

基于 Efficient-YOLO 的轻量化轴承缺陷检测^①



娄瑶迪, 岳俊峰, 周迪斌, 刘文浩

(杭州师范大学 信息科学与技术学院, 杭州 311121)

通信作者: 周迪斌, E-mail: dibinz@zju.edu.cn

摘要: 针对现有深度模型在工业轴承外观缺陷检测领域, 存在模型参数量大、特征融合不充分以及对小目标检测精度低等问题, 提出了一种轻量化自适应特征融合检测网络 (Efficient-YOLO). 首先, 该网络采用嵌入 CBAM 注意力机制的 EfficientNetV2 结构进行基本特征提取, 便于确保模型精度同时显著优化模型参数量; 其次, 设计了一种自适应特征融合网络 (CBAM-BiFPN), 用来增加网络对有效特征信息的提取; 接着, 在下游特征融合网络引入 Swin Transformer 机制, 同时配合上游网络引入的 Ghost 卷积, 大幅度提高模型对轴承外观缺陷的全局感知能力; 最后, 在推理阶段运用改进的非极大值抑制方法 (Soft-CIoU-NMS), 加入距离有关的权重评价因素, 减少了重叠框的漏检. 实验结果表明: 与现有主流检测模型相比, 此方法在轴承表面缺陷数据集上, mAP 达到了 90.1%, 参数量降低至 1.99M, 计算量为 7 GFLOPs, 对轴承缺陷小目标识别率显著提升, 满足工业现场轴承外观缺陷检测需求.

关键词: 轴承缺陷检测; 深度学习; EfficientNetV2; YOLOv5; 特征融合

引用格式: 娄瑶迪, 岳俊峰, 周迪斌, 刘文浩. 基于 Efficient-YOLO 的轻量化轴承缺陷检测. 计算机系统应用, 2024, 33(2): 265-275. <http://www.c-s-a.org.cn/1003-3254/9406.html>

Lightweight Bearing Defect Detection Based on Efficient-YOLO

LOU Yao-Di, YUE Jun-Feng, ZHOU Di-Bin, LIU Wen-Hao

(School of Information Science and Technology, Hangzhou Normal University, Hangzhou 311121, China)

Abstract: Since the existing deep model faces many problems such as a large number of model parameters, insufficient feature fusion, and low detection accuracy for small targets in the field of industrial bearing appearance defect detection, a lightweight adaptive feature fusion detection network (Efficient-YOLO) is proposed. First of all, the network uses the EfficientNetV2 structure embedded in the CBAM attention mechanism for basic feature extraction to ensure model accuracy and significantly optimize the model parameters. Secondly, an adaptive feature fusion network (CBAM-BiFPN) is designed to strengthen the network's extraction of effective feature information. Then, the Swin Transformer mechanism is introduced in the downstream feature fusion network, and the Ghost convolution introduced by the upstream network is used to greatly improve the model's global perception of bearing appearance defects. Finally, the improved non-maximum suppression method (Soft-CIoU-NMS) is applied in the inference phase, with distance-related weight evaluation factors added, so as to reduce missed detection of overlapping frames. The experimental results show that compared with the existing mainstream detection models, the method has a mAP of 90.1% on the bearing surface defect dataset. The number of parameters is reduced to 1.99M. and the calculation amount is 7 GFLOPs. The recognition rate of small targets with bearing defects is significantly improved, which meets the needs of industrial bearing appearance defect detection.

Key words: bearing defect detection; deep learning; EfficientNetV2; YOLOv5; feature fusion

① 基金项目: 国家自然科学基金联合重点项目 (U21A20466)

收稿时间: 2023-08-17; 修改时间: 2023-09-26; 采用时间: 2023-10-09; csa 在线出版时间: 2023-12-26

CNKI 网络首发时间: 2023-12-28

轴承是机械设备中的重要组成部件,其外观缺陷往往影响自身品质。然而,在生产和保存过程中,轴承易受制造工艺和环境因素影响,出现各类缺陷,如锈斑、擦伤、划痕和缺口等。不同类型的缺陷特征之间存在显著差异,因此难以设计一个适用于所有缺陷评估的通用算子,这一问题造成了该领域进展缓慢的现象。同时,外观缺陷可能出现在正反端面、内外侧面或倒角等位置,给图像采集和检测增加了复杂性。

随着人工智能技术的不断进步,深度学习技术在工业零件表面缺陷检测领域得到广泛应用。主流的深度学习目标检测算法可分为单阶段和两阶段检测算法。两阶段目标检测网络通常需要多次候选框生成、再分类和目标定位过程,导致较长的计算时间。相比之下,YOLO算法是一种单阶段目标检测算法,仅需一次前向传播,大大降低了计算复杂度,提高了检测速度。这使得YOLO在工业轴承缺陷检测中更适用于快速检测目标信息的需求。此外,工业缺陷数据通常包含多个角度和尺度的图像,以涵盖不同方向和大小的缺陷表现。对于这种多样化的缺陷数据,YOLO算法通过引入多尺度的特征图来检测不同大小的目标。多尺度特征表示有助于更好地捕捉小目标的细节和上下文信息,提高对小目标的检测准确性。为进一步提升模型性能,研究人员提出了一系列基于YOLO算法的改进,例如:Zhao等^[1]提出了一种轴承缺陷检测算法(GRP-YOLOv5),在特征提取阶段将ResC2Net与残差状结构相结合,在融合部分增加了PConv卷积,提高了模型捕获缺陷的能力,在化工设备缺陷轴承数据集上精度达到93.5%;Li等^[2]提出了基于改进YOLOv5算法的深度学习缺陷检测方法,结合了卷积注意力机制,并在颈部网络引入双向特征金字塔网络(BiFPN)模块,在金属轴数据集上精度达到93.6%;Niu等^[3]提出了一种改进的YOLOv5模型,用于检测印刷电路板缺陷。首先使用K-means++方法改进了先验锚框的位置匹配,然后引入了ECA-Net模块,以增强模型对通道特征的敏感性。在公开的PCB数据集上,该模型达到了99.1%的精度;Chen等^[4]提出一种Transformer-YOLO模型,使用Transformer作为特征提取网络,模型在电路板数据集上精度达到97.04%。上述模型虽然在各自的数据集上达到了较高的精度,但是还存在模型参数量大以及检测实时性较差的问题。因此,为降低模型的体积并提高检测速度,Qian等^[5]提出了LFF-YOLO模型用于多尺度缺陷检

测,首先使用ShuffleNetv2^[6]作为特征提取网络,其次提出了轻量级特征金字塔网络(LFPN),实现了检测速度的提升,简化后的模型参数减少74.6%;Zhao等^[7]提出了一种LSD-YOLOv5模型用于钢带表面损伤识别,在骨干网络中使用MobileNetV2,并且引入坐标注意力机制,颈部网络使用了一个更小的双向特征金字塔网络(BiFPN-S),成功减少了模型的体积,参数量降低了61.5%。Fan等^[8]构建了高精度低延迟的缺陷检测算法(GCB-Net),模型基于改进的YOLOv4目标检测网络,引入了卷积注意力模块和Ghost轻量级网络,完成对缺陷类别的预测,模型参数量压缩了82.3%。Xue等^[9]提出一种基于改进YOLO算法和ResNet18骨干特征提取网络的煤矸石检测算法,研究了模型的特征尺度缩减和非结构化修剪,模型体积可以压缩到原来的28.5%。上述算法虽然实现了模型的轻量化,但是由于模型参数和结构的减少,导致了检测精度的下降,以及对小目标检测能力的降低。

针对上述问题,结合轴承缺陷检测工业场合低成本的目标实时检测需求,本文基于YOLOv5基模型提出了一种轻量化自适应特征融合的Efficient-YOLO模型用于轴承外观检测,并从以下方面进行改进:(1)为解决目前轴承检测中模型参数量较大且占用资源较多的问题,采用改进的EfficientNetV2作为骨干网络,在保证算法精度的情况下实现算法的轻量化;(2)设计了一种自适应特征融合网络CBAM-BiFPN,用于提高模型特征融合能力,增强对轴承缺陷中边缘目标和不明显目标的检测能力,同时减少模型参数量;(3)为提高对轴承缺陷小目标的检测能力,使用C3STR模块替代C3模块,并且用GhostConv^[10]替代传统的深度分离卷积;(4)为减少轴承缺陷图像中重叠缺陷的漏检情况,在后处理阶段,通过使用Soft-CIoU-NMS替代非极大值抑制算法(non maximum suppression, NMS)^[11],解决了NMS存在的重叠框选择难题。

1 YOLOv5 模型结构

YOLO是一种目标检测算法,且不断迭代更新。本文基础网络部分使用的是YOLOv5 6.0版本,除输出端外,主要包括3个部分:Backbone、Neck、Head。

输入端主要由Mosaic图像增强、自适应锚框计算以及自适应图片缩放等模块组成,用于增加训练集多样性,提高模型的泛化能力。

Backbone 结构负责特征提取, 主要部分为卷积层, C3 结构和 SPPF (spatial pyramid pooling fast) 结构. 标准的卷积层结构用于对特征图的压缩和下采样; C3 结构拼接两部分特征图, 作为下层的输入; SPPF 结构对不同大小的特征图进行池化, 增强模型对不同尺度目标的感知能力.

Neck 颈部结构包括特征金字塔网络 (feature pyramid network, FPN)^[12] 路径聚合网络 (path aggregation network, PAN)^[13], 以实现特征融合和尺度适应. FPN 结构通过自上而下的上采样操作, 传递语义信息. PAN 结构则通过自下而上的下采样操作, 传递定位信息. 这两种结构相互补充, 聚合不同层次的特征信息, 实现对不同尺寸和不同大小目标的适应, 提高模型的

检测效果.

Head 头部结构包含 3 个检测头, 后处理中使用非极大值抑制来消除冗余框.

2 本文方法

2.1 Efficient-YOLO 模型结构

Efficient-YOLO 的整体结构如图 1 所示, 输入的图片调整为 $640 \times 640 \times 3$ 的统一形状, 首先通过 Efficient-NetV2^[14] 骨干网络进行特征提取, 生成 $10 \times 10 \times 512$ 的特征图, 然后输入到自适应特征融合网络 (convolutional block attention module-bidirectional feature pyramid network, CBAM-BiFPN) 中, 最后将输出的 3 种特征图传递给耦合头部网络, 生成检测框.

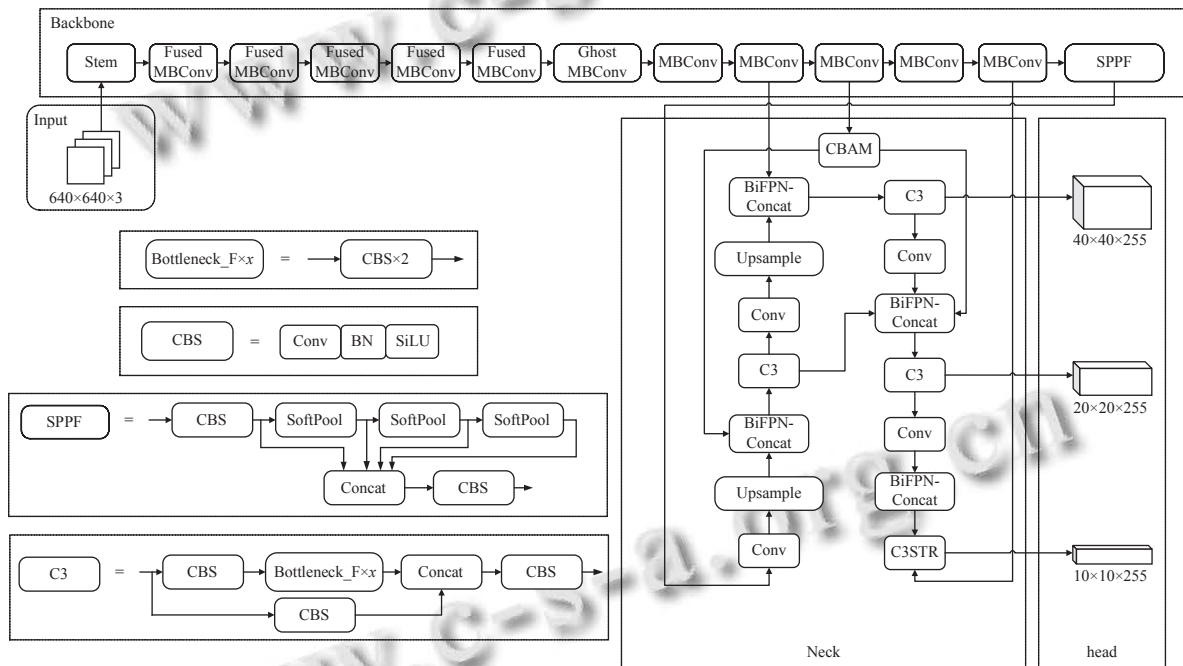


图 1 Efficient-YOLO 结构图

2.2 改进的 EfficientNetV2 模块

在工业轴承检测任务中, 传统的 YOLOv5 骨干网络使用大量卷积操作, 导致高度的计算资源消耗, 需要花费较长的时间来执行检测任务, 检测实时性能较差. 因此, 本文使用轻量化 EfficientNetV2 结构替换原有的骨干网络, 降低参数量, 减少计算复杂度, 提高模型的检测速度.

EfficientNetV2 是在 EfficientNet^[15] 的基础上进行了改进, 解决了原始结构中存在的复杂度大和计算成本高的问题, 包括 2 个组成部分: MBConv 模块和

Fused-MBConv 模块, 如图 2(a) 和图 2(b) 所示. 其中, MBConv 模块是 EfficientNet 中的基础结构, 实现了深度神经网络的高效训练; 而 Fused-MBConv 由升维卷积和降维卷积块组成, 可以在很小的计算量及很低的参数量条件下, 提高训练和推理的效率.

原 MBConv 模块引入了通道注意力模块 (squeeze-and-excitation, SE)^[16], 该模块对每个通道进行加权, 提高网络对重要的特征通道的关注度, 可在学习特征时更加灵活, 相比于原模型中的 SE 注意力模块只关注于通道域, 卷积注意力模块 (convolutional block attention

module, CBAM)^[17]包括两个注意力分支,分别是针对通道和空间维度的注意力.特征图被输入之后,先通过通道注意力获得新的特征图,再通过空间注意力,得到最终的特征图,此过程可以表示为式(1)和式(2),其中 F' 表示通过通道注意力获得的新特征, F'' 为最后输出的特征,符号 \otimes 表示元素相乘.

$$F' = M_c(F) \otimes F \tag{1}$$

$$F'' = M_s(F') \otimes F' \tag{2}$$

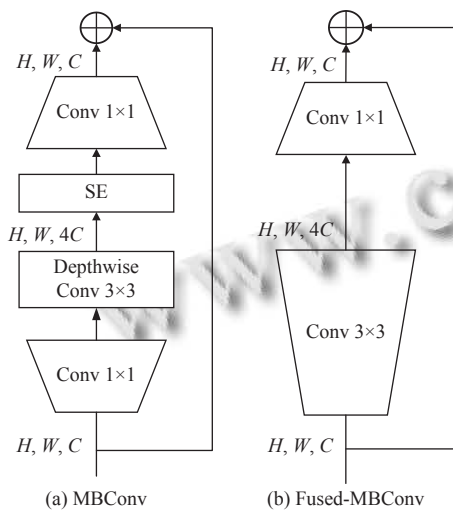


图2 MBCConv、Fused-MBCConv 结构图

CBAM 模块如图3所示.

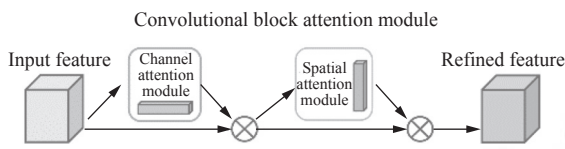


图3 CBAM 模块图

为了在大特征图提升小目标的关注度,本文在浅层网络将CBAM注意力模块嵌入到MBCConv模块中,替换了原有的SE注意力模块.在降维Conv之前嵌入CBAM,可有效筛选升维后深度分离卷积输出的高维度语义特征,少量增加参数和计算量,并且在保留重要细节特征图的同时,额外为网络提供筛选重要度较高的语义信息,将浅层的强定位特征传递到深层网络,改进后的MBCConv结构如图4(a)所示.

由于浅层特征提取网络可保留更多的小目标特征信息和位置信息,在稍微增加参数量以及计算量的情况下,本文将MBCConv模块中的深度分离3x3卷积替换为Ghost 5x5卷积,结构如图4(b)所示.使用Ghost

卷积的方法,可以仅通过少量的线性变换就能生成新的特征图,在轻量化的网络中能够保留更多冗余特征.Ghost卷积分为3个部分,第1部分是普通卷积操作,先生成少部分通道数的特征图,第2部分是分组卷积操作,通过对第1部分生成的特征图进行线性运算生成剩余特征图以保证通道数不变.第3部分就是将前两部分得到的特征图拼接.替换为Ghost 5x5卷积可以在有限的感受野范围内保留更多边缘信息,以此增强对小目标缺陷的定位准确性,改善图像边缘部分缺陷的遗漏情况, Ghost卷积结构如图5所示.

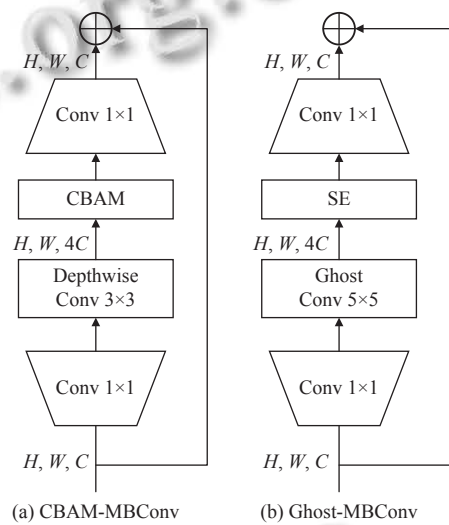


图4 CBAM-MBCConv、Ghost-MBCConv 结构图

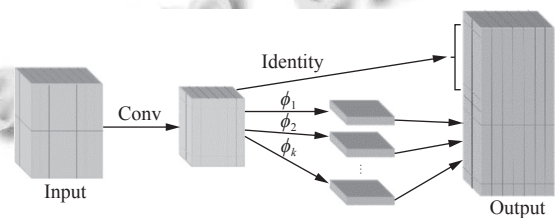


图5 Ghost卷积结构图

2.3 CBAM-BiFPN 模块

YOLOv5模型通过建立双向特征融合网络增强了特征等级结构的局部建模能力.然而,FPN和PAN模型未考虑特征融合层之间的语义信息差异,存在多尺度信息丢失的问题,缺乏对边缘小目标和密集目标的特征利用.为了补偿信息损失,EfficientDet检测算法提出了加权双向特征金字塔网络BiFPN(bidirectional feature pyramid network, BiFPN),在同一层的输入节点和输出节点之间新加一条边,并且在双向网络中简化单一节点,以便在不增加太多成本的情况下融合更多

的特性。

YOLOv5 在 FPN 和 PAN 结构的基础上, 把 P4 特征层也加入到特征融合中, 如图 6(a) 所示. 此结构可以额外融合中层提取网络的特征, 但是由于 PAN 是作为整个 Neck 部分重要的信息融合金字塔结构, 因此对输入分辨率和目标大小敏感且不够灵活. 额外的加权操作不仅给 PAN 部分增加了计算量, 而且容易使模型产生过拟合的情况。

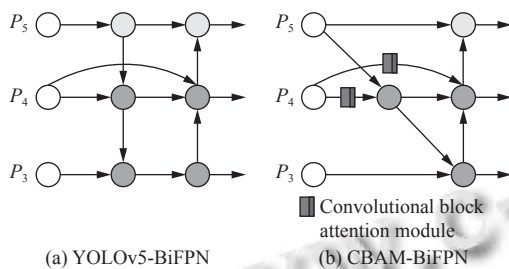


图 6 YOLOv5-BiFPN 和 CBAM-BiFPN 结构图

为适用于本文的轻量化网络设计目标, 并且提高对轴承细节特征的关注度, 设计了一种自适应特征融合网络 CBAM-BiFPN, 如图 6(b) 所示. 通过引入 CBAM 注意力机制于加权边以自适应抽取重要度较高的特征, 这种特性能够有效抑制不重要特征, 为加权操作提供了更为丰富的语义信息流. 此外, 引入 CBAM 模块的加权边可有效规制模型的过拟合倾向, 使得模型的泛化性能更强, 能够更好地适应不同样式和种类的轴承缺陷检测任务. 与 YOLOv5 提出的 BiFPN 改进方法相比, 本文提高模型对特征细节感知能力的同时, 简化了上采样过程的通路节点, 优化了卷积窗口大小, 实现了节点之间的计算量大幅度减少、整体网络参数量的下降。

2.4 C3STR 模块

在深层网络的高级特征图中, 轴承图像中小目标的特征信息逐渐丢失, 因此本文借鉴 Swin Transformer^[18] 的思想, 在原模型的大尺寸检测头中, 引入 C3STR 结构, 借助窗口自注意力模块增强小目标的语义信息和特征表示, 为底层带来更丰富的语义信息, 尤其是全局的语义信息。

基于 Swin Transformer 改进的 C3STR 结构如图 7 所示. 其中 STR 模块是 Swin Transformer block 的封装, 由两种 Swin Transformer block 成对组合而成. 这两种结构一个使用窗口多头自注意力模块 W-MSA, 另一

个使用滑动窗口多头自注意力模块 SW-MSA, 两者的配合使用使得相邻窗口间的信息得到有效传递。

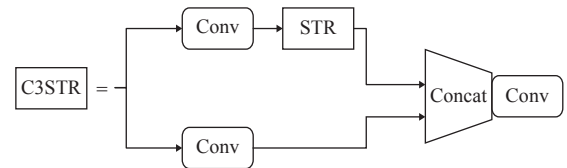


图 7 C3STR 结构图

本文使用 C3STR 加强卷积块替换了 C3 结构中的 Bottleneck 堆叠部分, 不仅能够通过自注意力机制使全局建模图像依赖关系, 提高模型对全局信息的特征提取能力, 而且 C3STR 结构支撑于小检测头部分, 能够在增加少量参数和计算量的情况下提高对全局目标的检测性能。

2.5 Soft-CIoU-NMS 模块

轴承缺陷图像中的部分缺陷往往以高密度和相互重叠的方式出现, 传统目标检测使用 NMS 方法进行后处理, 常常导致相邻检测框的分数强制设置为零, 会将重叠部分大于重叠阈值的检测框移除, 导致漏检的情况. 改进的 Soft-NMS 算法^[19] 的流程和 NMS 算法类似, 但是对于 IoU 大于阈值的检测框不是直接删除, 而是使用函数运算降低得分. 本文使用的 Soft-CIoU-NMS 结构, 是基于 Soft-NMS 和 Ciou 算法, 旨在解决传统 Soft-NMS 方法对距离度量不敏感的问题, 增加对检测框的考量, 减少对密集目标的错误抑制, 从而提高检测精度. Soft-CIoU-NMS 的公式可以表示为:

$$S_i^{\text{Soft-CIoU}} = S_i \left(1 - \max_{j \neq i} \frac{CIoU(b_i, b_j)}{\rho(i, j)} \right) \quad (3)$$

其中, S_i 表示第 i 个框的检测分数, b_i 表示第 i 个框的边界框, $\rho(i, j)$ 表示第 i 个框和第 j 个框的距离度量, $CIoU(b_i, b_j)$ 表示基于 Ciou 度量的重叠度量. 公式中的 $\max_{j \neq i}$ 表示对于第 i 个框, 降低最大 IoU 对应框的分数, $\frac{CIoU(b_i, b_j)}{\rho(i, j)}$ 表示基于距离度量的重叠度量。

3 实验结果

3.1 实验数据与设置

3.1.1 数据集采集

本次实验数据集为自制的轴承表面缺陷数据集, 从不同条件共采集 3 802 张分辨率为 1280×1024 的轴承图片, 经过数据增强扩充至 9 505 张, 并将缺陷分为

划痕、擦伤、缺口和锈斑等 4 个类别. 使用 LabelImg 软件对数据进行标注, 按照 8:1:1 的比例分为训练集、验证集和测试集, 其中训练集包含 7 604 张图片, 验证集包含 950 张图片, 测试集包含 951 张图片, 部分数据如图 8 所示.

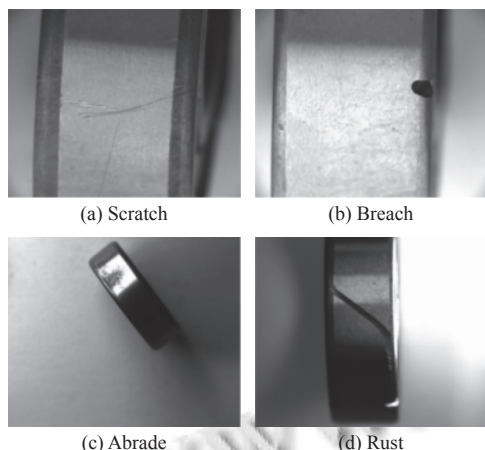


图 8 数据集抽样

3.1.2 模型训练与推理

实验基于 PyTorch 深度学习框架, 所有模型训练从零开始, 使用 GPU 进行训练, epoch 设置为 200 轮, 其中前 30 轮用于热身阶段. 优化器使用 SGD 随机梯度下降法. 初始学习率为 0.01, 动量为 0.937, 使用余弦退火策略将学习率降至 0.000 01, 权重衰减为 0.000 5. 推理验证测试统一使用 CPU 进行.

实验环境为 Ubuntu 20.04, PyTorch 2.0, CUDA 12.2. CPU 型号为 Intel I7 13700k, 显卡为 NVIDIA RTX4090. 使用 COCO 检测和评估标准中的指标进行性能评估, 包括平均精度 (AP), IoU 阈值为 0.5 的平均准确度 (mAP), 不同阈值 (从 0.5 至 0.95, 步长为 0.05) 下的平均准确度 (mAP@.5:.95), 参数量 (Params), 浮点运算 (GFLOPs) 和每秒帧数 (FPS). 其中 AP 和 mAP 计算公式为式 (4) 和式 (5), 其中 P 表示精确率, R 表示召回率.

$$AP = \int_0^1 P(R) dR \quad (4)$$

$$mAP = \frac{\sum_{i=0}^n AP(i)}{n} \quad (5)$$

3.2 实验结果对比

3.2.1 各种目标检测算法对比

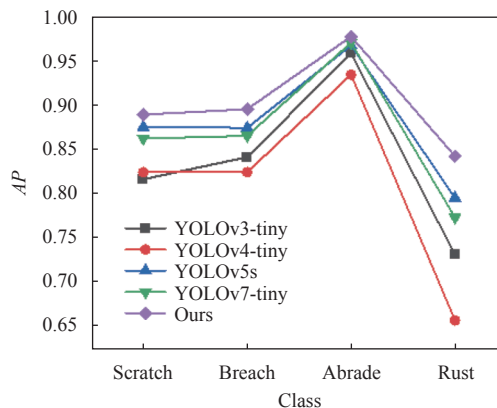
为验证 Efficient-YOLO 算法性能, 将其与主流目标检测算法的进行对比, 如表 1 所示.

表 1 各种目标检测算法对比实验

Model	mAP@.5	mAP@.5:.95	Params (M)	GFLOPs	FPS
Faster-RCNN (VGG16)	0.874	0.474	136.75	200.9	2
SSD (MobileNetV2)	0.755	0.310	3.94	6.3	6
YOLOv4-tiny	0.810	0.391	3.07	6.3	31
YOLOv3-tiny	0.836	0.428	8.67	12.9	31
YOLOv5s	0.876	0.500	7.02	15.8	19
YOLOv7-tiny	0.868	0.469	6.01	13.0	21
Efficient-YOLO (ours)	0.901	0.511	1.99	7.0	29

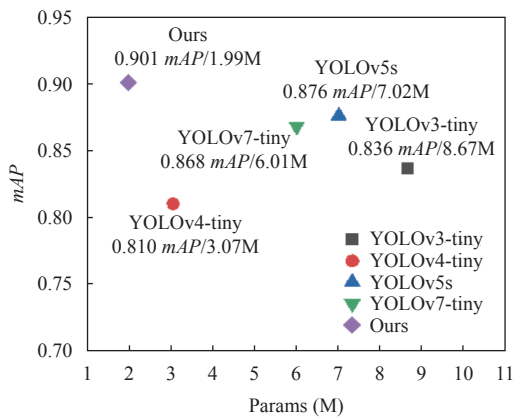
从表 1 中可看出, 相较于其他模型, 本文提出的改进模型在几个方面展现了卓越的高效性. 首先, 该模型在精度上取得了更高的表现, 即在目标检测任务中能够更准确地定位和识别目标. 其次, 该模型的检测速度更快, 即在相同的计算资源下能够更快地完成目标检测任务. 此外, 该改进模型的体积更小, 占用的计算资源也更少, 因此能够更容易地在嵌入式设备或有限计算资源的工业环境中部署和应用. 对比于传统两阶段检测算法 Faster-RCNN^[20]以及 SSD 模型^[21], 改进后的模型在精度、参数量方面都有大幅提升、计算量也大幅减少, 检测速度提高了数倍. 与原始模型 YOLOv5s 相比, Efficient-YOLO 在平均精度上提高了 2.5%, 达到了 90.1%, 参数量降低了 5.03M, 仅有 1.99M, 计算量降低了 8.8 GFLOPs, 检测速度提高了 10f/s, 在各类缺陷的检测精度上也都有不同程度的提升.

由图 9(a) 可得, 与同系列 YOLO 算法模型相比, 本文模型各类缺陷检测精度处于最高点, 说明在各类缺陷的检测精度均高于其他模型. 由图 9(b) 可得, 改进模型处于图的最左侧, 比 YOLOv4-tiny 的参数量还小, 在高度方面处于最高点, 说明本文模型在 mAP 上表现最好.



(a) 各类缺陷精度对比图

图 9 YOLO 系列检测网络效果对比



(b) 各类模型参数量与平均精度对比图

图9 YOLO 系列检测网络效果对比 (续)

为进一步证明本文提出的改进算法在轴承表面微小目标检测中的性能效果,参照于 YOLOv5s 基准模型测试结果,分别从 *AP*, *Recall* 及特征图可视化等方面进行评价分析。

在本文的轴承表面缺陷数据集中,缺口类缺陷作为显著性的小目标,在阈值 (threshold) 都为 0.6 的情况下,本文模型比 YOLOv5s 基准模型在缺口类 breach 上的 *Recall* 值提升了 3.5%, *AP* 值提高了 3.5%;在细微划痕类上的 *Recall* 值提升了 4.0%, *AP* 值提高了 4.4%。图 10 中分别展示了在小目标缺口和细微划痕类中, *IoU* 阈值随着步长为 0.1 变化时,本文模型 (橙线) 与原模型 (蓝线) 的 *AP* 值及 *Recall* 率曲线对比。

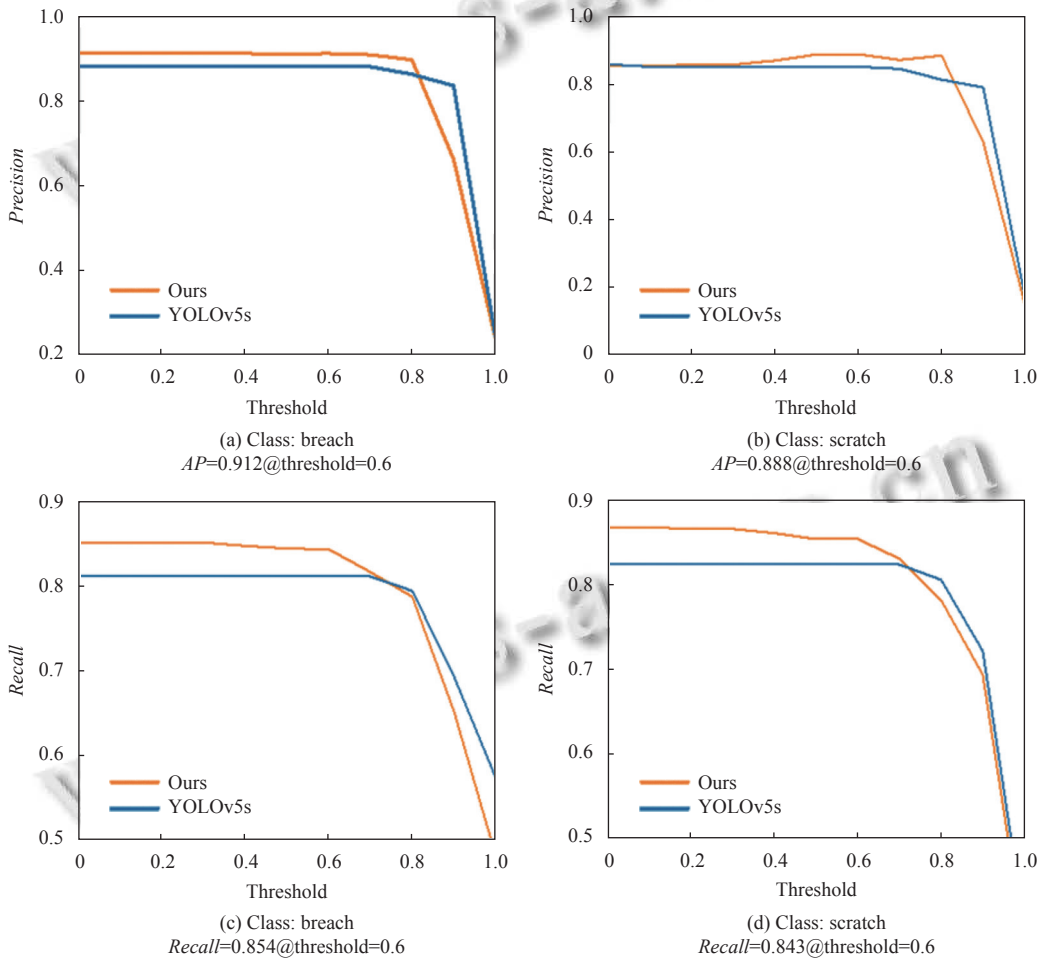


图 10 *AP* 和 *Recall* 对比

图 11 显示了原模型和改进模型的特征图可视化对比效果。由图可知,改进后的模型比原模型的热力点面积更大,说明对目标的关注度更高;改进后的模型比原模型的热力点更多,说明目标漏检情况得到了改善,

小目标的检测能力更强。

3.2.2 基于 YOLOv5s 的不同骨干网络对比

本文首先以 YOLOv5s 为基线进行轻量化改进,分别使用 EfficientLite^[15]、MobileNetV3^[22]、ShuffleNetV2、

GhostNet^[10]以及 EfficientNetV2 等主流特征提取网络重构 YOLOv5s 的骨干部分, 模型训练后对轴承表面缺陷数据集进行检测推理, 实验结果见表 2.

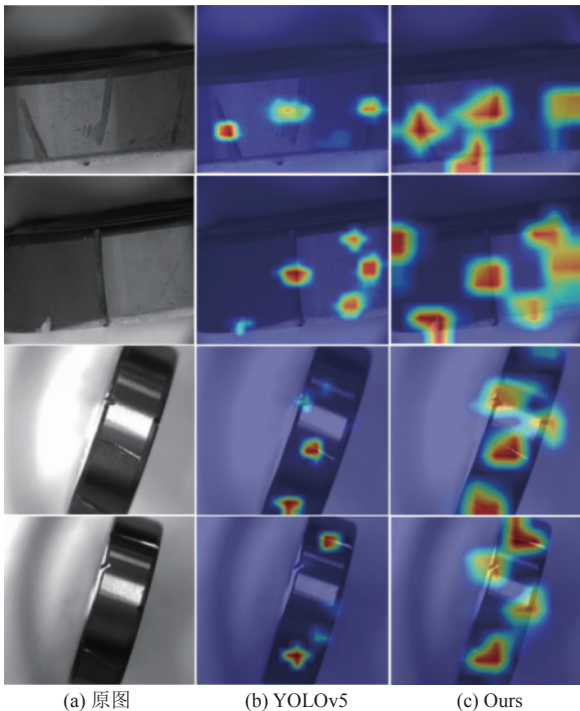


图 11 特征图可视化

表 2 不同骨干网络对比实验

Backbone	$mAP@.5$	$mAP@.5:.95$	Params (M)	GFLOPs	FPS
原YOLOv5s	0.876	0.500	7.02	15.8	19
EfficientLite	0.864	0.481	3.77	7.3	14
MobileNetV3	0.845	0.453	3.53	6.1	28
ShuffleNetV2	0.834	0.450	3.19	5.9	36
GhostNet	0.861	0.479	5.37	8.1	16
EfficientNetV2 (ours)	0.886	0.488	4.21	5.3	26

由表 2 可知, 以 EfficientLite、MobileNetV3、ShuffleNetV2、GhostNet 等作为骨干的模型, 对比于原 YOLOv5 模型在参数量以及计算量方面都有所下降, 但是牺牲了精确率, 而继承了 EfficientNetV2 作为骨干的模型, 在保证精度的情况下, 模型体积较小, 检测速度较快, 降低了资源占用率, 对比于其他算法性能更优.

3.2.3 嵌入不同注意力机制的 MBCConv 模块对比

为体现本文所用的 CBAM 注意力机制的性能效果, 将其与 SE^[23]、CA^[23]、ECA^[24]、CoT^[25]等几种经典注意力机制进行对比, 如表 3 所示.

由实验可得, 对比于其他注意力机制, 本文所使用

的 CBAM 注意力机制结合卷积和注意力机制, 能够充分利用空间和通道的信息, 关注不同空间位置的重要性. 在 $mAP@.5$ 上表现要好于其他注意力机制, 且在参数量以及检测速度方面均处于最优.

表 3 嵌入不同注意力机制的 MBCConv 模块对比实验

Model	$mAP@.5$	$mAP@.5:.95$	Params (M)	GFLOPs	FPS
MBCConv-SE	0.886	0.488	4.21	5.3	26
MBCConv-CA	0.881	0.487	4.20	5.3	26
MBCConv-ECA	0.878	0.485	4.21	5.3	27
MBCConv-CoT	0.883	0.491	4.21	5.3	27
MBCConv-CBAM (ours)	0.890	0.493	4.20	5.3	27

3.2.4 各种特征融合网络对比

为验证 CBAM-BiFPN 作为特征融合网络的性能表现, 本文以 EfficientNetV2 为骨干网络的 YOLOv5s 为基准, 将 CBAM-BiFPN 与 FPN、FPNPAN、BiFPN 这 3 种主流特征融合网络进行对比实验.

如表 4 所示, 在 mAP 表现上, 原模型的精度更高, 但也导致参数和计算量较大. 在参数量方面, 使用 CBAM-BiFPN 结构替换后的网络模型, 仅拥有 1.89M 参数量, 且整体网络的计算量大幅下降, 相较于使用 FPN 结构的模型, 参数量降低了 40.5%. 在测试速度方面, 本文的模型也稍快于其他模型.

表 4 各种特征融合网络对比实验

Neck	$mAP@.5$	$mAP@.5:.95$	Params (M)	GFLOPs	FPS
FPN	0.879	0.466	3.17	5.0	26
FPN PAN	0.890	0.493	4.20	5.3	27
BiFPN	0.883	0.484	3.71	5.1	27
CBAM-BiFPN (ours)	0.883	0.460	1.89	4.3	29

在第 2.2 节中, 阐述了通过作用于加权边的 CBAM 注意力模块可增强 BiFPN 网络的泛化性能, 为了证明这一特性, 本文列出了对比于 YOLOv5 提出的 BiFPN 改进方法的训练误差轮次下降结果, 分为 cls_loss, box_loss, obj_loss 这 3 部分, 如图 12 所示, 在 YOLOv5 训练 loss 计算方法中, 其中 cls_loss 用于监督类别分类, box_loss 用于监督检测框的回归, obj_loss 用于监督 grid 中是否存在物体, loss 的计算在每层 feature mAP 上进行. 由表 4 可知, 在相同 $mAP@.5$ 精度表现的两种加权特征融合网络, 本文所提出的 CBAM-BiFPN 训练误差结果下降和收敛均略高于 YOLOv5-BiFPN 水平, 可见轻量化模型过拟合倾向得到了有效抑制, 不难预见此种训练得到的模型泛化性能相比其他更强,

可期待在工业应用中得到更好的鲁棒性。

3.3 消融实验

为确认本文所提出的 Efficient-YOLO 的性能优

势, 在轴承表面缺陷数据集上, 设计了 5 组消融实验。消融实验中不使用迁移学习, 以准确比较设计模块的性能。如表 5 所示。

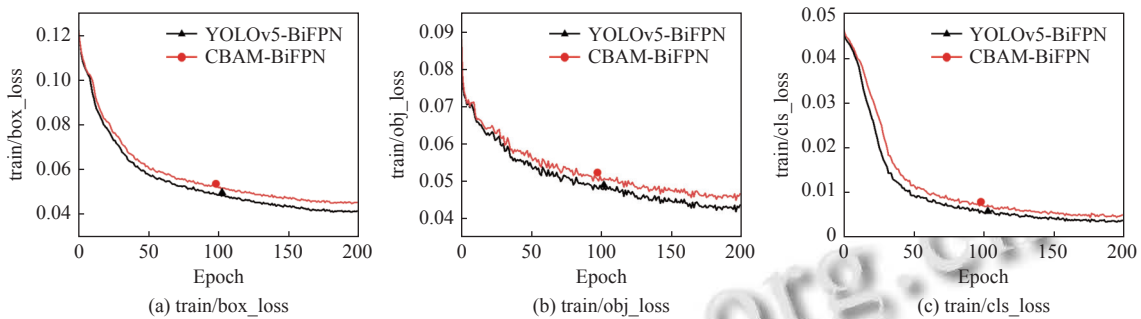


图 12 CBAM-BiFPN、YOLOv5-BiFPN 的损失曲线

表 5 消融实验

Number	Model	$mAP@.5$	$mAP@.5:.95$	Params (M)	GFLOPs	FPS	$AP50$ of Classes			
							Scratch	Breach	Abrade	Rust
①	EfficientNetV2 (改)+CBAM-BiFPN	0.883	0.460	1.89	4.3	29	0.865	0.870	0.974	0.825
②	①+GhostConv	0.880	0.461	1.96	4.5	27	0.878	0.871	0.968	0.801
③	①+C3STR	0.883	0.468	1.92	6.8	28	0.874	0.880	0.978	0.799
④	①+Soft-CIoU-NMS	0.890	0.499	1.89	4.3	28	0.871	0.894	0.969	0.826
⑤	①+GhostConv +C3STR	0.895	0.475	1.99	7.0	28	0.890	0.889	0.971	0.830
⑥	⑤+ Soft-CIoU-NMS (ours)	0.901	0.511	1.99	7.0	29	0.889	0.895	0.977	0.842

由表 5 可知, 实验①使用改进 EfficientNetV2 同时加入 CBAM-BiFPN, 并作为后续实验的对比基线, 对比于原模型大幅度降低了参数量以及计算量, 精度也得到小幅度改善。由实验分析可见, 用 Ghost Conv 替换 MBConv-CBAM 模块的深度分离 3×3 卷积, 获得了更多的位置信息, 保留了更多边缘信息, 使得小目标的检测精度提高, 在 scratch 类上表现良好。使用 C3STR 模块替代原先的 C3 结构, 由于 Swin Transformer 在深层网络处的语义信息获取能力, 在 scratch 类、breach 类以及 abrade 类中均有明显提升, 在增加少量参数的情况下, 增强了模块对各类缺陷的关注度。使用 Soft-CIoU-NMS 替换原先的 NMS 结构, 解决了重叠框的选择难题, 提升了对距离以及检测框的关注度, 使得检测精度得到了提高。综合原模型和最终改进的模型比较结果, 改进后的模型, 在 mAP 、GFLOPs 和 FPS 均优于原先的 YOLOv5s 模型。

针对小目标、特征不显著的 breach 和 scratch 缺陷类, 以及广泛分布的 rust 缺陷存在难检测、漏检、确信度不高等问题, 开启 Soft-CIoU-NMS 情况下, 本文测试了加入 GhostConv+C3STR 改进组合前后的模型

实际检测效果, 如图 13 所示, 第 1 行为改进前检测效果图, 第 2 行为改进后检测效果图。得益于 Ghost Conv 在特征提取层收获更大的感受野, 图像边缘处的缺陷检出率提高, 配合 Swin Transformer 在大尺寸整张图片的全局注意力运算时移动窗口的全局特征抽取的能力, 模型的下游部分在密集型预测型任务上置信度得到改善。

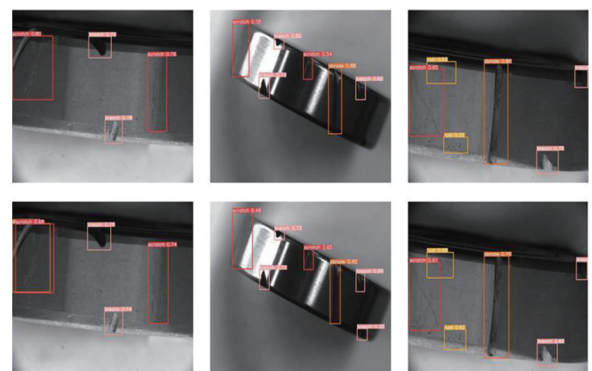


图 13 消融实验对比图

为更直观地展示 Efficient-YOLO 模型目标检测能力, 选取了部分实验结果, 如图 14 所示。

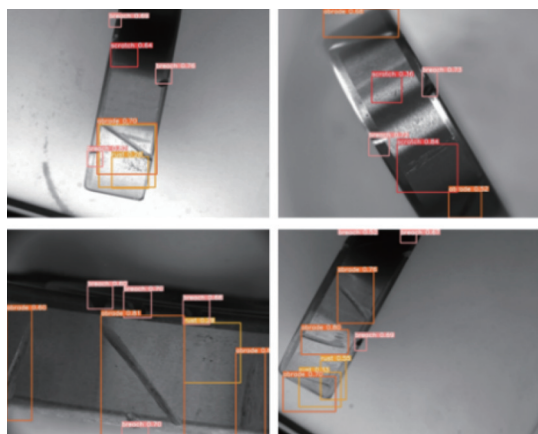


图14 模型检测效果展示

4 结论

本文针对轴承缺陷检测在工业应用中的需求,提出了一种轻量化自适应特征融合网络 Efficient-YOLO. 首先采用改进的 EfficientNetV2 作为骨干网络,大幅减少了参数量,其次在特征融合网络使用了嵌入 CBAM 注意力机制的双向加权特征融合网络,有效提高模型对浅层和深层语义信息的融合能力,然后引入 C3STR 配合 Ghost 卷积,提高对小目标的检测能力,最后在后处理阶段使用了 Soft-CIoU-NMS,提高了检测精度. 实验证明,本文方法在保持轴承外观缺陷检测精度的同时实现了检测模型的轻量化,并且性能表现优于现有基准模型. 对比于原模型,改进后的模型的参数量下降了 71.7%,计算量减少了 55.7%,检测速度提升了 34.5%,精度提高了 2.5%,进一步满足了对轴承表面缺陷检测的需求. 下一步的研究方向将聚焦于如何在嵌入式平台上部署低精度运算,并探索边缘计算的应用.

参考文献

- 1 Zhao Y, Chen BL, Liu BS, *et al.* GRP-YOLOv5: An improved bearing defect detection algorithm based on YOLOv5. *Sensors*, 2023, 23(17): 7437. [doi: 10.3390/s23177437]
- 2 Li B, Gao QJ. Defect detection for metal shaft surfaces based on an improved YOLOv5 algorithm and transfer learning. *Sensors*, 2023, 23(7): 3761. [doi: 10.3390/s23073761]
- 3 Niu J, Li HY, Chen X, *et al.* An improved YOLOv5 network for detection of printed circuit board defects. *Journal of Sensors*, 2023, 2023: 7270093.
- 4 Chen W, Huang ZT, Mu Q, *et al.* PCB defect detection method based on Transformer-YOLO. *IEEE Access*, 2022, 10: 129480–129489. [doi: 10.1109/ACCESS.2022.3228206]
- 5 Qian XH, Wang X, Yang SY, *et al.* LFF-YOLO: A YOLO algorithm with lightweight feature fusion network for multi-scale defect detection. *IEEE Access*, 2022, 10: 130339–130349. [doi: 10.1109/ACCESS.2022.3227205]
- 6 Ma NN, Zhang XY, Zheng HT, *et al.* ShuffleNet v2: Practical guidelines for efficient CNN architecture design. *Proceedings of the 15th European Conference on Computer Vision (ECCV)*. Munich: Springer, 2018: 122–138.
- 7 Zhao H, Wan F, Lei GB, *et al.* LSD-YOLOv5: A steel strip surface defect detection algorithm based on lightweight network and enhanced feature fusion mode. *Sensors*, 2023, 23(14): 6558. [doi: 10.3390/s23146558]
- 8 Fan BB, Li WX. Application of GCB-Net based on defect detection algorithm for steel plates. <https://www.researchsquare.com/article/rs-1550068/v1>. (2022-04-21).
- 9 Xue GH, Li SX, Hou P, *et al.* Research on lightweight YOLO coal gangue detection algorithm based on ResNet18 backbone feature network. *Internet of Things*, 2023, 22: 100762. [doi: 10.1016/j.iot.2023.100762]
- 10 Han K, Wang Y, Tian Q, *et al.* GhostNet: More features from cheap operations. *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020. 1580–1589.
- 11 Neubeck A, van Gool L. Efficient non-maximum suppression. *Proceedings of the 18th International Conference on Pattern Recognition (ICPR2006)*. Hong Kong: IEEE, 2006. 850–855.
- 12 Lin TY, Dollár P, Girshick R, *et al.* Feature pyramid networks for object detection. *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017. 936–944.
- 13 Liu S, Qi L, Qin HF, *et al.* Path aggregation network for instance segmentation. *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018. 8759–8768.
- 14 Tan MX, Le QV. EfficientNetV2: Smaller models and faster training. *Proceedings of the 38th International Conference on Machine Learning*. PMLR, 2021. 10096–10106.
- 15 Tan MX, Pang RM, Le QV. EfficientDet: Scalable and efficient object detection. *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle: IEEE, 2020. 10778–10787.
- 16 Hu J, Shen L, Sun G. Squeeze-and-excitation networks. *Proceedings of the 2018 IEEE/CVF Conference on Computer*

- Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 7132–7141.
- 17 Woo S, Park J, Lee JY, *et al.* CBAM: Convolutional block attention module. Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich: Springer, 2018. 3–19.
- 18 Liu Z, Lin YT, Cao Y, *et al.* Swin Transformer: Hierarchical vision Transformer using shifted windows. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE, 2021. 9992–10002.
- 19 Bodla N, Singh B, Chellappa R, *et al.* Soft-NMS-improving object detection with one line of code. Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017. 5562–5570.
- 20 Ren SQ, He KM, Girshick R, *et al.* Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137–1149. [doi: [10.1109/TPAMI.2016.2577031](https://doi.org/10.1109/TPAMI.2016.2577031)]
- 21 Liu W, Anguelov D, Erhan D, *et al.* SSD: Single shot multibox detector. Proceedings of the 14th European Conference on Computer Vision. Amsterdam: Springer, 2016. 21–37.
- 22 Howard A, Sandler M, Chu G, *et al.* Searching for MobileNetV3. Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019. 1314–1324.
- 23 Hou QB, Zhou DQ, Feng JS. Coordinate attention for efficient mobile network design. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 13708–13717.
- 24 Wang QL, Wu BG, Zhu PF, *et al.* ECA-Net: Efficient channel attention for deep convolutional neural networks. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 11531–11539.
- 25 Li YH, Yao T, Pan YW, *et al.* Contextual Transformer networks for visual recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(2): 1489–1500.

(校对责编: 孙君艳)