

基于深度学习的铸件表面缺陷检测： 系统综述与未来展望

杨信祥¹, *计效园¹, 余朋¹, 侯明君¹, 张子健¹, 殷亚军¹, 周建新¹

(1. 华中科技大学材料科学与工程学院, 材料成形与模具技术全国重点实验室, 湖北 武汉 430074)

第一作者: 杨信祥, 男, 博士生。E-mail: yxx_wyn327@163.com

*通讯作者: 计效园, 男, 教授, 博士。E-mail: jixiaoyuan@hust.edu.cn

摘要: 铸件缺陷检测是保障航天航空、汽车制造等领域产品质量的核心环节。传统方法因依赖人工或低效的机器学习模型难以满足高精度、实时性等需求。近年来, 基于深度学习驱动的目标检测技术为铸件缺陷智能检测带来了突破性进展。本文综述了铸件缺陷检测领域的发展历史, 涵盖了多种检测技术和评估指标, 回顾了有监督学习的两阶段和单阶段目标检测算法, 以及无监督学习方法在铸件缺陷检测中的研究进展, 同时探讨了当前面临的主要挑战, 并展望了未来研究的发展方向。

关键词: 铸件; 缺陷检测; 目标检测; 深度学习

Deep learning-based detection of casting surface defects: a systematic review and future perspectives

YANG Xin-xiang¹, *JI Xiao-yuan¹, YU Peng¹, HOU Ming-jun¹, ZHANG Zi-jian¹, YIN Ya-jun¹,
ZHOU Jian-xin¹

(1. National Key Laboratory of Materials Forming and Die Technology, Huazhong University of Science and Technology, Wuhan, Hubei Province, 430074, China)

Abstract: Casting defect detection is a core link to ensure product quality in fields such as aerospace and automotive manufacturing. Traditional methods are difficult to meet the requirements such as high precision and real-time performance because they rely on manual labor or inefficient machine learning models. In recent years, object detection technology driven by deep learning has brought breakthrough progress to the intelligent detection of casting defects. This paper reviews the development history in the field of casting defect detection, covering a variety of detection techniques and evaluation indicators. It examines the two-stage and single-stage object detection algorithms based on supervised learning, as well as the research progress of unsupervised learning methods in casting defect detection. Meanwhile, it discusses the main challenges currently faced and looks forward to the development direction of future research.

Keywords: casting; defect detection; target detection; deep learning

1 前言

铸造行业作为制造业的基础支柱产业, 是航空航天、轨道交通、能源装备等高端技术领域的重要支撑^[1-5]。铸件质量的核心评判标准在于其表面及内部缺陷的分布与严重程度, 这些缺陷可能引发结构

失效、疲劳断裂等灾难性事故, 直接威胁工业装备的服役安全^[6, 7]。因此, 高效精准的缺陷检测是保障工业安全的关键步骤。传统的缺陷检测主要依赖人工筛查, 存在操作流程繁琐、检测效率低、主观误差大、结果可追溯性差等问题, 难以满足现代工业对高精度、自动化、实时化的严苛需求。随着计算机视觉与人工智能技术的快速发展, 基于深度学习

的缺陷检测方法凭借其强大的特征学习能力和自动化分析优势,为铸件质量检测提供了革命性的解决方案^[8-10]。特别是卷积神经网络 (Convolutional Neural Networks, CNN) 在目标检测领域的不断演进,推动了深度学习算法在工业表面缺陷检测中的广泛应用,显著提升了检测的精度与鲁棒性^[11, 12]。

基于深度学习的表面缺陷检测主要分为有监督与无监督两类。有监督方法 (如 Faster R-CNN、YOLO 系列) 通过大规模标注数据驱动模型学习缺陷特征,已在工业场景中展现出优异的检测性能^[13-17];而无监督方法 (如自编码器、GAN) 则通过重构误差或特征分布对比实现缺陷定位,有效降低了对标注数据的依赖,为小样本与未知缺陷检测提供了新路径^[18-23]。近年来,国内外学者围绕这两类方法开展了大量创新性研究,通过算法优化、模型轻量化、多模态融合等策略不断突破技术瓶颈,推动着工业质检向智能化、实时化方向迈进。

本文系统综述了铸件表面缺陷检测领域的最新研究进展与技术挑战。全文内容组织概述如下:第 2 节深入剖析了铸件缺陷的主流检测技术 (光学、声学、电磁学) 及其评估指标体系;第 3 节聚焦有监督检测方法,对比分析了两阶段与单阶段目标检测算法的优化策略与实际应用;第 4 节探讨了无监督检测框架的核心原理与典型模型 (自编码器、生成对抗网络等);第 5 节总结现有技术的局限性,并从数据增强、模型轻量化、多模态融合等维度展望未来发展方向,为工业缺陷检测技术的迭代升级提供理论参考与实践指导。

2 检测技术和评估指标

2.1 缺陷检测技术

铸件表面缺陷的精准检测是保障装备制造安全性与可靠性的关键环节,直接影响工业装备的服役寿命与故障风险。基于非接触、无损伤的优势,无损检测技术通过避免与金属表面直接接触,最大程度降低对检测结果与工件完整性的干扰,在提升检测安全性的同时显著降低工业维护成本,成为缺陷

检测的首选方法^[24]。主流的无损检测技术根据物理场作用原理可分为三类:基于光学 (如 X 射线成像)、声学 (如超声波探伤) 及电磁学 (如涡流检测) 的技术体系,通过分析光波衰减、声波反射或电磁感应变化,实现铸件表面与内部缺陷的无损定位与定量评估^[25-28]。本文总结比较了各类铸件缺陷无损检测技术如表 1 所示。

2.1.1 基于光学的无损检测技术

X 射线检测技术使用 X 射线机或加速器作为射线源,从铸件一侧照射并穿透铸件。密度较大的区域吸收更多射线,导致射线强度减弱,而缺陷区域因密度较低,射线强度相对增高。通过荧光屏、胶片或数字影像接收器,可以记录射线穿透后的图像,揭示铸件内部缺陷,对气孔、夹杂物等体积型缺陷具有较高灵敏度。Lu 等人^[25]使用同步辐射成像技术对铸件成型过程中的组织演变进行实时成像,能够清晰地显示铸件内部的缺陷,如富集区和气泡,并通过不同冷却速率下的成像对缺陷的演化进行监测,从而为调整冷却速率以避免缺陷提供数据支持。Fayad 等人^[29]提出的新的 X 射线数字图像技术通过合成参考图像解决了 X 射线路径集成问题,有效跟踪多个物体的运动,为复杂环境中的 X 射线成像应用提供了新方法。

2.1.2 基于声学的无损检测技术

超声波检测通过压电换能器发射高频超声波,当声波在遇到铸件内部缺陷 (如裂纹、孔洞) 时产生反射或折射,检测系统通过分析反射波的强度和时间差,确定缺陷的存在、位置、深度和形态。Bai 等人^[30]提出的通过全跳变超声数据和贝叶斯推断改进裂纹缺陷表征的方法,结合直接和全跳变路径的散射矩阵,有效提高了裂纹尺寸和角度的表征精度。Tkocz 等人^[31]设计了一种高功率四通道相控阵 EMAT 脉冲系统,通过相位延迟控制聚焦超声波束,并成功完成了在 22.5 cm 厚的粗糙铸钢板上的缺陷检测实验,实验结果表明该系统能有效增强信号强度,并检测到深达 16 cm 的人工缺陷。

表 1.1 铸件缺陷无损检测技术评估对比

Tab. 1 Evaluation and comparison of Non-destructive testing techniques for casting defects

检测技术	物理学分类	检测指标	铸件类型	缺陷类型	优点	应用场景
X 射线成像技术	光学	图像质量	各种材质铸件	气孔、夹杂类等体积型缺陷	可探测复杂铸件、直观、便于存储、三维缺陷表达强	适用于内部缺陷检测,复杂形状铸件
超声波检测技术	声学	反射波强度与时间差	各种材质铸件	裂纹、孔洞等内部缺陷	快速、成本低、穿透能力强、无污染	适用于内部缺陷检测,适合多种铸件材料
涡流检测技术	电磁学	阻抗与信号强度变化	铁磁性铸件	裂纹、盲孔等表面或近表面缺陷	无需耦合剂、高温检测、快速、便于存储	适用于表面缺陷检测,尤其是高温环境下

2.1.3 基于电磁学的无损检测技术

涡流检测技术基于电磁感应原理，利用交变电流的线圈在铸件表面激发涡流，铸件的形状、尺寸和缺陷会导致涡流的分布和强度变化，从而影响线圈的阻抗或感应电压。通过分析这些变化，可以判断缺陷的位置、大小和性质。Betta 等人^[32]提出的多频涡流激励信号和谱分析方法，通过多频信号激励涡流传感器并分析其输出信号的谱图变化，有效提高了缺陷的分类与检测性能。Farag 等人^[33]提出的基于涡流传感器的无损检测方法，通过绝对探头和反射探头比较检测不同大小和深度的人工缺陷，评估了其在裂纹和盲孔检测中的性能。Xu 等人^[34]提出了一种基于正交磁场激励的新型高灵敏度远场涡流传感器，能够显著提高缺陷检测灵敏度，尤其在检测深度小于壁厚 40% 的轴向裂纹时表现出优异的检出能力。

2.2 评估指标

深度学习模型的性能评估是机器学习研究中的重要部分。真阳性 (TP) 表示模型正确预测为正类的样本数，真阴性 (TN) 表示模型正确预测为负类的样本数，假阳性 (FP) 表示模型错误地将负类样本预测为正类的样本数，假阴性 (FN) 表示模型错误地将正类样本预测为负类的样本数。基于这些基本分类结果，模型评估中用于缺陷检测的关键指标包括准确性 (Accuracy)、精度 (P)、平均精度 (AP)、平均精度均值 (mAP)、召回率 (Recall) 和 F1 分数。这些指标的合理评估和综合考量可以确定模型的优劣，指导后续的优化和改进工作，其定义如下：

(1) Accuracy: 衡量模型预测正确的样本数占总样本数的比例，由以下公式确定：

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.1)$$

(2) P: 在单幅图像中准确检测的框占有所有预测框的比例，由以下公式确定：

$$P = \frac{TP}{TP + FN} \quad (2.2)$$

(3) Recall: 正确预测出的正类样本数占实际正类样本数的比例，由以下公式确定：

$$Recall = \frac{TP}{TP + FN} \quad (2.3)$$

(4) AP: 平均精度通过对精确率-召回率曲线下的面积进行积分，量化模型的检测性能，由以下

公式确定：

$$AP = \int P(R) dR \quad (2.4)$$

其中 P 代表 Precision, R 代 Recall。

(5) mAP: 平均精度均值是所有类别 AP 的平均值，由以下公式确定：

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (2.5)$$

(6) F1 分数: 综合考虑了精度和召回率的调和平均数，用于平衡模型的查准率和查全率，由以下公式确定：

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (2.6)$$

(7) 参数量和 FPS: 在深度学习模型中，参数量 (Params) 指模型在训练过程中需要优化的学习变量，直接影响模型的复杂化度、存储需求和计算开销。帧率 (FPS) 用于衡量模型处理图像的速度，它测量系统每秒可以处理的帧数，反映了模型的实时响应能力。

3 有监督缺陷检测方法

有监督缺陷检测方法依赖于标记数据集进行训练，通过学习输入数据与相应标签之间的关系，进而实现对新样本的准确预测。在监督学习中，训练集和测试集是不可或缺的，其中训练集中的样本必须被标记，模型通过分析标记数据中的规律，将其应用于测试集进行预测。监督缺陷检测方法主要基于目标检测算法，分为两阶段算法和单阶段算法两类，图 1 展示了两阶段和单阶段目标检测深度学习算法的比较。

两阶段目标检测算法通常围绕候选区域的概念进行处理，首先生成候选框，提取其中的特征，再进行对象分类和位置回归等步骤，代表性的算法包括 R-CNN、Faster R-CNN 和 Mask R-CNN 等，这些算法通过逐步优化候选区域，提高了检测精度。而单阶段目标检测算法则省略了候选区域生成的中间步骤，直接输出目标的类别概率和位置坐标，典型的单阶段算法如 YOLO 系列和 SSD 等，这些算法由于其更高的检测速度和实时性，在工业缺陷检测中得到了广泛应用。

3.1 两阶段缺陷检测方法

两阶段目标检测方法通过分步处理实现高精度

的缺陷检测。首先由区域提议网络（RPN）生成候选区域，利用全图像卷积特征提取多尺度特征，并筛选潜在目标位置。随后，通过级联检测器对候选区域进行分类与边界框精修，结合特征金字塔网络（FPN）融合不同层级的语义信息，完成目标分类与边界框精修。

2014 年，Girshick 等人^[35]提出 R-CNN 算法，通过选择性搜索生成候选区域，结合卷积神经网络（CNN）提取特征，并使用支持向量机（SVM）分类及边框回归，显著提高了目标检测的精度。然而，R-CNN 也面临着需要逐一从 2000 个候选框中提取特征、计算冗余等问题。为了解决这些问题，2016 年，Ren 等人^[36]提出了 Faster R-CNN 算法，过引入区域提议网络（RPN）解决了传统方法中区域提议计算的效率瓶颈问题，RPN 与检测网络共享全图像卷积特征，以近乎零成本生成高质量候选区域，并通过端到端训练同时优化目标边界和目标性得分。图 2 描述了 Faster R-CNN 算法结构。Du 等人^[37]提出基于改进 Faster R-CNN 的汽车铝制铸件缺陷识别系统，针对 X 射线图像检测场景，引入特征金字塔网络（FPN）优化小缺陷检测，并在训练阶段结合数据增强缓解数据集不足。实验表明，改进算法小缺陷检测平均精度均值（mAP）较原模型提升 40.9%，借助 ROI 池化技术使边界盒定位准确率提高 23.6%，验证了多尺度特征融合与数据增强策略在工业检测中的有效性。2017 年，Mask R-CNN 算

法由 He 等人^[38]提出，通过在 Faster R-CNN 基础上引入掩码预测分支，并使用 ROI 对齐技术替代传统的 ROI 池化，进一步提升了边界框定位精度和支持实例分割。虽然 Mask R-CNN 在目标检测和像素级分割方面表现出色，但需要依赖大规模标注数据，且计算复杂度较高，导致在实时检测场景中的应用受限。Yang 等人^[39]提出改进的 Mask R-CNN 算法用于金属结构件微裂纹任务多维度优化，通过重构 FPN 强化底层特征细节特征的利用效率；在 ResNet 骨干网络中嵌入 DCNv2 与 CBAM 适应复杂形态；采用 DIOU Loss 提升定位精度。实验显示，改进算法在自建数据集上 mAP 达 78.6%，较原始模型提升 40.9%，边界框定位准确率提升 23.6%，定位误差降 9.78%。消融实验证实各模块有效，其中 FPN 优化对性能贡献最大，mAP 提升 4.4%。2020 年，Wang 等人^[40]提出了一种基于自注意力引导模块（SGM）的两阶段深度学习模型，用于铝合金铸件 X 射线图像的细微缺陷检测。该模型通过通用特征网络（GFN）提取全局特征，再通过嵌入 SGM 的细微特征网络（SFN）强化复杂背景下的细微缺陷特征提取能力。实验结果表明，改进后的模型在自建数据集上实现了 91.17% 的准确率，较传统模型 VGG-16（58.57%）和 ResNet-50（58.66%）提升约 32.6%；精度、召回率和 F1 分数分别达到 89.53%、89.17% 和 89.31%，较基线模型提升超过 40%。

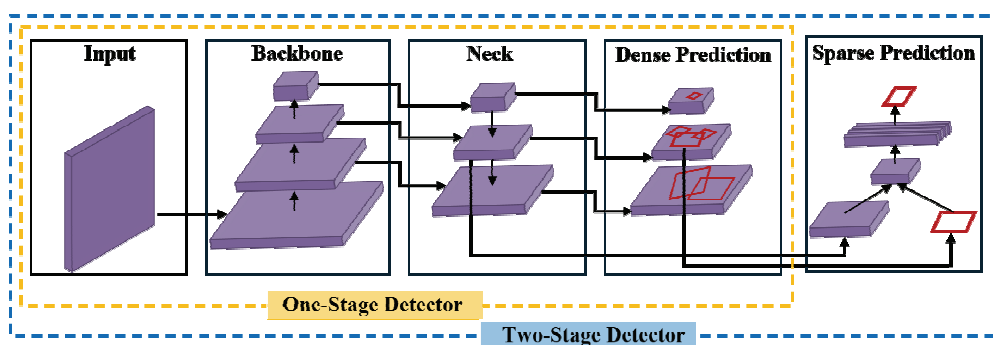


图 1 两阶段和单阶段目标检测深度学习算法的比较

Fig.1 Comparison of two-stage and one-stage object detection deep learning algorithms

综上所述，两阶段目标检测方法凭借其高精度特性，在缺陷检测中展现出显著优势，尤其在识别如微裂纹、气孔等微小缺陷及如噪声、遮挡等复杂干扰条件下表现卓越。然而，此类方法仍面临模型复杂度高、推理速度受限等挑战。未来研究需进一步优化两阶段算法的架构精简性与参数效率，在维持高精度的前提下提升实时性，以满足工业在线检

测对速度与可靠性的双重需求。

3.2 单阶段缺陷检测方法

单阶段目标检测算法通过端到端框架直接预测物体的位置和类别，无需生成候选区域，实现高效的单步处理。它将目标检测重新定义为回归和分类的联合优化问题，简化了检测过程并显著提升了推理速度。典型算法如 YOLO 和 SSD 系列，以其高

效性和准确性成为该领域的基准。

YOLO 系列已经从 YOLOv1 发展到最新的迭代 YOLOv11。2015 年, YOLOv1 首次将目标检测转化为端到端回归任务, 实现了实时检测, 但单网格单目标的限制导致多目标漏检率高^[41]。2016 年, YOLOv2 引入锚框 (Anchor Boxes) 机制和使用多尺度训练方法, 以改进锚框预测并提高召回率^[42]。2018 年, YOLOv3 采用 Darknet-53 骨干网络与 FPN 多尺度特征融合, 通过 3 种尺度预测框实现小目标检测性能提升, 但模型参数量增加, 推理速度下降^[43]。2020 年, YOLOv4 引入了 CSPDarknet53 骨干网络、SPP 模块、PANet 路径聚合模块以及 Mish 激活函数、Mosaic 数据增强、CIoU 损失等多种改进技术, 显著提升了目标检测的精度和速度, 同时优化了模型的训练效率, 使其能够在常规 GPU 上实现快速、高质量的实时目标检测^[44]。2022 年, YOLOv6 采用了先进的网络设计和量化技术, 进一步提升了模型的精度和速度, 在不同规模的工业应用中实现了良好的平衡。通过引入自蒸馏策略、改进的损失函数和量化训练, 能够在工业环境中提供高效的实时检测性能^[45]。2022 年, YOLOv7 提出可训练的 “Bag-of-Freebies” 方法, 采用 E-ELAN 骨干网络与复合缩放策略, 通过计划重参数化卷积和粗到细标签分配优化训练过程, 在 5-160 FPS 范围内实现速度与精度的全面提升^[46]。2024 年, Wang 等人推出了 YOLOv9, 融合了可编程梯度信息 (PGI) 调整权重, 并将 CSPNet 和 ELAN 合并到广义高效层聚合网络 (GELAN), 与 YOLOv8 相比, 深度模型的参数量减少了 49%, 计算量减少了 43%^[47]。2024 年, Wang 等人推出了 YOLOv10, 通过一致的双任务方法实现了无需非极大值抑制 (NMS) 的训练, 同时提出了一种整体效率-精度驱动模型设计策略^[48]。2024 年, Khanam 等人^[49]推出了 YOLOv11, 它引入了 C3k2、SPPF 和 C2PSA 等组件, 这些创新提高了模型在特征提取等方面的性能。YOLOv11 在 COCO 数据集上实现了约 54.5% 的 AP, 与 YOLOv10 相比, 在相同性能下, YOLOv11 的延迟更低, 参数更少, 优于许多现有的最先进的方法。图 3 为 YOLO 系列目标检测模型的发展脉络。

随着 YOLO 系列算法的持续演进, 在工业质检领域的应用潜力不断扩大, 特别是在铸件表面缺陷检测场景中展现出显著优势。2021 年, Duan 等人^[50]通过改进 YOLOv3 网络, 提出专用于铸件缺陷检测的 YOLOv3_134 模型, 创新性地引入双密度卷积层结

构和扩展预测尺度, 使气孔、缩松等小缺陷的检测 mAP 较原模型提升 26.1%, 同时保持 39 FPS 实时性, 显著优于 Faster R-CNN 和 YOLOv2。同年, Li 等人^[51]针对航空发动机部件缺陷, 通过 K-means 聚类优化先验框参数并改进参数调整算法以增强先验框与特征层的匹配度, 提升了特征提取能力, 使改进的 YOLOv4 在专用数据集上达到 82.67% mAP, 较原模型提升 4.55%。2023 年 Parlak 团队^[52]将 YOLOv5 应用于铝合金铸件 X 射线图像检测, 在 Al-Cast 数据集上实现 0.956 mAP, 其中 YOLOv5n 模型更以 132 FPS 的速度展现卓越效率。2024 年, Ge 等人^[53]提出了一种改进的 YOLOv5 算法, 用于检测小型铝铸件涡轮表面的小目标缺陷。通过数据增强策略平衡不同类别样本的分布; 使用 K-means++ 算法优化先验框, 使其更适合本数据集; 在主干网络中引入 CA 注意力机制模块, 以增强模型对小目标特征的提取能力; 以及添加小目标检测层, 专门针对小尺寸缺陷进行优化, 使改进的 YOLOv5 在铝铸涡轮缺陷检测中达到 97.8% mAP。2024 年, Liu 团队^[54]提出了一种改进的 YOLOv7 算法 (REPBB-YOLOv7), 通过引入 Rep-MBConv 模块的重新参数化主干网络 (Rep-backbone), 以增强特征提取能力并减少网络参数; 采用双向特征金字塔网络 (BiFPN) 实现加权特征融合; 并在关键网络位置集成注意力机制, 以抑制无关背景细节, 在连铸板坯检测中实现 97.8% mAP@50。2024 年, Jiang 等人^[55]提出了一种改进的 YOLOv8 算法, 用于检测航空发动机叶片的微观缺陷。引入了 MS-Block 模块以增强多尺度特征融合, 替换了 C2f 模块的瓶颈结构为扩张残差 (DWR) 模块以增强感知能力, 并添加了高效多尺度注意力 (EMA) 机制以提高对小目标的检测效果; 将损失函数替换为 Inner-IoU, 以提高模型的泛化能力和收敛速度, 改进后的 YOLOv8 在 EMT 数据集上实现了 97.3% 的 mAP, 在 MMT 数据集上实现了 95.6% 的 mAP, 相较于原始 YOLOv8 分别提高了 1.5% 和 1.8%。2024 年, Wang 团队^[56]提出改进 YOLOv8 算法 SLGA-YOLO 用于铸件表面缺陷检测, 通过 SlimNeck 模块简化颈部结构、融合 SimAM/LSKA 注意力机制增强特征聚焦, 结合 Ghostnet 的 GCML 模块和 Alpha-EIoU 损失函数优化特征提取与收敛效率, 该算法在自建数据集和 NEU 数据库上的 mAP@0.5 分别提升 3% 和 1.6%, 同时显著降低模型参数量与计算成本。2025 年 Wang 等人^[57]推出的

Edge-YOLO 通过边缘增强骨干网络,在多个工业数据集上实现 mAP@50 最高 10.3%的提升,同时降低 69.8%计算复杂度。这些持续创新不仅印证了 YOLO 系列架构的强扩展性,更通过针对性优化解决了工

业缺陷检测中小目标识别、复杂背景干扰、实时性要求等核心难题,为智能制造提供了可靠的技术支撑。

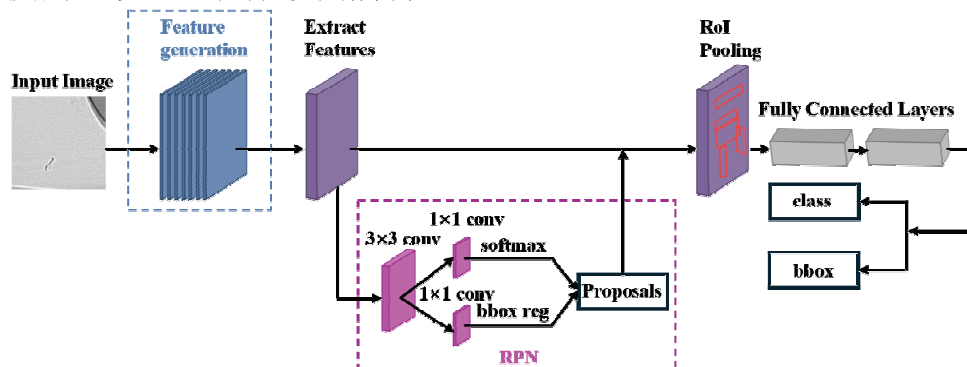


图 2 Faster R-CNN 算法的结构图
Fig.2 The structure diagram of Faster R-CNN

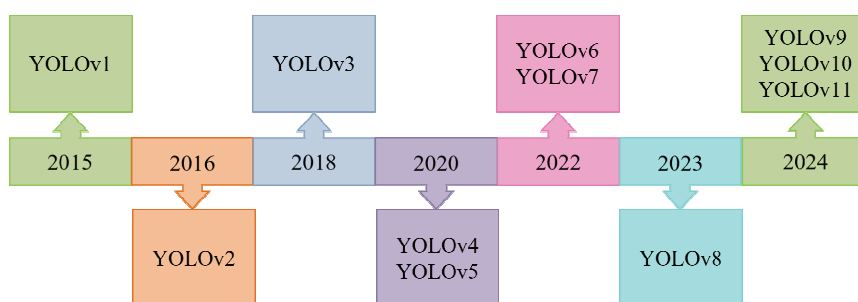


图 3 YOLO 系列目标检测模型的发展脉络

Fig.3 The development context of the YOLO series of object detection models

单阶段目标检测算法凭借高效实时特性在缺陷检测中优势显著,通过端到端框架与多尺度特征融合,结合优化策略平衡速度与精度,适配多场景数据集。但复杂背景下小目标检测能力不足,且部分改进模型因引入复杂模块导致参数数量和计算量增加,影响推理速度。未来研究需聚焦优化特征提取、降低计算复杂度,通过轻量化网络、注意力机制及自适应学习策略,在保持高效性的同时提升复杂场景鲁棒性与实时性,以满足工业在线检测对速度与精度的双重需求。

4 无监督缺陷检测方法

针对有监督缺陷检测方法对标注数据的高度依赖及其在未知缺陷检测中的局限性,研究者逐步转向无监督缺陷检测方法。通过自编码器 (AutoEncoder) 重构误差分析、生成对抗网络 (GAN) 的特征分布对比等方法,直接从无标注数据中挖掘潜在缺陷模式,突破监督方法受限于标注类别的固有瓶颈。图 4 描述了 GAN 算法结构。

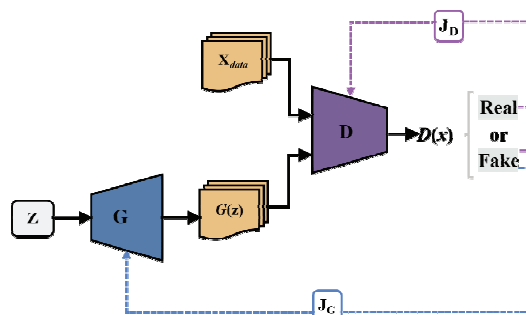


图 4 GAN 算法结构图

Fig.4 The structure diagram of GAN

Autoencoder 通过编码器压缩数据至低维潜在空间提取特征,再经解码器重构以最小化重建误差,其无监督特性在表面缺陷检测中优势显著:通过对比正常样本的重建误差定位异常区域或利用编码特征分类缺陷,有效降低对标注数据的依赖,适用于小样本和实时检测场景。2019 年, Yang 等人^[58]提出基于多尺度特征聚类的全卷积自动编码器 (MS-FCAE),通过多尺度子网络运行,有效重建纹理背景图像,提升特征映射的判别能力,在自建

数据集上达成 92.0% 的检测精度,检测时间仅 82 毫秒。2021 年, Tang 等人^[59]的基于卷积自编码器(CAE)的无监督框架用于压铸 X 射线图像检测,结合霍夫变换提取感兴趣区域,仅用无缺陷图像训练,实现高精度分类与定位,分类准确率达 97.45%,且在小训练集下表现优异,优于传统人工检查。2023 年,Chamberland 等人^[60]利用 CNN 自编码器(CAE),仅用无缺陷图像训练,通过图像重建损失聚类,实现铸件表面缺陷的无监督检测,平均 F1-score 达 89.14%,准确率为 88.52%,媲美监督学习方法。2024 年, Zhang 等人^[61]提出基于双向特征提取的自动编码器(BiSDE),结合特定特征提取架构和 Wasserstein 距离损失函数,优化小目标边界定位,检测精度和效率显著提升,在小目标检测中取得至少 5.8% 的 mAP 提升。GAN 凭借生成器与鉴别器的动态博弈机制,生成器拟合真实数据分布以生成高保真样本,鉴别器区分真实与生成数据,既能合成逼真缺陷样本缓解数据稀缺,又能通过重构差异识别异常。GAN 在表面缺陷检测方面取得了显著的成功,研究人员提出了各种改进方法来提高其性能。2023 年, He 等人^[62]提出基于生成对抗网络(DG-GAN)的高质量缺陷图像生成方法,引入了进阶生成对抗、D2 对抗损失函数、循环一致性损失函数、数据增强模块和自注意力机制,以提高网络的生成能力和训练稳定性,生成高质量、多样化的缺陷图像,有效扩充原始数据集,缓解缺陷图像采集的困难。结果显示使用 DG-GAN 生成的缺陷图像在 YOLOX 检测模型中取得了显著提升, YOLOX 的检测精度分别提高了 6.1% 和 20.4%,并且 FID 得分比现有方法显著降低。2024 年, García-Pérez 等人^[63]提出基于可扩展条件 Wasserstein GAN(SCWGAN)的合成缺陷生成方法,用于提升 GDxRay 铸件数据集的自动缺陷识别(ADR)。该方法将无缺陷图像与 SCWGAN 生成的合成缺陷结合,通过掩码图像确定合成缺陷的位置、噪声水平及大小等参数。实验表明,与仅用真实图像训练的基线模型相比,在 mAP@IoU=0.5:0.95 指标上提升 17%, mAP@IoU=0.50 指标达 96.0%,创该数据集准确率新高。

5 总结与展望

5.1 总结

本文全面综述了基于深度学习的铸件表面缺陷检测算法,重点介绍了有监督和无监督两类方法的

核心原理、发展脉络、关键算法及实际应用案例,并对相关检测技术和评估指标进行了梳理。基于有监督的缺陷检测方法已经逐渐成熟,并展示了优异的性能。然而,这些方法依赖于大量带有精确标注的训练数据,而数据的收集与标注成本高昂,导致其难以与实际工业场景完全契合。此外,有监督方法无法有效处理新类别缺陷的检测问题。尽管无监督方法显著降低了数据标注需求,并能够检测未知类型的缺陷,但在定位精度和可解释性方面,仍难以替代有监督方法。

5.2 展望

在当前研究的不足之外,该领域还蕴藏着诸多极具潜力的研究方向,亟待进一步的深入探索与挖掘。

(1) 增强的数据合成与增强策略:加强生成对抗网络(GAN)等数据合成方法,以生成高质量的缺陷图像和多样化的样本。结合自动数据增强策略,增加数据的多样性,缓解标注数据不足的问题,进一步提升检测模型的鲁棒性和性能。

(2) 结合监督与无监督学习提升模型的泛化能力:通过结合有监督学习和无监督学习方法,可以有效缓解数据标注不足的问题。无监督学习有助于发现新的缺陷类型,提升模型的泛化能力,而有监督学习则能够提高检测精度。探索如何将两者有机结合,以增强模型的适应性。

(3) 优化模型的轻量化和实时性:随着工业应用对实时性的需求增加,需要进一步优化深度学习模型的轻量化设计,以减小计算开销并提高处理速度。结合边缘计算和云计算技术,可以使模型能够在生产线附近的边缘设备上快速推理,实现低延迟的实时缺陷检测。

(4) 迁移学习与特征转移:随着工业现场环境的不断变化,基于已有训练数据的迁移学习将变得更加重要。探索如何通过深度迁移学习和特征转移技术,更好地应对新缺陷类型和不同工业环境中的适应性问题,提升模型在不同领域的泛化能力。

参考文献:

- [1] X. Ji, Q. Yan, D. Huang, et al. Filtered selective search and evenly distributed convolutional neural networks for casting defects recognition[J]. Journal of Materials Processing Technology, 2021, 292: 117064.
- [2] 周建新, 计校园, 侯明君, 等. 高端装备金属构件成形

- 缺陷智能检测技术进展与展望[J]. 前瞻科技, 2025: 108-117.
- [3] S. Pattnaik, D. B. Karunakar.P. K. Jha. Developments in investment casting process — A review[J]. Journal of Materials Processing Technology, 2012, 212 (11): 2332-2348.
- [4] 武博, 计效园, 孙晓龙, 等. 基于 X 射线图像的铸件缺陷智能检测与识别方法研究进展[J]. 铸造, 2022, 71: 1001-4977
- [5] F. Paolo. Casting and Solidification Processing[J]. Metals, 2022, 12 (4): 559.
- [6] 韩恒恒, 李晓棠, 曹殿涛, 等. 铝合金压铸铸件缺陷分析及对策和压铸技术新发展[J]. 机电技术, 2014: 1672-4801.
- [7] C. Hu.Y. Wang. An Efficient Convolutional Neural Network Model Based on Object-Level Attention Mechanism for Casting Defect Detection on Radiography Images[J]. IEEE Transactions on Industrial Electronics, 2020, 67 (12): 10922-10930.
- [8] W. Fu, X. Jiang, B. Li, et al. Rolling bearing fault diagnosis based on 2D time-frequency images and data augmentation technique[J]. Measurement Science and Technology, 2023, 34 (4): 045005.
- [9] G.-h. Lian, Q.-h. Sun, X.-m. Liu, et al. Automatic recognition and intelligent analysis of central shrinkage defects of continuous casting billets based on deep learning[J]. Journal of Iron and Steel Research International, 2023, 30 (5): 937-948.
- [10] H. Wei, C. T. Wu, W. Hu, et al. LS-DYNA Machine Learning - Based Multiscale Method for Nonlinear Modeling of Short Fiber - Reinforced Composites[J]. Journal of Engineering Mechanics, 2023, 149 (3): 04023003.
- [11] Z. Tang, E. Tian, Y. Wang, et al. Nondestructive Defect Detection in Castings by Using Spatial Attention Bilinear Convolutional Neural Network[J]. IEEE Transactions on Industrial Informatics, 2021, 17 (1): 82-89.
- [12] N. Yousef.A. Sata. Implementing Deep Learning-Based Intelligent Inspection for Investment Castings[J]. Arabian Journal for Science and Engineering, 2023, 49 (2): 2519-2530.
- [13] S. Cheng, J. Lu, M. Yang, et al. Wheel hub defect detection based on the DS-Cascade RCNN[J]. Measurement, 2023, 206: 112208.
- [14] J. Lin, Y. Yao, L. Ma, et al. Detection of a casting defect tracked by deep convolution neural network[J]. The International Journal of Advanced Manufacturing Technology, 2018, 97 (1-4): 573-581.
- [15] W. Liu, S. Shan, H. Chen, et al. X-ray weld defect detection based on AF-RCNN[J]. Welding in the World, 2022, 66 (6): 1165-1177.
- [16] Y. Xie, W. Hu, S. Xie, et al. Surface Defect Detection Algorithm Based on Feature-Enhanced YOLO[J]. Cognitive Computation, 2022, 15 (2): 565-579.
- [17] K. Pan, H. Hu.P. Gu. WD-YOLO: A More Accurate YOLO for Defect Detection in Weld X-ray Images[J]. Sensors, 2023, 23 (21): 8677.
- [18] T. Chen, C. Cai, J. Zhang, et al. RER-YOLO: improved method for surface defect detection of aluminum ingot alloy based on YOLOv5[J]. Optics Express, 2024, 32 (6): 8763.
- [19] W. Wu, W. Wang, X. Jia, et al. Transformer Autoencoder for K-means Efficient clustering[J]. Engineering Applications of Artificial Intelligence, 2024, 133: 108612.
- [20] B. Wang.C. Yang. Video Anomaly Detection Based on Convolutional Recurrent AutoEncoder[J]. Sensors, 2022, 22 (12): 4647.
- [21] M. Frid-Adar, I. Diamant, E. Klang, et al. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification[J]. Neurocomputing, 2018, 321: 321-331.
- [22] M.-j. Hou, H. Dong, X.-y. Ji, et al. I-DCGAN and TOPSIS-IFP: A simulation generation model for radiographic flaw detection images in light alloy castings and an algorithm for quality evaluation of generated images[J]. China Foundry, 2024, 21 (3): 239-247.
- [23] D. Mery. Aluminum Casting Inspection Using Deep Learning: A Method Based on Convolutional Neural Networks[J]. Journal of Nondestructive Evaluation, 2020, 39 (1): 12.
- [24] 张辉, 张邹铨, 陈煜嵘, 等. 工业铸件缺陷无损检测技术的应用进展与展望[J]. 自动化学报, 2022, 48 (4): 935-956.
- [25] W. Lu, S. Zhang, W. Zhang, et al. Imaging of Structure Evolution in Solidifying Al - Bi Immiscible Alloys by Synchrotron Radiography[J]. Journal of Materials Science & Technology, 2016, 32 (12): 1321-1325.
- [26] X. Li, S. K. Tso, X.-P. Guan, et al. Improving Automatic

- Detection of Defects in Castings by Applying Wavelet Technique[J]. IEEE Transactions on Industrial Electronics, 2006, 53 (6): 1927-1934.
- [27] B. Wu, Y. Huang, S. Krishnaswamy. A Bayesian approach for sparse flaw detection from noisy signals for ultrasonic NDT[J]. NDT & E International, 2017, 85: 76-85.
- [28] N. Munir, J. Huang, C.-N. Wong, et al. Machine learning based eddy current testing: A review[J]. Results in Engineering, 2025, 25: 103724.
- [29] S. S. Fayad, E. M. C. Jones, C. Winters. Path-Integrated X-Ray Digital Image Correlation using Synthetic Reference Images[J]. Experimental Techniques, 2024, 48 (6): 941-951.
- [30] L. Bai, J. Zhang, A. Velichko, et al. The use of full-skip ultrasonic data and Bayesian inference for improved characterisation of crack-like defects[J]. NDT & E International, 2021, 121: 102467.
- [31] J. Tkocz, D. Greenshields, S. Dixon. High power phased EMAT arrays for nondestructive testing of as-cast steel[J]. NDT & E International, 2019, 102: 47-55.
- [32] G. Betta, L. Ferrigno, M. Laracca, et al. An experimental comparison of multi-frequency and chirp excitations for eddy current testing on thin defects[J]. Measurement, 2015, 63: 207-220.
- [33] H. E. Farag, E. Toyserkani, M. B. Khamesee. Non-Destructive Testing Using Eddy Current Sensors for Defect Detection in Additively Manufactured Titanium and Stainless-Steel Parts[J]. Sensors, 2022, 22 (14): 5440.
- [34] X. Xu, M. Liu, Z. Zhang, et al. A Novel High Sensitivity Sensor for Remote Field Eddy Current Non-Destructive Testing Based on Orthogonal Magnetic Field[J]. Sensors, 2014, 14 (12): 24098-24115.
- [35] R. Girshick, J. Donahue, T. Darrell, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[J]. In Proceedings of the IEEE conference on computer vision and pattern recognition, 2014: 580-587.
- [36] S. Ren, K. He, R. Girshick, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39 (6): 1137-1149.
- [37] W. Du, H. Shen, J. Fu, et al. Approaches for improvement of the X-ray image defect detection of automobile casting aluminum parts based on deep learning[J]. NDT & E International, 2019, 107: 102144.
- [38] K. He, G. Gkioxari, P. Dollar, et al. Mask R-CNN[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42 (2): 386-397.
- [39] F. Yang, J. Huo, Z. Cheng, et al. An Improved Mask R-CNN Micro-Crack Detection Model for the Surface of Metal Structural Parts[J]. Sensors, 2023, 24 (1): 62.
- [40] Y. Wang, C. Hu, K. Chen, et al. Self-attention guided model for defect detection of aluminium alloy casting on X-ray image[J]. Computers & Electrical Engineering, 2020, 88: 106821.
- [41] J. Redmon, S. Divvala, R. Girshick, et al. You Only Look Once: Unified, Real-Time Object Detection[J]. In Proceedings of the IEEE conference on computer vision and pattern recognition, 2016: 779-788.
- [42] J. Redmon, A. Farhadi. YOLO9000: Better, Faster, Stronger[J]. In Proceedings of the IEEE conference on computer vision and pattern recognition, 2017: 7263-7271.
- [43] J. Redmon, A. Farhadi. Yolov3: An incremental improvement[J]. arxiv preprint arxiv, 2018: 02767.
- [44] A. Bochkovskiy, C.-Y. Wang, H.-Y. M. Liao. YOLOv4-Optimal Speed and Accuracy of Object Detection[J]. arxiv preprint arxiv, 2020: 10934.
- [45] C. Li, L. Li, H. Jiang, et al. YOLOv6-A Single-Stage Object Detection Framework for Industrial[J]. arxiv preprint arxiv, 2022: 02976.
- [46] C.-Y. Wang, A. Bochkovskiy, H.-Y. M. Liao. YOLOv7-Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[J]. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2023: 7464-7475.
- [47] C.-Y. Wang, I.-H. Yeh, H.-Y. M. Liao. YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information[J]. In European conference on computer vision, 2024: 1-21.
- [48] A. Wang, H. Chen, L. Liu, et al. YOLOv10-Real-Time End-to-End Object Detection[J]. Advances in Neural Information Processing Systems, 2024, 37: 107984-108011.
- [49] R. Khanam, M. Hussain. YOLOv11-An Overview of the Key Architectural Enhancements[J]. arxiv preprint arxiv, 2024: 17725.
- [50] L. Duan, K. Yang, L. Ruan. Research on Automatic

- Recognition of Casting Defects Based on Deep Learning[J]. IEEE Access, 2021, 9: 12209-12216.
- [51] 李彬, 汪诚, 吴静, 等. 改进 YOLOv4 算法的航空发动机部件表面缺陷检测[J]. 激光与光电子学进展, 2021, 58:
- [52] İ. E. Parlak.E. Emel. Deep learning-based detection of aluminum casting defects and their types[J]. Engineering Applications of Artificial Intelligence, 2023, 118:
- [53] 葛前峰, 浩, 袁., 渊, 王., 等. 基于 YOLOv5 的小型铝铸件涡轮缺陷检测[J]. 特种铸造及有色合金, 2024, 44: 760-765.
- [54] L. Liu, F. Huang, L. Gu, et al. Improved YOLOv7 Based on Reparameterization and Attention Mechanism for Continuous Casting Slab Detection[J]. IEEE Transactions on Instrumentation and Measurement, 2025, 74: 1-12.
- [55] Y. Jiang, P. Li, B. Lin, et al. Microscopic Defect Detection on Aircraft Engine Blades via Improved YOLOv8[J]. IEEE Access, 2025, 13: 4932-4946.
- [56] C. Wang,Y. Wang. SLGA-YOLO: A Lightweight Castings Surface Defect Detection Method Based on Fusion-Enhanced Attention Mechanism and Self-Architecture[J]. Sensors, 2024, 24 (13): 4088.
- [57] W. Guiqiang, C. Junbao, L. Chengzhang, et al. Edge-YOLO: Lightweight Multi-Scale Feature Extraction for Industrial Surface Inspection[J]. IEEE Access, 2025, 13: 48188-48201.
- [58] H. Yang, Y. Chen, K. Song, et al. Multiscale Feature-Clustering-Based Fully Convolutional Autoencoder for Fast Accurate Visual Inspection of Texture Surface Defects[J]. IEEE Transactions on Automation Science and Engineering, 2019, 16 (3): 1450-1467.
- [59] W. Tang, C. M. Vian, Z. Tang, et al. Anomaly detection of core failures in die casting X-ray inspection images using a convolutional autoencoder[J]. Machine Vision and Applications, 2021, 32 (4):
- [60] O. Chamberland, M. Reckzin.H. A. Hashim. An Autoencoder with Convolutional Neural Network for Surface Defect Detection on Cast Components[J]. Journal of Failure Analysis and Prevention, 2023, 23 (4): 1633-1644.
- [61] S. Zhang, H. Li, P. Ren, et al. A detection method for small casting defects based on bidirectional feature extraction[J]. Scientific Reports, 2025, 15 (1): 6362.
- [62] X. He, Z. Luo, Q. Li, et al. DG-GAN: A High Quality Defect Image Generation Method for Defect Detection[J]. Sensors, 2023, 23 (13): 5922.
- [63] A. García-Pérez, M. J. Gómez-Silva.A. d. I. Escalera-Hueso. Improving automatic defect recognition on GDXRay castings dataset by introducing GenAI synthetic training data[J]. NDT & E International, 2025, 151: 103303.