

面向发动机叶片缺陷检测的轻量化 YOLOv11 改进方法研究

敖良忠 马晋仟

(中国民用航空飞行学院航空工程学院 成都 618307)

摘要: 航空发动机叶片的缺陷检测是保障飞行安全的关键技术环节。传统工业内窥镜检测方法严重依赖人工经验,存在效率低、主观性强及微小缺陷易漏检等问题。为此,本文提出一种轻量化、高精度的 YOLOv11 改进模型,专用于发动机叶片缺陷的实时检测任务。研究中采集并构建了包含 4 类典型缺陷(弯曲、烧蚀、裂纹、材料缺失)在内的高质量工业图像数据集,并针对小目标、复杂背景及多尺度特性,构建 CFES 主干网络来增强语义信息的整合能力和减少计算量,采用 ShuffleNetV2 替换原始主干以减轻计算负担,引入 BiFormer 注意力机制提升特征表达能力,同时结合 Dynamic-DCNv3 增强检测头对复杂纹理与小尺寸目标的建模能力。实验结果表明,改进模型较基线模型在 mAP @0.5 提升至 85.0%,参数量降低至 1.7 M,具备更优的检测性能与边缘部署适应性,最终成功部署于内置芯片为 RK3588 的工业内窥镜平台,帧数维持在 30 帧左右,实现了高效、稳定的缺陷自动识别。该研究为航空维修现场的智能检测提供了可行方案,也为轻量化目标检测模型在工业嵌入式场景的应用拓展提供了技术支撑。

关键词: 航空发动机叶片;缺陷检测;实时目标检测;轻量化模型;YOLOv11

中图分类号: TN911.73; TP391.41 文献标识码: A 国家标准学科分类代码: 590.6099

Research on a lightweight improved YOLOv11 method for aero-engine blade defect detection

Ao Liangzhong Mou Jinqian

(School of Civil Aviation Engineering, Civil Aviation Flight University of China, Chengdu 618307, China)

Abstract: Defect detection of aero-engine blades is a critical technological step to ensure flight safety. Traditional industrial borescope inspection methods rely heavily on manual expertise, which often leads to low efficiency, strong subjectivity, and missed detections of subtle defects. To address these issues, this paper proposes a lightweight and high-precision improved YOLOv11 model, specifically designed for real-time defect detection of engine blades. A high-quality industrial image dataset was constructed, comprising four typical defect types—bending, ablation, cracks, and material loss. Aiming at the characteristics of small targets, complex backgrounds, and multi-scales, a CFES backbone network is constructed to enhance the integration ability of semantic information and reduce the amount of computation. Specifically, ShuffleNetV2 is employed as the backbone network instead of the original one to alleviate computational overhead, while the BiFormer attention mechanism is integrated to strengthen the feature representation capability. Additionally, a Dynamic-DCNv3-based detection head is employed to improve the modeling capability for complex textures and small-sized objects. Experimental results show that the improved model achieves an mAP@0.5 of 85.0%, surpassing the baseline model, while reducing parameters to only 1.7 M. This demonstrates superior detection performance and adaptability for edge deployment. The model was successfully deployed on an industrial borescope platform equipped with the RK3588 chip, where the frame rate remained at approximately 30 frames per second, achieving efficient and stable automatic defect recognition. This study provides a practical solution for intelligent on-site inspection in aviation maintenance and offers technical support for the application of lightweight object detection models in industrial embedded scenarios.

Keywords: aero-engine blade; defect detection; real-time object detection; lightweight model; YOLOv11

0 引言

可靠性直接关系到航空安全。叶片作为发动机中的关键受力部件,长期工作于高温高压环境中,极易产生裂纹、变形、涂层剥落等多种表面缺陷^[1]。这些缺陷若未能及时发现,可

能导致发动机性能下降,甚至引发严重飞行事故。因此,叶片缺陷检测已成为发动机健康管理中的关键环节^[2]。目前,航空发动机的叶片检测主要依赖工业内窥镜进行人工目视检查。这种方式虽然成本低、操作灵活,但严重依赖检测人员经验,不仅检测效率低、主观性强,还易造成微小缺陷的漏检。据文献统计,人工内窥检测平均耗时超过 45 min,清洁条件下识别准确率为 86.8%,而在油污污染、金属反光等复杂工况下准确率降至 66.8%^[3]。此外,尺寸小于 0.5 mm 的缺陷难以从低对比度图像中辨识,使得现有方法难以满足高精度、高效率的检测需求^[4]。

深度学习在计算机视觉领域的发展为航空发动机叶片缺陷的智能检测提供了新的技术路径。近年来,基于 YOLO(you only look-once)、Faster R-CNN(region-based convolutional neural network)、SSD(single shot multibox detector)等目标检测框架的研究在叶片检测任务中取得显著进展。其中,YOLO 算法凭借实时性与精度的平衡和端到端检测的高效性等优势,在航空发动机叶片检测领域展现出显著应用价值。在早期研究中,学者聚焦于基础框架的适应性验证。例如,Zhang 等^[5]将 YOLOv3 成功应用于风扇叶片损伤检测,证明了深度模型在检测速度与准确率之间的平衡能力,但对长条裂纹等弱边缘目标的检测仍然不足;Li 等^[6]提出的改进 YOLOv5 模型通过引入 ECA 注意力和 BiFPN 特征融合结构,检测精度提升了 1.0%;魏永超等^[6]通过对 YOLOv7 引入 CA 注意力机制和改进上采样算子,精度相较于基准网络提升了 7.4%,但裂纹精度仅为 71.87%;李文龙等^[7]提出的改进 YOLOv8 模型,通过在主干网络嵌入 EMA 注意力机制并采用 EIoU 损失函数,较基线模型平均精度提升 2.5%。

尽管上述方法在实验环境下取得了可观成果,但正如相关研究^[8-9]指出,未来航空叶片检测的关键在于提升微小缺陷的检测能力与嵌入式部署效率。因此,上述针对叶片缺陷识别的研究仍局限于算法层面的验证,尚未开展模型在嵌入式边缘设备的部署探索,难以评估模型经量化压缩后在实际工业环境中的真实性能表现。为此,已有研究者围绕嵌入式部署需求,在实现模型轻量化的同时保持优良检测精度方面展开了探索。如余肖飞等^[10]通过引入 Hard-mish 激活函数和 Dense Block 模块,将 YOLOv4 的精度提升至 79.95%,并在 Jetson Xavier NX 设备上实现 12 FPS 实时检测与端到端部署,但仍存在小目标检测精度不足和检测卡顿等问题。

YOLO 系列算法在实时应用场景中虽然展示了高效的目标检测能力,但在复杂场景中,特别是在小目标检测的场景中表现不佳,发动机叶片缺陷的数据集也存在正负样本差异大、样本稀缺等问题,限制了 YOLO 算法的泛化能力。尽管当前尚未有直接将 YOLOv11 部署至航空发动机孔探仪的研究,但从相关领域及对 YOLOv11 的研究拓展中可获取有价值参考。在工业检测任务中,YOLOv11 展

现出优于旧版本 YOLO 等模型的准确率与实时性^[11],尤其在边缘设备部署场景中表现稳定^[12]。针对航空发动机叶片缺陷检测中小目标密集、背景干扰强、部署平台受限等特点,YOLOv11 提供了良好结构基础。本研究旨在将模型部署于搭载国产边缘设备的航空孔探仪,实现航空发动机叶片缺陷的实时检测;算法层面以 YOLOv11n 为基础,提出一套兼顾精度与轻量化的边缘部署改进策略,围绕轻量化主干构建、跨尺度语义增强、注意力机制引入及检测头优化 4 维度展开改进:

1)跨尺度特征增强结构(cross-scale feature enhancement structure, CFES):为提升 YOLOv11 在边缘设备中对小尺寸缺陷检测的表现,本文对其颈部结构进行了轻量化改造,提出了跨尺度特征增强结构以替代传统特征金字塔网络(feature pyramid network, FPN)与路径聚合特征金字塔网络(path aggregation feature pyramid network, PAFPN)结构。该结构结合了最近邻上采样、通道对齐与 C3k2 卷积块,实现深浅层特征的有效融合。同时引入可分离卷积下采样模块(separable convolution down sampling, SCDown)用于增强语义表达并降低信息损失,从而在保持检测精度的同时,显著降低计算负担,提升边缘设备的实时推理能力。

2)检测头结构改进:针对航空发动机叶片缺陷在实际检测中呈现出的复杂形态变化特征,传统检测头在建模目标几何变形时存在一定局限,易导致微小缺陷漏检与定位偏差。为提升模型对非刚性结构与细节边界的建模能力,本文引入 Dynamic Head^[13] 结构,该检测头集成了尺度感知、空间感知与任务感知三类注意力机制,能够在不显著增加推理开销的前提下,有效增强检测精度与鲁棒性。在此基础上,进一步将原结构中的标准卷积模块替换为更高阶的 DCNv3^[14] 可变形卷积模块,相比于 DCNv2^[15],DCNv3 引入了偏移去噪与可学习采样机制,可更精细地建模复杂目标的局部形变与纹理细节。这种检测头结构在保持模型计算复杂度可控的同时,显著提升了对航空发动机叶片中微小缺陷与不规则区域的感知能力,增强了模型在复杂工况下的检测鲁棒性与精度。

3)注意力机制优化:为增强模型对微小缺陷与复杂背景纹理的感知能力,本文在特征提取阶段引入 BiFormer^[16] 模块。BiFormer 提出了双层路由注意力,即在注意力操作中引入区域引导策略,从而显著减少计算冗余。该机制能够动态聚焦于图像中的高响应区域,在不显著增加推理成本的前提下,增强了模型对叶片裂纹、材料缺失等微小目标的特征建模能力,尤其适用于密集小目标检测任务。BiFormer 的引入使得模型具备更强的区域选择性与判别力,为航空发动机叶片缺陷检测提供了结构支持。

4)轻量化主干网络重构:为实现模型在边缘设备上的高效部署,本文将 YOLOv11 原主干网络替换为 ShuffleNetV2^[17] 结构。ShuffleNetV2 基于通道分割与通

道重排机制,在有效减少参数规模与计算复杂度的同时,增强了跨通道的信息交互能力。该结构遵循多项硬件友好设计原则,显著降低内存访问开销与延迟,在保持良好特征提取能力的前提下,实现了对工业内窥镜平台的高效适配。

1 YOLOv11 模型概述

YOLO 系列算法经过多轮迭代,已发展至第十一代 YOLOv11,相较前代版本,其在检测精度、推理速度与模型泛化能力方面均实现显著提升。如图 1 所示,YOLOv11 引入了 C3k2 模块替代传统 C2f 结构,有效提升特征表达能力与梯度传播效率;同时在主干网络中加入 C2PSA (cross-stage partial spatial attention) 模块,增强对多尺度特征与遮挡目标的感知能力。此外,YOLOv11 延续了 YOLOv9 提出的路径引导融合机制(PGI)与 YOLOv10 的双标签分配策略,从结构与训练机制两方面提升了整体性能^[11]。

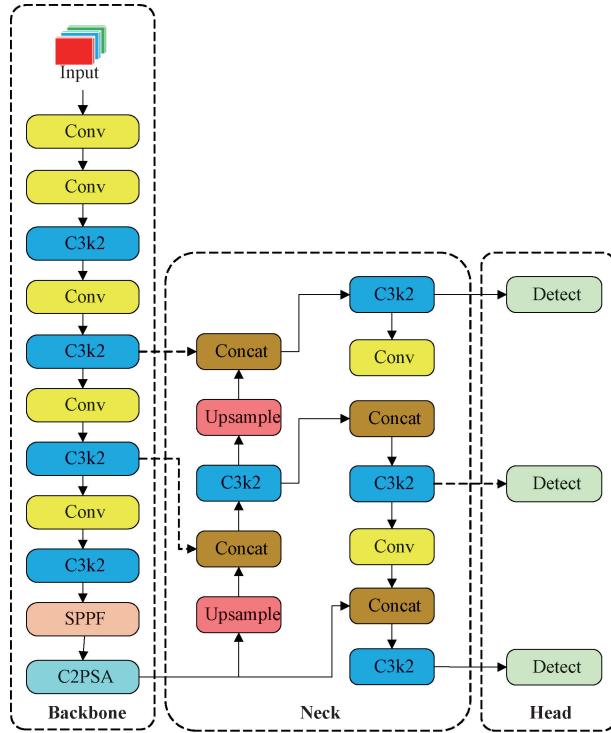


图 1 YOLOv11 网络结构图

Fig. 1 The network structure diagram of YOLOv11

2 本文方法

2.1 跨尺度特征增强结构

针对航空发动机叶片缺陷在图像表现上普遍存在尺寸小、边缘模糊、纹理弱等特性,传统特征融合结构在目标检测任务中易出现信息衰减与浅层特征语义不足的问题。为此,本文设计了一种适配嵌入式平台、兼顾高效性与多尺度表达能力的轻量级跨尺度特征融合结构。该结构整体遵循

YOLOv11 中“特征金字塔 + 路径聚合”机制的核心思想,在轻量化卷积基础上实现浅层细节与高层语义的充分交互融合。结构如图 2 中的 Neck 所示,该路径分为上采样和下采样特征增强两个阶段。

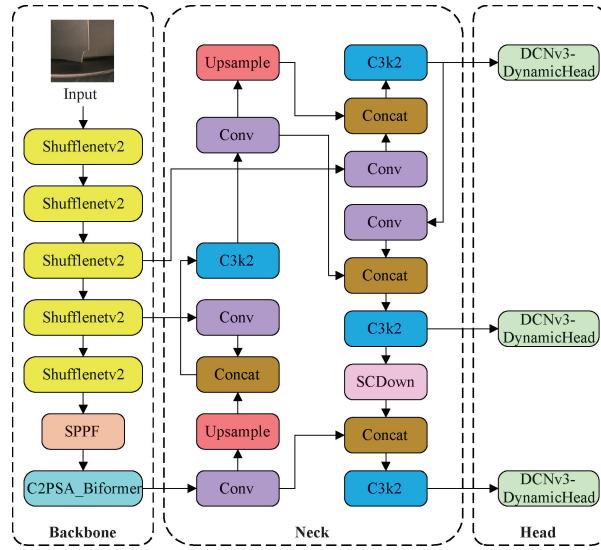


图 2 改进的 YOLOv11 架构图

Fig. 2 The diagram of the improved YOLOv11 architecture

上采样特征增强路径: 主干网络提取出的 3 个语义特征层首先分别经过 1×1 卷积通道变换,统一调整为 256 通道,用于特征对齐。最高语义层首先通过最近邻插值方式进行 2 倍上采样,使其尺寸与中间层保持一致,并将二者进行通道拼接后,输入至轻量化 C3k2 模块进行信息融合,生成语义增强后的中间特征图。上述融合后的中间特征图再次进行上采样,与最低层特征融合,经 C3k2 模块进一步提取浅层语义,形成语义增强后的浅层特征图。整个上采样路径利用高层信息向低层反向注入语义,有效弥补了浅层特征表达能力不足的问题,提高了模型对边缘模糊、小尺寸叶片缺陷的感知能力。

下采样特征整合路径: 为实现浅层与高层特征的双向流动,进一步提升多尺度信息整合能力,本文在上采样路径基础上引入下采样整合分支。具体地,浅层语义增强特征图通过 3×3 标准卷积以步长为 2 进行空间降采样后,与中间语义增强特征图进行通道拼接,再经由 C3k2 模块进行融合,形成新的中间融合特征。随后,该中间融合特征通过 YOLO10 中的 SCDown 模块再次降采样。SCDown 模块是一种面向轻量化目标检测网络设计的高效下采样单元,其设计初衷在于在保证特征表达能力的前提下,显著降低参数量与计算开销,提升模型在边缘设备上的推理效率。该模块主要结合了逐点卷积(pointwise convolution)与深度可分离卷积(depth wise convolution)的优势,具备良好的空间压缩能力与通道表达保留能力。

如图 3 所示,SCDown 模块首先利用核尺寸为 1×1 的逐点卷积对输入特征图进行通道维度的重构,实现特征通

道的压缩或扩展,从而增强通道表达能力并提升与后续结构的兼容性。随后,该模块引入带有步长为二的 3×3 深度可分离卷积,对各个通道进行独立的空间下采样操作。在该过程中,卷积操作仅在单一通道内部执行,避免了跨通道的参数耦合,有效降低了冗余计算与内存访问开销。相比标准卷积,SCDown 以串联式的逐点卷积与深度卷积结构为核心,显著减少了模型复杂度,具备更高的运算效率,特别适用于边缘计算场景下的资源受限设备如嵌入式系统或 NPU 平台。

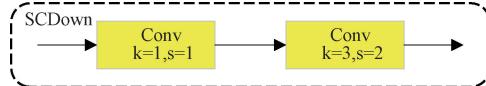


图 3 SCDown 结构图

Fig. 3 SCDown structure diagram

在本文所提出的改进 YOLOv11 架构中,SCDown 被集成于 Neck 结构的自底向上路径,用于对中层语义特征图进行高效下采样,并与高层语义特征进行融合。该设计

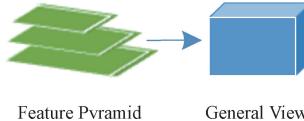


图 4 Dynamic Head 原理图

Fig. 4 Schematic diagram of Dynamic Head

在航空发动机叶片缺陷自动化检测任务中,本文引入 Dynamic Head 旨在增强模型对微小缺陷与复杂背景干扰的感知能力。然而,原始 Dynamic Head 所采用的 DCNv2 模块在建模精细形变与纹理细节方面仍存在一定局限,难以充分应对实际工业场景中存在的反光、污渍等非理想因素。为此,本文在不改变整体结构的前提下,聚焦于内部卷积方式的优化:将原结构中统一使用的 DCNv2 卷积,替换为更具空间感知能力的 DCNv3,从而构建出改进版

不仅增强了跨尺度语义信息的整合能力,还在确保模型检测精度的同时,有效抑制了网络参数量的增长,为模型的轻量化部署提供了关键支撑。

2.2 DCNv3-Dynamic 检测头

Dynamic Head 是一种将尺度感知、空间感知与任务感知有效融合于统一框架的检测头结构,其核心思想是将输入特征视为一个包含层级、空间位置和通道的三维张量,并在 3 个维度上分别引入轻量级注意力机制,以实现对多尺度目标、空间结构变化及不同检测任务的动态适应^[13]。如图 4 所示,Dynamic Head 依次堆叠了 3 个注意力模块:尺度感知注意力(Scale-aware Attention)作用于特征层级维度,通过统一不同层级特征的尺度,使网络更好地适配多尺度目标;空间感知注意力(Spatial-aware Attention)聚焦于图像中关键区域的空间位置,动态调整注意力权重以突出辨别性区域;任务感知注意力(Task-aware Attention)对通道维度进行建模,根据分类、中心点回归、边界框回归等任务动态激活不同通道,提高特征的任务适应性。

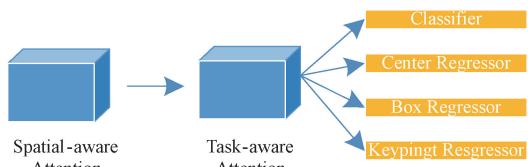


图 4 Dynamic Head 原理图

Fig. 4 Schematic diagram of Dynamic Head

DCNv3-Dynamic Head 检测头。

如图 5 所示,本文改进主要体现在:将当前层路径和上采样路径中的卷积操作由 DCNv2 替换为 DCNv3,以提升高分辨率特征对细节的建模能力;下采样路径则出于计算效率与稳定性考虑,仍保留 DCNv2。相比之下,DCNv3 在保留可变形感知优势的同时,具备更精细的可变形采样能力与中心特征调控能力,能更有效建模目标边缘、纹理与复杂空间变形,显著增强了模型在小目标与复杂背景下的鲁棒性。

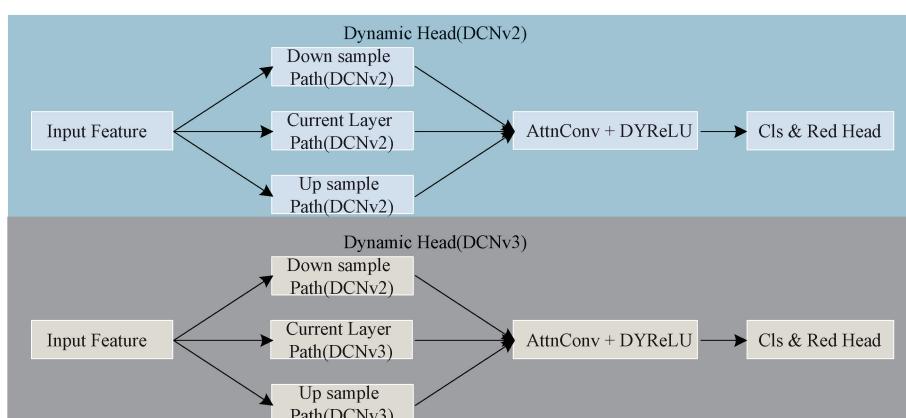


图 5 改进前后结构对比图

Fig. 5 The contrast chart of the structure before and after improvement

如表 1 所示,为采用 DCNv3 版本改进的 Dynamic Head 与原 DCNv2 版本在替换 YOLOv11 原始 Detect Head 后,于相同数据集及实验环境下的验证结果对比。实验表明,改进后的 DCNv3 版 Dynamic Head 在平均精度均值、准确率及召回率上均优于原 DCNv2 版本。

表 1 DCNv2 与 DCNv3 对比

Table 1 Comparison between DCNv2 and DCNv3

Models	mAP@0.5/%	Precision/%	Recall/%
Detect Head	83.1	88.2	79.5
DCNv2	84.5	88.7	81.2
DCNv3	85.0	89.1	82.5

2.3 注意力机制优化

BiFormer 核心机制为双层路由注意力,即在注意力操作中引入区域引导策略,将注意力机制分为区域级路径选择与局部 token 级信息交互两个阶段:首先根据全局语义对 query 进行路由,筛选出与其最相关的若干 key-value 区域,仅在这些高相关区域内执行精细的 token-to-token 注意力计算^[16]。相比传统 Transformer 的全局注意力或 YOLOv11 原 C2PSA 中的静态通道注意力, BiFormer 在降低计算复杂度的同时提升了注意力分配的动态性与内容相关性。

如图 6 所示,本文集成的 BiFormer 结构由局部增强模块(DWConv 3×3)、层归一化(LN)、双层路由注意力(Bi-Level Routing Attention)以及前馈网络(MLP)共同构成。模块采用多层残差连接结构,有效缓解深层网络信息衰减问题,实现局部纹理增强,从而在全局建模与局部细节表

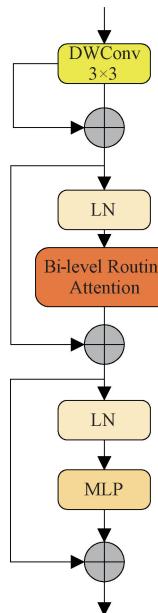


图 6 BiFormer 注意力模块

Fig. 6 BiFormer attention module

达之间达成平衡。在 YOLOv11 主干结构中,本文将 BiFormer 嵌入 Backbone 末端,用于对提取后的高级语义特征进行增强。

2.4 轻量化主干网络重构

本文引用 ShuffleNetV2 作为主干网络替代原有的主干结构,其核心结构如图 7 所示。ShuffleNetV2 是一种专为移动设备设计的高效卷积神经网络架构,其设计不仅关注 FLOPs(浮点运算次数),更强调实际推理速度、内存访问成本(MAC)和平台特性等直接指标。该架构提出了 4 条高效网络设计的实用准则:1) 输入和输出通道数应相等,以最小化内存访问成本;2) 避免过多的分组卷积,以减少 MAC;3) 减少网络碎片化,提升并行度;4) 降低逐元素操作的使用,减少计算开销。在结构设计上,如图 7(a)所示,ShuffleNetV2 引入了通道分割和图 7(b)通道重排机制。具体而言,输入特征图首先在通道维度上被平均分为两部分:一部分直接作为残差连接,另一部分经过一系列卷积操作后与前者在通道维度上拼接。随后,通过通道重排操作实现不同通道间的信息交互,增强特征表达能力。此外,ShuffleNetV2 避免使用分组卷积,减少了内存访问成本,提高了模型的实际推理速度。

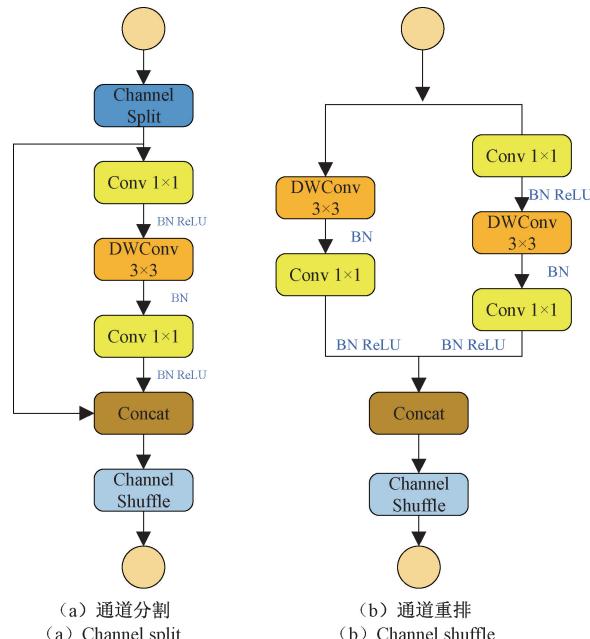


图 7 ShuffleNetV2 结构

Fig. 7 ShuffleNetV2 structure

3 实验结果和分析

3.1 数据集及实验平台

本文实验所用数据集为民航发动机叶片缺陷真实图像数据集,原始图像共计 3 228 张,涵盖 4 种缺陷类型,部分典型缺陷样本如图 8 所示。其中,图 8(a)弯曲缺陷表现为叶片在形态上出现明显的弯曲变形;图 8(b)烧蚀缺陷是叶片

表面因高温等因素,出现材料被侵蚀、表面粗糙且有烧蚀痕迹的情况;图 8(c)裂纹缺陷呈现为叶片上存在细微或较明显的裂隙,是材料内部或表面的断裂痕迹;图 8(d)材料缺失缺陷则是叶片局部出现材料的缺损,有部分区域的材料不

复存在,形成缺失的形态。为增强模型对实际复杂工况的适应性,对数据集进行了多轮增强处理,包括缺陷动态过程模拟、视角变化模拟、旋转变换及光照扰动等操作,最终扩充至 7 623 张图像,具体缺陷标签数量及解释如表 2 所示。

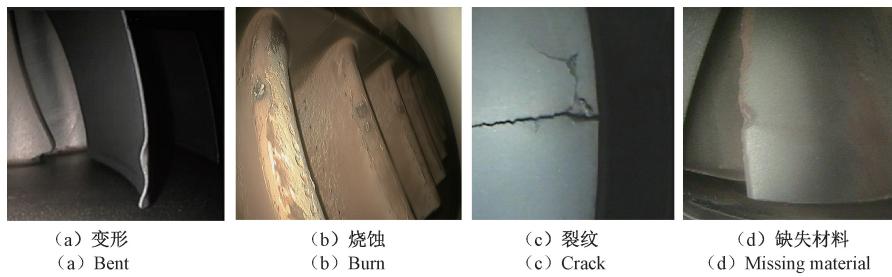


图 8 四种损伤类型

Fig. 8 Four types of damage

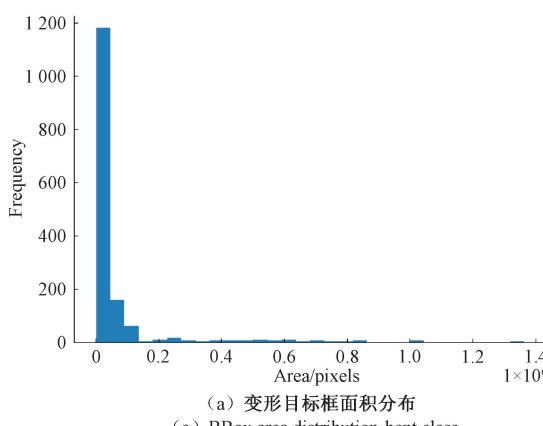
表 2 叶片缺陷标签数据集说明

Table 2 Description of the blade defect label dataset

图像数量	训练集/验证集/测试集	标签名称	标签含义	标签数量
7 623	6 099/762/762	Bent	变形	1 493
		Burn	烧蚀	2 269
		Crack	裂纹	2 630
		Missing Material	缺失材料	5 941

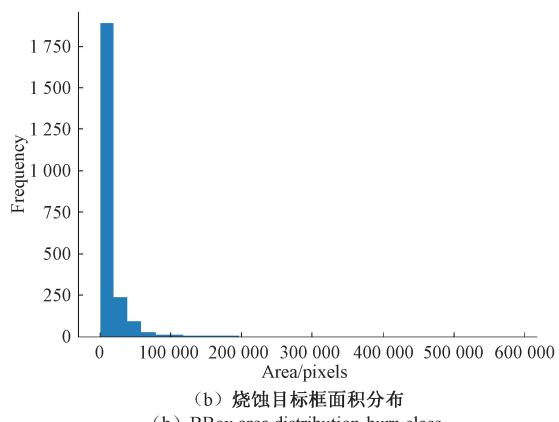
进一步对目标框面积分布进行统计分析,如图 9 所示,发现所有缺陷目标主要集中在小面积区间。如图 9(a)和(b)所示,Bent 和 Burn 类目标大多小于 0.05 M 像素;而图 9(c)和(d)中的 Crack 与 Missing Material 类尽管存在大尺寸样本,但整体呈现长尾分布。因此该数据特性对小目标检测能力和多尺度特征提取提出更高要求。

所有训练图像统一调整为 640 pixel \times 640 pixel,batch size 设置为 24,总训练轮次为 300。本文所有实验均在统一环境下完成:操作系统为 Windows 10,处理器为 Intel i9-12900K(3.20 GHz),显卡为 NVIDIA RTX 3090Ti,使用深度学习框架 PyTorch1.12.1,CUDA 版本为 11.8。为保证结果的一致性与可比性,所有实验采用相同的超参数设置。



3.2 模型评价指标

为全面评估本文提出的轻量化发动机叶片缺陷检测模型在精度与效率方面的综合性能,本文选取准确率(precision, P)、召回率(recall, R)、平均精度均值(mean average precision, mAP)、浮点运算量(GFLOPs)作为主要评价指标。其中,准确率用于衡量模型对缺陷目标的识别精度,反映其误报控制能力;召回率用于评估模型的漏检风险,尤其适用于对检测完整性要求极高的航空发动机应用场景;mAP@0.5 作为目标检测领域的核心评价指标,综合反映模型在各类缺陷上的检测性能;GFLOPs 用于衡量模型在推理过程中的计算复杂度,直接影响其在资源受限环境下的运行效率。上述多维指标可系统评估模型在准确性、轻量化及部署性能等方面的表现。



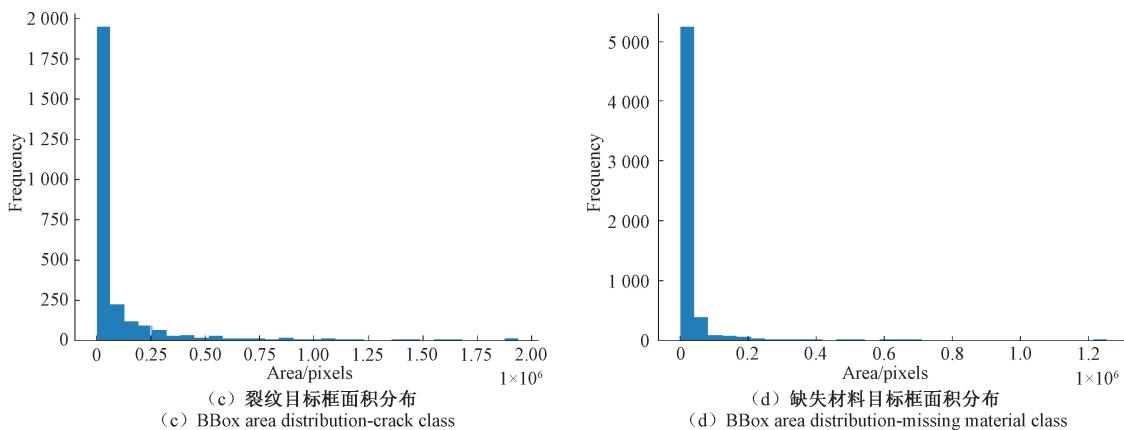


图 9 目标框面积分析
Fig. 9 Analysis of bounding box area

3.3 对比试验

为了验证所提出的改进 YOLOv11 模型在航空发动机叶片缺陷检测任务中的有效性与部署可行性,本文在统一的数据集与超参数设置下,与多种主流目标检测模型进行了系统对比实验,涵盖 Faster-RCNN^[18]、EfficientDet^[19]、YOLOv5s^[20]、YOLOv7-tiny^[21]、YOLOv8n^[22]、YOLOv10s^[23]、YOLOv11n^[24]、YOLOv12n^[25]等典型结构,结果如表 3 所示。本文提出的改进 YOLOv11 模型在综合精度与模型效率方面表现最为优异。其在 mAP@0.5 上达到 85.0%,相比于 YOLOv5s、YOLOv8n、YOLOv12n 与基线 YOLOv11n 均有提升,且接近 Faster-RCNN 的 87.8%。考虑到改进模型仅使用了 1.7 M 参数量与 4.6 G FLOPs,相比 Faster-RCNN 高达 137.1 M 参数与 370.6 G FLOPs 的计算代价,其精度和复杂度比具有显著优势。在

准确率与召回率方面,改进 YOLOv11 分别达到 89.3% 与 81.0%,较基线 YOLOv11n 有稳步提升;与 YOLOv12n 相比,其准确率虽略低 1.1%,但召回率表现出显著优势;而相较于 YOLOv7-tiny,其召回率虽稍显不足,准确率却具备明显优势。同时,与其他 YOLO 版本相比,该改进模型在准确率与召回率的综合性能上更优,这验证了其在检测精度与目标召回能力两方面的提升效果。这表明该模型在应对复杂纹理干扰与微小缺陷识别任务中具有更好的鲁棒性与检出能力。在模型轻量化方面,改进 YOLOv11 的参数量仅为 1.7 M,为所有对比模型中最小;其 FLOPs 也仅为 4.6 G,显著低于 YOLOv5s(16.0 G)和 YOLOv8n(8.1 G),在确保精度前提下显著降低了运算开销。这一特性使其更适合部署于如 RK3588 等国产高性能边缘计算平台,满足实时性与功耗约束。

表 3 不同算法模型的实验结果对比

Table 3 Comparison of experimental results among different algorithm models

模型	mAP@0.5/%	Precision/%	Recall/%	Parameters/M	GFLOPs
Faster-RCNN	87.8	89.1	84.9	137.1	370.6
EfficientDet	84.4	86.4	84.9	3.8	5.2
YOLOv5s	83.1	87.8	76.7	7.0	16.0
YOLOv7-tiny	78.5	70.1	84.6	13.1	6.0
YOLOv8n	84.0	88.6	77.9	3.0	8.1
YOLOv10s	77.3	83.2	72.1	7.2	21.4
YOLOv11n	83.1	88.2	79.5	2.6	6.4
YOLOv12n	82.9	90.4	74.2	2.5	5.8
Ours	85.0	89.3	81.0	1.7	4.6

此外,如图 10 所示,为改进前后模型的关键参数对比。在图 10(a)精度对比中,改进模型精度提升速度更快,能更早地达到较高的精度水平;图 10(b)召回率对比里,改进模型召回率显著更高,有效减少漏检率;图 10(c)平均精度对比显示,改进模型整体检测性能更佳;图 10(d)

mAP50:95 对比表明,改进模型在多 IoU 阈值下仍具高检测精度。综上,改进 YOLOv11n 各项指标均优于基线模型,验证了改进方法的有效性。

3.4 消融实验

为系统验证本文所提出各个结构改进模块对整体检测

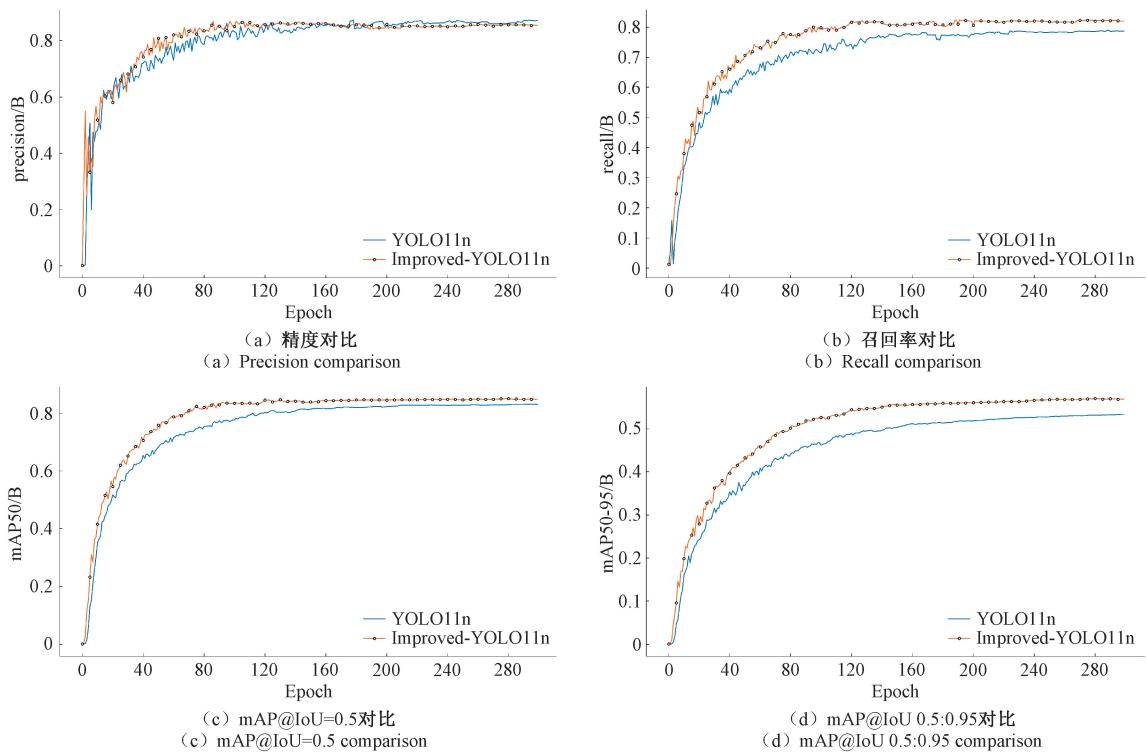


图 10 改进 YOLOv11n 与原始模型的精度、召回率、平均精度均值 mAP50, mAP50:95 对比

Fig. 10 Comparison of precision, recall, and mAP (mAP50, mAP50:95) between the improved YOLOv11n and the original model

性能与模型效率的影响,本文在 YOLOv11n 基线模型上依次引入 ShuffleNetV2 主干替换、特征融合结构 CFES、BiFormer

注意力机制、以及集成动态可变形卷积的 Dynamic DCNv3 检测头,并设计六组对照模型进行消融实验,如表 4 所示。

表 4 消融实验
Table 4 Ablation experiment

模型	ShuffleNetV2	CFES	BiFormer	DCNv3Dynamic	mAP@0.5/%	P/%	R/%	Parameters/M	GFLOPs
YOLOv11n					83.1	88.2	79.5	2.6	6.4
YOLOv11n-1	✓				81.5	85.9	78.6	1.8	5.3
YOLOv11n-2		✓			83.6	88.0	81.2	2.1	5.4
YOLOv11n-3			✓		84.1	88.5	82.2	2.9	7.3
YOLOv11n-4				✓	85.0	89.1	82.5	2.4	6.5
YOLOv11n-5	✓	✓			84.2	86.5	79.0	0.9	3.1
YOLOv11n-6	✓	✓	✓		84.6	89.0	81.5	1.5	4.4
改进 YOLOv11n	✓	✓	✓	✓	85.0	89.3	81.0	1.7	4.6

在单模块测试阶段,原始 YOLOv11n 模型的 mAP@0.5 为 83.1%,参数量为 2.6 M,GFLOPs 为 6.4。引入 ShuffleNetV2 后,模型参数显著降低至 1.8 M,GFLOPs 降至 5.3,但精度略有下降,说明轻量化虽可减小模型负载,但存在信息损失风险。替换 CFES 结构后,mAP@0.5/% 提升至 83.6%,P/% 和 R/% 分别提升至 88.0% 和 80.0%,验证其在多尺度特征融合方面的有效性。单独加入 BiFormer 模块,mAP@0.5/% 达 84.1%,R 提升至

82.2%,说明其对小目标与纹理区域具有良好建模能力。引入 DCNv3 Dynamic 检测头时,mAP@0.5/% 达 85.0%,R/% 提升至最高值 82.5%,进一步凸显其对复杂目标尺度建模的优势。

在多模块组合实验中,ShuffleNetV2 与 CFES 结构联合使用时,模型参数仅为 0.9 M,GFLOPs 降至 3.1,mAP@0.5/% 达 84.2%,在极低复杂度下仍保持良好检测性能。进一步引入 BiFormer 后,mAP@0.5/% 提升至

84.6%, R/% 达 81.5%。最终,在融合全部改进模块的改进 YOLOv11n 模型中,综合性能达到最优, mAP@0.5 提升至 85.0%, P/% 与 R/% 分别为 89.3% 和 81.0%, 在保持较低参数(1.7 M)与计算量(4.6 GFLOPs)的前提下,实现了最优的检测效果。

综上所述,各模块在小目标检测、尺度适应性与计算效率方面各具优势,组合使用可进一步提升模型的整体性能,验证了本文提出改进策略的有效性与工程实用性。



图 11 工业孔探仪

Fig. 11 Industrial borescope

页面如图 12 所示,在真实发动机上开展测试后,其检测帧数在 30 帧左右浮动,满足实际孔探需求。

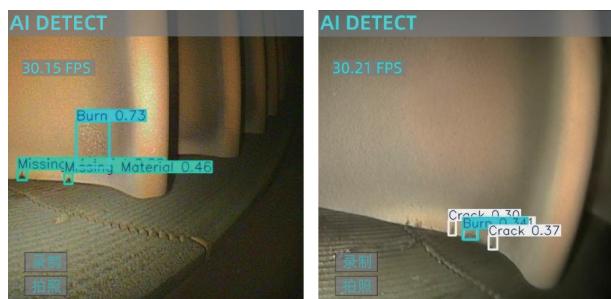


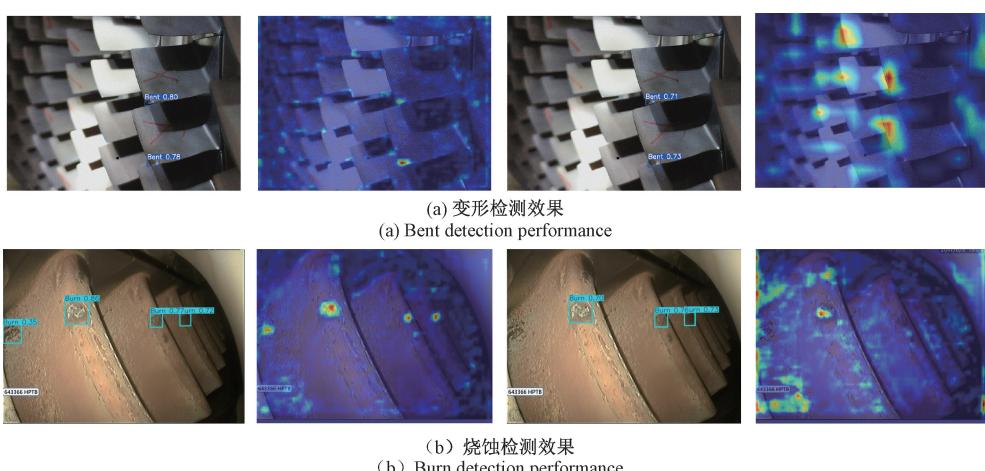
图 12 实时检测软件页面

Fig. 12 Real-time detection software interface

3.5 检测效果分析

为全面评估改进 YOLOv11n 模型在航空发动机叶片缺陷检测中的实际性能,本文将其与基线 YOLOv11n 模型一同部署至如图 11 所示的工业孔探仪中。其中,图 11(a)为搭载国产 RK3588 芯片的主机,作为操作与显示的核心部件,用于控制检测流程并呈现缺陷检测结果;图 11(b)为管线模块,负责传输探测信号与图像数据;图 11(c)为电源模块,为整个工业孔探仪提供电力支持。实时检测的软件

在相同拍摄位置下采集叶片缺陷图像,开展端到端的实时检测效果对比。本文不仅对比检测框结果,还引入 Grad-CAM^[26]热力图进行可视化分析,通过热力图的关注区域分布进一步探究两种模型在检测性能上的差异成因,如图 13 所示。在图 13(a)中,改进模型在两个缺陷区域的置信度更高,且热力图关注区域紧贴弯曲边缘。在图 13(b)中,改进模型可识别到包括微小烧蚀的所有缺陷,热力图中的关注区清晰集中;而基线模型存在漏检,注意力不集中,背景干扰明显。在图 13(c)中,改进模型对微小裂纹识别的置信度更高,且热力图准确聚焦裂纹位置;相比之下,基线模型置信度偏低,注意力分散。最后在图 13(d)中,改进模型识别到 3 处微小的缺失材料区域,而基线模型仅检测最明显一处。



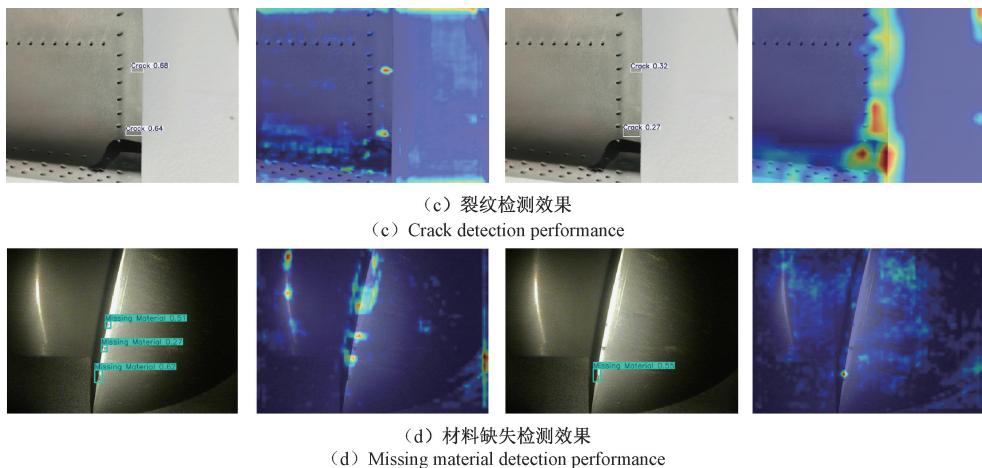


图 13 改进 YOLOv11n 与 YOLOv11n 检测效果对比可视化

Fig. 13 Visualization of detection effect comparison between the improved YOLOv11n and YOLOv11n

综上所述,改进模型的注意力更加集中于缺陷区域,而基线模型存在大量无效激活,漏检现象明显。引入 Grad-CAM 热力图分析不仅提升了模型性能对比的可解释性,也进一步验证了改进 YOLOv11n 模型在设计上的有效性与针对性。整体来看,改进模型在小目标感知、特征聚焦和背景干扰抑制等方面均显著优于基线模型,展现出更强的工程适用性。

4 结 论

本文围绕航空发动机叶片缺陷的智能检测需求,提出了一种融合轻量化与高性能的 YOLOv11 模型改进方法,并实现了其在 RK3588 工业内窥镜平台上的成功部署。通过构建覆盖多种缺陷类型的数据集,采用 ShuffleNetV2 主干网络、跨尺度特征增强结构 CFES、BiFormer 注意力机制与 Dynamic DCNv3 检测头,有效提升了模型对微小目标和复杂背景的适应能力。消融实验证明了各改进模块在精度、参数量和计算量上的协同优势,最终模型在保持较低计算开销的前提下实现了 85.0% 的 mAP@0.5 检测精度,参数量控制在 1.7 M,满足工业现场对实时性和可靠性的双重要求。通过 Grad-CAM 可视化分析进一步证明了模型关注区域的合理性与检测结果的可解释性。本文提出的轻量化改进模型已成功部署于一款搭载国产 RK3588 芯片的工业孔探仪,并在真实工业环境下完成现场测试,帧数维持在 30 帧左右,能够稳定实现航空发动机叶片缺陷的实时检测,充分验证了其在边缘设备上的实际应用能力。经过验证,该模型在检测精度、推理速度与部署适配性之间达成良好平衡,具备突工程实用性与推广前景。

参考文献

[1] ABDULRAHMAN Y, ELTOUM M A M, AYYAD A, et al. Aero-engine blade defect detection: A

systematic review of deep learning models[J]. IEEE Access, 2023, 11: 53048-53061.

- [2] 程亚茹,李涵,薛辉,等.航空发动机叶片表面损伤与检测研究进展[J].航空发动机,2024, 50(2):32-44.
CHENG Y R, LI T, XUE H, et al. Research progress on aeroengine blade surface damage and inspection[J]. Aeroengine, 2024, 50(2):32-44.
- [3] ZHOU N, LIANG R, SHI W. A lightweight convolutional neural network for real-time facial expression detection [J]. IEEE Access, 2021, 9: 5573-5584.
- [4] ZHANG D, ZENG N, LIN L. Detection of blades damages in aero engine[C]. 2020 Chinese Automation Congress(CAC). Shanghai: IEEE, 2020: 6129-6134.
- [5] LI X, WANG CH, JU H, et al. Surface defect detection model for aero-engine components based on improved YOLOv5 [J]. Applied Sciences, 2022, 12(14): 7235.
- [6] 魏永超,刘嘉欣,朱泓超,等.改进 YOLOv7 的航空发动机叶片损伤检测方法[J].航空发动机,2025, 51(1):133-139.
WEI Y CH, LIU J X, ZHU H CH, et al. Improved YOLOv7 damage detection method for aeroengine blade[J]. Aeroengine, 2025, 51(1):133-139.
- [7] 李文龙,王欣威,慕丽.基于改进 YOLOv8 的航空发动机叶片表面缺陷检测[J].组合机床与自动化加工技术,2024(12): 46-50.
LI W L, WANG X W, MU L. Detection of surface defects on aircraft engine blades based on improved YOLOv8[J]. Modular Machine Tool & Automatic Manufacturing Technique, 2024(12):46-50.
- [8] ZOU F. Review of aero-engine defect detection technology [C]. 2020 IEEE 4th Information

- Technology, Networking, Electronic and Automation Control Conference (ITNEC). Chongqing: IEEE, 2020: 1524-1527.
- [9] 史广,贺福强,史红燕,等.基于深度学习的航空发动机叶片缺陷检测技术综述[J/OL].航空动力学报,2025,1-21[2025-06-30].<https://doi.org/10.13224/j.cnki.jasp.20240411>. SHI G, HE F Q, SHI H Y, et al. Review of deep learning-based defect detection techniques for aero-engine blade[J/OL]. Journal of Aerospace Power, 2025, 1-21[2025-06-30]. <https://doi.org/10.13224/j.cnki.jasp.20240411>.
- [10] 余肖飞.基于嵌入式设备的航空发动机孔探损伤智能检测技术研究[D].广汉:中国民用航空飞行学院,2023.
- YU X F. Research on intelligent detection technology for aero engine borescope inspection damage based on embedded device[D]. Guanghan: Civil Aviation Flight University of China, 2023.
- [11] HE ZH, WANG K, FANG T, et al. Comprehensive performance evaluation of YOLOv11, YOLOv10, YOLOv9, YOLOv8 and YOLOv5 on object detection of power equipment [J]. ArXiv preprint arXiv: 2411.18871, 2024.
- [12] KHANAM R, HUSSAIN M. YOLOv11: An overview of the key architectural enhancements[J]. ArXiv preprint arXiv: 2410.17725, 2024.
- [13] DAI X. Dynamic head: Unifying object detection heads with attentions[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 7369-7378.
- [14] WANG W, DAI J, CHEN Z, et al. Internimage: Exploring large-scale vision foundation models with deformable convolutions[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 14408-14419.
- [15] ZHU X, HU H, LIN SH, et al. Deformable ConvNets v2: More deformable, better results[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 9308-9316.
- [16] ZHU L, WANG X, KE ZH, et al. Bifomer: Vision transformer with bi-level routing attention[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 10323-10333.
- [17] MA N, ZHANG X, ZHENG H T, et al. Shufflenet v2: Practical guidelines for efficient cnn architecture design[C]. European Conference on Computer Vision (ECCV), 2018: 116-131.
- [18] REN SH Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 39(6): 1137-1149.
- [19] 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(CVPR). Seattle: IEEE, 2020.
- [20] JOCHER G. Ultralytics YOLOv5 (7.0) [CP/OL]. 2022. <https://github.com/ultralytics/yolov5.html>.
- [21] WANG CH 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.
- [22] JOCHER G, CHAURASIA A, QIU J. Ultralytics YOLOv8 (8.0.0) [CP/OL]. 2024. <https://github.com/ultralytics/ultralytics.html>.
- [23] 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.
- [24] JOCHER G, QIU J. Ultralytics YOLO11 (11.0.0) [CP/OL]. 2024. <https://github.com/ultralytics/ultralytics.html>.
- [25] TIAN Y, YE Q, DOERMANN D. YOLOv12: Attention-centric real-time object detectors[J]. ArXiv preprint arXiv:2502.12524, 2025.
- [26] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: Visual explanations from deep networks via gradient-based localization [C]. 2017 IEEE International Conference on Computer Vision(ICCV). Venice, Italy, 2017: 618-626.

作者简介

敖良忠,硕士,教授,主要研究方向为航空发动机维护、状态监控和故障诊断。

E-mail:aolz@163.com

牟晋仟(通信作者),硕士研究生,主要研究方向为航空发动机缺陷智能检测。

E-mail:1014811105@qq.com