

基于改进 YOLOv10 的航拍车辆目标检测

周 鹏¹ 院老虎^{1*}

ZHOU Peng YUAN Laohu

摘 要

针对无人机视角下背景复杂、车辆尺寸易变导致现有目标检测算法精度偏低的问题,文章提出了一种改进型目标检测算法 DEW-YOLOv10n。为提升模型特征提取能力,首先在 YOLOv10n 网络的 C2f 结构中嵌入 ECA 注意力机制,并采用多分支 DBB 模块替换原有 Bottleneck 结构;其次,在 Neck 部分构建 BIFPN-P2 小目标特征增强金字塔,强化模型对小目标特征的提取与融合效能;最后,引入 WASIoU 损失函数,解决复杂背景下锚框扩展难题,提升目标回归精度。实验结果表明,相较于基线网络 YOLOv10n,所提 DEW-YOLOv10n 算法的检测精确度 P 提升 2.99%, $mAP@0.5$ 提升 4.31%,有效改善了无人机视角下车辆目标的检测性能。

关键词

深度学习; 车辆检测; YOLOv10; 多尺度目标; 复杂背景

doi: 10.3969/j.issn.1672-9528.2026.02.004

0 引言

近年来,深度学习发展迅速,基于深度学习的目标检测成为学者们的研究热点^[1]。但目前已有算法常受限于复杂的场景和目标尺寸大小易变化等因素,对航拍的检测效果不够理想。因此,提升基于无人机视觉的多尺度目标及复杂背景的检测性能,在预防和应对城市交通拥堵方面具有重要的意义。

目前基于深度学习的方法主要分为一阶段和两阶段算法两大类,两阶段算法先生成候选框然后再进行分类与回归,代表模型有 R-CNN 系列^[2-3]。一阶段算法直接提取特征进行分类,代表模型有 SSD^[4]、YOLO^[5-10] 系列。相比之下,一阶段算法的检测速度更快,因此本文使用一阶段算法。Hu 等^[11]使用基于 VGG-16 的 Faster R-CNN 对移动目标的运动特征和视觉特征进行关联,降低漏检率与误检率。蒋康等^[12]设计 REPELAN 采用多支路来优化特征提取和聚合表现,并在主干网络中添加 CA 注意力机制,实现减少模型的参数量的同时,关注特征的重要信息。Sun 等^[13]提出一种轻量高效的双重上下文解析网络,有效提升了无人机图像中车辆与道路信息的识别率,但在面对尺度变化剧烈或背景复杂区域时仍存在一定精度不足的问题。周飞等^[14]通过在 YOLOv8 中更换 Wise-IoU 损失函数,采用 FasterNet 替换骨干特征提取网络,改善了小目标车辆检测效果不佳的问题。上述模型性能上有所提高,但是在进行航拍车辆检测时的主要困难是车辆遮挡和尺度变化,这可能导致在检测时精度低,出现漏检、误检。针对这些问题,本文通过改进 YOLOv10n 算法来进行车辆检测。

1 模型概述

YOLOv10 于 2024 年 5 月由清华大学的研究团队发布,相比之前的 YOLO 版本有进一步的提升。本文以轻量化 YOLOv10n 版本为基线模型,其架构主要分为骨干网络、颈部网络和检测头,分别负责特征提取、特征层融合以及分类和定位预测输出。但基线网络在应对航拍背景复杂,多尺度目标时仍有不足,本文提出 DEW-YOLOv 架构(如图 1 所示)。

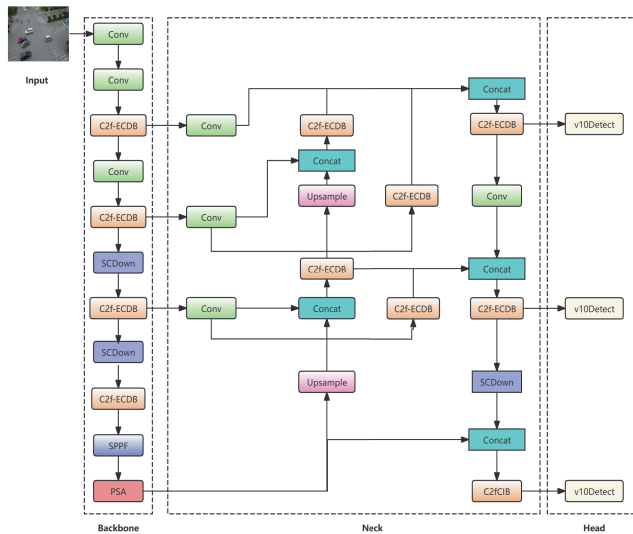


图 1 DEW-YOLOv10n 网络结构图

该算法通过嵌入 C2f-ECDB 模块替换基线网络中的 C2f 模块,加强特征能力的提取。引入 BIFPN 架构,利用双向特征融合和权重分配机制提高不同尺度特征融合的效率。并将 WASIoU 损失函数替代 CIoU 损失函数,来提高多尺度目标及复杂背景下的定位能力。

1. 沈阳航空航天大学航空宇航学院 辽宁沈阳 110136

1.1 C2f-ECDB

在 YOLOv10n 模型中, C2f 模块虽通过堆叠卷积与残差连接实现特征提取, 但受限于单一的卷积核尺寸, 对不同尺度、形态的目标特征捕捉能力存在明显局限。具体到航拍车辆检测场景中, 由于目标尺度差异显著 (从小型轿车到大型货车), 且 C2f 模块无法根据目标尺度动态调整感受野, 导致其对多尺度车辆特征的适应性不足, 难以在复杂航拍背景下实现精准检测。

针对这一问题, 本文提出一种结合注意力机制的融合多样化分支 C2f-ECDB 模块, 以替代原来的 C2f 模块, 其网络结构如图 2 所示。



图 2 C2f-ECDB 模块结构图

该模块的核心设计在于引入多样化分支块 (DBB), 通过整合多维度分支结构来增强特征表征能力。DBB 通过结合卷积序列、多尺度卷积 (如 1×1 与 $K \times K$ 卷积组合) 及平均池化等多样化分支, 丰富特征空间的维度。不同分支分别聚焦于目标的细节信息 (小卷积核)、全局轮廓 (大卷积核) 及背景抑制减少冗余 (池化操作), 使单个模块能够同时覆盖多尺度特征需求。DBB 采用“训练-推理分离”的结构设计: 训练阶段利用多分支复杂结构强化特征学习, 通过差异化路径的互补性提升模型对航拍场景中车辆尺度变化、姿态差异的适应能力; 推理阶段则将多分支等效融合为单一卷积层, 在保持推理速度不受影响的前提下, 保留训练阶段学到的丰富特征表达, 如图 3 所示。

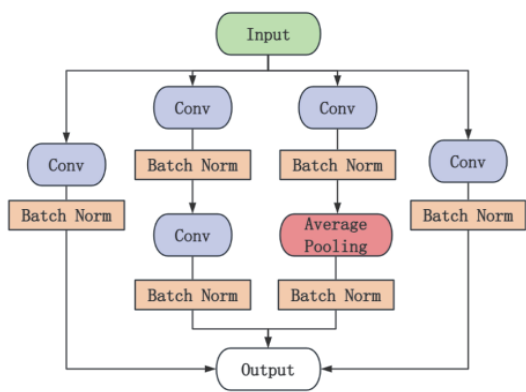


图 3 DBB 模块结构图

1.2 ECA 注意力机制

近年来, 注意力机制模块在深度学习中被广泛使用, 其作用是让模型知道更应该关注哪些内容与位置。然而在轻量级网络中, 注意力机制会受到一定的限制, 原因是轻量级网络担负不起太多注意力机制带来的额外计算, 因此为了提高网络的精度, 本文引入一种几乎没有额外计算的坐标注意力机制 ECA (ECA-Net), 可以对不同通道的特征进行加权,

从而使网络能够更好地捕捉长距离依赖关系, 且参数量小, 计算效率高。将 ECA 注意力机制添加到 C2f-EDBB 中的 Conv 后可直接对刚生成的特征图进行冗余过滤, 避免冗余信息输入到后续分支增加计算量。

ECA 模块如图 4 所示。在通过全局平均池化 (GAP) 获取聚合特征后, ECA 通过执行大小为 k 的快速 1D 卷积 (其中 k 通过通道维度 C 的映射自适应确定) 来生成通道权重。

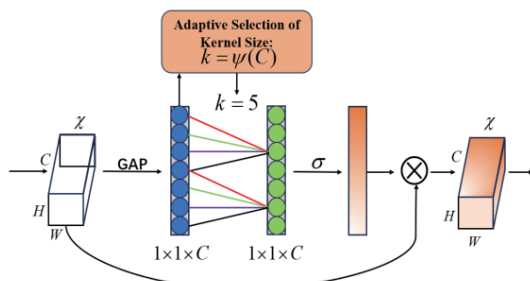


图 4 ECA 模块结构图

1.3 BIFPN 特征金字塔网络

BIFPN 创新集成双向特征交互机制与自适应权重分配策略, 突破传统 FPN 单向信息传递局限。引入该结构, 可通过双向通路 (自底向上 + 自顶向下) 构建特征传递闭环, 让高低层特征实现深度交互, 同时以可学习权重替代固定平均加权, 动态校准各层特征贡献度, 解决传统方法特征融合固定问题。此外, BiFPN 通过精简冗余节点, 在抑制计算开销的同时保留核心特征表达能力, 达成模型轻量化与检测性能的平衡, 为目标检测场景提供高效特征处理能力。如图 5 所示。

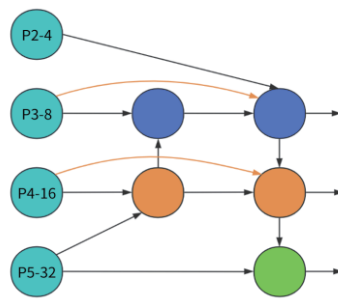


图 5 BIFPN 模块结构图

1.4 WASIoU 损失函数

传统的 IoU 损失函数仅关注边界框的中心点距离、重叠度和宽高比例, 未考虑边界框的方向差异, 而 SIoU 通过角度损失项显式优化预测框与真实框的方向对齐, 尤为适用于航拍图像顺序一致的车辆。SIoU 损失函数进一步考虑了真实框和预测框之间的向量角度, 具体包含 4 个部分: 角度损失 Λ 、距离损失 Δ 、形状损失 Ω 、IoU 损失, SIoU 损失函数的计算公式为:

$$\Lambda = 1 - 2 \cdot \sin^2 \left(\arcsin \left(\frac{c_h}{\sigma} \right) - \frac{\pi}{4} \right) = \cos \left(2 \cdot \left(\arcsin \left(\frac{c_h}{\sigma} \right) - \frac{\pi}{4} \right) \right) \quad (1)$$

$$\Delta = 2 - e^{-\gamma \rho_x} - e^{-\gamma \rho_y} \quad (2)$$

$$\rho_x = \left(\frac{b_{c_x}^{gt} - b_{c_x}}{c_w} \right)^2, \rho_y = \left(\frac{b_{c_y}^{gt} - b_{c_y}}{c_h} \right)^2, \gamma = 2 - \Delta \quad (3)$$

$$\Omega = (1 - e^{-w_w})^\theta + (1 - e^{-w_h})^\theta \quad (4)$$

$$w_w = \frac{|w - w^{gt}|}{\max(w, w^{gt})}, w_h = \frac{|h - h^{gt}|}{\max(h, h^{gt})} \quad (5)$$

$$IoU = \frac{|B_{pre} \cap B_{gt}|}{|B_{pre} \cup B_{gt}|} \quad (6)$$

$$Loss_{SIOU} = 1 - IoU + \frac{\Delta + \Omega}{2} \quad (7)$$

式中： c_h 和 σ 为预测框与真实框中心点的高度差和距离；IoU 为预测框和真实框的交并比。虽然 SIOU 再收敛速度与定位精度优于 CIOU，但 SIOU 在多尺度目标回归中仍有不足，引入尺度形状 Wasserstein 距离进行优化，提出 WASIoU 损失函数。其先利用预测框与真实框的中心偏移误差和宽高误差构建对角协方差矩阵，将边界框几何特征转化为分布特征。

$$p_1 = (x_{p1} - x_{g1})^2 + (y_{p1} - y_{g1})^2 \quad (8)$$

$$p_2 = \frac{(w_p - w_t)^2}{4} + \frac{(h_p - h_t)^2}{4} \quad (9)$$

$$W^2 = (x_{p1} - x_{g1})^2 + (y_{p1} - y_{g1})^2 + \frac{(w_p - w_t)^2}{4} + \frac{(h_p - h_t)^2}{4} \quad (10)$$

式中： (x_{p1}, y_{p1}) 和 (x_{g1}, y_{g1}) 分别为预测框 B_{pre} 和真实框 B_{gt} 的中心点坐标。

通过引入尺度敏感因子对高斯 - Wasserstein 距离进行归一化，得到基于高斯分布假设的紧凑近似形状尺度差异度量，用于刻画尺度与形状差异项。用公式表示为：

$$\gamma_s = \frac{2}{w_p h_p + w_t h_t} \quad (11)$$

$$W_{norm}^2 = \gamma_s \cdot W^2 \quad (12)$$

为增强小目标形状差异的权重，通过设计平滑惩罚项，将差异映射至 $[0,1]$ 区间以避免梯度震荡，用公式表示为：

$$\mathcal{L}_{SSWD} = 1 - \exp(-W_{norm}^2) \quad (13)$$

为动态平衡损失权重，引入自适应系数，用公式表示为：

$$\lambda = \min(1.0, 0.5 + 0.1 \ln(\tau_s + 1)) \quad (14)$$

式中： τ_s 为尺度比例相关因子，实现不同尺度下形状损失与位置损失的合理配比。最终 WASIoU 损失函数为：

$$\mathcal{L}_{WASIoU} = \mathcal{L}_{SIOU} + \lambda \cdot \mathcal{L}_{SSWD} \quad (15)$$

2 实验结果与分析

本实验在 Windows11 操作系统下，使用 Python 3.8，PyTorch 学习框架。实验参数如下：

实验数据集为 Vis-Drone，训练为 250 轮，批次大小为 32，学习率为 0.001，优化器为 SGD。

2.1 消融和对比实验

本文实验按照 8:1:1 的比例将 VisDrone 航拍数据集随机划分成训练集、验证集和测试集。每组实验均使用相同的训练参数，不同改进方法影响模型检测性能的结果如表 1 所示。

表 1 模型消融研究

方法	C2f-ECDB	BIFPN	WASIoU	Precision/%	Recall/%	mAP@0.5/%
YOLOv10n	×	×	×	79.4	71.6	74.2
改进 1	√	×	×	80.87	72.52	75.62
改进 2	√	√	×	81.67	74.16	77.39
DEW-YOLOv10n	√	√	√	82.39	76.64	78.51

注：“√”表示使用此方法，“×”表示未使用此方法

由表 1 可知，原始 YOLOv10n 算法的精确率、召回率以及平均精确率分别为 79.4%、71.6% 和 74.2%。在原始模型的基础上，每改进一处指标都有一定的提升，进一步验证了改进方案的可行性。同时将最终改进后的算法与主流算法进行对比，结果如表 2 所示，所改进的方法优于现流行算法，能有效提升目标检测的整体精度与定位质量。

表 2 模型泛化实验

Models	Precision/%	Recall/%	mAP@0.5/%	mAP@0.5:0.95/%	Parameters/ 10^6
YOLOv8n	77.6	71.4	72.4	49.6	3.01
YOLOv9t	78.7	70.5	73.2	50.2	1.97
YOLOv10n	79.4	71.6	74.2	52.1	2.70
YOLO11n	78.1	72.9	73.4	51.3	2.58
DEW-YOLOv10n	82.39	76.64	78.51	56.2	2.64

2.2 边界框损失函数分析

表 3 展示了在 YOLOv10n Head 中替换不同边界框回归函数后的性能变化，用来对比 WASIoU 损失函数的有效性。

表 3 不同边界框损失函数的性能比较

Baseline Model	Methods	Precision/%	Recall/%
YOLOv10n	CIoU	79.4	71.6
	DIoU	77.8	70.2
	GIOU	78.3	69.4
	EIoU	79.2	70.6
	SIOU	80.5	71.4
	WASIoU	81.2	73.2

相较于 CIoU, 引入方向量的 SIoU 精度有所提升, 表明在车辆检测中引入方向量能够更好地定位。在此基础上, 为解决多尺度目标的问题, 引入尺度形状 Wasserstein 距离, 精确度提升了 0.7%, 表明改进方法能够更加精确地约束边界框地几何条件, 提升模型地整体性能。

2.3 检测效果对比

部分检测效果如图 6 所示。左侧为原始 YOLOv10n 的检测效果图, 右侧为改进 DEW-YOLOv10n 检测效果图, 通过对比可以发现, 在复杂环境和多尺度目标的情况下, 改进的模型改善了漏检、误检的情况, 提升了模型的检测精度。



原 YOLOv10n 结果图 DBW-YOLOv10n 结果图

图 6 检测结果可视化

3 结论

本文以 YOLOv10n 为基线网络提出 DEW-YOLOv10n 检测模型, 实现了对车辆目标的精确检测。与原始的 C2f 相比, C2f-ECDB 不仅关注了信息的冗余, 还通过多分支结构丰富了特征信息, 加强了特征信息的提取。设计 BIFPN-P2 网络结构, 在不额外增加小目标检测头的同时, 利用 P2 特征层, 在不增加计算的同时, 关注了小目标特征。WASIoU 损失函数相较于 CIoU、GIoU 等通过引入方向量及尺度形状 Wasserstein 距离, 更好地关注了多尺度目标, 提升锚框的精度。利用实验结果表明, 改进后的 YOLOv10n 模型检测效果更好, 为基于无人机航拍图像的车辆检测提供了一种有效方法。

参考文献:

- [1] 江波, 屈若锐, 李彦冬, 等. 基于深度学习的无人机航拍目标检测研究综述 [J]. 航空学报, 2021, 27(11):158-129.
- [2] REN S Q, HE K M, 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.
- [3] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic seg-

mentation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2014: 580-587.

- [4] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//Computer Vision-ECCV 2016. Berlin: Springer, 2016: 21-37.
- [5] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2016: 779-788.
- [6] REDMON J, FARHADI A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08)[2025-12-30]. <https://doi.org/10.48550/arXiv.1804.02767>.
- [7] ZHANG Y, GUO Z Y, WU J Q, et al. Real-time vehicle detection based on improved YOLOv5[J]. Sustainability, 2022, 14(19): 12274.
- [8] WANG C Y, YEH I H, MARK LIAO H Y. YOLOv9: learning what you want to learn using programmable gradient information[EB/OL]. (2024-02-29)[2025-12-30]. <https://doi.org/10.48550/arXiv.2402.13616>.
- [9] WANG A, CHEN H, LIU L H, et al. YOLOv10: real-time end-to-end object detection[J]. Advances in neural information processing systems, 2024, 37: 107984-108011.
- [10] TIAN Y J, YE Q X, DOERMANN D. YOLOv12: attention-centric real-time object detectors[EB/OL]. (2025-02-18)[2025-12-30]. <https://www.semanticscholar.org/paper/YOLOv12%3A-Attention-Centric-Real-Time-Object-Tian-Ye/ae1d5360f2f556139cfd10d6e9d2e0241c937e0>.
- [11] HU H G, ZHOU L L, GUAN Q, et al. An automatic tracking method for multiple cells based on multi-feature fusion[J]. IEEE access, 2018, 6:69782-69793.
- [12] 蒋康, 孙仁云, 李卓霖, 等. 改进 YOLOv8 的轻量化车辆目标检测算法研究 [J]. 重庆理工大学学报 (自然科学), 2025, 39(7): 35-42.
- [13] SUN Y M, SHAO Z F, CHENG G, et al. Road and car extraction using UAV images via efficient dual contextual parsing network[J]. IEEE transactions on geoscience and remote sensing, 2023:102-110.
- [14] 周飞, 郭杜杜, 王洋, 等. 基于改进 YOLOv8 的交通监控车辆检测算法 [J]. 计算机工程与应用, 2024, 60(6):110-120.

【作者简介】

周鹏 (1999—), 男, 辽宁东港人, 硕士研究生, 研究方向: 深度学习、目标检测。

院老虎 (1978—), 通信作者 (email: ylhhit@126.com), 男, 山西太原人, 博士, 副教授、硕士生导师, 研究方向: 深度学习、飞行器控制。

(收稿日期: 2025-07-25 修回日期: 2026-02-02)