

# 基于边界特征与多尺度特征的车辆目标检测<sup>\*</sup>

李天林<sup>1</sup> 安毅<sup>1,2</sup> 陈岩<sup>1</sup>

(1. 新疆大学电气工程学院 乌鲁木齐 830017; 2. 大连理工大学控制科学与工程学院 大连 116024)

**摘要:** 车辆目标检测技术在智能驾驶、智能交通和公共安全领域中具有重要意义。然而,真实环境中存在背景干扰、小目标检测难度大以及车辆密集时出现相互遮挡的问题。针对以上问题,提出一种基于边界特征与多尺度特征融合改进 YOLOv8 的车辆目标检测方法 EM-YOLO。首先,设计了一种边界引导的多尺度特征块,以结合边界特征与多尺度特征,用于改进原始的骨干网络,增强抑制背景干扰的能力。其次,特征在网络流动的过程中会出现细节信息的损失,而小目标车辆能够提取到的有效特征较少,加重了细节信息的损失。为此提出细节特征增强块,通过充分结合不同层级的特征,缓解细节信息的损失,进而改善小目标问题。随后,分析了车辆相互遮挡导致检测性能下降的原因,并针对此问题提出了一种检测头。最后,结合 PIoU、Focaler-IoU 和 WIoU,构造了 WPF-IoU,以优化训练过程,进而提升检测性能。经实验验证,改进后模型的精度和召回率相比原始模型分别提升了 1.9% 和 4.1%,mAP50 和 mAP50:95 分别提升了 4.4% 和 3.3%。与其他方法相比,本文提出的方法在各项性能指标上表现更优,具有一定的实际应用价值。

**关键词:** 车辆目标检测; 小目标; 多尺度特征; 细节信息; 损失函数

中图分类号: TP391; TN919.8 文献标识码: A 国家标准学科分类代码: 520.6040

## Vehicle detection based on boundary and multi-scale feature

Li Tianlin<sup>1</sup> An Yi<sup>1,2</sup> Chen Yan<sup>1</sup>

(1. School of Electrical Engineering, Xinjiang University, Urumqi 830017, China;

2. School of Control Science and Engineering, Dalian University of Technology, Dalian 116024, China)

**Abstract:** Vehicle target detection is crucial for intelligent driving, intelligent transportation, and public safety. However, challenges like background interference, small targets, and vehicle occlusion in dense traffic affect detection accuracy. For these problems, we propose EM-YOLO, which improves YOLOv8 by fusing boundary features and multi-scale features. First, we design a boundary-guided multi-scale feature block. It combines boundary and multi-scale features to improve the backbone network and enhance its ability to suppress background interference. Second, features lose details information as they flow through the network. Small vehicles extract fewer effective features, which worsens this issue. We propose a feature enhancement block that combines features from different layers to reduce detail loss and improve small target detection. Then, we analyze the performance drop caused by occlusion in dense vehicles and propose a detection head to address this issue. Finally, WPF-IoU is constructed by combining PIoU, Focaler-IoU, and WIoU. It optimizes the training process and improves detection performance. Experimental results show that the improved model achieved a 1.9% increase in precision and a 4.1% increase in recall compared to the original model. The mAP50 and mAP50:95 improved by 4.4% and 3.3%. Compared with other advanced methods, the proposed method outperforms in all performance metrics and has significant practical application value.

**Keywords:** vehicle object detection; small targets; multi-scale features; detail information; loss function

## 0 引言

我国汽车保有量不断增加,车辆目标检测技术已经成为研究者们关注的焦点。车辆目标检测的对象为车辆,其

得到的最终结果是已捕获的视频或图像中车辆位置的坐标信息。这一技术极大的推进了智能驾驶、智能交通和公共安全领域的发展。因此,车辆目标检测技术有着重要的意义。

随着计算机性能的不断提升,将深度学习技术应用于车辆目标检测已经成为主流的方法。基于深度学习的目标检测方法可以分为两大类,分别是以前 Faster R-CNN<sup>[1]</sup>为代表的两阶段目标检测方法和以单阶段目标检测器(single shot multi-box detector, SSD)<sup>[2]</sup>和 YOLO<sup>[3]</sup>系列为代表的一阶段目标检测方法。两阶段目标检测方法在第一阶段通过区域提议网络从输入的图像中筛选出可能包含待检测目标的区域,然后对这些候选区域进行分类和定位。魏子洋等<sup>[4]</sup>提出了一种基于 Faster R-CNN 的车辆目标检测方法,通过 K-means 算法对锚框的宽高值进行聚类,确定中心点坐标后重新设置区域提议网络的锚框尺寸及比例。虽然该方法提高了检测精度,但未能充分融合多尺度特征。Zhang 等<sup>[5]</sup>对 Faster R-CNN 进行了优化,通过计算帧间差分来提取目标的兴趣区域,并利用目标区域外的关联信息构建时空上下文信息,增强了目标特征的表达能力,提高了目标检测的准确率。然而,这些改进在提升检测性能的同时也增加了网络的复杂性。虽然两阶段目标检测方法的精度得到了保证,但其运行速度却不及一阶段目标检测方法。一阶段的目标检测方法简化了推理流程,仅使用一次前向传播就完成了目标分类和回归的任务,不需要生成候选区域。在保证精度的同时,达到了实时的标准。因此,一阶段的检测方法更适合于车辆目标检测的任务。目前学者们基于一阶段检测方法,针对如何实现高精度的车辆检测进行了深入的研究。Chen 等<sup>[6]</sup>使用了一种动态图拼接的方法,以缓解检测过程中正负样本分布不均的影响。虽然改善了检测效果,但存在计算量过大、推理速度慢的问题。梁天添等<sup>[7]</sup>在 YOLOv8 的主干网络中引入了可扩张残差结构和高效多尺度注意力机制,增强了网络在复杂环境下特征提取能力。此外,他们还使用 Wise-IoU 替换了网络原始的 CIoU,提升了检测性能。但是该网络在小目标检测方面还需进一步优化。Kang 等<sup>[8]</sup>提出了一种模糊注意力机制,并集成到 YOLOv8 中,以减少环境中不确定因素对检测结果的影响。同时,采用混合深度卷积提取多尺度特征,进而提升了检测精度,但耗费了更多的计算资源。陈梓延等<sup>[9]</sup>对 YOLOv8 进行了优化,首先将其主干网络替换为更轻量的结构,接着用双向特征金字塔代替了原有的颈部网络,以增强特征融合能力。最后,他们设计了一种新的损失函数,相较于原始的 YOLOv8,提升了精度。Bao 等<sup>[10]</sup>基于 YOLOv8 进行了改进,通过在原有的特征金字塔网络中引入可变形卷积来解决感受野不足的问题,并提出了一种新的 IoU 损失函数计算方法,提升了模型的泛化能力和收敛速度,但没有考虑如何增强网络抵抗背景特征干扰的能力。代少升等<sup>[11]</sup>在 YOLOv8 的主干网络中引入了频率自适应膨胀卷积以提升特征融合能力,同时在检测头使用了共享卷积以达到轻量化的目的。霍爱清等<sup>[12]</sup>在 YOLOv5 主干网络中引入了卷积模块注意力机制(convolutional block attention module, CBAM)。此外,他们对检测头进行

了改进,实现了回归和分类任务的分离。同时提出了 Alpha-IoU 损失函数,提升了模型的检测精度并加快了模型的收敛速度。虽然这些改进提升了网络的性能,但并未针对小目标问题进行优化。彭杰等<sup>[13]</sup>通过重参数化重构了 YOLOv8 的主干网络,并对原始的颈部网络进行了替换,同时使用一种更为轻量的检测头,实现了网络的强量化。顾杨海等<sup>[14]</sup>首先对 YOLOv8 原始的主干网络进行了改进,实现了轻量化。此外,设计了一种特征融合模块。最后,对训练阶段使用得损失函数进行了优化。但并未从本质上解决遮挡导致检测精度下降的问题。向雷等<sup>[15]</sup>对主干网络进行了优化,并将原始的主干网络替换为双向特征金字塔。实现了轻量化的同时,增强了主干网络特征融合的能力。此外,为了增强小目标的检测能力,他们添加了小目标检测头。综合以上改进,检测性能有所提升,但并未考虑背景信息对目标特征的影响。

尽管研究人员对车辆目标检测进行了深入的研究并取得了一定的成果,但仍然面临一些挑战。首先,在真实环境中,车辆与摄像头之间的距离有时会较远,导致车辆在图像中所占比例较小。因小目标能够提取到的特征较少,这增加了车辆目标检测任务的难度。此外,真实环境较为复杂,背景特征会对目标车辆的特征产生干扰。最后,当车辆密集时,会出现车辆相互遮挡的现象,这会导致目标回归框不准确,从而降低检测精度。

由于 YOLO 系列检测算法具有检测精度高、推理速度快和计算量小的特点,成为了众多研究的基础。本文同样将 YOLOv8n 作为基础模型,提出了检测方法 EM-YOLO。在网络结构方面,EM-YOLO 对原始的主干网络、颈部网络以及检测头进行了改进,主要包括以下内容:首先,提出边缘特征引导的多尺度特征块(edge guide multi-scale enhanced block, EGMEB),通过结合边缘特征与多尺度特征改进原始的主干网络,增强主干网络抑制背景特征干扰的能力。对于颈部网络,提出细节特征增强块(detail feature enhancement module, DFEM),以构建细节特征增强特征金字塔(detail feature enhancement feature pyramid network, DFEFPN)。该方法充分利用主干网络提取的不同层级的特征,缓解特征在网络流动过程中细节信息的损失,进而缓解小目标对检测器性能的影响。车辆目标检测任务包括边界框回归和分类两个子任务。随后,从这两项任务对特征的需求以及其对特征图的响应区域两个角度分析了车辆相互遮挡导致检测性能下降的原因,并提出一种检测头来缓解该问题。对于训练过程,则结合 WIoU、PIoU 和 Focaler-IoU 的优点,构建了 WPF-IoU 改善了训练效果。

## 1 基于边界特征与多尺度特征融合的车辆目标检测

### 1.1 EM-YOLO 网络结构

YOLOv8 网络模型由主干网络、颈部网络和检测头 3

部分组成。主干网络包含 Conv、C2f 和快速空间特征金字塔(spatial pyramid pooling fast, SPPF)3 种模块, 其中 Conv 和 C2f 交替分布, SPPF 位于末端。Conv 模块由卷积、批归一化和 SiLU 激活函数组成, 用于初步提取图像特征; C2f 模块用于进一步提取语义信息; SPPF 模块用于整合特征并扩展感受野。颈部网络结合了特征金字塔(feature pyramid network, FPN)和路径聚合网络(path aggregation network, PAN), 实现了自顶向下和自底向上的特征融合。检测头的分类分支使用二进制交叉熵损失(binary cross-entropy loss, BCE), 回归分支采用分布焦点损失(distribution focal loss, DFL)和 CIOU 损失。

EM-YOLO 的网络结构如图 1 所示, 主要包含主干网络、颈部网络和检测头。对于主干网络, 为抑制背景信息对

目标特征的干扰, 提出了 EGMEB, 并使用其对原始的 YOLOv8 主干网络进行重构。对于颈部网络, 为缓解小目标车辆对检测器性能的影响, 提出了 DFEM, 并基于该模块重构了颈部网络。对于检测头部分, 提出了一种新的方法, 用于替换 YOLOv8 原始的检测头。网络的具体流程如下: 首先, 输入的形状为  $640 \times 640 \times 3$  的图像经主干网络进行特征变换后, 将得到的第 2、4、6 和 9 层特征作为其输出, 其形状分别为  $160 \times 160 \times 32$ 、 $80 \times 80 \times 64$ 、 $40 \times 40 \times 128$  和  $20 \times 20 \times 256$ 。接着, 颈部网络以主干网络的输出特征为输入, 进行特征融合, 并将融合得到的第 14、17 和 20 层特征作为颈部网络的输出, 其形状分别为  $80 \times 80 \times 128$ 、 $40 \times 40 \times 128$ 、 $20 \times 20 \times 128$ 。最后, 检测头以颈部网络的输出为输入, 进行边界框的回归和分类, 最终得到检测结果。

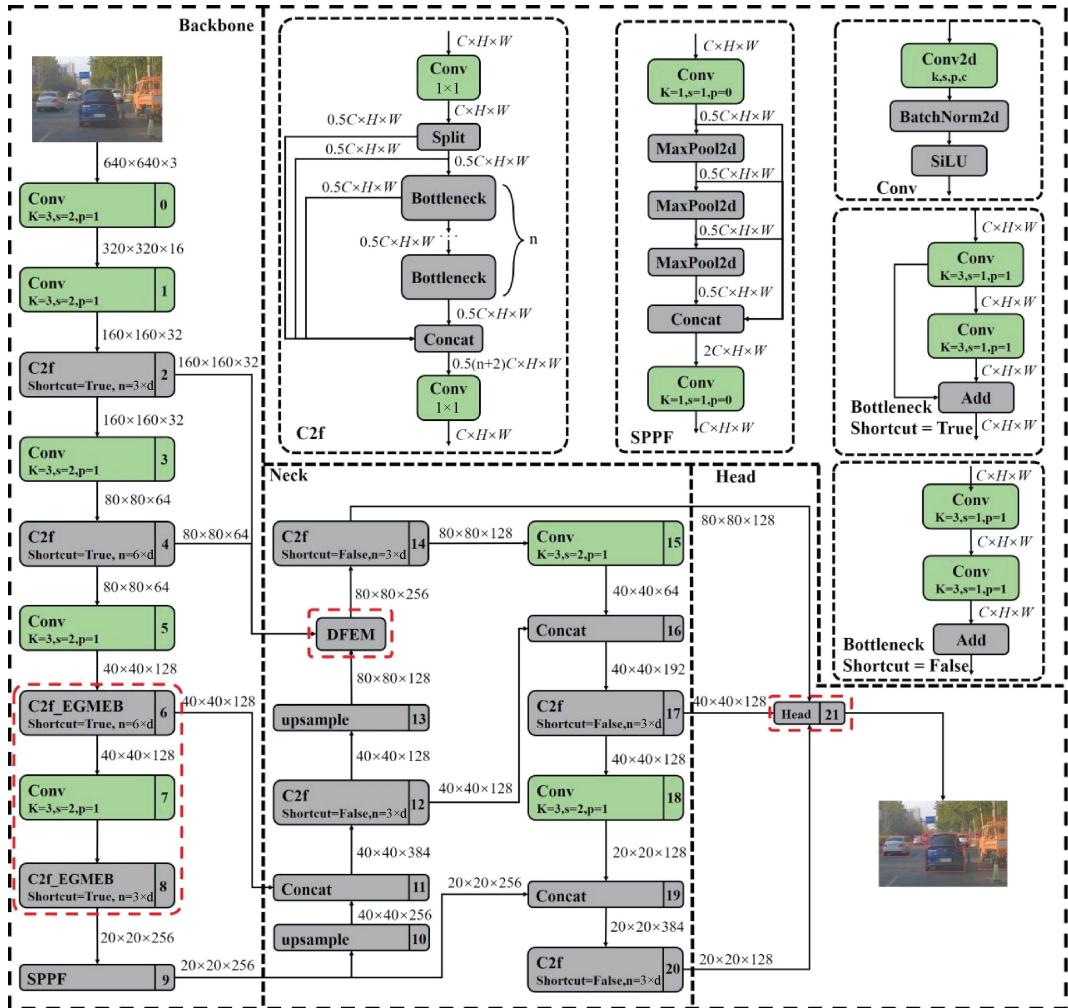


图 1 EM-YOLO 网络整体结构  
Fig. 1 Overall network structure of EM-YOLO

## 1.2 边缘特征引导的多尺度特征块

通过卷积神经网络提取的图像特征包含空间信息, 而图像的边缘特征保留了图像的重要结构属性。将两者结合可以抑制背景特征的干扰, 从而得到更加稳健的特征。

本文通过 EGMEB 实现了两者的结合, 并基于该模块构建了 C2f\_EGMEB, 将其添加到主干网络中, 从而增强了主干网络抵抗背景干扰的能力。C2f\_EGMEB 结构如图 2 所示。

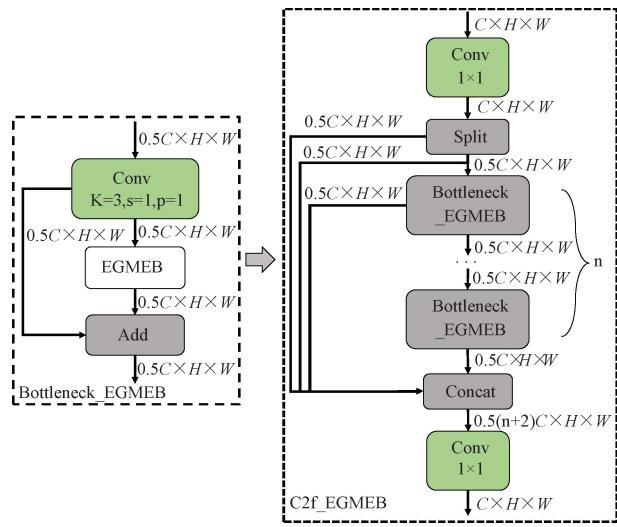


图 2 C2f\_EGMEB 结构

Fig. 2 C2f\_EGMEB structure

EGMEB 主要包含 MS 分支、EG 分支和 Guide stage 3 部分组成, 结构如图 3 所示。

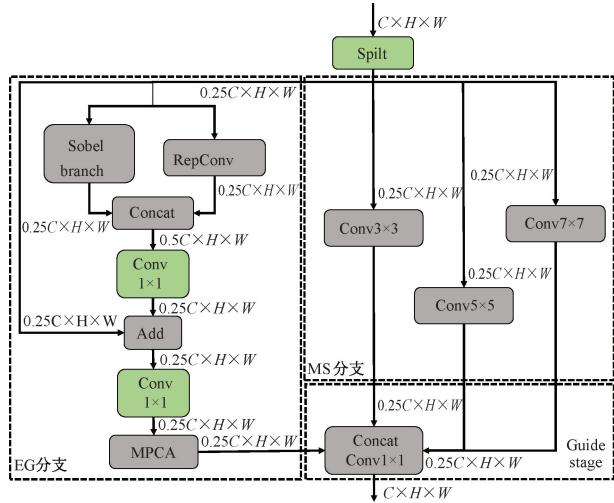


图 3 EGMEB 结构

Fig. 3 EGMEB structure

MS 分支由并行的  $3 \times 3$ 、 $5 \times 5$  和  $7 \times 7$  卷积构成, 用于捕获图像不同尺度的特征。EG 分支首先通过组卷积实现了 Sobel 算子, 以提取边缘特征, 其结构如图 4 中的 Sobel branch。除此之外, EG 分支为获取图像的空间信息使用了 RepConv, 其结构如图 4 所示。RepConv 在训练阶段使用多分支卷积层, 这样可以获得更加准确的特征表示。而在推理阶段, 则重参数化为单分支结构, 从而减少了计算量和内存消耗。

随后, EG 分支将提取的空间特征与边缘特征进行拼接, 拼接后通过卷积核大小为 1 的卷积进行通道变换, 变换后与原始输入至 EG 分支特征相加。为增强网络的非线性能力并强化网络提取到的边缘特征, 将相加后的特征

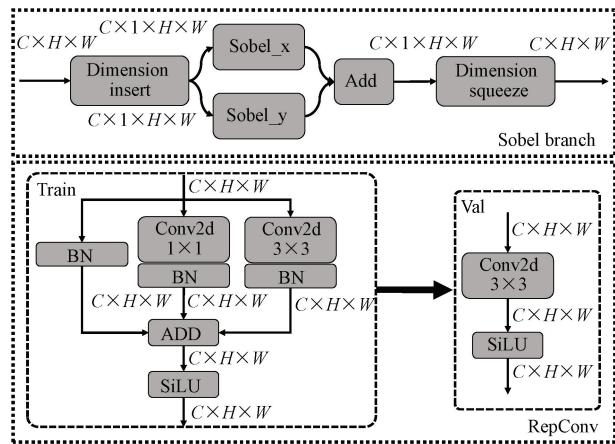


图 4 Sobel branch 和 RepConv 结构

Fig. 4 Sobel branch and RepConv structures

依次通过卷积核大小为 1 的卷积和多路径通道注意力机制(multi-path channel attention, MPCa)注意力机制<sup>[16]</sup>, 后者通过强化位置信息来提高边缘特征的准确性。MPCA 的结构如图 5 所示。

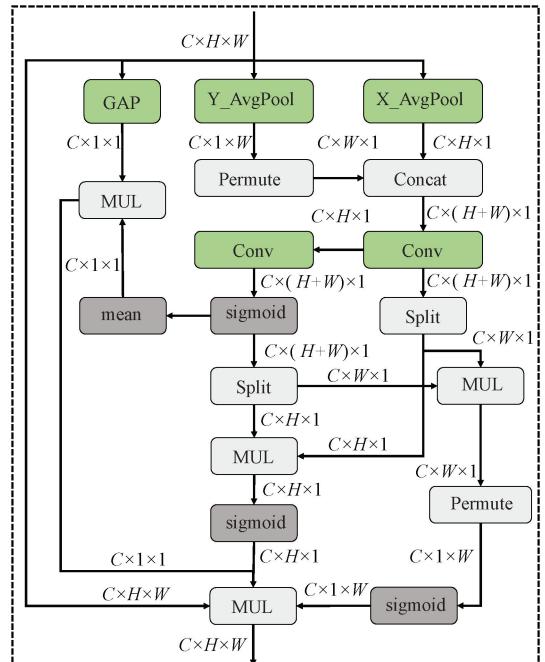


图 5 MPCa 结构

Fig. 5 MPCa structure

最终得到的特征即为 EG 分支输出的特征, 包含了图像中目标的边缘特征。逐点卷积<sup>[17]</sup>也称为  $1 \times 1$  卷积, 可以实现输入特征内部每一部分之间的关联。EGMEB 为实现边缘特征对多尺度特征的引导, 在网络的最后阶段对 EG 分支和 MS 分支输出的特征进行拼接, 并对拼接后的特征使用逐点卷积进行变换, 得到的特征即为最后的输出特征, 也即 Guide stage 部分的功能。EGMEB 实现了边缘特征与多尺度特征的结合, 将其用于主干网格, 可以增强

网络抵抗背景干扰的能力。

### 1.3 细节特征增强块

现有目标检测网络通过卷积的叠加实现特征提取,但随着网络的加深,提取的可用特征中细节信息逐渐减少。然而,在车辆目标检测任务中,目标车辆在图像中占比小的问题不可避免,这导致提取到的可用特征减少,进一步加剧了小目标车辆细节信息的损失,影响了检测精度。为此,本文对原始颈部网络进行了重构。考虑到原始颈部网络未使用主干网络提取的第 2 层特征,本文提出了 DFEM,并以此为基础对颈部网络进行优化。DFEM 整体结构以及通过其构造的细节特征增强 FPN (DFEFPN) 如图 6 所示。

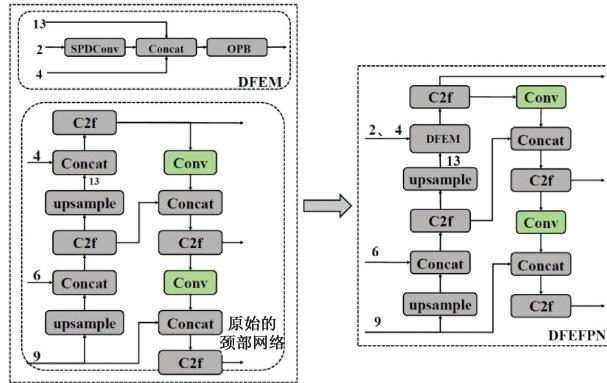


图 6 DFEFPN 结构

Fig. 6 DFEFPN structure

DFEM 首先使用 SPD-Conv<sup>[18]</sup> 对来自主干网络的第 2 层特征进行下采样,将空间维度转换为深度维度,以实现下采样的同时尽可能的保留细节信息,结构如图 7 所示。具体操作过程为:首先,对输入的尺寸为  $C \times H \times W$  的特征图使用 2 倍尺度因子进行隔行隔列的下采样,得到四个尺寸为  $C \times H/2 \times W/2$  的特征图。随后,将这 4 个特征图在通道维度上进行拼接,得到一个形状为  $4C \times H/2 \times W/2$  的特征图。最后,使用普通卷积对其进行通道维度的变换。

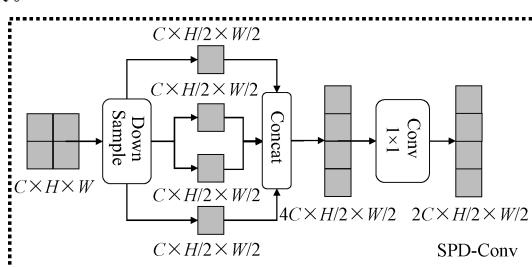


图 7 SPD-Conv 结构

Fig. 7 SPD-Conv structure

接下来,将下采样后的特征与主干网络的第 4 层和第 13 层提取的特征进行拼接,通过这种方法实现特征细节信息的补充。最后,为增强特征的表达能力并尽可能的

压缩计算量,本文基于特征的冗余性<sup>[19]</sup> 提出了全尺度部分卷积(omni-scale partial convolution block, OPB),使用其对拼接得到的特征进行变换,图 8 展示了其结构。特征的冗余性是指不同通道间的特征图存在高度相似性。在进行特征变换时可以选取部分通道的特征代表整个特征图进行变换,而其余通道保持不变。本文提出的 OPB 对特征进行处理时为提高运算效率,首先选择连续的前 1/4 通道的特征图作为整个特征图的代表,剩余的 3/4 通道的特征不做任何处理,以实现连续或规则的内存访问。随后,通过(multi-scale feature enhancement, MFE)模块<sup>[20]</sup>对选取的前 1/4 通道特征图进行变换,进一步提取特征中的有用信息,强化特征的表达能力。最后,将变换后的前 1/4 通道的特征图与未进行任何变换的 3/4 通道特征图进行拼接,将拼接得到的特征作为全尺度部分卷积 OPB 的输出特征,也即 DFEM 的输出特征。

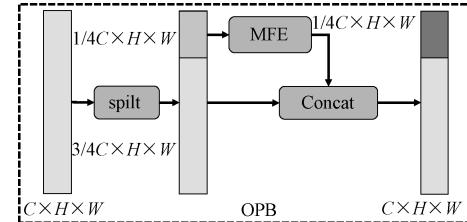


图 8 OPB 结构

Fig. 8 OPB structure

其中,MFE 模块通过扩大感受野、增强局部信息和全局建模能力来增强特征的表达能力,其结构如图 9 所示。

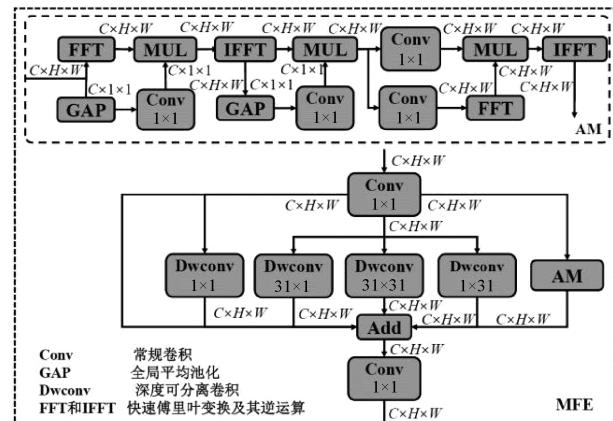


图 9 MFE 结构

Fig. 9 MFE structure

DFEM 融合了低层特征中的细节信息,减少了特征在网络流动过程中细节信息的损失,提升了检测器的整体性能。将其与原始的颈部网络进行结合,便得到了细节特征增强 FPN。与常规方法不同,细节特征增强 FPN 并未通过在颈部网络中单独引出特征分支来构建小目标检测层,而是通过融合低层特征的细节信息的方法提升小目标检测能力,进而提升检测网络的整体性能。

## 1.4 检测头

在车辆目标检测任务中,需要实现目标车辆边界框的回归和分类。从特征的本质来看,分类任务侧重于语义信息,而边界框回归任务侧重于空间信息<sup>[1]</sup>。目前,主流检测头通常采用分类分支和回归分支特征共享的结构,但由于两种任务对特征的需求不一致,它们之间相互影响。当车辆因密集而出现相互遮挡时,分类和回归任务对特征的要求有时无法同时满足,两种任务间的相互影响会被加剧。此外,这两个任务对目标图像的响应区域不同。分类任务关注的是目标图像中最具有判别性的区域,而回归任

务的响应区域在目标图像的边缘<sup>[21]</sup>。当车辆因密集而出现相互遮挡时,如果对回归任务响应区域的影响小于对分类任务响应区域的影响,模型会更加关注回归任务,相应的分类性能会受到影响。反之,模型会倾向于分类任务,相应的回归性能会受到影响。

针对以上问题,本文提出了一种检测头,其结构如图 10 所示。主要包含共享特征提取与分解以及分类子任务与回归子任务关联两个阶段。它分别从两个子任务特征需求不同和特征图响应区域不同的角度,缓解车辆相互遮挡时检测性能变差的问题。

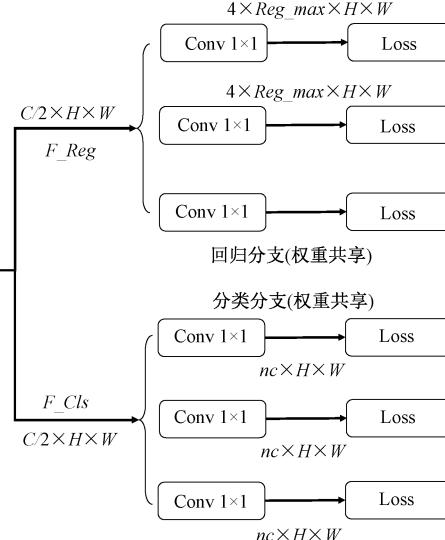
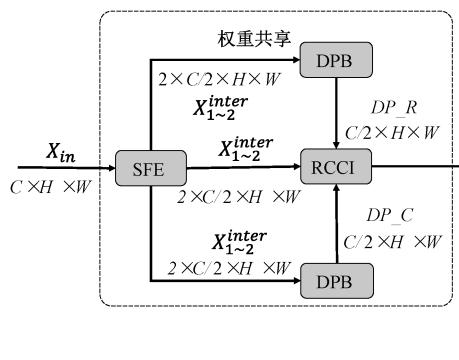


图 10 检测头结构

Fig. 10 Detection head structure

### 1) 共享特征提取与分解

首先,对颈部网络输出的 3 个特征分支使用共享特征提取器(shared feature extractor, SFE),其结构如图 11 所示,图中的  $X_{in}$  表示输入特征。SFE 由两个串联的 CGS 块构成,第一个 CGS 块将输入的特征的形状由  $C \times H \times W$  变为  $C/2 \times H \times W$ ,经过第 2 个 CGS 块后特征形状仍为  $C/2 \times H \times W$ 。本文将经由每一个 CGS 块变换后的特征构成的集合表示为  $X_{1~2}^{inter}$ ,并作为 SFE 的输出。其中每个 CGS 块依次由一个卷积核大小为 3 的常规卷积、一个组归一化层和一个 SiLu 激活函数构成。SFE 作为 3 个分支共享结构,实现了参数量和计算量的压缩。

$X_{1~2}^{inter}$  为分类任务和回归任务的共享特征,为解决分类任务和回归任务相互影响的问题,本文引入了一种解耦结构<sup>[22]</sup>,在文中表示为特征分解块(decouple block,DPB),其结构如图 11 所示。特征分解块 DPB 以  $X_{1~2}^{inter}$  作为输入,首先通过拼接得到形状为  $C \times H \times W$  的特征。随后,经过全局平均池化得到形状为  $C \times 1 \times 1$  的特征,再通过  $1 \times 1$  卷积与 Relu 激活函数进行变换,得到形状为  $C/16 \times 1 \times 1$  的特征。接着,通过  $1 \times 1$  卷积和 Sigmoid 激活函数得到集

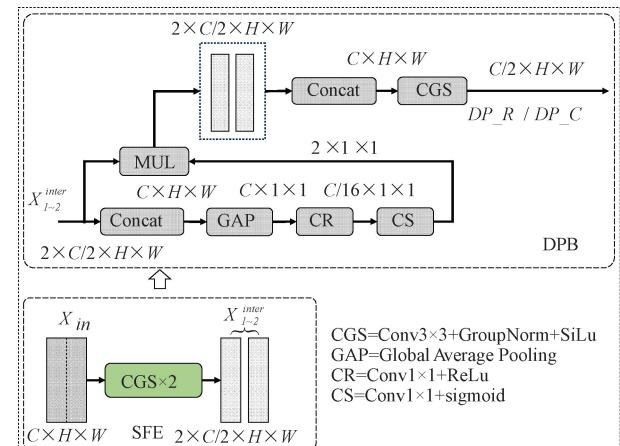


图 11 SFE 和 DPB 结构

Fig. 11 SFE and DPB structures

合  $X_{1~2}^{inter}$  中每一个特征的权重。随后,将权重与原始的  $X_{1~2}^{inter}$  的每一个元素相乘,并对相乘后得到的特征进行拼接,得到形状为  $C \times H \times W$  的特征。最后,对拼接后的特征进行卷变换,得到形状为  $C/2 \times H \times W$  的特征,将其作

为 DPB 的输出。以上是 DPB 对输入的特征  $X_{1-2}^{inter}$  进行处理的整个流程。本文将  $X_{1-2}^{inter}$  分别输入到两个 DPB 中实现特征的分解, 以获得用于回归任务和分类任务的特征  $DP\_R$  和  $DP\_C$ , 缓解了不同任务对特征需求不同的问题。

## 2) 分类子任务与回归子任务关联

实现共享特征分解后, 将得到的特征  $DP\_R$ 、 $DP\_C$  以及原始特征  $X_{1-2}^{inter}$  作为回归分类关联 (regress-classify cross-interaction, RCCI) 的输入, 其结构如图 12 所示, 图中  $X_{concat}$  表示特征  $X_{1-2}^{inter}$  拼接后的结果。在回归分类关联 RCCI 中, 特征  $DP\_R$  和  $DP\_C$  分别作为回归关联 (regress cross-interaction, RCI) 和分类关联 (classify cross-interaction, CCI) 的输入, 而  $X_{1-2}^{inter}$  拼接后的特征则作为 RCI 和 CCI 的共同输入, 为关联分类任务和回归任务提供了基础。RCI 和 CCI 下文将详细介绍。经过关联后, 得到的特征  $F\_Reg$  和  $F\_Cls$  用于最终的回归和分类。RCCI 缓解了因响应区域受到影响而导致检测性能下降的问题。

常规卷积的卷积核形状是固定的, 并且只会以固定的步幅和填充进行移动, 无法根据输入特征图对卷积核的形状进行动态调整<sup>[23]</sup>。相对于常规卷积, 可变形卷积在进行采样时能够动态的调整卷积核的形状, 以更贴近物体的形状和尺寸, 从而获得更准确的边缘特征。前文已探讨了目标检测中回归任务和分类任务的响应区域。因此, 在 RCI

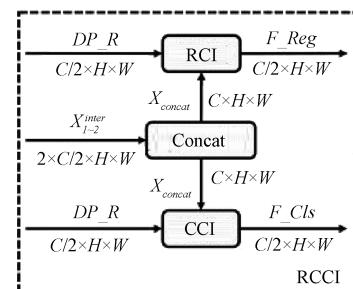


图 12 RCCI 结构

Fig. 12 RCCI structure

中使用可变形卷积 DCN, 其结构如图 13 所示。具体流程如下: 首先, 使用  $3 \times 3$  的卷积对拼接后的  $X_{1-2}^{inter}$  的进行变换, 并将得到的形状为  $27 \times H \times W$  的特征在通道维度以 2:1 的比例划分成形状为  $19 \times H \times W$  和  $8 \times H \times W$  的两部分, 其中前者保持不变作为可变形卷积的偏置项。随后, 对后者使用 Sigmoid 激活函数, 将得到的权重作为动态卷积的掩码。最后, 将得到的偏置、掩码以及特征  $DP\_R$  一同输入到可变形卷积中, 输出的结果为关联后的回归特征  $F\_Reg$ , 其形状为  $C/2 \times H \times W$ 。RCI 中使用了对共享特征  $X_{1-2}^{inter}$  进行拼接得到的特征生成可变形卷积的偏置项和掩码, 强化了特征  $DP\_R$  的边缘部分, 有利于边界框的回归任务。

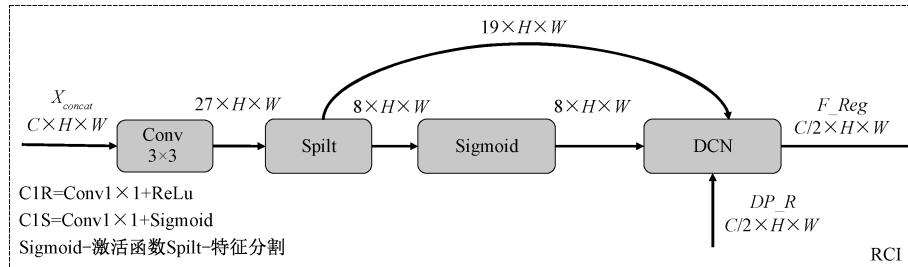


图 13 RCI 结构

Fig. 13 RCI structure

分类任务专注于目标图像中最具有特点的区域, 特征  $X_{1-2}^{inter}$  包含了全部的目标信息。因此, 在 CCI 中, 使用拼接后的  $X_{1-2}^{inter}$  对特征  $DP\_C$  进行强化, 以增强特征  $DP\_C$  中关于目标的特征信息, 其结构如图 14 所示。CCI 的输入为拼接后的特征  $X_{1-2}^{inter}$  和特征  $DP\_C$ , 输出的结果为关联后用于边界框分类任务的特征  $F\_Cls$ 。具体流程如下: 首先, 使用 C1R 对拼接后的  $X_{1-2}^{inter}$  进行变换, 特征形状由  $C \times H \times W$  变为  $C/4 \times H \times W$ 。随后, 使用 C1S 进行变换, 得到形状为  $1 \times H \times W$  的特征。最后, 将输入的特征  $DP\_C$  与其相乘, 得到的结果为关联后的用于边界框分类任务的特征  $F\_Cls$ 。其中, C1R 代表  $1 \times 1$  的卷积和 ReLu 激活函数的串联, C1S 代表  $1 \times 1$  的卷积和 Sigmoid 函数的串联。

## 1.5 损失函数

根据目标被检测的难易程度, 可将目标划分为简单样本与困难样本。由于小目标检测难度大, 因此被定义为困

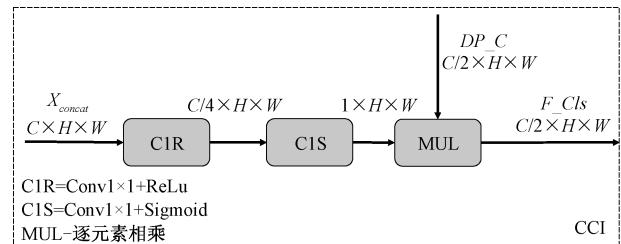


图 14 CCI 结构

Fig. 14 CCI structure

难样本。在车辆目标检测的任务中, 经常出现小目标车辆, 这会对模型的回归效果产生负面影响。Focaler-IoU<sup>[24]</sup>关注不同的回归样本, 考虑样本分布对回归任务的影响, 使用线性间隔映射的方法重新构造 IoU, 有效降低了样本分布对检测任务的影响, 提升了检测性能。计算公式如下:

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

式中:  $IoU$  表示计算得到的交并比,  $d$  和  $u$  表示预设的阈值。通过调节  $d$  和  $u$ , 可以关注不同的回归样本。在本文中,  $d$  和  $u$  使用 Focaler-IoU 原文设定的默认值。常用的  $IoU$  在锚框回归过程中会出现锚框膨胀的现象, 会影响模型的收敛过程。为解决这一问题,  $PIoU^{[25]}$  引入了适应目标大小的惩罚因子和基于锚框质量的梯度调整函数, 计算过程如图 15 所示。 $L_{IoU}$  和  $L_{PIoU}$  分别为常规方法计算得到的损失以及  $PIoU$  的计算结果。

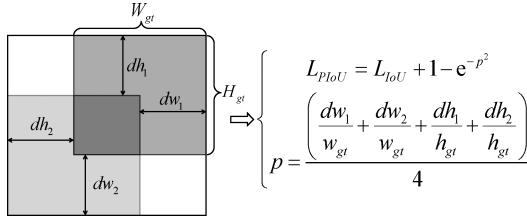


图 15 PIoU 计算过程

Fig. 15 PIoU calculation process

此外, 在训练过程中, 低质量样本也会影响模型的最终精度。WIOU<sup>[26]</sup>构造了一个非单调聚焦系数, 为高质量锚框和低质量锚框分配一个较小的梯度增益。通过这种方法, 实现了梯度的合理分配, 改善了训练效果。非单调聚焦系数的计算方式如下:

$$\begin{cases} \beta = \frac{L_{IoU}^*}{L_{IoU}} \in [0, +\infty) \\ r = \frac{\beta}{\delta\alpha^{\beta-\delta}} \end{cases} \quad (2)$$

式中:  $\beta$  为离群度, 通过损失  $L_{IoU}^*$  及其相应的滑动平均值  $\overline{L_{IoU}}$  计算得到。 $\alpha$  和  $\delta$  均为常值, 在本文中采用 WIOU 设定的默认值。非单调聚焦系数  $r$  会随着  $\beta$  动态变化, 能够在每一时刻根据当前状态制定最适合的增益分配策略。本文结合以上思想构造了 WPF-IoU, 表达式如下:

$$L = r(L_{Focaler} + 1 - e^{-p^2}) \quad (3)$$

WPF-IoU 缓解了样本分布以及样本锚框的质量对模型回归的影响, 提升了检测精度。

## 2 实验结果分析

### 2.1 数据集

本文使用的数据集包含 6 939 张图片, 训练集、验证集和测试集的比例为 7:1:2。该数据集源于清华大学智能产业研究院发布的 DAIR<sup>[27]</sup> 数据集, 涵盖了城市道路、高速公路以及路口场景。数据集由在不同光照条件下从驾驶视角拍摄的图像组成, 包含目标车辆密集、遮挡以及目标较小等真实环境中常见的情况。数据集的图像示例如图 16

所示, 展示了数据集中的部分场景。数据集中标注了 Car(乘用车)、Truck(货车)、Van(面包车)、Bus(公交车)四类车辆, 本文将这几类车辆统一归为一类。

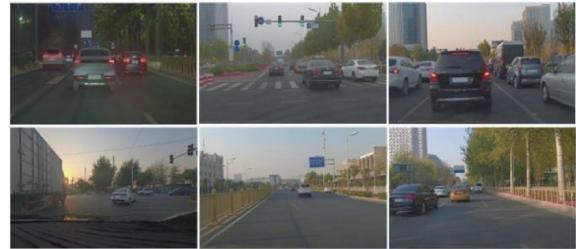


图 16 数据集图像示例

Fig. 16 Dataset image examples

### 2.2 实验环境及参数设置

本文的实验在操作系统为 Ubuntu 的云端服务器上进行, 使用两张 NVIDIA RTX3090 GPU, Python 版本为 3.8, Pytorch 的版本为 2.0.0。网络输入的图像尺寸为  $640 \times 640$ , 批处理大小设置为 16, 最大迭代次数设置为 250, 优化器使用 SGD, 初始学习率为 0.002。

### 2.3 评估指标

本文使用准确率  $P$ 、召回率  $R$  以及平均准确率  $mAP_{50}$  和  $mAP_{50:95}$  作为评价模型精度的主要指标。准确率和召回率的计算公式如下:

$$\begin{cases} P = \frac{TP}{TP + FP} \\ R = \frac{TP}{TP + FN} \end{cases} \quad (4)$$

式中:  $FP$  是模型将样本错误的预测为正样本的数量,  $FN$  是样本被模型错误的预测为反例的数量,  $TP$  是模型正确的预测结果。计算出某一类别的  $P$  和  $R$  后, 对  $P$  和  $R$  绘制成的曲线进行积分, 可以得到该类别的精度  $AP$ , 再次对所有类别的  $AP$  取平均值的结果即为平均准确率  $mAP$ 。计算公式如下:

$$\begin{cases} AP = \int_0^1 P(r) dr \\ mAP = \frac{\sum_{i=1}^c AP_i}{c} \end{cases} \quad (5)$$

式中:  $AP_i$  代表的是某一类别的准确率,  $P(r)$  表示由  $P$  和  $R$  绘制成的函数曲线,  $c$  表示类别总数, 本文中为 1。本文使用的指标  $mAP_{50}$  表示的是当  $IoU$  阈值设为 0.5 时  $mAP$  的大小, 而  $mAP_{50:90}$  则是  $IoU$  的阈值从 0.5 到 0.95 以 0.05 的步长变化时计算得到的平均精度, 计算公式如下:

$$mAP_{50:95} = \frac{1}{10} \sum_{i=0}^9 mAP_i \quad (6)$$

式中:  $mAP_i$  表示的是  $IoU$  的阈值为  $0.5 + 0.05 \times i$  时  $mAP$  对应的数值, 其中  $i$  的取值范围是 0~9。

## 2.4 消融实验

为了验证所提出模块的有效性,本文进行了消融实验,实验结果如表 1 所示。表 1 中 A、B、C、D 分别代表本文提出的 EGMEB、DFEM、检测头以及损失函数 WPF-IoU。在基线模型中添加 A 模块后,在维持参数量基本不变的前提下,计算量减少了 0.1 GFLOPs, 同时  $P$ 、 $R$ 、 $mAP50$ 、 $mAP50:95$  指标均得到提升。其原因在于对输入特征在通道维度进行均匀分割,并将分割的结果作为各分支的输入,压缩了模型的计算量。同时该模块结合了边界信息与多尺度特征,使模型抵抗背景干扰的能力得到了增强。尽管使用 B 模块后参数量和计算量有所增加,但在维持  $P$  不变的情况下, $mAP50$ 、 $mAP50:95$  以及  $R$  分别提升了 2.8%、1.7% 和 2.2%。其原因在于 B 模块能够充分的利用主干网络的每一层特征,缓解特征在网络流动过程中细节信息的损失,提升了检测器的性能。使用 C 模块后,模型的  $P$ 、 $R$ 、 $mAP50$  和  $mAP50:95$  分别提升了 0.3%、

2.1%、1.3% 和 1.7%。由于该网络与其他层级特征共享,使用 C 模块后,网络的参数量得到了降低。此外,C 模块通过特征分解,缓解了目标检测任务中边界框回归和分类子任务在遮挡情况下特征需求无法被同时满足的问题。同时,C 模块建立了两种子任务的关联,在出现车辆相互遮挡的现象时,能够缓解因两个子任务特征响应区域受影响程度不同导致的任务偏向,避免了检测性能变差的问题。使用 D 模块后,指标  $P$ 、 $R$ 、 $mAP50$  和  $mAP50:95$  分别提升了 0.5%、2%、1.7% 和 1%。这一提升的原因在于 D 模块缓解了样本分布对检测任务的影响,并且实现了训练过程中梯度的合理分配。除验证每个模块单独使用时的效果外,本文还进行了多模块组合的实验。多模块组合的实验表明了本文提出的方法的有效性,当同时使用 4 种模块时  $P$ 、 $R$ 、 $mAP50$  及  $mAP50:95$  分别达到了 87.5%、69.8%、81% 和 55.6%,相对于基准模型分别提升了 1.9%、4.1%、4.4% 和 3.3%。

表 1 消融实验

Table 1 Results of ablation experiments

基准	A	B	C	D	P	R	$mAP50$	$mAP50:95$	Params	GFLOPs
✓					85.6	65.7	76.6	52.3	3 005 843	<b>8.1</b>
✓	✓				86.1	66.4	77.2	52.8	3 005 331	8.0
✓		✓			85.5	68.5	78.3	54.5	3 897 555	11.8
✓			✓		85.9	67.8	77.9	54.0	<b>2 240 100</b>	8.6
✓				✓	86.1	67.7	78.3	53.3	3 005 843	8.1
✓	✓	✓			87.5	68.6	79.3	54.5	3 897 043	11.7
✓		✓	✓		87.0	69.4	79.5	55.3	3 135 908	12.4
✓	✓		✓		85.9	68.0	78.2	54.3	2 239 588	8.5
✓	✓			✓	87.0	67.1	78.4	53.4	3 005 331	8.0
✓		✓		✓	87.7	67.9	79.2	54.5	3 897 555	11.8
✓		✓	✓	✓	87.6	69.0	80.2	55.0	3 135 908	12.4
✓	✓		✓	✓	86.5	68.3	79.3	54.1	2 239 588	8.5
✓	✓	✓	✓		87.1	69.7	80.2	55.4	3 135 396	12.3
✓	✓	✓	✓	✓	<b>87.5</b>	<b>69.8</b>	<b>81.0</b>	<b>55.6</b>	3 135 396	12.3

## 2.5 对比试验

将本文改进后的办法与其他方法进行对比,实验结果如表 2 所示。本文改进后的办法在  $P$ 、 $R$ 、 $mAP50$  和  $mAP50:90$  四个指标上分别达到了 85.7%、69.8%、81% 和 55.6%,在表 2 中表以 Ours-n 表示,其性能优于表中其他计算量相近的模型,如 YOLOv3-tiny、YOLOv6n 和 YOLOv5s。其原因在于本文改进的模型能够更加充分的利用主干网络的各层特征,缓解特征在网络流动过程中细节信息的损失,并通过结合边界特征增强了模型抵抗背景干扰的能力。此外,检测头能够应对车辆相互遮挡时,因特征需求无法满足以及响应区域受影响导致的检测性能下降。相对于一些轻量化模型,如表中的 YOLOv5n、在其

基础上引入注意力机制的 MLCA 以及在 YOLOv8n 的检测头中引入重参数化卷积的 MIF-YOLO,虽然这些模型的计算量较小,但极大牺牲了模型的性能。进一步地,将本文提出的改进办法应用于用 YOLOv8s 模型,在表 2 中表示为 Our-s。虽然  $P$  指标维持不变,但  $R$ 、 $mAP50$  和  $mAP50:95$  性能指标得到了提升,且在部分指标上超越了计算量更大的 YOLOv8m 和 YOLOv10 m。综上所述,本文提出的方法是有效的。

此外,本文还与已有的改进 YOLOv8 的车辆目标检测方法进行了对比,他们分别是 LVD-YOLO<sup>[9]</sup>、LCD-YOLO<sup>[11]</sup> 和 RHL-YOLO<sup>[13]</sup>,实验结果如表 2 所示。

本文方法的  $P$ 、 $R$ 、 $mAP50$  和  $mAP50:95$  指标分别为

表 2 不同方法的结果对比

Table 2 Comparison of results from different methods

方法	P	R	mAP50	mAP50:95	Params	GFLOPs
YOLOv3-tiny <sup>[28]</sup>	82.5	57.9	67.2	41.6	8.66	12.9
YOLOv5n <sup>[29]</sup>	84.9	65.6	75.8	48.7	<b>1 760 518</b>	<b>4.1</b>
MLCA <sup>[30]</sup>	84.7	62.8	73.9	47.2	3 082 588	6.2
YOLOv6n <sup>[31]</sup>	85.2	64.6	74.8	50.7	4 233 843	11.8
YOLOv5s <sup>[29]</sup>	86.9	69.8	79.7	53.8	7 012 822	15.8
YOLOv6s <sup>[31]</sup>	88.3	68.6	78.9	54.8	16 297 619	44.0
YOLOv8n <sup>[32]</sup>	85.6	65.7	76.6	52.3	3 005 843	8.1
MIF-YOLO <sup>[33]</sup>	86.5	66.9	77.5	53.4	3 005 843	8.1
YOLOv9t <sup>[34]</sup>	86.2	64.7	75.9	52.0	1 970 979	7.6
YOLOv10n <sup>[35]</sup>	84.3	64.6	75.9	51.4	2 694 806	8.2
RHL-YOLO <sup>[13]</sup>	84.0	64.1	74.7	51.1	1 247 492	4.7
Ours-n	87.5	69.8	81.0	55.6	3 135 396	12.3
LVD-YOLO <sup>[9]</sup>	87.1	67.9	78.1	53.6	5 384 543	18.7
LCD-YOLO <sup>[11]</sup>	87.5	69.6	80.2	56.9	9 465 443	25.4
YOLOv8s <sup>[32]</sup>	88.3	69.8	79.9	56.3	11 125 971	28.4
YOLOv8m <sup>[32]</sup>	87.9	73.0	82.4	59.6	25 840 339	78.7
YOLOv10s <sup>[35]</sup>	88.0	69.4	80.2	55.8	8 035 734	24.4
YOLOv10m <sup>[35]</sup>	88.5	71.7	81.9	57.8	16 451 542	63.4
Ours-s	<b>88.3</b>	<b>73.4</b>	<b>83.2</b>	<b>58.8</b>	12 302 036	46.1

87.5, 69.8, 81 和 55.6, 均优于 LVD-YOLO 和 RHL-YOLO。

LVD-YOLO 颈部网络使用了加权双向特征金字塔, 虽然能够提高特征融合的能力, 但忽略了主干网络中的低级特征。而本文的方法融合了主干网络的低级特征, 有效的缓解了小目标的特征在网络流动过程中细节信息损失。此外, LVD-YOLO 引入了带有注意力机制的动态检测头, 虽然能够加强对目标的关注, 对遮挡问题有改善作用, 但并未从本质上解决遮挡问题。出现遮挡时检测性能下降的本质原因在于, 检测过程中的分类子任务和回归子任务因特征响应区域受影响程度不同, 从而导致任务偏向。本文通过关联分类子任务和回归子任务实现两者的平衡, 进而改善了检测效果。以上为本文方法的检测效果优于 LVD-YOLO 的主要原因。

RHL-YOLO 通过引入注意力机制改进的混合结构特征金字塔网络增强了特征融合能力, 改善了小目标问题, 但同样忽略了主干网络低级特征中的细节信息。本文的颈部网络充分的利用低级特征中的细节信息来补充小目标特征在网络流动过程中细节信息的损失。此外, RHL-YOLO 通过重参数化改进主干网络, 虽然减小了计算量, 但忽略了背景信息会对目标特产生干扰, 影响检测效果。而针对这一问题, 本文在主干网络部分通过边界特征与多尺度特征结合, 以提高抵抗背景干扰能力。RHL-YOLO 检测头使用了一种轻量化的结构, 无法缓解遮挡导致的检测性能下将的问题。本文通过关联分类子任务与回归子

任务, 能够改善这一问题。以上因素是本文方法在检测效果上优于 RHL-YOLO 的主要原因。

对于 LCD-YOLO, 虽然本文方法的指标 P 与其一致且 mAP50:95 指标低于 LCD-YOLO, 但指标 R 和 mAP50:95 均高于 LCD-YOLO, 并且计算量仅为 LCD-YOLO 计算量的一半。主要原因如下: 首先, LCD-YOLO 在主干网络中引入了自适应膨胀卷积, 通过动态调整膨胀率获取多尺度特征, 提高了抗干扰能力。本文方法则将多尺度特征与边缘特征结合, 具有更强的抗干扰能力。其次, LCD-YOLO 为了减小计算量, 将检测头中的部分卷积变为共享卷积, 并没有针对遮挡导致的检测性能下降问题进行优化。本文则通过关联检测过程中的分类和回归子任务, 缓解了遮挡条件下检测性能下降的问题。此外, LCD-YOLO 训练过程采用 WIoU 作为损失函数, 通过调节梯度分配优化训练过程, 而本文构造的损失函数在此基础上, 进一步缓解了样本分布对模型回归的影响。以上因素为本文方法在计算量仅为 LCD-YOLO 一半的情况下, 能够在指标 P 和 R 上超过 LCD-YOLO 的主要原因。

本文方法的优势主要体现在以下几个方面: 首先, 边缘特征保留了图像的重要结构属性, 本文的方法在主干网络中实现了多尺度特征与边缘特征的结合, 相比于对比的方法, 具有更强的抵抗背景干扰能力。其次, 针对小目标问题, 本文充分利用了主干网络的低级特征, 有效弥补了小目标特征在网络流动过程中细节信息的损失, 而不仅仅

是增强颈部网络特征融合能力。最后,本文从检测过程中分类子任务和回归子特征图响应区域不同的角度出发,通过关联这两项任务,缓解了出现遮挡时检测性能下降的问

题,而非单纯的依赖注意力机制。为便于读者比较,本文将 3 种对比方法与本文方法的改进策略及优劣性列举于表 3 中。

表 3 不同方法的改进策略以及优劣性

Table 3 Improvement strategies and pros and cons of different methods

方法	改进策略	优劣性分析
RHL-YOLO <sup>[13]</sup>	通过重参数化改进主干网络,减少计算量和参数数量。 设计轻量化检测头,减小计算量。 引入注意力机制优化多尺度特征融合网络,替换原始颈部网络。	改进后的颈部网络增强了特征融合能力,有助于小目标检测,但忽略了低级特征。虽然计算量减少,但精度有所牺牲。
LVD-YOLO <sup>[9]</sup>	主干网络使用计算量更小的 FasteNet。 颈部网络替换为双向特征金字塔。 引入融合注意力机制的动态检测头。	替换主干网络降低了计算量,但削弱了抵抗背景干扰能力。颈部网络采用双向特征金字塔增强特征融合,有助于小目标检测,但未充分利用低级特征。动态检测头增强了对目标的关注,但未有效解决遮挡问题。
LCD-YOLO <sup>[11]</sup>	主干网络引入频率自适应膨胀卷积。 检测头将部分卷积共享,以减少计算量。 训练过程使用 WIoU 作为损失函数。	主干网络引入自适应膨胀卷积,以增强多尺度特征和细节信息,改善小目标检测和抗干扰能力。颈部网络未融合低级特征,不利于小目标检测。检测头使用共享卷积,减小了计算量。
本文方法	提出了 EGMEB,旨在实现边界特征和多尺度特征融合,用于改进主干网络。设计了 DFEM,并基于该模块构建了细节特征增强 FPN,融合主干网获取的低级特征。提出了一种检测头,实现分类子任务和回归子任务的关联。结合已有损失函数,对训练过程中使用的损失函数进行了重构。	本文方法的优势主要体现在以下几点:首先,主干网络结合了多尺度特征与边界特征,有效抑制了背景干扰。其次,颈部网络融合了主干网络的低层特征,弥补了小目标特征细节的损失,其他方法未考虑这一点。最后,针对遮挡问题,本文通过关联分类和回归子任务解决任务偏向,而其他方法通过注意力机制,未能从本质上解决此问题。

为了更直观的展示本文模型改进的效果,本文将改进前后的 PR 曲线绘制在一个坐标系下,如图 17 所示。图中虚线和实线分别为基准模型和改进后模型的 PR 曲线。通过观察曲线与坐标轴围成的面积,可以看出改进后模型的性能优于基准模型。

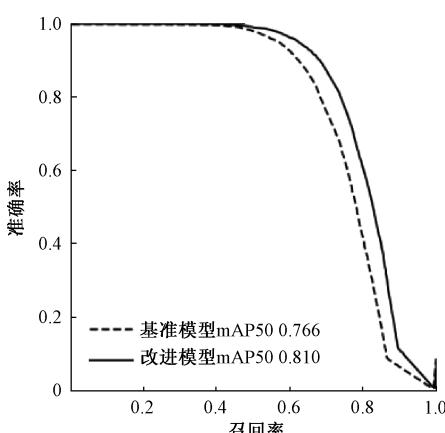


图 17 PR 曲线对比

Fig. 17 Comparison of PR curves

图 18 展示了本文提出的方法与基准方法 YOLOv8n 以及 YOLOv5s 模型的检测效果。第 1 行为原始输入图像,第 2 行和第 3 行分别为 YOLOv5s 的检测效果和热力图可视化结果,第 4 行和第 5 行分别为 YOLOv8n 的检测效果和热力图的可视化结果,其余两行为本文所提出方法的检测效果图以及热力图的可视化结果。在第 1 张图像中,存在小目标车辆。尽管 YOLOv5s 检测到了所有目标,但目标框的定位存在较大偏差,而 YOLOv8n 的目标边界框的定位较准确,但出现了漏检和错检。相比之下,本文方法的检测结果更加准确,体现了在小目标检测方面的优势。并且根据相应的热力图可以看出,本文的方法更加聚焦于目标对象。在第 2 张图像中,YOLOv5s 和 YOLOv8n 均出现了漏检,而本文的方法检测到了全部的目标且无错检。同时,根据热力图可以看出,本文的方法对目标的定位更加准确。在第 3 张图像中,背景环境较为复杂,增加了目标检测的难度。在这张图像中 YOLOv5s 出现了漏检和错检,而 YOLOv8n 仅出现了错检。相比之下,本文方法检测到了全部的目标且无错检,说明了本文方法具有较强的抵抗背景干扰的能力,通过对比热力图可以进一步证明

这一点。第 4 张图像中的场景较为复杂,出现了车辆相互遮挡的情况。YOLOv5s 将行人误识别为车辆,并且因目标车辆部分被遮挡而漏检。YOLOv8n 虽然检测到了所有目标,但在车辆密集区出现了错检,并且将路边的行人误

检为车辆。本文方法检测到了所有目标,并且没有错检,说明了本文方法具有较强的抗遮挡的能力。此外,热力图还表明本文的方法更加聚焦于待检测的目标,进一步证明了其在复杂场景中的优越性。

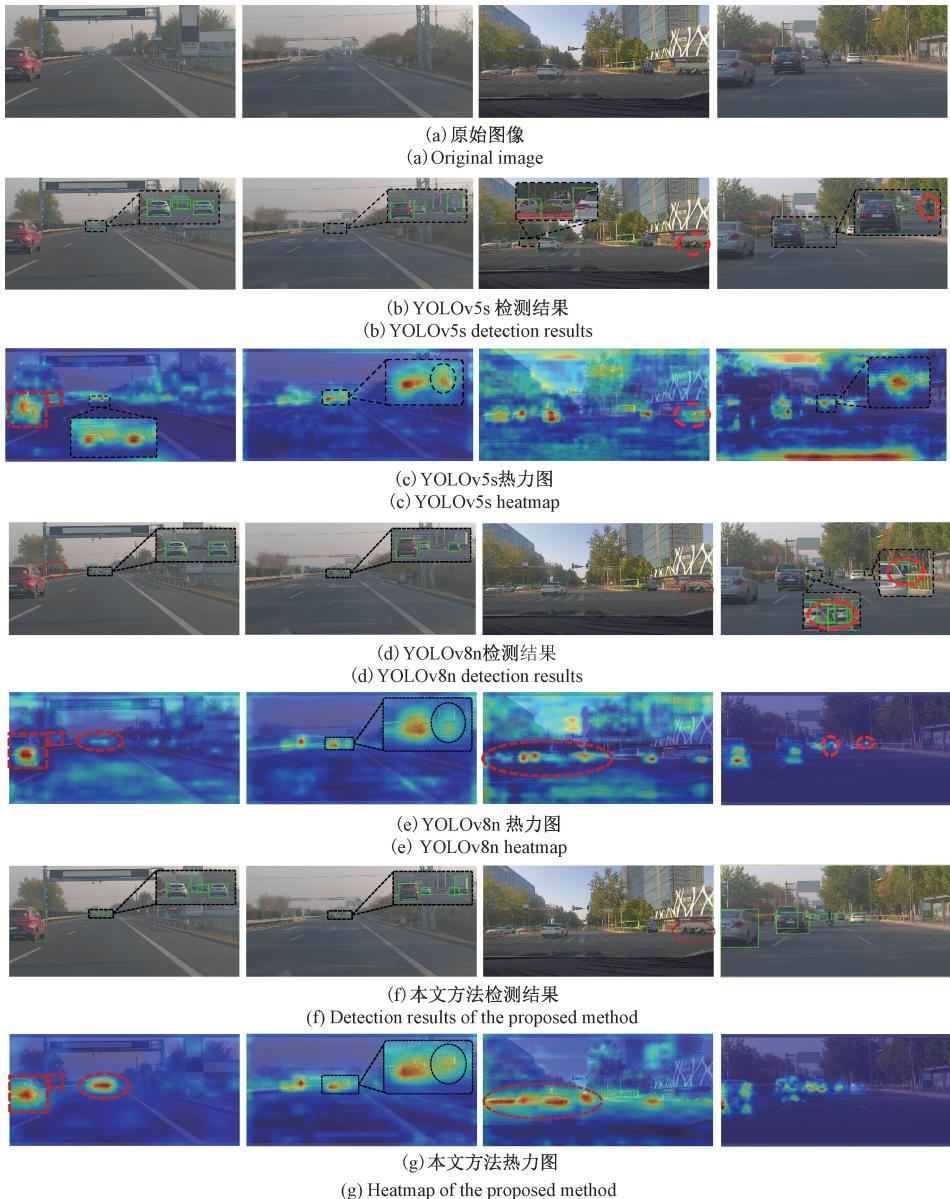


图 18 检测结果及热力图的可视化

Fig. 18 Visualization of detection results and heatmap

### 3 结 论

为解决背景特征对目标车辆特征的干扰、检测过程中出现小目标车辆以及车辆密集而出现相互遮挡导致的检测效果不佳的问题,提出了一种交通环境下的车辆检测方法。首先,结合边界特征与多尺度特征对主干网络进行重构,以增强网络抵抗背景干扰的能力,获得更加准确的目标车辆特征,进而改善检测效果。其次,对颈部网络进行

改进,充分结合主干网络提取的不同层级的特征,以缓解网络流动过程中目标细节信息的损失,提高小目标检测能力。随后,从目标检测中分类子任务和回归子任务对特征的需求,以及它们在特征图上的响应区域这两个角度,对车辆相互遮挡影响检测精度的机理进行了分析,并提出一种新的检测头以提升检测性能。最后,针对训练过程中样本分布和锚框质量对训练效果的影响,构造了一种新的 IoU 以优化训练过程。改进后的模型在  $P$ 、 $R$ 、 $mAP_{50}$  和

*mAP50:95* 指标上, 分别比原始模型提高了 1.9%、4.1%、4.4% 和 3.3%。

## 参考文献

- [1] 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.
- [2] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]. Computer Vision-ECCV, Amsterdam, Netherlands, 2016: 21-37.
- [3] 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, 2016: 779-788.
- [4] 魏子洋, 赵志宏, 赵敬娇. 改进 Faster R-CNN 算法及其在车辆检测中的应用[J]. 应用科学学报, 2020, 38(3): 377-387.  
WEI Z Y, ZHAO ZH H, ZHAO J J. Improved faster R-CNN algorithm and its application on vehicle detection [J]. Journal of Applied Science, 2020, 38(3): 377-387.
- [5] ZHANG H, SHAO F M, CHU W J, et al. Faster R-CNN based on frame difference and spatiotemporal context for vehicle detection[J]. Signal, Image and Video Processing, 2024, 18(10): 7013-7027.
- [6] CHEN Y K, LI Y W, KONG T, et al. Scale-aware automatic augmentation for object detection[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 9563-9572.
- [7] 梁天添, 杨淞淇, 钱振明. 基于改进 YOLOv8s 的恶劣天气车辆行人检测方法[J]. 电子测量技术, 2024, 47(9): 112-119.  
LIANG T T, YANG S Q, QIAN ZH M. Improved YOLOv8s method for vehicle and pedestrian detection in adverse weather [J]. Electronic Measurement Technology, 2024, 47(9): 112-119.
- [8] KANG L, LU ZH W, MENG L Y, et al. YOLO-FA: Type-1 fuzzy attention based YOLO detector for vehicle detection [J]. Expert Systems with Applications, 2024, 237: 121209.
- [9] 陈梓延, 王晓龙, 何迪, 等. 基于改进 YOLOv8 的轻量化车辆检测网络[J]. 计算机工程, 2025, 51(5), 314-325.  
CHEN Z Y, WANG X L, HE D, et al. Lightweight vehicle detection network based on improved YOLOv8[J]. Computer Engineering, 2025, 51(5), 314-325.
- [10] BAO D CH, GAO R J. YED-YOLO: An object detection algorithm for automatic driving[J]. Signal, Image and Video Processing, 2024, 18 (10): 7211-7219.
- [11] 代少升, 代佳伶, 余自安. 基于 LCD-YOLO 的车辆检测算法[J]. 半导体光电, 2024, 45(6): 1039-1046.  
DAI SH SH, DAI J L, YU Z AN. Research on vehicle detection algorithm based on lightweight car detection-YOLO[J]. Semiconductor Optoelectronics, 2024, 45(6): 1039-1046.
- [12] 霍爱清, 郭岚洁, 冯若水. 面向密集场景的多目标车辆检测算法[J]. 电子测量技术, 2024, 47 (9): 129-136.  
HUO AI Q, GUO L J, FENG R SH, Multi-objective vehicle detection algorithms for dense scenes [J]. Electronic Measurement Technology, 2024, 47 (9): 129-136.
- [13] 彭杰, 苏盈盈, 杜谦, 等. RHL-YOLOv8:一种轻量级的交通车辆检测算法[J/OL]. 电讯技术, 1-12 [2025-02-25 ]. <https://doi.org/10.20079/j.issn.1001-893x.240905002>.
- [14] PENG J, SU Y Y, DU Q, et al. RHL-YOLOv8: A lightweight traffic vehicle detection algorithm[J/OL]. Telecommunication Technology, 1-12 [ 2025-02-25 ]. <https://doi.org/10.20079/j.issn.1001-893x.240905002>.
- [15] 顾杨海, 李富, 陈德基, 等. 基于多尺度特征融合与交互的路侧目标检测算法[J]. 电子测量技术, 2024, 47(23): 152-161.  
GU Y H, LI F, CHEN D J, et al. Roadside object detection algorithm with multi-scale feature fusion and interaction[J]. Electronic Measurement Technology, 2024, 47(23): 152-161.
- [16] 向雷, 蒋文波. 改进 YOLOv8 的城市行车道路障碍物检测算法研究[J]. 电子测量技术, 2025, 48(1): 29-38.  
XIANG L, JIANG W B. Research on improved YOLOv8 urban driving road obstacle detection algorithm[J]. Electronic Measurement Technology, 2025, 48(1): 29-38.
- [17] 王雪秋, 高焕兵, 郑泽萌. 改进 YOLOv8 的道路缺陷检测算法[J]. 计算机工程与应用, 2024, 60(17): 179-190.  
WANG X Q, GAO H B, JIA Z M. Improved road defect detection algorithm based on YOLOv8 [J]. Computer Engineering and Applications, 2024, 60(17): 179-190.
- [18] HOWARD A G, ZHU M L, CHEN B, et al. MobileNets: Efficient convolutional neural networks

- for mobile vision applications [J]. ArXiv preprint arXiv: 1704. 04861, 2017.
- [18] SUNKARA R, LUO T. No more strided convolutions or pooling: A new CNN building block for low-resolution images and small objects [C]. Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Grenoble, France, 2022: 443-459.
- [19] CHEN J R, KAO SH H, HE H, et al. Run, don't walk: Chasing higher flops for faster neural networks[C]. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 2023: 12021-12031.
- [20] CUI Y N, REN W Q, KNOLL A. Omni-kernel network for image restoration [C]. 2024 AAAI Conference on Artificial Intelligence, Vancouver, Canada, 2024: 1426-1434.
- [21] SONG G L, LIU Y, WANG X G. Revisiting the sibling head in object detector[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 2020: 11563-11572.
- [22] FENG CH J, ZHONG Y J, CAO Y, et al. Tood: Task-aligned one-stage object detection [C]. 2021 IEEE/CVF International Conference on Computer Vision, QC, Canada, 2021: 3490-3499.
- [23] ZHU X ZH, HU H, LIN S, et al. Deformable convnets v2: More deformable, better results [C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 2019: 9300-9308.
- [24] ZHANG H, ZHANG SH J. Focaler-IoU: More focused intersection over union loss [J]. ArXiv preprint arXiv: 2401. 10525, 2024.
- [25] LIU C, WANG K G, LI Q, et al. Powerful-IoU: More straightforward and faster bounding box regression loss with a nonmonotonic focusing mechanism [J]. Neural Networks, 2024, 170: 276-284.
- [26] TONG Z J, CHENG Y H, XU Z W, et al. Wise-IoU: Bounding box regression loss with dynamic focusing mechanism[J]. ArXiv preprint arXiv: 2301. 10051, 2023.
- [27] YU H B, LUO Y ZH, SHU M, et al. Dair-v2x: A large-scale dataset for vehicle-Infrastructure cooperative 3D object detection[C]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 2022: 21329-21338.
- [28] REDMON J, FARHADI A. Yolov3: An incremental improvement[J]. ArXiv preprint arXiv: 1804. 02767, 2018.
- [29] GLENN J, ALEX S, JIRKA B, et al. Ultralytics/YOLOv5 [EB/OL]. (2022-06-18) [2022-06-22]. <http://github.com/ultralytics/YOLOv5>.
- [30] WAN D H, LU R SH, SHEN S Y, et al. Mixed local channel attention for object detection[J]. Engineering Applications of Artificial Intelligence, 2023, 123: 106442.
- [31] LI CH Y, LI L L, JIANG H L, et al. YOLOv6: A single-stage object detection framework for industrial applications [J]. ArXiv preprint arXiv: 2209. 02976, 2022.
- [32] JOCHER G, QIU J, CHAURASIA A. Ultralytics YOLO(Version8.0.0)[EB/OL]. 2023[2024-06-28]. <https://github.com/ultralytics/ultralytics>.
- [33] WAN D H, LU R SH, HU B T, et al. YOLO-MIF: Improved YOLOv8 with multi-information fusion for object detection in gray-scale images[J]. Advanced Engineering Informatics, 2024, 62: 102709.
- [34] WANG C Y, YEH I H, MARK LIAO H Y. YOLOv9: Learning what you want to learn using programmable gradient information [C]. 2024 European Conference on Computer Vision, Milan, Italy, 2024: 1-21.
- [35] WANG A, CHEN H, LIU L H, et al. YOLOv10: Real-time end-to-end object detection [J]. ArXiv preprint arXiv: 2405. 14458, 2024.

## 作者简介

**李天林**,硕士研究生,主要研究方向为深度学习、计算机视觉。

E-mail:LiTianlin98524@163.com

**安毅**(通信作者),博士,教授,主要研究方向为三维环境感知与认知、模式识别与人工智能等。

E-mail:anyi@dlut.edu.cn

**陈岩**,硕士研究生,主要研究方向为人员异常行为识别。

E-mail:2275761529@qq.com