DOI: 10.13382/j.jemi.B2508611

DPM-YOLO:面向自动驾驶复杂场景的多尺度自适应目标检测算法*

吴文欢[1,2]　陈　杰[1]　王文舒[1]

(1 湖北汽车工业学院智能网联汽车学院　十堰　442000;2. 湖北汽车工业学院汽车动力传动与电子控制湖北省重点实验室　十堰　442000)

摘　要:目标检测是自动驾驶环境感知的核心技术,其任务是识别并定位车辆周围环境中的关键物体。针对自动驾驶场景中小目标漏检、遮挡误判及动态适应性不足的挑战,提出基于 YOLOv11n 的算法模型 DPM-YOLO。首先,在骨干网络(Backbone)中,构建动态注意力金字塔模块(DAP)替换原快速空间金字塔池化模块(SPPF),利用参数共享空洞卷积与分解式空间注意力高效地捕获不同粒度特征,从而更好地聚合目标的上下文信息;其次,在颈部网络(Neck)中,设计并行融合卷积模块(PFConv),采用异构多分支卷积动态解耦通道并融合多尺度特征;最后,提出一种改进的 MPDWIoU 损失函数,融合动态梯度聚焦与角点约束,结合批次统计特征自适应策略优化定位精度。实验结果表明,DPM-YOLO 在 KITTI 数据集上的 mAP@0.5 达 92.7%,较基准模型 YOLOv11n 提升 2.0 个百分点,在 Cityscapes 和 BDD100K 上的 mAP@0.5 较 YOLOv11n 分别提升了 1.6 和 0.8 个百分点,在极端天气数据集 Foggy-Cityscapes 上的 mAP@0.5 较 YOLOv11n 提升了 1.2 个百分点。本研究方法较好地实现了对复杂场景检测的精度和实时性平衡。

关键词:机器视觉;YOLOv11n;动态注意力金字塔;并行融合卷积;MPDWIoU 损失函数

中图分类号:TP391.41;TN911.7　　**文献标识码**:A　　**国家标准学科分类代码**:520.6040

DPM-YOLO: A multi-scale adaptive object detection algorithm for complex scenarios in autonomous driving

Wu Wenhuan[1,2]　Chen Jie[1]　Wang Wenshu[1]

(1. School of Intelligent Connected Vehicle, Hubei University of Automotive Technology, Shiyan 442000, China;
2. Key Laboratory of Automotive Power Train and Electronics, Hubei University of Automotive Technology, Shiyan 442000, China)

Abstract: Object detection is a core technology in the environmental perception of autonomous driving, and its task is to identify and locate key objects in the surrounding environment of the vehicle. To address the challenges of missed detection of small targets, misjudgment due to occlusion, and insufficient dynamic adaptability in autonomous driving scenarios, an improved object detection algorithm DPM-YOLO based on YOLOv11n is proposed. Firstly, in the Backbone network, a dynamic attention pyramid module is constructed to replace the original rapid spatial pyramid pooling structure. By utilizing parameter-shared dilated convolutions and a decomposed spatial attention mechanism, multi-granularity features are efficiently captured and the context information of each target can be better aggregated. Secondly, in the Neck network, a parallel fusion convolution module is designed. It can achieve channel-wise dynamic decoupling and multi-scale features fusion by heterogeneous multi-branch convolutions. Finally, an improved MPDWIoU loss function is proposed, which integrates a dynamic gradient focusing mechanism with corner constraints and optimizes bounding box localization accuracy by combining a batch statistics adaptive strategy. Experimental results demonstrate that the mAP@0.5 of DPM-YOLO on the KITTI dataset is 92.7%, which is 2.0 percentage points higher than that of YOLOv11n. The mAP@0.5 of DPM-YOLO on the Cityscapes and BDD100K datasets is improved by 1.6 and 0.8 percentage points respectively compared with YOLOv11n. The mAP@0.5 of DPM-YOLO on the extreme weather dataset Foggy-Cityscapes increases by 1.2 percentage points over YOLOv11n. The

收稿日期: 2025-07-28　　Received Date: 2025-07-28

*基金项目:湖北省自然科学基金创新发展联合基金项目(2025AFD239)、汽车动力传动与电子控制湖北省重点实验室开放基金项目(ZDK12025A02)资助

results indicate that the proposed method effectively balances detection accuracy and real-time performance in complex driving scenarios.
Keywords: machine vision; YOLOv11n; dynamic attention pyramid; parallel fusion convolution; MPDWIoU loss

0 引　言

随着自动驾驶技术从辅助驾驶(L2)向高度自动化(L4)演进,环境感知系统对实时性、鲁棒性与泛化性的要求日益严苛[1]。目标检测作为感知任务的核心模块,其性能直接决定自动驾驶系统的决策安全边界。然而,真实道路场景中目标尺度跨度极大、空间关系高度复杂,且受动态遮挡、极端环境等影响,不同目标极易混淆,实现同时满足高精度与低延迟的实时检测仍面临严峻挑战。

近年来,随着深度学习的快速发展,目标检测算法也取得显著进展。基于深度学习的目标检测算法主要分为两阶段算法和单阶段算法。两阶段算法,如 Faster R-CNN (region-based convolutional neural networks, R-CNN)[2]、Mask R-CNN[3],先获得候选区域,然后对候选区域进行定位和分类。两阶段算法通常具有较高检测精度,但计算复杂度高,难以满足实时需求。单阶段算法,如 SSD(single shot multibox detector)[4],YOLO (you only look once)[5-12]系列,利用端到端的网络直接生成目标框,并进行位置修正和分类,因其计算效率高成为自动驾驶系统的优先选择。

YOLO 系列方法近年来发展迅速,YOLOv3[5]引入多尺度预测增强不同尺寸目标检测。YOLOv4[6]通过 CSPDarknet53 骨干网络、PANet 特征聚合结构和 Mosaic 数据增强等策略提升检测性能。YOLOv5[7]优化工程易用性,通过自适应锚框计算、Focus 切片结构和混合数据增强提升易用性与效率。也有学者提出了无锚框检测算法。Law 等[13]提出 CornerNet 角点检测框架,通过检测目标的左上角与右下角关键点对实现无锚框定位,并设计角点池化机制增强边缘特征判别能力,但小目标角点特征区分度低,检测召回率受限。Tian 等[14]提出全卷积单阶段框架(fully convolutional one-stage object detection, FCOS),利用特征级约束与中心度分支机制抑制低质量预测,从而规避锚框超参数设计依赖。上述方法未能有效捕获目标的上下文信息,容易受背景、遮挡等因素干扰,导致出现小目标漏检、遮挡目标误检等问题。

为了解决上述问题,研究者们尝试引入注意力机制更好地汇聚目标上下文,进而增强目标的表征能力,提升检测精度。Chung 等[15]提出 YOLO-SLD 模型,通过通道与空间注意力动态强化纹理特征,有效抑制了背景干扰,实现高精度目标检测。Wang 等[16]在 YOLOv5 中嵌入自适应方向注意力模块(adaptive orientation attention module, AOAM),通过动态生成方向敏感权重强化目标的纹理特征,实现各尺寸目标的高效检测。Yang 等[17]在 YOLOv8 中融合可变形卷积与坐标注意力机制,利用坐标注意力对空间位置的敏感性,强化了关键特征区域的聚焦,增强了对小目标的辨识能力。邝先验等[18]设计了一种并行级联注意力机制改进 YOLOv7-tiny 网络,通过并行 3 分支网络与空间注意力模块相结合,增强了网络对多尺度目标特征的提取能力。尽管上述方法取得不错的效果,但难以在精度与速度之间取得良好平衡。

自 Vision Transformer (ViT)[19]提出以来,研究者们开始利用 Transformer 增强模型对全局语义信息的捕获能力,提升模型在密集遮挡等复杂环境下的检测性能。Zhang 等[20]提出 ViT-YOLO 模型,通过在 YOLOv4 骨干网络最深层引入多头自注意力机制,构建卷积和 Transformer 混合架构以捕获全局上下文,显著增强了对遮挡目标与微小目标的识别能力。Zhu 等[21]提出 TPH-YOLOv5,通过在 YOLOv5 的预测头中引入 Transformer 编码器模块,利用自注意力机制增强全局上下文建模能力,有效提升了无人机航拍场景下对密集遮挡目标的检测精度。陈俊生等[22]在 YOLOv8 主干网络中嵌入 C2f-ST (CSPDarknet53 to 2-stage FPN-swin transformer)特征提取模块,通过 Swin Transformer 的窗口自注意力机制,增强微小缺陷的局部-全局特征关联,结合残差连接保留浅层细节特征,提升了细粒度特征提取能力。这些基于 Transformer 的方法能实现复杂的全局上下文建模,但计算复杂度随序列长度呈平方级增长,即使仅在最低分辨率特征图应用,仍显著增加内存占用与推理延迟,难以满足实时性要求。

针对小目标检测不足、遮挡检测困难与实时性问题,本文以 YOLOv11n[23]为基准模型,提出一种改进 DPM-YOLO 模型,主要贡献如下:

1) 构建动态注意力金字塔模块(dynamic attention pyramid module, DAP)替换原快速空间金字塔池化模块(spatial pyramid pooling fast, SPPF),利用参数共享空洞卷积与分解式空间注意力协同捕获不同粒度特征,提升对遮挡目标的特征聚合精度。

2) 提出了并行融合卷积(parallel fusion Conv, PFConv)模块重构骨干网络标准卷积路径,通过异构多分支卷积同步提取局部至全局特征,结合动态通道融合策略突破单感受野限制,增强多尺度目标表征能力。

3) 提出 MPDWIoU 损失函数,融合动态梯度聚焦与角点距离约束,通过批统计量自适应的权重分配抑制离群样本干扰,同时强化边界框角点对齐精度,平衡复杂场景下的检测稳定性与定位准确性。

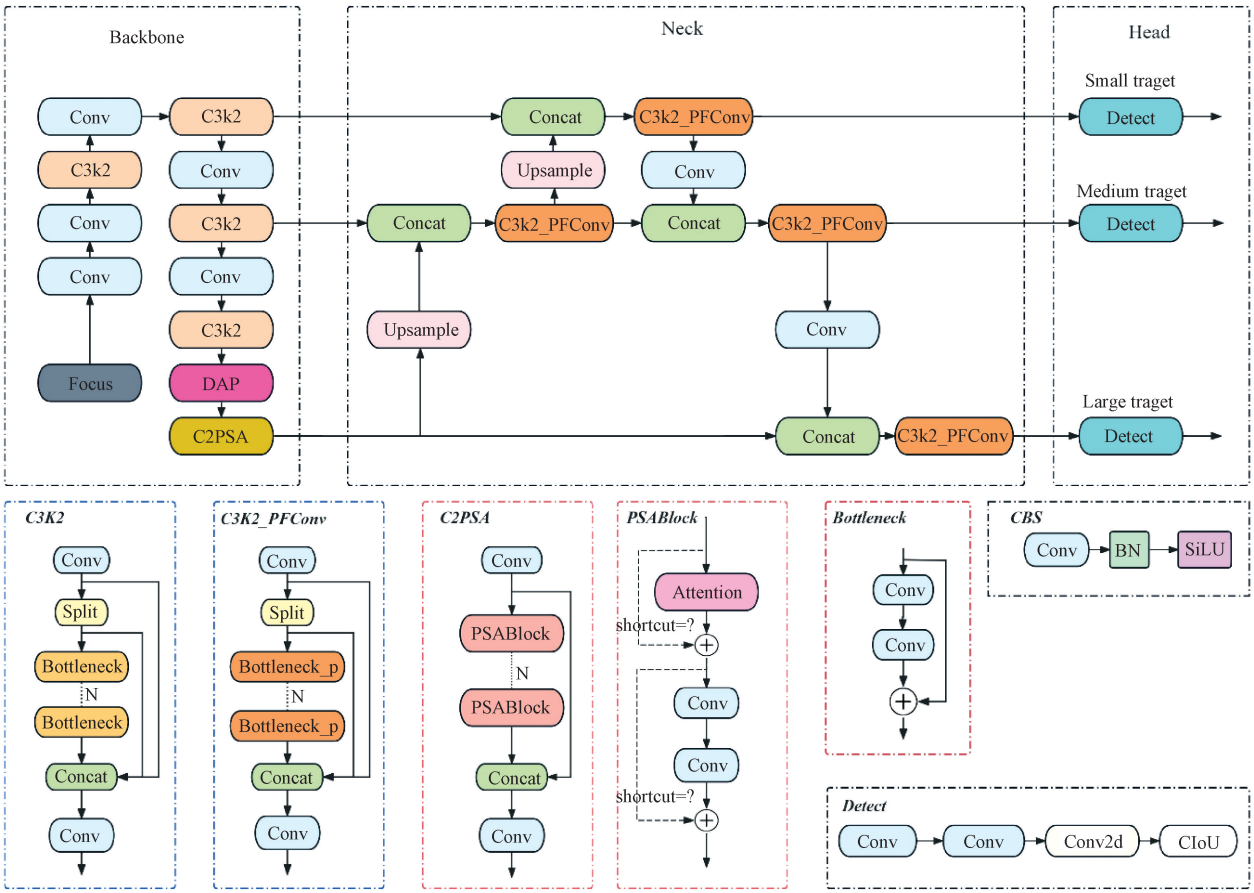

图 1　DPM-YOLO 网络结构图

Fig. 1　Network structure diagram of DPM-YOLO

1　YOLOv11 算法

YOLO 系列算法近年来仍在持续迭代。YOLOv6[8] 采用 EfficientRep 可重参数化骨干与无锚框检测头，实现训练-推理解耦及 GPU 部署优化。YOLOv7[9] 提出动态标签分配策略和可训练重参数化网络结构，解决训练时正样本分配冲突问题。YOLOv8[10] 通过任务解耦头支持多任务适配持续推动轻量化与多场景适配能力。YOLOv9[11] 通过高效特征融合架构、自适应样本分配策略及动态参数优化机制，提升特征利用效率，优化训练样本质量，平衡训练与推理性能，进一步增强实时目标检测在复杂场景下的鲁棒性与部署适配能力。YOLOx[12] 实现端到端（non-maximum suppression-free，NMS-free）训练，同时采用双分配策略（结合预测框与真实框的多维度匹配）优化样本分配，提升了训练效率与检测精度。

Ultralytics 团队在 2024 年继续推出了 YOLOv11[23]，直接继承 YOLOv8 的统一网络架构，延续了经典的 Backbone-Neck-Head 架构，并在此框架上进行了针对性优化与创新。模型 Backbone（骨干网络）负责多尺度特征提取，通过堆叠卷积层和专用模块 实现高效下采样，在降低空间维度的同时强化特征表达能力；Neck（颈部结构）集成高级特征融合技术，聚合不同层级的特征图，显著提升多尺度目标的检测能力；Head（检测头）采用统一预测机制（支持 C2PSA 并行空间注意力等组件），不仅完成目标定位与分类，还扩展至实例分割、姿态估计、定向目标检测等任务，实现多任务协同优化。

2　YOLOv11 算法改进

2.1　DAP 模块

为解决传统空间金字塔池化（SPPF）在特征聚合灵活性和计算效率方面的固有矛盾，本文提出了动态注意力金字塔（DAP），其核心思想是通过参数共享的空洞卷积与多尺度动态注意力机制组协同作用，实现上下文感知的特征融合。该模块采用级联式处理流程：首先利用通道压缩层对输入特征进行降维，在保留关键信息的同时降低后续计算开销；随后构建三路并行的空洞卷积分支，采用共享权重的 3×3 卷积核，并配合差异化膨胀

率(dilation rate = 1/3/5),在统一参数空间中捕捉局部细节、中等区域和全局语义等多层次上下文信息,同时每个分支引入了 SegNeXt[24] 方法的多尺度卷积注意力(multi-scale convolutional attention, MSCA)机制,将大尺寸卷积核分解为级联的非对称卷积操作(例如,将 7×7 核分解为 1×7 与 7×1 卷积序列),在保持大感受野优势的同时避免计算量呈平方级增长,进而实现对关键区域的动态聚焦。最终,通过跨分支的通道拼接和非线性投影,完成多尺度特征的深度融合。DAP 结构如图 2 所示。

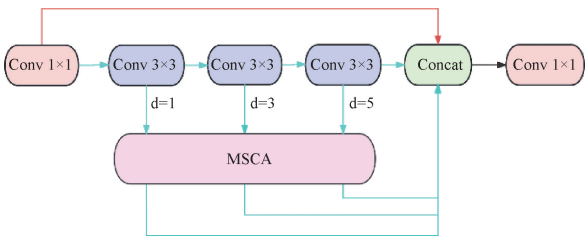

图 2　DAP 模块结构图

Fig. 2　Structural diagram of the DAP modul

MSCA 包含 3 个部分:一个深度卷积,用于聚合局部信息;多分支深度带状卷积,用于捕捉多尺度上下文;一个 1×1 卷积,用于模拟不同通道之间的关系。1×1 卷积的输出直接用作注意力权重,以重新权衡 MSCA 的输入。

MSCA 的数学表达为:

$$\boldsymbol{Att} = Conv_{1\times1}\left(\sum_{i=0}^{3} Scale_i(DW-Conv(\boldsymbol{F}))\right) \quad (1)$$

$$\boldsymbol{Out} = \boldsymbol{Att} \otimes \boldsymbol{F} \quad (2)$$

式中:$\boldsymbol{F}$ 代表输入特征;$\boldsymbol{Att}$ 和 $\boldsymbol{Out}$ 分别代表注意力图谱和输出;⊗是元素矩阵乘法运算;$DW\text{-}Conv$ 表示深度卷积;$Scale_i(i \in \{0,1,2,3\})$ 表示第 i 个分支。$Scale_0$ 分支采用恒等连接,$Scale_1$、$Scale_2$ 和 $Scale_3$ 分支使用 2 个深度条状卷积(depth-wise strip convolution)近似大核的标准深度卷积。本文每个分支的内核大小分别设置为 7、11 和 21。选择深度条状卷积的原因有 2 个:一方面,条状卷积是轻量级的,要模拟核大小为 7×7 的标准二维卷积,只需要一对 7×1 和 1×7 卷积;另一方面,在场景中存在一些条状物体,条状卷积可以作为 2D 卷积的补充,帮助提取条状特征。MSCA 结构如图 3 所示。

DAP 模块设计的核心创新体现在 3 个层面的技术突破:在特征提取层,借助共享权重的空洞卷积组构建参数高效的尺度空间,其膨胀率的逐级增大设计既确保了多尺度覆盖的完备性,又抑制了空洞卷积固有的背景噪声放大效应;在注意力生成层,采用多路径卷积组合策略生成空间权重图,其独特的非对称卷积分解设计将大尺寸注意力场的计算成本降低至线性复杂度;在特征融合层,

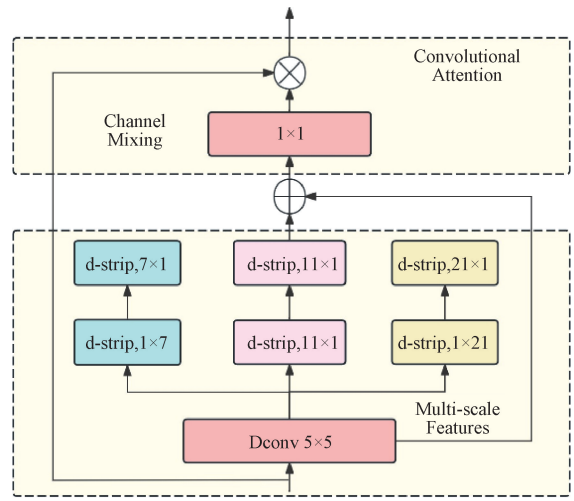

图 3　MSCA 结构图

Fig. 3　Structural diagram of MSCA

将原始卷积特征和注意力加权特征在通道维度上进行拼接,通过后续 1×1 卷积实现跨通道信息整合,相较于 SPPF 的固定尺寸池化核(5×5 最大池化),显著增强了对不规则目标的几何适应性。

2.2　PFConv 模块

本文提出的并行融合卷积模块 PFConv,旨在解决 C3k2 结构中 Bottleneck 块的特征表达能力不足问题,借助多尺度并行融合机制对其核心计算路径进行重构。

PFConv 模块的运作机理与参数设置深度耦合,其核心逻辑可概括为 3 级递进过程。首先,通过分组卷积策略将输入特征解耦至 4 个独立子空间,每个子空间采用专属尺寸的卷积核(1×1、3×3、5×5、7×7)同步提取特征——1×1 卷积核聚焦局部细节,3×3/5×5 捕捉区域纹理,7×7 扩大感受野以建立远距离语义关联,从而突破传统单分支卷积固定感受野的限制,实现从细粒度到粗粒度的多尺度特征覆盖。其次,通过通道维度重组与 1×1 卷积达成跨分支特征融合,在保留原始通道压缩结构(首个 1×1 卷积)的基础上,增强多尺度特征的语义交互能力。具体结构如图 4 所示。

设输入特征张量为 $\boldsymbol{X} \in \boldsymbol{R}^{B\times C\times H\times W}$,其中 B 为批次大小,C 为输入通道数,$H \times W$ 为特征图分辨率,该张量源于首个 1×1 卷积的通道压缩输出。

动态通道解耦的数学表达为:

$$\boldsymbol{X}_{group} = \mathcal{R}(\boldsymbol{X}) \in \boldsymbol{R}^{B\times d\times H\times W\times g}(d = C/g) \quad (3)$$

式中:d 表示每组通道数,将特征图均匀拆分为 g(对应异构卷积核数量),为异构卷积提供单独处理空间;R 表示张量重塑操作。此步骤显著降低后续计算复杂度,同时保留多尺度特征的多样性基础。

随后用 3 路并行卷积核同步处理分组特征,每路对应特定尺度建模:

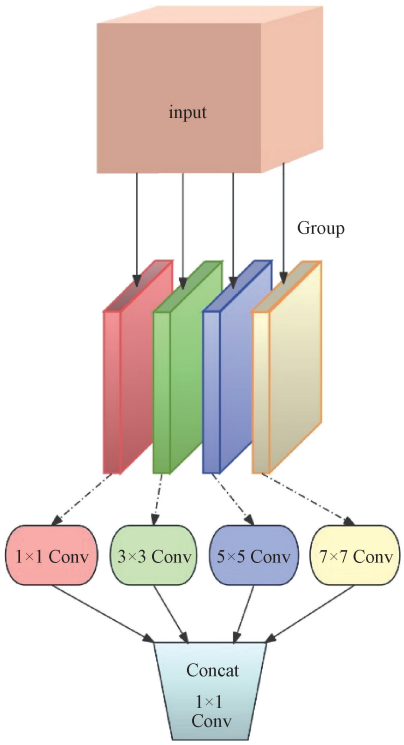

图4　PFConv模块图

Fig. 4　Schematic diagram of the PFConv module

$$\boldsymbol{Y}_i = SiLU(BN(Conv^{(i)}_{k_i\times k_i}(\boldsymbol{X}^{(i)}_{group}))) \tag{4}$$

式中：k_i 为第 i 组异构卷积核尺寸；$\boldsymbol{X}^{(i)}_{group}$ 是第 i 组特征切片；$Conv^{(i)}_{k_i\times k_i}$ 代表第 i 组所用的卷积核，通过分组参数共享策略，实现3路卷积核权重共享，在扩大感受野范围的同时将计算开销控制在一定范围内。

最后，多分支输出通过可学习的加权融合机制整合：

$$\boldsymbol{Y}_{concat} = \bigoplus_{i=0}^{2} \boldsymbol{Y}_i \in \boldsymbol{R}^{B\times d\times H\times W\times g} \tag{5}$$

$$\boldsymbol{Y}_{fused} = \mathcal{F}(\boldsymbol{Y}_{concat}) \in \boldsymbol{R}^{B\times C\times H\times W} \tag{6}$$

融合函数 F 本质上是通道维度的自适应重组，通过端到端训练调整各分支的权重分配。最终经1×1卷积压缩至目标通道数，并通过残差连接与Bottleneck输入相加，保留参数权重传播路径以加速收敛。

如图5所示，本文将PFConv模块替换C3k2中Bottleneck第2个卷积层，重构了标准Bottleneck的卷积路径，并命名Bottleneck_P，其特点是异构卷积分支与轻量化。

与原结构相比，PFC模块通过通道拆分再合并策略实现参数共享，在几乎不增加计算复杂度的情况下，有效平衡了特征多样性与计算效率。此外，模块化的设计使其能够无缝融入YOLOv11主干网络，通过残、差连接保留梯度传播路径，确保训练的稳定性。这一改进尤其适用于存在显著尺度变化的检测场景（例如监控中的近、远景车辆），其多粒度特征融合机制可显著提高复杂背景下小目标的边缘清晰度以及遮挡目标的特征判别性，为平衡轻量化检测模型的高精度与实时性提供了新的思路。

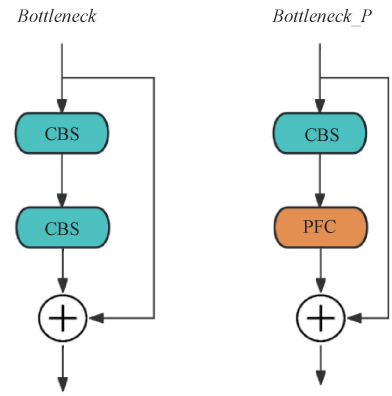

图5　Bottleneck结构改进

Fig. 5　Improvement of the Bottleneck structure

2.3　改进定位损失函数

1）CIoU损失函数

YOLOv11的目标检测损失函数建立在经典的多任务损失框架之上，其核心边界框定位损失采用了完整交并比（complete intersection over union，CIoU）这一损失函数，表示预测边界与真实边界框之间的整体差异。损失值越小，表示预测框与真实框的匹配度越高，其表达式如式(7)所示。

$$\mathcal{L}_{CIoU} = 1 - IoU + \frac{\rho^2(\mathrm{Box},\mathrm{Box}^{gt})}{c^2} + \alpha\times v \tag{7}$$

式中：IoU计算预测框Box和真实框Box^{gt}的交集与并集的比值，范围在[0,1]；$\frac{\rho^2(\mathrm{Box},\mathrm{Box}^{gt})}{c^2}$表示中心点距离惩罚项，用于惩罚预测框中心与真实框中心之间的距离；$\rho(\mathrm{Box},\mathrm{Box}^{gt})$是欧氏距离函数；$c$为最小封闭矩形的对角线长度；$\alpha$表示宽高比差异度量；$v$是平衡权重，$\alpha v$用于惩罚预测框与真实框在宽高上的差异。

CIoU定位损失有2个核心劣势。首先，角点约束缺失，CIoU仅通过中心点距离惩罚和宽高比惩罚优化边界框，缺乏对关键角点（左上/右下）的直接位置约束，导致预测框边界容易偏移。尤其在小目标、细长目标和高密度场景中，像素级定位精度受限。其次，离群点敏感导致训练不稳定，CIoU对IoU接近0的低质量样本（如遮挡目标、噪声标注）极为敏感，导致1－IoU和$\frac{\rho^2(\mathrm{Box},\mathrm{Box}^{gt})}{c^2}$会产生剧烈波动，显著降低训练稳定性，并干扰正常样本优化。

2）MPDWIoU损失函数

针对现有CIoU损失函数存在的离群点敏感性高和边界框角点对齐精度不足的问题，本文提出了一种新的

MPDWIoU 损失函数。原 MPDIoU[25] 通过协同设计动态梯度聚焦机制和角点距离优化模块,显著提升了目标检测模型在复杂场景下的鲁棒训练能力和定位精度。

首先是动态梯度聚焦 IoU 项,引入基于 Wise-IoU v3[26] 框架的离群因子 r_i 动态调制样本损失权重。

$$L_{\mathrm{WIoUv3}} = \frac{r_i}{\bar{r}} \times (1 - \mathrm{IoU}) \tag{8}$$

$$r_i = \frac{(\mathrm{IoU}_i - \mu_{\mathrm{IoU}})^2}{\sigma_{\mathrm{IoU}^2} + \delta} \tag{9}$$

式中:μ_{IoU} 和 σ_{IoU} 为当前批次 IoU 的均值与标准差;δ 是一个极小的常数用于数值稳定性(防止分母为 0);$\bar{r}$ 是归一化因子以确保权重分布合理。该机制的核心思想是:IoU 显著低于批次平均水平的低质量样本(离群点)会被赋予较大的 r_i 值,从而导致其梯度 $\frac{r_i}{\bar{r}}$ 被抑制(相对于高质量样本被放大),有效减轻低质量样本对模型训练的干扰。

然后是角点距离惩罚项,直接优化预测框与真实框 4 个角点之间的空间一致性,沿用了 MPDIoU 中对角点距离的精确建模方式。

$$P_{corner} = \frac{d_1^2 + d_2^2}{w^2 + h^2} \tag{10}$$

式中:d_1 和 d_2 为预测框与真实框左上、右下角点的欧氏距离;w 和 h 为图像宽和高。

最终损失函数 MPDWIoU 通过可学习的权重参数 λ 将上述 2 项有机结合:

$$\mathrm{L}_{\mathrm{MPDWIoU}} = 1 - (\mathrm{WIoUv3} - \lambda \times P_{\mathrm{corner}}) \tag{11}$$

在训练过程中,参数 λ 会自适应调整,其核心作用机理是通过梯度下降实现自适应调整,依据 2 项的实时状态动态分配权重。当动态 IoU 项损失更大时,λ 增大以加强对整体边界框交并比的优化。当角点惩罚项误差更显著时,λ 调整以提升角点惩罚项的权重。通过数据驱动的动态平衡,实现整体边界框 IoU 与局部角点定位精度的协同优化,其表达式如下:

$$\lambda = \frac{\exp(WIoUv3/\tau)}{\exp(WIoUv3/\tau) + \exp(P_{corner}/\tau)} \tag{12}$$

式中:参数 τ 控制权重对损失差异的敏感程度,取定值 $\tau=2$,τ 越小,权重分化越显著。

3　实验与结果分析

3.1　实验设置

本文实验环境和实验参数具体如表 1 所示。

3.2　数据集

本文主要在无人驾驶公开数据集 KITTI[27] 上开展实验,另外,利用 Cityscapes[28] 和 BDD100K[29] 以及 Foggy-Cityscapes[30] 等公共数据集对所提模型的泛化性能进行检验。KITTI 数据集由卡尔斯鲁厄理工大学与丰田技术研究院联合构建,包含 7 481 张城市、乡镇及高速公路场景的标注图像(分辨率 1 242×375 pixels),涵盖车辆、行人等关键目标,复杂场景下的目标混淆与残缺特性使其成为自动驾驶研究中的一个重要基准测试集,数据按 8∶1∶1 划分为 5 985∶748∶748 的样本。BDD100K 数据集由加州大学伯克利分校开发,采集自美国全境 10 万张多气候道路图像,包括晴天、阴天、雨天等多个气象状况,采用 7∶1∶2 划分策略形成 70 000∶10 000∶20 000 的训练、测试、验证集。Cityscapes 聚焦城市街景,提供 50 个城市 5 000 张高分辨率图像,通过保留汽车、行人等 9 类核心交通要素,按 6∶1∶3 比例构建 2 975∶500∶1 525 数据集,其精细语义标注支持复杂城市场景的跨任务验证。Foggy-Cityscapes 以 Cityscapes 精细子集为基础,通过大气散射模型生成不同程度雾效,并附加 8 类目标的 2D 边界框标注,是目前雾天目标检测领域的主流基准。多个不同场景数据集的协同应用可有效提升模型在目标多样性、环境复杂度等方面的泛化性能。

表 1　实验环境和实验参数设置

Table 1　Experimental environment and experimental

项目	内容
CPU	AMD EPYC 9754 128-Core
RAM	16 GB
GPU	RTX 4090D(24 GB) * 1
CUDA 版本	Cuda11.3 and Cudnn8.1
Pytorch 深度学习框架	torch-2.0.0+cu118
操作系统	Ubuntu 18.04.6
Python 版本	Python-3.8.10
Batch-size	32
Epoch	300
输入图片大小/pixel	640×640
学习率	0.01

3.3　模型评估指标

本文实验采用精确率(Precision,P)、召回率(Recall,R)、mAP@0.5、mAP@0.5:0.95、模型参数量(Param)、每秒检测帧数(FPS)以及每秒执行的浮点运算次数(GFLOPs)作为评估指标。

在本文实验中,mAP 是衡量模型性能关键指标,备受关注。mAP 可以表示为:

$$AP = \int_0^1 P(R), \mathrm{mAP} = \frac{\sum_{i=1}^{C} AP_i}{C} \tag{13}$$

式中:AP 表示每个类别在不同召回率下的精度,并取所有 C 个类别的平均值;mAP 是不同 IoU 阈值下的平均精度。其中,mAP@0.5 表示交并比阈值为 0.5 时的平均精

度,mAP@0.5:0.95 表示交并比阈值在 0.5~0.95 范围内(步长 0.05)的平均精度。

精确率 P 和召回率为 R 的计算公式如下:

$$P = \frac{TP}{TP + FP}, R = \frac{TP}{TP + FN} \tag{14}$$

式中:TP 是指预测框与真实框的 IoU>0.5 的正确检测数量;FP 表示预测框与真实框的 IoU≤0.5 的错误检测数量,而 FN 指的是未能检测出的真实目标的数量。

3.4 实验结果

1)主流算法实验对比

证明模型有效性和性能最直接的方法是在基准数据集上将其与最先进的方法进行比较,本文采用几种具有代表性的单阶段和多阶段检测器进行比较。

表 2 展示了不同目标检测模型在 KITTI 数据集上的性能对比。本文方法(记为 Ours)在 mAP@0.5 上达到 92.7%,优于主流模型如 YOLOv8s、YOLOv7-tiny 和 Cascade R-CNN[31]等。尽管参数量和计算量略高于基线模型 YOLOv11n,但得益于动态注意力金字塔(DAP)和多分支并行融合卷积(PFConv),以及对损失函数的改进,使得 mAP@0.5 提升了 2.0 个百分点。此外,本方法的推理速度在参数量相近的模型中表现更优,验证了其高效性。

表 2　KITTI 数据集上不同模型性能的对比实验结果

Table 2　Model comparison experiments on the KITTI dataset

模型	Param/($\times10^6$)	GFLOPs	FPS	*R*%	*P*%	mAP@0.5/%	mAP@0.5:0.95%
YOLOv3[5]	61.5	23.6	92.5	65.4	72.1	88.1	70.1
YOLOv4[6]	63.9	28.5	167.8	64.0	75.6	86.8	68.4
YOLOv5n[7]	3.2	**4.6**	154.3	82.5	90.3	88.2	61.5
YOLOv5s[7]	12.6	16.8	136.2	86.2	93.1	89.1	65.9
YOLOv7-tiny[9]	6.2	13.2	122.1	78.2	90.9	90.2	72.5
YOLOv8n[10]	6.1	8.1	170.2	80.7	89.7	88.9	72.3
YOLOv9-tiny[11]	**2.1**	7.7	48	78.1	87.3	86.4	70.1
YOLOx-tiny[12]	5.03	13.3	126.1	75.5	88.2	87.5	68.9
SSD[4]	23.6	60.9	35.3	45.2	68.3	50.2	42.7
Faster R-CNN[2]	28.4	83.2	11.6	68.9	79.4	75.6	58.3
Cascade R-CNN[31]	23.5	106.8	77.2	80.2	88.6	91.2	**74.5**
NanoDet-plus2.0[32]	4.7	8.1	89.1	72.6	83.4	82.7	64.1
RTMDet-tiny[33]	4.9	18.4	110.3	78.5	89.3	88.9	73.3
CrowdDet[34]	43.7	108.6	47.1	79.3	85.4	88.2	69.6
YOLOv11n[23]	5.2	6.3	**259.2**	86.4	92.1	90.7	70.1
Ours	5.8	6.5	231.5	**89.3**	**93.4**	92.7	72.6

2)消融实验

为了验证拟议模块的具体贡献并评估其性能,设计并进行了消融实验,实验结果如表 3 所示。以 YOLOv11n 为基线模型,通过添加不同模块逐步构建了一系列消融模型,然后定性分析每个模块对算法检测性能的影响。根据消融实验的结果,得出以下结论:单独引入 PFConv 模块后,mAP@0.5 提升 0.2%,参数量仅增加 0.1×10^6 验证了其通过多分支并行卷积增强特征表达的轻量化优势。采用 DAP 模块替换传统空间金字塔后,mAP@0.5 提升 0.5%,但因动态注意力计算引入额外参数,推理速度略微下降。而 MPDWIoU 损失通过动态聚焦机制优化定位精度,使 mAP@0.5 提升 0.7%,且不增加模型复杂度。模块组合实验进一步揭示了设计逻辑的协同性:当 PFConv 与 DAP 联合使用时,mAP@0.5 较基线提升 1.2%。而 DAP 与 MPDWIoU 组合则通过注意力机制与损失函数的跨层级优化,使 mAP@0.5 达到 92.1%。最终,完整模型集成 PFConv、DAP 和 MPDWIoU,mAP@0.5 达到 92.7%,较基线绝对提升 2.0%。改进模型和基线模型以及主流模型在 300 轮训练中 mAP@0.5 值的变化以折线图式绘制,如图 6 所示。

3)PFConv 卷积核组合消融实验

为验证 PFConv 模块中异构卷积核配置的合理性,在 DPM-YOLO 框架下设计了卷积核组合的消融实验。固定分组数 $G=4$,通过对比同构卷积、缺失特定尺度卷积以及增加冗余分支等多种配置,系统评估不同感受野组合对检测性能的影响。实验基于 KITTI 数据集,采用统一的训练策略与评估指标,重点分析各配置在参数量、计算开销与 mAP@0.5 之间的权衡关系。由表 4 知{1×1,3×3,5×5,7×7}组合的性能最优。

4)注意力机制对比实验

为验证 DynaAttnPyramid(DAP)模块中多尺度卷积(MSCA)通过融合空洞卷积的多尺度先验与动态通道注意力 MSCA(Multi-Scale Convolutional Attention)的有效性,在 KITTI 数据集上对比了 SE(Squeeze-and-Excitation)、CBAM(Convolutional Block Attention Module)、TrA(TripleAttention)等 7 类方法(如表 5 所

示)。DP-YOLO 是本文模型 DPM-YOLO 中的 DAP 模块不添加注意力机制。实验表明,本文提出的多尺度卷积注意力加权机制,提升 mAP@0.5 至 92.7%,较无注意力机制的基线(DP-YOLO)绝对提升 1.5 个百分点,且推理速度保持在 236.1 FPS(仅损失 3.8%),验证了跨尺度上下文建模对复杂驾驶场景的适应性。

表 3　在 KITTI 数据集上的消融试验结果

Table 3　Ablation study on the effectiveness of the design on the KITTI

模型	DAP	PFConv	Mpdwiou	Param/($\times10^6$)	FPS	P%	mAP@0.5/%
YOLOv11n				5.2	259.2	92.1	90.7
YOLOv11n	√			5.5	245.3	92.4	91.2
YOLOv11n		√		5.3	246.5	91.8	90.9
YOLOv11n			√	5.2	251.2	92.5	91.4
YOLOv11n	√	√		5.7	236.1	92.6	91.9
YOLOv11n	√		√	5.5	245.3	93.0	92.1
YOLOv11n		√	√	5.3	**247.5**	92.6	91.7
Ours	√	√	√	5.8	231.5	**93.4**	92.7

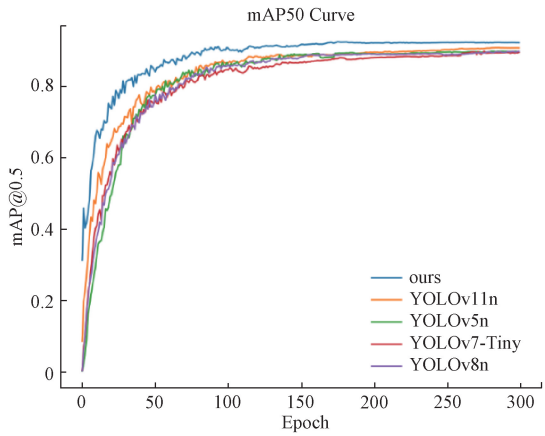

图 6　实验训练过程对比

Fig. 6　Comparison of the experimental training processes

表 4　卷积核组合消融实验

Table 4　PFConv kernel combination ablation study

卷积核配置	Param/($\times10^6$)	GFLOPs	mAP@0.5/%
{3×3,3×3, 3×3,3×3}	**89.5**	13.2	91.3
{3×3, 5×5, 7×7,7×7}	89.9	14.8	92.1
{1×1,3×3, 5×5,5×5}	89.7	13.3	91.8
{1×1,3×3, 5×5,7×7}	89.8	**13.1**	**92.7**
{1×1,3×3, 5×5,9×9}	89.9	14.1	92.5

5)损失函数对比实验

针对边界框回归损失函数的优化,本文使用集成 DAP 与 PFConv 的 YOLOv11n 模型(DPCI-YOLO)在 KITTI 数据集上对比了 GIOU、SIOU 等多种损失函数(如表 6 所示)。实验结果显示,一般损失函数(如 GIOU、CIOU)通过引入方向感知或长宽比约束,逐步提升 mAP@0.5 至 91.7%(MPDIoU)。而本文提出的动态梯度聚焦损失 MPDWIoU 通过加权困难样本梯度,进一步将 mAP@0.5 推至 92.7%,较 CIoU 提升 1.8 个百分点,且 mAP@0.5∶0.95 达到 68.0%。实验表明,MPDWIoU 在遮挡目标(Recall 提升 1.5 个百分点)和低对比度目标(Precision 提升 1.8%)上的优化效果最为显著,其动态梯度调整机制有效缓解了复杂场景下的定位模糊问题。

表 5　本文模型使用不同注意力机制的性能对比

Table 5　Comparison of the model in this paper using different attention mechanisms　(%)

模型	*P*	*R*	mAP@0.5
DP-YOLO	91.9	89.2	91.2
DP-YOLO-SE[35]	90.7	88.5	89.7
DP-YOLO-CBAM[36]	91.6	87.8	91.1
DP-YOLO-TrA[37]	89.7	87.3	88.5
DP-YOLO-CA[38]	91.5	89.8	91.9
DP-YOLO-ECA[39]	91.7	89.5	91.4
DP-YOLO-EMCA[40]	92.3	90.1	91.5
DP-YOLO-DEA[41]	91.9	89.7	91.8
Ours	**93.7**	**90.5**	**92.7**

表 6　使用不同损失函数的对比实验结果

Table 6　Comparative experiments using different loss functions　(%)

模型	*P*	*R*	mAP@0.5
DPCI-YOLO	89.5	88.0	91.6
DPCI-YOLO-GIoU	89.6	88.1	91.7
DPCI-YOLO-DIoU	89.7	88.2	91.8
DPCI-YOLO-CIoU	89.8	88.3	91.9
DPCI-YOLO-EIoU	89.9	88.5	91.0
DPCI-YOLO-SIoU	90.0	88.7	91.2
DPCI-YOLO -WIoU	90.2	89.0	91.5
DPCI-YOLO-MPDIoU	90.5	89.5	91.7
Ours	**91.0**	**90.0**	**92.7**

6)泛化实验

为评估模型的跨域泛化能力,本文在 Cityscapes 和 BDD100K 以及 Foggy-Cityscapes 数据集上进行了对比实验。如表 7 显示,本方法在 Cityscapes 上实现 46.2%的 mAP@0.5,较 YOLOv11n 提升 1.6 个百分点。在 BDD100K 多样化天气场景下,模型 mAP@0.5 达到 57.2%。在 Foggy-Cityscapes 极端天气场景下,mAP@0.5 依然较基准模型提高 1.2 个百分点,验证了 PFConv 多尺度特征与 DAP 动态注意力的强泛化性。尽管推理速度略有下降,但仍优于基准模型。

7)效果可视化

为直观展示改进模型相对于基线模型的优势,本节在 KITTI 测试集上进行了可视化对比分析。图 7 提供了热力图对比的可视化结果,其中图 7(a)为输入原图,图 7(b)和(c)分别为基线模型与改进模型的识别热力图。

表 7　Cityscape、BDD100K、Foggy-Cityscapes 数据集泛化实验

Table 7　Generalization experiments on the Cityscapes and BDD100K datasets and Foggy-Cityscapes

数据集	算法模型	FPS	mAP@0.5/%
Cityscapes	YOLOv11n	**218.8**	44.6
	Ours	199.2	**46.2**
BDD100K	YOLOv11n	**221.5**	56.4
	Ours	203.1	**57.2**
Foggy-Cityscapes	YOLOv11n	195.6	26.7
	Ours	**196.5**	**27.9**

通过对比可知,本文模型通过 DAP 模块的动态注意力机制与 PFConv 的多尺度梯度融合,使模型更精准地聚焦于目标关键区域,增强了检测图像的辨识度与抗干扰能力,有效抑制了背景噪声干扰,增强了特征的可区分度。

(a)　(b)　(c)

图 7　热力图可视化

Fig. 7　Visualization of the heatmap

图 8 展示了本文模型 DPM-YOLO 与基线模型 YOLOv11n 在 KITTI 数据集上的最终检测效果对比。图 8(a)为原图,图 8(b)为基线模型识别效果图,图 8(c)为改进模型识别效果图。由图 8 可以看出,本文模型有效检出了基线模型忽略的微小目标和重叠目标(如远处行人/汽车),降低了漏检率,提升了整体目标识别精度,在密集场景和小目标检测方面表现突出。

(a)　(b)　(c)

图 8　模型检测可视化

Fig. 8　Visualization of model detection

4　结　论

本文针对自动驾驶场景中目标检测面临的复杂挑战,提出了一种改进的目标检测框架 DPM-YOLO,该框架基于 YOLOv11n 架构进行三重创新优化。多分支并行融合卷积(PFConv),通过增强特征表达能力,显著提高了模型对多尺度目标的适应能力。动态注意力金字塔(DAP),优化了多尺度特征的融合过程,增强了对小目标和极端尺度目标的检测鲁棒性。改进定位损失函数(MPDWIoU),融合动态梯度聚焦机制和角点距离优化,有效提升边界框定位精度,增强模型训练稳定性。在 KITTI 数据集上的实验表明,改进模型的 mAP@0.5 达到 92.7%,较基线模型 YOLOv11n(90.7%)提升 2.0 个百分点,同时 mAP@0.5:0.95 提升 2.5 个百分点。在 BDD100K 多气候数据集上达到 mAP@0.5 为 57.2%,超越 YOLOv11n 的 56.4%。在 Cityscapes 精细城市场景中 mAP@0.5 取得 46.2%,较 YOLOv11n44.6%更高。在 Foggy-Cityscapes 极端天气场景下 mAP@0.5 比 YOLOv11n 提高 1.2 个百分点。

实验结果表明,本文提出的目标检测算法 DPM-YOLO 能够在极端尺度目标、密集遮挡等复杂场景下对道路车辆与行人实现精准实时检测,该算法为自动驾驶系统提供了新的环境感知方案,可降低复杂路况下的误判风险,提升行车安全冗余度。

参考文献

[1] 侯冠宇,赵万鑫,徐政. 新质生产力赋能无人驾驶汽车产业发展:理论、问题与前景[J]. 科学与管理,2025,45(2):17-23.
HOU G Y, ZHAO W X, XU ZH, et al. Empowering the development of the unmanned vehicle industry with new-quality productivity: Theory, problems and prospects [J]. Science and Management, 2025, 45(2): 17-23.

[2] GIRSHICK R. Fast R-CNN [C]. Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448.

[3] HE K, GKOXARI G, DOLLÁR P, et al. Mask R-CNN[C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2961-2969.

[4] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]. Proceedings of the European Conference on Computer Vision, 2016: 21-37.

[5] REDMON J, FARHADI A. YOLOv3: An incremental improvement [J]. ArXiv preprint arXiv: 1804.02767, 2018.

[6] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection[J]. ArXiv preprint arXiv:2004.10934, 2020.

[7] JOCHER G, STOKEN A, BOROVECJ, et al. Ultralytics/YOLOv5: v3.0[EB/OL]. https://github.com/ultralytics/yolov5/tree/v3.0.

[8] LI C, LI L, JIANG H, et al. YOLOv6: A single-stage object detection framework for industrial applications[J]. ArXiv preprint arXiv:2209.02976, 2022.

[9] 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]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 7464-7475.

[10] SOHAN M, SAI R T, RAMI R C V. A review on YOLOv8 and its advancements[C]. Proceedings of the International Conference on Data Intelligence and Cognitive Informatics, 2024: 529-545.

[11] WANG C Y, YEH I H, MMARK LIAO H Y. YOLOv9: Learning what you want to learn using programmable gradient information[C]. Proceedings of the European Conference on Computer Vision, 2024: 1-21.

[12] GE Z, LIU S, WANG F, et al. YOLOx: Exceeding YOLO series in 2021 [J]. arXiv preprint arXiv: 2107.08430, 2021.

[13] LAW H, DENG J. Cornernet: Detecting objects as paired keypoints [C]. Proceedings of the European Conference on Computer Vision, 2018: 734-750.

[14] TIAN Z, SHEN C, CHEN H, et al. Fcos: Fully convolutional one-stage object detection[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 9627-9636.

[15] CHUANG M A, LIN Y J, LIN C W. YOLO-SLD: An attention mechanism-improved YOLO for license plate detection[J]. IEEE Access, 2024, 12: 89035-89045.

[16] WANG X, HAN C, HUANG L, et al. AG-Yolo: Attention-guided YOLO for efficient remote sensing oriented object detection[J]. Remote Sensing, 2025, 17(6): 1027.

[17] YANG W, WU J, ZHANG J, et al. Deformable convolution and coordinate attention for fast cattle detection[J]. Computers and Electronics in Agriculture, 2023, 211: 108006.

[18] 邝先验,程福军,吴翠琴,等. 基于改进 YOLOv7-tiny 的

高效轻量遥感图像目标检测方法[J]. 电子测量与仪器学报,2024,38(7):22-33.
KUANG X Y, CHENG F J, WU C Q, et al. An efficient and lightweight object detection method for remote sensing images based on improved YOLOv7-tiny[J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(7): 22-33.

[19] DOSOVITSKIY A. An image is worth 16x16 words: Transformers for image recognition at scale[J]. ArXiv preprint arXiv:2010.11929, 2020.

[20] ZHANG Z, LU X, CAO G, et al. ViT-YOLO: Transformer-based YOLO for object detection [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 2799-2808.

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

[22] 陈俊生,陈沂蒙,刘明杰,等. 基于 MFES-YOLOV8n 的光伏电池缺陷检测方法[J]. 仪器仪表学报, 2025, 46(6):251-262.
CHEN J SH, CHEN Y M, LIU M J, et al. A photovoltaic cell defect detection method based on MFES-YOLOv8n[J]. Chinese Journal of Scientific Instrument, 2025, 46(6): 251-262.

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

[24] GUO M H, LU C Z, HOU Q, et al. Segnext: Rethinking convolutional attention design for semantic segmentation[J]. Advances in neural information processing systems, 2022, 35: 1140-1156.

[25] MA S, XU Y. Mpdiou: A loss for efficient and accurate bounding box regression [J]. arxiv preprint arxiv: 2307.07662, 2023.

[26] TONG Z, CHEN Y, XU Z, et al. Wise-IoU: Bounding box regression loss with dynamic focusing mechanism[J]. arxiv preprint arxiv:2301.10051, 2023.

[27] GEIGER A, LENZ P, STILLER C, et al. Vision meets robotics: The kitti dataset[J]. The International Journal of Robotics Research, 2013, 32(11): 1231-1237.

[28] CORDTS M, OMRAN M, RAMOS S, et al. The cityscapes dataset for semantic urban scene understanding [C]. Proceedings of the IEEE Conference on Computer Cision and Pattern Recognition, 2016: 3213-3223.

[29] YU F, CHENG H, WANG X, et al. Bdd100k: A diverse driving dataset for heterogeneous multitask learning[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 2636-2645.

[30] SAKARIDIS C, DAI D, VAN G L. Semantic foggy scene understanding with synthetic data [J]. International Journal of Computer Vision, 2018, 126(9): 973-992.

[31] CAN Z, VASCONCELOS N. Cascade R-CNN: Delving into high quality object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 6154-6162.

[32] MIAO D, WANG Y, YANG L, et al. Foreign object detection method of conveyor belt based on improved Nanodet[J]. IEEE Access, 2023, 11: 23046-23052.

[33] LYU C, ZHANG W, HUANG H, et al. Rtmdet: An empirical study of designing real-time object detectors[J]. ArXiv preprint arXiv:2212.07784, 2022.

[34] CHU X, ZHENG A, ZHANG X, et al. Detection in crowded scenes: One proposal, multiple predictions[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 12214-12223.

[35] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.

[36] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module[C]. Proceedings of the European Conference on Computer Vision, 2018: 3-19.

[37] MISRA D, NALAMADA T, ARASANIPALAI A U, et al. Rotate to attend: Convolutional triplet attention module [C]. Proceedings of the IEEE/CVF winter Conference on Applications of Computer Vision, 2021: 3139-3148.

[38] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13713-13722.

[39] WANG Q, WU B, ZHU P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11534-11542.

[40] RAHMAN M M, MUNIR M, MARCULESCU R. Emcad: Efficient multi-scale convolutional attention decoding for medical image segmentation [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 11769-11779.

[41] SUEYOSHI T, YUAN Y, GOTO M. A literature study for DEA applied to energy and environment[J]. Energy Economics, 2017, 62: 104-124.

作者简介

吴文欢,2020 年于西安理工大学获得博士学位,现为湖北汽车工业学院智能网联汽车学院副教授,主要研究方向为计算机视觉、图像处理及人工智能。

E-mail:wuwenhuan5@ 163. com

Wu Wenhuan received his Ph. D. degree from Xi′ an University of Technology in 2020. Now he is an associate professor in School of Intelligent Connected Vehicle, Hubei University of Automotive Technology. His main research interests include computer vision, image processing, and artificial intelligence.

陈杰(通信作者),2023 年于成都理工大学工程技术学院获得学士学位,现为湖北汽车工业学院智能网联汽车学院硕士研究生,主要研究方向为深度学习与目标检测。

E-mail:2472814733@ qq. com

Chen Jie (Corresponding author) received his B. Sc. degree from the Engineering & Technical College of Chengdu University of Technology in 2023. Now he is a M. Sc. candidate at School of Intelligent Connected Vehicle, Hubei University of Automotive Technology. His main research areas are deep learning and object detection.

王文舒,2020 年于运城学院获得学士学位,现为湖北汽车工业学院智能网联汽车学院硕士研究生,主要研究方向为计算机视觉和图像处理。

E_mail:mzjl1240504@ 163. com

Wang Wenshu received her B. Sc. degree from Yuncheng University in 2020. Now she is a M. Sc. candidate at School of Intelligent Connected Vehicle, Hubei University of Automotive Technology. Her main research interests include computer vision and image processing.