

基于计算机视觉的工业金属表面缺陷检测综述

伍麟¹ 郝鸿宇¹ 宋友¹

摘 要 针对平面及三维结构金属材料的工业表面缺陷检测, 概述了视觉检测技术的基本原理和研究现状, 并总结出视觉自动检测系统的关键技术包括光学成像技术、图像预处理技术与缺陷检测器。首先介绍了如何根据检测对象的光学特性选择合适的二维、三维光学成像技术; 其次介绍了图像降噪、特征提取、图像分割和拼接等预处理技术的重要作用; 然后根据缺陷检测器的实现原理将其分为模板匹配、图像分类、图像语义分割、目标检测和图像异常检测五类, 并对其中的经典算法进行了归纳分析。最后, 探讨了工业场景下金属表面缺陷检测技术实施中的关键问题, 并对该技术的发展趋势进行了展望。

关键词 表面缺陷检测, 计算机视觉, 金属表面缺陷, 自动化检测

引用格式 伍麟, 郝鸿宇, 宋友. 基于计算机视觉的工业金属表面缺陷检测综述. 自动化学报, 2024, 50(7): 1261–1283

DOI 10.16383/j.aas.c230039

A Review of Metal Surface Defect Detection Based on Computer Vision

WU Lin¹ HAO Hong-Yu¹ SONG You¹

Abstract Focusing on the industrial surface defect detection of metal planar and three-dimensional structural materials, this paper summarizes the basic principle and research status of visual defect detection technology, and summarizes the key technologies of visual automatic detection system including optical imaging technology, image pre-processing technology and defect detector. Firstly, this paper introduces how to select suitable 2D and 3D optical imaging technology according to the optical characteristics of the test object. Secondly, the important functions of image denoising, feature extraction, image segmentation and image Mosaic are introduced. Then, according to the implementation principle of defect detector, it is divided into five categories: Template matching, image classification, image semantic segmentation, target detection and image anomaly detection, and the classical algorithms are summarized and analyzed. Finally, this paper discusses the key problems in the implementation of surface defect detection in the industrial scene, and looks forward to the development trend of this technology.

Key words Surface defect detection, computer vision, metal surface defect, automatic inspection system

Citation Wu Lin, Hao Hong-Yu, Song You. A review of metal surface defect detection based on computer vision. *Acta Automatica Sinica*, 2024, 50(7): 1261–1283

金属表面缺陷是产品生产工艺的不确定性造成的结果, 对于金属铸造、机加工制品而言, 表面缺陷是常见的质量问题, 一般的表面缺陷包括磕碰、划痕、夹杂、气孔、开裂、麻点、翘曲和脆化等^[1]。这些缺陷不仅会影响产品的外观, 更重要的是它们还会影响产品的机械性能。对于某些产品而言, 表面缺陷是致命的。

表面检测是指为了确定产品是否满足质量标准和技术要求而进行的测试或测量活动。传统方法主要通过人工目视检测, 不仅会出现人员疲劳、评价标准不一致等情况, 同时人眼的时间及空间分辨率有限, 在许多工业场景下只能进行抽检, 难以满足

生产过程中高速在线实时检测的需求。热轧、浇铸等工艺过程对人员还有一定危险性, 因此人工抽检往往在工艺过程结束后进行, 即使发现缺陷, 过程损失也已发生, 无法对工艺参数进行及时调整。

自动光学检测技术 (Automated optical inspection, AOI)^[2] 是一种以计算机视觉为基础, 通过自动光学系统获取检测目标图像, 运用算法进行分析决策, 判断目标是否符合检测规范的非接触式检测方法。计算机视觉检测技术相对于人工检测, 具备实时性强、效率高、节省人力成本、不易受主观因素干扰的优势, 逐渐成为金属制品表面缺陷检测的重要手段, 目前在工业领域, 尤其是热轧钢板、带钢、精密铝带为代表的金属板带类产品中应用广泛。

表面缺陷检测系统的基本原理是利用金属表面的光学物理性质, 在特定的光学成像条件下, 令缺陷表现出区别于背景的图像特征, 然后通过图像处

收稿日期 2023-02-06 录用日期 2023-05-18

Manuscript received February 6, 2023; accepted May 18, 2023

本文责任编辑 桑农

Recommended by Associate Editor SANG Nong

1. 北京航空航天大学软件学院 北京 100191

1. School of Software, Beihang University, Beijing 100191

理技术和缺陷检测算法进行缺陷识别和定位. 其基本流程如图 1 所示, 大致可分为以下步骤:

光学成像: 通过光学图像传感方式来获取产品表面二维灰度图像或三维深度图像信息.

图像预处理: 对采集到的图像进行降噪、增强、分割、拼接等图像预处理操作, 以便获得感兴趣区域 (Region of interest, ROI) 或更适用于图像检测算法的输入尺寸, 并通过各种图像处理算法或它们的组合, 提取出最容易区分缺陷和背景的特征.

缺陷检测: 采用图像对比、统计模式识别、机器学习或深度学习等算法, 从背景中识别出可能存在的缺陷, 对其进行分类和定位.

根据金属表面缺陷检测的背景和成像方式, 通常可分为二维 (2D) 视觉检测和三维 (3D) 视觉检测. 前者主要针对具有平整表面或圆弧型连续规则结构的表面, 又可进一步分为均匀背景、周期性背景和随机纹理背景下的缺陷检测, 在冷轧、热轧金属板带表面缺陷、精密加工件和金属外壳外观件瑕疵检测等领域得到了广泛应用^[3]. 后者针对具有三维复杂结构背景的金属制成品表面缺陷进行检测. 二维视觉检测技术发展至今已经有 30 余年, 目前技术已经十分成熟, 在自动化生产和产品质量控制体系中发挥着重要作用. 但二维视觉仅仅能获得二维灰度图像信息, 难以获得金属表面的深度、细微纹理、法线方向等信息. 随着成像技术、感光元器件技术的快速发展, 成像技术已经从二维灰度图像发展到多光谱、三维成像技术, 与之对应的数字图像处理技术也在飞速发展, 近几年来取得了极大的进步, 为金属表面缺陷检测提供了越来越丰富的手段.

工业金属表面缺陷检测技术是一项融合了光学技术、传感器技术、图像处理技术、人工智能和深度学习等多门学科的复杂任务, 在这一领域仍有许多

理论和关键应用技术亟待解决. 例如:

1) 由于检测对象的表面光学特性复杂多变、种类繁多, 如何设计合适的视觉成像系统来捕获检测对象的图像特征, 突出缺陷与背景的特征差异, 使之达到并超越肉眼检测的效果, 是所有工业表面缺陷检测任务的关键问题.

2) 针对二维图像的缺陷检测算法已经相对成熟, 但基于三维深度信息的缺陷检测技术尚未普及. 而工业金属制成品实际大都具有复杂的表面结构, 需求比二维背景表面检测更加广泛, 这些领域背后蕴藏着巨大的商业价值有待发掘.

3) 金属表面缺陷检测数据集相对匮乏, 尤其是三维深度图像数据. 基于商业或其他因素影响, 许多研究未能公开它们的数据集. 另一方面, 由于工业产品不断改进生产工艺, 缺陷样本数量稀少, 并且样本类别极度不均衡. 同时在生产工艺改进的过程中, 还有可能出现未曾定义过的缺陷模式, 这进一步增加了检测任务的难度.

目前该领域的综述大多以研究针对现有图像、视频数据的缺陷检测算法为主. 文献 [2] 虽然以光学成像技术为主要综述内容, 但缺乏对三维成像、复合成像多模态数据融合特征技术的探讨. 文献 [4] 综述了金属平面材料二维三维缺陷检测的常用方法, 但主要介绍传统方法. 文献 [5-6] 分别对基于有监督、无监督和半监督的深度学习方法进行了系统性的综述, 但局限于基于卷积神经网络 (Convolutional neural network, CNN) 的方法, 而近年来基于视觉转换器 (Vision transformer, ViT)^[7] 的方法在计算机视觉领域大放异彩, 在很多任务上超越了基于 CNN 的方法. 目前尚无针对工业金属表面缺陷检测关键技术的全面细致的文献综述, 因此本文重点着眼于该领域的新方法进行介绍. 其中光学成像技术、图像预处理技术和缺陷检测器是表面缺

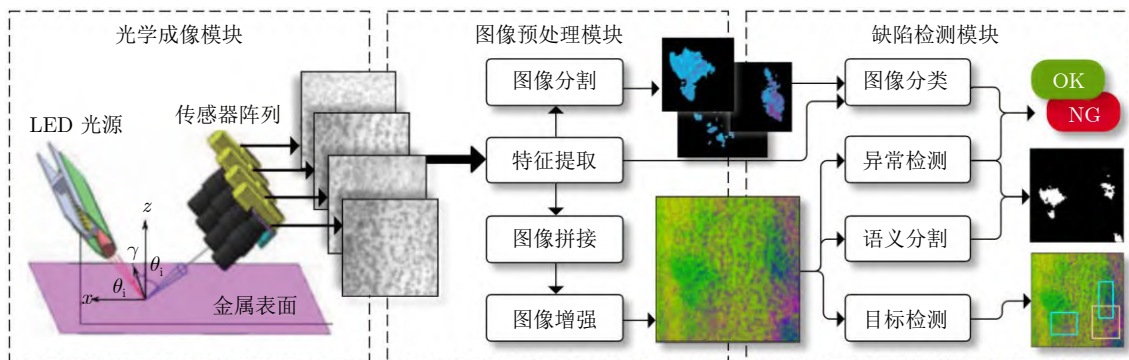


图 1 金属表面缺陷检测基本流程

Fig.1 Pipeline of metal surface defect detection

陷检测系统的关键技术,通过详细梳理这些技术的常用方法和研究现状,本文分析并指出了该领域存在的研究难点,并对未来的发展趋势进行了分析和展望。

1 光学成像技术

传统人工目视检测法存在效率低、主观性强、肉眼的时间和空间分辨能力有限等问题。自动光学技术通过光学传感技术模拟人眼的视觉成像功能,针对金属表面的物理反射性质实现缺陷特征增强显示,辅助人工进行视觉外观检测,同时通过多个基本单元组合,可以实现大视场、高分辨率、快速并行的在线检测。自动光学技术是金属表面异常检测技术的前提和基础,不仅可以辅助人工检测,提高实时性和稳定性,而且在某些场景中已经逐渐取代人工检测,成为金属表面缺陷检测的重要手段^[6]。

一种常见的误解是:如果人工肉眼检测可以看到一个缺陷特征,则可以设计一个光学成像系统来捕获完全相同的特征。但是在实际操作中,检验员可以从多个角度和不同的光源方向下查看产品表面以进行判断。而静态成像系统则难以做到大范围的角度调整,在检测具有复杂结构的三维金属表面时,光学成像系统会面临更大的挑战^[9]。通过研究金属表面的物理和光学特性,构建一个足以在图像处理任务上超越人眼的光学成像系统是一项艰巨的任务。

光学成像系统如图 2 所示,其基本组成包括光源、光学镜头、相机和运动控制器。光学成像系统不仅仅是简单对被测物体进行照明并用相机获取图像,根据金属表面和光的物理特性,合理选用不同的成像技术和光路来获取被测缺陷的特征,并使之与背景的特征差异最大化,是实现金属表面缺陷检测的前提。根据光学成像方式的不同,光学检测技术可分为基于二维灰度图像信息的检测技术和基于三维深度图像信息的检测技术^[10]。二维检测技术主要用于冷轧、热轧金属板带的生产过程,而三维检测常用于金属制成品、精密加工件、金属外观件的生产过程中检测。

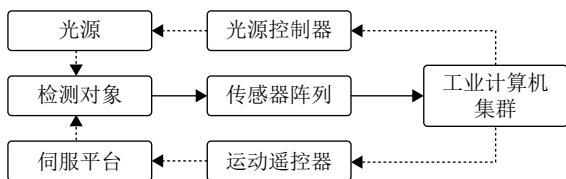


图 2 自动光学成像系统

Fig. 2 Automated optical inspection system

1.1 二维成像检测技术

二维成像技术主要包括角度分辨技术^[11]、色彩分辨技术^[12]和光谱分辨技术^[13]。由于金属制品表面往往颜色、材质单一,因此色彩分辨技术和光谱分辨技术很少用于金属表面缺陷检测,最常用的是角度分辨技术。

角度分辨技术的基本原理是表面散射模型^[14],如图 3 所示。当光束照射金属表面时,根据产品表面粗糙度的不同会发生朗伯反射、方向反射和镜面反射现象,通过成像光路获取二维灰度图像,然后根据缺陷和背景的敏明暗特征来检测缺陷。朗伯反射又叫均匀反射,在各个角度的反射光线能量强度几乎相等,可用双向散射分布函数(Bidirectional scattering distribution function, BSDF)来表达;方向反射发生在光线波长与金属表面粗糙度投影尺寸相耦合的情况下,反射光线的能量强度在某一空间角内最大;镜面反射服从镜面反射定律,其能量分布受金属表面粗糙度和起伏程度的影响。除此之外,光线反射还受到微观几何尺度上的表面起伏对光线遮挡的影响。根据上述效应的影响,散射光场可分为漫反射光瓣、镜面反射光瓣、镜面反射峰和微观纹理形成的菊花瓣。

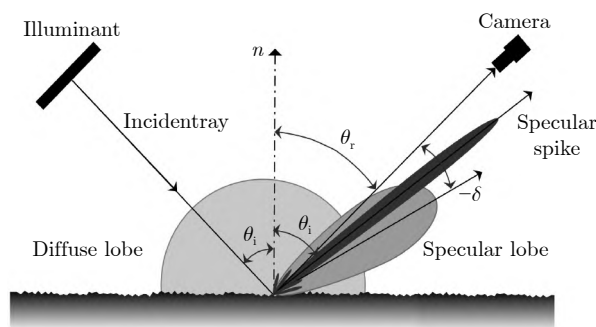


图 3 表面散射模型

Fig. 3 Light scattering model on surface

根据光的物理反射特性,针对不同类型的金属表面背景以及缺陷特点,产生了不同的光学角度散射分辨技术。美国 Westinghouse 公司通过高强度线光源和线阵 CCD (Charge coupled device) 相机采集钢板表面图像,并提出了明场、暗场、漫反射照明技术用于金属表面缺陷检测^[15]。随着光学传感技术的发展,又出现了同轴、背光等多种照明技术,这一类技术被称为角度分辨技术,如图 4。

a) 明场照明成像方式主要针对均匀的漫反射金属表面背景,背景成像结果为均匀的明亮图像。当金属表面存在缺陷时,反射光将在其他方向发生散射,从而在明亮的背景中产生暗点,常用于表面

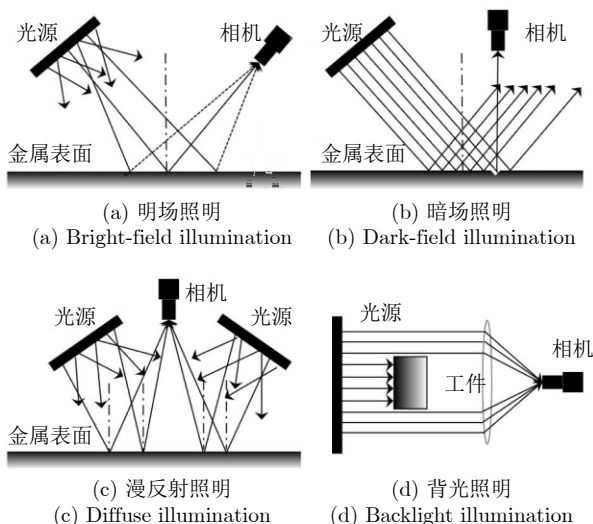


图 4 照明光路类型

Fig. 4 Types of lighting path

粗糙度较高的铸坯、热轧钢带等产品。

b) 暗场照明成像方式主要针对镜面金属表面, 此时背景的散射光线十分微弱, 背景成像结果为暗灰色图像。当背景中存在缺陷时, 光线会在多个方向发生散射, 在灰暗的背景中产生明亮的特征, 常用于冷轧钢带、不锈钢等平滑金属表面^[16]。

c) 漫反射照明成像主要针对凹凸不平或有压制纹理的金属表面, 光线从不同角度照明被测产品表面形成平滑过渡的图像, 褶皱或划痕则会形成阴影或高亮区域, 从而产生区别于背景的特征。

d) 背光照明成像方式针对金属产品外围轮廓, 相机与背光照明光源位于被测金属产品两侧, 光线被金属遮挡产生投影, 在相机中形成清晰的高对比度图像。

成像方式的选取与检测的缺陷和背景特性紧密相关。一般金属制品表面的反射性质介于漫反射和镜面反射之间, 而不同类型的缺陷在特定光照条件下更容易分辨。在角度分辨检测技术中, 光源照射方向、表面法线方向与相机光轴方向之间的角度, 决定了缺陷特征的主要因素, 包括对比度、空间分辨率、光照一致性和曝光度, 这些因素对系统捕获与背景明显区分的缺陷特征至关重要^[17]。例如划伤缺陷, 宽度尺寸通常小于 0.2 mm, 其光学特性有明显的方向性, 存在大范围的视觉盲区, 仅在特定照明角度下呈现高度反射, 当光源方向平行于划痕时则几乎不可见。又如水渍、油渍, 在金属表面干涸后呈现圆形亮斑, 在单一照明条件下极易与圆形凹痕相互混淆。大量研究表明不同的成像方式之间具有互补性, 采用复合光源成像系统能够有效提高表面检测的识别精度与稳定性。Parsytec^[18] 公司提出了

一种复合成像方案, 通过漫反射光源与面阵相机组合成像、平行光源与线阵相机组合成像同时进行表面缺陷检测。李松等^[19] 提出将明场照明与暗场照明结合的双打光成像方式对金属冲压件表面进行缺陷检测, 可以更好地区分磕碰、划伤以及锈斑等多种不同性质的表面缺陷。多种光学成像方式组合以提升系统的检测性能, 成为该领域的一个重要发展方向。

1.2 三维成像检测技术

二维成像方式在检测具有凹凸结构的金属表面时缺少深度、法线方向、曲率等微观特征信息, 对于例如气孔、夹渣等微小缺陷检测困难、漏检率高。对于氧化皮、水渍等伪缺陷与裂纹、辊印等具有深度信息的缺陷, 在二维灰度图像上特征十分相似, 极易混淆, 而在生产过程中非缺陷出现比例往往远高于真实缺陷, 带来了大量的误检信号, 显著降低了表面缺陷检测系统的实用性^[20]。三维视觉成像技术可以获取裂纹、辊印等缺陷的深度、法线方向等信息, 如图 5, 能够显著提高表面缺陷检测系统的分辨能力。

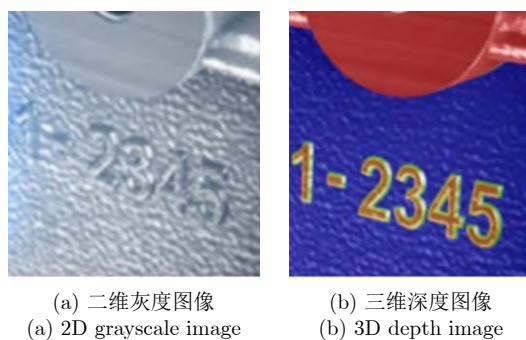


图 5 二维成像与三维成像对比

Fig. 5 2D imaging versus 3D imaging

三维视觉成像技术在工业领域是一种重要的方法, 近几年来在机器人控制、机器测量、表面和内部缺陷检测领域取得了显著的成果。三维成像检测技术包括光学方法和非光学方法, 其中光学方法应用最为广泛, 主要包括光度立体法、飞行时间法、扫描法、立体视觉法、结构光法和光场三维成像等。

光度立体视觉 (Photometric stereo, PS)^[21] 来自于阴影恢复形状 (Shape from shading, SFS) 方法的改进^[22]。经典光度立体法通常由 3 个位置不同的光源、被测产品和同一个相机组成, 其布置如图 6 所示。相机拍摄被测表面在多个空间角度的光源照射形成的一组图像, 根据 Lambertian 反射模型计算得到被测表面的法向量以及水平梯度信息, 从而实现三维表面重建^[23]。

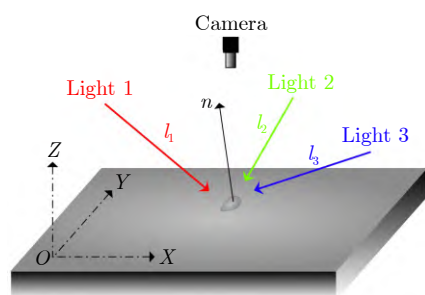


图 6 光度立体法

Fig.6 Photometric stereo

飞行时间 (Time of flight, TOF) 法利用每个像素接收的光的飞行时间差来获取金属表面深度信息^[24], 又可分为直接 TOF (Direct time-of-flight, D-TOF) 和间接 TOF (Indirect time-of-flight, I-TOF), 前者直接计算光脉冲的飞行时间, 后者则根据光强度相位偏移来间接估算距离, 目前已经出现了很多成熟的商业化 TOF 面阵相机产品. TOF 成像适用于大视场、工作距离远、检测速度要求较快的场景, 但深度精度为毫米级别, 不适用于表面检测精度要求较高的场景.

扫描法利用光束扫描整个目标表面来实现 3 维测量^[25], 包括单点飞行时间法、激光散射干涉法、色散共焦法等, 其中色散共焦法在目标表面同时具有粗糙和光滑的透明或不透明区域时, 具备十分突出的表现, 例如反射镜面、透明玻璃面等. 色散共焦法包括点扫描和线扫描, 测量精度高, 尤其适合光滑的金属表面, 目前在手机盖板三维检测等领域应用广泛. Huang 等^[26]通过三维激光扫描传感器捕获曲面点云并转换为二维灰度形图像, 然后用支持向量机进行像素级分类, 该方法不仅可以检测到金属 3D 增材制造零件表面广泛存在的凸起和凹陷缺陷, 而且可以检测到像素级的孔隙缺陷. 但基于扫描的方法速度慢、效率较低, 不适用于需要高速在线实时检测的场景.

立体视觉法源于对人类双目视觉的研究, 通过空间视差、阴影、运动视差等感知线索来获取目标的深度图像信息. 典型的双目立体视觉利用两个相机从不同视点拍摄同一个目标, 通过匹配左右两个图像中的特征点以及相机标定参数, 即可求解对图像的三维深度估计. 计算过程通常包括四个步骤: 矫正畸变、立体图像对校正、图像配准和重投影视差图计算. 文献 [27] 提出了一种基于双目立体视觉的实时表面缺陷三维检测技术, 对 6 500 多个缺陷数据集进行分类, 深度和面积的平均精度分别达到了 98.9% 和 98.0%. 文献 [28] 提出了一种基于双目

立体视觉的高反射焊接表面三维重建技术, 该方法通过条纹投影轮廓法实现了二维图像与三维点云特征的映射和对齐, 以消除图像特征中的“空洞”, 显著提高了高反射金属表面缺陷的检测精度.

如图 7 所示, 结构光法通过激光投影仪向目标表面投射调制后的光栅, 由相机拍摄投影图案来求解目标表面的 3 维深度信息, 又可分为瞬间投影成像和多次投影成像两类^[29]. 其中多次投影成像采用时间复用编码方式, 例如二进制编码、多频相移编码、格雷码和混合编码技术, 理论上可获得亚像素级的高精度 3 维点云深度信息, 目前已经成为行业主流方案. 由于硬件光栅成本高、灵活性差, 数字投影仪逐渐成为最流行的光栅投影方案. 通过格雷码与正弦相移条纹投影结合, 既可解决格雷码空间离散划分的缺点, 又可解决数字投影仪非线性响应导致的相位误差, 在大幅度提高测量精度的同时兼顾了测量速度, 成为了十分成熟的商用解决方案, 但结构光 3D 相机技术复杂, 硬件成本相对较高.

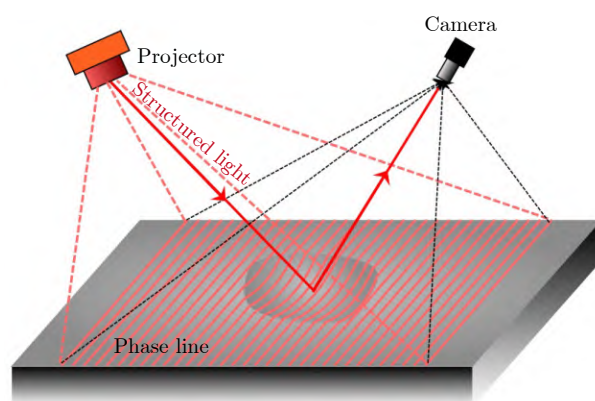


图 7 结构光法

Fig.7 Structured light illumination

光场 3 维成像技术模仿昆虫的复眼结构, 光线经主镜头入射, 分别穿过不同焦距的微透镜阵列后在感光阵列上成像, 使得不同远近的目标各自在对应的感光阵列上聚焦成像, 经算法处理估计出目标的 3 维深度信息. Raytrix 光场 3D 相机^[30]是目前唯一商用的光场相机, 具备横纵方向 20 μm 和深度方向毫米量级的空间分辨率, 但硬件成本最为昂贵.

在 3D 成像检测技术中, 光度立体法具有显著优势: 1) 空间尺度分辨能力不受视场限制, 其分辨率采样密度等同于像素密度, 能够有效捕捉微小的缺陷特征; 2) 采用多照明光路、多成像通道互补, 避免表面凸起阴影对局部细节造成的影响, 能够更有效地获取局部缺陷深度、梯度、法线方向和曲率等信息; 3) 硬件成本具备显著优势, 并可兼容传统

二维检测算法. 与仅有二维灰度图像信息的检测技术相比, 光度立体法能够更好地区分水渍、油污等伪缺陷与凹陷、刮痕等有深度变换的真实缺陷, 在具有复杂结构纹理的金属表面具备更有效的分辨能力^[23]. 文献 [31] 提出了一种基于光度立体视觉的自动金属零件质量控制系统, 该系统将光度立体图像组合成 RGB 图像以便操作员进行交互和操作, 并通过深度语义分割网络进行缺陷分割, 在各种不同反射率金属表面的检测实验中 DSC (Dice similarity coefficient) 指标均达到了 0.83 以上.

1.3 混合成像检测技术

在二维成像和三维成像技术的基础上, 还有很多研究多种成像技术混合的方法, 在实践中取得了显著成果. 文献 [32] 提出了一种融合二维灰度和三维深度图像信息的检测方法, 该方法通过计算感兴趣区域对象之间的潜在关系, 然后通过 ROI 种子提取和局部对比操作, 得到精化的 ROI 显著性图像, 在灰度图像上实现了像素级缺陷分割, 对表面凹陷缺陷具有良好的检测性能. Niu 等^[33] 提出了一种基于双目线扫描系统的无监督立体显著性检测方法, 将二维灰度图像显著性信息与三维深度异常

值检测结果协同融合, 同时避免了结构光重建方法的解码失真. 实验结果表明, 该方法的平均绝对误差为 0.09, ROC (Receiver operator characteristic curve) 曲线下面积为 0.94, 优于 15 种最先进的算法.

二维成像技术通常可获取高分辨率以及高对比度的图像, 能够更好地捕获缺陷的纹理细节特征信息, 对夹杂、气孔、划痕等微小瑕疵有更好的检测效果. 然而二维成像局限于金属表面形状的投影, 无法获取深度和法线方向等信息, 对于油污、锈迹等伪缺陷容易发生误判; 与此同时三维成像可以提供深度、褶皱、凹凸纹理等形状信息, 但计算量较大且空间分辨率有限, 对微小缺陷特征不够敏感. 混合成像检测技术能够克服两者的局限性, 结合二维纹理细节和三维深度信息更准确的判别缺陷, 在相同成像条件下准确率高于单一成像方式检测技术, 如图 8 所示. 但这一类技术实现成本较高, 目前尚难以推广.

2 图像预处理技术

图像预处理技术在表面缺陷检测中起着至关重要的作用, 它有助于提高图像质量和信噪比, 使检

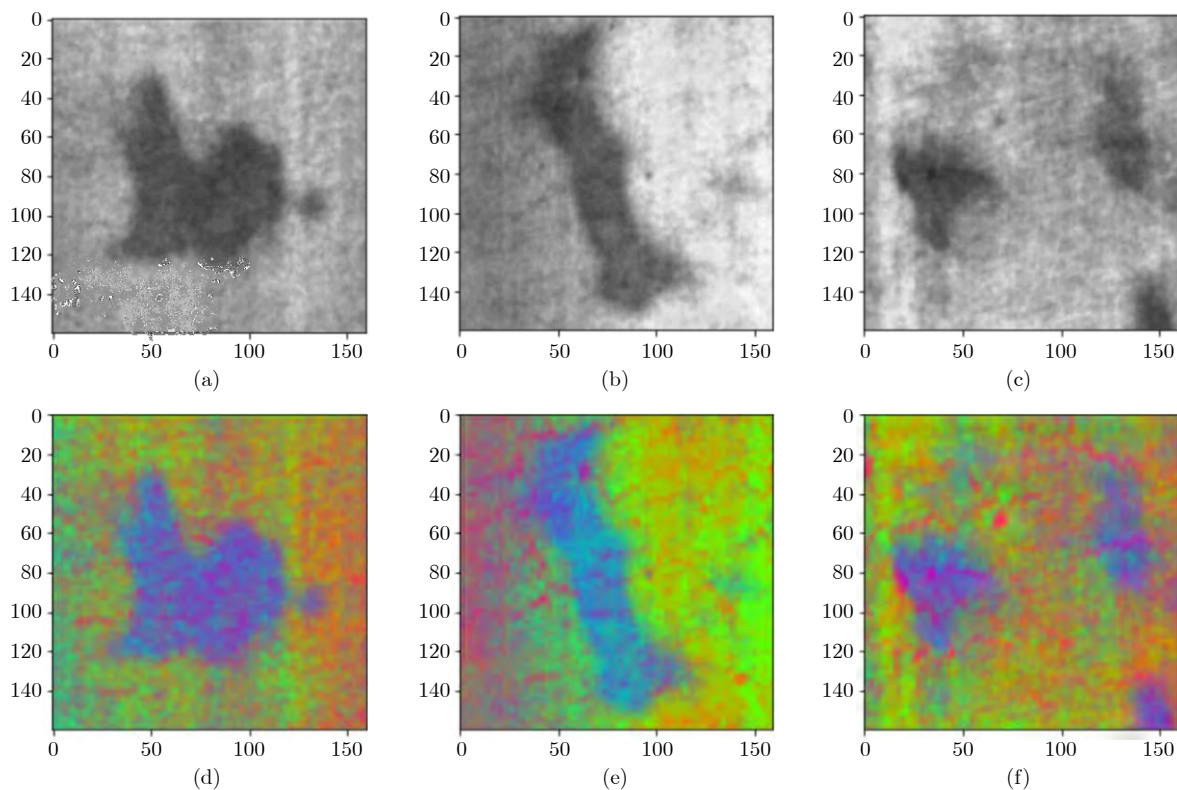


图 8 混合成像技术 ((a) ~ (c) 二维灰度图像; (d) ~ (f) 具有三维深度信息表示的图像)

Fig. 8 Hybrid imaging technique ((a) ~ (c) 2D grayscale images; (d) ~ (f) Images with 3D depth information represented)

测算法更容易准确识别缺陷^[34]。图像预处理包括图像增强、特征提取、图像分割和拼接等技术。图像增强有助于消除任何可能掩盖缺陷的不必要的视觉干扰。图像增强技术如直方图均衡化和对比度拉伸可以帮助提高缺陷的可见性和对比度。图像分割有助于将缺陷从图像的其余部分分离出来,使它们更容易分离和分析。在传统方法中,通过图像处理技术对通过各种成像技术获取的图像信息进行预处理,以获取拼接图像、感兴趣区域、复原图像和分割子图等,是表面缺陷检测的关键步骤。在基于深度神经网络的方法中,通过图像预处理技术来获取适用于网络输入的图像,能够显著提高检测方法的效率和稳定性。

2.1 图像增强

在金属表面缺陷检测系统中,图像预处理的主要任务是抑制噪声并增强缺陷区域的特征,令图像更容易被人眼观察或者被算法检测。由于工业现场工作环境恶劣、成像系统和信号传输线路的物理性质,图像不可避免地会产生噪声,因此需要进行图像预处理来抑制噪声或复原图像。图像增强的常用方法可分为基于滤波器的方法、基于模型的方法和基于学习的方法。

基于滤波器的方法利用了噪声能量和正常图像的频谱分布在一定范围区间的特点^[35],通过人工设计的带通滤波器来抑制噪声,常见的方法包括快速傅里叶变换、小波方法^[36]等。基于模型的方法尝试对噪声和复原图像的分布进行建模,并使用模型分布作为先验来获得清晰的复原图像和优化算法^[37],包括非局部自相似模型 (Non-local self similarity, NSS)、梯度模型、稀疏模型和马尔科夫随机场模型。基于学习的方法则将图像去噪任务定义为学习一个有噪声图像到正常图像的映射,一般可分为基于传统机器学习和基于深度学习的方法。在最近的研究中,基于深度学习的方法比传统方法获得了更好的性能,已经逐渐成为主流方法。基于深度学习的方法可以很好地整合进端到端深度学习方法,因此在缺陷检测网络中已经不是一个必要的步骤。然而传统图像去噪增强算法有着计算量小、易于硬件实现等优点,因此常常集成在工业相机硬件中以改善图像质量。

2.2 图像特征提取

特征提取将高维图像空间的样本特征映射到低维特征空间,以便缺陷检测算法进行处理。这就要求提取出的特征不仅可以很好地描述图像,同时可

以更好地区分不同类别的对象。目前针对二维成像检测技术的特征提取已经相对成熟。而针对三维点云数据则有两种处理方式,一种将三维点云数据变换到二维图像域并进行二维图像特征提取,另一种方式则直接从三维点云数据中捕获形状特征用于表面缺陷识别任务。

2.2.1 二维图像特征提取

统计法把纹理和形状看作随机变量,用统计量来表述纹理的随机变量分布,包括直方图特征、灰度共生矩特征 (Gray-level co-occurrence matrix, GLCM)^[38]、局部二值模式 (Local binary patterns, LBP)^[39]、方向梯度直方图 (Histogram of Oriented Gradient, HOG) 特征^[40]、尺度不变特征变换 (Scale-invariant feature transform, SIFT) 特征^[41-42]、Haar-like 特征^[43-44]等。信号处理法把图像当做二维分布的信号,通过设计滤波器将纹理转到变换域,并采用响应的能量准则提取纹理特征,主要包括傅里叶变换、Gabor 滤波器^[45-46]、小波变换、Laws 纹理等。模型法尝试拟合一个模型,将不同取值的模型参数作为纹理特征,包括随机马尔科夫场 (Markov random field, MRF)、分形模型和自回归模型等,这一类方法的关键在于如何估计模型的参数。

GLCM 是最常用的纹理特征,在许多研究中^[47]与基于 Gabor 滤波器和基于滤波的 LBP 方法对比性能略佳,目前已经成为一种非常流行的方法。Liu 等^[48]提出了一种新的 Haar-Weibull-variance (HWV) 模型,用来表征图像中各个局部斑块的纹理分布,在无监督金属表面缺陷数据集上的检出率达到 96.2%,优于以往算法。Pernkopf 等^[49]提出了一种基于多种特征提取技术的钢板表面缺陷检测系统,通过结合 GLCM、Gabor 小波和 HOG 特征并采用随机森林作为分类器,在 Kaggle 钢板缺陷检测数据集^[50]中获得了 95% 的准确率。

与手工设计的传统特征提取方法相比,深度学习方法从大量数据中自动学习特征表示,相对于传统方法表现出了显著的优势。基于深度学习的图像特征提取方法主要包括基于深度卷积网络 CNN 的方法和基于 ViT^[51]的方法。前者用堆叠的卷积层将图像逐层映射到特征图 (Feature map) 中,在堆叠的过程中对图像进行下采样以减少特征尺寸,同时通道数量成倍增加,以获得更加丰富的深层语义特征。后者是一种基于编码器-解码器 (Transformer) 结构的图像特征提取器,它把图像划分成 $K \times K$ 个图像块 (patch),同时添加一个可学习的类别块,将每个 patch 投影为固定长度的向量,加入位置编码后输入 Transformer 结构,之后的处理流程与原

始 Transformer 完全相同. 通常当数据集规模较小时, CNN 的表现通常更好; 而 ViT 则可在一个很大的数据集上进行预训练, 然后在数据量较小的下游任务中进行微调, 在最近的研究中取得了比 CNN 更好的效果.

2.2.2 三维点云特征提取

在三维成像检测技术中, 三维点云数据提供额外的空间信息, 从而有助于获取更多缺陷特征信息. 然而大多数图像处理技术并不是针对三维点云数据来设计, 因此需要将点云数据从三维空间投影到二维平面进行处理. 这类方法目前已经取得了显著的成果, 但仍然存在丢失空间信息以及不必要的冗余信息表示的问题^[52]. 另一些研究者则研究直接对三维点云数据提取特征的方法, 主要包括手工特征和基于深度学习的特征等.

直接提取三维点云特征的早期研究集中于手工特征, 例如曲率、法向量和几何描述符, 这些特征被设计用于捕获点云的特定属性, 这些属性对于区分表面缺陷与正常背景十分重要. 表面法线是垂直于物体表面的向量, 这些向量可用于识别点云中物体的方向; 曲率是衡量一个物体偏离平面的程度, 它可以用来识别物体的形状; 关键点是在点云中被认为对缺陷识别任务很重要的点, 可用于识别物体上对缺陷识别很重要的特定点; 描述符是一组描述对象的局部形状和纹理的特性. 它们可用于识别和匹配不同点云中的相似形状^[53].

深度学习能够从大型非结构化数据集中学习和提取复杂特征, 因此已经成为从三维点云数据中提取特征的流行方法. 斯坦福大学的 Qi 等^[54]提出了 PointNet 模型, 它可以直接输入点云向量, 通过学习一个变换矩阵来对齐, 从而获得具备空间变换不变性的空间编码表示, 然后通过最大池化融合所有点的特征来提取全局点云特征. PointNet 可以很好地完成缺陷分类任务, 但局部特征提取能力较差, 缺乏复杂背景的分析能力, 因此 PointNet++^[55]加入多层次特征提取结构对局部特征提取能力进行了改进. PointMLP^[56]提出了一种简单的三维点云特征提取架构, 仅使用残差前馈多层感知机 (Residual feed-forward, MLP) 和一个轻量化的几何变换提取器, 在 ScanObjectNN 数据集上取得了 3.3% 的精度提升和推理速度的显著提高. PointNeXt^[57]提出了一套改进的训练策略, 在没有架构变化的情况下将 PointNet++ 的分类准确率从 77.9% 提高到 86.1%, 优于 PointMLP 的 85.4%, 并在 PointNet++ 中引入了反向瓶颈结构和可分离 MLP, 实现了最先进的效果.

2.3 图像分割

在大多数工业场景下仅对产品某一块区域是否存在缺陷感兴趣, 因此需要先定位 ROI 后, 再对 ROI 进行缺陷检测和分类, 以降低计算消耗并提高检测效率. 图像分割的原则是使划分后的子图保持内部相似性最大、同时子图之间相似性最小. 经典分割方法主要包括阈值法、区域生长法、分水岭算法、边缘检测法、小波变换法^[58]和基于主动轮廓模型的方法.

文献 [59] 提出了一种基于三维形态学和分水岭算法的锈蚀钢结构表面锈蚀坑识别、提取和评价的新方法. 分水岭算法对微弱边缘具有良好的响应, 并且可以保证得到封闭的边缘, 但是存在过度分割的问题. 郭皓然等^[60]用全局自适应阈值法对高反射金属零件表面进行缺陷区域分割和识别, 利用空域和值域信息的滤波方式对图像进行处理, 并以高斯函数一阶导数构建边缘检测器, 实验结果验证了方法的有效性和可靠性. 魏爱东^[61]用大津法 (Otsu's method) 与最大熵分割算法结合, 在金属热图像缺陷识别和区域分割任务上取得了较好的效果. Prabha 等^[62]利用彩色通道对缺陷图像进行预处理, 检测光照不均匀, 对缺陷图像进行后处理, 利用边缘检测和轮廓将缺陷部分从原始图像中分离出来, 从而检测出产品中的缺陷, 比现有方法更加稳定, 过程如图 9 所示. Zhang 等^[63]基于小波变换、大津法、光谱测量 (Spectral measure) 和支持向量机 (Support vector machine, SVM) 开发了一种针对高反射金属表面的缺陷检测系统. 实验表明该方法可以有效地对高反射金属表面的七种典型缺陷进行表征, 优于 BP 神经网络方法.

2.4 图像拼接

在工业金属板、带加工场景, 由于检测对象往往尺寸较大, 为了同时满足大视场、高分辨率成像的需求, 往往采用扫描成像和阵列成像, 或者两种方法的组合来实现. 在阵列成像系统中, 需要采用图像拼接技术将多个光学传感器获取的图像通过图像匹配、重投影和融合技术来生成更大画布的图像, 以满足大幅面金属表面在线缺陷检测的需求.

图像拼接可分为三个步骤: 图像匹配 (Registration)、重投影 (Reprojection) 和融合 (Blending)^[64]. 图像匹配通过求解多个图像坐标系之间的几何变换矩阵来寻找几何对应关系; 重投影通过几何变换将多张图像投影到同一坐标系; 最后融合图像之间重叠的像素, 获取在空间上或者通道上扩展后的图像. 在工业金属表面检测领域, 通常对空间精度有较高

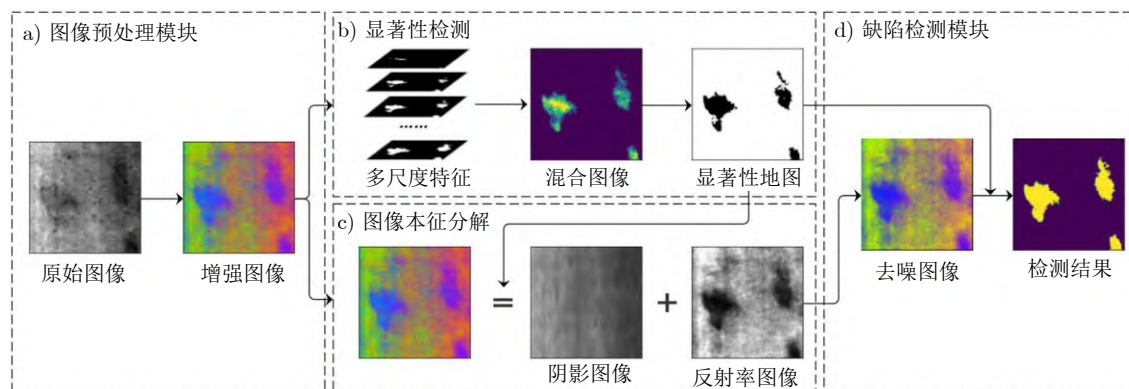


图9 基于图像分割的缺陷检测

Fig.9 Defect detection based on image segmentation

的要求, 因此往往使用图像像素的特征进行图像匹配。

图像拼接不仅可以在空间上进行, 同时也可以通道上进行。通过不同方向布置多个不同类的二维、三维传感器, 并用图像拼接技术来获取单一融合图像, 以实现多通道角度分辨技术。通过融合多个通道, 可以复合多种不同特征, 不仅可以检测金属表面凹凸纹理, 还可以分别检测油污、指纹、轻微划痕、曲率变化等缺陷, 以实现缺陷特征的光学方法分离, 减少缺陷检测算法的难度^[65]。

3 缺陷检测器

缺陷检测器的主要功能是从特征图中识别缺陷, 并对缺陷进行分类和定位。根据表面缺陷检测算法的任务目标和实现原理, 大致可分为模板匹配 (Template matching)^[66]、图像分类、目标检测、图像语义分割和图像异常检测^[67] 五类。图像分类把缺陷检测视为分类问题, 即判断图像窗口内是否为缺陷, 或属于哪一类缺陷^[68]。目标检测可以一次性检测出缺陷的类别、位置和大小, 其中缺陷的位置和大小可以用一个包围盒 (Bounding box) 来描述^[69]。图像语义分割需要找出背景中可能存在的缺陷区域, 并输出像素级的缺陷边界^[70]。这些算法会将已知的缺陷归纳成明确的类别, 通过标签对缺陷图像的位置、大小、边缘、类别进行标注, 因此这类方法更关注缺陷的特征, 通常用有监督学习的方式来构造和训练模型。而异常检测则关注正常产品的特征分布范围, 将与正常产品有明显不同的样本都看作缺陷样本, 通常只需要无缺陷样本即可进行训练, 该方法又被称为异常检测, 更关注正常样本的特征^[71]。

3.1 模板匹配

在工业表面缺陷检测任务中, 最理想的情况是正常样本都高度相似, 缺陷样本与正常样本仅在小

范围有差异^[72]。模板匹配是解决这类任务最有效的检测方法。模板匹配根据已知模板图像到另一待检图像中寻找与模板图像相似的子图像, 然后通过相似度度量计算模板与待检样本之间的相似性, 即可判断样本是否存在缺陷。模板匹配方法一般可分为两大类, 一类是基于灰度匹配的方法, 另一类是基于特征匹配的方法。

基于灰度匹配的方法^[73] 也称为图像相关匹配算法, 是一种逐像素地把一个匹配窗口与模板图像的所有可能窗口的灰度阵列按某种相似性度量方法进行搜索比较的匹配方法。图像相关匹配算法具有计算量小、易于硬件实现等优点, 但是要求两幅图像有大量重复像素。

基于特征的方法大致可分为点特征匹配^[74]、边缘特征匹配^[75] 和区域匹配三大类别。点匹配通常用人工设计的特征提取方法, 例如 Harris 特征、ASIFT (Affine Scale-Invariant Feature Transform) 特征、ASIFT (Affine Scale-Invariant Feature Transform) 特征、光流法等^[76]。边缘特征包括 LOG (Laplacian of Gaussian) 算子、Robert 算子、Sobel 算子、Canny 算子等。由参与匹配的点特征、边缘特征构成特征空间, 在特征空间中通过相似度度量来确定模板与待检样本特征之间的相似性, 它通常定义为某种代价函数或者距离函数。经典的相似度度量包括 Minkowski 距离、Hausdorff 距离、互信息等。

除此之外, 由于光学成像条件和待检目标定位方式带来的影响, 还需使用图像几何变换用来解决两幅图像之间的几何位置差异, 包括刚体变换、仿射变换、投影变换、多项式变换等。搜索策略是用合适的搜索方法在搜索空间中找出几何变换参数的最优估计, 使得图像经过变换之后的相似度最大, 包括穷尽搜索、分层搜索、模拟退火、动态规划、遗传算法和神经网络等。遗传算法采用非遍历寻优搜索

策略,可以保证结果具有全局最优性,并且计算量最小^[77];神经网络具有分布式存储和并行处理方式、自组织和自学习的功能以及很强的容错性和鲁棒性,因此这两种方法在图像匹配中得到了更为广泛的使用^[78]。

基于模板的计算机视觉检测方法原理简单,对算法要求不高,在大规模生产、产品一致性较高的金属机加工零配件生产中应用十分广泛。但这类方法需要提前制作模板,仅能用于检测重复性强的相同部件或产品,应用场景受到了一定限制,对于金属铸坯这一类正常样本之间也会存在较大差异的产品,难以通过模板匹配进行缺陷检测。为了解决这个问题,文献[79]提出了一种基于导向模板的带钢表面缺陷检测算法,通过大量无缺陷图像获得正常纹理的统计特征,对每个测定对象生成唯一模板,最后在像素级检测的基础上对图像和模板进行差分、自适应阈值等操作以定位缺陷。该方法在包含 1500 张图像的测试集上达到了 96.2% 的平均检出率。

3.2 图像分类

图像分类是计算机视觉领域的基础问题。基于图像分类的表面缺陷检测算法将已知的缺陷归纳成明确的类别,通过算法给样本分配一个标签,以实现最小的分类误差。通过与滑动窗口法组合,图像分类可以实现较粗粒度的缺陷定位。通过图像分割与图像分类组合,可以实现像素级的缺陷边界检测。根据模型结构设计中是否有神经网络参与,通常又可分为传统机器学习方法和基于深度学习的方法两大类。

3.2.1 传统机器学习方法

在金属表面缺陷检测任务中,统计模式识别是最基本的传统机器学习方法。统计模式识别就是利用有限数量的给定样本集,在已知总体样本统计模型或已知判别函数类的条件下,根据一定准则学习一个模型,将样本特征空间划分为 c 个区域,根据样本落入哪一个区域来判别样本所属的类别。通过图像预处理技术抑制噪声并通过特征提取和特征选择来控制样本特征空间,使样本在特征空间中的分布尽可能满足以上条件。在将原始图像信息映射到特征空间后,即可通过各种准则函数设计分类器,常用的分类器包括贝叶斯分类器、决策树、线性判别器、 K 近邻法、支持向量机、集成分类方法等。

文献[80]提出了一种具有三维特征的钢块表面缺陷检测方法,在构建贝叶斯网络结构时,采用浮动搜索算法,实现了分类性能和结构学习计算效率的良好平衡,分类准确率达到 98%,优于朴素贝

叶斯和树增广贝叶斯策略。文献[81]通过使用决策树,将主成分分析(Principal component analysis, PCA)和引导聚合(Bootstrap aggregation, bagging)应用于基于局部二值模式的算子提取的特征上,用于钢铁表面缺陷的两类分类。该方法快速准确,实现了钢轨在线实时自动表面检测。在文献[82]中作者基于六轴机器人上的二维激光传感器采集的图像数据,采用 GLCM 和其他方法来提取特征来训练 SVM 用于缺陷检测,检测精度超过了 96%。在缺乏先验知识或总体样本统计模型的情况下, K -means 是一种有效的无监督分类方法,能够有效解决工业金属表面缺陷检测缺乏标注数据时“冷启动”的问题。文献[83]利用 SURF 算法提取气缸图像的特征点,结合粒子群优化和 K -means 算法,验证了该算法在实际过程中的适用性,在钢制气缸密封件的检测准确率达到 98%。集成分类方法是一种将多个弱分类器进行加权融合,从而构成强分类器的方法。文献[84]提出了一种基于 AdaBoost 多分类器组合的钢轨表面缺陷识别方法,通过提取缺陷区域的几何形状和灰度特征描述缺陷属性,利用 Relief 算法过滤掉无关特征,然后采用 AdaBoost 多分类器组合方法,以分类回归树(Classification and regression tree, CART)决策树为弱分类算法设计组合分类器,能够有效识别钢轨表面剥落块、胎面裂纹和鱼鳞剥落裂纹三种常见缺陷类型。

3.2.2 基于深度学习的方法

传统机器学习方法往往依赖人工设计的特征模板和对数据分布的假设,而在真正的工业生产环境中,待检图像往往受到形状、位置、光照、背景和成像设备差异带来的各种噪声影响。这使得传统机器学习模型往往只能在十分苛刻的环境下使用。随着深度学习技术的发展,基于 CNN 的分类网络目前已经成为表面缺陷检测中非常重要的技术。一般而言,表面缺陷分类往往采用经过预训练的 VGG、ResNet、DenseNet、SENet 等网络作为骨干网络,然后针对实际问题搭建简单的网络结构,输入一幅测试图像到网络中,输出缺陷分类及其置信度,或者一组特征向量。根据分类网络的实现方法差异,可将其细分为直接用网络分类、利用网络进行缺陷定位和利用网络作为特征提取器三种方式。

直接利用 CNN 进行缺陷分类是最早应用于表面缺陷检测任务中的深度学习方法。2014 年,奥地利科技研究所最早训练 CNN 来实现轨道表面空洞缺陷检测,整个网络仅包含两个卷积层,两个池化层和一个全连接层,在钢轨表面缺陷数据集上达到了 1.108% 的错误率^[85]。Masci 等^[86]提出采用最

大池化层的 CNN 用于钢带表面缺陷分类任务, 错误率仅 7%, 远远超过了多种基于手工设计特征的传统机器学习模型, 包括 SIFT、SURF 和 SVM 方法的效果。

与基于手工特征提取的早期方法相比, 深度度量学习 (Deep metric learning) 使用深度神经网络直接学习相似度度量^[87]。这种方法对于个体级别的细粒度识别非常有用, 常见的应用是图像匹配、行人重识别^[88] 或人脸识别^[89]。常见的度量学习架构包括孪生网络 (Siamese network) 和三元网络^[90]。不同于表征学习输入单幅图像来进行分类、定位或分割任务, 孪生网络的输入为两幅成对图像, 通过网络学习输入图片的相似度, 来判断是否属于同一类别。如图 10 所示, 在三元网络架构中^[91], 由三个不同图像组成的三元组被输入同一个网络, 其中两张属于同一类, 一张属于不同的类, 然后训练网络以创建一个特征空间, 在该空间中同类别样本之间的距离比不同类别样本之间的距离要小。

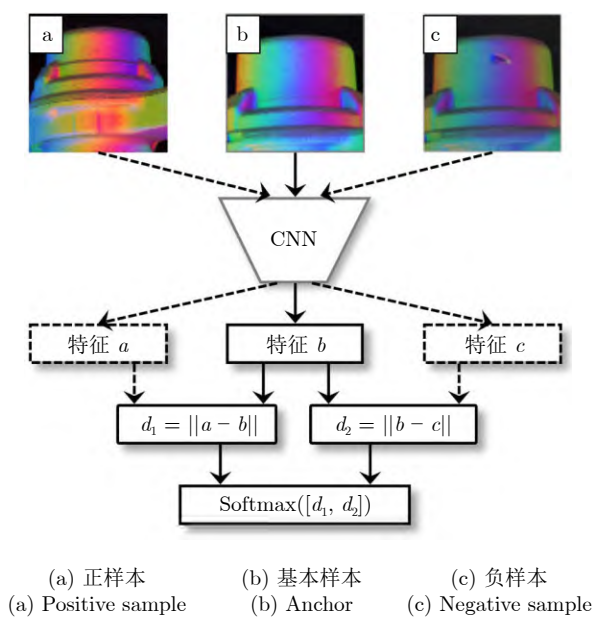


图 10 三元网络结构

Fig. 10 The structure of triplet network

深度度量学习用于表面缺陷检测的一个基本策略是计算正常样本与待检样本之间的距离。Kim 等^[91]设计了一个基于 CNN 的孪生网络对钢表面缺陷进行分类, 将两幅图像输入到共享权值的 CNN 中完成特征提取, 然后用基于相似度函数的对比损失计算图像特征的差异程度, 该方法在 NEU-CLS 数据集上仅用 10 张图片样本进行学习, 即可在 9 种缺陷的数据集上达到 86.5% 的分类准确率。Wu 等^[92]提出了一种基于孪生网络的度量学习模型来对纽扣

缺陷进行分类, 该方法在包含凹陷、裂缝、斑点、气孔、凹凸不平等多种缺陷的数据集上达到了 98% 的分类精度。Li 等^[93]采用三元网络架构进行工业制品表面缺陷检测, 在 Cifar100 数据集上进行预训练, 达到了 84.5% 的分类精度。与表征学习相比, 度量学习可以看作学习样本在特征空间中的流形分布, 而表征学习则可看作学习样本在特征空间的分界面。在工业金属表面缺陷检测领域, 经常使用人工生成正样本来训练网络以解决正样本稀少的问题。

样本分类是工业缺陷检测中最基础的任务, 分类指标通常可以用混淆矩阵 (Confusion matrix) 来表示, 通常分类性能包括准确率 (Precision)、召回率 (Recall)、精准率 (Accuracy)。

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

而在工业表面缺陷检测领域, 通常更关注误检率 (False drop rate, FPR) 和漏检率 (False negatives rate, FNR), 它们直接衡量表面缺陷检测系统的不足。在统计质量管理体系中, 用户对于 FNR 往往要求低于 0.03%, 这对于依赖大量数据驱动的深度机器学习模型来说是极大的挑战。

$$FPR = \frac{FP}{FP + TN} \quad (4)$$

$$FNR = \frac{FN}{FN + TP} \quad (5)$$

在实际工作场景中, FPR 和 FNR 存在逆相关性, 当 FPR 提升时 FNR 往往会不可避免的降低。式 (4) 只能在某一具体的置信度阈值下评估这两个性能指标, 因此可以设置从 0 至 1 的一系列阈值, 通过两个指标在不同置信度阈值下的关系曲线来表示模型性能, 常用的方法包括 PR (Precision-recall curve) 曲线和 ROC 曲线, 并可用 ROC 曲线下面积 (Area under the receiver operating characteristic curve, AUROC) 来度量对应曲线所表示的性能。平均精度值 (Average precision, AP) 是一种常用的方法, 对 PR 曲线进行平滑处理后使用积分来计算。

3.3 目标检测

一般意义而言, 目标检测任务是最接近于表面缺陷检测的任务, 在工业生产领域应用十分广泛。它将目标的识别与定位合并成为一个任务, 准确性

和实时性是系统的重要能力,尤其是在一个场景中对多个可能发生重叠的目标进行实时处理时.常用的基于深度学习的目标检测模型从结构上可以分为两类:二阶段 (Two-stage) 模型和一阶段 (One-stage) 模型.两者的主要区别在于,二阶段模型需要首先提议可能包含缺陷的候选框,然后再对候选框中的子图进行目标分类;而一阶段模型则可以通过 CNN 提取的特征,直接预测缺陷的位置、大小、类别和置信度.

3.3.1 二阶段网络

二阶段网络算法的经典算法是 R-CNN. R-CNN 首先基于提议候选框网络 (Region proposal network, RPN) 产生区域候选框,最后对每个候选框回归位置框并分类. 2014 年提出的 R-CNN 模型将目标检测技术成功运用于工业场景^[94],这一类方法可以同时找出缺陷的位置、大小和类别,实现了端到端的外观缺陷检测. R-CNN 候选区域缩放后的畸变问题和提取特征时的重复计算导致了模型性能和速度的瓶颈.为了解决这些问题,2015 年,He 等^[95]提出了 SPP-net,在保证性能的同时,检测速度也有了较大的提升.在其基础上,Girshick^[96]提出了 Fast R-CNN,该网络吸收了 SPP-net 的特点,使得目标检测的速度大幅提升.随后又提出了 Faster R-CNN,将特征抽取、获取建议框和目标分类器整合到一个模型中,大大提高了模型的性能^[97],网络结构如图 11(a) 所示. He 等^[98]于 2017 年进一步提出了 Mask R-CNN,采用残差网络 (ResNet) 作为基础网络以及 FPN 用于挖掘多尺度特征信息,并加入全连接的分割子网,由原来的分类-回归两个任

务变成了三个分支任务:分类、回归、分割.以上方法的检测器通常只在一个较小的交并比 (Intersection over union, IoU) 阈值内取得最佳的性能,因此怎样选取合适的阈值是十分重要的.为了解决这个问题,文献 [99] 提出了 Cascade R-CNN,通过级联的检测器,在每个阶段采用递增的 IoU 阈值,从而使提议框 (Proposal box) 的质量逐渐提高,成为了二阶段网络的一种重要的范式.

2021 年 Guo 等^[100]采用 Mask R-CNN 对铁轨表面缺陷进行检测和缺陷区域的语义分割.训练时采用了 1140 张经过标注的图片,并在 16 张图片上进行了测试,实验表明 Mask R-CNN 可以有效检测缺陷,具有应用价值,如图 12. 文献 [101] 提出了一种基于 Mask R-CNN 的隧道表面缺陷检测方法,针对隧道表面背景复杂的问题,在 Mask R-CNN 上增加了一个路径增强特征金字塔网络 (Path aggregation network for feature pyramid networks, PAFPN) 和一个边缘检测分支,实验表明它们在隧道表面缺陷检测和分割任务上具有很好的鲁棒性和准确性,AP 达到了 85.35%,高于原始版本 Mask R-CNN 的 74.35%. Fang 等^[102]将注意力机制引入 cascade R-CNN,提出了 ACRM (Attention cascade R-CNN with mix-nms) 用于缺陷检测,在真实工业金属缺陷数据集上取得了最先进的性能.

3.3.2 一阶段网络

YOLO (You only look once)^[103] 的名称意为“你只需要看一次”即可识别出所有的目标.相比二阶段网络,YOLO 不需要提议候选框,可直接根据全图背景信息识别出所有的物体的类别和位置,因此又被称为 Region-free 方法,流程如图 11(b). 2020 年 Bochkovskiy 等^[104]提出了 YOLO v4,引入 PAN (Path aggregation network) 作为网络的 Neck 部分,通过将不同尺度的特征来回融合两次,显著提升了模型的性能.另外 YOLO v4 采用了更多的数据增强方法,例如切片混合 (CutMix)、马赛克 (Mosaic)、标签平滑 (Label smoothing) 等.文献 [105] 将注意力机制嵌入主干网络中,将 PAN 结构修改为定制化感受野块 (Customised receptive field block) 结构,对四类钢带缺陷检测的平均精度达到了 85.41%. Usamentiaga 等^[106]将 YOLO v5 应用于金属表面缺陷检测,在 NEU-1800 金属缺陷数据集上以 28 ms/帧的检测速度达到了最高的检测精度,效果远远超过 SSD、Fast R-CNN、YOLO v2 和 YOLO v3 的修改版本,在检测效率和精度上均达到了工业在线实时检测的要求.文献 [107] 利用生成对抗网络 (Generative adversarial networks, GAN) 对数据集进行扩展,然后用 EfficientDet-D3

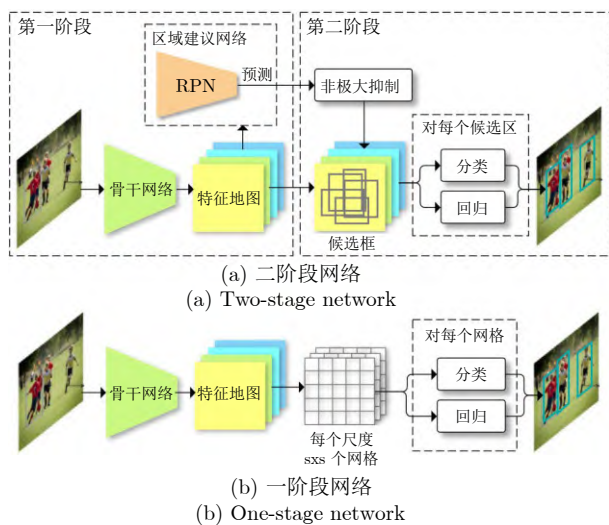


图 11 二阶段网络和一阶段网络对比

Fig. 11 Comparison of two-stage and one-stage networks

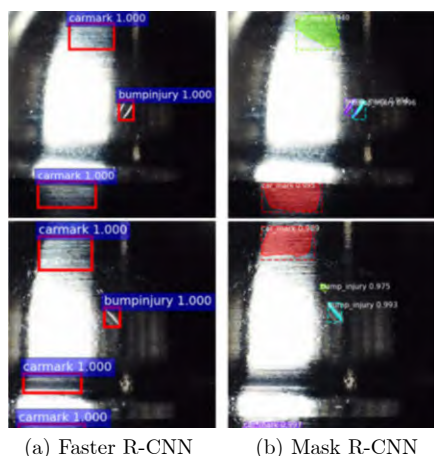


图 12 基于二阶段网络的金属表面缺陷检测^[100], 经许可转载自文献 [100], ©Sage, 2021

Fig.12 Metal surface defect detection based on two-stage networks^[100], reproduced with permission from reference [100], ©Sage, 2021

网络进行轨道表面缺陷检测, 在检测速度和精度之间取得了最佳的平衡。

在以往研究的基础上^[108]提出了融合多重检测的端到端金属表面缺陷检测方法 (Defect detection network, DDN). 为了实现强大的分类能力, 该模型采用骨干网络 (Backbone) 在每个阶段生成特征图, 然后用多级特征融合网络 (Multimodal factorization network, MFN) 将多层特征融合, 并采用 RPN 来生成 ROI, 对于每个 ROI 用分类器和边界框回归器组成的检测器来输出最终的检测结果. 该模型在缺陷检测数据集 NEU-DET (North-eastern University surface defect database)^[109]上实现了 74.8/82.3 的平均精度均值 (mean AP, mAP) 以及 20 FPS 的检测速度, 使模型达到了工业化应用的水平. 文献 [110] 在 NEU-DET 数据集上对当时最先进的五种目标检测模型进行了基准测试, 包括 Faster R-CNN、Deformable DETR、Double Head R-CNN^[111]、RetinaNet^[112] 和 Deformable convolutional network (DCN)^[113], 实验结果表明 DCN 模型有着最高的准确率. DCN 通过在可变形卷积网络中的空间采样点随机添加偏移量, 并在训练阶段学习偏移量, 从而实现了所有缺陷类别上的最高精度。

3.3.3 最新的方法

上述一阶段目标检测方法均基于锚框 (Anchor-based), 这一类方法均需要对正负样本进行密集采样, 不仅计算量大, 同时正负样本的划分依赖一定的先验假设, 例如中心先验原则假设物体的像素分布总是在物体中心附近. 但这些先验假设未必是最

优的, 例如细长弯曲的划痕和连续分布的裂纹, 这会带来额外的误报或漏检. 文献 [114] 提出了一种新的无锚框 (Anchor-free) 目标检测框架, 通过基于中心点的关键点检测算法来寻找目标的中心点, 并回归其他对象属性, 比基于锚框的方法更快、更准确. 2021 年旷视科技^[115]提出了 YOLOX 算法, 采用了 Anchor-free 方式, 每个位置只需要预测一个包围盒并直接预测四个值, 同时采用 multi positives 技术, 将每个目标的中心点 3×3 邻域作为正样本进行采样, 以改善单个网格中目标密集的问题. AutoAssign^[116]提出了一种完全可微的动态标签加权分配方法, 通过动态生成锚点并不断迭代来更好地匹配物体的形状, 同时用一种新的损失函数来鼓励模型生成高质量的边界框, 在目标检测领域实现了最先进的性能 (State-of-the-art). 文献 [117] 提出了 CA-FPN 结构, 通过学习不同尺度特征之间的关系实现了多尺度特征的自适应融合, 并结合 AutoAssign 提出了改进的 CA-AutoAssign 网络, 在 NEU-DET 数据集上达到了 82.7% 的平均精度 (mAP), 超过了专门为 NEU-DET 数据集设计的 DDN, 并且检测时间几乎没有增加。

2020 年^[118]提出的 DETR 将基于 Transformer 的方法成功应用于目标检测领域. 如图 13 所示, 图片经过 CNN 骨干网络获取特征图, 加入位置编码后输入编码器 (encoder), 与一系列可训练的目标查询向量 (Object queries) 在解码器 (decoder) 中进行交叉注意力计算, 输出的结果经过 FFN (Prediction feed-forward networks) 层后直接获得回归框和类别, 不需要非极大抑制 (Non-maximum suppression, NMS) 计算, 实现了真正的端到端 (End to end). 文献 [119] 将 DCN 与 Transformer 结合, 提出了 deformable DETR (Detection) 网络, 舍弃了每个查询向量 (Query) 与键 (Key) 点乘的思想, 通过邻近较小集合内的元素采样进行相关性刻画. 同时 Deformable attention 模块能够通过注意力机制聚合多尺度特征图, 从而解决了 DETR 小目标检测效果不佳的问题. 文献 [120] 在 DETR 的基础上利用 Transformer 结构实现了三维点云目标检测。

目标检测算法在金属表面缺陷检测任务中的性能, 通常基于 IoU 阈值 0.5 时的 mAP. 多 IoU 评估方法将 IoU 阈值划分为 0.5 至 0.95 的十个值, 并计算每个 IoU 阈值下的平均 AP, 得到最终的性能度量标准. 另外在工业高速在线检测时, 还要评估检测速度, 通常用每秒处理帧数 (Frame per second, FPS) 来表示. 在这一类任务中, 常用的数据集包括 NEU-DET^[109] 和 GC10-DET^[121], 其中 NEU-DET 提供了包含 6 种缺陷的钢带表面缺陷图像, 每种

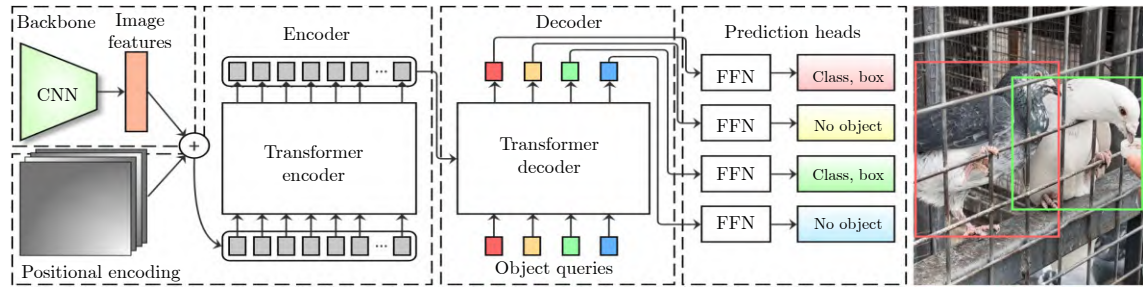


图 13 DETR 网络结构

Fig. 13 The network architecture of DETR

表 1 目标检测模型在 NEU-DET 上的表现
Table 1 Performance of object detection models on NEU-DET

Method	Backbone	Neck	AP_{50}
Faster R-CNN ^[97]	ResNet-50	FPN	74.7
Cascade R-CNN ^[99]	ResNet-50	FPN	75.8
YOLOX ^[115]	CSPDarknet	PA-FPN	70.9
YOLOv4 ^[105]	CSPDarknet	FPN	76.4
AutoAssign ^[116]	ResNet-50	FPN	76.6
AutoAssign ^[116]	Swin-Tiny	FPN	78.3
DDN ^[108]	ResNet-50	MFN	82.3
CA-AutoAssign ^[117]	CSPDarknet	CA	82.7

类别提供 300 张标注了边界框的灰度图片. 表 1 中展示了几种常用方法在 NEU-DET 数据集上的表现.

3.4 语义分割

目标检测网络可以检测缺陷的类别、位置和大小,但在某些工业场景中缺陷的形态和分布是不定的,此时用一个包围盒很难准确描述缺陷的分布,因此还要对每一个缺陷以像素级的精度标记出缺陷的边界.

语义分割网络将表面缺陷检测任务转化为像素级分割任务,一种简单的分割网络可以将正常背景与缺陷区域分割;另一种网络可以将多个缺陷分割成不同的实体.语义分割网络不仅能分割出缺陷,而且可以精确描述缺陷的位置、类别以及像素级的集合属性,例如长度、宽度、面积、轮廓、几何中心等等.根据分割网络的原理不同,可以分为基于全卷积神经网络的方法和基于上下文知识的方法.

3.4.1 全卷积神经网络

全卷积神经网络 (Full connected network, FCN) 是一种经典的端到端图像语义分割方法,网络中所有的层都是卷积层,故称为全卷积神经网络^[122].全卷积神经网络由三种结构组成:卷积 (Convolutional)、上采样 (Upsample) 和跳级 (Skip layer).

在传统的 CNN 中,输入的图像经过多次卷积和池化以后,尺寸越来越小,通道数越来越多,最终得到高维特征图,然后将特征图展平成一维向量,通过全连接网络和 Softmax 层进行分类.而在 FCN 中,高维特征图通过上采样操作将图像放大回原图像的尺寸,输出像素级分类.但是在多次下采样过程中损失了位置信息,因此得到的分割结果比较粗糙,所以作者通过跳级,将倒数几层的输出也进行上采样,并与最后一层的输出结果混合.2015 年 Ronneberger 等^[123]在 FCN 的基础上提出了 U-net,是一种具有 U 形结构的编码器-解码器型网络.相比 FCN, U-net 对跳层结构进行了改进,对所有下采样的特征图都与对应倍数的上采样特征图都进行了融合,同时用 concat 方法在通道维度进行堆叠,替代了 FCN 特征融合时的加法,虽然增加了计算量但是缓解了梯度消失的问题,加快了收敛速度,同时有利于分辨率的恢复.剑桥大学于 2017 年提出了 SegNet,该网络的解码器采用编码器在最大池化中的池化索引 (Pooling indices) 进行非线性的上采样,以避免下采样时的位置精度损失,从而提升边界轮廓的质量,在保持端到端训练特性的同时减少参数的数量.

在最近的研究中 Sabet 等^[124]在多个金属表面缺陷数据集上测试了 U-net 和 FCN-8 的表现,实验采用 U-net 和 FPN^[125]架构以及 ResNet、SE-ResNeXt^[126]和 EfficentNet^[127]作为骨干网络来寻找最佳的编码器.实验结果表明 EfficientNet-B0 表现出了最佳的效果.而 U-net 和 FCN-8 则分别实现了 0.97 和 0.98 的 Dice 系数,其中 U-net 在分类任务上表现更佳,而 FCN-8 对复杂缺陷形状的分割效果更好.

3.4.2 基于上下文信息的方法

语义分割是一种像素级标注任务,一般 FCN 的一元损失函数只考虑了标注中的独立像素,而忽略了与相邻像素之间的上下文语义关系.同时由于卷积层的感受野只能随层数线性增加的限制, CNN 并不擅长掌握图像的全局上下文信息.为了

解决这个问题, 通常采用以下几种方法: 条件随机场 (Condition random field, CRF)、循环神经网络 (RNN)、多尺度特征融合、图神经网络 (GCNs) 和基于记忆 (Memory-based) 的方法等。

CRF 常用于语言模型和序列标注任务, 引入 CRF 层作为后处理模块使 CNN 获得了相邻像素的空间约束, 它将相邻的相似点标记为同一类别, 强化了底层特征信息相互依赖, 实现了像素级预测和推理输出的结合。文献 [128] 提出了 DeepLab 模型, 采用全连接 CRF 模型作为后处理步骤以优化分割结果。该方法将每个像素视为无向图的节点, 无论像素是否相邻都能两两考虑其语义关系, 因此又被称为全连接因子图。CRFasRNN^[129] 则通过平均场近似, 将密集 CRF 近似为 RNN 结构, 将其作为网络的模块整合进 FCN, 从而实现了一个完整端到端语义分割网络。

多尺度特征融合是向 FCN 输入上下文信息的重要途径, 它将一个下采样倍数更大的特征图全局信息与一个下采样倍数较小的特征图局部信息结合, 通常可分为早融合和晚融合两种方式。早融合方式在相邻的层直接融合多个尺度的特征, 然后在融合特征上训练预测器以得到上下文信息, 然后输入下一层。U-net 是一种典型的晚融合结构, 通过跳级链接进行延迟特征融合, 即在每层独立预测, 然后逐层对预测结果进行融合和上采样, 最终输出原始尺寸的预测结果。

图卷积网络将数据视为无向图, 并对图的边和顶点应用卷积滤波器。这些方法已被证明可以有效地从 3D 点云数据中提取特征, 并已应用于三维表面缺陷检测任务^[130]。

在金属表面缺陷检测领域, 常用的语义分割数据集包括 MT (Magnetic tile defect datasets)^[131] 和 RSDDs (Rail surface discrete defects datasets)。MT 数据集包含不同光照条件下采集的共 5 类磁瓦表面图像, 每类提供大量正常表面图像和数十张带有像素级标注的缺陷图像。RSDDs 数据集则提供了 195 张包含像素级标注的钢轨表面缺陷图像。

3.5 异常检测

在实际工业生产场景中, 大量数据是缺乏标注的, 随着生产工艺的变化, 未定义过的缺陷类型也可能出现, 这就给表面缺陷检测带来了极大的挑战。基于无监督或弱监督学习的图像异常检测技术是解决这个问题的有效手段。传统异常检测方法通常会训练一个模型来拟合正常图像, 然后在检测阶段根据待检样本与模型之间的差异来进行异常检测。基于深度学习的方法尝试重构一个正常样本, 并根据

重构误差来检测可能存在的表面缺陷。

3.5.1 传统方法

传统方法主要包括基于统计的方法、基于图像低秩分解的方法和基于稀疏编码重构的方法。基于统计的方法通过学习一个模型来描述正常样本中像素点或特征向量在特征空间中的分布, 对于远离该分布的样本区域视为异常样本。常用的统计模型包括基于域的方法和基于距离的方法: 基于域的方法试图围绕正常数据的分布定义一个边界, 以便将超出边界的节点标记为异常点。基于距离的方法通过度量学习来训练样本点与正常样本之间的距离, 并通过统计它们与正常样本的较大距离来检测异常。基于图像低秩分解的方法利用周期性背景纹理低秩性的特点, 将原始图像分解成为代表背景的低秩矩阵和代表异常区域的稀疏矩阵, 无需任何训练过程, 但是计算量较大, 因此速度相对较慢, 难以进行实时检测。基于稀疏编码重构的方法对正常样本进行拟合, 在训练阶段学习一个完备字典来存储有代表性的特征, 并通过线性组合字典中存储的元素来重构输入图像。而对有缺陷样本进行重构时, 由于字典仅包含正常样本的特征, 因此会产生较大的重构误差, 从而检测出表面缺陷。相对于前几种方法而言, 该方法不需要任何先验假设, 仅需要足够的样本就可以拟合正常样本的特征表示, 使该方法在应用场景中受到更小的限制, 对于周期性纹理和随机纹理背景的样本均可检测, 在工业领域有许多成功应用。

基于传统方法的表面缺陷检测算法在很大程度上基于手动设计的特征, 最常见的方法是基于统计和过滤器的, 但设计特征的过程非常费时费力, 并且可能需要对每个新产品都要重复以上工作, 因此传统方法现在已经很少使用。但是这些传统方法的理论基础和设计思路, 对于我们设计基于深度学习的表面缺陷检测方法有着重大的指导意义。

3.5.2 基于图像重建的网络

图像重建模型的核心思想与传统机器学习中稀疏编码重构的方法类似。首先对正常样本的图像进行编解码, 并以重构输入为目标进行训练, 以此学习正常样本的分布。然后在检测阶段, 通过分析重构损失来进行缺陷检测。图像重构网络的训练方法主要包括变分自编码器 (Variational auto encoder, VAE)、GAN 和标准化流 (Normalizing flows)。

自编码器是一种编码器-解码器 (Encoder-decoder) 结构, 只需要正常样本就可以训练模型, 并通过计算重构图像与输入图像的差异来训练自编码器, 从而解决了工业生产场景下缺陷样本难以获取、缺陷形态不定、未知缺陷无法提前定义等问题。变

分自编码器是在自编码机的基础上让图像编码的潜在向量服从高斯分布从而实现图像的生成,优化了数据对数似然的下界 (Evidence lower bound, ELBO), 如图 14(a). 然而当训练样本的分布更加多元化时, 自编码机会表现出较强的特征表达能力, 从而对潜在的异常样本产生适应性, 尽管异常样本中的部分形状在正常样本中从未出现过, 但是网络仍然能够重构出与输入接近的图像, 导致重构差异不大, 使异常样本被误判为正常样本. 为了解决这个问题, Tian 等^[132] 尝试利用隐变量来存储正常样本的典型特征, 从而避免网络重构出包含缺陷的图像, 如图 14(c).

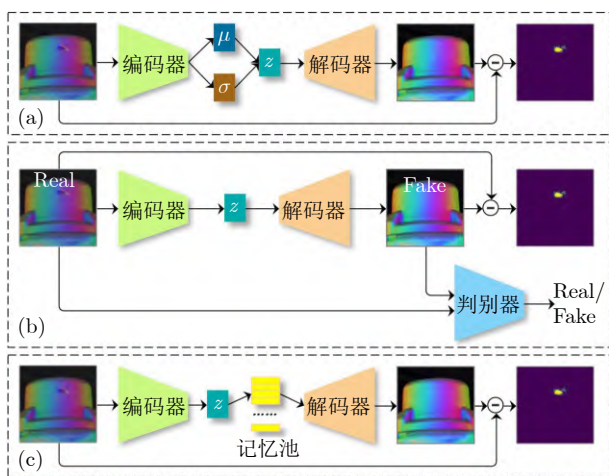


图 14 基于图像重建的缺陷检测 ((a) 变分自编码器; (b) GAN 结合 AE; (c) 基于记忆池的模型)

Fig. 14 Defect detection based on image reconstruction ((a) VAE; (b) GAN associated with Auto-Encoder; (c) Memory based model)

Mei 等^[133] 利用高斯金字塔混合自编码器 (Multi-scale-convolutional denoising autoencoder, MS-CDAE) 来进行纹理图像的异常区域定位, 该方法将样本分割成许多小图片分别进行重构, 并通过融合多个尺度的自编码器有效提高了异常检测的定位精度. ITAE (Inverse-transform autoencoder)^[134] 通过一组选定的变换 (例如随机灰度和旋转) 以从图像中删除特定目标信息, 然后仅用正常样本训练自动编码器来重构原始图像的目标信息, 并通过重构误差来区别正常样本和异常样本, 并在 MVTec AD 上验证了方法的有效性.

除了重构误差, 另一些方法利用图像重构中的重构损失梯度信息来进行表面缺陷检测. 重构损失梯度信息通常用于在反向传播时通过梯度下降进行模型权值的优化, 因此梯度实际上代表着朝向正常样本的特征表达的优化方向, 从而为重构损失提供

更加完备的信息. Zimmerer 等^[135] 首次将梯度信息引入异常检测任务中, 通过计算 VAE 中 ELBO 相对于输入图像的梯度来进行异常检测. Kwon 等^[136] 利用余弦相似度来计算正常样本梯度向量之间的角度, 以此来构建正常样本梯度向量的方向一致性约束. 在测试阶段出现不满足该一致性约束的梯度向量时, 则认为该图像为异常图像.

GAN 可以生成比自编码器方法更逼真的图像^[137]. 生成对抗网络一般由一个生成器和一个判别器组成. 生成器的作用是通过学习训练集的特征, 将随机噪声分布尽量拟合成训练数据的分布, 从而生成类似于训练数据分布的数据. 而判别器则负责区分输入的数据是真实的还是生成器生成的假数据, 并反馈给生成器. 两个网络交替训练, 能力同时提高, 直到生成器生成的数据可以以假乱真, 并与判别器的能力达到均衡.

另一类方法将自编码器与 GAN 方法结合, 如图 14(b) 所示, 在自编码机的基础上增加一个判别器, 通过判别器指导自编码机的对抗训练来提高重构图像的质量. 文献 [138] 在深度自编码机的基础上增加了对抗损失, 来实现无监督异常检测. Schlegl 等^[139] 直接用 GAN 中训练好的生成器替换了原始 AE 中的解码器, 以一种更为直接的方式利用 GAN 强大的图像生成能力来进行图像重构.

标准化流 (Normalizing flow) 的主要思想是通过一系列可逆变换将简单的基础分布转换为更复杂的分布, 可用于生成模型、强化学习、变分推断等领域, 原理如图 15. 但是传统的流模型忽略了整体与局部的关系, 无法将样本特征映射到一个明确的分布, 同时计算复杂度高, 推理速度有限, 难以用于工业场景实时检测. 文献 [140] 提出了一种全卷积跨尺度标准化流 (Cross-scale-flow, CS-flow), 通过联合处理不同尺度的多个特征图, 用标准化流为样本

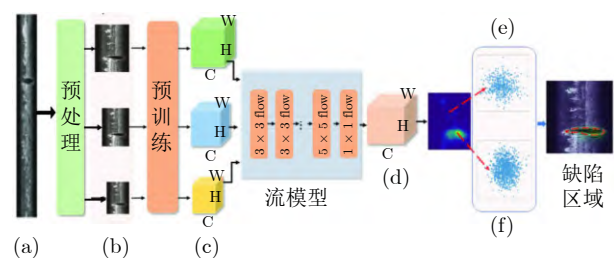


图 15 基于标准化流的缺陷检测 ((a) 原始图像; (b) 多尺度输入; (c) 图像特征分布; (d) 简单分布; (e) 标准分布; (f) 异常分布)

Fig. 15 Defect detection based on normalizing flow ((a) Origin image; (b) Multiscale input; (c) Feature distribution; (d) Simple distribution; (e) Normalized distribution; (f) Anomalous distribution)

输出有意义的概率分布, 实现了有效的缺陷检测, 在 MVTEC AD 数据的 15 个类别中的 4 类实现了 100% 的 AUROC 性能. FastFlow^[141] 将输入特征映射到一个特定分布, 并维持了特征的二维空间位置关系, 实现了端到端的图片级实时推理.

基于图像重构的方法通常可以直接实现异常区域的定位, 但难以实现缺陷分类. 这一类方法仍然有许多值得研究的内容, 例如正常图像区域的重构误差问题. 由于沙漏状的结构, 重构过程中容易丢失图像细节, 导致重构前后在正常样本区域可能出现较大差异, 这一缺点在结构复杂的金属结构表面尤其突出.

3.5.3 最新的方法

除了基于图像重构的方法, 还有一类基于切片 (Patch) 特征的方法用于异常检测. Yi 等^[142] 基于支持向量数据描述法 (SVDD)^[143] 提出了 Patch SVDD, 其主要思想是将图像分割成很多块 (Patches), 将其特征投影到一个较小的超球面上, 并以所有正常样本为中心规定一个半径 r , 将所有落在半径 r 以外的离群点视为异常样本, 在 MVTEC AD 数据集上的异常检测和分割任务上实现了 9.8% 和 7% 的性能提升. 文献 [144] 提出了一种基于记忆的端到端分割网络 MemSeg (Memory-based segmentation network). 该方法通过引入合成缺陷样本和内存样本辅助网络训练, 并用记忆池存储正常样本的一般模式, 从而在推理阶段直接定位图像的异常区域, 并且具有 31.2 FPS 的推理速度, 在工业金属表面缺陷分割领域有显著优势. PatchCore^[145] 提出了一种基于最具代表性特征记忆库的方法, 该方法结合了基于 ImageNet 预训练的图像嵌入表示和离群点检测模型, 如图 16(b) 所示. 该方法在 MVTEC AD 基准测试中实现了高达 99.6% 的 AUROC 分数, 错误率比当时最好的方法降低了一半. 基于预训练网络的方法通过非参数建模来估计正常样本的编码分布, 但是这些方法忽略了切片之间的位置和邻域信息. 为了解决这个问题, 文献 [146] 提出了 PNI (Position and neighborhood information) 模型, 利用条件概率估计领域特征在给定正常分布下的表示, 并用多层感知机对其进行建模. 同时模型引入一个基于随机合成异常图像进行训练的网络来提高异常检测的准确性, 成为该领域最先进的方法. 另一类基于教师-学生 (Teacher-student) 的方法^[147] 通过训练一个学生网络来预测正常样本的特征, 当学生网络预测的特征与教师网络不同时即可检测出图像中的异常, 如图 16(a) 所示. EfficientAD^[148] 提出了一个训练损失函数来阻止学生网络从教师网络中学习

到正常样本以外的特征以及正常特征的无效组合, 以高达 600 FPS 的检测速度在 MVTEC AD 数据集上取得了 98.8% 的准确率, 成为了该领域速度最快的异常检测方法. 表 2 中展示了这些方法在 MVTEC AD 数据集上的性能对比.

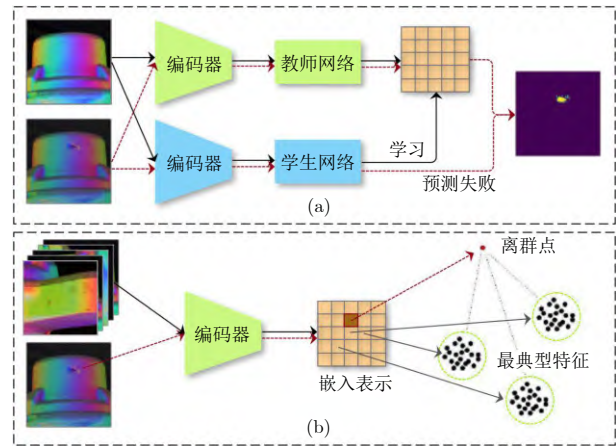


图 16 (a) 基于教师-学生网络的方法;
(b) 基于最典型嵌入表示的方法

Fig. 16 (a) Method based on teacher-student network;
(b) Method based on the most typical embedding representation

表 2 异常检测方法对比

Table 2 Comparison of abnormal detection

Method	Detection AUROC	Segmentation AUROC	FPS
PatchCore Large ^[145]	99.6	98.2	5.9
PNI ^[146]	99.5	99.0	—
MemSeg ^[144]	99.5	99.6	31.3
Fastflow ^[141]	99.4	98.5	21.8
EfficientAD-M ^[148]	99.1	96.9	269.0
EfficientAD-S ^[148]	98.8	96.8	614.0
CS-Flow ^[140]	98.7	—	—
Patch SVDD ^[142]	92.1	95.7	2.1
VAE-Grad ^[136]	89.2	—	—

综上, 基于不同工业应用场景的检测需求和任务目标不同, 不同类的方法各有其优缺点以及适用条件限制, 具体如表 3 所示.

4 关键问题和研究趋势

4.1 光学成像方案

基于光学成像的自动表面缺陷检测技术目前在工业制造领域正发挥着越来越重要的作用, 针对金属表面图像的缺陷检测算法研究也受到了学界和产业界的广泛关注. 然而以光学成像技术为基础的 AOI 技术不仅是一个算法问题, 同时也是一个复杂的

表 3 缺陷检测方法对比
Table 3 Comparison of defect detection methods

方法	基本原理	应用场景	优缺点
模板匹配	比较模板与待检样本的差异来判断是否存在缺陷	产品高度一致的金属精密加工制成品, 例如手机外壳、汽车零件等	方法简单有效, 但需要提取制作模板, 仅适用于一致性强的产品
分类网络	直接用 CNN 提取特征, 通过 Softmax 或距离度量来预测类别	公差较大、尺寸较小的金属制品, 例如螺母、金属盖等零件	结构简单, 是其他网络的基础, 准确率依赖缺陷样本数量, 难以定位缺陷位置
目标检测	对每个提议候选框或者每个网格进行密集预测, 从背景中找出所有目标的分类和位置	适用于绝大多数缺陷类别可事先定义的工业场景	速度快, 适用范围广, 但网络结构复杂, 依赖大量缺陷样本进行训练
语义分割	通过卷积提取高阶语义特征, 然后通过上采样输出像素级的缺陷边界划分	大面积金属板、带制品, 缺陷具有成片连续区域、形态不定的场景	可以进行像素级缺陷分割, 但是依赖大量像素级标注数据, 标注成本很高
异常检测	通过自编码器、GAN、标准流等生成模型学习正常样本的表达方式, 根据重建误差、梯度或分布差异来进行缺陷检测	缺乏缺陷样本, 只有正常样本可以用于训练的场景	无需缺陷样本和标注, 可以检测未事先定义的缺陷类别, 但准确率尚达不到有监督学习的效果

工程问题, 光学成像技术对于细微特征的获取也至关重要. 精心设计的成像方案能够有效降低检测算法的实现难度, 如何针对具体的应用场景选择合适的光学成像方案, 在硬件成本、算法精度和检测速度之间取得最佳的平衡, 仍然是一个亟待解决的问题.

同时随着产业智能化升级和精密制造技术的不断发展, AOI 技术对于检测精度、速度、分辨率的要求越来越高, 直接利用光源、相机、计算机等通用工业设备搭建的简单二维光学检测设备已经无法满足越来越多的细分场景的检测需求. 针对各种具体的工业生产场景的特点, 不断发展特殊光学成像技术也是十分重要的研究方向. 基于三维视觉的成像技术, 能够突破二维灰度图像的限制, 获取金属表面的法线方向、深度等信息, 对于具有纵向深度特征的微小缺陷取得了更好的检测效果, 具有良好的应用前景. 但目前三维成像技术更多的用于产品外形几何测量、定位、姿态估计等场景, 针对三维图像数据的缺陷检测方法研究较少, 很多方法仍需将图像数据从三维域转换到二维域才能处理, 在实际项目应用中的效果仍有待验证.

4.2 金属表面缺陷数据集

在不同场景下需要采用不同的光学成像方案, 不仅增加了技术的复杂度, 还带来了一个新的问题, 即工业金属表面检测领域难以构建一个大而统一的数据集. 与 ImageNet、PASCAL-VOC 和 COCO 等拥有数百万图片的通用计算机视觉数据集相比, 工业金属表面检测领域数据集例如 NEU-DET 和 MVTec AD 往往仅有数百至数千个样本, 另外还有许多金属表面缺陷数据集没有公开.

在很多视觉任务中, 通过在一个较大的数据集上进行预训练, 以获得一个具有良好特征表达能力的模型, 然后在较小的数据集上进行迁移学习 (Tran-

sfer learning) 和微调 (Fine-tuning) 往往能获得很好的效果. 许多研究表明, 预训练可以将模型从旧领域 (Domain) 学习到的知识迁移到新领域, 从而解决标注数据不足、训练代价过高的问题, 同时性能也显著优于直接训练的模型. 然而由于成像方式各不相同, 在不同类型的金属表面缺陷数据集之间几乎没有相似性, 因此预训练和迁移学习难以在这个领域获得很好的效果. 这也使得很多具体应用场景下的金属表面检测面临着冷启动的问题. 虽然在实施层面需要大量研究者的付出, 同时也面临着许多商业因素的阻力, 但构建一个公开的大型金属表面缺陷数据集将是推动该领域研究的有效手段.

4.3 样本不均衡问题

在公开数据集难以获取的同时, 研究者构建一个真实的金属表面缺陷数据集时, 还要面临着样本类别严重不均衡的问题. 在通常的机器学习任务中, 理想状态下多分类任务需要每类样本数量巨大且均衡. 但在大多数工业生产场景中仅能获得大量的正常样本, 缺陷样本的统计分布则会随着生产工艺的改善而不断变化. 未事先定义的异常样本或者在以往生产中极少出现的缺陷样本, 使得传统的基于监督学习的缺陷分类或目标检测算法难以满足实际应用所需的性能要求.

通常在监督学习中可以使用数据增广、合成和生成等方式来解决样本不均衡问题. 常规方法包括空间变换、色彩相移、剪切拼接、叠加等方式. 另一些研究者用 GAN 生成逼真的表面缺陷样本的同时, 可以获得一个强大的鉴别器用来进行缺陷检测. 近年来在生成领域大获成功的扩散模型 (Diffusion model) 则为这一类研究方法构建了新的范式.

单类别分类 (One-class classification, OCC) 是一类仅依靠正样本即可训练的二分类算法, 然而

这一类方法不仅可能遗漏人眼会明显认为是异常的缺陷,同时也存在异常的定义不够清晰、微小缺陷与正常样本的界限不够清晰的问题,在工业实际应用中存在较大的风险.小样本学习(Few-shot learning)则研究仅用极少量的标注样本来识别一个新类别的缺陷,已经有许多研究者进行了初步尝试,并取得了一定的效果.如何提高弱监督或半监督学习模型的可靠性仍需要进一步探索.相比之下,人类可以仅从少量样本中概括出一个新类别的明显特征,是人类智慧明显区别于人工智能的特点,同时也是机器学习中最具挑战性和最有意义的研究方向.

5 总结与展望

随着计算机视觉技术的高速发展,各种二维、三维成像技术的硬件成本不断降低,工业金属表面自动缺陷检测技术在智能制造领域必将发挥越来越广泛的作用.同时视觉检测系统要作为工业信息系统中不可缺少的设备,打通机联网数据,形成数据闭环上的关键节点,通过实时信息反馈,指导生产过程管理进行智能决策.

工业生产不同于实验室研究,必须综合考虑实施条件、硬件成本、算法精度、人工参与度低等要求,选择合适的光学成像技术、图像处理技术和缺陷检测算法来达到理想的效果.光学成像技术可分为二维、三维两类,其中二维成像技术已经广泛用于产品外观、尺寸测量、定位和分类;而三维成像技术对具有表面微观凹凸纹理特征的缺陷具有更好的检测效果,但三维图像缺陷检测技术尚不成熟,其中基于光度立体法的三维表面缺陷检测技术具有明显的成本优势,并且可以兼容目前的二维缺陷检测算法,有着良好的市场前景和研究价值.在获取图像信息以后,还需通过图像处理技术进行预处理,获取增强图像、缺陷特征和感兴趣区域,以便提高缺陷检测算法的计算效率和稳定性.最后再根据检测任务的目标,选择合适算法进行样本分类、缺陷定位、缺陷区域分割和异常检测.

随着产业智能化和精密制造的不断发展,通用工业光源、相机、检测算法的简单组合已经无法满足复杂多样的生产场景需求,必须根据检测目标的光学性质、表面纹理特点、缺陷的统计分布,并充分考虑生产环境的限制,选择针对性的成像方法和图像处理技术,并不断发展新的检测算法,这也是本文详细综述的重点内容.

在统计质量控制(Statistical quality control, SQC)体系中,表面缺陷检测方法仅仅是手段和工具,自动表面缺陷检测系统的最终目的是运用质量数据使质量控制可量化和科学化,保证所有工序产

出尽可能接近期望值,提高生产过程的稳定性.当缺陷数量较少时,以统计推断为基础的线性分类器效果往往不比深度学习算法效果差,同时可解释性和可信度更强,能够更好地指导我们建立机理模型,进行生产工艺优化.但当样本量增加、特征维度增大时,特征选择和相关性分析变得极为困难,而深度学习则尝试自动从数据中学习高级特征,在模型设计过程中尽量避免引入强假设,这毫无疑问会降低模型的可解释性.因此在实际工程应用中,传统方法和深度学习方法都有很大的研究价值,只有综合多种方法,在数据标注成本、算法可信度和预测精度之间取得平衡,才能满足实际工业生产需求.

References

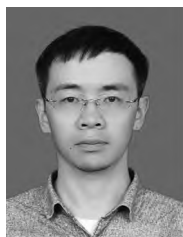
- 1 Neogi N, Mohanta D K, Dutta P K. Review of vision-based steel surface inspection systems. *EURASIP Journal on Image and Video Processing*, 2014, **2014**(1): Article No. 50
- 2 Lu Rong-Sheng, Wu Ang, Zhang Teng-Da, Wang Yong-Hong. Review on automated optical (visual) inspection and its applications in defect detection. *Acta Optica Sinica*, 2018, **38**(8): Article No. 0815002
(卢荣胜, 吴昂, 张腾达, 王永红. 自动光学(视觉)检测技术及其在缺陷检测中的应用综述. *光学学报*, 2018, **38**(8): Article No. 0815002)
- 3 Lv Cheng-Kan, Shen Fei, Zhang Zheng-Tao, Zhang Feng. Review of image anomaly detection. *Acta Automatica Sinica*, 2022, **48**(6): 1402-1428
(吕承侃, 沈飞, 张正涛, 张峰. 图像异常检测研究现状综述. *自动化学报*, 2022, **48**(6): 1402-1428)
- 4 Li Wei-Chuang, Yin Bai-Qiang. Research progress of automated visual surface defect detection for industrial metal planar materials. *Journal of Electronic Measurement and Instrumentation*, 2021, **35**(6): 1-16
(李维创, 尹柏强. 工业金属板带材表面缺陷自动视觉检测研究进展. *电子测量与仪器学报*, 2021, **35**(6): 1-16)
- 5 Tao Xian, Hou Wei, Xu De. A survey of surface defect detection methods based on deep learning. *Acta Automatica Sinica*, 2021, **47**(5): 1017-1034
(陶显, 侯伟, 徐德. 基于深度学习的表面缺陷检测方法综述. *自动化学报*, 2021, **47**(5): 1017-1034)
- 6 Luo Dong-Liang, Cai Yu-Xuan, Yang Zi-Hao, Zhang Zhe-Yan, Zhou Yu, Bai Xiang. Survey on industrial defect detection with deep learning. *Scientia Sinica Informationis*, 2022, **52**(6): 1002-1039
(罗东亮, 蔡雨萱, 杨子豪, 章哲彦, 周瑜, 白翔. 工业缺陷检测深度学习方法综述. *中国科学: 信息科学*, 2022, **52**(6): 1002-1039)
- 7 Khan S, Naseer M, Hayat M, Zamir S W, Khan F S, Shah M. Transformers in vision: A survey. *ACM Computing Surveys*, 2022, **54**(10s): Article No. 200
- 8 Huang Y C, Hung K C, Liu C C, Chuang T H, Chiou S J. Customized convolutional neural networks technology for machined product inspection. *Applied Sciences*, 2022, **12**(6): Article No. 3014
- 9 Liu S, Wang Q, Luo Y. A review of applications of visual inspection technology based on image processing in the railway industry. *Transportation Safety and Environment*, 2019, **1**(3): 185-204
- 10 Lu R S, Forrest A K. 3D surface topography from the specular lobe of scattered light. *Optics and Lasers in Engineering*, 2007, **45**(10): 1018-1027
- 11 Smith B. Geometrical shadowing of a random rough surface. *IEEE Transactions on Antennas and Propagation*, 1967, **15**(5): 668-671
- 12 Garcia-Lamont F, Cervantes J, López A, Rodríguez L. Seg-

- mentation of images by color features: A survey. *Neurocomputing*, 2018, **292**: 1–27
- 13 Foucher B. Infrared machine vision: A new contender. In: Proceedings of the SPIE 3700, Thermosense XXI. Orlando, USA: SPIE, 1999. 210–213
 - 14 Nayar S K, Ikeuchi K, Kanade T. Surface reflection: Physical and geometrical perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1991, **13**(7): 611–634
 - 15 Porter T F, Sylvester R A, Bouyoucas T W, Kolesar M P. Automatic strip surface defect detection system. *Iron and Steel Engineer*, 1988, **65**(12): 17–20
 - 16 Tian G Y, Lu R S, Gledhill D. Surface measurement using active vision and light scattering. *Optics and Lasers in Engineering*, 2007, **45**(1): 131–139
 - 17 Lu Rong-Sheng. State of the art of automated optical inspection. *Infrared and Laser Engineering*, 2008, **37**(S1): 120–123 (卢荣胜. 自动光学检测技术的发展现状. 红外与激光工程, 2008, **37**(S1): 120–123)
 - 18 Rinn R, Thompson S A, Foehr R, Luecking F, Torre J. Parsytec HTS-2: Defect detection and classification through software vs. dedicated hardware. In: Proceedings of the Electronic Imaging '99. San Jose, CA, United States: SPIE, 1999. 110–121
 - 19 Li Song, Zhou Ya-Tong, Zhang Zhong-Wei, Chi Yue, Han Chun-Ying. Surface defect detection of stamping parts based on double light pattern matching. *Forging & Stamping Technology*, 2018, **43**(11): 137–145 (李松, 周亚同, 张忠伟, 池越, 韩春颖. 基于双打光模板匹配的冲压件表面缺陷检测. 锻压技术, 2018, **43**(11): 137–145)
 - 20 Landstrom A, Thurley M J. Morphology-based crack detection for steel slabs. *IEEE Journal of Selected Topics in Signal Processing*, 2012, **6**(7): 866–875
 - 21 Woodham R J. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 1980, **19**(1): Article No. 191139
 - 22 Berthold K P H. Shape From Shading: A Method for Obtaining the Shape of a Smooth Opaque Object from One View, Technical Report AITR-232, Artificial Intelligence Laboratory, MIT, USA, 1970.
 - 23 Wang L, Xu K, Zhou P. Online detection technique of 3D defects for steel strips based on photometric stereo. In: Proceedings of the 8th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA). Macao, China: IEEE, 2016. 428–432
 - 24 Delpy D T, Cope M, van der Zee P, Arridge S, Wray S, Wyatt J. Estimation of optical pathlength through tissue from direct time of flight measurement. *Physics in Medicine & Biology*, 1988, **33**(12): 1433–1442
 - 25 Bologna F, Tannous M, Romano D, Stefanini C. Automatic welding imperfections detection in a smart factory via 2-D laser scanner. *Journal of Manufacturing Processes*, 2022, **73**: 948–960
 - 26 Huang C, Wang G L, Song H, Li R S, Zhang H O. Rapid surface defects detection in wire and arc additive manufacturing based on laser profilometer. *Measurement*, 2022, **189**: Article No. 110503
 - 27 Li J L, Liu T, Wang X F. Advanced pavement distress recognition and 3D reconstruction by using GA-DenseNet and binocular stereo vision. *Measurement*, 2022, **201**: Article No. 111760
 - 28 Li B Z, Xu Z J, Gao F, Cao Y L, Dong Q C. 3D reconstruction of high reflective welding surface based on binocular structured light stereo vision. *Machines*, 2022, **10**(2): Article No. 159
 - 29 Zhou P, Xu K, Wang D D. Rail profile measurement based on line-structured light vision. *IEEE Access*, 2018, **6**: 16423–16431
 - 30 Guillo L, Jiang X R, Lafruit G, Guillemot C. Light Field Video Dataset Captured by a R8 Raytrix Camera (With Disparity Maps), Technical Report hal-01804578, International Organisation for Standardisation, San Diego, USA, 2018.
 - 31 Saiz F A, Barandiaran I, Arbelaz A, Graña M. Photometric stereo-based defect detection system for steel components manufacturing using a deep segmentation network. *Sensors*, 2022, **22**(3): Article No. 882
 - 32 Wen X, Song K C, Huang L M, Niu M H, Yan Y H. Complex surface ROI detection for steel plate fusing the gray image and 3D depth information. *Optik*, 2019, **198**: Article No. 163313
 - 33 Niu M H, Song K C, Huang L M, Wang Q, Yan Y H, Meng Q G. Unsupervised saliency detection of rail surface defects using stereoscopic images. *IEEE Transactions on Industrial Informatics*, 2021, **17**(3): 2271–2281
 - 34 Ren Z H, Fang F Z, Yan N, Wu Y. State of the art in defect detection based on machine vision. *International Journal of Precision Engineering and Manufacturing-Green Technology*, 2022, **9**(2): 661–691
 - 35 Niblack W. *An Introduction to Digital Image Processing*. Birkeroed: Strandberg Publishing Company, 1985.
 - 36 Luisier F, Blu T, Unser M. A new SURE approach to image denoising: Interscale orthonormal wavelet thresholding. *IEEE Transactions on Image Processing*, 2007, **16**(3): 593–606
 - 37 Gu S H, Zhang L, Zuo W M, Feng X C. Weighted nuclear norm minimization with application to image denoising. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA: IEEE, 2014. 2862–2869
 - 38 Mohanaiah P, Sathyanarayana P, GuruKumar L. Image texture feature extraction using GLCM approach. *International Journal of Scientific and Research Publications*, 2013, **3**(5): 1–5
 - 39 Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, **24**(7): 971–987
 - 40 Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). San Diego, USA: IEEE, 2005. 886–893
 - 41 Lowe D G. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 2004, **60**(2): 91–110
 - 42 Lindeberg T. Scale invariant feature transform. *Scholarpedia*, 2012, **7**(5): Article No. 10491
 - 43 Oren M, Papageorgiou C, Sinha P, Osuna E, Poggio T. Pedestrian detection using wavelet templates. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Juan, USA: IEEE, 1997. 193–199
 - 44 Lienhart R, Maydt J. An extended set of Haar-like features for rapid object detection. In: Proceedings of the International Conference on Image Processing. Rochester, USA: IEEE, 2002.
 - 45 Wang X W, Ding X Q, Liu C S. Gabor filters-based feature extraction for character recognition. *Pattern Recognition*, 2005, **38**(3): 369–379
 - 46 Raheja J L, Kumar S, Chaudhary A. Fabric defect detection based on GLCM and Gabor filter: A comparison. *Optik*, 2013, **124**(23): 6469–6474
 - 47 Siew L H, Hodgson R M, Wood E J. Texture measures for carpet wear assessment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1988, **10**(1): 92–105
 - 48 Liu K, Wang H Y, Chen H Y, Qu E Q, Tian Y, Sun H X. Steel surface defect detection using a new Haar-Weibull-Variance model in unsupervised manner. *IEEE Transactions on Instrumentation and Measurement*, 2017, **66**(10): 2585–2596
 - 49 Pernkopf F, O'Leary P. Image acquisition techniques for automatic visual inspection of metallic surfaces. *NDT & E International*, 2003, **36**(8): 609–617
 - 50 Kaggle. Severstal: Steel defect detection [Online], available: <https://www.kaggle.com/c/severstal-steel-defect-detection>, March 19, 2023
 - 51 Qi S, Yang J, Zhong Z. A review on industrial surface defect detection based on deep learning technology. In: Proceedings of the 3rd International Conference on Machine Learning and Machine Intelligence. Hangzhou, China: ACM, 2020. 24–30
 - 52 Zheng Kai, Li Jian-Sheng. An overview of image semantic segmentation based on deep learning. *Geomatics & Spatial In-*

- formation Technology, 2020, **43**(10): 119–125
(郑凯, 李建胜. 基于深度神经网络的图像语义分割综述. 测绘与空间地理信息, 2020, **43**(10): 119–125)
- 53 Belongie S, Malik J, Puzicha J. Shape context: A new descriptor for shape matching and object recognition. In: Proceedings of the 13th International Conference on Neural Information Processing Systems. Denver, USA: MIT Press, 2000. 798–804
 - 54 Qi C R, Yi L, Su H, Guibas L J. PointNet: Deep learning on point sets for 3D classification and segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE, 2017. 652–660
 - 55 Qi C R, Yi L, Su H, Guibas L J. PointNet++: Deep hierarchical feature learning on point sets in a metric space. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc., 2017. 5105–5114
 - 56 Ma X, Qin C, You H X, Ran H X, Fu Y. Rethinking network design and local geometry in point cloud: A simple residual MLP framework. In: Proceedings of the 10th International Conference on Learning Representations (ICLR). OpenReview.net, 2022.
 - 57 Qian G C, Li Y C, Peng H W, Mai J J, Al Kader Hammoud H A, Elhoseiny M, et al. PointNeXt: Revisiting PointNet++ with improved training and scaling strategies. In: Proceedings of the 35th Conference on Neural Information Processing Systems (NeurIPS). 2022. 23192–23204
 - 58 Jain P, Tyagi V. LAPB: Locally adaptive patch-based wavelet domain edge-preserving image denoising. *Information Sciences*, 2015, **294**: 164–181
 - 59 Li A B, Ma H W, Xu S H. Three-dimensional morphology and watershed-algorithm-based method for pitting corrosion evaluation. *Buildings*, 2022, **12**(11): Article No. 1908
 - 60 Guo Hao-Ran, Shao Wei, Zhou A-Wei, Yang Yu-Xiang, Liu Kai-Bin. Novel defect recognition method based on adaptive global threshold for highlight metal surface. *Chinese Journal of Scientific Instrument*, 2017, **38**(11): 2797–2804
(郭皓然, 邵伟, 周阿维, 杨宇祥, 刘凯斌. 全局阈值自适应的高亮金属表面缺陷识别新方法. 仪器仪表学报, 2017, **38**(11): 2797–2804)
 - 61 Wei Ai-Dong. Thermal image defect extraction based on two image segmentation algorithms. *Electronic Test*, 2020, **34**(7): 56–59
(魏爱东. 基于脉冲涡流热成像的钢板缺陷检测研究. 电子测试, 2020, **34**(7): 56–59)
 - 62 Prabha P A, Bharathwaj M, Dinesh K, Prashath G H. Defect detection of industrial products using image segmentation and saliency. *Journal of Physics: Conference Series*, 2021, **1916**: Article No. 012165
 - 63 Zhang X W, Ding Y Q, Lv Y Y, Shi A Y, Liang R Y. A vision inspection system for the surface defects of strongly reflected metal based on multi-class SVM. *Expert Systems With Applications*, 2011, **38**(5): 5930–5939
 - 64 Ghosh D, Kaabouch N. A survey on image mosaicing techniques. *Journal of Visual Communication and Image Representation*, 2016, **34**: 1–11
 - 65 Lu R S, Shi Y Q, Li Q, Yu Q P. AOI techniques for surface defect inspection. In: Proceedings of the Applied Mechanics and Materials. Trans Tech Publications Ltd, 2010. 297–302
 - 66 Kong Q, Wu Z, Song Y. Online detection of external thread surface defects based on an improved template matching algorithm. *Measurement*, 2022, 195: Article No. 111087
 - 67 Pang G, Shen C, Cao L, Hengel V A D. Deep learning for anomaly detection: A review. *ACM Computing Surveys*, 2021, **54**(2): 1–38
 - 68 Konovalenko I, Maruschak P, Brevus V, Prentkovskis O. Recognition of scratches and abrasions on metal surfaces using a classifier based on a convolutional neural network. *Metals*, 2021, **11**(4): Article No. 549
 - 69 Jiang Q S, Tan D P, Li Y B, Ji S M, Cai C P, Zheng Q M. Object detection and classification of metal polishing shaft surface defects based on convolutional neural network deep learning. *Applied Sciences*, 2019, **10**(1): Article No. 87
 - 70 Tabernik D, Šela S, Skvarč J, Skočaj D. Segmentation-based deep-learning approach for surface-defect detection. *Journal of Intelligent Manufacturing*, 2019, **31**(3): 759–776
 - 71 Yang Jie. Unsupervised Visual Anomaly Detection and Recognition Based on Deep Learning [Ph.D. dissertation], University of Chinese Academy of Sciences, China, 2021.
(杨洁. 基于深度学习的无监督图像异常模式检测与识别研究 [博士学位论文], 中国科学院大学, 中国, 2021.)
 - 72 Chetverikov D, Khenokh Y. Matching for shape defect detection. In: Proceedings of the 8th International Conference on Computer Analysis of Images and Patterns (CAIP). Ljubljana, Slovenia: Springer, 1999. 367–374
 - 73 Wang L, Pavlidis T. Direct gray-scale extraction of features for character recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1993, **15**(10): 1053–1067
 - 74 Chui H, Rangarajan A. A new point matching algorithm for non-rigid registration. *Computer Vision and Image Understanding*, 2003, **89**(2): 114–141
 - 75 Borgefors G. Hierarchical chamfer matching: A parametric edge matching algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1988, **10**(6): 849–865
 - 76 Zhang H J, Hu Q. Fast image matching based-on improved SURF algorithm. In: Proceedings of the International Conference on Electronics, Communications and Control (ICECC). Ningbo, China: IEEE, 2011. 1460–1463
 - 77 Crispin A J, Rankov V. Automated inspection of PCB components using a genetic algorithm template-matching approach. *The International Journal of Advanced Manufacturing Technology*, 2007, **35**(3): 293–300
 - 78 Hashemi N S, Aghdam R B, Ghiasi A S B, Fatemi P. Template matching advances and applications in image analysis. arXiv preprint arXiv: 1610.07231, 2016.
 - 79 Wang H Y, Zhang J W, Tian Y, Chen H Y, Sun H X, Liu K. A simple guidance template-based defect detection method for strip steel surfaces. *IEEE Transactions on Industrial Informatics*, 2019, **15**(5): 2798–2809
 - 80 Pernkopf F. Detection of surface defects on raw steel blocks using Bayesian network classifiers. *Pattern Analysis and Applications*, 2004, **7**(3): 333–342
 - 81 Aghdam S R, Amid E, Imani M F. A fast method of steel surface defect detection using decision trees applied to LBP based features. In: Proceedings of the 7th IEEE Conference on Industrial Electronics and Applications (ICIEA). Singapore: IEEE, 2012. 1447–1452
 - 82 Samy M P, Foong S, Soh G S, Yeo K S. Automatic optical & laser-based defect detection and classification in brick masonry walls. In: Proceedings of the IEEE Region 10 Conference (TENCON). Singapore: IEEE, 2016. 3521–3524
 - 83 Li X G, Zhu J, Shi H R, Cong Z J. Surface defect detection of seals based on K-means clustering algorithm and particle swarm optimization. *Scientific Programming*, 2021, **2021**: Article No. 3965247
 - 84 Yue B, Wang Y P, Min Y Z, Zhang Z H, Wang W R, Yong J. Rail surface defect recognition method based on AdaBoost multi-classifier combination. In: Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC). Lanzhou, China: IEEE, 2019. 391–396
 - 85 Soukup D, Huber-Mörk R. Convolutional neural networks for steel surface defect detection from photometric stereo images. In: Proceedings of the 10th International Symposium on Visual Computing (ISVC). Las Vegas, USA: Springer, 2014. 668–677
 - 86 Masci J, Meier U, Cireşan D, Schmidhuber J, Fricout G. Steel defect classification with max-pooling convolutional neural networks. In: Proceedings of the International Joint Conference on Neural Networks (IJCNN). Brisbane, Australia: IEEE, 2012. 1–6

- 87 Staar B, Lütjen M, Freitag M. Anomaly detection with convolutional neural networks for industrial surface inspection. *Procedia CIRP*, 2019, **79**: 484–489
- 88 Liao S C, Shao L. Graph sampling based deep metric learning for generalizable person re-identification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, USA: IEEE, 2022. 7349–7358
- 89 Alzu'bi A, Albalas F, Al-Hadhrani T, Younis L B, Bashayreh A. Masked face recognition using deep learning: A review. *Electronics*, 2021, **10**(21): Article No. 2666
- 90 Mordia R, Verma A K. Visual techniques for defects detection in steel products: A comparative study. *Engineering Failure Analysis*, 2022, **134**: Article No. 106047
- 91 Kim M S, Park T, Park P. Classification of steel surface defect using convolutional neural network with few images. In: Proceedings of the 12th Asian Control Conference (ASCC). Kitakyushu, Japan: IEEE, 2019. 1398–1401
- 92 Wu S L, Wu Y B, Cao D H, Zheng C Y. A fast button surface defect detection method based on Siamese network with imbalanced samples. *Multimedia Tools and Applications*, 2019, **78**(24): 34627–34648
- 93 Li X, Yang X, Ma Z, Xue J H. Deep metric learning for few-shot image classification: A selective review. arXiv preprint arXiv: 2105.08149, 2021.
- 94 Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA: IEEE, 2014. 580–587
- 95 He K M, Zhang X Y, Ren S Q, Sun J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, **37**(9): 1904–1916
- 96 Girshick R. Fast R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE, 2015. 1440–1448
- 97 Ren S Q, He K M, Girshick R, Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, **39**(6): 1137–1149
- 98 He K M, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017. 2980–2988
- 99 Cai Z W, Vasconcelos N. Cascade R-CNN: Delving into high quality object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 6154–6162
- 100 Guo F, Qian Y, Rizos D, Suo Z, Chen X B. Automatic rail surface defects inspection based on mask R-CNN. *Transportation Research Record: Journal of the Transportation Research Board*, 2021, **2675**(11): 655–668
- 101 Xu Y Y, Li D W, Xie Q, Wu Q Y, Wang J. Automatic defect detection and segmentation of tunnel surface using modified mask R-CNN. *Measurement*, 2021, **178**: Article No. 109316
- 102 Fang J T, Tan X Y, Wang Y H. ACRM: Attention cascade R-CNN with Mix-NMS for metallic surface defect detection. In: Proceedings of the 25th International Conference on Pattern Recognition (ICPR). Milan, Italy: IEEE, 2021. 423–430
- 103 Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection. In: Proceedings of the 29th IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE, 2016. 779–788
- 104 Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: Optimal speed and accuracy of object detection. arXiv preprint arXiv: 2004.10934, 2020.
- 105 Li M J, Wang H, Wan Z B. Surface defect detection of steel strips based on improved YOLOv4. *Computers and Electrical Engineering*, 2022, **102**: Article No. 108208
- 106 Usamentiaga R, Lema D G, Pedrayes O D, Garcia D F. Automated surface defect detection in metals: A comparative review of object detection and semantic segmentation using deep learning. *IEEE Transactions on Industry Applications*, 2022, **58**(3): 4203–4213
- 107 Xu Y H, Wang X J, Li S Y. Track surface defect detection based on EfficientDet. In: Proceedings of the 5th International Conference on Electrical Engineering and Information Technologies for Rail Transportation (EITRT) 2021: Rail Transportation Information Processing and Operational Management Technologies. Singapore: Springer, 2022. 56–66
- 108 Wang C Y, Bochkovskiy A, Liao H Y M. Scaled-YOLOv4: Scaling cross stage partial network. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, USA: IEEE, 2021. 13024–13033
- 109 He Y, Song K C, Meng Q G, Yan Y H. An end-to-end steel surface defect detection approach via fusing multiple hierarchical features. *IEEE Transactions on Instrumentation and Measurement*, 2020, **69**(4): 1493–1504
- 110 Anthony A, Ho E S L, Woo W L, Gao B. A review and benchmark on state-of-the-art steel defects detection [Online], available: <http://dx.doi.org/10.2139/ssrn.4121951>, March 19, 2023
- 111 Song G L, Liu Y, Wang X G. Revisiting the sibling head in object detector. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE, 2020. 11560–11569
- 112 Lin T Y, Goyal P, Girshick R, He K M, Dollár P. Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017. 2999–3007
- 113 Ma P F, Ma J, Wang X J, Yang L C, Wang N N. Deformable convolutional networks for multi-view 3D shape classification. *Electronics Letters*, 2018, **54**(24): 1373–1375
- 114 Law H, Deng J. CornerNet: Detecting objects as paired keypoints. In: Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich, Germany: Springer, 2018. 765–781
- 115 Ge Z, Liu S T, Wang F, Li Z M, Sun J. YOLOX: Exceeding YOLO series in 2021. arXiv preprint arXiv: 2107.08430, 2021.
- 116 Zhu B J, Wang J F, Jiang Z K, Zong F H, Liu S T, Li Z M, et al. AutoAssign: Differentiable label assignment for dense object detection. arXiv preprint arXiv: 2007.03496, 2020.
- 117 Lu H T, Fang M Y, Qiu Y, Xu W Q. An anchor-free defect detector for complex background based on pixelwise adaptive multiscale feature fusion. *IEEE Transactions on Instrumentation and Measurement*, 2023, **72**: Article No. 5002312
- 118 Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A, Zagoruyko S. End-to-end object detection with transformers. In: Proceedings of the 16th European Conference on Computer Vision (ECCV). Glasgow, UK: Springer, 2020. 213–229
- 119 Zhu X, Su W, Lu L, Li B, Wang X, Dai J. Deformable detr: Deformable transformers for end-to-end object detection. arXiv preprint arXiv: 2010.04159, 2020.
- 120 Misra I, Girdhar R, Joulin A. An end-to-end transformer model for 3D object detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, Canada: IEEE, 2021. 2886–2897
- 121 Lv X, Duan F, Jiang J, Fu X, Gan L. Deep metallic surface defect detection: The new benchmark and detection network. *Sensors*, 2020, **20**(6): Article No. 1562
- 122 Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, USA: IEEE, 2015. 3431–3440
- 123 Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. In: Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). Munich, Germany: Springer, 2015. 234–241
- 124 Sabet D N, Zarifi M R, Khoramdel J, Borhani Y, Najafi E. An automated visual defect segmentation for flat steel surface using deep neural networks. In: Proceedings of the 12th International

- tional Conference on Computer and Knowledge Engineering (ICCKE). Mashhad, Islamic Republic of Iran: IEEE, 2022. 423–427
- 125 Lin T Y, Dollar P, Girshick R, He K, Hariharan B, Belongie S. Feature pyramid networks for object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE, 2017. 2117–2125
- 126 Xie S N, Girshick R, Dollár P, Tu Z W, He K M. Aggregated residual transformations for deep neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE, 2017. 5987–5995
- 127 Koonce B. EfficientNet. *Convolutional Neural Networks With Swift for Tensorflow: Image Recognition and Dataset Categorization*. Berkeley: Apress, 2021. 109–123
- 128 Chen L C, Papandreou G, Kokkinos I, Murphy K, Yuille A L. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, **40**(4): 834–848
- 129 Zheng S, Jayasumana S, Romera-Paredes B, Vineet V, Su Z Z, Du D L, et al. Conditional random fields as recurrent neural networks. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE, 2015. 1529–1537
- 130 Ma L F, Li J. SD-GCN: Saliency-based dilated graph convolution network for pavement crack extraction from 3D point clouds. *International Journal of Applied Earth Observation and Geoinformation*, 2022, **111**: Article No. 102836
- 131 Huang Y B, Qiu C Y, Guo Y, Wang X N, Yuan K. Surface defect saliency of magnetic tile. In: Proceedings of the 14th International Conference on Automation Science and Engineering (CASE). Munich, Germany: IEEE, 2018. 612–617
- 132 Tian H, Li F. Autoencoder-based fabric defect detection with cross-patch similarity. In: Proceedings of the 16th International Conference on Machine Vision Applications (MVA). Tokyo, Japan: IEEE, 2019. 1–6
- 133 Mei S, Yang H, Yin Z P. An unsupervised-learning-based approach for automated defect inspection on textured surfaces. *IEEE Transactions on Instrumentation and Measurement*, 2018, **67**(6): 1266–1277
- 134 Huang C Q, Cao J K, Ye F, Li M S, Zhang Y, Lu C W. Inverse-transform Autoencoder for anomaly detection. arXiv preprint arXiv: 1911.10676, 2019.
- 135 Zimmerer D, Petersen J, Kohl S A A, Maier-Hein K H. A case for the score: Identifying image anomalies using variational autoencoder gradients. arXiv preprint arXiv: 1912.00003, 2019.
- 136 Kwon G, Prabhushankar M, Temel D, AlRegib G. Backpropagated gradient representations for anomaly detection. In: Proceedings of the 16th European Conference on Computer Vision (ECCV). Glasgow, UK: Springer, 2020. 206–226
- 137 Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial networks. *Communications of the ACM*, 2020, **63**(11): 139–144
- 138 Baur C, Denner S, Wiestler B, Navab N, Albarqouni S. Autoencoders for unsupervised anomaly segmentation in brain MR images: A comparative study. *Medical Image Analysis*, 2021, **69**: Article No. 101952
- 139 Schlegl T, Seebock P, Waldstein S M, Langs G, Schmidt-Erfurth U. f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks. *Medical Image Analysis*, 2019, **54**: 30–44
- 140 Rudolph M, Wehrbein T, Rosenhahn B, Wandt B. Fully convolutional cross-scale-flows for image-based defect detection. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Waikoloa, USA: IEEE, 2022. 1829–1838
- 141 Yu J W, Zheng Y, Wang X, Li W, Wu Y S, Zhao R, et al. FastFlow: Unsupervised anomaly detection and localization via 2D normalizing flows. arXiv preprint arXiv: 2111.07677, 2021.
- 142 Yi J H, Yoon S. Patch SVDD: Patch-level SVDD for anomaly detection and segmentation. In: Proceedings of the 15th Asian Conference on Computer Vision (ACCV). Kyoto, Japan: Springer, 2021. 375–390
- 143 Tax D M J, Duin R P W. Support vector data description. *Machine Learning*, 2004, **54**(1): 45–66
- 144 Yang M H, Wu P, Feng H. MemSeg: A semi-supervised method for image surface defect detection using differences and commonalities. *Engineering Applications of Artificial Intelligence*, 2023, **119**: Article No. 105835
- 145 Roth K, Pemula L, Zepeda J, Schölkopf B, Brox T, Gehler P. Towards total recall in industrial anomaly detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, USA: IEEE, 2022. 14298–14308
- 146 Bae J, Lee J H, Kim S. Image anomaly detection and localization with position and neighborhood information. arXiv preprint arXiv: 2211.12634, 2022.
- 147 Bergmann P, Fauser M, Sattlegger D, Steger C. Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE, 2020. 4182–4191
- 148 Batzner K, Heckler L, König R. EfficientAD: Accurate visual anomaly detection at millisecond-level latencies. arXiv preprint arXiv: 2303.14535, 2023.



伍麟 北京航空航天大学硕士研究生. 主要研究方向为计算机视觉, 目标检测和表面缺陷检测.

E-mail: zf2021349@buaa.edu.cn

(WU Lin Master student at the School of Software, Beihang University. His research interest covers

computer vision, object detection and surface defect detection.)



郝鸿宇 北京航空航天大学硕士研究生. 主要研究方向为计算机视觉, 神经网络和少样本学习.

E-mail: JoeyHao@buaa.edu.cn

(HAO Hong-Yu Master student at the School of Software, Beihang University. His research interest

covers computer vision, graph neural network and few-shot learning.)



宋友 北京航空航天大学教授. 主要研究方向为软件工程, 异常信号检测, 算法分析与设计. 本文通信作者.

E-mail: songyou@buaa.edu.cn

(SONG You Professor at the School of Software, Beihang University. His research interest covers soft-

ware engineering, anomaly signal detection, algorithm analysis and design. Corresponding author of this paper.)