

基于YOLOv8改进的轴承表面缺陷检测方法

于明源, 周景亮, 曾绍锋, 易思敏

(福建理工大学机械与汽车工程学院, 福建福州 350118)

摘要: 针对深度学习模型在轴承表面缺陷检测过程中漏检率高、模型复杂度高的问题, 提出了一种基于YOLOv8的缺陷检测改进算法。在主干网络中引入GSConv轻量化卷积模块, 用GSConv模块代替普通卷积, 在不影响模型精度的情况下, 减少模型的计算量; 引入CBAM卷积注意力模块, 通过改进网络特征提取技术, 提高了检测的准确性。实验结果表明, 改进的模型在自建轴承表面缺陷检测数据集上的准确率为92.6%, 较原模型准确率88.8%提高了3.8%; 在提升准确率的同时, 计算量也从8.2 GFLOPs减少到8.0 GFLOPs, 证明了改进后的模型对轴承缺陷检测的有效性。

关键词: 轴承表面缺陷检测; YOLOv8; 注意力机制; 深度学习; 缺陷检测

中图分类号: TP391.4

文献标志码: A

文章编号: 2097-3853(2024)03-0280-06

Improved bearing surface defect detection method based on YOLOv8

YU Mingyuan, ZHOU Jingliang, ZENG Shaofeng, YI Simin

(School of Mechanical and Automotive Engineering, Fujian University of Technology, Fuzhou 350118, China)

Abstract: Aiming at the problems of high missing rate and high model complexity of deep learning model in the process of bearing surface defect detection, an improved defect detection algorithm based on YOLOv8 was proposed. Firstly, the GSConv lightweight convolutional module is introduced into the backbone network, and the GSConv module is used to replace the ordinary convolutional module, which reduces the calculation amount of the model without affecting the accuracy of the model. Secondly, CBAM convolutional attention module is introduced to improve the detection accuracy by improving the network feature extraction technology. Experimental results show that the accuracy of the improved model on the self-built bearing surface defect detection dataset is 92.6%, which is 3.8% higher than that of the original model (88.8%). While the accuracy is improved, the computational cost is also reduced from 8.2GFLOPs to 8.0 GFLOPs, which proves the effectiveness of the improved model for the detection of bearing defects.

Keywords: bearing surface defect detection; YOLOv8; attention mechanisms; deep learning; defect detection

轴承是机械传动系统的重要零部件, 其质量将直接影响机械装备的使用寿命与可靠性。而在轴承制造过程中会造成划痕、凹坑、擦伤等^[1]各种缺陷, 因此缺陷检测在轴承生产过程中是不可或缺的^[2]。陈金贵^[3]、陈硕^[4]等人曾对传统的轴承缺陷检测进行改进, 但由于实际生产过程中轴承表面缺陷不同, 传统检测方法仍然存在效率低和可行性差的问题。

与传统检测方法相比, 深度学习技术在缺陷检测领域^[5]发挥着越来越重要的作用。Liu等^[6]利用图像处理和深度学习技术对轴承缺陷方法进行研究, 完成缺陷检测任务。袁天乐等^[7]提出基于YOLOv5的自动提取检测区域预处理方法, 并改进Transformer中的多头自注意力机制模块, 提高了模型训练时的收敛速度和检测精度。为了提高轴承环形表面缺陷的检测精度, 马忠平等^[8]提

收稿日期: 2023-12-11

基金项目: 福建省科技计划项目(2022L3014)

第一作者简介: 于明源(1999—), 男, 湖北仙桃人, 硕士研究生, 研究方向: 机器视觉。

通信作者: 曾绍锋(1974—), 男, 山西长治人, 副教授, 博士, 研究方向: 机械设计及理论、材料塑性加工、先进结构材料。

出了一种基于机器视觉的轴承环形表面缺陷检测方法,设计了轴承环形表面强反光特性的照明方案,提高了缺陷检测的准确率,但针对缺陷分类部分还有待完善。苗荣慧等^[9]提出了一种改进的轻量化YOLOv7模型的樱桃番茄果实成熟度检测方法,有效实现了模型轻量化和性能的平衡,为实际工业检测提供了新的思路。

YOLOv8模型在颈部网络中,因不能高效提取并融合前后图片特征信息,导致在小目标检测方面存在漏检等问题,而且由于模型参数量大,在实际工业检测中,检测率较低。因此,本文基于YOLOv8模型进行改进,引入轻量化卷积模块以

及注意力机制模块来提升模型检测准确率。

1 相关模型及优化

YOLOv8目标检测算法是YOLO系列模型中的最新版本,支持图像分类、物体检测和实例分割任务,该算法引入了全新的功能,提高了模型性能和灵活性,更加便于使用。

1.1 改进的YOLOv8模型

由于YOLOv8模型在轴承表面缺陷检测中存在模型精度与检测速度不平衡的问题,无法满足实际工业检测。因此,本文基于YOLOv8模型进行改进,改进后的网络结构如图1所示。

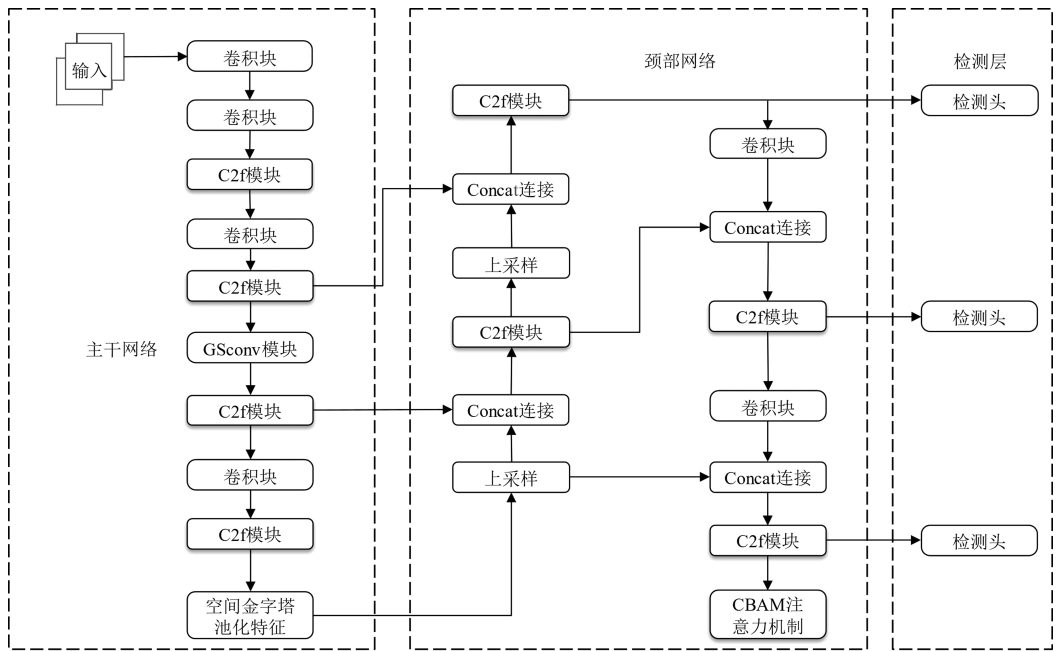


图1 YOLOv8网络改进结构图

Fig.1 Improvement structure of YOLOv8 network

1.2 GSConv 模块

GSConv (group shuffle convolution) 模块是一个轻量化卷积模块^[10],结构如图2所示。输入一个普通卷积的下采样,将传统卷积模块生成的信息渗透到深度可分离卷积生成信息的每个部分,输入通道数为 C_1 ,输出通道数为 C_2 。经过一个标准卷积后通道数变为 $C_2/2$,通过深度可分离卷积处理后,通道数不变,然后将第一次卷积后的结果与深度可分离卷积后的结果进行 Concat 连接和混洗。在最后的混洗操作中,通道信息被均匀打乱,将图像缺陷融合,提高图像特征的表达能。本文引入轻量级 GSConv 卷积模块,用 GSConv 模块替代标准卷积,缩减特征通道数,使

模型的准确性和速度之间达到平衡,从而降低模型的参数与计算。

GSConv 对应的时间复杂度为:

$$T_{\text{GSConv}} = O\left(W \times H \times K_1 \times K_2 \times \frac{C_2}{2} (C_1 + 1)\right) \quad (1)$$

式中, W 为输出特征图的宽度; H 为输出特征图的高度; K_1 、 K_2 分别为卷积核大小; C_1 为每个卷积核的通道数量,与输入特征图的通道数相等; C_2 为输出特征图的通道数。每个网络层中都使用它,模型的网络会更深,计算更加复杂,因此本文选择仅在主干网络中引入 GSConv 模块,以减少计算量。

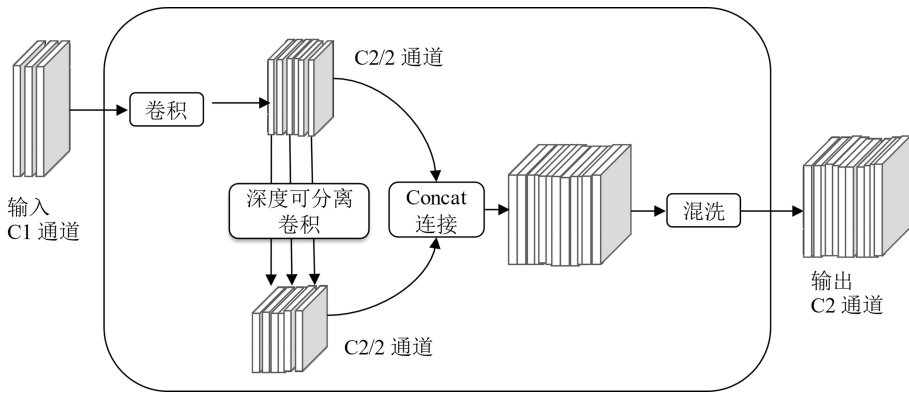


图 2 GSConv 卷积模块

Fig.2 GSConv convolution module

1.3 CBAM 注意力机制

在网络中引入 CBAM (convolutional block attention module) 卷积注意力机制^[11], 该模块分别使用全局最大和全局平均两种池化, 通过全连接网络对这两种池化的结果进行非线性变换, 将变换后的结果相加并经过激活函数得到注意力权重。CBAM 模块给定中间特征图后, 通过推断通道和空间注意力图, 将其与输入特征图相乘实现自适应特征优化。该模块通过提高网络的局部特征提取能力来提高模型的性能, 从而增强模型泛化能力, 取得更好的效果。结构示意图如图 3 所示。

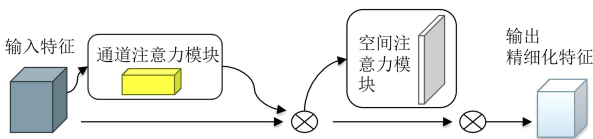


图 3 CBAM 注意力模块

Fig.3 CBAM attention module

通道注意力机制压缩空间维度, 生成一维矢量并操作, 结合平均池化和最大池化, 如图 4 所示。空间注意力模块保持空间维度不变, 压缩通道维度, 关注目标位置信息, 如图 5 所示。通过通道和空间上的注意力相结合, 提高模型计算效率。

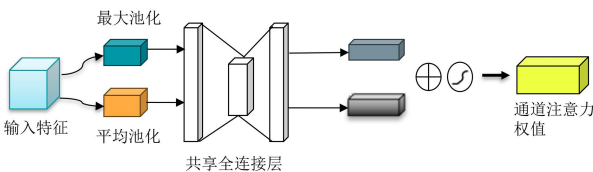


图 4 通道注意力模块

Fig.4 Channel attention module

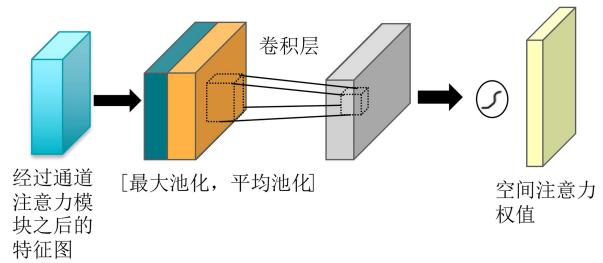


图 5 空间注意力模块

Fig.5 Spatial attention module

2 实验与结果分析

2.1 实验环境

实验在 Windows 11 系统下进行, 服务器的 GPU 为 NVIDIA GeForce RTX3060 Laptop, 内存为 16 GB。所使用的软件配置包括 PyTorch 2.1、Python 3.9、Anaconda、CUDA 11.8、cuDNN 开发库以及 Pycharm 编译器。

训练时选用模型自带的默认参数, 即采用随机梯度下降优化器, 动量参数为 0.937, 学习率为 0.01, 选择交叉熵损失, 对其他参数微调, 总迭代次数 300 次。

2.2 数据集构建

数据集于实验室环境下拍摄, 将轴承表面缺陷划分为划痕、擦伤和凹槽 3 类。涵盖了单一缺陷图片及多缺陷图片, 为了扩充数据集以及提高实验的鲁棒性, 随机对部分图像进行旋转、翻转、模糊、增加噪声和改变亮度等处理, 得到轴承缺陷图片 1 500 张。使用 LabelImg 软件对轴承缺陷图片标注。图 6 为部分数据集的样例图片。

以 YOLOv8 为基准模型。将自制的数据集按



图 6 部分数据集图片

Fig.6 Part of dataset images

9 : 1 的比例划分为训练集和验证集,并进行模块消融实验,最终确定最优模型。

2.3 评价指标

选取准确率 A 、精确率 P 以及损失值等指标,用以全面评估检测模型的性能。准确率越大表示模型整体精度越高。在讨论轻量化与性能相平衡的问题中,也要考虑到模型计算量,因此还需要引入浮点运算次数作为参数指标,浮点运算次数越小表示模型计算量越小。各指标计算公式为:

$$A = \frac{TP+TN}{TP+FP+TN+FN} \quad (2)$$

$$P = \frac{TP}{TP+FP} \quad (3)$$

式中,TP 为正确预测的正样本数量;TN 为正确预测的负样本数量;FN 为误判的负样本数,FP 为误判的正样本数。

2.4 模型训练

经过 300 次迭代训练后,得出原模型准确率为 88.8%,改进后模型的准确率为 92.6%,提高了 3.8%。可以看出从训练开始,改进后的模型相对于原模型一直都有更好的表现。如图 7 所示。

从训练开始到结束,模型的分类损失值逐步下降。由图 8 可知,改进后模型分类损失曲线

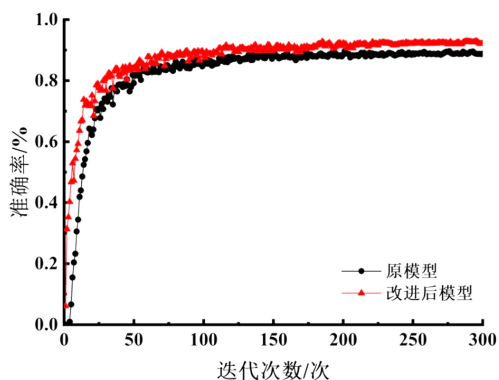


图 7 改进前后验证集精确率曲线对比

Fig.7 Comparison of accuracy curves of validation set before and after improvement

相较于原模型分类损失曲线不仅能更快收敛,而且在整个训练过程中,损失值更小,效果更优。

在进行 300 次训练后,得出每类缺陷检测准确率如图 9 所示。在凹槽、擦伤、划痕 3 类缺陷中,改进后的模型对比原模型分别提高了 3.7%、2.8%、4.7%。

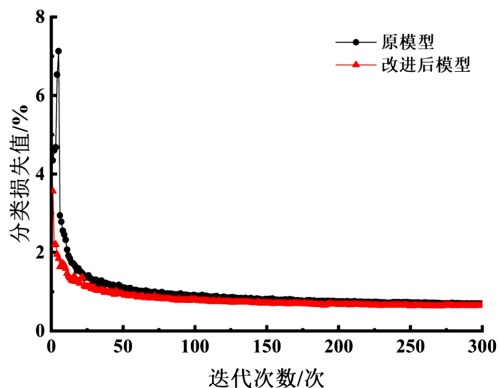


图 8 改进前后损失值对比

Fig.8 Comparison of loss values before and after improvement

2.5 消融实验

消融实验结果如表 1 所示。分析可知,加入 GSConv 模块后,计算量也从 8.2 GFLOPs 减少到 7.9 GFLOPs,在模型性能上准确率有 1.4% 提升,尽管模型性能提升不明显,但是证明了 GSConv 在 YOLOv8 上轻量化的可行性;为了提高模型对小目标缺陷的识别能力,加入 CBAM 注意力机制模块使得模型各类指标都有所提升,准确率提高了 3.2%,但是模型的计算量有所增加;通过同时将两种模块加入 YOLOv8 模型,检测准确率提高了 3.8%,而计算量却降低到了 8.0 GFLOPs,使得模

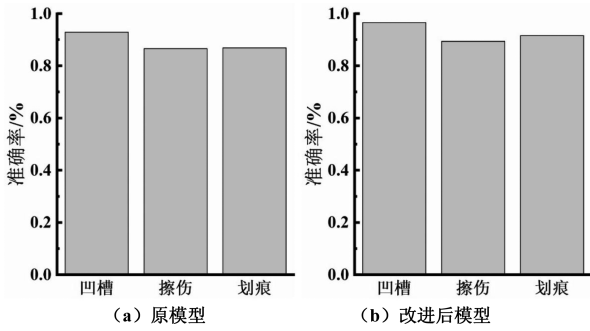


图 9 改进前后各类缺陷的精度曲线

Fig.9 Accuracy curves of various defects before and after improvement

型在性能提升的同时变得更加轻量化。说明改进后的 YOLOv8 模型有更优的检测效果。

表 1 消融对比实验

Tab.1 Ablation comparison experiments

实验内容	准确率/%	计算量/GFLOPs	迭代次数/次
YOLOv8(基线)	88.8	8.2	300
YOLOv8+GSCConv	90.2	7.9	300
YOLOv8+CBAM	92.0	8.1	300
YOLOv8+GSCConv+CBAM	92.6	8.0	300

2.6 实验分析

将本文改进模型分别与 Faster RCNN、YOLOv5、YOLOv7-tiny 在同一数据集下进行对比实验。结果曲线如图 10 所示。

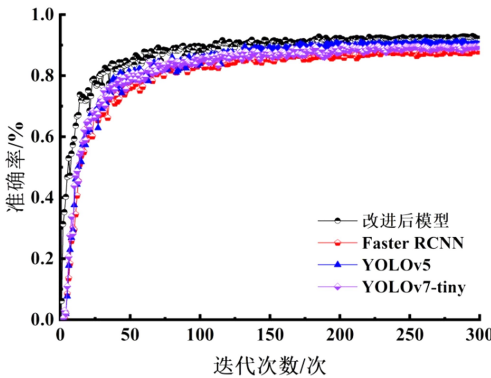


图 10 不同模型精度对比

Fig.10 Precision comparison of different models

本文基于 YOLOv8 改进后的模型相较于 Faster RCNN 和 YOLOv5、YOLOv7-tiny 模型,检测准确率分别高出 4.5%、2.4%、3%,可以看出,在训

练过程中,改进后的模型相较于其他模型,都有更好的表现。

改进后的模型对凹槽识别率最高,达到 96.6%,如表 2 所示。由表 2 可知,这几类算法对凹槽缺陷的检测效果都较好,体现出模型的良好性能。在擦伤和划痕这两类缺陷上,所选模型的检测效果都不是很好,但本文改进后的模型在这两类缺陷检测中,擦伤识别率提升明显,并且划痕识别率达到了 90% 以上。

表 2 不同模型准确率对比

Tab.2 Accuracy comparison of different models

实验内容	凹槽识别率	擦伤识别率	划痕识别率	准确率 %
原模型	92.9	86.6	86.9	88.8
Faster RCNN	91.0	86.8	86.4	88.1
YOLOv5	94.6	87.0	89.1	90.2
YOLOv7-tiny	93.3	88.0	87.6	89.6
改进后模型	96.6	89.4	91.6	92.6

2.7 检测效果可视化

通过上述实验可以得出,YOLOv8 在轴承表面缺陷检测中存在一定不足,加入注意力机制模块后,改进后的模型对轴承表面缺陷识别效果得到了有效提升,如图 11、图 12 所示,原模型对轴承缺陷有漏检现象,改进后的模型可以完整识别出缺陷,更加验证了改进后模型的有效性。

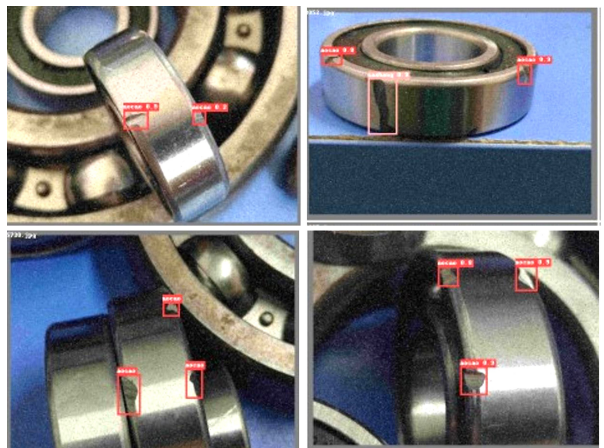


图 11 原模型识别效果

Fig.11 Recognition effect of original model

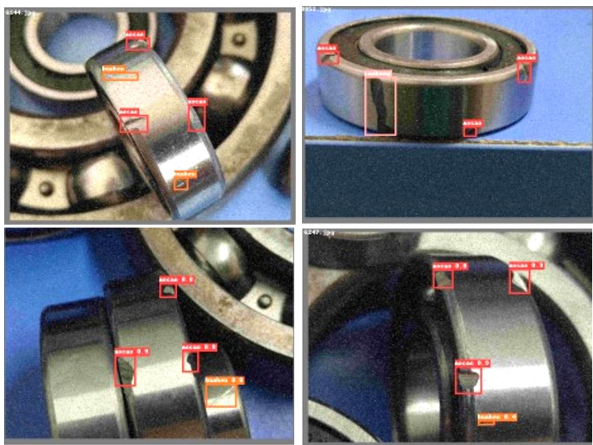


图12 改进后模型识别效果

Fig.12 Recognition effect of improved model

3 结论

1)在主干网络中引入 GSCConv 轻量化卷积模块,替代原模型的标准卷积,在不影响精度的情况下,减小计算量,提升检测速度。

2)在颈部网络添加一层 CBAM 卷积注意力,提高对图片缺陷的提取能力,从而提高检测精度。

3)将改进模型与 Faster RCNN、YOLOv5、YOLOv7-tiny 进行对比实验可知,本文改进模型准确率最高,进一步证明了改进模型的有效性,提高了实际工业中对轴承表面缺陷检测的准确率。

参考文献:

- [1] 袁野,张群. 轴承套圈裂纹缺陷分析[J]. 金属世界,2021(3):49-51.
- [2] 程锦锋,方贵盛,高惠芳. 表面缺陷检测的机器视觉技术研究进展[J]. 计算机应用研究,2023,40(4):967-977.
- [3] 陈金贵,陈昊,张奔. 基于改进 Niblack 算法的轴承滚子表面缺陷检测[J]. 组合机床与自动化加工技术,2018(12):82-85,97.
- [4] 陈硕,林志敏,吴岳彬,等. 轴承套圈端面缺陷在线视觉检测的研究与实现[J]. 轴承,2022(2):48-54.
- [5] 罗会兰,陈鸿坤. 基于深度学习的目标检测研究综述[J]. 电子学报,2020,48(6):1230-1239.
- [6] LIU Y,ZHANG C S,DONG X J. A survey of real-time surface defect inspection methods based on deep learning[J]. Artificial Intelligence Review,2023,56(10):12131-12170.
- [7] 袁天乐,袁巨龙,朱勇建,等. 基于改进 YOLOv5 的推力球轴承表面缺陷检测算法[J]. 浙江大学学报(工学版),2022,56(12):2349-2357.
- [8] 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.
- [9] 苗荣慧,李志伟,武锦龙. 基于改进 YOLO v7 的轻量化樱桃番茄成熟度检测方法[J]. 农业机械学报,2023,54(10):225-233.
- [10] 赵佰亭,吴俊东,贾晓芬. 融合特征增强的轻量化罐道缺陷检测算法[J]. 电子测量与仪器学报,2023,37(6):159-168.
- [11] WOO S,PARK J,LEE J Y,et al. CBAM:convolutional block attention module[C]//European Conference on Computer Vision. Cham:Springer,2018:3-19.

(责任编辑:陈雯)