

文章编号 1004-924X(2024)11-1746-13

融合多尺度特征的蜗杆表面缺陷检测

王 蕾, 郭文平*, 陈欣慰, 夏 珉

(华中科技大学 光学与电子信息学院, 湖北 武汉 430070)

摘要: 为了改善蜗杆齿顶表面缺陷检测依赖人工检测, 检测效率低下以及检测成本高昂的现状, 对基于机器视觉的自动化检测方法展开研究。设计了蜗杆缺陷采集系统, 针对不同缺陷发生率不同提出了数据增广策略。在 YOLOv7 的基础上改进算法, 针对不同缺陷的尺寸分布差异, 引入渐进特征金字塔重构颈部网络, 提升模型的多尺度特征融合能力; 为减少蜗杆非缺陷位置对检测结果的干扰, 引入注意力机制进一步加强模型的缺陷关注能力; 最后改进回归损失函数为 SIOU, 在网络训练中加入考虑标注框和预测框的方向, 进一步提升网络的检测精度。通过对比消融实验证明了上述改进的有效性。在参数量下降 20.7% 的情况下, 文本所提出的算法相较于 YOLOv7 算法精度提升了 3.3%; 和 YOLOR, YOLOv5m 等多种算法相比, 本算法的检测性能最优, 基本满足蜗杆表面缺陷自动化检测需求。

关键词: 机器视觉; 表面缺陷; YOLOv7; 特征金字塔

中图分类号: TP394.1; TH691.9 **文献标识码:** A **doi:** 10.37188/OPE.20243211.1746

Worm surface defect detection with fusion of multi-scale features

WANG Lei, GUO Wenping*, CHEN Xinwei, XIA Min

(School of Optical and Electronic Information, Huazhong University of Science and Technology,
Hubei 430070, China)

* Corresponding author, E-mail: wpguo@hust.edu.cn

Abstract: To tackle the challenges of reliance on manual inspection, low detection efficiency, and high costs in detecting surface defects on worm gear teeth, automated methods utilizing machine vision were re-searched. A defect collection system was designed to capture worm gear defects, and data augmentation strategies were introduced to handle varying defect occurrence rates. Enhancements were made to the YOLOv7 algorithm. Firstly, to address the differences in defect size distribution, a progressive feature pyramid was incorporated to reconstruct the neck network, improving the model's multi-scale feature fusion capability. Secondly, an attention mechanism was added to minimize interference from non-defective areas and bolster the model's focus on defects. Lastly, the regression loss function was modified to SIOU, and orientation consideration was included during network training to boost detection accuracy. Ablation experiments demonstrated the effectiveness of these improvements. With a 20.7% reduction in parameter count, experimental results show that the proposed algorithm achieves a 3.3 percentage point increase in accuracy compared to the YOLOv7 algorithm. Additionally, when compared to other algorithms like YOLOR and YOLOv5m, this algorithm provides optimal detection performance, effectively meeting the requirement

收稿日期: 2024-01-08; 修订日期: 2024-02-20.

基金项目: 湖北省科技厅重点研发计划(重点项目)(No. 2020BAA019)

for automated detection of surface defects in worm gears.

Key words: machine vision; surface defect detection; YOLOv7; feature pyramid

1 引 言

《中国制造 2025》行动纲领中明确指出需要加速推进工业化和信息化的深度融合^[1]。而工业的发展离不开对工业产品质量的严格把控,其中工业表面缺陷检测旨在发现各种工业制品的外观瑕疵,可以阻止带有缺陷的工业品流入市场。小模数蜗杆上面有一个或多个螺旋齿,可以和涡轮啮合形成传动结构^[2],常用在航天、轮船、汽车等设备中,是工业生产中重要的机械基础零件。蜗杆在生产加工过程中,难免会产生裂纹、凹坑等缺陷将直接降低蜗杆的机械性能,应用中会埋下巨大的安全隐患,因此,蜗杆表面缺陷检测成为工业界十分重视的问题。目前,小模数蜗杆产线上的缺陷检测主要依赖人工目测的方式,检测结果容易受主观因素的影响,并且人眼容易疲劳也会影响检测结果,无法长时间重复性的工作,从而极大地制约了蜗杆生产过程的智能化发展。

在早期,与蜗杆形状类似的螺纹钢的无损检测方法有涡流检测法、漏磁检测法等。涡流检测法是通过涡流电磁感应变化来判定螺纹表面是否存在缺陷,相比于螺纹钢,小模数蜗杆的自身尺寸和缺陷尺寸非常小,并不适合这种检测方式,而待测蜗杆并不存在磁性,也不适合漏测检测法。基于机器视觉的检测方式因为其非接触式测量、成本低、部署便捷而得到广泛的应用。2018 年^[3],辛露等针对蜗杆设计了线阵采集系统,使用形态学处理提取特征后,利用支持向量机进行缺陷分类实现对小模数蜗杆表面缺陷的检测。该检测是利用传统图像处理进行定制化检测,在普适性上表现不足。2012 年,ALexNet^[4]网络在图像分类任务中表现出色,能够抽象提取特征、普适性更强的深度学习在图像处理领域强势崛起。如今基于深度学习的目标检测已经相当成熟,主流的深度学习检测框架有以 Faster R-CNN 为代表的两阶段网络和以 YOLO 系列网络为代表的单阶段网络。其中,单阶段网络则可以实现端到端的检测,直接对图片特征进行提取,

可以实现在线、实时检测。乔健^[5]等在 YOLOX 的基础上融入新的注意力机制,并应用在金属圆柱形锅的表面缺陷检测上,成功识别了 6 种不同形态的缺陷。李闻^[6]等提出了一种改进的 YOLOv3 网络的 PCB 缺陷检测算法,对残差模块、聚类算法进行改进,并引入 SE 注意力机制,实验结果表明,改进算法在 PCB 缺陷检测上有较好的表现,检测时间为 0.139 s。2023 年,YOLOv4 的作者再次提出 YOLOv7^[7],提出了 E-ELAN 结构和 MP 模块,以及一些训练技巧,是当时目标检测领域的最优检测模型。

针对人工目测的低效和传统图像处理的普适性不足,本文基于 YOLOv7 针对蜗杆表面缺陷检测提出改进算法,并通过实验验证了算法的检测效力,该算法能够实现蜗杆表面缺陷的快速高精度检测。

本文设计蜗杆表面缺陷采集系统,制作蜗杆缺陷数据集,并针对蜗杆不同类别表面缺陷发生率不同导致缺陷样本数量不平衡,提出了数据集扩充策略。针对不同缺陷的尺寸分布差异明显,对 YOLOv7 的颈部网络进行改进,引入渐进特征金字塔重构颈部网络,提升模型的多尺度特征融合能力。引入注意力机制进一步加强模型的特征提取能力,让模型更加关注缺陷位置对结果的贡献。将损失函数中回归损失函数更换为 SIOU,让网络训练时考虑标注框和预测框的方向,进一步提高了检测精度。

2 AFP-YOLO 模型

单阶段的目标检测网络可以实现端到端的检测,从 YOLOv3^[8]开始基本奠定了由背部特征提取网络、颈部特征融合网络和检测头三部分组成的网络格局。本文在 YOLOv7 的基础上进行改进,提出基于渐进特征金字塔的网络(AFP-YOLO)。首先,针对不同缺陷的尺寸分布差异明显,使用渐进特征金字塔网络(Asymptotic Feature Pyramid Network,AFPN)实现特征融合,通过不同层特征

进行反复融合实现网络传输中同时保留抽象特征和细节特征,提升模型的多尺度特征融合能力。其次,在 Backbone 插入高效多尺度注意力机制(Efficient Multi-Scale Attention,EMA),引导网络提取特征时关注缺陷区域。并且将回归损失函数替换为 SIOU,考虑标注框和预测框之间不匹配的方向,避免模型陷入振荡,进一步提升了蜗杆缺陷检测的准确性。

本文提出的网络模型的结构如图 1 所示,使用 CSPDarkNet53 针对输入图片进行下采样完成特征提取。分别取 8 倍下采样、16 倍下采样和 32 倍下采样之后的输出作为 Neck 网络的三层输入。

三层 Neck 输入后经过普通卷积或者 SPPCSPC 进行通道整合,较低的两层先进行融合,而后加入高层特征,并且每次特征融合时都对输入的特征进行反复交叉融合。金字塔同样包含三层输出以适应多尺度检测的需求,接入检测头。图中还标注了可以插入注意力机制的位置,经过消融实验验证最终选择在位置 1,2 插入 EMA。

2.1 渐进特征金字塔

Neck 网络的主要作用是融合不同下采样层的特征来实现多尺度检测。2016 年,Liu^[9]等从图像金字塔中获得灵感构建了特征金字塔网络 (Feature Pyramid Network,FPN),采用自顶向下

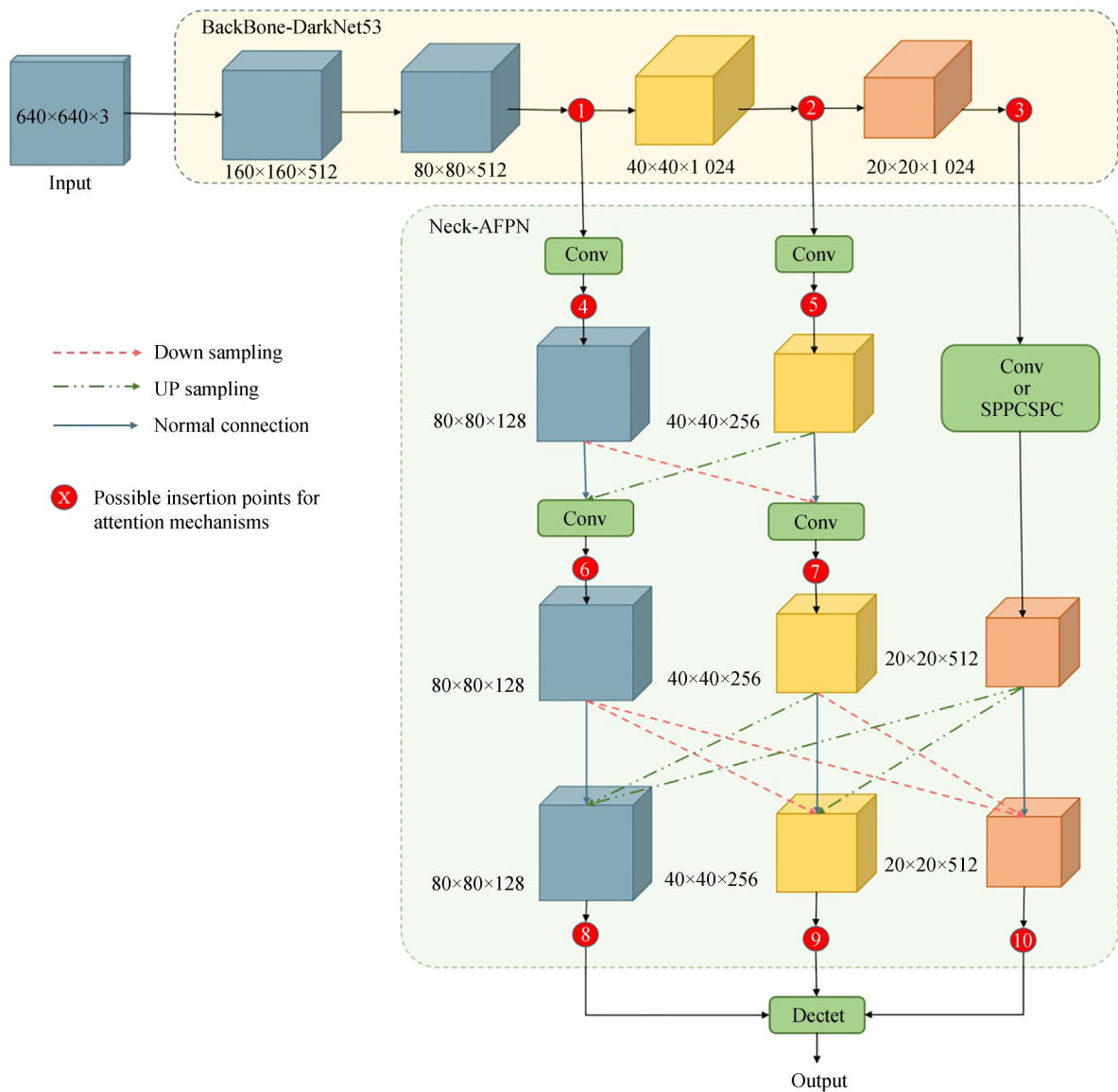


图1 APF-YOLO网络模型结构
Fig. 1 Structure of APF-YOLO network model

的方式将高层特征逐层上采样并融入低层特征,成功在网络传输的过程中保留了高层抽象信息。为了补充 FPN 低层细节信息的丢失,Liu 等在 FPN 的基础上再次添加了一条自底向上的特征融合路径,构建成双金字塔网络,称之为 PANet。无论是在自顶向下还是自底向上的特征金字塔中,特征层需要经过单向的上采样或者下采样后才进行融合,并且非相邻层无法直接进行融合。这会造成原始信息的丢失和退化。受 HRnet^[10]采用反复融合的方式保留不同尺度的特征信息的启发,AFPN^[11]应运而生,该网络以两层底层特征作为金字塔的起点,逐渐加入高层抽象特征信息,加入时进行反复的特征融合。其结构如图 1 所示,以输入图片的 8 倍和 16 倍下采样为金字塔的起点,先将两者进行交叉融合后,加入 Backbone 32 倍下采样后获得特征,三者再次进行交叉融合。融合过程中,上采样操作利用双线性插值实现, k 倍下采样利用 k 个 $k \times k$ 的、步长为 2 的卷积实现。在蜗杆表面缺陷检测任务中,更多的缺陷集中为中小目标,因此,本算法将 SPPCSPC 模块替换为更简洁的普通卷积,避免过度学习的同时极大地节省了参数量。

2.2 高效多尺度注意力机制

为了节约计算资源,在不加深卷积网络的情况下提升网络性能,在自然语言处理领域发挥重要作用的注意力机制被引入图像处理领域。注意力机制的本质是模仿人眼在目视的时候会无意识地关注某些部分,通过给输入的信息添加不同的权重来增加网络对特定信息的关注。注意力机制的一般流程是针对 n 个输入的键值对 $V_1, V_2, \dots, V_n$ 和 $K_1, K_2, \dots, K_n$, 根据需关注的信息确定查询 Q , 利用 $F(Q, K_i)$ 计算出第 i 个关键字 K_i 所对应的注意力得分,而后利用归一化函数 $S(\cdot)$ 获得每个关键字的权重系数,最后将权重系数乘上输入值就得到注意力机制的输出 $A(V)$ 。在视觉任务中通常不指定 Q , 并且 K 和 V 相同, 即有:

$$4(V) = S(F(Q, K)) * V. \quad (1)$$

图像处理领域内注意力机制主要有三种:通道注意力机制、空间注意力机制,以及跨通道

和空间的注意力机制。经典的通道注意力机制是 SENet^[12], 对输入通道添加不同的权重。而空间注意力机制则是通过不同的方式对位置信息进行编码,让网络关注不同的空间位置。CANet^[13]是将位置信息进行编码后嵌入到通道中,利用 X 和 Y 方向上的一维池化操作分解位置信息,再经过卷积降维将其嵌入到通道信息中实现跨通道交互建立通道依赖关系。卷积降维操作是对网络提取特征能力的反向作用^[14], 因此,EMA 机制^[15]在 CA 的基础上引入了特征分组、并行子网络,在不增加参数量的情况下将位置编码信息嵌入通道之中。图 2 展示了 EMA 网络结构。

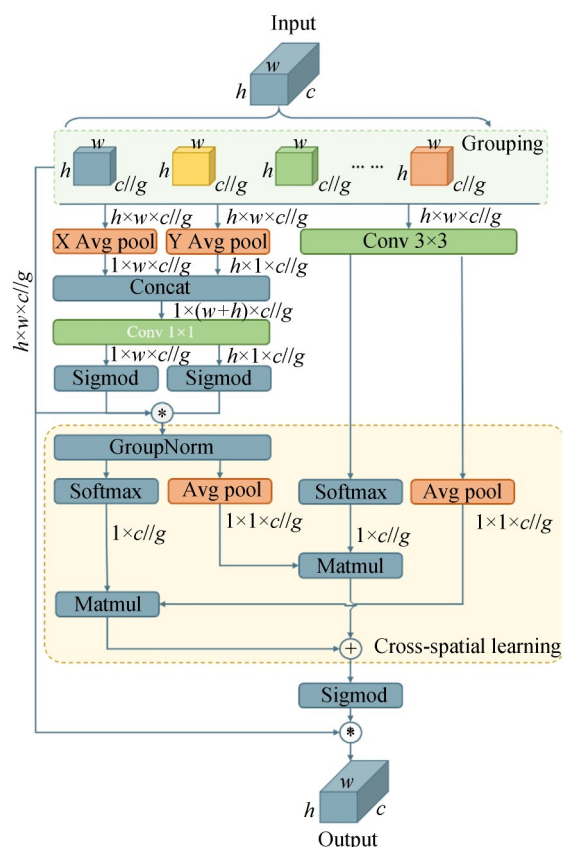


图2 EMA网络结构

Fig. 2 Structure of EMA network

2.3 SIOU回归损失函数

目标检测的损失函数 L 由置信度损失 L_{obj} 、回归损失 L_{iou} 和分类损失 L_{cla} 共同组成, 如式(2)所示。其中, a_1, a_2, a_3 是不同损失的权重系数, 损失函数是模型训练的判别标准, 很大程度上决定

了模型训练的效果。

$$L = a_1 L_{\text{obj}} + a_2 L_{\text{iou}} + a_3 L_{\text{cla}}. \quad (2)$$

IoU Loss^[16]通过计算标注框和预测框的交并比,充分考虑了坐标框4个点的相关性,将分立的4个坐标框回归问题整合为一个坐标框的整体回归。在此基础上,发展出GIoU^[17],DIoU^[18],CIoU^[19],WIoU^[20]及其变体等回归损失函数。上述损失函数都没有考虑标注框与预测框之间角度不匹配的问题,这可能会使网络陷入“来回游荡”,性能变得更差。因此,SIoU^[21]由角度成本、距离成本、形状成本和交并比成本4部分组成,其定义为:

$$L_{\text{SIoU}} = 1 - \text{IoU} + \frac{\Delta + \Omega}{2}, \quad (3)$$

式中: Δ 为距离成本, Ω 为形状成本,IoU为标注框和预测框的交并比,角度成本内嵌进距离成本中。

角度成本为:

$$\Delta = 1 - 2 \times \sin^2 \left(\arcsin \left(\frac{c_h}{\sigma} \right) - \frac{\pi}{4} \right) = \cos \left(2 \times \left(\arcsin \left(\frac{c_h}{\sigma} \right) - \frac{\pi}{4} \right) \right), \quad (4)$$

其中: c_h 为标注框和预测框中心的高度差, σ 为标注框和预测框的中心点的距离。 $(x_{\text{gtc}}, y_{\text{gtc}})$ 为标注框的中心坐标, $(x_{\text{prec}}, y_{\text{prec}})$ 为预测框的中心坐标,则 σ, α 满足:

$$\sigma = \sqrt{(x_{\text{gtc}} - x_{\text{prec}})^2 + (y_{\text{gtc}} - y_{\text{prec}})^2}, \quad (5)$$

$$\sin(\alpha) = \frac{c_h}{\sigma}. \quad (6)$$

可以注意到 α 为0或 $\frac{\pi}{2}$ 时,角度成本为0。

将距离成本定义为:

$$\Delta = \sum_{l=x,y} (1 - e^{-\gamma \rho_l}) = 2 - e^{-\gamma \rho_x} - e^{-\gamma \rho_y}, \quad (7)$$

这里的 c_h 和 c_w 分别为标注框和预测框的外接矩形的高和宽, $\gamma = 2 - \Delta$ 。 $(x_{\text{gtc}}, y_{\text{gtc}})$ 为标注框的中心坐标, $(x_{\text{prec}}, y_{\text{prec}})$ 为预测框的中心坐标, ρ_x 和 ρ_y 可以表示为:

$$\rho_x = \left(\frac{x_{\text{gtc}} - x_{\text{prec}}}{c_w} \right)^2, \rho_y = \left(\frac{y_{\text{gtc}} - y_{\text{prec}}}{c_h} \right)^2. \quad (8)$$

形状成本定义为:

$$\Omega = \sum_{l=w,h} (1 - e^{-w_l})^\theta = (1 - e^{-w_w})^\theta + (1 - e^{-w_h})^\theta, \quad (9)$$

其中 θ 用于控制损失函数对形状损失的权重,一般取值2~6。 $(w_{\text{gt}}, h_{\text{gt}})$ 和 $(w_{\text{pre}}, h_{\text{pre}})$ 分别为标注框和预测框的宽度和高度,满足:

$$\begin{cases} w_w = \frac{|w_{\text{pre}} - w_{\text{gt}}|}{\max(w_{\text{pre}}, w_{\text{gt}})} \\ w_h = \frac{|h_{\text{pre}} - h_{\text{gt}}|}{\max(h_{\text{pre}}, h_{\text{gt}})} \end{cases}. \quad (10)$$

3 实验

3.1 实验数据集制作

实验数据来自如图3所示的针对蜗杆的圆柱形状设计缺陷采集系统,大恒黑白相机MER-503-36U3M/C配合视清科技远心镜头DTCM230-26组成拍摄元件,拍摄参数如表1所示。运动方案为:在放置蜗杆后启动拍摄程序,索雷博旋转台DDR100带动蜗杆自转20°后,触发相机采集,随后旋转台继续旋转,旋转台和相机交替工作直到完成单个蜗杆的拍摄;打光方案使用白色环形光源,相机采集经过光源发出达到蜗杆齿顶表面后反射的光信号。拍摄到的缺陷图片放大后展示在图4中。

根据缺陷的特征和产生原因,将蜗杆齿顶常见的缺陷定义为磕碰伤(Impact Indentation, Ii)、裂纹(Surface Crack, Sc)、夹伤(Foreign Contaminant, Fc)和异物(Squeeze Deformation, Sd)四类

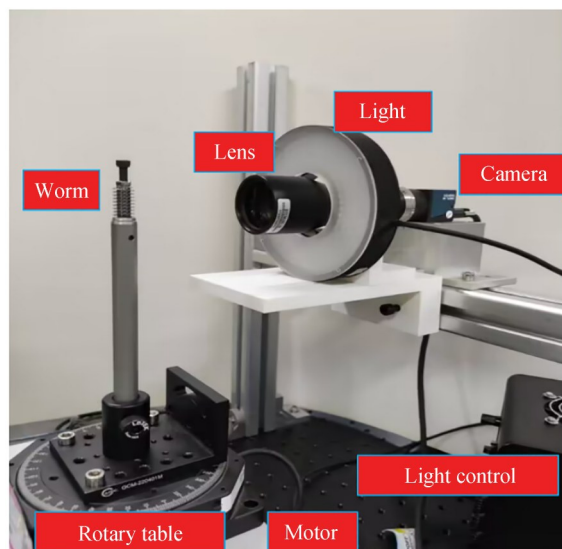


图3 蜗杆缺陷采集系统

Fig. 3 Worm defect collection system

表 1 蜗杆拍摄相机和镜头参数表

Tab. 1 Worm shooting camera and lens parameters

Property	Value
Resolution	2 448×2 048
Pixel size/ μm	3.45
Working distance/mm	73±1
Frame rate/(frame·s ⁻¹)	36
Field of view/mm ²	19.3×16.2
Magnification	0.438
Unit pixel represents physical distance/ μm	7.89

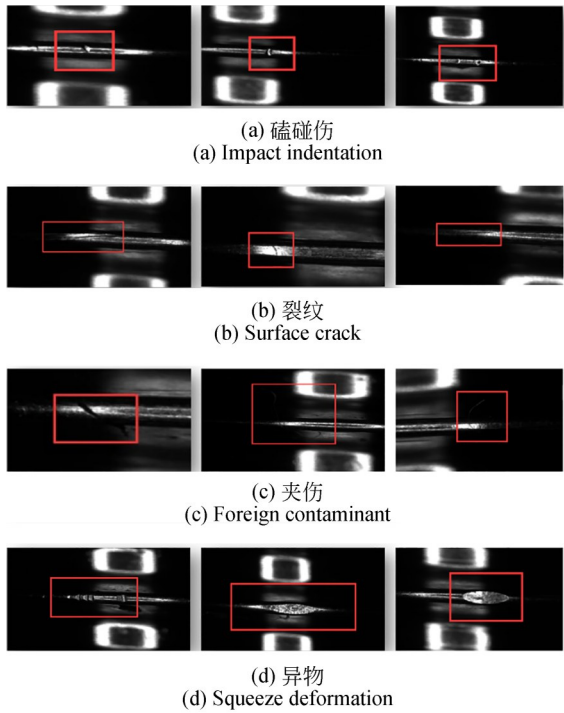


图 4 图片放大后的蜗杆缺陷

Fig. 4 Worm defects after image enlargement

缺陷。一方面,为了减少冗余信息的干扰,需要去除视场中的非齿顶区域;另一方面,为了压缩网络模型的训练时间和推理时间,需要将所采集到的图片进行裁剪。对 207 个蜗杆进行拍摄和图像处理,共得到包含缺陷的图片 2 235 张,使用 Labelme 软件进行人工标注得到带标注信息的原始数据集,其中缺陷的分布情况如表 2 所示。可以看到,原始数据集存在严重的类别数量不平衡,其中磕碰伤的数量远大于其他三类数据,磕碰伤和异物的数量比达到了 6.38:1。这使得网络模型

表 2 蜗杆原始数据缺陷分布

Tab. 2 Distribution of worm raw data defect

Defect type	Defect code	Defect quantity
Ii	1	1 609
Sc	2	414
Fc	3	252
Sd	4	669

过分关注磕碰伤,干扰其他缺陷特征的学习。这种缺陷样本数量不平衡是由蜗杆齿顶不同缺陷发生率决定的。因此,在无法获取新样本的情况下,使用扩增数据集的方式缓解此类不平衡的影响。扩增的原则是尽可能模拟现实存在的情况,在不引入和原始数据差异较大的数据分布的同时增加图片信息。对含有裂纹、异物、夹伤的图片添加随机微小平移和旋转后,进行翻转得到扩充后的数据集。其中,含有裂纹和异物的图像进行水平、垂直和对角线 3 种翻转,含有夹伤的图片仅进行水平、垂直翻转。

将扩增后的数据集按照 8:1:1 划分训练集、验证集和测试集,划分后的数据集分布如表 3 所示,不同缺陷样本处于同一量级。磕碰伤和异物的数量比例下降为 1.98:1。将图片尺寸定义为 S_0 ,在 $0<S<0.005S_0$ 内的标签定义为小目标 (Small Target, ST),在 $0.005S_0<S<0.01S_0$ 内的标签为中目标 (Middle Target, MT),在 $0.01S_0<S<S_0$ 内定义为大目标 (Large Target, LT),如图 5 所示。可以看到,不同类别的尺寸分布差异较大,对于磕碰伤,多为小目标,裂纹中尺寸分布比较均匀,异物和夹伤中大目标缺陷占主流。

表 3 蜗杆表面缺陷数据集数据分布

Tab. 3 Distribution of surface defect data in worm dataset

Defect type	Train	Val	Test	Total
Ii	1 548	242	205	1 995
Sc	1 372	140	144	1 656
Fc	804	108	96	1 008
Sd	1 674	171	144	1 989
Image quantity	3 862	480	463	4 805

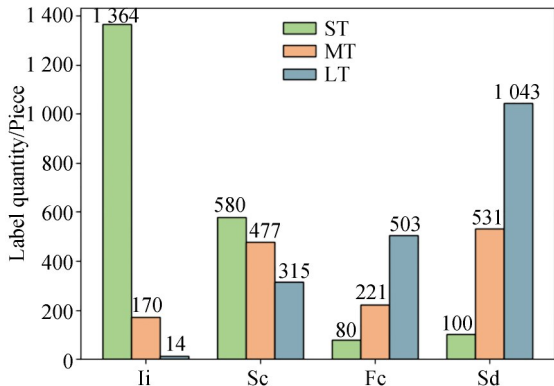


图5 蜗杆表面缺陷训练集标签尺寸分布

Fig. 5 Distribution of label sizes in worm surface defect training dataset

3.2 实验环境及评价指标

实验配置如下:CPU为Intel(R) Xeon(R) Sliver 4114 CPU@2.20 GHz, GPU为NVIDIA GeForce RTX 3090,使用两张GPU进行并行运算来加速模型训练,单张GPU的显存为24 GB,在Ubuntu 22.04 LTS操作系统中安装CUDA 12.1,基于Pytorch框架进行训练。

训练时,输入的图片尺寸设置为 640×640 ,优化器使用随机梯度下降(Stochastic Gradient Descent, SGD),动量因子设置为0.937,权重衰减系数设置为0.0005,在初始几个训练周期中快速增长至 10^{-3} ,随后使用余弦退火算法将学习率衰减至 10^{-5} 。训练周期设置为300。在训练前启动随机缩放系数为0.9, Mosaic数据增强系数1和0.15的mixup数据增强策略。

本文采用平均精度均值(mean Average Precision, mAP)来评价模型训练的多类别缺陷识别结果。平均精度(Average Precision, AP)为识别单类别时的精度(Precision, P)-召回(Recall, R)曲线所覆盖的范围,见式(11),可以用于衡量对单一类别的检测精度,其值越高,检测效果越好。而mAP则是不同待检测类别的平均精度的均值,见式(12)。N为缺陷的类别数,用以衡量模型对不同类别的检测效果,体现了模型的综合检测能力。这里用帧每秒(Frame Percent Second, FPS)来衡量模型的检测速度,表征模型在一秒内能够处理的图片数量。

$$AP = \int_0^1 P(R) dR, \quad (11)$$

$$mAP = \frac{\sum AP}{N}. \quad (12)$$

在实际任务中,模型检测效果则使用更直观的检测准确率(Accuracy, A)和漏检率(Miss, M)来评价。假设在一批样品中检出缺陷的个数为 J ,误检的个数为 W ,漏检的个数为 L ,则检测准确率 A 定义为检测出来的缺陷中真实缺陷所占的比率,由式(13)计算;漏检率 M 定义为漏检的缺陷占有所有真实缺陷的比率,由式(14)计算。实际应用中,要求检测准确率大于95%,漏检率小于5%。

$$A = \frac{J - W}{J}, \quad (13)$$

$$M = \frac{L}{J - W + L}. \quad (14)$$

3.3 实验结果及分析

AFP-YOLO是在YOLOv7的基础上进行改进,保留了YOLOv7的Backbone网络和检测头,将Neck网络改进为AFPN,并且Backbone中引入注意力机制以及采用Siou参与回归损失函数的计算。为了验证算法改进的有效性,本文对改进不同的Neck网络、不同位置添加EMA注意力机制、采用不同的回归损失函数、消融实验和AFP-YOLO与其他经典YOLO版本进行了对比实验,以上实验均是在相同环境配置下,在蜗杆表面数据集上进行的。

首先验证数据增广的有效性,使用YOLOv7模型训练数据增广前后的数据进行实验。训练结果的 $mAP_{0.5}$ 如图6所示。数据增广前 $mAP_{0.5}$ 为66.6%,数据增广后为73.6%。在数据增广前因为不同缺陷类别样本数量严重失衡,齿顶磕碰伤数量最多,所以模型过分关注磕碰伤的特征,对与磕碰伤特征类似的裂纹和夹伤缺陷检测能力不足, $mAP_{0.5}$ 分别为54.5%和63.7%。但在数据增广后不同样本数量趋向均衡,模型对裂纹和夹伤的检测能力有明显提升, $mAP_{0.5}$ 分别为67.5%和77.3%。同时异物缺陷的检测精度只有非常微小的提升,证明数据增广前后实验的差异是由样本数量不平衡造成的,而不是样本数量的增加。

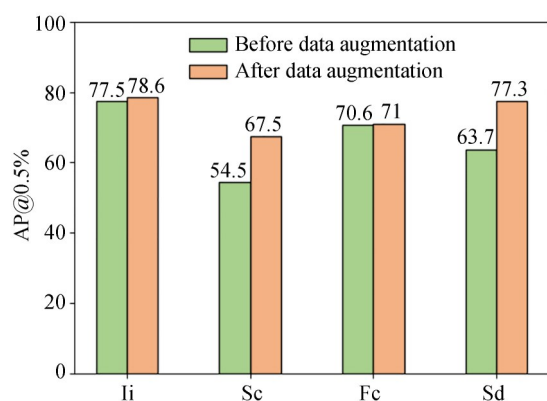


图 6 数据增广对模型 mAP 的影响

Fig. 6 Impact of data augmentation on mAP of model

在表 4 中, YOLOv7-SPPAFPN 在 Backbone 的高层输出的特征图经过 SPPCSPC 模块后再进入 Neck 网络, 而 YOLOv7-AFPN 中 Backbone 的

高层输出特征图经过普通卷积后送入 Neck 网络。可以看到, 在采用 AFPN 作为 Neck 网络实现特征融合的时候, 插入 SPPCSPC 模块在蜗杆表面缺陷数据集上仅取得了 73.9% 的 $mAP_{0.5}$, 相比 PAN 提升了 0.3%。而直接接入 AFPN 模块则将参数量从 36.49M 降低至 29.73M 的同时取得了 75.5% 的 $mAP_{0.5}$, 相比 PAN 提升了 1.9%。深入分析可以发现, AFPN 网络虽然损失了一定准确率, 但是更大幅度地提升了模型的召回率, 综合提升了网络的检测精度。这在对漏检标准更为严格的工业缺陷检测领域是可以接受甚至是更重要的。并且因为参数量的减少, 网络模型处理图片的速度由 60.24 frame/s 提升至 75.75 frame/s, 这表示 AFPN 相比于 PAN 能够更快地完成特征融合, 压缩推理时间。

表 4 添加不同 Neck 网络的模型实验结果

Tab. 4 Experimental results adding different neck networks to model

Neck	Parameters/M	P/%	R/%	mAP@0.5/%	FPS
PAN	36.49	80.89	68.4	73.6	60.24
SPPAFPN	35.81	76.12	71.8	73.9	65.79
AFPN	28.73	76.03	74.5	75.5	75.75

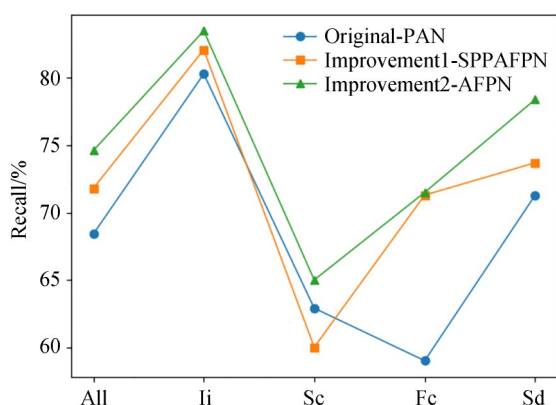


图 7 不同 Neck 网络对不同类别召回率的影响

Fig. 7 Impact of different neck networks on recall for various categories

图 7 展示了不同的 Neck 网络对不同缺陷的召回率结果。可以看到, 采用渐进特征金字塔式的特征融合方式对尺寸较大的异物和夹伤缺陷的召回率都有不同程度的提升。但是, SP-

PAFPN 对裂纹缺陷的召回率有所下降, 并且对其他类别缺陷的召回率的提升也不如 AFPN。因此, 采用 AFPN 作为本算法的颈部网络。AFPIN 在异物和夹伤的召回率分别为 71.5% 和 78.4%, 相对 PAN 分别有 21% 和 9.9% 的提升。对所有的缺陷综合评估, AFPN 的召回率为 74.5%, AFPN 相较于 PAN 综合召回率提升了 8.9%。综上所述, AFPN 在减少网络参数数量的同时能够检出更多的缺陷, $mAP_{0.5}$ 提升了 1.9%。

注意力机制能够使得网络关注更重要的部分, 为了深入地探究在不同位置引入注意力机制对网络模型检测精度的影响, 对不同位置引入注意力机制进行了测试, 测试结果如表 5 所示。一般的注意力机制会添加在模型的中间层, 因为在起始层可能会造成关注错误信息, 在靠后层可能会因为网络过分学习造成过拟合。方案编号为 1, 2, 3 的注意力机制插入的位置为 Neck 网络, 可

表 5 在不同位置插入 EMA 的实验结果

Tab. 5 Experimental results of inserting EMA at different positions

Project code	Insertion position	Parameters	mAP@0.5/%
—	—	28.73	75.5
1	4,5	28.75	75.9
2	6,7	28.75	75.6
3	8,9,10	28.79	75.6
4	2,3	29.61	74.7
5	1,2	28.94	76.5

以看到,不管是插入到 APFN 的输出段(方案 3)或中间(方案 2)注意力机制的收效甚微。但是当注意力机制插入起始段(方案 1),网络模型的检测精度提升了 0.4%。进一步,将注意力机制插入 Backbone 的输出层,插入低层和中层输出(方案 5)获得了 1% 的提升,仅付出了 0.21M 的参数量代价。而将注意力机制添加在高层输出时(方案 4),检测精度不升反降。图 8 展示了插入注意力方案 5 和没有插入注意力机制的热力图对比(彩图见期刊电子版),颜色越红表示对此次检测结果的贡献越大,反之则越小。可以看到,添加了注意力机制以后网络模型会更加关注缺陷附近,缺陷附近的区域特征对检测结果的权重增大,减少其他位置对结果的干扰。以上结论说明,将注意力机制引入在位置 1,2 能够有效促进网络进行特征提取,提升检测精度。

在 AFPN+EMA 的网络模型进行了回归损失函数的对比实验,探究不同损失函数对检测结果的影响,测试结果如表 6 所示。CIoU 和 SIoU 是最适合蜗杆缺陷检测任务的回归损失函数,CIoU 作为损失函数时 $mAP_{0.5}$ 为 76.5%, $mAP_{0.5}$ 为 26.5%。采用 DIoU, GIoU, EIoU, MPDIoU, WIoU 及其变形版本时 $mAP_{0.5}$ 均有不同程度的下降,当采用 SIoU 作为回归损失函数时 $mAP_{0.5}$ 提升了 0.04%,为 76.9%。因此,在阈值为 0.5 时采用 SIoU 作为回归损失函数在蜗杆表面缺陷检测任务精度是最高的。

为了验证 AFPN,EMA,SIoU 3 种改进方式的耦合效果,在蜗杆表面数据集上使用同样的超

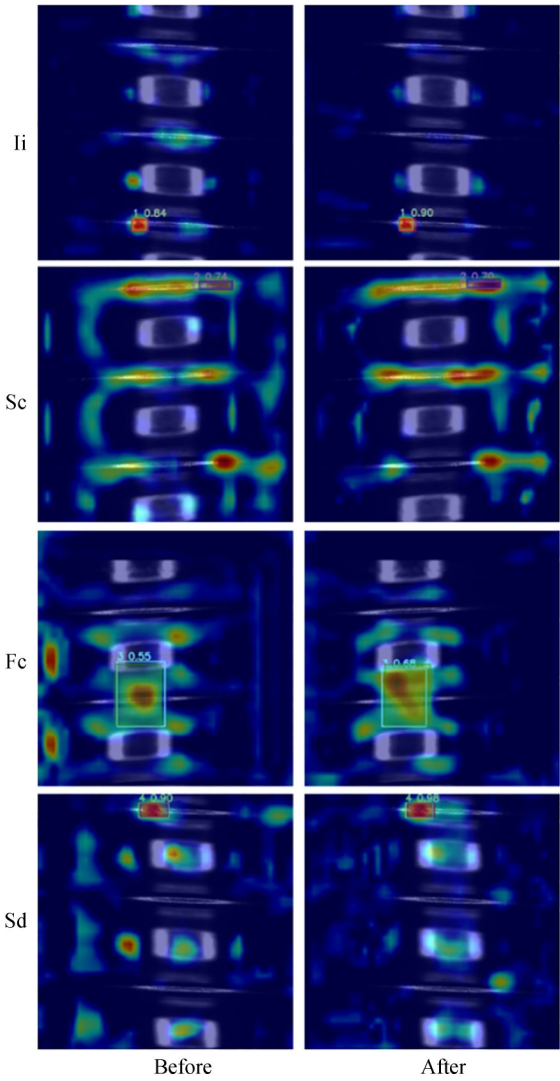


图 8 插入注意力机制前后热力图

Fig. 8 Heatmaps before and after inserting attention mechanism

表 6 使用不同损失函数的实验结果

Tab. 6 Experimental results using different loss functions

Index	Loss function	mAP@0.5	mAP@0.75
1	CIoU	76.5	26.5
2	DIoU	74.4	25.5
3	GIoU	73.7	25.2
4	EIoU	73.4	24.8
5	MPDIoU	74.2	26.4
6	WIoUv1	74.4	25.3
7	WIoUv2	75.3	26.7
8	WIoUv3	74.8	24.6
9	SIoU	76.9	25.3

参数进行了消融实验,实验结果如表 7 表示。1 表示采用改进方式,0 表示不采用此改进方式。对比第一组和第二组结果可以验证 AFPN 的性能在蜗杆检测上优于 PAN,这个改进使检测精度提升了 1.9%。对比第三组和第五组的结果可以证明 EMA 注意力机制本身的有效性,以及在基于 AFPN 的网络上添加 EMA 机制能够带来更大的提升,这也侧面印证了 AFPN 的有效性。分析其他组的实验数据表明,当同时采用 3 种改进策略时改进效果最佳,mAP_{0.5}获得了 3.3%的提升。并且相较于 YOLOv7,本文提出的模型参数量下降至 28.94M。

为了横向验证 AFP-YOLO 的检测性能,选择经典的 YOLO 版本 v3,R,v5,v7,v8 中公开权重中参数量最接近 AFP-YOLO 的变体算法,即 YOLOv3, YOLOR-p6, YOLOv5m, YOLOv7, YOLOv8m 在蜗杆表面缺陷检测数据中进行对比实验,实验结果如表 8 所示。AFP-YOLO 在阈值为 0.5 时精度最高为 76.9。在阈值为 0.75 时,精度仅次于 YOLOv8m。图 9 展示了不同算法在蜗杆表面缺陷检测任务中的效果,其中红色检测框为磕碰伤缺陷检测,绿色检测框为异物检测,蓝色检测框为裂纹检测,紫色检测框为夹伤检测。可以看出,YOLOv3, YOLOR, YO-

表 7 三种改进方式的消融实验
Tab. 7 Ablation experiments of three improvement approaches

AFPN	EMA	SIoU	Paramete/ M	mAP @0.5	Variation
0	0	0	36.49	73.6	—
1	0	0	28.7	75.5	+1.9
0	1	0	36.5	74.3	+0.7
0	0	1	36.49	73.8	+0.2
1	1	0	28.94	76.5	+2.9
0	1	1	36.51	75.1	+1.5
1	0	1	28.7	74.5	+0.9
1	1	1	28.94	76.9	+3.3

LOv5m, YOLOv7, YOLOv8m 对磕碰伤都有不错的检测能力,但是对于裂纹和异物会存在漏检,而本文提出的算法因为采用了渐进式的特征金字塔融合方式,不同的特征层进行反复融合,所以保留了更多的缺陷信息,能够有效地检出裂纹和异物,这一点和原始算法 YOLOv7 相比更加明显。

最后,为了验证 AFP-YOLO 的实际检测效果。使用图 3 中的采集系统另外拍摄了 30 个蜗杆的图片,送入 AFP-YOLO 进行检测。经过统

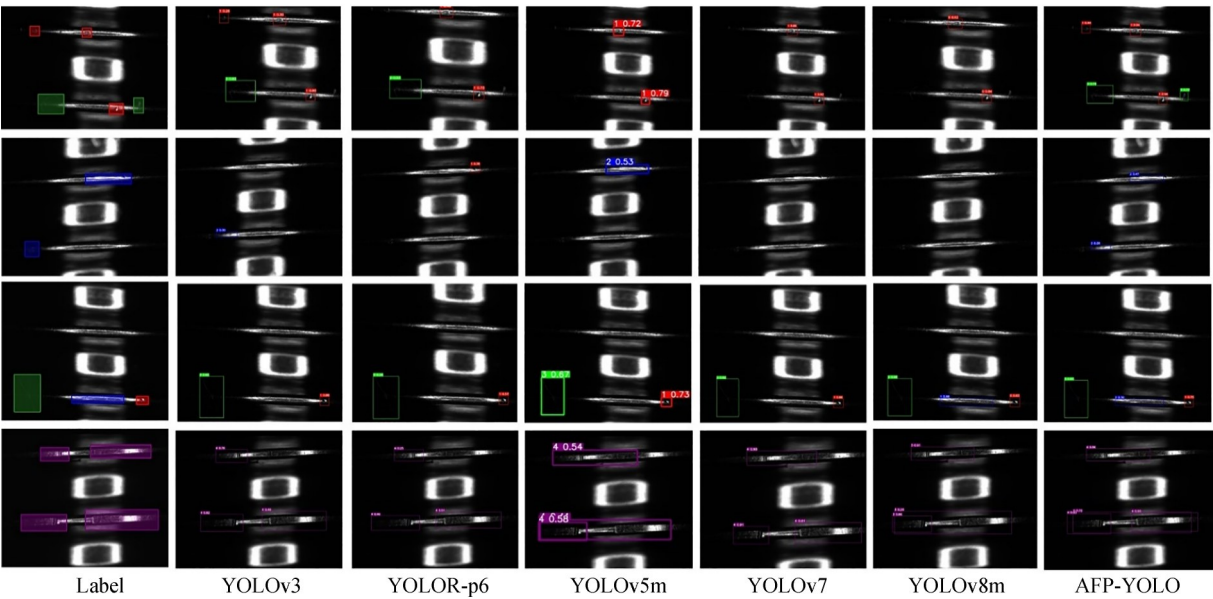


图 9 不同算法对蜗杆表面缺陷的检测效果

Fig. 9 Effectiveness of different algorithms on worm gear surface defect detection

表 8 蜗杆表面缺陷数据集上不同算法的测试结果

Tab. 8 Test results of different algorithms on worm gear surface defect dataset

Algorithm	Image Size	Parameter/M	mAP@ 0.5	mAP@ 0.75
YOLOv3	640×640	61.51	69.9	18.1
YOLOR-p6	640×640	36.82	45.9	9.5
YOLOv5m	640×640	20.8	73.3	—
YOLOv7	640×640	36.49	73.6	23.1
YOLOv8m	640×640	25.84	72.2	35.5
AFP-YOLO	640×640	28.94	76.9	25.3

计,这 30 个蜗杆中包含磕碰伤缺陷 316 个、裂纹缺陷 137 个、异物缺陷 135 个、夹伤 229 个。统计不同缺陷的检出、误检和漏检情况如表 9 所示,AFP-YOLO 的综合检测准确率为 99.25%,漏检率为 2.2%。其中,裂纹的误检比率和漏检比率均略高于其他缺陷,因为裂纹一般位于齿顶中央,为细长状,更易被打光情况干扰,但是其他 3 种缺陷的检测准确率和漏检率均符合应用标准。总体而言,AFP-YOLO 的检测效果满足实际应用需求。

表 9 AFP-YOLO 在 30 个蜗杆中的实际检测效果

Tab. 9 Actual detection performance of AFP-YOLO among 30 worms

Defect	Defect quantity	Detect quantity	False detect	Miss detect
Ii	316	312	1	5
Sc	137	131	2	8
Fc	135	133	1	3
Sd	229	229	2	2
Total	817	805	6	18

参考文献:

- [1] 罗东亮,蔡雨莹,杨子豪,等.工业缺陷检测深度学习方法综述[J].中国科学:信息科学,2022,52(6):1002-1039.
LUO D L, CAI Y X, YANG Z H, et al. Survey on industrial defect detection with deep learning[J]. *Scientia Sinica (Informationis)*, 2022, 52(6): 1002-1039. (in Chinese)

4 结 论

本文基于机器视觉技术针对蜗杆齿顶的表面缺陷实现了缺陷的快速识别定位。针对蜗杆的形状特征设计了图片采集系统,并对 207 个蜗杆进行数据采集,结合合适的数据增广策略制作了数据集,数据增广前后的类别间最大数量比由 6.38:1 下降至 1.98:1,减小了缺陷发生率不同引起的检测难度。而后针对不同缺陷的尺寸分布差异,重构了 YOLOv7 的 Neck 网络,利用 Backbone 的两个低层作为特征金字塔的起点,随后加入更高层的抽象特征,并在加入过程中进行反复的融合,在参数量减少 20.7% 的情况下有效提升了网络对蜗杆不同尺寸缺陷的检测能力,使得检测精度提升 1.9%;其次,针对图片中非缺陷位置对检测结果的干扰,引入注意力机制改进网络的特征提取能力,并通过消融实验确定了注意力机制插入至 Backbone 网络的两个低层输出后端,加强了网络对缺陷位置的关注,检测精度进一步提升 1%;最后,采用 SIoU 回归损失函数,在网络训练中考虑了标注框与预测框的方向差异,增加方向成本,检测精度再次提升 0.4%。最终,改进的检测算法相较于 YOLOv7,参数量下降了 7.55M,精度提升了 3.3%,mAP_{0.5} 达到了 76.9%。与 YOLOv3, YOLOR-p6 和 YOLOv5m 等检测算法进行对比,本文提出的检测算法训练性能最优。在实际样品中进行测试,AFP-YOLO 的综合检测准确率为 99.25%,漏检率为 2.2%。本文提出的蜗杆表面缺陷检测方法基本能够满足蜗杆表面自动检测的应用需求,后续研究将进一步完善算法,重点提升检测速度。

- [2] 郑立平.渐开线蜗杆机构的参数化设计及动态接触特性研究[D].合肥:合肥工业大学,2012.
ZHENG L P. *Parametric Design and Dynamic Contact Characteristics Research of Involute Worm Gears*[D]. Hefei: Hefei University of Technology, 2012. (in Chinese)
- [3] 辛露,李睿.小模数蜗杆外观缺陷的识别与分类[J].工具技术,2018,52(9):137-140.

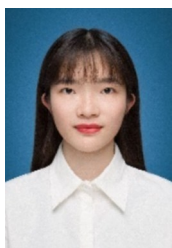
- XIN L, LI R. Identification and classification of appearance defect of small module worm[J]. *Tool Engineering*, 2018, 52(9): 137-140. (in Chinese)
- [4] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. *Communications of the ACM*, 2017, 60(6): 84-90.
- [5] 乔健, 陈能达, 伍雁雄, 等. 融合注意力机制的金属蜗圆柱表面缺陷检测[J]. *光学精密工程*, 2023, 31(3): 404-416.
- QIAO J, CHEN N D, WU Y X, *et al.* Defect detection of cylindrical surface of metal pot combining attention mechanism [J]. *Opt. Precision Eng.*, 2023, 31(3): 404-416. (in Chinese)
- [6] 李闻, 李小春, 闫昊雷. 基于改进 YOLO v3 的 PCB 缺陷检测[J]. *电光与控制*, 2022, 29(4): 106-111.
- LI W, LI X CH, YAN H L. PCB defect detection based on improved YOLO v3[J]. *Electronics Optics & Control*, 2022, 29(4): 106-111. (in Chinese)
- [7] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors [C]. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Vancouver, BC, Canada. IEEE, 2023: 7464-7475.
- [8] REDOM J, FARHADI A. Yolov3: An incremental improvement [J]. *arXiv preprint arXiv: 1804.02767*, 2018.
- [9] LIU S, QI L, QIN H F, *et al.* Path aggregation network for instance segmentation[C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT, USA. IEEE, 2018: 8759-8768.
- [10] SUN K, XIAO B, LIU D, *et al.* Deep high-resolution representation learning for human pose estimation[C]. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, CA, USA. IEEE, 2019: 5686-5696.
- [11] YANG G Y, LEI J, ZHU Z K, *et al.* AFPN: asymptotic feature pyramid network for object detection [EB/OL]. 2023: *arXiv: 2306.15988*. <http://arxiv.org/abs/2306.15988>
- [12] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT, USA. IEEE, 2018: 7132-7141.
- [13] HOU Q B, ZHOU D Q, FENG J S. Coordinate attention for efficient mobile network design [C]. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Nashville, TN, USA. IEEE, 2021: 13708-13717.
- [14] ZHANG Z, WU Y, ZHANG J, *et al.* Efficient channel attention for deep convolutional neural networks [C]. *Proceedings of the AAAI Conference on Artificial Intelligence*. 2021.
- [15] OUYANG D L, HE S, ZHANG G Z, *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)*. Rhodes Island, Greece. IEEE, 2023: 1-5.
- [16] YU J H, JIANG Y N, WANG Z Y, *et al.* UnitBox: an advanced object detection network [C]. *Proceedings of the 24th ACM international conference on Multimedia*. Amsterdam The Netherlands. ACM, 2016: 516-520.
- [17] REZATOFIGHI H, TSOI N, GWAK J, *et al.* Generalized intersection over union: a metric and a loss for bounding box regression[C]. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, CA, USA. IEEE, 2019: 658-666.
- [18] ZHENG Z H, WANG P, LIU W, *et al.* Distance-IoU loss: faster and better learning for bounding box regression[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, 34(7): 12993-13000.
- [19] ZHENG Z H, WANG P, REN D W, *et al.* Enhancing geometric factors in model learning and inference for object detection and instance segmentation [J]. *IEEE Transactions on Cybernetics*, 2022, 52(8): 8574-8586.
- [20] TONG Z J, CHEN Y H, XU Z W, *et al.* Wise-Iou: Bounding Box Regression Loss with Dynamic

Focusing Mechanism [EB/OL]. 2023: *arXiv*: 2301.10051. <http://arxiv.org/abs/2301.10051>.

[21] GEVORGYAN Z. *Siou Loss: More Powerful*

Learning for Bounding Box Regression [EB/OL]. 2022: *arXiv*: 2205.12740. <http://arxiv.org/abs/2205.12740>.

作者简介:



王 蕾(2000—),女,云南昭通人,硕士研究生,2021年于重庆大学获得学位,主要从事机器视觉及目标检测方向的研究。E-mail: m202172882@hust.edu.cn

通讯作者:



郭文平(1984—),男,武汉湖北人,副教授,博士生导师,2007年、2012年于华中科技大学分别获得学士和博士学位,主要从事光电系统设计,光电检测,机器视觉等领域的研究。E-mail: wpguo@hust.edu.cn