



计算机工程
Computer Engineering
ISSN 1000-3428, CN 31-1289/TP

《计算机工程》网络首发论文

题目: 基于 YOLOv8 的边缘端轻量级多尺度目标检测方法
作者: 郝金隆, 张祯, 李秀华, 曾胡爽, 黄河澎, 蔡春茂
DOI: 10.19678/j.issn.1000-3428.0252340
网络首发日期: 2025-08-28
引用格式: 郝金隆, 张祯, 李秀华, 曾胡爽, 黄河澎, 蔡春茂. 基于 YOLOv8 的边缘端轻量级多尺度目标检测方法[J/OL]. 计算机工程.
<https://doi.org/10.19678/j.issn.1000-3428.0252340>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

基于 YOLOv8 的边缘端轻量级多尺度目标检测方法

郝金隆, 张 祯, 李秀华*, 曾胡爽, 黄河澎, 蔡春茂

(重庆大学大数据与软件学院, 重庆 400044)

摘 要: 针对资源受限场景中多尺度目标检测模型难以兼顾高精度与低参数量、低计算复杂度的问题, 提出了一种基于 YOLOv8 的边缘端轻量级多尺度目标检测方法。首先, 在主干网络的跨阶段部分双特征融合模块中嵌入了自主设计的轻量级通道-空间注意力模块, 通过融合高效通道注意力机制和多尺度深度可分离低秩卷积, 在模块复杂度较低的同时实现通道和空间双维度特征增强。其次, 设计了跨层自适应加权融合模块, 建立跨层连接, 通过自适应加权机制融合浅层细节特征与深层语义信息。再次, 将颈部网络跨阶段部分双特征融合模块中的瓶颈结构替换为通用倒置瓶颈, 在维持检测精度的同时降低计算复杂度。最后, 提出了聚焦式尺度自适应动态交并比损失, 通过尺度自适应调制项与聚焦机制, 动态调整误差惩罚力度并强化对难检测目标的关注。在 BDD100K 数据集上, 与当前先进的 YOLO11m 模型相比, LMS-YOLO-m 在 mAP@50 与 mAP 分别提升了 0.5% 和 0.1%, 参数量减少了 2.4%, 计算量降低了 5.8%, 结果表明该方法在检测精度更高的同时具有更低的参数量和更低的计算复杂度。

关键词: 计算机视觉; 目标检测; 注意力机制优化; 多尺度特征融合; 模型轻量化; 损失函数优化

源代码链接: <https://github.com/Jayne1031/LMS-YOLO/tree/main>

DOI: 10.19678/j.issn.1000-3428.0252340

Lightweight Multi-Scale Object Detection Method for Edge Devices Based on YOLOv8

HAO Jinlong, ZHANG Zhen, LI Xiuhua*, ZENG Hushuang, HUANG Hepeng, CAI Chunmao

(School of Big Data & Software Engineering, Chongqing University, Chongqing 400044, China)

[Abstract] To address the challenge of balancing high accuracy with low parameter count and computational complexity in multi-scale object detection for resource-constrained edge scenarios, this paper proposes LMS-YOLO, a lightweight multi-scale detection method based on YOLOv8. The approach introduces four key innovations: a lightweight channel-spatial attention module integrated into the backbone's CSP blocks that combines efficient channel attention with multi-scale depthwise separable low-rank convolutions for dual-dimensional feature enhancement; a cross-layer adaptive weighted fusion module establishing skip connections to dynamically integrate shallow detail features with deep semantic information; replacement of standard bottlenecks with generalized inverted bottlenecks in the neck network to reduce computation while maintaining accuracy; and a novel focal scale-aware dynamic IoU loss that adaptively adjusts error penalties based on target scale and detection difficulty. Comprehensive evaluations on the BDD100K dataset demonstrate that LMS-YOLO-m achieves superior performance compared to YOLOv8m, with 0.5% and 0.1% improvements in mAP@50 and mAP respectively, while reducing parameters by 2.4% and computation by 5.8%, making it particularly suitable for deployment in resource-constrained edge computing environments where efficiency and accuracy must be carefully balanced.

[Key words] Computer vision; Object detection; Attention mechanism optimization; Multi-scale feature fusion; Model lightweight; Loss function optimization

0 引言

随着人工智能技术的飞速发展, 深度学习模型广泛部署于边缘设备, 用于环境实时感知与智能决策。多尺度目标检测作为环境感知关键环节, 在安防监控、无人

机巡检、智能家居及自动驾驶^[1,2]等场景至关重要。现有方法面临挑战: 高精度模型计算量大、参数多, 难以满足边缘设备实时需求; 轻量级模型在复杂多尺度场景检测精度不足。有限资源下实现高效准确的多尺度目标检

基金项目: 国家重点研发计划项目(2024YFC3014900); 国家自然科学基金项目(62372072); 重庆市技术创新与应用发展专项重大项目(CSTB2023TIAD-STX0049 和 CSTB2024TIAD-STX0024)

通信作者 E-mail: lixiuhua@cqu.edu.cn

测^[3]成为边缘计算领域亟待解决的问题。

现有方法在目标检测中取得进展但仍存挑战。特征增强方面,现有模型主要依赖通道注意力机制^[4,5],难以充分捕捉空间信息,采用通道-空间增强会增加模型参数与计算量,影响轻量化效果。特征融合方面,简单加权融合难以捕捉特征间复杂关系,融合模块增加计算开销。模块轻量化方面,剪枝和量化可能导致精度下降影响安全性。多尺度检测中,不同尺寸目标对定位误差敏感度差异显著:小尺度目标对定位误差敏感性高,大尺度目标容忍度高,传统交并比(IoU)损失及其变体^[6-8]采用统一误差度量标准,未充分考虑不同尺度目标对定位误差的敏感性差异。

针对以上问题,本文提出基于YOLOv8的边缘端轻量级多尺度目标检测方法。针对通道注意力机制难以捕捉空间信息且复杂度高的问题,设计轻量级通道-空间注意力模块,通过融合高效通道注意力机制和多尺度深度可分离低秩卷积,实现通道和空间双维度特征增强且模块复杂度低;针对简单加权融合难以建模复杂特征关系的问题,设计跨层自适应加权融合模块,通过建立主干与颈部网络的多层跨越连接和自适应加权机制,有效融合不同层级特征信息;针对剪枝等导致的精度下降问题,采用通用倒置瓶颈结构替代标准瓶颈,在保证轻量化的同时维持模型精度;针对传统交并比损失无法适应多尺度检测需求的问题,提出聚焦式尺度自适应动态交并比损失,通过尺度自适应调制项与聚焦机制,动态调整误差惩罚力度并强化对难检测目标的关注,实现对不同尺度目标的差异化优化。

1 国内外研究现状

目标检测是计算机视觉领域的核心任务之一,在过去的十年里,目标检测方法主要沿着三条技术路线向前演进,分别是两阶段检测、单阶段检测以及基于Transformer的检测。

1.1 两阶段检测模型

该类方法通过区域提议与目标识别分离实现高精度检测。2014年, Girshick 等人^[9]提出的基于区域的卷积神经网络(Region-based Convolutional Neural Network, R-CNN)首次将CNN特征提取器与选择性搜索(Selective Search)区域生成结合,但存在重复计算的缺陷。2015年 Girshick^[10]提出的快速区域卷积神经网络(Fast

R-CNN)通过引入感兴趣区域池化层(RoI Pooling)实现特征图共享,显著提升效率。同年 Ren 等人^[11]提出更快的区域卷积神经网络(Faster R-CNN),创新性地采用区域提议网络(Region Proposal Network, RPN),实现端到端候选框生成,成为两阶段检测的里程碑。此后,2017年 He 等人^[12]扩展 Faster R-CNN 框架,提出掩膜区域卷积神经网络(Mask R-CNN),新增掩码预测分支,推动多任务检测发展。

1.2 单阶段检测模型

单阶段模型通过回归直接输出检测结果,在实时性场景优势显著。2016年,Redmon 团队推出YOLOv1^[13],开创网格化预测范式,实现端到端检测。同年, Liu 等人^[14]提出单次多框检测器(Single Shot MultiBox Detector, SSD),通过多尺度特征图与默认框机制增强尺度适应性。2017年, Lin 等人^[15]提出RetinaNet,设计了焦点损失函数(Focal Loss)和特征金字塔网络(Feature Pyramid Network, FPN),解决了单阶段方法中样本不平衡的问题。2019年, Duan 等人^[16]提出CenterNet,进一步简化检测流程,通过中心点热力图回归实现了无锚框检测(Anchor-free)。2020年, Bochkovski 等人^[17]推出YOLOv4,结合跨阶段局部Darknet53网络(Cross Stage Partial Darknet53, CSPDarknet53)和马赛克数据增强(Mosaic),显著提升了模型的泛化能力。2023年, Ultralytics 团队发布YOLOv8,引入无锚框检测头和动态任务分配策略,减少了超参数依赖。2024年, Wang 等人^[18]提出YOLOv10,通过一致双标签分配策略消除非极大值抑制(Non-Maximum Suppression, NMS)依赖。同年, Khanam R 等人^[19]提出了YOLO11,创新融合C3k2模块、快速空间金字塔池化(Spatial Pyramid Pooling-Fast, SPPF)及并行空间注意力卷积块(Convolutional block with Parallel Spatial Attention, C2PSA),优化多尺度特征提取效率。

1.3 基于Transformer的检测模型

该类方法摒弃传统手工设计组件,利用自注意力机制实现全局建模。2020年, Carion 等人^[20]提出检测Transformer(Detection Transformer, DETR),首次将Transformer架构引入检测领域,通过可学习对象查询替代锚框机制与NMS。2021年, Zhu 等人^[21]提出可变形检测Transformer(Deformable DETR),设计可变形注意力机制,提升小目标检测效率。同年, Meng 等人^[22]提

出了条件检测 Transformer (Conditional detection Transformer, Condition-al DETR), 通过条件空间查询机制加速模型收敛。2022 年, Liu 等人^[23]提出了动态锚框检测 Transformer (Dynamic Anchor Boxes detection Transformer, Dab-DETR), 将目标框坐标作为动态查询强化位置感知。2024 年, Zhao 等人^[24]提出了实时端到端目标检测 Transformer (Real-Time Detection Transformer,

RT-DETR), 计了高效混合编码器 (Hybrid Encoder) 并支持解码器层数的灵活调整, 首次在 Transformer 架构上实现了高精度的实时端到端目标检测。

2 轻量级多尺度目标检测模型设计

2.1 网络框架

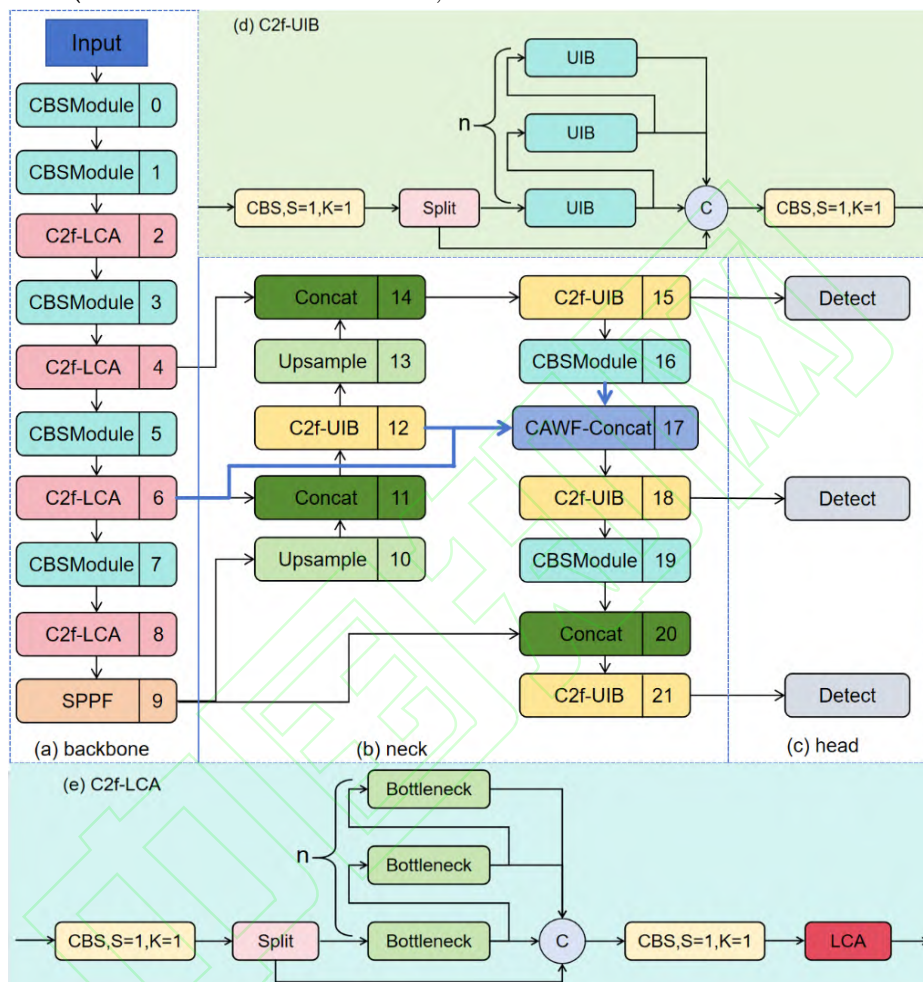


图 1 LMS-YOLO 模型总体框架图

Fig 1 Overall framework of the LMS-YOLO model

为满足边缘设备上多尺度目标检测模型高精度、低参数量、低计算复杂度的需求, 本文基于 YOLOv8 提出了一种轻量级多尺度目标检测方法 (Lightweight Multi-Scale YOLO, LMS-YOLO), 其结构如图 1 所示。该框架的核心创新点体现在以下四个方面: 首先, 在主干网络的跨阶段部分双特征融合网络 (Cross Stage Partial Network with 2 Fusion, C2f) 模块中嵌入了自主研发的轻量级通道-空间注意力模块 (Lightweight Channel-spatial Attention, LCA), 该模块通过高效通道注意力机制 (Efficient Channel

Attention, ECA)^[4]自适应校准通道权重, 结合多尺度深度可分离低秩卷积提取多层次空间特征, 利用双域注意力机制实现特征图的全面增强, 显著提升了特征表达能力; 其次, 借鉴双向特征金字塔网络 (Bidirectional Feature Pyramid Network, BiFPN)^[25]的思想, 设计了跨层自适应加权融合模块 (Cross-layer Adaptive Weighted Fusion Module, CAWF), 在主干网络与颈部网络的第 6、12、16 层之间建立跨层连接, 通过自适应加权机制融合浅层细节特征与深层语义信息, 在提升检测精度的同时保持了较高的计算效

率；再次，将颈部网络 C2f 模块中的瓶颈结构替换为 MobileNetV4^[26] 的通用倒置瓶颈（Universal Inverted Bottleneck, UIB），在显著降低计算复杂度的同时维持了较高的检测精度；最后，提出了聚焦式尺度自适应动态交并比损失（Focal Scale-Adaptive Dynamic IoU Loss, Focal-SADIoU Loss），通过尺度自适应调制项与聚焦机制的双重优化，动态调整误差的惩罚力度并进一步强化了对难检测目标的关注，实现了对不同尺度目标的差异化精准检测。

2.2 基于轻量级通道-空间注意力的增强方法

为增强特征提取能力，本文提出了 LCA 模块，其结

构如图 2 所示。该模块首先通过 ECA^[4] 自适应调整通道权重，以捕捉通道间的全局依赖关系；其次，通过多尺度深度可分离低秩卷积提取多层次空间特征，生成空间注意力权重；最后，将通道注意力权重与空间注意力权重逐元素相乘，实现特征图在通道和空间两个维度上的协同增强，提升特征图的表达能力。如图 1 (e) 所示，在 C2f 模块后面加入 LCA 模块，其中， n 表示模块的堆叠次数，设置为 3， K 表示卷积核尺寸，设置为 1， S 表示卷积步长设置为 1。

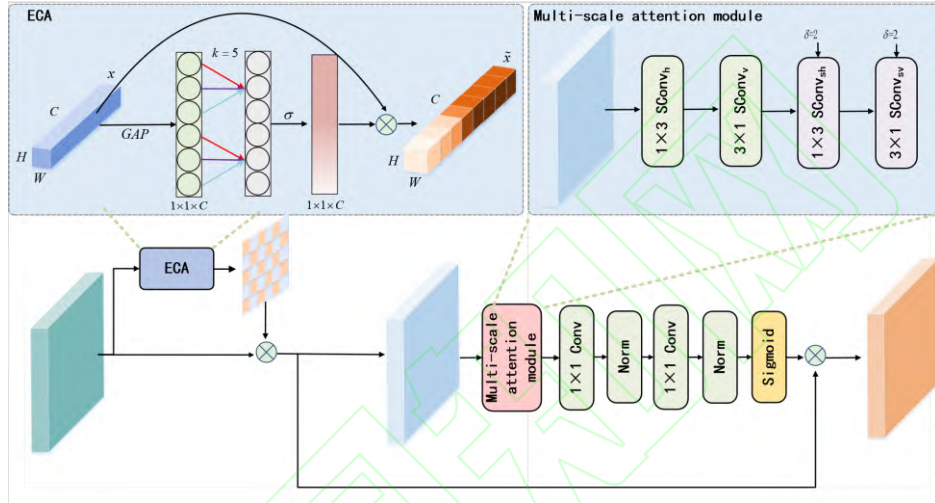


图 2 LCA 模块示意图

Fig 2 Schematic diagram of the LCA module

2.2.1 基于 ECA 的通道注意力机制

通道注意力可以理解对各通道特征分布进行加权分配的过程。它可以突出具有判别意义的特征通道，同时削弱冗余或无关特征的影响。具体步骤如下：

首先，对输入特征图 X 进行全局平均池化，生成每个通道的全局描述向量 z ：

$$Z = \text{AvgPool}(X) \in \mathbb{R}^{B \times C \times 1 \times 1} \quad (1)$$

其中，AvgPool 表示自适应平均池化操作，将每个通道维度 $H \times W$ 压缩为 1×1 ， z 表示全局平均池化后的特征图，形状为 $B \times C \times 1 \times 1$ 。

将 z 从 $B \times C \times 1 \times 1$ 转换为 $\mathbb{R}^{B \times 1 \times C}$ ，通过一维卷积捕捉通道间关系，生成通道注意力权重 A_c ：

$$A_c = \sigma(\text{Conv}(Z)) \in \mathbb{R}^{B \times C \times 1 \times 1} \quad (2)$$

其中，Conv 表示一维卷积操作， $\sigma(\cdot)$ 表示 sigmoid 激活函数， A_c 表示通道注意力权重，形状为 $B \times C \times 1 \times 1$ 。

将生成的通道注意力权重 A_c 与原始输入特征图 X 逐元素相乘，得到加权后的特征图：

$$X_c = X \odot A_c \in \mathbb{R}^{B \times C \times H \times W} \quad (3)$$

其中， \odot 表示逐元素相乘操作， X_c 表示通道加权后的特征图，形状与输入 X 相同。

2.2.2 基于多尺度深度可分离低秩卷积的空间注意力机制

LCA 模块中的空间注意力机制采用多尺度深度可分离低秩卷积增强特征表达。其核心是通过空间位置重要性加权优化特征图，具体如下：

(1) 空间方向特征解耦

使用深度可分离卷积^[27]分解空间维度，分别提取水平与垂直方向特征：

$$A_h = \text{SConv}_h(X_c) \quad (4)$$

$$A_v = \text{SConv}_v(X_h) \quad (5)$$

其中， SConv_h 和 SConv_v 分别表示水平方向和垂直方向的深度可分离低秩卷积操作， $A_h, A_v \in \mathbb{R}^{B \times C \times H' \times W'}$ 表示水平方向和垂直方向的深度可分离低秩卷积输出。深度可分离卷积分为两步：深度卷积和逐点卷积（ 1×1 卷积）。

深度卷积对每个输入通道独立进行卷积操作，减少了通道之间的耦合度，从而降低了计算量；随后，逐点卷积在通道维度上进行信息融合，进一步优化特征提取。

(2) 低秩近似压缩计算量

逐点卷积采用低秩近似^[28]进一步优化，将权重矩阵 $W_{pw} \in \mathbb{R}^{C \times C \times 1 \times 1}$ 分解为两个低秩矩阵的乘积：

$$W_{pw} \approx U_r V_r^T \quad (6)$$

其中， $U_r \in \mathbb{R}^{C \times r}$ 和 $V_r^T \in \mathbb{R}^{r \times C}$ 表示低秩近似矩阵， $r \ll C$ ，表示低秩近似的秩。低秩分解将原本 $C \times C$ 的参数数量降至 $C \times r + r \times C$ ，显著减少了计算开销。

(3) 多尺度感受野扩展

使用带有膨胀因子的深度可分离低秩卷积进一步提取多尺度特征：

$$A_{sh} = \text{SConv}_{sh}(A_v) \quad (7)$$

$$A_{sv} = \text{SConv}_{sv}(A_{sh}) \quad (8)$$

其中， SConv_{sh} 和 SConv_{sv} 分别表示带有膨胀因子^[29]的水平方向和垂直方向的深度可分离低秩卷积操作。膨胀因子通过扩大卷积核间距获取多尺度上下文，在保持分辨率的同时捕捉远距离依赖关系。 A_{sh} ,

$A_{sv} \in \mathbb{R}^{B \times C \times H' \times W'}$ 表示带有膨胀因子的水平方向和垂直方向的深度可分离低秩卷积输出。

(4) 空间注意力权重生成

通过低秩逐点卷积进一步处理多尺度卷积输出，并结合批归一化和激活函数，生成空间注意力权重 $A_{spatial}$ ：

$$A_{spatial} = \sigma \left(\text{BN}_2 \left(\text{Conv}_2 \left(\text{ReLU} \left(\text{BN}_1 \left(\text{Conv}_1(A_{sv}) \right) \right) \right) \right) \right) \quad (9)$$

其中， Conv_1 和 Conv_2 表示 1×1 低秩逐点卷积，分别用于通道降维和升维， BN_1 和 BN_2 表示归一化层， ReLU 表示激活函数， $\sigma(\cdot)$ 表示 *sigmoid* 激活函数， $A_{spatial}$ 表示最终的空间注意力权重，形状为 $B \times C \times H \times W$ 。

2.2.3 融合通道与空间的综合注意力机制

在 LCA 模块中，通过融合通道注意力和空间注意力机制，实现了对特征图的深度增强。这一过程可表示为以下公式：

$$Y = X_s = X_c \odot A_{spatial} \quad (10)$$

其中， $A_{spatial}$ 表示空间注意力权重，形状为 $B \times C \times H \times W$ ， Y 表示 LCA 模块的最终输出特征，形状为 $B \times C \times H \times W$ 。

2.3 基于跨层自适应加权的特征融合方法

为了增强特征融合的能力，本方法将 BiFPN 思想引

入 LMS-YOLO 中，设计出 CAWF。BiFPN 的核心思想在于通过高效的双向跨尺度连接与加权特征融合，实现对多尺度特征的充分整合，CAWF 的示意图如图 3 所示。

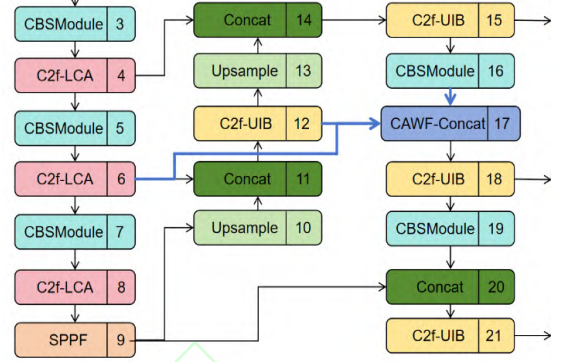


图 3 CAWF 示意图

Fig 3 Schematic diagram of the CAWF

第 6 层的特征图通常包含细节信息，但语义信息较少，通常关注较小的目标。第 12 层的特征图包含中等大小目标的特征信息，关注中等大小的目标。第 16 层的特征图包含较强的语义信息，主要关注大目标。在第 6 层、12 层和 16 层之间加入跳跃连接，使得这些不同尺度的特征能够直接进行信息传递，将低层的细节信息与高层的语义信息相结合，避免高层特征的过度抽象导致细节丢失，且使用自适应进行加权融合，根据每个尺度特征的重要性赋予不同的权重。

CAWF 中使用的 CAWF-Concat 是一种自适应加权的特征融合模块，该模块通过可学习的归一化权重对三个输入特征进行自适应融合，并沿指定维度拼接。该模块数学描述如下：

权重归一化：

$$w_i' = \frac{w_i}{\sum_{j=1}^3 w_j + \epsilon} \quad (11)$$

加权特征融合：

$$\hat{x}_i = w_i' \cdot x_i \quad (12)$$

维度拼接：

$$\text{Output} = \text{Concat}_d(\hat{x}_1, \hat{x}_2, \hat{x}_3) \quad (13)$$

其中， x_i 表示第 i 个输入张量， w_i 表示第 i 个可学习权重标量， ϵ 表示数值稳定性常数， w_i' 表示归一化权重， \hat{x}_i 表示加权后特征张量， Concat_d 表示沿维度 d 的拼接操作。

2.4 基于倒置瓶颈结构的模块轻量化方法

为实现轻量化设计，本方法将 LMS-YOLO 颈部网络的 C2f 模块中的瓶颈结构替换为 MobileNetV4^[26]的 UIB

模块,如图 1(d)示意。这一替换在有效降低计算复杂度的同时,仍能维持较高的检测精度。

UIB 通过融合深度可分离卷积与反向瓶颈结构实现效率与性能的平衡。其设计包含四部分:起始深度卷积仅对输入进行空间特征混合,保持通道数不变,以低成本抑制噪声、扩大感受野;扩展卷积采用 1×1 点卷积将通道数扩展至输入的 4 到 6 倍(并取整至 8 倍数),在高维空间增强特征表达能力;中间深度卷积对扩展后各通道独立进行卷积,深度挖掘空间特征;投影卷积再通过 1×1 点卷积压缩通道至目标维度,同时保留前置跳跃连接以缓解梯度消失,在降低后续计算量的同时保留关键信息。

2.5 基于尺度自适应的损失优化方法

针对传统传统的交并比损失函数在多尺度目标检测的局限性,本方法提出了两项创新改进:尺度自适应动态交并比损失(Scale-Adaptive Dynamic IoU Loss, SADIoU Loss)和 Focal-SADIoU Loss。

2.5.1 尺度自适应动态交并比损失

SADIoU Loss 引入了尺度自适应调制项 $\alpha(S)$,动态调整宽高误差权重,实现了对不同尺度目标的差异化优化,进一步提升模型在多尺度目标检测中的性能。具体而言:当目标尺度较小(S 较小)时, $\alpha(S)$ 较大,对宽高误差的惩罚力度增强,促使模型更加精确地拟合小目标的边界。当目标尺度较大(S 较大)时, $\alpha(S)$ 较小,降低对宽高误差的惩罚力度,避免过度关注无关紧要的微小偏差。SADIoU Loss 综合考虑了重叠度、中心点距离、宽高比例差异以及目标尺度因素,其数学表达式如下:

$$L_{\text{SADIoU}} = 1 - \text{IoU} + \frac{\rho^2(b, b^{gt})}{(w^c)^2 + (h^c)^2} + \alpha(S) \left(\frac{\rho^2(w, w^{gt})}{(w^c)^2} + \frac{\rho^2(h, h^{gt})}{(h^c)^2} \right) \quad (14)$$

$$\text{IoU} = \frac{\text{Area}(B \cap B^{gt})}{\text{Area}(B \cup B^{gt})} \quad (15)$$

其中, IoU 表示预测框 B 与真实框 B^{gt} 的交并比, $\rho^2(b, b^{gt}) = (b_x - b_x^{gt})^2 + (b_y - b_y^{gt})^2$ 表示预测框与真实框中心点的欧氏距离平方, $w^c = \min(w, w^{gt})$ 和 $h^c = \min(h, h^{gt})$ 分别表示预测框与真实框在宽度和高度

上的最小值, $\rho^2(w, w^{gt}) = (w - w^{gt})^2$ 表示预测框与真实框在宽度上的差异平方, $\rho^2(h, h^{gt}) = (h - h^{gt})^2$ 表示预测框与真实框在高度上的差异平方。 $\alpha(S)$ 为尺度自适应调制项,定义为:

$$\alpha(S) = \frac{1}{1 + e^{\lambda(S - S_0)}} \quad (16)$$

其中, $S = \sqrt{w^{gt} h^{gt}}$ 表示目标物体的尺度, λ 为自适应的调节参数, S_0 为尺度阈值。

2.5.2 聚焦式尺度自适应动态交并比损失

为了更好地平衡处理难例和易例,本文结合了 SADIoU Loss 和 Focal 机制^[15],提出了 Focal-SADIoU Loss。该损失函数通过引入 IoU^γ ,为那些低 IoU 值的预测框(即难例)赋予了更高的权重,使得模型在训练过程中能够更加关注难以检测的目标,从而提高其定位的精确度。通过调整 γ 参数,模型可以在易例和难例之间实现动态平衡,避免对易检测目标的过度拟合。具体的定义如下所示:

$$L_{\text{Focal-SADIoU}} = \text{IoU}^\gamma \cdot L_{\text{SADIoU}} \quad (17)$$

其中, γ 是调节参数,用于控制 Focal 机制的作用强度, L_{SADIoU} 表示尺度自适应动态交并比损失。

3 实验结果与分析

3.1 实验配置

实验环境包括运行 Ubuntu20.04 操作系统的服务器,双卡 RTX NVIDIA 4090 GPU,总显存为 48GB,并使用 PyTorch1.9.0 搭建深度学习环境。如表 1 所示,实验选择了两个权威的公开道路检测数据集 Cityscapes 和 BDD100K。对于数据集,训练批量大小设置为 64,图像尺寸为 640×640 像素。使用随机梯度下降(SGD)优化器,初始学习率为 0.01,动量为 0.937,权重衰减设置为 0.0005。训练轮次大小为 250。训练时,使用数据增强手段对数据进行预处理,左右翻转概率设置为 0.5,马赛克增强设置为 1,缩放设置为 0.5。

表 1 Cityscapes 和 BDD100K 数据集的对比表

Table 1 Comparison table of the Cityscapes and BDD100K datasets

Dataset	Quantity	Image Resolution	Publication Year
Cityscapes	5,000	2048 × 1024	2016
BDD100K	100,000	1280 × 720	2020

3.2 性能分析

为了全面验证所提出的检测算法的有效性,进行了深入的对比分析,涵盖了现有的多种先进检测方法。在对比实验中,将 LMS-YOLO 模型与一系列经典的目标检测模型进行了比较,包括 Faster R-CNN^[11]、SSD^[14]、EfficientDet^[25]等。此外,还与 YOLO 系列的最新模型,包括 YOLOv8、YOLOv10^[18]和 YOLO11^[19]以及基于 Transformer 的先进方法,包括 DETR^[20]、Deformable-DETR^[21]、Conditional-DETR^[22]、DAB-DETR^[23]和 RT-DETR^[24]在相同的参数条件下进行了对比。实验选取了平均精度均值(mean Average Precision, mAP)作为检测精度指标,同时引入参数量(Parameter)、浮点数运算次数(Floating Point Operations, FLOPs)作为性能的辅助评判指标。

3.2.1 Cityscapes 数据集实验结果

表 2 展示了在 Cityscape 数据集上的结果,LMS-YOLO-n 在 mAP@50 和 mAP 这两个关键指标上均取得了优异的成绩,分别达到了 48.5%和 27.6%。与基线算法 YOLOv8 相比,在精度上提升了 5.1%,同时在参数量上减少了 21.3%,在计算量上减少了 16.8%。这表明 LMS-YOLO 在提高检测精度的同时,也实现了模型的轻量化和计算效率的提升。

表 2 在 Cityscapes 数据集上性能比较

Table 2 Performance comparison on the Cityscapes dataset

Method	mAP@50	mAP	Params(M)	FLOPs(G)
Faster R-CNN-r50	43.0	21.6	41	201
SSD-VGG16	39.8	19.3	26.8	31
EfficientDet-D0	36.2	21.2	3.9	2.5
YOLOv5n	42.0	24.7	2.6	7.7
YOLOv8n	43.4	24.7	3.2	8.9
YOLOv10n	45.1	26.9	2.3	6.7
YOLO11n	47.2	27.4	2.6	6.4
DETR	42.3	23.9	28.8	50.0
RT-DETR-r18	46.1	26.7	19.9	57.0
LMS-YOLO-n	48.5	27.6	2.5	7.4

3.2.2 BDD100K 数据集实验结果

表 3 展示了 LMS-YOLO 在 BDD100K 数据集上的性能比较。实验结果表明,LMS-YOLO-m 在 mAP@50 与 mAP 两个关键指标上均达到最优水平,分别为 53.5%和 30.0%。与基线算法 YOLOv8m 相比,LMS-YOLO-m 在 mAP@50 与 mAP 指标上分别提升 2.8%和 0.9%,同时参

数量减少 22.1%,计算量降低了 17.7%。与当前最先进的 YOLO11m 模型相比,LMS-YOLO-m 在 mAP@50 与 mAP 分别提升了 0.5%和 0.1%,参数量减少了 2.4%,计算量降低了 5.8%。值得注意的是,与基于 Transformer 架构的 RT-DETR-r50 相比,LMS-YOLO-m 在 mAP@50 与 mAP 指标上均提升了 1%,同时参数量大幅减少 52.1%,计算量降低 50%。此外,LMS-YOLO-n 虽然检测精度略低于其他模型,但参数量和计算量显著低于其他模型,展现了优异的性能。图 4 气泡图展示了在 BDD100K 数据集上 LMS-YOLO 与其他方法进行对比,展现出显著的优势。

表 3 在 BDD100K 数据集上性能比较

Table 3 Performance Comparison on the BDD100K Dataset

Method	mAP@50	mAP	Params(M)	FLOPs(G)
RT-DETR-r18	50.0	27.6	19.9	57.0
RT-DETR-r34	51.1	28.0	30.0	87.2
RT-DETR-r50	52.5	29.0	42.0	129.6
DETR	39.4	18.0	28.8	50.0
Dab-DETR	44.4	21.3	43.7	102
Conditional-DETR	48.6	21.6	43.5	101
Deformable-DETR	52.0	27.6	40.1	193
YOLOv5m	47.9	26.5	20.1	48.0
YOLOv6s	46.7	25.9	17.2	21.89
YOLOv7	49.5	26.2	36.9	104.7
YOLOv8m	50.7	29.1	25.8	78.7
YOLO11m	53.0	29.9	20.6	68.7
LMS-YOLO-n	45.6	24.7	2.5	7.4
LMS-YOLO-m	53.5	30.0	20.1	64.7

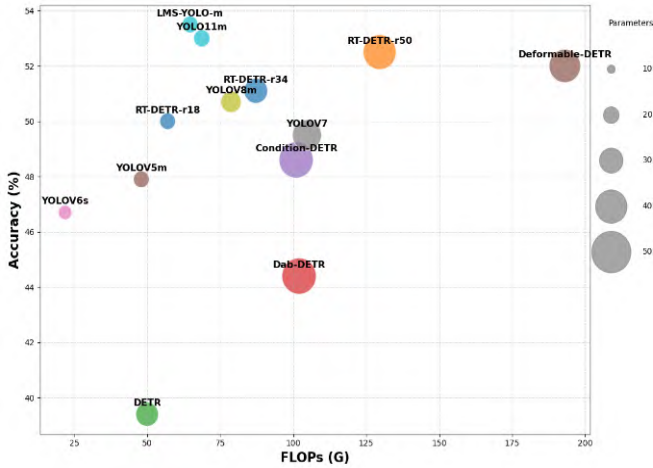


图 4 BDD100K 数据集上与其他先进方法对比图

Fig 4 Performance comparison with other advanced methods on the BDD100K dataset

3.3 消融实验

3.3.1 模块有效性消融实验

如表 4 所示, 在 Cityscapes 数据集上对 C2f-LCA、C2f-UIB 以及 CAWF 模块的有效性进行了详细的消融实验验证。为了确保实验的公平性, 所有实验均采用

Focal-SADIoU 作为损失函数。实验以 YOLOv8n 模型未添加任何模块时的实验结果作为基准, 从检测精度、参数量和计算量三个维度对模块的有效性进行分析。

表 4 在 Cityscapes 数据集上各个模块有效性消融实验

Table 4 Ablation study on the effectiveness of each module on the Cityscapes dataset

C2f-LCA	C2f-UIB	CAWF	mAP@50	mAP	Params(M)	FLOPs(G)
			46.5	26.0	3.1	8.9
√			47.4	26.6	3.1	9.0
	√		45.6	25.0	2.1	6.9
		√	47.9	26.7	3.1	8.9
√	√		47.2	26.4	2.5	7.4
√		√	48.7	27.8	3.1	9.0
	√	√	47.4	26.3	2.1	6.9
√	√	√	48.5	27.6	2.5	7.4

注: “√”表示添加该模块。

实验结果表明, 将原始主干网络中的 C2f 模块替换为 C2f-LCA 模块后, mAP@50 和 mAP 分别提升 0.9% 和 0.6%, 同时模型参数量未出现明显增加, FLOPs 仅增加 0.1G。这一结果证明了 C2f-LCA 模块在保持轻量化特性的同时, 能够有效提升模型性能。图 5 展示了 LMS-YOLO 第 4 阶段的 C2f-LCA 与 YOLOv8 第 4 阶段的 C2f 和 YOLO11 第 4 阶段的 C3K2 可视化对比, 可以看出 C2f-LCA 的特征图注意力更集中在目标。将原始颈部网络中的 C2f 模块替换为 C2f-UIB 模块后, 虽然检测精度略有下降, 但模型参数量和计算量均显著减少。考虑到

C2f-UIB 模块的设计初衷是实现模型轻量化, 检测精度的适度下降是可接受的。此外, 加入 CAWF 模块后, 模型参数量和计算量基本保持不变, mAP@50 和 mAP 分别提升 1.4% 和 0.7%。图 6 展示了 LMS-YOLO 第 17 层的 CAWF-Concat 与 YOLO11 对应位置的 Concat 特征可视化对比。可以看出原始 Concat 出现漏检现象, CAWF-Concat 检测更加全面。当这些模块全部被整合到模型中时, 模型的 mAP@50 和 mAP 分别达到了 48.5% 和 27.6%, 相比基准模型分别提升了 2.0% 和 1.6%, 且模型的参数量减少了 19.3%, 计算量降低了 16.8%。

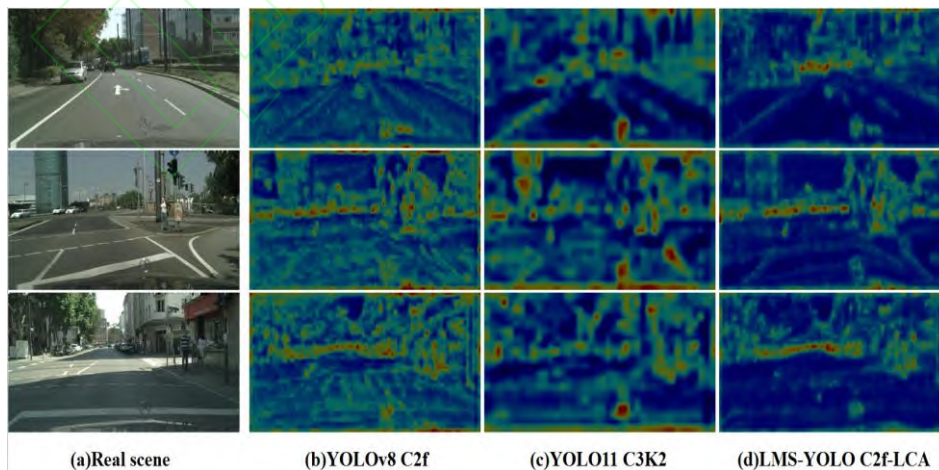


图 5 C2f-LCA 与对应位置 C2f 和 C3K2 可视化对比

Fig.5 Visualization comparison of C2f-LCA with corresponding C2f and C3K2 modules

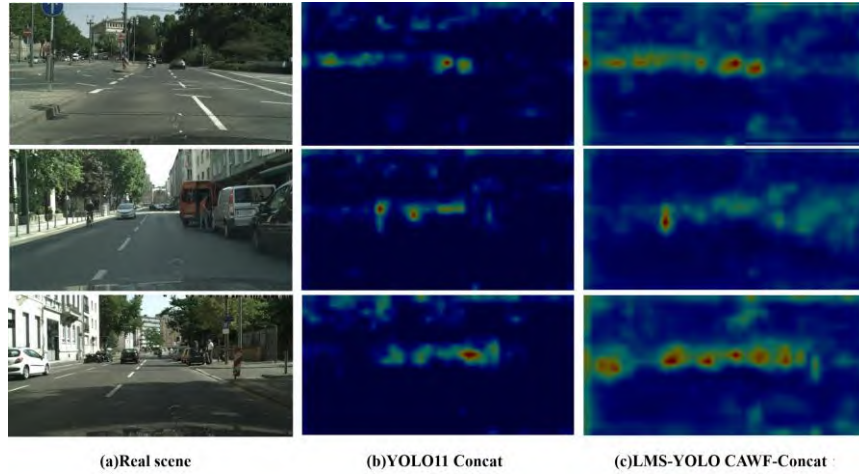


图 6 CAWF-Concat 与 YOLO11 对应的 Concat 特征可视化对比

Fig 6 Visual comparison of the Concat features in CAWF-Concat and YOLO11

在 Cityscapes 数据集评估中, 面积小于 32×32 像素的为小目标, 32×32 至 96×96 像素为中目标, 大于 96×96 像素为大目标。评估采用 APs、APm、API 作为核心指标, 分别衡量小、中、大三个尺度的检测精度。实验设置三组对照: YOLOv8n 作为基线, 固定权重 CAWF (权重强制均等), 以及本文提出的自适应权重 CAWF。表 5 展现了自适应加权融合的优势, 相比固定权重策略, 自适应 CAWF 在小、中、大目标检测上均取得显著突破。

表 5 分尺度精度统计表

Table 5 Object Detection Accuracy Statistics by Scale

Model	APs	APm	API	mAP@50	mAP
YOLOv8n	6.1	25.3	42.5	43.4	24.7
+CAWF (Fixed Weights)	8.5	25.8	42.6	46.2	25.7
+CAWF(Adaptive)	10.9	29.1	43.6	47.9	26.7

3.3.2 损失函数有效性消融实验

在 Cityscapes 数据集上对 Focal-SADIoU Loss 的有效性进行验证, 实验保留了 C2f-LCA 模块、C2f-UIB 模

块和 CAWF 模块, 将 YOLOv8 所使用的默认检测损失作为基准。如表 6 所示, 引入 Focal-SADIoU 损失后, 模型的两个精度指标均得到了显著提升。

表 6 在 Cityscapes 数据集上损失函数有效性消融实验

Table 6 Ablation study on the effectiveness of loss functions on the Cityscapes dataset

Focal-SADIoU	mAP@50	mAP
	47.4	26.3
✓	48.5	27.6

图 7 展示了损失函数的对比示意图, 蓝色曲线表示原始损失函数, 红色曲线表示 Focal-SADIoU 损失函数。具体来看, 在训练和验证阶段的边界框回归损失上, 使用 Focal-SADIoU 损失函数明显下降更快、更稳定, 而在类别损失与分布焦点损失两种损失函数的表现差别不大。在评价指标方面, 平均精度 mAP50 与 mAP50-95 指标均显示 Focal-SADIoU 损失函数在收敛速度、检测精度和平稳性方面都优于原始损失函数。

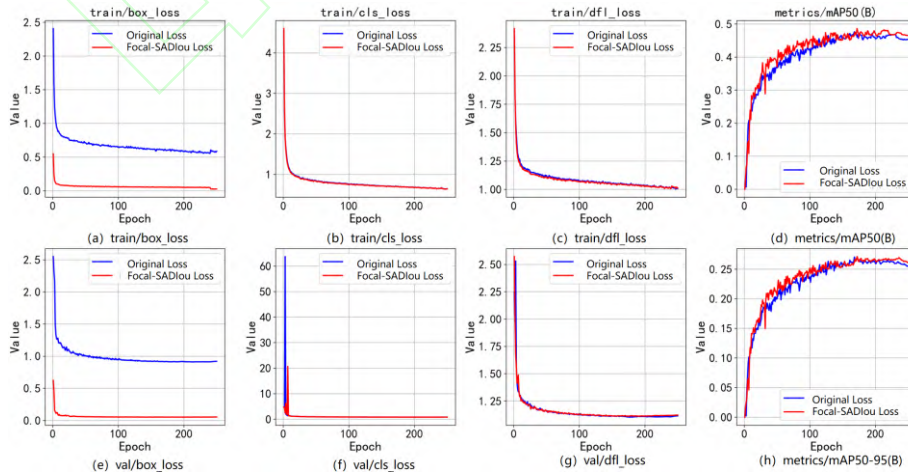


图 7 损失函数对比示意图

Fig.7 Loss function comparison schematic

4 结束语

本文提出基于 YOLOv8 的轻量级多尺度检测框架 LMS-YOLO, 采用轻量通道-空间注意力、跨层自适应加权融合及通用倒置瓶颈, 有效降低计算复杂度并提升检测精度。同时, 通过聚焦式尺度自适应动态交并比损失, 针对不同尺度目标实现精准定位。提出的 LMS-YOLO 框架主要基于视觉传感器数据, 边缘设备通常需要融合多模态数据, 例如激光雷达和毫米波雷达, 以实现更全面的感知。因此, 未来可深入研究基于 Transformer 架构的多模态特征融合方案, 提升多源信息融合效率与准确性。

参考文献

- [1] 中国汽车工程学会. 节能与新能源汽车技术路线图 2.0[M]. 2.机械工业出版社: 539.
China Society of Automotive Engineers. Technology Roadmap for Energy-Saving and New Energy Vehicles 2.0 [M]. 2nd ed. Beijing: Mechanical Industry Press, p. 539.
- [2] 郭延永, 刘佩, 袁泉, 等. 网联自动驾驶车辆道路交通安全研究综述[J]. 交通运输工程学报, 2023, 23(5): 19-38.
Guo Yanyong, Liu Pei, Yuan Quan, et al. A Review of Road Traffic Safety Research for Connected and Automated Vehicles [J]. Journal of Traffic and Transportation Engineering, 2023, 23(5): 19-38.
- [3] 张阳婷, 黄德启, 王东伟, 等. 基于深度学习的目标检测算法研究与应用综述[J]. 计算机工程与应用, 2023, 59(18).
Zhang Yangting, Huang Deqi, Wang Dongwei, et al. Review of Research and Applications of Deep Learning-Based Object Detection Algorithms [J]. Computer Engineering and Applications, 2023, 59(18).
- [4] 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(CVPR). 2020: 11534-11542.
- [5] Qin Z, Zhang P, Wu F, et al. Fcanet: Frequency channel attention networks[C], Proceedings of the IEEE/CVF international conference on computer vision(ICCV). 2021: 783-792.
- [6] Zheng Z, Wang P, Liu W, et al. Distance-IoU loss: Faster and better learning for bounding box regression[C], Proceedings of the AAAI conference on artificial intelligence(AAAI). 2020: 12993-13000.
- [7] Yu J, Jiang Y, Wang Z, et al. Unitbox: An advanced object detection network[C], Proceedings of the 24th ACM international conference on Multimedia(MM). 2016: 516-520.
- [8] Zhang Y-F, Ren W, Zhang Z, et al. Focal and efficient IOU loss for accurate bounding box regression[J]. Neurocomputing, 2022, 506: 146-157.
- [9] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C], Proceedings of the IEEE conference on computer vision and pattern recognition(CVPR). 2014: 580-587.
- [10] Girshick R. Fast r-cnn[C], Proceedings of the IEEE international conference on computer vision(ICCV). 2015: 1440-1448.
- [11] Ren S, He K, 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.
- [12] He K, Gkioxari G, Dollár P, et al. Mask r-cnn[C], Proceedings of the IEEE international conference on computer vision(ICCV). 2017: 2961-2969.
- [13] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection[C], Computer Vision & Pattern Recognition(CVPR). 2016: 779-788.
- [14] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C], Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. Springer, 2016: 21-37.
- [15] Lin T-Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C], Proceedings of the IEEE international conference on computer vision(ICCV). 2017: 2980-2988.
- [16] Duan K, Bai S, Xie L, et al. Centernet: Keypoint triplets for object detection[C], Proceedings of the IEEE/CVF international conference on computer vision(ICCV). 2019: 6569-6578.
- [17] Bochkovskiy A, Wang C-Y, Liao H-Y M. YOLOv4: Optimal speed and accuracy of object detection[EB/OL]. https:

- [//arxiv.org/abs/2004.10934](https://arxiv.org/abs/2004.10934), 2020-04-23.
- vision (ECCV). 2018: 116-131.
- [18] Wang A, Chen H, Liu L, et al. YOLOv10: Real-time end-to-end object detection[EB/OL]. <https://arxiv.org/abs/2405.14458>, 2024-10-30.
- [19] Khanam R, Hussain M. YOLO11: An overview of the key architectural enhancements[EB/OL]. <https://arxiv.org/abs/2410.17725>, 2024-10-23.
- [20] Carion N, Massa F, Synnaeve G, et al. End-to-end object detection with transformers[C], European conference on computer vision(ECCV). Springer, 2020: 213-229.
- [21] Zhu X, Su W, Lu L, et al. Deformable detr: Deformable transformers for end-to-end object detection[EB/OL]. <https://arxiv.org/abs/2010.04159>, 2021-03-18.
- [22] Meng D, Chen X, Fan Z, et al. Conditional detr for fast training convergence[C], Proceedings of the IEEE/CVF international conference on computer vision(ICCV). 2021: 3651-3660.
- [23] Liu S, Li F, Zhang H, et al. Dab-detr: Dynamic anchor boxes are better queries for detr[EB/OL].<https://arxiv.org/abs/2201.12329>, 2022-03-30.
- [24] Zhao Y, Lv W, Xu S, et al. Detsr beat yolos on real-time object detection[C], Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2024: 16965-16974.
- [25] Tan M, Pang R, Le Q V. Efficientdet: Scalable and efficient object detection[C], Proceedings of the IEEE/CVF conference on computer vision and pattern recognition(CVPR). 2020: 10781-10790.
- [26] Qin D, Lechner C, Delakis M, et al. MobileNetV4: universal models for the mobile ecosystem[C], European Conference on Computer Vision(ECCV). Springer, 2024: 78-96.
- [27] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions[EB/OL]. <https://arxiv.org/abs/1511.07122>, 2016-04-30.
- [28] Chollet F. Xception: Deep learning with depthwise separable convolutions[C], Proceedings of the IEEE conference on computer vision and pattern recognition(CVPR). 2017: 1251-1258.
- [29] Ma N, Zhang X, Zheng H-T, et al. Shufflenet v2: Practical guidelines for efficient cnn architecture design[C], Proceedings of the European conference on computer