

DOI:10.19651/j.cnki.emt.2518634

# 基于改进 YOLOv10n 的滚动轴承表面缺陷检测 \*

王海群 陈晓宇 于海峰

(华北理工大学电气工程学院 唐山 063200)

**摘要:** 针对现有算法在滚动轴承表面缺陷检测时检测精度较低,模型参数量大,实时性较差等问题,提出一种改进 YOLOv10n 的滚动轴承表面缺陷检测算法。在主干网络上,利用 GhostConv、MSMHSA 模块和 CGLU 模块对 C2f 重新设计,构建 CGMC2f 模块,增强模型的特征提取能力,降低模型的参数量;在 SPPF 中,结合 GroupConv、Residual-Conv 和 Fusion-Conv 对 SPPF-LSKA 模块进行设计,构建新的 GRFSPPF-LSKA 模块,有效解决了信息丢失问题,提升模型的多尺度特征提取和融合能力;在 Neck 网络上,结合 BIFPN 的多尺度特征加权融合、MAF-YOLO 网络和 EMCAD 模块,构建 EMBS-FPN 网络,提高模型的检测精度,降低了模型的参数量,使模型轻量化;借鉴 Focal-loss 思想,优化 CIoU 损失函数为 Focaler-CIoU,加快模型的收敛速度。实验结果表明,改进后的 YOLOv10n 的 mAP 达到了 92.6%,相较于原模型提高了 2.7%,参数量降低了 0.45 M,计算量降低了 0.6 GFLOPs,更好的满足滚动轴承表面缺陷实时性检测要求。

**关键词:** YOLOv10n; 缺陷检测; 特征提取; 注意力机制; 损失函数

中图分类号: TP391.4; TN919.8 文献标识码: A 国家标准学科分类代码: 520.6040

## Surface defect detection of rolling bearings based on improved YOLOv10n

Wang Haiqun Chen Xiaoyu Yu Haifeng

(School of Electrical Engineering, North China University of Science and Technology, Tangshan 063200, China)

**Abstract:** Aiming at the problems of low detection accuracy, large model parameters, and poor real-time performance of existing algorithms in surface defect detection of rolling bearings, an improved YOLOv10n rolling bearing surface defect detection algorithm is proposed. On the backbone network, redesign C2f using GhostConv, MSMHSA module, and CGLU module, construct CGMC2f module to enhance the model's feature extraction capability and reduce the model's parameter count; in SPPF, the SPPF-LSKA module is designed by combining GroupConv, Residual-Conv, and Fusion-Conv modules to construct a new GRFSPPF-LSKA module, effectively solving the problem of information loss and improving the model's multi-scale feature extraction and fusion capabilities; on the Neck network, combining the multi-scale feature weighted fusion of BIFPN, MAF-YOLO network, and EMCAD module, an EMBS-FPN network is constructed to improve the detection accuracy of the model, reduce the number of model parameters, and make the model lightweight; drawing on the Focal-loss approach, optimize the CIoU loss function to Focaler-CIoU to accelerate the convergence speed of the model. The experimental results showed that the improved YOLOv10n achieved a mAP of 92.6%, an increase of 2.7% compared to the original model, a reduction of 0.45 M in parameter count, and a decrease of 0.6 GFLOPs in computational complexity, better meeting the real-time detection requirements of rolling bearing surface defects.

**Keywords:** YOLOv10n; defect detection; feature extraction; attention mechanism; loss function

## 0 引言

滚动轴承作为机械设备中减小摩擦的重要零部件,在实际生产过程中,由于工艺、材料和环境等多方面的影响,

其表面容易出现划痕、凹槽、擦伤等缺陷,这些缺陷不仅会削弱滚动轴承的性能和使用寿命,还会对整个机械设备造成严重损害。随着国家大力推进重点领域设备的高质量发展,对滚动轴承等基础机械零部件的质量要求变得格外严

收稿日期:2025-04-19

\* 基金项目:河北省自然科学基金(D2024209006)、河北省教育厅科学研究项目(QN2024147)资助

格,因此,滚动轴承表面缺陷检测研究具有重要意义。谢凌峰等<sup>[1]</sup>将轴承缺陷图像进行灰度化预处理,再利用阈值分割、颗粒分析和轮廓分析对轴承缺陷进行识别。马忠平<sup>[2]</sup>提出了一种基于机器视觉的轴承环形表面缺陷检测方法,该方法解决了传统检测方法成本高、误检、漏检率高的问题,但针对缺陷分类还有待完善。与传统检测算法相比,深度学习技术应用到滚动轴承缺陷检测领域显著提高了检测性能。目前,深度学习目标检测算法主要分为两阶段目标检测算法和单阶段目标检测算法,两阶段目标检测算法主要有 Fast R-CNN<sup>[3]</sup>、Faster R-CNN<sup>[4]</sup>、R-CNN<sup>[5]</sup>等;单阶段目标检测算法主要有 YOLO<sup>[6]</sup> 系列和 SSD<sup>[7]</sup> 等。石炜等<sup>[8]</sup>利用 ZF Net 卷积神经网络模型对列车轴承表面缺陷图像进行检测,并与传统的 Canny 算法进行比较,实验结果表明,Faster R-CNN 算法大幅提高了检测精度。吴飞等<sup>[9]</sup>设计轻量化双瓶颈结构模块 DBM,强化特征提取能力,降低模型的复杂度;在主干网络中引入 WTConv 作为下采样因子,增强模型对纹理和噪声的抗干扰能力;设计 Alpha-MPDIoU 联合损失函数,利用幂变换机制提高边界框的定位精度;采用辅助头训练策略,加快模型的收敛速度,提高了模型的检测精度。姚景丽等<sup>[10]</sup>提出一种改进 YOLOv8 的轻量级轴承缺陷检测算法 YOLO-SSW,降低模型参数量的同时,提高了模型的检测精度。通过在 YOLOv8 主干网络上引入 3-D 注意力机制 SimAM,增强

算法的特征提取能力;在颈部网络上嵌入 C2f-SCConv 模块,减少冗余参数,起到轻量化作用;并在颈部网络中添加小目标检测层,提高模型对小目标的检测能力;最后引入 WIoU 损失函数,加快网络的收敛速度。

上述研究成果虽用不同方法提高了轴承表面缺陷的检测性能,但在实际应用中仍存在挑战,不能很好的满足实时性检测要求。追求高检测精度时模型的参数量和计算量增多,这对资源受限设备的平稳运行不利;此外,光照不均、轴承表面反光会导致缺陷特征不明显,使缺陷成像不清晰;轴承堆叠致使缺陷被遮挡,缺陷无法被识别,都会提高误检、漏检的概率。为了更好的满足实时性检测需求,算法仍需不断优化,本文旨在降低模型参数冗余的同时提高对滚动轴承表面缺陷检测的检测精度,因此,提出一种改进 YOLOv10n 算法模型。

## 1 YOLOv10n 目标检测算法

YOLOv10<sup>[11]</sup>作为目前应用于目标检测任务中的新算法,提出一致的双重分配策略,消除 YOLO 系列的非极大值抑制训练依赖。YOLOv10 按模型尺度不同共分为 YOLOv10n、YOLOv10s、YOLOv10m、YOLOv10b、YOLOv10l、YOLOv10x 六种模型。为提高滚动轴承表面缺陷的检测精度,更好的满足实时性检测要求,本研究选择轻量化更好的 YOLOv10n 网络模型作为主体网络。YOLOv10n 的网络结构如图 1 所示。

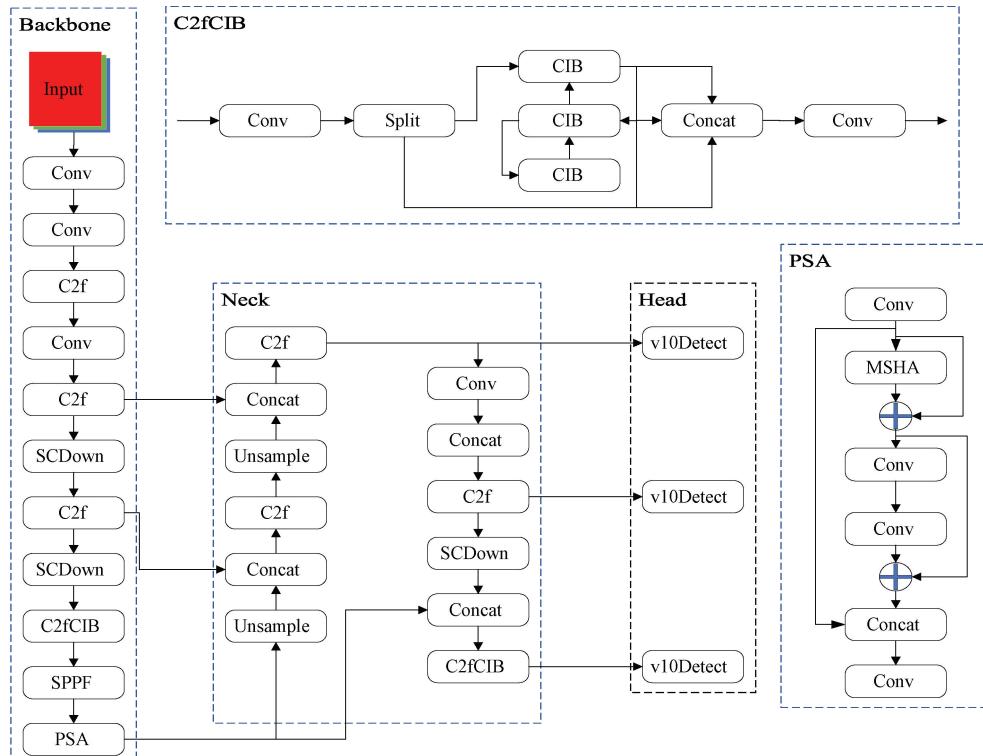


图 1 YOLOv10n 网络结构

Fig. 1 YOLOv10n network structure

YOLOv10n 网络主要由主干网络、颈部网络、检测头三部分组成, 主干网络在 YOLOv8 的基础上添加了空间与通道解耦下采样 SCDown ( spatial channel downsample)、C2fCIB 和部分自注意力 PSA (partial self-attention) 三种新的模块。SCDown 模块通过点卷积调整通道维度, 再通过深度卷积进行空间下采样, 实现空间通道解耦下采样; C2fCIB 模块中的 CIB ( compact inverted bottleneck) 结构采用了高效的深度卷积进行空间特征混合, 使用点卷积进行通道特征混合操作; PSA 模块是一种高效的部分自注意力机制, 通过逐步细化空

间注意力来提高模型的特征提取能力。颈部网络保留了 FPN 和 PANet 结构, 用于汇聚不同尺度的特征, 实现特征融合。检测头部分引入双重标签分配, 采用一对一 Head 和一对多 Head 同时训练, 提高模型的学习能力和检测的准确性。

## 2 改进 YOLOv10n 目标检测算法

为了提升滚动轴承表面缺陷检测精度, 使模型更加轻量化, 本文提出一种改进 YOLOv10n 的目标检测算法。改进后的 YOLOv10n 网络结构如图 2 所示。

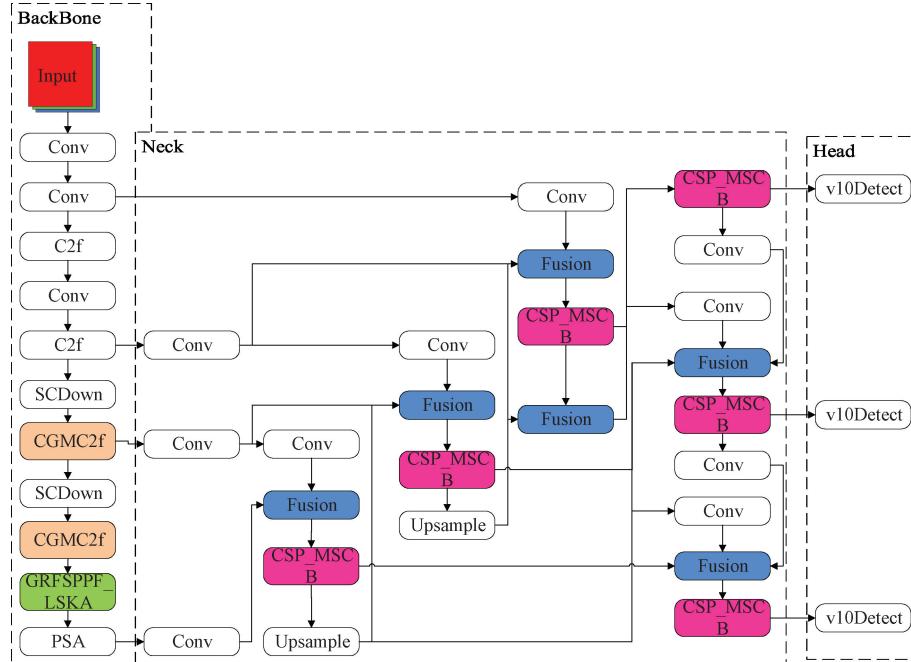


图 2 改进后的 YOLOv10n 网络结构

Fig. 2 Improved YOLOv10n network structure

主干网络中, 采用多尺度多头自注意力 (multi-scale multi-head self-attention, MSMHSA)<sup>[12]</sup> 模块、卷积门控线性单元 (convolutional gated linear unit, CGLU)<sup>[13]</sup> 和 Ghost 卷积 (GhostConv) 对 C2f 进行改进, 构建 CGMC2f 模块, 具有轻量化效果, 增强模型的特征提取能力; 在 SPPF 中引入大核分离卷积注意力机制 (large separable kernel attention, LSKA)<sup>[14]</sup>, 并结合分组卷积 (GroupConv)、残差连接模块 (Residual-Conv) 和特征融合模块 (Fusion-conv), 构建新的 GRFSPPF-LSKA 模块, 减少信息丢失, 提升模型的多尺度特征提取和融合能力; 在颈部网络中设计全新的高效多分支特征金字塔网络 (efficient multi-branch & scale FPN, EMBS-FPN), 进一步增强模型的多尺度特征提取和融合能力, 提高模型检测精度; 借鉴 Focal-loss 思想, 优化 CIoU 损失函数为 FocalCIoU, 增强边界框损失的拟合能力, 加快模型的收敛速度。

### 2.1 CGMC2f 模块

为增强滚动轴承表面缺陷检测模型的特征提取能力,

在主干网络中设计 CGMC2f 模块, 采用卷积门控线性单元 (CGLU) 和 Ghost 卷积对 MSMHSA 模块进行二次改进, 并将全新的 GhostMSMHSAs\_CGLU 模块引入到 C2f 模块中。其中 GhostMSMHSAs 模块结构如图 3 所示, CGLU 模块结构如图 4 所示。

MSMHSAs 模块是一种结合了多尺度特征提取和多头自注意力机制 (multi-head self-attention, MHSA) 的模块, 与传统多头自注意力 (MHSA) 机制不同的是, MSMHSAs 模块可以使用不同的卷积核 (如  $3 \times 3$ 、 $5 \times 5$ ) 或空洞率提取不同感受野的特征, 并将多头自注意力机制的多个头分配到不同尺度的特征上, 让不同注意力头关注不同尺度的特征, 增强模型对多尺度信息的融合能力, 平衡全局和局部信息的捕捉, 但也会导致其在多尺度处理时出现重复的卷积操作, 在特征拼接与融合时进一步增加通道维度, 进而增加计算量和参数量, 因此, 将 MSMHSAs 模块中的传统卷积替换为成 GhostConv 模块, 以减少其计算量和参数量。

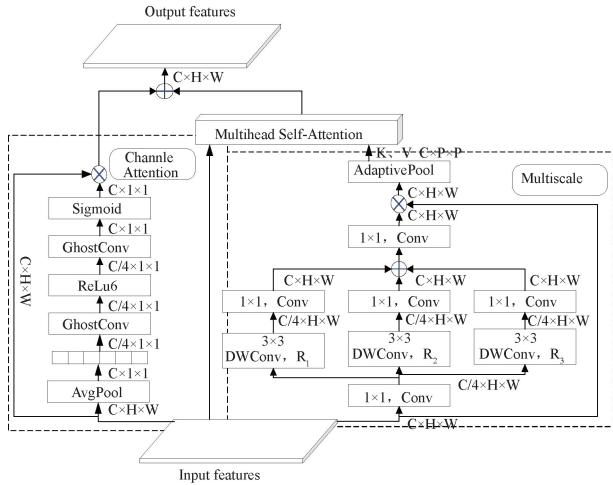


图 3 GhostMSMHSAs 模块

Fig. 3 GhostMSMHSAs module

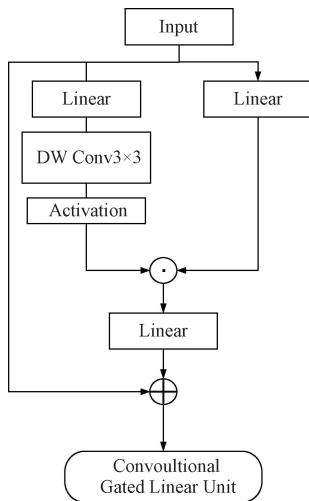
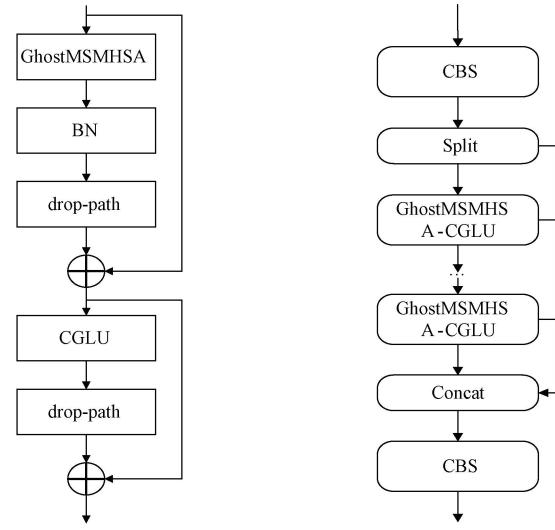


图 4 CGLU 模块

Fig. 4 CGLU module

如图 3 所示,输入特征首先经过一个 AvgPool 层进行下采样,然后通过一个  $C/4 \times 1 \times 1$  的 GhostConv 层,此时的通道数减少为原来的  $1/4$ ,从而降低了计算复杂度;在 GhostConv 层之后接 ReLU6 激活函数和另一个  $C/4 \times 1 \times 1$  的 GhostConv 层,这两次 GhostConv 操作共同作用于减少后的通道数上,生成更少的基础特征图,并通过线性变换生成更多的“幽灵”特征图,这样避免了直接使用大量参数来生成所有特征图,从而大大减少了参数量。

GhostMSMHSAs\_CGLU 模块结合了 MSMHSAs 模块、CGLU 线性单元模块和 Ghost 卷积,旨在同时利用自注意力机制强大的特征建模能力、Ghost 卷积的轻量化操作和门控机制对特征的动态调整能力,降低模型的计算量和参数量,增强模型在处理复杂数据时的特征表达和提取能力,其网络结构图如图 5(a)所示。



(a) GhostMSMHSAs\_CGLU 模块

(a) GhostMSMHSAs\_CGLU module

(b) CGMC2f 模块

(b) CGMC2f module

图 5 GhostMSMHSAs\_CGLU 和 CGMC2f 模块结构图

Fig. 5 GhostMSMHSAs\_CGLU and CGMC2f module  
structure diagram

先输入特征图,通过 GhostMSMHSAs 模块捕捉不同尺度的特征信息,然后由批量归一化 BN (Batch Normalization) 模块和 drop\_path 模块对捕捉到的特征信息进行归一化和正则化处理,再通过残差连接,将处理后的特征信息和原始特征信息相加,这有助于在特征增强的过程中保持原始信息的完整性,避免在深度网络中出现梯度消失或信息丢失的问题,使得模型能够更好地学习到滚动轴承表面缺陷的特征,由 CGLU 模块动态调整信息位置,增强模型对重要特征的表达,并将特征信息进一步正则化,最后将两次正则化处理的特征信息一起输出到下一环节。

采用 CGMC2f 模块,通过 C2f 和 GhostMSMHSAs 的结合,模块能够在轻量化的基础之上,有效的融合不同尺度的特征信息,使模型能更好的适应滚动轴承表面不同大小和形状的缺陷,提高检测的准确性;再由 CGLU 模块引入门控机制可以自适应的调整特征的位置,根据不同的输入特征动态的选择重要的信息进行保留和增强,这使得模块能更好地处理复杂的特征模式,提高模型的特征提取能力,CGMC2f 模块的网络结构图如图 5(b)所示。

CGMC2f 模块的具体计算公式:

$$R = \text{Conv}_{n+1}(\text{SiLU}(\text{Conv}_n(Y_{n-1}))) \quad (1)$$

$$F_{c2f} = \text{Conv}_{final}([X_1; R]) \quad (2)$$

$$F_{si} = \text{Conv}(F_{c2f}, K_{si}) \quad (3)$$

$$O_h = \text{softmax}\left(\frac{\mathbf{Q}_h \mathbf{K}_h^T}{\sqrt{d_k}}\right) \times \text{Linear}_v([F_{s1}, F_{s2}, \dots, F_{sm}]) \quad (4)$$

$$F_{MSMHSAs} = \text{Linear}_{out}([O_1, O_2, \dots, O_H]) \quad (5)$$

$$F_{output} = \text{Conv}_A(F_{MSMHSAs}) \odot \sigma(\text{Conv}_B(F_{MSMHSAs})) \quad (6)$$

$$Y = F_{shallow} = GhostConv(X) \quad (7)$$

$$F_{fusion} = Conv([Y]) \quad (8)$$

$$F_{output} = [F_{fusion}[:, : \frac{C}{2}] + F_{shallow}; F_{fusion}[:, \frac{C}{2}:]] \quad (9)$$

其中,  $SiLU$  是激活函数,  $Y_{n-1}$  是第  $n-1$  次卷积操作,  $R$  是残差特征,  $X_i$  是输入特征张量,  $K_{si}$  是不同大小的卷积核,  $F_{si}$  是不同尺度的特征图,  $Q_h$ 、 $K_h$  分别是查询 ( $Q$ )、键 ( $K$ ) 的第  $h$  个头,  $d_k$  是键的维度,  $H$  是头的数量,  $\sigma$  是 Sigmoid 函数,  $\odot$  表示逐元素相乘。

## 2.2 GRSPPF\_LSKA 模块

为增强滚动轴承表面缺陷检测模型对多尺度特征的提取和融合能力, 在空间金字塔池化模块 (SPPF) 上引入 LSKA 注意力机制, 并对 SPPF\_LSKA 模块进行二次改进, 改进后的 GRSPPF\_LSKA 模块结构图如图 6(a) 所示。

在 SPPF 模块进行池化操作之前, 添加一个  $1 \times 1$  卷积层, 对输入的特征图进行降维, 减少后续池化和拼接操作的计算量; 在拼接同尺度的池化结果后, 使用了 GroupConv 将输入通道和输出通道分别分成两组进行卷积, 降低模型的复杂度和计算量。

I-LSKA 网络结构图如图 6(b) 所示, 使用一个  $3 \times 3$  卷积层对输入特征图进行初步的特征提取和通道数调整, 降低模型的参数量; 引入轻量级注意力机制, 即使用全局平均池化层 (avg-Pool) 将特征图的空间维度压缩为  $1 \times 1$ , 得到每个通道的全局信息。然后通过全连接层对全局信息进行非线性变换。该全连接层包含两个  $1 \times 1$  卷积层, 中间使用 ReLU 激活函数增加非线性, 最后使用 Sigmoid 函数将输出值映射到  $[0, 1]$  区间, 得到每个通道的注意力权重, 增强模型的特征提取能力。

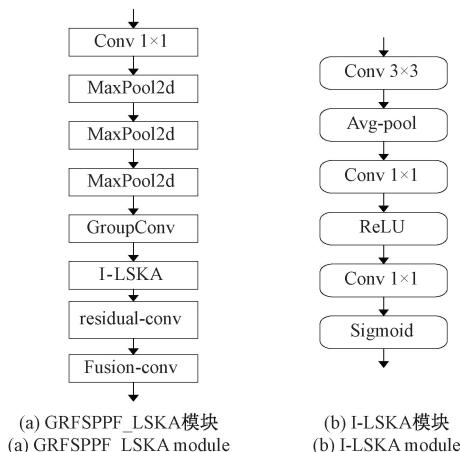


图 6 GRSPPF\_LSKA 模块和 I-LSKA 模块结构图

Fig. 6 Structure diagram of GRSPPF\_LSKA module and

I-LSKA module

在原 SPPF-LSKA 模块中, 经过多次最大池化和分组卷积后, 特征图已包含丰富的多尺度信息, 其维度和信息

量会发生较大变化, 导致反向传播过程中梯度不稳定, 可能在深层网络中面临梯度消失或爆炸的问题, 且对不同尺度的特征图, 缺乏有效的融合机制, 导致模型对不同尺度轴承缺陷目标的检测能力受限。引入 Residual-Conv 模块, 其结构图如图 7 所示, 通过多层卷积和跳跃连接, 可以进一步增强滚动轴承表面缺陷图像特征的表达能力, 同时也可以有效解决模型训练过程中出现的梯度消失或梯度爆炸问题。

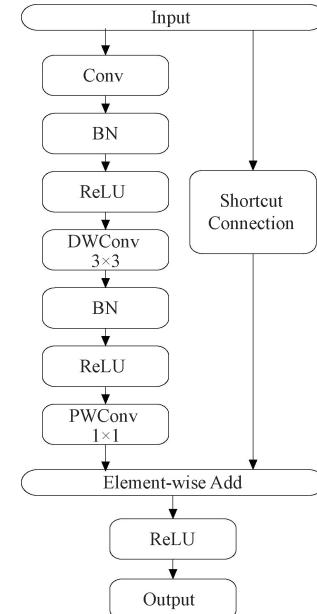


图 7 Residual-Conv 结构

Fig. 7 Residual-Conv structure

引入 Fusion-Conv 模块, 其结构图如图 8 所示, 通过将来自不同路径或不同尺度的特征图进行高效融合, 提高了模型对不同尺度滚动轴承缺陷目标的检测能力, 进一步增强了多尺度特征的表达和融合能力。

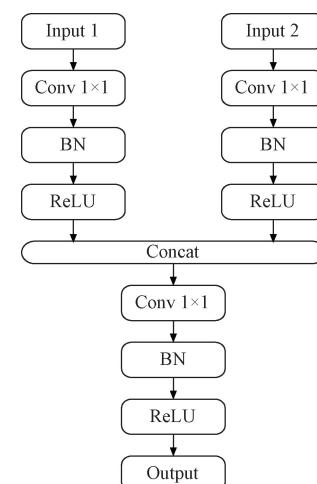


图 8 Fusion-Conv 结构

Fig. 8 Fusion-Conv structure

GRFSPPF\_LSKA 模块具体公式:

$$P_i = \text{MaxPool}(X, k_i, s_i) \quad (10)$$

$$X_{sppf} = \text{Concat}(X, P_1, P_2, \dots, P_n) \quad (11)$$

$$X_{sppf\_out} = \text{Conv}(X_{sppf}, W_{sppf}) + b_{sppf} \quad (12)$$

$$X_{lksa\_out} = X_{sppf\_out} \odot A \quad (13)$$

$$R = \text{Conv}_{1 \times 1}(X, W_r) + b_r \quad (14)$$

$$X_{concat\_final} = \text{Concat}(X_{lksa\_out}, R) \quad (15)$$

$$Y = \text{Conv}(X_{concat\_final}, W_{fusion}) + b_{fusion} \quad (16)$$

其中,  $P_i$  是池化后的特征图,  $k_i$  是池化核大小,  $s_i$  是步长,  $\text{Concat}$  是通道维度上的拼接操作,  $W_{sppf}$  是一个卷积层,  $b_{sppf}$  是偏置项,  $A$  是注意力图,  $\odot$  是逐元素相乘操作,  $W_r$  是卷积核,  $b_r$  是偏置项,  $R$  是残差分支的特征图,  $W_{fusion}$  是融合卷积层,  $b_{fusion}$  是偏置项。

### 2.3 EMBS-FPN 网络

输入滚动轴承表面缺陷图像, 经过 Backbone 网络输出 C2、C3、C4、C5 多个不同尺度的特征图, 并对每个特征图进行多尺度卷积(MSCB)操作, 生成更丰富的特征图, 然后由自上而下的路径, 将最高层的特征图(D5), 通过  $3 \times 3$

卷积操作进一步平滑特征图, 生成新的特征图 P5; 将 D5 特征图通过高效上采样模块(EUCB)上采样到与 D4 特征图相同的分辨率, 再将上采样后的 D5 特征图与 D4 特征图进行逐元素相加, 并对相加后的特征图进行卷积操作, 生成新的特征图 P4; 将 P4 上采样到与 D3 特征图相同的分辨率, 再将上采样后的 P4 特征图与 D3 特征图进行逐元素相加, 并进行卷积操作, 生成新的特征图 P3; 将 P3 上采样到与 D2 特征图相同的分辨率, 再将上采样后的 P3 特征图与 D2 特征图进行逐元素相加, 并进行卷积操作, 生成新的特征图 P2。其中 P2、P3、P4、P5 每个尺度通过多分支架构分成 4 个并行分支, 分支一进行普通卷积操作, 增强特征的抽象能力; 分支二进行深度可分离卷积操作, 减少计算量; 分支三进行空洞卷积操作, 扩大感受野, 分支四, 进行池化操作, 提取更宏观的特征。而在每个分支中, 引入 RepHELAN<sup>[15]</sup>模块, 通过动态大小卷积核实现自适应感受野, 动态调整特征图的分辨率。最后将每个尺度的多个分支特征图通过加权融合的方式进行聚合, 生成最终的多尺度特征图像。EMBS-FPN 网络结构如图 9 所示。

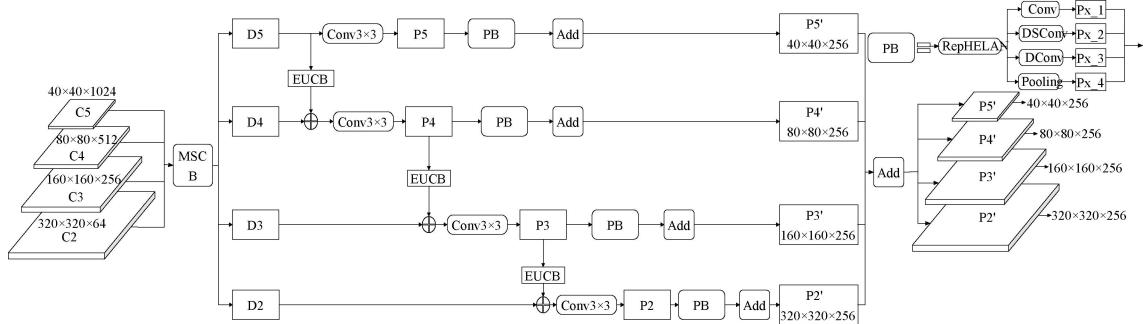


图 9 EMBS-FPN 网络结构

Fig. 9 EMBS-FPN network structure

EUCB<sup>[16]</sup>模块使用高效上卷积模块来逐步上采样当前阶段的特征图, 以匹配下一个连接层中特征图的维度和分辨率。EUCB 模块结构图如图 10(a)所示。

EUCB 模块具体公式:

$$\text{EUCB}(x) = C_{1 \times 1}(\text{ReLU}(\text{BN}(\text{DWC}(\text{Up}(x))))) \quad (17)$$

先使用比例因子为 2 的上采样 Up(·)来放大特征图, 再通过  $3 \times 3$  的深度卷积 DWC(·), 然后是 BN(·)和 ReLU(·)激活函数增强放大特征图, 最后使用  $1 \times 1$  卷积 C1  $\times$  1(·)来减少通道数以匹配下一个阶段。

MSCB<sup>[16]</sup>模块是引入的高效多尺度卷积模块, 用以增强级联扩展路径生成的特征, MSCB 模块的结构图如图 10(b)所示。

MSCB 模块的具体公式:

$$\text{MSCB}(x) = \text{BN}(\text{PWC}_2(\text{CS}(\text{MSDC}(\text{R6}(\text{BN}(\text{PWC}_1(x))))))) \quad (18)$$

MSCB 在多尺度上进行深度卷积, 并使用通道随机排

序来跨组对通道进行随机排序。具体来讲, 先使用逐点( $1 \times 1$ )卷积层 PWC<sub>1</sub>(·); 然后是批量归一化层 BN(·)和 ReLU 激活层 R6(·)扩展通道数; 再使用多尺度深度卷积 MSDC(·)来捕获多尺度和多分辨率背景; 由于深度卷积忽略了通道之间的关系, 因此使用通道随机操作来合并通道之间的关系, 接下来, 使用另一个逐点卷积 PWC<sub>2</sub>(·)后跟一个批量化归一层 BN 来转换回原始的通道数。其中, 多尺度深度卷积模块(MSDC)的结构图如图 10(c)所示。

MSDC<sup>[16]</sup>模块的具体公式为:

$$\text{MSDC}(x) = \sum_{ks \in KS} \text{DWCB}_{ks}(x) \quad (19)$$

$$\text{DWCB}_{ks}(x) = \text{R6}(\text{BN}(\text{DWC}_{ks}(x))) \quad (20)$$

$$x = x + \text{DWCB}_{ks}(x) \quad (21)$$

DWC<sub>ks</sub>(·)是内核大小为  $ks$  的深度卷积; BN(·)和 R6(·)分别是批量归一化和 ReLU6 激活函数。此外, 使用递归更新的输入  $x$ , 与之前的 DWCB<sub>ks</sub>(·)残差连接, 以实现更好的正则化。

RepHELAN 模块作为 EMBS-FPN 网络中的全局异

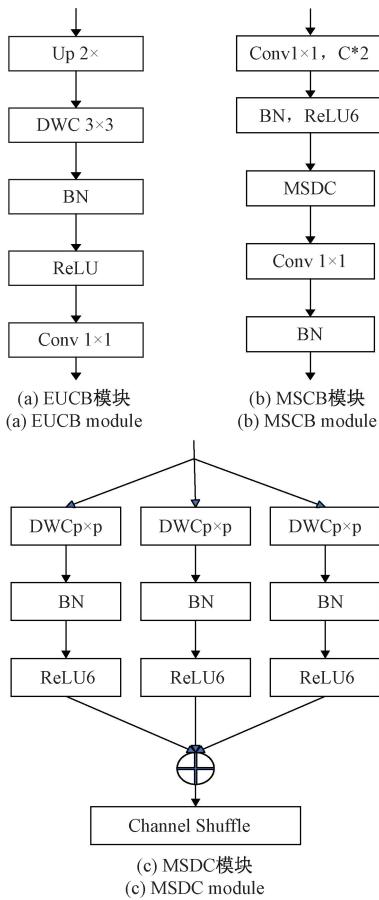


图 10 EMBS-FPN 各模块结构

Fig. 10 Structure of EMBS-FPN modules

构核选择机制,利用动态大小的卷积核来实现自适应感受野,更有效的提取不同尺度的特征。RepHELAN 模块结构图如图 11 左侧所示。

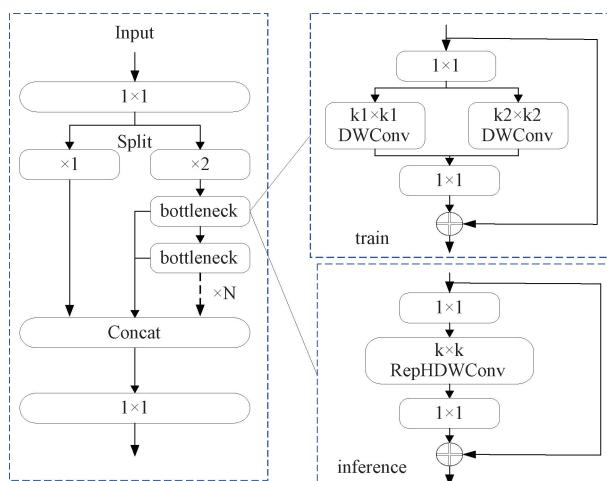


图 11 RepHELAN 和倒置瓶颈结构

Fig. 11 RepHELAN and inverted bottleneck structure

输入信息会经过  $1 \times 1$  卷积和 Split 操作,分成两个分支。一个分支保留原始信息,直接执行合并操作;另一个

分支,通过  $N$  个倒置瓶颈单元进行处理,由于高效特征聚合的机制,通过每个 Inverted Bottleneck 的分支和输出被保留并最终连接在一起。倒置瓶颈的具体结构如图 11 右侧结构所示,其中输入依次通过  $1 \times 1$  卷积扩大通道数,然后是  $k \times k$  RepHDWConv 操作,最后是  $1 \times 1$  点卷积以缩小通道数并补偿 DWConv 可能造成的信息丢失。

## 2.4 Focaler-CIoU 边界损失函数

为了加快模型的收敛速度,提高边界框的定位精度,解决训练样本不平衡问题,采用 Focaler-CIoU 损失函数作为本文的损失函数,其通过调整正负样本的权重,提高了模型识别稀有目标类别的能力,使模型更加关注难以分类的正样本,并减少相对容易分类的负样本的权重。使用线性间隔映射方法重构 IoU,这有助于提高边缘回归,其公式如下:

$$IoU^{focaler} = \begin{cases} 0, & IoU < d \\ \frac{IoU - d}{u - d}, & d \ll IoU \ll u \\ 1, & IoU > u \end{cases} \quad (22)$$

式中: $IoU^{focaler}$  是重构的 Focaler-IoU; $IoU$  是传统的交并比;[ $d, u$ ]是[0,1],通过调整  $d$  和  $u$  的值,可以使  $IoU^{focaler}$  更加关注滚动轴承表面缺陷检测的回归样本。其损失定义如下:

$$L_{Focaler-IoU} = 1 - IoU^{focaler} \quad (23)$$

将 Focaler-IoU 损失应用于 CIoU 边界框回归损失函数的具体公式如下:

$$L_{Focaler-CIoU} = L_{CIoU} + IoU - IoU^{focaler} \quad (24)$$

## 3 实验结果与分析

### 3.1 实验环境及参数设置

使用 Ubuntu 20.04 系统作为实验的操作系统。深度学习框架为 PyTorch 2.1.0;Python 版本为 3.10;CUDA 版本为 12.1;GPU 采用 NVIDIA GeForce R-TX4090,显存为 24 GB;CPU 为 15vCPU Intel(R) Xeon(R) Platinum 8474c。输入的图片尺寸为  $640 \times 640$ ,epochs 设置为 200,batch-size 设置为 32,IOU 阈值设置为 0.7,初始学习率设为 0.01。

### 3.2 数据集介绍

数据集在目标检测和识别中起着至关重要的作用,它在很大程度上影响着算法的检测效果。本文使用从工业现场采集的真实滚动轴承表面缺陷图像作为数据集,共 2 000 张图像,根据其表面特征将缺陷分为擦伤、划痕和凹槽 3 种常见的类型,然后通过调整图像亮度和增加高斯白噪声对数据集进行扩充,扩充后的数据集共 5 824 张图像,并按照 7:1 的比例分成 5 106 张训练集和 718 张验证集,其中,训练集中擦伤缺陷共 1 866 张图像,划痕缺陷共 1 750 张图像,凹槽缺陷共 1 490 张图像。

### 3.3 评价指标

本实验采用召回率(recall,R)、准确率(precision,P)、

均值平均精度 (mean average precision, mAP)、参数量 (parameters, Params) 和计算量 (giga floating point operations per second, GFLOPs) 作为算法检测精度的评价指标, 评价指标公式如式(25)~(27)所示。

$$P = \frac{T_p}{T_p + F_p} \quad (25)$$

$$R = \frac{T_p}{T_p + F_N} \quad (26)$$

$$mAP = \frac{\sum_{i=1}^n AP(i)}{n} \quad (27)$$

其中, 准确率  $P$  表示模型在实际检测任务中, 正确检测到的样本数量占总正确样本数的比例; 召回率  $R$  表示正确检测到的正样本数量占总正样本数的比例; 均值平均精度  $mAP$  表示所有类别的平均检测精度的平均值;  $TP$  表示正确预测的缺陷图像数量;  $FP$  表示将非缺陷图像错误预测成缺陷图像的数量;  $FN$  表示将缺陷图像预测成非缺陷图像的数量。

### 3.4 消融实验

为了验证各个改进点的有效性, 本研究以 YOLOv10n 模型为基线, 依次引入改进后的模块进行消融实验。实验结果如表 1 所示, 其中“√”表示采用对应的改进模块。

表 1 消融实验

Table 1 Ablation experiment

实验	CGMC2f	GRFSPPF-LSKA	EMBSFPN	Focaler-CIoU	P/%	R/%	mAP@0.5/%	Params/M	GFLOPs/G
基线					86.0	83.3	89.9	2.26	6.5
1	√				87.2	83.9	90.6	2.21	6.4
2		√			88.8	84.3	90.6	2.30	6.6
3			√		88.0	85.1	91.4	1.85	5.9
4				√	87.0	82.6	90.0	2.26	6.5
5	√	√			88.0	83.7	91.0	2.24	6.4
6	√	√	√		89.2	85.7	91.8	1.81	5.9
7	√	√	√	√	<b>90.0</b>	<b>87.7</b>	<b>92.6</b>	<b>1.81</b>	<b>5.9</b>

由表 1 的前 4 组实验结果可知, 单独引入对应的改进模块较基线相比均有不同程度的提升。与基线 YOLOv10n 相比, 将二次改进后的 GhostMSMHSACGLU 模块引入到 C2f 模块后, 模型的均值平均精度 (mAP) 提升了 0.7%, 模型的参数量减少了 0.05 M, GFLOPs 降低了 0.1 G; 在 SPPF 模块中引入 LSKA 注意力机制, 并对 SPPF\_LSKA 模块进行二次改进, 即 GRFSPPF\_LSKA 模块, 虽然模型的参数量增加了 0.04 M, 但模型的 mAP 提升了 0.7%; 采用 EMBS-FPN 网络, 模型的 mAP 提升了 1.5%, 模型的参数量降低 0.41 M, GFLOPs 降低了 0.6 G; 采用 Focaler-CIoU 边界损失函数, 在不增加模型计算成本的基础上, mAP 提升了 0.1%; 其中, 单独引入 EMBSFPN 网络相较与其他单独模块提升最优。通过后 3 组实验依次叠加改进策略来验证 4 种改进模块组合后对算法检测性能的影响, 实验 5 在原 YOLOv10n 模型中同时引入 CGMC2f 模块和 GRFSPPF-LSKA 模块, 模型的 mAP 提升了 1.1%, 参数量减少了 0.02 M, GFLOPs 降低了 0.1 G; 实验 6 在实验 5 的基础上引入 EMBSFPN 网络, 无论相较于实验 5 还是单独引入 EMBSFPN 网络, 模型的 mAP 都得到了提升, 参数量和 GFLOPs 均减少, 较基线的 mAP 提高了 1.9%, 参数量减少了 0.45 M, GFLOPs 降低了 0.6 G; 当综合以上 4 种改进方案后, 模型的 mAP 提高了 2.7%, 参数量减少了

0.45 M, GFLOPs 降低了 0.6 G, 有效降低了模型的参数量和计算成本, 使模型更加轻量化, 同时提高了模型的检测精度, 综上所述, 实验 7 的综合检测性能最佳, 验证各个改进方案的有效性。

### 3.5 SPPF 模块对比实验

为了探究在 SPPF 中引入不同注意力机制对模型检测性能的影响, 本文在 SPPF 中分别引入 CAA<sup>[17]</sup>、ELA<sup>[18]</sup>、EMA<sup>[19]</sup> (exponential moving average)、MLCA (multi-level channel attention)、MPCA (multipath coordinate attention)、SE<sup>[20]</sup> (squeeze-and-excitation)、LSKA 以及使用的二次改进后的 LSKA 注意力机制进行对比实验, 实验结果如表 2 所示。

其中, 引入 CAA 注意力机制相较于原 SPPF 模块, 模型的均值平均精度 (mAP) 降低了 1.9%, 参数量增加了 0.54 M, GFLOPs 提高了 0.5 G; 引入 ELA 注意力机制, 模型的 mAP 提高了 0.6%, 参数量增加了 1.84 M, GFLOPs 提高了 0.2 G, 相较于原 SPPF 模块, 虽然检测精度提升, 但模型的参数量大幅增加, 计算量 GFLOPs 也有所增大, 因此不适用于资源受限的边缘设备部署; 引入 EMA 注意力机制, mAP 提高了 0.1%, 参数量增加了 0.04 M, GFLOPs 提高了 0.3 G; 引入 MLCA 注意力机制, mAP 降低了 0.5%, 计算量和参数量都没有变化; 引入 MPCA 注意力机制, mAP 提高了 0.4%, 参数量增加了 1.31 M, GFLOPs 提

表 2 SPPF 模块对比实验

Table 2 SPPF module comparison experiment

模块	P/%	R/%	mAP@0.5/%	Params/M	GFLOPs/G
SPPF	86.0	83.3	89.9	2.26	6.5
+CAA	87.9	80.8	88.0	2.80	7.0
+ELA	87.6	84.6	90.5	4.10	6.7
+EMA	88.6	82.5	90.0	2.30	6.8
+MLCA	87.5	82.1	89.4	2.26	6.5
+MPCA	88.6	82.4	90.3	3.57	6.6
+SE	88.2	80.9	89.0	2.30	6.5
+LSKA	88.3	82.4	90.5	2.53	6.7
GRFSPPF_LSKA	<b>88.8</b>	<b>84.3</b>	<b>90.6</b>	<b>2.30</b>	<b>6.6</b>

高了 0.1 G, 虽然检测精度有所提升, 但参数量的大幅增加, 使模型的适用环境受到限制, 如: 不适合部署在资源受限的边缘设备上; 引入 SE 注意力机制, 模型的 mAP 降低了 0.9%, 参数量增大了 0.04 M, GFLOPs 没有变化; 引入 LSKA 注意力机制, 模型的 mAP 提高了 0.6%, 参数量增加了 0.27 M, GFLOPs 提升了 0.2 G, 而采用 GRFSPPF\_LSKA 模块, 模型的 mAP 提高了 0.7%, 参数量仅增加 0.04 M, GFLOPs 也仅提高了 0.1 G, 相较于其他几个模块, 该模块在提高模型检测精度的同时, 参数量和计算量

增加的最少, 因此, 采用二次改进后的 GRFSPPF\_LSKA 模块能更有效的提升模型的多尺度特征提取和融合能力, 从而提高模型的检测精度。

### 3.6 FPN 网络对比实验

为了探究不同的 FPN 网络对模型检测性能的影响, 本文分别引入 CGRFPN (context-guided spatial feature reconstruction FPN)、FreqFPN、GFPN<sup>[21]</sup>、MAFPN 以及使用的 EMBSFPN 网络作为 YOLOv10n 的颈部网络进行对比实验, 实验结果如表 3 所示。

表 3 FPN 网络对比试验

Table 3 FPN network comparison test

FPN 网络	P/%	R/%	mAP@0.5/%	Params/M	GFLOPs/G
YOLOv10n	86.0	83.3	89.9	2.26	6.5
CGRFPN	87.5	82.4	89.2	2.82	6.7
GFPN	89.9	83.3	91.2	3.18	8.2
MAFPN	88.6	84.4	90.7	3.05	7.9
FreqFPN	88.3	84.0	90.3	2.12	6.2
EMBSFPN	<b>88.0</b>	<b>85.1</b>	<b>91.4</b>	<b>1.85</b>	<b>5.9</b>

其中, 引入 CGRFPN 网络作为 YOLOv10n 的颈部网络, 相较于原模型的颈部网络, 模型的均值平均精度 (mAP) 降低了 0.7%, 参数量增加了 0.56 M, GFLOPs 提升了 0.2 G; 引入 GFPN 网络, 模型的 mAP 提高了 1.3%, 参数量增加了 0.92 M, GFLOPs 提高了 1.7 G, 引入 MAFPN 网络, 模型的 mAP 提高了 0.8%, 参数量增加了 0.79 M, GFLOPs 提高了 1.4 G, 虽然引入 GFPN 网络和 MAFPN 网络, 都使模型的检测精度有所提高, 但其参数量和计算量也大幅增加, 不符合模型的实时性检测需求; 引入 FreqFPN 网络, 模型的 mAP 提高了 0.4%, 参数量降低了 0.14 M, GFLOPs 降低了 0.3 G; 而采用 EMBSFPN 网络, mAP 提高了 1.5%, 参数量降低了 0.41 M, GFLOPs 降低了 0.6 G, 相较于其他几个 FPN 网络, 该 FPN 网络不仅提高了模型的检测精度, 而且降低模型的参数量和计算

量, 使模型更加轻量化, 因此, 采用 EMBSFPN 网络作为原模型的颈部网络, 可以在提高模型检测精度的同时, 更好的满足模型实时性检测要求。

### 3.7 主流算法对比实验

为了进一步验证本文改进后的算法在滚动轴承表面缺陷检测中的有效性和先进性, 将改进后的算法同其他几种主流算法进行对比实验, 包括 SSD、Faster R-CNN、YOLOv5n、YOLOv5s、YOLOv7-tiny、YOLOv8n、YOLOv10n、文献[9]以及文献[10]。实验结果如表 4 所示。

改进后的算法, 与 SSD 算法相比较, mAP 提高了 1.8%, 参数量降低了 22.87 M, 计算量 (GFLOPs) 降低了 337.87 G; 与 Faster R-CNN 算法相比, mAP 提高了 5%, 参数量降低了 39.32 M, GFLOPs 降低了 200.8 G; 与 YOLOv5n 算法相比, mAP 提高了 3.7%, 参数量增加了 1.3 M, GFLOPs 降低了 1.3 G。

表 4 主流算法对比实验

Table 4 Comparison experiments of mainstream algorithms

模型	R/ %	mAP@0.5/ %	Params/ M	GFLOPs/ G
SSD	94.4	90.8	24.68	343.77
Faster R-CNN	90.7	87.6	41.13	206.7
YOLOv5n	82.8	88.9	1.76	4.1
YOLOv5s	82.2	88.9	7.01	15.8
YOLOv7-tiny	81.1	89.5	6.02	13.2
YOLOv8n	82.9	90.0	3.01	8.1
YOLOv10n	83.3	89.9	2.26	6.5
文献[9]	83.7	89.7	2.60	7.4
文献[10]	87.5	91.5	2.70	11.8
Ours	<b>86.0</b>	<b>92.6</b>	<b>1.81</b>	<b>5.9</b>

0.05 M, GFLOPs 提高了 1.8 G;与 YOLOv5s 算法相比,

mAP 提高了 3.7%,参数量降低了 5.2 M,GFLOPs 降低了 9.9 G;与 YOLOv7-tiny 算法相比,mAP 提高了 3.1%,参数量降低了 4.21 M,GFLOPs 降低了 7.3 G;与 YOLOv8n 算法相比,mAP 提高了 2.6%,参数量降低了 1.2 M,GFLOPs 降低了 2.2 G;与 YOLOv10n 算法相比,mAP 提高了 2.7%,参数量降低了 0.45 M,GFLOPs 降低了 0.6 G;具备检测精度高和模型轻量化的特点,与文献[9]中提出的改进算法相比,mAP 提高了 2.9%,参数量降低了 0.79 M,GFLOPs 降低了 1.5 G;与文献[10]中提出的改进算法相比,mAP 提高了 1.1%,参数量降低了 0.89 M,GFLOPs 降低了 5.9 G。通过对这些主流算法的对比试验,验证了本文改进后算法在滚动轴承表面缺陷检测中的先进性和有效性。

### 3.8 可视化分析

为了更清晰的展示改进后算法的检测效果,将其与原 YOLOv10n 算法在现有数据集上进行检测,使用不同颜色的锚框区分不同缺陷的类别,可视化结果如图 12 所示。

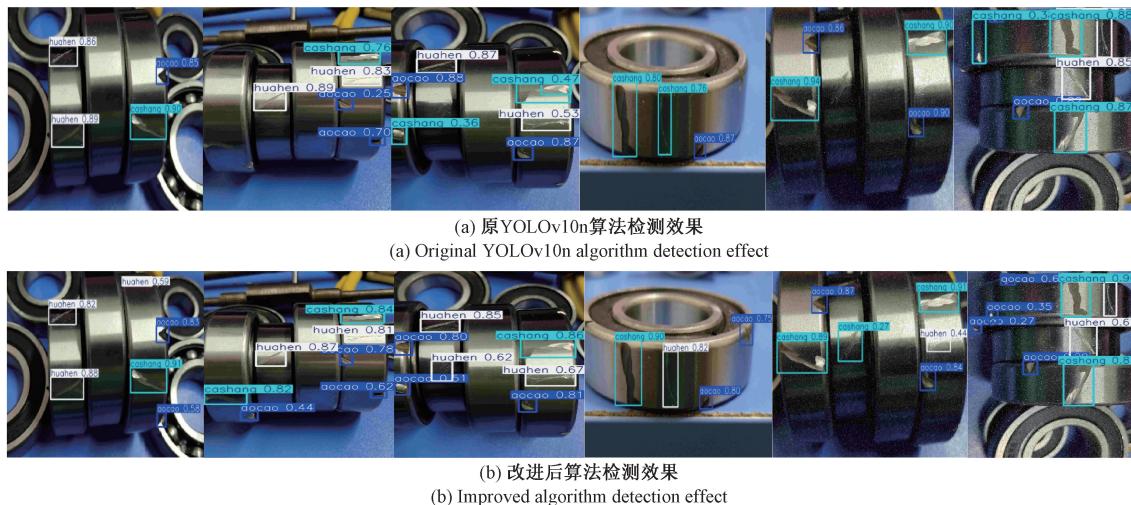


图 12 改进前后算法检测效果对比

Fig. 12 Comparison of algorithm detection performance before and after improvement

通过对比改进前后的检测效果图可以看出,原 YOLOv10n 算法在检测凹槽、擦伤、划痕缺陷时出现误检、漏检和检测精度偏低的问题;而改进后的算法在对各个缺陷进行检测时,提高了模型的检测精度,减少了误检和漏检情况的发生。对比结果表明,改进后的算法在滚动轴承表面缺陷检测中具有更好的检测效果。

## 4 结 论

针对在滚动轴承表面缺陷检测时出现的误检、漏检,模型参数量和计算量大的问题,本文提出了一种改进 YOLOv10n 的滚动轴承表面缺陷检测算法。在主干网络中,构建 CGMC2f 模块,使模型更加轻量化的同时,增强模型的特征提取能力;在 SPPF 模块中引入 LSKA 注意力机制,并对 SPPF-LSKA 模块进行二次改进,构建新的

GRFSPPF-LSKA 模块,增大感受野,有效增强了模型对多尺度特征的提取和融合能力,减少信息丢失问题;在颈部网络中,设计了 EMBSFPN 网络,进一步增强模型对不同尺度特征的提取和融合能力,并对多尺度特征进行自适应加权融合,提高了模型的检测精度;引入 Focaler-CIoU 边界损失函数,提高边界框的定位精度,加快模型的收敛速度。

## 参考文献

- [1] 谢凌峰,王娟.基于 LabVIEW 的轴承表面缺陷检测系统研究[J].湖州师范学院晚报,2023,45(10):49-56.
- XIE L F, WANG J. Research on bearing surface defect detection system based on LabVIEW [J]. Huzhou Normal University Evening News, 2023, 45(10): 49-56.

- [2] 马忠平. 一种基于机器视觉的轴承环形表面缺陷检测[J]. 中国科技信息, 2023(19):129-132.  
MA ZH P. A bearing ring surface defect detection based on machine vision [J]. China Science and Technology Information, 2023(19): 129-132.
- [3] GIRSHICK R. Fast R-CNN [J]. 2015 IEEE International Conference on Computer Vision. Santiago: IEEE, 2015: 1440-1448.
- [4] REN SH Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards realtime object detectionwith region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [5] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.
- [6] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]. IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 779-788.
- [7] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]. European Conference on Computer Vision. Cham: Springer International Publishing, 2016: 21-37.
- [8] 石炜, 李嘉楠, 张惠丽, 等. 基于 Faster R-CNN 算法的列车轴承表面缺陷检测研究[J]. 机床与液压, 2021, 49(11):103-108.  
SHI W, LI J N, ZHANG H L, et al. Research on surface defect detection of train bearing based on fast R-CNN algorithm[J]. Machine Tools and Hydraulics, 2021, 49(11): 103-108.
- [9] 吴飞, 范鹏主, 马一凡. 融合双瓶颈结构的轴承外圈缺陷检测算法[J]. 电子测量技术, 2025, 48(6):53-64.  
WU F, FAN P ZH, MA Y F. Algorithm for detecting outer ring defects of bearings using a dual bottleneck structure fusion [J]. Electronic Measurement Technology, 2025, 48(6):53-64.
- [10] 姚景丽, 程光, 万飞, 等. 改进 YOLOv8 的轻量化轴承缺陷检测算法[J]. 计算机工程与应用, 2024, 60(21): 205-214.  
YAO J L, CHENG G, WAN F, et al. An improved lightweight bearing defect detection algorithm based on YOLOv8[J]. Computer Engineering and Applications, 2024, 60(21): 205-214.
- [11] WANG AO, CHEN H, LIU L H, et al. YOLOv10: Real-time end-to-end object detection[J]. Advances in Neural Information Processing Systems, 2024, 37: 107984-108011.
- [12] WU H, HUANG P, ZHANG M, et al. CMTFNet: CNN and multiscale transformer fusion network for remote-sensing image semantic segmentation [J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 1-12.
- [13] SHI D. Transnext: Robust foveal visual perception for vision transformers [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 17773-17783.
- [14] LAU K W, PO L M, REHMAN Y A U. Large separable kernel attention: Rethinking the large kernel attention design in CNN [J]. Expert Systems with Applications, 2024, 236: 121352.
- [15] YANG Z, GUAN Q, ZHAO K, et al. Multi-branch auxiliary fusion yolo with re-parameterization heterogeneous convolutional for accurate object detection [C]. Chinese Conference on Pattern Recognition and Computer Vision (PRCV). Singapore: Springer Nature Singapore, 2024: 492-505.
- [16] RAHMAN M M, MUNIR M, MARCULESCU R. Emcad: Efficient multi-scale convolutional attention decoding for medical image segmentation[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 11769-11779.
- [17] CAI X, LAI Q, WANG Y, et al. Poly kernel inception network for remote sensing detection[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 27706-27716.
- [18] XU W, WAN Y. ELA: Efficient local attention for deep convolutional neural networks [J]. ArXiv preprint arXiv:2403.01123, 2024.
- [19] OUYANG D, HE S, ZHANG G, et al. Efficient multi-scale attention module with cross-spatial learning[C]. ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP). IEEE, 2023: 1-5.
- [20] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.
- [21] JIANG Y, TAN Z, WANG J, et al. GiraffeDet: A heavy-neck paradigm for object detection[J]. ArXiv preprint arXiv:2202.04256, 2022.

### 作者简介

王海群,硕士,副教授,主要研究方向为智能控制与应用、深度学习。

E-mail: wanghq0604@163.com

陈晓宇(通信作者),硕士研究生,主要研究方向为深度学习、目标检测。

E-mail: 1754336236@qq.com

于海峰,博士,主要研究方向为深度学习、图像处理。