DOI:10.19651/j.cnki.emt.2520226

ECG-YOLO:基于 YOLO11n 改进的自动驾驶场景目标检测算法*

杨伟豪 徐晓龙 董敏聪 卞俊良 王惠亭
(河海大学信息科学与工程学院 常州 213200)

摘 要:针对自动驾驶复杂场景下远景小目标与遮挡目标漏检问题,提出了一种基于 YOLO11n 的改进算法。首先引入高效多尺度注意力机制替换原 Backbone 网络中的 C2PSA 模块,在不降低通道维度的情况下进行分组重塑和并行多尺度融合,增强模型对小目标的关注度;同时增加了大小为 160×160 的小目标检测层,利用浅层特征精确定位小目标;引入上下文引导模块,设计 C3k2-CGB,通过全局上下文信息增强特征融合,提升对遮挡目标的判别能力;最后采用 Wise-IoU 优化边界框回归,抑制低质量样本的梯度干扰。在自动驾驶数据集 KITTI 上的实验结果显示,改进后模型的 mAP 提升 5.9%,召回率提高 10.3%,参数量降低 10.1%,在多类 YOLO 算法中表现最优,显著改善了小目标与遮挡目标的检测效果。

关键词:目标检测;小目标;遮挡目标;多尺度注意力;上下文引导模块;轻量化存储
中图分类号:TN911.73 **文献标识码**:A **国家标准学科分类代码**:510.4030

ECG-YOLO: Improved object detection algorithm for autonomous driving scene based on YOLO11n

Yang Weihao Xu Xiaolong Dong Mincong Bian Junliang Wang Huiting
(College of Information Science and Engineering, Hohai University, Changzhou 213200, China)

Abstract: To address the challenge of missed detections for distant small objects and occluded targets in complex autonomous driving scenarios, this paper proposes an improved algorithm based on YOLO11n. First, an efficient multi-scale attention mechanism is introduced to replace the original C2PSA module in the backbone network. It performs group reshaping and parallel multi-scale fusion without reducing channel dimensions, thereby enhancing the model's focus on small objects. Second, an additional 160×160 detection layer is incorporated to leverage detailed spatial information from shallow features for the precise localization of small targets. Furthermore, a context-guided module, designated C3k2-CGB, is designed to augment feature fusion with global contextual information, improving the recognition capability for occluded objects. Finally, the Wise-IoU loss is adopted for bounding box regression, which suppresses the harmful gradients from low-quality examples. Experimental results on the KITTI dataset demonstrate that the improved model achieves a 5.9% increase in mAP and a 10.3% gain in recall, while reducing the number of parameters by 10.1%. It outperforms several mainstream YOLO variants, showing significant improvements in detecting small and occluded objects.

Keywords: object detection; small objects; occluded objects; multi-scale attention; context guided module; lightweight-storage

0 引 言

复杂城市场景下的目标检测技术面临两大关键挑战:部分遮挡目标识别与远距离小目标感知。自动驾驶系统能否准确捕捉被遮挡物体的有效信息,直接关系到车辆的决策安全;而对远景中的微小目标实现早期探测,则可为系统规划与控制预留更充裕的响应时间。因此,提升遮挡目标与小目标[1]的检测能力,对构建安全可靠的自动驾驶感知体系具有关键意义。

在传统目标检测方法中,基于方向梯度直方图(histogram

收稿日期:2025-11-14
*基金项目:国家重点研发计划(2022YFB4703404)、国家自然科学基金(61671202)项目资助

of oriented gradients，HOG)与支持向量机(support vector machine，SVM)的组合方法曾得到广泛应用。然而，这类方法在面对复杂场景变化及多样化环境条件时，往往表现出有限的适应能力与泛化性能。并且由于仍依赖于人工设计的特征工程，其性能上限明显，且计算效率往往难以满足实时处理需求。

近年来，随着深度学习技术的持续突破，基于卷积神经网络的目标检测算法[2]在性能与架构上均取得了显著进展。当前，主流目标检测方法普遍基于深度学习框架实现，根据检测机制的不同，可划分为两类典型范式：双阶段(two-stage)检测与单阶段(one-stage)检测。区域卷积神经网络 (regions with convolutional neural network features，R-CNN)系列[3-4]作为双阶段算法的典型代表，其检测流程可分为 3 个关键步骤：首先生成候选区域提议，继而利用分类器对候选框进行识别与筛选，最后通过非极大值抑制操作去除冗余检测框并进一步优化预测框的定位精度。R-CNN 系列的算法在检测精度上表现出色，但由于其复杂的结构，训练和推理时间较长，因此，该系列的模型多用于对精度要求高、时间要求相对宽松的应用场景。单阶段目标检测算法以 YOLO(you only look once)系列[5-8]及单次多框检测器(single shot multibox detector，SSD)等为代表，通过单一神经网络完成前向推理，直接输出预测框的位置与类别信息。该类算法在检测速度方面具有显著优势，然而在遮挡目标与小尺度目标的检测精度方面仍存在明显不足，这一局限已成为制约其在自动驾驶场景中广泛应用的主要瓶颈。

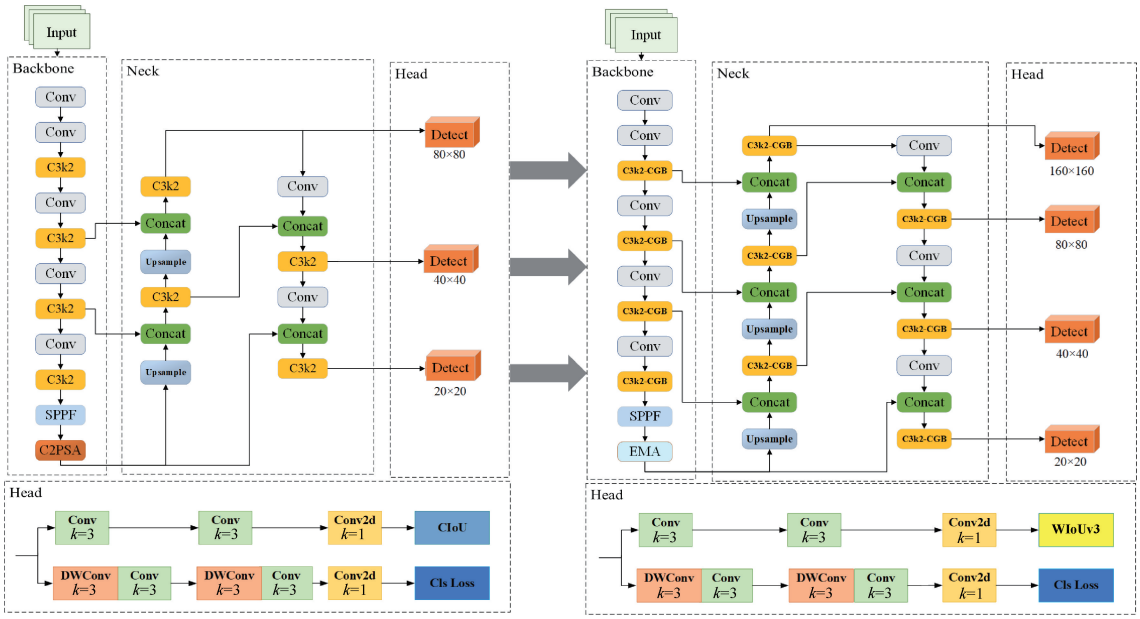
尽管基于深度学习的目标检测算法已取得显著进展，但在实际部署中，现有的改进策略仍面临严峻挑战。为解决特征提取不足的问题，部分研究引入了高复杂度的模块。例如，Zhu 等[9]提出的 TPH-YOLOv5 和钱承山等[10]通过引入 Transformer 架构，利用自注意力机制增强了全局感知能力。然而，Transformer 的计算复杂度随序列长度呈二次方增长，导致参数量和计算量激增，难以在算力受限的车载嵌入式平台上实时运行。Wang 等[11]提出的 YOLOv8-QSD 算法通过引入 Query 机制与 BiFPN 特征金字塔(feature pyramid network，FPN)，有效改善了小目标检测性能，但增加了复杂的网络架构，对硬件算力提出了过高要求。杜运亮等[12]通过结合半监督域适应与混合注意力机制，缓解了微弱光照条件下行人特征提取困难的问题，但其双模型训练机制显著增加了计算资源消耗，且未充分解决密集遮挡下的语义缺失问题。为此，一些研究者致力于模型的轻量化部署，Edge-YOLO[13]、BGR-YOLO[14]、MF-YOLO[15]等算法通过剪枝、重参数化或引入 ViT (vision transformer)混合架构实现了模型压缩，但过度精简往往以牺牲模型对微弱特征的捕捉能力为代价，导致在面对远距离小于 32×32 pixels 的极小目标时，特征易被清洗，造成漏检。为了进一步改善小目标与遮挡漏检问题，Zhou 等[16]通过改进 FPN 增强特征图中的上下文信息表达，改善了模型对小目标的表征能力；Deng 等[17]设计了一种扩展特征金字塔结构(extended feature pyramid network，EFPN)融合相邻层级的特征信息，增强模型对小尺度目标的表征能力；唐彪等[18]在检测头前引入注意力机制，利用双向 FPN 网络重构颈部，增强多层特征的融合能力。然而，这些方法大多侧重于局部特征的堆叠，缺乏对全局上下文信息的有效建模，在自动驾驶复杂的交通流中，当面对行人或自行车这类细长且易被遮挡目标时，单纯的局部特征增强往往失效，缺乏上下文引导的模型容易将其误检为行人或背景，从而导致语义信息的断裂。现有的一些自动驾驶场景改进算法，如杨磊等[19]提出 FAN-YOLOv8n 和 Luo 等[20]提出的 YOLOv8-ghost-EMA，虽然在通用数据集上表现尚可，但由于其锚框分配机制和损失函数难以平衡高质量样本与低质量样本的权重，导致模型在处理姿态多变的行人与背景混淆的骑行者等难例时，泛化能力仍稍显不足。

综上所述，当前的算法改进主要面临着引入复杂模块导致无法部署或过度轻量化导致特征丢失的两难困境，且在遮挡与小目标并存的极端场景下缺乏鲁棒的上下文推理能力。针对上述难题，本文旨在探寻一种在保证一定的轻量化前提下，能有效解决小目标漏检与遮挡语义推断的平衡方案。以 YOLO11n 为基础，本文提出 ECG-YOLO 算法，主要通过以下 4 个针对性的改进来突破现有瓶颈：1)针对主干网络特征提取不足，引入高效注意力模块(efficient multi-scale attention，EMA)，利用跨空间学习机制自适应校准通道权重，在维持低计算负载的前提下显著抑制复杂城市背景噪声；2)针对下采样导致的细粒度信息缺失，增加 P2 微小目标检测层，利用浅层高分辨率特征解决远景小目标漏检问题；3)针对遮挡造成的局部特征失效问题，集成 C3k2-CGB 上下文引导模块，通过聚合多尺度环境上下文信息，填补遮挡导致的语义缺失，增强特征的完整性与鲁棒性；4)针对低质量锚框对训练产生的负面影响，采用 WIoUv3(wise-IoUv3)损失函数，通过动态非单调聚焦机制，平衡简单与困难样本的梯度贡献，提升模型对行人和骑行者等难检目标的泛化性能。

1　YOLO11n 检测算法

YOLO11n 架构的核心由 3 个核心组件构成，如图 1(a)所示。主干网络作为基础特征提取器，通过卷积神经网络将输入图像转换为多尺度特征表征。颈部网络采用 FPN 与路径聚合(path aggregation network，PAN)相结合的结构，通过对骨干网络不同层级的特征进行拼接融合，并利用 C3k2 模块实现跨尺度特征增强与信息整合。头部部分有 3 个检测头，采用双分支解耦头，通过两个并行的分支来进行分类和回归，提升了模型的检测效果。YOLO11n 建立在先前 YOLOv8n 模型的优势之上，代表了这个进化

系列中的最新迭代,其优异的特征提取能力已在多个精密感知的工业领域得到验证[21]。相比较于 YOLOv8n 模型,YOLO11n 提出了 C3k2 模块,取代了之前版本中使用的 C2f 模块。其次,YOLO11n 在 SPPF 模块后面添加了一个 C2PSA 模块用于增强特征提取。最后,将原先的解耦头中的分类检测头增加了两个 DWConv,并且将原先两个普通的 Conv 的卷积和由 3 变为了 1,形成了两个深度可分离 Conv,减少了冗余计算,提高了效率。



(a) YOLO11n (b) ECG-YOLO

图 1 YOLO11n 和 ECG-YOLO 网络模型结构

Fig. 1 Network architectures of YOLO11n and ECG-YOLO

2 ECG-YOLO 算法

针对自动驾驶场景中车辆与行人检测任务常面临的远距离小目标漏检及密集遮挡目标识别困难等挑战,现有 YOLO11n 模型表现稳定性不足,而更大规模模型又面临计算负载与参数成本过高的问题。为进一步压缩模型参数的同时有效提升道路目标的检测精度与效率,本文对 YOLO11n 架构进行了针对性改进,优化后的 ECG-YOLO 模型结构如图 1(b)所示。图 2 进一步展示了 ECG-YOLO 从 Backbone 特征提取、Neck 多尺度融合到 Loss 优化策略的整体设计,阐明了各改进模块如何在轻量化设计与高精度检测之间实现高效协同。

2.1 EMA 注意力机制模块

由于自动驾驶图像中存在大量小目标物体,且目标特征弱、信息量少。原 YOLO11n 骨干网络末端使用的 C2PSA 模块存在两个显著缺陷。首先,C2PSA 虽然引入了空间注意力,但其生成的注意力图在不同通道间具有一致性,这意味着模型无法区分某个空间位置的激活是来自背景还是远处的行人,从而导致小目标特征的通道语义缺失。其次,C2PSA 内部复杂的像素重排与多分支深度卷积显著增加了计算量与内存访问成本,这与自动驾驶车载嵌入式平台对实时性的严苛要求相悖。基于上述问题,本文引入高效多尺度注意力机制 EMA[22]来置换

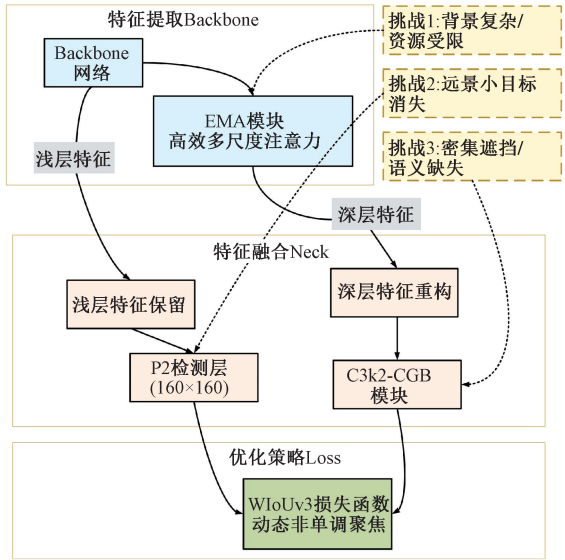


图 2 ECG-YOLO 整体设计策略

Fig. 2 The overall design strategy of ECG-YOLO

C2PSA,如图 3 所示。在 EMA 模块的示意图中,"G"表示输入通道被分成的组数。"X Avg Pool"和"Y Avg Pool"分别代表一维水平和垂直的全局池化操作。任何输入特征图 $\boldsymbol{F}\in\mathbf{R}^{C\times H\times W}$ 都被分为 $G(G<<C)$ 个子特征

$\boldsymbol{F}=[\boldsymbol{F}_0, \boldsymbol{F}_1, \cdots, \boldsymbol{F}_{G-1}], \boldsymbol{F}_i \in \mathbf{R}^{C//G\times H\times W}$ 用于学习不同的语义。通过将输入特征按通道维度分组，并且采用两个并行分支，一个分支采用一维全局池化和 1×1 卷积处理全局信息，另一个分支使用 3×3 卷积捕捉局部空间信息，两分支的输出通过矩阵乘法融合，生成注意力图，最终与输入特征相结合。

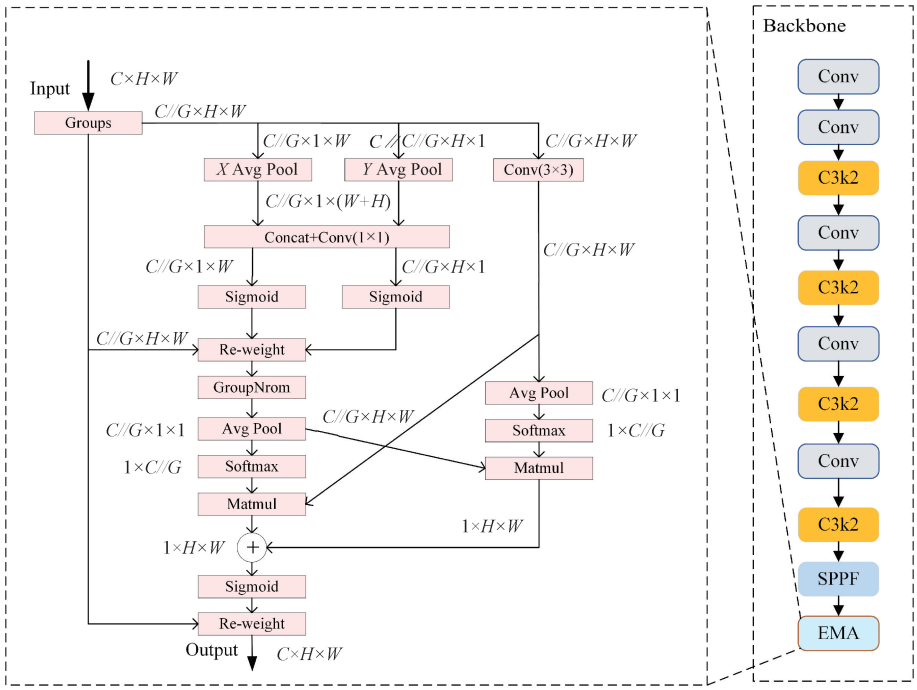


图 3　EMA 模块

Fig. 3　The EMA module

针对小目标易丢失的问题，传统的注意力机制通常会对通道进行降维以减少参数量，导致关键信息在降维中丢失。EMA 的独特优势在于其可以在不进行通道降维的情况下，通过分组处理策略，利用跨空间学习让模型针对特定的通道赋予更高的空间权重，从而在复杂的城市背景噪声中提炼出微弱的小目标信号，确保远处行人或车辆特征的完整性与显著性。相较于其他注意力机制，EMA 解决了传统注意力机制的缺陷，且在结构设计上摒弃了 C2PSA 中高计算量的复杂算子，转而采用分组卷积与并行子网络结构，最大程度地保留了小目标在原始通道中的微弱特征信号，有效弥补了原模块速度与精度平衡上的缺陷。

2.2　P2-Detect

在自动驾驶的感知任务中，远距离微小目标与高度遮挡目标是引发安全事故的高危因素。YOLO11n 模型中默认配置了 3 个检测头，分别是 P3-Detect、P4-Detecet 和 P5-Detect。这种设计在常规目标检测中表现良好，但在自动驾驶场景下却存在显著缺点。由于 YOLO11n 采用较大的下采样倍数，经过多次下采样后，对于图像中尺寸小于 32×32 pixels 的极小目标，经过 P3 层的 8 倍下采样后，其在特征图上的投影仅剩不到 4×4 pixels，导致 P3 层的检测头难以有效识别这些小目标的细节特征。

为解决这一问题，本文在 YOLO11n 模型的基础上，增加了一个更浅层的检测头 P2-Detect，如图 4 所示。该检测头通过利用 P2 层特征图来提升对小目标的检测能力，P2 层特征图的尺度为 160×160，仅经过 4 倍下采样，相较于 P3 层特征图，P2 层拥有更丰富的底层细节信息，使得模型能够捕捉到远处行人或骑行者的细微轮廓，显著提升了对非刚性小目标的感知灵敏度。

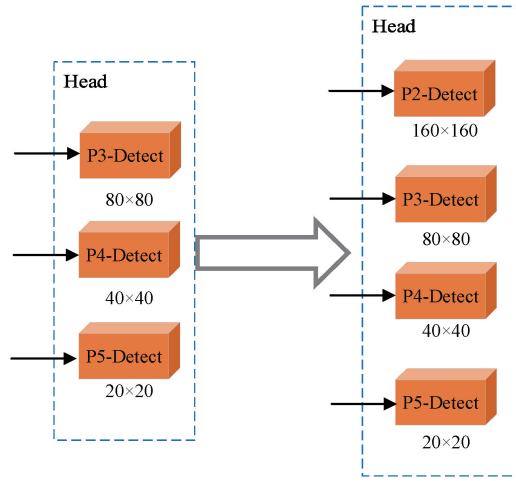


图 4　增加 P2 检测头

Fig. 4　Add the P2 detection head

2.3　C3k2-CGB 模块设计

C3k2 模块作为 YOLO11n 骨干网络的核心特征提取

单元,其内部依赖 Bottleneck 结构进行非线性变换。然而,Bottleneck 主要通过堆叠 3×3 标准卷积来隐式获取特征,其感受野相对受限。当面对严重遮挡目标或纹理信息匮乏的小目标时,缺乏显式的上下文捕获机制,致使模型在自动驾驶的复杂城市场景中难以利用周围环境信息来辅助推断,从而造成误检和漏检。

相关研究表明,通过全局特征交互能够有效增强模型对复杂场景的感知能力[23]。为了更好地推理目标遮挡部分以及理解目标与场景的关系,本文设计了 C3k2-CGB 模块。当参数 c3k 设置为 False 时,该模块将原有的 Bottleneck 分支替换为上下文引导模块(context guided block,CGBlock)[24],CGBlock 结构如图 5 所示。

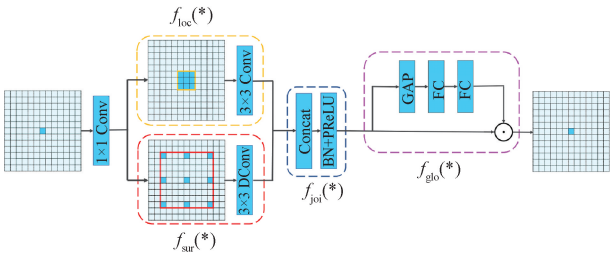


图 5 CGBlock 结构

Fig. 5 CGBlock structure

CGBlock 主要由 4 个子模块构成:

1)局部特征提取器 $f_{loc}(*)$ 使用标准卷积学习局部特征,该组件采用一个标准的 3×3 卷积层从图像中的局部区域提取特征,用于保留车辆轮廓、行人姿态等物理特征。

2)周围上下文提取器 $f_{sur}(*)$,针对自动驾驶中常见的遮挡问题,该组件引入空洞卷积,在不显著增加参数量的前提下扩大感受野,有效感知目标周围的区域上下文,解决因目标遮挡造成的语义缺失,从而降低漏检率。

3) 联合特征融合器 $f_{joi}(*)$ 通过连接层将局部特征 $f_{loc}(*)$ 与周围上下文特征 $f_{sur}(*)$ 进行拼接,并依次经过批量归一化与 PReLU 激活函数,从而确保模型在关注目标细节的同时,不丢失对环境关系的理解。

4) 全局上下文提取器 $f_{glo}(*)$,首先借助全局平均池化层对空间信息进行压缩与聚合,提取图像级全局上下文表征,随后通过轻量级多层感知器进一步编码高阶语义信息。对于远景小目标,引入全局上下文能帮助模型根据整体场景逻辑来过滤背景噪声,从而降低小目标的误检率。基于上述组件的特性,构建 C3k2-CGB,其结构如图 6 所示。

在本网络架构设计中,C3k2-CGB 模块代替 Backbone 和 Neck 网络中所有的 C3k2 模块。该策略不仅保证了模型轻量化存储的优势,同时改善了自动驾驶中因信息缺失和信息微弱带来的检测难题。

2.4 改进的损失函数

针对解耦头中的回归检测,原有的 YOLO11n 采用的是 CIoU (complete IoU)损失函数,该函数综合考虑了重叠

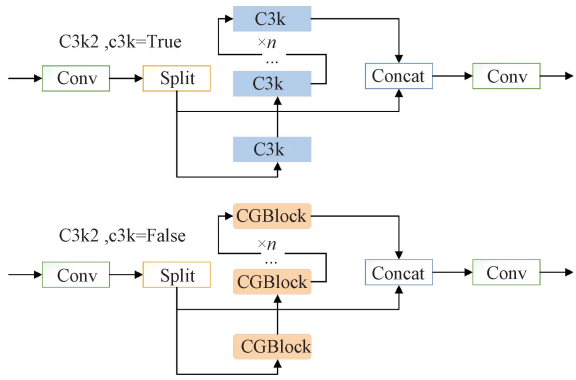


图 6 C3k2-CGB 结构

Fig. 6 C3k2-CGB structure

面积、中心点距离及长宽比的一致性,充分量化了预测框与真实框的形状差异,在处理中大尺度目标时表现优异。CIoU 的计算公式如式(1)~(3)所示。

$$L_{CIoU} = 1 - IoU + \frac{\rho^2(\boldsymbol{b},\boldsymbol{b}^{gt})}{(c_w)^2 + (c_h)^2} + \alpha v \tag{1}$$

$$\alpha = \frac{v^2}{1 - IoU + v} \tag{2}$$

$$v = \frac{4}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h}\right)^2 \tag{3}$$

其中,IoU 表示预测框与实际框的交并比,$\rho^2(\boldsymbol{b},\boldsymbol{b}^{gt})$ 表示实际框和预测框的质心之间的欧氏距离,h 和 w 表示高度和预测盒,h^{gt} 和 w^{gt} 表示真实框的高度和宽度,c_h 和 c_w 表示由预测框和真实框组成的最小围框的高度和宽。式中涉及的部分参数如图 7 所示。

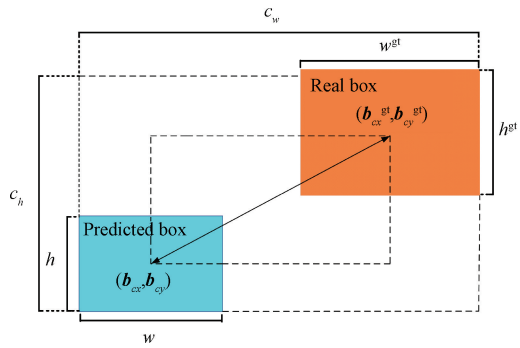


图 7 损失函数参数示意

Fig. 7 The parameters of the loss function

然而,在面对自动驾驶场景检测时,CIoU 存在明显的局限性。当车辆或行人被部分遮挡时,预测框的几何形状往往与真实框存在显著差异。CIoU 将长宽比作为强惩罚项,会导致模型在处理低质量遮挡样本时产生过大的惩罚梯度,迫使模型去拟合错误的几何特征,从而破坏了整体的泛化能力。并且由于自动驾驶场景中存在海量的简单背景样本与少量困难的小目标样本,CIoU 缺乏针对样本难易程度的动态调节机制,导致训练过程易被大量简单样本的主导梯度淹没,使得小目标的回归精度难以进一步提升。

为了解决上述梯度分配不合理的问题，本文引入 WIoU 损失函数，其核心在于采用“离群度”指标替代传统 IoU 度量，并基于此构建动态梯度调制机制。该机制通过降低高质量锚框的梯度增益，同时抑制低质量样本的梯度贡献，使模型聚焦于中等质量样本的优化，从而显著提升检测器的泛化性能与收敛稳定性。Tong 等[25]提出了 3 个版本的 WIoU。WIoUv1 采用基于注意力的预测框损失设计，WIoUv2 和 WIoUv3 则在此基础上增加了聚焦系数。

WIoUv1 通过引入距离感知机制重构边界框损失函数。当预测框与真实框达到一定重叠程度时，该方法通过适度放宽几何约束来降低惩罚强度，从而使模型获得更优秀的泛化性能。WIoUv1 的计算公式如式(4)～(6)所示。

$$\mathcal{L}_{\mathrm{WIoUv1}} = \mathcal{R}_{\mathrm{WIoU}}\,\mathcal{L}_{\mathrm{IoU}} \tag{4}$$

$$\mathcal{R}_{\mathrm{WIoU}} = \exp\left(\frac{\rho^2(\boldsymbol{b},\boldsymbol{b}^{\mathrm{gt}})}{((c_w)^2+(c_h)^2)^*}\right) \tag{5}$$

$$\mathcal{L}_{\mathrm{IoU}} = 1-\mathrm{IoU} \tag{6}$$

WIoUv2 在 WIoUv1 基础上构建了单调聚焦系数 $\mathcal{L}^*_{\mathrm{IoU}}$ 通过降低简单样本在损失计算中的权重来优化训练过程。但是，考虑到 $\mathcal{L}^*_{\mathrm{IoU}}$ 在模型训练过程中会随着 $\mathcal{L}_{\mathrm{IoU}}$ 的减小而衰减，从而影响模型收敛速度。因此引入 $\mathcal{L}_{\mathrm{IoU}}$ 的平均值对 $\mathcal{L}^*_{\mathrm{IoU}}$ 进行归一化处理。WIoUv2 的计算公式如式(7)～(9)所示。

$$\mathcal{L}_{\mathrm{WIoUv2}} = \mathcal{L}^{\gamma *}_{\mathrm{IoU}}\,\mathcal{L}_{\mathrm{WIoUv1}},\gamma>0 \tag{7}$$

$$\frac{\partial\,\mathcal{L}_{\mathrm{WIoUv2}}}{\partial\,\mathcal{L}_{\mathrm{IoU}}} = \mathcal{L}^{\gamma *}_{\mathrm{IoU}}\frac{\partial\,\mathcal{L}_{\mathrm{WIoUv1}}}{\partial\,\mathcal{L}_{\mathrm{IoU}}},\gamma>0 \tag{8}$$

$$\mathcal{L}_{\mathrm{WIoUv2}} = \left(\frac{\mathcal{L}^*_{\mathrm{IoU}}}{\mathcal{L}_{\mathrm{IoU}}}\right)^{\gamma}\mathcal{L}_{\mathrm{WIoUv1}} \tag{9}$$

WIoUv3 通过定义离群度 β 来量化锚框质量，并基于此构建非单调聚焦因子 r 对 WIoUv1 进行加权。该方法采用动态梯度分配机制：β 值越大，对应的梯度增益越小，从而有效抑制低质量样本产生的有害梯度影响。这种策略通过平衡不同质量样本的优化权重，使模型聚焦于普通质量样本的学习，最终提升整体检测性能。WIoUv3 公式如式(10)～(12)所示。

$$r = \frac{\beta}{\delta\alpha^{\beta-\delta}} \tag{10}$$

$$\beta = \frac{\mathcal{L}^*_{\mathrm{IoU}}}{\mathcal{L}_{\mathrm{IoU}}}\in[0,+\infty) \tag{11}$$

$$\mathcal{L}_{\mathrm{WIoUv3}} = r\,\mathcal{L}_{\mathrm{WIoUv1}} \tag{12}$$

其中，δ 和 α 是超参数，可根据不同模型进行调整。

本文最终选择在物体边界框回归损失中引入 WIoUv3。WIoUv3 通过动态非单调聚焦，增加了对中等难度及高质量样本的关注，防止模型在无法收敛的极端样本上浪费梯度，从而提升模型的泛化性。

3　实验结果与分析

3.1　实验数据准备与处理

本实验采用由德国卡尔斯鲁厄理工学院与丰田工业大学芝加哥分校联合创建的 KITTI-2D 数据集。作为自动驾驶领域最具影响力的基准数据集之一，KITTI 以其大规模、多模态特性为计算机视觉算法提供全面评估平台，是目前该领域公认的标准数据集。本研究从原始数据集中选取了 4 862 张街道场景较多、行人更为密集的图像。该子集不仅更具针对性地面向城市复杂道路下的自动驾驶任务，其包含的目标也具有尺度更小、分布更密集的特点，显著提升了任务的挑战性与其实际应用价值。按照 8∶2比例将数据集划分为训练集与验证集，其中 3 889 张图片用于训练，973 张图片用于验证，实验在数据预处理进行了类别分类和整理，将数据类别分为 3 类：Car、Pedestrian 和 Cyclist。

3.2　实验环境

实验环境是在操作系统为 Windows11、处理器为 Intel i5-13400F、内存大小为 32 GB 的计算机上运行，并装配有一张 24 GB 显存的 NVIDIA GeForceRTX 3090 显卡来加速训练。开发工具是 PyCharm，开发语言是 Python 3.10.8，深度学习框架为 Pytorch 2.1.2，CUDA 版本为 12.1。实验训练参数如表 1 所示。

表 1　实验参数

Table 1　Experimental parameters

参数	参数值
输入图像大小	640×640
初始学习率	0.01
优化器	SGD
批量大小	32
迭代轮数	200
动量	0.937

3.3　评估指标

为了测试提出的改进模型的检测性能，本文使用召回率 Recall、mAP@0.5、mAP@0.5∶0.95、模型参数数量作为评估指标。以下参数用于上述一些评估指标的公式中：TP(预测为正样本，实际上也是正样本)、FP(预测为正样本，实际上是负样本)和 FN(预测为负样本，实际上是正样本)。IoU 表示模型检测出的区域与实际目标区域的重叠程度。

精度(precision，P)是模型预测的正样本数与所有检测到的样本数之比，计算方法如式(13)所示。

$$P = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \tag{13}$$

召回率(recall，R)衡量的是模型检测正确目标的能力，可以体现漏检率，对于自动驾驶场景尤为关键，因为错过任何一个目标都可能会导致严重的后果。召回率的计算方法如式(14)所示。

$$R = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{14}$$

平均精密度(AP)等于精度-召回率曲线下的面积,计算公式如式(15)所示。

$$AP = \int_0^1 P(R)\mathrm{d}R \tag{15}$$

mAP(mean average precision)是将所有样本类别的 AP 值加权平均得到的结果用来衡量模型在所有类别下的检测性能,其公式如式(16)所示。

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{16}$$

式中:AP_i 表示指目标第 i 个类别的平均精度,N 表示训练数据集中样本的类别数。mAP@0.5 表示检测模型 IoU 设置为 0.5 时的平均精度,mAP@0.5:0.95 表示检测模型 IoU 设置为 0.5～0.95 时的平均精度。

3.4 实验结果分析

1)模型改进前后对比

根据表 2 的实验结果,改进后的 YOLO11n 算法在目标检测性能上取得了显著提升。具体而言,该算法的 mAP 指标较基线模型提高了 5.9%。进一步分析各类别的检测精度,Car、Pedestrian 和 Cyclist 这 3 大类别的 AP 值分别提升了 1.2%、9.4%和 7.1%。值得注意的是,改进后的算法对 Cyclist 和 Pedestrian 类别的检测性能提升最为显著。在 KITTI 数据集中,这两类目标通常具有尺寸较小、空间分布密集且易重叠的特点,属于典型的难检测目标。实验结果表明,改进后的 YOLO11n 算法通过结构优化有效提升了小目标检测能力,这为解决自动驾驶场景中的关键检测难题提供了新的技术途径。

表 2 改进模型与 YOLO11n 比较

Table 2 Comparison between the improved model and YOLO11n

模型	Params/M	mAP@0.5/%	mAP@0.5:0.95/%	AP/%		
				Car	Pedestrian	Cyclist
YOLO11n	2.58	82.9	55.9	94.8	72.4	81.6
本文	**2.32**	**88.8**	**59.4**	**96.0**	**81.8**	**88.7**

2)消融实验

为了验证本文提出的改进策略的有效性,本文对基线模型进行了系统的消融实验,实验结果如表 3 所示(表中"√"表示采用对应的改进策略)。

表 3 消融实验

Table 3 Ablation experiment

组别	EMA	Layer	C3k2-CGB	WIoUv3	Params/M	mAP@0.5/%	AP/%			APs	APm	APl	R/%
							Car	Pedestrian	Cyclist				
组 1					2.58	82.9	94.8	72.4	81.6	23.8	50.9	67.2	71.4
组 2	√				2.33	83.5	94.7	72.9	82.9	27.1	50.8	66.3	73.2
组 3	√	√			2.42	87.7	96.6	78.4	88.1	31.7	51.8	**67.7**	77.5
组 4	√	√	√		2.32	87.8	**96.6**	78.5	88.4	31.8	53.2	66.2	78.0
组 5	√	√	√	√	**2.32**	**88.8**	96.0	**81.8**	**88.7**	**32.4**	**53.8**	66.3	**81.7**

由于 KITTI 数据集中 Pedestrian 和 Cyclist 主要分布在中小尺度范围,且常存在遮挡,因此本文选取这两类目标的 AP 变化以及 APs 和 APm 的数值作为衡量模型对中小目标检测能力的主要依据。实验结果表明,各改进模块均对模型检测性能产生了积极的提升效果。在基线模型基础上,首先,针对模型的特征提取与轻量化,采用 EMA 模块替换原骨干网络中的 C2PSA 模块。实验结果显示,在模型参数量显著降低 9.6%的前提下,mAP@0.5%仍提升了 0.6%。其中,Pedestrian 和 Cyclist 的 AP 值分别提升 0.5%和 1.3%且 APs 从 23.8% 提升至 27.1%。表明 EMA 模块在压缩模型规模的同时,有效抑制了复杂背景中的冗余信息,使小目标的特征提取更加纯净。随后,为进一步增强对小目标的检测能力,在组 3 中引入 P2 检测层。该改进使 mAP@0.5%显著提升 4.2%,观察各类别的 AP 变化,以小目标为主的 Pedestrian 和 Cyclist 的 AP 值分别激增 5.5%(从 72.9%提升至 78.4%)和 5.2%(从 82.9%提升至 88.1%),而目标较大的 Car 仅提升 1.9%。同时,APs 和 APm 分别达到了 31.7%和 51.8%的高点。验证了 P2 层成功找回了在深层下采样中丢失的 Pedestrian 和 Cyclist 特征,保留了更多高分辨率的空间细节,显著提升了模型对微小目标的感知能力。在加入 EMA 模块和 P2 检测头的基础上,组 4 将网络中所有 C3k2 模块替换为本文设计的 C3k2-CGB 模块,对于常被遮挡的 Cyclist 和 Pedestrian,其 AP 值依旧维持在 88.4%和 78.5%的高位,Recall 提升至 78.0%,同时进一步降低了模型参数量与复杂度。最终,将原 CIoU 损失函数替换为 WIoUv3,实验结果表明,Pedestrian 的 AP 值再次大幅跃升 3.3%(从 78.5%提升至 81.8%),APs 达到最优的 32.4%。同时 mAP@0.5%提升 1.0%,

Recall 提升 3.7%，由此可见，WIoUv3 通过动态非单调聚焦机制迫使模型专注于优化 Pedestrian 这类低质量的难例，从而降低漏检，提升检测精度。

3）对比试验

为检验 ECG-YOLO 算法在自动驾驶场景下的综合性能，将其与当前较为流行的几种一阶段目标检测算法 YOLOv8n、YOLOv9s、YOLOv10n 进行对比实验。实验均采用相同训练参数，并将上述算法在本文中所筛选的 KITTI 数据集上进行训练与测试，表 4 展示了不同算法在 KITTI 数据集目标检测上的实验结果。

表 4　各类模型结果对比实验

Table 4　Comparison of results of various models

模型	AP/%			mAP@0.5/%	R/%			R/%	Params/M	GFLOPs	FPS
	Car	Pedestrian	Cyclist		Car	Pedestrian	Cyclist				
YOLOv8n	94.4	74.0	83.1	84.0	88.4	60.3	72.2	73.6	2.68	6.8	184
YOLOv9s	**96.1**	78.8	86.7	87.2	90.9	65.3	77.5	77.9	6.19	22.1	69
YOLOv10n	94.3	69.7	79.9	81.3	87.5	56.3	69.1	71.0	2.69	8.2	145
YOLO11n	94.8	72.4	81.6	82.9	88.4	57.2	68.7	71.4	2.58	**6.3**	**196**
ECG-YOLO	96.0	**81.8**	**88.7**	**88.8**	**92.7**	**70.8**	**81.8**	**81.7**	**2.32**	9.4	170

可以看出，本文所提方法相较于其他主流网络，检测精度 mAP@0.5 和召回率 R 都是最高，达到 88.8% 和 81.7%，比原网络提高了 5.9% 和 10.3%，明显高于其他算法，大幅减少小目标漏检现象。然而值得注意的是，表 4 数据显示 ECG-YOLO 的计算量为 9.4 G，相较于基线模型增加了约 3.1 G，且 FPS 从 196 下降至 170。这是引入高分辨率 P2 微小目标检测层所带来的代价，由于 P2 层直接处理 160×160 尺度的特征图，其卷积运算密度显著高于深层网络，从而给系统造成一定的推理延迟增加。但在自动驾驶逻辑中，漏检的风险远高于毫秒级的延迟。漏检一个远处的行人或遮挡车辆可能直接导致致命交通事故，ECG-YOLO 以适度的计算量增加为代价，换取了召回率高达 10.3% 的显著提升。特别是针对 Pedestrian 和 Cyclist 这类高危弱势目标，其召回率分别大幅度提升了 13.6% 和 13.1%，且其推理速度仍远超自动驾驶 30 fps 的实时性及格线。这意味着驾驶系统能更早发现危险，守住安全底线。

综合来看，ECG-YOLO 虽然因 P2 层的引入带来了计算量的适度增长，但成功实现了高召回、低参数、强鲁棒的设计目标。在保证模型轻量化存储与可部署性的前提下，显著解决了城市交通中行人与车辆的漏检难题，展现了更优的工程实用价值。

4）检测效果显示对比

为了更直观地感受改进效果，分别用改进前后的模型在部分验证集进行测试，改进后的 YOLO11n 网络整体精度有所提高，且针对漏检情况有所改善，能够准确对遮挡以及小目标进行检测，在不同复杂场景下检测出来精度也有所提高，可适用性高。测试结果图 8～10 所示。

(a) YOLO11n检测结果
(a) Detection results of YOLO11n

(b) ECG-YOLO检测结果
(b) Detection results of ECG-YOLO

图 8　小目标漏检

Fig. 8　Missed detection of small targets

改进后的 ECG-YOLO 模型在复杂道路场景的感知任务中表现出显著优越性，尤其在对基准模型构成挑战的小目标与部分遮挡目标的检测性能上实现了实质性突破。模型通过引入多尺度特征增强模块与上下文注意力机制，有效强化了对远距离小目标的细微特征提取能力，同时利用场景上下文信息对遮挡区域进行了精准推理。这一优

(a) YOLO11n检测结果
(a) Detection results of YOLO11n

(b) ECG-YOLO检测结果
(b) Detection results of ECG-YOLO

图 9 目标遮挡导致的漏检

Fig. 9 Missed detection due to object occlusion

(a) YOLO11n检测结果
(a) Detection results of YOLO11n

(b) ECG-YOLO检测结果
(b) Detection results of ECG-YOLO

图 10 行人密集与昏暗场景下的漏检

Fig. 10 Missed detection in dense and dim pedestrian scenes

化为下游的预测与规划模块提供了更可靠、更及时的环境感知信息,提升了自动驾驶系统在复杂场景下的安全性与可靠性。

4 结 论

针对现有的目标检测算法在自动驾驶场景易出现小目标及遮挡目标漏检问题,本研究提出了一种基于 YOLO11n 网络的系统性优化方法 ECG-YOLO。首先,在骨干网络引入 EMA 注意力机制,在降低参数量的同时强化有效特征的提取,为后续 Neck 处理提供更加优质的特征图。其次,针对 Neck 端特征融合的不足,构建 P2 微小目标检测层,配合设计的 C3k2-CGB 模块,通过保留浅层特征中的高分辨率信息与融合深层多尺度上下文信息,有效提升对遮挡与极小目标的定位精度,显著缓解漏检误检问题。最后,配合 WIoUv3 损失函数,通过动态非单调聚焦机制优化高质量与低质量样本的梯度分配,进一步提升模型在复杂自动驾驶路况下的鲁棒性与泛化能力。

综上所述,实验数据充分证明了这些改进有效提升了 YOLO11n 在自动驾驶复杂场景目标检测任务中的性能,大幅减少了漏检现象,模型的召回率 R、平均精确度 mAP 对比基线模型分别提高 10.3%、5.9%,参数量相较于基线模型减少 10.1%,达到了预期的实验效果。此外,还与其他模型进行了比较,在检测结果上优于其他模型。尽管优化后 YOLO11n 在模型参数和检测性能方面表现出色,然而,由于引入了额外的小目标检测层,该层需要对高分辨率特征图进行大量卷积运算,导致模型的计算复杂度会有所增加,推理 FPS 从 196 降至 170,但这种计算成本的增加在自动驾驶安全逻辑中是极具性价比的。在系统实时性允许的范围内,最大限度地降低了对远距离行人和骑行者的漏检风险,为实现更安全的自动驾驶感知提供了有效方案。

参考文献

[1] 李翔宇,王伟,王峰萍,等. 面向密集场景结合 TC-YOLOX 的小目标检测方法[J]. 电子测量技术,2023,46(15):133-142.
LI X Y, WANG W, WANG F P, et al. Small target detection method for dense scenes based on TC-YOLOX[J]. Electronic Measurement Technology, 2023, 46(15): 133-142.

[2] 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 A Q, GUO L J, FENG R S. Multi-target vehicle detection algorithm for dense scenes [J]. Electronic Measurement Technology, 2024, 47(9): 129-136.

[3] LIU Z, YEOH J K W, GU X Y, et al. Automatic pixel-level detection of vertical cracks in asphalt pavement based on GPR investigation and improved mask R-CNN[J]. Automation in Construction, 2023, 146: 104689.

[4] CHEN M Q, YU L J, ZHI C, et al. Improved faster R-CNN for fabric defect detection based on gabor filter with genetic algorithm optimization[J]. Computers in Industry, 2022, 134: 103551.

[5] LI C 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.

[6] 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, 2023: 7464-7475.

[7] WANG A, CHEN H, LIU L H, et al. YOLOv10: Real-time end-to-end object detection [J]. ArXiv preprint arXiv:2405.14458, 2024.

[8] KHANAM R, HUSSAIN M. YOLOv11: An overview of the key architectural enhancements[J]. ArXiv preprint arXiv:2410.17725, 2024.

[9] ZHU X K, LYU S C, WANG X, et al. TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios[C]//IEEE/CVF International Conference on Computer Vision, 2021: 2778-2788.

[10] 钱承山,沈有为,孙宁,等. 基于 Transformer 改进 YOLOv5 的山火检测方法研究[J]. 电子测量技术, 2023,46(16):46-56.
QIAN C S, SHEN Y W, SUN N, et al. Research on improved YOLOv5 for forest fire detection method based on Transformer[J]. Electronic Measurement Technology, 2023, 46(16): 46-56.

[11] WANG H, LIU C Y, CAI Y F, et al. YOLOv8-QSD: An improved small object detection algorithm for autonomous vehicles based on YOLOv8[J]. IEEE Transactions on Instrumentation and Measurement, 2024, 73: 1-12.

[12] 杜运亮, 王明甲. 基于半监督域适应的微弱光环境下行人检测研究[J]. 电子测量与仪器学报, 2024, 38(1): 106-113.
DU Y L, WANG M J. Research on pedestrian detection in low-light conditions based on semi-supervised domain adaptation [J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(1): 106-113.

[13] LIANG S Y, WU H, ZHEN L, et al. Edge YOLO: Real-time intelligent object detection system based on edge-cloud cooperation in autonomous vehicles [J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(12): 25345-25360.

[14] 黄崇庆, 徐慧英, 张晓雷, 等. BGR-YOLO: 基于 YOLOv8 改进的交通场景下目标检测算法[J/OL]. 计算机工程与科学, 1-13[2025-04-10]. https://link.cnki.net/urlid/43.1258.TP.20250408.1455.002.
HUANG C Q, XU H Y, ZHANG X L, et al. BGR-YOLO: An improved object detection algorithm under traffic scenarios based on YOLOv8[J/OL]. Computer Engineering & Science, 1-13[2025-04-10]. https://link.cnki.net/urlid/43.1258.TP.20250408.1455.002.

[15] WANG C, LIN M H, SHAO L, et al. MF-YOLO: A lightweight method for real-time dangerous driving behavior detection [J]. IEEE Transactions on Instrumentation and Measurement, 2024, 73: 1-13.

[16] ZHOU Y, WEN S J, WANG D L, et al. MobileYOLO: Realtime object detection algorithm in autonomous driving scenarios[J]. Sensors, 2022, 22(9): 3349.

[17] DENG C F, WANG M M, LIU L, et al. Extended feature pyramid network for small object detection[J]. IEEE Transactions on Multimedia, 2022, 24: 1968-1979.

[18] 唐彪, 贾小军, 骆淑云, 等. BL-YOLO: 无人机航拍图像目标检测算法[J/OL]. 光电子·激光, 1-12[2025-04-20]. https://link.cnki.net/urlid/12.1182.o4.20250423.1727.002.
TANG B, JIA X J, LUO S Y, et al. BL-YOLO: Object detection algorithm for UAV aerial images[J/OL]. Journal of Optoelectronics·Laser, 1-12[2025-04-20]. https://link.cnki.net/urlid/12.1182.o4.20250423.1727.002.

[19] 杨磊,陈艳菲,李海鸣,等. 基于改进 YOLOv8 的自动驾驶场景目标检测算法[J]. 计算机工程与应用, 2025, 61(1):131-141.
YANG L, CHEN Y F, LI H M, et al. Object detection algorithm for autonomous driving scenes based on improved YOLOv8 [J]. Computer Engineering and Applications, 2025, 61(1): 131-141.

[20] LUO Y C, CI Y S, JIANG S X, et al. A novel lightweight real-time traffic sign detection method based on an embedded device and YOLOv8 [J].

Journal of Real-Time Image Processing, 2024, 21(2): 24.

[21] 杨思念, 曹立佳, 郭川东, 等. 基于 PXD-YOLO11s 的陶瓷电路板缺陷检测方法研究[J]. 仪器仪表学报, 2025, 46(7): 307-318.
YANG S N, CAO L J, GUO C D, et al. Research on ceramic circuit board defect detection method based on PXD-YOLO11s [J]. Chinese Journal of Scientific Instrument, 2025, 46(7): 307-318.

[22] OUYANG D L, HE S, ZHANG G Z, et al. Efficient multi-scale attention module with cross-spatial learning [C]//IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2023: 1-5.

[23] 刘明杰,魏宇,陈俊生,等. 基于全局体素特征交互的 3D 目标检测算法[J]. 仪器仪表学报,2025,46(9): 146-158.
LIU M J, WEI Y, CHEN J S, et al. Global voxel feature interaction-based 3D object detection [J]. Chinese Journal of Scientific Instrument, 2025, 46(9): 146-158.

[24] WU T Y, TANG S, ZHANG R, et al. CGNet: A light-weight context guided network for semantic segmentation[J]. IEEE Transactions on Image Processing, 2021, 30: 1169-1179.

[25] 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.

作者简介

杨伟豪,硕士研究生,主要研究方向为目标检测、机器视觉。

E-mail:18914985673@163.com

徐晓龙(通信作者),硕士生导师,高级实验师,主要研究方向为机器视觉、机器学习、水下探测与成像、嵌入式系统设计与应用。

E-mail:xuxl@hhu.edu.cn