

文章编号:1671-6833(2024)03-0064-08

# 基于改进 YOLOv5s 的复杂道路交通目标检测算法

汤林东<sup>1,2</sup>, 云利军<sup>1,2</sup>, 罗瑞林<sup>3</sup>, 卢琳<sup>3</sup>

(1. 云南师范大学 信息学院, 云南 昆明 650500; 2. 云南师范大学 云南省教育厅计算机视觉与智能控制技术工程研究中心, 云南 昆明 650500; 3. 云南省烟草烟叶公司, 云南 昆明 650500)

**摘要:** 针对目前自动驾驶场景下交通目标检测算法抗复杂背景干扰能力弱, 导致检测性能不足的问题, 提出了一种改进 YOLOv5s 的复杂道路交通目标检测算法。首先, 在特征提取区域, 采用多头自注意残差模块(MHSARM)来强化待检目标特征信息, 弱化复杂背景干扰; 其次, 在特征融合区域, 采用 CoordConv 代替传统 Conv, 使网络具备空间信息感知能力, 提升网络检测精度。在开源数据集 Kitti 及 BDD100K 上的实验结果表明: 改进 YOLOv5s 算法在复杂道路中具备更强的特征提取能力及良好的泛化能力,  $mAP_{0.5}$  分别达到 93.3% 和 47.4%, 与 YOLOv5s 相比, 分别提升了 0.9% 和 1.4%。另外, 改进 YOLOv5s 相较于目前最新的目标检测算法 YOLOv7、YOLOv8,  $mAP_{0.5}$  分别提高了 1.3% 和 2.2%, 与在 Kitti 数据集上最新的研究成果 Sim-YOLOv4 算法相比,  $mAP_{0.5}$  提高了 2.2%。

**关键词:** 自动驾驶; 目标检测; YOLOv5s; MHSARM; CoordConv

**中图分类号:** TP391 **文献标志码:** A **doi:** 10.13705/j.issn.1671-6833.2024.03.016

随着大数据、人工智能等技术的蓬勃发展, 自动驾驶已在全球掀起浪潮<sup>[1]</sup>。目标检测在自动驾驶感知系统中发挥着重要作用。

基于深度学习的目标检测算法通过卷积神经网络自动学习目标的高级语义特征, 具备优异的检测性能。目前主流的深度学习目标检测算法分为单阶段以及两阶段, 其中两阶段目标检测算法代表有 R-cnn<sup>[2]</sup>、Fast R-cnn<sup>[3]</sup>。此类算法首先从图像中提取出感兴趣的候选区域, 固定大小后送入 CNN 中进行特征提取, 最后进行目标分类与回归。但是由于该类算法时间成本高, 不能做到实时检测。2015 年, Redmon 等<sup>[4]</sup>提出了 YOLOv1 网络, 将目标检测算法引入到单阶段的研究中, 此算法将输入分割成一个  $S \times S$  的网格, 让每一个网格单元回归边界框的位置以及目标中心落入该单元的置信度得分。YOLOv1 展示了单阶段目标检测算法的潜力, 自此单阶段目标检测算法的研究迎来新的篇章。

本文以单阶段目标检测算法 YOLOv5s<sup>[5]</sup> 为基础网络, 提出了改进 YOLOv5s 算法, 主要贡献有以下两方面: ①针对复杂的交通背景带来的干扰, 提出

基于多头注意力残差模块(multi-head self-attention residuals module, MHSARM); ②针对传统卷积层的空间感知能力较弱, 导致模型特征提取能力不足的问题, 采用 CoordConv<sup>[6]</sup> 来取代 YOLOv5s 中特征融合层中的传统卷积层。

## 1 相关工作

在交通目标检测领域, 袁志宏等<sup>[7]</sup>在 YOLOv3 中引入 SPP 结构, 进一步提升了检测精度, 但是此算法通过牺牲图片分辨率来换取检测速度的提升, 此方式容易丢失细节特征信息, 在复杂的交通场景并没有起到较好的效果。刘丽伟等<sup>[8]</sup>在 YOLOv4 残差模块中嵌入轻量级 SimAM 注意力模块, 提升网络对重要特征的提取能力, 改善了小目标检测精度不足的问题, 但对于特征信息较弱的交通目标存在大量漏检的现象。院老虎等<sup>[9]</sup>通过在 YOLOv5s 中加 CBAM 注意力机制, 提升网络对遮挡目标与小目标的检测性能。Cai 等<sup>[10]</sup>使用可变形卷积代替 CSP-Daeknet 中最后一层, 通过自适应动态调整卷积核大小, 捕捉更加丰富的局部特征信息, 提升了网络对小

收稿日期: 2023-10-23; 修订日期: 2023-11-26

基金项目: 国家自然科学基金资助项目(62265017)

通信作者: 云利军(1973—), 男, 内蒙古呼和浩特人, 云南师范大学教授, 博士, 博士生导师, 主要从事视频图像处理、计算机视觉、深度学习算法研究, E-mail: yunlijun@ynnu.edu.cn。

引用本文: 汤林东, 云利军, 罗瑞林, 等. 基于改进 YOLOv5s 的复杂道路交通目标检测算法[J]. 郑州大学学报(工学版), 2024, 45(3): 64-71. (TANG L D, YUN L J, LUO R L, et al. Complex road traffic target detection algorithm based on improved YOLOv5s[J]. Journal of Zhengzhou University (Engineering Science), 2024, 45(3): 64-71.)

目标的检测性能。Shi 等<sup>[11]</sup>在原有检测头的基础上增加额外的检测头,并将注意机制与检测器头部融合,进一步提高目标的检测精度。Gao 等<sup>[12]</sup>在 YOLOv5 特征提取区域加入接收场,更好地关注小目标物体,且引入了空间通道注意力机制,减少背景信息干扰,增强网络关注能力。

上述算法虽然能实现交通目标检测,但是对于复杂交通场景下的检测仍然存在以下挑战。①背景多样性。对比其他应用场景,交通场景富含丰富的信息,如天气、光照、背景等,交通目标容易受到各个复杂场景的干扰,导致特征信息衰减。②局部遮挡现象普遍发生。在交通目标密集的交通场景,由于目标之间相互遮挡会使目标特征信息丢失,从而导致被遮挡目标存在重检、漏检现象。

2 改进 YOLOv5s

2.1 YOLOv5s 网络架构

YOLOv5s<sup>[5]</sup>网络为全卷积网络,整体架构如图 1 所示,其网络架构由 Backbone、Neck 和 Head 三个部分组成。Backbone 主要用于特征提取,通过自上而下的路径获取多尺度的特征;Neck 结构是一个特征金字塔(FPN),其目的是把浅层的图像特征与深层的语义特征融合,以获取更加完整的特征;Head 层为 Detect 模块,由卷积层构成,对应三个特征检测层。

YOLOv5 通过调整深度 *depth* 和宽度 *width* 两个参数,可分为五个版本,分别为 YOLOv5n、YOLOv5s、YOLOv5m、YOLOv5l、YOLOv5x。为了平衡精度和速度,本实验采用轻量级版本 YOLOv5s 作为基础模型。

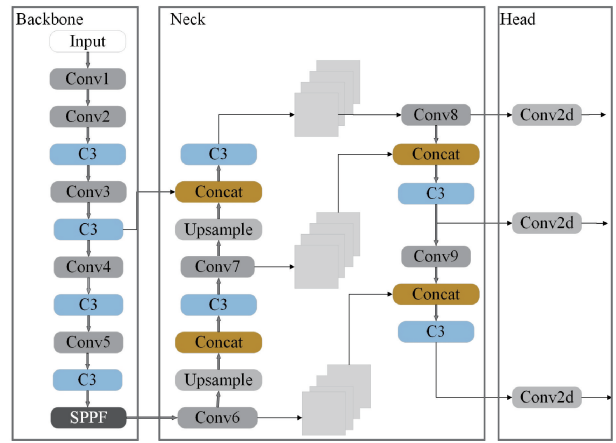


图 1 YOLOv5s 网络架构

Figure 1 YOLOv5s network architecture

2.2 多头自注意力机制 (MHSA)

Self-attention<sup>[13]</sup>更多地关注输入的关键信息,注重特征内部相关性,减少对外部信息的依赖。首

先,通过输入向量  $X$  乘以对应的权重矩阵获取  $Q$ 、 $K$ 、 $V$ ;其次,通过  $Q$  跟  $K$  相乘计算自注意力分数;再次,通过注意力分数除以根号下向量  $K$  的维度调整自注意力分数幅值,并使用 softmax 函数对注意力得分进行操作得到权重系数,确保分数都为正数且分数之和为 1;最后,根据权重系数对  $V$  值进行加权求和,得到最终的输出,如式 (1)<sup>[13]</sup> 所示。

对比 Self-attention, MHSA<sup>[14]</sup>使模型拥有多个表示子空间,进一步丰富了信息的特征表达,其网络结构如图 2 所示。MHSA 使用多组不同注意力权重  $W$  对  $Q$ 、 $K$ 、 $V$  进行线性变换,随后将多组变换结果并行进行自注意机制计算,最后对产生的多组结果进行拼接,并通过线性变化得到最终输出结果,如式 (2)、(3)<sup>[14]</sup> 所示:

Attention(Q,K,V) = softmax(QK^T / sqrt(d\_k)) V。 (1)

{ h\_1 = Attention(QW\_1^Q, QW\_1^K, QW\_1^V); ... h\_i = Attention(QW\_i^Q, QW\_i^K, QW\_i^V)。 (2)

MultiHead(Q,K,V) = Concat(h\_1, h\_2, ..., h\_i)。 (3)

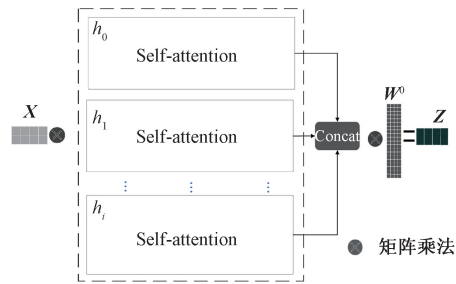


图 2 MHSA 结构图

Figure 2 MHSA structure diagram

2.3 CoordConv

CoordConv 继承传统卷积优异特性,能有效解决传统卷积获取空间信息能力差的问题,其结构如图 3 所示。

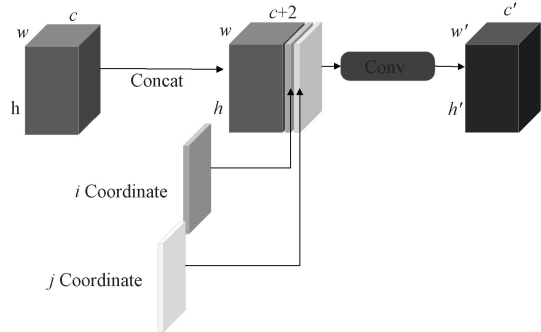


图 3 CoordConv 结构图

Figure 3 CoordConv structure diagram

与传统卷积相比,CoordConv 在输入的 Feature Map 中增加 2 个通道,分别代表  $X$  坐标、 $Y$  坐标,两者的加入使卷积具备了空间信息感知能力,增强交通目标在网络中的信息表达。

### 2.4 MHSA-Bottleneck 方法

Bottleneck<sup>[5]</sup> 是一种精简的网络结构。首先,Bottleneck 利用  $1\times 1$  的卷积降低特征通道;其次,通过  $3\times 3$  的卷积将通道数加倍;最后,采用残差结构将输入输出相加。该模块通过先降维再升维的方式,不仅减小模型参数量,降低模型的推理时间,还使卷积核能更好地理解特征信息,有利于提取到更多的细节特征。MHSA 通过相关像素之间的相似函

数动态计算注意力权重,以适应关注不同的区域,并捕捉到更多的特征。

如图 4 金色区域所示,MHSA-Bottleneck 的原理是在 Bottleneck 中  $3\times 3$  卷积后接入 MHSA,考虑实时性跟检测精度之间的平衡,本次研究采用的是 4 头 MHSA,其优势体现在以下三点:①保证模型不会增加过多的参数量导致实时性受到影响;②不会因为自注意力机制层数的加深反而导致特征学习能力的下降;③将 Bottleneck 优异的特征学习能力与 MHSA 优秀的抗干扰能力结合,在复杂的交通场景具备更好的检测性能。

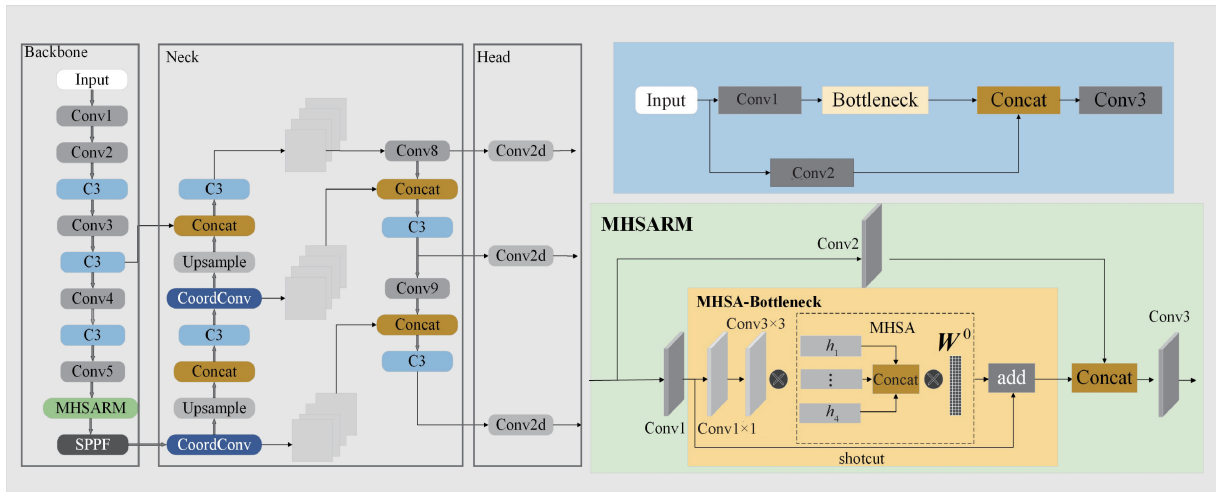


图 4 改进 YOLOv5s 整体架构

Figure 4 The overall architecture of improved YOLOv5s

### 2.5 MHSARM 方法

C3<sup>[5]</sup> 模块是 YOLOv5s 主干网络和特征融合网络重要组成部分,其综合了深度可分离卷积和空洞卷积的优势。C3 由 3 个卷积层和 Bottleneck 模块组成,结构如图 4 浅蓝色区域所示,其中 Conv1 步幅为 2,其目的是增加网络的感受野,使网络能更好地保留全局信息。Conv2、Conv3 步幅为 1,其目的为保留特征图的空间分辨率,从而更好地保留目标的局部特征信息。但 C3 模块因缺乏对目标的关注能力,导致在复杂场景中的检测性能不足,存在检测精度低、漏检、错检现象。

MHSARM 结构如图 4 浅绿色区域所示,该模块原理为 C3 中的 Bottleneck 模块用 MHSA-Bottleneck 替换,既继承了 MHSA-Bottleneck 的优势,又结合了 C3 优异的不同尺度特征提取能力,使其具备优异的特征提取能力的同时,也可根据特征相关性,为特征分配权重系数。在自动驾驶目标检测领域,MHSARM 将更多的注意力集中于交通目标,以此达到弱化背景干扰、强化目标特征信息的目的。经过实

验证明,MHSARM 面对复杂的各类场景(强光、背光、密集遮挡),能有效提升检测性能。

### 2.6 整体架构

改进 YOLOv5s 整体架构如图 4 所示,改进的内容主要包含 MHSARM 替换 C3 模块和 CoordConv。在 YOLOv5s 特征提取区域,SPPF 结构是特征提取区域的最后一个模块,作用是将任意大小的特征图转换成固定大小的特征向量,其包含图像的强语义信息,降低复杂背景对 SPPF 结构带来的干扰,能有效提升算法的检测效率,因此使用 MHSARM 代替基础算法中获取 SPPF 的 C3 模块。

特征金字塔实现了浅层网络的图像特征和深层网络的语义特征融合。但针对空间特征信息,传统卷积学习难度大,并且在经过一系列的卷积、池化操作后,空间信息进一步丢失,因此本文使用 CoordConv 来代替基础网络特征融合区域中的 Conv6、Conv7 卷积层,从而弥补基础网络的空间信息提取能力不足,提升网络检测性能。

### 3 仿真实验与结果分析

#### 3.1 数据集介绍

实验采用由德国卡尔斯鲁厄理工学院和丰田美国技术研究院联合创办的自动驾驶数据集 (Kitti)。Kitti 数据集包含真实图像数据共 7 481 张图片,分为 8 个类别,考虑到自动驾驶场景中的情况,将 8 个类别改为 3 个类别:汽车、货车、箱式汽车合并为汽车;坐着的人和行人合并为人;骑车的人单独做一个类别,删除电车及其他 2 个类别。

#### 3.2 实验平台

本次实验采用的实验平台的配置分别为操作系统 Windows10、显卡 RTX 3090、CUDA11.6、深度学习框架 Pytorch1.12。

#### 3.3 评价指标

本次实验采用  $mAP$ 、 $FPS$  来衡量该网络的检测性能和速度。 $mAP$  为所有类别的平均精度值,其值为以召回率 ( $Recall$ ) 为横坐标和精确率 ( $Precision$ ) 为纵坐标得出的  $P-R$  曲线的面积。精确率又叫查准率,表示在所有被预测为正的样本中实际为正的样本的概率;召回率也叫查全率,表示在实际为正的样本中被预测为正样本的概率。 $Precision$  与  $Recall$  的计算公式<sup>[15]</sup>分别为

$$Precision = \frac{TP}{TP + FP};$$

(4)

$$Recall = \frac{TP}{TP + FN}。$$

(5)

式中:  $TP$  表示正样本分类为正样本数量;  $FP$  表示负样本分类为正样本数量;  $FN$  表示正样本分类为负样本数量。

#### 3.4 训练参数对比实验

超参数在控制模型结构、功能、效率方面发挥着至关重要的作用,本文以 YOLOv5s 为基础算法,采用控制变量法分析学习率 ( $Lr0$ )、余弦退火参数 ( $Lrf$ )、学习率动量 ( $Momentum$ ) 的取值对模型性能的影响,如表 1 所示。表 1 中加粗数值为 YOLOv5s 默认值,也是本文所用参数值及最优值。学习率 ( $Lr0$ ) 是优化算法中的关键参数,该参数确定每次迭代中的步长,  $Lr0$  太小,会导致下降速度慢,需要长时间才能找到最优解;反之,虽然收敛的速度加快,但是模型会忽略损失的最小值,导致损失来回震荡无法收敛甚至直接造成梯度爆炸。表 1 中,第 1 组实验为  $Lr0$  对于模型检测性能的影响实验,本文以 0.01 为基础数值,分别朝着更大及更小的取值方向进行 4 次实验,结果表明,  $Lr0 = 0.01$  时的模型取

得最优的检测性能。 $Lrf$  是一种优化参数,其主要思想是使用余弦函数动态调整学习率。 $Lrf$  可以将学习率从初始值逐步降低到最小值,从而提升模型的训练稳定性和泛化能力。表 1 中第 2 组实验以 0.01 为基础值,双向调整  $Lrf$  值的变化以分析  $Lrf$  对模型检测性能的影响。如表 1 所示,  $Lrf = 0.01$  时模型具备最优异的检测性能。 $Momentum$  用于解决损失函数在训练过程中出现局部最小值而没有达到全局最优的问题,其原理是由动量和梯度共同决定损失函数的运动方向,减少损失震荡现象,更快找到全局最优解。第 3 组实验结果表明,  $Momentum = 0.937$  时模型取得最佳的检测性能。

表 1 训练参数对比实验

Table 1 Experiment with comparison of training parameters					
组别	$Lr0$	$Lrf$	$Momentum$	$mAP_{0.5}$	$mAP_{0.5:0.95}$
1	0.001	0.010	0.937	0.875	0.553
	0.005	0.010	0.937	0.915	0.639
	<b>0.010</b>	<b>0.010</b>	<b>0.937</b>	<b>0.924</b>	<b>0.677</b>
	0.050	0.010	0.937	0.911	0.635
	0.100	0.010	0.937	0	0
2	0.010	0.001	0.937	0.920	0.658
	0.010	0.005	0.937	0.922	0.654
	<b>0.010</b>	<b>0.010</b>	<b>0.937</b>	<b>0.924</b>	<b>0.677</b>
	0.010	0.050	0.937	0.916	0.649
	0.010	0.100	0.937	0.916	0.655
3	0.010	0.010	0.900	0.914	0.642
	0.010	0.010	0.930	0.920	0.661
	<b>0.010</b>	<b>0.010</b>	<b>0.937</b>	<b>0.924</b>	<b>0.677</b>
	0.010	0.010	0.950	0.920	0.659
	0.010	0.010	1.000	0	0

#### 3.5 消融实验结果与分析

采用控制变量法对 Kitti 实验数据进行消融实验,目的是分析不同模块对改进 YOLOv5s 算法性能的影响,结果如表 2 所示。其中 CoordConv-Y、MH-SARM-Y 分别对应两种改进措施,  $mAP_{0.5}$  是 IOU 阈值为 0.5 的所有类别的平均精度值,  $mAP_{0.5:0.95}$  是不同 IOU 阈值(从 0.5 到 0.95)的  $mAP$  平均值。由表 2 可知,两种改进措施都提升了网络的检测性能。CoordConv-Y 与基础算法相比,  $Precision$  提高了 1.1 百分点,  $Recall$  下降了 0.1 百分点。两者数据变化说明具备空间感知能力 CoordConv 能大幅度提升网络的检测精确度,但是针对漏检、错检问题,效果并不明显。MHSARM-Y 与基础算法相比,  $Precision$  提高了 0.3 百分点,  $Recall$  上升 0.2 百分点,  $mAP_{0.5}$  提高了 0.8 百分点,该数据表明, MH-SARM-Y 具备更强的检测性能。



表 2 消融实验数据对比

Table 2 Comparison of ablation experimental data

算法	Precision	Recall	mAP_0.5	mAP_0.5:0.95	FPS
YOLOv5s <sup>[5]</sup>	0.938	0.877	0.924	0.677	97
MHSA-Y	0.934	0.868	0.922	0.660	95
CoordConv-Y	<b>0.949</b>	0.876	0.929	0.673	86
MHSARM-Y	0.941	<b>0.880</b>	0.932	<b>0.684</b>	90
本文算法	0.944	0.879	<b>0.933</b>	<b>0.684</b>	84

为对比 MHSARM 模块与 MHSA 模块之间的性能差异,采用控制变量对 MHSA-Y、MHSARM-Y 两种模型的结果进行数据分析,其中 MHSA-Y 是将 C3 模块替换成 MHSA。由表 2 可知,与 MHSA-Y 相比,MHSARM-Y 的 *Precision* 提高了 0.7 个百分点,*Recall* 提高了 1.2 百分点,*mAP*\_0.5 提高了 1.0 百分点。该数据表明,MHSA 特征学习能力有限,其与 Bottlentck 组合而成的 MHSARM,具备更强的检测性能。

由表 2 可知,改进 YOLOv5s 在检测性能方面的数据都优于原基础算法,整体检测精度上升明显,*mAP*\_0.5 提高了 0.9 百分点,*Precision* 提高了 0.6 百分点,*Recall* 提高了 0.2 百分点,*mAP*\_0.5:0.95 提高了 0.7 百分点。整体数据表明,YOLOv5s 虽在 Kitti 数据集上取得了较好的效果,但是在复杂交通场景下的检测精度有限,漏检、错检等问题依然存在,改进 YOLOv5s 算法可以有效改善网络在各种复杂场景中的特征提取能力,使算法在自动驾驶交通目标检测领域的检测性能得到进一步提升。最后,比较 *FPS* 可以看出,就本次研究的平台环境而言,本文中各个改进点对实时性能造成的影响并不明显,仍可以达到自动驾驶实时性的要求。

3.6 Kitti 数据集检测效果分析

选取复杂背景典型样本(强光、背光、密集遮挡)的检测效果分析改进 YOLOv5s 对交通目标的检测能力,如图 5 所示。

改进 YOLOv5s 的各种复杂场景下的目标框拥有更高的置信度。由图 5(a)可以看出,在强光环境中,受强光照射的物体会导致物体色淡化、特征信息减弱,改进 YOLOv5s 网络改善了特征减弱的目标的检测性能。由图 5(b)可以看出,在背光场景中,待检目标呈现黑色灰暗的状态,能见度、清晰度大幅降低,导致模型特征学习难度大,存在漏检现象(绿色方框区域),改进 YOLOv5s 通过强化对目标的关注度及学习目标空间维度特征的方式,提升网络对该类目标的特征提取能力,改善漏检的问题。由图 5(c)可以看出,在交通场景中,交通目标密集拥堵



图 5 实际检测效果对比

Figure 5 Comparison of actual detection results

现象造成待检目标仅保留局部特征,使得网络提取到的特征不全、不准,导致漏检问题(黄色区域所示),改进 YOLOv5s 通过 MHSARM 模块增强模型对关键点的特征提取能力,改善了漏检现象。

3.7 不同注意力机制改进模型对比

为了验证 MHSARM 的优势,分别采用多种表现优异的注意力机制对 YOLOv5s 中 C3 结构进行改进,改进方式与 MHSARM 一致,得到改进的模块 BAM-C3、GAM-C3 等,如表 3 所示。由表 3 可知,对比其他注意力残差结构,MHSARM 的 *Precision*、*Recall*、*mAP*\_0.5、*mAP*\_0.5:0.95 均为最大值,表明由 C3 模块跟多头自注意力机制(MHSA)组合的 MHSARM 在目标检测自动驾驶场景中有着更好的适用性,对交通目标的特征学习能力更强。此外,该实验数据再一次证明,对比其他表现优异的注意力机制,MHSA 在应对复杂背景带来的干扰时抗性更强。

3.8 模型对比

为了进一步验证改进 YOLOv5s 的创新性及优

势,本文从两个角度进行对比实验:①与主流的目标检测算法(YOLO 系列)对比;②与在 Kitti 数据集上最新的(2022 年至 2023 年)研究成果对比。如表 4 所示,对比当前先进的目标检测网络,改进 YOLOv5s 算法的 *Precision*、*Recall*、*mAP\_0.5* 均取得最好的效果,在自动驾驶交通目标检测任务中具备最优异的性能。

表 3 不同注意力机制改进模型对比

Table 3 Different attention mechanisms comparison of improved models				
模块	<i>Precision</i>	<i>Recall</i>	<i>mAP_0.5</i>	<i>mAP_0.5:0.95</i>
YOLOv5s <sup>[5]</sup>	0.938	0.877	0.924	0.677
BAM <sup>[16]</sup> -C3	0.938	0.871	0.928	0.677
GAM <sup>[17]</sup> -C3	0.933	0.878	0.927	0.681
SE <sup>[18]</sup> -C3	0.938	0.879	0.927	0.682
CBAM <sup>[19]</sup> -C3	0.940	0.880	0.931	0.680
SimAM <sup>[20]</sup> -C3	0.933	0.868	0.926	0.677
本文模块	<b>0.941</b>	<b>0.880</b>	<b>0.932</b>	<b>0.684</b>

表 4 模型性能比较

Table 4 Model performance comparison				
算法	<i>Precision</i>	<i>Recall</i>	<i>mAP_0.5</i>	<i>mAP_0.5:0.95</i>
LBGE and MSLB <sup>[21]</sup>			0.874	
DSGIoU <sup>[22]</sup>			0.897	
K+Means++ <sup>[23]</sup>			0.838	
改进 SSD <sup>[24]</sup>			0.685	
Res101-FcaNet <sup>[25]</sup>			0.765	
Sim-YOLOv4 <sup>[8]</sup>			0.911	
YOLOv5s <sup>[5]</sup>	0.938	0.877	0.924	0.677
YOLOv6s <sup>[26]</sup>	0.945	0.824	0.906	0.676
YOLOv7s <sup>[27]</sup>	0.926	0.869	0.920	0.647
YOLOv8s <sup>[28]</sup>	0.929	0.849	0.911	<b>0.686</b>
本文算法	<b>0.944</b>	<b>0.879</b>	<b>0.933</b>	<b>0.684</b>

3.9 模型泛化能力验证

为进一步验证模型能力,排除数据集偏差的影响,在开源自动驾驶数据集 BDD100K 中分别对 YOLOv5s 和改进 YOLOv5s 进行测试和训练。BDD100K 是迄今为止规模最大、内容最具多样性的开源数据集,总共包含 10<sup>5</sup> 张图像数据,本次实验采用的是 BDD100K 轻量级数据集,共包含 10 000 张图像数据,包含 10 个类别,为了排除误差,删除样本数极少的类别,最终保留 6 个类别。与 Kitti 数据集相比,该数据集数据类型更多,采集环境丰富,背景干扰强度也更大。

表 5 为两种模型在数据集的性能比较,从表 5 中可以看出,改进 YOLOv5s 对比原基础模型,各个参数数值均取得提升,其中 *Precision* 提高了 1.6%, *Recall* 提高了 0.6%, *mAP\_0.5* 提高了 1.4%, *mAP\_*

0.5:0.95 提高了 1.5%。

表 5 两种网络模型性能比较

Table 5 Performance comparison of the two network models				
算法	<i>Precision</i>	<i>Recall</i>	<i>mAP_0.5</i>	<i>mAP_0.5:0.95</i>
YOLOv5s <sup>[5]</sup>	0.651	0.417	0.460	0.249
本文算法	<b>0.677</b>	<b>0.423</b>	<b>0.474</b>	<b>0.264</b>

3.10 BDD100K 检测效果分析

图 6 为 YOLOv5s 和改进 YOLOv5s 在 BDD100K 数据集上各类场景的检测效果对比。由图 6 可知,在各类环境中,改进 YOLOv5s 的目标框拥有更高的置信度。如图 6 中红色边框所示,YOLOv5s 对特征信息不明显以及局部丢失的交通目标存在错检、漏检、重检等现象。改进 YOLOv5s 弱化了背景干扰,在关注目标信息的同时,通过学习空间信息特征优化检测性能,缓解了上述现象。改进 YOLOv5s 除了在 Kitti 数据集有着较好的检测效果,在类别更多、场景更复杂的数据集中同样拥有更好的检测效果。

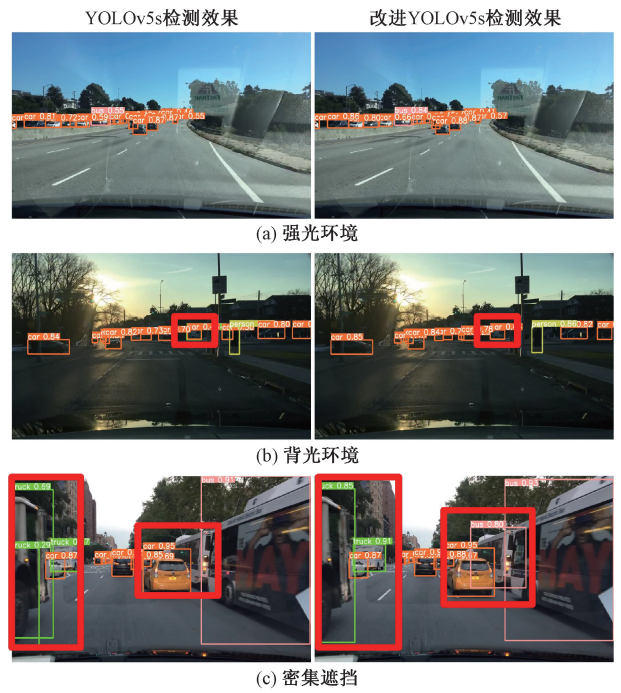


图 6 在 BDD100K 数据集上的实际检测效果对比  
Figure 6 Comparison of actual detection results on BDD100K dataset

4 结论

本文提出改进 YOLOv5s 复杂道路交通目标检测算法。在特征提取区域,引入基于多头自注意力机制残差模块(MHSARM),在含有强语义信息的深层特征层采用分配权重的方式,强化目标信息,抑制背景干扰。经实验证明,MHSARM 应对复杂的交通

场景的特征提取能力明显超越原 C3 模块。特征融合方面,使用 CoordConv 代替传统卷积,使网络具备感知空间位置信息的能力,实现更丰富的特征融合。经实验证明,在自动驾驶领域,具备空间感知能力的检测网络能更精准地检测出目标。

改进 YOLOv5s 网络虽然在强光、背光、密集遮挡等场景下取得良好的效果,但是面对更复杂的背景,如下雨天、大雾天气、下雪天的检测性能依然有限。同时,据检测效果分析,改进 YOLOv5s 面对跨度大的目标时,存在严重的重检现象,面对局部特征丢失的小目标存在较严重漏检现象。下一步工作主要解决三个问题:对跨度大的目标重检率较高的问题;局部特征的小目标漏检问题;面对更复杂的场景检测能力有限的问题。

参考文献:

[1] 崔家悦. 浅谈自动驾驶的研究现状和发展[J]. 科技资讯, 2021, 19(13): 83-85.

CUI J Y. Research status and development of automatic driving[J]. Science & Technology Information, 2021, 19(13): 83-85.

[2] GIRSHICK R, DONAHUE J, DARRELL T, et al. Region-based convolutional networks for accurate object detection and segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(1): 142-158.

[3] CHEN Y H, LI W, SAKARIDIS C, et al. Domain adaptive faster R-CNN for object detection in the wild[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 3339-3348.

[4] 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). Piscataway: IEEE, 2016: 779-788.

[5] ULTRALYTICS. Yolov5[EB/OL]. [2023-08-11]. <https://github.com/ultralytics/yolov5>.

[6] LIU R, LEHMAN J, MOLINO P, et al. An intriguing failing of convolutional neural networks and the Coord-Conv solution[C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems. New York: ACM, 2018: 9628-9639.

[7] 袁志宏, 孙强, 李国祥, 等. 基于 Yolov3 的自动驾驶目标检测[J]. 重庆理工大学学报(自然科学), 2020, 34(9): 56-61.

YUAN Z H, SUN Q, LI G X, et al. Automatic driving target detection based on Yolov3 [J]. Journal of Chongqing University of Technology (Natural Science),

2020, 34(9): 56-61.

[8] 刘丽伟, 侯德彪, 侯阿临, 等. 基于 SimAM-YOLOv4 的自动驾驶目标检测算法[J]. 长春工业大学学报, 2022, 43(3): 244-250.

LIU L W, HOU D B, HOU A L, et al. Automatic driving target detection algorithm based on SimAM-YOLOv4 [J]. Journal of Changchun University of Technology, 2022, 43(3): 244-250.

[9] 院老虎, 常玉坤, 刘家夫. 基于改进 YOLOv5s 的雾天场景车辆检测方法[J]. 郑州大学学报(工学版), 2023, 44(3): 35-41.

YUAN L H, CHANG Y K, LIU J F. Vehicle detection method based on improved YOLOv5s in foggy scene[J]. Journal of Zhengzhou University (Engineering Science), 2023, 44(3): 35-41.

[10] CAI Y F, LUAN T Y, GAO H B, et al. YOLOv4-5D: an effective and efficient object detector for autonomous driving[J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 4503613.

[11] SHI T, DING Y, ZHU W X. YOLOv5s\_2E: improved YOLOv5s for aerial small target detection[J]. IEEE Access, 2023, 11: 80479-80490.

[12] GAO T Y, WUSHOUER M, TUERHONG G. DMS-YOLOv5: a decoupled multi-scale YOLOv5 method for small object detection[J]. Applied Sciences, 2023, 13(10): 6124.

[13] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: ACM, 2017: 6000-6010.

[14] WU C H, WU F Z, GE S Y, et al. Neural news recommendation with multi-head self-attention[C]//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Stroudsburg: Association for Computational Linguistics, 2019: 6389-6394.

[15] ZHANG Y F, REN W Q, ZHANG Z, et al. Focal and efficient IOU loss for accurate bounding box regression[J]. Neurocomputing, 2022, 506: 146-157.

[16] PARK J, WOO S, LEE J Y, et al. BAM: bottleneck attention module[EB/OL]. (2018-07-17) [2023-08-11]. <http://arxiv.org/abs/1807.06514>.

[17] LIU Y C, SHAO Z R, HOFFMANN N. Global attention mechanism: retain information to enhance channel-spatial interactions[EB/OL]. (2021-10-10) [2023-08-11]. <http://arxiv.org/abs/2112.05561>.

[18] 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.

[19] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module [C] // European Conference on Computer Vision. Cham: Springer, 2018: 3–19.

[20] YANG L X, ZHANG R Y, LI L D, et al. Simam: a simple, parameter-free attention module for convolutional neural networks[C]//Proceedings of the 38th International Conference on Machine Learning. United States: International Conference on Machine Learning, 2021: 11863–11874.

[21] SU T, SHI Y, XIE C J, et al. A hybrid loss balancing algorithm based on gradient equilibrium and sample loss for understanding of road scenes at basic-level [J]. Pattern Analysis and Applications, 2022, 25(4): 1041–1053.

[22] 马素刚, 李宁博, 侯志强, 等. 基于 DSGIoU 损失与双分支坐标注意力的目标检测算法[J/OL]. 北京航空航天大学学报, 2023: 1–14. <https://bhxb.buaa.edu.cn/bhbk/cn/article/doi/10.13700/j.bh.1001-5965.2023.0192>.

MA S G, LI N B, HOU Z Q, et al. Object detection algorithm based on DSGIoU loss and dual branch coordinate attention[J/OL]. Journal of Beijing University of Aeronautics and Astronautics, 2023: 1–14. <https://bhxb.buaa.edu.cn/bhbk/cn/article/doi/10.13700/j.bh.1001-5965.2023.0192>.

[23] 丁剑洁, 安雯. 基于改进锚点框与 Transformer 架构的目标检测算法[J]. 现代电子技术, 2023, 46(15): 37–42.

DING J J, AN W. Object detection algorithm based on improved anchor box and transformer architecture [J]. Modern Electronics Technique, 2023, 46(15): 37–42.

[24] 李娟娟, 侯志强, 白玉, 等. 基于空洞卷积和特征融合的单阶段目标检测算法[J]. 空军工程大学学报(自然科学版), 2022, 23(1): 97–103.

LI J J, HOU Z Q, BAI Y, et al. Single-stage object detection algorithm based on dilated convolution and feature fusion[J]. Journal of Air Force Engineering University (Natural Science Edition), 2022, 23(1): 97–103.

[25] 侯志强, 郭浩, 马素刚, 等. 基于双分支特征融合的无锚框目标检测算法[J]. 电子与信息学报, 2022, 44(6): 2175–2183.

HOU Z Q, GUO H, MA S G, et al. Anchor-free object detection algorithm based on double branch feature fusion [J]. Journal of Electronics & Information Technology, 2022, 44(6): 2175–2183.

[26] LI C Y, LI L L, JIANG H L, et al. YOLOv6: a single-stage object detection framework for industrial applications [EB/OL]. (2022–09–07) [2023–08–11]. <http://arxiv.org/abs/2209.02976>.

[27] 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 [C] //2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2023: 7464–7475.

[28] ULTRALYTICS. YOLOv8 [EB/OL]. (2023–01–01) [2023–08–11]. <https://github.com/ultralytics/ultralytics>.

Complex Road Traffic Target Detection Algorithm Based on Improved YOLOv5s

TANG Lindong<sup>1,2</sup>, YUN Lijun<sup>1,2</sup>, LUO Ruilin<sup>3</sup>, LU Lin<sup>3</sup>

(1. College of Information, Yunnan Normal University, Kunming 650500, China; 2. Yunnan Provincial Department of Education Computer Vision and Intelligent Control Technology Engineering Research Center, Yunnan Normal University, Kunming 650500, China; 3. Yunnan Tobacco Leaf Company, Kunming 650500, China)

**Abstract:** A complex road traffic object detection algorithm was proposed to address the issue of traffic target detection algorithms' inability to resist complex background interference and insufficient detection performance in the current autonomous driving scenario. At first, the multi-head self-attention residual module (MHSARM) was used to improve the feature information of the target to be inspected while decreasing the complex background interference. Secondly, in the feature fusion area, CoordConv was used instead of traditional Conv, so that the network could perceive spatial information and improve network detection accuracy. The improved YOLOv5s algorithm had stronger feature extraction ability and good generalisation ability in complex roads, and *mAP*<sub>0.5</sub> reached 93.3% and 47.4%, respectively, which was higher than that of YOLOv5s 0.9% and 1.4%. In addition, compared with the latest target detection algorithms YOLOv7 and YOLOv8, the *mAP*<sub>0.5</sub> of improved YOLOv5s improved by 1.3% and 2.2%, respectively. Compared with the latest research results of Sim-YOLOv4 algorithm on Kitti dataset, *mAP*<sub>0.5</sub> improved 2.2%.

**Keywords:** automatic driving; target detection; YOLOv5s; MHSARM; CoordConv