

DOI:10.19651/j.cnki.emt.2416964

基于多尺度特征融合与交互的路侧目标检测算法^{*}

顾杨海^{1,2} 李富^{2,3} 陈德基^{2,3} 王泉²

(1.南京信息工程大学计算机学院 南京 210044; 2.同济大学嵌入式系统与服务计算教育部重点实验室 上海 201804;
3.无锡学院物联网工程学院 无锡 214105)

摘要: 针对路侧视角下目标检测任务中,小目标密集,多尺度变化以及复杂天气背景干扰等挑战,提出基于多尺度特征融合与交互的目标检测算法——MF-YOLO。设计 C2f-CAST,通过星操作将来自不同子空间的特征进行交互与变换,并引入 MLCA 捕捉远距离像素之间局部,全局特征以及通道和空间特征,多尺度信息聚合加强对遮挡目标显著语义信息关注,消除背景影响;针对颈部层在上下文信息融合效率较低的问题,加入轻量级卷积 GSConv 对传统卷积进行优化,并设计跨级部分网络模块,降低模型复杂度和参数量。构造跨层级融合模块 SDFM,对浅层特征图进行自校准操作,并融合深层特征图语义信息,解决小目标漏检的问题;最后,设计基于自适应惩罚因子和锚框质量的梯度调整函数,并结合动态聚集机制改进的 WPIoU 损失函数,提升边界框回归性能和检测鲁棒性。实验结果显示,MF-YOLO 在 DAIR-V2X-I 和 UA-DETRAC 数据集上 mAP@0.5 指标分别达到 85.1%,92.3%,与原 YOLOv8s 相比分别提升 4.4%和 1.8%,计算量 GFLOPs 下降了 19.8%,参数量下降 8.18%。检测速度达到 152 fps,满足实时要求。

关键词: 路侧图像;星操作;特征融合;目标检测;注意力机制

中图分类号: TP391;TN791 **文献标识码:** A **国家标准学科分类代码:** 520.60

Roadside object detection algorithm with multi-scale feature fusion and interaction

Gu Yanghai^{1,2} Li Fu^{2,3} Chen Deji^{2,3} Wang Quan²

(1. School of Computer Science and Technology, Nanjing University of Information Science and Technology, Nanjing 210044, China;

2. Key Laboratory of Embedded System and Service Computing Ministry of Education, Tongji University, Shanghai 201804, China;

3. School of Internet of Things Engineering, Wuxi University, Wuxi 214105, China)

Abstract: In view of the challenges of dense small targets, multi-scale variations, and complex weather background interference in roadside perspective target detection tasks, a multi-scale feature fusion and interaction-based target detection algorithm, MF-YOLO, is proposed. Design C2f-CAST, interact and transform features from different subspaces through star operations, and introduce MLCA to capture local, global, channel, and spatial features between distant pixels. Multi-scale information aggregation enhances attention to significant semantic information of occluded objects and eliminates background influence; to address the problem of low efficiency in context information fusion for the neck layer, we add lightweight convolution GSConv to optimize traditional convolution, and design a cross-level partial network module VoV-GSCSP to reduce model complexity and parameter count. Construct a cross-level fusion module SDFM to perform self-calibration on shallow feature maps and fuse semantic information from deep feature maps to solve the problem of missed detection of small targets; finally, the design is based on an adaptive penalty factor, a gradient adjustment function for anchor box quality combined with a dynamic clustering mechanism to improve the WPIoU loss function, enhancing the performance of bounding box regression and detection robustness. The experimental results show that MF-YOLO achieves mAP@0.5 of 85.1% and 92.3% on DAIR-V2X-I and UA-DETRAC datasets, respectively, which is 4.4% and 1.8% higher than the original YOLOv8s, with a reduction of 19.8% in computational complexity and 8.18% in parameter count. The detection speed reaches 152 fps, meeting the real-time requirements.

Keywords: roadside images; star operation; feature fusion; object detection; attention mechanism

0 引言

随着智能网联技术和自动驾驶技术的发展应用,车路

协同系统已经成为提升交通安全和效率的关键驱动力之一^[1]。车路协同系统通过整合车、路、设施等信息,实现智能交通的高效管理。其中,路侧感知技术利用摄像机、毫米

收稿日期:2024-09-23

* 基金项目:国家自然科学基金青年项目(42305158)资助

波雷达等传感器,实现对该路段的交通参与者、路况等的瞬时智能感知^[2]。然而,在日益复杂的交通道路场景中,系统需要准确识别和定位各种动态和静态的目标,如车辆、行人、交通标志等。同时,路侧设备受限于边缘计算服务器的计算资源,对检测的实时性提出更高要求。因此,研究一种平衡精度和轻量化的路侧目标检测算法具有实际意义。

早期车路协同系统中的路侧目标检测算法包括:光流法、帧间差分法、背景差分法。这些传统方法主要基于手工设计的特征和分类器,但在复杂场景下会有特征工程复杂,检测精度有限等问题。为解决以上问题,深度学习方法被应用于车路协同系统中的目标检测任务。深度学习领域的目标检测算法大致划分为两个流派:一是采用候选区域策略的两步检测法,以 R-CNN 系列为代表;二是直接基于回归的单步检测算法,代表算法是 YOLO 系列(you only look once)后者因其检测速度快,逐步成为主流。

YOLO 系列包含 3 个部分:主干网络、颈部网络和检测头。Jocher 等^[3]提出 YOLOv5,引入了 CSPNet 来取代 DarkNet,并结合了数据增强策略、增强的 PAN 结构以及更多的模型尺度选择等。Li 等^[4]设计 YOLOv6,针对颈部和主干网络分别提出了 BiC 和 SimCSPPPF,引入了基于锚点的训练和自蒸馏策略。Wang 等^[5]提出 YOLOv7 通过引入 E-ELAN 以增强梯度流路径,并探索了多个可训练的方法。Jocher 等^[6]提出 YOLOv8,引入了 C2f 构建块,用于高效的特征提取与融合。Wang 等^[7]提出 Gold-YOLO,设计先进的 GD 机制,增强了多尺度特征融合能力。Wang 等^[8]提出 YOLOv9,针对 GELAN 架构改进,和 PGI 训练过程增强机制。Wang 等^[9]设计 Mamba-YOLO,引入状态空间模型 SSM,并设计 ODSSBlock。Wang 等^[10]提出 YOLOv10,采用双重分配策略解决了后处理中的冗余预测问题,并结合双重标签分配技术实现了无 NMS。

在检测精度方面,卫策等^[11]提出了一种轻量化解耦预测头,用于分类和回归不同尺寸的特征层,提升了小目标在道路场景中的检测能力,但随着模型复杂度的增加,计算资源消耗随之增加。许德刚等^[12]改进了 YOLOv8 的城市车辆检测算法,设计了 GAM-C2f 模块,以平衡计算效率和准确性,并加入了 SPPFAPGC 模块,避免最大池化操作引起的局部特征丢失。然而,该方法在低光和雾天等复杂场景中的检测效果仍需进一步验证。梁天添等^[13]针对恶劣天气中的图像模糊和光照不均的问题,改进了 C2f 结构,融入了区域残差化和语义残差化 DWR 模块,捕捉不同尺度的信息,并通过单分支解耦结构优化检测头,提升了推理速度。然而,该方法在恶劣天气下的小目标检测中仍存在漏检问题,需要进一步优化。刘若璿等^[14]提出了一种基于多尺度序列融合的轻量级路侧目标检测算法 MQ-YOLO,将 NWD 损失函数引入边界框回归损失中,降低对目标尺度和小目标框位置偏差的敏感性,并设计多尺度序列融合结构 PSF,实现多层级的特征交互融合,但在高性能设备上的

表现略有不足。

在轻量化方面,古佳欣等^[15]提出 YOLOv8-DEL,结合 GroupNorm 和共享卷积,设计轻量化的检测头,并使用动态组卷积 Shuffle Transformer 模块替代 C2f 模块,提升了模型的特征提取能力并减轻网络的计算负担,但该算法仅在 BDD100K 和 KITTI 数据集上进行了实验,缺乏在极端天气以及复杂交通场景中的进一步验证。王泉等^[16]在 YOLOv8s 中引入 FasterBlock 替代 C2f 中的瓶颈组件,减少了模型的浮点运算量和参数量,并通过通道剪枝修剪网络中的冗余连接,缩小模型规模,提升检测速度。但由于该研究采用的是单一路口的自制数据集,场景和目标相对单一,数据集规模有限,因此模型在多目标检测和复杂场景中的泛化能力仍有待进一步优化。闫建红等^[17]在 YOLOv8s 的颈部网络设计全新的聚焦扩散特征金字塔结构,使每层特征图都聚焦三层特征信息,减少特征冗余,提升模型的检测速度,并设计共享卷积检测头,其在无人机小目标视角下提高了实时性,但在路侧复杂场景中的检测鲁棒性仍需验证。

尽管上述研究取得了一定的进展,但在面对复杂路侧视角下无法同时平衡模型的精度和实时性,存在漏检误检率高,实时性差,泛化能力弱等问题。因此,本文在 DAIR-V2X-I 和 UA-DETRAC 数据集上实验,提出一种基于 YOLOv8s 的改进算法。本文的主要贡献有:

受 Ma 等^[18]提出的 StarNet 的启发本文在骨干网络中设计一种星注意力模块替换 C2f 的 Bottleneck 模块,改进的 C2f-CAST 通过星操作将来自不同子空间的特征交互与变换,再添加混合局部通道注意力 MLCA^[19],解决在低照度路侧环境下,对遮挡模糊目标误检和漏检问题,相比已有方法^[12],更加注重全局与局部多尺度信息的聚合,提升对遮挡和小目标的特征提取。

本文在已有方法的基础上进一步优化^[17],对颈部网络进行重构,使用组混洗卷积 GSConv^[20]替换标准卷积 Conv,设计跨级部分网络模块 VoV-GSCSP 模块,提高上下文信息的融合效率,降低模型参数量和计算量。

重写基于通道-空间注意力机制表面细节融合模块融合模块 SDFM^[21]对浅层特征图进行自校准操作,并对深层特征进行自适应的加权融合,缓解因层数堆叠而丢失的小目标细节特征。与已有方法相比^[6],本文侧重于细粒度的小目标检测。

设计 WPIoU 损失函数,使用自适应惩罚因子和基于锚框质量的梯度调整 PLOUv2^[22]来改进使用动态非单调机制和梯度增益分配策略的 WIoUv3^[23]损失函数,解决收敛速度慢和锚框放大问题并过滤低质量锚框。

1 本文方法

本文深入研究并改进 YOLOv8s 网络模型,主要改进在主干网络中设计特征交互模块、重构特征强化网络,添加

跨层特征融合模块、优化损失函数等。通过以上改进本文解决了路侧感知系统面对复杂交通环境时常见的小目标检测困难，特别是低光、遮挡以及计算资源受限的真实场景，实现了在复杂条件下的高效、多目标实时检测，为智能网联汽车和智慧交通提供了更加可靠的技术支持。相对于改进后的网络结构如图 1 所示，蓝色模块是原 YOLOv8 网络的模块，彩色模块是本文改进的模块，上标是改进的已有模块。

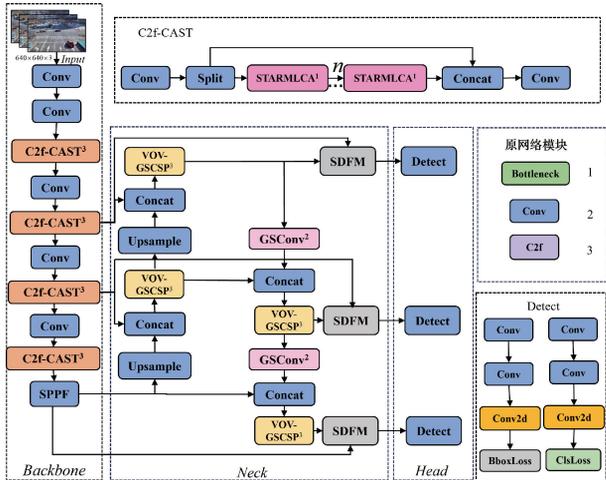


图 1 改进的网络结构

Fig. 1 Improved network architecture

1.1 特征交互模块

在 YOLOv8 骨干网络中，C2f 模块通过与 Conv 模块、Bottleneck 等其模块协同工作，完成对输入图像的特征提取任务。C2f 中的卷积层和其他特征提取层具有局部感受野，在遮挡情况下，局部感受野无法捕获到足够的特征，遮挡不仅会导致特征信息的直接损失，还会引起特征之间的混叠。且 C2f 模块计算量与参数量较大，影响模型的检测速度。所以本文设计轻量级 C2f-CAST，该模块能够更全面地捕捉图像中的特征信息，增强遮挡目标显著语义特征，并降低整体的计算量与参数量。

受 Starnet 的启发，设计一种融入元素级乘法（星操作）的 STARMLCA 模块，并替换掉 C2f 模块中的 Bottleneck 层，其原理如图 2 所示。首先进入深度可分离卷积进行特征提取，深度卷积通过独立处理每个通道来提取空间特征，逐点卷积则通过跨通道组合这些特征来生成新的特征表示。两个轻量的 GSCov 组卷积层进行特征的映射变换，通道信息被均匀打乱，有效保留通道信息，增强语义信息。接着使用星操作来交互这些特征，星操作通过元素级乘法将来自不同子空间的特征进行深度整合，在不拓宽网络结构的前提下，将输入数据映射至高维、非线性的特征空间，这种特征变换和交互有助于增强网络的非线性能力，从而捕捉更复杂的特征关系未引入任何额外计算成本，实现了特征表示能力与计算效率的提升。

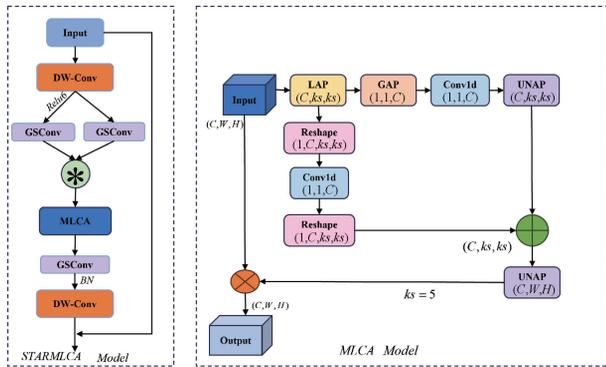


图 2 STARMLCA 和 MLCA 模块结构

Fig. 2 The structure of STARMLCA and MLCA module

星操作更侧重于局部信息提取，因此引入混合局部通道注意力机制 MLCA 可以进一步帮助模型捕捉远距离像素之间通道，空间，局部和全局信息。这种多尺度的信息聚合，有助于模型更好地理解目标与周围环境之间关系，增强空间信息捕获能力，提高对目标显著关注。MLCA 通过局部平均池化等操作，能够关注到遮挡目标所在局部区域，从而强化对这些区域的特征提取。结合全局信息，能够在保持整体上下文感知的同时，突出遮挡目标的特征。MLCA 注意力模块的原理如图 2 所示，在 MLCA 中，输入特征向量首先经历两步池化过程。第一步，通过局部池化操作，将输入特征转换为 $1 \times C \times ks \times ks$ 的向量形式，捕获局部空间信息。随后，输入特征被分流到两个并行处理的分支中。第 1 个分支专注于提取全局信息，而第 2 个分支则负责处理局部空间信息。两个分支在各自完成一维卷积操作后，通过反池化技术恢复原始分辨率，确保信息在维度上保持一致。最终，两个分支信息在融合阶段结合，形成混合注意力机制，从而实现对特征的高效利用和增强。其中，GAP（全局平均池化）是一种将特征图调整为 1×1 大小的自适应平均池化操作。与之相对，LAP（局部平均池化）不关注整个特征图的总特征，而是将其划分为 $k \times k$ 个特征模块，并对每个模块进行平均池化操作。这种局部化的处理方式有助于捕捉更精细的特征信息。而 UNAP（反平均池化）则专注于根据图像属性调整尺寸。它根据每个补丁所分配的权重来填充值，实现对特征图的尺寸调整。星操作在 d 维空间内实现了高效的计算，MLCA 仅增加少量参数量，所以 C2f-CAST 让特征提取网络更加轻量且高效。

1.2 重构颈部网络

对于路侧计算终端设备，计算资源相对有限，因此采用轻量化的卷积设计尤为重要。为解决这样的问题，LiHulin^[20]提出了 GSCov 轻量化卷积模块，主要结构如图 3 所示。该模块融合混洗技术与卷积架构实现了传统卷积与深度可分离卷积信息流的深度交互。首先，输入特征图（通道数为 $C1$ ）通过一标准卷积层初步处理，将通道数缩减至 $C2/2$ ，以精简信息。随后，此特征图进入深度可分离

卷积模块,其通道数维持不变,专注于空间特征的提取。将标准卷积与深度可分离卷积的输出进行 Concat(级联)操作,实现信息的初步整合。接着,引入混洗机制,对级联后的通道进行随机重排,这一过程不仅确保了通道信息的均匀分布,还促进了不同来源特征间的交互与融合,有效增强了特征图的语义丰富度和表达能力。本文依据 slim-neck 设计范式,在 Neck 层用 GSConv 代替标准卷积,以降低参数数量和计算量,同时在最大程度上保证采样效果。

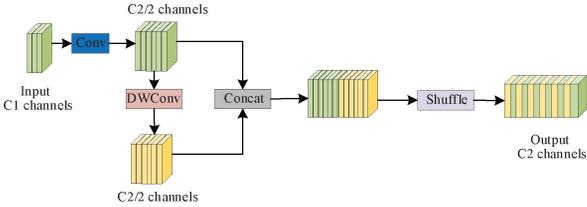


图 3 GSConv 模块结构

Fig. 3 The structure of GSConv module

GSConv 虽然降低了计算复杂度,但要同时减少推理时间并保证准确性,需要在颈部引入更先进的模型结构。为此,引入基于 GSConv 的跨级部分网络模块(VoVGSCSP),替代原有的 C2f 模块。图 4 展示了 VoVGSCSP 的结构。VoVGSCSP 模块进一步增强特征信息的融合与表达能力,为 MF-YOLO 的性能提升奠定了坚实基础。

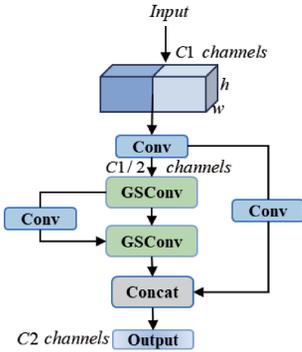


图 4 VOVGSCSP 模块的结构

Fig. 4 The structure of VOVGSCSP module

1.3 多尺度特征融合层

在特征金字塔结构中,随着下采样操作的逐步深入,多层卷积堆叠,特征图的细节信息逐步丢失,而语义信息逐步增强。浅层特征图具有丰富的纹理和边缘等特征,因此浅层网络更适合提取小尺度目标的细节信息。在路径增强特征融合金字塔网络 PANet 中,自顶向下的特征融合方式,可能会让原本像素较少的小目标在经过上采样与逐层融合后,特征分辨率降低,导致小尺度目标的特征难以被提取。为了更好地利用浅层特征图的细节语义信息,引入一个基于通道-空间注意力机制的表面细节融合模块 SDFM,对浅层特征图进行自校准操作,并根据通道和空间位置的重要性对深层特征图进行自适应的加权融合。在添加 SDFM

表面细节融合层时,若直接添加在骨干网络特征提取层后,会引入噪声导致性能下降和计算复杂度升高,所以将该层添加到强化特征提取网络层后面。此时,深层的特征图经过一系列卷积堆叠和降采样的操作,原始的特征已经被压缩,并且冗余信息进一步减少。SDFM 内部包含多个卷积层、池化层、激活函数,增加一定参数数量和计算复杂度,因此,在 SDFM 模块中引入 GSConv 组卷积替换点卷积。改进的 SDFM 的原理如图 5 所示。

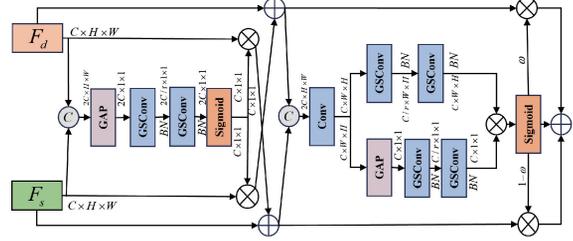


图 5 SDFM 模块架构

Fig. 5 The structure of SDFM

在通道注意力模块内,利用卷积与池化操作生成注意力权重。随后,这些权重通过元素级乘法操作被应用于浅层特征上,并将得到的特征添加到来自另一个分支的浅层特征上,以增强它们的表示能力。特征强化过程可以总结如式(1)和(2)。

$$F'_d = F'_d \oplus (F'_s \otimes \delta(\text{GSC}^n(\text{GAP}(C(F'_s, F'_d)))))) \quad (1)$$

$$F'_s = F'_s \oplus (F'_d \otimes \delta(\text{GSC}^n(\text{GAP}(C(F'_d, F'_s)))))) \quad (2)$$

其中, F_d 表示深层特征, F_s 表示浅层特征。操作符 \oplus 表示对特征进行逐元素的求和,操作符 \otimes 用于逐元素的乘法, GSC^n 表示 n 个级联的组卷积层, $C(\cdot)$ 表示一个通道维度上的连接操作,在特征处理过程中, $\delta(\cdot)$ 即 sigmoid 函数来进行非线性激活, $\text{GAP}(\cdot)$ 表示全局平均池化来减少特征图的空间维度。优化后的特征通过通道维度加以组合,随后并行地传递给通道注意机制和空间注意机制模块,以实现融合权重生成。

1.4 优化损失函数

目标检测中边界框回归性能受损失函数的影响。YOLOv8 采用的 $\text{CIoU}^{[24]}$ 损失函数对于小标注框和低质量标注框的回归,仍面临一定挑战。Focal $\text{EIoU}^{[25]}$ 通过引入焦点损失来优化这一问题,但其单调聚焦函数并未充分发挥聚焦机制的潜力。最新研究 WIoU 则提出一种动态非单调的 FM 损失函数,专注于中高质量的锚框,并有效抑制低质量样本的有害梯度。因此本文引入 WIoUv3 作为模型的边界框损失函数, WIoUv3 相较于 WIoUv1 增加梯度增益分配策略。

WIoUv3 计算公式如式(6)~(8)所示。

$$L_{\text{WIoUv3}} = rL_{\text{WIoUv1}} \quad (3)$$

$$L_{wIoUv3} = \exp\left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_{gt}^2 + H_{gt}^2)^*}\right)(1 - IoU) \quad (4)$$

$$r = \frac{\beta}{\delta \alpha^{\beta-\delta}} \quad (5)$$

式中： r 是根据离群度 β 动态计算的梯度增益， α 和 δ 是用于控制聚焦强度的超参数，本文实验中分别将 α 和 δ 设置为 1.7 和 2.7。 W_{gt} 和 H_{gt} 分别表示最小锚框的宽和高， x 和 y 分别表示预测框中点的横纵坐标， x_{gt} 和 y_{gt} 分别表示真实框的中心点坐标。为防止产生阻碍收敛的梯度，上标 * 表示将 W_{gt} 和 H_{gt} 从梯度计算中分离出来。

基于 IoU 损失函数系列受到不合理的惩罚因素的影响，导致锚框在回归过程中膨胀，显著减缓了收敛速度，为解决锚框增大问题。本文借鉴 PIOUSv2^[14] 损失函数思想，PIOUSv2 结合了自适应的惩罚因子和基于锚框质量的梯度调整函数，有效解决了锚框在回归过程中放大的问题。此外，PIOUSv2 通过单个超参数控制的非单调注意力层，显著增强了其对中等质量锚框的聚焦能力，进一步提升目标检测的准确性。预测框与真实框示意图如图 6 所示。

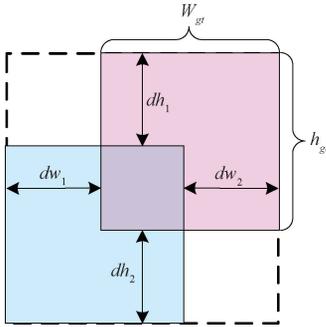


图 6 预测框与真实框示意图

Fig. 6 Diagram of predicted box and ground truth box

为解决锚框放大问题，PIOUSv2 提出一个适应目标大小的惩罚因子 P 。惩罚因子定义如式(7)：

$$P = \left(\frac{dw_1}{W_{gt}} + \frac{dw_2}{W_{gt}} + \frac{dh_1}{H_{gt}} + \frac{dh_2}{H_{gt}}\right)/4 \quad (6)$$

式中： P 是惩罚因子， dw_1, dw_2, dh_1, dh_2 是预测框和目标框对应边缘之间距离的绝对值。 W_{gt}, H_{gt} 是预测框和目标框对应边缘之间距离的绝对值。

结合 WIoUv3 和 PIOUSv2 的优点，设计 WPIOUS。加快边界框的回归过程的同时，使边框回归更为精确，并更好地适应路侧目标的特征。

基于以上惩罚项 P ，设计惩罚函数，得到下面的 WPIOUS 的损失函数，在 WPIOUS 损失中添加一个注意力层，增强对中等和高质量锚框的聚焦，注意力函数和 WPIOUS 损失定义式如下：

$$u(x) = 3x \cdot e^{-x^2} \quad (7)$$

$$L_{wPIoU} = u(\lambda e^{-P}) \cdot (L_{wIoUv3} + 1 - e^{-P^2}) \quad (8)$$

式中： $u(x)$ 代表非单调聚焦函数， λ 是控制注意力函数行为的超参数，在实验中设置为 1.3。惩罚因子替换成 e^{-P} 。

2 实验分析

2.1 实验环境与参数配置

实验环境：Nvidia GeForce RTX A6000，训练平台 Python 3.9，深度学习框架 PyTorch 2.0.1。模型训练参数设置：Epoch 大小为 200，批大小 batch_size 为 32，初始学习率是 0.001，循环学习率是 0.1，其余是默认值。

2.2 实验数据集

实验选用两个公共路侧视角下数据集。首先，采用清华大学开源路端数据集 DAIR-V2X-I，该数据集专注于车-路协同感知任务，提供车载视角和路侧视角的多源数据集，涵盖多种类别的对象，这些对象在同一类别中展现不同形态差异，为模型提供丰富的训练样本，其中夹杂多种背景干扰，包括不同的天气条件、光线变化、拥堵情况、车辆遮挡等。从中选择了 8 634 张作为训练集，1 129 张作为验证集，以及 1 050 张作为测试集。目标的类别共分为 9 类，分别是 Car、Van、Truck、Bus、Pedestrian、Cyclist、Motorcyclist、Barrow 以及 Tricyclist。

此外，为了验证本文提出算法鲁棒性，采用 UA-DETRAC 基准数据集。该数据集涵盖了多云、夜间、晴天和雨天 4 种天气条件，以及 4 种不同类型的车辆(car, bus, van, others)。由于 UA-DETRAC 数据集相邻视频帧中的车辆差异和背景变化并不显著，因此，为避免数据冗余，本文通过每 10 帧提取 1 帧的方式，最终选取了 10 870 张车流量较大、背景变化复杂的图像。随后按照 8:1:1 的比例对数据集进行了划分，以确保训练、验证和测试集的均衡分布。

相比于别的文献中使用的 BDDK100 和 KITTI 等使用单一静态车载视角的数据集，场景多集中在单一的城市街道和高速公路等宽车道，DAIR-V2X-I 和 UA-DETRAC 进一步补充了路侧视角下更加复杂和多样化的场景，车辆遮挡、夜间行驶、低照度环境、交叉路口，乡村街道等场景。更加适用于道路监控、车路协同等应用任务。

2.3 评价方法

本文采用一系列综合指标来全面评估模型的整体性能，包括平均精度均值(mean average precision, mAP)、网络参数量(Parameters)、网络计算量(GFLOPs)以及每秒检测图片数量(frames per second, fps)，作为有效性和实时性的指标。

精确率(P)和召回率(R)的计算公式如式(9)、(10)。

$$precision = \frac{TP}{TP + FP} \quad (9)$$

$$recall = \frac{TP}{TP + FN} \quad (10)$$

AP 是评估模型对于某一类别样本的识别精度，mAP

是模型对于所有类别样本的平均识别精度,其计算公式如式(11)所示。

$$mAP = \frac{1}{N} \sum_{n=0}^1 \int_0^1 P_n(r) dr \quad (11)$$

2.4 对比实验

为评估模型性能,使用 UA-DETRAC 和 DAIR-V2X-I 数据集,依据 2.3 节中指标作为评价基准,在相同的环境下与当前主流的目标检测算法进行对比实验。实验结果如表 1 所示,首先将模型与经典检测模型 Faster-RCNN^[26]、RT-DETR-R18^[27]对比。在 YOLO 目标检测模型部分则选取与本文改进模型参数量规模接近的模型 YOLOv6-s、YOLOv7-tiny、YOLOv5-S、YOLOv8-N、YOLOv8-S、YOLOv8-M 进行对比。以及与目前流行的先进改进算法 YOLOv9-S、YOLOv10-S、Gold-YOLO-S 模型 Mamba-

YOLO-T 进行对比。上述对比模型在多目标检测和车载领域都有的广泛应用场景,且针对资源受限设备的适应性,能够为评估模型的实际性能提供较为全面的参考。由表 1 可得,双阶段算法 Fast R-CNN 远比单阶段检测的性能低,在数据集上分别生成候选框和进行目标分类以及位置回归,导致产生大量的冗余框,从而增加计算负担不能满足实际的检测要求。RT-DETR-R18 系列虽然检测性能表现良好,但是参数量和计算量不满足实时部署的要求;主流的 YOLO 系列算法虽然精度有所提升,但在参数量、计算量等综合方面都远不如本文的改进算法。而与目前流行的先进改进算法 Mamba-YOLO-T、Gold-YOLO-S、YOLOv9-S、YOLOv10-S 相比,在 DAIR-V2X-I 数据集和 UA-DETRAC 数据集平均精度分别提升了 0.9%、1.8%、1.7%、1.2%和 0.5%、1.2%、0.8%、0.7%。

表 1 不同模型对比实验结果

Table 1 Comparison of experimental results of different models

模型	Params	GFLOPs	mAP@0.5UA	map@0.5Ve
YOLOv5-S ^[3]	9.11	15.8	84.6	79.1
YOLOv6-S ^[4]	16.30	65.0	88.6	79.4
YOLOv7-tiny ^[5]	6.00	132.5	90.8	80.3
YOLOv8-N ^[6]	3.01	8.2	82.4	78.5
YOLOv8-S ^[6]	11.10	28.8	90.5	80.7
YOLOv8-M ^[6]	25.90	79.4	91.7	81.4
Gold-YOLO-S ^[7]	21.52	30.5	91.2	83.3
YOLOv9-S ^[8]	8.90	27.8	91.5	83.5
Mamba-YOLO-T ^[9]	7.10	16.6	91.8	84.2
YOLOv10-S ^[10]	8.40	21.6	91.6	83.9
Faster-RCNN ^[26]	28.48	829.5	87.7	78.8
RT-DETR-R18 ^[27]	20.40	64.0	91.7	82.3
本文	10.210	23.1	92.3	85.1

通过比较图中各模型的 mAP@0.5 值,可以得出结论,MF-YOLO 算法相对于其他基准模型在路侧场景下表现出更出色的检测性能。

综上可知,本文改进模型实现更好的检测精度。图 7(a)和图 8(a)展示了基础模型在 DAIR-V2X-I 数据集和 UA-DETRAC 上每个类以及总的 mAP@0.5 值和 PR 曲线,图 7(b)和图 8(b)展示了改进后的模型在 DAIR-V2X-I 数据集和 UA-DETRAC 上每个类以及总的 mAP@0.5 值和 PR 曲线,每个类别对应的曲线按数值排列,标注在图中。从曲线的面积可知,每个类实现不同程度的提升,彰显改进模型良好的性能。

2.5 消融实验

为验证不同改进模块对模型提升程度,设计消融实验,使用 DAIR-V2X-I 数据集进行测试和评价。消融实验

结果如表 2 所示。其中√表示有加入该模块,一则表示没有加入该模块。表 2 的消融实验结果证明各改进模块的有效性。改进后模型在平均精度 mAP@0.5 上和检测速度上均有所提升,且模型的参数量和计算复杂度有所下降,平衡了检测性能和速度。

组②结果表明,使用 STARMLCA 替换 C2f 的 Bottleneck 模块,在评价指标 mAP@0.5 和 FPS 分别提高了 1.5%、8.3%,计算量 GFLOPs 下降 3.5 G,参数量下降了 130 万,特征提取性能提高,并使模型更加轻量高效。组③结果显示对颈部进行重构,参数量和计算量分别下降 81 万、3.5 G, mAP 提高 0.6%,提升了信息融合的效率,并降低了模型复杂度。组④显示添加 SDFM 层,在 mAP@0.5 评价指标上提高了 1.8%,由于 SDFM 内部模块的堆叠,以及模型层数加深,参数量和计算量分别增加

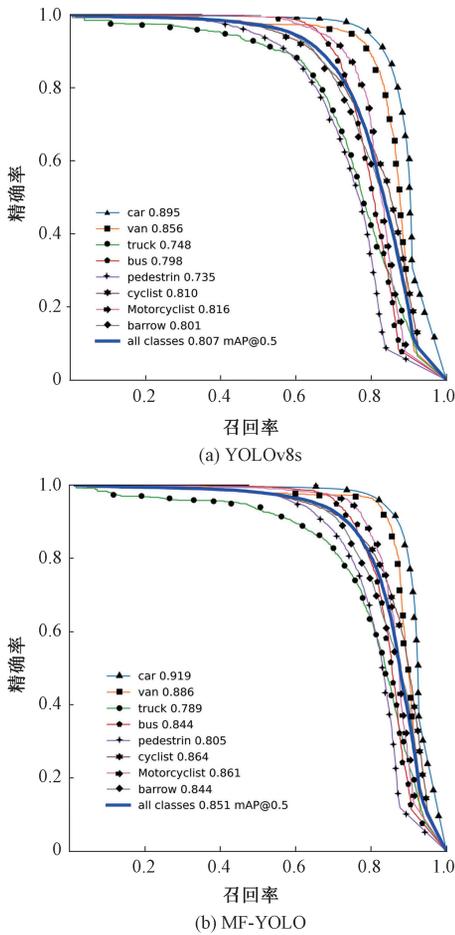


图 7 YOLOv8s 和 MF-YOLO 在数据集 DAIR-V2X-I 上的 PR 曲线

Fig. 7 Precision-recall (PR) curves of YOLOv8s and MF-YOLO on the UA-DETRAC dataset

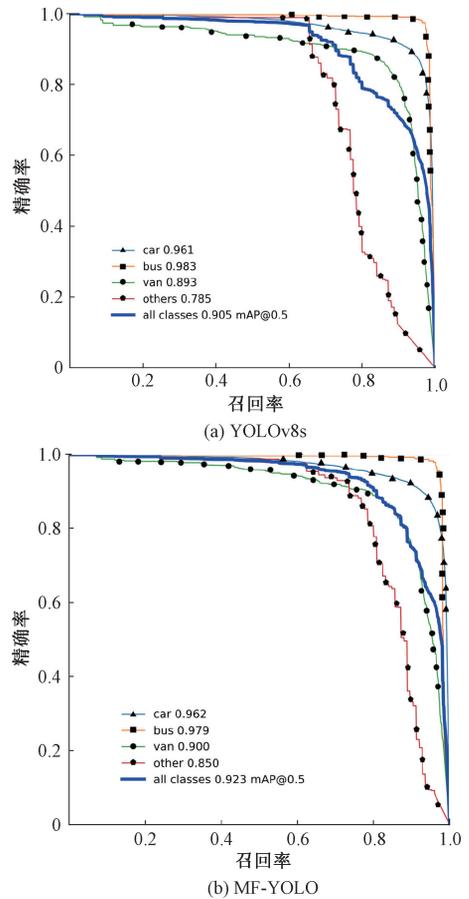


图 8 YOLOv8s 和 MF-YOLO 在数据集 UA-DETRAC 上 PR 曲线

Fig. 8 Precision-recall (PR) curves of YOLOv8s and MF-YOLO on the UA-DETRAC dataset

表 2 消融实验结果

Table 2 Ablation experiment results

组	STARMLCA	Slimneck	SDFM	WPIoU	mAP@0.5	Params	GFLOPs	FPS
①	—	—	—	—	80.7	11.12	28.8	144
②	✓	—	—	—	82.3	9.82	25.3	156
③	—	✓	—	—	81.3	10.31	25.3	162
④	—	—	✓	—	82.5	12.07	29.6	122
⑤	—	—	—	✓	82.0	11.12	28.6	139
⑥	✓	✓	—	✓	84.4	9.96	22.0	172
⑦	✓	—	✓	✓	84.7	10.58	25.3	143
⑧	✓	✓	✓	✓	85.1	10.65	24.4	152

95 万, 0.7 G, 检测速度下降明显。组⑤结果表明更换 WPIoU 损失函数后的模型, mAP@0.5 分别提升 1.3%, 检测速度略有下降, 计算复杂度有所提高。对比组①和⑥, 重构颈部网络并引入 STARMLCA 改进网络, 在 mAP@0.5 上分别提升 3.6%, 检测速度也达到 172 fps, 在计算量和参数量降低了 5.8 G, 116 万。对比组①和⑦,

mAP@0.5 和检测速度 FPS 两评价指标上大幅度提升, 参数量和计算量进一步降低。对比组①和组⑧, 在融合 4 个改进模块后, 改进模型在 mAP@0.5 提高了 4.4%, 计算量和参数量分别下降了 19.8% 和 8.18%, 且 FPS 达到了 152 fps。实验结果表明本文方法在提高模型性能和实时性方面的潜力。

2.6 实验效果展示

为了直观的观察训练网络学习到特征情况,对基准模型 YOLOv8s 和 MF-YOLO 进行热力图分析,结果如图 9。其中第 2 列是 MF-YOLO 模型所得热力图,第 1 列基准模

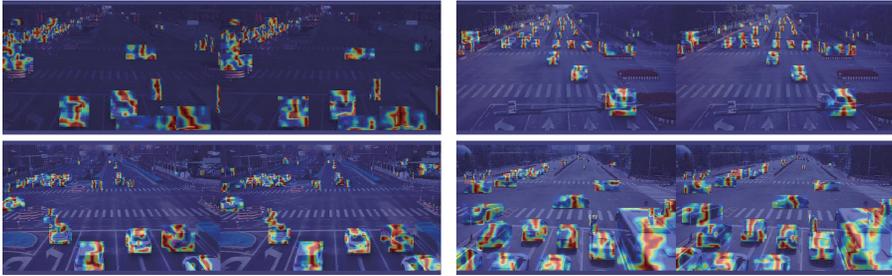


图 9 YOLOv8s 与 MF-YOLO 检测结果对比热力图

Fig. 9 Heatmap comparison of detection results between YOLOv8s and MF-YOLO

图 10 和 11 展示了在 DAIR-V2X-I 数据集和 UA-DETRAC 数据集图像可视化检测结果,左列为原模型检测结果,右列是 MF-YOLO 的检测结果。图 10 上对照组中,对于具有严重遮挡和密集分布的模糊目标车辆,原模型无法检测远处处于遮挡的车辆和摩托车,本方法可以检测出被遮挡的车辆,说明 MF-YOLO 良好特征提取能力以及对遮挡情况的鲁棒性。在图 10 下对照组, MF-YOLO 能够识别原模型漏检的行人小目标,多层级的特征融合模块优化对小目标的提取能力。在图 11 上对照组中,雨天环境下, YOLOv8s 将一辆面包车和一辆厢式货车误检为普通

型以及所得热力图。基准模型存在学习特征不够充分的情况。MF-YOLO 能够学习到更多小目标的纹理特征,并且能够给予足够的关注。

车辆,而 MF-YOLO 能够精确识别,没有产生误检,在复杂环境下泛化能力表现良好。图 11 下对照组中,在夜晚环境下,由于灯光的直射,监控视角下可见度下降, MF-YOLO 能够检测出原模型漏检的目标物体。综上所述,该方法优于 YOLOv8s。

3 结 论

本文提出一种基于多尺度特征融合与交互的路侧目标检测算法 MF-YOLO。首先在骨干网络中设计 C2f-CAST,基于星操作将来自不同子空间的特征进行交互变换,并添加混合局部通道注意力 MLCA 结合局部和全局特征以及通道和空间特征的信息,减少背景噪声干扰,解决遮挡模糊目标检错漏检的问题;对颈部网络进行重构,使用 GSConv 替换传统 Conv,设计 VoV-GSCSP 模块,降低模型参数量和计算量,提高模型检测的速度。重写基于通道-空间注意力机制表面细节融合模块 SDFM,增强对浅层特征细节的关注,解决小目标漏检的问题。设计 WPIoU 损失函数,使用自适应惩罚因子和基于锚框质量的梯度调整的 PIOUv2 改进 WIoUv3 损失函数,提升边界框的回归性能和模型收敛速度并过滤低质量锚框。通过实验可知 MF-YOLO 在 UA-DETRAC 和 DAIR-V2X-I 数据集上,在减少了 90 万参数量的情况下, mAP@0.5 分别提升 1.8% 和 4.4%, 计算量降低 19.2%, 检测速度达到 152 fps。与原模型相比更加精准和迅速。本研究完全能满足路侧设备的需求。尽管本文在降低模型复杂度方面做了相关改进,但与通道剪枝策略相比,还有进一步降低参数量和适应资源受限设备的可能空间。未来工作会进一步扩充训练场景,将结合实际应用场景来验证其鲁棒性,并使用通道剪枝量化等策略对模型进行进一步优化。

参考文献

[1] 李炳萱. 面向路侧感知的车辆与行人检测及跟踪算法研究[D]. 北京: 北京交通大学, 2023.

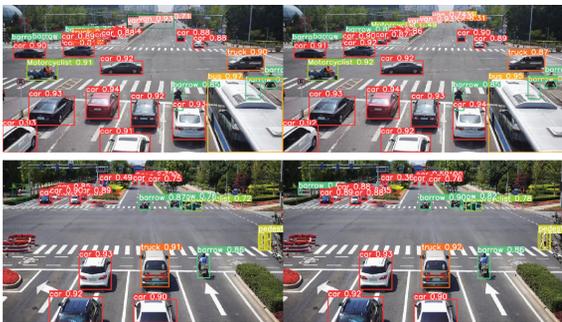


图 10 DAIR-V2X-I 数据集的可视化对比图

Fig. 10 Visualization comparison on the DAIR-V2X-I dataset

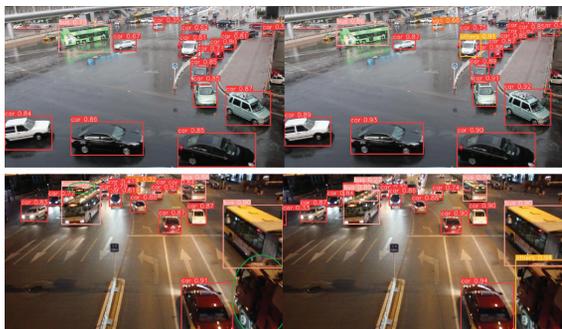


图 11 UA-DETRAC 数据集的可视化对比图

Fig. 11 Visualization comparison on the UA-DETRAC dataset

- LI B X. Research on vehicle and pedestrian detection and tracking algorithms oriented to roadside perception[D]. Beijing: Beijing Jiaotong University, 2023.
- [2] 李晓晖, 夏芹, 张强. 基于路侧视觉感知的交通目标检测及跟踪方法研究[J/OL]. 汽车工程学报, 2024: 1-10 [2024-09-20]. <https://link.cnki.net/urlid/50.1206.U.20240707.1822.002>.
- LI X H, XIA Q, ZHANG Q. Research on traffic target detection and tracking method based on roadside visual perception [J/OL]. Journal of Automotive Engineering, 2024: 1-10 [2024-09-20]. <https://link.cnki.net/urlid/50.1206.U.20240707.1822.002>.
- [3] JOCHER G, CHAURASIA A, STOKEN A, et al. Ultralytics/YOLOv5: v7.0-YOLOv5 SOTA realtime instance segmentation[EB/OL]. (2022-11-22) [2024-09-20]. <https://zenodo.org/record/3908559>.
- [4] LI CH Y, LI L L, JIANG H L, et al. YOLOv6: A single-stage object detection framework for industrial applications[C]. Twelfth International Conference on Learning Representations(ICLR), 2023.
- [5] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), 2023: 7464-7475.
- [6] JOCHER G, QIU J, CHAURASIA A. Ultralytics YOLO[EB/OL]. (2023) [2024-09-20]. <https://github.com/ultralytics/ultralytics>.
- [7] WANG C, HE W, NIE Y, et al. Gold-YOLO: Efficient object detector via gather-and-distribute mechanism [C]. Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems(NeurIPS), 2023.
- [8] WANG C Y, YE H I, LIAO H Y M. YOLOv9: Learning what you want to learn using programmable gradient information [J]. ArXiv preprint arXiv: 2402.13616, 2024.
- [9] WANG Z Y, LI CH, XU H Y, et al. Mamba YOLO: SSMs-based YOLO for object detection[J]. ArXiv preprint arXiv: 2406.05835, 2024.
- [10] WANG AO, CHEN H, LIU L H, et al. YOLOv10: Real-time end-to-end object detection [C]. 38th Conference on Neural Information Processing Systems (NeurIPS), 2024.
- [11] 卫策, 吕进, 曲晨阳. 改进 YOLOv5s 的复杂交通场景下目标检测算法[J]. 电子测量技术, 2024, 47(2): 121-130.
- WEI C, LYU J, QU CH Y. An improved target detection algorithm for complex traffic scenarios based on YOLOv5s[J]. Electronic Measurement Technology, 2024, 47(2): 121-130.
- [12] 许德刚, 王双臣, 王再庆, 等. 改进 YOLOv8 算法的城市车辆目标检测[J]. 计算机工程与应用, 2024, 60(18): 136-146.
- XU D G, WANG SH CH, WANG Z Q, et al. Urban vehicle target detection based on an improved YOLOv8 algorithm [J]. Computer Engineering and Applications, 2024, 60(18): 136-146.
- [13] 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. Vehicle and pedestrian detection method under adverse weather conditions based on improved YOLOv8s [J]. Electronic Measurement Technology, 2024, 47(9): 112-119.
- [14] 刘若瓔, 黄妙华, 王量子, 等. 基于多尺度序列融合的路侧目标检测算法[J/OL]. 激光与光电子学进展, 2024: 1-15 [2024-09-20]. <http://kns.cnki.net/kcms/detail/31.1690.tn.20240529.0917.004.html>.
- LIU R Y, HUANG M H, WANG L Z, et al. Roadside target detection algorithm based on multi-scale sequence fusion [J/OL]. Laser & Optoelectronics Progress, 2024: 1-15 [2024-09-20]. <http://kns.cnki.net/kcms/detail/31.1690.tn.20240529.0917.004.html>.
- [15] 古佳欣, 陈高华, 张春美. YOLOv8-DEL: 基于改进 YOLOv8n 的实时车辆检测算法研究[J/OL]. 计算机工程与应用, 2024: 1-13 [2024-09-20]. <http://kns.cnki.net/kcms/detail/11.2127.tp.20240926.1133.010.html>.
- GU J X, CHEN G H, ZHANG CH M. YOLOv8-DEL: Research on real-time vehicle detection algorithm based on improved YOLOv8n [J/OL]. Computer Engineering and Applications, 2024: 1-13 [2024-09-20]. <http://kns.cnki.net/kcms/detail/11.2127.tp.20240926.1133.010.html>.
- [16] 王泉, 曹心雨, 陈祺东. 面向车路协同的路侧交通目标检测模型及部署[J/OL]. 计算机应用, 2024: 1-13 [2024-09-20]. <http://kns.cnki.net/kcms/detail/51.1307.TP.20240808.1431.006.html>.
- WANG Q, CAO X Y, CHEN Q D. Roadside traffic target detection model and deployment oriented to vehicle-road collaboration [J/OL]. Computer Applications, 2024: 1-13 [2024-09-20]. <http://kns.cnki.net/kcms/detail/51.1307.TP.20240808.1431.006.html>.

- 006.html.
- [17] 闫建红, 冉同霄. 基于YOLOv8的轻量化无人机图像目标检测算法[J/OL]. 图学学报, 2024;1-9[2024-09-20]. <http://kns.cnki.net/kcms/detail/10.1034.T.20241010.2055.004.html>.
- YAN J H, RAN T X. Lightweight UAV image target detection algorithm based on YOLOv8 [J/OL]. Journal of Graphics, 2024;1-9[2024-09-20]. <http://kns.cnki.net/kcms/detail/10.1034.T.20241010.2055.004.html>.
- [18] MA X, DAI X Y, BAI Y, et al. Rewrite the stars[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), 2024; 5694-5703.
- [19] WANG 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.
- [20] LI H L, LI J, WEI H B, et al. Slim-neck by GSConv: A lightweight-design for real-time detector architectures [J]. Journal of Real-Time Image Processing, 2024, 21(3): 62.
- [21] TANG L F, ZHANG H, XU H, et al. Rethinking the necessity of image fusion in high-level vision tasks: A practical infrared and visible image fusion network based on progressive semantic injection and scene fidelity[J]. Information Fusion, 2023, 99: 101870.
- [22] 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.
- [23] TONG Z J, CHEN Y H, XU Z W, et al. Wise-IoU: Bounding box regression loss with dynamic focusing mechanism [J]. ArXiv preprint arXiv: 2301.10051, 2023.
- [24] ZHENG ZH H, WANG P, LIU W, et al. Distance-IoU loss: Faster and better learning for bounding box regression [J]. AAAI Conference on Artificial Intelligence, 2020, 34(7): 12993-13000.
- [25] ZHANG Y F, REN W Q, ZHANG ZH, et al. Focal and efficient IoU loss for accurate bounding box regression[J]. Neurocomputing, 2022, 506: 146-157.
- [26] GIRSHICK R. Fast R-CNN[C]. IEEE International Conference on Computer Vision (ICCV), 2015: 1440-1448.
- [27] ZHAO Y, LYU W Y, XU SH L, et al. DETRs beat YOLOs on real-time object detection[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), 2024; 16965-16974.

作者简介

顾杨海, 硕士, 主要研究方向为图像目标检测、图像融合、车路协同。

E-mail: 202312490346@nuist.edu.cn

李富(通信作者), 硕士, 副教授, 硕士生导师, 主要研究方向为嵌入式系统、车路协同、分心驾驶。

E-mail: lifu@cw Xu.edu.cn

陈德基, 博士, 首席科学家, 硕士生导师, 主要研究方向为工业物联网、机器视觉。

E-mail: dejichen@cw Xu.edu.cn

王泉, 博士, 正高级工程师, 硕士生导师, 主要研究方向为车联网、物联网理论及应用。

E-mail: wangquan@cw Xu.edu.cn