计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2020,29(10):267−273 [doi: 10.15888/j.cnki.csa.007566]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于公路监控视频的车辆检测和分类①

曹富奎[1], 白　天[1], 许晓珑[2]

[1](中国科学技术大学 软件学院, 合肥 230027)
[2](福建省 厦门市公路局 信息处, 厦门 361008)
通讯作者: 曹富奎, E-mail: sa517006@mail.ustc.edu.cn

摘　要: 在学习了已有的检测与分类算法以后, 设计了一种将改进的高斯混合模型 (GMM) 与分类网络 (GoogLeNet) 融合的方案用于车辆的检测和分类. 针对高斯混合模型存在模型初始化速度慢和计算复杂的问题, 改进了初始化模型的算法提升初始化效率. 运用五帧差法做车辆初提取, 在提取到的车辆区域上运用高斯混合模型获得车辆图片, 把五帧差法和高斯混合模型结合起来减小了建模的区域, 提升了检测速度, 提高了系统实时性. 最后使用 GoogLeNet 对车辆分类. 实验证明相较于现有的车辆检测分类方法, 本文所提方法在检测速度和分类准确性上都有很大提升, 满足了现实场景下对监控视频的车辆检测和分类的实时性要求.

关键词: 车辆检测; 高斯混合模型; 目标识别; 分类网络; 实时性

引用格式: 曹富奎,白天,许晓珑.基于公路监控视频的车辆检测和分类.计算机系统应用,2020,29(10):267–273. http://www.c-s-a.org.cn/1003-3254/7566.html

Vehicle Detection and Classification Based on Highway Monitoring Video

CAO Fu-Kui[1], BAI Tian[1], XU Xiao-Long[2]

[1](School of Software Engineering, University of Science and Technology of China, Hefei 230027, China)
[2](Information Office, Highway Administration of Xiamen, Fujian Province, Xiamen 361008, China)

Abstract: Having studied the existing detection and classification algorithms, we design a scheme of fusion of improved Gaussian Mixture Model (GMM) and classification network (GoogLeNet) for vehicle detection and classification. In view of the inaccurate initialization and complex computation of GMM, we improve the algorithm of initialization models to increase the initialization efficiency. The five-frame difference method is used to execute the preliminary vehicle extraction. In the extracted vehicle area, GMM is used to get vehicle images, the five-frame difference method is combined with GMM to reduce the area of modeling and to increase the speed of vehicle detection and improve the real-time performance of the system. At last, we use GoogLeNet to execute the vehicle classification. The results show that the proposed methods have greatly improved the detection speed and recognition accuracy, and satisfy the real-time requirement of vehicle detection and recognition for surveillance video in real scenario.

Key words: vehicle detection; GMM; object recognition; classification network; real-time performance

中国经济高速发展带来机动车数目的急剧增加, 造成道路交通中堵车和交通事故现象呈上涨形势, 改善交通治理策略逐渐提上日程. 随着科技的进步, 诞生了智能交通理念, 即通过分析道路交通信息得出科学的治理策略, 车辆检测作为智能交通工作的基础成为了研究的热点. 最早有帧差法[1] 和光流法[2] 进行车辆检测, 这两种方法原理简洁, 运算快速, 但是在应对复杂的道路状况时会出现检测偏差. 后来出现了基于统

① 基金项目: 福建省交通运输厅科技发展项目 (201431)
Foundation item: Science and Technology Development Project of Fujian Provincial Department of Transportation (201431)
收稿时间: 2020-01-15; 修改时间: 2020-02-13, 2020-02-25; 采用时间: 2020-03-11; csa 在线出版时间: 2020-09-30

计的模型方法, 运用最多的是基于多态高斯分布的背景模型法[3], 该方法能够很好的应对各种变化, 能在复杂的道路场景里提取出运动车辆的图片. 随着现代深度学习理论的产生和成熟, 围绕着目标检测问题诞生了很多检测算法, 主要有: RCNN 算法[4]、Mask-RCNN 算法[5]、Fast-RCNN 算法[6]、SPP-Net 算法[7]、Faster-RCNN 算法[8]、R-FCN 算法[9], Singh[10] 改进了 R-FCN 算法, 实现了多类别的检测, Hu 等[11] 创建的的关系网络刷新了检测的精确度, Cai 等[12] 创建了级联 RCNN 将其扩展到多段. 深度学习技术在目标检测中获得了不俗的成绩, 但是却太依赖于高性能硬件, 训练模型也需要花费大量时间, 因此本文在车辆检测阶段不采用深度学习的方法, 选择高斯混合模型法来进行车辆检测, 传统的高斯混合模型法由于初始化模型速度慢、计算量大在实时性上表现不佳, 本文对高斯混合模型法做了改进, 降低了模型计算量, 提高了检测实时性.

车辆分类属于目标识别问题. 目标识别问题解决方案主要有两种, 一种是采用手动设计图像特征, 如 HOG 特征[13], 然后使用分类器做分类. 另外一种是使用深度学习技术做分类, 主要是运用卷积神经网络 (CNN) 做目标识别. 2012 年 AlexNet 网络[14] 模型在图像识别的应用中取得了巨大成功, 使得 CNN 成为了目标识别的主流方法, 因此本文采用深度学习技术, 使用 GoogLeNet[15] 分类网络对车辆进行分类.

1 基于高斯混合背景模型的目标检测

1.1 建立和维护背景模型

在实际的道路场景里有很多不可控的变量会导致场景背景发生改变, 给背景模型的建立和维护增加了难度. 目前主要有以下 3 个方面引起背景发生变化:

(1) 光照突然变化. 例如突然转换光源, 场景背景像素发生剧烈变化, 导致检测结果出现差错.

(2) 物体阴影. 运动物体由于光线投影产生阴影, 阴影属于背景但是却始终伴随着运动目标, 导致背景和目标的界限变得模糊, 增加了检测的难度.

(3) 动态的背景. 现实场景中的背景区域存在运动变化的现象, 可能是周期性的运动, 也可能是不规则的运动, 比如树叶的不规则晃动, 信号灯的规律闪烁.

为解决上述 3 方面带来的背景变化问题, Stauffer 等[16] 首次利用高斯混合模型进行视频监控的前景检测, 取得了较好的效果. 此后, 高斯混合模型经过不断发展和完善, 成为目标检测领域的经典模型.

1.2 高斯混合模型

在真实的监控环境下, 道路两侧摄像头拍摄到的视频影像中会夹杂很多的噪声, 一般我们认为从监控设备引入的噪声服从正态分布, 对于输入的每一视频帧, 按照式 (1) 进行预处理.

$$h(x,y)=\frac{1}{2\pi\sigma^2}\mathrm{e}^{-\frac{x^2+y^2}{2\sigma^2}} \tag{1}$$

用 k 个高斯分布组合表示每个背景像素, 假设 t 时刻的背景像素为x_t, 由 k 个高斯分布如式 (2) 表示:

$$f(x_t)=\sum_{i=1}^{k} w_{i,t}*\eta(x_t,\mu_{i,t},\Sigma_{i,t}) \tag{2}$$

其中, $u_{i,t}$是第 i 个高斯分布在 t 时刻的均值向量. 将像素值x_t与组合分布中的 k 个高斯分布做比较, 当x_t和$u_{i,t}$满足式 (3) 时, 认为x_t属于其中一个高斯分布.

$$\left|x_t-u_{i,t}\right|<\delta*\Sigma_{i,t} \tag{3}$$

δ通常取 3. 若x_t属于其中的一个高斯分布, 则将该高斯分布的参数按式 (4) 更改, 剩下 k–1 个高斯分布的参数则按式 (5) 衰减, 其中α表示模型的学习率. 与x_t相符的高斯分布的$u_{i,t}$和 $\sigma_{i,t}{}^2$ 按照式 (6) 和式 (7) 更新, 其他 k–1 个高斯分布则保持不变, $\rho=\alpha/w_{i,t}$表示参数的学习率.

$$w_{i,t}=(1-\alpha)w_{i,t-1}+\alpha \tag{4}$$

$$w_{i,t}=(1-\alpha)w_{i,t-1} \tag{5}$$

$$u_{i,t}=(1-\rho)u_{i,t-1}+\rho x_t \tag{6}$$

$$\sigma_{i,t}^2=(1+\rho)\sigma_{i,t-1}^2+\rho(x_t-u_{i,t-1})^T(x_t-u_{i,t-1}) \tag{7}$$

$p=w_{i,t}/\sigma_{i,t}$, 我们将 k 个高斯分布按 P 值排序, 选择前 B 个分布作为背景模型, 如式 (8) 所示:

$$B=\arg\min_b\left\{\Sigma_t^b w_{i,t}>V\right\} \tag{8}$$

式中, V 通常取 0.75.

建立好了背景模型, 当新的视频帧输入系统, 将像素值x_t与背景模型中的 B 个高斯分布进行比较, 如果x_t属于其中的一个分布则认为其是背景, 反之则认为是运动目标.

1.3 改进模型初始化算法

高斯混合模型法初始化背景模型时最常用的方法是用视频开始的帧图像作为背景来初始化模型的参数值, 当视频的开始部分只有背景时, 该方法能很好的进行下去, 但如果视频一开始就出现了运动目标, 该方法

会把运动目标设置为背景, 故本文提出新的初始化方法, 使用视频前 4,9,16,$\cdots$,$(K+1)^2$帧图像, K 是高斯混合模型中包含的高斯分布的个数, 将$(K+1)^2$帧图像的像素值作为第 K 个高斯分布的平均值, 再初始化方差和相同的权值. 具体步骤如下:

(1) 将视频输入系统, 并提取前 4,9,16,$\cdots$,$(K+1)^2$帧图像, 做预处理降低图像中的噪声, 记作: $f_1, f_2, f_3, \cdots, f_k$;

(2) 将f_i的像素值作为第 i 个高斯分布的均值, 初始化一个方差, 并赋权值为 1/N;

(3) 输入下一帧图像, 用建好的背景模型检测车辆.

试验分析: 图 1 是用传统高斯混合模型初始化的背景模型图像, 可以看出使用传统方法得到的背景模型图像模糊难辨, 与真实的场景背景有着十分大的差别, 基于这样的背景模型检测车辆, 检测结果往往会不准确. 图 2 是改进方法的初始化的背景模型图像, 相比于图 1, 这幅背景模型图像显的十分清晰, 能够正确的表现场景的背景, 由于初始化的背景模型的图像是由多个高斯分布的权重和均值之积组合得到的, 造成了图像中出现车辆重影的现象, 不影响车辆的检测.

图 1 传统高斯混合模型背景图像

图 2 改进的高斯混合模型背景图像

1.4 改进建模区域

通过实际观察可以看出, 在视频帧中运动车辆只占了图像的小部分区域, 图像的大部分区域都是没有发生变化的背景区域, 为这些背景区域的像素建立模型是不必要的, 由此带来的计算过程也是浪费计算资源, 故本文提出新的建模方法, 通过对视频帧的初提取, 先把包含汽车的区域提取出来, 然后使用高斯混合模型法在提取到的图像上进行建模, 缩小了建模区域, 减少了建立背景模型过程耗费的时间, 提高了车辆检测的实时性.

对视频帧进行初提取得到包含车辆的区域, 初提取的精度影响着后续检测结果的精度, 当初提取出现漏检时, 后续的检测也一定会漏检, 因此初提取要能正确提取出包含车辆的区域, 同时初提取的速度要快, 不能耗费过多时间, 因此选择帧差法做车辆的初提取. 常用的帧差法有两种, 分别是两帧差法和三帧差法, 但两帧差法会出现“重影”和“空洞”现象, 在实际应用中面对复杂场景无法正确检测出目标, 三帧差法可以消除“重影”, 但是依然存在“空洞”现象和部分物体轮廓不连续.

通过对本文采集的视频的研究, 发现在大部分情况下, 车辆从在视频里出现到占据视频的中心, 大多是经过 5 帧, 因此本文在发现这种现象后, 提出用相邻 5 帧图像进行运动车辆的初提取. 工作流程如下:

第 1 步. 假设当前帧为f_t, 将f_t与f_{t+2}按像素点逐一相减, 得到二值图像 B1.

第 2 步. 将f_{t+2}与f_{t+4}按像素点逐一相减, 得到二值图像 B2.

第 3 步. 对 B1 和 B2 取并集, 在经过形态学处理和填充, 得到车辆区域 B3.

分别使用两帧差法、三帧差法和本文提的五帧差法对拍摄的视频进行前景提取, 如图 3 所示.

通过实验结果可以直观的看出, 本文的方法相较于传统的两帧差法和三帧差法, 在前景提取精度上有了很大的提升, 并且检测速度快, 非常适合应用于对前景的初提取.

使用五帧差法得到提取好的车辆区域图像, 在初提取的车辆图像上运用高斯混合模型法获得最终的车辆图像. 改进后的结合五帧差法的高斯混合模型的车辆检测算法的工作流程如图 4.

(a) 原始图像 (b) 两帧差法

(c) 三帧差法 (d) 五帧差法

图 3 差分法仿真图

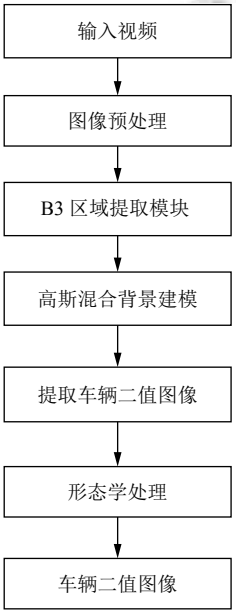

图 4 改进的高斯混合模型算法

改进后的基于高斯混合模型的检测方法步骤如下:

(1) 使用五帧差法对输入的视频进行前景区域初提取, 得到前景的二值图像, 通过形态学操作得到前景区域.

(2) 使用改进的背景模型初始化方法, 对前景区域的像素建立背景模型.

(3) 建立完背景模型, 用当前帧减去背景图片得到目标的二值图像, 并更新模型相关参数.

(4) 对得到目标二值图像进行形态学操作, 得出最终的目标区域.

(5) 对下一帧重复上述步骤, 直至算法完结.

2 GoogLeNet 分类网络

谷歌公司 2014 年提出来 GoogLeNet 分类网络, 并在当年的图像分类竞赛中取得了第一名, 刷新了图像识别的最好成绩. GoogLeNet 网络运用了模块化设计, 提出了新的网络单元结构 Inception, 网络整体是由多个网络单元结构按照一定次序组合而成, 图 5 是 Inception 模块的设计图.

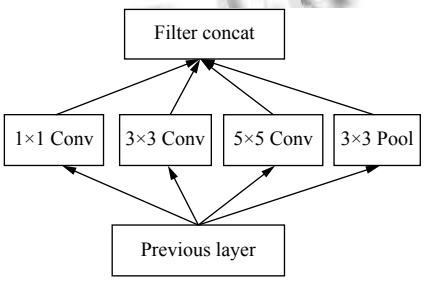

图 5 Inception 模块

不同的卷积神经网络会选择不同大小的卷积核构建卷积层, 每一卷积层会使用同一大小卷积核来对特征图进行卷积计算, Inception 模块最大的创新是在一个卷积层中采用了多个卷积核, 不同卷积核会提取不同的图像特征, 在模块最后将这些特征有机融合形成下一层的特征图, 这样融合多特征的设计延深了网络宽度, 提高了对图像识别的精度, 但计算复杂度也随之增加, 降低了识别速度. 为了解决计算成本高的问题, 在模块的设计上做了创新, 图 6 是改进后的模块. 新模块最大的特点是对 1×1 卷积核的巧妙运用, 在进行卷积计算和池化操作之前先用 1×1 卷积核来降低特征图维度, 这样在下一层的卷积计算中运算次数会降低一个量级, 从而有效的降低了计算成本. 经过模块有序组合而成的 GoogLeNet 网络层数达到 22 层, 相比于之前的网络模型有大幅度的提升.

3 实 验

本文所提的车辆检测和车型分类方法基于 OpenCV 和 TensorFlow 实现, 试验所用的视频包含了不同天气和光照的复杂场景, 实验环境的信息如表 1.

3.1 实验系统架构

根据交通视频把车辆型号分成 4 类分别为: car、bus、motor 和 truck, 系统要实现的就是检测到交通视

频里的车辆然后识别出车辆是上述 4 种类型中的哪一种, 为了实现这一功能, 我们设计的系统工作流程为: 先通过系统的运动车辆检测模块将视频里的运动车辆检测出来, 然后人工标注好图片的类别, 训练目标分类模块, 使用该模块执行分类任务, 系统架构如图 7.

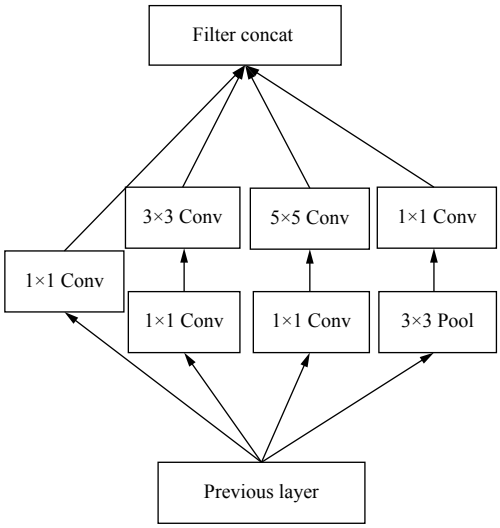

图 6 改进的 Inception 模块

表 1 实验环境信息

类型	名称	性能
处理器	Intel Xeon Silver	2.10 GHz
内存	DDR3L	32 GB
显卡	Tesla V100	16 GB
系统	Ubuntu16.04	—
开发框架	TensorFlow	—

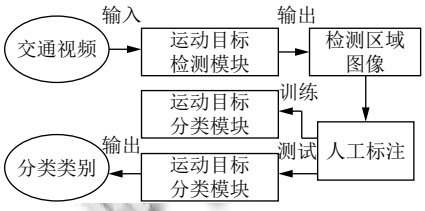

图 7 系统架构图

3.2 试验数据集

本文以厦门市各路段摄像头拍摄的监控视频为系统实验数据, 所有的视频均为现实场景下的真实监控视频, 通过车辆检测模块总共截取了 50 000 张不同类型的车辆图片. 其中各种车型对应的数量见表 2.

表 2 车辆检测模块检测结果

类型	car	bus	motor	truck	总计
数量(张)	20 000	16 000	4000	10 000	50 000

将得到的全部图片按照 1:1 划分为测试集和训练试集, 训练集中 car 10 000 张, bus 8000 张, motor 2000 张, truck 5000 张. 由表 2 可知, 由于 motor 数量比其他车型的数量少了很多, 如果直接用原始的数据集训练分类模型得到的模型分类准确性会有很大的偏差, 所以要进行数据集扩充, 将 2000 张 motor 图片使用图像处理技术增加到 10 000 张, 对其他 3 种车的图片执行相同的操作. 扩充后的数据集如表 3 所示. 4 类车型样本图片如图 8 所示.

表 3 运动目标分类模块数据集

车型	训练集	测试集	类测比
car	10 000	10 000	2/5
bus	10 000	8000	8/25
motor	10 000	2000	2/25
truck	10 000	5000	1/5
合计	40 000	25 000	1

图 8 4 种车型样本

3.3 实验结果及分析

针对基于监控视频的车辆检测和分类的问题, 本文改进了高斯混合模型算法, 并用 GoogLeNet 网络做车辆分类, 在实验中设计实现了算法. 由于实验效果需要对比, 在车辆检测阶段, 本文使用 Faster R-CNN 和 YOLO 对相同的视频做了实验[17], 实验结果对比如表 4 所示.

表 4 车辆检测结果对比

算法	准确率(%)	召回率(%)	速度(f/s)
本文算法	97.4	89.8	45
YOLO	89.3	81.0	40
Faster R-CNN	88.9	83.5	10

表 4 给出了检测阶段的数据, 结果表明相比于深度学习的方法, 本文所提方法在检测准确率和检测速

度上都更加优秀.

在车型分类阶段, 本文训练了 GoogLeNet 模型进行分类, 同时也训练了经典的 AlexNet 模型用作对比[18]. 在实验中我们对原始数据集进行了扩充, 但得到的训练集也只有 40 000 张图片, 数据集过小会导致模型过拟合现象, 因此在实验中采用了迁移学习[19] 技术, 将经典模型卷积层的参数直接迁移到试验模型上作为模型初始化参数. 使用 ImageNet 数据集训练好的 GoogLeNet 模型和 AlexNet 模型能够很好的提取图像的特征, 因此在实验中用训练好的模型的参数作为试验模型的初始化参数, 然后使用自采集的训练集 finetune[16] 试验模型. 图 9 为 4 种车型的分类结果, 从置信度上可以看出分类的准确性非常高.

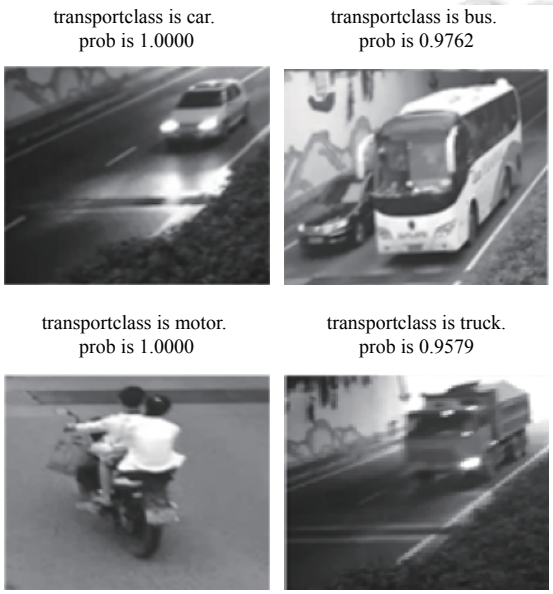

图 9 车型分类结果

表 5 展示了 GoogLeNet 模型和 AlexNet 模型在分类准确率上的对比, 可以看出 GoogLeNet 不仅在分类准确率上比 AlexNet 稍高一些, 而且在运算时间上比 AlexNet 降低了 22%. 试验证明, GoogLeNet 分类网络在基于监控视频的车型分类上具有很好的实用性.

表 5 车型分类结果对比

车型	GoogLeNet		AlexNet	
	分类/测试	正确率	分类/测试	正确率
car	9923/10 000	0.992	9891/10 000	0.989
bus	7712/8000	0.964	7696/8000	0.962
motor	1962/2000	0.981	1960/2000	0.980
truck	4635/5000	0.927	4675/5000	0.935
合计	24 232/25 000	0.9692	24 222/25 000	0.9689
时间(s)	845.3		1 083.7	

4 结 语

本文针对公路交通监控视频中车辆检测和车型分类的具体应用场景, 改进了高斯混合模型建立背景模型的方法, 提升了系统的实时性, 并和 GoogLeNet 分类网络结合起来. 通过建立背景模型和背景相减法得到运动车辆, 然后使用训练好的分类模型对车辆进行分类. 在具体场景的实验中, 该算法取得了较快的检测速度和分类准确率, 比现有方法更加适合公路监控视频这一应用场景, 具有很高的实用价值.

参考文献

1 高凯亮, 覃团发, 王逸之, 等. 一种基于帧差法与背景减法的运动目标检测新方法. 电讯技术, 2011, 51(10): 86–91. [doi: 10.3969/j.issn.1001-893x.2011.10.018]
2 袁国武, 陈志强, 龚健, 等. 一种结合光流法与三帧差分法的运动目标检测算法. 小型微型计算机系统, 2013, 34(3): 668–671. [doi: 10.3969/j.issn.1000-1220.2013.03.047]
3 王永忠, 梁彦, 潘泉, 等. 基于自适应混合高斯模型的时空背景建模. 自动化学报, 2009, 35(4): 371–378.
4 Girshick R, Donahue J, Darrell T, *et al*. Rich feature hierarchies for accurate object detection and semantic segmentation. Proceeding of the 2014 IEEE conference on Computer Vision and Pattern Recognition. Columbus, OH, USA. 2014. 580–587.
5 He KM, Gkioxari G, Dollar P, *et al*. Mask R-CNN. Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy. 2017. 2980–2988.
6 Girshick R. Fast R-CNN. arXiv preprint arXiv: 1504.08083, 2015.
7 He KM, Zhang XY, Ren SQ, *et al*. Spatial pyramid pooling in deep convolutional networks for visual recognition. Proceedings of the 13th European Conference on Computer Vision. Zurich, Switzerland. 2014. 346–361.
8 Ren SQ, He KM, Girshick R, Sun J. 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]
9 Dai JF, Li Y, He KM, *et al*. R-FCN: Object detection via region-based fully convolutional networks. Proceedings of the 30th International Conference on Neural Information Processing Systems. Red Hook, NY, USA. 2016. 379–387.
10 Singh B, Li HD, Sharma A, *et al*. R-FCN-3000 at 30fps: Decoupling detection and classification. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern

Recognition. Salt Lake City, UT, USA. 2018. 1081–1090.
11 Hu H, Gu JY, Zhang Z, *et al*. Relation networks for object detection. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. 2018. 3588–3597.
12 Cai ZW, Vasconcelos N. Cascade R-CNN: Delving into high quality object detection. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. 2018. 6154–6162.
13 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, CA, USA. 2005. 886–893.
14 Krizhevsky A, Sutskever I. Hinton GE. ImageNet classification with deep convolutional neural networks. Proceedings of the 25th International Conference on Neural Information Processing Systems. Red Hook, NY, USA. 2012. 1097–1105.
15 Szegedy C, Liu W, Jia YQ, *et al*. Going deeper with convolutions. Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA. 2015. 1–9.
16 Stauffer C, Grimson WEL. Adaptive background mixture models for real-time tracking. Proceedings of 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Fort Collins, FL, USA. 1999. 246–252.
17 陈伟星, 白天, 许晓珑. 基于公路监控视频的车辆检测和识别. 信息技术与网络安全, 2018, 37(11): 64–68.
18 胡鹏, 白天, 许晓珑. 一种快速的车型识别方法. 信息技术与网络安全, 2018, 37(5): 41–45.
19 Pan SJ, Yang Q. A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(10): 1345–1359. [doi: 10.1109/TKDE.2009.191]