

多维度协作注意力机制与加权特征融合的 零件检测方法*

方雁雁¹ 林志凌¹ 李宇宇¹ 张志俭² 李旦¹

(1. 复旦大学未来信息创新学院 上海 200433; 2. 南京大学电子科学与工程学院 南京 210023)

摘要: 针对复杂堆叠零件因遮挡导致的检测精度下降问题,提出了多维度协作注意力(MCA)机制与加权特征融合改进型 YOLO11 目标检测算法 YOLO11-MBM。该算法通过三重协同优化提升遮挡鲁棒性:首先,在骨干网络的 C3k2 模块中引入多维度协作注意力机制,融合通道、空间与尺度注意力,增强纹理相似零件的区分能力,有效抑制纹理混淆性误检;其次,在特征融合网络中设计加权特征融合架构 BiFPN-SFF-Concat,通过双向特征交互与尺度敏感特征动态筛选加权策略,强化遮挡场景下的特征互补与细节表征,降低遮挡性漏检;最后,采用最小点距离交并比(MPDIoU)作为损失函数,提升定位精度与收敛速度。采用百度飞桨工业零件扩展数据集验证表明,YOLO11-MBM 在保持实时推理速度(V100 GPU 达 110 fps)与轻量化设计(参数量仅增 1.1×10^6)的同时,检测精度 mAP50 达到了 87.8%,较基准提升了 6%;mAP50:95 达到了 58.0%,较基准提升了 9%。研究结果验证了其在复杂工业检测场景中的实用价值,为智能制造的视觉检测系统提供了可行的技术方案。

关键词: 目标检测;多维度协作注意力;加权特征融合

中图分类号: TN911.73 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Multi-dimensional collaborative attention and weighted feature fusion for part detection

Fang Yanyan¹ Lin Zhiling¹ Li Ningyu¹ Zhang Zhijian² Li Dan¹

(1. College of Future Information Technology, Fudan University, Shanghai 200433, China;

2. School of Electronic Science and Engineering, Nanjing University, Nanjing 210023, China)

Abstract: To address the accuracy degradation in detecting occluded and complex stacked parts, an improved YOLO11-based object detection algorithm named YOLO11-MBM is proposed. The method integrates a multidimensional collaborative attention (MCA) mechanism and employs weighted feature fusion. The main strategies include three aspects: Firstly, an MCA module is incorporated into the C3k2 module of the backbone network to fuse channel, spatial and scale attention, which enhances the ability to distinguish texture-similar parts and effectively suppresses false positives caused by texture confusion. Secondly, a weighted feature fusion architecture BiFPN-SFF-Concat is designed in the feature fusion network. Through bidirectional feature propagation and a scale-sensitive dynamic feature weighting strategy, it improves feature complementarity and detail representation in occluded scenarios, reducing missed detections. Finally, the minimum point distance intersection over union (MPDIoU) is adopted as the loss function to improve localization accuracy and convergence speed. Experimental results on the extended Baidu Paddle industrial parts dataset show that YOLO11-MBM achieves an mAP50 of 87.8% and an mAP50:95 of 58.0%, outperforming the baseline model by 6% and 9% respectively, while maintaining real-time performance (110 fps on V100 GPU) and lightweight design (with only a 1.1×10^6 parameter increase). The experimental results confirm its practicality in complex industrial inspection scenarios and furnish a feasible technical solution for visual inspection systems in smart manufacturing.

Keywords: object detection; multidimensional collaborative attention; weighted feature fusion

0 引言

工业智能化升级依赖高精度视觉分拣,然而堆叠场景中

零件遮挡引发的特征提取困难、尺度差异和纹理混淆等问题严重制约了检测性能。早期 AlexNet^[1]虽开创了深度学习目标检测先河,但缺乏多尺度特征建模能力。区域卷积神经网络

络(region-based convolutional neural network, R-CNN)^[2]系列通过候选区域提升了定位精度,但因 ROI Pooling 的量化误差导致小目标定位不准。基于 Transformer 架构的 DETR (detection transformer, DETR)^[3]、RT-DETR^[4] 及 MFSF-DETR^[5]虽具有全局建模能力,但计算开销大,实时性较差,且面临轻量化部署挑战^[6]。一些改进型 YOLOv5^[7-9]虽在特定场景中取得良好效果,但对复杂遮挡的特征保持仍具挑战,近期针对金属表面缺陷的 GMS2-YOLO^[10]结构优化模型也显示出类似局限。而引入 C2f 与坐标注意力(coordinate attention, CA)的 YOLOv7^[11]方案虽然提升了钢板缺陷识别精度,但在高度遮挡和复杂纹理下泛化能力仍有限。YOLO11^[12]在常规场景中表现优异,但其 C3k2 和 C2PSA 模块对遮挡与尺度变化的适应性仍存在局限。

为此,本文提出 YOLO11-MBM 算法,通过三重协同优化构建适应复杂堆叠场景的检测框架:首先,在骨干网络的 C3k2 模块中嵌入多维度协作注意力(multidimensional collaborative attention, MCA)机制,实现跨维度特征校准,提升对纹理相似零件的辨识能力。其次,设计加权特征融合架构 BiFPN-SFF-Concat(bidirectional feature pyramid

network-selective feature fusion-concat),实现多尺度特征动态筛选与加权融合,缓解遮挡导致的语义断层问题。最后,采用最小点距离交并比(minimum point distance intersection over union, MPDIoU)损失函数,优化边界框回归过程,抑制定位漂移。综合以上改进,YOLO11-MBM 在遮挡鲁棒性、多尺度适应性和实时检测效率之间实现更优平衡,为复杂堆叠零件的高精度实时检测提供解决方案。

1 YOLO11-MBM 算法

1.1 YOLO11-MBM 网络结构

YOLO11-MBM 在 YOLO11 框架基础上进行系统性优化(图 1),通过骨干网络与特征融合网络的协同提升堆叠遮挡零件的检测性能:在骨干网络(Backbone)部分,设计了 C3k2-MCA 模块增强纹理相似零件区分及遮挡目标识别能力;在特征融合网络(Neck)部分,构建了 BiFPN-SFF-Concat 特征融合架构实现遮挡下的特征互补。该方案显著提升目标召回率与多尺度检测精度,为工业视觉系统提供了一种高效实用的检测方案。

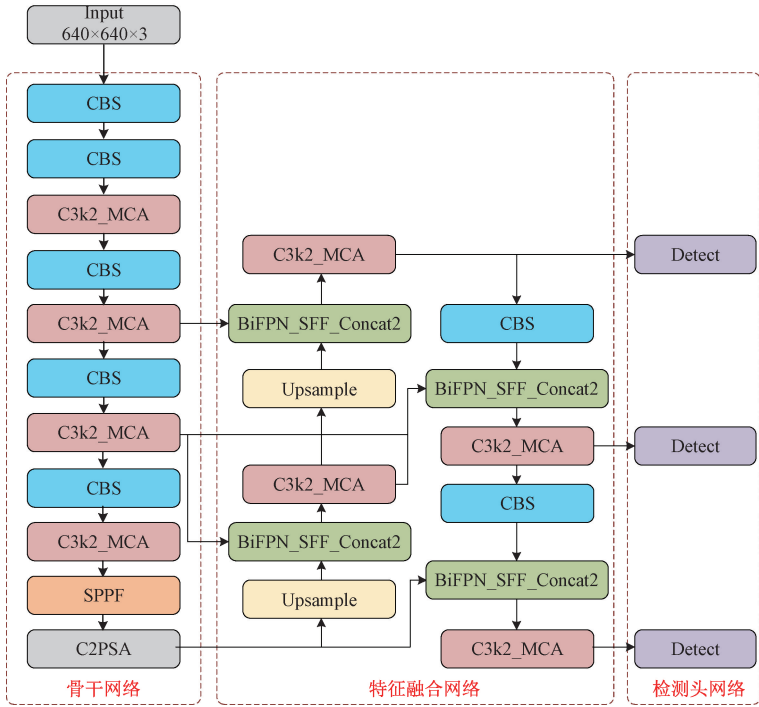


图 1 YOLO11-MBM 网络结构
Fig. 1 YOLO11-MBM network architecture

1.2 C3k2-MCA 结构

为了应对工业堆叠场景中零件遮挡导致的特征提取难题,本文在骨干网络中设计了 C3k2-MCA 模块(如图 2 所示,其中 h 、 w 、 c 分别为特征图的垂直像素、水平像素和通道数, in 和 out 表示输入与输出)。该模块融合多维度协作注意力机制与瓶颈(BottleNeck)结构,构建具备跨维度

特征感知能力的增强单元。

C3k2-MCA 模块先经卷积层提取输入特征的多尺度信息,再用 split 操作将特征划分为两支路送入瓶颈层处理。在瓶颈结构中嵌入了 MCA 机制,并设计了 3 种融合策略(图 3),其目的是通过不同融合方式探索最优的多维度注意力集成路径,在增强纹理相似零件区分能力的同

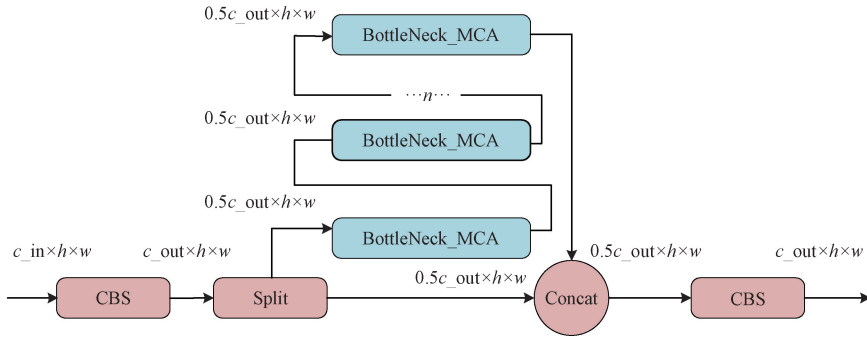


图 2 C3k2-MCA 模块结构示意图

Fig. 2 Schematic of the C3k2-MCA module

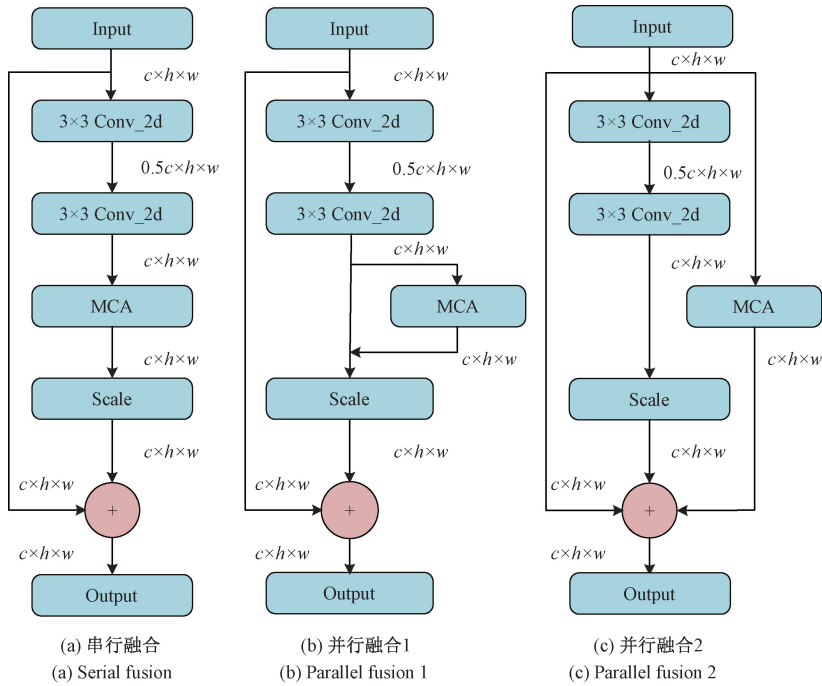


图 3 MCA 注意力机制 3 种融合方式示意图

Fig. 3 Schematic of three fusion methods for the MCA attention mechanism

时,平衡特征强化效果与计算开销。串行融合是特征经 1×1 卷积降维和 3×3 卷积提取后,再进行 MCA 重标定和尺度调整,再与残差分支融合,该策略结构简单且计算效率高。并行融合 1 是 MCA 特征重标定与特征尺度调整并行处理,然后通过逐元素相加融合两个分支输出。并行融合 2 采用双路并行结构,主支路按标准卷积提取特征,侧支路引入 MCA 机制以增强尺度特征感知,提升局部与全局特征的协同表达能力。最终,任意一个策略的 Bottleneck 输出均与残差分支特征融合,生成模块输出。后续实验将比较 3 种融合策略的性能表现,以确定最优结构。

MCA^[13]注意力机制融合多尺度信息与通道、空间注意力,以增强网络对不同尺度特征的自适应选择能力。该机制能抑制冗余信息干扰,提升网络对关键特征的提取

能力。

1.3 BiFPN-SFF-Concat 结构

为解决工业堆叠零件检测中因遮挡导致的特征语义差异与细节丢失问题,本文提出了 BiFPN-SFF-Concat 架构,该架构融合双向特征金字塔、动态权重融合拼接机制以及尺度敏感特征筛选策略,旨在增强低层特征的语义表达能力,降低漏检率。

在 BiFPN^[14]基础上,引入动态加权融合机制,通过 1×1 卷积统一各层通道数,为不同输入特征分配可学习权重实现自适应融合,还采用注意力机制引导低层特征关键区域增强,抑制无效背景。最终将融合后的特征与 BiFPN 输出拼接,提升整体特征表达能力。该结构按特征融合尺度层级数,分为双尺度的 BiFPN-SFF-Concat2 和三尺度的 BiFPN-SFF-Concat3。

BiFPN-SFF-Concat2 融合了高层与低层特征(图 4(a)):先将 f_{high} 上采样至与 f_{low} 空间分辨率一致,经 1×1 卷积统一通道数,分配可学习权重系数 W_{high} 和 W_{low} 。再用 CA 机制处理 f_{high} ,生成尺度敏感的注意力权重 $W_{\text{attn_high}}$ 引导 f_{low} 的关键区域增强。最后将增强后的 f_{low} 与加权 f_{high} 相加,再与 BiFPN-Concat 的输出拼接,得到兼具语义与细节的融合特征。

BiFPN-SFF-Concat3 扩展至高、中、低三尺度特征融合(图 4(b)):第 1 级将 f_{high} 上采样匹配中层特征 f_{mid} ,统一通道后加权融合,通过 CA 增强 f_{mid} 关键区域;第 2 层将第 1 级输出上采样至 f_{low} 的分辨率,再次统一通道并加权,同时用注意力机制引导 f_{low} 增强关键细节。最后将两级融合结果与 BiFPN-Concat 输出拼接,获得覆盖多尺度的综合特征,缓解多尺度语义断层问题。

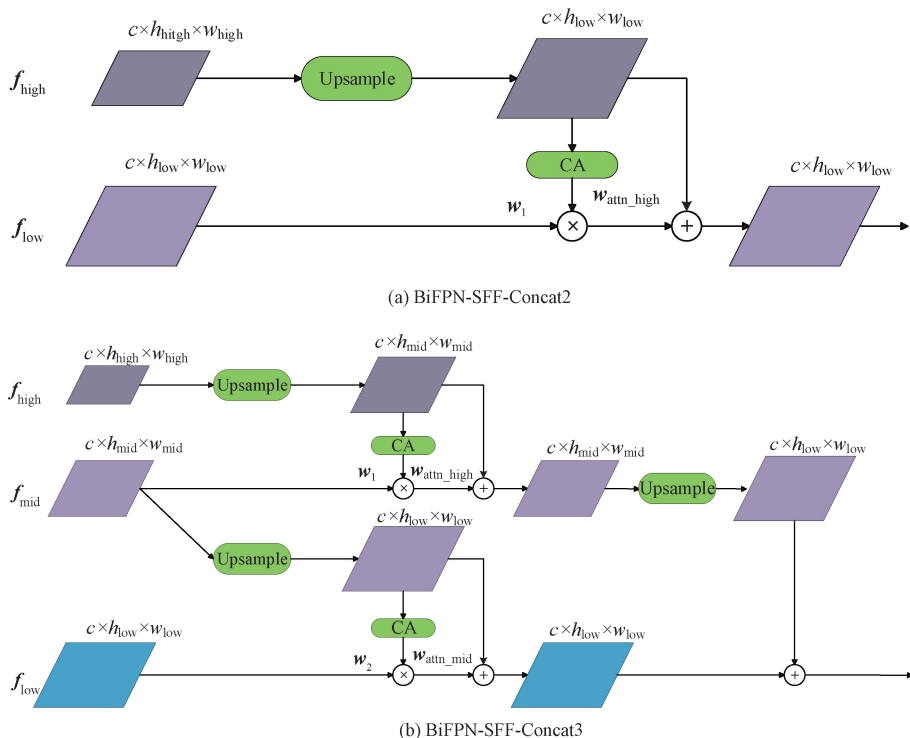


图 4 BiFPN-SFF-Concat2 网络结构

Fig. 4 BiFPN-SFF-Concat2 network architecture

该结构保留 BiFPN 的双向跨尺度连接优势,SFF^[15]模块通过转置卷积生成高层注意力图,引导低层特征的选择性融合,强调关键区域,抑制干扰。最终拼接 BiFPN 的输出与 SFF 处理结果,生成语义丰富、细节完备的特征表达,支撑高精度工业检测。

1.4 损失函数

工业堆叠零件检测中,传统交并比 (Intersection over union, IoU) 损失函数在严重遮挡时易失效。因为其仅基于预测框与真实框的重叠区域计算损失,当两者无重叠时梯度消失,且无法区分相同重叠度下不同几何关系的预测框,易引发边界框漂移,影响检测精度。

为此,本文采用 MPDIoU 损失函数^[16]替代传统 IoU。其核心是最小化预测框与真实框左上角和右下角对应点的欧氏距离之和,能更全面衡量边界框间相似性,适用于不同尺度的目标(尤其小目标检测),可有效缓解遮挡带来的定位误差,提升检测精度并加快模型收敛,实现精度与效率的双提升。

2 实验结果与分析

2.1 实验环境与评估指标

实验硬件为 Intel Core i7-13700KF 处理器 (24 核) 和 NVIDIA RTX-4080 显卡。软件环境基于 Ubuntu 20.04, 采用 Python 3.8 和 PyTorch 1.8.0 开发。训练超参数:初始学习率 0.01, 预热周期为 3 轮,余弦退火方法在训练过程中逐步降低学习率。训练轮数设为 200, batch_size = 16, 模型输入尺寸为 640 × 640。

本文构建的数据集以百度飞桨公开的工厂 13 类零件数据集为基础,利用 Labelimg 工具对其进行了精细化修正与系统性扩充。修正内容包括调整原始数据中不准确的边界框、纠正错误的类别标签,并对样本数量不足的类别进行了补充标注。调整后的数据集包含了 5 000 张图片,其中标注框总数约 18 500 个,涵盖六角螺丝、T 形螺丝、六角钢柱、键条、水平泡、长方形螺母、六角铜柱、双通六角柱、法兰螺母、圆头螺丝、六角螺母、弹簧垫圈和塑料垫柱

在内的 13 类典型工业零件,覆盖堆叠与非堆叠两种装配场景,数据按 8:1:1 的比例划分为训练集、验证集和测试集。在模型训练阶段,采用了随机旋转、亮度对比度调整、裁剪、高斯模糊及 Mosaic 等多种随机数据增强方法以提升模型的泛化能力。

参考文献[17-19]采用的评估体系,本文使用精确率 (precision, P)、召回率 (recall, R)、平均精度均值 (mean average precision, mAP)mAP50 和 mAP50:95 作为模型的评价指标,计算公式为:

$$P = \frac{TP}{TP + FP} \tag{1}$$

$$R = \frac{TP}{TP + FN} \tag{2}$$

$$AP = \int_0^1 P dR \tag{3}$$

$$mAP = \frac{1}{n} \sum_n AP_i \tag{4}$$

式中:TP 为模型正确识别的正类样本数;FP 为被错误预

测为正类的负类样本数;FN 为被错误判断为负类的正类样本数;AP 表示 P-R 曲线下的面积,用于评估模型在单个类别上的性能;mAP 为所有类别 AP 的平均值,体现模型在多类别任务中的整体表现。mAP50 代表 IoU 阈值设为 50%时所有类别平均检测精度;mAP50:95 代表 IoU 阈值设为 50%~95%之间,步长为 0.05 的整体平均检测精度。

2.2 消融实验

为验证 YOLO11-MBM 模型各模块的有效性,采用控制变量法设计多组消融实验,分析 C3k2-MCA 特征增强模块、BiFPN-SFF-Concat 加权特征融合架构及 MPDIoU 损失函数对检测性能的影响,并探究模块间协同作用机制。实验均在统一数据集与硬件环境下进行。

首先,针对 C3k2-MCA 模块的结构优化,设计两组对比实验。

1)注意力融合方式对比(图 3):测试串行融合、并行融合 1 和并行融合 2 这 3 种结构,结果如表 1 所示。

表 1 MCA 机制 3 种融合方式对应性能对照

Table 1 Performance comparison of 3 fusion strategies for MCA mechanism

模型	mAP50	mAP50:95	精确率	召回率	GFLOPs	平均推理时间/ms	参数量/(×10 ⁶)
YOLO11	0.818	0.490	0.816	0.788	6.3	6.1	25.9
YOLO11-MCA-串行	0.822	0.537	0.760	0.790	7.3	10.8	25.9
YOLO11-MCA-并行 1	0.827	0.555	0.792	0.789	6.3	10.0	25.9
YOLO11-MCA-并行 2	0.850	0.560	0.847	0.777	6.3	10.6	25.9

2)注意力机制类型对比:以坐标注意力(CA)、单映射注意力 (simple attention module, SimAM)为对照,验证

MCA 的优越性,结果如表 2 所示。特指 IoU 阈值设为 50%时计算得到的平均 AP 值。

表 2 不同注意力机制性能对照

Table 2 Performance comparison of different attention mechanisms

模型	mAP50	mAP50:95	精确率	召回率	GFLOPs	平均推理时间/ms	参数量/(×10 ⁶)
YOLO11	0.818	0.490	0.816	0.788	6.3	6.1	25.9
YOLO11-CA	0.825	0.520	0.810	0.819	6.4	7.9	26.0
YOLO11-SimAM	0.827	0.531	0.814	0.793	6.3	10.4	26.0
YOLO11-MCA	0.850	0.560	0.847	0.777	6.3	10.6	25.9

表 1 数据表明,C3k2-MCA 模块的 3 种注意力融合方式中,并行融合 2 在 mAP50 与 mAP50:95 上均取得最优性能,分别达到 0.850 与 0.560,且计算开销保持较低。并行融合 1 将 mAP50 提升至 0.827,mAP50:95 提升至 0.555,说明并行使用 MCA 注意力对提升堆叠场景性能有一定帮助,但效果有限。串行结构 mAP50 降至 0.822,mAP50:95 降至 0.537,这是因为两次卷积后再进行 MCA 操作,会使网络更关注抽象空间特征,忽略目标尺寸特征,还可能导致梯度消失。

为了深入理解不同融合策略对特征感知机制的影响,

进一步对 3 种策略的注意力特征图进行可视化分析,结果如图 5 所示。其中图 5(a)为输入原图及其检测结果,图 5(b)~(d)分别对应串行融合、并行融合 1 与并行融合 2 这 3 种策略的注意力特征图。从图 5(b)~(d)可知,不同融合策略下模型关注区域存在差异:串行融合(图 5(b))注意力分布较分散,部分响应落在背景区域,导致对堆叠目标的区分能力不足;并行融合 1(图 5(c))增强了对目标主体区域的关注,但其定位精度不高,且对重叠相似物体存在注意力混淆;而并行融合 2(图 5(d))的注意力更集中于目标的边缘与结构关键点,尤其在遮挡交界处表现出更强的语义

区分能力。该可视化结果与表 1 的定量结论一致,表明并行融合 2 通过双路结构有效融合了局部细节与全局上下文,从

而在遮挡场景下取得最优的检测性能。因此,本文采用并行融合 2 作为 C3k2-MCA 模块的最终融合策略。

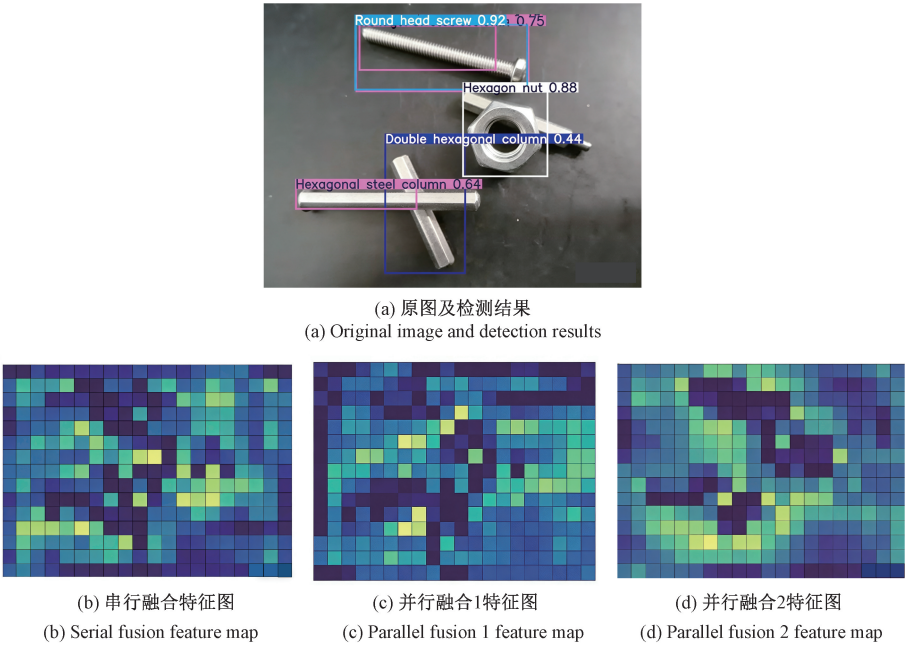


图 5 不同注意力融合策略特征图可视化对比

Fig. 5 Feature map visualization comparison of different attention fusion strategies

表 2 数据表明,3 种注意力机制均能提升原模型性能,引入 CA 注意力机制后,mAP50 从 0.818 提高至 0.825,mAP50:95 从 0.490 提升至 0.520。SimAM 注意力机制的 mAP50 达 0.827,mAP50:95 达 0.531,略优于 CA,其轻量级设计能较好平衡性能和参数量。YOLO11-MCA 模型将 mAP50 提升至 0.850,同时 mAP50:95 提升至 0.560,且召回率和精确率均有改善,分别为 0.777 和 0.847。虽然 MCA 计算开销更高,但其更复杂的多维度注

意力机制能更好地捕捉目标的细节特征,适配复杂堆叠场景需求。

其次,在原始 YOLO11 基础上,通过逐步引入创新模块构建了 8 组对比模型(表 3),包括原版 YOLO11(基线)、分别集成 C3k2-MCA (YOLO11-MCA)、BiFPN-SFF-Concat (YOLO11-BiFPNSFF) 和 MPDIoU (YOLO 11-MPDIoU)的单一模块组;两两组合的双模块组合,以及同时集成全模块的完整模型 YOLO11-MBM。

表 3 不同 YOLO11 改进方式对应性能对照

Table 3 Performance comparison of different YOLO11 improvement strategies

模型	mAP50	mAP50:95	精确率	召回率	GFLOPs	平均推理时间/ms	参数量/($\times 10^6$)
YOLO11	0.818	0.490	0.816	0.788	6.3	6.1	25.9
YOLO11-MCA	0.850	0.560	0.847	0.777	6.3	10.6	25.9
YOLO11-BiFPNSFF	0.848	0.542	0.841	0.769	6.5	7.5	27.0
YOLO11-MPDIoU	0.831	0.515	0.780	0.830	6.3	7.6	25.9
YOLO11-MCA-BiFPNSFF	0.866	0.568	0.812	0.838	6.5	10.6	27.0
YOLO11-MCA-MPDIoU	0.868	0.576	0.882	0.768	6.3	10.1	25.9
YOLO11-BiFPNSFF-MPDIoU	0.856	0.560	0.754	0.854	6.5	7.2	27.0
YOLO11-MCA-BiFPNSFF-MPDIoU	0.878	0.580	0.829	0.828	6.5	11.8	27.0

从表 3 可知,融合了 MCA、BiFPN-SFF-Concat 与 MPDIoU 的 YOLO11-MBM 模型较原 YOLO11 提升显著,mAP50 达 0.878,mAP50:95 也提升至 0.580。在表 3 所示的消融实验中,MCA 与 BiFPN-SFF-Concat 结合使

mAP50 提升至 0.866,mAP50:95 提升至 0.568,增强了多尺度特征融合与关键区域感知能力;加入 MPDIoU 可以进一步优化定位精度,mAP50 升至 0.878,mAP50:95 升至 0.580,有效减少了遮挡与重叠导致的定位误差。以上实

验结果证明了 YOLO11-MBM 算法的有效性。

2.3 不同算法对比实验

为验证 YOLO11-MBM 的综合性能,将其与主流目标检测算法在相同条件下对比(表 4)。由表 4 可知,

YOLO11-MBM 表现出色,召回率显著高于 RT-DETR, mAP50 略高于 RT-DETR,虽精确率稍低,但参数量仅 27.0×10^6 ,远低于 RT-DETR 42.0×10^6 的参数量。因此, YOLO11-MBM 实现了低计算开销与高检测性能的兼顾。

表 4 YOLO11-MBM 模型与现有其他模型性能对比

模型	mAP50	mAP50:95	精确率	召回率	GFLOPs	平均推理时间/ms	参数量/($\times 10^6$)
Faster-RCNN	0.635	0.342	0.696	0.618	180.0	38.0	42.0
YOLOv8	0.788	0.489	0.744	0.747	8.1	7.0	29.0
YOLOv10	0.801	0.503	0.769	0.725	8.3	8.0	27.0
RT-DETR	0.870	0.582	0.883	0.818	125.7	18.0	42.0
YOLO11	0.818	0.490	0.816	0.788	6.3	6.1	25.9
YOLO11-MBM	0.878	0.580	0.829	0.828	6.5	11.8	27.0

2.4 检测结果可视化

图 6(a)~(d)的检测效果对比表明,在复杂堆叠场景中, YOLO11-MBM 的检测效果显著优于 YOLOv10,后者受光照等因素干扰,难以区分目标与背景,存在严重误检与漏检问题。同时, YOLO11-

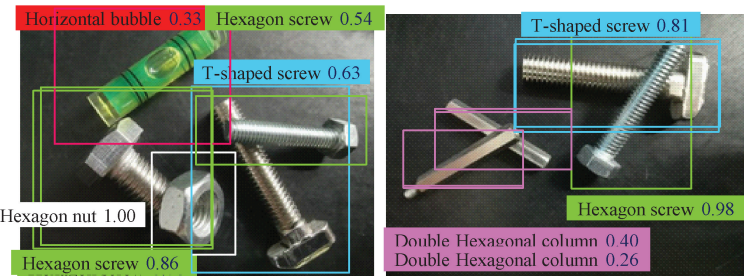
MBM 与凭借强大全局特征建模能力实现较高检测精度的 RT-DETR 表现相近。更为重要的是, YOLO11-MBM 在达到接近 RT-DETR 精度的同时,还有效兼顾了更优的推理速度,因此在实际场景测试中展现出理想的综合性能。



(a)YOLOv10检测效果
(a) Detection results of YOLOv10



(b) 原版YOLO11检测效果
(b) Detection results of original YOLO11



(c) RT-DETR检测效果
(c) Detection results of RT-DETR



(d) YOLO11-MBM检测效果
(d) Detection results of YOLO11-MBM

图 6 不同算法检测效果

Fig. 6 Visual comparison of detection results across different algorithms

3 结 论

针对工业堆叠零件检测中误检、漏检与定位精度不足问题,本文提出 YOLO11-MBM 算法,通过 C3k2-MCA 模块增强纹理相似零件区分能力、BiFPN-SFF-Concat 架构优化遮挡特征表达、MPDIoU 损失提升定位精度,构成三重协同优化。实验表明,该算法在百度飞桨工业零件扩展数据集上实现 87.8% mAP50,较 YOLO11 提升 6%,召回率提高 4%,同时保持 110 fps 实时性(V100 GPU)与轻量化(参数量仅增 1.1×10^6),有效平衡精度与效率,为复杂堆叠场景的零件检测提供了实用解决方案。

参考文献

[1] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6):84-90.

[2] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.

[3] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]//31st International Conference on Neural Information Processing Systems, 2017: 6000-6010.

[4] ZHAO Y, LYU W Y, XU S L, et al. Detrs beat YOLOs on real-time object detection[C]//2024 IEEE/ CVF Conference on Computer Vision and Pattern Recognition, 2024:16965-16974.

[5] 张永宏,许鑫豪,尹贺峰,等. MFSF-DETR:一种基于多尺度特征移位融合的 PCB 缺陷与元件检测算法[J]. 仪器仪表学报, 2025,46(8):266-285.

ZHANG Y H, XU X H, YIN H F, et al. MFSF-DETR: A PCB defect and component detection algorithm based on multi-scale feature shift fusion[J].

Chinese Journal of Scientific Instrument, 2025,46(8): 266-285.

[6] 杨洋,吴一全. 视觉 Transformer 在表面缺陷检测中的应用:研究进展与挑战[J]. 仪器仪表学报, 2025, 46(12):1-22.

YANG Y, WU Y Q. Applications of vision Transformer in surface defect detection: Research progress and challenges [J]. Chinese Journal of Scientific Instrument, 2025, 46(12):1-22.

[7] 余松森,张明威,杨欢. 改进 YOLOv5 的瓷砖表面缺陷检测[J]. 计算机系统应用, 2023,32(8):151-161.

YU S S, ZHANG M W, YANG H. Detection of ceramic tile surface defects based on improved YOLOv5 [J]. Computer Systems & Applications, 2023, 32(8): 151-161.

[8] 闫彬,邵军. 基于深度学习的苹果树果实目标检测实验设计[J]. 实验室研究与探索, 2025, 44(4): 5-8.

YAN B, SHAO J. Object detection experiment design of apple tree fruit based on deep learning[J]. Research and Exploration in Laboratory Science, 2025, 44(4): 5-8.

[9] 苏国用,帅洪铨,邓海顺,等. 基于机器视觉的煤矿多场景关键目标检测方法研究[J]. 电子测量与仪器学报, 2025,39(7):212-226.

SU G Y, SHUAI H X, DENG H SH, et al. Research on multi-scene key target detection method for coal mine based on machine vision[J]. Journal of Electronic Measurement and Instrumentation, 2025,39(7):212-226.

[10] 严赫,史晓亮,吴超华,等. GMS2-YOLO:融合改进 MANet 和检测头的金属表面缺陷检测方法[J]. 电子测量与仪器学报, 2026, 40(3):176-188.

YAN H, SHI X L, WU CH H, et al. GMS2-YOLO: A metal surface defect detection method integrating improved MANet and detection head[J]. Journal of Electronic Measurement and Instrumentation, 2026, 40(3):176-188.

[11] 孙超,刘均学,陈正超,等. 改进 YOLOv7 算法的钢

板表面缺陷检测方法[J]. 实验室研究与探索, 2025, 44(1): 19-23.

SUN CH, LIU J X, CHEN ZH CH, et al. Surface defect detection method of steel plate based on improved YOLOv7 algorithm [J]. Research and Exploration in Laboratory, 2025, 44(1): 19-23.

[12] SHI Y, HE Q, MAO C, et al. The object detection in remote sensing images based on improved YOLO11 with LDConv[C]//2024 6th International Conference on Robotics, Intelligent Control and Artificial Intelligence, 2024: 605-609.

[13] YU Y, ZHANG Y, CHENG Z Y, et al. MCA: Multidimensional collaborative attention in deep convolutional neural networks for image recognition[J]. Engineering Applications of Artificial Intelligence, 2023, 126: 107079.

[14] TAN M X, PANG R M, LE Q V. EfficientDet: Scalable and efficient object detection [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020:10778-10787.

[15] CHEN Y F, ZHANG C Y, CHEN B, et al. Accurate leukocyte detection based on deformable-DETR and multi-level feature fusion for aiding diagnosis of blood diseases [J]. Computers in Biology and Medicine, 2024, 170: 107917.

[16] MA S L, XU Y. MPDIoU: A loss for efficient and accurate bounding box regression[J]. ArXiv preprint arXiv:2307.07662, 2023.

[17] 梁列全,李想,何永华,等. 基于改进 YOLOv10n 的泊车图像旋转目标检测算法[J]. 电子测量技术, 2025, 48(19):205-216.

LIANG L Q, LI X, HE Y H, et al. Rotating target detection algorithm in parking images based on improved YOLOv10n [J]. Electronic Measurement Technology, 2025, 48(19):205-216.

[18] 牟涛,王代强. 基于改进 YOLOv11 光伏热斑小目标缺陷检测[J]. 电子测量技术, 2026, 49(6):229-238.

MU T, WANG D Q. Defect detection of small targets based on the improved YOLOv11 photovoltaic hot spot[J]. Electronic Measurement Technology, 2026, 49(6):229-238.

[19] 郭莉,张雪松,李萌萌,等. 改进 YOLOv10 的复杂场景人体跌倒检测方法[J]. 电子测量技术, 2026, 49(3): 204-212.

GUO L, ZHANG X S, LI M M, et al. Improved YOLOv10 method for human fall detection in cluttered scenes [J]. Electronic Measurement Technology, 2026, 49(3):204-212.

作者简介

方雁雁, 博士, 工程师, 主要研究方向为计算机视觉。
E-mail: yyfang@fudan. edu. cn

林志凌, 硕士研究生, 主要研究方向为计算机视觉。
E-mail: lzljelly@126. com

李宁宇, 硕士研究生, 主要研究方向为多模态大模型、计算机视觉。
E-mail: 24210720214@m. fudan. edu. cn

张志俭, 本科, 高级工程师, 主要研究方向为嵌入式系统、物联网应用。
E-mail: njuzzj@nju. edu. cn

李旦(通信作者), 博士, 教授, 主要研究方向为信号处理及其应用。
E-mail: lidan@fudan. edu. cn