

基于改进 Faster RCNN 的城市道路货车检测^①



任杰, 李钢, 赵燕姣, 姚琼辛, 田培辰

(长安大学 信息工程学院, 西安 710064)
通信作者: 任杰, E-mail: 839562424@qq.com

摘要: 针对货车利用躲避摄像头等手段在城市道路中不按规定时间、规定线路行驶, 使得车辆不能被准确识别的问题, 提出基于改进 Faster RCNN 的城市道路货车检测方法. 该方法以 Faster RCNN 为基础模型, 通过对传入主干网络的车辆图片进行卷积和池化等操作来提取特征, 其中增加特征金字塔网络 (FPN) 提升对多尺度目标检测的精度; 同时将 K-means 聚类算法应用在数据集上以获取新的锚点框; 利用 RPN (region proposal network) 生成建议框; 并使用 CIoU (complete-IoU) 损失函数代替原算法的 *smoothL1* 损失函数以提升检测车辆的精确性. 实验结果显示, 改进后的 Faster RCNN 相比原算法对货车检测的平均精度 (AP) 提高 7.2%, 召回率 (recall) 提高 6.1%, 减少了漏检的可能, 在不同场景下具有良好的检测效果.

关键词: Faster RCNN; CIoU; 特征金字塔网络; RPN; 车辆检测

引用格式: 任杰, 李钢, 赵燕姣, 姚琼辛, 田培辰. 基于改进 Faster RCNN 的城市道路货车检测. 计算机系统应用, 2022, 31(12): 316-321. <http://www.c-s-a.org.cn/1003-3254/8900.html>

Detection of Urban Trucks Based on Improved Faster RCNN

REN Jie, LI Gang, ZHAO Yan-Jiao, YAO Qiong-Xin, TIAN Pei-Chen
(School of Information Engineering, Chang'an University, Xi'an 710064, China)

Abstract: Trucks cannot be accurately identified when they do not follow the prescribed time and route on urban roads by avoiding cameras and other means. In view of this, an urban road truck detection method based on improved Faster RCNN is proposed. Features are extracted by performing convolution and pooling operations on the vehicle images passed into the backbone network. The feature pyramid network (FPN) is added to improve the accuracy of multi-scale target detection. At the same time, the K-means clustering algorithm is applied to the dataset to obtain new anchor boxes. Region proposal network (RPN) is utilized to generate proposal boxes and complete-IoU (CIoU) loss function is used for replacing the *smoothL1* loss function of the original algorithm to improve the accuracy of vehicle detection. The experimental results show that the improved Faster RCNN increases the average precision (AP) for truck detection by 7.2% and the recall by 6.1%. The improved method reduces the possibility of missed detection and has a good detection effect in different scenarios.

Key words: Faster RCNN; complete-IoU (CIoU); feature pyramid network (FPN); region proposal network (RPN); network truck detection

1 引言

近年来, 随着汽车产业的快速发展, 道路上的车辆不断增加, 政府也出台了一系列针对道路车辆行驶

政策. 因为大型货车的载货量大, 排放大, 车身过宽, 所以对于进入城市的货车也给出了规定驾驶的时间和线路, 但一些违法货车通过各种手段来逃避监控, 例如利

^① 基金项目: 中央高校学生创新实践能力提升子计划 (300102242806)

收稿时间: 2022-04-07; 修改时间: 2022-06-01; 采用时间: 2022-07-06; csa 在线出版时间: 2022-08-26

用电子狗或导航等工具在距离摄像头较远时提前预知前方有监控而调整路线,从而可能导致监控拍摄的货车画面较小,影响识别效果.这种大型货车的违法行为不但加剧了交通拥堵,也造成了空气污染,并带来一系列道路交通安全问题.近年来,在计算机技术和人工智能的推动下,目标检测技术得到快速发展,在各个领域都能够得到充分应用.其中,车辆检测在智能交通领域中得到十分广泛的应用,车辆检测技术可以精确地捕捉到每个汽车,有助于道路交通管理、车辆跟踪,从而有效地保护道路设施和人身安全.

2014年之前,车辆检测主要以传统的目标检测算法来进行研究,比如2005年Dalal等人^[1]提出的方向梯度直方图(histogram of oriented gradient, HOG)算法.这种算法利用车辆局部形状通过梯度或车辆边缘方向的密度分布检测目标.通过HOG特征与支持向量机(support vector machines, SVM)分类器的结合在车辆检测算法中取得良好的检测效果.2008年Felzenszwalb等人^[2]提出DPM(deformable parts model)算法,对于车辆检测的多视角问题,通常采用多组件(component)方法.对于车辆的形变问题,通常采用基于图结构(pictorial structure)的部件模型方法.虽然这些算法还存在精确度不高,工作量过大,效率低下等问题,但为后来基于深度学习的目标检测算法奠定了坚实的基础.

计算机视觉的快速发展使得基于深度学习的目标检测算法在检测精度和速度上都超过了传统算法.智能交通领域引起了众多学者的关注,将车辆检测与深度学习相结合成为社会研究一大热门方向.当前基于深度学习的目标检测算法主要有one-stage系列的YOLO(you only look once)^[3]、SSD(single shot multibox detector)^[4]和RetinaNet^[5]等,以及two-stage系列的Fast RCNN^[6]、Faster RCNN^[7]和Mask RCNN^[8]等.两者的区别在于two-stage算法要先在待选目标中生成一个粗略建议框,最后进行边界框回归以及分类.one-stage直接在网格中进行目标定位和分类,省略了生成建议框这一步骤.实时性上one-stage耗费时间较短,在识别精度上two-stage占据一定优势,这些算法为未来的自动驾驶、车辆跟踪提供了可靠依据.

Faster RCNN是具有代表性的一种基于卷积神经网络的目标检测模型,在使用传统的Faster RCNN检测货车时可能存在以下问题:(1)因车辆在不断行驶,而摄像头的位置固定不变,所以在检测车辆时,需要进

行多尺度检测,而传统的Faster RCNN对于小目标的检测精度较差.(2)不同区域交通视频监控的分辨率区别较大.(3)通常使用smoothL1作为损失函数,但真实的评估框检测指标是使用IoU,两者之间存在差异,当多个车辆检测框Loss相同时,车辆检测框和真实车辆位置之间的IoU可能差异较大.

本文以Faster RCNN模型为基础,做出以下几点改进:(1)将特征金字塔网络(feature pyramid networks, FPN)^[9]与ResNet50相结合作为特征提取网络用来改善对不同尺度车辆的检测效果;(2)根据数据集使用K-means聚类算法调整默认的锚点框比例;(3)原算法的损失函数为smoothL1,文献[10]提出CIoU损失函数,本文使用CIoU作为新的损失函数,可以有效提升Loss的收敛速度以及检测框与真实框的重叠程度,使车辆检测精确性得到进一步提高.

2 Faster RCNN 检测方法

以Fast RCNN模型为基础加上区域建议网络(region proposal network, RPN)候选框提取模块形成Faster RCNN模型.其中,RPN替代了Fast RCNN采用的可选择性搜索(selective search)方式来产生建议框.数据集传入到主干特征网络得到特征层,一部分特征层进入RPN,另一部分进入ROI Pooling层.RPN对目标图片生成建议框并和特征层结合传入ROI Pooling层,ROI Pooling层将所有特征图进行分区域池化调整统一尺寸传到全连接层分别进行分类预测和边界框回归预测,得到最终的预测框和目标类别.总体架构图如图1.

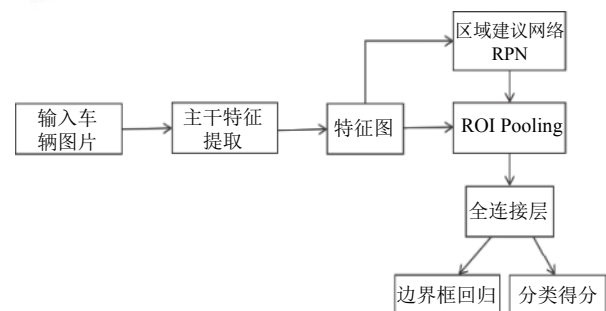


图1 Faster RCNN 总体结构图

2.1 主干特征提取网络

卷积神经网络(CNN)是让机器学习提取特征的途径,能从局部获取全局特征.另一方面,其权值共享性也减少了训练参数^[11].相较于传统特征提取方法,提高

了精度和效率。

本文基于 Faster RCNN 模型, 运用 ResNet-50-FPN 来提取车辆特征, 可以解决梯度消失问题, 训练更深的网络, 提升多尺度目标的检测精度。

2.2 RPN 区域建议网络

RPN 网络是一个全卷积神经网络. 其主要功能: 一是区分前景和背景; 二是对被检测车辆生成区域建议框。

原始图像输入到特征提取网络, 其最后一层特征图输出到 RPN 网络. 运用滑动窗口在特征图上进行滑动, 滑动窗口的大小为 3×3. 对于每个滑动窗口, 生成一组特定的锚点, 每一个位置共生成 9 个锚点框. 经过一个 3×3 的卷积对特征图进行处理, 然后再经过 1×1 的卷积, 进行一个二分类, 把有车辆的归为前景, 其他的归为背景. RPN 网络会对归为前景的图片生成候选框, 再经过一个 1×1 的卷积调整候选框 4 个位置坐标参数, 进行回归. 最后再通过非极大值抑制 (NMS) 算法^[12] 筛选多余的候选框, 计算锚点框与真实框的重叠面积^[13], 设置交并比 (IoU) 参数 0.7, 若 IoU 大于 0.7 则对候选框进行训练, 其中, $IoU = \text{Overlap} / \text{Union}$, RPN 和 IoU 结构图如图 2 所示。

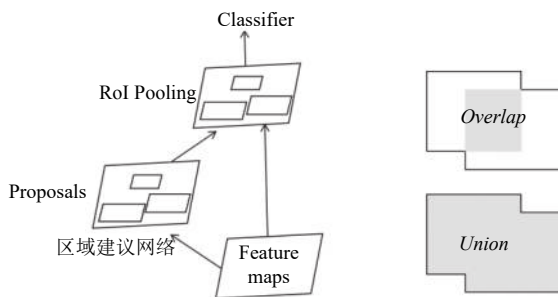


图 2 RPN 和 IoU 结构图

3 改进后的 Faster RCNN

3.1 特征金字塔网络

为了进行多尺度检测, 可以通过构建一个特征金字塔网络 (FPN)^[9] 以提升检测速度和精度, 获得更加鲁棒的语义信息. FPN 由自底向上到自顶向下的路径以及两者间的横向连接组合而成, 如图 3 所示。

自底向上是将处理过的图片输入到特征提取网络构建正向传播过程, 自顶向下是将语义更强的高层特征图进行上采样, 然后将该特征与自底向上的特征图

进行相加, 每个阶段的融合构成一个金字塔, 来提取不同尺度的特征。

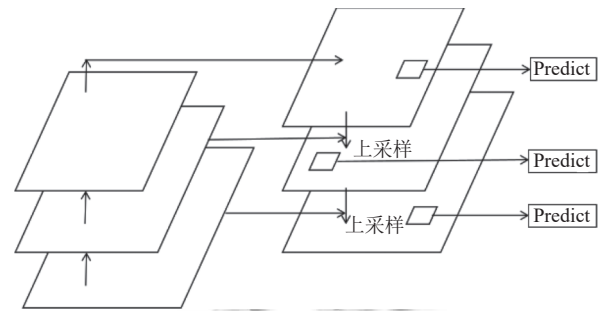


图 3 特征金字塔网络示意图

3.2 锚点框 K-means 聚类

传统的 Faster RCNN 锚点框默认有 3 种宽高比 (1:1, 1:2, 2:1) 和 3 种尺度 (128, 256, 512)^[14], 可以得出每一个位置共生成 9 个锚点框, 这些数据是根据 VOC2007 数据集所得出的参数. 由于数据集不同, 直接使用默认参数可能会导致检测精度下降, 识别货车时出现检测框和实际目标误差过大的情况. 故使用 K-means 聚类算法, 重新计算合适的锚框宽高比. 其中 K-means 是一种常用的无监督学习算法^[15], 以距离为指标, 根据不同数据集标定的目标得出新的比例和尺度参数. 本文将对已经标定好的数据集进行 K-means 聚类, 经过重新计算, 将得到的参数替代原始参数。

3.3 CIoU 损失函数

Faster RCNN 的损失函数 (loss function) 由分类损失函数和回归损失函数两部分组成^[16], 如式 (1):

$$L(p_i, t_i) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*) \quad (1)$$

其中, i 为候选框编号; p_i 为候选框分类的概率, 候选框为正样本时, p_i^* 为 1, 负样本时, p_i^* 为 0; t_i 为预测边界框的 4 个参数坐标, t_i^* 为真实边界框的 4 个参数坐标。

$L_{cls}(p_i, p_i^*)$ 表示区分前景, 背景的分类损失, 公式如式 (2):

$$L_{cls}(p_i, p_i^*) = -\log[p_i p_i^* + (1 - p_i)(1 - p_i^*)] \quad (2)$$

$L_{reg}(t_i, t_i^*)$ 表示边界框回归损失, 公式如下:

$$L_{reg}(t_i, t_i^*) = \text{smoothL1}(t_i - t_i^*) \quad (3)$$

其中, smoothL1 公式如下:

$$\text{smoothL1}(x) = \begin{cases} 0.5x^2, & |x| < 1 \\ |x| - 0.5, & \text{其他} \end{cases} \quad (4)$$

实际筛选候选框的指标使用 IoU , 当使用 $smoothL1$ 作为回归损失函数时, 多个检测框可能出现大小相同的 $Loss$, 但 IoU 存在很大差异. 为了解决这一问题, 本文使用 $CIoU$ (complete- IoU)^[10] 来代替 $smoothL1$ 作为边界框回归损失函数. $CIoU$ 计算过程如下:

常规的 IoU Loss 表示为:

$$L_{IoU} = 1 - IoU + R(B, B^{gt}) \quad (5)$$

其中, B 为预测框, B^{gt} 为真实框, $R(B, B^{gt})$ 为预测框和真实框的惩罚项.

计算 $CIoU$ 惩罚项:

$$R_{CIoU} = \frac{p^2(b, b^{gt})}{c^2} + \alpha v \quad (6)$$

其中,

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2 \quad (7)$$

$$\alpha = \frac{v}{(1 - IoU) + v} \quad (8)$$

如图 4 所示, 其中 b 和 b^{gt} 表示预测框和真实框的中心点, p 表示欧氏距离, c 是覆盖预测框和真实框的最小矩形的对角线距离. α 表示平衡比例参数, v 表示预测框和真实框的长宽比的一致性.

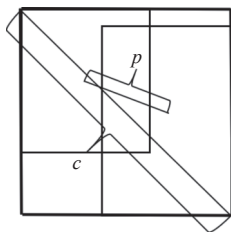


图 4 $CIoU$ 构造图

$CIoU$ 损失函数表示为:

$$L_{CIoU} = 1 - IoU + \frac{p^2(b, b^{gt})}{c^2} + \alpha v \quad (9)$$

相对于 $smoothL1$ 作为损失函数, $CIoU$ 可以直接让预测框和真实框之间的距离最小化. 如果出现两个边界框包含、水平、垂直的情况, $CIoU$ 使预测框可以快速回归, 并且还考虑了边界框的纵横比, 使得目标检测框的精度进一步提高.

4 实验分析

4.1 实验环境配置

实验所使用的环境配置为: CPU: i7-9700, GPU:

GTX1660TI, 显存 6 GB, 运行内存 64 GB, CUDA 10.0, CUDNN 7.4.1, 实验环境: Torch 1.2.0.

4.2 数据集

实验使用 KITTI 数据集^[17], 其中包含小汽车、面包车、货车、行人等, 共计 5600 张图片. 对数据集进行筛选, 只选择存在货车的图片, KITTI 数据集默认分辨率为 1242×375 , 根据实际监控情况, 将分辨率调整为 700×500 、 960×740 进行训练, 并在数据集中添加 500 张实际场景中监控视频下的不同分辨率的图像, 以增加模型的泛化性. 选取其中 90% 为训练集, 10% 为测试集.

4.3 实验结果及分析

利用召回率 ($recall$) 和平均精度 ($average\ precision, AP$)^[18] 来评估改进后的算法. 其中, TP (true positive) 为预测为正的样本的数量, FP (false positive) 为预测为正的负样本的数量, FN (false negative) 为预测为负的正样本的数量, TN (true negative) 为预测为负的负样本的数量.

$$recall = \frac{TP}{TP + FN} \quad (10)$$

$$precision = \frac{TP}{TP + FP} \quad (11)$$

$$AP = \frac{p_1 + p_2 + \dots + p_n}{n} \quad (12)$$

(1) 模型改进前后对比

为了验证改进模型后的效果, 分别对改进前后的模型进行训练和验证, 学习率设置为 $1E-4$; 锚点框宽高比设置为 (0.3, 0.7, 1, 1.5), 实验结果如表 1 所示.

表 1 改进前后的两种模型检测结果 (%)

检测方法	召回率	AP50
Faster RCNN	83.47	85.36
改进的Faster RCNN	89.59	92.61

表 1 的结果显示, 改进后的 Faster RCNN 模型在货车召回率和 AP50 指标上都优于传统 Faster RCNN. 其中目标召回率提升约 6.1%, AP50 提升了约 7.2%. 改进后的模型既有效地提升了货车目标检测的精度, 也降低了漏检的产生.

(2) 消融实验

为进一步评估所作改进的有效性, 以 Faster RCNN 模型为基础进行消融实验. 表中的结果显示, 添加特征金字塔网络对 AP50 与召回率分别提升 2.7% 和 4.2%;

改进损失函数对 AP50 与召回率分别提升约 2.5% 和 3.5%。表明两种改进方案均对模型性能有一定提升。消融实验结果如表 2 所示。

表 2 消融实验结果 (%)

方案	检测方法	召回率	AP50
A	ResNet50+smoothL1	83.47	85.36
B	ResNet50-FPN+smoothL1	87.67	88.06
C	ResNet50+CIoU	87.02	87.93
D	ResNet50-FPN+CIoU	89.59	92.61

(3) 多种模型对比

当前主流目标检测模型主要分为 one-stage 和 two-stage 两个方向, 将 SSD、YOLO、原 Faster RCNN 以及改进后的 Faster RCNN 利用本文数据集进行训练、测试。将各个模型分别从精确度、召回率、实时性上进行对比, 实验结果如表 3 所示。

表 3 多种模型检测结果

检测方法	召回率 (%)	AP50 (%)	时间 (ms)
SSD	77.68	78.55	82
YOLOv4	80.2	83.0	75
Faster RCNN	83.47	85.36	191
改进的 Faster RCNN	89.59	92.61	196

YOLO 的训练与检测在一个单独网络中进行, 没有生成建议框的过程。输入目标图像, 经过主干网络就能得到目标的位置及所属类别。这一优势使得 YOLO 模型在低性能设备中可以顺利运行, 在检测速度上占据明显优势。SSD 结合 Faster RCNN 模型的优点, 引入预先设定好的边界框, 但没有生成建议框这一步骤, 使得 SSD 的检测速度依然较快。改进后的 Faster RCNN 模型由表 3 可以看出, 相对于其他模型, 在检测精度上具有一定优势。在检测速度上, 本文方法略次于其他方法。由于本文应用于对城市道路货车的检测, 货车相对行驶速度较慢, 所以在实际场景中本文方法具有可行性。

(4) 复杂条件下的货车检测效果

从测试集中选出不同场景下的图像来进行验证模型检测效果, 结果如图 5 所示, 可以看出改进后的模型在可视环境较差, 货车细节不充分, 距离较远, 背景复杂等情况下均有较好的检测效果。实验结果表明, 改进后的模型满足在不同场景下的需要, 在对货车的检测中具有较高的精确性和鲁棒性。

5 结语

针对城市道路的货车检测精度不高、漏检的问题,

本文以 Faster RCNN 模型为基础, 使用 CIoU 作为边界框损失函数, 并构建特征金字塔网络, 结合 KITTI 数据集和实际监控视频进行训练, 使得改进后的 Faster RCNN 无论在召回率还是在平均精度上对于城市道路货车检测效果都得到显著优化。因为本文模型相对复杂, 检测实时性不高, 后续应进一步提高模型的轻量化。

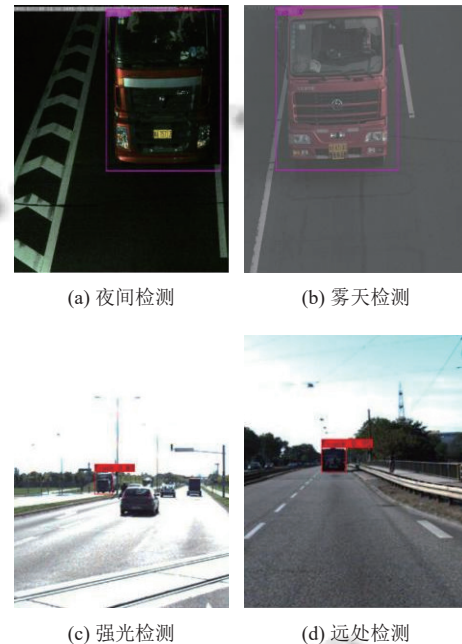


图 5 复杂条件下的货车检测效果

参考文献

- 1 Dalal N, Triggs B. Histograms of oriented gradients for human detection. Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego: IEEE, 2005. 886–893.
- 2 Felzenszwalb P, McAllester D, Ramanan D. A discriminatively trained, multiscale, deformable part model. Proceedings of 2008 IEEE Conference on Computer Vision and Pattern Recognition. Anchorage: IEEE, 2008. 1–8.
- 3 Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 779–788.
- 4 Liu W, Anguelov D, Erhan D, et al. SSD: Single shot MultiBox detector. Proceedings of the 14th European Conference on Computer Vision. Amsterdam: Springer, 2016. 21–37.

- 5 Tran VP, Tran TS, Lee HJ, *et al.* One stage detector (RetinaNet)-based crack detection for asphalt pavements considering pavement distresses and surface objects. *Journal of Civil Structural Health Monitoring*, 2021, 11: 205–222. [doi: [10.1007/s13349-020-00447-8](https://doi.org/10.1007/s13349-020-00447-8)]
- 6 Girshick R. Fast R-CNN. *Proceedings of 2015 IEEE International Conference on Computer Vision*. Santiago: IEEE, 2015. 1440–1448.
- 7 Ren SQ, He KM, Girshick R, *et al.* Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137–1149. [doi: [10.1109/TPAMI.2016.2577031](https://doi.org/10.1109/TPAMI.2016.2577031)]
- 8 He KM, Gkioxari G, Dollár P, *et al.* Mask R-CNN. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(2): 386–397. [doi: [10.1109/TPAMI.2018.2844175](https://doi.org/10.1109/TPAMI.2018.2844175)]
- 9 Lin TY, Dollár P, Girshick R, *et al.* Feature pyramid networks for object detection. *Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu: IEEE, 2017. 936–944.
- 10 Zheng ZH, Wang P, Liu W, *et al.* Distance-IoU loss: Faster and better learning for bounding box regression. *Proceedings of the 34th AAAI Conference on Artificial Intelligence*. New York: AAAI, 2020. 12993–13000.
- 11 周飞燕, 金林鹏, 董军. 卷积神经网络研究综述. *计算机学报*, 2017, 40(6): 1229–1251. [doi: [10.11897/SP.J.1016.2017.01229](https://doi.org/10.11897/SP.J.1016.2017.01229)]
- 12 Neubeck A, van Gool L. Efficient non-maximum suppression. *Proceedings of the 18th International Conference on Pattern Recognition*. Hong Kong: IEEE, 2006. 850–855.
- 13 康庄, 杨杰, 李桂兰, 等. 基于改进 YOLOv3 的站口行人检测方法. *铁道科学与工程学报*, 2021, 18(1): 55–63. [doi: [10.19713/j.cnki.43-1423/u.T20200236](https://doi.org/10.19713/j.cnki.43-1423/u.T20200236)]
- 14 徐守坤, 王雅如, 顾玉宛, 等. 基于改进 Faster RCNN 的安全帽佩戴检测研究. *计算机应用研究*, 2020, 37(3): 901–905.
- 15 李荟娆. K-means 聚类方法的改进及其应用 [硕士学位论文]. 哈尔滨: 东北农业大学, 2014.
- 16 张金, 冯涛. 基于改进的 Faster RCNN 的手势识别. *信息通信*, 2019, (1): 44–46.
- 17 Fan QF, Brown L, Smith J. A closer look at Faster R-CNN for vehicle detection. *Proceedings of 2016 IEEE Intelligent Vehicles Symposium (IV)*. Gothenburg: IEEE, 2016. 124–129.
- 18 徐义臻, 贺鹏, 任东, 等. 基于改进 Faster RCNN 的木材运输车辆检测. *计算机应用*, 2020, 40(S1): 209–214. [doi: [10.11772/j.issn.1001-9081.2019081431](https://doi.org/10.11772/j.issn.1001-9081.2019081431)]

(校对责编: 牛欣悦)