

复杂交通场景下的目标检测方法

濮志远, 罗素云

上海工程技术大学机械与汽车工程学院, 上海 201620

基金项目: 国家自然科学基金项目(62101314)

通信作者: 罗素云, lsyluo@163.com 收稿/录用/修回: 2024-07-04/2024-09-02/2025-01-22

摘要

针对复杂交通场景目标检测方法的不足,特别是对小目标和遮挡目标的漏检及多尺度目标检测和模型鲁棒性方面的不足,提出了一种改进后的YOLOv8s-SRCEM(You Only Look Once version 8 small model with Small object detection head, Residual Convolutional block attention module, Efficient channel attention module, and Multi-scale block)模型:引入小目标检测头,使模型能够更加敏感地捕捉小尺寸目标,提高对小目标的检测能力;在小目标检测头上集成Res-CBAM(Residual Convolutional Block Attention Module),进一步提高特征学习的显著性;在骨干网络中加入ECA(Efficient Channel Attention)模块,强化模型对特征通道重要性的关注,提升特征选择和模型的鲁棒性;将原始的SPPF(Spatial Pyramid Pooling-Fast)模块替换为MS-Block(Multi-Scale Block),模型在不同尺度上的特征捕捉和融合能力得到增强。在KITTI数据集上,改进后的模型相比于YOLOv8s模型mAP(mean Average Precision)值提高了6.6%。实验结果表明,多种改进方案的组合使模型在复杂交通场景中的检测性能得到全面提升。

关键词

复杂场景
小目标检测
MS-Block(Multi-Scale Block)
ECA(Efficient Channel Attention)
注意力模块
中图法分类号: TP391
文献标志码: A

Object Detection Method in Complex Traffic Scenarios

PU Zhiyuan, LUO Suyun

School of Mechanical and Automotive Engineering, Shanghai University of Engineering Science, Shanghai 201620, China

Abstract

To address the limitations of target detection methods in complex traffic scenarios, particularly missed detections of small and occluded targets and challenges in multi-scale target detection and model robustness, we propose an improved YOLOv8s-SRCEM model. We introduce a small target detection head to enhance the sensitivity of the model to small-sized targets, improving its detection capability for such objects. Additionally, we integrate the Res-CBAM attention module into the small target detection head to further enhance feature learning salience. We incorporate the ECA module into the backbone network to strengthen the attention to important feature channels of the model and improve feature selection and model robustness. Furthermore, by replacing the original SPPF module with MS-Block, we enhance the feature capture and fusion capabilities of the model across different scales. On the KITTI dataset, the improved model achieves a 6.6% increase in mAP compared to the YOLOv8s model. Experimental results demonstrate that the combination of these enhancements substantially improves the detection performance of the model in complex traffic scenarios.

Keywords

complex scenario;
small target detection;
MS-Block;
ECA;
attention module

0 引言

随着自动驾驶技术^[1-2]的飞速发展, 目标检测作为自动驾驶感知系统^[3-4]的重要组成部分, 已成为确保车辆安全行驶的关键环节。在复杂的交通环境中, 如何准确高效地识别和定位行人、车辆等交通参与者, 对于实现真正的自动驾驶系统至关重要。

近年来, 深度学习技术在目标检测领域取得了长足的进步, 尤其是在目标检测速度和准确性方面取得了重大突破。然而, 在复杂的交通场景中, 如何进一步提高目标检测的速度和准确性, 仍然是亟待解决的挑战。为了解决这些问题, 许多研究者提出了各种改进方法。HUANG 等^[5]提出了 DC-SPP-YOLO(Dense Connection and Spatial Pyramid Pooling-based YOLO)方法, 该方法将密集连接(DC)和空间金字塔池化(SPP)技术进行融合, 密集连接通过在网络层之间建立短连接, 增强了特征传播和梯度传递, 减轻了梯度消失问题, 而空间金字塔池化则能够在不增加计算量的前提下, 提取多尺度特征, 有效提升了基于 YOLOv2 的车辆检测精度; LIN 等^[6]提出一种特征金字塔网络(Feature Pyramid Network), 通过融合高低分辨率的特征, 构建了一个顶到底的架构, 提高了目标检测的性能, 尤其在小目标检测方面表现突出。ZHANG 等^[7]通过在 YOLOv5(You Only Look Once version 5)算法上引入 Flip-Mosaic 数据增强技术, 提出了一种专注于小目标感知能力提升的实时车辆检测方法, 显著提高了在不同交通场景下车辆检测的准确性, 并降低了误检率。LI 等^[8]通过改进 YOLOv5, 主要针对小目标引入注意力机制和 SoftPool(Softmax-based differentiable Pooling), 提高了检测车辆和人行横道等道路物体的检测精度。ZHOU 等^[9]提出了一种基于 YOLOv8 的轻量级车辆检测算法, 通过使用 FasterNet 替换原有的骨干网络, 减少了计算复杂度的同时保持了高效的特征提取能力。此外, 还添加了 SimAM(a Simple, parameter-free Attention Module for convolutional neural networks)注意力模块, 自适应地调整特征图的权重, 从而增强对关键区域的关注。

尽管上述方法在不同程度上提升了目标检测的性能, 但现阶段复杂交通场景下的目标检测模型仍然面临小目标和遮挡目标漏检严重的问题, 并且多尺度目标检测和模型鲁棒性方面也存在不足。针对这些挑战, 本文在综合考虑检测精度和检测速度等因素后, 以 YOLOv8 模型为基础, 提出了一种改进

的复杂交通场景目标检测算法。

本文的主要贡献包括: 1) 引入了专门针对小目标设计的检测头, 使得模型能够更加敏感地捕捉到图像中的小尺寸目标; 2) 在小目标检测头中, 集成了注意力模块, 有助于模型集中学习目标的显著特征; 3) 在骨干网络中融入了通道注意力机制, 使网络能更加高效地筛选和强化有用的特征通道; 4) 将传统的快速空间金字塔池化(SPPF)模块替换为多尺度特征融合模块, 有助于模型在不同尺度上捕捉和融合特征。

通过调整特征提取和预测策略, 该改进算法在保持高检测速度的同时, 提高了对小目标的检测性能, 从而提升了算法在复杂交通场景中的整体表现。

1 YOLOv8s 模型算法原理

YOLOv8s 是由 Ultralytics 公司推出的基于深度学习的一阶段目标检测算法, 是目前最先进的 YOLO 系列模型之一。它在 YOLOv5 的基础上进行了网络架构、特征提取、预测策略和训练方式等多项改进, 在速度和精度方面取得了完美的平衡, 使其成为智能视频监控、自动驾驶等实时性要求较高的应用场景的理想选择。YOLOv8s 模型由骨干网络、颈部网络及预测头组成, 详细的模型结构如图 1 所示。模型中的 Conv 模块由 Conv2d(2 维卷积)、BatchNorm2d(2 维批次归一化)和 SiLU(激活函数)组成。YOLOv8s 延用了 YOLOv5 骨干网络架构, 并将 CSPlayer(Cross Stage Partial layer)模块^[10]更换成了更加轻量化的 C2f(Cross stage partial bottleneck with 2 convolutions-fast version)模块。C2f 模块吸收了 C3(Cross stage partial bottleneck with 3 convolutions)模块与 ELAN(Efficient Layer Aggregation Network)的理念, 通过并行更多的梯度流分支来获得更丰富的梯度信息, 从而提升模型的精度和性能。在 C2f 模块中, Bottleneck Shortcut 是一种通过捷径连接将输入和输出进行有效整合的设计方式, 能够缓解深层网络中的梯度消失问题, 可以通过 Shortcut 参数选择是否使用该方式。为了将骨干网络提取的特征进行进一步的整合, YOLOv8s 的 Neck 模块采用 PANet(Path Aggregation Network)^[11]特征金字塔结构, 它将高级别的语义特征与空间细节相融合来丰富特征表示, 提升了模型检测不同尺寸目标的能力。其中 Concat 操作用于将小尺寸特征图 Upsample(上采样)后和大尺寸特征图 Downsample(下

采样)后拼接在一起,以便模型可以同时利用多尺度的特征信息,从而增强检测能力。MaxPool2d 则

是在网络中用于进行最大池化操作的模块,它有助于下采样并提取最显著的特征。

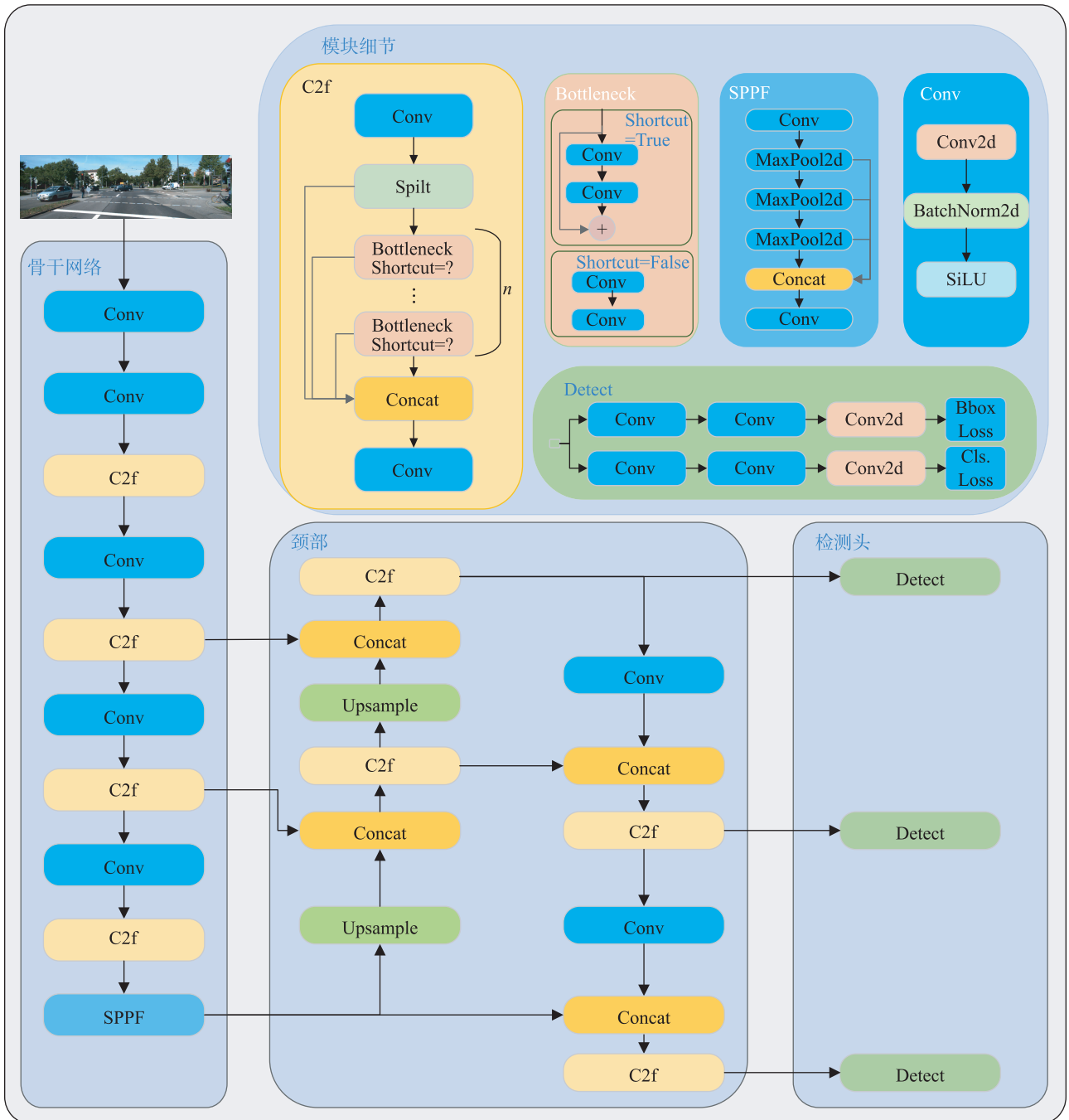


图1 YOLOv8s 结构图

Fig.1 Structure diagram of YOLOv8s

在检测头上, YOLOv8s 摒弃了传统 YOLO 算法中的锚框机制,直接预测目标的中心点、宽高和类别。这种无锚框设计使得模型能够更加灵活地适应不同尺寸和形状的目标,提高了模型的鲁棒性。损失函数方面, YOLOv8s 采用了 CIoU^[12] (Complete

Intersection over Union) + DFL(distribution focal loss) 损失函数。Bbox.Loss 和 Cls.Loss 是 YOLOv8s 用来计算边界框回归和分类损失的部分,它们通过结合 CIoU 损失函数和 DFL 损失,能够更加精确地衡量预测框与真实框之间的差距。

2 改进的 YOLOv8s 模型

本文选择 YOLOv8 的 S 版本作为基础模型进行改进。S 版本在计算复杂度和模型性能之间取得了良好的平衡, 同时在保持较高检测精度的情况下, 能够提供相对较快的推理速度, 这对于自动驾驶场景中实时检测任务尤为重要。

2.1 加入通道注意力的骨干网络

考虑到 YOLOv8s 模型骨干网络存在特征提取能力不足的缺陷, 会导致模型难以捕捉图像中关键信息, 影响目标检测精度, 本文选择对骨干网络中的 Conv 模块进行改进, 通过融合 ECA 模块^[13]形成 ECA_Conv 模块。

传统的 Conv 模块在特征提取方面存在特征选择不够精准的问题, 而 ECA 模块能够计算通道间的相关性并生成权重, 动态地调整不同通道的特征贡献度, 使模型更加聚焦于最具信息量的特征通道。将 ECA 模块与 Conv 模块结合, 可以在特征提取过程中提高模型对重要特征的关注, 有助于模型在复杂交通场景中提升对目标的识别和定位

精度。

ECA 模块通过全局平均池化操作对输入特征进行通道级别的聚合, 生成每个通道的综合特征。与 SENet (Squeeze-and-Excitation Network)^[14] 中的 SE 块不同, ECA 模块没有进行维度缩减, 以保持通道与其权重之间的直接对应关系。接着通过 1D 卷积实现局部交叉通道交互, 考虑每个通道及其 K 个邻居的局部交互, 从而学习通道注意力。该 1D 卷积的核大小 K 通过一种自适应的方法确定, 使其与通道维度 C 成比例, 并通过非线性映射来确定卷积核大小:

$$K = \left\lfloor \frac{\ln C}{\gamma} + \frac{b}{\gamma} \right\rfloor \quad (1)$$

其中, γ 和 b 是用于确定核大小的参数, 实验过程中将 γ 和 b 分别设置为 2 和 1。由式(1)得到的 K 大小的卷积核有利于在不同的网络结构和任务中找到最优的交互范围。1D 卷积后使用 Sigmoid 激活函数生成每个通道的权重, 从而动态调整特征通道的贡献度, 增强对重要特征的关注。ECA 模块的结构如图 2 所示。

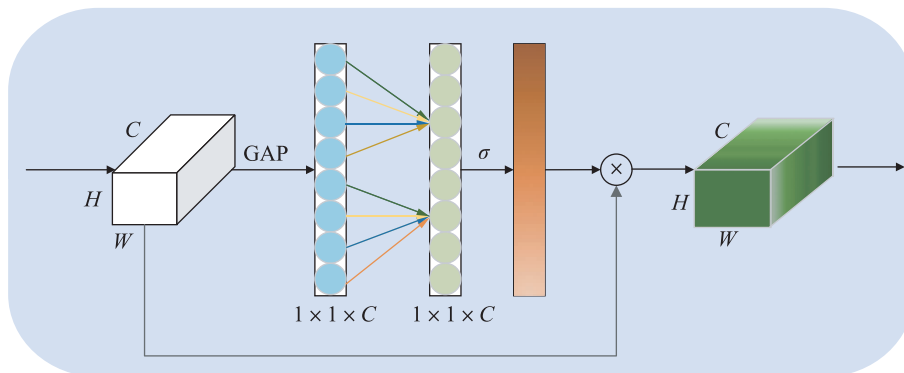


图 2 ECA 结构图

Fig.2 ECA structure diagram

2.2 加入小目标检测层

原始的 YOLOv8s 模型在目标检测任务中表现良好, 但其配置的 3 个检测头主要侧重于中等和大尺寸目标的检测。这 3 个检测头虽然在多尺度目标检测上取得了较好的平衡, 但在检测小目标时却面临一些局限, 尤其是当小目标在特征图上所占空间较小时, 容易导致漏检和误检的现象, 一定程度上限制了 YOLOv8s 在处理小目标密集场景时的应用潜力。

针对复杂交通场景数据集中小目标占据大多数, 并且小目标由于其尺寸小导致特征不明显, 通

常更难被准确识别和定位的问题, 本文在 YOLOv8s 模型中针对分辨率为 160×160 的特征图引入小目标检测头, 具体结构如图 3 所示。

小目标检测头的特征图来源于 PAN-FPN (Path Aggregation Network with Feature Pyramid Network) 中的高层特征图。这些高层特征图包含了全局上下文信息和抽象特征, 能够有效地捕捉大型和中型目标。然而, 由于高层特征图的分辨率较低, 缺乏小目标的细节捕捉能力。因此, 本文采用上采样层将这些高层特征图恢复到较低尺度 (即更高的分辨率), 以更好地保留小目标的细节信息。为了进一

步丰富特征图的信息，将上采样后的高层特征图与骨干网络中的低尺度特征图进行融合。骨干网络的

低尺度特征图通常包含较多的边缘和纹理等细节信息，适合用于小目标的检测。

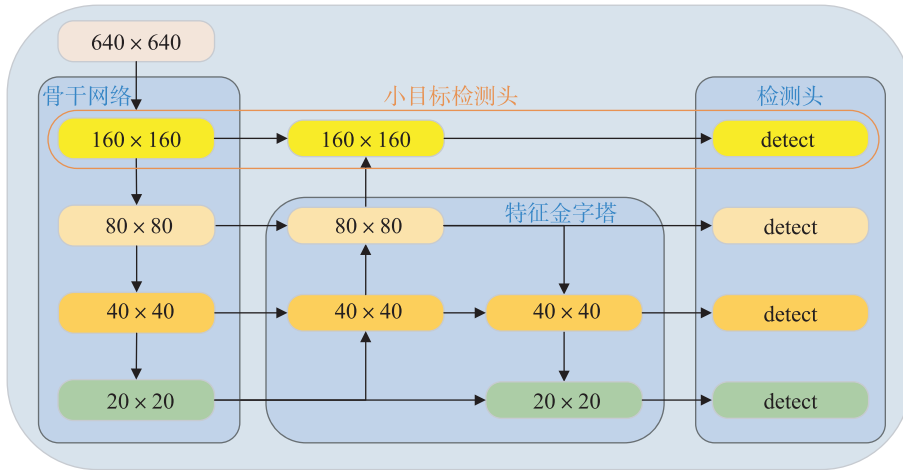


图3 集成小目标检测头结构图

Fig.3 Structure diagram of integrated small target detection head

这一改进使模型可以更好地利用较低尺度特征图中的丰富细节，增强对小目标的感知能力，提高小目标检测的准确性和鲁棒性。

2.3 残差注意力模块

为了进一步提升 YOLOv8s 模型在小目标检测上的表现，本文在小目标检测头上加入了 Res-CBAM 模块。Res-CBAM 是将残差连接和卷积注意力模块相结合的一种改进策略。

目前，注意力机制已广泛应用于深度学习的多个领域，模块通过对信息特征进行相关和不相关的

选择，构建动态权重，帮助网络记录位置关系，估计不同信息的重要性^[15]。CBAM (Convolutional Block Attention Module) 模块^[16]根据通道注意力和空间注意力的计算结果，动态地调整特征的权重，以突出重要特征，抑制无关特征。但这种操作削弱了原始特征的重要性，而在本文提出的 Res-CBAM 中，残差连接的引入解决了注意力模块带来的特征信息损失问题。

Res-CBAM 的结构如图 4 所示。首先，输入特征进行全局平均池化(global average pooling)和全局

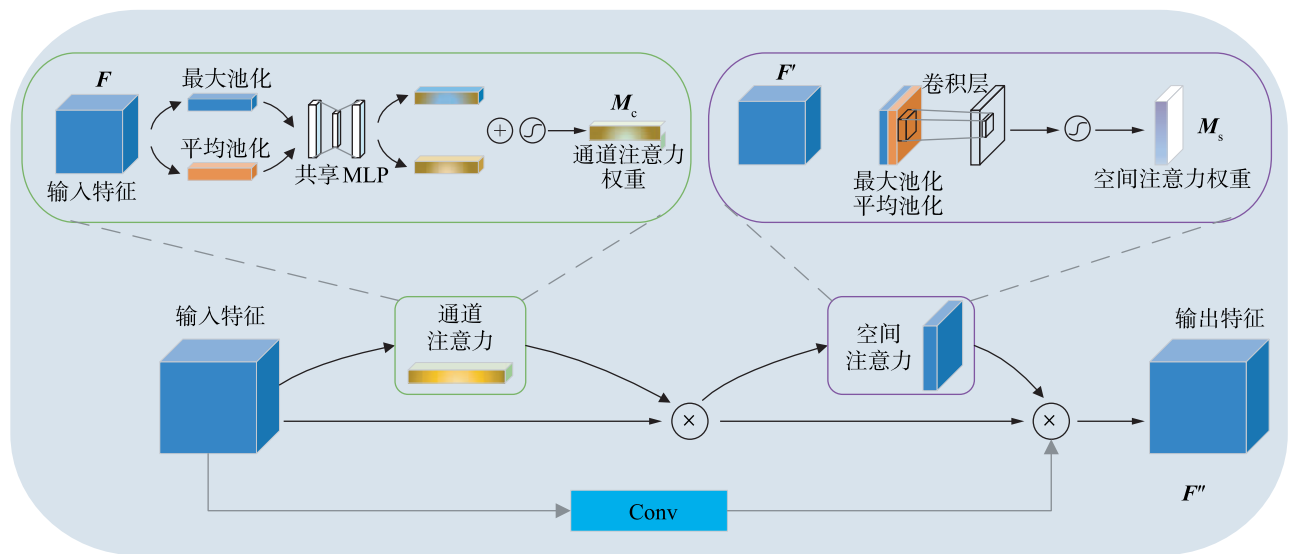


图4 Res-CBAM 结构图

Fig.4 Res-CBAM structure diagram

最大池化(global max pooling), 生成两个特征图。然后, 通过全连接层(fully connected layer)和 Sigmoid 激活函数生成通道注意力图 M_c :

$$M_c = \sigma(F_c(G_{AP}(F)) + F_c(G_{MP}(F))) \quad (2)$$

其中, σ 表示 Sigmoid 激活函数, F_c 表示全连接层, G_{AP} 表示全局平均池化, G_{MP} 表示全局最大池化。最后, 原始特征图与通道注意力图相乘, 得到通道细化后的特征图 F' :

$$F' = FM_c \quad (3)$$

通道细化后的特征图通过全局平均池化和全局最大池化, 生成两个特征图, 接着通过卷积层(convolutional layer)和 Sigmoid 激活函数生成空间注意力图 M_s :

$$M_s = \sigma(f^{7 \times 7}([G_{AP}(F'), G_{MP}(F')])) \quad (4)$$

其中, σ 表示 Sigmoid 激活函数, $f^{7 \times 7}$ 表示卷积核大小为 7×7 的卷积操作, G_{AP} 表示全局平均池化, G_{MP} 表示全局最大池化。

将空间注意力图与通道细化特征图相乘后再与输入特征图相乘, 实现残差连接, 得到经过通道和空间双重细化的特征图 F'' :

$$F'' = f^{3 \times 3}(F)F'M_s \quad (5)$$

其中, $f^{3 \times 3}$ 表示卷积核大小为 3×3 的卷积操作。

该结构不仅充分利用了通道和空间注意力机制, 也保留了原始输入特征的优势, 同时增强了梯度传播的稳定性, 提高了检测模型对小目标的敏感度和鲁棒性。

2.4 MS 模块改进

YOLOv8s 模型中的 SPPF 模块主要通过不同大小的卷积核来增大感受野。尽管这样能增强模型对空间特征的捕捉能力, 但其多尺度特征融合策略不够灵活, 会导致对多尺度信息的处理不足。

因此本文在 YOLOv8s 模型上使用 MS 模块来解决这一问题。MS 模块通过层次化特征融合策略和异构卷积核选择协议, 能够更好地处理不同尺度的特征。同时, MS 模块采用多分支结构和多尺度特征融合, 显著提升了模型对细粒度和粗粒度特征的感知能力及对小目标的敏感度。

MS 模块采用倒置瓶颈层(inverted bottleneck layer)和深度可分离卷积(depth-wise convolution)作为核心组件, 在保持计算效率的同时增加了模型的深度和复杂度, 模型结构如图 5 所示。首先, 输入特征图 $X \in \mathbb{R}^{H \times W \times C}$ 通过 1×1 卷积将通道数增加到 $n \times C$ 。然后, 将特征图分割成多个组, 并对每个组应用倒置瓶颈层处理(除第 1 组外), 倒置瓶颈层有

一个通道扩展比 r 。实验表明, 适当的通道扩展比可以增强深度可分离卷积的表达能, 并且倒置瓶颈层的设计允许在保持计算效率的同时使用大卷积核, 从而获得丰富的特征表示。最后, 将这些组的特征拼接, 进行特征融合。

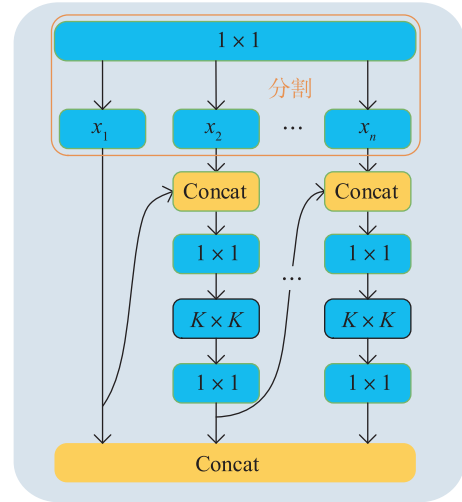


图 5 MS 模块

Fig.5 MS module

2.5 改进后的 YOLOv8s

通过对 YOLOv8s 模型的 4 个主要改进, 模型的整体性能得到了提升。加入通道注意力模块增强了特征选择性; 增加小目标检测头, 提高了模型对小目标的感知能力; 采用残差注意力模块解决了特征信息损失的问题; MS 模块替换 SPPF 模块, 通过层次化特征融合和异构卷积核选择增强了多尺度特征感知能力。改进后的 YOLOv8s 模型如图 6 所示。

3 实验结果与分析

3.1 数据集

本文聚焦于提升自动驾驶系统中小目标检测的性能。KITTI 数据集^[17]为这一目的提供了一个理想的测试平台, 其包含从城市街道到乡村小径的广泛场景, 并且特别标注了小目标物体, 包括行人、自行车等, 这些小目标由于尺寸小、对比度低且常被部分遮挡, 使得其在复杂交通场景中的检测尤其具有挑战性。

本文采用了 KITTI 数据集中的 7 481 张训练图像及其相应的标注信息和 7 518 张测试图像。这些图像涵盖了多种类别的目标, 包括轿车、货车、行人、骑自行车的人等。为了更专注于小目标的检测, 本文进一步筛选了包含较多小目标的图像。具

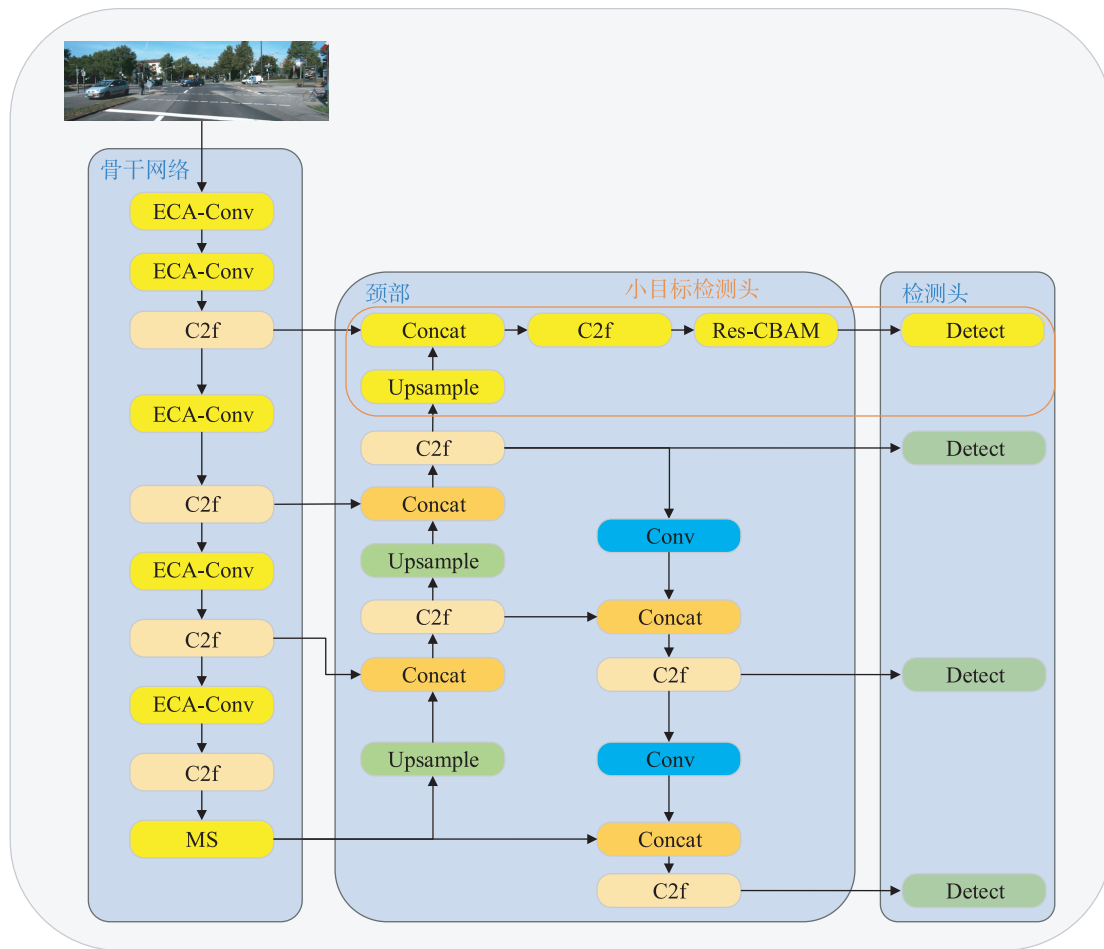


图6 改进后的 YOLOv8s 模型
Fig.6 Improved YOLOv8s model

体方法是根据锚框在图像中所占大小的比例进行筛选,以确保所选图像中小目标的数量较多,从而更准确地衡量模型在小目标检测上的改进效果。

3.2 实验环境

本文实验的设备及环境配置如表 1 所示。

表 1 实验设备及环境配置

Tab.1 Experimental equipment and environment configuration

名称	配置信息
操作系统	Ubuntu 22.04.4 LTS x86_64
Python 版本	3.10.13
框架	Pytorch2.3.0 + CUDA12.1
CPU	Intel Xeon Platinum 8352V(16)
GPU	NVIDIA GeForce RTX 4090
内存	90G

3.3 部分实验参数及评价指标

本文实验部分对模型性能的评估基于精度 (precision)、召回率 (recall) 和平均准确率 (mean

Average Precision, mAP) 三个指标:

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}} \quad (6)$$

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}} \quad (7)$$

$$P_{mAP} = \frac{1}{n} \sum_{i=1}^n P_A(i) \quad (8)$$

其中, P 为精度, R 召回率, P_{mAP} 为平均准确率, N_{TP} 代表模型正确预测 (true positive) 为正类的样本数, N_{FP} 代表模型错误预测 (false positive) 为正类的样本数, N_{FN} 代表模型错误预测为负类 (false negative) 的样本数, n 代表类别的数量, $P_A(i)$ 代表第 i 个类别的平均准确率。

在实验过程中, 本文训练模型 150 个周期, 批次大小为 32。模型训练采用了从 0 开始的方式, 没有利用预训练权重。训练过程中, 初始学习率设置为 0.01, 并且通过 0.000 5 的权重衰减来帮助模型保持良好的泛化性, 避免过拟合。

3.4 实验结果

3.4.1 增加小目标检测头实验分析

本文在 YOLOv8s 模型的基础上增加了小目标检测头, 提出了改进版的 YOLOv8s-S 模型。为了全面评估改进效果, 本文不仅将改进后的 YOLOv8s-S 模型与原始 YOLOv8s 模型进行了直接对比, 还进行了消融实验, 分别测试了单独使用上采样特征图 (YOLOv8s-US) 和单独使用骨干网络的低尺度特征图 (YOLOv8s-LS) 的检测性能。

实验结果如表 2 所示, YOLOv8s 模型的精度、召回率和 mAP 分别为 0.962、0.905 和 0.801; 改进后的 YOLOv8s-S 模型精度、召回率和 mAP 分别提高至 0.963、0.940 和 0.833。与之相比, YOLOv8s-US 模型的精度为 0.962, 召回率为 0.910, mAP 为 0.798; YOLOv8s-LS 模型的精度为 0.963, 召回率为 0.916, mAP 为 0.815。这些结果表明, 虽然单独使用上采样特征图或低尺度特征图的性能有所提升, 但通道拼接融合策略在小目标检测中表现更为显著, 能够在保持较高精度的同时, 显著地提高召回率和整体检测性能。

表 2 添加小目标实验结果

Tab.2 Experimental results with small object detection

模型	P	R	mAP50: 95
YOLOv8s	0.962	0.905	0.801
YOLOv8s-US	0.962	0.910	0.798
YOLOv8s-LS	0.963	0.916	0.815
YOLOv8s-S	0.963	0.940	0.833

注: mAP50: 95 表示在 IoU 阈值从 0.5 到 0.95 之间 (步长为 0.05) 计算出的平均精度值。

3.4.2 注意力检测头实验分析

本文进一步对小目标检测头进行了改进, 尝试引入不同类型的注意力机制 SA (Spatial Attention)^[18]、

GAM (Global Attention Mechanism)、CBAM、Res-CBAM 以提升模型性能。为评估这些改进的效果, 本文对每种注意力机制与原始小目标检测头进行了对比, 实验数据见表 3。

表 3 注意力机制对模型性能影响

Tab.3 Impact of attention mechanisms on model performance

模型	P	R	mAP50: 95
YOLOv8s-S	0.963	0.940	0.833
YOLOv8s-S + SA	0.952	0.940	0.835
YOLOv8s-S + GAM	0.965	0.940	0.836
YOLOv8s-S + CBAM	0.964	0.941	0.834
YOLOv8s-S + Res-CBAM	0.962	0.938	0.837

实验结果表明, 采用不同的注意力机制对模型的性能有不同程度的影响。在 YOLOv8s-S 模型引入 SA 注意力机制后, 精度略有下降至 0.952, 但 mAP 略有提高至 0.835; 在引入通道和空间注意力相结合的 GAM 机制后, 精度提高至 0.965, 召回率保持不变, mAP 进一步提高至 0.836; CBAM 的引入也表现出类似的效果, 精度为 0.964, 召回率略有提高至 0.941, mAP 为 0.834。最后, YOLOv8s-SRC 作为 YOLOv8s-S 添加 Res-CBAM 模块的模型, mAP 值达到 0.837, 领先于其他的注意力机制。

为了直观地评估不同注意力机制对模型性能的影响, 本文采用热力图来展示特征图的可视化结果。在使用的测试集中随机选取了一帧图像, 并利用热力图对其特征进行了直观的展示。这些热力图能够清晰地反映出各种注意力机制在处理数据集时的效果差异。如图 7 所示, 可以观察到不同模型在特征提取方面的显著变化。Res-CBAM 展现出了更优的特征提取能力, 其特征图的激活分布更为均匀且集中。

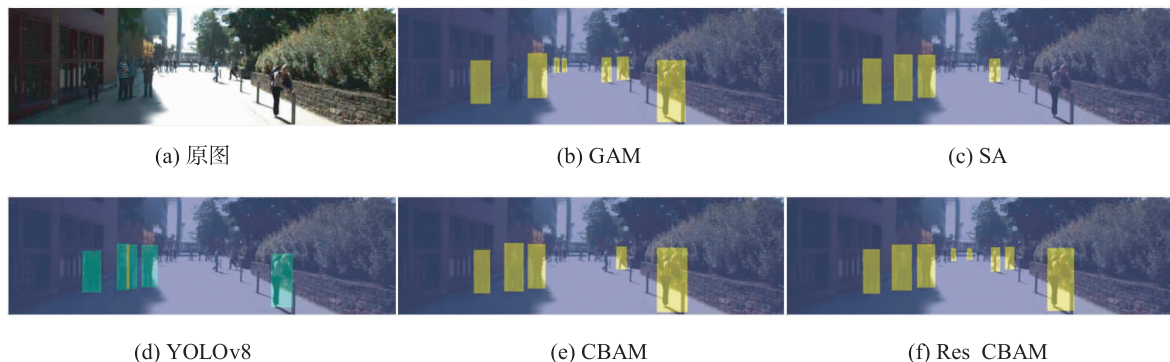


图 7 不同注意力对比

Fig.7 Contrast of different attention

综合来看,引入注意力机制在总体上提高了模型在复杂交通场景中的检测性能,特别是 Res-CBAM 表现出更高的 mAP,证明了这种结合残差连接的注意力机制的潜力。

3.4.3 骨干网络添加注意力实验分析

本文对骨干网络进行了进一步的改进,尝试在 YOLOv8s-SRC 模型的骨干网络中添加注意力机制。实验中对比了 ECA 模块和 CoT (Co-Transformation) 模块^[19]对模型的影响,实验结果见表 4。

表 4 骨干网络添加注意力实验结果
Tab.4 Experiment results of adding attention to the backbone network

模型	P	R	mAP50: 95
YOLOv8s-SRC	0.962	0.938	0.837
YOLOv8s-SRC + CoT	0.963	0.936	0.838
YOLOv8s-SRC + ECA	0.968	0.943	0.839

在骨干网络上添加了 CoT 注意力机制后,精度提高 0.001,但召回率下降了 0.002, mAP 则从 0.837 提高至 0.838。在引入 ECA 注意力机制后,精度从 0.962 提高到 0.968,召回率从 0.938 提高到 0.943, mAP 从 0.837 提高到 0.839。这些结果显示,在 YOLOv8s-SRC 模型基础上添加 ECA 模块的 YOLOv8s-SRCE 模型,能够增强模型的特征选择性,提高模型在复杂交通场景下的检测准确性和鲁棒性。

3.4.4 改进 SPPF 模块实验分析

本文对 YOLOv8s 模型中的 SPPF 模块进行了改进,引入了多种替代方案,包括 MS 模块、BasicRFB (Basic Receptive Field Block) 模块^[20]、SimSPPF (Simplified Spatial Pyramid Pooling-Fast) 模块和 AS-

PP (Atrous Spatial Pyramid Pooling) 模块^[21],并对这些替代方案的性能进行了对比。表 5 展示了这些模型在实验中的性能表现。

表 5 替代 SPPF 模块的不同方案性能对比
Tab.5 Performance comparison of different alternatives to the SPPF module

模型	P	R	mAP50: 95
YOLOv8s-SRCE	0.968	0.943	0.839
YOLOv8s-SRCE + BasicRFB	0.965	0.946	0.842
YOLOv8s-SRCE + SimSPPF	0.962	0.942	0.838
YOLOv8s-SRCE + ASPP	0.965	0.944	0.841
YOLOv8s-SRCE + MS	0.975	0.951	0.867

引入 BasicRFB 模块的改进方案在召回率和 mAP 上分别提高了 0.003,在精度上下降了 0.003。然而,在加入 SimSPPF 模块和 ASPP 模块后,模型的性能与 YOLOv8s-SRCE 相当。相较之下,在 YOLOv8s-SRCE 模型基础上引入 MS 模块的 YOLOv8s-SRCEM 模型中,所有评价指标有明显的提高:精度从 0.968 提高到 0.975,召回率从 0.943 提高到 0.951, mAP 从 0.839 提高至 0.867。

3.4.5 对比实验

本文通过对比实验评估了 YOLOv8s 及其改进版本与 YOLOv5、YOLOX 和其他小目标检测改进模型的性能差异,结果如表 6 所示,其中, S 、 R_c 、 E 、 M 分别表示小目标检测头、检测头通道注意力模块、骨干网络注意力机制和多尺度特征融合。YOLOv8s 在基准性能上表现优异,精度、召回率和 mAP 分别为 0.962、0.905 和 0.801,且其模型大小为 21.4 MB, GFLOPs (Giga Floating Point Operations per second) 为 28.8, FPS (frames per second) 为 85。

表 6 目标检测模型性能对比与 YOLOv8s 改进效果消融实验

Tab.6 Performance comparison of target detection models and ablation experiment of YOLOv8s improved effect

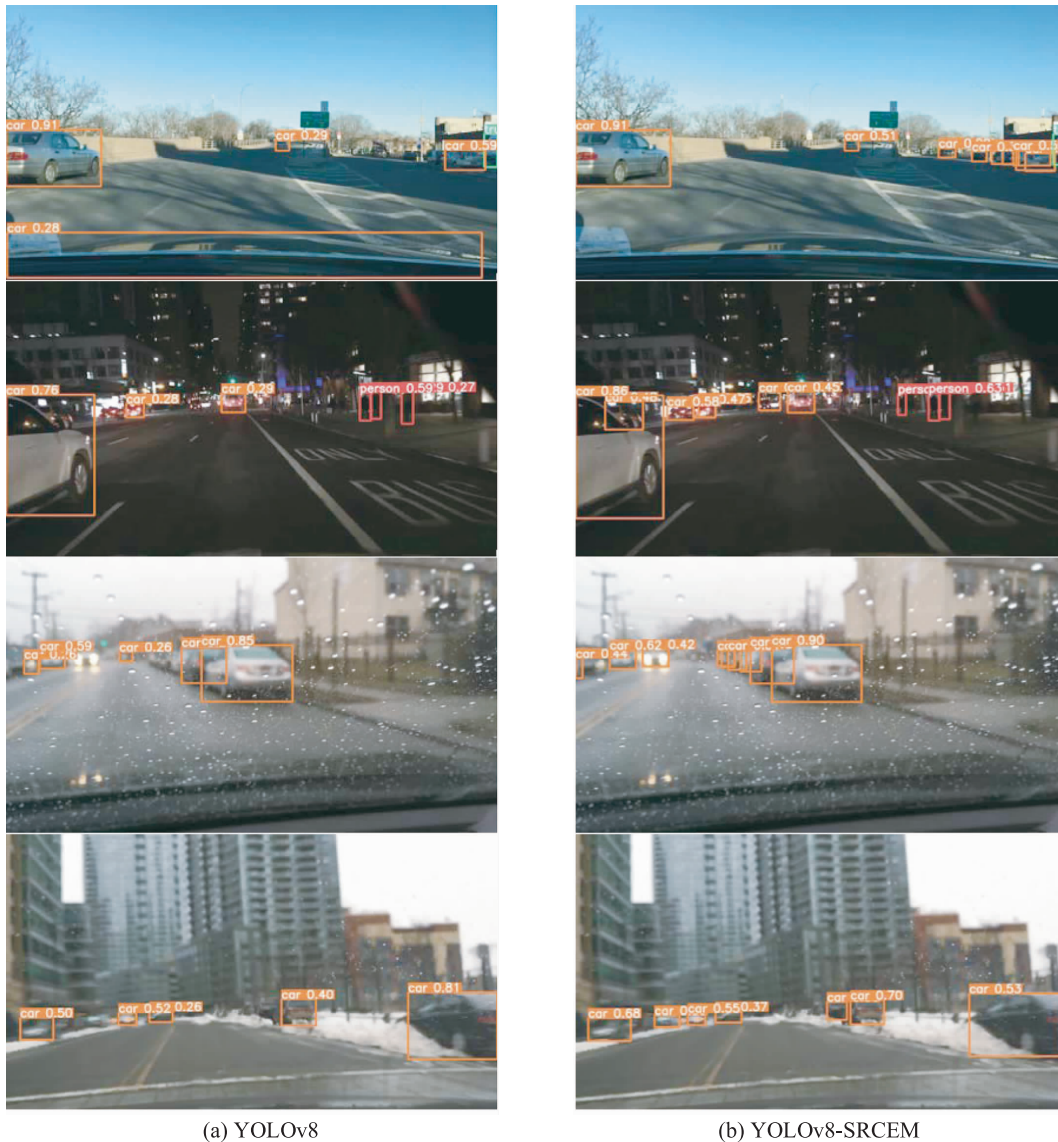
模型	S	R_c	E	M	P	R	mAP50: 95	模型大小 /MB	GFLOPs	FPS
YOLOX					0.940	0.892	0.650	34.4	46.0	52
YOLOv5s					0.934	0.881	0.658	13.7	16.8	82
文[22]模型					0.931	0.890	0.674	12.9	22.5	38
文[9]模型					0.893	0.834			38.7	88
文[23]模型					0.950	0.861	0.700			
YOLOv8s					0.962	0.905	0.801	21.4	28.8	85
YOLOv8s-S	√				0.963	0.940	0.833	22.5	31.6	78
YOLOv8s-SRC	√	√			0.962	0.938	0.837	22.7	32.3	76
YOLOv8s-SRCE	√	√	√		0.968	0.943	0.839	23.1	33.1	72
YOLOv8s-SRCEM	√	√	√	√	0.975	0.951	0.867	25.9	35.2	66

与 YOLOv5 和 YOLOX 相比, YOLOv8s-SRCCEM 模型表现更佳。YOLOv5 的精度、召回率和 mAP 分别为 0.934、0.881 和 0.658, YOLOX 的精度、召回率和 mAP 分别为 0.940、0.892 和 0.650, 而 YOLOv8s-SRCCEM 的精度、召回率和 mAP 分别达到了 0.975、0.951 和 0.867, 提高明显。文[9, 22-23]的模型尽管在小目标检测上有一定进展, 但总体性能仍不如 YOLOv8s-SRCCEM。由于文[9]和文[23]的研究重点更偏向于检测精度的提升或特定场景下的模型表现, 而非推理速度或计算复杂度的优化, 因此其实实验部分未详细报告 GFLOPs 和 FPS 等指标。其中, 表现最佳的文[23]的精度、召回率和 mAP 分别为 0.950、0.861 和 0.700, 均低于 YOLOv8s-SRCCEM。

YOLOv8s-SRCCEM 模型在 YOLOv8s 基础上, 通

过增加小目标检测头、检测头通道注意力模块、骨干网络注意力机制和多尺度特征融合, 实现了显著性能提升。尽管模型大小增加了 4.5 MB, GFLOPs 增加了 6.4, FPS 下降了 19, 但精度、召回率和 mAP 分别提高了 0.013、0.046 和 0.066, 表现优异。

为了展示 YOLOv8s-SRCCEM 模型在小目标检测方面的改进效果, 本文对改进前后的模型检测效果进行对比。图 8 展示了 YOLOv8s 和 YOLOv8s-SRCCEM 在白天、夜晚、雨天和雪天等不同场景下的小目标检测结果对比。从图 8 可以看出, YOLOv8s 在没有针对小目标进行优化的情况下, 存在漏检和检测准确性不足的问题, 特别是在复杂场景(如夜晚和恶劣天气)下更为明显。而 YOLOv8s-SRCCEM 模型显著提升了对小目标的检测能力, 不仅减少了漏



(a) YOLOv8

(b) YOLOv8-SRCCEM

图 8 模型改进前后对比

Fig.8 Comparison before and after model improvement

检现象,还能够更准确地定位和识别更多小目标。对比结果显示, YOLOv8s-SRCEM 在目标定位、边界框的准确度及整体检测质量上均有显著提高,尤其在复杂场景下的表现尤为突出。

4 结论

通过对 YOLOv8s 模型进行多项改进,本文显

著提升了其在目标检测任务中的性能。加入小目标检测头增强了模型对小目标的感知能力,注意力机制提高了特征选择性,通道注意力模块优化了模型的精度和召回率,多尺度特征融合模块提升了对不同尺度特征的感知能力。结合这些改进后的模型在精度、召回率和 mAP 等指标上取得了显著提升,更适用于复杂交通场景中的目标检测任务。

参考文献

- [1] YURTSEVER E, LAMBERT J, CARBALLO A, et al. A survey of autonomous driving: Common practices and emerging technologies[J]. *IEEE Access*, 2020, 8: 58443 – 58469.
- [2] BADUE C, GUIDOLINI R, CARNEIRO R V, et al. Self-driving cars; A survey[J/OL]. *Expert Systems with Applications*, 2021, 165 [2024 – 07 – 01]. <https://www.sciencedirect.com/science/article/pii/S095741742030628X>. DOI: 10.1016/j.eswa.2020.113816.
- [3] LIU L, OUYANG W, WANG X, et al. Deep learning for generic object detection; A survey[J]. *International Journal of Computer Vision*, 2020, 128: 261 – 318.
- [4] MAO J, SHI S, WANG X, et al. 3D object detection for autonomous driving: A comprehensive survey[J]. *International Journal of Computer Vision*, 2023, 131(8): 1909 – 1963.
- [5] HUANG Z, WANG J, FU X, et al. DC-SPP-YOLO: Dense connection and spatial pyramid pooling based YOLO for object detection[J]. *Information Sciences*, 2020, 522: 241 – 258.
- [6] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//*IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway, USA: IEEE, 2017: 2117 – 2125.
- [7] ZHANG Y, GUO Z Y, WU J Q, et al. Real-time vehicle detection based on improved YOLO v5[J/OL]. *Sustainability*, 2022, 14(19) [2024 – 07 – 01]. <https://www.mdpi.com/2071-1050/14/19/12274>. DOI: 10.3390/su141912274.
- [8] LI Z Z. Road aerial object detection based on improved YOLOv5[J/OL]. *Journal of Physics: Conference Series*, 2022, 2171 [2024 – 06 – 29]. <https://iopscience.iop.org/article/10.1088/1742-6596/2171/1/012039/meta>. DOI: 10.1088/1742-6596/2171/1/012039.
- [9] 周飞, 郭杜杜, 王洋, 等. 基于改进 YOLOv8 的交通监控车辆检测算法[J]. *计算机工程与应用*, 2024, 60(6): 110 – 120.
ZHOU F, GUO D D, WANG Y, et al. Traffic surveillance vehicle detection algorithm based on improved YOLOv8[J]. *Computer Engineering and Applications*, 2024, 60(6): 110 – 120.
- [10] WANG C Y, LIAO H Y M, WU Y H, et al. CSPNet: A new backbone that can enhance learning capability of CNN[C]//*IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. Piscataway, USA: IEEE, 2020: 390 – 391.
- [11] WANG K, LIEW J H, ZOU Y, et al. Panet: Few-shot image semantic segmentation with prototype alignment[C]//*IEEE/CVF International Conference on Computer Vision*. Piscataway, USA: IEEE, 2019: 9197 – 9206.
- [12] ZHENG Z H, WANG P, LIU W, et al. Distance-IoU loss: Faster and better learning for bounding box regression[C]//*Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, 34(7): 12993 – 13000.
- [13] WANG Q, WU B, ZHU P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks[C]//*IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway, USA: IEEE, 2020: 11534 – 11542.
- [14] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//*IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway, USA: IEEE, 2018: 7132 – 7141.
- [15] 袁磊, 唐海, 陈彦蓉, 等. 改进 YOLOv5 的复杂环境道路目标检测方法[J]. *计算机工程与应用*, 2023, 59(16): 212 – 222.
YUAN L, TANG H, CHEN Y R, et al. Improved YOLOv5 method for road object detection in complex environments[J]. *Journal of Computer Engineering & Applications*, 2023, 59(16): 212 – 222.
- [16] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module[C]//*European conference on computer vision*. Berlin, Germany: Springer, 2018: 3 – 19.
- [17] GEIGER A, LENZ P, STILLER C, et al. Vision meets robotics; The KITTI dataset[J]. *The International Journal of Robotics*

- Research, 2013, 32(11): 1231 – 1237.
- [18] ZHANG Q L, YANG Y B. SA-Net: Shuffle attention for deep convolutional neural networks[C]//IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway, USA: IEEE, 2021: 2235 – 2239.
- [19] LI Y, YAO T, PAN Y, et al. Contextual transformer networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(2): 1489 – 1500.
- [20] LIU S, HUANG D. Receptive field block net for accurate and fast object detection[C]//European Conference on Computer Vision. Berlin, Germany: Springer, 2018: 385 – 400.
- [21] 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, 2017, 40(4): 834 – 848.
- [22] 盛博莹, 侯进, 李嘉新, 等. 面向复杂交通场景的道路目标检测方法[J]. 计算机工程与应用, 2023, 59(15): 87 – 96.
SHENG B Y, HOU J, LI J X, et al. Road target detection method for complex traffic scenarios[J]. Journal of Computer Engineering & Applications, 2023, 59(15): 87 – 96.
- [23] 董恒祥, 潘江如, 董芙楠, 等. 基于改进 YOLOv5s 模型的车辆及行人检测方法[J]. 北华大学学报(自然科学版), 2024, 25(2): 244 – 254.
DONG H X, PAN J R, DONG F N, et al. Vehicle and pedestrian detection method based on improved YOLOv5s model[J]. Journal of Beihua University (Natural Science Edition), 2024, 25(2): 244 – 254.

作者简介

濮志远(2001 –), 男, 硕士。研究领域为无人驾驶汽车环境感知。

罗素云(1975 –), 女, 硕士, 副教授。研究领域为无人驾驶汽车环境感知及控制。

(上接第 631 页)

- [20] SANDLER M, HOWARD A, ZHU M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2018: 4510 – 4520.
- [21] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]//International Conference on Machine Learning. New York, USA: PMLR, 2015: 448 – 456.
- [22] NAIR V, HINTON G E. Rectified linear units improve restricted boltzmann machines[C]//International Conference on Machine Learning. New York, USA: PMLR, 2010: 807 – 814.
- [23] ROHMER E, SINGH S P N, FREESE M. V-REP: A versatile and scalable robot simulation framework[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, USA: IEEE, 2013: 1321 – 1326.
- [24] 赵亮, 杨铁, 于鹏, 等. 一种用于机械臂拟人化控制的学习框架[J]. 机器人, 2023, 45(5): 513 – 522.
ZHAO L, YANG T, YU P, et al. A Learning framework for anthropomorphic control of robotic arms[J]. Robot, 2023, 45(5): 513 – 522.
- [25] 孙先涛, 陈文杰, 陈伟海, 等. 一种机械手定位装置及其定位方法: CN112873176B[P]. 2021 – 09 – 07.
SUN X T, CHEN W J, CHEN W H, et al. A Mechanical hand positioning device and its positioning method: CN112873176B [P]. 2021 – 09 – 07.
- [26] TANG B, CORSARO M, KONIDARIS G, et al. Learning collaborative pushing and grasping policies in dense clutter[C]//IEEE International Conference on Robotics and Automation. Piscataway, USA: IEEE, 2021: 6177 – 6184.

作者简介

李德平(1987 –), 男, 博士, 讲师。研究领域为计算机视觉, 3 维视觉, 机器人视觉感知。

洪楷宣(2000 –), 男, 硕士生。研究领域为计算机视觉。