

基于 YOLOV3 改进的实时车辆检测方法

李汉冰*, 徐春阳, 胡超超

江苏大学汽车与交通工程学院, 江苏 镇江 212013

摘要 针对原始 YOLOV3 目标检测算法在车辆检测任务中存在的实时性不高的问题, 提出了一种改进的车辆检测模型。该模型使用反残差网络作为基础特征提取层, 以减少参数量, 降低计算复杂度, 解决梯度消失和梯度爆炸问题。并且使用组归一化降低批量大小对模型准确性的影响, 同时用软化非极大值抑制降低漏检率, 使用 Focal-loss 改进损失函数, 使模型在训练时聚焦于难分类样本。改进后的模型参数量为 YOLOV3 的 36.23%, 每帧检测时间较 YOLOV3 降低了 13.8 ms, 平均类别精度提高了 1.15%。结果表明, 本文算法兼顾实时性和准确性, 为车辆的实时性检测提供参考。

关键词 机器视觉; 车辆; 目标检测; YOLOV3; 反残差网络; 实时检测

中图分类号 TP391.4

文献标志码 A

doi: 10.3788/LOP57.101507

Improved Real-Time Vehicle Detection Method Based on YOLOV3

Li Hanbing*, Xu Chunyang, Hu Chaochao

School of Automotive and Traffic Engineering, Jiangsu University, Zhenjiang, Jiangsu 212013, China

Abstract Aiming at the problem of low real-time performance of original YOLOV3 target detection algorithm in vehicle detection tasks, this paper proposes an improved vehicle detection model. In order to reduce the number of parameters, reduce the computational complexity, and solve the problem of gradient disappearance and gradient explosion, the model uses the inverted residual network as the basic feature extraction layer. In addition, group normalization is used to reduce the impact of batch size on the accuracy of the model. At the same time, softening non maximum suppression is used to reduce the rate of missed detection. Finally, the Focal-loss is used to improve the loss function so that the model focuses on the difficult-to-classify samples in the process of training. The parameter amount of the improved model is 36.23% of YOLOV3 model. The detection time per frame is reduced by 13.8 ms compared with YOLOV3, and the average category accuracy is improved by 1.15%. The results show that the proposed algorithm ensures both real-time performance and accuracy, and providing a reference for real-time detection of vehicles.

Key words machine vision; vehicle; target detection; YOLOV3; inverted residual network; real-time detection

OCIS codes 150.0155; 150.1135; 100.4996

1 引言

随着城市智能交通的不断完善, 对车辆的快速检测成为了自动驾驶中一个非常重要的工作。车辆检测是高级驾驶辅助系统(ADAS)的重要组成部分, 在减少交通事故方面有着十分重要的作用。

传统的车辆检测算法在图像预处理后, 使用滑动窗口对整幅图像进行遍历, 判断车辆可能出现的

位置, 接着人工选取某个特征, 例如, 方向梯度直方图特征^[1](HOG)、加速稳健特征^[2](SURF)等, 通过这些特征对目标进行识别。例如, 黎向锋等^[3]提出的基于 Haar 特征的车辆检测算法, 周行等^[4]提出的基于 HOG 特征和局部纹理特征(LBP)特征融合检测算法。传统车辆检测算法主要存在两方面的问题: 1) 基于滑动窗口的区域提取时间复杂度较高, 冗余窗口太多, 严重影响特征提取和分类的速度; 2) 由

收稿日期: 2019-08-05; 修回日期: 2019-09-18; 录用日期: 2019-10-22

基金项目: 国家自然科学基金(51275212)

* E-mail: 1340733996@qq.com

于目标的形态、光照、背景等多样性,人工选择的特征鲁棒性较差。

随着深度学习的快速发展,卷积神经网络在计算机视觉领域取得了巨大的成功,出现了许多优秀的目标检测算法。目前的目标检测框架一般分为两种:1)基于回归的 One-stage 算法(SSD^[5-6]、YOLO^[7-9]等);2)基于候选区域提取的 Two-stage 算法(R-CNN^[10]、Fast R-CNN^[11]、Faster R-CNN^[12]等)。基于深度学习,李浩等^[13]提出了利用卷积神经网络进行多角度车辆检测,降低了复杂场景下背景噪声的影响;程雅慧等^[14]提出了基于组合目标框提取结构的扩张卷积神经网络,提高了车辆检测的准确性;王得成等^[15]提出了神经网络融合彩色与深度图像的车辆检测算法。这些方法旨在提高车辆检测准确性,但是忽略了实时性,然而在道路上进行车辆检测时,需要在保证一定准确性的同时尽可能地增加其检测速度,及时为车主提供周围车辆信息。

本文在 YOLOV3 的基础上进行改进,改进后的模型使用反残差网络^[16]构建基础特征提取层,反残差网络包含深度可分离卷积^[17]和残差结构^[18](ResNet),使用深度可分离卷积替换标准卷积,降低模型参数量,结合残差结构,解决深层网络训练过程中梯度消失和梯度爆炸的问题。此外使用组归一化^[19](GN)降低批量尺寸较小引起的预测误差,使用软化非极大值抑制^[20](SoftNMS)代替非极大值抑制(NMS),将交并比大于一定阈值的框得分降低,而不是直接置 0,降低模型对于重叠物体的漏检率,最后通过 Focal-loss^[21]改进损失函数,增加难分类样本在损失中的权重,提高模型检测准确度。实

验结果表明,改进后的算法降低了模型的运算复杂度及参数,提高了车辆检测速度及准确性。

2 改进的车辆检测模型

2.1 网络结构

本文用反残差网络改进 YOLOV3 的基础卷积模块,反残差网络如图 1 所示,包含深度可分离卷积和残差结构,不同于标准卷积,深度可分离卷积实现了通道和区域的分离,打破输出通道数与卷积核大小之间的相互连接作用。深度可分离卷积包括深度卷积和逐点卷积,深度卷积负责滤波,逐点卷积就是 1×1 卷积,负责转换通道。深度可分离卷积相对于标准卷积参数量压缩率为

$$\frac{D_K \cdot D_K \cdot M \cdot D_F \cdot D_F + M \cdot N \cdot D_F \cdot D_F}{D_K \cdot D_K \cdot M \cdot N \cdot D_F \cdot D_F} = \frac{1}{N} + \frac{1}{D_K^2}, \quad (1)$$

式中: D_K 表示卷积核尺寸; D_F 表示输入的特征尺寸; M 为输入通道数; N 为输出通道数;分母为标准卷积参数量;分子为深度可分离卷积参数量。由于深度卷积的计算量较小,因此在深度卷积前添加 1×1 卷积以提升通道数,获得更多特征,接着通过融合残差网络,使用跳跃连接,构成残差块,解决梯度消失和梯度爆炸问题,使模型加深网络的同时保证良好的性能。此外,当特征通道数较低时,激活函数 Relu 会使某个维度的张量值为 0,导致特征信息损失,因此在降维后用线性变换层代替通道数较少层的 Relu 激活函数。图中,conv 表示常规卷积,DWconv 表示深度卷积, stride 表示卷积步长,Concat 表示通道融合。

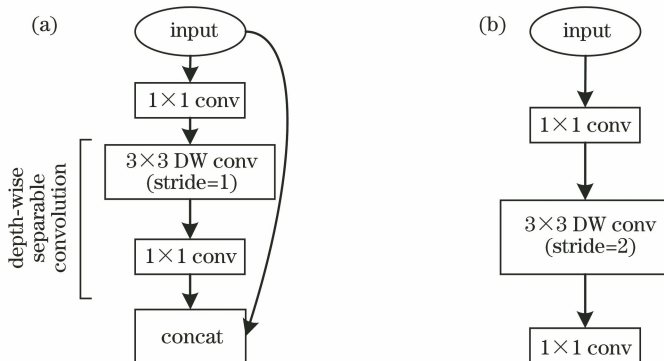


图 1 反残差网络。(a)步长为 1;(b)步长为 2

Fig. 1 Inverted residual network. (a) Stride is 1; (b) stride is 2

如图 2 所示,高层特征语义信息比较丰富,能够准确地识别目标,感受野较大,适合预测大物体,而底

层特征具有高分辨率,目标位置信息更加准确,感受野较小,适合预测小物体,因此本文采用特征金字塔

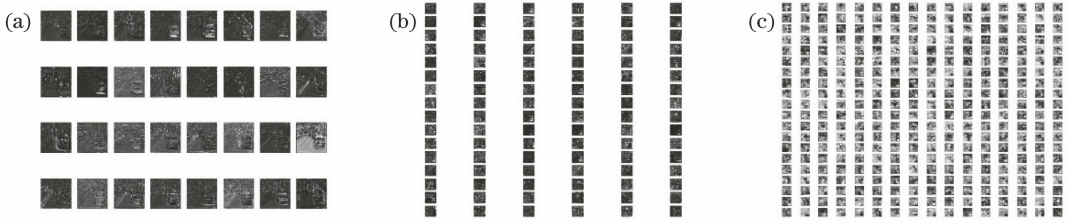


图 2 网络最后三层不同尺寸特征图。(a) 52×52;(b) 26×26;(c) 13×13

Fig. 2 Feature maps of different sizes in the last three layers of network. (a) 52×52; (b) 26×26; (c) 13×13

结构,融合候选目标的上下文信息,将 13×13 特征图上采样与 26×26 特征图进行特征融合,26×26 特征图上采样与 52×52 特征图进行特征融合,在融合后采用 3×3 的卷积核对每个融合结果进行卷积,消除

上采样的混叠效应,增加模型对尺寸变化的鲁棒性。

改进后的模型结构如图 3 所示,图中 upsampling 表示上采样,conv set 表示连续卷积模块。

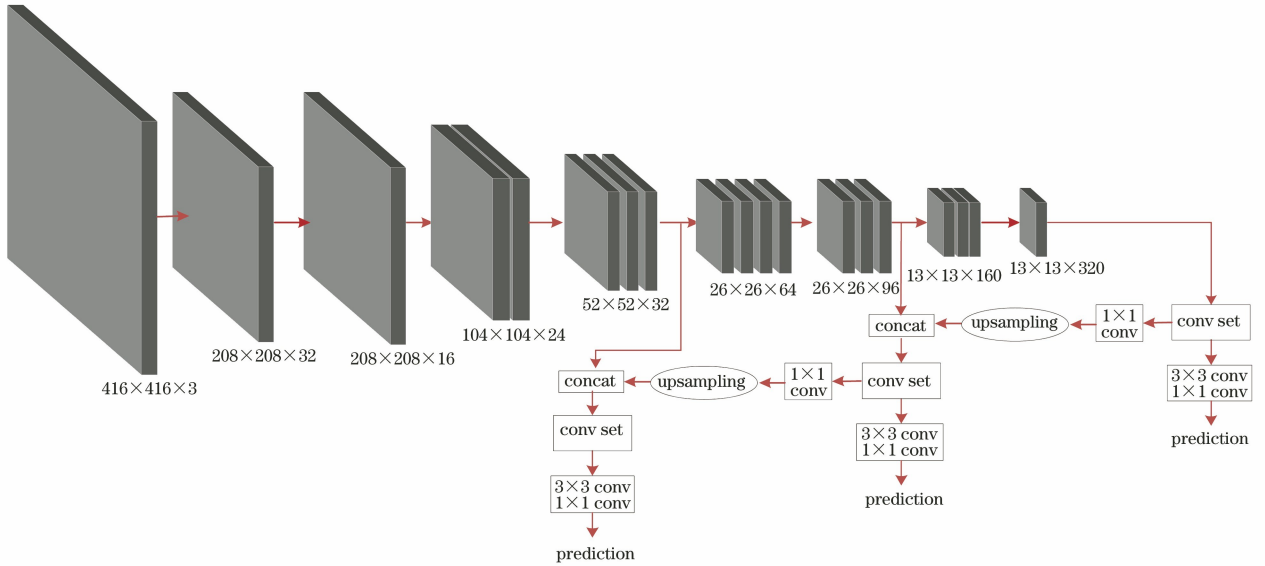


图 3 改进后的网络结构

Fig. 3 Improved network structure

2.2 组归一化

目标检测中常用的归一化方法是批量归一化 (BN),但 BN 也存在一定的局限性,BN 在 batch 的维度做归一化,而这个维度是变化的,训练过程中通过滑动平均计算数据的均值和方差,测试时直接调用训练集已经计算好的数据,但当训练集和测试集分布不同时,会导致误差产生。在目标检测中,由于输入的图像数据较大,为了节省显存,batchsize 只能设为较小值,而较小的 batchsize 会导致计算出的均值和方差不准确,使得 BN 性能下降。本文采用 GN 代替 BN,在通道维度进行归一化,公式为

$$S_i = \left\{ k \mid k_N = i_N, \left\lfloor \frac{k_C}{C/G} \right\rfloor = \left\lfloor \frac{i_C}{C/G} \right\rfloor \right\}, \quad (2)$$

式中:G 表示组的数量,为预定义的超参数;C/G 是

每组的通道数; $\left\lfloor \frac{k_C}{C/G} \right\rfloor = \left\lfloor \frac{i_C}{C/G} \right\rfloor$ 表示索引 i 和 k 在同一组通道中,每组通道沿 C 轴按顺序存储,GN 沿 (H, W) 轴和沿着的一组 C/G 通道计算均值和方差, H 和 W 是空间高度和宽度轴。

2.3 软化非极大值抑制

由于车辆在行驶过程中,车载摄像头拍摄的图像中包含许多重叠的车,而普通的非极大值抑制会将重叠度较大的预测框得分置为 0,容易造成车辆的漏检,因此本文采用软化非极大值抑制,利用高斯加权来平滑函数,对重叠度较高的预测框降低其得分,而不是将其直接置 0,降低模型的漏检率,改进后的 NMS 公式为

$$s_i = s_i \exp \left[- \frac{\text{iou}(M, b_i)^2}{\sigma} \right], \quad (3)$$

式中: σ 为常数; s_i 表示检测框的得分; M 表示得分

最高的框; b_i 表示第*i*个检测框。

2.4 损失函数

在计算损失过程中,YOLOV3将模型所有的预测框分为正样本(与标注框区域交并比 >0.5)与负样本(与标注框区域交并比 <0.4),一般图片中目标占的比例都远小于背景占的比例,因此,正负样本数

相差较大,并且大部分都是易分类的负样本,此时的损失函数在大量简单样本的迭代过程中比较缓慢且可能无法优化至最优。本文针对难易分类样本数不平衡问题,使用Focal-loss改进损失函数,Focal-loss是在标准交叉熵损失基础上修改得到的,交叉熵损失函数为

$$E_{CE} = -y \log y' - (1-y) \log(1-y') = \begin{cases} -\log y', & y=1 \\ -\log(1-y'), & y=0 \end{cases}, \quad (4)$$

式中: y' 是预测的概率,在 $0\sim 1$ 之间; y 为真实标签,可以看出,在交叉熵损失中正样本的输出概率越大损失越小,负样本的输出概率越小损失越小。修改后的公式为

$$L_{FL} = \begin{cases} (1-y')^\gamma [-y \log y' - (1-y) \log(1-y')] & y=1 \\ y'^\gamma [-y \log y' - (1-y) \log(1-y')] & y=0 \end{cases} = \begin{cases} -(1-y')^\gamma \log y', & y=1 \\ -y'^\gamma \log(1-y'), & y=0 \end{cases}, \quad (5)$$

式中: γ 为调节参数,值设为2,用于调节简单样本权重降低的速率,使得模型在训练时更注重难分类样本。用Focal-loss改进后的损失函数为

$$\begin{aligned} L = & \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B l_{ij}^{\text{obj}} [-t_{x_i} \log \hat{t}_{x_i} - (1-t_{x_i}) \log(1-\hat{t}_{x_i})] + \\ & \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B l_{ij}^{\text{obj}} [-t_{y_i} \log \hat{t}_{y_i} - (1-t_{y_i}) \log(1-\hat{t}_{y_i})] + \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B l_{ij}^{\text{obj}} [(t_{w_i} - \hat{t}_{w_i})^2 + (t_{h_i} - \hat{t}_{h_i})^2] + \\ & \lambda_{\text{obj}} \sum_{i=0}^{S^2} \sum_{j=0}^B l_{ij}^{\text{obj}} [-c_i \log \hat{c}_i - (1-c_i) \log(1-\hat{c}_i)] + \\ & \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^B l_{ij}^{\text{noobj}} [-c_i (1-\hat{c}_i)^\gamma \log \hat{c}_i - (1-c_i) (\hat{c}_i)^\gamma \log(1-\hat{c}_i)] + \\ & \lambda_{\text{class}} \sum_{i=0}^{S^2} l_i^{\text{obj}} \sum_{c \in \text{classes}} \{-p_i(c) \log \hat{p}_i(c) - [1-p_i(c)] \log [1-\hat{p}_i(c)]\}, \end{aligned} \quad (6)$$

式中: S^2 为输出特征图的全部网格单元; B 为每个网格预测的边框数; l_{ij}^{obj} 、 l_{ij}^{noobj} 判断第*i*个网格中第*j*个边界框是否负责物体的预测; l_{ij}^{obj} 判断物体的中心是否落在第*i*个网格; \hat{t}_{x_i} 、 \hat{t}_{y_i} 、 \hat{t}_{w_i} 、 \hat{t}_{h_i} 为预测的边界框相对位置; t_{x_i} 、 t_{y_i} 、 t_{w_i} 、 t_{h_i} 为真实框位置参数; c_i 为真实框置信度; \hat{c}_i 为预测的边界框置信度; $p_i(c)$ 为真实框类别概率; $\hat{p}_i(c)$ 预测的边界框类别概率; λ_{coord} 为坐标损失在总损失中的权重,值设为5; λ_{obj} 为正样本在置信度损失中的权重,值设为1; λ_{noobj} 为负样本在置信度损失中的权重,值设为0.5; λ_{class} 为类别损失在总损失中的权重,值设为1。

3 车辆检测实验

本次实验平台配置为:操作系统 Ubuntu16.04, Intel Core i7-8700@3.2 GHz 六核处理器,16 G 内存,Nvidia GTX 1080Ti 11GB GPU,深度学习框架

选用 Tensorflow。

3.1 数据集及预处理

本文采用的数据集是由不同的地区采集的5000张图片组成,以及从VOC数据集中抽取包含车辆的2500张图片,其中绝大部分样本中包含多个检测目标。本文采用在线数据增强的方式来扩充数据集,在送入网络中进行训练前,图片都会进行随机数据增强,包括光照变化、旋转、翻转、剪切,平移等,使传入网络的图片场景具有多样化。

3.2 训练

本文采用迁移学习的方法训练模型,先使用ImageNet数据集对模型的backbone进行预训练,再用本文数据集对整个模型进行微调式训练。本文使用随机梯度下降法优化损失函数,学习率设为0.001,动量参数为0.9,权重衰减因子为0.005,总迭代次数为2000。在训练过程中,模型会对每一个预测框输出一个是否包含物体的置信度预测,用来区

分物体和背景,当某个预测框与标注框的重叠度大于其他所有的预测框时,则增加其置信度损失权重,对每个标注框只分配一个预测框与其对应,如果某个预测框没有任何标注框与其对应,则对边框位置的回归和类别的预测没有做任何贡献,模型会试图减小其置信度,从而降低其在置信度损失中的权重。

4 结果与分析

本文选用的评价指标为平均类别精度(mAP)和检测时间,平均类别精度由P-R曲线求出,P代表Precision即准确率,R代表Recall即召回率,两者的计算公式为

$$P_{\text{precision}} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FP}}}, \quad (7)$$

$$R_{\text{recall}} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FN}}}, \quad (8)$$

式中: N_{TP} 为正确划分为正样本的个数; N_{FP} 为错误划分为正样本的个数; N_{FN} 为错误划分为负样本的个数。

P-R曲线是以召回率和准确率作为横、纵坐标的曲线,P-R曲线围起来的面积就是类别精度

表2 不同改进策略对mAP的影响

Table 2 Influence of different improvement strategies on mAP

Improvement strategy	Inverted residuals	GN	SoftNMS	Focal-loss
Change of mAP / %	-3.17	1.15	1.39	1.78

接着对比了不同算法的P-R曲线,如图4所示,可以看出,SSD和YOLOV2的召回率和精确率都低于YOLOV3和本文算法,而在相同召回率的情况下,改进后的模型精确率高于YOLOV3模型,表明模型的漏检率和误检率都低于YOLOV3模型,具有更好的检测效果。

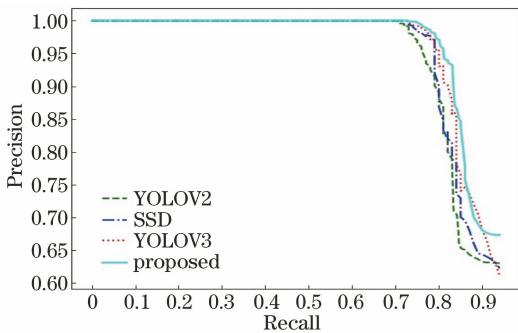


图4 不同模型的P-R曲线

Fig. 4 P-R curves for different models

为了进一步验证本文算法的有效性,对比了不同算法的检测时间和平均类别精度,实验结果如表3所示,本文算法通过特征金字塔(FPN)结构融

(AP),mAP就是所有类的AP平均值。

首先,进行网络层数和模型大小对比,结果列于表1,可以看出,本文改进的模型网络层数高于SSD和YOLOV2,低于改进前的YOLOV3模型,由于使用了反残差网络,网络层数较多并不会引起梯度消失和梯度爆炸,且模型尺寸均小于其他3个模型,参数量仅为改进前的36.23%,有效地降低了模型的计算复杂度,提升了模型的检测速度。

表1 不同模型的网络层数及大小对比

Table 1 Comparison of network layers and sizes of different models

Model	Layer	Parameter	Size /MB
SSD	88	27188676	103
YOLOV2	75	50983561	194
YOLOV3	256	61587112	235
Proposed	223	22314120	85.6

其次,研究不同改进策略对mAP的影响,如表2所示,反残差网络的使用虽然减少了网络层数及参数,提升了模型的实时性,但mAP降低了3.17%,而GN、SoftNMS和Focal-loss提升了模型检测的准确性。

合了上下文信息,增加了候选框的数量,同时使用非软化极大值抑制和Focal-loss,有效地降低了模型的漏检率,改进后的模型平均类别精度相对于YOLOV3模型提升了1.15%。改进后的模型检测单张图片所需时间为28.5ms,低于YOLOV3模型的42.3ms。综合考虑,本文改进后的模型兼顾实时性和准确性,能够及时为高级辅助驾驶系统提供周围环境感知信息。

表3 不同模型检测结果对比

Table 3 Comparison of test results of different models

Model	mAP / %	Time /ms
SSD	89.88	48.8
YOLOV2	89.60	30.2
YOLOV3	91.91	42.3
Proposed	93.06	28.5

图5为改进后的模型在不同场景下的检测效果对比,图5(a)为原图,光照强度有明显的变化,并且图中的车辆有不同程度的遮挡,图5(b)为YOLOV3检测结果,图5(c)为本文改进后的模型检测结果,从第二行和第四行的局部放大图可以看

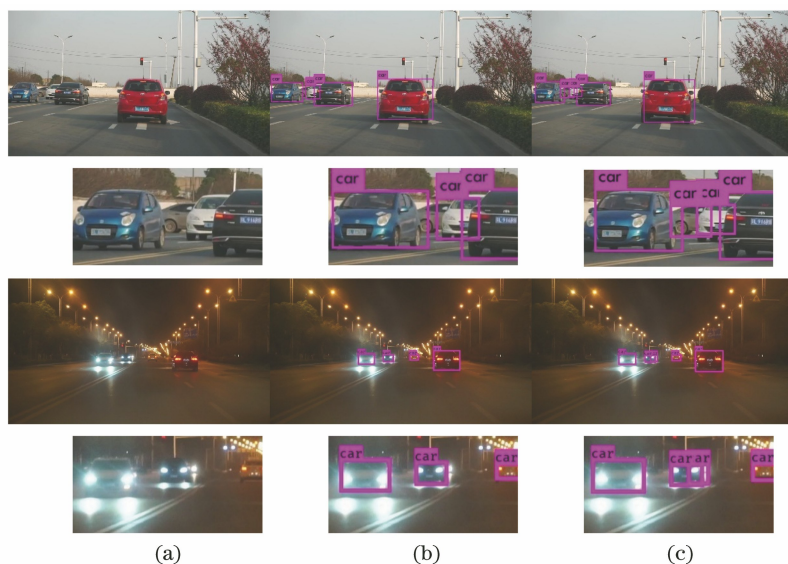


图5 模型在不同场景下的检测结果。(a)原图;(b)YOLOV3检测结果;(c)本文改进后的模型检测结果
Fig. 5 Model detection results in different scenarios. (a) Original images; (b) detection results of YOLOV3;
(c) detection results of improved model

出,在一些较小的车辆目标以及遮挡严重的场景,本文算法具有更好的鲁棒性。

5 结 论

为了提高车辆检测的实时性和鲁棒性,本文使用反残差网络作为基础卷积结构,降低特征提取过程中的运算量,解决网络加深引起的梯度消失和梯度爆炸问题,并且通过FPN结构提升对小目标物体的检测能力,使用GN和SoftNMS进一步提高模型检测准确性,最后通过Focal-loss修改损失函数,减少简单样本对模型的影响,使模型训练更加精准有效。实验结果表明,改进后的模型尺寸更小,速度更快,检测性能更好。本文工作可以扩展到通用的目标检测模型,也为后续的车辆测距、监测车辆违法行为、交通管理等奠定了基础。

参 考 文 献

- [1] 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: 886-893.
- [2] Bay H, Ess A, Tuytelaars T, et al. Speeded-up robust features (SURF) [J]. Computer Vision and Image Understanding, 2008, 110(3): 346-359.
- [3] Li X F, Zhao W K, Dou X Y, et al. Vehicle detection algorithm based on improved AdaBoost and haar [J]. Measurement & Control Technology,

2019, 38(2): 42-45.

黎向锋, 赵伟康, 豆心愿, 等. 基于改进 AdaBoost+Haar 的车辆检测算法[J]. 测控技术, 2019, 38(2): 42-45.

- [4] Zhou X, Chen S R. A vehicle detection algorithm based on multi-feature fusion and cascade classifier [J]. Modern Computer, 2018(17): 38-43.
周行, 陈淑荣. 一种基于多特征融合级联分类器的车辆检测算法[J]. 现代计算机, 2018(17): 38-43.
- [5] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[M] // Computer Vision-ECCV 2016. Cham: Springer International Publishing, 2016: 21-37.
- [6] Song Y L, Pang Y W. Backbone network for object detection task [J]. Laser & Optoelectronics Progress, 2020, 57(4): 041021.
宋雅麟, 庞彦伟. 针对目标检测任务的基础网络[J]. 激光与光电子学进展, 2020, 57(4): 041021.
- [7] 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), June 27-30, 2016. Las Vegas, NV, USA. IEEE, 2016: 779-788.
- [8] Redmon J, Farhadi A. YOLOV3: an incremental improvement[J]. arXiv preprint arXiv: 1804.02767, 2018.
- [9] Ju M R, Luo H B, Wang Z B, et al. Improved YOLOV3 algorithm and its application in small target detection [J]. Acta Optica Sinica, 2019, 39 (7):

- 0715004.
- 鞠默然, 罗海波, 王仲博, 等. 改进的 YOLOV3 算法及其在小目标检测中的应用[J]. 光学学报, 2019, 39(7): 0715004.
- [10] 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, June 23-28, 2014. Columbus, OH, USA. IEEE, 2014: 580-587.
- [11] Girshick R. Fast R-CNN [C] // 2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015. Santiago, Chile. IEEE, 2015: 1440-1448.
- [12] Ren S Q, He K M, 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] Li H, Zhang Y S, Lian J, et al. A multi-aspect method for vehicle dynamic detection based on deep learning[J]. Journal of Transport Information and Safety, 2017, 35(5): 37-44.
李浩, 张运胜, 连捷, 等. 基于深度学习的多角度车辆动态检测方法[J]. 交通信息与安全, 2017, 35(5): 37-44.
- [14] Cheng Y H, Cai X, Feng R. Dilated fully convolutional network with grouped proposals for vehicle detection [J]. Computer Systems & Applications, 2019, 28(1): 107-112.
程雅慧, 蔡烜, 冯瑞. 面向车辆检测的扩张全卷积神经网络[J]. 计算机系统应用, 2019, 28(1): 107-112.
- [15] Wang D C, Chen X N, Zhao F, et al. Vehicle detection algorithm based on convolutional neural network and RGB-D images [J]. Laser & Optoelectronics Progress, 2019, 56(18): 181003.
王得成, 陈向宁, 赵峰, 等. 基于卷积神经网络和 RGB-D 图像的车辆检测算法[J]. 激光与光电子学进展, 2019, 56(18): 181003.
- [16] Sandler M, Howard A, Zhu M L, et al. MobileNetV2: inverted residuals and linear bottlenecks [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018. Salt Lake City, UT. IEEE, 2018: 4510-4520.
- [17] Chollet F. Xception: deep learning with depthwise separable convolutions [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017. Honolulu, HI. IEEE, 2017: 1800-1807.
- [18] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016. Las Vegas, NV, USA. IEEE, 2016: 770-778.
- [19] Wu Y X, He K M. Group normalization [J]. arXiv preprint arXiv: 1803.08494, 2018.
- [20] Bodla N, Singh B, Chellappa R, et al. Soft-NMS: improving object detection with one line of code [C] // 2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017. Venice. IEEE, 2017.
- [21] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 318-327.