2024 年 第 24 卷 第 16 期
2024, 24(16):06757-09

科 学 技 术 与 工 程
Science Technology and Engineering

ISSN 1671—1815
CN 11—4688/T

DOI:10.12404/j.issn.1671-1815.2303779

引用格式:高昕, 甄国涌, 储成群, 等. 基于改进 YOLOv5 的自动驾驶目标检测方法[J]. 科学技术与工程, 2024, 24(16): 6757-6765.
Gao Xin, Zhen Guoyong, Chu Chengqun, et al. Autonomous driving target detection method based on improved YOLOv5[J]. Science Technology and Engineering, 2024, 24(16): 6757-6765.

自动化技术、计算机技术

基于改进 YOLOv5 的自动驾驶目标检测方法

高昕, 甄国涌*, 储成群, 王子硕
(中北大学仪器与电子学院, 太原 030051)

摘 要 针对目前自动驾驶领域的目标检测算法在交通场景下的漏检目标,目标定位不精确、目标特征表达不充分及目标识别效果欠佳等问题,提出一种基于 TPH-YOLOv5(transformer prediction heads-YOLOv5)的道路目标检测方法。首先为了减轻物体尺度急剧变化带来的漏检风险,增加了用于微小物体检测的检测头,为在高密度场景中精确定位对象,使用 Transformer 预测头来捕获全局信息;其次为了增强模型的特征表达能力,用选择性注意力机制(selective interactive module with affinity learning,SIMAM)模块对卷积层的输出进行加权;最后,为了提高目标识别的精度,网络颈部增加了 4 个金字塔池化模块(spatial pyramid pooling,SPP)块来进行多尺度融合,为了加快收敛速度和提高回归精度采用 EIOU(Euclidean distance-IOU)作为边界框损失函数。通过消融、对比和可视化验证实验表明,提出的算法比 YOLOv5 在平均精度上提高了 8.1%,漏检率明显减少,目标检测效果明显增强。

关键词 目标检测; 自动驾驶; YOLOv5; 多尺度检测; 损失函数
中图法分类号 TP391; **文献标志码** A

Autonomous Driving Target Detection Method Based on Improved YOLOv5

GAO Xin, ZHEN Guo-yong*, CHU Cheng-qun, WANG Zi-shuo
(School of Instrument and Electronics, North University of China, Taiyuan 030051, China)

[**Abstract**] Aiming at the problems of missed targets, inaccurate target localization, insufficient target feature expression, and unsatisfactory target recognition effects in the current object detection algorithms for autonomous driving in traffic scenarios, a road object detection method based on transformer prediction heads-YOLOv5 (TPH-YOLOv5) was proposed. Firstly, to reduce the risk of missed objects caused by drastic changes in object scales, a detection head for detecting small objects was added, and a Transformer prediction head was used to capture global information for precise object localization in high-density scenes. Secondly, to enhance the feature expression ability of the model, the output of the convolutional layer was weighted using the selective interactive module with affinity learning(SIMAM) module. Finally, in order to improve the accuracy of target recognition, four spatial pyramid pooling(SPP) blocks were added to the neck of the network for multi-scale fusion, and the Euclidean distance-IOU(EIOU) was used as the bounding box loss function to accelerate convergence speed and improve regression accuracy. Through ablation, comparison, and visualization experiments, it is shown that the proposed algorithm improves the average precision by 8.1% compared to YOLOv5, significantly reduces the missed detection rate, and enhances the object detection performance.

[**Keywords**] object detection; autonomous driving; YOLOv5; multi-scale detection; loss function

现今城市建设日益复杂,使得自动驾驶汽车在应对多变的交通场景时面临着挑战。为了解决由于路况感知不足所导致的城市交通问题,目前广泛采用摄像头搭建环境感知系统,以提供自动驾驶汽车全面、丰富的周边环境信息。而目标检测技术则是实现准确的道路目标检测的最基础和核心技术。目标检测技术是一种通过分析目标的几何和统计特征,在图像中确定目标所在位置并识别其属于哪一类的图像处理技术。在驾驶任务中,自动驾驶汽车利用摄像头捕捉道路场景,并通过车载电脑内的目标检测算法进行定位和识别,提供充足的路况信息,从而协助自动驾驶汽车进行规划和决策,保障行车安全。

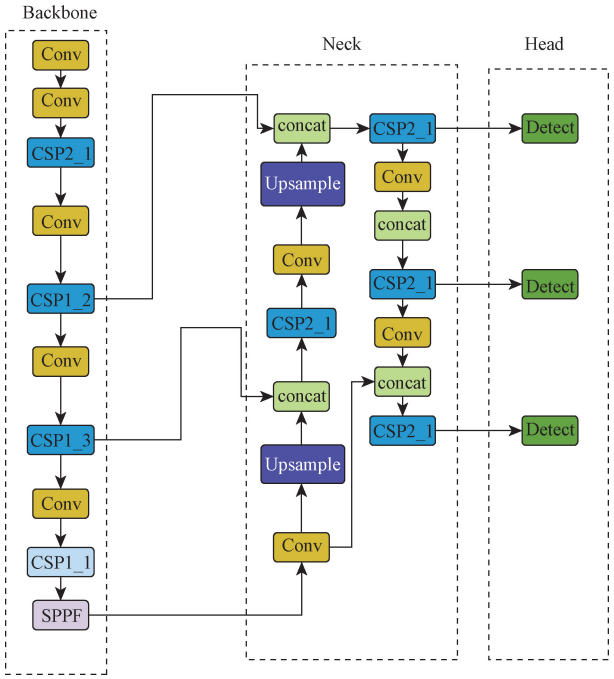
收稿日期: 2023-05-23 **修订日期**: 2024-03-11
基金项目: 国家自然科学基金重点项目(62131018);山西省基础研究计划(202103021222012)
第一作者: 高昕(1998—),男,汉族,山西忻州人,硕士研究生。研究方向:图像处理,计算机视觉。E-mail:869118922@qq.com。
* **通信作者**: 甄国涌(1971—),男,汉族,山西阳泉人,博士,教授。研究方向:应用软件开发、机器视觉。E-mail:zengguoyong@nuc.edu.cn。

目标检测算法主要是利用图像特征进行目标检测,如尺度不变特征转换(scale-invariant feature transform,SIFT)算法[1]等。在此基础上,出现了一些基于灰度图像的目标检测方法,如 Viola-Jones 算法。但是这些算法仅适用于相对简单的场景,并且无法处理遮挡、光照变化和姿态变化等复杂情况。随着深度学习的发展,特别是卷积神经网络(convolutional neural network,CNN)的出现,目标检测算法得到了重大突破。首先是 R-CNN(region with CNN feature)[2]系列算法的提出,包括 R-CNN、Fast R-CNN[3]、Faster R-CNN 等,但是,这些方法有很多不足之处,包括速度慢、复杂度高等。随后,YOLO(you only look once)算法的出现极大地简化了目标检测算法的流程,并且速度更快,效果更好,SSD[4](single shot multibox detector)算法也是同样如此。近年来,有学者对 YOLOv5 进行了全方位的改进。王志斌等[5]添加 RFB(receptive field block)模块用于增加网络感受野,有助于模型捕捉全局特征,将无参数的 SimAM 注意力机制添加 Bottleneck 中,在不增加参数的情况下,提高网络的特征提取能力,但是在检测速度方面与原模型基本一致,只是提高了检测精度,并不能完美的适应实际的工业需求。郭朦等[6]提出一种 YOLOv5 与视觉转换器(vision transformer, vit)结合的检测与识别方法。首先采用 YOLOv5 对目标进行检测,得出交通标志的位置信息,再将其输入 vit 进行分类识别,其中特征连接部分引入 DenseNet 网络模块,来实现原始特征和卷积后特征映射的密集连接,加强特征的传递性,提高识别率。但是由于该算法为了提升检测精度和检测时间,缩小模型复杂度时将检测的目标类别进行减少,只有对部分交通标志的识别检测,这与自动驾驶的目标检测领域检测目标种类繁多的现状未完全贴合。鉴于此,针对以上的缺陷,增加被检测目标的类别,优化模型的复杂度,减少计算量,在提高检测精度的基础上提升检测速度,以满足实际的工业需求。

针对自动驾驶目标检测过程中由于检测密度较大导致的漏检和检测效果变差的问题,基于 YOLOv5[7]的网络结构改进来解决上述问题,使用 CSPDarknet53 和路径聚合网络(PANet[8])作为 TPH-YOLOv5 的骨干和颈部。首先,为了减轻物体尺度急剧变化带来的漏检风险,增加了用于微小物体检测的检测头 TPH(Transformer prediction heads),可以同时完成在高密度场景中精确定位对象和捕获全局信息的任务;其次,为了在覆盖范围较大的图像中找到注意区域,采用卷积注意力模块[9](convolutional block attention module, CBAM)沿通道维度和空间维度依次生成注意图;再次,为了增强模型的特征表达能力和提高目标识别的精度,用选择性注意力机制(selective interactive module with affinity learning,SIMAM)模块对卷积层的输出进行加权且在 neck 层增加了 4 个 SPP(spatial pyramid pooling)块来进行多尺度融合;最后,为了加快收敛速度和提高回归精度采用 EIOU 作为边界框损失函数。通过以上网络结构的调整和改进,在保持基本的检测速度同时提升对密集交通场景目标的检测精度,满足实际应用的需求。研究成果对提升智能汽车自动驾驶技术具有重要意义。

1 YOLOv5 目标检测算法

YOLO(you only look once)是著名的单级目标检测算法,是单级检测发展史上的里程碑。其后续系列改进版本进一步提高了检测性能。目前,YOLOv5 仍然是主流的目标检测算法之一,被广泛应用于各种场景。YOLO 的网络结构如图 1 所示,由 3 个主要部分组成。



Backbone 为主干网络;Neck 为颈部网络;Head 为头部网络;Conv 为卷积块;CSP 为交叉阶段分离网络;SPPF 为空间金字塔池化;Concat 为连接层;Upsample 为上采样层;Detect 为检测层

图 1 YOLOv5 网络结构

Fig. 1 YOLOv5 network structure

1.1 Backbone 网络

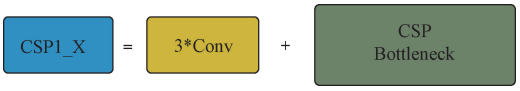
Backbone 网络使用 CSPDarknet53 作为骨干网络,提取输入图像的特征表示。CSPDarknet53 是轻量级的卷积神经网络,由若干个 CSPBlock 组成,结

构如图 2 所示,其中每个 CSPBlock 包含 3 个卷积块和 1 个残差块,以减少参数数量和计算量。

YOLOv5 是一种目标检测模型,采用 CSPDarknet 作为主干网络,并通过 5 次卷积层 CBS(convolutional block stage)提取输入图像的特征。每次卷积后,使用 CSP 模块对特征进行堆叠,同时在主干网络末尾添加 SPPF(spatial pyramid pooling-fast)模块来池化特征,最终输出 3 个不同尺度的特征层。CBS 由卷积层 Conv、标准化层 BatchNorm 和激活函数 Silu 构成,可以映射目标特征,增强检测模型的非线性拟合能力。

CSP(cross stage partial)模块借鉴了 ResNet[10] 残差网络思想,在处理特征时将输入分成两部分。主干部分利用 CBS 和两次卷积逐步提取特征并进行深度堆叠,分支部分仅使用单次卷积来调整通道数,经过处理后再由 CBS 卷积提取堆叠的新特征层。通过使用残差结构提取特征,可以增加网络宽度,减少冗余的梯度信息,从而提高模型性能。

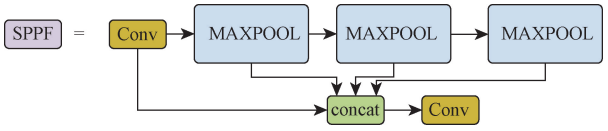
SPPF 模块采用串行方式进行池化操作,包括 1 次 ConV 卷积和 3 次 5×5 的池化层。经过这些处理后,再由 ConV 卷积层提取并堆叠新的特征层。SPPF 模块能够进一步提取特征,增强主干网络的深层次特征表达能力,并扩大模型的感受野,从而提升检测性能,SPPF 结构如图 3 所示。



CSP Bottleneck 为交叉阶段分离瓶颈块;* 为卷积

图 2　CSP 结构图

Fig. 2　CSP structure diagram



MAXPOOL 为最大池化

图 3　SPPF 结构图

Fig. 3　SPPF structure diagram

1.2　Neck 网络

采用 SPP 和 PAN(path aggregation network)模块构建 Neck 网络,将来自不同尺度的特征图融合成更具代表性的特征图。SPP 模块用于从不同尺度的特征图中提取信息,而 PAN 模块用于融合来自不同尺度的特征图。

1.3　Head 网络

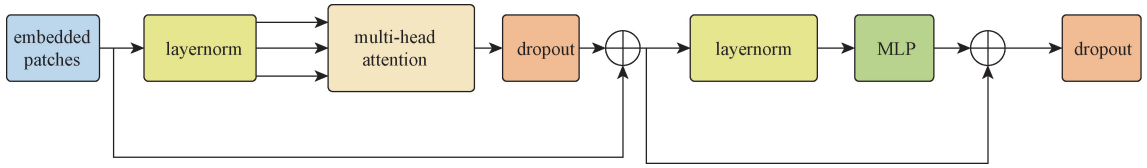
Head 网络作为分类网络,不能完成定位任务,头部负责通过骨干网提取的特征图检测目标的位置和类别。磁头一般分为两种:一级目标检测器和二级目标检测器。两级检测器一直是目标检测领域的主流方法,其中最具代表性的是 RCNN 系列。与两级检测器相比,一级检测器同时预测边界框和目标类别,速度优势明显,但精度较低。

2　改进 YOLOv5 目标检测算法

2.1　目标识别不充分改进

YOLOv5 原网络存在当场景目标密度变大时识别量降低和物体尺度急剧变化时容易漏检目标的问题,导致目标识别效果欠佳。因此在网络中增加了用于微小物体检测的检测头,与其他 3 个探测头相结合,四探测头结构可以减轻物体尺度急剧变化带来的负面影响。其次使用 TPH 在高密度场景中精确定位对象。此外,由于分类结果较差,对于一些容易混淆的类别,使用自训练分类器来提高分类能力。Transformer 编码器模块替换 Neck 中的一些 CSP 模块,将原始预测头转换成 TPH,实现具有多头注意力机制的预测潜力,捕获全局信息和充足的背景信息。

Transformer 预测头的结构包括两个主要模块:第一个子层是多头注意层,第二个子层(MLP)是全连接注意层。每个子层之间使用残差连接。Transformer 预测头不仅增加了捕获不同本地信息的能力还可以利用自注意机制探索特征表示潜力。LayerNorm 和 Dropout 层有助于网络更好地收敛,防止网络过拟合,结构如图 4 所示。



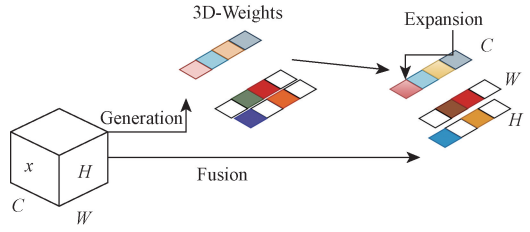
embedded patches 为嵌入片段;layernorm 为层归一化;multi-head attention 为多头注意力机制;dropout 为随机失活;MLP 为多层感知机

图 4　Transformer 预测头结构图

Fig. 4　Transformer prediction head structure diagram

2.2 目标特征表达不充分改进

SIMAM[11]是一种交互式的注意力机制,主要用于增强图像特征的相关性,从而提高图像识别的准确性。它包含相似度计算和特征交互。首先,SIMAM通过计算两个特征向量之间的相似度来衡量它们之间的相关性。接着,SIMAM 将相似度作为权重,通过特征交互来增强特征向量之间的相关性。特别地,SIMAM 采用基于图像分割的注意力机制,将图像分成多个区域,分别计算区域之间的相似度,然后将相似度应用于区域内的特征向量交互,以增强区域之间的相关性。SIMAM 可以用于各种图像识别任务,如分类、检测和分割等。实验证明,SIMAM 可以提高模型的准确性和鲁棒性,尤其是在处理具有复杂场景和多个目标的图像时表现优异。因此将 SIMAM 注意力机制添加到 TPH-YOLOv5 的 Backbone 层,以增强模型对特征的关注能力。具体而言,常常将 SIMAM 模块插入 CSP-Darknet53 网络中的每个残差块(residual block)的最后一个卷积层中,在该结构中,SIMAM 模块紧跟在一个卷积层之后,对该卷积层的输出进行加权,强调重要的特征通道,抑制无用的特征通道。这样做可以增强模型的特征表达能力,提高检测精度。SIMAM 结构如图 5所示。



Generation 为生成模型;Fusion 为特征融合;3D-Weights 为 3D 权重;Expansion 为维度扩张

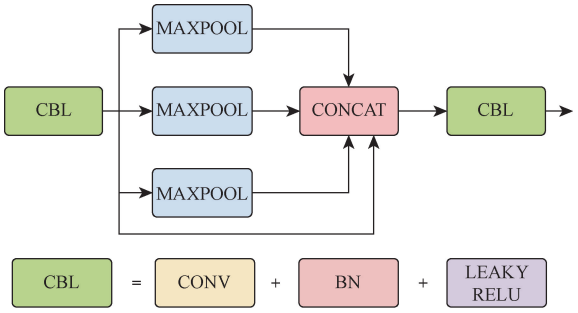
图 5 SIMAM 结构图

Fig. 5 SIMAM structure diagram

2.3 目标定位不精确改进

SPP[12]的主要作用是解决输入图像大小不均匀的问题,在 SPP 中融合不同大小的特征有利于检测图像中物体大小差异较大的情况。在网络颈部增加 4 个 SPP 块,使得特征图经过多个多尺度融合步骤,丰富了特征图的表达能力。在网络中采用均匀步幅(stride = 1)但大小不同的核来实现 SPP,并通过 concat 来实现特征融合。SPP 块的结构如图 6 所示。

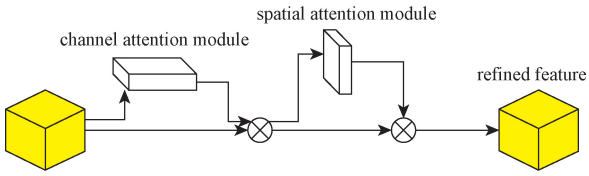
卷积块注意模块(CBAM)是一个轻量级模块,将 CBAM 沿通道和空间两个独立维度依次推断注意力图,然后将注意力图与输入特征图相乘,进行自适应特征细化,CBAM 模块的结构如图 7 所示。



CBL 为卷积块层;BN 为批归一化;LEAKY RELU 为修正线性单元

图 6 SPP 块结构图

Fig. 6 SPP block structure diagram



channel attention module 为通道注意力模块;spatial attention module 为空间注意力模块;refined feature 为处理后特征

图 7 CBAM 结构

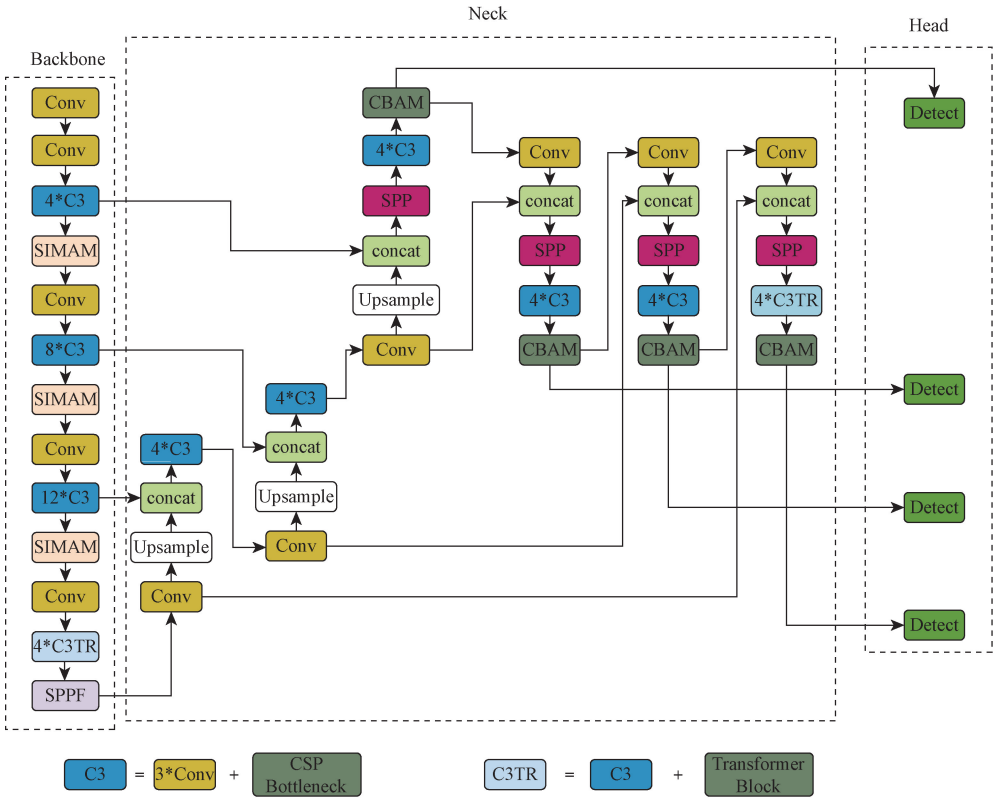
Fig. 7 CBAM structure

2.4 目标识别效果改进

损失函数是在训练深度学习模型时用于衡量模型预测输出与真实标签之间差异的指标。在训练过程中,优化器通过最小化损失函数来调整模型参数,以使模型的预测输出与真实标签更加接近。在 YOLOv5 中,默认使用的是 CIOU[13](complete intersection over union)损失函数。EIOU[14](Euclidean distance-IOU)是一种用于计算物体检测中边界框回归损失的函数,该函数将边界框预测与真实边界框进行比较,并计算它们之间的差异性损失,包括:重叠损失(LEIOU)、中心距离损失(Ldis)和宽高损失(Lasp),其中前两部分继承于 CIOU 的方法。相对于 CIOU,EIOU 可以同时增大或减小边界框的宽度和高度,并且收敛速度更快。这表明使用 EIOU 方法可以更快地训练物体检测模型。与 CIOU 相比,EIOU 考虑了重叠部分的数量,同时还考虑边界框之间的距离,因此在边界框回归任务中可以更好地表示两个边界框的相似性。使用 EIOU 代替 CIOU,来获得更好的目标检测效果。

2.5 YOLOv5 网络模型改进

对 YOLOv5 进行多方面的优化和改进,如图 8 所示,首先在 Backbone 层,YOLOv5 只有 3 个 CSP 卷积块,而本算法在 Backbone 层有 12 个 C3 层,卷积运算更多,检测精度更大,且在 SPPF 之前加入 C3TR 模块,减少了参数量,优化了计算,保持了原有精度。在 Neck 层,由于在 Head 层添加了一个微



C3 为 3 层卷积块;SIMAM 为选择注意力模块;C3TR 为 3 层卷积转换头模块

图 8　改进后网络结构图

Fig. 8　Improved network structure diagram

小物体检测头,使得在 Neck 层增加了一次上采样过程,且将原始的4 个CSP 卷积块增加到 16 个 C3 卷积块,因为计算量的提升对目标特征提取的工作量变大,所以在输出到 4 个 Detect 头之前,在每一部分的尾部添加了 SPP 块,这一模块的主要作用是对高层特征进行提取并融合,在融合的过程中多次运用最大池化,尽可能多地去提取高层次的语义特征。

3　实验与结果分析

3.1　实验环境及参数设置

本实验使用的操作系统为 Windows10,GPU 型号为 NVIDIA GeForce GTX 1650,编译语言为 Python 3.7,深度学习框架为 Pytorch 1.7.0,CUDA 版本为 11.0。

整个过程采用随机梯度下降法训练 100 个epoch(周期),输入图片大小为 640×640,batch_size(批尺寸)设为 8,学习率设为 0.01,余弦退火超参数设为 0.01,学习率动量设为 0.937,权重衰减系数设为 0.000 5。在模型的初始训练中,首先进行 3 个 epoch 的预热训练,预热学习率动量设为 0.8,预热学习率设为 0.1。置信度阈值设为 0.25,IOU 阈值设为 0.45。

3.2　实验数据集

使用伯克利大学 AI 实验室发布的 BDD100K 数据集中的 10K 样本来模拟道路场景。该数据集提供了多种不同的昼夜、天气和地理环境变化,可以训练模型以适应各种实际应用场景并具有更强的泛化能力。数据集被分为 2 500 张训练集、900 张验证集和 1 000 张测试集,涵盖了 9 类道路检测目标,数据集样本如图 9 所示。

3.3　实验评价指标

(1)精确率(precision)。指被分类器正确识别为正例的样本数占被分类器判定为正例的样本总数的比例。

(2)召回率(recall)。指被分类器正确识别为正例的样本数占所有正例的样本总数的比例。

选取单类精度 AP(average precision)和平均精度 mAP(mean average precision)来评价算法的检测精度。其中 mAP 是目标检测任务中常用的评价指标之一,用于衡量模型在不同置信度阈值下的精度。相应计算公式分别为

$$P = \frac{TP}{TP + FP} \tag{1}$$

$$R = \frac{TP}{TP + FN} \tag{2}$$

$$\mathrm{AP} = \int_1^0 P(R)\,\mathrm{d}R \tag{3}$$

$$\mathrm{mAP} = \frac{1}{n}\sum_{i=1}^{n}\mathrm{AP}_i \tag{4}$$

式中:AP(average precision)为单个检测类别的精度;P(precision)为所有被预测为正的样本中 实际为正样本的概率;R(recall)为实际为正的样本中被预测为正样本的概率;n 为检测类别的总类别数;TP 为真正例数量;FP 为假正例数量;FN 为假反例数量;mAP 为所有检测类别的平均精度;AP_i为第 i 个检测类别的精度。

3.4 实验结果分析

以 IOU0.5 时的平均精度mAP@0.5 和IOU0.5-0.95 的平均精度 mAP@0.5-0.95 为评价指标 来验证各改进措施的有效性。将原 YOLOv5 算法和本文算法在相同条件下进行对比实验,经过训练后,检测 9 个交通类别的性能如表 1 所示。本文算法的 mAP@0.5 经过 100 个 epoch 后从 0.678 涨到 0.950,mAP@0.5-0.95 从 0.448 涨到 0.766,而原算法 mAP@0.5 从 0.525 提升到0.869,mAP@0.5-0.95从0.327 升到0.671,本文算法相较于基础的 YOLOv5 平均精确率提高 8.1 个百分点,仿真结果对比如图 10 所示。

(a)雪天 (b)雨天

(c)夜晚 (d)白天

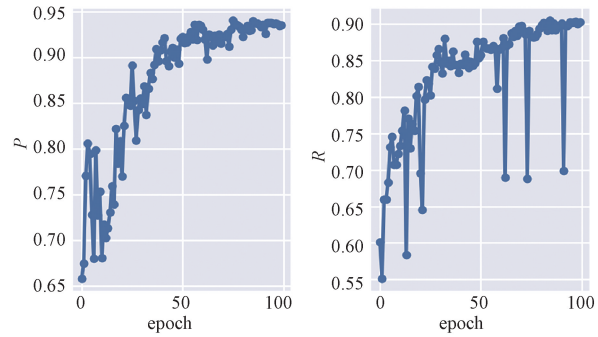
图 9 交通场景数据集 BDD100K

Fig. 9 Traffic scenario dataset BDD100K

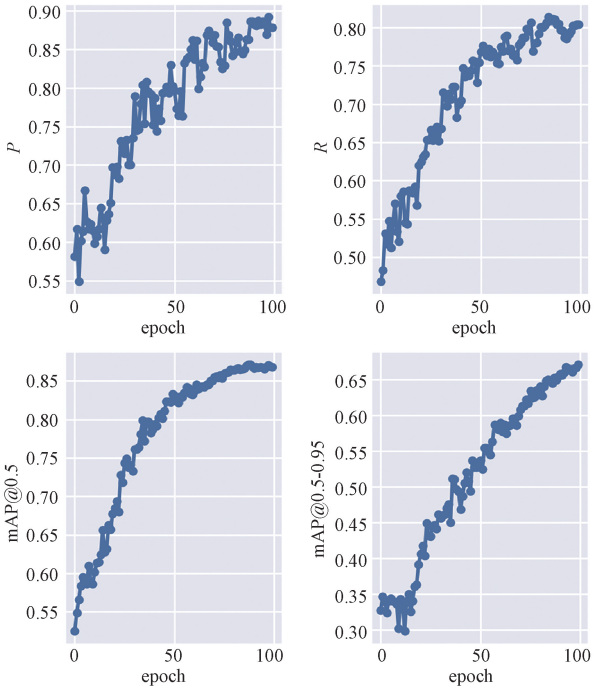
表 1 本文算法(A)与 YOLOv5(B)的 9 个类别性能测试对比

Table 1 Comparison of 9 categories of performance tests between the proposed algorithm (A) and YOLOv5 (B)

类别	*P*		*R*		mAP@0.5		mAP@0.5-0.95	
	A	B	A	B	A	B	A	B
行人	0.991	0.887	0.903	0.802	0.957	0.878	0.765	0.651
自行车	0.973	0.891	0.863	0.792	0.951	0.885	0.714	0.646
汽车	0.911	0.823	0.878	0.785	0.948	0.852	0.726	0.648
摩托车	0.941	0.827	0.899	0.724	0.937	0.788	0.735	0.639
大巴车	0.951	0.806	0.865	0.793	0.941	0.815	0.721	0.628
火车	0.974	0.886	0.855	0.776	0.945	0.786	0.719	0.625
货车	0.981	0.885	0.861	0.787	0.949	0.788	0.706	0.619
交通灯	0.936	0.798	0.831	0.779	0.927	0.779	0.715	0.614
停止标志	0.886	0.785	0.829	0.720	0.918	0.774	0.702	0.609



(a)本文算法仿真结果



(b)YOLOv5仿真结果

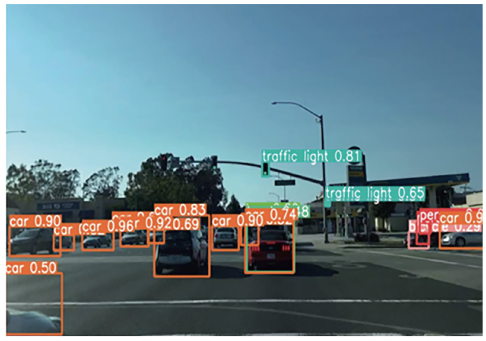
epoch(周期)表示模型训练过程中数据集的完整遍历次数

图 10　仿真结果对比

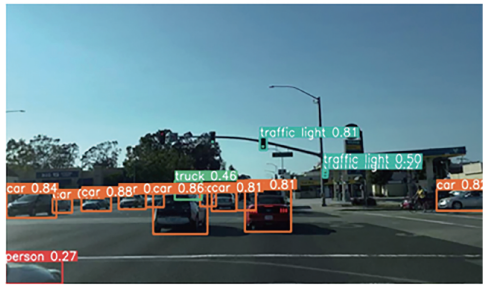
Fig. 10　Comparison of simulation results

3.5　算法效果验证

使用以上两种模型分别对相同图片进行处理,得到可视化的结果,由图片上的置信度可以容易得知,改进后的算法在交通场景识别上具有更高的置信度,证明本文算法能改善目标 特征表达不充分及目标定位不准确的问题,有明显的提升效果,如图 11所示。



(a)本文算法(图片1)



(b)原算法(图片1)

(c)本文算法(图片2)

(d)原算法(图片2)

图 11　道路目标检测结果对比

Fig. 11　Comparison of road object detection results

3.6　消融实验

为了探究本文改进点对 YOLOv5 算法的提升效果,对改进点逐一进行消融实验分析,实验结果如表 2 所示。

由表 2 可知,增加了用于微小物体检测的检测头的算法 1 在 mAP 上有小幅度提升,但是增加了检测内容,导致 FPS 变小,检测速度变慢。算法 2 使用改进 TPH-YOLOv5 网络作为主体网络,增加网络了复杂度,检测精度提升了 2.6 个百分点,检测速度降

表 2 在数据集 BDD100K 上的消融实验

Table 2 Ablation experiments on dataset BDD100K

算法	改进点						mAP@0.5	FPS/(帧·s^{-1})
	A	B	C	D	E	F		
YOLOv5							0.869	126.8
YOLOv5-head	√						0.874	103.6
YOLOv5-TPH		√					0.895	107.1
YOLOv5-SIM			√				0.902	99.6
YOLOv5-SPP				√			0.884	112.3
YOLOv5-CBAM					√		0.926	101.9
YOLOv5-EIoU						√	0.872	121.8
本文算法	√	√	√	√	√	√	0.950	98.9

注:A 为增加了用于微小物体检测的检测头;B 为使用改进 TPH-YOLOv5 网络作为主体网络;C 为使用 SimAM 添加到 Backbone 层来增强模型的特征表达能力;D 为使用 4 个 SPP 来进行多尺度融合;E 为使用 CBAM 轻量级模块来进行自适应特征细化;F 为使用 EIOU 损失函数作为边界框损失函数;FPS(frames per second)为每秒传输帧数。

低 19.7 帧/s。算法 3 增加了注意力机制 SimAM 来增强模型表达能力,mAP 显著提升,检测速度显著下降。算法 4 使用 4 个 SPP 进行多尺度融合,在保持较高的检测速度下,提升了检测精度。算法 5 使用 CBAM 轻量级模块来进行自适应特征细化,mAP 提升到 0.926,证明该模块对特征提取的显著作用。算法 6 使用 EIOU 损失函数作为边界框损失函数,网络结构没有显著变化,检测速度与 YOLOv5 相差不大,略微提升了检测精度。本文算法集成了以上 6 种算法的改进,检测精度提升到 0.950,提升了 8.1 个百分点,但是由于本文网络复杂度的增加,检测速度为 98.9 帧/s,略慢于 YOLOv5 检测速度,但是完全可以满足工业视觉的需求。

3.7 对比实验

为了进一步验证本文算法的检测性能,将本文算法与 Faster RCNN[15]、SSD、YOLOv4[16]、YOLOX[17]、YOLOv7[18]、DetectoRS[19]等目标检测算法进行对比,均采用 BDD100K 数据集进行性能测试,同时使用 mAP@0.5 和 FPS 作为评价指标,检测结果如表 3 所示。

由表 3 可知,首先因为 Faster RCNN 卷积提取网络是单层的且分辨率也很小,这些都不利于小物体及多尺度的物体检测,而本文算法拥有多层融合的特征图和特征融合,因此在检测精度和在检测速度方面都有大幅度的提升;其次,由于 SDD(single shot multibox detector)算法存在重复框和对小目标检测不够鲁棒的问题,导致它虽然比起 Faster RCNN 有所进步,但还是在检测精度上低于本文算法 34.6 个百分点,检测速度慢了 30.7 帧;再次,由于 YOLO 系列算法的不断更迭,YOLO 算法在检测精度上有了大幅度的提升,但是检测速度由模型的复杂度决定,也就是检测精度越高,检测速度相应就会变慢,在此基础上只能进行检测精度和速度之间的良好平衡;由于本文算法增加了一个预测头且在多尺度融合和特征表达能力方面的优化,本文算法复杂度提升,检测精度比起 YOLOv7 算法低了 1.1 个百分点,但 98.9 帧/s 的检测速度明显优于 YOLOv7,完全满足工业视觉的需要;在精度方面,由于 DetectoRS 是两阶段算法,本身就是为了提升检测精度大幅度牺牲了检测速度的算法,虽然在精度上高于本文算法 1.3 个百分点,但是检测速度低于本文算法 83.3 帧。本文算法在检测速度遥遥领先的情况下检测精度基本与二阶段算法持平,基于表 3 的结果表明,本文算法能在实时检测速度和精度方面达到平衡,符合自动驾驶对道路目标检测的实际需求。

表 3 与主流算法的对比实验

Table 3 Comparative experiments with mainstream algorithms

算法	mAP@0.5/%	FPS/(帧·s^{-1})
Faster RCNN	56.7	15.9
SDD	60.4	68.2
YOLOv4-tiny	79.8	94.5
YOLOX	85.3	49.7
YOLOv7	96.1	70.3
DetectoRS	96.3	15.6
本文算法	95.0	98.9

4 结论

以 YOLOv5 算法为基础,针对复杂交通场景检测目标密度变大导致的漏检和密集人群时道路目标检测效果较差的问题,使用微小物体检测的检测头来增加检测内容和避免漏检,通过使用 SIMAM 来增强模型的特征表达能力,并通过 4 个 SPP 来进行多尺度融合,还运用 CBAM 轻量级模块来进行自适应特征细化,提升检测精度,最后,使用 EIOU 损失函数作为边界框损失函数来提升检测速度。在数据集 BDD100K 上的实验结果显示,本文算法在保持目标检测精度的同时兼具足够快的实时检测速度,检测精度提升了 8.1 个百分点,满足工业需要,可为提升自动驾驶汽车的视觉感知能力提供指导意义。

参 考 文 献

[1] Gao X, Wu Y, Yang K, et al. Vehicle bottom anomaly detection algorithm based on SIFT[J]. Optik, 2015, 126(23): 3562-3566.

[2] Agrawal P, Girshick R, Malik J. Analyzing the performance of multilayer neural networks for object recognition[C]//Computer Vision-ECCV 2014: 13th European Conference. Zurich: Springer International Publishing, 2014: 329-344.

[3] 吴雪, 宋晓茹, 高嵩, 等. 基于深度学习的目标检测算法综述[J]. 传感器与微系统, 2021, 40(2): 4-7.
Wu Xue, Song Xiaoru, Gao Song, et al. Review of target detection algorithms based on deep learning[J]. Transducer and Microsystem Technologies, 2021, 40(2): 4-7.

[4] 侯庆山, 邢进生. 基于 Grad-CAM 与 KL 损失的 SSD 目标检测算法[J]. 电子学报, 2020, 48(12): 2409-2416.
Hou Qingshan, Xing Jinsheng. SSD object detection algorithm based on KL loss and Grad-CAM[J]. Acta Electronica Sinica, 2020, 48(12): 2409-2416.

[5] 王志斌, 冯雷, 张少波, 等. 基于改进 YOLOv5 和视频图像的车型识别[J]. 科学技术与工程, 2022, 22(23): 10295-10300.
Wang Zhibin, Feng Lei, Zhang Shaobo, et al. Vehicle type recognition based on improved YOLOv5 and video images[J]. Science Technology and Engineering, 2022, 22(23): 10295-10300.

[6] 郭朦, 陈紫强, 邓鑫, 等. 基于 YOLOv5l 和 ViT 的交通标志检测识别方法[J]. 科学技术与工程, 2022, 22(27): 12038-12044.
Guo Meng, Chen Ziqiang, Deng Xin, et al. Traffic sign detection and recognition method based on YOLOv5l and ViT[J]. Science Technology and Engineering, 2022, 22 (27): 12038-12044.

[7] Zhu X, Lü S, Wang X, et al. TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. New York: IEEE, 2021: 2778-2788.

[8] Liu S, Qi L, Qin H, et al. Path aggregation network for instance segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2018: 8759-8768.

[9] Xie L, Huang C. A residual network of water scene recognition based on optimized inception module and convolutional block attention module[C]//2019 6th International Conference on Systems and Informatics(ICSAI). Shanghai: IEEE, 2019: 1174-1178.

[10] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2016: 770-778.

[11] Qin X, Li N, Weng C, et al. Simple attention module based speaker verification with iterative noisy label detection[C]//IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). New York: IEEE, 2022: 6722-6726.

[12] Cholakkal H, Johnson J, Rajan D. Backtracking spatial pyramid pooling-based image classifier for weakly supervised top-down salient object detection[J]. IEEE Transactions on Image Processing, 2018, 27(12): 6064-6078.

[13] Wang X, Song J. ICIoU: improved loss based on complete intersection over union for bounding box regression[J]. IEEE Access, 2021, 9: 105686-105695.

[14] Ahmed F, Tarlow D, Batra D. Optimizing expected intersection-over-union with candidate-constrained CRFs[C]//Proceedings of the IEEE International Conference on Computer Vision. New York: IEEE, 2015: 1850-1858.

[15] Sun X, Wu P, Hoi S C H. Face detection using deep learning: an improved faster RCNN approach[J]. Neurocomputing, 2018, 299: 42-50.

[16] Hu X, Liu Y, Zhao Z, et al. Real-time detection of uneaten feed pellets in underwater images for aquaculture using an improved YOLO-v4 network[J]. Computers and Electronics in Agriculture, 2021, 185: 106135.

[17] Ji W, Pan Y, Xu B, et al. A real-time apple targets detection method for picking robot based on ShufflenetV2-YOLOX[J]. Agriculture, 2022, 12(6): 856.

[18] Wang C Y, Bochkovskiy A, Liao H Y M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[J]. arXiv Preprint, 2022: arXiv: 2207.02696.

[19] Qiao S, Chen L C, Yuille A. Detectors: detecting objects with recursive feature pyramid and switchable atrous convolution[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2021: 10213-10224.