

小目标车辆及行人轻量化检测算法改进:DSWR-YOLO^{*}蔡惠强¹ 方 道^{1,2} 王 震¹ 姚 聪¹

(1. 厦门理工学院机械与汽车工程学院 厦门 361024; 2. 厦门大学航空航天学院 厦门 361005)

摘 要: 针对自动驾驶场景下小目标车辆与行人检测精度低、模型计算复杂度高等问题,提出了一种基于 YOLOv10n 改进的 DSWR-YOLO 轻量化小目标行人与车辆检测算法。首先引入 DWR 模块,改进其扩张残差为 DSConv,并嵌入 SimAM 注意力机制重构 C2f 模块,增强小目标特征保留能力;其次新增 160×160 检测层,采用多专家融合和分离设计并结合 DynamicConv,设计全新的动态检测头 Detect-dyHead-P2,显著降低模型参数量并提高对小目标的检测能力;最后融合 Focaler-SDIoU 损失函数,动态调整损失权重,以解决样本不均衡和边界框回归不稳定等问题。在 VisDrone2019 数据集上进行验证,改进模型(DSWR-YOLO)在参数量和浮点运算分别减少 25.9%和 33.3%的情况下,mAP@0.5、Precision 和 Recall 分别提升 3.7%、2.9%和 3.3%,轻量化的同时提升模型的检测精度,适用于资源受限的嵌入式设备。在 UA-DETRAC 数据集进行泛化实验,改进模型比原模型 mAP@0.5、Precision 和 Recall 分别提升 0.9%、1.2%和 2.0%,表明模型的泛化能力良好,鲁棒性强。

关键词: 行人与车辆检测;YOLOv10n;轻量化;DSWR;动态检测头;Focaler-SDIoU

中图分类号: TP391.41; TN914 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Improvement of lightweight detection algorithm for small
target vehicles and pedestrians: DSWR-YOLOCai Huiqiang¹ Fang Qiu^{1,2} Wang Zhen¹ Yao Cong¹

(1. School of Mechanical and Automotive Engineering, Xiamen University of Technology, Xiamen 361024, China;

2. School of Aerospace Engineering, Xiamen University, Xiamen 361005, China)

Abstract: To address the challenges of low detection accuracy for small-target vehicles/pedestrians and high computational complexity in autonomous driving scenarios, this paper proposes DSWR-YOLO—an improved lightweight object detection algorithm based on YOLOv10n. First, a DWR module is introduced, where its dilated residual structure is enhanced to DSConv. The SimAM attention mechanism is embedded to reconstruct the C2f module, strengthening feature retention for small targets. Second, an additional 160×160 detection layer is incorporated. A novel dynamic detection head, Detect-dyHead-P2, is designed using multi-expert fusion and separation combined with DynamicConv, significantly reducing model parameters while improving small-target detection capability. Finally, the Focaler-SDIoU loss function is integrated to dynamically adjust loss weights, mitigating sample imbalance and unstable bounding box regression. Validation on the VisDrone2019 dataset demonstrates that DSWR-YOLO achieves a 25.9% reduction in parameters and 33.3% decrease in FLOPs, while improving mAP@0.5 by 3.7%, Precision by 2.9%, and Recall by 3.3%. This delivers enhanced accuracy with reduced computational costs, making it suitable for resource-constrained embedded devices. Generalization experiments on the UA-DETRAC dataset show that DSWR-YOLO outperforms the baseline by 0.9% in mAP@0.5, 1.2% in Precision, and 2.0% in Recall, confirming its robust generalization capability.

Keywords: pedestrian and vehicle detection; YOLOv10n; lightweight; DSWR; dynamic detection head; Focaler-SDIoU

0 引 言

随着自动驾驶和智能交通系统的快速发展,行人及车

辆检测算法成为计算机视觉领域的关键研究方向。为满足自动驾驶的安全需求,对于行人和车辆目标检测算法精度提出更高的要求。在复杂天气背景及交通环境下,目标车

辆及行人的尺寸较小,在整个画面中只占有很少的像素,导致检测过程中存在漏检和误检的问题^[1]。因此,如何快速准确检测出复杂交通环境下的小目标车辆和行人,实现高精度检测已成为当前自动驾驶领域的重要方向。

近年来,深度学习技术的突破推动目标检测领域的飞速发展,基于深度学习的目标检测算法可分为两大类:两阶段(Two-Stage)方法和一阶段(One-Stage)方法。两阶段目标检测算法包括区域卷积网络(regions with convolutional neural network, R-CNN)^[2]、Fast R-CNN^[3]、Faster R-CNN^[4]、Mask R-CNN^[5]等。两阶段检测算法通过候选框筛选与精细化回归,对遮挡、小目标及复杂背景的鲁棒性更强,但其两阶段架构导致计算冗余高、推理速度慢,且依赖大量标注数据训练,难以适配自动驾驶等实时性要求高的场景。而一阶段目标检测算法,如单次多边框检测(single shot multi-box detector, SSD)^[6]算法, YOLO (you only look once)系列算法^[7],采用端到端单次预测框架,直接回归目标位置与类别。特别是 YOLO 算法,作为一种端到端的深度学习模型,能够直接从原始图像中提取特征,简化目标检测流程,并增强模型的直观性与易理解性。然而在面对复杂环境条件下,小目标车辆和行人信息极易被遮挡,导致模型准确率不高。

为提升小目标检测性能,李爽等^[8]在骨干网络中加入注意力机制增加模型的特征提取能力,为缓解训练样本的不平衡问题,使用 Focal-EIOU-Loss 作为模型的损失函数。张祺等^[9]基于 YOLOv10 构建特征提取模块 Deepstar Block 保持计算效率的基础上实现结构稳定和表达丰富,提出注意力模块 SCAF,增强模型局部信息感知能力同时降低计算复杂度。黄毅等^[10]引入多通道逆残差模块 MCIR 以及 CBAM 注意力机制,提升模型在复杂背景中对多尺度目标的特征提取能力和特征目标的定位能力。陈海秀等^[11]提出了 C2f-MCLU 模块并构建了 DyHead-DCNv4 检测头,针对小目标设计了增强的双向融合金字塔结构来重构颈部网络,使模型能提取更多的细微特征。翟永杰等^[12]通过在图像层面引入 Deep Lab V3 结合 3 种损失函数提取显著前景以削弱背景干扰。Fan 等^[13]引入一种新的多尺度特征融合模式,通过在特征金字塔网络中引入上采样和渐进特征金字塔网络,解决特征传播和交互的退化问题;Li 等^[14]在 YOLOv8 中引入有效通道注意力(ESCA)机制,显著增强模型识别小目标的能力;Zhong 等^[15]提出一种基于倒残差的多阶门控聚集算法,该算法围绕感兴趣流形的概念进行设计,有效地适应多尺度变化,减少小尺度目标的信息损失;Bai 等^[16]设计一个细粒度信息提取模块(FIEM)来代替标准卷积,并且在双向特征金字塔网络(BiFPN)的基础上增加一个额外的跳过连接分支,以保留细粒度信息并改善多尺度特征融合;杨沙沙等^[17]提出了一种批量正则化注意力模块和 QARepC2f 模块来提高 YOLOv8 的特征提取能力。以上研究虽然提高了模型对小目标的检测精度,

然而,上述方法往往依赖较深的网络结构,导致模型参数量与计算复杂度较高,限制了其在资源受限场景中的应用。

为降低模型复杂度,提升在嵌入式设备中的部署可行性,赵森等^[18]设计轻量化特征金字塔网络(Tiny-FPN),结合上下文锚点注意力机制与优化的 C3k2 模块,在控制计算成本的同时提高检测精度。杨迪等^[19]提出轻量级特征对齐检测头 LSCSBD,通过共享卷积层提升参数利用效率。马俊杰等^[20]将 Re-VGG 结构融入 C3 模块,构建 Re-C3,并采用轻量下采样操作 SCDwn,在减少计算冗余的同时增强特征融合能力。江灏等^[21]将 TCN 和 GCN 模块结合并在此基础上计算重构与真实信号间的残差使模型能够适应不同的工况变化。黄建硕等^[22]使用组混洗卷积(GSConv)替代传统卷积,通过通道分组与特征重排实现效率与性能的平衡。尽管这些方法在模型压缩方面取得一定成效,但在复杂场景下对小目标的检测精度仍有待提高。

综上所述,现有的研究在车辆和行人检测领域已取得不错的效果,但是仍存在以下问题:1)尽管提高模型检测精度,但是模型的参数量较大,计算复杂度高,不适用于资源受限的嵌入式设备中;2)在复杂环境下,小目标车辆和行人信息少,在特征提取阶段容易丢失,造成漏检、误检的情况;3)数据集包含多尺度样本,样本复杂度高且不同样本之间平衡性差,对于小目标样本,在预测框回归阶段模型稳定性差,导致模型检测精度受限制。针对以上问题本文提出一种改进 YOLOv10n 的轻量化小目标车辆和行人检测网络模型。主要改进包括:

1)引入扩张残差(dilation-wise residual, DWR)模块替代标准 Bottleneck,采用两步残差特征提取(区域+语义残差化),增强多尺度信息捕获;用 DSConv 替换空洞卷积降低计算开销;嵌入 SimAM 注意力机制,聚焦关键区域提升小目标特征保留能力;

2)新增 160×160 小目标检测层,解决深层特征图信息丢失问题;在回归分支集成 DynamicConv,通过动态路由机制融合多专家卷积核,自适应增强多尺度特征表示。采用分类静态+回归动态分离设计,平衡性能与计算效率;

3)融合 SDIoU 与 Focaler-IoU;SDIoU 基于目标尺度动态调整损失权重,缓解小目标边界框回归不稳定问题;Focaler-IoU 通过线性区间映射聚焦困难样本,解决复杂场景中样本不均衡问题,共同提升小目标检测鲁棒性。

1 YOLOv10 模型介绍

实时目标检测在计算机视觉领域应用广泛, YOLO 系列因其卓越的性能和效率平衡备受欢迎。YOLOv10^[23]由清华大学多媒体智能组(THU-MIG)在 2024 年 5 月发布。YOLOv10 的架构建立在以前 YOLO 模型的优势之上,通过消除非最大抑制(non-maximum suppression, NMS)和优化各种模型组件,实现最先进的性能,并显著降低计算开销。其网络结构如图 1 所示。与 YOLOv8 相比, YOLOv10

主要有以下 3 个创新点:

- 1) 无 NMS 训练: 利用一致的双重分配来消除对 NMS 的需求, 从而减少推理延迟。
- 2) 整体模型设计: 从效率和精度两个角度对各种组件

进行全面优化, 包括轻量级分类头、空间通道解耦下采样和引导块设计。

- 3) 增强的模型功能: 整合大核卷积和部分自注意力模块, 可在不增加大量计算成本的情况下提高性能。

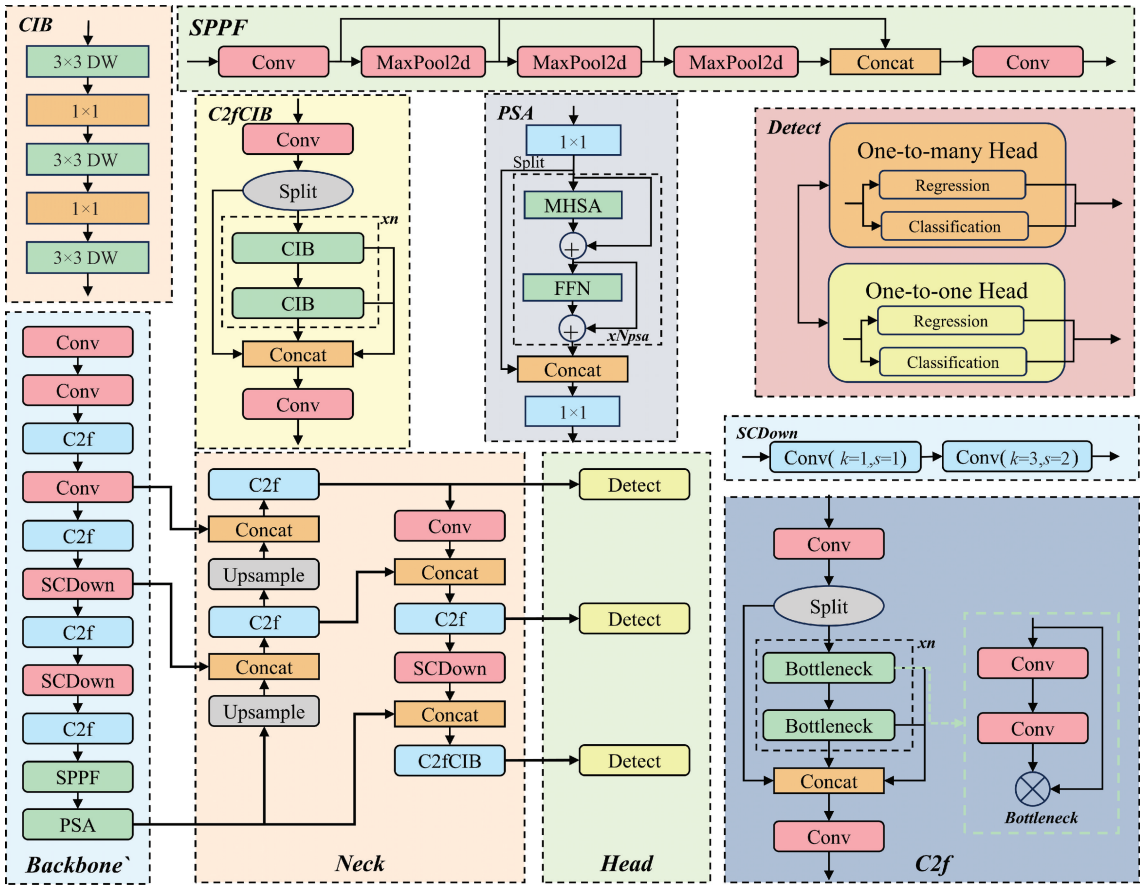


图 1 YOLOv10 网络结构
Fig. 1 YOLOv10 network structure

2 改进 DSWR-YOLO 算法

在目标检测领域,小目标检测是一个具有挑战性的问题,YOLOv10 模型是一种新的实时目标检测模型,然而在面对小目标时,仍存在一定的局限性。为提升小目标车辆及行人检测精度的同时轻量化模型,本文提出一种轻量化小目标行人及车辆检测改进 YOLOv10n 算法,其网络结构如图 2 所示。首先引入 DWR 模块,使用 DSCov 替换 DConv,以此设计全新的分布偏移残差(distribution shift-wise residual, DSWR)模块改进 Bottleneck 模块,并嵌入 SimAM 注意力机制重构 C2f;其次,通过动态路由机制结合 DynamicConv 设计全新的检测头 Detect-dyHead-P2,并增加 160×160 小目标检测头,增强对小目标的检测能力;最后将 Focaler-SDIoU 作为边界框回归损失函数,解决不同样本之间均衡性问题,提升模型对小目标预测的稳定性。

2.1 重构 C2f 模块(C2f-DSWRAttn)

在传统的目标检测算法中,骨干网络的核心作用在于提取输入图像中多层次、多尺度的特征。在传统卷积神经网络中,随着网络深度的增加,特征图的分辨率随之逐级下降。

目前,很多研究者采用基于多速率深度方向的线性卷积,使得模块从一个输入特征映射中捕获多尺度上下文信息,以提高特征提取效率,实现实时语义分割,导致深度可分离扩张卷积中的大量权重参数(尤其是扩张率较大的卷积层)难以被有效学习,从而导致小目标的细节信息在深度特征图中逐渐丢失,影响检测精度。扩张残差 DWR 模块^[24]主要应用于网络的高阶段,采用设计的两步残差特征提取方法(区域残差化-语义残差化),首先生成多尺度的区域特征表示,称为区域残差化(region residuals, RR);其次采用单一深度可分离膨胀卷积核对区域特征图执行语义形态学滤波,称为语义残差化(semantic residuals, SR)。如图 3

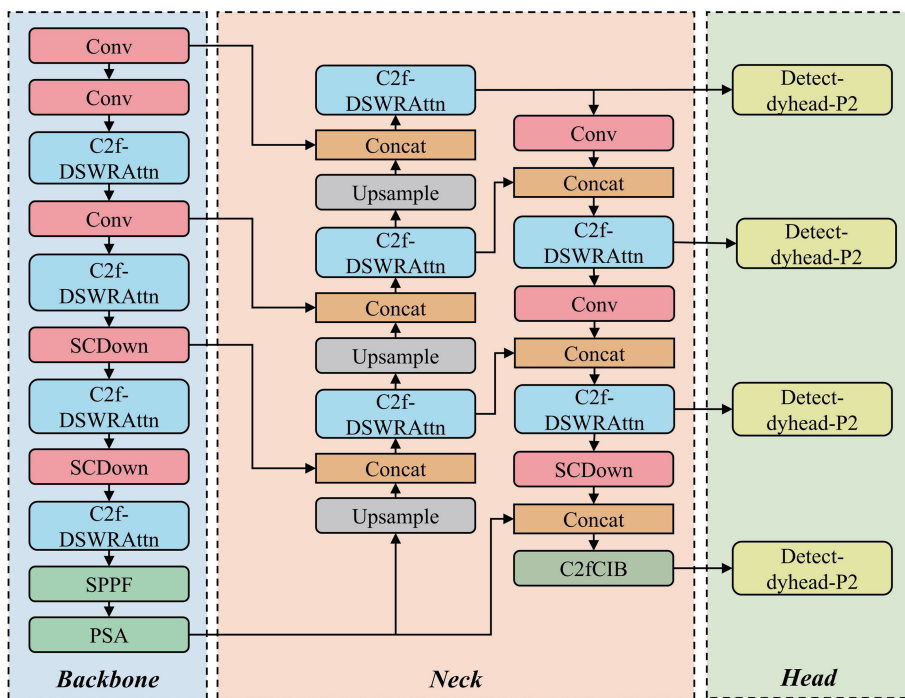


图 2 DSWR-YOLO 网络结构

Fig. 2 DSWR-YOLO network structure

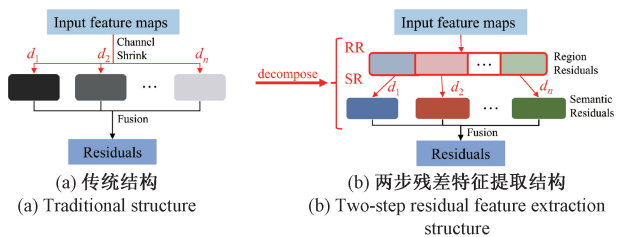


图 3 DWR 两步残差特征提取

Fig. 3 Two-step residual feature extraction of DWR

所示, DWR 能够有效提高实时语义分割中多尺度信息捕获的效率。

在原有的语义残差中,空洞卷积(DConv)采用偏移量学习,增加计算开销,使得模型在计算复杂度上更高,而分布偏移卷积(DSConv)^[25]采用可量化内核(variable quantized kernel, VQK)和分布偏移(kernel distribution shift, KDS),动态适配网络特征提取任务,降低模型复杂度并加速推理。其中可量化内核尺寸与原始卷积核完全相同,即尺寸为 (C_o, C_i, K_h, K_w) 。将每个值量化到 b 比特,其取值范围为:

$$[-2^{b-1}, 2^{b-1} - 1] \quad (1)$$

VQK 通过将原始权重按块(block)划分为大小为 B 的组,对每组进行拉伸和取整,使其适应上述整数区间。这使得所有运算可以使用补码整数算术,极大降低了计算和存储开销。通过简单地将正常卷积更改为 DSConv,每个张量权重节省的内存为:

$$p = \frac{b}{32} + \frac{\left\lceil \frac{C_i}{B} \right\rceil}{C_i} \quad (2)$$

KDS 是一个高精度的缩放因子张量,其尺寸为:

$$\left(C_o, \left\lceil \frac{C_i}{B} \right\rceil, K_h, K_w\right) \quad (3)$$

每个 KDS 值 ξ 负责对一组大小为 B 的 VQK 值进行缩放,以恢复其原始分布。采用最小化 L2 范数(欧几里得距离)最小化量化误差,即:

$$\xi = \min_{\xi} \sum_{i=0}^{B-1} (w_{q_i} \hat{\xi} - w_i)^2 \quad (4)$$

对于 $B = 64$,经 VQK 和 KDS 权重过滤之后,FP MACs(浮点数乘法运算)将从 1 350 减少至 27,如图 4 所示。

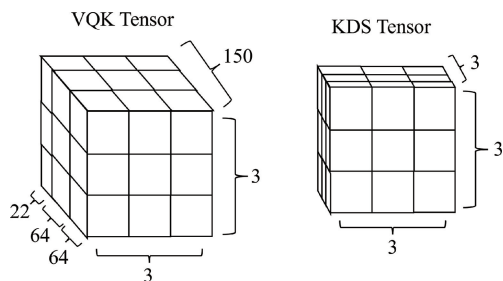


图 4 经 VQK 和 KDS 权重过滤后内存计算量

Fig. 4 Memory computation after VQK and KDS weight filtering

对于输入特征图尺寸 $H \times W \times C$ (其中 C 为通道数),卷积核尺寸 $K \times K$, 输出特征图尺寸 $H_{out} \times W_{out} \times C$,

DConv 的 FLOPs 计算公式为:

$$\text{FLOPs}_{\text{DConv}} = H_{\text{out}} \times W_{\text{out}} \times C \times K \times K \quad (5)$$

其中, $H_{\text{out}} = \left\lceil \frac{H + 2P - K}{S} \right\rceil + 1$, P 为填充, S 为步长。

对于 DSConv, 其 FLOPs 由整数操作等效 FLOPs 和浮点缩放操作 FLOPs 两部分组成, 公式如式(6)所示。

$$\text{FLOPs}_{\text{DSConv}} = H_{\text{out}} \times W_{\text{out}} \times C \times K \times K \times \frac{b}{32} + H_{\text{out}} \times W_{\text{out}} \times C \times \left\lceil \frac{C}{B} \right\rceil \quad (6)$$

其中, b 为量化比特数, $\frac{b}{32}$ 表示整数操作相对于浮点操作的精度比例, $\left\lceil \frac{C}{B} \right\rceil$ 为缩放因子的数量, 应用于每个输出元素。

将原有结构中的 DConv 替换为 DSConv, 所节省计算量为:

$$\Delta \text{FLOPs} = \text{FLOPs}_{\text{DConv}} - \text{FLOPs}_{\text{DSConv}} = H_{\text{out}} \times W_{\text{out}} \times C \times K \times K \times \left(1 - \frac{b}{32}\right) - H_{\text{out}} \times W_{\text{out}} \times C \times \left\lceil \frac{C}{B} \right\rceil \quad (7)$$

因此, 将原有结构中的 DConv 替换为 DSConv, 在适配 DWR 设计功能的基础上, 优化特征提取的效率, 降低模型的计算开销, 设计全新的 DSWR 模块, 如图 5 所示。将 DSWR 模块替换原有 C2f 中的 Bottleneck 中的 Conv 构成 Bottleneck-DSWR。在特征提取输出之前嵌入 SimAM 注意力机制^[26]对 DSWR 的多尺度输出进行空间加权, SimAM 注意力机制能够在不额外增加网络参数的情况

下, 让模型聚焦于关键区域, 提高模型的准确率。从而重构 C2f 模块为 C2f-DSWRAttn, 其网络结构如图 6 所示。

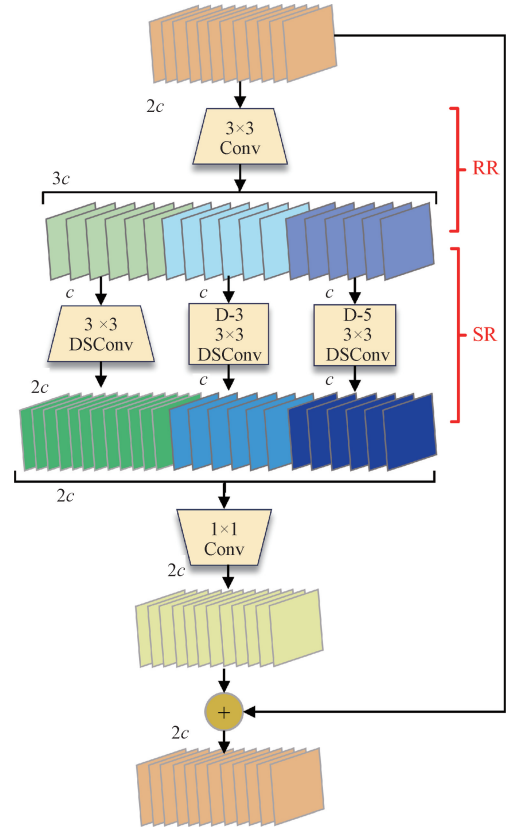


图 5 DSWR 模块
Fig. 5 DSWR module

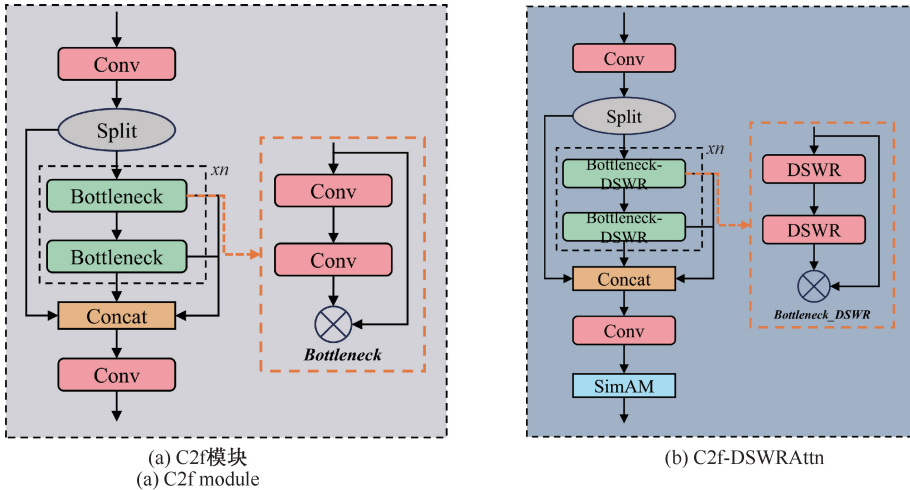


图 6 重构 C2f
Fig. 6 Refactor C2f

2.2 设计小目标动态检测头

在车辆目标检测数据中, 复杂的交通环境导致车辆及行人在密集场景下遮挡问题十分严重, 对于小目标车辆和

行人, 由于其所占像素很小, 很容易在采样过程中遗失特征信息导致检测精度下降。许德刚等^[27]通过添加小目标检测头, 提升模型对小目标的检测精度, 从而提高模型的

准确率和查全率等指标,为本文模型精度的提升提供思路。YOLOv10 的检测层分别输出 20×20 、 40×40 和 80×80 这 3 种不同尺寸的特征图,当输出的特征图尺寸为 20×20 时,由于小目标车辆所占图像大小很小,极易导致局部信息丢失。因此,在基准模型 YOLOv10 的基础上添加一层 160×160 尺寸的小目标检测层,增强对小目标的检测能力,此外,由于添加小目标检测头,使模型更加复杂,计算成本上升,内存占用增加,因此在检测头的回归分支设计中参考 ParameterNet^[28] 的方法,将动态路由机制融合多专家卷积核设计全新的轻量化检测头 Detect-dyHead-P2。

假设给定输入特征 $\mathbf{X} \in \mathbb{R}^{C_{in} \times H \times W}$ 和权重张量 $\mathbf{W} \in \mathbb{R}^{C_{out} \times C_{in} \times K \times K}$, 传统卷积层的操作为:

$$\mathbf{Y} = \mathbf{X} \times \mathbf{W} \quad (8)$$

其中, $\mathbf{Y} \in \mathbb{R}^{C_{out} \times H' \times W'}$ 是输出, $\times$ 是卷积运算。

ParameterNet 引入参数增强函数:

$$\mathbf{W}' = f(\mathbf{W}) \quad (9)$$

因此具有 M 个动态专家的 DynamicConv^[29] 可以表示为:

$$\mathbf{Y} = \mathbf{X} \times \mathbf{W}' \quad (10)$$

$$\mathbf{W}' = \sum_{i=1}^M \alpha_i \mathbf{W}_i \quad (11)$$

其中, $\mathbf{W}_i \in \mathbb{R}^{C_{out} \times C_{in} \times H \times W}$ $\mathbf{X}$ 是第 i 个卷积权重张量,对于输入 $\mathbf{X}$,首先应用全局平均池化将信息融合成一个向量,然后使用一个两层的带有 *softmax* 激活的 MLP 模块来动态地产生系数 α ,即:

$$\alpha = \text{softmax}(\text{MLP}(\text{Pool}(\mathbf{X}))) \quad (12)$$

其中, $\alpha \in \mathbb{R}^M$, 对应动态卷积核心设计如图 7 所示。

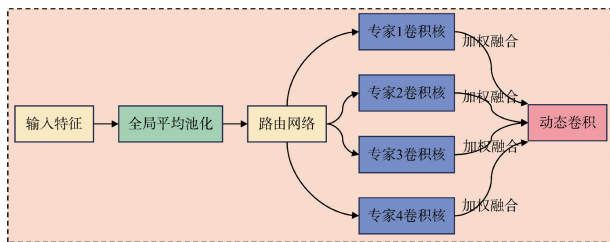


图 7 动态卷积核心

Fig. 7 Dynamic convolution core

动态路由机制通过动态调整以适应不同输入特征,提升多尺度目标检测能力;并且 DynamicConv 能够为分布焦点损失(distribution focal loss, DFL)提供更丰富的特征表示,提升分布参数预测精度;其次通过专家融合增加模型容量,同时避免计算量爆炸;最后采用分离设计(回归动态+分类静态)平衡性能与效率,改进检测头网络结构如图 8 所示。本文设计的 Detect-dyHead-P2 能够适应复杂场景中的目标检测,提升在遮挡目标、多尺度目标等任务下模型的检测精度。

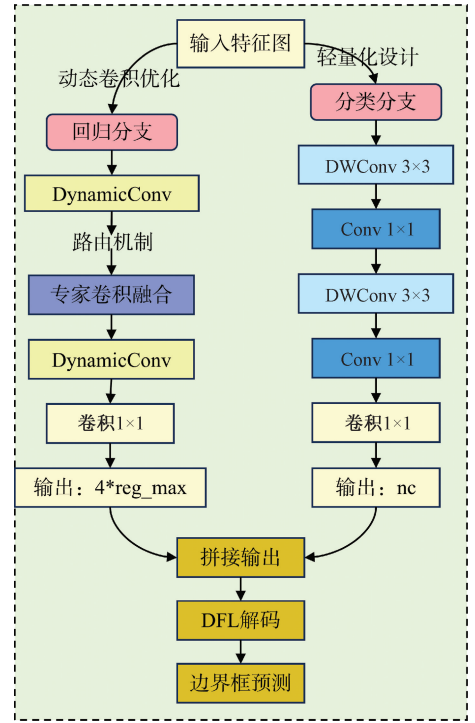


图 8 Detect-dyHead-P2 检测头网络结构

Fig. 8 The network structure of the Detect-dyHead-P2 detection head

2.3 损失函数优化(Focaler-SDIoU)

在车辆和行人检测中,提高小目标检测能力是如今面临的难题,CIoU 损失函数在处理边框尺寸差异以及样本不均衡问题方面存在缺陷,小目标的长宽比受微小尺寸变化影响显著,而且对于小目标检测通常需要密集预测和快速收敛,而复杂的计算可能增加训练耗时,尤其在资源受限场景下表现更明显,CIoU 需同时计算重叠面积、中心点距离、长宽比等多项指标,其复杂度较高。

针对小目标预测框在 IoU 损失中表现出的极大不稳定性,从而对模型的稳定性和回归产生负面影响,文献[30]提出基于尺度的动态损失函数(scale-based dynamic loss, SDIoU),SDIoU 根据目标尺度动态调整 Sloss 和 Lloss 的影响系数,减少标签不准确对损失函数稳定性的影响。其计算公式如式(13)所示。

$$\begin{cases} L_{BS} = 1 - IoU + \alpha v \\ L_{BL} = \frac{\rho^2(b_p, b_{gt})}{c^2} \end{cases} \quad (13)$$

其中, L_{BS} 和 L_{BL} 分别为 Sloss 和 Lloss 的影响系数, IoU 表示预测 BBox 和真实 BBox 的交并比, αv 衡量 BBox 的纵横比一致性, $\rho(\cdot)$ 是欧几里得距离, b_p 和 b_{gt} 是预测 BBox B_p 和目标 BBox B_{gt} 的质心, c 是两个 BBox 的对角线长度。

对于掩码的分割损失函数,参考 SLS^[31] 损失,并定义掩码尺度损失(L_{MS})和掩码位置损失(L_{ML})如式(14)所示。

$$L_{MS} = 1 - \omega \frac{|M_p \cap M_{gt}|}{|M_p \cup M_{gt}|} \quad (14)$$

$$L_{ML} = 1 - \frac{\min(d_p, d_{gt})}{\max(d_p, d_{gt})} + \frac{4}{\pi^2} (\theta_p - \theta_{gt})^2 \quad (15)$$

其中, M_p 和 M_{gt} 是预测像素和目标像素的集合, d_p 和 d_{gt} 是预测像素和目标像素在极坐标中与原点距离的平均值, θ_p 和 θ_{gt} 是预测像素和目标像素在极坐标中的平均角度, ω 描述 M_p 和 M_{gt} 之间的差异。当模型缩放图像或对特征图进行下采样时,目标大小会发生变化。为确定真实目标大小,计算原始图像与当前特征图之间的比例:

$$R_{oc} = \frac{w_o \times h_o}{w_c \times h_c} \quad (16)$$

其中, w_o, h_o 是原始图像的宽度和高度, w_c, h_c 是当前特征图的宽度和高度。BBBox(β_B)和掩码(β_M)的影响系数计算为:

$$\beta_B = \min\left(\frac{B_{gt}}{B_{gt\max}} \times R_{oc} \times \delta, \delta\right) \quad (17)$$

$$\beta_M = \min\left(\frac{M_{gt}}{M_{gt\max}} \times R_{oc} \times \delta, \delta\right) \quad (18)$$

其中, $B_{gt\max} = M_{gt\max} = 81$ 是由国际光学工程学会定义的红外小目标的最大尺寸。损失的影响系数基于当前目标框的面积,其范围限制在 δ 内,并且 δ 是可调的。最终基于尺度的动态 BBBox 损失(SDB 损失)为:

$$\beta_{L_{BS}} = 1 - \delta + \beta_B, \beta_{L_{BL}} = 1 + \delta - \beta_B \quad (19)$$

$$L_{SDB} = \beta_{L_{BS}} \times L_{BS} + \beta_{L_{BL}} \times L_{BL} \quad (20)$$

当目标 BBBox 面积 > 81 时, L_{SDB} 退化为 CIoU 损失。基于尺度的动态掩码损失(SDM 损失)计算为:

$$\begin{cases} \beta_{L_{MS}} = 1 + \beta_M \\ \beta_{L_{ML}} = 1 - \beta_M \end{cases} \quad (21)$$

$$L_{SDM} = \beta_{L_{MS}} \times L_{MS} + \beta_{L_{ML}} \times L_{ML} \quad (22)$$

其中, $\beta_{L_{BS}}$ 和 $\beta_{L_{BL}}$ 分别是 L_{BS} 和 L_{BL} 的影响因子。

同时,VisDrone2019 数据集中样本占比不平衡,导致其检测难度上升,其各目标占比如表 1 所示。

表 1 VisDrone2019 数据集各尺寸目标占比

Table 1 The proportion of targets of different sizes in VisDrone2019 dataset

目标大小/pixel	目标占比/%
< 20	60
$20 \sim 30$	25
> 30	15

基于不同检测任务中对于简单样本和困难样本的分布不同,文献[32]提出加强边界框回归 Focaler-IoU 损失函数,Focaler-IoU 通过关注不同的回归样本,提高检测器在不同检测任务中的性能,使用线性区间映射的方法重构 IoU 损失,以实现对不同回归样本的关注,其公式为:

$$IoU^{focaler} = \begin{cases} 0, & IoU < d \\ \frac{IoU - d}{u - d}, & d \ll IoU \ll u \\ 1, & IoU > u \end{cases} \quad (23)$$

其中, $IoU^{focaler}$ 是重建的 Focaler-IoU, IoU 是原始 IoU 值, $[d, u] \in [0, 1]$ 。通过调整 d 和 u 的值,可以使 $IoU^{focaler}$ 聚焦于不同的回归样本。其损失定义为:

$$L_{Focaler-IoU} = 1 - IoU^{focaler} \quad (24)$$

结合以上两种损失函数,构建新的损失函数,即 Focaler-SDIoU 损失函数:

$$L_{Focaler-SDIoU} = L_{SDIoU} + IoU - IoU^{focaler} \quad (25)$$

由于在复杂背景环境中,样本预测存在简单样本与困难样本之间不平衡的情况,同时小目标预测框的极大不稳定性容易对模型稳定性和准确率造成影响,而 Focaler-SDIoU 损失函数能较好的关注不同的回归样本,解决预测框不稳定的问题,降低低质量样本对描框回归的影响,从而提高对小目标车辆检测的稳定性。

3 实验结果与分析

3.1 实验环境与评价指标

本文实验平台具体配置为:Windows 10(专业版)64 位操作系统,GPU 型号为 NVIDIA GeForce RTX 3090,显存大小为 24 G, CPU 为 Inter(R) Core(TM) i7-9700 @ 3.0 GHz,编程语言版本为 Python 3.9.21,深度学习框架为 PyTorch 2.4.0,CUDA 版本为 11.8。实验环境超参数设置如表 2 所示。

表 2 实验环境参数配置

Table 2 Configuration of experimental environment parameters

参数	设定值
Epochs	200
Batch-Size	8
Optimizer	SGD
Image-Size	640
Momentum	0.973
Lr0	0.01
Weight-Ecay	0.000 5
Pre-training Weights File	YOLOv10n.pt

本文实验采用平均精度(mAP@0.5, mAP@0.5:0.95)、精确率(Precision)、召回率(Recall)、模型参数量(Parameters)以及模型大小(Model-Size)作为模型评估指标,各评价指标的计算式如式(26)~(29)所示。

$$P = \frac{TP}{TP + FP} \quad (26)$$

$$R = \frac{TP}{TP + FN} \quad (27)$$

$$AP = \sum_{i=1}^{n-1} (r_{i+1} - r_i) \times P_{\text{inter}}(r_i + 1) \tag{28}$$

$$mAP = \frac{\sum_{i=1}^k AP_i}{k} \tag{29}$$

其中, TP 表示正确将正样本预测为正类的数量, FP 指模型错误地将负类样本预测为正类的数量, FN 指模型未能检测到实际存在的目标, r_i 是 Precision 第一个插值处对应的 Recall 值, P_{inter} 为下一个 Recall 值 r_{i+1} 与当前值 r_i 之间的最大 Precision 值, k 表示类别数。

3.2 数据集

本文采用 VisDrone2019 数据集^[33]和 UA-DETRAC 数据集^[34]作为实验数据集,其中 VisDrone2019 数据集作为对比实验和消融实验数据集,UA-DETRAC 数据集作为泛化实验数据集。

VisDrone2019 数据集是由天津大学机器学习和数据挖掘实验室 AISKEYEY 团队收集的。该数据集包含 288 个视频片段和 10 209 幅静态图像,涵盖从中国 14 个不同城市收集的广泛方面,如不同的位置、环境、对象类型和场景密度。这些图像和视频由各种无人机摄像头捕获,记录多种天气和光照条件下的场景。将 10 209 张静态图像按照 8:1:1 的比例随机划分为训练集、验证集和测试集。

UA-DETRAC 数据集是一个具有挑战性的多目标检测和多目标跟踪数据集。该数据集包含在北京和天津的 24 个不同地点拍摄的 10 h 视频。视频以 25 fps 的速度录制,分辨率为 960 pixel×540 pixel。UA-DETRAC 数据集中有超过 14×10^4 帧和 8 250 辆手动标注的车辆,总共有 121×10^4 个标记的对象边界框,其中训练集约 82 085 张图片,测试集约 56 167 张图片。通过对训练集和测试集每 10 张图片抽一张,制作得到新的训练集 8 209 张图片,测试集 5 617 张图片。

3.3 对比实验

本文所有的对比实验以 YOLOv10 轻量级模型 YOLOv10n 为基准模型验证不同改进模块的检测效果。

1) 不同 C2f 改进模块对比实验

为验证基准模块 DWR 中将 DConv 替换成 DSConv 后对整体模型计算量的贡献度,将 DWR 与 DSWR 进行对比实验,实验结果如表 3 所示。由表 3 可知,在将 DWR 模块中的 DConv 替换成 DSConv 后,模型整体参数量略微下降,GFLOPs 下降 11.7%,验证了重构 DWR 模块的有效性。

表 3 DSConv 对模型计算量影响

Table 3 Impact of DSConv on model computational load		
模型	GFLOPs/G	Parameters/M
DWR	9.4	2 986 124
DSWR	8.3	2 984 140

为探究 C2f-DSWRAttn 改进模块的有效性,以 YOLOv-10n 为基准模型,选用基准模型的原始模块 C2f 以及 C2f-SAConv、C2f-GhostModule-DynamicConv、RCSOSA、C2fCIBAttention 与 C2f-DSWRAttn 进行对比实验。

由表 4 和图 9 可知,C2f-GhostModule-DynamicConv 表现最差;RCSOSA 的平均精度表现不足;C2f-SAConv 检测精度有提升,但过多的参数量以及浮点运算加重模型的计算负担;C2fCIBAttention 虽然平均精度有不错的表现,但是精度和召回率方面过低。相比较其他改进 C2f 模块,C2f-DSWRAttn 在不额外增加模型复杂度情况下,相比于原模型 mAP@0.5 提升 0.4%,Recall 提升 0.6%,这表明对引进的 DWR 模块进一步的设计有效帮助模型提高对于小目标车辆行人的特征提取能力,同时加强对感兴趣区域的注意力,防止小目标特征信息在深度特征图中丢失。

表 4 不同 C2f 改进模块对比实验

Table 4 Comparative experiments of different C2f improvement modules

模块	mAP@0.5/	mAP@0.5;0.95/	Precision/	Recall/	Parameters/	GFLOPs/
	%	%	%	%	M	G
C2f	32.0	18.5	43.7	32.5	2.7	8.4
C2f-SAConv	32.2	18.5	42.9	32.7	3.1	8.9
C2f-GhostModule-DynamicConv	30.9	17.8	40.8	31.9	2.6	7.7
RCSOSA	31.8	18.6	42.5	31.9	6.2	22.6
C2fCIBAttention	32.2	18.7	43.4	32.4	2.7	8.4
C2f-DSWRAttn	32.4	18.9	43.3	33.1	2.9	8.3

2) 不同损失函数对比实验

为验证 Focaler-SDIoU 损失函数的模型检测能力,将其与 CIoU、DIoU、GIoU、NWD、WIoU v3 进行对比实验,实验结果如表 5 所示,实验表明 DIoU 表现最差,其检测精度低于基准损失函数 CIoU;SDIoU 虽然能显著提升模型

预测稳定性,但是在处理样本不平衡问题时表现较差,导致检测精度有所下降;GIoU 和 NWD 表现中等,对模型性能的提升效果有限;WIoU v3 虽然表现良好,检测精度都有所提升,但是 Focaler-SDIoU 效果最佳,与基础损失函数 CIoU 相比,其 mAP@0.5 和 Recall 分别提升 0.6%和

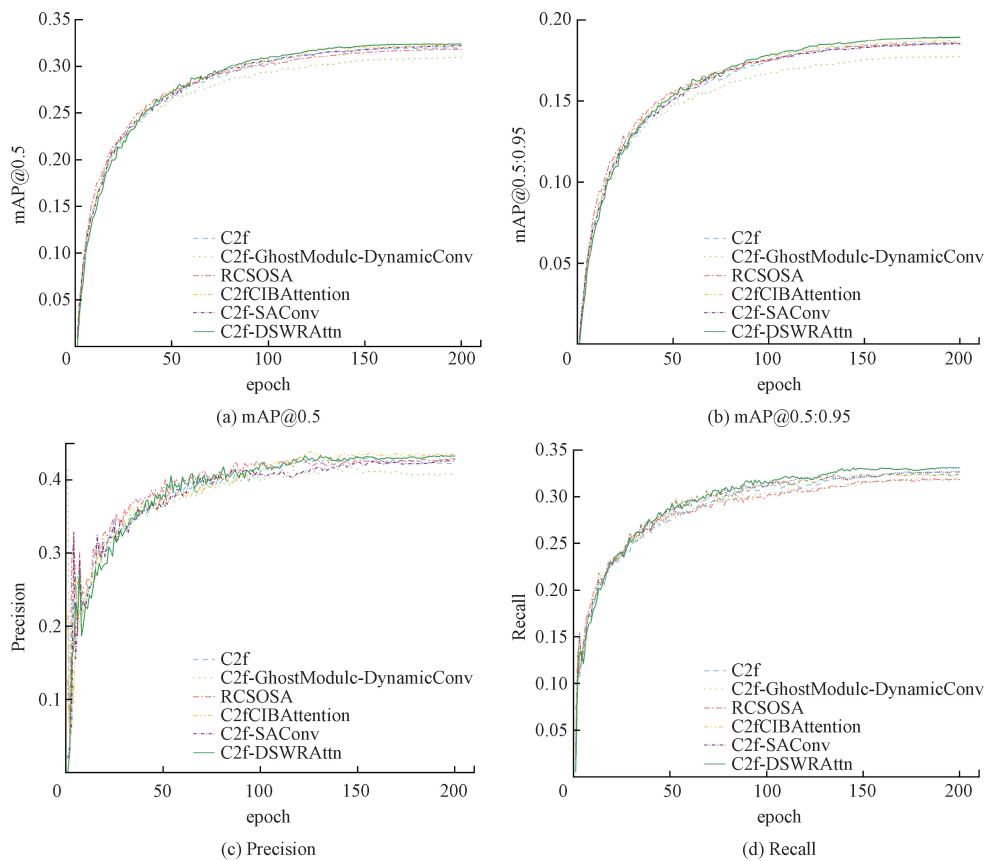


图 9 不同 C2f 改进模块可视化对比

Fig. 9 Visual comparison of different C2f improvement modules

表 5 不同损失函数对比实验

Table 5 Comparative experiments of different loss functions

%

损失函数	mAP@0.5	mAP@0.5:0.95	Precision	Recall
Base(CIoU)	32.0	18.5	43.7	32.5
DIoU	31.8	18.5	42.4	32.1
GIoU	32.2	18.6	44.0	31.8
NWD	32.0	18.3	43.0	31.8
WIoU v3	32.3	18.8	42.3	32.7
SDIoU	31.9	18.4	42.6	32.5
Focaler-SDIoU	32.6	18.7	42.9	33.2

0.7%,提升效果显著,该改进损失函数有效缓解了样本不平衡问题和小目标预测框不稳定性问题。此外,从图 10 的可视化结果中,可以直观的看到改进后的模型在训练过程中,其预测框损失较原模型有显著降低,进一步验证了本文所提改进损失函数的优越性。

3) 不同检测头对比实验

为验证设计的 Detect-dyhead-P2 检测头的优势,将其与 V10detect、P2(采用原模型检测头并添加 160×160 小目标检测头)、Detect-dyhead、RTDETRDecoder、Detect-MBConv、Detect-PPA 和 Detect-Conv2Formers 进行对比

实验,实验结果如表 6 和图 11 所示,实验结果表明 RTDETRDecoder 表现显著落后,其不仅模型复杂度高而且精度很低;P2 结果表明,添加小目标检测层对于模型的检测精度提升效果明显,但其大大增加了模型计算复杂度;Detect-dyhead 虽然大幅降低了模型复杂度,但其检测精度较差;Detect-MBConv 检测精度和模型复杂度较原始模型有所提升,但提升幅度有限;Detect-PPA 的 $mAP@0.5$ 相比基准模型提升 1.5%,综合精度表现良好,但其模型复杂度有所增加;本文采用多专家融合和分离设计并结合 DynamicConv 提出的 Detect-dyhead-P2 检测头表现最佳,

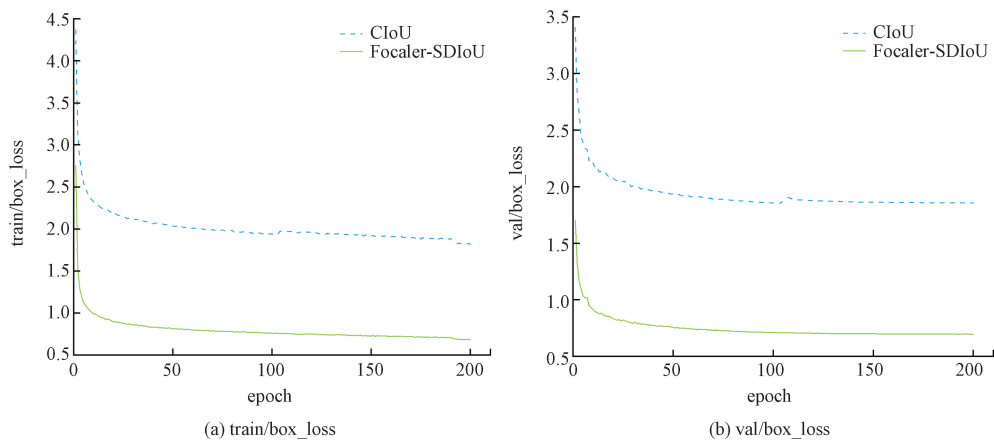


图 10 不同损失函数可视化对比

Fig. 10 Visual comparison of different loss functions

表 6 不同检测头对比实验

Table 6 Comparative experiments of different detection heads

检测头	mAP@0.5/%	mAP@0.5:0.95/%	Precision/%	Recall/%	Parameters/M	GFLOPs/G
V10detect	32.0	18.5	43.7	32.5	2.7	8.4
P2	34.0	20.2	43.8	34.8	2.8	15.7
Detect-dyhead	31.6	17.6	43.7	32.2	2.0	5.2
RTDETRDecoder	15.4	8.0	34.9	18.7	9.1	16.5
Detect-MBConv	32.8	18.6	44.6	33.2	2.0	5.4
Detect-PPA	33.5	19.2	43.7	33.9	2.9	9.7
Detect-Conv2Formers	32.3	18.2	42.7	32.4	2.1	6.2
Detect-dyhead-P2(ours)	35.7	19.8	45.2	36.0	2.1	6.5

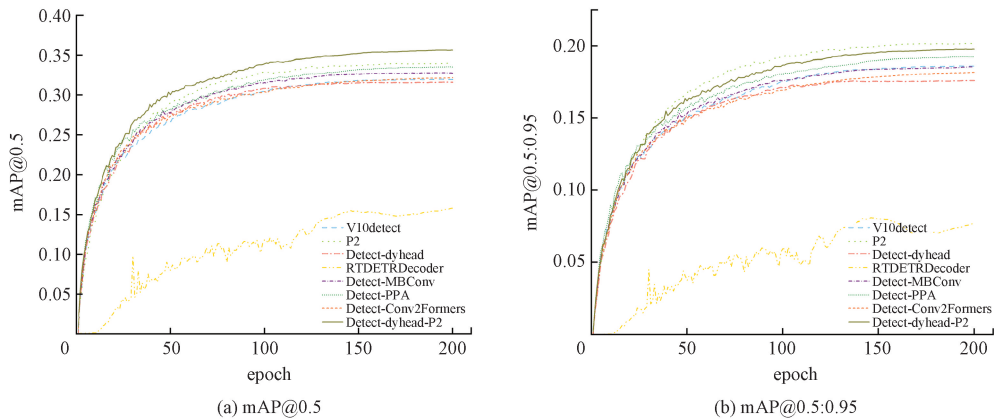
相对于原模型参数量减少 22.2%，浮点运算降低 22.6%，不仅大大降低模型的参数量和模型复杂度，而且在精度上，其 mAP@0.5 和 mAP@0.5:0.95 相比原模型分别提升 3.7% 和 1.3%，实现了多维度精度全面提升，有效解决在复杂环境下遮挡目标以及多尺度目标检测难度大的问题，进一步验证了本文所提检测头的优越性。

4) 不同算法对比实验

为进一步证明本文改进算法的优越性，将改进模型与

目前主流的深度学习算法模型进行对比实验，包括 Faster R-CNN、YOLOv5n、YOLOv8n、YOLOv10n(baseline) 以及近几年在小目标车辆和行人检测方向提出的先进方法 RGGE-YOLOv8^[35]、QARep-YOLOv8n 和 MELF-YOLOv5-S^[36]。

实验结果如表 7 所示，本文提出的 DSWR-YOLO 算法相比于其他主流的一阶段检测算法 YOLOv5n、YOLOv8n 和 YOLOv11n，mAP@0.5 分别更高 3.3%、2.0% 和 3.1%，参数量分别更低 25%、50% 和 30%，浮



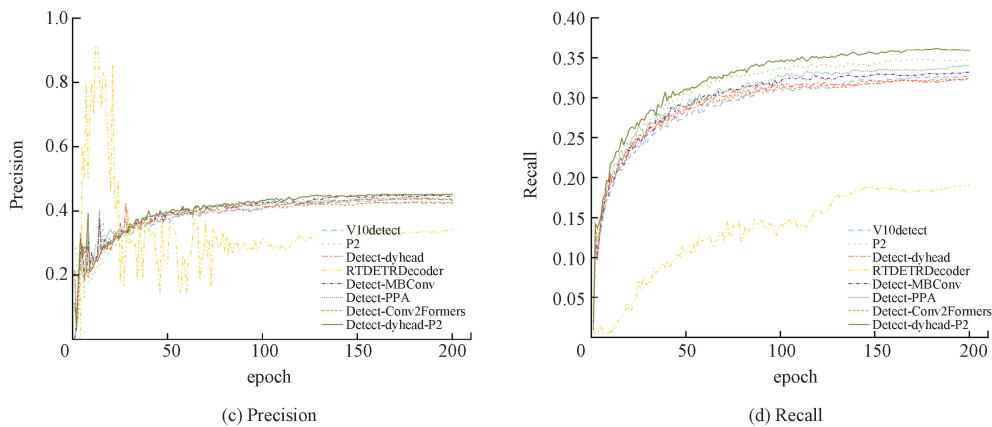


图 11 不同检测头可视化对比

Fig. 11 Visual comparison of different detection heads

表 7 不同算法对比实验

Table 7 Comparative experiments of different algorithms

算法	mAP@0.5/%	mAP@0.5:0.95/%	Precision/%	Recall/%	Parameters/M	GFLOPs/G
YOLOv5n	32.4	18.3	43.0	33.1	2.5	7.2
YOLOv8n	33.7	19.1	45.1	33.7	3.0	8.2
YOLOv11n	32.6	18.5	43.3	33.0	2.6	6.5
YOLOv10n(baseline)	32.0	18.5	43.7	32.5	2.7	8.4
RGGE-YOLOv8	34.8	18.9	33.7	66.5	14.3	20.2
QARep-YOLOv8n	35.2	19.7	—	35.5	3.0	8.1
MELF-YOLOv5-S	34.8	18.7	—	—	3.4	9.8
Faster-RCNN+ARFP	33.9	20.4	—	—	—	—
DSWR-YOLOv10n(ours)	35.7	19.8	46.6	35.8	2.0	5.6

点运算分别更低 28.6%、46.4% 和 16.1%。对比二阶段检测算法 Faster-RCNN+ARFP, 虽然在 mAP@0.5:0.95 低 0.6%, 但是 mAP@0.5 更高 1.8%。对比 RGGE-YOLOv8、QARep-YOLOv8n 和 MELF-YOLOv5-S 等最近提出的先进改进算法, 本文提出的算法不仅在所有模型中具有最高的检测精度达到 35.7%, 而且模型的参数量和浮点运算都是最低的, 在兼顾轻量化的同时, 提升模型的准确率, 有利于将改进算法布置在受限

制嵌入式设备中, 证明本文所提改进算法的实用性和优越性。

3.4 消融实验

为验证本文提出的各改进模块方法的有效性, 以 YOLOv10n 为基准模型, 使用 C2f-DSWRAttn、Detect-dyhead-P2 和 Focaler-SDIoU 在 VisDrone2019 数据集上分别进行消融实验, 实验结果如表 8 所示, 表中“√”表示某一组改进中使用哪些改进模块。

表 8 模块消融实验

Table 8 Module ablation experiment

C2f- DSWRAttn	Detect- dyhead-P2	Focaler- SDIoU	mAP@ 0.5/%	mAP@0.5: 0.95/%	P/ %	R/ %	Parameters/ M	GFLOPs/ G
Baseline	32.0	18.5	43.7	32.5	2.7	8.4		
✓			32.4	18.9	43.3	33.1	2.9	8.3
	✓		35.7	19.8	45.2	36.0	2.1	6.5
		✓	32.6	18.7	42.9	33.2	2.7	8.4
✓		✓	32.7	19.0	42.9	33.3	2.9	8.3
✓	✓		34.6	19.3	44.4	35.2	2.0	5.6
	✓	✓	35.6	19.9	46.0	35.7	2.1	6.5
✓	✓	✓	35.7(↑3.7)	19.8(↑1.3)	46.6(↑2.9)	35.8(↑3.3)	2.0(↓0.7)	5.6(↓2.8)

重构的 C2f-DSWRAttn 模块尽管 Precision 下降 0.4%,但是在 $mAP@0.5$ 、 $mAP@0.5:0.95$ 和 Recall 分别提升 0.4%、0.4% 和 0.6%,说明 C2f-DSWRAttn 模块能够有效解决小目标的细节信息在深层特征图逐渐丢失的问题,从而提高模型的检测精度;使用设计的 Detect-dyhead-P2 检测头代替原模型的 V10detect 检测头, $mAP@0.5$ 、 $mAP@0.5:0.95$ 、Precision 和 Recall 分别提升 3.7%、1.3%、1.5% 和 3.5%,表明本文设计的检测头在小目标检测上的优越性,其不仅有效降低模型的参数量以及浮点运算,而且大幅度提升模型的检测精度,有效解决在复杂环境下对遮挡目标和多尺度目标检测难度大的问题;用 Focaler-SDIoU 作为新的边界框回归损失函数,在 Precision 尽管下降 0.8%,但是在 $mAP@0.5$ 、 $mAP@0.5:0.95$ 和 Recall 分别提升 0.6%、0.2% 和 0.7%,验证了在参数量和计算复杂度不变的情况下,该损失函数能够有效改善小目标预测框不稳定以及简单样本与复杂样本之间的不平衡等问题,提升对车辆和行人检测的稳定性。融合 C2f-DSWRAttn 和 Focaler-SDIoU 后,模型在 $mAP@0.5$ 、 $mAP@0.5:0.95$ 和 Recall 分别提升 0.7%、0.5% 和 0.8%,这主要是这两个模块在融合过程中加强对小目标的细节特征提取能力以及预测框的稳定性,从而提升模型

的检测精度和稳定性;融合 C2f-DSWRAttn 和 Detect-dyhead-P2 后,模型的参数量和计算复杂度进一步下降,而且检测性能分别提升 2.6%、0.8%、0.7% 和 2.7%,这主要是融合过程中车辆和行人信息多尺度融合得到增强,并且使模型的网络聚焦与有用的参数上,有效缓解模型的计算冗余;融合 Detect-dyhead-P2 和 Focaler-SDIoU 后,检测性能提升 3.6%、1.4%、2.3% 和 3.2%,表明引入损失函数提高设计检测头的检测性能;当所有改进模块同时融入,相比较与原始 YOLOv10n 模型,改进模型检测性能分别提升 3.7%、1.3%、2.9% 和 3.3%,同时参数量下降 0.7 M,浮点运算下降 2.8 G,模型大小仅 4.64 MB,综合性能得到全面提升。综上所述,本文提出的改进模块在车辆行人检测中不但显著提升对小目标车辆及行人的检测精度,而且有效降低模型复杂,使其更适用于资源受限的场景。

为直观地展示改进算法的实际效果,本文将 YOLOv10n 和改进后的 DSWR-YOLO 模型进行可视化结果展示,如图 12 所示,第 1 列为原始图像,第 2 列为 YOLOv10n 模型检测效果,第 3 列为 DSWR-YOLO 模型检测效果,从图 12 中直观可以看到在不同场景和背景环境下,DSWR-YOLO 模型能够更好的检测出图像中的小目标,降低原始模型的漏检率和错检率,提升模型的检测精度。

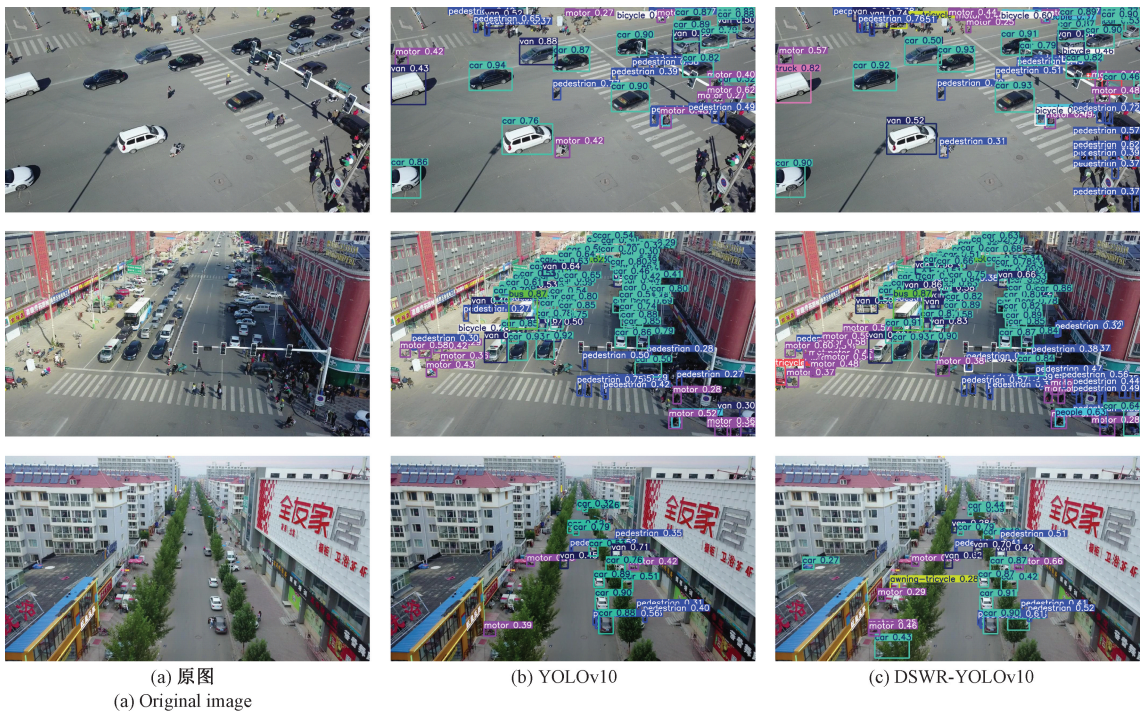


图 12 改进模型与未改进模型效果对比

Fig. 12 A comparison of the effects of the improved model and the unimproved model

3.5 泛化能力实验

为验证本文设计 DSWR-YOLO 算法的泛化能力,将训练好的模型与基准模型 YOLOv10n 在处理后的 UA-DETRAC 数据集上进行实验验证。实验结果如

图 13 和表 9 所示,改进前后算法检测性能分别提升 0.9%、0.4%、1.2% 和 2.0%,证明本文改进算法对于车辆及行人目标检测精度的提升效果良好,具有一定的泛化能力。



图 13 泛化实验结果可视化
Fig. 13 Visualization of generalization experiment results

表 9 泛化实验结果
Table 9 Generalization experiment results

模型	mAP@0.5/%	mAP@0.5:0.95/%	Precision/%	Recall/%	Parameters/M	GFLOPs/G
YOLOv10n	58.3	43.1	59.5	54.8	2.7	8.4
DSWR-YOLOv10n(本文)	59.4	43.5	60.7	56.8	2.0	5.6

3.6 嵌入式设备部署实验

为验证改进模型在嵌入式设备中的实际效果,如图 14 所示,将本文设计模型 DSWR-YOLO 与 YOLOv5n、YOLOv8n 以及 YOLOv10n 分别部署在 NVIDIA Jetson Xavier NX 平台上,基于 VisDrone2019 数据集,比较其检测精度和检测速度。

实验结果如表 10 所示,改进后的模型在检测精度上为 mAP@0.5 为 35.9%,Recall 为 36.0%,在所有模型中表现最好,同时检测推理速度为 53.4 ms,表现最佳,相比于原始模型,检测速度提升 11.0 ms,mAP@0.5 和 Recall 分别提升 3.2%和 3.5%。因此本文提出的算法在检测精度和检测速度上都有明显的提升,性能优于其他模型,证明有部署在嵌入式设备中应用的潜力。

表 10 不同模型在嵌入式设备上部署性能比较
Table 10 Performance comparison of the same model deployed on embedded devices

模型	mAP@0.5/%	Recall/%	Parameters/M	Detection speed/(ms/frame)
YOLOv5n	32.5	33.2	2.5	67.5
YOLOv8n	33.5	33.6	3.0	56.8
YOLOv10n (baseline)	32.7	32.5	2.7	64.4
DSWR-YOLO	35.9	36.0	2.0	53.4



图 14 NVIDIA Jetson Xavier NX 嵌入式平台测试
Fig. 14 NVIDIA Jetson Xavier NX embedded platform test

4 结 论

针对自动驾驶场景下,小目标行人及车辆目标检测精度低,本研究提出一种轻量化的行人及车辆检测 DSWR-YOLO 算法,该算法有效改善原模型存在的问题,提升复杂环境背景下小目标的检测精度。首先,设计 C2f-DSWRAttn 代替原网络中的 C2f 模块,解决小目标的细节信息在深度特征图中逐渐丢失的问题,有效提高实时语义分割中多尺度信息捕获的效率;其次,在网络结构中增加 160×160 检测头,并设计全新的 Detect-dyhead-P2 检测头替换原始网络中的 V10detect 检测头,增强模型小目标检测能力并显著降低模型计算复杂度;最后采用 Focaler-SDIoU 改进原始损失函数,较好的解决样本不平衡和小目标边界框回归不稳定的问题,提升对行人和车辆目标检测的稳定性。相较于基准模型 YOLOv10n,本文改进模型在 VisDrone2019 数据集上 mAP@0.5、mAP@0.5:0.95、Precision 和 Recall 为 35.7%、19.8%、46.6%和 35.8%,分别提升 3.7%、1.3%、2.9%和 3.3%,同时参数量下降 25.9%,浮点运算降低 33.3%,模型体积减少 18.4%。在模型轻量化的同时提高模型的检测精度和检测速度,适用于算力受限制的嵌入式设备,充分说明本文改进算法有应用于行人及车辆目标检测的潜在能力。在 UA-DETRAC 数据集上进行泛化实验,改进模型检测精度分别提升 0.9%、0.4%、1.2%和 2.0%,表明模型泛化能力良好,鲁棒性强。

参考文献

- [1] 赵倩,杨一聪.多重金字塔的轻量化遥感车辆小目标检测算法[J].电子测量技术,2023,46(13):88-94.
ZHAO Q, YANG Y C. Small object detection algorithm for lightweight remote sensing vehicles with multiple pyramids [J]. Electronic Measurement Technology, 2023,46(13):88-94.
- [2] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.
- [3] GIRSHICK R. Fast R-CNN[C]. IEEE International Conference on Computer Vision, 2015:1440-1448.
- [4] REN SH 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, 2016, 39(6): 1137-1149.
- [5] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN [C]. IEEE International Conference on Computer Vision, 2017: 2961-2969.
- [6] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multi box detector [C]. European Conference on Computer Vision. Cham: Springer, 2016:21-37.
- [7] 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), 2016: 779-788.
- [8] 李爽,曹阳,沈琴琴,等.改进 YOLOv5s 的小目标车辆检测算法[J].计算机仿真,2025,42(2):155-161.
LI SH, CAO Y, SHEN Q Q, et al. Small target vehicle detection algorithm based on improved YOLOv5s[J]. Computer Simulation, 2025, 42(2): 155-161.
- [9] 张祺,宋永献,阎妍,等. YOLOv10s-STD: 一种无人机视角下高精度目标检测算法[J].激光与光电子学进展,2026,63(6):563-574.
ZHANG Q, SONG Y X, YAN Y, et al. YOLOv10s-STD: A high-precision object detection algorithm from a drone's perspective [J]. Laser & Optoelectronics Progress, 2026, 63(6):563-574.
- [10] 黄毅,周纯,刘欣军,等.基于 YOLOv10 的无人机复杂背景下多尺度检测模型[J].光通信研究,2025(5): 45-52.
HUANG Y, ZHOU CH, LIU X J, et al. Multiscale detection model for complex backgrounds in UAV images based on YOLOv10 [J]. Study on Optical Communications, 2025(5):45-52.
- [11] 陈海秀,陈子昂,陈丁熊,等.改进 YOLOv10-n 的密集行人检测模型[J/OL].红外技术,1-10[2025-10-03].
<https://link.cnki.net/urlid/53.1053.tn.20250805.1001.002>.
- [12] 翟永杰,吴梓泮,周迅琪,等.基于双层级显著性驱动的车辆部件检测方法[J].仪器仪表学报,2025,46(5): 226-241.
ZHAI Y J, WU Z F, ZHOU X Q, et al. A dual-level saliency-driven vehicle component detection method [J]. Chinese Journal of Scientific Instrument, 2025, 46(5): 226-241.
- [13] FAN Q S, LI Y T, DEVECI M, et al. LUD-YOLO: A novel lightweight object detection network for unmanned aerial vehicle [J]. Information Sciences,

- 2025, 686: 121366.
- [14] LI J, ZHANG ZH Y, SUN H CH. DiffuYOLO: A novel method for small vehicle detection in remote sensing based on diffusion models [J]. Alexandria Engineering Journal, 2025, 114: 485-496.
- [15] ZHONG J L, ZHANG J X. MIR-YOLO: Remote sensing small target detection network based on visible-infrared dual modality [J]. Digital Signal Processing, 2025, 162: 105158.
- [16] BAI CH X, ZHANG K X, JIN H ZH, et al. SFPEF-YOLO: Small object detection network based on fine-grained feature extraction and fusion for unmanned aerial images[J]. Image and Vision Computing, 2025, 156: 105469.
- [17] 杨沙沙,徐新源. 基于 QARep-YOLOv8n 的城市物联网车辆目标检测方法[J]. 电子测量技术, 2025, 48(8): 46-54.
- YANG SH SH, XU X Y. Urban Internet of things vehicle object detection method based on QARep-YOLOv8n[J]. Electronic Measurement Technology, 2025, 48(8): 46-54.
- [18] 赵森,赵健康,龙海辉,等. DarkTiny-YOLO: 低光照场景下无人机航拍小目标检测算法[J]. 光电与控制, 2026, 33(2): 20-26, 34.
- ZHAO S, ZHAO J K, LONG H H, et al. DarkTiny-YOLO: Small object detection algorithm for UAV aerial photography in low light scenes[J]. Electronics Optics & Control, 2026, 33(2): 20-26, 34.
- [19] 杨迪,张喜龙,王鹏. LDD-YOLO: 改进 YOLOv8 的轻量级密集行人检测算法[J]. 计算机科学与探索, 2026, 20(1): 251-265.
- YANG D, ZHANG X L, WANG P. LDD-YOLO: An improved lightweight dense pedestrian detection algorithm based on YOLOv8[J]. Journal of Frontiers of Computer Science and Technology, 2026, 20(1): 251-265.
- [20] 马俊杰,张继红,王强,等. PGS-YOLO: 一种轻量高效的带钢表面缺陷检测模型[J]. 电子测量与仪器学报, 2025, 39(8): 156-167.
- MA J J, ZHANG J H, WANG Q, et al. PGS-YOLO: A lightweight and efficient strip surface defect detection model[J]. Journal of Electronic Measurement and Instrumentation, 2025, 39(8): 156-167.
- [21] 江灏,叶铭新,林蔚青,等. 基于时空动态检测的核电厂堆外中子探测器故障检测方法[J]. 仪器仪表学报, 2024, 45(9): 131-144.
- JIANG H, YE M X, LIN W Q, et al. Fault detection for ex-core neutron detectors in nuclear power plants using the spatial-temporal dynamic detection model [J]. Chinese Journal of Scientific Instrument, 2024, 45(9): 131-144.
- [22] 黄建硕,赵宇清,王佳,等. 夜视海面红外舰船目标轻量化检测算法[J/OL]. 计算机工程与科学, 1-13 [2025-10-03]. <https://link.cnki.net/urlid/43.1258.TP.20250903.1617.002>.
- HUANG J SH, ZHAO Y Q, WANG J, et al. Night vision infrared surface ship target lightweight detection algorithm[J/OL]. Computer Engineering & Science, 1-13 [2025-10-03]. <https://link.cnki.net/urlid/43.1258.TP.20250903.1617.002>.
- [23] WANG AO, 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.
- [24] WEI H R, LIU X, XU SH CH, et al. DWRSeg: Rethinking efficient acquisition of multi-scale contextual information for real-time semantic segmentation [J]. ArXiv preprint arXiv: 2212.01173, 2022.
- [25] NASCIMENTO M G, FAWCETT R, PRISACARIU V A. DSConv: Efficient convolution operator [C]. IEEE/CVF International Conference on Computer Vision, 2019: 5148-5157.
- [26] YANG L X, ZHANG R Y, LI L D, et al. SimAM: A simple, parameter-free attention module for convolutional neural networks[C]. International Conference on Machine Learning. PMLR, 2021: 11863-11874.
- [27] 许德刚,王双臣,尹柯栋,等. 改进 YOLOv8 的城市车辆目标检测算法[J]. 计算机工程, 2025, 51(11): 377-391.
- XU D G, WANG SH CH, YIN K D, et al. Improved YOLOv8 urban vehicle target detection algorithm[J]. Computer Engineering, 2025, 51(11): 377-391.
- [28] HAN K, WANG Y H, GUO J Y, et al. ParameterNet: Parameters are all you need for large-scale visual pretraining of mobile networks[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 15751-15761.
- [29] CHEN Y P, DAI X Y, LIU M CH, et al. Dynamic convolution: Attention over convolution kernels[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11030-11039.
- [30] YANG J N, LIU SH L, WU J J, et al. Pinwheel-shaped convolution and scale-based dynamic loss for infrared small target detection[C]. AAAI Conference on Artificial Intelligence, 2025, 39(9): 9202-9210.
- [31] LIU Q K, LIU R, ZHENG B L, et al. Infrared small

- target detection with scale and location sensitivity[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 17490-17499.
- [32] ZHANG H, ZHANG SH J. Focaler-IoU: More focused intersection over union loss [J]. ArXiv preprint arXiv:2401.10525, 2024.
- [33] DU D W, ZHU P F, WEN L Y, et al. VisDrone-DET2019: The vision meets drone object detection in image challenge results[C]. IEEE/CVF International Conference on Computer Vision Workshops, 2019.
- [34] WEN L Y, DU D W, CAI ZH 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.
- [35] 火久元,苏泓瑞,武泽宇,等. 基于改进 YOLOv8 的道路交通小目标车辆检测算法[J]. 计算机工程, 2025, 51(1):246-257.
- HUO J Y, SU H R, WU Z Y, et al. Road traffic small target vehicle detection based on improved YOLOv8[J]. Computer Engineering, 2025, 51(1): 246-257.
- [36] CAO L J, SONG P D, WANG Y CH, et al. An improved lightweight real-time detection algorithm based on the edge computing platform for UAV images[J]. Electronics, 2023, 12(10): 2274.

作者简介

蔡惠强, 硕士研究生, 主要研究方向为图像处理、汽车电子技术。

E-mail:1319274849@qq.com

方遁(通信作者), 教授, 硕士生导师, 主要研究方向为智能网联汽车、汽车电子技术、红外激光。

E-mail:fangq@xmut.edu.cn

王震, 硕士研究生, 主要研究方向为车辆动力学控制、汽车电子技术。

E-mail:1519393154@qq.com

姚聪, 硕士研究生, 主要研究方向为图像处理、汽车电子技术。

E-mail:y13698364976@163.com