

改进 YOLOv5s 的自动驾驶汽车目标检测^①



余以春, 李明旭

(西南交通大学 计算机与人工智能学院, 成都 611756)
通信作者: 余以春, E-mail: yuyichun@my.swjtu.edu.cn

摘要: 在自动驾驶领域, 由于道路背景复杂以及小目标信息缺失, 现有目标检测算法存在检测精度低的问题. 由于车载摄像头视角较为固定, 道路上的目标在图像空间中的分布具有一定的规律, 可以为自动驾驶汽车进行目标检测提供更为丰富的信息. 因此, 提出一种改进 YOLOv5s 的空间特征增强网络 (SE-YOLOv5s). 在 YOLOv5s 的颈部网络中添加位置注意力模块 (location attention module, LAM), 该模块能够根据道路目标在图像中的分布特征进行加权, 增强网络对目标类别位置分布的感知和定位能力. 设计一种小目标增强模块 (small target enhancement module, STEM), 将浅层特征和深层特征进行融合, 可以获得更丰富的小目标语义信息和空间细节信息, 提高小目标检测效果. 实验结果表明, 改进模型对不同尺度目标检测精度均有所提高, AP_S 提高 2.8%, AP_M 提高 2.5%, AP_L 提高 2%.

关键词: 目标检测; 注意力机制; 小目标; YOLOv5

引用格式: 余以春, 李明旭. 改进 YOLOv5s 的自动驾驶汽车目标检测. 计算机系统应用, 2023, 32(9):97-105. <http://www.c-s-a.org.cn/1003-3254/9198.html>

Improved YOLOv5s for Autonomous Vehicle Target Detection

YU Yi-Chun, LI Ming-Xu

(School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu 611756, China)

Abstract: In the field of automatic driving, existing target detection algorithms are haunted by low detection precision due to complicated road backgrounds and insufficient information about small targets. Since the onboard camera has fixed viewing angles, and targets on the road are somewhat regularly distributed in the image space, richer information can be provided to autonomous vehicles for target detection. Therefore, a spatial feature augmentation network (SE-YOLOv5s) to improve YOLOv5s is proposed. A location attention module (LAM) is added to the neck network of YOLOv5s, which can be weighted according to the distribution characteristics of road targets in the image and enhance the network's perception and localization ability for the target category location distribution. A small target enhancement module (STEM) is designed to fuse shallow features and deep ones, so as to obtain richer semantic information and detailed space information about small targets, thereby improving the detection effect of small targets. The results of the experiment show that the improved model witnesses an increase in detection precision against targets of different scales, with AP_S increased by 2.8%, AP_M increased by 2.5%, and AP_L increased by 2%.

Key words: target detection; attention mechanism; small target; YOLOv5

随着经济的发展, 道路上的车辆数量也显著增加, 道路上的行车压力也越来越大, 随之而来的交通事故

也逐年递增. 自动驾驶能够辅助甚至代替驾驶员控制车辆安全运行, 能够减少交通事故的发生, 缓解和改善

① 基金项目: 国家自然科学基金 (61961038)

收稿时间: 2023-02-12; 修改时间: 2023-03-08; 采用时间: 2023-03-14; csa 在线出版时间: 2023-07-14

CNKI 网络首发时间: 2023-07-17

道路拥堵状况,是未来的发展趋势^[1].环境感知系统利用雷达和摄像头等多种硬件传感器为自动驾驶系统提供可靠的外部环境判断,是自动驾驶系统进行控制和决策的基础.通过车载前视摄像头进行目标检测是环境感知的重要组成部分,已经成为目前研究热点^[2].

在自动驾驶场景中,由于车载摄像机视角相对固定,拍摄得到的道路图像具有特殊的空间性质.图像底部主要由道路组成,中间部分包含各种相对较小的物体,而顶部主要包括建筑物和天空,并且不同类别在图像中的分布也不均匀^[3].本文统计了BDD100K数据集^[4]中不同类别在图像中的位置分布,如图1所示.从图1中可以看到,交通信号灯(traffic light)和交通标识(traffic sign)主要集中在图像顶部区域,汽车(car)在图像中下部道路区域出现的概率更高,而行人(person)主要分布在图像两侧的非行驶区域或道路边缘.此外,

因为车载摄像头存在透视投影,导致同样大小的物体离摄像头越近成像越大,而离摄像头越远成像越小.在这种情况下,道路目标的尺寸与其在图像中的位置有很强的相关性^[5].如图2所示,小目标主要集中在图像高度比例为0.4–0.7,宽度比例为0.3–0.6区域,这意味着小目标往往出现在图像的中上区域.

目前流行的目标检测算法分为两类:以R-CNN系列^[6]为代表的两阶段检测算法和以YOLO系列^[7]、SSD^[8]、RetinaNet^[9]为代表的单阶段检测算法.两阶段目标检测算法先进行区域提议再进行分类和定位,而单阶段目标检测算法直接由图像特征得到目标的类别概率和检测框位置.虽然两阶段目标检测算法检测精度高但检测效率较差,无法满足自动驾驶高实时性要求.同时,单阶段检测算法速度较快但检测精度低于两级检测器^[10].

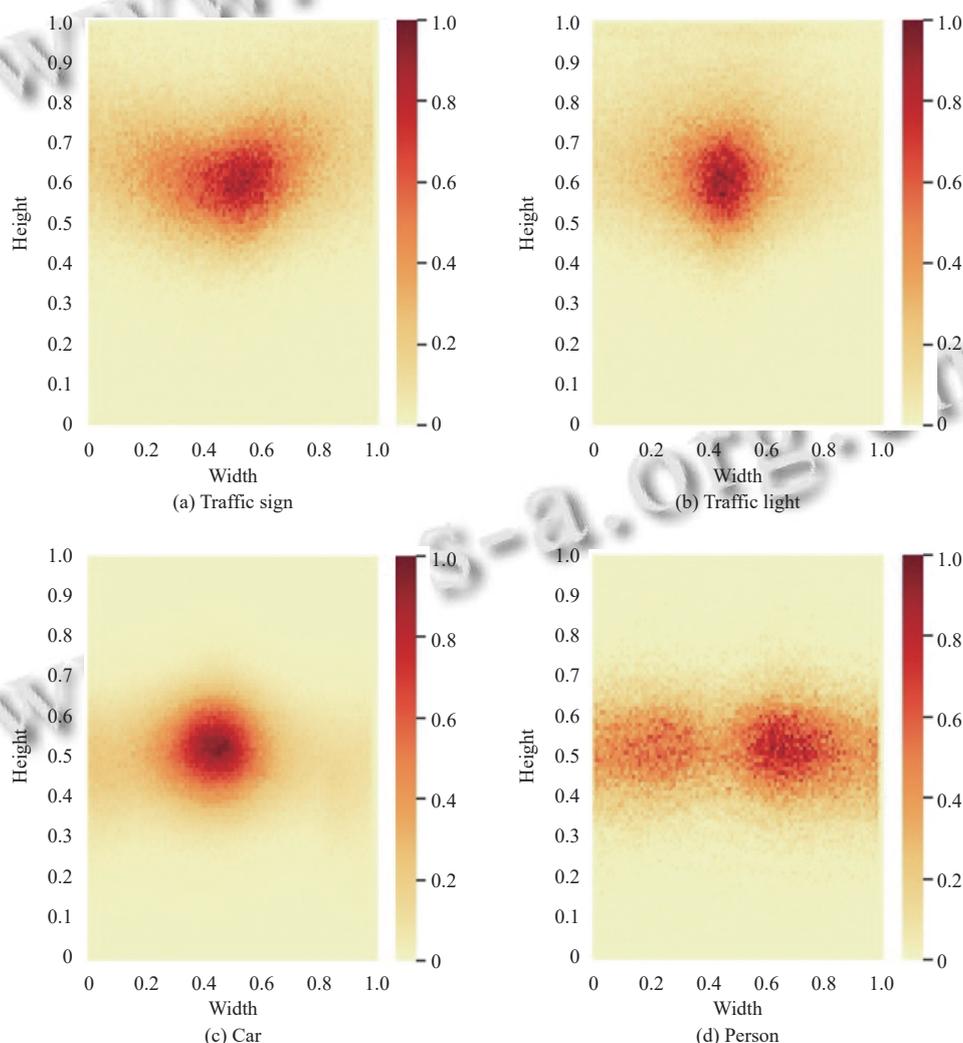


图1 BDD100K 类别分布热力图

自动驾驶计算平台的计算资源有限,需要同时处理检测、跟踪、决策等多种传感和计算任务.这要求应用在车端的检测算法需要具有较小的内存占用率和计算资源占用率^[11].因此,如何在保持实时目标检测能力的同时,提升检测精度是近几年的研究重点. Cai 等人^[11]改进了YOLOv4算法,使用可变形卷积代替特征提取网络的最后一个输出层并且设计了一个新的特征融合模块进行特征融合,提高了平均检测精度. Zhu 等人^[12]提出一种新的目标检测网络TPH-YOLOv5.在YOLOv5的基础上集成卷积注意力模块,使模型在检测不同尺度物体时能关注重要的特征,提升特征提取能力. Chen 等人^[13]在YOLOv3-tiny上进行裁剪,量化网络中的参数,降低了算法运行复杂度. 刘革等人^[14]用MobileNet V3替换RetinaNet的特征提取网络,引入GIoU改进损失函数,提高了网络特征提取能力. 李彦辰等人^[15]以Efficientdet为基础框架,设计多尺度注意力机制模块应用于特征融合网络,提高了对小目标的检测精度.

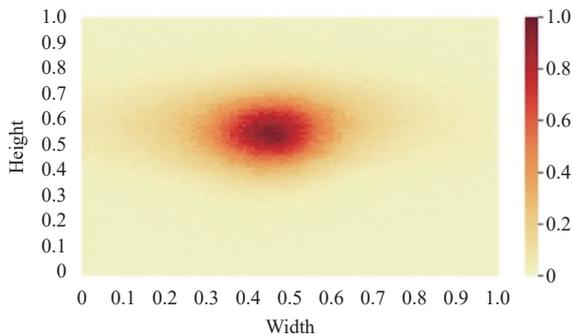


图2 BDD100K 小目标分布热力图

综上所述,虽然对于自动驾驶领域目标检测已经取得一定成果,但是缺少针对自动驾驶场景特有的空间特征进行优化设计,没有很好地利用这一关键特征.并且由于小目标面积小、图像模糊、易受噪声影响、网络卷积层数过深等因素,导致很难有效提取小目标特征,检测精度有待进一步提高.因此,提出一种改进YOLOv5的空间特征增强网络SE-YOLOv5,主要改进点如下.

(1) 为了增强网络对目标类别位置分布的感知和定位能力,提出了一种位置注意力模块.它能够提取元素级空间位置特征图,根据像素的空间位置自适应调整图像特征权重,增强网络特征提取能力.

(2) 针对小目标信息缺失的问题,提出一种小目标增强模块.通过自下而上以及自上而下的双向特征调制,结合深层特征的语义信息以及浅层特征的细节信息,获得更丰富的小目标特征,提高网络对小目标的检测性能.

1 网络改进

1.1 改进YOLOv5s网络结构

YOLOv5是一种单阶段目标检测算法,具有部署成本低、灵活度高和检测速度快的特点.其网络结构可以分为输入端(Input)、主干网络(Backbone)、颈部网络(Neck)、检测网络(Head)这4个部分.输入端主要完成图像增强、锚框计算、图像缩放等预处理操作;主干网络使用CSPDarknet53和SPPF提取图像特征;PANet结构作为颈部网络用于聚合特征;最后在检测网络对3种尺度特征图进行目标的分类和预测.此外,YOLOv5根据网络深度和宽度可以分为依次递增的4个版本,分别为YOLOv5s、YOLOv5m、YOLOv5l以及YOLOv5x.

综合考虑算法检测速度和精度,本文选用YOLOv5s作为基础网络,在颈部网络中添加位置注意力模块,帮助网络感知道路目标在图像中的空间位置分布,增强网络特征提取能力;在主干网络中添加小目标增强模块,丰富检测网络中小目标特征信息,提高网络小目标检测精度,改进后的SE-YOLOv5s网络结构如图3所示.

1.2 位置注意力模块

本文在HANet^[3]的基础上设计了一种位置注意力模块,它能够提取自动驾驶场景中空间位置注意力特征,对不同通道特征以及空间位置进行加权,增强网络对目标类别和位置分布的感知和定位能力,模块结构如图4所示.给定特征图 $X \in \mathbb{R}^{C \times H \times W}$,位置注意力特征图A可由以下步骤得到.

(1) 特征压缩.首先,经过全局平均池化聚合高度和宽度方向上的上下文,接着通过 1×1 的卷积在通道维度上对特征图进行压缩,剔除冗余通道特征,减少参数量,如式(1)所示.

$$\hat{X} = \text{Conv}(\text{AvgPool}(X)) \quad (1)$$

其中,AvgPool为全局平均池化函数;Conv为卷积函数; $\hat{X} \in \mathbb{R}^{C/r \times \hat{H} \times \hat{W}}$ 为压缩特征, \hat{H} 和 \hat{W} 为全局平均池化后特征图宽高, r 为通道压缩倍率.

(2) 计算位置注意力特征图. 为了生成像素级位置注意力特征图, 同时使用通道注意力和空间注意力, 以捕获像素间的空间上下文相关性和通道间的语义相关性. 同时, 为了提高计算效率, 使用并联提取注意力特征, 减少计算时间成本. 其中, 通道注意力计算过程如式(2)–式(4)所示.

$$G_{avg}(X) = \sigma(FC(\delta(FC(AvgPool(X)))))) \quad (2)$$

$$G_{max}(X) = \sigma(FC(\delta(FC(MaxPool(X)))))) \quad (3)$$

$$A_c(X) = G_{avg}(X) + G_{max}(X) \quad (4)$$

其中, FC 为全连接层; δ 为 ReLU 函数; σ 为 Sigmoid 函数; $MaxPool$ 为最大池化.

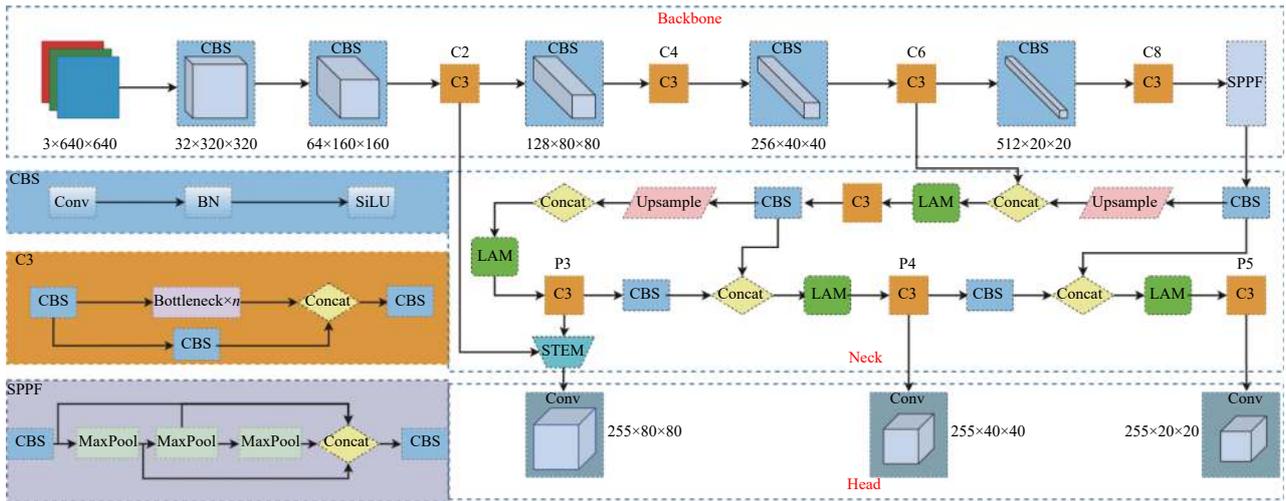


图3 SE-YOLOv5s 网络结构

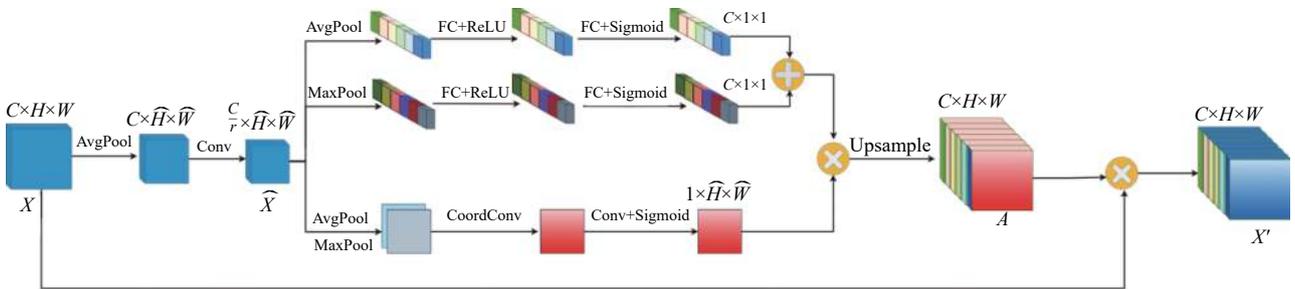


图4 位置注意力模块结构

在计算空间注意力权重时, 引入 CoordConv 模块^[16]. CoordConv 通过在输入特征中增加两个坐标通道, 增强卷积的空间特征提取能力, 能更好地学习自动驾驶场景中的空间位置分布. 空间注意力可用如式(5)表示.

$$A_s(X) = \sigma(Conv(Cd(AvgPool(X), MaxPool(X)))) \quad (5)$$

其中, Cd 为 CoordConv 模块. 最后, 结合通道注意力特征图和空间注意力特征图, 经过逐元素相乘以及上采样生成与位置注意力特征图 $A \in \mathbb{R}^{C \times H \times W}$.

$$A = Upsample(A_c(X) \otimes A_s(X)) \quad (6)$$

其中, $Upsample$ 为上采样函数; \otimes 为矩阵点乘.

1.3 基于上下文的小目标增强模块

为了提高小目标的检测精度, 本文在非对称上下文调制模块^[17]的基础上进行改进, 设计了一种小目标增强模块, 如图5所示.

如图5所示, 小目标增强模块主要由自下而上的逐点通道上下文调制模块和自上而下的全局上下文调制模块构成, 通过对浅层高分辨率特征以及深层高级语义特征进行多尺度交换, 以获得包含小目标细节信息以及语义信息的增强特征.

1.3.1 自下而上的通道上下文调制

如图6所示, 自下而上的上下文调制为一个逐点通道注意力, 其中每个空间位置的通道特征上下

文被单独聚合. 该模块利用浅层特征中丰富的小目标细节信息自下而上的生成空间细节上下文调制权

重, 弥补深层特征中缺失的小目标细节信息, 主要流程如下.

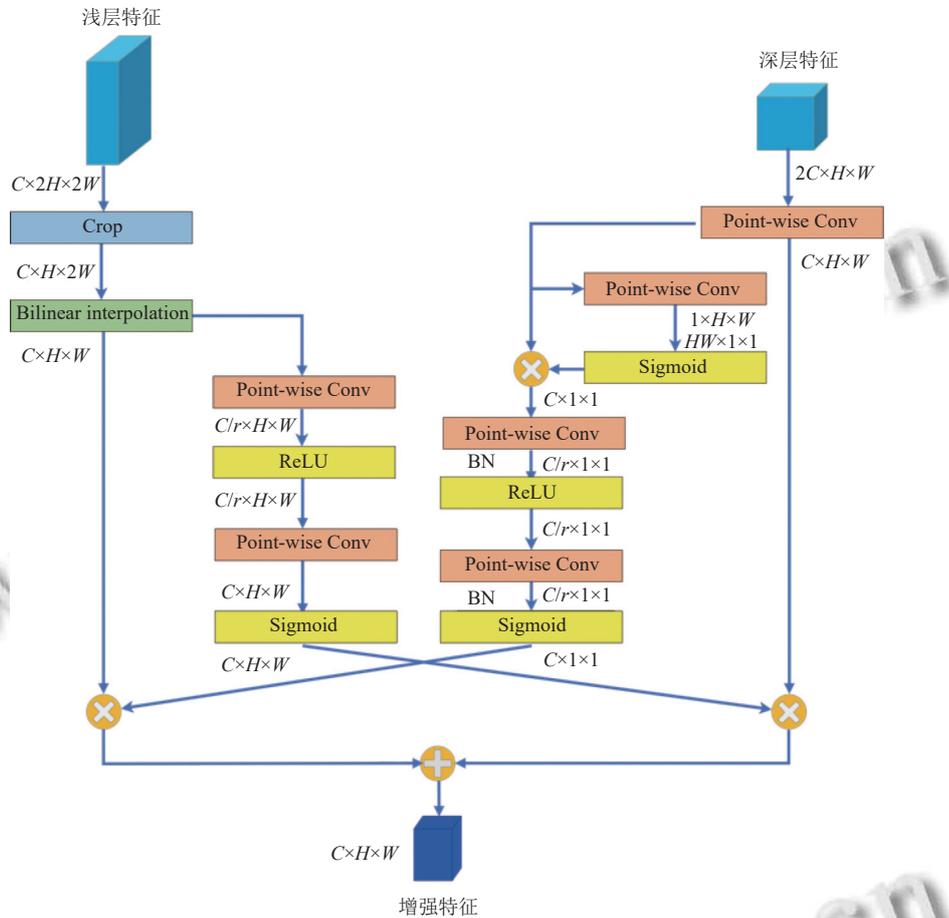


图5 小目标增强模块结构

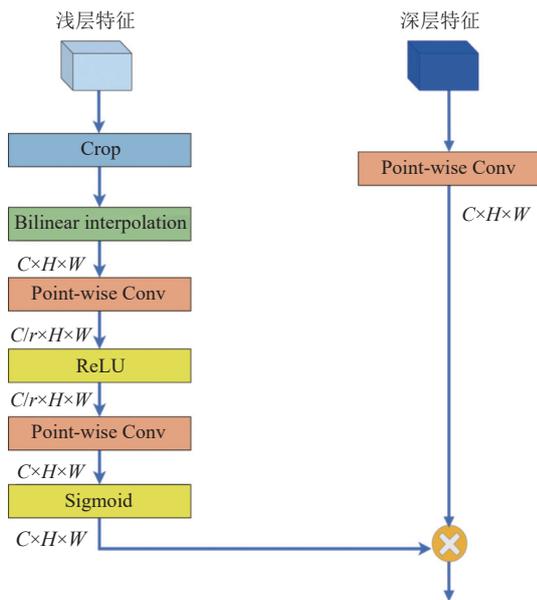


图6 自下而上调制模块结构

(1) 特征裁剪. 首先对浅层特征在高度比例为 0.4-0.7, 宽度比例为 0.3-0.6 区域进行裁剪, 获取包含小目标的特征区域, 并减少冗余特征和参数量.

(2) 尺度匹配. 考虑到特征尺寸匹配问题, 采用双线性插值算法生成与 Y 同样宽高的特征图, 如式 (7) 所示.

$$\alpha(X) = BI(Crop(X)) \quad (7)$$

(3) 计算调制权重. 经过两次逐点卷积和激活函数获得调制权重, 最终经过逐点相乘得到调制后的高级特征 $Y' \in \mathbb{R}^{C \times H \times W}$.

$$Y' = \sigma(PWC(\delta(PWC(\alpha(X)))) \otimes PWC(Y) \quad (8)$$

其中, $X \in \mathbb{R}^{C \times 2H \times 2W}$ 为浅层特征, $Y \in \mathbb{R}^{2C \times H \times W}$ 为深层特征; $Crop$ 为特征裁剪函数; BI 为插值函数; PWC 为逐点卷积.

1.3.2 自上而下的全局上下文调制

自上而下的上下文调制模块通过全局上下文信息来指导浅层特征,并将全局语义特征嵌入到浅层特征中,减少浅层特征噪声信息的干扰.模块结构如图7所示,调制步骤如下.

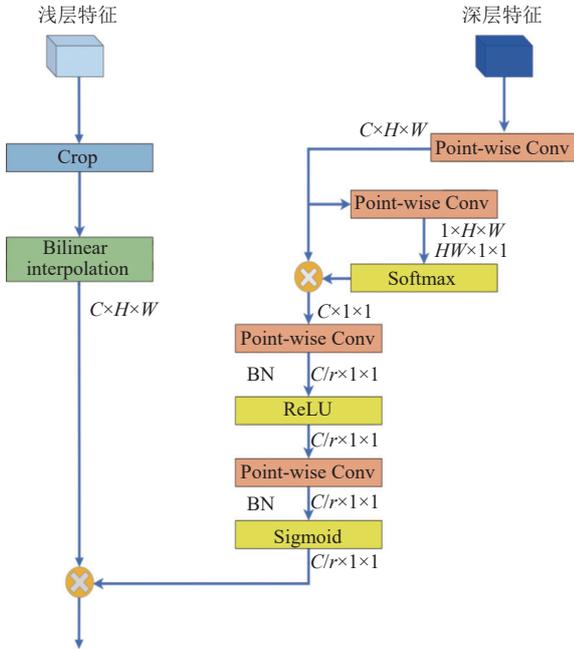


图7 自上而下调制模块结构

1) 建模全局上下文信息. 使用了简化版的全局上下文模块(global context, GC)^[18]来建模全局上下文信息,如式(9)所示.

$$\beta(Y) = PWC(Y) \otimes \rho(PWC(PWC(Y))) \quad (9)$$

2) 瓶颈转换. 将全局上下文特征通过卷积捕获通道间依赖,如式(10)所示.

$$\gamma(Y) = \sigma(PWC(\delta(PWC(Y)))) \quad (10)$$

3) 特征融合. 使用乘法融合策略将上下文特征嵌入到浅层特征中,实现特征融合.调制特征可以由式(11)得到.

$$X' = \alpha(X) \otimes \gamma(\beta(Y)) \quad (11)$$

其中, ρ 为Softmax函数, $X' \in \mathbb{R}^{C \times H \times W}$ 为调制特征.

2 实验分析

2.1 实验数据集和实验设置

BDD100K数据集是2018年由伯克利大学发布的大规模、多样化自动驾驶数据集,涵盖一天中不同时间、天气条件和驾驶场景.本文在BDD100K数据集

的基础上筛选一些不均衡的类别,最终选取其中46500张图片进行训练,按照7:2:1比例随机分配,其中训练集32550张图片,验证集9300张图片,测试集4650张图片,重新分类为汽车、大车、人、两轮车、交通灯和交通标识共6种道路目标.

实验使用NVIDIA GeForce RTX 3090显卡进行模型训练和测试,CPU为Intel(R) Xeon(R) Platinum 8255C.系统和软件环境为Ubuntu 22.04,Python 3.8.10,深度学习框架为PyTorch 1.10.2,训练配置如表1所示.

表1 模型训练配置

参数名称	参数值
输入图像尺寸	640×640
批处理大小	16
优化器	SGD
初始学习率	0.01
权重衰减系数	0.0005
动量	0.937
IoU阈值	0.5
训练轮次	200

2.2 评价指标

为了更好地观察网络对不同尺度目标的检测效果,使用COCO评价指标进行评价,具体名称和含义如表2所示.

表2 模型训练配置

指标名称	含义
AP _{0.5:0.95}	IoU在[0.5:0.05:0.95]条件下的mAP
AP _{0.5}	网络在IoU为0.5条件下的mAP
AP _{0.75}	网络在IoU为0.75条件下的mAP
AP _S	面积小于32×32目标的mAP
AP _M	面积在32×32到96×96目标的mAP
AP _L	面积大于96×96目标的mAP

2.3 消融实验

为了证明改进模块的有效性,设置了4组对比实验,如表3所示,其中“√”表示在基准网络中加入该模块.从表3中可以看到,YOLOv5s+LAM修改YOLOv5s的Neck部分,在每个Concat模块后添加位置注意力模块,相比于YOLOv5s,AP_S,AP_M,AP_L均有一定提高,分别为0.7%,1.6%,1.2%.说明位置注意力模块能够学习道路目标的空间位置分布规律,增强网络特征提取能力,整体提高网络检测性能.YOLOv5s+STEM在YOLOv5s中增加小目标增强分支,与YOLOv5s相比,AP_S,AP_M分别提高2.6%以及1.2%,AP_L基本保持不变.这表明小目标增强模块在不影响中大目标检测的

情况下,能够充分结合浅层丰富的小目标空间细节信息以及深层高级语义信息,得到更为有效的小目标特征,提高对小对象的检测性能.结合位置注意力模块和小目标增强模块的改进网络检测性能有显著的提高,AP_S提高了2.8%,AP_M提高了2.5%,AP_L提高了2%.由此可见,改进后的SE-YOLOv5s网络能够有效提高对不同尺度道路目标的检测精度.

表3 消融实验结果(%)

模块		AP _{0.5:0.95}	AP _{0.5}	AP _{0.75}	AP _S	AP _M	AP _L
LAM	STEM						
—	—	25.6	50.0	23.3	12.4	32.1	41.0
√	—	26.5	51.6	24.1	13.1	33.7	42.2
—	√	27.1	51.9	25.0	15.0	33.3	41.1
√	√	27.9	53.0	25.9	15.2	34.6	43.0

表4 网络对比结果

模型	主干网络	输入尺寸	mAP@0.5 (%)	计算量(10 ⁹)	参数量(10 ⁶)	帧率(FPS)
RetinaNet	ResNet-50	608×1024	39.1	61.82	36.43	20
SSD	VGG-16	300×300	35.7	30.74	24.40	44
YOLOv3-tiny ^[19]	Darknet53	416×416	24.5	13.01	8.60	285
YOLOv4-tiny ^[20]	CSPDarknet53	416×416	28.2	16.20	5.80	182
YOLOXs ^[21]	Darknet53	640×640	50.6	26.65	8.94	73
YOLOv5m	CSPDarknet53+SPPF	640×640	54.6	48.10	20.90	45
SE-YOLOv5s	CSPDarknet53+SPPF	640×640	53.0	16.56	7.26	63

2.5 可视化实验

本文在BDD100K数据集中挑选出包含不同类别的道路图片输入YOLOv5s网络,然后对网络检测层前一层3个卷积特征图进行Grad-CAM^[22]可视化.可视化得到的热力图对比结果如图8所示,其中左侧为YOLOv5s检测热力图,右侧为添加LAM模块后检测热力图.热力图中绿色目标框中的目标为预测目标,像素对预测目标贡献的高低用冷暖颜色区分,颜色较暖的红色表示对预测目标贡献度高,颜色较冷的蓝色表示对预测目标贡献度低.

从热力图对比结果中可以看出YOLOv5s网络的注意力为分散,并且对一些目标类别不会出现的地方也会产生响应.比如在图8(a)–(c)中,需要识别的对象为道路上的目标,然而网络在其不会出现的天空区域产生了错误的响应;而在图8(d)中,需要预测的目标为交通标识,它往往出现在图像顶部区域,而图像底部地面区域却分散了网络的注意力.这些错误位置的特征响应会降低特征的有效性和鲁棒性,导致预测目标分类和定位准确度降低,甚至出现违反常理的预测结果.

2.4 对比实验

为了检验改进网络在道路目标检测算法中的性能,选择目前主流的道路目标检测网络进行比较,在相同数据集上的实验结果如表4所示.从表4可知,YOLOv3-tiny和YOLOv4-tiny网络检测速度很快,但检测精度较低;YOLOv5m在YOLOv5s网络的基础上加宽加深,其检测效果最好,但网络计算量和参数量较大,检测帧率相较于YOLOv5s下降明显,无法满足自动驾驶的实时性需求;改进网络与RetinaNet、SSD以及YOLOXs网络相比,计算量和参数量更低,并且平均精度有明显领先.由此可见,改进的空间特征增强网络在显著提高检测精度的同时检测帧率仍满足实时性要求,综合性能优于目前其他主流检测算法.

添加了LAM模块改进后的网络对预测目标的注意力更为集中,显著减少了各类别对错误区域的特征响应.这表明LAM模块能够很好地学习数据集中各类别的空间分布特征,并根据目标类别抑制错误空间区域的响应,帮助网络聚焦关键区域,从而提高网络的分类和定位的准确度.

2.6 网络改进前后对比

为了更好地分析改进模块对网络整体性能的影响,改进前后网络的mAP@0.5以及总损失(totalLoss)对比图如图9所示,其中总损失为分类损失、置信度损失、定位损失之和.从图9(a)、(b)中的曲线变化可以发现,改进后的网络在训练125轮时收敛,mAP@0.5达到最大,并且在训练过程中mAP@0.5整体高于改进前的YOLOv5s网络,总体损失值也更低.

改进前后网络在测试集上的实际检测效果对比图如图10所示.从图10可以看到,改进后的SE-YOLOv5s网络有更好的检测效果,可以检测出YOLOv5s漏检和错检的道路目标,并且在不同光照以及复杂道路场景下鲁棒识别目标.

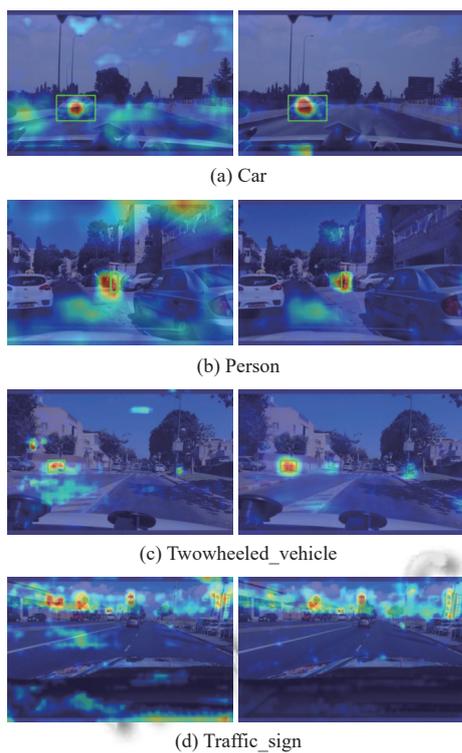
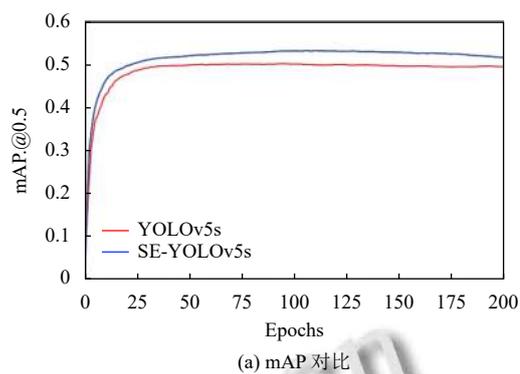
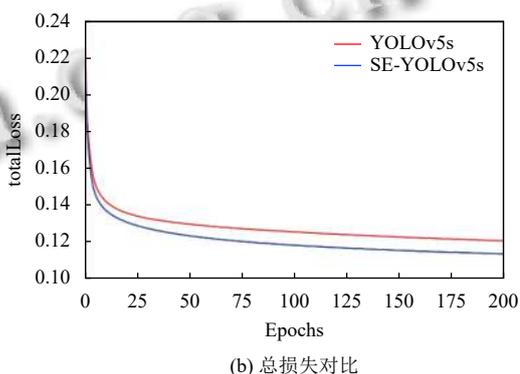


图 8 热力图对比



(a) mAP 对比



(b) 总损失对比

图 9 网络改进前后 mAP 和总损失对比

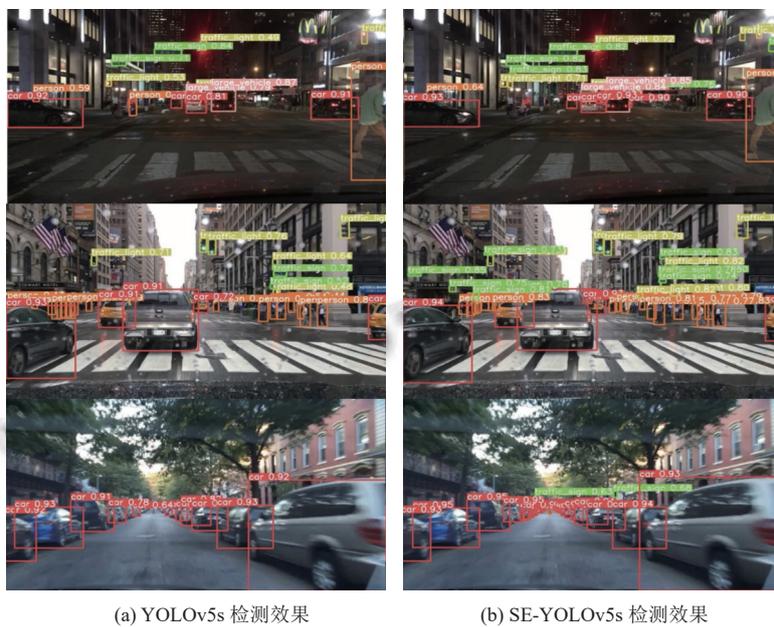


图 10 模型检测效果对比

3 结论与展望

本文在 YOLOv5s 网络的基础上,通过分析自动驾驶场景下目标类别和尺寸的空间分布规律,提出了一种空间特征增强网络 SE-YOLOv5s,成功提高了网络

对不同尺度道路目标的检测性能.为了增强网络的空间特征提取能力,在颈部网络中添加位置注意力模块,帮助网络感知道路目标在图像中的空间位置分布.针对检测层中小目标细节信息缺失的问题,添加小目标

增强模块, 结合浅层和深层特征进行多尺度融合, 丰富检测层的小目标特征信息. 实验结果表明, 改进后的 SE-YOLOv5s 网络在没有过多增加计算量和参数量的情况下, 显著提高了对道路目标的检测精度, 优于改进前的 YOLOv5s 网络.

参考文献

- 1 章军辉, 陈大鹏, 李庆. 自动驾驶技术研究现状及发展趋势. 科学技术与工程, 2020, 20(9): 3394–3403. [doi: [10.3969/j.issn.1671-1815.2020.09.005](https://doi.org/10.3969/j.issn.1671-1815.2020.09.005)]
- 2 Dai XR. HybridNet: A fast vehicle detection system for autonomous driving. Signal Processing: Image Communication, 2019, 70: 79–88. [doi: [10.1016/j.image.2018.09.002](https://doi.org/10.1016/j.image.2018.09.002)]
- 3 Choi S, Kim JT, Choo J. Cars can't fly up in the sky: Improving urban-scene segmentation via height-driven attention networks. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 9373–9383.
- 4 Yu F, Chen HF, Wang X, et al. BDD100K: A diverse driving dataset for heterogeneous multitask learning. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 2636–2645.
- 5 Carranza-García M, Lara-Benítez P, García-Gutiérrez J, et al. Enhancing object detection for autonomous driving by optimizing anchor generation and addressing class imbalance. Neurocomputing, 2021, 449: 229–244. [doi: [10.1016/j.neucom.2021.04.001](https://doi.org/10.1016/j.neucom.2021.04.001)]
- 6 Girshick R. Fast R-CNN. Proceedings of the 2015 IEEE International Conference on Computer Vision. Santiago: IEEE, 2015. 1440–1448.
- 7 Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 779–788.
- 8 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.
- 9 Lin TY, Goyal P, Girshick R, et al. Focal loss for dense object detection. Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017. 2980–2988.
- 10 Zou ZX, Chen KY, Shi ZW, et al. Object detection in 20 years: A survey. Proceedings of the IEEE, 2023, 111(3): 257–276. [doi: [10.1109/JPROC.2023.3238524](https://doi.org/10.1109/JPROC.2023.3238524)]
- 11 Cai YF, Luan TY, Gao HB, et al. YOLOv4-5D: An effective and efficient object detector for autonomous driving. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 4503613.
- 12 Zhu XK, Lyu SC, Wang X, et al. TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision Workshops. Montreal: IEEE, 2021. 2778–2788.
- 13 Chen SB, Lin W. Embedded system real-time vehicle detection based on improved YOLO network. Proceedings of the 3rd IEEE Advanced Information Management, Communicates, Electronic and Automation control Conference. Chongqing: IEEE, 2019. 1400–1403.
- 14 刘革, 郑叶龙, 赵美蓉. 基于 RetinaNet 改进的车辆信息检测. 计算机应用, 2020, 40(3): 854–858.
- 15 李彦辰, 张小俊, 张明路, 等. 基于改进 Efficientdet 的自动驾驶场景目标检测. 计算机工程与应用, 2022, 58(6): 183–191. [doi: [10.3778/j.issn.1002-8331.2109-0190](https://doi.org/10.3778/j.issn.1002-8331.2109-0190)]
- 16 Liu R, Lehman J, Molino P, et al. An intriguing failing of convolutional neural networks and the coordconv solution. Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montréal: Curran Associates Inc., 2018. 9628–9639.
- 17 Dai YM, Wu YQ, Zhou F, et al. Asymmetric contextual modulation for infrared small target detection. Proceedings of the 2021 IEEE Winter Conference on Applications of Computer Vision. Waikoloa: IEEE, 2021. 949–958.
- 18 Cao Y, Xu JR, Lin S, et al. GCNet: Non-local networks meet squeeze-excitation networks and beyond. Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision Workshop. Seoul: IEEE, 2019. 1971–1980.
- 19 Redmon J, Farhadi A. YOLOv3: An incremental improvement. arXiv:1804.02767, 2018.
- 20 Bochkovskiy A, Wang CY, Liao HYM. YOLOv4: Optimal speed and accuracy of object detection. arXiv:2004.10934, 2020.
- 21 Ge Z, Liu ST, Wang F, et al. YOLOX: Exceeding YOLO series in 2021. arXiv:2107.08430, 2021.
- 22 Selvaraju RR, Cogswell M, Das A, et al. Grad-CAM: Visual explanations from deep networks via gradient-based localization. Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017. 618–626.

(校对责编: 孙君艳)