

基于迁移学习与动态权重局部金字塔注意力机制的无人机航拍车辆目标检测方法

周永倩^{1,2},任燕杰^{1,2},李志刚^{1,2*},曾强云³,沙特·吾斯曼³

(1 石河子大学信息科学与技术学院,新疆 石河子 832003;2 计算智能与网络信息安全兵团重点实验室,新疆 石河子 832003;3 新疆商贸物流(集团)有限公司,新疆 乌鲁木齐 830000)

摘要:我国汽车保有量持续增长导致交通拥堵、违停和事故频发,严重制约社会经济发展。为缓解传统路面监控摄像头因监控盲区造成的管理难题,本文基于无人机航拍图像,提出了一种高效的车辆目标检测方法。针对图像中存在的遮挡、模糊、畸变和复杂场景等问题,设计了一种融合迁移学习与动态权重局部金字塔注意力(Dynamic Weight Local Pyramid Attention, DWLPA)机制的改进YOLOv8-s方法。通过引入迁移学习,缓解数据样本有限和分布不均的问题,提升模型对未见样本的泛化能力。设计DWLPA模块,结合动态权重分配与多尺度特征优化,提高模型在复杂环境下的目标检测能力。本文基于公开无人机航拍车辆图像数据开展实验,结果表明,改进模型在mAP@0.5:0.95、mAP@0.5和mAP@0.75上分别达到67.6%、98.3%和81.4%,相较YOLOv8-s基线模型和YOLOv13-s、YOLOv12-s、YOLOv11-s等轻量级模型均有明显提升,在复杂场景下展现出更强的检测精度和鲁棒性。

关键词:无人机影像;车辆检测;YOLOv8-s;迁移学习;注意力机制

中图分类号:TP391.41;TP751

文献标志码:A

Vehicle detection in UAV imagery with transfer learning and Dynamic Weight Local Pyramid Attention mechanism

ZHOU Yongqian^{1,2},REN Yanjie^{1,2},LI Zhigang^{1,2*},ZENG Qiangyun³,WUSIMAN·Shate³

(1 College of Information Science and Technology, Shihezi University, Shihezi, Xinjiang 832003, China; 2 Xinjiang Production and Construction Corps Key Laboratory of Computing Intelligence and Network Information Security, Shihezi, Xinjiang 832003, China; 3 Xinjiang Commercial Logistics (Group) Co., Ltd., Urumqi, Xinjiang 830000, China)

Abstract: With the continuous growth of automobile ownership in China, issues such as traffic congestion, illegal parking, and traffic accidents have become increasingly severe, constituting significant factors constraining social and economic development. To address management challenges from blind spots in traditional roadside surveillance cameras, this study proposes an efficient vehicle detection method using UAV aerial imagery. For occlusion, blurring, distortion, and complex backgrounds in UAV aerial imagery, we design an improved YOLOv8-s approach integrating transfer learning with Dynamic Weight Local Pyramid Attention (DWLPA). By introducing transfer learning, we mitigate issues of limited samples and uneven data distribution to enhance model generalization for unseen instances. The designed DWLPA module improves detection capability in complex environments through dynamic weight allocation and multi-scale feature refinement. Experiments on public UAV aerial imagery datasets demonstrate our improved model achieves 67.6% mAP@0.5:0.95, 98.3% mAP@0.5, and 81.4% mAP@0.75, consistently outperforming the YOLOv8-s baseline model and light-weight models (YOLOv13-s, YOLOv12-s, YOLOv11-s) with superior precision and robustness in complex environments.

Key words: UAV imagery; vehicle detection; YOLOv8-s; transfer learning; attention mechanism

随着科技进步和经济快速发展,我国汽车保有量激增,交通拥堵、事故频发等问题凸显。传统道路

监控系统因摄像头密度不足存在盲区。因此,亟需结合计算机视觉技术与无人机设备,实现道路车辆

收稿日期:2025-07-31

基金项目:国家自然科学基金项目(62262057);新疆维吾尔自治区重点研发计划项目(2023B01030)

作者简介:周永倩(2002—),女,硕士研究生,专业方向为计算机视觉。

*通信作者:李志刚(1970—),男,教授,从事大数据技术及应用的研究,e-mail:Lzg_inf@shzu.edu.cn。

的全自动化检测与管理。

早期车辆目标检测方法主要依赖手工特征提取(如 Haar、HOG 等)与传统机器学习算法(如 SVM^[1]、决策树^[2]等)。Cao 等^[3]改进 HOG 特征结合 SVM,在低空视频中实现 90% 准确率。Wen 等^[4]结合 AdaBoost 算法,优化 Haar-like 特征选择和归一化算法,提升车辆检测性能。Rios-Cobrerá 等^[5]融合多摄像头与置信度评分提升检测实时性。尽管上述方法在简单场景中具有一定有效性,但在复杂场景下鲁棒性不足,难以满足现代交通管理需求。

相较传统机器学习方法,深度学习通过自动特征学习显著提升检测性能。主流框架分为二阶段与一阶段方法,二阶段以 Faster R-CNN 为代表,通过 RPN 生成候选框实现高精度检测,但计算效率低,不适用于实时场景。Othmani 等^[6]优化卷积特征以提升检测灵敏度。Yin 等^[7]融合域适应技术减少小目标车辆漏检,提高检测性能 4.8%。Arora 等^[8]结合形态学操作与卡尔曼滤波器,增强车辆检测精度和鲁棒性。Luo 等^[9]结合 Retinex 增强与神经架构搜索优化车辆遮挡检测。张莹等^[10]采用 ResNet 替代 VGG-16,结合改进 Faster R-CNN,提升检测精度但检测速度受限。一阶段检测方法以 YOLO(You Only Look Once)系列^[11]为代表,通过直接回归目标位置和类别,兼顾检测精度与实时性,适用于高效实时应用。Zhu 等^[12]设计 TPH-YOLOv5,通过 Transformer 预测头提升车辆定位精度。Qiu 等^[13]设计 YOLO-GNS 算法,基于 GhostNet 优化 YOLOv7 骨干网络,增强小目标检测能力。Farid 等^[14]通过迁移学习优化 YOLOv5,提升车辆检测精度和执行效率。江德港等^[15]结合 EfficientNet 与 CIoU 损失优化无人机车辆检测的精度与漏检控制。张利丰等^[16]提出 RBT-YOLO,融合 RevColNet 与 BiFPN 以提升多尺度车辆检测。Bakirci 等^[17]通过引入 GELAN 与可编程梯度信息(PGI)模块改进 YOLOv9 模型,提升了检测精度与速度。Ou 等^[18]设计 Drone-TOOD 模型,构建轻量化 CSPRegNet 主干,融合 EVC 模块及高效任务分解注意力(ETDA)机制,提升了车辆检测精度与效率。Zhang 等^[19]提出轻量化 OSD-YOLOv10 算法,结合在线卷积重参数化(OCRConv)模块、轻量级特征提取结构(SPCC)及双馈混合注意力(DFMA)网络,在减少模型参数数量的同时提升了小目标检测精度。白俊卿等^[20]提出 Atiny-YOLO 算法,通过引入小目标检测层,改进 C3 与 Transformer 模块并优化特征融合,将车辆小目标检测精度提升 2.9%。

现有深度学习方法虽已在车辆目标检测中取得显著进展,但应用于无人机航拍图像时仍面临多重挑战。高空俯瞰视角下的目标遮挡、平台抖动导致的图像模糊及镜头畸变,严重干扰了特征提取;车辆目标因像素占比小且常密集重叠,导致检测精度不足;复杂多变的地面背景与光照、天气等因素引起的数据特征复杂性增加,使模型鲁棒性受限;同时,高质量、大规模数据的获取与标注存在显著难度。提升复杂环境下的目标检测精度与模型鲁棒性,同时减少对大规模标注数据的依赖,已成为该领域亟待解决的核心问题。

针对上述挑战,本文提出融合迁移学习与动态权重局部金字塔注意力(Dynamic Weight Local Pyramid Attention, DWLPA)的无人机车辆检测方法。迁移学习通过迁移通用特征参数并利用无人机车辆数据来微调模型,直接缓解数据稀缺问题,同时增强模型对尺度、视角差异等分布不均场景的泛化性与稳定性。针对图像识别困难(遮挡、模糊、小目标)和复杂背景干扰带来的检测精度和适应性挑战,改进的 DWLPA 模块在局部金字塔注意力(Local Pyramid Attention, LPA)模块^[21]基础上引入动态权重与多尺度优化;其局部金字塔结构专注于挖掘多尺度特征以提升对小目标的敏感性;动态分配注意力权重的机制则强化了对关键区域的聚焦并抑制无关背景干扰。本文集成迁移学习与 DWLPA 对 YOLOv8-s 进行优化,迁移学习有效克服数据瓶颈并增强基础泛化能力, DWLPA 则显著提升了目标检测精度及复杂环境的鲁棒性。2 种方法的结合提升了模型在无人机航拍影像中的目标检测精度和鲁棒性。

1 材料与方 法

1.1 材料准备

1.1.1 数据获取

本文采用 GitHub 公开的无人机航拍车辆图像数据(<https://github.com/DataSharing997/UrbanDroneVehiclesImages>),该数据涵盖多种环境下的 1 604 张 1 920×1 080 分辨率图像。为确保标注的准确性,采用 Labelimg 1.8.6 工具进行人工标注,并划分为训练集(962 张)、验证集(321 张)和测试集(321 张)。基于车辆尺寸及功能用途,将检测目标划分为常规车辆(vehicle)和大型车辆(bigvehicle)2 类。数据集中共标注 38 554 个 vehicle 目标和 2 548 个 bigvehicle 目标。其中,vehicle 类包括微型和小型载人轿车,而 bigvehicle 类涵盖中型及以上载人或载货

客车(如公交车、大货车等)及作业专用车(如混凝土搅拌车、高空作业车等)。

此外,本文中检测目标类别间差异显著,vehicle类别的样本数量远多于bigvehicle,同时2类目标的尺寸分布跨度较大,使得本文数据存在分布不均的情况。图1展示了数据集中部分的车型,可以证实本文涉及车型种类的多样性和差异性。



图1 数据集中的部分车型

Figure 1 Partial vehicle types in the dataset

1.1.2 数据预处理

为应对无人机航拍影像中的多重挑战,本文采用线下基础增强与线上动态增强相结合的数据预处理方案。线下增强阶段旨在系统性地应对目标遮挡、图像畸变及复杂背景等基础性挑战。如图2所示,通过像素擦除模拟目标遮挡场景,引导模型学习局部特征。采用随机旋转与平移变换,以应对因无人机姿态变化引起的图像畸变,同时,利用亮度调整来模拟不同光照条件下的复杂背景。在此过程中,每张图像至少经历1种变换。该策略将训练集规模从962张扩充至1924张,显著提升了数据多样性,为模型训练提供了稳健的基础。在此基础上,为解决小目标密集与图像模糊等更为复杂的挑战,本文在训练过程中动态引入了Mosaic增强^[22]、随机仿射变换、高斯模糊与中值模糊、对比度受限自适应直方图均衡化和HSV色彩空间随机增强线上策略。线上与线下策略的结合构成了全面的预处理方案,有效降低了模型过拟合风险,提升了模型在复杂条件下的综合性能。经过上述线上线下数据增强策略,最终用于模型训练、验证和测试的总数据集规模扩充至2566张图像。

1.2 实验方法

1.2.1 YOLOv8-s

YOLO系列算法最早由Redmon等^[23]提出,其核心思想是将检测任务视为回归问题,通过单次前向传播实现目标定位与分类。与依赖区域建议的两阶段方法相比,YOLO兼具较高的实时性和检测精度,并通过持续迭代优化提升性能表现。

YOLOv8通过改进网络架构与增强多尺度特征融合,使检测精度与推理速度进一步优化,其轻量化设计也更符合边缘部署需求。

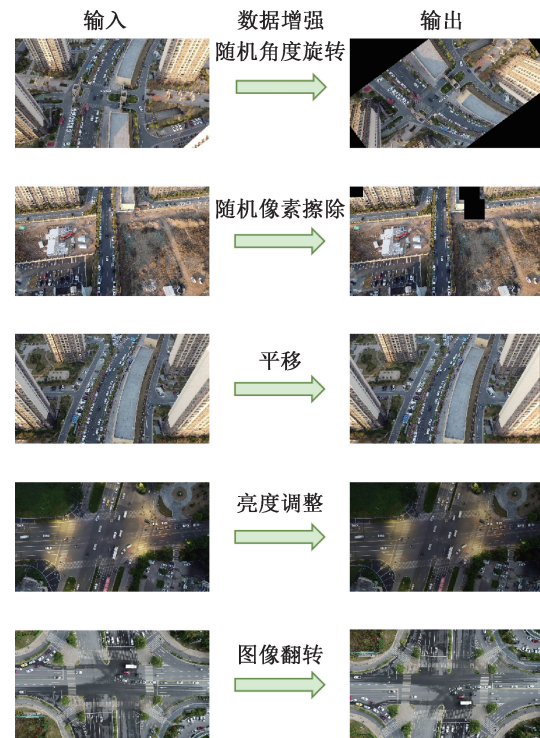


图2 线下数据增强操作

Figure 2 Offline data augmentation operations

YOLOv8-s是YOLOv8系列中兼具轻量化和高性能的检测模型,其高效架构特别适用于边缘计算设备的车辆检测任务。网络架构由骨干、颈部和头部3部分组成,协同完成特征提取、融合与目标检测。骨干网络从输入图像中提取丰富的特征信息,初始阶段采用1个卷积块,由二维卷积层、批量归一化层(BN)和激活函数组成。之后,骨干网络进入4个阶段的逐层构建,每个阶段包括1个ConvModule模块和1个C2f模块。C2f模块集成Darknet瓶颈块,利用小卷积核提升特征提取效率并降低计算复杂度。随着网络加深,特征通道数逐级递增,Darknet瓶颈块数量也随之变化,并在最终阶段引入快速空间金字塔池化(Spatial Pyramid Pooling-Fast, SPPF)模块增强多尺度特征提取能力。颈部模块采用路径聚合网络(Path Aggregation Feature Pyramid Network, PAFPN),通过上采样和下采样实现多尺度特征融合。上采样层采用最近邻插值进行2倍上采样,提升空间分辨率,并结合C2f模块优化跨阶段特征融合。下采样层利用ConvModule缩小特征图尺寸,提取更高层次语义信息,再通过C2f模块增强特征表达能力。最终,融合后的特征图传递至后续模

块,实现多尺度特征的高效利用。YOLOv8-s 的头部设计采用分离式结构,包含 2 个并行分支:类别预测分支输出目标类别概率,边界框回归分支预测目标位置坐标。训练阶段采用多种优化损失函数:二元交叉熵(BCE)损失用于类别预测,完全交并比损

失(CIoU)用于边界框回归,分布焦点损失(DFL)提升定位精度。此外,训练阶段采用任务对齐分配器(TAA)作为标签分配策略,依据任务对齐动态匹配预测框与真实框,确保训练过程中标签分配的准确性与高效性。其网络结构如图 3 所示。

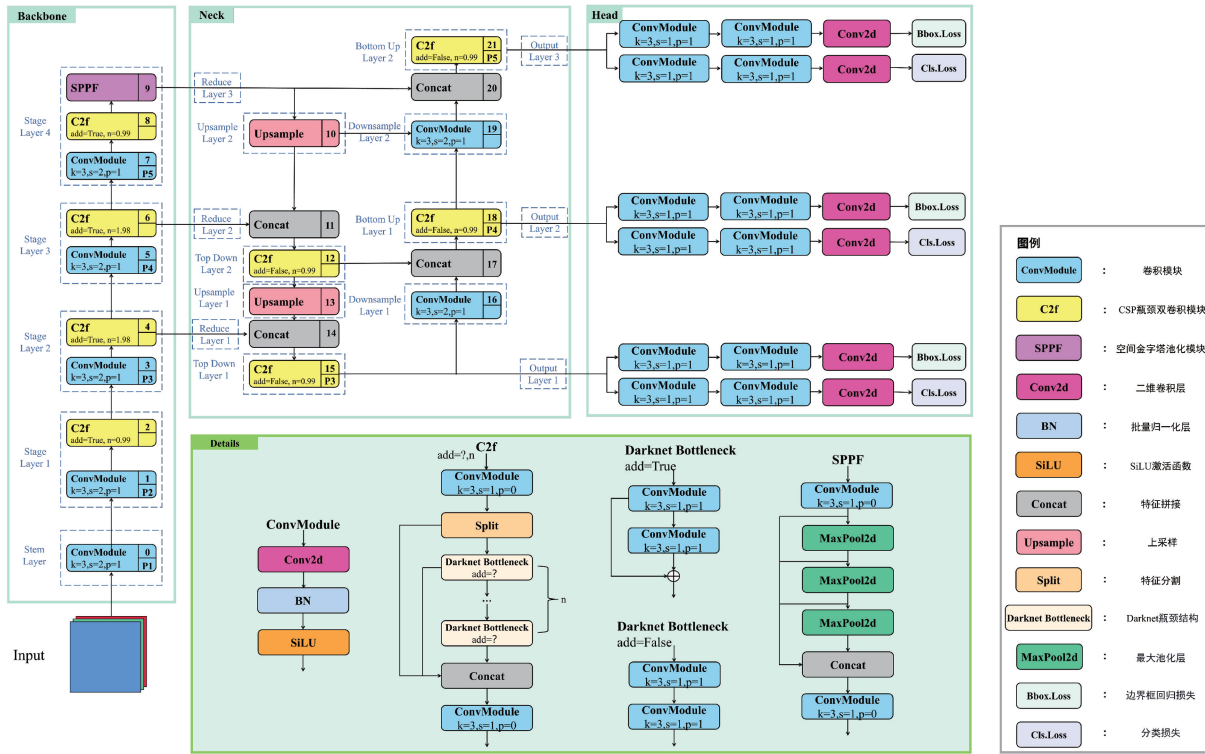


图 3 YOLOv8-s 的模型结构
Figure 3 Model structure of YOLOv8-s

1.2.2 改进的 YOLOv8-s 网络

1) DWLPA

在计算机视觉领域,注意力机制通过引导模型自动聚焦输入数据的关键区域,显著提升模型性能与效率。该方法被广泛应用于目标检测任务,通过强化目标区域特征表示并抑制背景干扰,有效提高检测精度与鲁棒性。

注意力机制的成效主要源于其对特征的灵活筛选与强化能力。例如,通道注意力(Channel Attention, CA)通过为不同通道分配权重,帮助模型关注重要特征;空间注意力(Spatial Attention, SA)着重定位关键区域,减少背景干扰,从而改善视觉任务的表现。

在现有方法中,LPA 模块利用多尺度特征聚合策略和高效加权方式,提供了一种新的特征增强方法,强化了模型的特征表达能力。

LPA 模块结合了 CA 模块和 SA 模块,通过优化不同尺度和层次的特征表示来进一步增强模型的特征表达。SA 模块通过空间信息定位与任务最相关的特征,其输出 A_s 被定义为:

$$A_s = \sigma(\xi(\text{Concat}(P_{\text{avg}}(X), P_{\text{max}}))) \quad (1)$$

式(1)中的 σ 代表 sigmoid 函数, P_{max} 和 P_{avg} 分别表示通道维度上的自适应最大池化和自适应平均池化, $\text{Concat}(P_{\text{avg}}(X), P_{\text{max}})$ 表示将自适应平均池化和最大池化的输出在通道维度上进行拼接, ξ 表示卷积操作。

CA 模块通过给通道分配不同的权重来测量特征的重要性。由 1 个全局平均池化层和 2 个全连接层组成, CA 模块的输出 A_c 被定义为:

$$A_c = \sigma(\text{FC}_2 \text{ReLU}(\text{FC}_1 \text{Pool}_{\text{avg}}(X))) \quad (2)$$

式(2)中的 FC_1 和 FC_2 代表 2 个全连接层, Pool_{avg} 代表平均池化层。

LPA 模块擅长处理多尺度特征,适用于无人机航拍车辆检测中的复杂背景与多尺度目标。其即插即用特性便于模型集成,有效降低开发成本。传统 LPA 模块通过全局与局部分支捕获特征,采用固定加权融合策略。但固定权重难以自适应不同输入特征,影响特征重要性分配;同时多尺度处理灵活性不足,限制关键信息捕获能力。

为此,本文提出 DWLPA 模块,在 LPA 基础上引入动态权重机制,使特征融合更具适应性和灵活性,从而提升目标检测性能。其模块结构如图 4 所示。

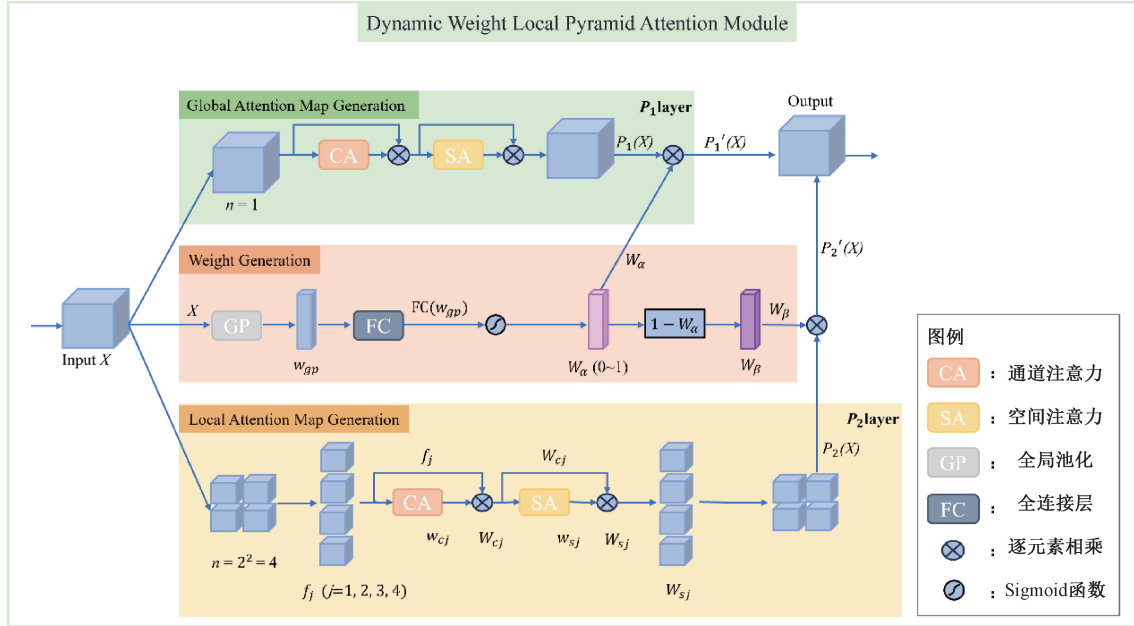


图 4 DWLPA 模块结构图

Figure 4 Structure diagram of DWLPA module

DWLPA 模块通过引入动态权重融合机制改进传统 LPA 模块的特征融合,使得特征图融合的权重能依据输入特征进行动态调整。在 DWLPA 模块中,全局注意力图生成(Global Attention Map Generation, GAMG)分支和局部注意力图生成(Local Attention Map Generation, LAMG)分支分别用于捕获全局和局部特征,权重生成(Weight Generation, WG)分支则根据全局和局部特征的重要性,动态调整融合比例。

GAMG 分支采用 CA 和 SA 依次处理特征图 X , 先由 CA 模块生成通道维度权重 W_c , 再通过 SA 模块生成空间维度权重 W_s , 两者融合后得到全局注意力特征图 $P_1(X)$, 增强全局信息感知。WG 分支首先通过全局池化(Global Pooling, GP)将输入特征图 X 压缩为全局特征 w_{gp} , 随后经全连接层与 Sigmoid 激活函数得到动态权重 $W_\alpha \in [0, 1]$, 进而令 $W_\beta = 1 - W_\alpha$, 分别作为局部与全局注意力特征图的融合权重, 以动态平衡全局与局部信息。LAMG 分支将 X 分割成 4 个子块, 分别计算通道注意力权重 w_{cj} 和空间注意力权重 w_{sj} , 每个子块经注意力加权与特征拼接, 生成局部注意力特征图 $P_2(X)$, 用于捕捉局部上下文信息。最终, $P_1(X)$ 与 W_α 相乘得到 $P_1'(X)$; $P_2(X)$ 与 W_β 相乘得到 $P_2'(X)$; 输出特征图通过 $P_1'(X)$ 和 $P_2'(X)$ 的加权融合得到, 实现对全局和局部特征的动态调整和优化。

DWLPA 模块通过结合 GAMG 分支、WG 分支和 LAMG 分支, 能够有效地提取全局和局部的特征, 并通过动态权重机制自适应地调整特征图的融合比例, 使模型能在多尺度环境下灵活处理特征, 提高特征表达能力和检测性能。

2) 迁移学习

迁移学习通过迁移源任务知识提升目标任务的学习效率, 在图像分类与目标检测等计算机视觉任务中缓解数据稀缺问题。本文采用 COCO 数据集^[24]进行迁移学习, 该数据集包含 33 万张图像和 80 类目标, 提供边界框和分割掩码等注释, 其数据多样性和标准评估指标(AP, mAP)适用于无人机航拍车辆检测任务。

首先, 使用 MS COCO 数据集对 YOLOv8-s 模型进行预训练, 让模型学习通用特征表示。预训练过程中, Backbone、Neck 和 Head 结构一致的卷积层与 BN 层直接加载权重; 由于类别数从 COCO 数据集的 80 个类别减少到车辆目标检测任务的 2 类, 类别预测分支权重维度不匹配而不被加载; 新增 DWLPA 模块因未包含于预训练模型亦不加载。未加载模块采用 PyTorch 默认参数初始化方法, 以确保权重参数的完整性。完成预训练后, 将特征权重迁移至车辆检测任务进行微调, 模型结合车辆目标检测任务的具体数据和需求进一步优化其参数, 使模型适配航拍车辆检测的需求。

1.2.3 损失函数

在模型训练过程中,改进的 YOLOv8-s 模型的损失函数由分类损失、边界框回归损失和分布焦点损失 (Distribution Focal Loss, DFL) 的加权求和计算得到。

分类损失函数使用二元交叉熵 (BCE) 逐类别评估模型预测的误差,计算公式如下:

$$L_{cls}(i) = -[y_i \ln p_i + (1 - y_i) \ln(1 - p_i)], \quad (3)$$

其中, $L_{cls}(i)$ 表示第 i 个样本的分类损失, $y_i (y_i \in \{0, 1\})$ 是第 i 个样本的真实标签, $y_i = 1$ 表示目标存在, $y_i = 0$ 表示目标不存在。 $p_i (0 \leq p_i \leq 1)$ 是第 i 个样本属于正类的预测概率。

边界框回归损失是预测边界框与真实边界框之间的完全交并比 (CIoU) 损失,计算公式如下:

$$L_{bbox}(i) = 1 - CIoU_i, \quad (4)$$

其中, $L_{bbox}(i)$ 表示第 i 个样本的边界框回归损失,

$$CIoU_i = IoU_i - \frac{\rho^2(b_i, b_i^{gt})}{c_i^2} - \alpha_i v_i, IoU_i = \frac{|b_i \cap b_i^{gt}|}{|b_i \cup b_i^{gt}|},$$

IoU_i 是第 i 个样本的预测边界框和真实边界框之间的交并比, b_i 是预测边界框的面积, b_i^{gt} 是真实边界框的面积。 $\rho^2(b_i, b_i^{gt})$ 表示预测边界框和真实边界框中心点之间的欧几里得距离的平方。 c_i^2 是第 i 个样本中包含预测边界框和真实边界框的最小外接矩形对角线长度的平方。

$$\alpha_i = \frac{v_i}{(1 - IoU_i) + v_i},$$

$$v_i = \frac{4}{\pi^2} \left(\arctan \frac{w_i^{gt}}{h_i^{gt}} - \arctan \frac{w_i}{h_i} \right)^2,$$

其中 α_i 是正样本权衡参数, v_i 用于衡量预测边界框和真实边界框中心点之间的宽高比的一致性, w_i^{gt} 和 h_i^{gt} 分别表示真实边界框的宽度和高度, w_i 和 h_i 则是预测边界框的宽度和高度。

DFL 用于比较边界框中心点到各边的预测距离概率分布与实际距离之间的差异,计算公式如下:

$$L_{DFL}(i) = -\frac{1}{4} \sum_{j \in \{l, t, r, b\}} [(y_{i,j,k+1} - y_{i,j}) \ln(p_{i,j,k+1}) + (y_{i,j} - y_{i,j,k}) \ln(p_{i,j,k})]. \quad (5)$$

其中, $L_{DFL}(i)$ 表示第 i 个样本的 DFL 损失, j 表示边界框的 4 个边: 左 (l)、上 (t)、右 (r)、下 (b), k 为预定义离散区间的索引, 用于将第 i 个样本边界框中心点到第 j 边的连续距离值映射到离散区间。 $y_{i,j}$ 表示第 i 个样本边界框中心到第 j 边的

实际距离, $y_{i,j,k}$ 表示第 i 个样本边界框中心到第 j 边距离在第 k 个离散区间下界, $y_{i,j,k+1}$ 表示第 i 个样本边界框中心到第 j 边距离在第 $k+1$ 个离散区间上界。 $p_{i,j,k}$ 表示模型预测第 i 个样本中心到第 j 边距离落于第 k 个区间的概率, $p_{i,j,k+1}$ 表示模型预测第 i 个样本中心到第 j 边距离落于第 $k+1$ 个区间的概率。

2 结果与分析

2.1 实验平台

本实验服务器在 Windows 11 专业版上运行,使用 Python 3.8 进行编译与执行。模型的训练、验证与测试依赖于 CUDA 11.3 和 PyTorch 1.10.2, 确保高效的计算性能。配备 NVIDIA GeForce RTX 4060 显卡, 以加速深度学习任务的执行, 且搭载 12th Gen Intel(R) Core(TM) i3-12100F 处理器与 16 G 内存, 提供稳定而迅速的处理能力, 满足复杂模型的训练与数据处理需求。

2.2 模型训练

2.2.1 训练方法

本文未设置独立预训练阶段, 模型架构直接加载 YOLOv8-s 在 MS COCO 通用目标检测数据集上预训练完成的权重参数进行初始化, 充分继承其通用特征提取能力。其初始权重来源于 GitHub 上公开的 YOLOv8-s 预训练模型权重。COCO 数据集包含 80 类通用目标, 覆盖多样化的目标类别与场景。通过此预训练过程, 模型学习通用目标的特征表示能力, 为后续任务迁移提供参数初始化基础。

在微调阶段, 采用基于城市道路无人机航拍车辆公开数据进行人工标注的自建数据集进行模型调优。该数据集包含 2 566 张分辨率 1 920×1 080 的无人机航拍图像, 覆盖复杂城市道路场景, 车辆目标呈现小尺度目标的多尺度特性, 其像素尺寸分布在 15×15 至 53×53 区间。该尺度差异主要源于车辆本体尺寸差异及遮挡程度变化。为平衡高分辨率细节保留与计算效率, 将输入尺寸设置为 1 280×1 280 像素, 可在 RTX 4060 上实现 batch_size = 2 的稳定训练。在 80 轮训练中, 从第 10 轮开始每 1 轮都在验证数据集进行一次评估, 以定期监测模型的性能。通过这种策略, 可以对模型的学习进度进行全面检查, 并确保对训练的最后阶段进行密切监控, 以优化模型的性能。微调阶段的超参数详情如表 1 所示。

表 1 微调阶段的详细超参数

超参数	值
损失函数	BCE 损失、CIoU 损失、DFL
输入尺寸	1 280×1 280 像素
NMS 的 IoU 阈值	0.65
学习率调度策略	带预热的余弦衰减策略
初始学习率	0.001
学习率缩放因子	0.25
优化器	SGDM
SGDM 动量值	0.937
SGDM 权重衰减系数	0.000 5
批量大小	2
数据加载进程数	4
最大训练轮数	80

2.2.2 模型评价标准

为了评估该方法的检测性能,本文使用了平均精度均值(mAP)和平均精度(AP)度量。mAP 是指不同类别和 IoU 阈值的 AP 值的平均值,而 AP 是在特定的 IoU 阈值下计算的。在特定类别 j 的给定 IoU 阈值 t 的 AP 被定义为:

$$AP_{j, IoU=t} = \int_0^1 p_j(r) dr. \quad (6)$$

其中, $p_j(r)$ 是 j 类召回 r 的函数。这个积分表示在阈值 t 下 j 类的精度-召回(P-R)曲线下的面积。较高的 IoU 阈值对应于预测值和地面真实值之间的边界框重叠程度的判定更严格,反映了模型检测的准确性。

本文中使用的 mAP 指标的定义如下:

$$mAP@0.5:0.95 = \frac{1}{10} \sum_{i=0}^9 \left(\frac{1}{C} \sum_{j=1}^C AP_{j, IoU=0.5+0.05i} \right), \quad (7)$$

$$mAP@0.5 = \frac{1}{C} \sum_{j=1}^C AP_{j, IoU=0.5}, \quad (8)$$

$$mAP@0.75 = \frac{1}{C} \sum_{j=1}^C AP_{j, IoU=0.75}. \quad (9)$$

在式(7)、(8)和(9)中, $mAP@0.5:0.95$ 表示 0.5~0.95 IoU 阈值内计算的 mAP 值, $mAP@0.5$ 和 $mAP@0.75$ 分别为固定 IoU 阈值为 0.5 和 0.75 时的 mAP 值。 C 表示对象类别的数量。

2.3 实验结果分析

2.3.1 损失分析

在微调阶段,改进的 YOLOv8-s 模型采用 3 个关键损失指标进行监督:分类损失、边界框回归损失及分布焦点损失,三者共同构成总损失。这些指标

对车辆精准检测与定位具有重要作用。在训练过程中,批次大小设置为 2,因此每个训练轮次迭代 962 次。模型共训练 80 轮,总迭代次数达到 76 960 次(图 5)。如图 5a 所示,总损失反映模型整体优化情况。初始损失为 10.91,经 20 000 次迭代快速下降后稳定于 4.824,表明模型收敛良好,成功学习了车辆检测的核心特征。下降过程无剧烈波动,表明训练稳定性较高。

分类损失直接决定模型区分车辆与非车辆目标的准确性,影响误检率和分类定位的协同能力。根据图 5b 显示,分类损失从 3.833 降至 0.853,下降 77.7%。在前 10 000 次迭代时曲线大幅下降,表明模型初始化合理,使得梯度方向能快速引导网络学习车辆分类特征。

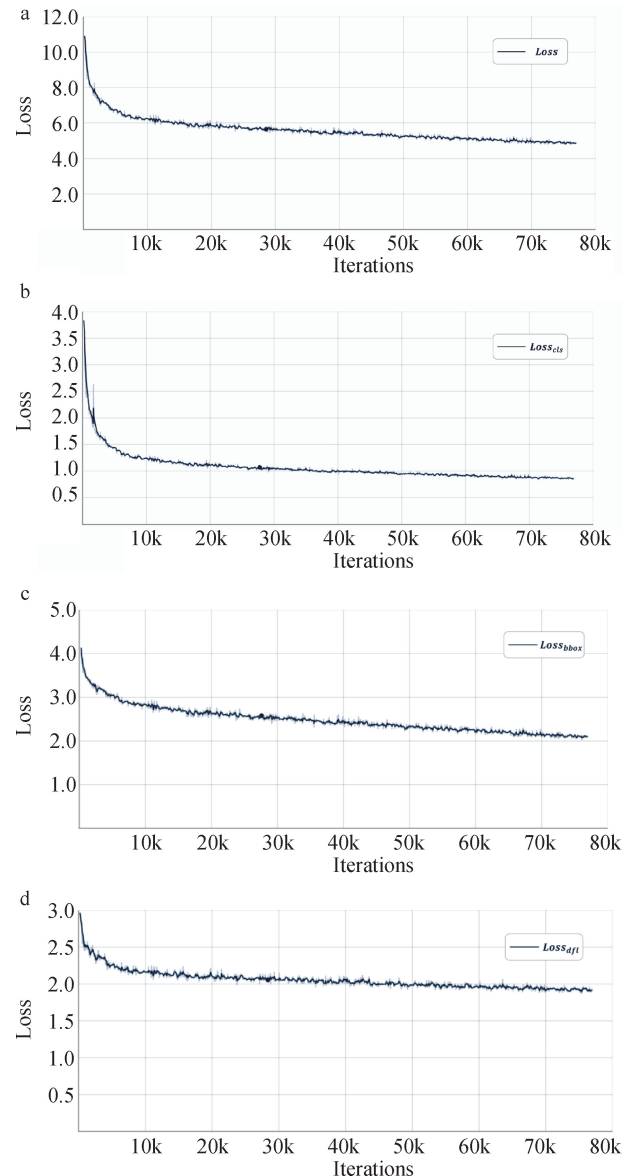


图 5 损失曲线

Figure 5 Loss curves

边界框回归损失关系到检测框与真实车辆位置的匹配精度,决定模型对车辆位置和大小预测能力。如图5c所示,损失值从初始4.113逐步下降至2.093,这表明模型通过整体的迭代有效优化了定位参数,使得检测框更加精准,提高了模型对不同尺度车辆的适应能力。

DFL主要用于优化检测框的位置预测,提升密集车辆或小目标的检测稳定性。图5d显示DFL从初始值2.966逐步下降至1.902,降幅约35.9%,表明模型逐步学习到车辆位置分布的统计规律,在无人机航拍等高视角复杂环境下,对密集车辆的定位和检测的精度有所增强。

改进后的YOLOv8-s模型在训练过程中,各损失项均呈现合理的收敛趋势,验证了模型优化的有效性。

2.3.2 消融实验

本文通过消融实验评估DWLPA模块和迁移学习(TF)对YOLOv8-s模型的独立和联合效果。如表2所示,联合使用DWLPA和TF在测试集上所有关键指标均表现最佳,mAP@0.5:0.95提升至67.6%(增加1.8%),mAP@0.5升至98.3%(增加0.2%),mAP@0.75升至81.4%(增加2.8%)。DWLPA增强局部特征提取能力,在保持高召回率的前提下改善了中高阈值检测精度;TF则通过迁移学习增强模型泛化性,使mAP@0.5:0.95和mAP@0.75分别提高1.5%和1.9%。两者结合显著提升了复杂场景下的检测性能,验证了多模块融合的有效性。

表2 不同改进方法之间的对比

方法	mAP@ 0.5:0.95	mAP@0.5	mAP@0.75
YOLOv8-s	65.8%	98.1%	78.6%
YOLOv8-s+TF	67.3%	98.0%	80.5%
YOLOv8-s+DWLPA	66.6%	98.2%	79.8%
YOLOv8-s+DWLPA+TF	67.6%	98.3%	81.4%

1) TF方法的消融研究

bigvehicle类别样本数量过少导致数据分布不均,使得模型对该类目标学习不充分。TF能够有效缓解此类样本失衡问题,提高该类别的检测效果。表3结果显示,引入迁移学习后,bigvehicle在测试集的AP@0.5:0.95达到67.1%,较YOLOv8-s提升1.9%,在更严格的AP@0.75指标上增幅达到2.7%,明显超过vehicle类别在这2个指标上的提升。这表明迁移学习方法有效提升了本文数据集中

少数类别的检测性能,改善了模型在数据分布不均条件下的鲁棒性。

表3 不同类别车辆的检测对比

Table 3 Detection comparison of different vehicle categories

指标	类别	YOLOv8-s	YOLOv8-s+TF
AP@0.5:0.95	vehicle	66.4%	67.4%
	bigvehicle	65.2%	67.1%
AP@0.5	vehicle	98.4%	98.5%
	bigvehicle	97.7%	97.6%
AP@0.75	vehicle	79.0%	80.0%
	bigvehicle	78.3%	81.0%

2) DWLPA模块的消融研究

为验证DWLPA模块在不同骨干网络架构上的泛化能力与鲁棒性,本文选取了YOLOv5-s、YOLOv7-tiny及YOLOv8-s作为测试模型。这3个模型的骨干网络设计分别代表了YOLO系列的3次关键架构范式跃迁。如表4所示,在这3个具有显著架构差异的骨干网络上,集成DWLPA模块后均实现了性能的稳定提升。其中,mAP@0.5:0.95指标分别提升了0.2%、0.1%和0.8%,mAP@0.75指标则分别提升了0.8%、0.2%和1.2%。这充分证明了DWLPA作为一种即插即用模块,具有良好的架构无关性,能够有效适配从奠基性的CSP结构到现代化的C2f结构等多种网络范式,并持续为模型带来性能增益。

表4 DWLPA在不同骨干网络上的对比

Table 4 Comparison of DWLPA on different backbone networks

方法	mAP@ 0.5:0.95	mAP@0.5	mAP@0.75
YOLOv5-s	63.5%	98.1%	74.0%
YOLOv5-s+DWLPA	63.7%	98.2%	74.8%
YOLOv7-tiny	63.8%	98.2%	75.3%
YOLOv7-tiny+DWLPA	63.9%	98.3%	75.5%
YOLOv8-s	65.8%	98.1%	78.6%
YOLOv8-s+DWLPA	66.6%	98.2%	79.8%

本文将DWLPA与3种基于小目标检测具有代表性的注意力模块(HFP^[25]、SEAM^[26]、DSAM^[27])在YOLOv8-s骨干网络上进行了对比实验,以验证其性能优势。如表5所示,本文提出的DWLPA模块在mAP@0.5:0.95、mAP@0.5和mAP@0.75三项关键指标上均取得最优性能,分别达到66.6%、98.2%和79.8%。这表明,DWLPA通过有效融合全局与局部特征,并利用动态权重机制自适应优化多尺度特征表达,显著提升了小目标的检测精度,其综合性能与定位精度均优于其余3个小目标注意力

模块。

表 5 不同注意力模块的对比

方法	mAP@ 0.5;0.95	mAP@ 0.5	mAP@ 0.75
YOLOv8-s+HFP	66.0%	98.0%	78.9%
YOLOv8-s+SEAM	66.2%	98.0%	79.0%
YOLOv8-s+DSAM	66.0%	98.1%	79.2%
YOLOv8-s+DWLPA	66.6%	98.2%	79.8%

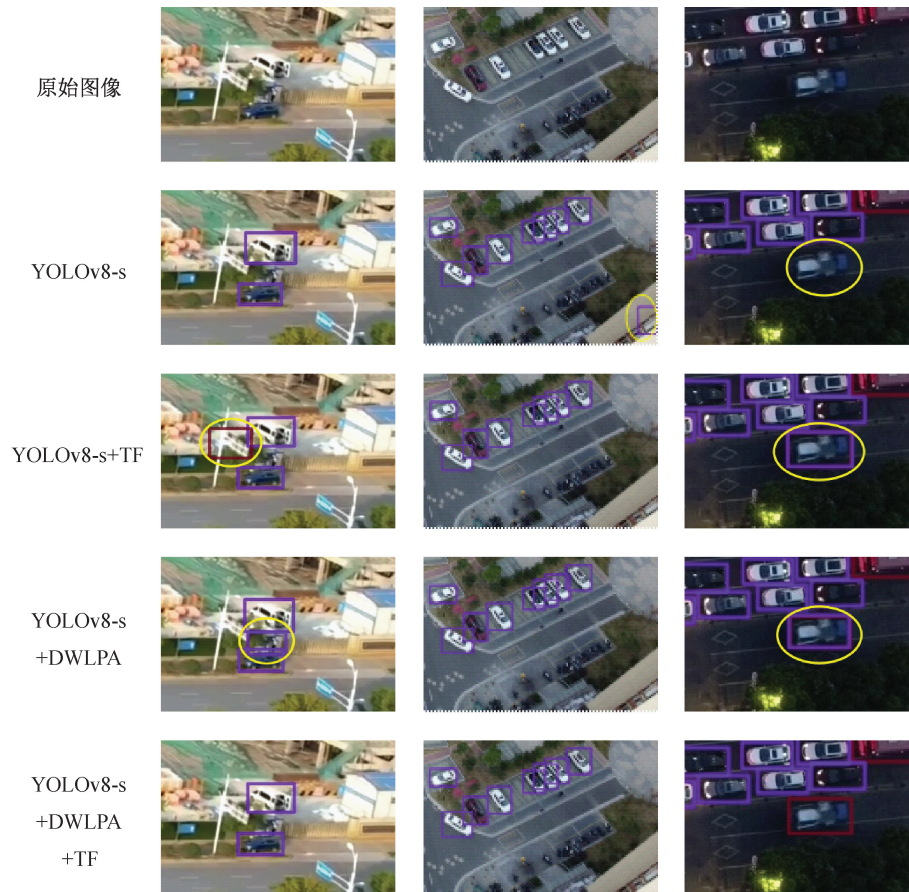
为评估改进模型效率,结合其计算成本与推理速度分析。如表 6 所示,本文方法 FPS 从 48.5 帧降至 41.2 帧,只降低 7.3 帧。FLOPs 只增加了 0.005 G,因此计算量的增加可忽略不计,但如表 2 所示,其精度明显提升,mAP@ 0.5 :0.95 增长 1.8%,mAP@ 0.75 增长 2.8%。需要说明的是,表 6 中的 YOLOv8-s + DWLPA 即为本文提出的完整方法(YOLOv8-s+DWLPA+TF)。由于 TF 仅作用于训练阶段,并未改变最终模型的网络结构,因此其 FPS、Param. 和 FLOPs 与仅加入 DWLPA 的模型完全一致。对于无人机航拍等精度敏感场景,以轻微速度损失换取明显的精度提升是合理权衡,且 41.2 帧

FPS 的推理速度仍能满足实时检测需求,证明了本方法在保持实用性的同时有效提升了检测性能。

表 6 改进模型与基线模型的效率对比

方法	FPS/帧	Param./M	FLOPs/G
YOLOv8-s	48.5	11.136	57.089
YOLOv8-s+DWLPA	41.2	11.824	57.093

为更加直观地展示不同改进方法的性能差异,图 6 呈现了 3 种场景及光线条件下的无人机航拍图像及其对应改进方法的推理结果。图中紫框代表 vehicle(常规车辆),红框代表 bigvehicle(大型车辆),黄框圈出了误检或漏检现象。YOLOv8-s 在第二、三张图像中存在明显错误,将建筑物误识别为常规车辆或未能检测到货车。YOLOv8-s+TF 在第一、三张图像中发生误检,将交通标志杆错误识别为大型车辆以及将货车误分类为常规车辆。YOLOv8-s+DWLPA 在第一张图像中错误检测建筑物与交通标志杆的重叠部分,在第三张图像中未能正确分类货车。这些方法在复杂背景与低光照环境下车辆特征



紫框:常规车辆;红框:大型车辆;黄框:误检或漏检现象。

图 6 不同改进方法的推理结果对比

Figure 6 Comparison of inference results of different improved methods

模糊时均存在局限性,影响检测精度和鲁棒性。相比之下,本文的改进方法在全部场景中均表现优异,能够精准检测车辆目标并正确区分不同类别车辆,且无明显误检或漏检情况。结果证明本文方法能够有效应对复杂背景和不同光照条件,提高目标检测的准确性和稳定性。

2.3.3 与经典目标检测模型的对比

本文将改进的 YOLOv8-s 与 YOLOv8-s 基线、YOLOv7-tiny^[28]、YOLOv6-s^[29] 和 YOLOv5-s 模型在本文数据集上的性能进行对比。结果如表 7 所示,本文方法在 mAP@0.5:0.95、mAP@0.5 和 mAP@0.75 三项指标在测试集上均表现最佳,分别达到 67.6%、98.3% 和 81.4%。YOLOv7-tiny 和 YOLOv5-s 虽运行高效,但精度较低,mAP@0.5:0.95 分别为 63.8% 和 63.5%。结果表明,本文方法能够明显提

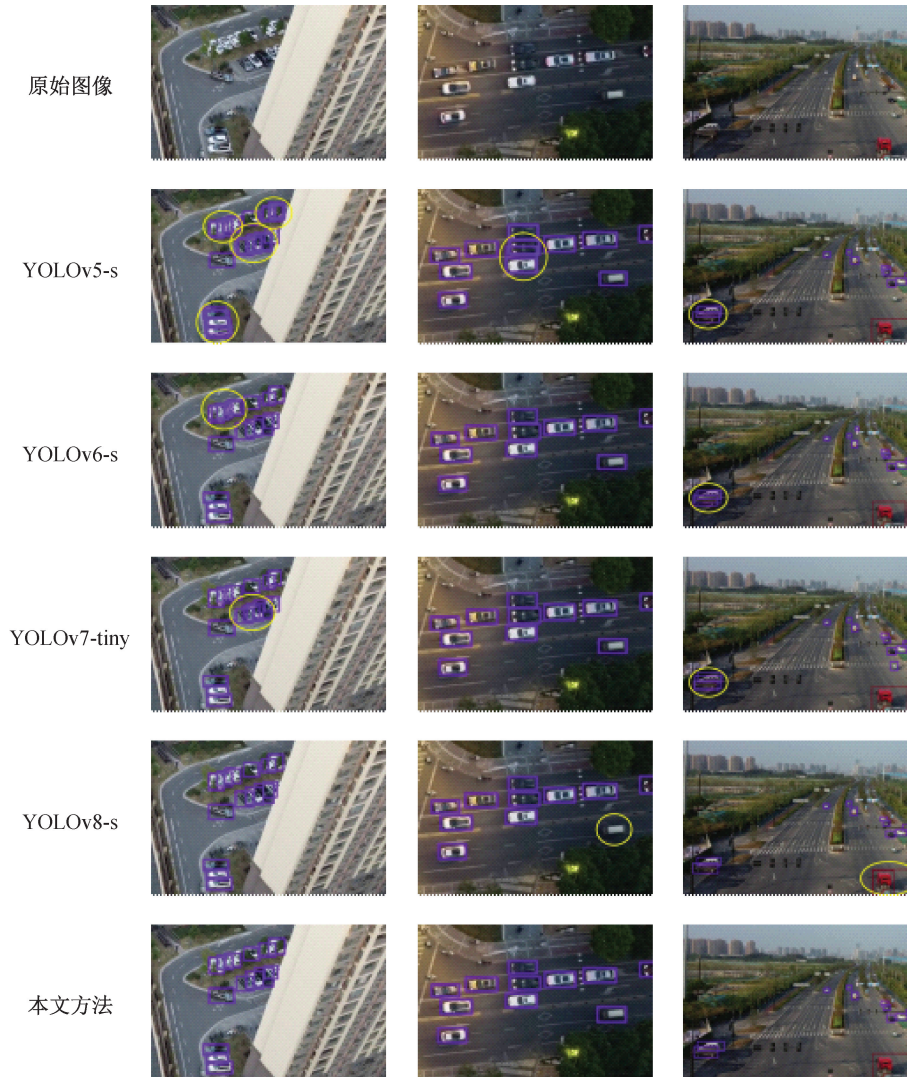
升模型性能,改进后的 YOLOv8-s 在复杂场景下表现出更强的适应性和鲁棒性。

表 7 与经典目标检测模型的对比

Table 7 Comparison with classic object detection models

方法	mAP@0.5:0.95	mAP@0.5	mAP@0.75
YOLOv5-s	63.5%	98.1%	74.0%
YOLOv6-s	64.9%	98.1%	77.4%
YOLOv7-tiny	63.8%	98.2%	75.3%
YOLOv8-s	65.8%	98.1%	78.6%
本文方法	67.6%	98.3%	81.4%

图 7 展示了改进方法与多种主流模型在复杂场景下的性能差异。YOLOv5-s 在车辆密集场景和低光照环境下出现重复检测,同一车辆被多个检测框覆盖,该现象在全部测试图像中均可观察到。YOLOv6-s 和 YOLOv7-tiny 在低光照环境下表现优于 YOLOv5-s,但在高密度目标区域仍存在重复检



紫框:常规车辆;红框:大型车辆;黄框:误检或漏检现象。

图 7 与经典目标检测模型的推理结果对比

Figure 7 Comparison of inference results with classic object detection models

测。YOLOv8-s 基线模型在暗光环境下易漏检,并在遮挡场景中对大货车存在重复检测问题。上述主流模型在低光照、目标密集及遮挡场景中均存在显著局限性。相比之下,本文方法能够有效减少冗余检测框,并提升暗光环境和车辆密集区域目标重叠和遮挡情况下的检测能力,同时减少漏检和误检。

2.3.4 与先进目标检测模型的对比

为进一步验证本方法在最新 YOLO 系列中的竞争力,本文将改进的 YOLOv8-s 与 YOLOv9-s^[30]、YOLOv10-s^[31]、YOLOv11-s、YOLOv12-s^[32]和 YOLOv13-s^[33]进行对比。如表 8 所示,本文方法的综合性能指标 mAP@0.5:0.95 与高精度指标 mAP@0.75 分别达到 67.6% 和 81.4%,均领先于所有对比模型。尽管其 mAP@0.5 为 98.3%,与 YOLOv9-s 等模型并列,但这反映了在宽松检测标准下,先进模型的基础检测能力已接近极限。本文方法的优势在于,通过 DWLPA 和 TF 模块显著提升了目标定位的精确度,这在更严格的评估指标中得到了充分体现,证明了其在复杂场景下的优越性与鲁棒性。

2.3.5 可视化分析

Grad-CAM^[34]是一种通过梯度加权生成类激活

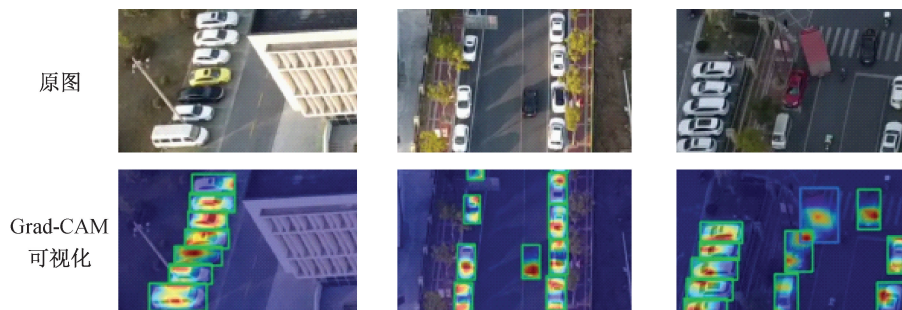


图 8 Grad-CAM 可视化

Figure 8 Grad-CAM visualization

3 讨论与结论

为提升复杂场景下无人机航拍影像的车辆检测性能,本文基于 YOLOv8-s 模型,设计并实现了融合迁移学习与动态权重局部金字塔注意力的改进方法。该方法利用迁移学习缓解数据分布不均,增强模型对稀疏类别的识别能力与未见样本的泛化性,同时通过 DWLPA 模块优化多尺度特征融合,提高模型对复杂环境的适应性。实验验证,改进模型在 mAP@0.5:0.95、mAP@0.5 和 mAP@0.75 指标上分别达到 67.6%、98.3% 和 81.4%,相较于 YOLOv8-s 基线及 YOLOv7-tiny 等轻量级模型均有明显提升。需要指出的是,该方法在提升检测精度的同时,也增

加了模型的计算负担。未来研究将致力于通过模型剪枝、知识蒸馏等技术实现轻量化优化,同时探索边缘协同计算机制,构建低计算开销与高推理速度的检测框架,以满足无人机航拍视频的端到端实时处理与精度需求。

表 8 与先进目标检测模型的对比

Table 8 Comparison with advanced object detection models

方法	mAP@0.5:0.95	mAP@0.5	mAP@0.75
YOLOv8-s	65.8%	98.1%	78.6%
YOLOv9-s	67.3%	98.3%	81.3%
YOLOv10-s	65.8%	97.8%	78.9%
YOLOv11-s	67.1%	98.3%	80.8%
YOLOv12-s	66.4%	98.2%	79.3%
YOLOv13-s	66.5%	98.3%	80.1%
本文方法	67.6%	98.3%	81.4%

加了模型的计算负担。未来研究将致力于通过模型剪枝、知识蒸馏等技术实现轻量化优化,同时探索边缘协同计算机制,构建低计算开销与高推理速度的检测框架,以满足无人机航拍视频的端到端实时处理与精度需求。

参考文献 (References)

- [1] CORTES C, VAPNIK V. Support-vector networks[J]. Machine Learning, 1995, 20(3): 273-297.
- [2] HUNT E B, MARIN J, STONE P J. Experiments in induction[M]. New York: Academic Press, 1966.
- [3] CAO X B, WU C X, YAN P K, et al. Linear SVM classification using boosting HOG features for vehicle detection in low-altitude airborne videos[C]//Proceedings of the 2011 18th IEEE International Conference on

- Image Processing. Brussels, Belgium: IEEE, 2011: 2421-2424.
- [4] WEN X Z, SHAO L, FANG W, et al. Efficient feature selection and classification for vehicle detection [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2015, 25(3): 508-517.
- [5] RIOS-CABRERA R, TUYTELAARS T, VAN GOOL L. Efficient multi-camera vehicle detection, tracking, and identification in a tunnel surveillance application [J]. Computer Vision and Image Understanding, 2012, 116(6): 742-753.
- [6] OTHMANI M. A vehicle detection and tracking method for traffic video based on faster R-CNN [J]. Multimedia Tools and Applications, 2022, 81(20): 28347-28365.
- [7] YIN G X, YU M, WANG M, et al. Research on highway vehicle detection based on faster R-CNN and domain adaptation [J]. Applied Intelligence, 2022, 52(4): 3483-3498.
- [8] ARORA N, KUMAR Y, KARKRA R, et al. Automatic vehicle detection system in different environment conditions using Fast R-CNN [J]. Multimedia Tools and Applications, 2022, 81(13): 18715-18735.
- [9] LUO J Q, FANG H S, SHAO F M, et al. Multi-scale traffic vehicle detection based on faster R-CNN with NAS optimization and feature enrichment [J]. Defence Technology, 2021, 17(4): 1542-1554.
- [10] 张莹, 刘子龙, 王伟. 基于 Faster R-CNN 的无人机车辆目标检测 [J]. 电子科技, 2021, 34(11): 11-20.
- ZHANG Y, LIU Z L, WAN W. UAV vehicle target detection based on faster R-CNN [J]. Electronic Science and Technology, 2021, 34(11): 11-20.
- [11] JIANG P Y, ERGU D, LIU F Y, et al. A review of yolo algorithm developments [J]. Procedia Computer Science, 2022, 199: 1066-1073.
- [12] ZHU X K, LYU S C, WANG X, et al. TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. Montreal, Canada: IEEE, 2021: 2778-2788.
- [13] QIU Z F, BAI H H, CHEN T Y. Special vehicle detection from UAV perspective via YOLO-GNS based deep learning network [J]. Drones, 2023, 7(2): 117.
- [14] FARID A, HUSSAIN F, KHAN K, et al. A fast and accurate real-time vehicle detection method using deep learning for unconstrained environments [J]. Applied Sciences, 2023, 13(5): 3059.
- [15] 江德港, 江智, 黄子杰, 等. 基于 Efficientnet 的无人机车辆目标检测算法 [J]. 计算机工程与应用, 2023, 59(12): 228-234.
- JIANG D G, JIANG Z, HUANG Z J, et al. UAV vehicle object detection algorithm based on efficientnet [J]. Computer Engineering and Applications, 2023, 59(12): 228-234.
- [16] 张利丰, 田莹. 改进 YOLOv8 的多尺度轻量级车辆目标检测算法 [J]. 计算机工程与应用, 2024, 60(3): 129-137.
- ZHANG L F, TIAN Y. Improved YOLOv8 multi-scale and lightweight vehicle object detection algorithm [J]. Computer Engineering and Applications, 2024, 60(3): 129-137.
- [17] BAKIRCI M. Enhancing vehicle detection in intelligent transportation systems via autonomous UAV platform and YOLOv8 integration [J]. Applied Soft Computing, 2024, 164: 112015.
- [18] OU K T, DONG C J, LIU X K, et al. Drone-TOOD: A lightweight task-aligned object detection algorithm for vehicle detection in UAV images [J]. IEEE Access, 2024, 12: 41999-42016.
- [19] ZHANG Y, CHEN X B, SUN S, et al. Vehicle detection in drone aerial views based on lightweight OSD-YOLOv10 [J]. Scientific Reports, 2025, 15(1): 25155.
- [20] 白俊卿, 王梦婷, 沈守婷. 基于 YOLOv5 的无人机遥感图像车辆小目标检测算法 [J]. 科学技术与工程, 2025, 25(12): 5110-5118.
- BAI J Q, WANG M T, SHEN S T. Vehicle small target detection algorithm for UAV remote sensing images based on YOLOv5 [J]. Science Technology and Engineering, 2025, 25(12): 5110-5118.
- [21] DU H, WANG J Z, LIU M, et al. SwinPA-net: Swin transformer-based multiscale feature pyramid aggregation network for medical image segmentation [J]. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35(4): 5355-5366.
- [22] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection [EB/OL]. arXiv, 2020 [2025-11-11]. <https://arxiv.org/abs/2004.10934>.
- [23] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016: 779-788.
- [24] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: Common objects in context [C] // European Con-

- ference on Computer Vision. Cham; Springer International Publishing, 2014: 740–755.
- [25] SHI Z C, HU J, REN J, et al. HS-FPN: High frequency and spatial perception FPN for tiny object detection [C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2025, 39(7): 6896–6904.
- [26] YU Z P, HUANG H B, CHEN W J, et al. YOLO-FaceV2: A scale and occlusion aware face detector[J]. Pattern Recognition, 2024, 155: 110714.
- [27] SONG J N, ZHOU M L, LUO J, et al. Boundary-aware feature fusion with dual-stream attention for remote sensing small object detection [J]. IEEE Transactions on Geoscience and Remote Sensing, 2025, 63: 5600213.
- [28] WANG C Y, BOCHKOVSKIY A, LIAO H M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 7464–7475.
- [29] LI C Y, LI L L, GENG Y F, et al. YOLOv6 v3.0: A full-scale reloading[EB/OL]. arXiv, 2023[2025-11-11]. <https://arxiv.org/abs/2301.05586>.
- [30] WANG C Y, YEH I H, MAPK L H Y. YOLOv9: Learning what you want to learn using programmable gradient information [C]//European Conference on Computer Vision. Cham; Springer Nature Switzerland, 2024: 1–21.
- [31] WANG A, CHEN H, LIU L H, et al. YOLOv10: Real-time end-to-end object detection[EB/OL]. arXiv, 2024 [2025-11-11]. <https://arxiv.org/abs/2405.14458>.
- [32] TIAN Y J, YE Q X, DOERMANN D. YOLOv12: Attention-centric real-time object detectors [EB/OL]. arXiv, 2025 [2025-11-11]. <https://arxiv.org/abs/2502.12524>.
- [33] LEI M Q, LI S Q, WU Y H, et al. YOLOv13: Real-time object detection with hypergraph-enhanced adaptive visual perception [EB/OL]. arXiv, 2025 [2025-11-11]. <https://arxiv.org/abs/2506.17733>.
- [34] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: Visual explanations from deep networks via gradient-based localization [C]//Proceedings of the IEEE International Conference on Computer Vision. 2017: 618–626.

(责任编辑:唐慧 郭芸婕)