

雾天场景下基于 RT-DETR 的轻量化车辆检测算法



黎鹏¹, 刘阳², 胡玛莉², 隆伟¹, 朱洲森², 廖雪花^{1,*}

¹ 四川师范大学计算机科学学院, 四川成都 610101

² 四川师范大学物理与电子工程学院, 四川成都 610101

摘要: 针对雾天场景下车辆检测任务中, 检测精度与轻量化难以平衡的问题, 提出了一种基于 RT-DETR-r18 的改进雾天车辆检测算法, 该算法克服了模型的计算瓶颈, 首先, 设计了一种改进的主干网络 (C2F_DYN CB) 模块, 结合动态卷积权重 (DKW) 机制和参数共享 (DyInConv), 构建了一种新型的动态卷积单元 (DAIN Mixer)。该模块能够根据输入特征自适应调节深度卷积与特征融合方式, 在显著减少参数量和计算量的同时, 仍保持对多尺度空间特征的有效建模能力。其次, 提出 PEMD 模块, 通过引入 Pola Attention 线性注意力机制与 EDFFN 结构, 在保持线性复杂度的同时大幅提高注意力图的判别性和模型性能; 然后设计一个轻量化适配块 MOEN 模块, 用于进一步整合局部信息并稳定特征分布。最后, 采用 WIoU-v3 作为回归损失, 通过在训练过程中自适应调整正负样本权重, 增加对细微框的关注。实验结果表明, 在 REFYG 雾天数据集上, 改进算法的 mAP50 比原算法提高了 3.05 个百分点, 模型的计算量与参数量分别降低了 39.4% 和 32.8%, 实现了显著的轻量化。

关键词: RT-DETR; 雾天检测; 轻量化模型; 注意力机制; 动态卷积

DOI: [10.57237/j.se.2026.01.001](https://doi.org/10.57237/j.se.2026.01.001)

Lightweight Vehicle Detection Algorithm Based on RT-DETR in Foggy Weather Scenarios

Li Peng¹, Liu Yang², Hu Mali², Long Wei¹, Zhu Zhoushen², Liao Xuehua^{1,*}

¹ School of Computer Science, Sichuan Normal University, Chengdu 610101, China

² School of Physics and Electronic Engineering, Sichuan Normal University, Chengdu 610101, China

Abstract: Addressing the difficulty in balancing detection accuracy and model lightweighting in vehicle detection tasks under foggy conditions, this paper proposes an improved foggy vehicle detection algorithm based on RT-DETR-r18. This algorithm overcomes the computational bottleneck of the model. First, an improved backbone network (C2F_DYN CB) module is designed, combining a dynamic convolution weight (DKW) mechanism and parameter sharing (DyInConv) to construct a novel dynamic convolution unit (DAIN Mixer). This module enables adaptive depth convolution and feature fusion, significantly reducing the number of parameters and computations while still maintaining detailed processing of input feature maps. Second, a PEMD module is proposed, introducing the linear attention mechanism Pola Attention and EDFFN, which significantly improves the discriminative power of the attention map and model performance while

基金项目: 国家社会科学基金一般项目 (20BMZ092).

*通信作者: 廖雪花, liao_xuehua@sicnu.edu.cn

收稿日期: 2025-12-11; 接受日期: 2026-01-13; 在线出版日期: 2026-02-03

<http://www.sciandeng.com>

maintaining linear complexity. Then, a lightweight adaptation block, the MOEN module, is designed to further integrate local information and stabilize feature distribution. Finally, WIoU-v3 is used as the regression loss, adaptively adjusting the weights of positive and negative samples during training to increase attention to subtle bounding boxes. Experimental results show that on the REFYG foggy dataset, the improved algorithm's mAP50 is 3.05 percentage points higher than the original algorithm, and the model's computational complexity and number of parameters are reduced by 39.4% and 32.8%, respectively, achieving significant lightweighting.

Keywords: RT-DETR; Foggy Weather Detection; Lightweight Model; Attention Mechanism; Dynamic Convolution

1 引言

随着自动驾驶和智慧交通的快速发展，车辆检测在全天候感知系统中发挥着至关重要的作用。然而在低照度、雾天等恶劣天气下，受能见度下降、目标特征被遮蔽及背景干扰增强等影响，现有检测算法的性能显著下降，增加了自动驾驶感知与决策的不确定性，带来潜在安全隐患。与此同时，许多针对复杂场景提出的深度学习检测模型虽能在一定程度上提升精度，但普遍存在结构臃肿、参数量庞大、推理速度慢等问题，难以满足自动驾驶边缘设备对实时性与轻量化的严格要求。

近年来，基于深度学习的目标检测算法取得了长足进展。两阶段检测方法（如 R-CNN、Faster R-CNN 等）[1]在精度方面表现良好，但由于需要生成候选框和复杂的特征处理流程，计算开销大、推理速度慢，不利于在资源受限的场景部署。单阶段检测算法（如 YOLO 系列、SSD、EfficientDet 等）具有端到端的优势，能显著提升计算效率和检测速度，但在低照度和雾天场景下的检测精度仍有待提高。与此同时，Transformer [2]结构凭借其全局上下文建模能力逐渐应用于目标检测任务，百度提出的实时目标检测模型 RT-DETR [3]通过高效混合编码器与 IoU 感知查询机制实现了较好的实时性，成为轻量化端到端检测的重要探索。但在低照度和雾天车辆检测中，RT-DETR 仍面临小目标特征难以捕获、背景复杂以及计算量较大的问题。

针对上述挑战，本文在继承 RT-DETR 端到端检测优势的基础上，提出一种轻量化雾天车辆检测算法。通过精简主干网络、优化特征建模与注意力机制，实现了结构的深度轻量化与多尺度特征增强，在显著减少计算量和参数量的同时，仍能在低能见度场景下保持优异的检测精度与实时性能。本文的主要创新如下。

- 1) 提出了一种基于 C2f 动态卷积模块的改进网络结构，设计了 DyInMixer 模块，通过多尺度

动态卷积实现更高效的特征融合与特征表达 [4]。该设计能够充分捕捉雾天等低能见度条件下目标轮廓模糊、尺度不均匀等复杂特征，从而显著提升目标的可分辨性与识别精度。同时，模块中结合 LayerScale 与 DropPath 技术，在增强梯度传播稳定性、避免训练初期梯度不平衡的基础上，有效提升了模型的收敛速度和泛化能力。

- 2) 在改进编码器层的基础上，引入 PolarLinearAttention [5]与 EDFFN [6]模块，以强化全局上下文捕捉与频域特征补充能力，并在归一化环节采用动态 Tanh 函数（DynamicTanh）[7]以获得更平滑稳定的梯度传播。在此基础上，设计了轻量化的 MOEN 模块，用于在注意力与前馈网络之间进行特征适配与重组，通过通道归一化、非线性激活与自定义卷积操作，MOEN 模块能有效提升特征表达的紧凑性与判别性，减少冗余计算，从而进一步改善低能见度场景下的目标检测性能。
- 3) 在边界框回归部分，本文采用 WIoU v3 [8]损失函数。该损失函数通过引入动态权重机制，对不同回归要素进行自适应平衡，有效缓解了定位偏差与长宽不一致对训练过程带来的不稳定性，使得目标框学习过程更加平滑和精确，从而进一步提升检测模型在雾天、低照度等复杂场景下的整体性能与鲁棒性。

2 RT-DETR 的相关理论

RT-DETR 的整体结构可划分为三个核心模块：主干网络、高效混合编码器以及解码器（整体架构如图 1 所示）。主干网络用于从输入图像中提取多尺度特征。

混合编码器在设计中融入了卷积操作，以强化局部特征感知，并在维持少量 Transformer 层的基础上，集成多尺度特征信息。该设计能够在控制计算复杂度的前提下，有效保持模型对全局上下文的建模能力。解码器部分采用了 IoU 感知的查询选择机制，能够从编码器输出中筛选出高质量的候选检测框，进而通过匈牙利

算法实现端到端的目标匹配与预测。这一机制避免了对传统非极大值抑制（NMS）后处理的依赖。

通过全局注意力建模与轻量化设计，RT-DETR 在精度与速度之间取得了良好平衡，并允许灵活调整编码器深度以适应不同实时或边缘计算场景，为资源受限环境下的高效部署提供了可行方案。

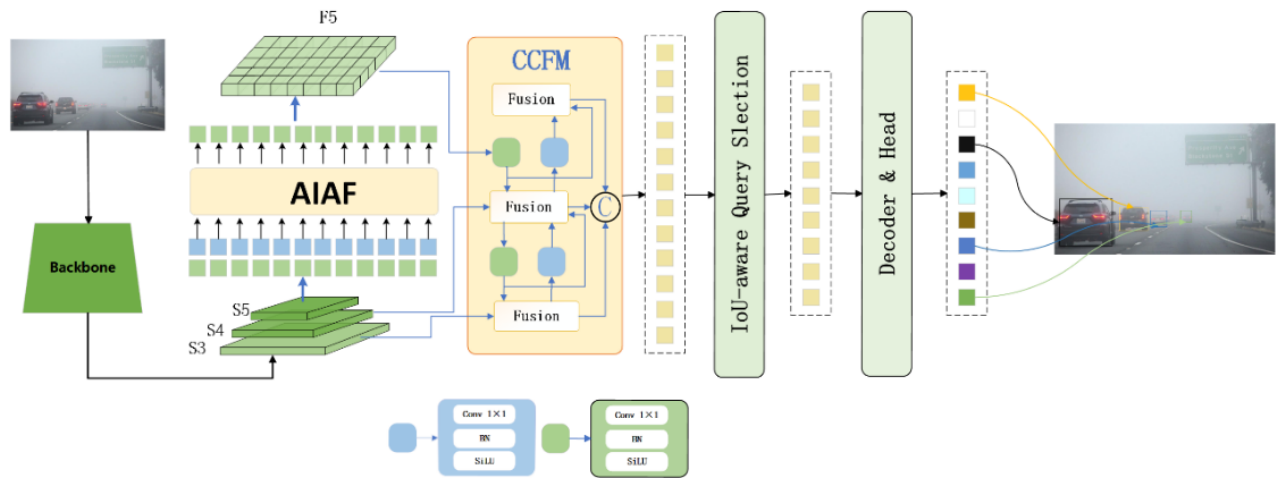


图 1 RT-DETR 网络结构

2.1 改进的 RT-DETR 网络结构

本文针对雾天车辆检测任务，构建了一种改进的 RT-DETR-R18 模型，其总体框架如图 2 所示。主干网络引入动态混合卷积模块 C2f_DYN CB，以增强多尺

度特征提取能力；编码器部分通过结合线性注意力机制 PolaAttention 与轻量化适配块 MOEN，设计出高效的 PEMD 结构，实现特征融合与判别能力的提升；在边界框回归过程中，采用 WIoU v3 作为定位损失函数，自适应优化正负样本权重，并强化对细粒度目标框的回归精度。

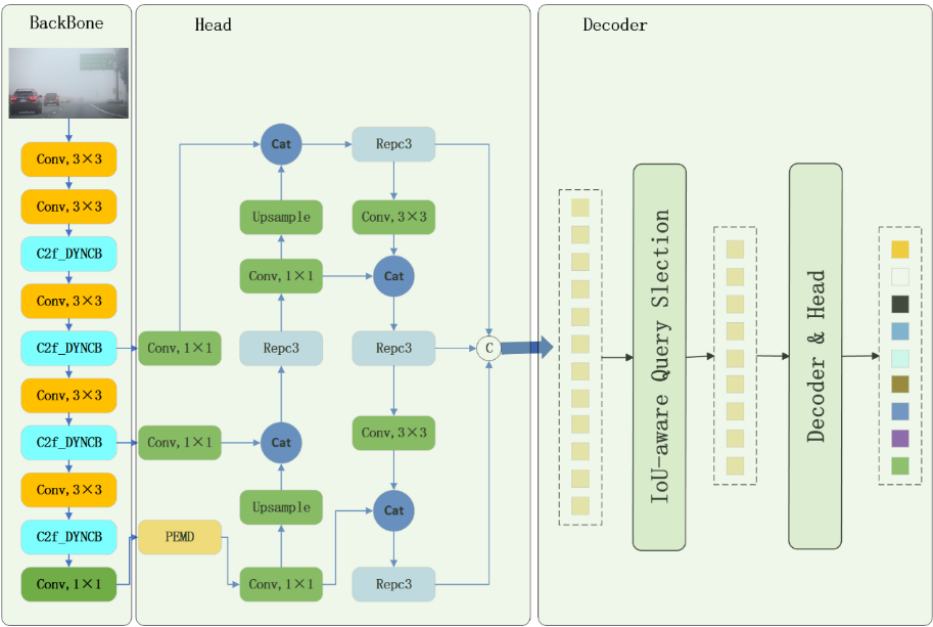


图 2 改进算法整体结构

2.2 C2F_DYN CB 主干网络改进

在雾天车辆检测任务中, 不同层次的特征图信息融合不足, 上下层语义信息未能充分利用, 从而限制了模型在复杂环境下的表现[9, 10]。此外, 网络卷积操作缺乏灵活性, 无法根据输入特征自适应调整, 降低了模

型在多变场景下的鲁棒性与适应性; 同时, 固定卷积核及冗余计算带来的参数量和计算量较大, 限制了模型的推理速度和部署效率, 难以满足车载端或边缘设备对实时性的需求[11]。针对上述问题, 本文在主干网络中引入 C2F_DYN CB 改进模块如图 3 所示。

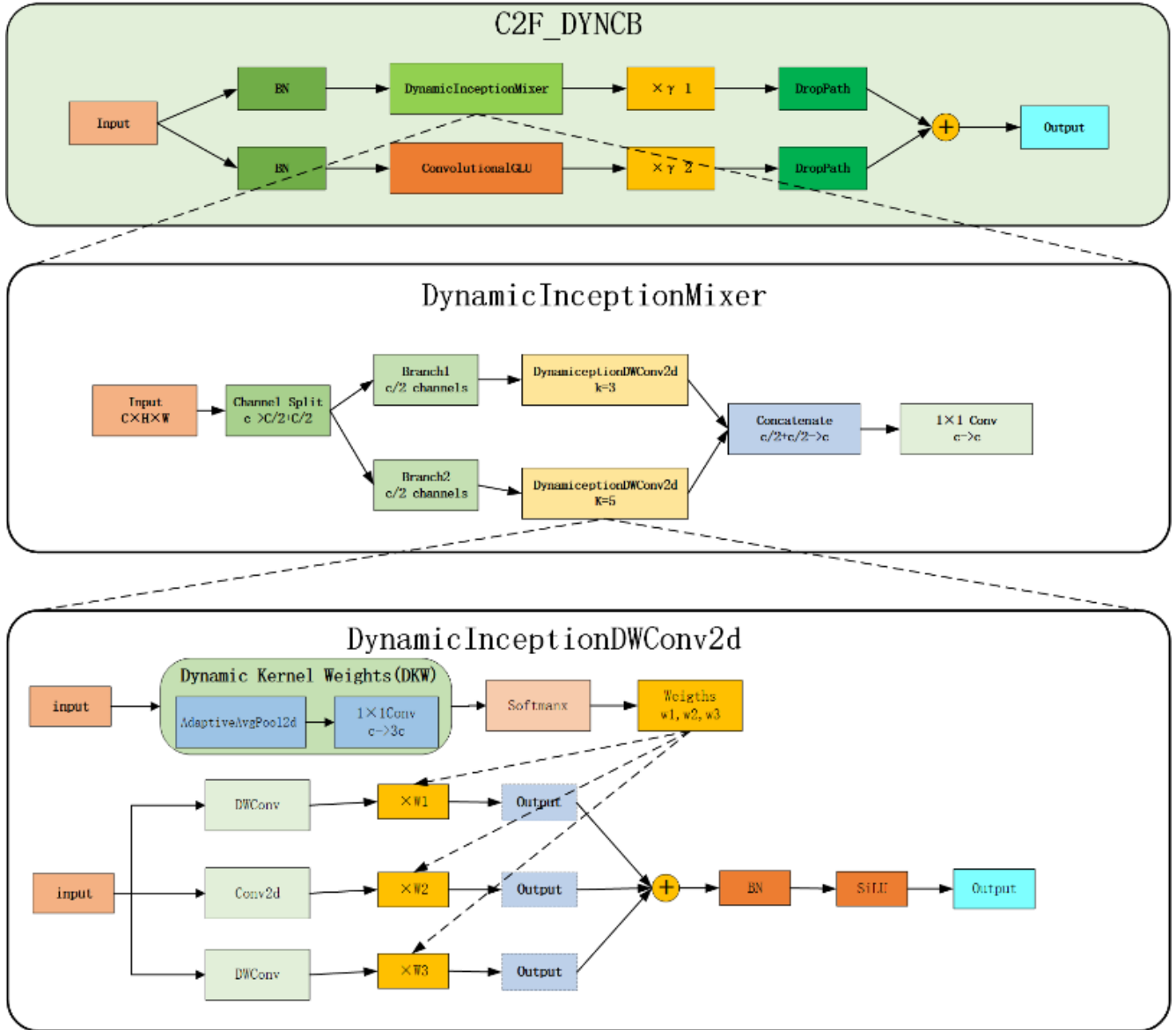


图 3 C2F_DYN CB 模块结构图

改进后的主干网络由 4 个 DynamicIncMixerBlock 级联组成, 每个 Block 内包括: 多尺度卷积核提取与融合 (M) 和一个局部特征增强与门控机制 (G), 整体流程公式如(1), (2)所示。

$$y_1 = x + \text{DropPath}(\gamma_1 \square M(BN_1(x))) \quad (1)$$

$$y_2 = y_1 + \text{DropPath}(\gamma_2 \square G(BN_2(y_1))) \quad (2)$$

其中 BN 为批量归一化, γ 为可学习的层缩放系数, DropPath 表示随机深度正则化, 用于在训练过程中随机丢弃部分路径以增强模型泛化能力, 在 M()中, 输入特征先在通道维分组, 再分别通过不同核尺寸的

动态深度卷积 (DWConv) 提取多尺度上下文信息, 最后拼接并用 1×1 的卷积融合, 如公式 3~6 所示。

$$X = [X_1, X_2] \quad X_1, X_2 \in \mathbb{R}^{B \times \frac{C}{2} \times H \times W} \quad (3)$$

$$Z_i = D_{ki}(X_i) \quad i=1,2 \quad (4)$$

$$Y = \text{Conv}_{1 \times 1}([Z_1, Z_2]) \quad (5)$$

其中 D_{ki} 表示动态深度卷积 (InDWConv), 具体公式如(6)~(8)所示。

$$w = \text{Soft max}(\text{Conv}_{1 \times 1}(\text{GAP}(X))) \in \mathbb{R}^{3 \times C} \quad (6)$$

$$Y = \sum_{i=1}^3 w_i \square (DWConv_{ki}(X)) \quad (7)$$

$$Z = \text{SiLU}(\text{BN}(Y)) \quad (8)$$

其中 $DWConv(ki)$ 分别为方形卷积 $K(s) \times K(s)$ 、条形卷积 $1 \times K(b)$ 与 $K(b) \times 1$, GAP 为全局平均池化, $W(i)$ 为动态权重。通过动态卷积核选择、跨层信息融合、以及高效的计算设计, DyInMixer 能够有效解决传统卷积

神经网络在多尺度特征提取、信息融合及计算效率方面的不足。该设计不仅提升了网络的适应性和灵活性, 还通过实验结果展示了更低的参数量和计算量以及高效的计算性能。

2.3 AIFI 改进

在 RT-DETR 模型中, AIFI 结构在标准场景下有良好的特征交互能力, 但在低照度与恶劣天气场景下, 由于 AIFI 的前馈网络主要基于点卷积和深度卷积, 无法有效提取被雾气、雨水、暗光遮挡后的频域特征导致空间频域建模能力的不足, 其次多头注意力模块偏重长距离依赖, 对局部结构和细粒度纹理 (车灯、车身轮廓) 捕捉能力不强, 且标准 LayerNorm 在不同光照或噪声水平下参数固定, 难以自适应复杂场景, 针对上述问题, 在 AIFI 结构的基础上引入极化线性注意力 (Pola Linear Attention)、增强型双域前馈网络 (EDFFN) 和设计的多通道非线性适配模块 MOEN, 构建了 PEMD 模块, 以提升特征的多尺度建模能力、上下文对齐和光照以及雾霾适应性, 模型结构图如 4 所示。

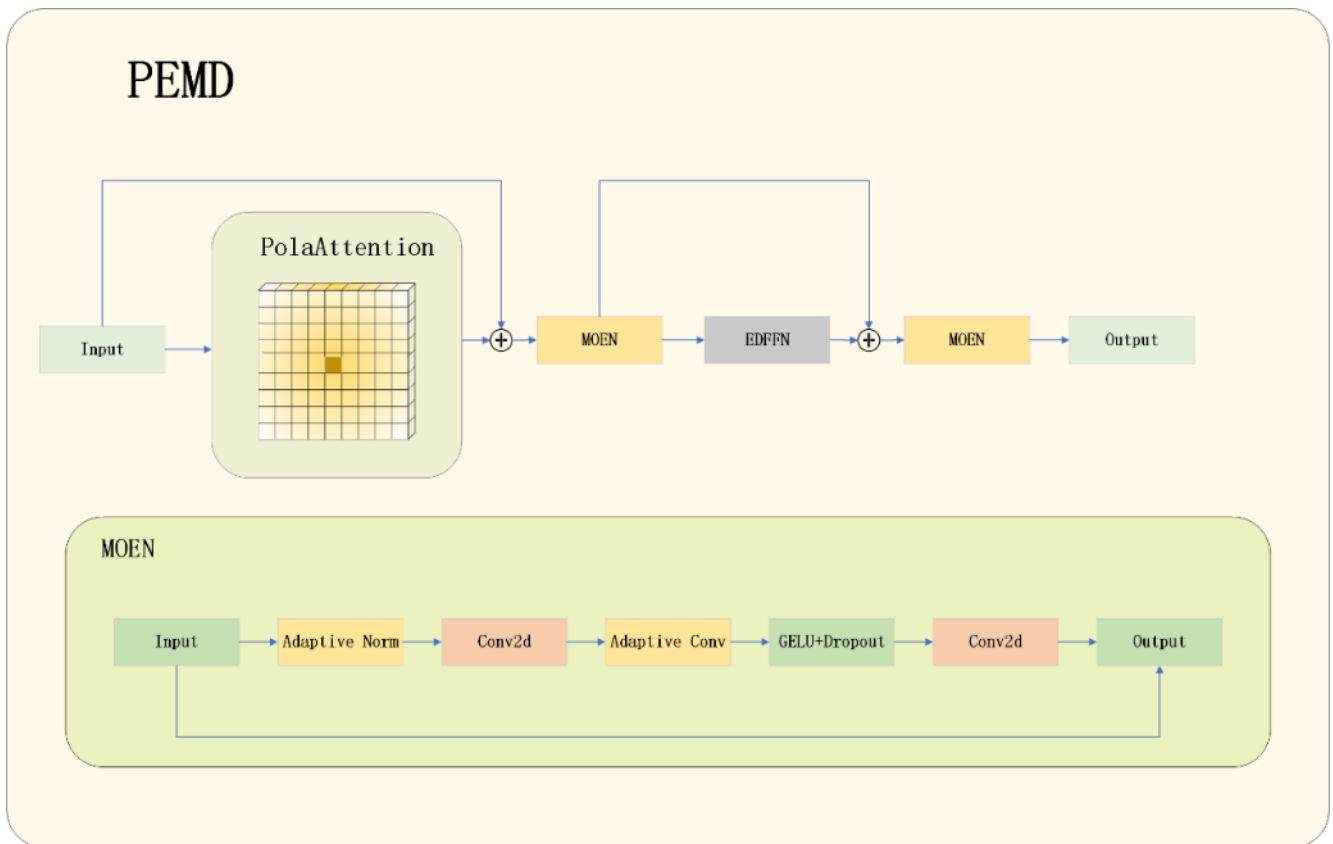


图 4 PEMD 模块结构图

PolaAttention 是一种线性多头注意力机制, 该注意力机制旨在通过核函数线性化来降低标准 Softmax 注意力的二次复杂度并利用 ReLU 激活将 Q 和 K 划分为正负极性分量 (Q_{pos}, Q_{neg}), 并引入自适应幂参数 P ($P=1+\alpha \cdot \text{Sigmoid}(\text{power})$) 来调节线性近似的准确性。同时, 通过深度卷积 DWConv 将局部特征注入到注意力输出中, 如公式(9)所示。

$$Q_{pos=\sigma}(Q)^P, K_{pos=\sigma}(K)^P, Q_{neg=\sigma}(-Q)^P, K_{neg=\sigma}(-K)^P \quad (9)$$

$$\text{Attention}(Q, K) \approx \frac{(Q_{pos}^P K_{pos}^T + Q_{neg}^P K_{neg}^T) V}{\text{Normalization}} + \text{DWConv}(V) \quad (10)$$

其中 $\sigma(\cdot)$ 为激活函数, $p=1+\alpha \cdot \text{sigmoid}(P)$ 为可学习幂次参数。最终将正、负分量拼接后计算注意力, 通过该机制, 模型在保持线性复杂度的同时提升了对雾天场景局部细节的敏感性。

其次, 设计了一个 MOEN 模块, 用于在注意力和 FFN 之间引入通道重标定和适应性变换。它先对输入进行 LayerNorm 和可学习缩放, 模块对输入特征 x 进行了初始转换。这一步结合了 LayerNormal 和两个可学习的尺度参数 γ 和 γ_x 。其数学表达如公式 11 所示。

$$X' = \text{LayerNorm}_{2d}(X) \odot \gamma + X \odot \gamma_x \quad (11)$$

其中, x 代表输入的特征图, $\odot$ 为元素级乘法。模块首先将归一化后的特征与原始特征按可学习比例进行融合, 使模型在训练过程中能够动态调节特征尺度与偏移, 从而在加速收敛的同时减弱训练初期残差连接带来的负面影响, 然后, 利用 1×1 卷积对特征进行降维, 并通过自定义的 MonaOp 完成通道级特征重组与增强, 以实现轻量化的多尺度适配。在此基础上, 引入 GELU 激活函数提升非线性表达能力, 并结合 Dropout 增强模型的泛化能力。最后投影回原通道维度, 并加残差, 如公式 12~15 所示。

$$X_1 = \text{MonaOp}(\text{Conv}_{1 \times 1}(X')) \quad (12)$$

$$X_2 = \text{Dropout}(\text{GELU}(X_1)) \quad (13)$$

$$X_{out} = \text{Conv}_{1 \times 1}(X_2) \quad (14)$$

$$Y = X + X_{out} \quad (15)$$

为增强特征的多尺度表达能力, 引入扩展频域前馈网络 (EDFFN)。首先利用深度可分离卷积对输入特

征进行通道扩展并分块, 随后在频域内通过快速傅里叶变换 (FFT) 对每个 patch 的频谱进行自适应加权, 其中 $F()$ 表示二维 FFT, Γ 为可学习频域权重。再通过逆变换恢复到时域并拼接得到输出特征, 实现局部细节与全局频率信息的有效融合, 如公式 16 所示。

$$\hat{X}_{FFT} = F(X_{patch}) \odot \Gamma \quad (16)$$

通过上述模块的组合, 改进后的 AIFI 层在特征交互上实现了全局与局部的互补: Pola 线性注意力保证全局信息的高效整合, EDFFN 在频域中强化多尺度特征, MOEN 模块突出细节特征, DynamicTanh 提供动态归一化与激活。多机制协同作用显著提升了模型在雾天场景下对车辆轮廓与细节的提取能力, 同时降低了冗余信息干扰, 为后续检测头提供了更具判别力的特征表示。

2.4 损失函数优化策略

RT-DETR 采用广义交并比损失 (Generalized Intersection over Union, GIoU) 作为边界框回归损失函数, 如公式 17~19 所示。尽管 GIoU 克服了传统 IoU 与 DIoU 的梯度消失问题, 其在复杂场景中仍存在显著局限。

$$IOU = \frac{|A \cap B|}{|A \cup B|} \quad (17)$$

$$GIOU = IOU - \frac{A^c - U}{A^c} \quad (18)$$

$$\text{Loss}_{\text{GIOU}} = 1 - GIOU \quad (19)$$

式中, IoU 为预测框与真实框的交并比; A^c 为最小闭包区; U 为预测框与真实框的并集面积。传统 IoU 损失相比, GIoU 在处理预测框与真实框不重叠的情况下, 能够更好地反映二者的重合程度。然而, 当预测框与真实框的交并比相同但位置不同步时, GIoU 难以区分预测框与真实框的相对位置关系, 导致边界框位置调整不够快速和精确。

针对上述问题, 本文在定位损失函数中引入了动态非单调聚焦机制 WIoU (Wise-IoU), WIoU 采用了合理的梯度分配策略来对锚框进行评估, 能够使其聚焦于普通的锚框, 提升模型的检测性能。WIoU 有 3 个版本, 其中 WIoUv3 是由离群度 β 构造的非单调聚焦系数 r 用在 WIoU v1 上得到的, 离群度 β 用来描述锚

框的质量, β 越小, 锚框的质量越高。当离群度较大时, 对质量较差的锚框分配较小的梯度增益, 能够减少其产生较大的有害梯度。WIoUv3 采用更加合理的梯度增益分配策略, 使模型在训练过程中能够聚焦于普通质量锚框, 进一步提升检测性能。WIoUv3 的计算公式如式 20~21 所示。

$$L_{WIoUv1} = R_{WIoU} L_{IoU} \quad (20)$$

$$R_{WIoU} = \exp\left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 + H_g^2)^*}\right) \quad (21)$$

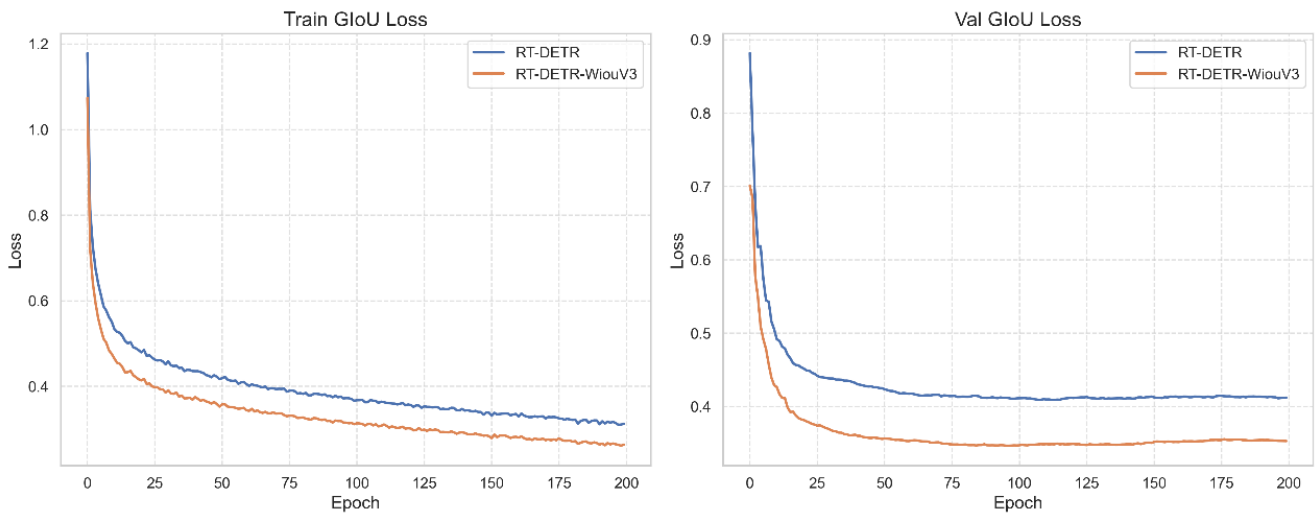


图 5 损失函数比较图

3 实验结果和分析

3.1 实验数据集建立

针对雾天复杂交通场景下的车辆检测问题, 本研究选取了真实雾天图像数据集 RTTS [12] 作为主要研究数据集。RTTS 共包含 4332 张图像, 涵盖汽车、公交车、自行车、摩托车及行人等类别。鉴于研究重点为车辆目标检测, 仅保留包含车辆的图像, 共计 3323 张, 用于模型训练与评估。由于单独使用 RTTS 数据量有限, 难以充分训练深度卷积神经网络以获得满意的检测性能, 本研究进一步整合了 Foggy_Driving [13] 数据集和 DAWN [14] 数据集中雾天部分, 以及部分其他雾天拍摄图像, 生成总计 4147 张真实雾天场景图像。该数据集包含不同时间段的图像, 包括白天和黑夜; 不同的场景, 如高速公路、城市街道和住宅; 涵盖了多种

图 5 展示了不同损失函数下 RT-DETR 模型在训练集和验证集上的 GIoU 定位损失变化情况。蓝色曲线为原始 RT-DETR 模型采用原始定位损失函数时的结果, 橙色曲线为将原损失替换为本文引入的 WIoU-v3 改进损失函数后的结果。从图中可以看出, 在训练与验证过程中, 使用 WIoU-v3 的改进模型整体收敛速度更快, 且在同等训练轮次下的 GIoU Loss 显著低于原始模型, 表明本文提出的改进损失函数能够有效提高模型的定位精度与收敛稳定性。

场景和低中高浓度的真实雾天场景, 满足模型训练和评估需求。

数据集按 7:1.5:1.5 的比例划分为训练集、验证集和测试集。训练集用于优化模型参数, 验证集用于独立评估训练效果并检测过拟合情况, 从而保证模型的泛化能力, 测试集则用于最终性能评测与比较。

最终数据集中保留的主要目标类别包括汽车、公交车、自行车、摩托车和行人, 其中车辆类别为研究主体, 以满足在雾天复杂场景下进行车辆目标检测的研究需求。

3.2 实验环境和训练策略

本文实验使用的硬件环境部署于云端计算平台, 使用 Ubuntu 22.04 LTS 操作系统, 处理器型号为 Intel Xeon Gold 6230 CPU @ 2.10GHz, 输入照片标准化为 640×640, 表 1 包含有关实验设置的特定信息, 使用

AdamW 用于优化训练阶段。它从学习率 0.0001 开始，动量因子 0.9 和权重衰减系数 0.0001，实验未加载预训练权重，确保改进模型的训练与原始模型的一致性，保证实验的公平性和可靠性。

表 1 实验参数及配置

Type	Version	Type	Value
GPU	RTX4090	Epoch	200
Python	3.12.3	Batch	4
CUDA	12.1	Lr0	0.0001
Pytorch	2.3.1	Optimizer	AdamW

3.3 评估指标

为了全面评估本文所提出改进算法的性能，本文选取了多种关键评价指标，包括参数量 (Parameters)、计算量 (GFLOPs)、平均精度均值 (mAP)、精确率 (Precision, P)、召回率 (Recall, R)、平均精度 (Average Precision, AP) 以及帧率 (Frames Per Second, FPS)。其中参数量指模型中可训练参数的总数，直接影响模型的存储需求与显存占用。参数量越大，模型的表达能力通常越强，但也可能带来过拟合风险和更高的训练复杂度。计算量，用于衡量模型在推理过程中所需执行的浮点运算次数，直接影响模型推理速度与计算资源消耗。计算量受特征图尺寸、通道数以及卷积核大小等因素影响。对于实时性要求较高的检测任务，较低的计算量有助于提升帧率 (FPS)，从而保证模型能够快速处理图像并实现及时检测。

精确率 (Precision, P) 衡量被模型判定为目标样本中真实为目标的比例，计算公式如(22)所示。

$$P = \frac{TP}{TP + FP} \tag{22}$$

其中，TP 表示正确检测到的目标数 (真阳性)，FP 表示被错误判定为目标样本数。

召回率 (Recall, R) 是衡量所有真实存在的目标中被模型成功检测出的比例，计算公式如(23)所示。

$$R = \frac{TP}{TP + FN} \tag{23}$$

其中，TP 表示正样本数 (正确检测的目标)，FP 表示负样本数 (错误检测为目标样本)，FN 表示漏检的正样本数。

平均精度均值 (mean Average Precision, mAP) AP 用于表示在特定类别上的平均精度；mAP 是对所有类别 AP 的平均，用以反映模型整体检测性能，计算公式如(24)所示。

$$mAP = \frac{1}{n} \sum_{i=1}^n \int_0^1 AP_i \tag{24}$$

其中，n 表示目标类别总数，AP(i)为第 i 类的平均精度。在该实验中，为了兼顾目标定位评价的不同严格程度，采用 mAP@0.5 作为评价指标，表述其交并比 (IoU) 阈值为 0.5 时计算平均精度，反映模型在较宽松匹配标准下的检测能力。

帧率 (Frames Per Second, FPS) 用于衡量模型在单位时间内能够处理的视频帧 (或图像帧) 数，是评价模型推理速度与实时性的重要指标。其计算公式如(25)所示。

$$FPS = \frac{Frames}{Time} \tag{25}$$

其中，N 表示在推理过程中处理的帧数，T 表示对应的总耗时 (秒)。FPS 值越高，表明模型具有更高的处理速度和更强的实时处理能力。

3.4 消融实验

进行消融研究以确认本文提出的改进方法的有效性，结果总结在 Table 2 中。消融实验分为 8 组，每组保持输入图像、训练超参数等的一致性。其中，C2f_DYN CB、PEMD、WiouV3 是本文提出的改进方法。

表 2 消融实验结果

Model	C2f_DYN CB	PEMD	WiouV3	Precision/%	mAP@0.5(%)	FPS	Paramters	GFLOPS
RTDETR-r18				75.4	68.4	34.0	19.8	57.0
1	✓			80.2	69.9	46.8	13.2	43.6
2		✓		79.0	69.7	146.0	20.0	57.2
3			✓	74.9	70.3	59.4	19.8	57.0
4	✓	✓		76.0	70.2	109.1	13.3	43.8
5	✓		✓	77.1	70.1	50.2	13.2	43.7
6		✓	✓	75.9	69.5	110.6	20.0	57.2
7	✓	✓	✓	77.5	71.5	119.0	13.3	43.8

为验证所提出各个改进模块对雾天车辆检测性能的贡献。实验中分别引入了 C2f_DYN CB 模块、PEMD 模块以及 WIoUv3 损失函数,并通过不同组合方式进行对比分析。由表 2 可见,基础模型(RT-DETR-R18)在 Precision、mAP@0.5 和 FPS 上分别达到 75.4%、68.4% 和 34 FPS,参数量与计算量分别为 19.8M 和 57 GFLOPs。在单独引入 C2f_DYN CB 模块(实验 1)后, Precision 提升至 80.2%, mAP@0.5 提升至 69.9%, FPS 增至 46.8,且参数量和计算量显著降低,表明该模块有效增强了特征提取能力并改善了推理速度。单独引入 PEMD 模块(实验 2)时, mAP@0.5 提升至 69.7%, FPS 大幅提升至 146,显示出对模型实时性的显著改善;单独引入 WIoU 损失(实验 3)则在 mAP@0.5 上取得 70.3% 的最高值,表明其在边界框回归优化方面的优势。在多模块组合方面,引入 C2f_DYN CB+PEMD(实验 4)或 C2f_DYN CB+WIoU(实验 5)均较基础模型取得明显提升,尤其是实验 5 在 Precision (77.1%) 和 mAP@0.5 (70.1%) 上实现双重改善;当 PEMD+WIoU 联合使用(实验 6)时,模型在 FPS 上进一步达到 110.6。最终共同改进结果表明,在同时集成 C2f_DYN CB、

PEMD 和 WIoU 三个模块(实验 7)后,模型在 mAP@0.5 上达到 71.5% 的最高值, Precision 提升至 77.5%, FPS 达 119,且参数量与计算量依旧保持在较低水平(13.3M, 43.8 GFLOPs),各改进模块在提升检测精度、推理速度及降低模型复杂度方面均具有积极作用,且多模块联合使用能够产生协同增益,实现了精度、速度与模型轻量化的综合平衡。充分验证该架构设计的合理性与工程实用性。

3.5 对比实验

本研究在相同硬件配置和软件环境下,对多种主流目标检测算法(SSD [15]、Faster-RCNN [16]、YOLO 系列(YOLOv8n [17]、YOLOv10n [18]、YOLOv11n [19]、YOLOv12n [20])、DETR [21]以及原始 RT-DETR-R18)进行了系统性对比实验。所有模型均采用统一输入尺寸 640×640 进行训练和测试,评价指标包括检测精度(Precision/%)、平均精度均值(mAP@0.5/%)、推理速度(FPS)、模型参数量(Parameters/M)以及计算量(GFLOPs),结果见表 3。

表 3 对比实验

Models	Precision/%	mAP@0.5/%	FPS	Paramters/	GFLOPS
SSD	53.7	63.1	23.1	24.4	275.6
Faster-Rcnn	61.2	67.3	64.1	28.3	401.7
YOLOv8n	72.6	67.6	98.4	3.2	8.6
YOLOv10n	71.3	64.7	111.0	2.3	6.7
YOLOv11n	69.7	66.7	98.4	2.6	6.3
YOLOv12n	72.6	68.4	94.6	2.5	6.3
DETR	65.3	64.1	67.5	40.9	86.2
RE-DETR-r18	75.4	68.4	34.0	19.8	57.0
OURS	77.5	71.5	119.0	13.3	43.8

实验结果表明,本文改进模型在精度与实时性方面均优于原始模型及其他轻量化检测器。在雾天车辆检测任务中,改进模型实现 77.5% Precision 和 71.5% mAP@0.5,分别较原始 RT-DETR-R18 提升 2.1 个百分点和 3.1 个百分点;在 640×640 输入下推理速度达 119 FPS,显著高于原始模型的 34 FPS;同时模型参数量仅 13.3M,计算量 43.8 GFLOPs,保持了良好的轻量化特性。

进一步比较发现,YOLO 系列轻量模型(YOLOv8n、YOLOv8s)虽推理速度较快,但精度略低于本文方法;SSD 与 Faster-RCNN 精度一般且推理速度偏慢;DETR 与 YOLOv10/11/12 系列在精度和速度上较为均衡,但

在计算成本或参数量上存在一定负担。综合精度、速度与计算资源消耗,本文提出的改进模型在雾天车辆检测中实现了更优的性能平衡,可有效满足实际部署对效率和检测精度的需求

4 结论与展望

本文提出一种基于改进 RT-DETR-R18 的雾天车辆检测算法,旨在解决雾天场景中车辆检测精度与模型轻量化之间的矛盾。通过引入 C2f 动态卷积模块并结合 DyInMixer,增强网络多尺度特征提取与融合能力,有效提升在低可见度条件下的目标识别度。改进的 AIFI 结构联合 PolaLinearAttention 和 EDFFN,在保持

轻量化的同时强化全局上下文建模能力, 并通过自主设计的轻量化 MOEN 模块优化注意力与前馈网络之间的特征适配, 从而进一步提高模型在复杂雾天场景下的鲁棒性。边界框回归部分采用 WIoU v3 损失函数, 通过动态权重机制自适应平衡不同回归要素, 缓解定位偏差并提升检测精度。实验结果表明, 改进后的算法在保持高精度的同时显著降低了计算量和参数量, 满足雾天车辆检测对实时性和性能的实际需求。未来工作将聚焦于进一步优化模型的轻量化设计, 尤其面向边缘设备部署, 探索模型剪枝、量化和知识蒸馏等技术以持续提升计算效率。

参考文献

- [1] Girshick R, Donahue J, Darrell T, et al. Region-based convolutional networks for accurate object detection and segmentation [J]. IEEE transactions on pattern analysis and machine intelligence, 2015, 38(1): 142-158.
- [2] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [J]. Advances in neural information processing systems, 2017, 30.
- [3] Zhao Y, Lv W, Xu S, et al. Detrs beat yolos on real-time object detection [C] // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2024: 16965-16974.
- [4] Chen Y, Dai X, Liu M, et al. Dynamic convolution: Attention over convolution kernels [C] // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 11030-11039.
- [5] Meng W, Luo Y, Li X, et al. PolaFormer: Polarity-aware linear attention for vision transformers [J]. arXiv preprint arXiv: 2501.15061, 2025.
- [6] Ciubotariu G, Vasluianu F A, Zhou Z, et al. AIM 2025 Challenge on High FPS Motion Deblurring: Methods and Results [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision. 2025: 5541-5550.
- [7] Zhu J, Chen X, He K, et al. Transformers without normalization [C] // Proceedings of the Computer Vision and Pattern Recognition Conference. 2025: 14901-14911.
- [8] Tong Z, Chen Y, Xu Z, et al. Wise-IoU: bounding box regression loss with dynamic focusing mechanism [J]. arXiv preprint arXiv: 2301.10051, 2023.
- [9] Cheng L, Zhang D, Zheng Y. Road object detection in foggy complex scenes based on improved YOLOv8 [J]. IEEE Access, 2024.
- [10] 孟志永, 吴晨曦, 王鹏, 等. 基于改进 RT-DETR 的井盖病害轻量化检测算法 [J/OL]. 计算机工程与应用, 1-13. (2025-09-28). <https://doi.org/10.3778/j.issn.1002-8331.2506-0079>
- [11] Chen Y, Wang Y, Zou Z, et al. GMS-YOLO: A Lightweight Real-Time Object Detection Algorithm for Pedestrians and Vehicles Under Foggy Conditions [J]. IEEE Internet of Things Journal, 2025.
- [12] Li B, Ren W, Fu D, et al. Benchmarking single-image dehazing and beyond [J]. IEEE transactions on image processing, 2018, 28(1): 492-505.
- [13] Sakaridis C, Dai D, Van Gool L. Semantic foggy scene understanding with synthetic data [J]. International Journal of Computer Vision, 2018, 126(9): 973-992.
- [14] Kenk M A, Hassaballah M. DAWN: vehicle detection in adverse weather nature dataset [J]. arXiv preprint arXiv: 2008.05402, 2020.
- [15] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector [C] // European conference on computer vision. Cham: Springer International Publishing, 2016: 21-37.
- [16] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks [J]. Advances in neural information processing systems, 2015, 28.
- [17] Reis D, Kupec J, Hong J, et al. Real-time flying object detection with YOLOv8 [J]. arXiv preprint arXiv: 2305.09972, 2023.
- [18] Wang A, Chen H, Liu L, et al. Yolov10: Real-time end-to-end object detection [J]. Advances in Neural Information Processing Systems, 2024, 37: 107984-108011.
- [19] Khanam R, Hussain M. Yolov11: An overview of the key architectural enhancements [J]. arXiv preprint arXiv: 2410.17725, 2024.
- [20] Tian Y, Ye Q, Doermann D. Yolov12: Attention-centric real-time object detectors [J]. arXiv preprint arXiv: 2502.12524, 2025.
- [21] Carion N, Massa F, Synnaeve G, et al. End-to-end object detection with transformers [C] // European conference on computer vision. Cham: Springer International Publishing, 2020: 213-229.