DOI:10.3969/j.issn.1672-0032.2025.05.012

基于 RSG-YOLOv10n 道路场景抗干扰与小目标检测方法

孔飞一,付振山*,王玉刚,付聪,戴先鑫,马栋
山东交通学院船舶与港口工程学院,山东 威海 264209

摘要:针对自动驾驶道路场景目标检测存在背景干扰、远距离小目标、特征缺失等问题,以 YOLOv10n(you only look once version 10n)模型为基准,引入感受视野卷积(receptive field attention convolution, RFAConv)模块替换 Conv 模块,提取多尺度感受野空间特征,结合注意力机制动态分配权重,增强模型复杂图像处理能力;引入小目标增强金字塔(small object enhance pyramid, SOEP)模块,采用改进的 CSP-OmniKernel(cross stage partial-omniKernel)模块进行特征信息整合,显著提高小目标检测性能;引入全局通道空间注意力(global channel-spatial attention,GCSA)模块,通过通道注意力、通道洗牌与空间注意力的耦合机制强化特征图表达,捕捉特征图全局依赖关系,增强特征提取能力,形成 RSG-YOLOv10n(RFAConv-SOEP-GCSA-YOLOv10n)复杂道路场景小目标检测模型,开展消融试验、模型性能对比试验、通用性验证试验和可视化检测效果试验。试验结果表明:引入 RFAConv、SOEP、GCSA 3 个模块形成的 RSC-YOLOv10n 模型的精确率 P、召回率 R、交并比阈值为 50 时的平均精度均值 E_{mAP50}、交并比阈值从 50 增至 95(步长 5)时的平均精度均值 $E_{\mathrm{mAP50-95}}$ 比 YOLOv10n 模型分别增大 6.3 百分点、2.5 百分点、3.5 百分点和 2.8 百分点,检测精度明显提高;与较低参数量模型(YOLOv5n、YOLOv8n、YOLOv10n)、相近参数量模型(YOLOv3-ting、YOLOv6n、YOLOv7-ting)及较高参数量模型(RT-DETR-L)相比,RSC-YOLOv10n 模型在较低参数量和浮点运算次数基础上,检测精度最高;RSC-YOLOv10n 模型对由沙尘、大雪、浓雾等多种恶劣气象图片组成的通用性验证数据集进行可视化检测,其 E_{mAP50}、$E_{\mathrm{mAP50-95}}$ 比 YOLOv10n 模型分别增大 0.9 百分点、1.3 百分点,鲁棒性和泛化能力较高;RSG-YOLOv10n 模型在可视化检测效果试验中对密集遮挡、复杂光照、恶劣天气及远近目标共存等道路场景的特征提取与小目标检测能力较强。

关键词:小目标检测;背景干扰;特征缺失;RSG-YOLOv10n;注意力机制
中图分类号:U495;TP391.41　　**文献标志码**:A　　**文章编号**:1672-0032(2025)05-0090-12
引用格式:孔飞一,付振山,王玉刚,等.基于 RSG-YOLOv10n 道路场景抗干扰与小目标检测方法[J].山东交通学院学报,2025,33(5):90-101.
KONG Feiyi, FU Zhenshan, WANG Yugang, et al. Anti-interference and small object detection method for road scene based on RSG-YOLOv10n[J]. Journal of Shandong Jiaotong University,2025,33(5):90-101.

0　引言

随着人工智能技术的高速发展,自动驾驶成为智能交通主要研究方向之一[1]。自动驾驶的关键技术是在复杂道路环境中准确识别小目标和特征不明显的物体,并有效应对环境干扰。目标检测技术[2]可为解决这一问题提供支持,特别是基于深度学习目标检测的模型[3],在提高目标识别的精确度和处理速度

收稿日期:2025-06-09
基金项目:山东省自然科学基金项目(ZR2022QF149)
第一作者简介:孔飞一(1999—),男,南京人,硕士研究生,主要研究方向为 SLAM 算法和图像处理,E-mail:2324294576@qq.com。
*通信作者简介:付振山(1976—),男,山东泰安人,教授,硕士研究生导师,工学博士,主要研究方向为工业机器人、啮合原理和机械电子设备等,E-mail:273482839@qq.com。

方面技术优势明显[4]。传统深度学习算法分为两阶段算法和单阶段算法[5-7]两大类,自动驾驶道路场景目标识别需具备高效性,一般采用单阶段算法中 YOLO(you only look once)模型完成目标检测任务[8-9]。

国内外学者在道路场景目标识别领域对 YOLO 模型开展相关研究:赵龙阳等[10]提出一种基于 YOLOv8 改进的算法 MDSD-YOLO(mlca dcnv4 seam dualConv-YOLO),通过适应不规则检测目标增强小目标检测精度;Li 等[11]提出 Attention-YOLOv4 模型,结合通道注意力机制与残差块,增强网络获取目标局部特征;Xue 等[12]提出 YOLO-FSE(YOLO-fasterNet shuffle attention EIoU)模型,主干网络采用 C3faster 模块和 Shuffle Attention 模块分别替换 YOLOv5 中 C3 模块和损失函数,使 YOLO-FSE 模型轻量化,提高模型泛化能力和特征提取能力;Fan 等[13]提出 YOLOv8 模型改进算法,采用 BIFPN(bidirectional feature pyramid network)模块提高小目标和遮挡目标的识别精度,采用 SimAM(simple attention module)模块抑制复杂背景干扰;Luo 等[14]通过融合 Ghost 模块和 EMA(efficient multi-scale attention)模块改进 YOLOv8 算法,显著提高模型的运算速度,并优化特征提取的效果。上述研究均为通过引入注意力机制、模块轻量化、多尺度特征融合等技术改进 YOLO 模型,提高模型在道路环境中的目标检测性能。此类模型识别道路环境复杂、噪声干扰以及小目标像素稀疏等情形的目标尚存在一定不足,检测精度较低[15-16]。

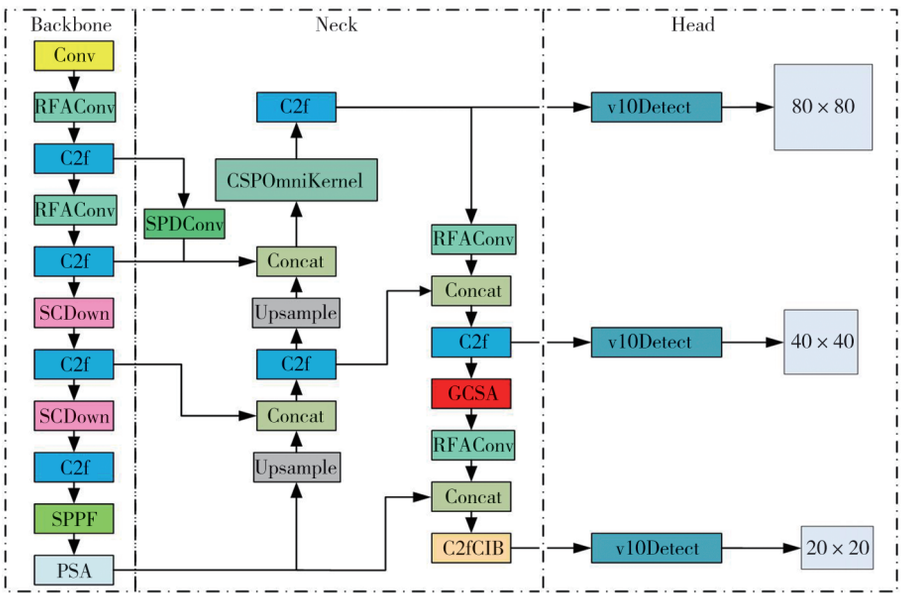
自动驾驶中道路场景目标检测区别于一般目标检测任务,主要体现在三个方面:1)检测环境常面临昼夜光照变化反差较大和恶劣天气条件(如雨、雪、雾等),光照、能见度、背景干扰等因素影响目标检测效果;2)存在大量远距离小目标,目标尺寸小、特征模糊,如远处的行人、车辆、交通标志等;3)交通场景中常伴随高密度、大面积遮挡目标,导致部分目标特征缺失,例如在拥堵路段或行人流量较大的交叉路口,目标间相互遮挡,导致部分目标特征缺失,存在漏检现象。针对自动驾驶交通场景下的目标检测任务,需考虑光照变化和恶劣天气引起的背景干扰、远距离小目标、检测目标密集导致特征缺失等问题。

为解决上述问题,在 YOLOv10n(you only look once version 10n)模型的基础上,针对光照变化和恶劣天气下检测鲁棒性问题,在 YOLOv10n 网络中引入感受视野卷积(receptive field attention convolution, RFAConv)模块,通过自适应感受野,减少噪声干扰,以期增强模型处理复杂图像的能力;针对检测远距离小目标问题,提出小目标增强金字塔(small object enhance pyramid, SOEP)模块,采用 SPDConv(space-to-depth convolution)提取 P2 层的小目标特征并融合至 P3 层,采用改进的 CSP-OmniKernel(cross stage partial-omniKernel)模块进行特征整合,以期提高对小目标的检测性能并降低检测网络的参数规模;针对检测目标密集导致特征缺失问题,采用全局通道空间注意力(global channel-spatial attention, GCSA)模块,通过通道支路和空间支路建立跨通道依赖并捕捉位置依赖,以期增强输入特征图的表征能力和全局依赖关系。构建 RSG-YOLOv10n(RFAConv-SOEP-GCSA-YOLOv10n)复杂道路场景小目标检测模型,以期增强对密集遮挡、复杂光照、恶劣天气及远近目标共存等道路场景特征提取与小目标检测能力,为自动驾驶在复杂道路环境下进行小目标检测提供参考。

1　RSG-YOLOv10n 模型

YOLOv10 模型保留了 YOLO 模型实时高效检测目标图像的特点,并在前代模型的基础上改进和优化图像处理算法,但对自动驾驶道路场景中环境干扰和小目标检测的效果较差。为解决这一问题,改进 YOLOv10 模型,提出 RSG-YOLOv10n 复杂道路场景小目标检测模型,模型结构如图 1 所示。

对 YOLOv10 模型的改进包括:1)在骨干网络(Backbone Network)中引入 RFAConv 模块,增强模型在复杂道路环境下抵抗环境噪声干扰的能力;2)在颈部网络(Neck Network)采用 SPDConv 卷积和拼接操作 Concat,对包含丰富小目标特征的 P2 层进行特征处理,采用改进的 CSPOmniKernel 模块进行特征整合,提高模型对小目标的检测性能;3)在颈部网络引入 GCSA 模块,增强输入特征图的表征能力和全局依赖关系;4)在头部网络(Head Network)采用 v10Detect 模块接收颈部网络的多尺度特征图作为输入内容,采用解耦头结构处理分类和回归任务,采用无锚框设计适应不同形状和尺寸的检测目标,提高模型预测的精确度和效率。

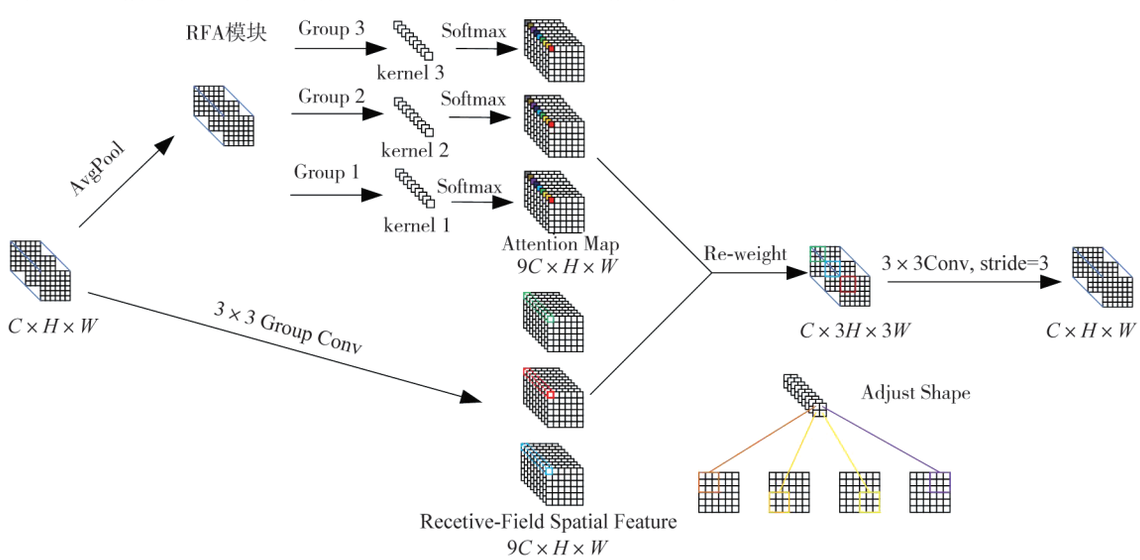


注：Conv 为卷积操作；RFAConv 为感受视野卷积操作；C2f 为跨阶段连接和特征融合模块；SCDown 为特征图下采样模块；SPPF 为快速空间金字塔池化模块；PSA 为高效注意力机制模块；CSPOmnikernel 为改进的特征整合模块；SPDConv 为下采样卷积操作；Upsample 为上采样操作；GCSA 为全局通道空间注意力模块；C2fCIB 为轻量化设计的特征增强模块；v10Detect 为根据提取的图像特征预测目标检测结果；80×80、40×40、20×20 为特征图检测结果大小（Byte）。

图 1 RSG-YOLOv10n 模型结构

1.1 RFAConv 模块

RFAConv 模块为新型卷积神经网络架构，核心为感受野注意力（receptive field attention，RFA）模块[17]，通过分组卷积高效提取多尺度感受野空间特征，并采用空间注意力机制评估输入特征图的重要性，通过生成权重掩膜、降低背景权重、增强前景目标权重的方法动态分配特征权重，实现抑制背景噪声并提高目标识别能力。RFAConv 模块增强关注感受野内空间特征信息，使 RSG-YOLOv10n 模型更好地理解局部上下文信息，进一步提高复杂背景下的目标识别精确率。RFAConv 模块结构如图 2 所示。



注：$C\times H\times W$ 为输入特征图数据，C 为通道数，H 为特征图高度，W 为特征图宽度；AvgPool 为平均池化操作；3×3 Group Conv 为分组卷积操作；Group 1、Group 2、Group 3 为输入特征图数据的 3 个分组；kernel 1、kernel 2、kernel 3 为 3 个 1×1 的卷积核；Softmax 为生成注意力图的函数；Attention Map 为注意力图；Recetive-Field Spatial Feature 为感受野空间特征；Re-weight 为重加权，通过调整特征权重优化模型训练策略；Adjust Shape 为变换特征图的几何结构；3×3Conv 为 3×3 卷积操作；stride 为跨距。

图 2 RFAConv 模块结构

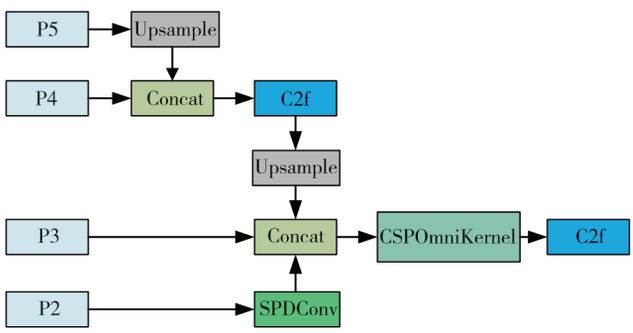
RFAConv 模块的工作流程为：1）通过快速分组卷积将输入特征图转换为感受野空间特征图，并调整其尺寸为原始尺寸的 9 倍；2）对输入特征图进行全局平均池化，提取每个感受野的特征信息，并通过 1×1 分组卷积和 Softmax 操作生成每个感受野的注意力权重；3）将注意力图与感受野空间特征相乘进行重加权，经 3×3 卷积操作后得到最终特征图输出结果：

$$\boldsymbol{F}=\mathrm{Softmax}\{\boldsymbol{g}^{k\times k}[\mathrm{AvgPool}(\boldsymbol{X})]\}\times\mathrm{ReLU}\{\mathrm{Norm}[\boldsymbol{g}^{k\times k}(\boldsymbol{X})]\}=\boldsymbol{A}_{\mathrm{rf}}\boldsymbol{F}_{\mathrm{rf}}\ ,$$

式中：$\boldsymbol{g}^{k\times k}$ 为 $k\times k$ 的分组卷积，k 为卷积核的大小；ReLU 为修正线性单元激活函数；Norm 为归一化函数，$\boldsymbol{X}$ 为输入特征图矩阵；$\boldsymbol{A}_{\mathrm{rf}}$ 为注意力图矩阵；$\boldsymbol{F}_{\mathrm{rf}}$ 为感受野空间特征矩阵。

1.2　SOEP 模块

在传统方法中添加 P2 检测层提高小目标检测性能，存在两方面不足：1）计算量增大，导致模型检测效率降低；2）后处理阶段的耗时问题严重影响模型整体运行的实时性。为解决这些问题，基于路径聚合特征金字塔网络（path aggregation feature pyramid network，PAFPN）构建 SOEP 小目标增强金字塔模块[18]，结合 P2 和 P3 特征图增强模型对小目标的识别能力，减少冗余计算。SOEP 模块结构如图 3 所示。



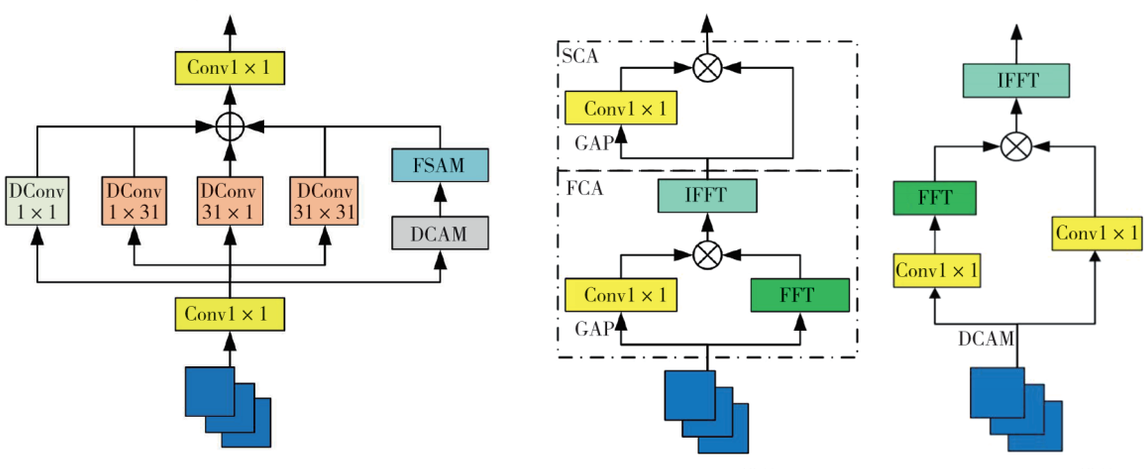
注：P2、P3、P4、P5 为对应不同分辨率或语义层次的特征检测层。

图 3　SOEP 模块结构

SOEP 模块工作流程为：1）通过 SPDConv 卷积操作将 P2 层的特征信息与 P3 层融合；2）采用 CSPOmnikernel 模块处理融合后的特征信息，通过多尺度卷积和频率域变换增强特征表示，增强模型对不同尺度目标的检测能力，减小计算量。

CSPOmnikernel 模块为 CSP 结构与 OmniKernel 结构[19]的结合，将输入特征图分为两个分支：一个分支通过 OmniKernel 模块进行复杂的特征变换，另一个分支则保留原始特征，两个分支通过 1×1 卷积层（cv2）进行通道调整和特征融合。CSPOmnikernel 模块既保留了特征图的原始特征，又融合复杂变换的特征，增强了模型的表达能力。

OmniKernel 模块特征变换的工作流程为：给定输入特征 $\boldsymbol{X}\in\mathrm{R}^{C\times H\times W}$，经过 1×1 卷积处理后，特征信息被送到 3 个分支，即局部分支、大分支和全局分支，3 个分支的处理结果通过加法融合并经 1×1 卷积进行调制。局部分支由 1×1 深度可分离卷积（DConv）组成，实现局部信号调制；大分支由 3 个并行的 1×31、31×1、31×31 的深度可分离卷积组成，捕获带状上下文信息；全局分支由双域通道注意力模块（dual channel attention module，DCAM））和频域空间注意力模块（frequency-based spatial attention module，FSAM）组成，增强从频域到空域的多层次特征，提高模型在图像重建任务中的全局建模能力。OmniKernel 模块结构如图 4 所示。



a）总结构　　b）DCAM 模块　　c）FSAM 模块

图 4　OmniKernel 模块结构

DCAM 模块由空间通道注意力模块(spatial channel attention, SCA)和频域通道注意力模块(frequency channel attention, FCA)两个模块构成,在频域和空域采用全局平均池化(global average pooling, GAP)、快速傅里叶变换(fast fourier transform, FFT)、逆快速傅里叶变换(inverse fast fourier transform, IFFT)和 1×1 卷积增强特征信息通道维度的粗粒度。采用 GAP 将特征图压缩为单维度,得到每个通道的全局信息,采用 FFT 将特征图信息从空域转换至频域,采用 IFFT 将经过注意力调制特征图信息从频域转换回空域。FSAM 模块是将 DCAM 模块输出的特征信息通过两个 1×1 卷积调制,经 FFT 变换后在频域逐元素相乘,增强高频信息,再经 IFFT 逆变换恢复至空域,通过关注特征信息多种频率成分,提高图像重建质量。

1.3 GCSA 模块

在传统卷积神经网络中,因常忽视通道间的关系,可能削弱模型捕捉全局信息的能力,采用通道注意力机制虽能增强特征表达,但通道间的信息融合不充分,模型无法有效利用空间信息,易忽略图像中的关键细节。采用融合通道注意力机制、通道洗牌机制和空间注意力机制的 GCSA 模块[20]增强输入特征图的表达能力和捕捉特征图的全局依赖关系,提高模型对关键特征的捕捉能力和整体性能表现。

通道注意力机制的工作流程为:1)将输入特征图的维度由 $C\times H\times W$ 变为 $W\times H\times C$;2)通过多层感知器(multilayer perceptron, MLP)捕捉通道间依赖关系,第一层感知器将通道数缩减为原来的 1/4,通过 ReLU 激活函数进行非线性处理,第二层感知器将通道数恢复到原始维度;3)将特征图维度恢复到 $C\times H\times W$,通过 Sigmoid 激活函数生成通道注意力图,将输入特征图和通道注意力图逐元素相乘得到增强的特征图。

通道洗牌机制的工作流程为:1)将增强后的特征图分为 4 组,每组包含 $C/4$ 个通道;2)对分组后的特征图进行转置操作;3)打乱各组内的通道顺序,并将打乱顺序后的特征图恢复为原始维度($C\times H\times W$)。通过通道洗牌机制能更好地混合特征信息,增强特征表达能力。

空间注意力机制的工作流程为:1)输入特征图,通过 7×7 卷积层操作,将通道数缩减为原来的 1/4;2)采用批归一化层和 ReLU 激活函数进行非线性变换;3)再次通过 7×7 卷积层操作,通道数恢复至原始维度;4)采用批归一化层和 Sigmoid 激活函数生成空间注意力图。将洗牌后的特征图和空间注意力图逐元素相乘,得到最终的输出特征图。

2 试验验证

2.1 试验准备

1)试验环境(软硬件)。进行试验时,配备 Windows10 操作系统、Core(TM)i7-13400F 处理器、RTX4060 显卡、16 GB 内存以及 8 GB 显存的高性能计算机,深度学习框架 Pytorch 1.13.1,并行计算平台 CUDA 11.7。

2)试验过程中的参数选择。图片为 640 像素×640 像素,优化器为 SGD,学习率为 0.01,Batchsize 为 8,迭代次数为 250。

3)试验数据。采用华为 SODA10M 数据集[21],从 32 个不同的城市中选取 10 000 张图片,包括不同天气、时间段的道路场景,划分并标注 6 种主要自动驾驶场景标签类别,分别为 Person(行人)、Bicycle(自行车)、Car(小型车)、Truck(重型卡车)、Electric vehicles(电动车)和 Tricycles(三轮车)。将图片数按 6∶2∶2 划分为训练集、测试集、验证集。

2.2 评价指标

评价模型性能的指标包括精确率 P、召回率 R、平均精度均值(mean average precision, mAP)E_{mAP}、参数量(param)N_{p}、浮点运算次数(floating point operations, FLOPs)N_{f}、帧数(frames per second, FPS)N_{fps}。

精确率反映模型预测结果的可靠性。召回率反映模型检测的全面性。平均精度(average precision, AP)反映模型对单类检测结果的准确性,平均精度均值为多类平均精度的平均值,通常取交并比

（intersection over union，IoU）阈值为 50 时的平均精度均值 E_{mAP50} 及 IoU 阈值从 50 增至 95（步长 5）时的平均精度均值 $E_{mAP50-95}$ 作为模型检测结果准确性的评价指标，E_{mAP50} 和 $E_{mAP50-95}$ 越大，模型检测结果的精确率越高。参数量反映模型的规模和学习能力，参数量越大，模型通常捕捉的图像特征信息越复杂。浮点运算次数为模型计算量参数，衡量模型的计算复杂度，计算量参数越大，对中央处理器（central processing unit，CPU）的计算能力要求越高。帧数决定模型检测性能能否满足实时应用需求（如自动驾驶需 $N_{fps}\geqslant 30$ 帧/s）[1]。

2.3　RFAConv 模块验证试验

针对光照变化和恶劣天气对道路环境目标检测产生背景干扰，导致识别准确率低的问题，在 YOLOv10n 模型中引入 RFAConv 模块替换原 Conv 模块。为验证 RFAConv 模块的有效性，采用 Conv、SPDConv、RFCBAMConv、RFCAConv 模块开展性能对比试验。SPDConv 模块采用无损下采样技术，RFCBAMConv 模块采用通道-空间双重注意力机制，RFCAConv 模块将注意力机制嵌入感受野，RFAConv 模块通过自适应扩张感受野并同步生成空间权重掩码，试验结果如表 1 所示。由表 1 可知：引入 RFAConv 模块的 E_{mAP50}、$E_{mAP50-95}$ 比 Conv 模块分别增大 1.3 百分点、1.4 百分点，比 SPDConv 模块分别增大 1.2 百分点、0.9 百分点，比 RFCBAMConv 模块分别增大 1.9 百分点、1.1 百分点，比 RFCAConv 模块分别增大 1.2 百分点、0.5 百分点。RFAConv 模块明显提高目标检测效果。

表 1　RFAConv 模块验证试验结果

模块	E_{mAP50}/%	$E_{mAP50-95}$/%
Conv	59.7	40.2
SPDConv[22]	59.8	40.7
RFCBAMConv[15]	59.1	40.5
RFCAConv[15]	59.8	41.1
RFAConv[15]	61.0	41.6

2.4　SOEP 模块验证试验

在交通道路场景中远处的行人和车辆等小目标较多，目标尺寸小、特征不明显，检测难度增大。为提高小目标的检测精度，在 YOLOv10n 模型中引入 SOEP 模块。该模块中 DConv 卷积核尺寸为 K（$K=2^n-1$，$n=4$、5、6），K 越大，检测精度越高，检测模型的计算量越大。为研究 K 对检测精度和计算量的影响，进行对比试验。Baseline 为基准模型，即 YOLOv10n 模型。定义 Z 为检测精度的提高量与浮点运算增加量之比，即

$$Z=\Delta E_{mAP50}/\Delta N_f\times 10^9,$$

式中：$\Delta E_{mAP50}=E_{mAP50}(K)-E_{mAP50}(\text{Baseline})$，$\Delta N_f=N_f(K)-N_f(\text{Baseline})$，$Z$ 越大，表明模型达到目标检测精度时付出的计算量越小。

试验结果如表 2 所示。由表 2 可知：当 K 分别为 15、31、63 时，引入 SOEP 模块的模型的 E_{mAP50} 比 Baseline 基准模型分别增大 1.6 百分点、2.5 百分点、2.7 百分点。综合考虑，当 $K=31$ 时，模型的平均精度均值提高较大，计算量增长较小，SOEP 模块性能最优。

表 2　SOEP 模块 K 值对比试验结果

参数	E_{mAP50}/%	$N_f/10^9$	Z
Baseline	59.7	8.4	
$K=15$	61.3	11.5	0.516
$K=31$	62.2	12.2	0.657
$K=63$	62.4	14.7	0.428

为验证引入 SOEP 模块的模型对小目标识别的效果，采用 Baseline 基准模型和引入小目标检测层（P2）的模型进行性能对比试验，试验结果如表 3 所示。由表 3 可知：引入 P2 的检测模型的 E_{mAP50}、$E_{mAP50-95}$ 比 Baseline 基准模型分别增大 0.8 百分点、0.7 百分点，引入 SOEP 的检测模型的 E_{mAP50}、$E_{mAP50-95}$ 比 Baseline 基准模型分别增大 2.5 百分点、1.6 百分点；引入 P2 的检测模型的浮点运算次数 N_f 比 Baseline 基准模型大 6.7×10^9，引入 SOEP 的检测模型的浮点运算次数 N_f 比 Baseline 基准模型大 3.8×10^9，引入 SOEP 的检测模型的平均精度均值优于引入 P2 的检测模型，且计算量增加较少。

表 3　SOEP 模块验证试验结果

模块	E_{mAP50}/%	$E_{mAP50-95}$/%	$N_f/10^9$
Baseline	59.7	40.2	8.4
P2	60.5	40.9	15.1
SOEP	62.2	41.8	12.2

2.5 GCSA 模块验证试验

在道路环境检测中,因目标密集、遮挡导致部分图像缺失,特征信息不全,检测难度增大,在 YOLOv10n 模型中引入 GCSA 模块,提高模型对图像特征信息的提取能力。为验证引入 GCSA 模块后模型的特征提取能力,选取 CBAM 模块、SE 模块、EMA 模块 3 种典型注意力机制模块及 Baseline 基准模型进行性能对比试验,试验结果如表 4 所示。

由表 4 可知:引入 GCSA 模块的模型平均精度均值 E_{mAP50}、$E_{mAP50-95}$ 比 Baseline 基准模型分别增大 0.6 百分点、0.7 百分点,比 CBAM 模块分别增大 0.6 百分点、0.3 百分点,比 SE 模块分别增大 0.9 百分点、0.8 百分点,比 EMA 模块分别增大 0.4 百分点、0.6 百分点。GCSA 模块采用卷积神经网络增强对特征图的表征能力和对全局依赖关系捕捉能力,提高了模型对道路环景中密集遮挡目标的检测精度。

表 4 GCSA 模块验证试验结果

模块	E_{mAP50}/%	$E_{mAP50-95}$/%
Baseline	59.7	40.2
CBAM[23]	59.7	40.6
SE[24]	59.4	40.1
EMA[25]	59.9	40.3
GCSA	60.3	40.9

2.6 消融试验

为验证引入 SOEP、RFAConv、GCSA 3 种模块后 RSG-YOLOv10n 模型的改进效果及对复杂道路环境的检测精度,以 YOLOv10n 模型为基准模型,对 3 种模块进行不同组合,进行消融试验的结果见表 5。

表 5 消融试验结果

基准模型	SOEP	RFAConv	GCSA	P/%	R/%	E_{mAP50}/%	$E_{mAP50-95}$/%	$N_p/10^6$	$N_f/10^9$	N_{fps}/(帧·s^{-1})
YOLOv10n				65.8	54.4	59.7	40.2	2.70	8.4	500
	√			64.4	57.6	62.2	41.8	3.01	12.2	285
		√		70.1	54.5	61.0	41.6	2.86	8.7	294
			√	69.2	53.6	60.3	40.9	3.11	9.7	396
	√	√		71.9	56.5	62.8	42.8	3.16	12.4	200
	√	√	√	72.1	56.9	63.2	43.0	3.58	13.7	169

注:√为在基准模型中采用改模块。

由表 5 可知:相较于 YOLOv10n 模型,单引入 SOEP 模块后,E_{mAP50}、$E_{mAP50-95}$ 分别增大 2.5 百分点、1.6 百分点,表明模型提高了小目标检测能力;单引入 RFAConv 模块后,P、E_{mAP50}、$E_{mAP50-95}$ 分别增大 4.3 百分点、1.3 百分点、1.4 百分点,表明 RFAConv 模块抑制了背景噪声,提高了检测精度;单引入 GCSA 模块后,P、E_{mAP50}、$E_{mAP50-95}$ 分别增大 3.4 百分点、0.6 百分点和 0.7 百分点,表明 GCSA 模块有助于提高道路场景中密集遮挡目标的检测精度;引入 3 种模块后,P、R、E_{mAP50}、$E_{mAP50-95}$ 分别增大 6.3 百分点、2.5 百分点、3.5 百分点、2.8 百分点,模型的检测精度明显提高,N_p 增大 32.6%,N_f 增大 63.1%,模型计算量增长较大,后期对模型进行轻量化处理。N_{fps} 减至 169 帧/s,满足自动驾驶的目标检测实时性要求。

2.7 模型性能对比试验

为验证 RSG-YOLOv10n 模型改进有效性,与目前广泛使用的 YOLO 系列模型及 RT-DETR-L 模型进行对比,试验结果如表 6 所示。

由表 6 可知:与较低参数量模型相比,RSC-YOLOv10n 模型的 E_{mAP50}、$E_{mAP50-95}$ 比

表 6 不同模型性能对比试验结果

模型	E_{mAP50}/%	$E_{mAP50-95}$/%	$N_p/10^6$	$N_f/10^9$
YOLOv3-ting	46.8	30.4	8.86	5.6
YOLOv5n	55.3	36.5	1.90	4.5
YOLOv6n	54.2	36.2	4.70	11.4
YOLOv7-ting	56.7	35.3	6.20	5.8
YOLOv8n	56.4	37.1	3.00	8.7
RT-DETR-L	55.4	35.3	20.00	60.0
YOLOv10n	59.7	40.2	2.70	8.4
RSC-YOLOv10n	63.2	43.0	3.58	13.7

YOLOv5n 模型分别增大 7.9 百分点、6.5 百分点，比 YOLOv8n 模型分别增大 6.8 百分点、5.9 百分点，比 YOLOv10n 模型分别增大 3.5 百分点、2.8 百分点；与相近参数量模型相比，RSC-YOLOv10n 模型的 E_{mAP50}、$E_{mAP50-95}$ 比 YOLOv3-ting 模型分别增大 16.4 百分点、12.6 百分点，比 YOLOv6n 模型分别增大 9.0 百分点、6.8 百分点，比 YOLOv7-ting 模型分别增大 6.5 百分点、7.7 百分点；与较高参数量模型相比，RSC-YOLOv10n 模型的 N_p、N_f 约为 RT-DETR-L 模型的 17.9%、22.8%，但 E_{mAP50}、$E_{mAP50-95}$ 比 RT-DETR-L 模型分别增大 7.8 百分点、7.7 百分点。RSC-YOLOv10n 模型在较低参数量和浮点运算次数基础上，检测精度较高。

2.8　模型通用性验证

为验证 RSC-YOLOv10n 模型的鲁棒性和泛化能力，以恶劣天气关联数据集（adverse conditions dataset with correspondences，ACDC）[26-27] 为基础，新增大雪、大雨、浓雾、沙尘等多种恶劣气象条件的图片如图 5 所示，组成通用性验证数据集。

a）大雪　　b）大雨　　c）浓雾　　d）沙尘

图 5　新增恶劣气象条件图片示例

通用性验证数据集共 3 400 张图片，其中 2 660 张来自 ACDC 数据集，新增 740 张图片，约占通用性验证数据集的 21.76%。ADCD 数据集中自带自动驾驶场景分类标签，标签类型与 SODA10M 数据集不同，为使图像数据标签一致，对通用性验证数据集进行重新标注。通用性验证数据集标签包含 SODA10M 数据集中标签类型，增补 Bus（公交车）、Motorcycle（摩托车）类型。将通用性验证数据集图片数按 6∶2∶2 划分为训练集、测试集、验证集。通用性验证数据集比 SODA10M 数据集能更真实地评估 RSC-YOLOv10n 模型在实际应用中的检测效果。进行模型通用性验证试验，RSC-YOLOv10n 模型对通用性验证数据集进行道路场景目标检测的 E_{mAP50}、$E_{mAP50-95}$ 比 YOLOv10n 模型分别增大 0.9 百分点、1.3 百分点，表明 RSC-YOLOv10n 模型对恶劣气象条件下道路场景目标检测具有较高的鲁棒性和泛化能力。

2.9　模型可视化检测效果试验

采用 SODA10M 数据集和通用性验证数据集进行可视化检测效果试验，对比分析 RSG-YOLOv10n 模型和 YOLOv10n 模型的检测效果，如图 6~9 所示。

a）检测场景 1　　b）检测场景 2　　c）检测场景 3　　d）检测场景 4

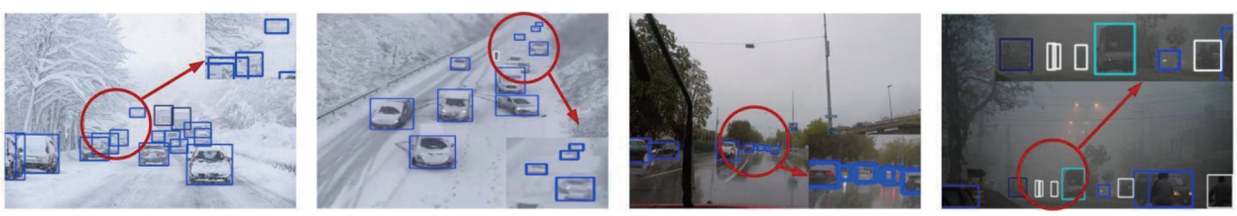
图 6　YOLOv10n 模型对 SODA10M 数据集的可视化检测效果

a）检测场景 1　　b）检测场景 2　　c）检测场景 3　　d）检测场景 4

图 7　RSC-YOLOv10n 模型对 SODA10M 数据集的可视化检测效果

a）检测场景 5 b）检测场景 6 c）检测场景 7 d）检测场景 8

图 8 YOLOv10n 模型对通用性验证数据集的可视化检测效果



a）检测场景 5 b）检测场景 6 c）检测场景 7 d）检测场景 8

图 9 RSC-YOLOv10n 模型对通用性验证数据集的可视化检测效果

由图 6、7 可知：在 SODA10M 数据集上，YOLOv10n 模型在检测远处小目标时普遍存在一定程度的漏检现象，尤其在密集遮挡、复杂光照和远近目标共存等具有挑战性的场景下表现不佳，如图 6a）、b）所示；RSG-YOLOv10n 模型表现更稳定准确，能更准确地识别远处小目标，鲁棒性和检测能力更强，如图 7a）、b）所示。在图 6b）、7b）中，夜间行驶过程中复杂的光照条件和明显的对比度变化使得目标模糊难辨，YOLOv10n 模型将巴士上的反光误判为车辆，而 RSG-YOLOv10n 模型未出现类似误检，在复杂光照下的鲁棒性更好。在图 6c）、7c）中，车辆行驶过程中存在远近目标共存的情况，远处目标的准确检测对于行车安全至关重要，RSG-YOLOv10n 模型成功识别远处行人，YOLOv10n 模型未能完全检测到此目标，证明前者在小目标检测方面的优势。

由图 8、9 可知：在通用性验证数据集中涵盖更多不同天气条件下的道路场景。相较于 SODA10M 数据集，通用性验证数据集包含大量恶劣天气，环境干扰更强且目标特征缺失更严重。在此道路环境条件下，改进后的 RSG-YOLOv10n 模型的特征捕捉能力和鲁棒性比 YOLOv10n 模型强。例如在图 8a）、b）中，雪天场景中车辆、远处小目标易被大雪覆盖而导致漏检，YOLOv10n 模型对此检测表现不佳；在图 9a）、b）中，RSG-YOLOv10n 模型成功检测出 YOLOv10n 模型漏检的目标。在图 9c）、d）中，RSG-YOLOv10n 模型表现较强的特征提取与小目标检测能力，体现良好的泛化性和应用潜力。

3 结论

针对交通道路场景中背景干扰、目标遮挡、小目标特征微弱等导致目标检测可靠性下降问题，提出 RSC-YOLOv10n 模型。

1）引入 RFAConv 模块替换 Conv 模块，通过分组卷积高效提取多尺度感受野空间特征，协同空间注意力机制，抑制背景噪声；引入 SOEP 模块，结合 CSP 结构和 OmniKernel 模块进行特征整合，显著提高小目标检测性能；引入 GCSA 模块，通过通道注意力、通道洗牌与空间注意力的耦合机制强化特征图表达，进一步优化全局上下文建模能力。

2）进行消融试验，相较于 YOLOv10n 模型，引入 RFAConv、SOEP、GCSA 3 种模块形成 RSC-YOLOv10n 模型的 P、R、E_{mAP50}、$E_{\mathrm{mAP50-95}}$ 分别增大 6.3 百分点、2.5 百分点、3.5 百分点、2.8 百分点，检测精度提高；N_{p} 增大 32.6%，N_{f} 增大 63.1%，模型计算量增大，后期对模型进行轻量化处理；N_{fps} 减小至 169 帧/s，满足自动驾驶的目标检测实时性要求。

3）与较低参数量模型（YOLOv5n、YOLOv8n、YOLOv10n）、相近参数量模型（YOLOv3-ting、YOLOv6n、

YOLOv7-ting)、较高参数量模型(RT-DETR-L)相比,RSC-YOLOv10n 模型在较低参数量和浮点运算次数基础上,检测精度最高。

4)进行模型通用性验证试验,新增沙尘、大雪、浓雾等多种恶劣气象条件的图片,组成通用性验证数据集。RSC-YOLOv10n 模型对通用性验证数据集进行道路场景目标检测的 E_{mAP50}、$E_{\mathrm{mAP50-95}}$ 比 YOLOv10n 模型分别增大 0.9 百分点、1.3 百分点,表明 RSC-YOLOv10n 模型对恶劣气象条件下道路场景目标检测具有较高的鲁棒性和泛化能力。

5)进行可视化检测效果试验,在 SODA10M 数据集中,RSC-YOLOv10n 模型在密集遮挡、复杂光照及远近目标共存等条件下比 YOLOv10n 模型表现更稳定准确;在通用性验证数据集的雪天、雨天和雾天等恶劣天气下,RSC-YOLOv10n 模型比 YOLOv10n 模型表现较强的特征提取与小目标检测能力,体现良好的泛化性和应用潜力,验证 RSC-YOLOv10n 模型在复杂道路场景下的检测性能和鲁棒性明显增强。

参考文献:

[1] 邓亚平, 李迎江. YOLO 算法及其在自动驾驶场景中目标检测综述[J]. 计算机应用, 2024, 44(6):1949-1958.

[2] 叶桐铖, 朱尤攀, 周永康, 等. 轻量级目标检测算法综述[J]. 红外技术, 2025, 47(3):289-298.

[3] 张浩晨, 张竹林, 史瑞岩, 等. 基于 YOLO-NPDL 的复杂交通场景检测方法[J]. 山东交通学院学报, 2025, 33(2):34-47.

[4] 董红召, 林少轩, 佘翊妮. 交通目标 YOLO 检测技术的研究进展[J]. 浙江大学学报(工学版), 2025, 59(2):249-260.

[5] DIWAN T, ANIRUDH G, TEMBHURNE J V. Object detection using YOLO:challenges, architectural successors, datasets and applications[J]. Multimedia Tools and Applications, 2023, 82(6):9243-9275.

[6] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN:towards real-time object detection with region proposal networks [C]//Proceedings of IEEE Transactions on Pattern Analysis and Machine Intelligence. California, USA:IEEE, 2017:1137-1149.

[7] 邵延华, 张铎, 楚红雨, 等. 基于深度学习的 YOLO 目标检测综述[J]. 电子与信息学报, 2022, 44(10):3697-3708.

[8] LIU W, ANGUELOV D, ERHAN D, et al. SSD:single shot MultiBox detector[C]//Computer Vision-ECCV 2016. Cham, Germany:Springer International Publishing, 2016:21-37.

[9] ZHU X, SU W, LU L, et al. Deformable detr:Deformable transformers for end-to-end object detectionn[EB/OL]. (2020-10-08)[2025-04-12]. https://arxiv.org/abs/2010.04159.

[10] 赵龙阳, 李天豪, 张会兵, 等. MDSD-YOLO:一种复杂道路场景目标检测方法[J]. 计算机技术与发展, 2025, 35(9):30-37.

[11] LI Y, LI J G, MENG P. Attention-YOLOV4:a real-time and high-accurate traffic sign detection algorithm[J]. Multimedia Tools and Applications, 2023, 82(5):7567-7582.

[12] XUE T, LIU Z W, LAN S W, et al. YOLO-FSE:an improved target detection algorithm for vehicles in autonomous driving [J]. IEEE Internet of Things Journal, 2025, 12(10):13922-13933.

[13] FAN Z X, LIU S L. An improved YOLOv8n algorithm and its application to autonomous driving target detection[C]// Proceedings of IEEE 6th International Conference on Power, Intelligent Computing and Systems. Shenyang, China:IEEE, 2024:1133-1139.

[14] LUO Y C, CI Y S, JIANG S X, et al. A novel lightweight real-time traffic sign detection method based on an embedded device and YOLOv8[J]. Journal of Real-Time Image Processing, 2024, 21(2):1-10.

[15] ZHANG X, LIU C, YANG D, et al. RFAConv:innovating spatial attention and standard convolutional operation[EB/OL]. (2023-04-06)[2025-04-10]. https://arxiv.org/abs/2304.03198.

[16] 高立鹏,周孟然,胡锋,等. 基于 REIW-YOLOv10n 的井下安全帽小目标检测算法[J/OL]. 煤炭科学技术. (2024-09-20)[2025-05-25]. https://link.cnki.net/urlid/11.2402.td.20240919.1902.003.

[17] 刘罡, 闫曙光, 刘钰, 等. 基于感受野增强的复杂道路场景目标检测研究[J]. 电子测量技术, 2024, 47(13):

157-166.

[18] 杨燕, 景嘉杰. 基于多尺度特征提取及域迁移优化的去雾网络[J]. 北京邮电大学学报, 2025, 48(2):46-53.

[19] CUI Y N, REN W Q, KNOLL A. Omni-kernel network for image restoration[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2024, 38(2):1426-1434.

[20] FAN Y K, ZHANG K, ZHENG B, et al. GCSA-ResNet:a deep neural network architecture for malware detection[J]. Scientific Reports, 2025, 15(1):24098.

[21] HAN J H, LIANG X W, XU H, et al. SODA10M:a large-scale 2D self/semi-supervised object detection dataset for autonomous driving[EB/OL]. (2021-06-21)[2025-05-21]. https://arxiv.org/abs/2106.11118.

[22] SUNKARA R, LUO T. No more strided convolutions orPooling:a new CNN building block forLow-resolution images andSmall objects[C]//Machine Learning and Knowledge Discovery in Databases. Cham, Germany: Springer Nature Switzerland, 2023:443-459.

[23] WOO S, PARK J, LEE J Y, et al. CBAM:convolutional block attention module[C]//Computer Vision-ECCV 2018. Cham,Germany:Springer International Publishing, 2018:3-19.

[24] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA:IEEE, 2018:7132-7141.

[25] OUYANG D, HE S, ZHAN J, et al. Efficient Multi-Scale Attention Module with Cross-Spatial Learning[EB/OL]. (2023-05-23)[2025-05-12]. https://arxiv.org/abs/2305.13563.

[26] LIANG S, WU H. Edge YOLO: Real-time intelligent object detection system based on edge-cloud cooperation in autonomous vehicles[EB/OL]. (2022-05-30)[2025-05-21]. https://arxiv.org/abs/2205.14942.

[27] SAKARIDIS C, WANG H, LI K, et al. ACDC:The adverse conditions dataset with correspondences for robust semantic driving scene perceptio[EB/OL]. (2021-04-27)[2025-05-21]. https://arxiv.org/abs/2104.13395.

Anti-interference and small object detection method for road scene based on RSG-YOLOv10n

KONG Feiyi, FU Zhenshan, WANG Yugang, FU Cong, DAI Xianxin, MA Dong*

Naval Architecture and Port Engineering College, Shandong Jiaotong University, Weihai 264209, *China*

Abstract: Addressing the issues of background interference, distant small objects, and feature loss in autonomous driving road scene object detection, an improved model based on You Only Look Once version 10n (YOLOv10n) is proposed. The receptive field attention convolution (RFAConv) module is introduced to replace the Conv module, extracting multi-scale receptive field spatial features and dynamically allocating weights through attention mechanisms to enhance the model's complex image processing capabilities. A small object enhance pyramid (SOEP) module is incorporated, utilizing an improved cross stage partial-omniKernel (CSP-OmniKernel) module for feature information integration, significantly improving small object detection performance. The global channel-spatial attention (GCSA) module is introduced to enhance feature map representation through the coupling mechanism of channel attention, channel shuffle, and spatial attention, capturing global dependencies in feature maps and enhancing feature extraction capabilities, forming the RSG-YOLOv10n complex road scene small object detection model. Ablation experiment, model performance comparison experiment, generalization validation experiment, and visualization detection effect experiment are conducted. Experimental results show that: after introducing the RFAConv, SOEP, and GCSA modules, the RSG-YOLOv10n model's precision P, recall R, mean average precision at 50 intersection over union threshold E_{mAP50}, and mean average precision averaged over intersection over union thresholds from 50 to 95 (with a step of 5) $E_{\mathrm{mAP50\text{-}95}}$ is improved by 6.3 percentage points, 2.5 percentage points, 3.5 percentage points, and 2.8

percentage points respectively compared to the YOLOv10n model, with significantly enhances detection accuracy; compared with lower parameter models (YOLOv5n, YOLOv8n, YOLOv10n), similar parameter models (YOLOv3-tiny, YOLOv6n, YOLOv7-tiny), and higher parameter model (RT-DETR-L), the RSG-YOLOv10n model achieves the highest detection accuracy with lower parameter count and floating-point operations; the RSG-YOLOv10n model performs road scene object detection on a generalization validation dataset composed of various adverse weather images including sandstorms, heavy snow, and dense fog, with E_{mAP50} and $E_{\mathrm{mAP50-95}}$ increasing by 0.9 percentage points and 1.3 percentage points respectively compared to the YOLOv10n model, demonstrating high robustness and generalization capability; the RSG-YOLOv10n model exhibits strong feature extraction and small object detection capabilities in visualization detection experiments for road scenes with dense occlusion, complex lighting, adverse weather, and coexisting near and distant objects.

Keywords: small object detectionn; background interference; feature loss; RSG-YOLOv10n; attention mechanism

(责任编辑:边文超)

(上接第 76 页)

absorption value and activation degree selected as evaluation indicators. An orthogonal experiment is conducted to explore the optimal modification process. The microstructural characteristics of ordinary mineral powder and aluminum ester modified mineral powder asphalt mastic are compared using fluorescence microscopy. Residual stability and freeze-thaw splitting strength tests are performed on the aluminum ester modified AC-20C specimens. The results show that the optimal modification conditions for aluminum ester modified mineral powder are a test temperature of 80 ℃, a mass fraction of aluminum ester mineral powder of 1.0%, and a modification time of 50 minutes. The activation degree of SBS modified AC-20C with aluminum ester modified mineral powder is 27.9 times that of ordinary mineral powder AC-20C, while the oil absorption value decreases by 58.7%. In terms of water stability performance, the residual stability of SBS modified AC-20C with aluminum ester modified mineral powder increases by 4.48% compared to ordinary mineral powder AC-20C, and the freeze-thaw splitting strength increases by 6.52%. Aluminum ester modified mineral powder can effectively improve the water stability performance against stripping of SBS modified AC-20C.

Keywords: asphalt mixture; Aluminate; limestone mineral powder; modification condition; asphalt mastic; water stability performance

(责任编辑:王惠)