文章编号:1007-2780(2026)03-0373-15

基于RT-DETR的轻量化车辆目标检测算法

张子轶[1,2],马　丽[2*],吕　帅[2],朱中宁[2]
(1. 南京信息工程大学 计算机学院,江苏 南京 210044;
2. 无锡学院 物联网工程学院,江苏 无锡 214105)

摘要:针对自动驾驶场景的硬件限制以及多尺度和遮挡现象导致检测性能不佳的问题,本文提出了一种用于车辆检测任务的轻量级目标检测算法RT-DETR-light。首先,提出利用CG Block模块改进骨干网络卷积模块,并基于此构建了轻量级特征提取网络CGResNet,实现了推理速度与检测精度的平衡。在特征融合阶段,引入双向特征金字塔网络BiFPN,通过双向信息传递实现精度的提升。最后,针对车辆目标检测任务中小目标与遮挡场景下定位精度不足的问题,设计一个改进的损失函数EPGIoU,通过多约束协同设计优化极端场景梯度稳定性。实验结果表明,本文算法在UA-DETRAC数据集上的mAP@0.5与精确率分别达到了75.0%、74.5%,相较于基线算法,参数量与计算量降低了26.4%与18.0%,检测速度提升了1.4个百分点。在BDD100K-Sub数据集上的跨数据集评估进一步验证了其泛化能力。本文提出的检测算法在检测精度、轻量化与推理速度上取得了显著优势,具备良好的泛化能力,为自动驾驶场景中实时车辆检测与边缘设备部署提供了更优的解决方案。

关　键　词:深度学习;RT-DETR算法;轻量化;车辆目标检测
中图分类号:TP391.4　**文献标识码:**A　**doi:**10.37188/CJLCD.2026-0005　**CSTR:**32172.14.CJLCD.2026-0005

Lightweight vehicle object detection algorithm based on RT-DETR

ZHANG Ziyi[1,2], MA Li[2*], LÜ Shuai[2], ZHU Zhongning[2]

(1. *School of Computer*, *Nanjing University of Information Science and Technology*, *Nanjing* 210044, *China*;
2. *School of Internet of Things Engineering*, *Wuxi University*, *Wuxi* 214105, *China*)

Abstract: To tackle degraded vehicle detection performance caused by hardware constraints, multi-scale objects, and occlusions in autonomous driving, this paper proposes RT-DETR-light, a lightweight detection algorithm. First, we design a CG Block to enhance the backbone network, forming the lightweight feature extractor CGResNet, which balances speed and accuracy. A bidirectional feature pyramid network BiFPN is then introduced for feature fusion to improve precision via bidirectional information flow. Furthermore, an enhanced loss function, EPGIoU, is proposed to improve localization accuracy for small and occluded vehicles by stabilizing gradient optimization via multi-constraint collaboration. Experiments on the UA-DETRAC dataset show a mAP@0.5 of 75.0% and a precision of 74.5%. Compared to the baseline, it reduces parameters and computation by 26.4% and 18.0%, respectively, while improving detection speed by 1.4 percentage points. Cross-dataset evaluation on

收稿日期:2026-01-12;**修订日期:**2026-02-15.
*通信联系人,E-mail:mali@cwxu.edu.cn

BDD100K-Sub confirms its strong generalization ability. The proposed algorithm offers superior accuracy, lightweight design, and inference speed, providing an effective solution for real-time vehicle detection and edge device deployment.

Key words: deep learning; RT-DETR algorithm; lightweight; vehicle object detection

1 引 言

在自动驾驶场景中,车辆作为道路交通中重要的参与方和交通研究的关键对象,往往需要进行快速准确的目标检测,以确定更好的驾驶路线,减少交通事故的发生率[1]。因此,车辆目标的检测研究对道路交通安全具有重要意义,成为当前重点研究课题[2]。得益于计算机技术的持续进步,基于深度学习的目标检测算法已成为该领域的核心技术路径。现有的深度学习目标检测算法主要包括基于候选区域的两阶段方法与基于回归的单阶段方法。在两阶段目标检测算法中,第一阶段通过区域提议网络在图像中生成多个提议框,第二阶段微调提议框。经典的算法有R-CNN(Region-based Convolutional Neural Network)[3]、Faster R-CNN[4]、Mask R-CNN[5]。尽管两阶段方法的检测精度较高,但其较慢的检测速度制约了其在实时场景中的应用。单阶段目标检测算法采用端到端的架构,无需区域提议,直接检测目标的类别与位置,经典的算法包括SSD(Single Shot MultiBox Detector)[6]、YOLO(You Only Look Once)[7]系列等。这种单阶段模型的推断速度比两阶段模型快很多,但其准确性略低。考虑到实时性要求在交通安全中具有重要意义,相关改进工作主要围绕单阶段模型展开[8]。Wang等人[9]提出的YOLOv7-E6E通过优化重参数化卷积结构、设计动态标签分配策略以及辅助头深度监督机制,实现了复杂环境下车辆的检测精度与召回率的显著提升。Li等人[10]提出面向动态小尺度行人检测场景的YOLO-FFRD,平衡优化推理效率与检测精度。

与此同时,基于Transformer的目标检测方法逐渐兴起。Carion等人[11]提出的DETR(Detection Transformer)首次将Transformer架构引入检测任务,摒弃了非极大值抑制与锚框等先验设计,将检测构建为一个集合预测问题,实现了端到端的检测。Zhu等人[12]提出了Deformable DETR,其利用可变形注意力模块,显著降低了计算开销,增强了对远距离小尺度目标的精准定位能力。Huang等人[13]提出的Swin DETR以Swin Transformer作为特征提取骨干,通过引入层级设计和移位窗口自注意力,增强骨干网络对于多尺度目标的表达能力,特征金字塔能更好地适应从近处车辆到远处车辆的尺度变化。百度团队提出的实时化改进模型RT-DETR(Real-Time Detection Transformers)[14]通过混合编码器和IoU感知查询选择机制,解耦尺度内交互和跨尺度融合两个过程,在保留Transformer全局建模能力的同时,大幅减少了计算冗余,首次使Transformer检测器能够满足自动驾驶等场景对车辆检测的实时性高要求。

当前研究虽然提升了车辆目标检测的基础性能,然而在自动驾驶的具体应用场景中,仍有一系列难题亟待解决。首先,高精度模型通常结构复杂、计算量大,非常依赖高性能图形处理器,因此很难应用到算力有限的边缘设备上。其次,待检测车辆的目标尺寸大小各异,密集目标难以精准识别。由于车辆在尺度上相差较大,导致在同一画面上存在不同尺度的多目标,使检测器对于不同目标的特征提取存在难度。同时在人流量或车流量较多时,密集的目标之间也极易出现遮挡现象,导致图像中遮挡目标的特征提取不完整,使检测器的检测性能变差[15]。为应对上述挑战,本文设计了一种面向自动驾驶场景的轻量级车辆检测模型RT-DETR-light,本文的主要贡献如下:

(1)为解决ResNet-18(Residual Network with 18 layers)骨干网络计算资源非必要消耗与多尺度目标层级特征融合不足的问题,提出新的CGResNet(Context Guided Residual Network)骨干网络,有效保持对车辆目标的检测能力并且降低原模型参数量与计算量,从而提高检测速度。

(2)在特征融合阶段引入双向特征金字塔网络(Bidirectional Feature Pyramid Network,BiFPN),在保持轻量化优势的基础上,通过多层级的特征金字塔和双向信息传递实现精度的提升。

（3）提出一种新的损失函数EPGIoU（Enhanced Precision and Generalized Inter-section over Union）进行预测框定位回归，解决小目标与遮挡场景下定位精度不足的问题。

2　轻量级RT-DETR-light算法

2.1　RT-DETR算法介绍

RT-DETR是一种基于Transformer架构的端到端实时目标检测算法，其实现了检测速度与定位精度的卓越平衡，并具备高效的实际场景部署能力，成为实时目标检测领域的一个标志性进展。RT-DETR采用创新的混合编码器，兼顾Transformer的全局建模能力与IoU（Intersection over Union）感知查询选择机制的高质量初始特征获取能力，无需后处理即可实现端到端检测，规避了非极大值抑制带来的计算开销与性能瓶颈。同时，可调解码层数设计支持推理时动态调整模型规模与性能且无需重训，为“精度-速度”权衡提供了前所未有的灵活性。为了满足边缘设备在计算能力和存储空间方面的严格要求，本文以RT-DETR系列模型中的轻量化版本RT-DETR-R18为基础网络进行优化，提出的轻量级车辆目标检测算法RT-DETR-light网络结构如图1所示。

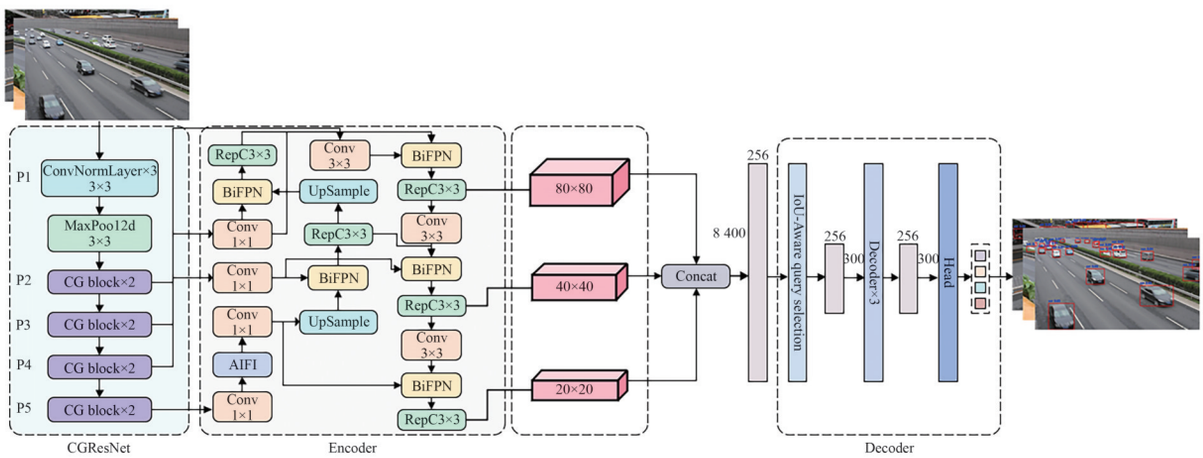


图1　RT-DETR-light模型网络结构图

Fig. 1　Network structure diagram of RT-DETR-light model

2.2　轻量化骨干网络设计

RT-DETR模型采用基于ResNet（Residual Network）[16]系列的骨干网络架构，本文使用的残差网络ResNet-18具有较浅的网络结构。现有骨干网络多采用全局一致的计算策略，造成背景区域非必要计算资源消耗。与此同时，又因固定感受野限制与层级特征融合不足，现有骨干网络难以精准聚合多尺度目标的层级特征。针对上述问题，本文设计了轻量级CGResNet网络，通过CG Block（Context Guided Block）的局部特征精炼与自适应上下文聚合双路径设计，在差异化分配计算资源以减少消耗的同时，强化了多尺度特征的捕捉与融合能力，在计算效率与检测精度上实现了权衡。

2.2.1　CG Block的双路径设计

CG Block的核心思想是在图像的不同区域与多尺度目标上同时捕获局部特征、周围上下文及全局语义信息，实现信息的有效整合与增强[17]。CG Block模块由局部特征提取器（Local Feature Extractor）、周围上下文提取器（Surrounding Context Extractor）、全局上下文提取器（Global Context Extractor）及联合特征提取器（Joint Feature Extractor）组成，其结构如图2所示。

CG Block采用“局部-上下文”双路径并行架构，通过功能互补的特征提取分支与自适应融合机制，实现局部细节、周围上下文及全局语义信息的高效聚合。其中一条路径为局部特征精炼分支，采用3×3标准卷积核构建提取器，借助受限感受野精准捕获目标边缘、纹理等细粒度特征；另一条路径为自适应上下文聚合分支，由“中程-全局”两级提取单元构成，先通过3×3膨胀卷积扩大感受野，捕获目标与周围环境的中程上下文关联，再经全局平均池化（Global Average Pooling，

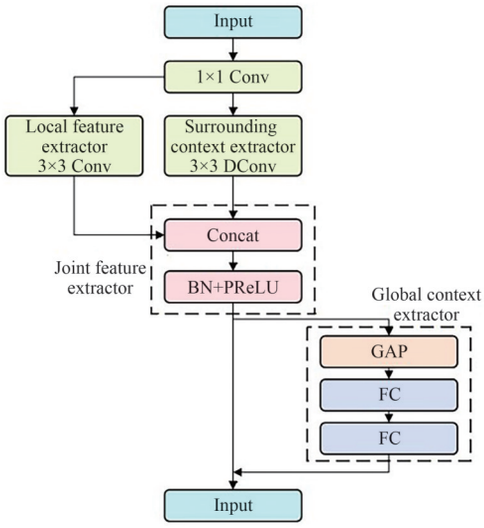


图2 CG Block基本结构图

Fig. 2 Basic structure diagram of CG Block

GAP)聚合全局信息,并通过非线性变换生成通道自适应权重向量,完成对全局结构的动态感知。

双路径设计以联合特征提取器为融合枢纽,对局部与周围上下文特征进行连接融合,同步实现自适应信息互补与维度压缩,输出兼具局部精确性与上下文语义完整性的强判别性特征表示。

CG Block的自适应上下文聚合双路径设计显著增强了多尺度特征的捕捉与融合能力:局部精炼分支精准捕捉小尺度车辆的细粒度特征,避免小目标特征被背景或大目标信息稀释;上下文聚合分支有效聚合大尺度车辆的全局语义特征,最终通过通道自适应权重实现不同尺度车辆特征的高效整合,提升多尺度车辆的表达与融合效果。

2.2.2 CG Block的轻量化与局部精炼策略

为实现模型轻量化的核心目标,CG Block在结构设计层面采用降维融合分离式计算等轻量化策略,通过简化冗余计算路径、优化特征交互机制,实现局部特征精炼。

(1)降维融合策略[18]聚焦有效计算:在联合特征提取器中,特征拼接操作会导致通道维度急剧增加。为实现高效融合,该模块采用了先降维后卷积的策略:首先通过1×1卷积大幅降低通道维度,再通过3×3卷积精准捕捉目标局部空间相关性以聚焦有效计算,如图3所示。相较于直接对拼接特征执行3×3卷积,该策略可将参数量减少K^2-1倍。

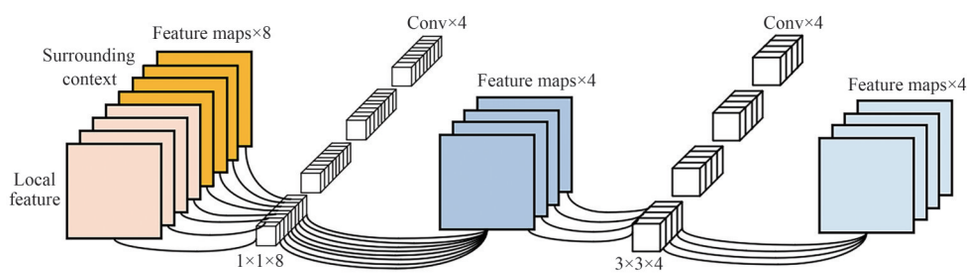


图3 降维融合策略结构图

Fig. 3 Structure diagram of the dimensionality reduction and fusion strategy

(2)深度可分离卷积[19]实现局部特征精准采样:局部特征提取器与周围上下文提取器基于深度可分离卷积构建,其通过解耦标准卷积的空间与通道维度变换,在聚焦局部特征、保持空间交互有效性的前提下,大幅降低参数量。对于卷积核大小为$K\times K$的卷积层,当输出通道数较大时,参数量可以减少近K^2倍。

(3)膨胀卷积[20]高效扩大感受野:周围上下文提取器通过膨胀的3×3卷积在保持计算效率的同时有效扩大感受野,从而高效捕获图像的中程上下文信息。

2.2.3 轻量级CGResNet网络

为精准捕捉与融合多尺度目标层级特征的同时实现模型轻量化,本文对骨干网络ResNet-18进行改进,设计了一种改进的上下文引导残差网络CGResNet,如图4所示。具体而言,选取骨干网络中语义由中级向高层逐步演化的P2~P5层级,采用融合上下文引导机制的CG Block对残差基础模块进行结构性替换。该层级选择能够在控制计算复杂度的前提下实现计算资源的精准分配,多尺度上下文感知使其能够自适应地捕捉跨层次特征,从而实现对大尺度车辆目标与小尺度车辆目标的

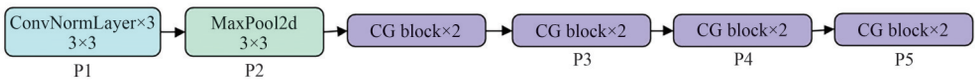


图4　CGResNet结构图

Fig. 4　Structure diagram of CGResNet model

协同精准检测。在显著降低参数量与计算量的前提下，通过重构特征提取的多尺度特征关联路径，CGResNet实现了从单一尺度特征映射到跨域上下文特征表达，既避免了浅层高分辨率特征带来的计算冗余，也消除了仅替换深层结构所造成的特征断层，从而提升特征表达的判别性与鲁棒性，为后续车辆检测任务提供了更具表征力的特征输入。

2.3　双向特征金字塔网络

为了缓解车辆多目标检测任务中目标尺度差异显著及复杂遮挡导致检测精度下降的问题，在模型的特征融合阶段引入一种高效的双向特征金字塔网络（BiFPN）[21]，核心创新包括双向特征融合以及加权融合机制。

双向特征融合机制允许特征层中的信息在自顶向下和自底向上两个方向上进行流动和融合，如图5所示。在BiFPN中，每一条双向路径被视作一个单独的特征层，这些层可以被重复多次。这种操作使得网络能够在不显著增加计算成本的前提下，有效整合跨尺度的特征信息，进一步优化检测效果。

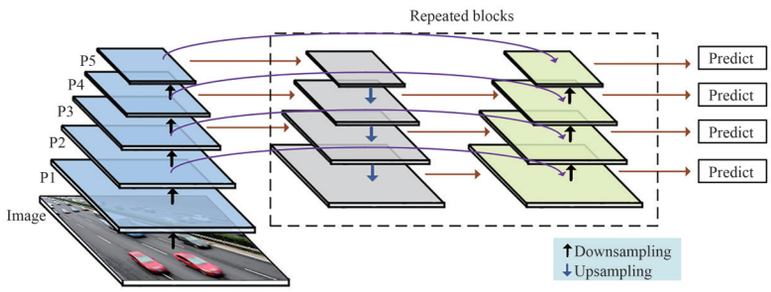


图5　BiFPN网络结构图

Fig. 5　Structure diagram of BiFPN

在实际任务中，小目标检测更依赖低层的细节特征，大目标检测则更依赖高层的语义特征[22]，不同分辨率的特征在表征质量上存在显著异质性，不加区分的融合策略会引入大量噪声，最终损害融合特征的判别性能。为了解决这个问题，加权融合机制为每个输入添加了一个权重，如式（1）所示：

$$O=\sum_i \frac{w_i}{\varepsilon+\sum_j w_j}\cdot I_i\,, \tag{1}$$

其中：w_i是第i个输入特征的权重；O是融合后的输出特征；ε是防零常数，避免分母为0导致数值不稳定；I_i是第i个输入特征；$\sum_j w_j$是所有输入特征权重的总和。通过加权融合机制，网络模型能够自适应地平衡不同尺度特征对检测结果的贡献。

在车辆多目标检测任务中，针对车辆目标尺度差异大的挑战，本文创新性地设计了“2个自顶向下＋3个自底向上”的级联融合策略，前2个BiFPN负责将高层语义信息注入中低层特征，后3个BiFPN则通过自底向上的加权融合，强化底层纹理与高层语义的关联。同时针对不同BiFPN模块设计差异化特征输入，如第1个BiFPN接收经AIFI增强的P5特征与原始P4特征，第3个BiFPN融合P2下采样、原始P3、自顶向下增强P3三路特征，精准适配多尺度车辆检测的特征需求。遮挡问题会导致目标特征不完整及语义信息模糊，BiFPN的加权融合机制通过提升未被遮挡区域的有效特征权重，抑制被遮挡区域的无效特征干扰。同时，双向路径设计促进不同层级特征间的交互，即使车辆被部分遮挡，也能基于可见部分和上下文信息辅助定位目标的完整边界框，使得模型在复杂交通场景下，稳定性与精度均得到了提升。

2.4　损失函数优化

作为实时端到端目标检测模型，检测框回归任务采用L1损失与GIoU（Generalized Intersection

over Union)损失[23]的加权组合。其中GIoU损失在传统IoU损失的基础上引入最小包围框的惩罚项,解决了预测框与真实框不重叠时梯度消失的问题,计算方式如式(2)所示:

$$\mathrm{Loss}_{\mathrm{GIoU}}=1-\mathrm{IoU}+\frac{|S_{\mathrm{C}}-S_{\mathrm{U}}|}{|S_{\mathrm{C}}|},\qquad(2)$$

其中:IoU为预测框A和真实框B的交并比,S_{C}为能够同时包含两框的最小轴对齐包围框的面积,S_{U}为两框并集面积。

GIoU损失虽然解决了IoU损失在非重叠情况下的梯度消失问题,但在两种特定场景时会退化为IoU损失:一是预测框与真实框存在包含关系;二是二者呈水平或垂直方向平行排列,如图6所示。

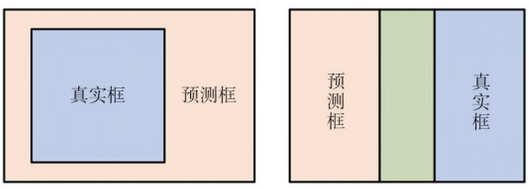


图6 GIoU退化为IoU的两种情况

Fig. 6 Two cases of GIoU degenerating to IoU

在车辆遮挡场景中,预测框与真实框往往存在大量重叠,GIoU的惩罚项失效导致优化力度不足,进而影响定位精度。同时,远距离小目标的像素占比极低,即便预测框有轻微偏移,最小包围框会因包含大量背景像素而急剧增大,使得惩罚项无法形成有效优化信号,最终导致梯度稀疏和优化停滞。

在小目标与遮挡车辆检测场景中,当前主流损失函数存在明显劣势。DIoU(Distance-IoU)[24]虽引入了中心距离惩罚,缓解了非重叠场景的梯度问题,但未对宽高比与面积差进行约束,遮挡场景下易导致预测框宽高比畸变,且小目标中心的轻微偏移难以被有效感知,最终造成定位精度下降。SIoU(Scylla-IoU)[25]聚焦于边界框的方向角度约束,但尺度适应性较差,远距离小目标的角度偏差信号微弱,难以形成有效约束。同时在遮挡场景下目标方向特征被掩盖,角度约束失效,难以保证定位稳定性。EIoU(Efficient-IoU)[26]在DIoU基础上增加了宽高比与面积惩罚,但未实现尺度自适应设计,小目标的宽高比偏差与面积偏差易被大目标信号掩盖,优化力度仍然不足。

针对上述问题,本文提出一种新的损失函数EPGIoU,通过中心距离归一化惩罚项、宽高比一致性惩罚项和面积差优化项的多约束协同设计,一方面为车辆小目标提供更强的针对性优化信号,另一方面缓解遮挡场景下的约束失效问题,同时通过归一化处理避免极端情况下的惩罚失衡,有效抑制了梯度波动,最终提升了小目标与密集车辆的定位精度。EPGIoU损失的计算方式如式(3)所示:

$$\mathrm{Loss}_{\mathrm{EPGIoU}}=1-\mathrm{IoU}+\frac{\rho^2(b,b^{\mathrm{g}})}{C^2}+\left(1-\mathrm{e}^{-\gamma\left|\frac{w}{h}-\frac{w^{\mathrm{g}}}{h^{\mathrm{g}}}\right|}\right)+\frac{\left|C_{\mathrm{area}}-\left|\mathrm{B}\cup\mathrm{B}^{\mathrm{g}}\right|\right|}{C_{\mathrm{area}}},\qquad(3)$$

其中:IoU为预测框A和真实框B的交并比;$\rho^2(b,b^{\mathrm{g}})$为两框中心欧氏距离的平方;C为预测框A和真实框B的最小包围框对角线长度;w和h是预测框A的宽高,w^{g}和h^{g}是真实框B的宽高;γ是尺度因子;C_{area}为能够同时包含预测框A和真实框B的最小轴对齐包围框的面积;$\mathrm{B}\cup\mathrm{B}^{\mathrm{g}}$为两框并集面积。

中心距离归一化惩罚项采用与目标尺度强关联的归一化设计,以最小包围框对角线长度平方为归一化因子,对两框中心欧氏距离进行归一化惩罚,针对交通场景中目标的多尺度特性实现了惩罚强度的尺度自适应。对于远距离小目标,EPGIoU的归一化机制可确保小目标的偏移信号被精准捕捉,避免定位梯度被稀释,显著提升远距离小目标车辆的定位精度。

不同于传统损失函数的线性宽高独立惩罚逻辑,宽高比一致性惩罚项创新性地采用指数平滑惩罚范式。该惩罚项通过指数函数对预测框与真实框的宽高比偏差进行平滑约束,引入尺度因子γ动态适配车辆目标的宽高比分布特性。在车辆遮挡场景中,被遮挡的部分目标轮廓易引发预测框宽高比畸变,指数平滑惩罚可温和且高效地驱动预测框向真实框贴合,避免因线性惩罚导致的梯度波动,显著降低遮挡车辆的预测框失真水平。

面积差优化项则突破了GIoU单一惩罚逻辑,以最小包围框面积为基准惩罚并集面积偏差,即便遮挡场景下预测框与训练框高度重叠,仍能提供稳定优化信号,同时避免了远距离小目标因最小包围框扩大导致的惩罚失衡。

总而言之，多约束协同设计可同时解决遮挡场景下惩罚项失效以及远距离小目标定位模糊的难题，为车辆目标检测提供更精准的回归导向。

3　实验设计

3.1　数据集与评价指标

本研究选取UA-DETRAC公开数据集[27]进行车辆检测实验，该数据集的图像分辨率为960像素×540像素，标注车辆被划分为car、van、bus和others四类。其中训练集包含60个视频序列，83 791张图片；测试集包含40个视频序列，56 340张图片。以10帧为间隔对视频序列进行均匀采样，得到约1.38万张图像。进一步筛选出1万张，按8∶1∶1的比例划分为8 000张训练集、1 000张验证集和1 000张测试集。

本研究的泛化性实验采用自行创建的BDD100K-Sub数据集[28]。BDD100K数据集是加州大学伯克利分校的人工智能实验室于2018年发布的大规模公开驾驶视频数据集，该数据集的图像分辨率为1 280像素×720像素，标注目标被划分为car、truck、bus、train、motor、bike、person、rider、traffic light和traffic sign十类。数据集BDD100K-Sub从BDD100K数据集中随机选取40 000张图像，其中训练集24 000张图像，验证集6 000张图像，测试集10 000张图像。同时，在数据处理过程中只保留车辆行人类别的标签，删除标签数量过少的类别。BDD100K-Sub数据集最终涵盖car、truck、bus、motor、bike、person和rider七类目标。

为了验证本文提出方法的性能，评估工作从均值平均精度（mean Average Precision，mAP）、精确率（Precision）、参数量（Parameters）、浮点运算次数（Giga Floating-Point Operations Per Second，GFLOPS）和帧率（Frames Per Second，FPS）5个维度展开。其中mAP@0.5表示在IoU为0.5的平均精度均值，mAP@0.5∶0.95指从IoU＝0.5到IoU＝0.95，以0.05为步长取10个不同阈值，分别计算每个阈值下的AP（Average Precision），再对这10个AP取平均值。精确率用于评估模型正类预测的准确性，具体指预测为正类的所有结果里，真实情况也为正类的比例。参数量表示模型在训练过程中的可训练变量总数。浮点运算次数表示对浮点数进行的加法、减法、乘法、除法及相关复合运算的总次数。FPS直接体现了算法的图像处理速度，其值越高代表检测效率越优。

3.2　实验设置

所有实验均在统一硬件环境下完成，核心配置如下：显卡为NVIDIAL40，其显存大小为40 GB；CPU为8核处理器；操作系统为Ubuntu22.04。本文所有实验均在Python开发环境中进行，实验环境为Python3.9、CUDA12.4，基于Pytorch2.6.0与Torchvision-0.21.0深度学习框架开展。实验环境参数设置如表1所示，本研究对所有模型采用一致的训练配置。

表1　实验环境参数设置

Tab.1　Experimental environment parameter settings

参数	设置	参数	设置
Epochs	100	Lr0	1e－4
Batch	16	Momentum	0.9
Workers	4	Weight_decay	1e－4
Images	640×640	Optimizer	AdamW
Random seed	42	Warm-up epochs	5
LR scheduler	Cosine annealing	Early stop patience	20

4　实验结果分析

4.1　消融实验

为了探究算法改进对RT-DETR模型关键指标的影响，在UA-DETRAC数据集上，针对算法的平均精度、精确率、参数量、计算量及推理速度5个维度，对改进效果进行了全面验证。现在以RT-DETR-R18模型为基础网络进行消融实验，表2展示了各实验设置与结果，其中“√”代表相应方法被纳入模型，不同模块的消融实验mAP@0.5与精确率指标变化折线图如图7所示。

由表2可知，实验RT-DETR为使用基线RT-DETR原模型的实验结果。对比实验RT-DETR与实验1可知，在使用CGResNet网络作为特征提取主干后，算法参数量和计算量分别下降了29.9%和26.1%，mAP@0.5仅下降了1.8%，精确率则提升了3.4%，FPS提升了3.4%。这一现象的本质原因是CGResNet网络的局部精炼设计聚焦目标关键细节、剔除冗余噪声，增强了目标与相似背景的区分度，有效抑制假阳性以提升

表2 消融实验结果

Tab. 2 Ablation experiment results

编号	ResNet-18	CGResNet	原编码器	BiFPN	GIoU	EP GIoU	mAP@0.5/%	Precision/%	参数量/10^6	计算量/GFLOPS	FPS
RT-DETR	√		√		√		71.6	72.6	20.09	58.3	39.8
实验1		√	√		√		69.8	**76.0**	**14.08**	**43.1**	**43.5**
实验2		√		√	√		74.2	75.9	14.79	47.8	41.2
实验3		√		√		√	**75.0**	74.5	14.79	47.8	41.2

注：粗体数字表示最优值

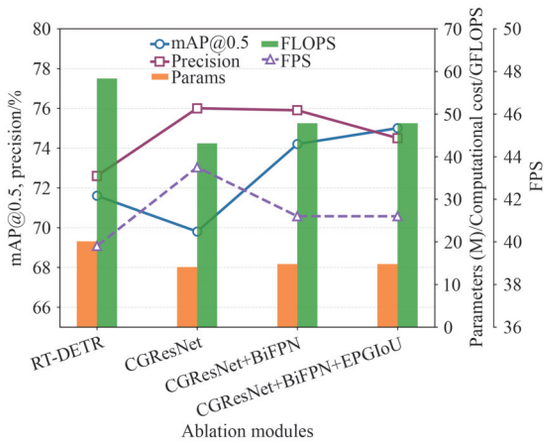


图7 不同模块的消融实验mAP@0.5与精确率指标变化折线图

Fig. 7 Line chart of mAP@0.5 and precision metrics variations in ablation experiments for different modules

精确率；与此同时上下文聚合设计可能损失了部分精细特征，如车辆边缘等定位关键信息，导致预测框偏移，最终表现为mAP@0.5指标下降。

对比实验1和实验2可知，特征融合阶段引入BiFPN模块后，在模型大小几乎不变的情况下，mAP@0.5提升了4.4%，精确率几乎持平，说明优化后的多尺度特征加权融合机制，能够显著提升模型对车辆目标的检测性能。BiFPN通过多尺度特征的加权双向融合与空间对齐，增强了模型对不同尺度目标的特征表达与召回能力，有效弥补了CGResNet精细特征信息不足的缺陷，缓解了检测框偏移问题，从而实现mAP@0.5的显著提升；同时，该模块在特征聚合过程中未引入过多背景噪声，保留了目标关键区域的判别能力，并有效抑制假阳性，因此精确率未出现明显下降。

实验2与实验3的对比结果显示，将损失函数替换为EPGIoU后，mAP@0.5提升了0.8%，精确率下降了1.4%。EPGIoU通过中心距离、宽高比一致性与面积差的多约束优化，强化了检测框的定位准确性，使正样本目标的预测框更贴合真实框，mAP@0.5指标得到提升。但EPGIoU对边界框的强约束会引导模型更充分地挖掘与目标特征相关的区域，在提升目标召回能力的同时，也使少量背景区域被判定为目标，导致假阳性预测略有增加，最终表现为精确率的小幅下降。

综上所述，在本文轻量化改进中，参数量和计算量分别降低26.4%、18.0%，mAP@0.5提升了3.4%，精确率提升了2.0%，检测速度提升了2.6个百分点，说明该轻量化策略以有限的参数量实现了对基准模型的性能超越。

为了系统分析模型改进前后的性能差异，对比了RT-DETR-R18模型改进前后在训练过程中mAP@0.5和精确率的变化情况，如图8所示。训练前期改进模型的收敛速度略慢于原始模型，其指标暂时低于原始模型，这一现象源于改进方法中自适应上下文聚合设计与双向特征金字塔需要更多训练迭代来完成参数优化。随着训练轮次的增加，RT-DETR-light模型的mAP@0.5与精确率曲线在训练后半程实现反超，最终mAP@0.5

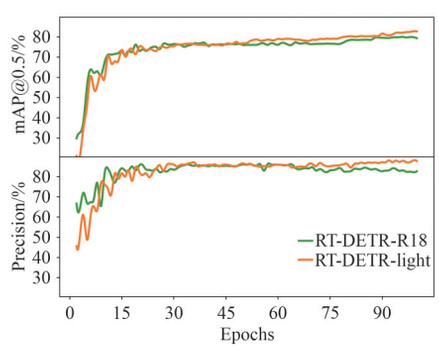


图8 mAP@0.5和精确率指标的训练过程对比曲线图

Fig. 8 Comparison curves of mAP@0.5 and precision metrics during training

指标较RT-DETR-R18模型提升了3.4个百分点，精确率指标提升了4.7个百分点，充分验证了改进设计在深度特征挖掘与目标定位精度提升上的长期有效性，实现了性能提升与检测可靠性的平衡。实验结果表明，本文改进方法在性能上得到了有效验证，其设计具有合理性与明显优势。

4.2　对比实验

Faster R-CNN作为双阶段目标检测领域的标志性算法，在泛化能力和检测精度上展现出优异的性能。YOLO系列模型是单阶段目标检测领域的代表性算法，在较短的推理时间内实现了较高的准确性。YOLOV5、YOLOV6与YOLOV8是YOLO系列算法中的典型代表，其中，YOLOV5、YOLOV8的应用范围最广、落地场景最多，而YOLOV11是YOLO系列在2024年推出的新一代目标检测模型。

表3展示了本文算法与多个主流检测模型Faster R-CNN（基于Torchvision-0.21.0实现）、YOLOV5S（v7.0）、YOLOV6S（v3.0）、YOLOV8（v8.2.0）以及YOLOV11N（v8.3.235）的对比结果，进一步验证其优势。

根据表3的结果分析得到，虽然双阶段算法Faster R-CNN的mAP@0.5和精确率较高，但其较高的参数量和计算量难以满足边缘设备在资源与效率上的严格约束。改进后的模型RT-DETR-light较Faster R-CNN的mAP@0.5仅下降0.2%，但参数量和计算量是其29%和35%，在mAP@0.5损失极小的前提下，通过轻量化主干与高效特征结构实现了模型的大幅精简。与注重实时性的YOLO系列算法相比，本文算法在mAP@0.5与精确率指标上取得了更好的性能。这是因为模型在特征提取、多尺度融合与框回归环节进行了针对性优化，提升了复杂交通场景下的目标判别与定位能力。且本文改进模型相较于原始模型，推理速度更快，不仅能够支持自动驾驶场景的实时计算需求，而且在参数量与计算量上更具优势，更适配硬件资源有限的边缘设备部署场景。在所有参与对比的模型中，本文提出的RT-DETR-light算法以75.0%的mAP@0.5领先于大部分模型，与规模相当的YOLOV8S模型相比，改进后mAP@0.5值高出2.6%，精确率高出2.7%，这得益于主干特征精炼、多尺度双向融合与EPGIoU定位损失的协同优化。与同精度级别的RT-DETR-R18模型相比，本文模型的参数量大小是其73.6%。

表3　不同算法性能对比结果

Tab. 3　Performance comparison results of different algorithms

Model	mAP@0.5/%	mAP@0.5:0.95/%	Precision/%	参数量/10^6	计算量/GFLOPS	FPS
Faster R-CNN	**75.2**	57.7	**76.2**	50.98	137.2	32.1
YOLOV5S	69.8	58.2	70.1	9.11	23.8	50.7
YOLOV6S	70.3	56.5	71.3	17.2	44.2	41.5
YOLOV8N	67.2	54.8	66.1	3.01	8.2	47.2
YOLOV8S	72.4	58.3	71.8	11.13	28.7	42.6
YOLOV11N	69.5	56.6	69.9	**2.59**	**6.4**	**54.4**
RT-DETR-R18	71.6	57.5	72.6	20.09	58.3	39.8
RT-DETR-light	75.0	**59.9**	74.5	14.79	47.8	41.2

注：粗体数字表示最优值

这充分证明了所提方法不仅能够保持较高的检测精度，还能在模型轻量化、推理效率与硬件适配性方面实现多重优化，为自动驾驶等实时性要求高且硬件资源有限的边缘计算场景提供了更加均衡和实用的解决方案。

4.3　损失函数对比实验

为了验证损失函数EPGIoU在复杂场景下的有效性，本文基于测试集分别构建了小目标子集与遮挡目标子集，对当前主流损失函数的场景适配能力进行对比分析。表4展示了EPGIoU与GIoU、DIoU、SIoU、EIoU在全目标、小目标及遮挡目标场景下的性能对比结果。在全目标测试集上使用EPGIoU后，模型的mAP@0.5较基线模型提高了0.8%，同时显著优于使用DIoU、SIoU以及EIoU的模型。在小目标子集上，由于目标尺度小、有效特征少[29]，各方法的mAP@0.5均出现

表4 IoU性能评估

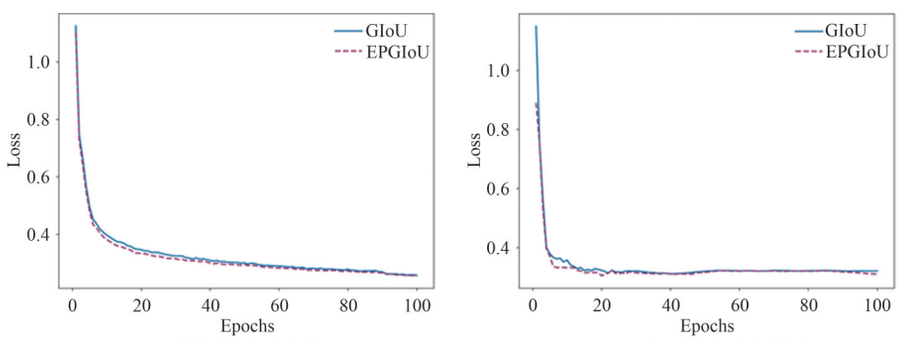
Tab. 4 IoU performance evaluation

IoU	GIoU (baseline)	EPGIoU	DIoU	SIoU	EIoU
全目标mAP@0.5/%	74.2	**75.0**	71.2	72.0	73.8
小目标mAP@0.5/%	41.8	**42.5**	40.9	41.2	41.5
遮挡目标mAP@0.5/%	17.8	**20.8**	17.2	18.4	18.8

注:粗体数字表示最优值

不同程度下降,使用EPGIoU后模型实现了42.5%的mAP@0.5,相比基线模型提升0.7%,表现出更强的小目标检测能力。在遮挡目标子集上,由于目标特征残缺、定位难度显著提升,所有方法的mAP@0.5均处于较低水平,充分体现了遮挡场景的挑战性。在遮挡场景下,EPGIoU依然取得20.8%的mAP@0.5,较基线GIoU提升3.0%,领先其余对比方法。上述结果表明,EPGIoU通过中心距离归一化、宽高比指数平滑与面积差优化的多约束协同设计,克服了遮挡场景下传统IoU损失优化力度不足及远距离小目标梯度稀疏的挑战,在提升整体检测精度的同时,显著增强了模型在小目标与遮挡目标等复杂场景下的定位鲁棒性与场景适配能力。

为了进一步对比损失函数的实际训练表现,将改进模型与基线模型的训练损失变化进行可视化,如图9所示。实验结果表明,无论是在训练阶段还是验证阶段,引入EPGIoU损失函数的模型均展现出显著的优化特性。相较于基线模型,其损失值下降速率加快,能够在更少的训练迭代轮次内进入稳定收敛区间,有效提升了模型的训练效率。同时,改进模型的最终收敛损失值更低,训练阶段可更精准地拟合任务数据分布,验证阶段则维持了良好的泛化性能,未出现过拟合现象。所以,使用EPGIoU作为RT-DETR-light的损失函数,有助于加速模型训练的收敛过程,保持泛化能力,显著提升了检测任务的综合表现。



(a)训练阶段损失变化图
(a) Training loss curves

(b) 验证阶段损失变化图
(b) Validation loss curves

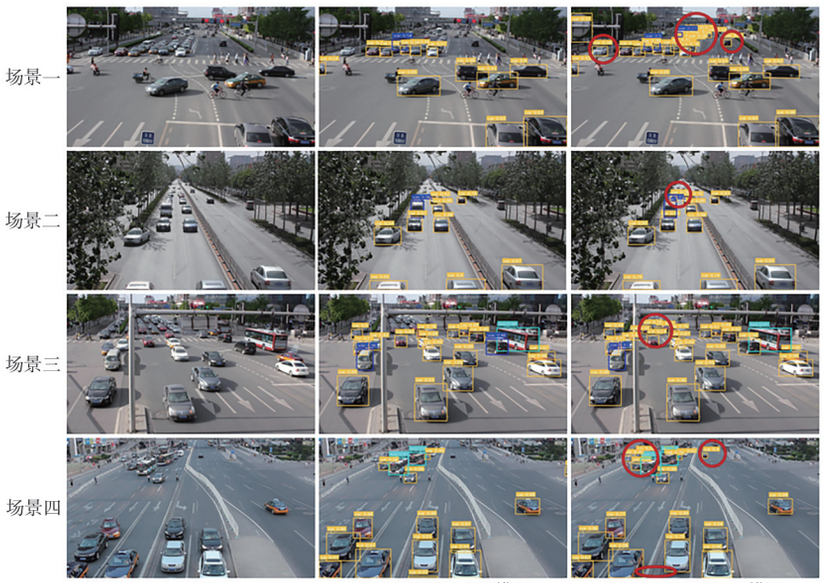
图9 GIoU与EPGIoU对比图

Fig. 9 Comparison of GIoU and EPGIoU

4.4 复杂遮挡场景可视化分析

为了进一步佐证本文改进算法在复杂遮挡场景下的有效性,在UA-DETRAC数据集中选取4张具有目标密集特性的图像样本,将改进后的模型与基线模型RT-DETR-R18进行识别可视化效果对比,结果如图10所示。其中第一列为数据集中选取的4张图片,第二、第三列分别为基线模型和改进模型RT-DETR-light的可视化结果。对比可知,在所有场景中基线模型RT-DETR-R18对遮挡车辆以及远处小目标车辆的检出能力不足,且面临误检问题,在场景四中将路面阴影等背景噪声误判为车辆。而改进后的轻量化模型RT-DETR-light在检测覆盖率和识别准确率上实现了提升,有效解决了复杂场景下的遮挡问题。

RT-DETR-light算法通过对特征提取、特征融合和损失计算3个环节的优化,既具备了遮挡



(a) 原图
(a) Original images

(b) RT-DETR-R18模型
(b) RT-DETR-R18

(c) RT-DETR-light模型
(c) RT-DETR-light

图 10　不同模型识别可视化对比图

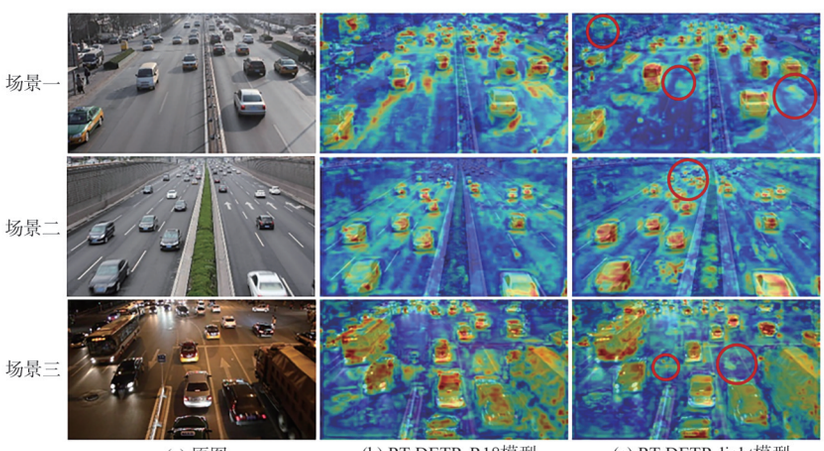
Fig. 10　Visual comparison of recognition performance among different models

场景下的高检测覆盖率，又保持了对背景噪声的抗干扰能力，最终在具体检测任务中的表现全面优于基线模型。

4.5　热力图可视化分析

为了更加直观地评估改进模型的检测效果，本文采用 Grad-CAM（Gradient-weighted Class Activation Mapping）可视化技术[30]生成注意力热力图，展示模型在图像中不同区域的关注程度，从而提升模型的可解释性。其中红色部分表示模型关注度较强的区域，蓝色部分则表示关注度较弱的区域。如图 11 所示的热力图对比揭示了不同算法在目标检测任务中的表现差异，其中第一列为在 UA-DETRAC 数据集中选取的 3 张图片，第二、第三列分别为基线模型和改进模型 RT-DETR-light 的热力图可视化结果。

由对比分析可知，RT-DETR-light 模型在多尺度目标聚焦、遮挡场景适应性以及抗干扰能力上更具优势。场景一中改进模型对于远距离小尺度车辆的注意力分布呈现出更为聚焦和清晰的边界，有效捕获了目标的结构特征，而未出现



(a) 原图
(a) Original images

(b) RT-DETR-R18模型
(b) RT-DETR-R18

(c) RT-DETR-light模型
(c) RT-DETR-light

图 11　热力图可视化对比

Fig. 11　Comparative visualization of heatmaps

明显的注意力扩散现象，且近距离大车辆与中小型车辆的注意力区域边界独立、互不干扰。而基线模型的红色高关注区域呈现明显的弥散状态，影响模型在复杂交通场景中的检测鲁棒性与定位精度。在遮挡场景二中，改进模型对重叠车辆的注意力聚焦性更优，为每个被遮挡目标提供独立且高辨识度的注意力区域，有效规避了遮挡场景下的注意力混叠问题。在场景一、场景三中，基线模型的道路空白区域存在浅红色注意力冗余区域，而改进模型的道路区域始终保持低注意力分布，展现出更优的抗干扰能力。

4.6 泛化性实验

为了进一步证明本文改进方法的适用性和鲁棒性，在BDD100K-Sub数据集上进行跨场景、多环境下的检测性能验证。将改进方法RT-DETR-light与其他主流轻量级模型YOLOV5S(v7.0)、YOLOV8（v8.2.0）、YOLOV9（v9.0.0）[31]、YOLOV11N(v8.3.235)以及基线模型进行对比，实验结果如表5所示，改进方法与主流轻量级模型对比图如图12所示。

表5 泛化性实验对比结果

Tab. 5 Comparative results of generalization experiments

Model	mAP@0.5/%	mAP@0.5:0.95/%	Precision/%	参数量/10^6	计算量/GFLOPS	FPS
YOLOV5S	52.8	30.0	57.0	9.11	23.8	51.9
YOLOV8S	53.4	31.9	57.4	11.13	28.5	49.8
YOLOV8M	55.6	33.4	59.7	25.84	78.7	36.1
YOLOV9S	55.1	33.2	59.6	7.17	26.7	51.5
YOLOV9M	56.5	33.5	60.9	20.01	76.5	38.3
YOLOV11N	47.8	28.3	60.4	**2.59**	**6.4**	**57.8**
RT-DETR-R18	57.5	33.9	61.2	20.09	58.3	38.1
RT-DETR-light	**57.8**	**34.0**	**61.2**	14.79	47.8	40.7

注：粗体数字表示最优值

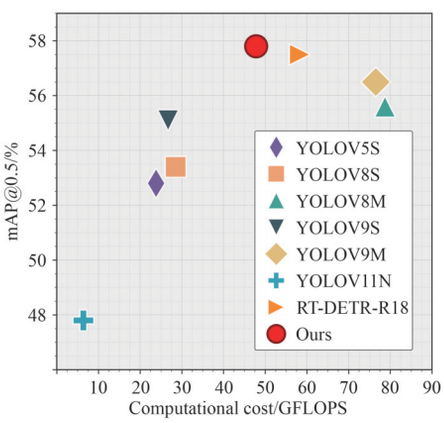


图12 改进方法与主流轻量级模型对比图

Fig. 12 Visual comparison diagram of improved method against mainstream lightweight models

根据表5的结果分析得到，在YOLO系列算法中，YOLOV9M以56.5%的mAP@0.5、60.9%的精确度成为YOLO系列模型中检测精度表现最优的模型，但其计算与存储开销较高，推理速度因此受限，FPS仅为38.3。YOLOV11N则呈现轻量化优势，FPS达到57.8，但检测精度显著低于其他模型。其余YOLO模型则呈现“精度越高、参数量与计算量越大、推理速度越慢”的权衡关系。在所有参与对比的模型中，本文提出的RT-DETR-light模型展现出了较高的mAP@0.5，达到了57.8%；检测精确率达到了61.2%。与同量级的YOLOV8S模型相比，指标mAP@0.5高出4.4%，精确率高出3.8%。相较于基线模型RT-DETR-R18，模型RT-DETR-light在该数据集上仍保持了优异的检测性能，其mAP@0.5与精确率指标未出现明显衰减，即便面对数据集特有的复杂环境干扰与场景差异仍能精准识别车辆目标。

图12直观展示了不同模型计算量与mAP@0.5的关系。在所有参与对比的模型中，YOLO系列模型最高mAP@0.5仅局限在56.5%，轻量级YOLOV11N模型虽计算成本极低，但mAP@0.5也相应大幅下降至47.8%。相比之下，RT-DETR-R18模型以中等计算成本实现了较高的mAP@0.5，尤其突出的是，本文提出的改进模型在计算量

仅为47.8 GFLOPS的情况下，达到了57.8%的mAP@0.5，不仅优于同计算区间的YOLO模型，也以更低的计算成本略超RT-DETR-R18的mAP@0.5。

这充分证明改进后的模型并非仅在特定数据集上过拟合，而是具备了较强的环境适应性与跨场景泛化能力，能够稳健应对真实驾驶场景中的多样化挑战，为其实际部署应用提供了有力支撑。

5　结　　论

本文针对自动驾驶场景的硬件限制以及多尺度和遮挡现象导致检测性能不佳的问题，设计了一种轻量级车辆目标检测算法RT-DETR-light。首先，提出了一种新的骨干网络CGResNet，通过轻量化设计在保持检测精度的同时显著降低了模型复杂度，从而实现了检测速度的优化。其次，在模型的特征融合阶段引入双向特征金字塔网络BiFPN，缓解了车辆多目标检测任务中目标尺度差异显著及复杂遮挡导致的检测精度下降问题。最后，提出了一种改进的损失函数EPGIoU，提升了小目标与密集车辆的定位精准度。

实验结果表明，模型RT-DETR-light在UA-DETRAC数据集和BDD100K-Sub数据集上，mAP@0.5分别达到了75.0%和57.8%，参数量减少了26.4%，计算量减少了18.0%。与其他模型相比，本文模型在实际边缘设备上的部署可行性更高，能够支持自动驾驶系统所需的实时车辆感知，且能有效缓解多尺度和遮挡现象导致的漏检问题。

参　考　文　献：

[1] 王龙春，方巍，张丽娟，等. 基于YOLOv8s改进的自动驾驶目标检测[J]. 液晶与显示，2025，40(5)：773-784.
WANG L C, FANG W, ZHANG L J, *et al*. Improved autonomous driving object detection based on YOLOv8s [J]. *Chinese Journal of Liquid Crystals and Displays*, 2025, 40(5): 773-784. (in Chinese)

[2] 杨晓寒，王峻，段中兴，等. 基于改进YOLOx的弱光照环境车辆检测方法[J]. 液晶与显示，2024，39(6)：801-812.
YANG X H, WANG J, DUAN Z X, *et al*. Improved YOLOx-based vehicle detection method for low light environment [J]. *Chinese Journal of Liquid Crystals and Displays*, 2024, 39(6): 801-812. (in Chinese)

[3] GIRSHICK R, DONAHUE J, DARRELL T, *et al*. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//*Proceedings of* 2014 *IEEE Conference on Computer Vision and Pattern Recognition*. Columbus: IEEE, 2014: 580-587.

[4] 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 the* 29*th International Conference on Neural Information Processing Systems*. Montreal: MIT Press, 2015: 91-99.

[5] HE K M, GKIOXARI G, DOLLÁR P, *et al*. Mask R-CNN [C]//*Proceedings of* 2017 *IEEE International Conference on Computer Vision*. Venice: IEEE, 2017: 2980-2988.

[6] LIU W, ANGUELOV D, ERHAN D, *et al*. SSD: single shot MultiBox detector [C]//*Proceedings of the* 14*th European Conference on Computer Vision*. Amsterdam: Springer, 2016: 21-37.

[7] REDMON J, DIVVALA S, GIRSHICK R, *et al*. You only look once: unified, real-time object detection [C]//*Proceedings of* 2016 *IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE, 2016: 779-788.

[8] 伍锡如，郝家琦，赵一波，等. 复杂天气条件下基于YOLO-CGT的自动驾驶车辆检测[J]. 光学 精密工程，2025，33(19)：3135-3149.
WU X R, HAO J Q, ZHAO Y B, *et al*. Autonomous vehicle detection in complex weather based on YOLO-CGT [J]. *Optics and Precision Engineering*, 2025, 33(19): 3135-3149. (in Chinese)

[9] WANG C Y, BOCHKOVSKIY A, LIAO H Y M, *et al*. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors [C]//*Proceedings of* 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver: IEEE, 2023: 7464-7475.

[10] LI S Q, WANG R, WANG S Y, *et al*. YOLO-FFRD: dynamic small-scale pedestrian detection algorithm based

on feature fusion and rediffusion structure [J]. *Sensors*, 2025, 25(16): 5106.

[11] CARION N, MASSA F, SYNNAEVE G, *et al*. End-to-end object detection with transformers [C]//*Proceedings of the* 16*th European Conference on Computer Vision*. Glasgow: Springer, 2020: 213-229.

[12] ZHU X Z, SU W J, LU L W, *et al*. Deformable DETR: deformable transformers for end-to-end object detection [C]//*Proceedings of the* 9*th International Conference on Learning Representations*. Online: OpenReview.net, 2021.

[13] HUANG Y F, PAN K X, WEI K W, *et al*. Swin Detr: Pedestrian and vehicle detection under low light conditions based on Swin Detr [C]//*Proceedings of the* 2024 3*rd International Conference on Artificial Intelligence and Intelligent Information Processing*. Tianjin: ACM, 2024: 292-297.

[14] ZHAO Y A, LV W Y, XU S L, *et al*. DETRs beat YOLOs on real-time object detection [C]//*Proceedings of* 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle: IEEE, 2024: 16965-16974.

[15] 于承峰,高松,王鹏伟,等. 改进的YOLOv11智能车辆动态环境目标检测算法[J]. 电子测量技术,2025,48(21): 55-66.

YU C Y, GAO S, WANG P W, *et al*. Improved YOLOv11 intelligent vehicle dynamic environment object detection algorithm [J]. *Electronic Measurement Technology*, 2025, 48(21): 55-66. (in Chinese)

[16] HE K M, ZHANG X Y, REN S Q, *et al*. Deep residual learning for image recognition [C]//*Proceedings of* 2016 *IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE, 2016: 770-778.

[17] WU T Y, TANG S, ZHANG R, *et al*. CGNet: a light-weight context guided network for semantic segmentation [J]. *IEEE Transactions on Image Processing*, 2021, 30: 1169-1179.

[18] SZEGEDY C, LIU W, JIA Y Q, *et al*. Going deeper with convolutions [C]//*Proceedings of* 2015 *IEEE Conference on Computer Vision and Pattern Recognition*. Boston: IEEE, 2015: 1-9.

[19] HOWARD A G, ZHU M L, CHEN B, *et al*. MobileNets: efficient convolutional neural networks for mobile vision applications [J/OL]. *arXiv*, 2017: 1704.04861.

[20] YU F, KOLTUN V. Multi-scale context aggregation by dilated convolutions [J/OL]. *arXiv*, 2016: 1511.07122.

[21] TAN M X, PANG R M, LE Q V. EfficientDet: scalable and efficient object detection [C]//*Proceedings of* 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle: IEEE, 2020: 10778-10787.

[22] HUANG Y X, LIU H L, SHUAI H H, *et al*. DQ-DETR: DETR with dynamic query for tiny object detection [C]//*Proceedings of the* 18*th European Conference on Computer Vision*. Milan: Springer, 2024: 290-305.

[23] REZATOFIGHI H, TSOI N, GWAK J Y, *et al*. Generalized intersection over union: a metric and a loss for bounding box regression [C]//*Proceedings of* 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach: IEEE, 2019: 658-666.

[24] ZHENG Z H, WANG P, LIU W, *et al*. Distance-IoU loss: faster and better learning for bounding box regression [C]//*Proceedings of the* 34*th AAAI Conference on Artificial Intelligence*. New York: AAAI, 2020: 12993-13000.

[25] GEVORGYAN Z. SIoU loss: more powerful learning for bounding box regression [J/OL]. *arXiv*, 2022: 2205.12740.

[26] 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.

[27] WEN L Y, DU D W, CAI Z W, *et al*. UA-DETRAC: a new benchmark and protocol for multi-object detection and tracking [J]. *Computer Vision and Image Understanding*, 2020, 193: 102907.

[28] YU F, CHEN H F, WANG X, *et al*. BDD100K: a diverse driving dataset for heterogeneous multitask learning [C]//*Proceedings of* 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle: IEEE, 2020: 2633-2642.

[29] 王伯霄,宋延嵩,董小娜. 难点注意力感知红外小目标检测网络[J]. 中国光学(中英文),2024,17(3):538-547.

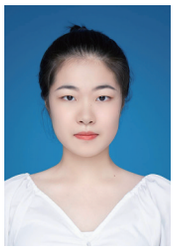
WANG B X, SONG Y S, DONG X N. Indistinguishable points attention-aware network for infrared small object detection [J]. *Chinese Optics*, 2024, 17(3): 538-547. (in Chinese)

[30] SELVARAJU R R, COGSWELL M, DAS A, *et al*. Grad-CAM: visual explanations from deep networks via gradient-based localization [C]//*Proceedings of* 2017 *IEEE International Conference on Computer Vision*. Venice:

IEEE, 2017: 618-626.

[31] WANG C Y, YEH I H, LIAO H Y M. YOLOv9: learning what you want to learn using programmable gradient information [C]//*Proceedings of the 18th European Conference on Computer Vision*. Milan: Springer, 2024: 1-21.

作者简介：



张子轶，女，硕士研究生，2024年于南京工业大学获得学士学位，主要从事交通多目标检测和图像处理方面的研究。E-mail：ziyi_5436@163.com

马　丽，女，博士，教授，2021年于吉林农业大学获得博士学位，主要从事图像处理与深度学习方面的研究。E-mail：mali@cwxu.edu.cn