

DOI:10.13364/j.issn.1672-6510.20220180

## 基于改进 YOLOv4 的车辆目标检测算法

苏 静, 刘伊智

(天津科技大学人工智能学院, 天津 300457)

**摘 要:** 针对公路上高速行驶的车辆检测常常存在错检、漏检的问题,对 YOLOv4 算法进行改进优化. 首先,将 CSPDarknet53-tiny 作为主干特征提取网络,并对网络中的 ResBolck\_body 模块中的残差边与 GhostModule 模块结合代替原始特征网络 CSPDarknet53,从而简化网络结构,同时提高其检测精度;然后,将原算法中的 SPPNet 模块结构替换为 ASPPNet,增大网络感受野,降低参数计算量,使模型能够在保持精准度的同时更加轻量化;最后,将注意力机制模块 SENet 结构嵌入特征金字塔 PANet 的两个不同位置,使其可对不同重要程度的特征进行相应处理. 在 BDD100K 数据集实验中,原 YOLOv4 算法训练后得到的模型的平均精度(AP)为 88.27%,改进优化后的 YOLOv4 模型 AP 为 90.96%,改进后的 YOLOv4 算法相比原算法 AP 提高了 2.69%. 在实际真实场景数据集实验中,改进优化后的 YOLOv4 算法比原算法 AP 提高了 3.31%. 实验结果表明,本文提出的方法可以有效提高 YOLOv4 算法对车辆目标检测的精度.

**关键词:** 目标检测; 车辆检测; YOLOv4; 注意力机制; 特征提取网络

中图分类号: TP389.1

文献标志码: A

文章编号: 1672-6510(2023)03-0061-07

## Vehicle Target Detection Algorithm Based on Improved YOLOv4

SU Jing, LIU Yizhi

(College of Artificial Intelligence, Tianjin University of Science & Technology, Tianjin 300457, China)

**Abstract:** Aiming at the problems of misdetection and missed detection, which often exist in the detection of high-speed vehicles on the highway, the YOLOv4 algorithm is improved and optimized in this article. Firstly, CSPDarknet53-tiny is used as its main feature extraction network, and the residual edge of ResBolck\_body module in the network is combined with GhostModule module to replace CSPDarknet53, so as to simplify the network structure and improve its detection accuracy. Then, the structure of SPPNet module in the original algorithm is replaced by ASPPNet to increase the network receptive field and reduce the calculation of parameters, so that the model could be more lightweight while maintaining the accuracy. Finally, the structure of the attention mechanism module SENet is embedded in two different positions of the feature pyramid PANet, so that it can deal with features of different importance. The AP value of the model obtained by the original YOLOv4 algorithm after training is 88.27%, and the AP value of the improved and optimized YOLOv4 model is 90.96%. The AP value of the improved YOLOv4 algorithm is 2.69% higher than that of the original algorithm. In actual real-world dataset experiments, the AP value of the improved YOLOv4 algorithm is 3.31% higher than that of the original algorithm. Experimental results show that the proposed method can effectively improve the accuracy of YOLOv4 algorithm for vehicle target detection.

**Key words:** object detection; vehicle detection; YOLOv4; attention mechanism; feature extraction network

随着城市的不断发展,人口流动更加频繁,堵车等交通拥挤问题也逐渐得到重视,发展智慧交通成为

了重中之重. 车辆的目标检测是实现城市智慧交通的关键技术. 视觉目标检测是计算机视觉的经典任

收稿日期: 2022-07-17; 修回日期: 2022-10-04

基金项目: 天津市自然科学基金项目(19JCYBJC15300); 天津市教委项目(2018KJ1105)

作者简介: 苏 静(1979—),女,北京人,副教授, [sujing@tust.edu.cn](mailto:sujing@tust.edu.cn)

务,旨在定位图像中存在物体的位置并识别物体的具体类别. 目标检测是许多计算机视觉任务及相关应用的基础与前提,直接决定相关视觉任务及应用的性能<sup>[1]</sup>.

传统的目标检测算法以手工设计特征为主,这些特征的泛化能力弱,对复杂场景的性能表现较差<sup>[2]</sup>,并且存在速度缓慢和准确率低等问题. 现如今基于深度学习的目标检测算法已经成为主流,此类目标检测算法主要分为两类:单阶段(one stage)算法和两阶段(two stage)算法. 两阶段算法的主要代表为 R-CNN<sup>[3]</sup>系列算法,单阶段算法最有代表性的是 YOLO 算法和 SSD 算法<sup>[4]</sup>,其中 YOLO 系列算法更是进行了多次迭代更新,后续分别提出了 YOLOv2<sup>[5]</sup>、YOLOv3<sup>[6]</sup>、YOLOv4 等算法. 在城市道路交通中,车辆行驶速度一般在 30 ~ 120 km/h 之间,在本文的实验部分,对 Faster RCNN 与 YOLOv4 进行了帧每秒(FPS)的对比实验,Faster RCNN 的平均 FPS 只有 32.4,而 YOLOv4 原算法 FPS 为 58.1. 由实验可知 Faser RCNN 算法检测速度缓慢,不适用于道路上高速行驶车辆的检测. YOLO 系列算法具有检测速度快、准确率高优点,可以更好地适用于公路上行驶车辆的目标检测. YOLO 系列的第四代算法 YOLOv4 具有良好的检测性能与检测速度,但是对于一些目标也存在识别率不高等问题.

李挺等<sup>[7]</sup>将 MobileNetv2 作为 YOLOv4 的主干网络,加入了 Bottom-up 连接,并融合了 CBAM 注意力机制,在分类网络中加入了 Inception 结构. 虽然在行人检测方面的结果有所提高,但是对行驶车辆的检测却会出现漏检等问题. 陈洋等<sup>[8]</sup>在 PANet 中加入 CBAM 模块,同时使用加强的 k 均值聚类(k-means)对目标真实框进行聚类,使算法检测锚框更适合于训练集,提高了船舰图像检测的精度,但是却增大了网络参数,使得模型训练缓慢,不够轻量. 王滢暄等<sup>[9]</sup>提出了多标签检测方法,建立其约束关系,并提出图像拼接检测方法,以提升网络的运行效率.

因此,本文提出改进 YOLOv4 车辆检测算法:将主干网络改进为轻量型的 CSPDarknet53-tiny 和 GhostModule 模块结合的网络结构进行特征提取,形成 YOLOv4-Ghost-tiny,提高模型的运行速度,减少网络的规模大小;将 SPPNet 替换为 ASPPNet<sup>[10]</sup>,增大网络感受野并减少模型参数;将注意力机制 SENet<sup>[11]</sup>结构嵌入其残差结构当中,使其能够对不同特征图进行相应处理.

## 1 相关工作

### 1.1 YOLOv4算法介绍

Alexey Bochkovskiy 继承了 YOLO 系列的观念,在 YOLOv3 的基础上进行了改进,发布了 YOLOv4<sup>[12]</sup>. 在与 EfficientDet 网络性能相当的情况下,该算法的推理速度是 EfficientDet 的 2 倍左右,比上一代 YOLOv3 算法的平均精度(AP)和 FPS 分别提高了 10% 和 12%.

该算法由主干特征提取网络 CSPDarknet53、特征金字塔 SPPNet 和检测结构 YOLO-Head 构成. 主干网络 CSPDarknet53 在 YOLOv3 主干网络 Darknet53 基础上,借鉴了 CSPNet 的思想,在减少参数计算量的同时保证了准确率. YOLOv4 算法在特征金字塔模块中采用了 SPPNet 结构,进一步提高了算法的特征提取能力,而 YOLO Head 特征层则继续使用 YOLOv3 的结构.

### 1.2 GhostModule网络模块

GhostModule<sup>[13]</sup>网络模块来自 GhostNet 网络,网络核心是开创了一种全新的 Ghost 模块,它可以在使用较少参数的同时能保留更多的特征. Ghost 模块可以替换任何 CNN 网络中的卷积操作,其优点是轻量、高效,效果优于轻量级网络 MobileNetV3<sup>[14]</sup>. 在神经网络中,Ghost 模块将一个普通的卷积块分成两部分. 首先将输入的特征图先经过卷积,再通过逐层深度可分离卷积生成其对应的冗余特征图,最后将卷积生成的特征图和深度分离卷积生成的特征图进行堆叠操作. 与传统卷积神经网络相比,该方法的参数量和计算复杂度均有所降低.

主要核心模块 Ghost Module 的原理主要是用 Ghost Module 代替传统卷积. 首先采用  $1 \times 1$  卷积对输入图片进行通道数的压缩,然后再进行深度可分离卷积<sup>[15]</sup>得到更多的特征图,最后对不同的特征图进行拼接操作,使其拼接到一起,组合成新的输出图. Ghost Module 结构如图 1 所示.

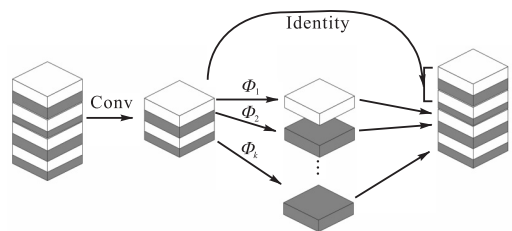


图 1 Ghost Module 结构  
Fig. 1 Ghost Module structure

### 1.3 ASPPNet

DeepLabv3<sup>[16]</sup>网络的ASPP模块结构包括1次 $1 \times 1$ 的普通卷积、3次 $3 \times 3$ 的空洞卷积和1个全局池化分支. 该网络模块受到空间金字塔池化的启发, 可以对不同特征的特征图进行有效采样以及分类操作.

ASPPNet利用空洞卷积<sup>[17]</sup>, 在不改变其他参数的情况下增大其感受野, 该模块的组成部分为1次 $1 \times 1$ 卷积和3次 $3 \times 3$ 卷积, 分别进行这些操作计算, 最终将4次卷积结果进行拼接叠加.

### 1.4 SENet注意力机制

通过引入注意力机制, 使众多输入信息能够聚焦于更为关键的信息, 降低对其余参数信息的敏感度, 过滤掉无关信息, 提高任务处理的效率与准确率. SENet是Momenta公司2017年在CVPR中提出, 并获得最后一届ImageNet图像识别冠军.

SENet模块结构如图2所示. 首先, 对给定输入特征图 $X$ 进行全局平均池化操作, 得到一个通道数为 $C$ 的特征图; 然后, 对通道数为 $C$ 的特征图进行两次全连接操作, 第一次全连接时的参数量相比第二次全连接时的参数量更少, 在两次全连接操作完成后, 对其特征层的每一层通道进行Sigmoid操作, 得到权重值; 最后, 将权重值与其原特征层进行参数相乘操作. 该模块可以用到其他网络结构中, 能够达到即插即用的效果, 对大多数网络都有所提升.

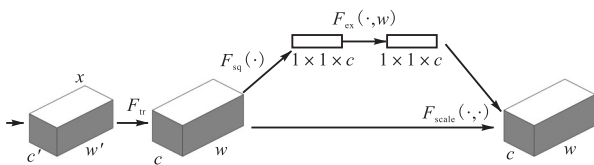


图2 SENet模块结构

Fig. 2 SENet module structure

## 2 算法改进

### 2.1 改进主干网络

原始的YOLOv4网络使用了CSPDarknet53作为其算法的主干网络, CSPDarknet53对图片特征有着很强的提取能力. 但是, 由于其网络结构复杂、参数繁杂, 导致其训练缓慢, 在轻量性方面有所欠缺. 为了提高速度, 将网络前驱部分DarknetConv2D的激活函数由Mish改为LeakyReLU. 由于模型一般在减少参数后准确率往往会降低, 因此对其网络中的Resblock\_body模块进行改进, 将该模块的残差部分

改为GhostModule结构, 形成ResGhost\_body模块, 弥补准确率降低的缺点. ResGhost\_body模块结构如图3所示. 主干网络最终提取到3个有效特征层( $52 \times 52$ 、 $26 \times 26$ 、 $13 \times 13$ ).

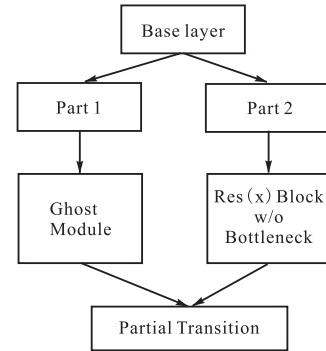


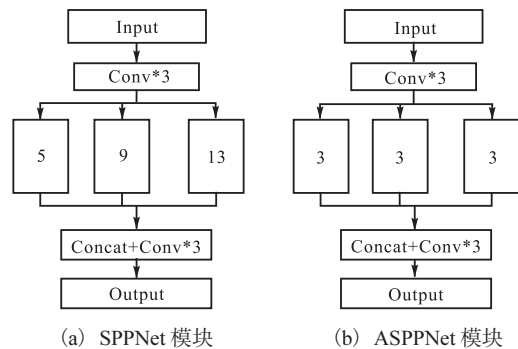
图3 ResGhost\_body模块结构

Fig. 3 ResGhost\_body module structure

### 2.2 改进特征融合模块

YOLOv4在其主干特征提取网络之后连接了特征融合模块SPPNet, 该网络结构由He等<sup>[18]</sup>提出, 起初是为了解决R-CNN速度缓慢的问题. 在原YOLOv4网络结构中, SPPNet接在了最后一层特征层的卷积里, 并对其主干网络中最后一个特征层进行3次卷积之后, 分别利用4个不同大小的池化层对其进行处理, 其池化核大小分别为 $1 \times 1$ 、 $5 \times 5$ 、 $9 \times 9$ 、 $13 \times 13$ , 最后将得到的4个特征模块图进行通道拼接, 结构如图4(a)所示. 但是, 此模块会导致参数量增大, 使网络结构训练减慢, 运行速度下降.

因此, 将SPPNet模块结构替换为与其相似的ASPPNet, 不仅能够减少参数量, 还可以增大网络的感受野. 该网络结构模块为1个 $1 \times 1$ 的普通卷积和3个 $3 \times 3$ 的空洞卷积. 首先对输入的特征图进行3次卷积, 之后对其进行1次 $1 \times 1$ 的普通卷积和3次 $3 \times 3$ 的空洞卷积, 最后将得到的4个模块特征图进行叠加, 其结构如图4(b)所示.



(a) SPPNet 模块

(b) ASPPNet 模块

图4 SPPNet改进图

Fig. 4 SPPNet improved diagram

### 2.3 嵌入 SENet 结构

原始特征图在经过了主干网络的特征提取及 ASPPNet 等结构的一系列卷积操作之后,随着参数数量的增大,产生梯度消失等问题,容易失去主要特征,使得最终的准确率下降.因此,在最后的 PANet<sup>[19]</sup>模块中加入注意力机制 SENet 结构,让网络聚焦更为关键的信息.在原始的 YOLOv4 网络结构中,主干特征提取网络有 2 个分支以及在特征融合模块之后有 1 个分支,这 3 个分支会进入 PANet 结构中,因此为了保留原始分支中的信息,在 PANet 模块中加入注意力机制 SENet.改进前模块和改进后模块的结构如图 5 所示,其中 P1、P2、P3 代表其 3 个分支.

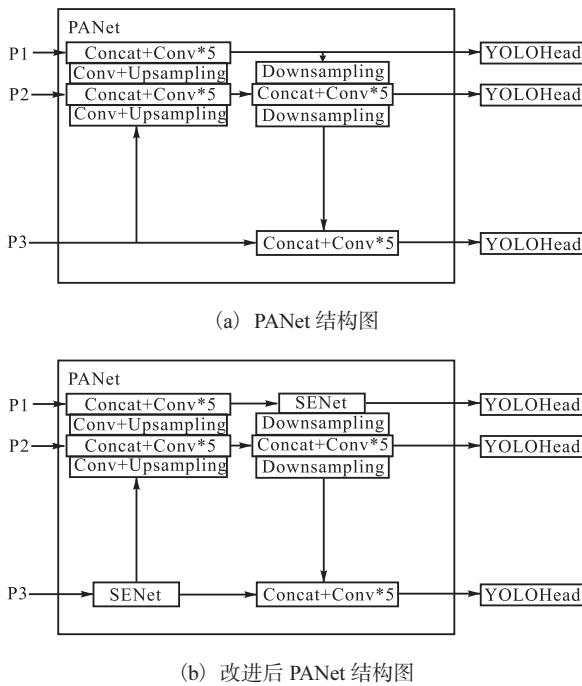


图 5 PANet 结构对比图

Fig. 5 Comparison of PANet structures

## 3 实验设计与结果分析

### 3.1 实验平台

本实验在 Windows10 操作系统下完成,使用语言为 Python,使用框架为 PyTorch 1.7.0,处理器为 Inter i5-9300H,内存为 16 GB, GPU 采用 GTX 1650 4 GB 显卡.

### 3.2 实验数据集

对于车辆的目标检测,采用自动驾驶 BDD100K 数据集,其中训练集 50 000 张,测试集 15 000 张.由于本文主要研究对车辆的目标检测,因此对数据集进行了筛选和清洗,最终训练集 36 400 张,测试集

9 256 张.

### 3.3 评价指标

用精确率 (precision)、召回率 (recall) 和平均精度 (AP) 对模型进行评估,公式为

$$P = N_{TP} / (N_{TP} + N_{FP}) \quad (1)$$

$$R = N_{TP} / (N_{TP} + N_{FN}) \quad (2)$$

$$P_{av} = \int_0^1 P(R) dR \quad (3)$$

式中:  $N_{TP}$  为被模型预测为正类的正样本,  $N_{FP}$  为被模型预测为正类的负样本,  $N_{FN}$  为被模型预测为负类的正样本,  $P$ 、 $R$ 、 $P_{av}$  分别表示精确率、召回率和平均精度.

### 3.4 实验结果比较

为了验证本文算法的有效性,实验在 BDD100K 数据集上进行了测试,结果见表 1.由表 1 可知:本文算法通过改进残差模块,在残差分支中加入了 GhostModule 模块;对原网络结构中的 SPP 结构进行改进,增大感受野,提高精度;PANet 网络结构添加了 SENet 注意力机制模块,使整个网络结构得到了优化.较原算法 YOLOv4 结构,AP 提升了 2.69%,帧每秒 (FPS) 也有相应提升,使算法具备更好的检测效果,AP、召回率、精确率结果对比如图 6 所示.

表 1 对比实验结果

Tab. 1 Results of contrast experiment

算法	平均精度/%	召回率/%	精确率/%	$F_1$	FPS
CenterNet	85.92	60.76	90.22	0.73	30.9
Faster-RCNN	86.64	66.86	83.57	0.74	32.4
YOLOv3	85.38	59.32	93.15	0.72	53.6
YOLOv4	88.27	62.11	95.76	0.75	58.1
本文算法	90.96	70.11	95.90	0.81	63.2

### 3.5 实验效果对比

本文模型与原算法 YOLOv4 在测试集上进行实验,部分检测结果如图 7 所示.从图 7 中可以看出,原算法对于部分车辆检测出现了漏检现象,而本文算法有着更好的检测性能.

### 3.6 消融实验

为了验证本文提出的改进 YOLOv4 算法对于原算法的提升效果,对本算法提出的 3 个改进点分别进行消融实验,分别为主干网络残差边融合 GhostModule、ASPP、SENet.消融实验结果见表 2.

由表 2 可知:在 AP 方面,改进后的算法均比原 YOLOv4 算法有不同程度的提升,加入 ASPP 后提升效果最好.

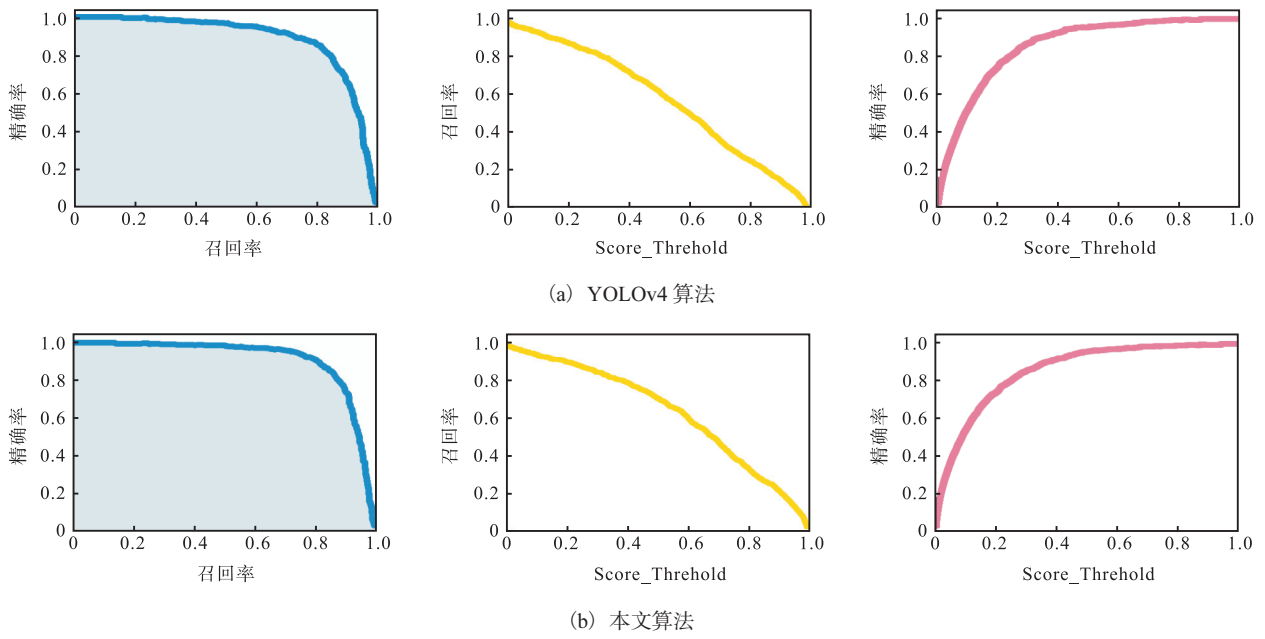
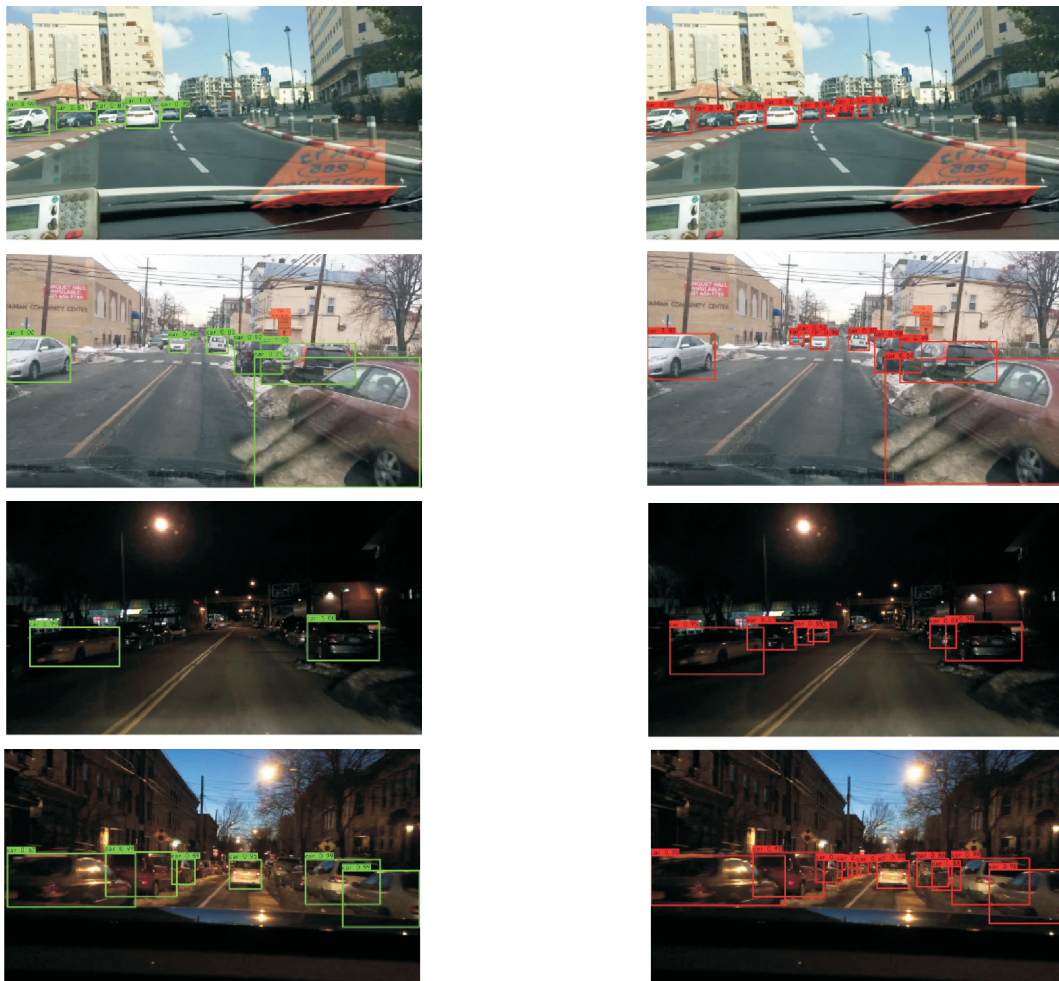


图6 不同算法 AP、召回率、精确率结果对比

Fig. 6 Contrast graph of AP, recall and precision of different algorithm



(a) YOLOv4 算法

(b) 本文算法

图7 YOLOv4算法和本文改进算法在测试集上的效果

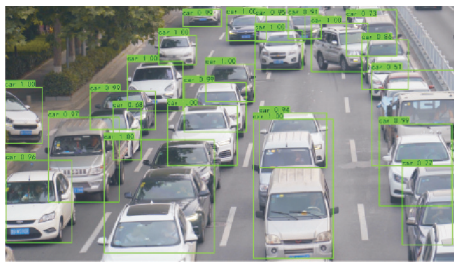
Fig. 7 Effects of YOLOv4 algorithm and improved algorithm on the test set

表 2 消融实验结果

Tab. 2 Ablation experimental result

算法	平均精度/%	召回率/%	精确率/%	$F_1$	FPS
YOLOv4	88.27	62.11	95.76	0.75	58.1
GhostModule	89.53	68.51	96.23	0.80	55.2
ASPP	90.26	67.49	93.84	0.78	67.9
SENet	89.71	69.36	94.68	0.80	62.4
本文算法	90.96	70.11	95.90	0.81	63.2

加入 SENet 模块之后召回率的提升效果最为明显,这主要是因为注意力机制可以更加关注重要的特征通道,提升网络的信息提取能力. ASPP 和 SENet 模块的精确率相对于 YOLOv4 算法并没有提升,而加入了 GhostModule 之后,精确率有所提升,这是因为在残差边部分加入 GhostModule 模块可以加强其特征提取能力,使得精确率提升. 在 FPS 方面,在将 SPP 改为 ASPP 之后,参数量减少,导致模型检测 FPS 提升最为明显. 因此在加入上述 3 处改进点之后,4 个指标都有不同程度的增长. 本文中的算法同时融合了 3 处改进点,相对于 GhostModule,牺牲了一部分精确率,但是平均精度和召回率均为上述实验



(a) YOLOv4 算法

中的最高值,并且精确率和 FPS 相对于 YOLOv4 算法也有所提升.

### 3.7 实际场景中的应用

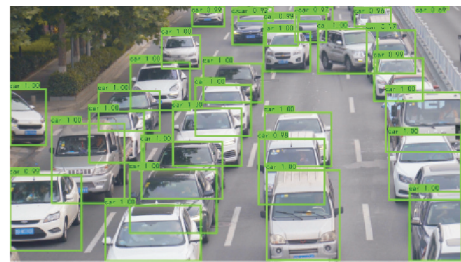
为了验证本算法在实际场景中的应用能力,自制搜集相关数据集 6 000 张,对其进行标注并将图片转换为 voc 格式数据集,在实际场景中对数据集进行实验,实验结果见表 3.

表 3 网络模型在实际场景下的实验结果

Tab. 3 Experimental result of network model in real scene

算法	平均精度/%	召回率/%	精确率/%	$F_1$	FPS
CenterNet	88.22	69.35	92.34	0.79	37.6
Faster-RCNN	90.98	75.26	94.66	0.84	37.3
YOLOv3	87.42	73.92	95.28	0.83	48.6
YOLOv4	92.57	79.75	95.84	0.87	60.4
本文算法	95.88	83.33	96.93	0.89	65.7

由表 3 可知,本文算法在实际场景下比 YOLOv4 算法表现更为优秀,平均精度提高了 3.31%,且其他指标都有不同程度的提高,YOLOv4 算法与本文算法在实际场景下检测效果如图 8 所示.



(b) 本文算法

图 8 实际场景下检测效果对比

Fig. 8 Comparison of detection effects in actual scenarios

## 4 结 语

(1)本算法在原始算法 YOLOv4 的主干特征提取网络的残差块中添加 GhostModule 模块,提高其检测精度;同时将特征金字塔中的 SPPNet 模块替换为 ASPPNet 模块,减少参数量的同时增大网络感受野;在 PANet 模块中的两个分支插入 SENet 结构,防止其因为网络结构复杂而出现梯度消失的问题,增强网络特征的表现能力.

(2)本文算法通过实验验证,解决了在交通道路中对于车辆的漏检、误检情况,经过改进后的 YOLOv4 算法已经可以满足道路场景下的实际应用需要. 但是,目前图片数据集还是不够全面,之后还需要增加一些处于极端天气下,比如暴雨和沙尘暴天

气下的车辆图片,从而让模型能够在任何条件下都拥有良好的检测性能.

### 参考文献:

[1] 曹家乐,李亚利,孙汉卿,等. 基于深度学习的视觉目标检测技术综述[J]. 中国图象图形学报, 2022, 27(6): 1697-1722.

[2] 吕璐,程虎,朱鸿泰,等. 基于深度学习的目标检测研究与应用综述[J]. 电子与封装, 2022, 22(1): 72-80.

[3] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//IEEE. 2014 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2013: 6909475.

[4] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single-

- shot multibox detector[C]//IEEE. IEEE Conference on Computer Vision and Pattern Recognition ECCV 2015. New York: IEEE, 2016: 21–37.
- [ 5 ] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//IEEE. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2017: 6517–6525.
- [ 6 ] REDMON J, FARHADI A. YOLOv3: an incremental improvement[EB/OL]. [2022–08–01]. <https://arxiv.org/pdf/1804.02767.pdf>.
- [ 7 ] 李挺,伊力哈木·亚尔买买提. 基于改进YOLOv4的行人检测算法[J]. 科学技术与工程, 2022, 22(8): 3221–3227.
- [ 8 ] 陈洋,张明,杨立东,等. 一种基于改进YOLOv4的SAR舰船检测算法[J/OL]. 电子测量技术: 1–6[2022–07–07]. <http://kns.cnki.net/kcms/detail/11.2175.TN.20220601.1914.078.HTML>.
- [ 9 ] 王滢暄,宋焕生,梁浩翔,等. 基于改进的YOLOv4高速公路车辆目标检测研究[J]. 计算机工程与应用, 2021, 57(13): 218–226.
- [ 10 ] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. IEEE Transactions on pattern analysis and machine intelligence, 2018, 40(4): 834–848.
- [ 11 ] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on pattern analysis and machine intelligence, 2020, 42(8): 2011–2023.
- [ 12 ] BOCHKOVSKIY A, WANG C Y, LIAO H. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. [2022–08–01]. <https://doi.org/10.48550/arXiv.2004.10934>.
- [ 13 ] HAN K, WANG Y, TIAN Q, et al. GhostNet: more features from cheap operations[C]//IEEE. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2020: 9157333.
- [ 14 ] HOWARD A, SANDLER M, CHU G, et al. Searching for MobileNetV3[EB/OL]. [2022–08–01]. <https://doi.org/10.48550/arXiv.1905.022442019>.
- [ 15 ] SIFRE L, MALLAT S. Rigid-motion scattering for texture classification[EB/OL]. [2022–08–01]. <https://doi.org/10.48550/arXiv.1403.1687>.
- [ 16 ] CHEN L C, PAPANDREOU G, SCHROFF F, et al. Rethinking atrous convolution for semantic image segmentation[EB/OL]. [2022–08–01]. <https://arxiv.org/pdf/1706.05587.pdf>.
- [ 17 ] YU F, KOLTUN V. Multi-scale context aggregation by dilated convolutions[EB/OL]. [2022–08–01]. <https://doi.org/10.48550/arXiv.1511.07122>.
- [ 18 ] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on pattern analysis & machine intelligence, 2014, 37(9): 1904–1916.
- [ 19 ] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]//IEEE. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2018: 8579011.

责任编辑: 郎婧