2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). San Diego, CA, USA: IEEE, 2005(1): 886-893. DOI:10.1109/ICACCT.2005.8529455.
- [13] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C] //2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH: IEEE, 2014: 580-587. DOI:10.1109/CVPR.2014.81.
- [14] BUZCU I, ALATAN A A. Fisher-selective search for object detection [C] //2016 IEEE International Conference on Image Processing (ICIP). Phoenix, AZ: IEEE, 2016: 3633-3637. DOI:10.1109/ICIP.2016.7533037.
- [15] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916. DOI:10.1109/TPAMI.2015.2389824.
- [16] GIRSHICK R. Fast R-CNN [C] //2015 IEEE International Conference on Computer Vision (ICCV). Santiago: IEEE, 2015: 1440-1448. DOI:10.1109/ICCV.2015.169.
- [17] REN S, HE K, 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. DOI:10.1109/TPAMI.2016.2577031.
- [18] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C] //2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016: 779-788. DOI:10.1109/CVPR.2016.91.
- [19] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C] //2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017: 6517-6525. DOI:10.1109/CVPR.2017.690.
- [20] FELZENSZWALB P F, GIRSHICK R B, MCALLESTER D, et al. Object detection with discriminatively trained part-based models [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(9): 1627-1645. DOI:10.1109/TPAMI.2009.167.
- [21] 宋欢欢, 惠飞, 景首才, 等. 改进的 RetinaNet 模型的车辆目标检测 [J/OL]. 计算机工程与应用: 1-8 [2019-03-29]. <http://kns.cnki.net/kcms/detail/11.2127.TP.20190314.0948.022.html>.

(责任编辑 朱雪莲 英文审校 黄振坤)