基于改进 YOLOv8 的道路交通小目标车辆检测算法

火久元，苏泓瑞，武泽宇，王婷娟
（兰州交通大学电子与信息工程学院，甘肃 兰州 730000）

摘　要：针对交通道路中小目标车辆存在的识别困难、检测精度低以及误检和漏检等问题，提出一种基于 YOLOv8 算法的大内核、多尺度梯度组合的道路交通小目标车辆检测模型 RGGE-YOLOv8。首先，使用 RepLayer 模型替换 YOLOv8 网络的主干部分，引入大内核深度可分离卷积结构，拓展上下文信息，以增强模型对小目标的信息捕获能力；其次，使用 GIoU 代替原损失函数，解决 IoU 在预测框与真实框没有重叠时存在的无法优化问题；然后，引入全局注意力机制（GAM），通过减少信息丢失并增强全局交互信息来提高网络的特征表达能力；最后，引入 CSPNet 并重参化梯度组合特征金字塔，使得模型具有较大感受野和高形状偏差。实验结果表明，RGGE-YOLOv8 在 Visdrone 数据集和自有数据集上 mAP@0.5 指标分别达到 34.8% 和 94.7%，相较于原始 YOLOv8n 算法精度分别提高了 2.2 和 5.51 百分点，证明了 RGGE-YOLOv8 模型对道路小目标车辆检测的有效性。

关键词：YOLOv8；小目标检测；深度学习；多尺度特征金字塔；注意力机制
中图分类号：TP391　　**文献标志码：**A　　**DOI：**10.19678/j.issn.1000-3428.0069825

Road Traffic Small Target Vehicle Detection Algorithm Based on Improved YOLOv8

HUO Jiuyuan, SU Hongrui, WU Zeyu, WANG Tingjuan
(School of Electronic and Information Engineering, Lianzhou Jiaotong University, Lanzhou 730000, Gansu, China)

【Abstract】To address the issues of identification difficulties, low detection accuracy, misdetection, and missing detection of small target vehicles on traffic roads, this study proposes a road traffic small target vehicle detection model, RGGE-YOLOv8, based on the YOLOv8 algorithm with a large kernel and multi-scale gradient combination. First, the RepLayer model replaces the backbone of the YOLOv8 network, and depthwise separable convolution is introduced to expand the context information, thereby enhancing the ability of the model to capture information on small targets. Second, the Complete IoU loss (GIoU) replaces the original loss function to address the issue where the IoU cannot be optimized when there is no overlap. Subsequently, a Global Attention Mechanism (GAM) is introduced to improve the feature representation capability of the network by reducing information loss and enhancing global interactive information. Finally, CSPNet is incorporated, and the gradient combination feature pyramid is parameterized to ensure that the model achieves a large receptive field and high shape deviation. The experimental results indicate that the mAP@0.5 index of the improved algorithm on the Visdrone dataset and the custom dataset reaches 34.8% and 94.7%, respectively. The overall accuracy of the improved algorithm is 2.2 percentage points and 5.51 percentage points higher than that of the original YOLOv8n algorithm. These findings demonstrate the practicability of the RGGE-YOLOv8 model for small target vehicle detection on traffic roads.

【Key words】YOLOv8; small target detection; deep learning; multi-scale feature pyramid; attention mechanism

0　引言

智慧交通系统作为现代城市管理的关键组成部分，在提升交通系统效率、保障系统安全性以及推动可持续发展方面发挥着举足轻重的作用[1]。但在现实的交通环境中，特别是在一些拥挤且复杂的路段，目标车辆的尺寸较小，不具备相应的形状和纹理特征[2]，导致检测过程面临误检和漏检的难题。因此，如何准确检测复杂道路环境下的小尺度目标，实现城市智慧交通中的车辆精确识别、车流统计分析，已成为当前智慧交通领域研究的重要方向。

在传统机器学习中，检测目标的特征需要手工标注，研究者利用支持向量机（SVM）[3]提取图像特征，通过使用标注的正负样本进行训练并学习决策边界，再测试图像中分离目标和非目标。文献[4]提出方向梯度直方图（HOG）特征，通过计算图像局部

收稿日期：2024-05-08　　**修回日期：**2024-07-19
基金项目：甘肃省教育厅优秀研究生“创新之星”项目（2023CXZX-508）。
通信作者 E-mail：huojy@qq.com

区域中梯度直方图来描述图像的局部结构和纹理信息并应用于道路车辆检测任务。这些方法在简单场景下表现良好,但在处理复杂多变的真实场景时,手工设计的特征对于不同尺度、姿态和遮挡的目标可能缺乏鲁棒性,影响识别的准确性。

近年来,深度学习被广泛应用于目标检测任务。基于深度学习的目标检测算法可分为两阶段(Two-Stage)方法和一阶段(One-Stage)方法这两类。两阶段目标检测算法包括 R-CNN[5]、Fast R-CNN[6]、Faster R-CNN[7]、Mask R-CNN[8]等。两阶段检测算法虽在检测精度上表现优越,但复杂的检测流程和较慢的处理速度限制了其在动态交通环境中的应用。一阶段目标检测算法包括 SSD[9]算法、YOLO 系列算法[10-12]等,它们提供了更为简化的处理流程,特别是 YOLO 算法,作为一种端到端的深度学习模型,其能够从原始图像中直接提取特征,简化了目标检测流程,并增强了模型的直观性与易理解性,2023 年发布的 YOLOv8 算法已经达到了迄今为止的最高精度。然而,在拥堵或遮挡等复杂场景下,YOLOv8 对于小目标物体的检测性能仍存在局限性[13]。

目前,众多学者采用深度学习方法来检测小目标物体。文献[14]通过将 CBAM 注意力机制整合到 YOLOX 主干网络中,有效解决了目标检测过程中存在的鲁棒性不足等问题,这一改进使得网络能够更专注于显著性信息,但在处理全局上下文依赖性和通道空间关系方面仍存在一定的限制,该方法可能导致模型在处理小目标识别问题时表现不佳。文献[15]提出 DecIoU 边界框回归损失函数,以检测更多被遮挡的道路车辆,并使用动态 anchor box 机制来提高置信度的准确率,从而改善没有锚框时检测模型标签不准确的问题。文献[16]提出一种改进 SFP-DETR 算法,结合膨胀卷积设计单级特征金字塔结构,构造新的边界框回归损失,实现无锚框小目标检测,并在解码器的交叉注意力上加入权重约束,将全局注意力计算限制在局部范围内。文献[17]使用 Ghost Hards wish Conv 模块来简化卷积运算,并引入坐标注意力结合加权双向特征金字塔(BiFPN)[18],以增强对小目标特征的表达能力,但 BiFPN 具有较复杂的结构和较多的参数,如果训练数据不足或数据分布不均衡,可能会导致模型过拟合,降低模型在测试集上的性能。文献[19]基于 YOLOv3 算法,通过嵌入 SE 和 CBAM 注意力模块,提升了红外图像中车辆目标检测的性能,减轻了对预训练权重的依赖。但是,预训练权重基于大规模数据集训练,包含丰富的先验知识,可以加速网络收敛并提高泛化能力,在小数据集或特定场景中,预训练权重的使用尤为重要。文献[20]提出一种基于渐进式多粒度 ResNet 的车型识别网络,通过局部卷积、随机通道丢弃和渐进式多粒度训练模块,有效解决了车辆姿态和视角变化导致的识别难题。文献[21]提出一种基于改进 YOLOv5 算法的车辆目标检测模型,该模型通过集成坐标注意力机制并采用 Focal-EIoU 作为损失函数,同时引入全维度动态卷积,显著提高了高密度目标环境下车辆的检测精度和定位速度,验证了其在车辆目标检测方面的有效性和优越性。

本文提出一种融合大内核、多尺度梯度组合的道路车辆检测模型 RGGE-YOLOv8,旨在提升道路交通中小目标车辆的检测精度,解决识别过程中可能出现的漏检和误检问题。本文的主要贡献如下:

1)针对小卷积在目标检测中感受野有限、特征提取不充分等缺点,本文将原有模型(YOLOv8n)的 C2f 模块替换为 RepLayer,引入深度可分离大内核(5×5)卷积,以此增强卷积性能,扩大感受野范围,并尽可能地保留检测目标的特征信息,以提高模型对小目标的理解和识别能力。

2)在 IoU 特性的基础上引入最小外接框 GIoU[22],改进边界框回归损失函数,解决检测框和真实框没有重叠时 loss 值等于 0 的问题,优化模型性能,有效提升小目标的定位精准度。

3)针对检测模型存在的局部信息聚焦过度、全局上下文感知不足等问题,引入全局注意力机制(GAM)[23],对融合后的特征进行优化,增强小目标特征的关注度。

4)在颈部引入 CSPNet[24],优化 RepCSP 模型,对卷积层实施高效灵活的特征融合机制,并进行重参数化梯度组合设计,提升模型对小目标的检测能力。

1 道路车辆检测改进方法

2023 年发布的 YOLOv8 算法在检测精度方面达到业界前沿,其核心优势在于可扩展性高,这一特性允许 YOLOv8 集成并充分利用所有版本的 YOLO 架构,为研究者提供了灵活的模型选择和切换能力。因此,本文选择单阶段 YOLOv8 版本作为基线,在满足实时性与高精度检测的同时完成交通视频下实时车辆检测计数任务。

本文在 YOLOv8 的基础上提出大内核、多尺度梯度组合的道路交通小目标车辆检测模型 RGGE-YOLOv8,其网络结构如图 1 所示(彩色效果见《计算机工程》官网 HTML 版,下同)。在主干 Backbone 部分用 RepLayer 模块替换 C2f 模块,引入大内核深度

可分离卷积,利用大卷积核捕获上下文信息;针对车辆在运行过程中出现重叠遮挡而导致的漏检、误检等情况,引入 GIoU 损失函数,通过在损失函数中融入最小外接框信息,使模型准确评估和纠正检测框与真实框之间的偏差,有效提升检测性能;在颈部网络加入全局注意力机制 GAM 模块,对重要特征的定位信息和语义信息进行强化学习;重参化梯度组合特征金字塔 ReConext,将原始特征金字塔 FPN 替换为 CSPNet,进一步将 CSPNet 中的传统 3×3 卷积特征融合方式替换为 ResCSP。ResCSP 不仅继承了 CSPNet 部分连接结构的优势,还能高效和灵活地融合特征,提升模型对复杂场景下多尺度目标的检测能力。

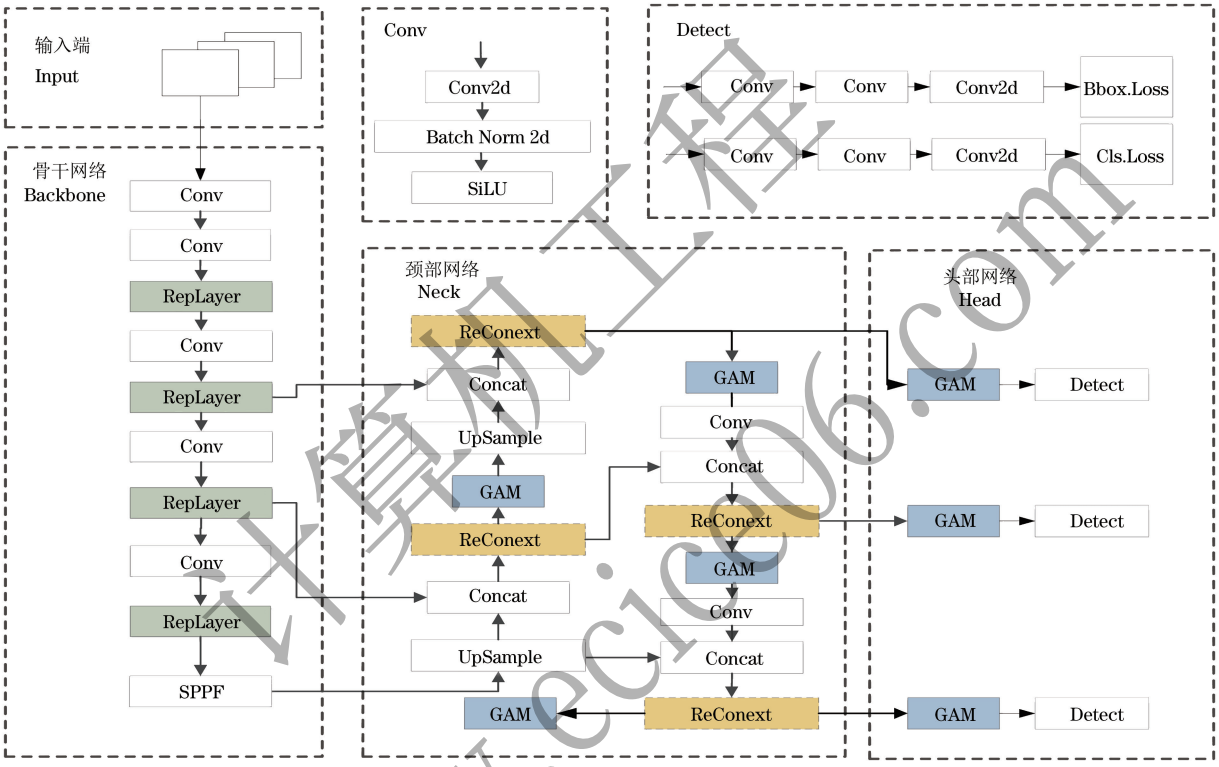


图 1 RGGE-YOLOv8 整体框架

Fig. 1 Overall framework of RGGE-YOLOv8

1.1 RepLayer

YOLOv8n 的网络结构采用 Darknet-53 作为模型骨干部分,网络结构相对较深且复杂,卷积神经网络通过多个卷积层进行特征提取时对输入图像进行下采样操作。由于小目标的像素点数量较少,下采样操作导致小目标在输出特征图上的映射区域变得非常小,甚至少于一个像素点,从而使其位置信息严重丢失,即使再通过上采样操作也不能有效地恢复缺失的位置信息。其次,Darknet-53 使用传统卷积操作,使用小卷积核对输入图像进行处理时会使感受野的大小固定。在处理小目标图像检测任务中只能看到目标周围的有限区域,缺乏足够的上下文信息。相反,大卷积核可以覆盖更大的输入区域,在更大的范围内寻找特征,捕获更多的上下文信息。通过增大卷积核大小,减少网络中的下采样次数,能够保留更多的空间信息,使得模型更好地定位小目标位置信息。

本文引入大内核深度可分离卷积(DW),在 YOLOv8n 中使用 RepLKNet[25] 模块改进骨干网络,以加强模型对深浅层特征的融合能力。RepLKNet 使用高性能大核 CNN,扩大模型的感受域,缩小形状偏差,缩小了 CNN 和 VIT 之间的性能差距,具有高效、低延迟和低计算成本等优势。模型的骨干结构可分为 10 个部分(A1~A10),其中 A1、A2、A4、A6、A8 包括一个 Conv 块、批量归一化(BN)模块以及 SiLU,针对低层次特征,使用 3×3 普通卷积提取特征。A3、A5、A7、A9 使用 RepLKNet 模块替换原本的小卷积核设计,替换为 5×5 的卷积核大小,捕捉图像中更广泛的上下文信息。应用 BN 模块和激活函数增加网络的非线性映射能力,使其能够捕获数据中的复杂模式。然后使用点卷积(1×1 卷积)融合通道的输出,增加模型深度并提高模型的表达能力。残差连接使模型成为由具有不同接受域的子模型组成的隐式集合,保留了捕捉小规模模式的能力。A10 模块对应原始 YOLOv8 模型中的 SPPF 层,SPPF 对网络上层的

特征图进行池化操作，拼接池化后的特征图进而识别各种尺度的物体，同时在特征图中保留高级语义信息。改进后的 Backbone 网络结构如图 2 所示，其中 PW 为逐点卷积。

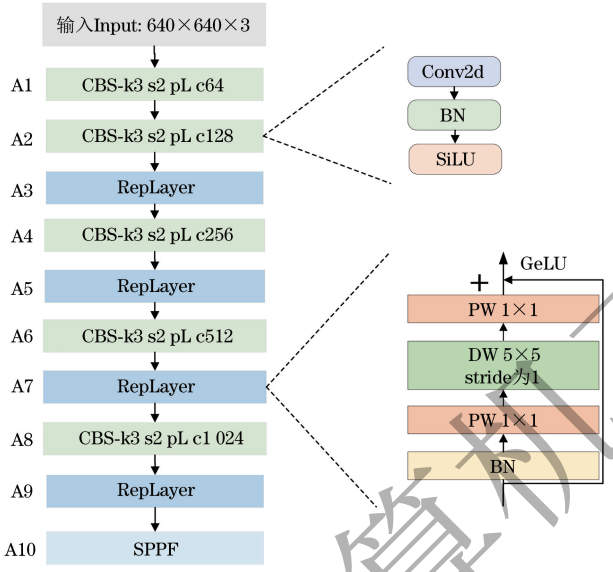


图 2 改进的 Backbone 结构

Fig. 2 Structure of the improved Backbone

1.2 损失函数

目标检测过程中一项重要的工作就是确定检测物体的边界框位置，边界框回归损失函数在目标检测中优化定位精度，帮助模型精确预测边界框的坐标，缩小预测边界框与真实边界框之间的差距，从而实现对目标的精准识别和定位。YOLOv8 的边界框回归损失函数为 CIoU，计算如式(1)～式(4)所示：

$$L_{\mathrm{CIoU}}=L_{\mathrm{IoU}}+\frac{\rho^2(b,b_{\mathrm{gt}})}{c^2}-\alpha\nu \tag{1}$$

$$\alpha=\frac{\nu}{1-L_{\mathrm{IoU}}+\nu} \tag{2}$$

$$\nu=\frac{4}{\pi^2}\left(\tan^{-1}\frac{w_{\mathrm{gt}}}{h_{\mathrm{gt}}}-r\tan^{-1}\frac{w}{h}\right)^2 \tag{3}$$

$$L_{\mathrm{IoU}}=\frac{|A\cap B|}{|A\cup B|} \tag{4}$$

式中：A 为真实框；B 是预测框；L_{IoU} 为真实框与预测框的交集与并集之比；b 为预测框中心点的横纵坐标点；b_{gt} 为真实框中心点的横纵坐标点；ρ 为预测框与真实框中心点之间的欧氏距离；w_{gt}、h_{gt} 分别为真实框的宽和高；w、h 分别为预测框的宽和高；α 为权重系数；ν 用来衡量预测框和真实框之间宽高比的一致性。

传统的 IoU 基准在检测框与真实框没有任何重叠时，$|A\cap B|=0$，$\mathrm{IoU}(A,B)=0$，在这种情况下，IoU 无法表明 2 个形状相对位置的远近，对目标物体的位置不敏感，无法直接优化没有重叠的部分，从而使得损失降为零，在实际应用中可能导致模型训练效率低下、检测精度不稳定等问题。为了解决这一问题，本文引入了最小外接框，通过计算检测框与真实框的最小外接框，从而确保即使在两者无重叠的情况下损失函数也能反映出它们之间的距离和差异。GIoU 作为模型的边界框损失函数，引入最小外接框，解决检测框和真实框没有重叠时 loss 值等于 0 的问题[22]，其公式如式(5)所示：

$$L_{\mathrm{GIoU}}=L_{\mathrm{IoU}}-\frac{|C-(A\cup B)|}{|C|} \tag{5}$$

式中：C 代表 A 和 B 最小闭合矩形的面积。如果 $|A\cup B|=|A\cap B|$，GIoU 是 IoU 的下界，在 2 个框无限重合的情况下，$L_{\mathrm{IoU}}=L_{\mathrm{GIoU}}=1$。当 2 个物体所占区域的并集 $|A\cup B|$ 与包围这 2 个区域的最小闭合区域 $|C|$ 的比值接近零时，GIoU 值趋于−1。

$$\lim_{\frac{|A\cup B|}{|C|}\to 0} L_{\mathrm{GIoU}}(A,B)=-1 \tag{6}$$

当 2 个对象完全覆盖对方时，GIoU 达到最大值 1；若 2 个对象不重叠且彼此之间的距离极大时，GIoU 达到最小值−1。与将重叠区域作为衡量标准的 IoU 不同，GIoU 通过同时考虑重叠与非重叠区域，提供了一个更为全面的相似度度量。将 IoU 替换为 GIoU，最终目标边界框 CGIoU 的公式如下：

$$L_{\mathrm{CGIoU}}=L_{\mathrm{GIoU}}+\frac{\rho^2(b,b_{\mathrm{gt}})}{c^2}-\alpha\nu \tag{7}$$

最终，CGIoU loss 的计算公式如下：

$$I_{\mathrm{CGIoU}}=1-L_{\mathrm{CGIoU}}+\frac{\rho^2(b,b_{\mathrm{gt}})}{c^2}+\alpha\nu \tag{8}$$

GIoU 对现有的 IoU 进行了推广，很好地解决了预测框与真实框没有重叠时所存在的梯度消失问题。通过确定预测边界和实际边界所形成的最小闭合矩形，衡量预测边界和实际边界在该区域内的相对占比，可以为拥挤场景下的目标检测提供更强的鲁棒性。

1.3 GAM 注意力机制

注意力机制能提升特征的表达能力、改善模型泛化性能、提高检测效率以及增强模型可解释性，因此，本文引入注意力机制以提升模型性能，常见的注意力机制有 SE[26]、CBAM[27]等。SE 通过通道权重调制特征级联方法，减弱贡献不显著的通道权重，但其在抑制次要像素效能方面存在局限。CBAM 在融合空间位置信息和通道信号强度这 2 个维度进行

了创新，但没有充分挖掘 2 种维度之间潜在的相互作用。本文引入全局注意力机制 GAM，通过减少信息丢失并提升全局交互能力，提高模型的整体性能。GAM 继承并优化了 CBAM 的架构，对通道及空间维度的子模块进行精炼，采用三维注意力模型，侧重于通道深度、空间横向宽度及纵向高度，从而提高处理效率并加强跨维度的相互作用。GAM 过程由式(9)～式(11)表示：

$$F_2=M_c(F_1)\otimes F_1=\text{Sigmoid}[\boldsymbol{K}_1\cdot\text{ReLU}(w_2y+b_2)]^{\mathrm{T}} \tag{9}$$

$$y=w_1\boldsymbol{K}_1^{\mathrm{T}}+b_1 \tag{10}$$

$$F_3=M_s(F_2)\otimes F_2=\text{Sigmoid}[\text{ConvBN}(\text{Conv ReLU}(\boldsymbol{K}_2))] \tag{11}$$

式中：F_1 为输入特征图；F_2 为通道注意力子模块的输出特征图；w_1、w_2 和 b_1、b_2 分别为多层感知机(MLP)[28]的初始权值和偏执项；M_c 为通道注意力函数；F_3 为 GAM 注意力输出的特征图；M_s 为空间注意力函数。引入全局注意力机制使模型放大全局交互信息，在关键区域分配更高的权重，捕捉图像中的关键特征，提高不同层级和尺寸之间的信息互动和学习过程。GAM 结构如图 3 所示。

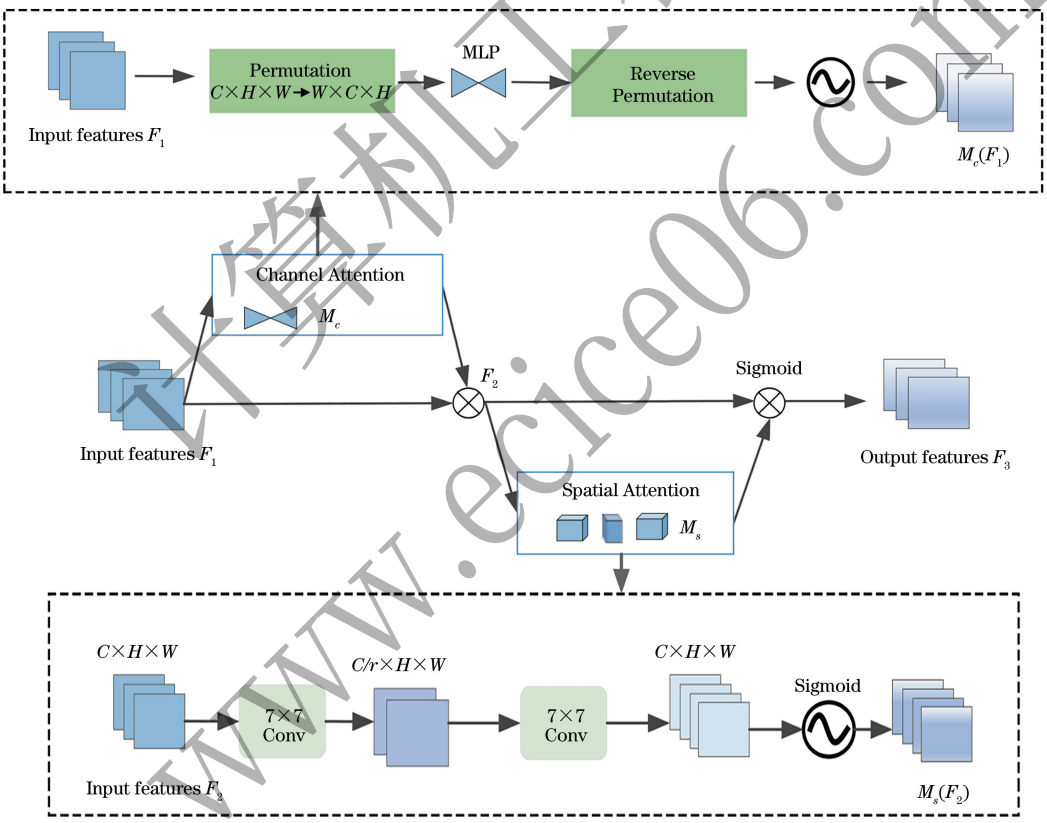


图 3 GAM 的结构

Fig.3 The structure of GAM

1.4 ReConext 网络

特征金字塔通过在不同尺度上提取图像特征，使得网络能够更全面地理解和处理不同尺寸的物体或场景。优化特征金字塔，使算法能够在检测和识别过程中同时融合细节信息和全局视角，从而提高检测结果的准确性和鲁棒性。传统的 FPN 采用横向连接以融合高层的语义信息和低层的精细信息。由于高层语义特征中可能会丢失细微的物体信息，因此传统 FPN 对于单一检测尺寸可能存在不足。PAFPN[29]通过引入更复杂的路径聚合机制来优化特征传递过程，可以自适应地选择和聚合路径，但是其增加了模型的计算复杂度，需要较多的计算资源。BiFPN 通过引入双向路径，同时考虑自上而下和自下而上的信息流动，更好地平衡了不同尺度的特征。CSPNet 采用一种特殊方法，利用两分支并行捕获不同层级的跨尺度信息，通过降低计算量来增强 CNN 的学习能力，CSPNet 在通道层面上进行分流，一部分通过特征提取单元向网络深层传递，另一部分直接在跨阶段的结构中与前者的输出相融合。这种设计增强了网络对信息的整合能力，从而提高了学习效率和特征表达的丰富性，实现了梯度组合。将 CSPNet 集成到 YOLOv8 模型后，模型的精度得到显著提升。CSPNet 通过引入部分连接结构，有

效减轻了网络参数负担，并提高了特征的利用效率。但其特征融合主要依赖于标准 3×3 卷积操作，在处理复杂多变的特征融合任务时可能不够灵活和高效。

为进一步提升模型的特征融合效果，本文对 CSPNet 模型进行优化，对卷积层进行重参数化梯度组合设计，该设计既保留 CSPNet 分支结构的优点，又通过高效灵活的特征融合机制提升模型对复杂场景下多尺度目标的检测能力。将 CSPNet 中的传统 3×3 卷积特征融合方式替换为 ResCSP，如图 4 所示，实现模型结构的优化。ResCSP 模型的 2 个分支分别构建了一个 3×3 内核和一个 1×1 内核，归一化 BN 层。其中，1×1 卷积层平行于 3×3 卷积层，并在 BN 层之后将结果相加输出。最终，小卷积核和 BN 参数被合并进大卷积核中，产生一个等效于训练模型但不依赖小卷积的结构，激活函数 Act 使用 ReLU 以提升效率。重新参数化的小内核[30]有助于优化问题处理，使模型在不同尺度上更加高效。对分支进行重参数化处理，在不增加额外计算量的同时提高模型的精度。如图 5 所示，ReConext 模块通过在不同尺度上提取图像特征，使网络更全面地理解和处理各种尺寸的场景或物体，提高复杂环境下车辆检测的精度，使模型能够准确识别和定位目标。

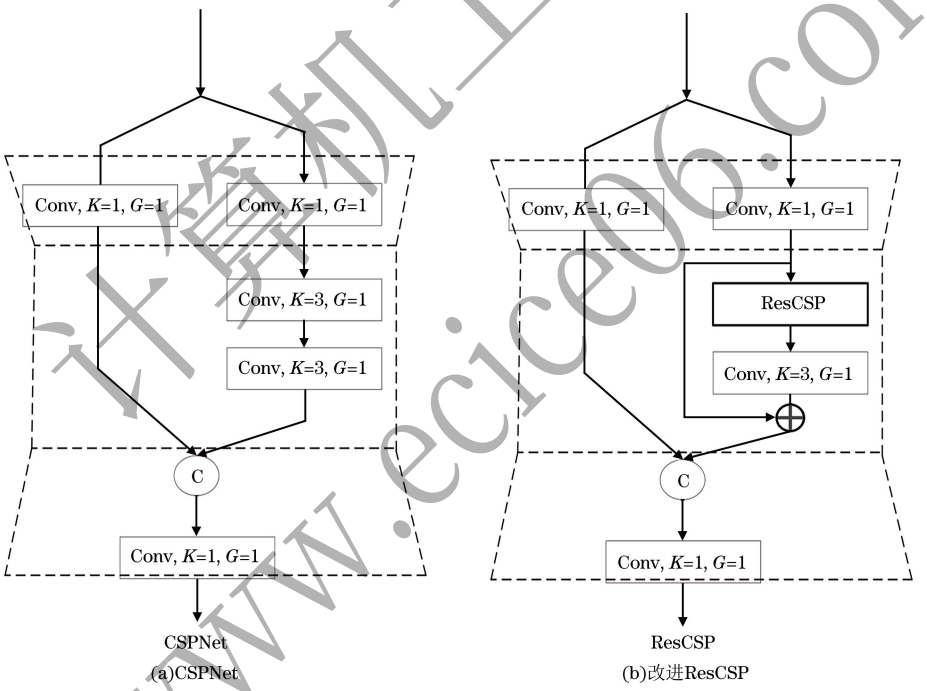


图 4 CSPNet 和改进 ResCSP 的结构

Fig. 4 The structures of CSPNet and improved ResCSP

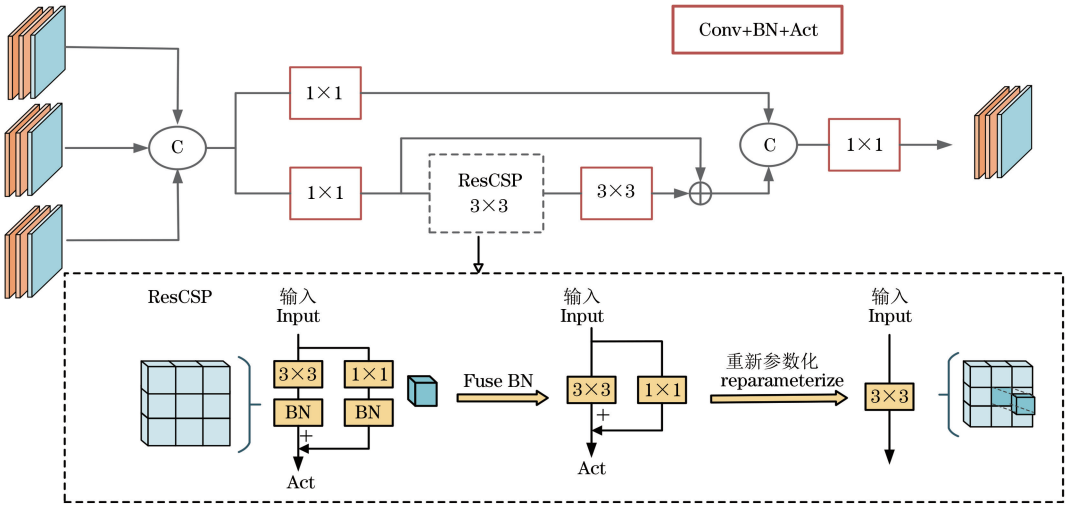


图 5 ReConext 网络模块结构

Fig. 5 ReConext network module structure

2 实验对比分析

2.1 数据集与实验参数

本文所使用的数据集来源于甘肃省兰州市的道路监控视频，由“甘肃爱城市”全天候路况直播应用提供实时的监控画面。该数据集包含来自城市街道、高速公路、公共汽车停车场等多种道路环境的图像，涵盖不同的交通道路和拍摄角度。数据集共有 2 700 个样本，集中反映不同的天气、光照条件和交通密集情况，其中包含晴朗天气 1 002 张、雾霾大雾天气 803 张、黑夜 895 张，并按照 8∶1∶1的比例划分为训练集、验证集和测试集，确保模型在各种环境条件下都能进行准确的目标检测，为实验提供多样化的训练数据。为了确保目标检测的准确性，每个图像均经过手动标注，并添加车辆大小和所在车道的边界框标签。数据集中的图像标签以特定的命名约定进行分类：数据集开头数字（1～6）代表从左至右的车道数；字母（s、m、l）根据车型进行分类，“s”代表小型车辆（如摩托车、电动车），“m”代表中型车辆（如轿车、中型货车），“l”代表大型车辆（如公交车、重型货车），该数据集共有“1s”、“1m”、“1l”、“2s”等 18 个类别。

数据集的图像分辨率为 430×430 像素，本文中的小目标具体定义为相对尺寸，依赖于摄像头的安装位置，基于目标边界框面积与图像面积的比例。同一类别中所有目标实例相对面积的中位数在 0.08%～0.58%之间定义为小目标。本文使用的数据集中目标的像素尺寸范围均符合相对尺度定义，除大型车辆 l 类（公交车、重型货车）外，其他检测目标满足绝对尺度的小目标定义。实验所采集的数据集存在大雾、强光等因素，致使数据质量较低，因此，对其进行图像增强、图像去雾等操作，使得图像清晰度与对比度提升，从而提高数据集的整体质量。图 6 为本实验预处理前后的数据集样本。

(a)预处理前

(b)预处理后

图 6 预处理前后的数据集样本

Fig.6 Dataset samples before and after preprocessing

本实验所用显卡采用英伟达的 NVIDIA GeForce RTX 3060 Laptop GPU，GPU 加速版本为 CUDA 11.3，编程语言为 Python，采用 Python 3.7 版本中的 OpenCV 库进行图像处理，深度学习框架为 PyTorch 1.12.1，在 Windows 11 操作系统上进行实验，显存为 16 GB，本文实验所用训练参数配置如表 1 所示。

表 1 实验环境配置及参数

Table 1 Experimental environment configuration and parameters

参数	设置	参数	设置
Optimizer	SGD	Batch	8
Close_mosaic	20	Workers	2
epoch	500	weight_decay	0.000 5
patience	50	lrf	0.01
images	640	warmup_momentum	0.8

2.2 实验结果分析

2.2.1 消融实验

消融实验应用控制变量的原理，在神经网络中控制一部分网络以便更好地理解网络作用。因此，本文逐个加入各个模块，验证改进模块对算法性能的影响，实验结果如表 2 所示。将 4 个模块逐个加入到原算法 YOLOv8n 中，并与原算法进行比较，得到每组实验的召回率、浮点计算数、参数量、精确率。在处理复杂场景中的目标检测任务中，改进后算法 RGGE-YOLOv8 的 mAP 比原算法 YOLOv8n 提高了 5.6 百分点，改进后的算法在所有评估阶段均表现出性能提升，每个阶段的召回率都有所增加，这表明改进策略提供了显著的性能优化空间。各模块具体提升如下：1）第一模块经更换大卷积核后，模型的召回率、mAP 和精确率分别提升了 1.1、3.1 和 3.2 百分点，这是因为相比传统卷积，大卷积核可以覆盖更大的输入区域，因此能够捕获更广阔的上下文信息，在处理一些全局结构时大卷积核通常更有优势，这验证了大卷积核在特征提取中的关键作用，并对网络性能产生积极的影响；2）添加 GIoU 损失函数后，模型的召回率、mAP 和精确率分别提升 1.0、4.3 和 2.8 百分点，能够更准确地评估目标框的预测质量，有效提升模型在复杂场景中的性能；3）在改进模型的基础上引入注意力机制，分别添加 GAM 和 CBAM 这 2 种不同的注意力机制，尽管 CBAM 注意力机制的浮点计算数和参数量相较于 GAM 注意力机制来说更低，但在全局注意力的捕获能力上稍显不足，因此，在召回率和精确率提升方面效果相对较差，GAM 能够更好地捕获全局特征信息，使模型在处理具有全局依赖关系的任务时表现出更强的性能；4）第 4 模块优化特征金字塔结构后，模型的精确率从 90.8%提升到 92.5%，梯度融

合后使得模型在检测和识别过程中能够同时考虑细节和全局信息。通过在不同尺度上提取图像特征，网络对不同尺寸的物体或场景有更全面的理解，从而提高了目标检测的准确性和鲁棒性。

表 2 消融实验结果

Table 2 Results of ablation experiment

RepLayer 大内核网络	CGIoU 损失函数	GAM 注意力模块	CBAM 注意力机制	ReConext 梯度重参网络	mAP@0.5/%	召回率/%	GFLOPs	参数量/10^6	精确率/%
—	—	—	—	—	89.1	74.8	8.2	3.0	87.3
√	—	—	—	—	92.2	75.7	16.4	7.4	90.5
√	√	—	—	—	93.4	75.8	17.8	8.2	90.1
√	√	—	√	—	87.4	63.3	18.5	9.6	89.6
√	√	√	—	—	92.8	74.0	19.5	11.8	90.8
√	√	√	—	√	94.7	77.6	20.2	14.3	92.5

综上所述，相较于原始 YOLOv8n 基线网络结构，改进后的 RGGE-YOLOv8 模型虽然参数量较高，但是 mAP@0.5、召回率以及精确率分别提升 5.6、2.8 和 5.2 百分点。实验结果充分表明，这些改进措施能够综合提高模型在复杂交通场景中的学习和泛化能力。

2.2.2 对比实验

表 3 记录了所提方法与 Faster R-CNN、YOLOv7、未改进的 YOLOv8n、在小目标检测中表现较好的模型 MAME-YOLOX[14]、YOLOX-s[15]、BiGA-YOLO[17]进行性能比较的结果。Faster R-CNN 作为两阶段算法代表，其 mAP 值高于 YOLOv7、MAME-YOLOX[14]和 YOLOX-s[15]这些一阶段算法，但是两阶段算法的参数量远高于一阶段算法，这也证实了一阶段和两阶段目标检测算法各自的优势。YOLOv7 在 YOLOv5 的基础上进行改进，采用 Darknet-19 作为基础网络架构，在速度和准确率之间取得了平衡，具有较好的实时性能。YOLOv8n 采用更先进的 Darknet-53 作为基础网络，在保持高性能的同时，通过减少计算量和参数量来提高推理速度，使得模型在处理复杂场景时能够提取更丰富的特征。在追求更高精度时，YOLOv7 的表现不如 YOLOv8n。

表 3 对比实验结果

Table 3 Comparative experimental results

模型	mAP@0.5/%	mAP@0.5:0.95/%	召回率/%	GFLOPs	参数量/10^6	帧率/(帧·s^{-1})	精确率/%
Faster R-CNN	86.7	61.5	71.2	230.8	40.8	30.5	87.5
YOLOv8n	89.1	63.9	74.8	8.2	3.0	87.2	87.3
YOLOv7	84.8	58.2	70.5	23.8	12.6	74.4	79.7
YOLOX-s[15]	91.6	65.7	76.6	27.5	16.8	69.5	87.2
MAME-YOLOX[14]	76.8	51.6	73.4	13.4	10.2	82.8	72.5
BiGA-YOLO[17]	85.1	63.1	77.9	26.2	15.5	71.4	88.5
RGGE-YOLOv8	94.7	66.7	77.6	20.2	14.3	74.1	92.5

在相同的软硬件环境和参数条件下，本文改进模型的 mAP 值达到 94.7%，帧率达到 74.1 帧/s，召回率达到 77.6%，略低于 BiGA-YOLO[17]，这是因为通过大卷积嵌入，使得模型更加关注上下文关键特征，从而在不同环境条件下表现出更高的泛化能力，导致召回率有所下降。小目标检测领域具有良好表现的轻量化模型 MAME-YOLOX[14]帧率快于改进模型 RGGE-YOLOv8。MAME-YOLOX[14]引入 CBAM，虽然结合了通道注意力和空间注意力，但它仍然是基于局部信息的注意力机制，没有像 GAM 那样全面理解图像中的全局结构与关系，其 mAP 远低于本文改进模型。YOLOX-s[15]的 mAP 值达到 91.6%，虽然能准确识别小目标物体，但其参数量大，影响了检测速度，导致帧率低于本文改进模型 4.6 帧/s。尽管改进模型 RGGE-YOLOv8 相较于 YOLOv8n 在参数量上有所提高，但在检测任务中精确率明显提高 5.2 百分点，证实了模型改进的有效性。

2.2.3 算法验证

通过对比实验结果可知 YOLOv8n 原模型是目前最先进的一阶段模型，其精确率高于其他一阶段模型，选用 YOLOv8n、RGGE-YOLOv8 模型在交通道路车辆检测画面中进行可视化对比。如图 7 所示，RGGE-YOLOv8 模型能够精准定位原模型难以识别的小尺寸目标，减少漏检和误检的情况，这种改进对于小型车辆在复杂道路场景中的准确检测具有重要的实际意义，并且强化了模型的检测精度。如图 8 所示，图 8(a)为原 YOLOv8n 模型的混淆矩阵，图 8(b)为改进 RGGE-YOLOv8 模型的混淆矩阵。在混淆矩

阵中，横轴（True）表示真实标签，纵轴（Predicted）表示模型预测的标签，对角线上的值表示正确预测的比例，非对角线上的值表示预测错误的情况，颜色越深表示比例越高。在原模型的混淆矩阵中，部分类别（如“2m”、“3m”和“4m”）在对角线上的值较低，相比之下，在改进模型的混淆矩阵中，上述类别在对角线上的值有所提升，显示出更深的颜色，这意味着预测的准确率有所增加。在非对角线上，改进模型的混淆矩阵中错误预测的值普遍比原模型更低，表明改进模型在减少误分类方面也有进步。综上所述，改进模型的混淆矩阵在多个类别上的预测准确率较原模型有所提升，且在减少误分类方面表现良好。

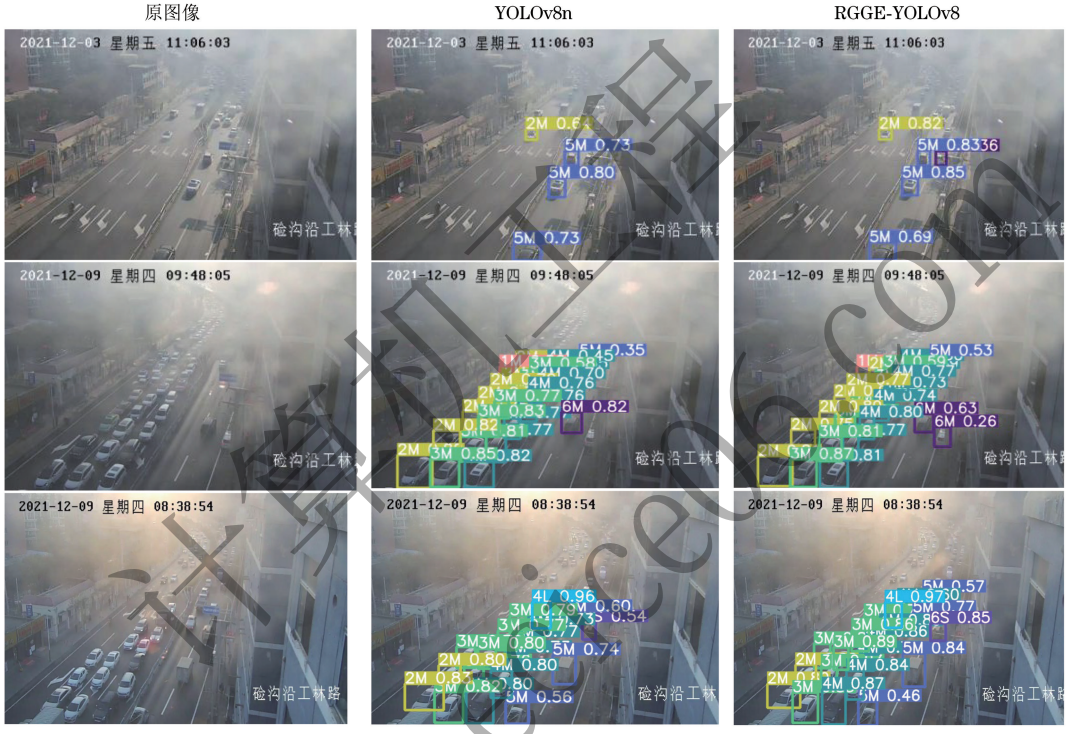
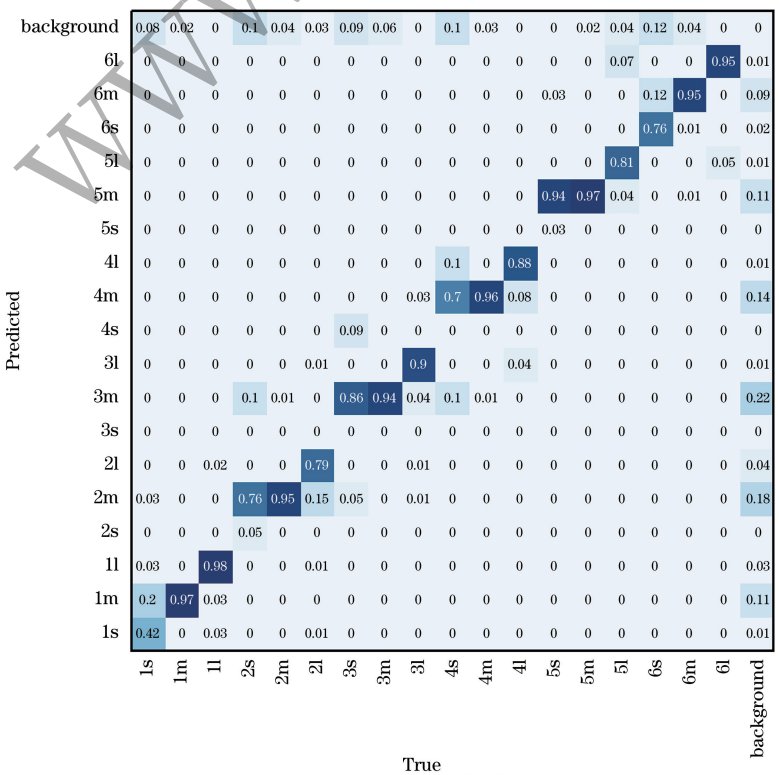
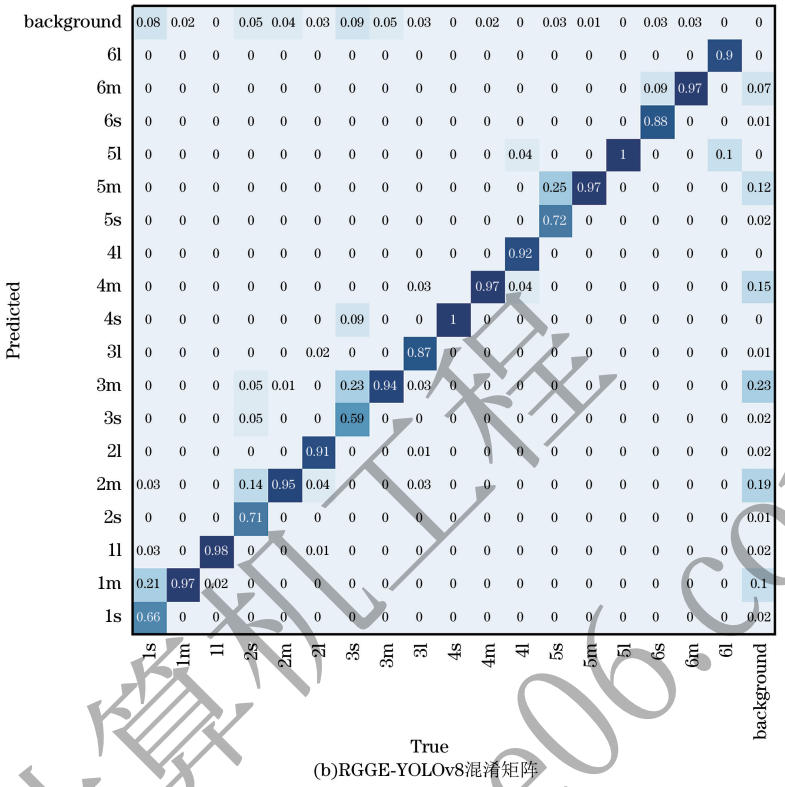


图 7 改进模型与未改进模型的对比

Fig. 7 Comparison between improved model and unimproved model



(a)YOLOv8n混淆矩阵



(b)RGGE-YOLOv8混淆矩阵

图 8　2 种模型的混淆矩阵

Fig. 8　Confusion matrix of two models

图 9 展示了在不同车道上改进模型 RGGE-YOLOv8 与原始 YOLOv8n 模型的精度对比。在该图中，横轴表示各类别的预测精度值，纵轴表示改进前后模型的对比。结果表明，在道路车辆检测任务中，尽管目标数量有限且检测难度较高，每个车道的识别精度仍然超过数据集的平均水平。特别地，改进模型不仅在大型车辆的识别上实现了稳定的检测性能，而且在检测小型车辆时检测精度显著提高。

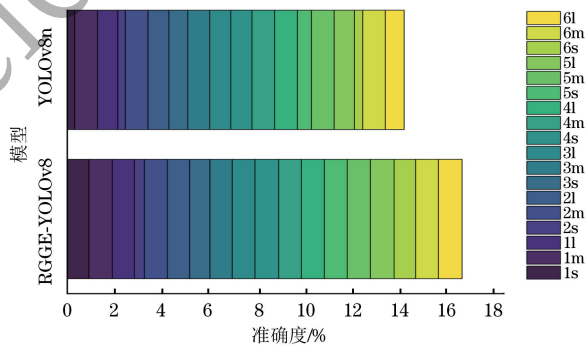
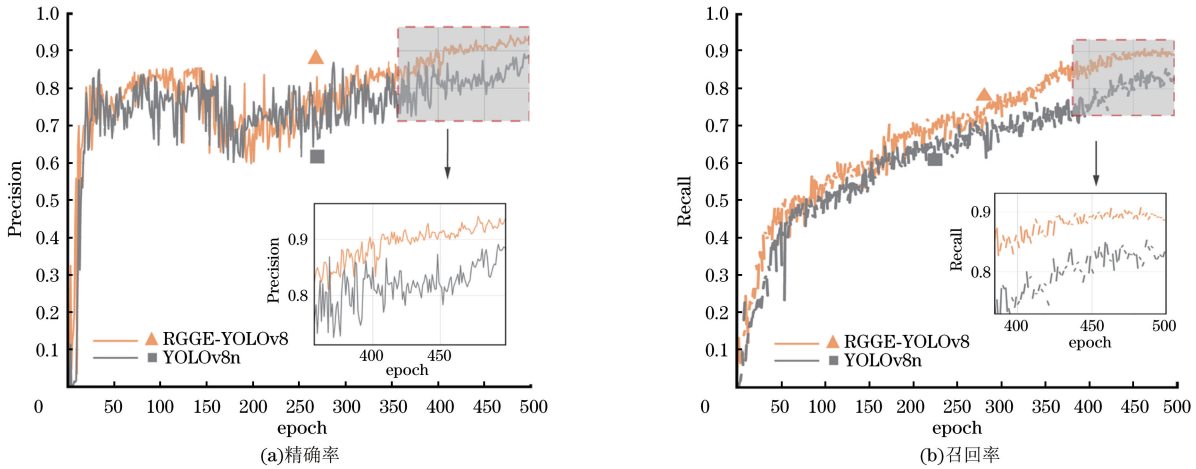


图 9　2 种模型的准确度对比堆叠图

Fig. 9　Stacked plot of accuracy comparison results between two models

图 10 展示改进前后模型在 500 轮训练后的性能对比，其中横轴表示训练轮数，纵轴分别代表预测精度和召回率。从测试集上获得的精确率和召回率



(a)精确率　(b)召回率

图 10　交通车辆数据集上的实验结果

Fig. 10　Experimental results on traffic vehicle dataset

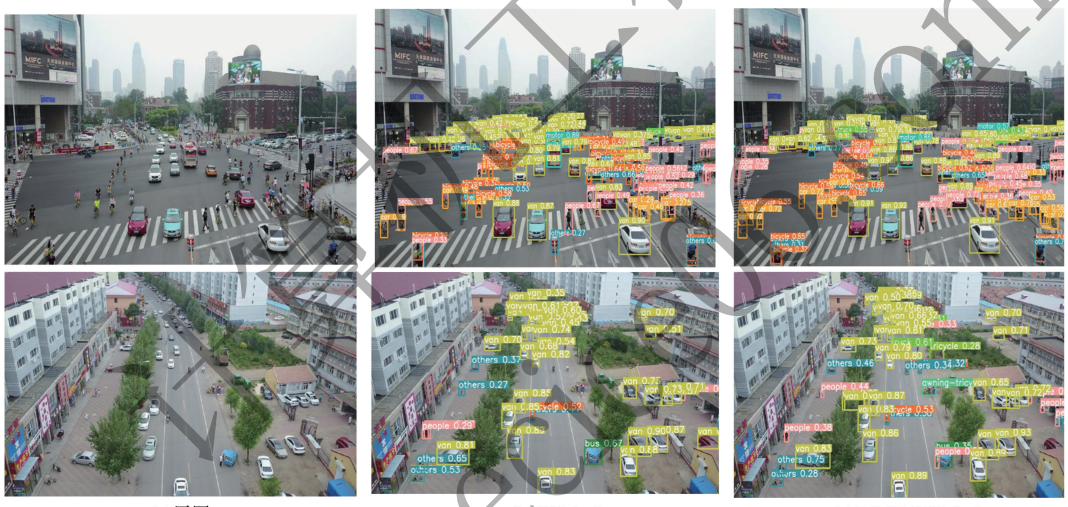
曲线可以观察到，相较于未改进的基线模型，本文提出的方法在这 2 个指标上都显示出更高的性能，从而得到更高的平均精度均值 mAP。这些结果反映出所提模型在处理复杂交通车辆数据集方面的良好性能。综上所述，通过在交通车辆数据集上训练，RGGE-YOLOv8 模型展现了优异的性能，能够提高对小目标车辆的检测能力，实验数据和视觉结果均验证了所提方法的有效性。

在验证 YOLOv8n 模型与 RGGE-YOLOv8 模型的泛化能力时，选用广泛使用的公共无人机视觉检测数据集 Visdrone。Visdrone 数据集涵盖了相隔数千公里的 14 个中国城市，包含不同的环境(城市和农村)、物体(行人、车辆、自行车等)、密度(稀疏和拥挤的场景)以及天气和光照条件，特别包含了大量遮挡和小目标的情况，是评估目标检测算法泛化能力的理想数据集。本文使用原始模型 YOLOv8n 与改进模型 RGGE-YOLOv8 进行对比实验，结果如表 4、图 11 所示。

表 4 Visdrone 数据集上的实验结果

Table 4 Experimental results on the Visdrone dataset %

模型	mAP@0.5	召回率	精确率
YOLOv8n	32.6	62.7	31.8
RGGE-YOLOv8	34.8	66.5	33.7



(a)原图 (b)YOLOv8n (c)RGGE-YOLOv8

图 11 Visdrone 数据集上改进模型与未改进模型的对比结果

Fig. 11 Comparison results of improved and unimproved models on the Visdrone dataset

具体来说，在 Visdrone 数据集上，RGGE-YOLOv8 的 mAP@0.5 达到 34.8%，相比原始 YOLOv8n 模型提升 2.2 百分点。在相同的背景下，改进模型 RGGE-YOLOv8 在遮挡小目标的识别上表现出了明显优势。通过引入新的特征提取机制和优化算法，RGGE-YOLOv8 能够更有效地捕捉目标的细节信息，并准确地区分遮挡和重叠的小目标。改进后的模型相较于原始模型召回率提升 3.8 百分点，这说明改进模型不仅提高了模型的识别准确率，还减少了误检和漏检的情况。在图 11 中观察到，改进模型相对未改进模型车辆的误检率和漏检率有所降低，这表明改进模型在复杂场景下的鲁棒性和泛化能力更强。

3 结束语

本文针对道路交通小目标车辆检测问题，提出了一种道路车辆检测模型 RGGE-YOLOv8，该模型结合大内核、多尺度梯度组合策略，旨在提高交通车辆的检测精度，并在多变的应用场景中准确且全面地提取车辆特征。该模型基于 YOLOv8n 架构，通过将 YOLOv8n 模型的 C2f 模块替换为 RepLayer，引入大内核卷积以增强模型的卷积性能，从而扩大感受野并尽可能地保留特征信息，增强模型的目标特征提取能力；通过引入最小外接框并优化边界框回归损失函数，提高目标重叠时的定位准确度；引入全局注意力机制，对融合特征进行优化，进一步提高目标特征的关注度；借鉴 CSPNet 的设计思想提出优化 ResCSP，通过重参数化设计来优化模型结构，提高特征融合的效果以及复杂环境下的车辆检测精度。本文所提 RGGE-YOLOv8 模型提高了检测的准确率与平均精度，能够满足对监控视频车辆检测准确性与实时性的要求。为了确保交通车辆在恶劣天气条件下依然能维持高检测精度，本文计划进一步对图像预处理模块进行研究，其中将聚焦于图像去雾技术的优化方面。同时，还将对算法进行轻量化处理，在保持检测精度不变的前提下有效缩减模型的参数规模，从而提升计算效率与部署能力，尤其是对于资源受限的应用场景，将使得算法部署得到有效改善。

参考文献

[1] ZOU Z, SHI Z, GUO Y, et al. Object detection in 20 years: a survey[EB/OL]. [2024-04-05]. https://arxiv.org/abs/1905.05055.

[2] 李嘉新，侯进，盛博莹，等. 基于改进 YOLOv5 的遥感小目标检测网络[J]. 计算机工程，2023，49(9)：256-264.
LI J X, HOU J, SHENG B Y, et al. Remote sensing small object detection network based on improved YOLOv5[J]. Computer Engineering, 2023, 49(9): 256-264. (in Chinese)

[3] 张华美，张皎洁. 基于人工智能的脱机手写数字识别研究综述[J]. 南京邮电大学学报(自然科学版)，2021，41(5)：83-91.
ZHANG H M, ZHANG J J. Summary of offline handwritten digit recognition research based on artificial intelligence[J]. Journal of Nanjing University of Posts and Telecommunications (Natural Science Edition), 2021, 41(5): 83-91. (in Chinese)

[4] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2022: 886-893.

[5] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[EB/OL]. [2024-04-05]. https://ieeexplore.ieee.org/document/6909475.

[6] QIU W C, YUILLE A. UnrealCV: connecting computer vision to unreal engine[EB/OL]. [2024-04-05]. https://arxiv.org/abs/1609.01326.

[7] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.

[8] HE K M, GKIOXARI G, DOLLAR P, et al. Mask R-CNN[C]// Proceedings of the IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2017: 2961-2969.

[9] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector[EB/OL]. [2024-04-05]. https://arxiv.org/abs/1512.02325.

[10] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[EB/OL]. [2024-04-05]. https://arxiv.org/abs/1506.02640.

[11] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[EB/OL]. [2024-04-05]. https://ieeexplore.ieee.org/document/8100173.

[12] ZHAO J W, TIAN G Z, QIU C, et al. Weed detection in potato fields based on improved YOLOv4: optimal speed and accuracy of weed detection in potato fields[J]. Electronics, 2022, 11(22): 3709.

[13] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. [2024-04-05]. http://arxiv.org/abs/2004.10934v1.

[14] SHEN C, MA C, GAO W. Multiple attention mechanism enhanced YOLOX for remote sensing object detection[J]. Sensors (Basel, Switzerland), 2023, 23(3): 1261.

[15] WU Y Y, YANG T, TANG Y. Research on road object detection algorithm based on improved YOLOX[C]// Proceedings of the 3rd International Conference on Neural Networks, Information and Communication Engineering. Washington D. C., USA: IEEE Press, 2023: 271-275.

[16] 张正，白佳华，田青. 基于单级特征金字塔的图像旋转目标检测[J]. 计算机工程与应用，2023，59(15)：235-242.
ZHANG Z, BAI J H, TIAN Q. Image rotating objects detection based on single level feature pyramid[J]. Computer Engineering and Applications, 2023, 59(15): 235-242. (in Chinese)

[17] LIU J, CAI Q Q, ZOU F M, et al. BiGA-YOLO: a lightweight object detection network based on YOLOv5 for autonomous driving[J]. Electronics, 2023, 12(12): 2745.

[18] CHEN J, MAI H S, LUO L B, et al. Effective feature fusion network in BiFPN for small object detection[C]// Proceedings of the IEEE International Conference on Image Processing. Washington D. C., USA: IEEE Press, 2021: 699-703.

[19] 陈皋，王卫华，林丹丹. 基于无预训练卷积神经网络的红外车辆目标检测[J]. 红外技术，2021，43(4)：342-348.
CHEN G, WANG W H, LIN D D. Infrared vehicle target detection based on convolutional neural network without pre-training[J]. Infrared Technology, 2021, 43(4): 342-348. (in Chinese)

[20] 徐胜军，荆扬，李海涛，等. 渐进式多粒度 ResNet 车型识别网络[J]. 光电工程，2023，50(7)：36-51.
XU S J, JING Y, LI H T, et al. Progressive multi-granularity ResNet vehicle recognition network[J]. Opto-Electronic Engineering, 2023, 50(7): 36-51. (in Chinese)

[21] 朱凯斌，吕红明，秦彦彬. 基于改进 YOLOv5 算法的车辆目标检测[J]. 自动化与仪表，2024，39(5)：78-83.
ZHU K B, LÜ H M, QIN Y B. Vehicle target detection based on improved YOLOv5 algorithm[J]. Automation & Instrumentation, 2024, 39(5): 78-83. (in Chinese)

[22] REZATOFIGHI H, TSOI N, GWAK J, et al. Generalized intersection over union: a metric and a loss for bounding box regression[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2019: 15-21.

[23] LIU Y C, SHAO Z R, HOFFMANN N. Global attention mechanism: retain information to enhance channel-spatial interactions[EB/OL]. [2024-04-05]. http://arxiv.org/abs/2112.05561v1.

[24] WANG C Y, LIAO H Y, WU Y H, et al. CSPNet: a new backbone that can enhance learning capability of CNN[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Washington D. C., USA: IEEE Press, 2020: 390-391.

[25] DING X H, ZHANG X Y, HAN J G, et al. Scaling up your kernels to 31 × 31: revisiting large kernel design in CNNs[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2022: 11963-11975.

[26] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2018: 7132-7141.

[27] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[EB/OL]. [2024-04-05]. https://arxiv.org/abs/1807.06521.

[28] TOLSTIKHIN I O, HOULSBY N, KOLESNIKOV A, et al. MLP-Mixer: an all-MLP architecture for vision[EB/OL]. [2024-04-05]. https://arxiv.org/abs/2105.01601.

[29] LIU S, QI L, QIN H F, et al. Path aggregation network for instance segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2018: 8759-8768.

[30] TAN M X, PANG R M, LE Q V. EfficientDet: scalable and efficient object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2020: 10781-10790.

编辑 吴云芳