



工业缺陷检测深度学习方法综述

罗东亮¹, 蔡雨萱¹, 杨子豪¹, 章哲彦¹, 周瑜^{1*}, 白翔^{2*}

1. 华中科技大学电子信息与通信学院, 武汉 430074

2. 华中科技大学人工智能与自动化学院, 武汉 430074

* 通信作者. E-mail: yuzhou@hust.edu.cn, xbai@hust.edu.cn

收稿日期: 2021-09-26; 修回日期: 2021-12-31; 接受日期: 2022-01-13; 网络出版日期: 2022-06-13

国家自然科学基金 (批准号: 62176098)、华中大前沿创新团队项目 (批准号: 2017QYTD08) 和湖北省自然科学基金资助创新团队项目 (批准号: 2019CFA022) 资助

摘要 基于深度学习的工业缺陷检测方法可以降低传统人工质检的成本, 提升检测的准确性与效率, 因而在智能制造中扮演重要角色, 并逐渐成为计算机视觉领域新兴的研究热点之一. 其被广泛地应用于无人质检、智能巡检、质量控制等各种生产与运维场景中. 本综述旨在对工业缺陷检测的任务定义、难点、挑战、主流方法、公共数据集及评价指标等进行全面归纳, 以帮助研究人员快速了解该领域. 具体而言, 本文首先介绍工业缺陷检测的背景与特点. 接着, 按照实际数据标注情况, 划分出缺陷模式已知、缺陷模式未知与少量缺陷标注 3 种研究任务设置, 并根据方法类型作进一步归纳与分析, 探讨了各方法的性能优劣与适用场景, 阐明了方法与实际应用需求的关联性. 此外, 本文还归纳了方法部署中的关键辅助技术, 总结了现有方法在实际产业落地中存在的局限性. 最后, 本文对该领域未来的发展趋势和潜在研究方向进行了展望.

关键词 缺陷检测, 异常检测, 计算机视觉, 工业视觉, 深度学习

1 引言

大到飞机机翼, 小到芯片晶粒, 工业制品在现代社会中无处不在. 工业缺陷检测, 旨在发现各种工业制品的外观瑕疵, 是保障产品质量、维持生产稳定的重要技术之一. 以往的缺陷检测需要人工筛查, 成本高、效率低, 难以覆盖大规模的质检需求. 近年来, 随着工业成像、计算机视觉和深度学习等领域的新技术层出不穷, 基于视觉的工业缺陷检测技术得到了长足的发展, 成为了针对产品外观质检的一种有效的解决方案, 引发了学术界和工业界的强烈关注. 工业缺陷检测不仅可以用于检测各种工业制品, 如金属、纺织物、半导体等, 而且具有优秀的检测精度与效率, 还能提供简便、安全的操作环境. 因此, 工业缺陷检测已成为智能制造领域重要的基础研究与技术之一, 并被广泛应用于无人质检、智能

引用格式: 罗东亮, 蔡雨萱, 杨子豪, 等. 工业缺陷检测深度学习方法综述. 中国科学: 信息科学, 2022, 52: 1002–1039, doi: 10.1360/SSI-2021-0336

Luo D L, Cai Y X, Yang Z H, et al. Survey on industrial defect detection with deep learning (in Chinese). Sci Sin Inform, 2022, 52: 1002–1039, doi: 10.1360/SSI-2021-0336

巡检、生产控制和异常溯源等场景。《中国制造 2025》行动纲领指出,建设制造强国任务艰巨而紧迫,需要加速推进信息化与工业化的深度融合,推进生产过程的智能化。因此,基于视觉的工业缺陷检测不仅具有非常重要的研究价值,同时也拥有广阔的应用前景。然而,相比于一般的目标检测任务,工业缺陷检测面临着诸多难点,如,缺陷样本匮乏、缺陷的可视性低、形状不规则、类型未知等,导致许多现有方法难以同时满足高精度、高速度的任务需求,因此在实现落地应用的道路上依然存在大量问题亟待解决。

由于工业缺陷可以视为工业产品的外观“异常”,因此也有部分工业缺陷检测方法采用了异常检测的思路。然而异常检测的定义与工业缺陷检测也有所区别。具体而言,异常检测的概念更加广泛与抽象,其中图像异常检测主要关注输入图像是否为异常实例,而工业缺陷检测更关注像素层面的检出任务。在像素层面上,异常与正常模式的差别更加细微,检测难度也大幅增加。因此直接使用异常检测方法难以满足工业缺陷检测的任务需求。

近年来,深度学习方法因其在处理背景复杂、缺陷微弱的工业图像时展现出卓越的性能优势,逐渐在该领域占据主导地位。鉴于此,本文对基于深度学习的工业缺陷检测方法进行综述,旨在帮助研究人员快速对该领域的任务设置、主流方法、评价体系等方面有系统性的了解。由于该任务具有较强的应用驱动性,本文按照实际情况中数据样本的标注与使用情况,划分出 3 种任务设置:缺陷模式已知、缺陷模式未知和少量缺陷标注。特别地,本文还归纳了促进方法落地的关键辅助技术,其有助于提升方法的实用性。本文关注各方法间的共性与不同,按照问题导向逐步剖析其发展脉络,并结合领域研究现状对未来发展趋势进行展望,希望帮助研究者们开拓思路。

目前国内外的综述^[1~5]大多探讨广义的异常检测领域的方法,试图涵盖图像、视频、表格和序列等各种数据形式。文献[2,3]对基于深度学习的异常检测方法进行了综合性的归纳与分析,但缺乏针对工业场景的探讨。文献[6,7]虽然以工业生产为背景来综述,但主要着眼于传统方法与系统控制。文献[8]对基于深度学习的表面缺陷检测方法进行了系统性的归纳,但主要梳理有监督方法。而近期,基于无监督、半监督等设置的研究同样涌现出许多新的成果,但目前尚无相应针对工业缺陷检测领域的全面而细致的综述文献。因此,本综述希望填补这一空缺,并着重对此类新方法进行介绍与总结。

本文后续内容的组织如下:第 2 节介绍工业缺陷检测问题的定义,分析研究难点与挑战;第 3 节基于 3 种任务设置介绍近年主流的工业缺陷检测方法,并按照方法的设计原理进一步归纳与分析;第 4 节梳理针对实际部署的关键辅助技术;第 5 节介绍常用的公开数据集与评价指标,并比较典型方法的性能;最后,第 6 节总结了当前研究的状况与局限性,并对未来发展趋势与潜在研究方向进行展望。

2 问题定义与研究现状

2.1 问题定义

基于视觉的工业缺陷检测旨在发现织物、芯片、药品乃至建筑材料等各种工业制品的外观可见缺陷。这些缺陷虽然微小,却可能严重危害产品的正常功能。它们可能发生在工业产品的生命周期中任何时期,如产品的生产、运输与使用过程等。

缺陷(defect)的概念可类比到异常(anomaly)。异常指超出预期模式范围的数据^[1],目前已有大量工作对其进行定义与归类^[1~5,9,10]。通常学者根据数据之间是否存在上下文关系,将异常分为点异常、上下文异常和集群异常^[1,5,10]。如图 1(a)所示,点异常又称为离群值(outliers)^[9],描述数值上偏离正常样本的独立数据;上下文异常同样描述数据点,其数值属于正常范围,但不符合局部上下文规律;集

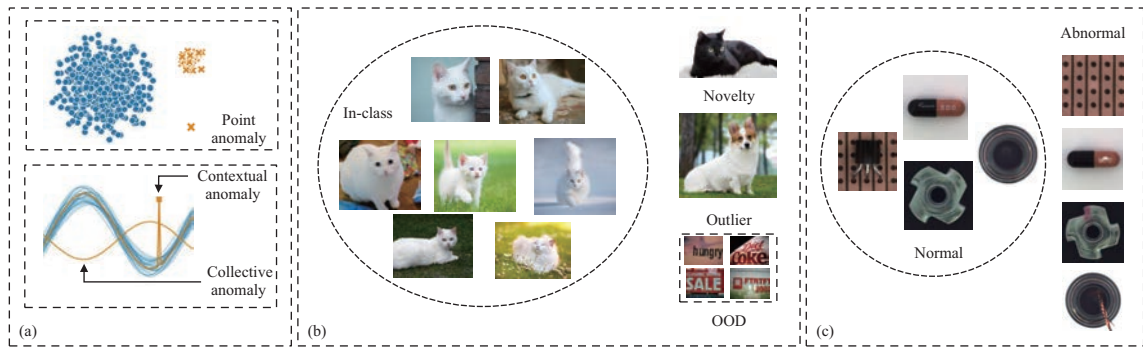


图 1 工业缺陷检测与异常检测中异常的关联与差异. (a) 一般异常的概念; (b) 图像异常检测相关概念; (c) 工业缺陷

Figure 1 The connections and differences of anomalies between industrial defect detection and anomaly detection scenarios. (a) Concepts of general anomalies; (b) concepts related to anomaly detection in images; (c) industrial defects

群异常描述一系列相关数据的集合, 集合中的每一个实例的数值在单独考察时都处于正常值域, 但集合整体的相关性特征不服从正常模式. 具体到图像数据, 文献 [5] 根据异常是否存在明显的语义性将图像异常分为低级纹理异常和高级语义异常. 与异常相近的概念还包括新颖点 (novelty) 和分布外数据 (out-of-distribution, OOD) [4]. 在图 1(b) 所示的图像分类任务中, 基于白猫样本定义猫类. 白狗即使颜色相近, 但因语义类别不同而属于离群值; 黑猫属于猫类, 语义类别相同, 但其颜色属性未在训练集中出现过, 而属于新颖点; OOD 则关注数据集合的分布差异, 文本数据集与自然场景中猫的数据集的分布呈现出明显差异. 类似地, 如图 1(c) 所示, 工业缺陷检测中, 正常样本包括多类产品, 缺陷可被视为其外观上的“异常”. 不同的是, 工业缺陷往往出现在图像中的小部分区域, 显著程度更低, 且语义概念模糊. 因而, 一般的图像异常检测往往仅需区分正常与异常样本, 而工业缺陷检测更关注于检测图像中的异常像素. 在实际工业场景中, 缺陷的定义更加主观, 因而学者们试图寻求其与异常检测的关联. 考虑到上述几种任务的相似性, 在部分缺陷检测方法中也采用了异常检测、新颖点检测与 OOD 检测的思路. 例如, 单独考虑缺陷的像素值, 可类比为新颖点或离群值; 考虑像素间的关联时, 又可将其类比上下文异常或集群异常.

虽然缺陷通常属于未知模式, 但仍然可以从已有的缺陷样本中发现一定的共性, 因此总结缺陷与背景的类型有助于针对性地设计检测方法. 如图 2 所示, 依据缺陷出现的位置与表现形式, 本文将工业缺陷分为表面缺陷与结构缺陷. 表面缺陷主要出现在产品表面的局部位置, 通常表现在纹理突变、异状区域、反规律模式或错误的图案. 例如, 表面裂纹、色块、织物的稀织以及商标文字的印刷错误等. 根据缺陷区域的像素值与周围背景的差异性可将其类比为离群值或集群异常: 离群值型缺陷的像素值通常与正常图像具有明显差异; 集群异常型缺陷的像素值与周围正常区域属于同一范围, 因而更难被发现. 结构缺陷主要是由产品整体的结构错误所致, 包括形变、错位、缺损与污染. 例如铁丝的弯曲、二极管的边缘缺损或处于错误的位置等. 相应地, 根据是否包含产品整体结构, 背景可分为纹理类与物体类. 纹理类聚焦产品的局部表面, 按照复杂程度依次划分为简单纹理、规则纹理与无规则纹理. 物体类包含产品整体, 结构更加复杂, 且存在产品之外的背景干扰. 此时, 不仅需要考虑表面缺陷, 也要考虑结构缺陷. 可见, 在不同的背景上, 不同种类的缺陷微弱程度不同. 即使是同种缺陷的不同实例之间, 可视性也可能有较大差异.

根据输出结果粒度的不同, 工业视觉缺陷检测任务一般包括分类和定位. 如图 3 所示, 对于一个待测图像实例, 分类任务首先将其二分类为正常样本或缺陷样本; 当缺陷类型已知时, 还可进一步对

Surface defects			Structural defects			Backgrounds		
Type	Outlier	Collective	Type	Normal	Defect	Type	Feature	Sample
Mutation			Deformation			Texture	Simple	
Region			Dislocation				Regular	
Pattern			Incomplete				Irregular	
Content			Contamination			Object	Complex	

图 2 工业视觉缺陷与背景的分类

Figure 2 Categories of industrial visual defects and backgrounds

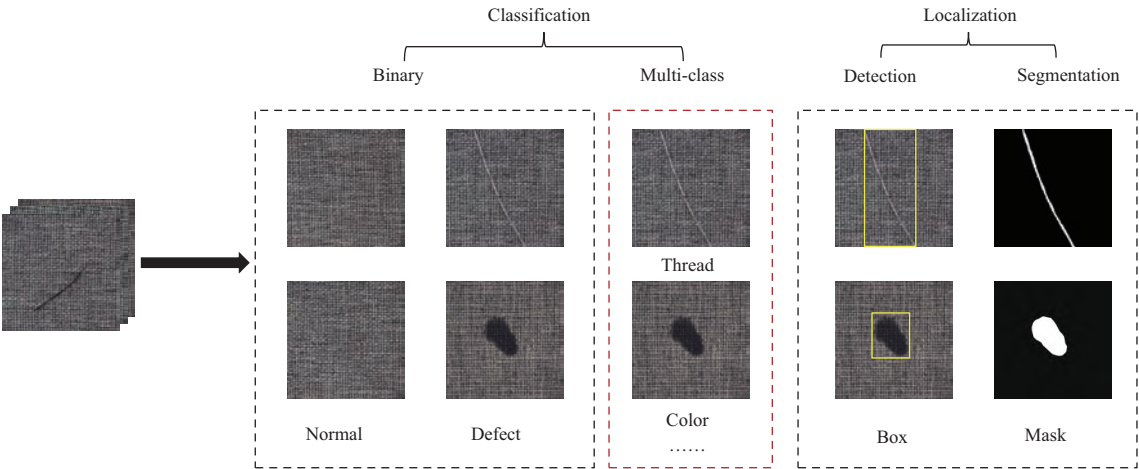


图 3 工业缺陷检测任务分类

Figure 3 Taxonomy of industrial defect detection tasks

缺陷类型进行判别, 例如, 织物上可能存在穿线、异色等多种缺陷. 定位任务的目标是找到缺陷在图像中的具体区域, 根据缺陷区域的描述方式可分为检测 (检测框) 与分割 (像素级). 事实上, 上述任务有时是可以同时进行的, 例如, 在分类模型上利用可视化方法^[11~13]可实现像素级的定位, 分割结果同样可以用来判断整图分类. 由于工业缺陷的形状不规则、尺寸变化较大, 检测框难以精确表示缺陷的位置, 而且容易引入较多无关的背景信息, 对缺陷检测性能的评估造成麻烦. 因而, 在实际的缺陷定位任务中, 研究者们更关注缺陷分割方法. 鉴于此, 本文将重点论述分类与分割任务的工作.

2.2 问题难点与挑战

与通用目标的分类、检测与分割任务不同, 工业缺陷检测需要考虑具体任务需求、数据样本与部署环境的特点, 因而带来了全新的挑战. 本文从数据难点与任务挑战两个角度进行分析.

2.2.1 数据难点

样本匮乏. 样本匮乏的问题主要体现在 3 个方面. 首先, 所提供的数据数量有限. 第二, 由于生产过程严格把控次品率, 缺陷在产品中是罕见的, 因此相比更易被获取的正常样本, 缺陷样本的数量很少. 第三, 对缺陷进行精确的标注成本高昂. 因此工业数据呈现明显的不均衡现象, 即大部分样本为正常样本, 只有少量为缺陷样本, 而具有精确标注的缺陷样本更加稀少.

未知性与主观性. 缺陷的模式往往无法预知, 因此难以直接对缺陷进行描述. 目前人为归纳的缺陷依然无法覆盖所有类型, 导致许多以往的方法无法识别新出现的缺陷类型. 同时, 缺陷的定义往往需要结合实际场景, 更加主观, 这也增加了数据标注的难度.

可视性低. 可视性低主要体现在 3 个方面. 第一, 缺陷往往仅为高分辨率图像中的一小部分区域; 第二, 缺陷与周围背景的对比度低, 显著性不足; 第三, 缺陷具有较低的信噪比, 成像引入的噪声也会干扰缺陷的检出. 因此, 相较一般异常检测, 工业缺陷检测中正负样本之间的差别非常小.

种类繁多. 缺陷通常是不规则的. 同一工业产品可能存在不同种类的缺陷, 同种缺陷也可能在形状、尺寸、颜色等特征上具有多样性. 以往的方法通常只能检测特定种类的缺陷. 现今, 方法需要具备检出各种样式的缺陷的能力, 避免出现漏检情况.

背景复杂. 以往的检测方法主要针对简单规则的产品背景, 而现在的待测产品具有更加复杂的结构与纹理. 一张高分辨率的产品图像中可能包含多种组件, 因而在不同区域风格迥异. 由于成像条件难以保持完全一致, 同种产品的不同待测图像之间也存在一定的差异. 在无人巡检任务中甚至存在多种视角. 对此, 模型需要足以容纳正常样本的类内差异, 避免将正常区域误检为缺陷.

2.2.2 任务挑战

高精度. 工业视觉缺陷检测要求方法具有高精度, 甚至能检测出人眼难以观测的细微缺陷. 不仅需要尽量减少漏检与误检, 还需要可以及时地调整检测性能. 不同的应用对于漏检与误检的需求也不同. 大部分工业场景下, 往往更倾向于避免漏检. 对于缺陷分割任务, 高精度的分割区域有助于后续的缺陷评估与修复.

低开销. 目前的成像技术已经可以为工业视觉系统提供高分辨率图像. 在实际部署时, 需要考虑到计算时间和存储资源的开销. 例如, 模型可能部署在嵌入式系统等算力有限的环境中. 为了与实际生产制造或智能巡检同步运行, 高速的、实时性的方法也成为研究重点之一.

2.3 研究概述

工业缺陷检测长期以来都是工业视觉领域最重要的研究之一. 近年来, 随着深度学习在计算机视觉任务中的普及, 基于深度学习的工业缺陷检测方法也得到了飞速发展, 并逐渐占据主流. 得益于卷积神经网络 (convolutional neural network, CNN) 强大的特征提取能力与对高维数据的表征能力, 基于深度学习的方法可以自动学习人工难以设计的特征, 不仅节约了手工设计特征的成本, 还令检测精度大幅提高. 相比基于图像处理与统计学习的传统方法, 它更擅长处理复杂的工业图像数据. 本文将近年国内外工业缺陷检测领域的热点研究进展组织为图 4 所示的结构. 检测算法依据实际工业场景中数据情况的不同, 被分为缺陷模式已知、缺陷模式未知与少量缺陷标注 3 种设置. 缺陷模式已知时, 一般采用有监督深度学习方法, 需要充足而精确的样本标注, 可以从分类、检测与分割 3 种角度进行方法设计. 缺陷模式未知时, 一般采用无监督深度学习方法构造比较对象. 根据比较对象维度的不同, 可分为在图像维度与在特征维度比较相似度, 并基于方法的原理进一步细分. 少量缺陷标注的场景贴近实际工业情况, 训练集中包含比例不均衡正负样本, 且只有少量的缺陷样本具有精确或不精确的标

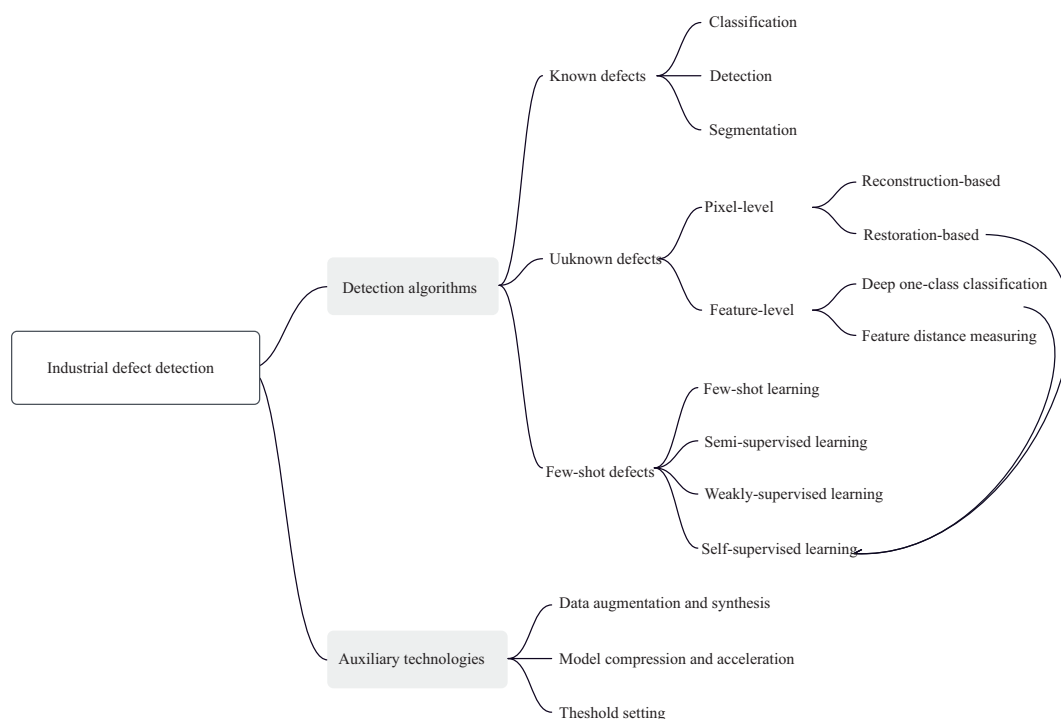


图 4 工业缺陷检测的研究进展与主流发展纵览

Figure 4 Overview of recent progress and dominant trends in industrial defect detection

注. 此时, 根据具体的数据标注情况, 分别采用小样本、半监督和弱监督等新兴方法来处理, 近年来这些方法也逐渐得到学术界与工业界的高度重视. 自监督学习属于无监督学习的一种, 其从无标注数据中挖掘自身的监督信息, 最近也在工业缺陷检测领域得到了广泛的应用. 因此, 本文从构建监督信息的角度对自监督方法予以归纳. 辅助技术主要用于提升检测方法的实用性, 本文主要从 3 个角度讨论. 数据增强与合成为数据贪婪的检测模型提供足够的训练集; 模型压缩与加速技术面向落地使用中的低存储开销与实时性需求; 阈值设置旨在找到推理阶段最合适的分类边界.

3 检测算法

3.1 缺陷模式已知

在许多实际工业场景中, 大部分缺陷的类型已事先经过专业的统计与归纳, 因而在方法设计时可以利用缺陷的特征直接检测, 或者利用先验知识搜集、标注数据集来训练模型.

3.1.1 传统方法简述

传统方法主要依据缺陷颜色、形状等特征, 利用图像处理方法或结合传统机器学习方法进行检测. 难点之一在于如何对缺陷进行描述. 从图像的角度, 由于工业缺陷往往表现在图像中像素突变的区域, 对于金属等背景简单的产品, 边缘检测方法可简单有效地定位缺陷区域. 常用的边缘检测算子包括 Prewitt, Sobel 和 Canny 等. 但是该方法难以处理复杂的背景或低信噪比的缺陷, 且对成像条件要求较高. 从频域的角度, 突变型缺陷在频谱中往往表现出高频特征. 因此对于具有简单或周期性背景

的产品, 可用傅里叶变换^[14]、Gabor 变换、小波变换等方法转换到频域来检测. Chetverikov 等^[15] 基于纺织品表面的纹理方向检测突变型缺陷. Hou 等^[16] 采用适合用于纹理表达的 Gabor 小波变换算子提取图像的频域信息, 然后利用支持向量机 (support vector machine, SVM) 进行分类. 然而, 区域型工业缺陷的内部往往较为平稳, 因此基于边缘检测的方法通常仅能检测到缺陷边界. 不过, 大面积缺陷会影响图像的统计特性, 因此可以利用灰度变化差异性^[17]、灰度直方图^[18]、颜色特性等基于统计的方法进行表征. 基于上述缺陷描述符, 还可进一步利用 SVM、随机森林^[19] 等传统机器学习方法进行分类.

传统方法需要手工设计缺陷的特征描述, 适用于简单规则的工业场景. 由于缺陷的微弱性与主观性, 手工特征难以准确地描述缺陷; 由于缺陷的未知性与多样性, 手工特征需要设计多组缺陷模板, 但仍只能描述种类有限的缺陷, 且无法探测新型缺陷. 当面对复杂无规则的数据时, 传统方法不仅难以应用, 还可能需要复杂的后处理过程.

3.1.2 深度学习方法

基于有监督设置的深度学习方法已经在许多视觉任务上取得了较为成熟的发展与应用. 当缺陷种类已知且具有充足的标注样本, 或者需要解决缺陷种类的多分类问题时, 工业缺陷检测往往采用有监督方法. 这些方法大多是将已有的通用目标的分类、检测、分割模型应用于工业场景, 并根据实际情况中缺陷的微弱性与模型的速度需求进行微调.

最基本的图像级分类方法^[20~24] 首先利用 CNN 提取待测样本的特征, 然后使用分类器进行分类. 这些方法在 PCB 板、塑料、金属、电池等各种工业生产领域均展现出较高的性能. 进一步, Cha 等^[25] 利用滑动窗口的方法在各个子图像块上进行分类, 实现了粗糙的缺陷定位. 由于工业缺陷数据集存在严重的不均衡现象, Xu 等^[26] 首先利用标签膨胀解决样本不均衡问题, 然后使用半监督数据扩增方式, 根据特征图的响应强度裁剪含缺陷的图像块, 为模型提供充分而均衡的训练集.

在实际应用中, 不仅需要判断图像是否存在缺陷, 还需要确定缺陷的位置. 对此, 研究者大多对通用的检测框架进行调整, 以适合工业数据的检测. Cha 等^[27] 基于双阶段检测器 Faster R-CNN^[28] 检测桥梁缺陷, 并将骨干网络改为 ZF-net 提升实时性. Tao 等^[29] 将 Faster R-CNN 级联, 实现电力巡检中绝缘子的缺陷检测, 第一阶段先从自然场景检测绝缘子, 以排除其他背景干扰, 第二阶段从裁剪的绝缘子区域检测缺陷, 整个网络采用端到端训练方式. He 等^[30] 通过融合多尺度特征提升检测精度. Li 等^[31] 基于单阶段检测器 YOLO 检测钢表面缺陷, 融合浅层特征提升对细小缺陷的检测能力; Zhang 等^[32] 采用 YOLOv3^[33] 检测桥梁缺陷, 并使用迁移学习、批再规范化与 Focal Loss 提升检测性能. Chen 等^[34] 将 DenseNet^[35] 与 YOLOv3 结合来检测 LED 缺陷. 可见, 对模型的调整主要在于两个思路: 引入轻量化网络提升检测速度; 利用多尺度融合与数据增强等方法来提升检测精度.

然而, 由于缺陷的无规则性, 其形状与尺度变化范围较大, 基于检测的方法因仅输出缺陷的包围盒, 难以准确地对复杂缺陷的位置与形态进行描述. 因此, 为了得到像素级的缺陷定位结果, 基于分割的方法逐渐成为研究的焦点, 更精确的缺陷描述也有助于后续的缺陷鉴定与恢复工作. 主流的分割网络如图 5 所示, 全卷积神经网络 (fully convolutional networks, FCN)^[36] 是最早使用 CNN 进行端到端分割的方法; U-Net^[37] 构建了完整的编码器 - 解码器结构, 使用解码器来还原分辨率, 并采用跨连接的方式融合细粒度特征. 在工业缺陷检测领域, 研究者往往基于这些基础的网络结构, 主要从多尺度与轻量化的角度设计模型.

Qiu 等^[38] 提出了一种三阶段有监督分割方法, 其首先基于 FCN 对输入图像初步分割, 接着采用全卷积神经网络对第一阶段的感兴趣区域 (region of interest, RoI) 进行二次判别以去除误检, 最后使

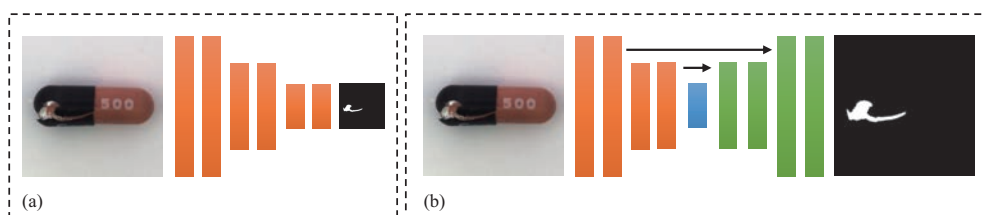


图 5 有监督分割网络的基本结构. (a) 常规 FCN 架构; (b) U-Net 架构

Figure 5 Basic architectures of supervised segmentation networks. (a) General FCN; (b) U-Net

用导向滤波器细化分割区域. 为了提升网络的效率, 其主要采用深度可分离卷积等网络结构优化的方法减少模型的计算复杂度. Tabernik 等^[39] 首先利用基于 FCN 的分割网络定位表面缺陷, 然后利用一个决策网络预测整张图像中存在缺陷的概率. 但是, 一般的 FCN 的模型输出图像经过了下采样, 尺寸减小, 需要后续通过上采样才能还原到原有分辨率. U-Net^[37] 采用编码器-解码器结构, 不仅可以输出原始的分辨率, 还采用跨连接融合多尺度特征以提升分割的精细度. Huang 等^[40] 在 U-Net 架构的基础上融入显著性检测, 输入经过多种显著性方式处理后的图像与原图的叠加, 以提升前后景分割的效果. 同时, 训练过程中, 在编码器之后引入额外的检测器分支, 令编码器学习的特征更加专注于缺陷区域, 从而提升显著性检测的效果. 然而, 面对显著程度不同的缺陷, 仅考虑空间域可能存在局限性. 对此, Xie 等^[41] 同时考虑了空间域与频域的信息. 其在主干分割网络的基础上, 使用一个 Secondary 分支, 基于离散小波变换提取图像的频域特征, 与主干网络的多尺度特征进行融合. 该方法有效地提升了模型对于细小的水泥裂纹的分割能力.

然而, 现实情况下难以提供充足且均衡的含标注数据集, 即使利用数据增强也无法彻底解决问题. 对此, 研究人员利用数据合成的方式生成缺陷样本及其标注, 然后用于训练有监督的方法. Valente 等^[42] 基于印刷品缺陷的分布先验合成数据, 用于训练分割网络. Zhang 等^[43] 基于生成对抗网络合成足够拟真而多样的缺陷数据来扩充训练集. 此外, 还可利用裁切^[44] 或剪贴^[45] 等自监督方法构建负样本, 然后按照有监督的方式训练模型.

有监督的方法可以解决缺陷多分类的问题, 适合于大部分缺陷的类型已知或缺陷特点鲜明的情况. 尽管需要高昂的标注成本, 但是此类方法在样本充足的情况下具有优良的性能, 并已在部分实际应用中检验了方法的有效性. 现有方法在简单规则的工业场景中已较为成熟, 而对于复杂背景及无规则微弱缺陷的检测仍有发展空间. 然而, 面对样本匮乏、缺陷模式未知的情况时, 有监督方法本身依然存在不足.

3.2 缺陷模式未知

在实际情况中, 含有缺陷的样本极难获取. 面对缺陷的未知性与无规则性, 基于缺陷先验知识的方法存在较大的局限性. 因此, 无监督的设置已经引起了广泛重视. 此类方法大多借鉴异常检测的思路, 对易于获取与描述的正常样本进行建模. 缺陷被定义为正常范围之外的模式. 无监督设置下的任务目标通常是判断待测样本是否包含缺陷, 或对缺陷区域进行定位, 此时定位结果的输出一般是像素级的分割结果.

3.2.1 传统方法简述

当缺陷模式未知时, 传统方法主要依靠正常样本呈现出的图像特征进行比较, 或基于传统机器学习模型对正常样本进行描述. 对于不同的背景, 存在相应合适的方法.

当各个样本在对应的非缺陷部位高度一致, 仅在缺陷区域存在差异时, 模板匹配法可以基于两图之差简单快速直接地实现缺陷的定位. 为了提升检测的精度, 往往通过对特征点^[46, 47]、局部线性轮廓^[48, 49]及局部区域^[50, 51]等手工设计特征进行匹配, 将模板与待测图像先尽量对齐. 只要能保证成像条件一致, 即使产品的背景排布复杂也能完成检测. 但也因此, 该方法对于具有多样性的背景, 如随机背景存在局限性. 同时, 模板的检索与匹配过程可能会导致检测时间大幅增加. 基于稀疏编码重构的方法利用了字典学习的思想, 构建一组“超完备”的基向量, 其线性组合可以表达足够多样的正常数据. 检测时迭代地寻找重构出最佳正常图像的稀疏矩阵, 随即实现模板匹配与缺陷分割. 该方法在氨气管道^[52]、织物表面^[53]、触摸屏^[54]与瓶盖^[55]等多种工业产品中得到应用. 此方法可以用更少的存储开销描述更多样的正常样本, 但是迭代寻找最优的稀疏系数的过程限制了检测速度.

对于存在周期性结构或规律性分布的正常图像, 可采用基于统计的方法, 利用图像的颜色、灰度^[17]、频谱^[56]等特征进行统计与比较. 与缺陷模式已知时不同, 此时主要统计正常样本的特征, 在测试时进行直方图相减等操作^[57], 消除正常背景区域^[58], 凸显缺陷区域. 同时, 这类正常图像往往具有低秩性的特点. 对此, 基于低秩分解的方法^[56, 59~63]将目标图片分解为低秩矩阵和稀疏矩阵. 基于缺陷具有稀疏性的假设, 利用稀疏矩阵表达图片中的缺陷区域. 然而该方法难以处理复杂而多样的纹理背景, 也容易将噪声误检为缺陷.

对于更加复杂的数据, 基于一类分类模型的方法将正常样本映射到特征空间使之聚集, 构建超平面^[64]或超球面^[65]对样本边界进行描述. 推理时将该边界作为样本分类的分界面. 处理高维的工业图像时, 使用核函数可以避免维度灾难. 针对分布复杂的工业数据, 后续方法^[66~68]主要着眼于提升模型对复杂数据覆盖与描述能力, 从而提升分类性能. 此类方法更多地被用于解决图像级分类问题.

综上所述, 面对成像条件稳定、背景简单或规则的工业产品, 传统方法可以有效地实现缺陷检测. 但随着场景逐渐复杂、缺陷愈发微弱、对检测精度与速度的要求不断提高, 传统方法的局限性也更加明显. 不过, 传统方法的思路一直在延续, 在许多基于深度学习的方法上得到了传承.

3.2.2 深度学习方法

基于无监督设置的深度学习方法仅需要易于获取的正常样本用于模型训练, 无需使用真实缺陷样本. 其不仅能解决有监督深度学习方法无法发现未知缺陷的问题, 而且拥有比传统方法更强的对图像特征的表达能力, 因而成为了当今的研究热点. 此类方法的核心思想是构建出一个与待测样本最相近的“模板”与之比较, 根据像素或特征的差异性实现缺陷的检出与定位. 根据比较维度的不同, 本文将方法划分为基于图像相似度的方法与基于特征相似度的方法.

(1) 基于图像相似度的方法

基于图像相似度的方法在图像像素层面进行比较, 其核心思想是重建出与输入样本最相近的正常图像, 两者仅在缺陷区域存在差别. 因而, 生成图与输入图像的差异图可表示缺陷存在的概率, 既可以用于判断整图是否包含异常, 也可以取阈值来得到缺陷的分割结果. 此类方法常使用自编码模型与生成式模型, 如图 6 所示, 包括自编码器 (auto-encoder, AE)^[69]、变分自编码器 (variational auto-encoder, VAE)^[70]、生成对抗模型 (generative adversarial networks, GAN)^[71]等, 最近则流行将 AE 与 GAN 模型结合, 以发挥各自的优势. 推理阶段时, 模型用于检测缺陷的步骤在图 6 中用蓝色箭头突出表示. 根据模型优化目标的不同, 可分为基于图像重建的方法与基于图像恢复的方法.

(i) 基于图像重建的方法. 基于图像重建的方法仅在正常样本上训练模型, 使其学习到足以用来重建出正常样本的分布特征. 此类方法假设: 由于模型的参数仅由正常样本训练得到, 模型只能较好地重建正常样本, 而在缺陷样本的缺陷区域则会产生较大的重建误差. 由于此类方法大多采用重建误

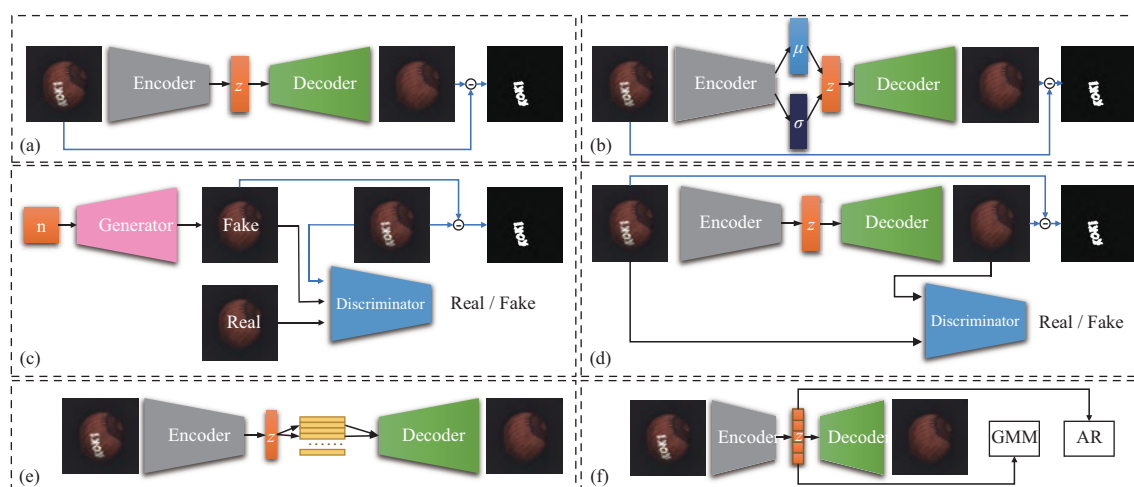


图 6 基于图像相似度的基础框架。(a) 自编码器; (b) 变分自编码器; (c) 生成对抗模型; (d) AE 与 GAN 结合。常见的改进方式: (e) 存储器机制; (f) 用高斯混合模型或自回归模型对隐变量建模

Figure 6 Basic frameworks based on pixel-level similarity. (a) AE; (b) VAE; (c) GAN; (d) combination of AE and GAN. Common improvements: (e) memory mechanism; (f) applying Gaussian mixture model or autoregressive model on latent variables

差实现缺陷的检出与定位, 重建图像的质量对缺陷检测性能影响巨大, 因而大部分研究者们尝试从提升重建质量的角度改进方法, 使得重建图像与输入图像的像素仅在缺陷区域存在明显差异, 而在正常区域几乎一致。

基于 AE 的方法采用编码器-解码器 (encoder-decoder) 结构的网络, 如图 6(a)。其中编码器将输入图像编码为隐空间变量 z , 解码器则利用 z 重建图像。根据输入图像与重建图像的重建误差即可实现缺陷定位。然而, AE 在重建时存在模糊的现象, 因而在计算重建误差时, 容易造成正常像素点的误检。对此, 一种思路是提升重建图像的质量。MS-FCAE^[72] 和 DFR^[73] 针对多尺度特征信息进行了相关设计, 为图像的重建提供了不同粒度的上下文信息, 使得重建图像更加准确清晰。MS-FCAE 在图像空间实现信息的多尺度融合, 该方法针对编码器不同层提取的特征分别设计对应的解码器进行重建, 得到多组重建结果, 然后将不同层的分割结果融合, 进行缺陷部位的定位分割。DFR 则在隐空间实现信息的多尺度融合, 将 VGG16 提取的底层特征与高层特征对齐并叠加后输入 AE 进行重建, 从而提升重建质量。另一种思路是反过来利用模糊效应, 将输入图像风格化。Chung 等^[74] 在 AE 的基础上引入了风格蒸馏分支, 该网络模拟了 AE 的模糊效应, 将原始图像转换为与重建图像具有同样风格的图像, 以达到对齐的目的。

基于 VAE 的方法通过编码器将正常图像映射到隐空间的先验分布, 从对应分布中随机采样得到变量 z , 再利用解码器将 z 映射回图像空间。假设缺陷区域的编码不服从该先验分布, 因而也无法较好地重建缺陷区域, 从而也可根据重建误差实现缺陷定位。VAE 可以构建比 AE 更加可控的结构化的隐空间流形, 因而可以从隐空间分布的角度来学习正常样本主要的共性特征。但是源于随机采样, VAE 难以获得清晰而一致的重建图像。针对该问题, Dehaene 等^[75] 采用了迭代逼近的思路, 首先使用正常样本训练 VAE, 训练完成后 VAE 将学习到正常样本在图像空间的流形分布。在测试阶段, 根据重建误差计算出的能量将待测样本沿分布的法线方向迭代更新, 使其逐渐映射到分布上, 此时 VAE 重建的图像即为最佳重建图像。为了提升检测性能, Zhou 等^[76] 综合利用图像空间与隐空间的信息进行织物表面的缺陷检测。其将 VAE 和高斯混合模型 (Gaussian mixture model, GMM) 结合, VAE 用于提取输

入图像的特征并重建图像, 为提升检测的鲁棒性而采用结构相似性 SSIM^[77] 来度量重建误差; GMM 根据 VAE 提取的特征进行概率密度估计, 最后将两者结果融合. 将学习隐空间分布的思想应用于 AE, Abati 等^[78] 通过自回归模型 (autoregressive model, AR) 学习隐空间变量的概率分布. AR 将隐空间变量中的各元素建模为条件概率分布, 其只与该元素之前的所有元素的预测结果有关. 对于测试图像中的异常区域, AR 将输出较低的似然. 相比于将隐空间约束为标准高斯分布的 VAE, 该方法考虑了隐空间向量元素间的关系, 可以更好地应对复杂数据.

基于 GAN 的方法利用了 GAN 强大的对分布建模的能力. 相比 AE 和 VAE, GAN 可以生成更加清晰的高质量图像. GAN 由生成器 (generator, G) 和判别器 (discriminator, D) 组成, 如图 6(c) 所示, 其中生成器-判别器的对抗训练机制是 GAN 生成清晰图像的关键. 具体而言, 生成器将从隐空间采样的变量映射为与监督样本尽可能相似的图像, 判别器则对生成样本的真实性进行判断, 旨在能够判别生成的假图, 两者互相对抗进行优化, 使得生成的图像趋于真实与清晰. 基于 GAN 的方法假设: 当训练集中均为正常样本时, 模型仅能学习到正常样本的分布, 因此隐空间变量经过生成器后只能得到正常图像. 此时, 可根据待测图像与重建图像的误差实现缺陷定位, 同时也可根据判别器的判别结果实现缺陷的检出. AnoGAN^[79] 是首个将 GAN 引入缺陷检测的方法. 由于 GAN 随机生成的图像可能与待测样本不匹配, 该方法采取了迭代优化的思路. 其首先在隐空间中随机采样一个变量, 输入到已在正常样本上训练完毕的 DCGAN^[80] 中得到生成图像, 计算重建误差并进行反向传播对隐空间变量进行更新, 旨在找到使生成图像与待测图像最相近的隐空间变量. 若 n 次迭代之后重建误差仍然较大, 则认为该样本为异常样本, 并可通过重建误差定位缺陷. 然而, 优化过程需要反复迭代, 大幅增加了模型推理时间, 使得方法缺乏实用性. 对此, Eff-GAN^[81], f-AnoGAN^[82] 在 GAN 的基础上额外引入了编码器以提取图像特征, 以引导生成器生成最匹配的图像. f-AnoGAN 采用了多阶段的训练方式, 第一阶段在正常样本上训练 WGAN^[83], 第二阶段训练编码器使其编码结果能够令 WGAN 生成与输入最接近的正常图像. Eff-GAN 则将生成器的结构倒置作为编码器, 训练时根据重建损失和判别损失共同约束进行端到端的训练. 引入编码器后无需进行繁复的迭代过程, 有效提升了推理速度. 类似地, Hu 等^[84] 也采用了多阶段的训练方式, 在测试阶段采用滑动窗口的方式将缺陷还原成正常织物纹理, 同时利用判别器对每个窗口进行判别得到似然图. 融合了似然图与重建误差图的异常分数图可用于缺陷定位, 有效减少了误检. 但是由于生成器与判别器用于检测的机制不同, 判别器的训练需要根据预设的阈值提前停止. Komoto 等^[85] 认为此类结构无法实现图像空间和隐空间的双向映射, 因而难以保证重建图像具有较高的一致性, 从而提出 CBI-GAN^[85] 将 BI-GAN^[86] 的生成器用 AE 模型代替, 以实现图像空间和隐空间双向映射, 编码器提取的特征即为图像在隐空间的最佳表达.

近年来, 将 AE 或 VAE 与 GAN 结合的框架^[77, 87~92] 成为了主流, 如图 6(d) 所示, GAN 中的生成器可作为 AE 或 VAE 的解码器, 以兼顾两者的优势. Baur 等^[88] 将 VAE 与 GAN 结合以重建出更加清晰且接近正常样本的图像, 有助于利用图像的重建误差分割缺陷区域. 作者进一步探索了编码器结构与编码形状的影响, 发现使用空间形状编码的 VAE 主干网络具有最佳的重建性能. GANomaly^[89] 除了对 AE 增加对抗训练, 还在生成器后接入了额外的编码器, 约束重建图像与原图的高层特征也保持相同, 从而提升重建的一致性. 推理阶段根据生成图像和输入图像的重构误差以及两者的编码误差来进行缺陷检测. 为了提升重建的质量, Skip-GANomaly^[90] 借鉴 U-Net 的结构引入了跨层连接, 以充分结合正常样本的多尺度的特征. 该方法相比 GANomaly 提升了检测能力, 但在不同类别的数据上性能差异较大. 为了使模型训练过程更加稳定、防止模式坍塌, DAGAN^[91] 借鉴了 BEGAN^[93] 的方法, 将判别器也改造成了 AE 结构, 从而可以适用于较小的工业数据集. 但是这些方法主要面向局部的比较, 且要求输入与输出足够对齐, 而缺乏对结构关系的考量. Bergmann 等^[77] 采用了基于结构相似性

的损失函数来考虑图像中的相关性,有效提升了模型对复杂真实工业数据的检测能力.为了强化结构与纹理的关系,Zhou等^[92]额外引入经过领域自适应方法训练好的结构提取模块.一方面,将待测图像与其结构信息的编码融合后送入解码器,可以提升重建质量;另一方面,待测图像与重建图像的结构差异也可作为缺陷定位的一种判断标准.

最近,研究者们将神经网络的可解释性工作用于无监督的工业缺陷检测.Liu等^[94]将基于梯度的注意力产生机制^[12]推广到VAE模型中,并根据输入图像编码后的隐空间分布与标准高斯分布产生的注意力图之差实现缺陷的定位.由于训练集中均为正常样本,Venkataramanan等^[95]则在训练时约束VAE的注意力图覆盖所有的正常区域,在推理阶段模型将对缺陷区域呈现出低注意力.该方法不仅促进模型充分学习正常样本的特征,而且对于阈值更加鲁棒,无需对不同类别的产品或缺陷精心设置阈值.Kimura等^[96]发现GAN的判别器的注意力集中在前景区域,因而利用注意力图可以抑制背景噪声的干扰,有效提升了对物体类产品检测的鲁棒性.

然而,基于图像重建的方法的假设并不完全可靠,即使仅在正常样本上训练,模型仍然可能将未见过的缺陷完整重建.研究人员认为原因主要有两点:AE等模型具有较大的容量;缺陷与正常区域的特征差别不明显.目前主要有两种思路来解决这一问题.一种方法是对隐空间分布施加约束^[78,97~99],使其只对正常样本具有良好表示,如图6(f)所示.Pidhorskyi等^[97]额外使用一个判别器对隐空间进行对抗训练,使之限制于标准正态分布.Perera等^[98]约束去噪AE模型的隐空间服从均匀分布,并用一个弱分类器挖掘与剔除隐空间中的缺陷特征,强化对隐空间的限制.Hong等^[99]提出离散损失将正常样本的特征分散充满整个隐空间,从而在推理阶段可以鲁棒地重建出与正常图像一致的效果,该方法有效地解决了结构缺失型缺陷的检测问题,且无需额外引入模块.另一种普遍的方法是引入特征存储器(Memory)来存储正常特征.

基于Memory的改进旨在利用额外的空间开销来保证模型重建正常图像.如图6(e)所示,研究人员在编码器后引入Memory存储正常训练样本的特征,推理时将待测图像的样本特征与Memory中的特征进行匹配,并用其进行重建.由于Memory中存储的都是正常样本特征,因此重建图像将不会包含缺陷.该思想最早应用于Gong等^[100]提出的MemoryAE,其编码器提取输入图像的特征 z 后,在Memory中检索若干项与 z 最相关内存项,并根据相似度进行重组后得到 \hat{z} ,将 \hat{z} 送入解码器中进行重建.该方法比直接存储图像更加灵活,且节约内存.后续的工作往往围绕Memory的结构与更新机制进行改进.Yang等^[101]利用多尺度的Memory机制提升了重建的效果.Hou等^[102]发现选择适当粒度的特征图作为Memory有助于增加重建图像在缺陷区域的差异性,因而提出基于图像块拆分与组装的Memory机制,提升了缺陷分割的能力.Park等^[103]借鉴了原型学习的思想,将正常特征区分为多个原型中心,拉近同种原型的特征的距离,增大不同原型的距离,以保证Memory具有足够的多样性,从而提升了对复杂数据的表达能力.在实际工业场景中,训练样本还可能被少量的异常样本污染,为了防止异常样本的特征存储到Memory中,TrustMAE^[104]提出了一种基于信任区间的Memory更新机制,设定一个阈值 r 作为每一个存储项的信任区间,只有当提取的特征与存储项的距离小于 r 时,才将提取的特征存入Memory中.实验证明,当训练集中混入了少量异常样本时,TrustMAE的缺陷检测性能远远优于MemoryAE.

基于图像重建方法简单直观,无需对训练集数据进行复杂的预处理.但是此类方法主要存在两大问题:第一,输入与输出难以保持对齐,例如图像风格不一致或存在偏移,且对噪声鲁棒性不佳,因此在计算输出与输入的差异时,容易造成像素级别的误检;第二,虽然基于正常样本训练,AE等模型仍可能泛化到缺陷,甚至退化成恒等映射,即输出和输入趋于一致,此时模型不具备异常检测能力.基于MemoryAE的方法虽然一定程度解决了恒等映射的问题,但是为了保证重建图像的质量,Memory中

通常会存储大量的正常样本特征, 这样不仅会消耗大量的内存, 在索引与输入图像特征最相关的内存项时, 也会消耗大量的时间, 导致模型运算效率低。

(ii) 基于图像恢复的方法. 基于图像恢复的方法将缺陷视为噪声, 将图像恢复视为去噪过程。此类方法的核心思想是在正常图像上加入缺陷后, 训练网络模型将其恢复为对应的原始图像。常用的模型包括 AE 和 U-Net 等。训练完成后, 模型具有根据上下文消除缺陷的能力。测试阶段利用恢复图像和输入图像的重建误差进行缺陷分割。由于模型的输入与输出不对等, 此类方法能一定程度上避免恒等映射的问题。

如何设计用于构造缺陷图像的变换方式是此类方法的核心问题之一。Sabokrou 等^[87]在原图上叠加高斯噪声。Fei 等^[105]采用了属性消除的方式, 作者认为选取的属性应当在正常样本内具有共性, 而在缺陷区域具有差异性, 且无需依靠额外数据集或标注。具体而言, 作者随机擦除图像的颜色与方向属性, 并依靠网络恢复这些属性。该方法有效避免了基于图像重建方法的恒等映射问题。Salehi 等^[106]考虑到图像中的位置关系, 借鉴拼图还原^[107]的方法打乱原始图像中的图像块顺序, 训练网络将其还原。为了防止网络学习到还原的捷径, 又对乱序图像增加了噪声。

更常用的方法是直接在原图上叠加掩膜, 训练网络恢复被掩膜覆盖的内容。Haselmann 等^[108]给正常样本随机添加矩形的掩膜以模拟真实缺陷, 但是该方法容易使模型过拟合到矩形缺陷上, 当测试集中出现其他形状的缺陷时, 模型可能无法将其消除。为了防止模型过拟合到训练阶段添加的缺陷上, Li 等^[109]随机选择图像的超像素作为掩膜用于训练模型。在推理阶段, 将超像素逐个遮住, 然后利用模型将其一一还原。然而, 一幅图像中可能存在数以万计的超像素点, 因而模型要进行数万次的前向传播, 计算量极高。基于此, Collin 等^[110]在重建误差的基础上, 引入了模型的预测不确定度进行缺陷区域分割, 在一定程度上缓解了计算效率低的问题。具体而言, 在 AE 的训练和测试阶段都引入 Dropout 操作^[111], 测试时对输入图像进行 N 次前向传播, 若图像中存在模型未见过的缺陷, 则 N 次重建结果间将表现出较高的不确定度。Yang 等^[112]利用隐空间聚类的思想, 训练时在原图上人工构造缺陷, 在隐空间将正常区域与缺陷区域的特征分别聚合成无缺陷簇和缺陷簇。测试时, 逐像素地在隐空间检测输入图像的异常特征, 进而对异常特征的像素进行定位和编辑, 实现缺陷特征的擦除, 使之尽可能接近无缺陷样本特征, 最后结合图像重建模块即可生成无缺陷样本, 通过重建误差实现缺陷部位的定位分割。

考虑到各种缺陷的尺度差异较大, 为了提升检测的鲁棒性和精确性, 相比于特别设计掩膜, 研究人员往往采用多尺度的常规样式掩膜。Mei 等^[113]使用图像金字塔结构, 将图像放缩到 3 个不同的尺度并随机加噪之后, 输入到对应的 AE 中进行恢复, 测试阶段将 3 个尺度的分割结果融合得到最终的缺陷分割图。Zavrtanik 等^[114]将图像划分为多个正方形区域, 并随机将其划为多个子集, 从而得到了多个包含随机正方形掩膜的图像。为了利用图像块之间的上下文信息恢复图像块, 作者额外对图像滤波, 并提出基于梯度相似度的损失进行约束。Yan 等^[115]则采用多尺度的水平与垂直方向的条纹状掩膜。在测试阶段, 图像中的缺陷被掩膜覆盖, 从而被恢复为正常图像。由于缺陷的尺度与形态未知, 通常需要将每一种掩膜都施加到图像上并进行恢复。

基于图像恢复的方法显式地约束模型, 使其将缺陷还原为正常像素。该过程可以促进模型更好地学习图像的上下文信息, 因此还能有助于检测结构缺陷。然而, 此类方法对人工缺陷的设计具有较高的要求, 不仅希望尽可能接近真实缺陷的分布, 而且需要具有足够的多样性, 否则容易过拟合到特定的缺陷形态上。虽然已有相关工作尝试利用多种更加通用的预处理方式解决该问题, 但在测试阶段需要对经过所有预处理方式变换后的图像进行多次前向传播, 因而效率较低, 在实际工业部署中, 难以满足实时性的要求。同时, 作为基于图像相似度的方法, 其性能与恢复图像的对齐程度直接相关。

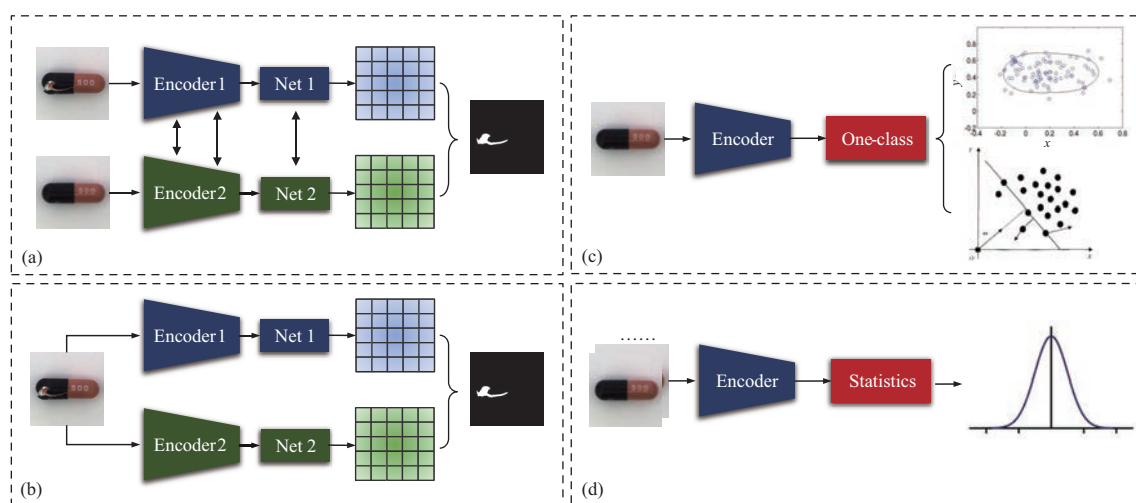


图 7 基于特征相似度的框架。(a) 孪生网络架构; (b) 教师-学生架构; (c) 深度一类分类模型; (d) 深度统计模型
Figure 7 Frameworks based on feature similarity. (a) Siamese network architecture; (b) teacher-student architecture; (c) deep one-class classification model; (d) deep statistical model

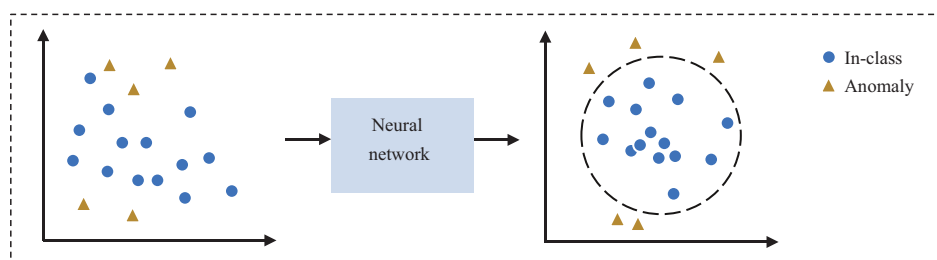


图 8 Deep SVDD 原理示意图
Figure 8 Illustration of Deep SVDD

(2) 基于特征相似度的方法

基于图像相似度的方法非常直观, 具有较强的可解释性。然而, 现有方法往往难以实现理想的重建效果。例如, 重建图像的像素可能未与输入图像对齐, 重建过程还可能同时改变图像的风格, 这些偏差都会导致检测错误, 限制了检测的性能。此外, 像素层面的比较容易受到噪声干扰, 导致检测的鲁棒性不佳。对此, 研究人员将目光投向特征空间, 通过比较高维特征以实现更加鲁棒的检测。借助深度神经网络的特征提取能力, 此类方法的核心目的是找到具有区分性的特征嵌入, 并减少无关特征的干扰。CNN 提取的特征包含了局部感受野的信息, 而非只考虑单独的像素, 因此不需要非常严苛的空间对齐, 同时增大了对噪声干扰的容忍能力。根据构建特征空间中正常“模板”的方式, 图 7 展示了 4 种常见的模型架构。

(i) 深度一类分类。深度一类分类方法利用深度神经网络构造质量更高的特征空间, 然后构造分类界面, 其主要框架如图 7(c) 所示。在训练时通过神经网络提取正常图像的特征, 并让提取出的正常图像的特征向量尽可能地分布紧凑, 从而在特征空间构建描述正常特征边界的分界面; 测试时, 模型提取待测样本的特征并映射到上述特征空间, 根据分界面判断测试样本是否为异常。

传统的一类分类方法在处理较大的数据集或在较高维的数据中检测异常时存在局限性。例如, 处理图像等复杂数据时容易失效; 基于核方法的一类分类需要大量内存来储存支持向量。对此, 引入深

深度学习有助于从高维的工业图像数据中筛选出区分性特征. 早期的混合式方法^[116]分两阶段依次使用深度学习与传统的分类方法. 其首先利用自编码器网络学习正常样本特征, 第二阶段固定自编码器参数, 再用 OC-SVM 等方法对特征编码构建分界面. Chalapathy 等^[117]提出的 OC-NN 提升了方法在复杂数据上的性能, 其在第二阶段用前馈神经网络替代传统分类模型, 用以提取有利于判别异常的特征, 并同时构造超平面将正常特征与子空间原点分离. Ruff 等^[118]提出了深度支持向量数据描述 (deep support vector data description, Deep SVDD) 方法, 将深度学习与 SVDD 结合. 如图 8 所示, 其原理为训练一个神经网络将正常样本特征映射到超球内. 训练阶段, 首先在特征空间中人为指定一中心点, 然后将正常样本特征到该点的距离作为主要损失函数来训练网络, 使其能将正常数据的特征映射到以中心 c 和半径 R 为最小体积的超球内. 测试阶段, 正常样本的特征会在上述中心点附近 (超球内), 异常样本的特征则会在超球之外, 据此可判断待测样本是否为异常样本.

然而上述方法不仅只针对分类任务, 还存在其他局限性. 首先, Deep SVDD 的特征中心点需要人为指定并且无法修改; 其次, Deep SVDD 假设正常样本在特征空间中落在人为指定的中心点附近, 但是分布复杂的数据集可能不满足上述假设; 第三, 上述方法可能存在模型退化的问题, 即模型将所有样本在特征空间中都映射到同一点. 这些问题都限制了方法在工业缺陷检测上的性能.

Deep SVDD 的局限性一定程度上可以通过优化特征提取过程规避^[119]. Perera 等^[120]以减小正常样本的特征之间的距离为目标, 可以做到无需人为指定特征中心, 并基于额外的多类数据集与描述性损失保证网络的区分力. Yi 等^[121]提出了 Patch SVDD 的方法, 解决了 Deep SVDD 只能判断图像是否为异常, 但是不能对异常区域定位的问题. 由于不同图像块 (patch) 的特征可能有明显差别, 即使正常样本的 patch 在特征空间中距离也可能非常远, 所以此方法用聚类的方式在训练时自动生成多个特征中心, 达到了图片异常检出和定位的效果. Liznerski 等^[122]提出了全卷积数据描述 (fully convolutional data description, FCDD), 其为 Deep SVDD 的一种修改形式. Deep SVDD 中的聚类中心在 FCDD 中由神经网络中的偏置项替代. FCDD 通过全卷积网络得到的特征图直接完成异常部位的定位. 由于 FCN 存在下采样, 其采用基于感受野的上采样提升定位能力. Wu 等^[123]通过保持特征的语义信息避免模型退化的问题, 即在网络中增加了解码模块, 要求正常样本的特征能够被解码为输入图像. 进一步, Massoli 等^[124]在此基础上引入多尺度来提升检测能力. 该方法将深度神经网络的运算视为一系列的顺序变换, 首先在图像重建任务上训练自编码器, 然后选取多层特征图分别构造超球面.

为了提升一类分类的性能, 自监督学习是另一种有效的手段. 其利用正常样本构造监督信息训练神经网络, 有助于挖掘正常样本的特征, 提升网络的判别能力. 这类方法的关键在于如何设计有利于缺陷检测的监督信息与自监督代理任务. Golan 等^[125]提出了一种基于 RotNet^[126]的方法, 该方法对输入数据进行几何变换, 并训练一个可以识别所使用变换的网络. 作者认为训练后的分类器可以很好地泛化到新的正常图像, 而不能正确地预测异常图像的变换, 因此可以将其用于异常检测. Bergman 等^[127]将上述思路 and Deep SVDD 结合, 提出了更加具有通用性的框架. 其首先采用正常样本进行多种几何变换, 然后分别将对应变换后的样本映射到特征子空间并使其紧凑. 在子空间中用超球对特征进行描述. 测试时, 若样本为异常, 则经过几何变换后样本的特征在超球之外. 此外, 研究者还将采用制造负样本的自监督方法^[44, 45, 128]提升一类分类模型的判别能力. Pourreza 等^[128]利用 GAN 在训练过程初期生成的图像作为负样本, Li 等^[45]通过剪贴其他的正常区域构造负样本, 从而可以用于训练有监督分类的代理任务.

深度一类分类方法大多面向异常检测任务, 通常用于图像级的异常分类, 可以直观地找到分界面. 其主要难点在于区分性特征的挖掘和优化问题. 由于任务的相似性, 相关方法也可以被应用于工业缺陷检测任务, 而且已有相关工作将其推广到缺陷分割中. 但是, 工业图像数据具有正负样本差别微小

的特点, 现有方法还需要对工业数据进行针对性的设计.

(ii) **特征距离度量.** 深度一类分类方法需要找到合适的子空间与分界面, 然而这对于复杂数据而言是困难的. 特征距离度量的方法无需优化分界面, 而是在特征空间匹配相应的“正常模板”, 直接与待测样本的特征比较. 因此, 合适的特征表达至关重要. 单纯使用正样本训练的特征提取器虽然可以学到正常样本的共性特征, 但是却无法学到缺陷的特征, 导致模型不具区分性. 因此, 相比于从零开始训练特征提取器, 研究者们选择直接使用预训练模型来描述图像的特征. 这些模型已经在大型的通用视觉任务数据集上预训练, 因而具备强大的表征能力. 度量形式主要有两种: 基于向量的距离、基于分布的距离.

最直观的方法是匹配到与待测图像最相近的正常模板进行比较. 如图 7(a) 所示, 该方法使用参数相同的具有区分力的特征提取网络将多个输入图像映射到特征空间, 它们的特征仅在缺陷区域差别明显. Napoletano 等^[129] 对正常图像划分区域, 依次采用预训练模型提取特征、降维与聚类等操作来构建正常字典, 从而可以处理具有多样性的正常图像. Cohen 等^[130] 提出的 SPADE 基于 k 最近邻 (k -nearest neighbor, k NN) 的方法, 首先使用已在 ImageNet^[131] 上预训练的 CNN 模型提取训练集的特征向量构建正常样本数据库, 然后使用 k NN 法检索与待测样本语义最相似的 k 个正常样本, 最后采用特征金字塔匹配的方法进行多尺度特征对齐, 从而提升了缺陷定位的鲁棒性. 这种方法简单直接, 无需复杂的训练过程却能实现良好而稳定的分割性能, 尤其适用于固定机位的工业检测场景. 但是, 较大的存储开销和较高的检索时间限制了其在大型复杂数据集上的应用.

为了减少上述开销, 基于深度统计模型的方法对正常样本的特征进行概率分布建模, 从而无需建立庞大的正常样本库. Rippel 等^[132] 首先利用预训练网络提取正常样本的多尺度特征, 并将各个特征图分别建模为多元高斯分布. 测试阶段用待测样本特征向量与“正常”分布的马氏距离来度量样本是否存在缺陷. Dedard 等^[133] 进一步将多元高斯分布建模于图像块的粒度上, 从而实现了像素级分割. 计算分布参数时, 利用多层特征图对齐组成的超列来融合多尺度特征. 相比于 SPADE^[130], 该方法减少存储和检索开销, 在大型数据上尤其显现出低复杂度的优势. 虽然它具有结构缺陷的检出能力, 但是限制了模板的多样性, 因而仅适用于成像位置尽量固定的场景, 而对于变化数据的泛化性不佳. Rudolph 等^[134] 引入归一化流 (normalizing flow, NF)^[135] 模型来提升灵活性. 但是由于 NF 关注低维特征而难以处理 OOD 问题^[136], 该方法首先利用预训练模型提取正常样本的特征, 然后基于 NF 模块将正常样本的特征映射到潜在空间 Z , 并保证正常样本的特征在特征空间 Z 上服从高斯分布. 异常样本的特征在 Z 空间将不服从高斯分布, 并具有较小的似然. 即使只使用小型轻量的特征提取网络, 该模型依然可以敏锐地检测出微弱的缺陷, 且无额外的存储开销. 虽然该方法为分类而设计, 但是通过梯度回传等技巧可以在一定程度上实现分割, 不过缺陷定位的准确性欠佳.

另一种思路是将同一张测试图像按照不同的方式映射到特征空间, 如图 7(b) 所示, 假设它们对正常区域的映射结果相似, 而对缺陷区域的映射结果差别较大. Bergmann 等^[137] 首次将教师-学生 (teacher-student) 框架应用于缺陷检测中, 其首先通过知识蒸馏的方式将大型预训练网络的表征能力迁移到轻量的教师网络中; 然后仅在正常数据集上训练多个随机初始化的学生网络, 使它们对正常样本的表达与教师网络相同. 该方法假设学生网络与教师网络对缺陷的表达具有较大的回归误差, 同时多个学生网络之间对缺陷的表达存在较大的不确定性, 从而实现像素级的缺陷分割. 面对尺度范围较大的缺陷, 该方法进一步利用多尺度感受野的网络提升分割性能. 但是网络的感受野是固定的, 且依靠经验设计, 不能自适应地调整. 同时, 模型数量的增加也会限制测试的实时性. Wang 等^[138] 利用自监督的方式训练教师网络以学习更优的特征, 并在蒸馏学生网络的同时显式地拉近各个学生网络输出之间的距离. Salehi 等^[139] 为了尽可能地保留网络对区分性特征的表达能力, 直接将大型预训练模

表 1 缺陷模式未知方法总结
Table 1 Summary of methods for unknown defects

Level	Method	Framework	Advantage	Limitation
Pixel	Reconstruct-ion	AE	Simple and intuitive to implement	Blurry output. May reconstruct anomaly
		VAE	Can build better latent manifold for further use	The reconstructed image is usually blurry
		GAN	Strong ability of modeling complex distribution	May need to find the best match iteratively
		AEGAN	The reconstructed image can be sharper, and the matched image is obtained without iteration	May reconstruct anomaly as well The training phase might be unstable like GAN
		Latent	Alleviate anomalous reconstruction implicitly	Difficult to model complex distribution
	Restoration ^{a)}	Memory	Alleviate anomalous reconstruction intuitively	Need extra storage and time overhead
		Synthetic defect	Alleviate anomalous reconstruction explicitly	Performance relies on the the design of defects. May overfit to synthetic defects
		Inpainting	Alleviate anomalous reconstruction explicitly	May need multiple forward propagation
Feature	Deep OCC	Deep SVDD	Intuitive to obtain classification boundary	Difficult to optimize when processing complex data
		Synthetic defect ^{a)}	High performance by introducing the idea of supervised learning using ‘anomalous samples’	Performance relies on the quality and diversity of synthetic data. May easily overfit
		Transformation ^{a)}	A more general approach to design supervision just using the properties of image itself	Performance relies on the design of transformations. Most are unable to localize defects
	Feature distance measuring	Template matching	No need for high quality reconstructions. Robust to noises. Suitable to fixed scenarios	Need extra storage. Performance relies on the quality and quantity of templates
		Student-Teacher	Robust to noises, less sensitive to alignment, more flexible than template matching	Performance relies on the pre-trained feature extractor. Based on uncertain hypothesis

a) This can also be categorized as self-supervised approaches.

型作为教师网络. 将学生网络设计为紧凑的轻量模型有助于提升推理速度, 并使其专注于区分性特征. 其对中间层特征向量的数值与方向同时蒸馏提升了知识迁移的效果, 进而提升缺陷分割性能. 针对文献 [137] 中感受野固定的问题, 其采用模型解释性方法, 如 SmoothGrad^[13] 等, 使模型可以适应多种尺度的缺陷. 该方法还极大地降低了计算复杂度, 实现了实时检测.

近年来无监督方法已取得令人振奋的进展, 并成为工业缺陷检测的研究重点之一. 在背景较为简单或固定的产品上, 相关方法已达到了较高的性能. 虽然在复杂背景的数据上依然无法达到有监督方法的精度, 但由于其对未知缺陷的检测能力和无需像素级精确标注的优势, 未来依然有很大的发展前景. 表 1 对缺陷模式未知设置下各方法的优缺点进行了总结与比较. 最近, 学者们开始尝试使用完全无监督的设置^[104, 140], 即训练集混杂了正常与缺陷样本, 但没有任何标注. 现有方法的主要思想是提升模型对异常样本的鲁棒性, 使其专注于数量较多的正常样本的特征. TrustMAE^[104] 利用基于信任区间的更新机制防止异常样本的特征被存入 Memory. Li 等^[140] 基于聚类的思想, 假设缺陷样本具有较大的方差. 训练时根据聚类结果仅采样部分被视为正常样本的数据来训练 AE, 并在训练过程中根据隐变量不断地调整聚类结果, 迭代地排除缺陷样本, 从而在训练后期仅采样正常样本来训练模型. 完全无监督的设置可以进一步节约标注成本, 并适用于样本极不均衡的工业数据. 目前相关研究仍处于起步阶段, 有待进一步发展.

3.3 少量缺陷标注

实际工业场景的情况往往介于上述两种设置之间. 不仅可以获得相对充分的正常样本, 还可以事先收集少量缺陷样本, 并进行标注. 即使缺陷样本只有粗粒度的标注也能为检测性能带来提升, 因此纯粹的有监督与无监督方法均难以充分利用提供的数据, 而不是最佳的解决方案. 此时, 面对样本稀少、数据不均衡、标注不精确等问题, 研究者们尝试基于小样本、半监督和弱监督的设置来设计更加

合理的方法.

3.3.1 小样本学习

无论是监督方法还是无监督方法都需要尽量充足的数据来支撑模型性能,有监督方法还希望训练集中正负样本均衡.然而,实际工业生产中可提供的样本总数较小,对于次品率的把控严苛,又直接导致缺陷样本的获取非常困难,从而难以构建理想的数据集.对此,小样本学习用于改善样本稀少时模型的性能.目前,解决思路大致可以分为以下 3 种.

(1) **网络优化.** 样本稀少往往会造成两大难题:当训练轮次受限时,模型难以快速充分地学习特征;当模型在训练集上充分训练时,又容易导致过拟合的问题.一方面,对于过拟合问题,可以选择更加合理的网络结构、度量方式和损失函数来抑制,如,利用正则项提升模型的泛化性;另一方面,合理的网络设计与稳定的优化过程有助于网络快速收敛^[91],使模型在小型数据集上也能有良好的性能.然而,网络与优化过程的设计具有一定的挑战性.

(2) **数据扩增.** 数据扩增是从训练数据的角度提升模型的训练效果与泛化性.其思想是充分利用所给数据,基于对缺陷的特点,有针对性地设计相应的缺陷数据与正常样本叠加进行缺陷样本的扩充.面向深孔零件铜丝数据, Tao 等^[141] 利用仿射变换、高斯模糊以及添加椒盐噪声等方法进行缺陷样本的生成;面向织物表面数据, Yang 等^[112] 通过生成随机掩模与正常样本叠加模拟真实缺陷模式进行缺陷样本的扩充;面向多类别工业产品数据, Lin 等^[142] 从缺陷样本中裁剪出缺陷部位与随机正常样本进行组合叠加进行缺陷数据的扩增.该类方法易于实现,但针对某一具体场景设计的缺陷形式难以泛化到其他场景,不能从根本上解决小样本问题.

(3) **知识迁移.** 一般情况下,使用正负样本失衡的训练集对模型从零开始训练极易导致模型失效,过拟合到训练集中数量较多的一方.在工业缺陷检测中,这种失衡的训练数据会导致模型产生大量的漏检.为了避免这种现象,常采用基于知识迁移的方法,具体可分为迁移学习与元学习.

基于迁移学习的方法指将用于其他任务的模型的知识迁移到工业缺陷检测任务中,即用少量数据对预训练模型进行微调. Gong 等^[143] 将预训练好的 DCNN 网络迁移到宇航材料表面缺陷的检测; Imoto 等^[144] 使用在海量弱监督数据上预训练的 CNN 网络,将其迁移到半导体数据上进行微调,完成了半导体材料缺陷检测的任务; Jing 等^[145] 采用在 MNIST 数据集上预训练的 CNN 模型,基于迁移学习的思路设计了一种局部特征检索的缺陷检测方法用于布匹缺陷的检测.迁移学习不仅有效地节省了模型的训练时间,还提高了少量样本的训练效果,增强了已有框架的泛用性.但另一方面,当源域与目标域差别过大时,迁移学习方法存在负迁移的可能性,导致模型整体性能的下降.因此,预训练模型的选择与使用也至关重要.基于元学习的方法的目标是训练一个初始模型,其可以利用少量数据快速地收敛到新任务中,从而能进行快速的调整. Lu 等^[146] 将元学习应用到自编码模型的训练中.第一阶段利用各种场景的正常样本,按照 MAML^[147] 的方式训练初始模型,第二阶段用少量新的待部署场景样本微调.工业缺陷检测方法往往部署于固定的场景,因此可以用该方法得到最适合部署场景的模型.值得注意的是,元学习也需要考虑优化方向的问题,在第一阶段希望各个场景对模型具有相似的优化方向.

3.3.2 半监督学习

在实际工业场景中,缺陷样本的搜集和标注成本高昂,因此存在大量的无标注数据.以往的任务设置未能充分利用这些无标注数据,基于此,研究人员将目光转向了基于半监督的设置.该设置下,训练集中同时包含正常样本和异常样本,但只有少量样本具有标注.对比有监督设置,其减少了对样本

标注的需求; 相比于无监督的设置, 其对少量的缺陷数据也加以利用。

Di 等^[148] 基于半监督的设置提出了一种用于钢材表面缺陷分类的方法。基于知识迁移的思想, 该模型的训练过程包括两个阶段, 第一阶段利用大量的无标注样本训练卷积自编码器 (convolutional autoencoder, CAE), 利用重建的优化目标使自编码器具有良好的特征提取能力。第二阶段在 CAE 的编码器后引入全连接层构成 SGAN^[149] 的判别器, 利用生成器的生成图和有标注样本训练判别器进行 $1 + N$ 类的分类, 其中 N 分类用于判断待检样本属于 N 类缺陷中的何种类型, 剩余一类用于判断待检样本是否属于真实样本。测试阶段将待测图像输入判别器中, 根据判别器的判别结果进行缺陷分类。

然而, 如果仅将少量有标注数据用于监督训练, 模型容易过拟合到有标注样本, 难以泛化到其他无标注的缺陷。为了利用无标注数据, 构造伪标签是常用的手段。Gao 等^[150] 简单地利用训练中的模型对无标签数据分类来构造伪标签, 并在训练过程中不断地调整伪标签数据参与的权重。当训练集本身较小时, 还可以事先合成数据来扩充数据集。He 等^[151] 提出了一种基于 cDCGAN 和以 ResNet18 为骨干网络的分类器的方法。其首先通过 cDCGAN 生成无标签样本, 扩充训练集的规模。然后利用判别器和分类器对这些样本分别进行分类, 当两者分类结果一致时, 为这些样本分配伪标签用于加强分类器的训练。基于伪标签的方法需要注意数据的可靠性, 错误的标签反而会干扰训练。

Chu 等^[152] 从另一个角度来考虑缺陷检测问题, 提出了一种利用损失函数曲线来度量样本是否存在缺陷的新颖方法。作者认为, 当训练集中正常样本占绝大多数时, 自编码器损失函数的收敛归因于正常样本, 而异常样本将导致损失函数振荡, 因此可以根据损失函数曲线的变化趋势判断对于像素是否属于缺陷。对此, 本文提出了一种新型的半监督设置: 利用少量有标注数据来实现大量无标注数据的缺陷分割, 而在训练时综合利用所有数据。具体而言, 该方法在训练阶段从全体数据中采样来训练 AE, 每次更新 AE 的参数后, 在全体数据上评估重建损失, 从而记录损失函数曲线。接着使用有标注数据集的损失曲线训练预测器实现二分类。经过多次 AE 的初始化与训练过程, 存储无标注数据集上最优的损失函数曲线, 用预测器实现缺陷分割。该方法与以往的缺陷检测方法不同, 其将无标注的测试数据加入到训练过程中, 而不是训练一个可以直接预测新样本的模型。但从另一角度而言, 该方法可以视为标签的自动生成, 从而扩充了有标注数据集。

相比于无监督方法, 如果能提供少量有标注的缺陷样本, 运用半监督学习可以显著提升模型的性能。基于像素级别的标注, CAM-UNet^[153] 利用 U-Net 进行缺陷的分割。该方法在第一阶段训练编码器令其可以进行图像级别的分类, 编码器训练完成后, 根据分类结果利用 CAM^[11] 得到缺陷的粗定位。第二阶段训练两个解码器, 首先将激活图作为先验知识与编码器提取的特征图融合, 引导解码器 0 实现精细的分割, 然后将此分割结果下采样, 再度与上述特征图融合, 输入解码器 1 进一步地细化分割。同时约束两解码器的分割结果, 使其互相促进, 从而提升模型整体的分割性能。Kimura 等^[96] 在其半监督的设置中利用少量有标注的缺陷样本训练一个额外的判别器, 使网络可生成更优的注意力图。

当训练集中包含无标注的有缺陷样本时, 大部分现有的半监督方法主要用于解决图像级别的缺陷分类问题, 而在实际工业产品的生产过程中会不断产生新类型的缺陷, 因此已有的训练集依然无法囊括所有的缺陷类型, 导致模型无法进行新类别缺陷的发现。对于缺陷定位任务, 半监督的方法大多为基于无监督方法的改良, 即除了大量正常样本之外, 还存在少量缺陷样本, 并都有像素级精确标注。而对于存在无标注缺陷样本的情况, 半监督分割的方法仍有待进一步的探索。

3.3.3 弱监督学习

在弱监督任务中, 缺陷样本通常只具有粗粒度的标注, 而在推理阶段模型需要输出细粒度的结果。在仅有图像级别标注的条件下, Niu 等^[154] 提出了一种基于 CycleGAN^[155] 的缺陷分割方法。类似图

像恢复,该方法在训练阶段将异常样本输入到自编码器结构的生成器,用正常样本对生成结果进行监督,使其逼近正常样本;推理阶段将待测样本输入生成器,根据重建误差进行缺陷分割。

神经网络可解释性的方法也被用在弱监督的设置中。此类方法通常利用图像级别的标注训练分类模型,然后使用 CAM^[11], Grad-CAM^[12] 等方法找到特征图中对分类结果贡献最大的区域,从而实现缺陷定位。Mayr 等^[156] 基于 Resnet50 的骨干网络设计二分类网络,基于 CAM 的思想,利用分类结果生成缺陷的激活图,从而得到分割结果。Venkataramanan 等^[95] 利用 VAE 中的注意力机制实现了弱监督设置下的分割,为了提升分割性能,训练时还对注意力图施加约束。在其无监督设置的基础上,增加一个二分类分支来判断样本是否存在缺陷。训练时,对正常与异常分类的分数梯度回传,扩展正常注意力,抑制异常注意力。在推理阶段,通过注意力图实现分割,通过分类器实现图像级的分类。即使只使用少量粗粒度标注的缺陷样本,也能有效提升模型的性能。

基于弱监督的方法利用粗粒度的标注实现了比无监督更高的性能,同时与有监督设置所需的精确标注样本相比,标注成本更低。这种权衡在实际情况中具有较大的实用性,因此,基于弱监督的方法在工业缺陷检测领域中也非常值得研究。

3.3.4 自监督学习

最近,自监督学习成为计算机视觉领域热点之一。自监督学习属于无监督学习的一种,利用无标注数据本身自动设计监督标签,接着借助类似有监督学习的范式训练模型。由于工业场景缺乏充足的人工标注的缺陷样本,自监督学习方法同样引发了广泛的关注,并展现出强大的缺陷检测与分割能力。因此,本文在此单独归纳基于自监督学习的方法,以帮助研究人员从该角度了解自监督方法在工业缺陷检测中的应用与发展。根据监督信息和代理任务设计方式的不同,主要可以分为 3 种:基于图像复原的方法、基于缺陷合成的方法和基于图像变换的方法。

基于图像复原的方法挖掘图像内在属性与内容的关联性,主要思想是显式地训练模型将经过变换的输入图像复原为原始正常图像,从而在测试阶段,正常区域的像素基本不变,而异常区域被擦除,进而根据重建误差实现缺陷的检测与分割。在第 3.2.2 小节的基于图像恢复的方法中已有比较详细的描述。Fei 等^[105] 采用属性擦除的方式,让模型可以恢复正常图像的颜色、方向等属性,并将异常像素复原。Salehi 等^[106] 利用拼图还原的代理任务,使网络具有恢复正常图像的能力。常见的方法还有给输入图像加入噪声^[113] 和添加人造缺陷^[112]。但是这些方法需要精心设计监督信息的构造方式。为了规避擦除的属性或缺陷设计的麻烦,研究者们考虑采用更加通用的复原任务^[109,110,114,115,157],即直接在图像中叠加掩膜,训练网络根据周围信息复原掩膜中的内容。在测试时,网络可以输出复原的图像,然后基于重建误差实现缺陷的检测与分割。这类方法展现出更强的泛化性,但在测试时也需要遍历图像的预处理过程,进行多次前向传播,影响了模型的效率。Ristea 等^[158] 则直接将复原的思路整合到卷积核中,使之可以适用于各种基于图像相似度的框架,并加入通道注意力提升复原效果。

基于缺陷合成的方法通过在正常图像上构造缺陷,得到缺陷样本及其图像或像素级精确的标签,从而可以利用对比学习或有监督的方式训练模型。最常见的异常样本构造方式是裁切区域^[44]。Tayeh 等^[159] 基于图像块进行度量学习,分别将其他图像块及其随机裁剪后的数据作为正样本对与负样本对,采用三元组损失使得网络具有区分性,实现了异常图像的检测。Pourreza 等^[128] 利用 GAN 训练初期的图像作为负样本训练分类模型,以实现异常检测。Li 等^[45] 将正常图像的其他区域随机裁切并粘贴到正常样本中,得到异常样本与像素级标注,从而利用其在分类任务上对模型进行表征学习。测试时,分别在图像与图像块维度进行高斯密度估计,以实现缺陷的检测与定位。然而,缺陷合成的方法对于合成方式非常敏感,往往容易过拟合到人造缺陷模式上。对此,一种思路是提升合成缺陷的真实

表 2 任务设置总结
Table 2 Summary of different task settings

Setting	Scenario	Advantage	Limitation
Supervised	Adequate and balanced data with annotations	Best performance when training set is adequate. Able to classify defect types	May overfit when training set is limited. Fail to detect unseen defects
Unsupervised	Normal data only, or unlabeled data only	Save the cost of manual annotating. Can detect unknown defects	Performance is limited due to the lack of defect prior. Can not classify defect types
Self-supervised ^{a)}	Annotations are generated on normal data	Comparable performance with supervised methods without manual annotating	Performance highly relies on the design of supervision and pretext
Few-shot learning	Limited annotated data, or limited normal data	Designed for limited training set or fast adaptation	Generalization still needs to be improved
Semi-supervised	Both labeled and unlabeled data can be accessed	Close to actual scenario. Make full use of every provided unlabeled data	Most are designed for classification task. May fail to detect unseen defects
Weakly-supervised	Annotations of data are inexact or inaccurate	Improve localization using coarse labels compared to unsupervised methods	Localization still needs to be improved compared to supervised approaches

a) Self-supervised learning can also be regarded as a sub-domain of unsupervised learning.

性, Schlüter 等^[160]在 CutPaste^[45]的基础上, 使用泊松图像编辑无缝混合多个图像块作为合成缺陷, 提升了缺陷的多样性与真实性. 另一种思路是与其他分支结合以减少拟真需求. Zavrtnik 等^[161]将图像修复网络与分割网络级联以提升泛化性. 其首先利用柏林 (Berlin) 噪声和额外的数据合成缺陷, 第一阶段还原成正常图像, 第二阶段根据输入图像与原图的不一致性来分割缺陷区域, 实现了具有较高精度与泛化性的分割能力. Song 等^[162]在 CutPaste^[45]的基础上, 在裁切前对源图像增加了随机旋转、位置打乱和颜色抖动等数据增强的方式提升合成缺陷的多样性, 并拼接坐标通道提升分割性能, 然后用两个分支分别进行正常区域重建与异常区域分割. 测试时, 融合重建误差与分割结果实现鲁棒的缺陷定位.

基于图像变换的方法通过代理任务训练判别式模型, 其假设模型能够正确判别正常样本的变换方式, 而无法预测异常样本的变换方式. 这类方法往往会与深度一类分类方法结合. 常见的变换方式包括旋转角度^[125, 127, 163]、颜色变化、顺序打乱^[164]等, 并将变换的编号或参数作为监督信息, 训练网络预测正常图像的变换方式. 测试时, 若预测结果与实际变换方式不一致, 则该输入图像包含缺陷. 然而, 由于图像变换往往是全局的, 此类方法通常用于图像级的缺陷检测. Sheynin 等^[165]基于 SinGAN^[166]学习图像的內部分布, 并将判别器改为 Patch GAN. 训练时, 生成图像将被按照定义变换进行数据增强, 训练判别器预测每个图像块的变换方式. 在测试阶段, 被错误预测的图像块则表明了缺陷的存在, 从而实现缺陷的定位.

表 2 总结了各种任务设置的定义与优缺点, 可根据实际数据情况与任务需求来选择合适的方法.

4 辅助技术

检测方法设计不仅要考虑方法本身的检测精度, 还需要根据实际情况, 利用一些辅助技术帮助其投入产业使用. 对此, 本文主要综述 3 种关键的辅助技术: 数据扩增旨在利用合成等手段增加样本的多样性, 以提升有监督和自监督方法的性能; 轻量化技术关注模型在实际部署时的时间与空间复杂度, 帮助模型在低功耗设备上达到实时检测的水平; 阈值设置是区分正常与异常样本的关键, 合适而可控的阈值设置有助于方法实现较高的实际性能.

4.1 数据增强与合成

由于工业数据往往缺乏具有精确标注的缺陷样本, 不足以支撑神经网络模型的训练, 对此适当的数

据扩增是提升模型性能的关键. 常用的图像扩增方法包括对原始的缺陷样本采用裁切、旋转、扭曲、仿射变换等. Huang 等^[40] 使用移动最小二乘法将缺陷变形来扩充数据集, 以防模型过拟合. 虽然这些方法可以改变缺陷的形状, 但是缺陷依旧处于原有背景, 而缺乏足够的多样性, 且存在图像失真的问题.

为了得到更加丰富的扩增数据, 缺陷合成技术希望在正常样本上构造拟真的缺陷. Lin 等^[142] 采用了一种直观的方法, 利用少量缺陷样本及其掩膜, 将真实的缺陷区域叠加到另一正常样本的相应位置而构成新的缺陷样本. 由于直接的剪切粘贴可能导致图像不真实, 用合成样本计算的损失被施加权重后再进行反向传播, 该权重基于新样本与真实样本的余弦相似度而得. 但是该方法依然缺乏灵活性, 无法在图像的任意区域合成足够真实的缺陷. Valente 等^[42] 基于印刷品的缺陷分布, 随机生成了色带与黑条纹缺陷, 并模拟了印刷过程的色彩变化. 该方法依靠人工事先对缺陷分布进行统计, 难以拓展到更加复杂的缺陷样式.

由于可以自动学习缺陷的分布, 生成足够逼真的图像, 基于 GAN 的方法被自然地用于缺陷合成, 但是其难以生成复杂而多变的缺陷样式. Pourreza 等^[128] 探究了 GAN 的训练过程, 将模型训练初期的生成图作为随机异常样本, 而训练后期的生成图用于扩充正常样本, 以辅助一类分类任务. Zhang 等^[43] 提出的 Defect-GAN 模拟缺陷的制造与修复过程, 不仅可以控制合成缺陷的位置与种类, 而且利用自适应的噪声添加实现了多样化的缺陷合成. 实验表明利用该方法合成的缺陷样本有效促进了缺陷检测模型的训练.

4.2 模型压缩与加速

现在深度学习方法已经被广泛地应用到工业缺陷检测领域中, 神经网络强大的学习能力也使得其在复杂数据上往往要比基于传统方法的表现更好. 但是神经网络在处理高分辨图片时往往也存在较高的计算量, 限制了模型的速度. 在实际的工业场景中, 实时性也是至关重要的, 模型需要配合生产速度及时地反馈结果. 同时, 有限的硬件资源也对方法的计算量与存储量提出了低开销的要求. 对此, 许多工业缺陷检测方法^[28,38,39] 也采用了更加轻量的网络模型, 来提升方法的实用性.

目前已有多种方法可以实现模型压缩与加速. 第一, 采用更高效的计算单元. Andrew 等^[167] 利用深度可分离卷积替代了传统的卷积操作, 显著降低了模型的计算复杂度; Christian 等^[168] 使用了多个小卷积核的堆叠替代了大卷积核, 可以在不损失精度的条件下减少网络的计算复杂度. Max 等^[169] 采用低秩分解来减少模型的计算复杂度. 基于卷积核的低秩性, 将原来大矩阵分解成多个小的矩阵. 因为小矩阵的计算量都明显比原来大矩阵的计算量小, 所以可以实现计算复杂度的降低. 第二, 减少模型结构或计算过程的冗余. 剪枝通过去除大型网络中冗余的神经元与连接, 在不影响性能的前提下, 得到一个更轻量而高效的网络^[170]. Rippel 等^[132] 对网络中用于区分正常与缺陷的特征进行分析, 发现存在大量冗余, 进而采用 NPCA 的方法筛选少量重要的通道用于后续计算. Dedard 等^[133] 发现直接随机选取特征通道的效果比 PCA 更好. Bergmann 等^[137] 在全卷积网络的前项传播中引入快速特征提取^[171] 避免了重复计算. 第三, 将大型网络的知识迁移到小网络中. Bergmann 等^[137] 利用知识蒸馏的方式将大而深的网络的知识迁移到轻量网络上以提升推理速度.

模型压缩方法在提升模型推理速度的同时也能减少模型的参数量. 对于需要额外存储正常样本的方法, 还需要在保证检测精度的同时, 尽量减少存储开销. Park 等^[103] 通过增加 Memory 中特征的表达效率, 用更少的存储容量就能实现更加多样的正常样本表达. Dedard 等^[133] 用多元高斯分布对正常样本的多尺度特征建模, 因而仅需存储在无缺陷样本的特征上计算的均值与协方差. 与直接存储无缺陷样本图像的 SPADE^[130] 相比, 其不仅减少了存储开销, 也免去了耗时的图像检索步骤.

4.3 阈值设置

在实际部署中, 需要设置阈值判断待测样本是否属于异常类. 有监督的方法通过 softmax 等分类器获得归一化的概率, 基于一类分类的方法构造出分类界面, 两者都具有明确的物理意义. 然而, 大部分基于图像或特征距离度量的无监督方法对缺陷发生概率的表达往往基于异常分数, 因此难以直接确定分界面. 这些方法的实际性能对阈值敏感, 促使阈值设置的问题成为迫在眉睫的待解难题.

Venkataramanan 等^[95] 不使用重建误差, 而是将归一化的注意力图作为异常分数图, 用于像素级的缺陷分割. 实验证明该方法对阈值的设置不敏感. Rippel 等^[132] 利用多元高斯分布对特征建模, 用马氏距离度量异常分数, 进一步可根据相关公式理论推导得出保证假阳性率在预期范围内的分界阈值. 但是上述的解决方法并不具有普适性, 大部分方法仍然需要借助异常样本来确定最佳阈值. 对此, Bergmann 等^[172] 提出了 4 种使用验证集来设置阈值的方法, 验证集中只包含正常样本.

最大缺陷分数法. 设定阈值最简单的方法, 是将验证集上预测的最大异常分数作为阈值, 从而保证模型对验证集上的正常样本恰好不会产生误检. 但由于不同种类的缺陷的分数范围不一致, 该方法过于保守, 可能会导致在测试时产生大量漏检.

p 分位数法. 如果验证集不够纯净, 上述方法会导致阈值过高. 为了提升对验证集的鲁棒性, 设置阈值时考察模型在验证集上输出的异常分数的整体分布, 并允许存在一定数量的异常像素点. 具体而言, 预先设置一个 p 分位数, 以该分位数选择一个阈值, 使 $p\%$ 的像素点分类为无异常部分.

k-Sigma 阈值法. 为了一定程度上预测假阳性率, 该方法考察模型在验证集上输出的异常分数, 并将其用均值为 μ 、标准差为 σ 的高斯分布建模, 然后将阈值定义为 $t = \mu + k\sigma$. 如果可以假定此分布完全遵循高斯分布, 则可以选择 k 以在验证集上满足一定的假阳性率, 并推广到测试数据. 然而在实际情况中, 数据很难恰好呈现高斯分布.

最大缺陷面积法. 上述方法仅考虑单个像素的层面, 但事实上, 缺陷图中可能存在被误检为缺陷的小面积噪点. 因此, 可预先设置最大缺陷面积, 当分割结果的连通域小于该面积时, 则将其滤除. 阈值同样设置为验证集恰好不被检出缺陷区域的异常分数.

5 数据集与性能评估

5.1 常用数据集

当前常被用于工业缺陷检测研究的数据集如表 3^{[19, 30, 39, 173~182]¹⁾} 所示. 本文具体介绍近年提出

1) 各数据集具体链接地址.

elpv-dataset: <https://github.com/zae-bayern/elpv-dataset>.

NEU-CLS: http://faculty.neu.edu.cn/yunhyan/NEU_surface_defect_database.html.

TianChi Fabric: <https://tianchi.aliyun.com/competition/entrance/231666/information>.

TianChi AP: <https://tianchi.aliyun.com/competition/entrance/231682/information>.

NEU-DET: http://faculty.neu.edu.cn/yunhyan/NEU_surface_defect_database.html.

GDXray Casting: <https://domingomery.ing.puc.cl/material/gdxray/>.

DeepPCB: <https://github.com/Charmve/Surface-Defect-Detection/tree/master/DeepPCB>.

Magnetic Tile: <https://github.com/abin24/Magnetic-tile-defect-datasets>.

KolektorSDD: <https://www.vicos.si/resources/kolektorsdd/>.

RSDDs: <http://icn.bjtu.edu.cn/Visint/resources/RSDDs.aspx>.

CrackForest: <https://github.com/cuilimeng/CrackForest-dataset>.

MVTec AD: <http://www.mvtec.com/company/research/datasets>.

AITEX: <https://www.aitex.es/afid/>.

BTAD: <http://avires.dimi.uniud.it/papers/btad/btad.zip>.

NanoTWICE: <http://www.mi.imati.cnr.it/ettore/NanoTWICE/>.

DAGM: <https://hci.iwr.uni-heidelberg.de/node/3616>.

表 3 工业缺陷检测常用的公开数据集

Table 3 Common open datasets for industrial defect detection

Task ^{a)}	Name	Scenario	Resolution	Size
Supervised CLS	elpv-dataset [173]	Solar panel	300×300	2624
	NEU-CLS [30]	Steel	200×200	1800
Supervised DET	TianChi Fabric [174]	Fabric	1920×2560	20000
	TianChi AP	Aluminium profile	1920×2560	4356
	NEU-DET [30]	Steel	200×200	1800
	GDXray Casting [175]	X-ray castings	$256 \times 256 - 768 \times 572$	2727
	DeepPCB [176]	PCB	640×640	1500
Supervised SEG	Magnetic Tile [40]	Magnetic Tile	614×248	1344
	KolektorSDD [39]	Metal	$1240 \times 500 - 1270 \times 500$	399
	RSDDs [177]	Rail steel	55×1250	4356
	CrackForest [19]	Road crack	480×320	118
Unsupervised SEG	MVTec AD [178]	Texture & object	$700 \times 700 - 1024 \times 1024$	5354
	AITEX [179]	Fabric	4096×256	245
	BTAD [180]	Texture & object	$600 \times 600 - 1600 \times 1600$	2830
	NanoTWICE [181]	SEM nanofibres	1024×696	45
Weakly-supervised SEG	DAGM [182]	Synthetic texture	512×512	1150

a) Classification, detection, and segmentation are abbreviated as CLS, DET, and SEG, respectively.

的契合不同任务设置的典型数据集。

天池布匹数据集^[174]由阿里云天池平台于 2019 年在天池布匹检测大赛上提出,用于有监督检测的任务设置。在布匹的实际生产流程中,会产生织稀、擦洞、毛斑等各种缺陷。数据集立足工业生产实际情况,包含简单的素色花布约 8000 张,复杂纹理的花色布约 12000 张,图片分辨率为 2560×1920 。共包含 10 种缺陷类型,所有缺陷数据均给出了矩形包围盒作为缺陷区域的标注。

KolektorSDD^[39]是 Kolektor 公司于 2019 年提出的面向有监督分割任务设置的工业缺陷检测数据集。该数据集收集了大量电子换向器的图像数据,包含 348 张无缺陷图片以及 52 张有缺陷图片,分辨率在 $1240 \times 500 - 1270 \times 500$ 之间。缺陷位于电子换向器表面,主要表现为微小的破损、裂纹等。所有图片均提供了像素级别的标注。

AITEX 织物数据集^[179]是由 AITEX 实验室于 2019 年提出的面向无监督分割任务的织物数据集。该数据集共包含 245 张高分辨率图像,分辨率均为 4096×256 。织物结构共有 7 种类别,每一类包含 20 个无缺陷样本。测试集由 105 张有缺陷样本构成,包含了 12 种典型的织物缺陷。相比于整图,缺陷的尺寸微小。对于缺陷样本,数据集均给出了像素级别的标注。

MVTec AD^[178]是 MVTec 公司于 2019 年提出的针对无监督分割任务设置的多类型工业产品数据集,共包含 15 种产品:5 种纹理类和 10 种物体类。每种产品也存在多种不同类型的缺陷。所有图像的分辨率均在 $700 \times 700 - 1024 \times 1024$ 之间,其中 3629 张正常图像用于训练与验证,1725 张图像用于测试。所有含缺陷的图片均提供了像素级别的标注,总共包含近 1900 个标注区域。

DAGM^[182]由德国模式识别协会于 2007 年提出,属于弱监督分割任务的数据集。所有数据均为合成数据,但其逼真程度不亚于从实际工业场景中获取的真实数据。数据集包含 6 类纹理,每类包含 1000 张无缺陷样本以及 150 张缺陷样本。其主要特点在于纹理背景复杂与缺陷微弱。图像的分辨率均

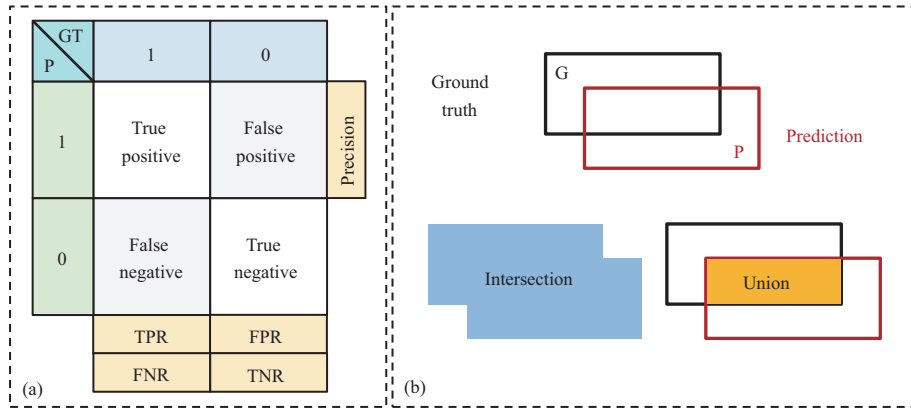


图 9 评估方法. (a) 二分类混淆矩阵; (b) 分割评价协议的可视化解释

Figure 9 Evaluation protocols. (a) Binary confusion matrix; (b) visual explanation of evaluation protocols for segmentation task

为 512×512 , 所有缺陷样本均给出了椭圆形的弱标注来表示缺陷的大致位置.

此外, 工业缺陷检测的模型也常会使用 CIFAR-10^[183], MINST 等小型分类数据集, 以及 STC^[184], UCSD-Pred2^[185] 和 CUHK Avenue^[186] 等视频异常数据集进行与 OOD 任务相关的性能评估.

5.2 评价指标

由于分类与分割通常在工业缺陷检测中具有较大的实用价值, 本文主要介绍这两类任务的评价指标.

5.2.1 分类指标

分类指标可以基于混淆矩阵 (confusion matrix) 计算. 图 9(a) 表示了二分类的混淆矩阵, 其中 1 代表异常类, 0 代表正常类. 多分类的混淆矩阵则按照相同的形式扩展. 分类的性能通常采用准确率 (accuracy)、精确率 (precision) 和召回率 (recall) 进行评估. 准确率指所有预测正确的类占全部的比重; 精确率也被称查准率, 指预测为缺陷的样本中正确预测所占的比例; 召回率也被称为查全率、真阳性率 (true positive rate, TPR), 指所有缺陷中被正确检出的比例. 此外, 还采用精确率和召回率的调和平均数 F1 值 (F1-measure), 综合两者对模型性能进行评估. 上述指标的计算方式如式 (1) 所示.

$$\begin{aligned}
 \text{Precision} &= \frac{TP}{TP + FP}, \\
 \text{Recall} = \text{TPR} &= \frac{TP}{TP + FN}, \\
 \text{F1-measure} &= \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \\
 \text{Accuracy} &= \frac{TP + TN}{TP + FP + TN + FN}.
 \end{aligned} \tag{1}$$

在实际应用中, 往往还会关注误检率 (false drop rate, FPR) 与漏检率 (false negatives rate, FNR), 它们衡量模型的不足之处, 用以帮助后续的改进, 其计算方式如式 (2) 所示. 通常模型在部署调试时会权衡误检率与漏检率.

$$\text{FPR} = \frac{FP}{FP + TN}, \quad \text{FNR} = \frac{FN}{FN + TP}. \tag{2}$$

使用式 (1) 和 (2) 需要事先设置分类阈值. 如果阈值的设置偏差较大, 评估的结果将与实际性能差异较大. 与阈值无关的指标不仅可以避免因阈值设置不当而导致评价偏差, 而且可以更加全面地评价模型, 帮助找到最优的工作点. 这类方法设置一系列阈值, 评价两个指标在不同阈值下的关系变化曲线. 在工业缺陷检测的分类任务中, 常用的曲线有两种: PR 曲线 (precision-recall curve) 描述了精确率和召回率的关系; ROC 曲线 (receiver operator characteristic curve) 描述了 TPR 与 FPR 的关系. 通常采用曲线下面积 (area under curve, AUC) 来度量相应曲线指标所描述的性能.

5.2.2 分割指标

像素级缺陷定位的任务可视为前景与背景的两类分割问题. 当输出的异常分数图是表示对应像素为缺陷概率的热力图时, 需要对其设定相应阈值获得二值化分割结果. 因此, 分割指标也可分为固定阈值的指标和与阈值无关的指标.

像素级评价指标可以视为图片级分类指标在像素层面中的体现. 此时, 是以像素为单位的二分类问题, 因而可以采用像素级的混淆矩阵来评估. 评价指标的计算方式与式 (1) 和 (2) 相同. 针对分割任务, 还可以从分割区域的角度评价方法的性能. 交并比 (intersection of union, IoU) 用来评价两个像素区域的重合情况, 一定程度上可以反映分割的位置与形状的准确性. 计算方式如式 (3) 所示, 其中物理量的含义如图 9(b) 所示. 因为 ROC 曲线中 TPR 值与缺陷面积存在加权关系, 所以一个较大的正确分割的区域会极大地弥补多个分割错误的小缺陷区域的量化指标, 因而该指标在工业缺陷检测中存在局限性. 对此, Bergmann 等^[137] 提出连通域级的指标 PRO (per region overlap). 如式 (3) 所示, 其首先将预测区域和 GT 区域按照连通域划分为 N 个区域, 然后依次计算每个真实缺陷区域 G_n 与缺陷区域 P 的交集, 以及该交集与该 GT 区域 G_n 的比值, 最后按照 N 个连通域加权平均. 这些指标也都可以根据相应的混淆矩阵算得.

$$\begin{aligned} \text{IoU} &= \frac{\text{Union}}{\text{Intersection}} = \frac{P \cap G}{P \cup G} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}, \\ \text{PRO} &= \frac{1}{N} \sum_{n=1}^N \frac{\text{Union}_n}{G_n} = \frac{1}{N} \sum_{n=1}^N \frac{P \cap G_n}{G_n} = \frac{1}{N} \sum_{n=1}^N \frac{\text{TP}_n}{\text{TP}_n + \text{FN}_n}. \end{aligned} \quad (3)$$

类似地, 阈值无关的分割指标同样为曲线指标. 除了前述的 PR 曲线与 ROC 曲线外, 还可采用 IoU 曲线和 PRO 曲线. 两者均为对应指标与 FPR 的关系曲线. 通过 IoU 曲线不仅可以检验模型性能, 度量模型对阈值的鲁棒性, 还有助于找到合适的工作点. 在 PRO 曲线度量下, 当 FPR 过高时, 大量的正常像素被检测为异常, 此时一个优秀的 PRO 值并不具有意义. 因此, 在用 AUC 评价 PRO 曲线时, 只选用令 FPR 处于 0-30% 之间的阈值, 并将 AUC 归一化为 PRO-Score.

5.3 性能评估

鉴于最近基于无监督的分割方法成为研究的焦点, 本文在最近提出的有关数据集 MVTec AD 上, 对代表性算法的性能予以归纳与比较, 其 ROCAUC 与 PRO-Score 指标分别如表 4 与 5²⁾所示.

2) 其中 MemAE 性能评估的数据来源于文献 [104], AnoGAN, AE-L2, AE-SSIM 与 CNN-Dict 的数据来源于文献 [178].

表 4 现有代表性算法在 MVTec AD^[178] 上的 ROCAUC 指标
Table 4 ROCAUC benchmark of typical existing algorithms on MVTec AD

Algorithm	Year	Description	Image-level ROCAUC (%)			Pixel-level ROCAUC (%)		
			Texture	Object	Mean	Texture	Object	Mean
AnoGAN [79]	2017	Reconstruction	54.2	56	55	57.6	82.6	74.0
AE-L2 [77]	2018	Reconstruction	—	—	—	69.6	87.8	81.7
AE-SSIM [77]	2018	Reconstruction	—	—	—	78.2	91.2	86.9
P-Net [92]	2020	Reconstruction	—	—	—	87.8	89.1	88.7
MemAE [100]	2019	Memory	89.03	78.26	81.85	85.17	86.03	85.74
TrustMAE [104]	2021	Memory	90.66	86.73	88.04	93.83	94.00	93.94
VE-VAE [94]	2019	VAE attention	—	—	—	80.6	88.8	86.07
CAVGA-Ru [95]	2020	VAE attention	—	—	—	88.6	89.9	89
RIAD [114]	2021	Restoration	95.1	89.9	91.7	93.9	94.3	94.2
SCADN [115]	2021	Restoration	78.12	83.7	81.8	71.14	77.18	75.2
Patch-SVDD [121]	2020	Deep OCC	94.6	90.9	92.1	93.7	96.7	95.7
FCDD [122]	2021	Deep OCC	—	—	—	92.8	91.0	92
CNN-Dict [129]	2018	Feature distance	—	—	—	80.4	76.9	78
SPADE [130]	2020	Feature distance	—	—	85.5	92.9	97.6	96.5
PaDiM [133]	2021	Feature distance	98.8	93.6	95.3	96.9	97.8	97.5
MKDAD [139]	2021	Feature distance	87.63	87.80	87.74	90.61	90.76	90.71
CutPaste [45]	2021	Self-supervised	97	94.3	95.2	96.3	95.8	96.0
DRAEM [161]	2021	Self-supervised	99.1	97.4	98.0	97.9	97.0	97.3

表 5 现有代表性算法在 MVTec AD^[178] 上的 PRO 指标 (%)
Table 5 PRO-score of typical existing algorithms on MVTec AD (%)

Algorithm	Year	Description	Texture	Object	Mean
AnoGAN [79]	2017	Reconstruction	27.4	53.3	44.3
AE-L2 [77]	2018	Reconstruction	69.6	83.8	79.0
AE-SSIM [77]	2018	Reconstruction	56.7	75.8	69.4
CNN-Dict [129]	2018	Feature distance	54.2	50.3	51.5
SPADE [130]	2020	Feature distance	88.38	93.4	91.7
PaDiM [133]	2021	Feature distance	93.14	91.59	92.1
U-Std [137]	2020	Feature distance	92.66	90.77	91.4

6 总结与展望

6.1 总结

作为工业视觉的核心技术, 基于视觉的工业缺陷检测因其广泛而重要的应用价值, 逐渐成为炙手可热的新兴研究问题之一. 近年来已经涌现出大量的新理论和新方法, 并已逐渐投入实际产业. 本文对该领域的任务定义、研究难点、检测方法、辅助技术、常用公开数据集、评价指标与典型方法性能等进行了综合梳理, 重点综述了深度学习方法的重要进展.

然而, 尽管目前工业缺陷检测方法已在部分工业数据上表现出较高的性能, 但现有研究依然存在很大的局限性, 限制了相关方法的进一步研究与落地。

6.1.1 更贴近实际的任务设置

基于有监督的缺陷检测方法已逐渐趋于成熟, 展现出了优越的性能。但是有监督方法依赖于大规模带有精确标注数据的支持, 而数据的搜集与标注过程成本高昂, 使之无法与实际工业场景完全吻合。同时, 有监督方法无法处理新类别缺陷的检测问题。无监督方法虽然大大降低了数据标注的需求, 也能检测未知型缺陷, 但是在定位性能与可解释性方面仍然无法替代有监督方法。

在实际生产环境中, 少量缺陷标注的情况更加常见。用户不仅可以提供大量的正常样本, 同时也能提供少量具有精确或粗糙标注的缺陷样本。相比于上述两种设置, 小样本、半监督和弱监督的方法更加贴合工业场景, 可以更加充分地利用资源, 因而其关注度迅速增加。目前, 这些方法仍然处于发展的初期, 仍有待进一步的研究与发展。

6.1.2 缺乏完备的数据集

数据集是深度学习方法研究的基础。目前常用于工业缺陷检测任务的数据集主要存在着背景类型单一、缺陷种类单一等不足, 无法模拟现实中复杂的检测场景。在实际工业生产过程中, 所拍摄的产品图像也可能来源于不同的成像条件、不同的拍摄角度或者不同制程, 在同一产品上也会出现不同类型的缺陷。构建更贴近实际工业场景与任务设置的数据集不仅有助于更加实用的算法的开发, 也有助于算法实际部署。例如, 多重成像的数据集可以将一种成像条件下难以被发现的缺陷在另一成像中凸显, 从而给缺陷发现带来了便利。工业缺陷检测方法在无人巡检、异常溯源等延伸性工作的性能也需要经过相应数据集的检验。因此, 新数据集的建立与完善也非常重要, 只有构造出合适的数据集, 才能支撑各种新场景、新任务和新设置下方法的发展。

6.2 展望

除了上述总结的目前研究中的不足, 该领域还有许多极具潜力的研究方向有待进一步探索。

6.2.1 神经架构搜索

现有的基于深度学习的方法大多依靠手工设计深度神经网络, 依照经验设置超参数。因此, 现有的模型结构可能并非面向工业缺陷检测的最优解。Rippel 等^[132]比较了多种基础网络对模型性能的影响, 其中基于自动机器学习 (automated machine learning, AutoML) 获得的 EfficientNet^[187]的性能优于人工设计的 ResNet^[188]。EfficientNet 高效的网络结构与 Swish 激活函数被认为是提升性能的主要原因。相比于其他通用的视觉检测任务, 大部分工业视觉的使用场景更加具体, 利用神经架构搜索 (neural architecture search, NAS) 的方法自动搜索针对特定任务场景的最优模型是一种可行的解决方案, 不仅有助于弥补人工经验设置的不足, 还能提升模型的效率。AutoOD^[189]基于 NAS 搜索 AE 架构的模型参数, 实现了优于手工模型的检测性能。然而, 目前 NAS 在无监督与半监督设置中的应用尚少, 未来可能在工业视觉领域的相关任务上得到发展。

6.2.2 Transformer 的应用

近年来, Transformer 模型^[190]被引入到计算机视觉领域, 并在多种视觉任务上展现出优越的性能^[191~195]。相比于 CNN 关注局部特征关系, 视觉 Transformer 利用注意力机制, 可以建模图像中的长距离关系。Xie 等^[196]将高效的 Swin Transformer^[194]改进后用于有监督的太阳能电池板的缺陷检测。

其增加窗间注意力来增强全局语义特征的提取, 用多尺度聚合模块结合不同层级的特征, 实现了较高的分类性能. 然而, Transformer 模型的性能优势通常需要足够的训练数据支撑, 而大部分工业缺陷检测场景难以获取充足的缺陷数据. 因此, 研究者们尝试在无监督设置下引入 Transformer, 一般将其用于基于图像相似度的方法. Mishra 等^[180] 将基于图像重建模型的编码器改为 Transformer 结构, 以增强全局特征的提取能力. 同时, 用 GMM 对 Transformer 的编码结果进行密度估计. 在测试阶段, 根据 GMM 对图像块的密度估计结果进行缺陷定位, 同时结合重建误差实现图像级分类. Pirnay 等^[157] 认为在基于图像复原任务的缺陷检测方法^[114] 中, 长距离信息有利于高质量的图像复原, 从而提升缺陷分割能力, 因而将 CNN 替换为 Transformer, 根据周围信息来复原图像块, 最终根据重建误差进行缺陷分割. Pinaya 等^[197] 首先训练 VQ-VAE^[198] 对图像进行离散编码, 然后利用多个高效的 Transformer 模型 Performer^[195] 对不同顺序的离散编码进行自回归建模. 最终基于重建误差和隐空间编码的似然进行异常分割, 从而减少误检.

Transformer 具有全局关系建模的能力, 且具有更好的可解释性^[199]. 但是其对计算资源的需求也相应增大, 对工业场景的低开销需求提出挑战. 目前也有相关工作继续在高效 Transformer 模型^[194, 195] 与小数据集^[199] 的应用上进行探索. 尽管目前 Transformer 在工业缺陷检测领域的实用性还尚未得到充分探索, 但现有研究已经指出了 Transformer 的潜力.

6.2.3 多重成像与多模态

早期的工业缺陷检测方法利用光学方法提升缺陷的显著性, 从而使其易于被检出. 事实上, 工业缺陷的种类千差万别, 可能存在各自适合的成像方式, 因此, 利用不同成像条件的差异产生区分力是一个具有潜力的方案. 在实际场景下, 红外线、X 射线等其他成像方式可能适合特定产品的质检任务; 结构光、激光等方式获得的 3D 信息还能进一步提升产品结构性缺陷的表达能力. 目前大部分的方法仅关注于单一的 RGB、灰度图像或其他成像形式的输入. 为了综合不同成像条件的感知能力, 模型可以进一步考虑融合多模态数据的特征来增强检测性能, 以应对更复杂的场景.

6.2.4 更具可解释性的方法

尽管深度学习极大地促进了工业缺陷检测的发展, 新兴的方法拥有卓越的检测性能, 但是不少方法仍缺乏完备的可解释性. 许多基于无监督的方法还依赖大量的经验性假设. 对模型解释性的研究不仅有助于人们理解模型工作机制, 还能促进新型方法的研发. 从辅助人工检测的角度而言, 建立人机互信在实际应用落地的推进中至关重要. 因此, 置信度预测^[200] 和模型可视化等模型理解的研究不只在工业缺陷检测领域, 也在整个人工智能领域极具意义.

6.2.5 在线学习与联邦学习

现有方法大多采用离线学习的方法, 即预先根据对应所给数据集训练模型, 然后部署到产线中. 但由于训练初期通常缺乏有标签的缺陷样本, 部署时模型的实际性能难达最佳. 在实际应用中, 产线可以不断地提供新样本, 也可能发现新的缺陷类型. 因此, 运用在线学习的方法对模型进行快速实时的调整也将成为后续研究与应用的重点之一. 由于某些产品及其缺陷存在一定的共性特征, 结合不同生产线上的样本有助于充分训练与调整模型. 考虑到工业数据往往容易涉及隐私问题, 引入联邦学习的框架有助于在保障各个合作者隐私安全的前提下, 有效地结合各方的样本数据.

参考文献

- 1 Chandola V, Banerjee A, Kumar V. Anomaly detection: a survey. *ACM Comput Surv*, 2009, 41: 1–58
- 2 Pang G, Shen C, Cao L, et al. Deep learning for anomaly detection: a review. 2020. ArXiv:2007.02500
- 3 Chalapathy R, Chawla S. Deep learning for anomaly detection: a survey. 2019. ArXiv:1901.03407
- 4 Bulusu S, Kailkhura B, Li B, et al. Anomalous instance detection in deep learning: a survey. In: *Proceedings of the 42nd IEEE Symposium on Security and Privacy Princeton*, 2020
- 5 Ruff L, Kauffmann J R, Vandermeulen R A, et al. A unifying review of deep and shallow anomaly detection. *Proc IEEE*, 2021, 109: 756–795
- 6 Czimmermann T, Ciuti G, Milazzo M, et al. Visual-based defect detection and classification approaches for industrial applications—a survey. *Sensors*, 2020, 20: 1459
- 7 Zhou D H, Wei M H, Si X S. A survey on anomaly detection, life prediction and maintenance decision for industrial processes. *Acta Automatica Sin*, 2013, 39: 711–722
- 8 Tao X, Hou W, Xu D. A survey of surface defect detection methods based on deep learning. *Acta Autom Sin*, 2021, 47: 1017–1034
- 9 Hawkins D M. *Identification of Outliers*. London: Chapman and Hall, 1980
- 10 Cook A A, Misirlı G, Fan Z. Anomaly detection for IoT time-series data: a survey. *IEEE Internet Things J*, 2020, 7: 6481–6494
- 11 Zhou B, Khosla A, Lapedriza A, et al. Learning deep features for discriminative localization. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016. 2921–2929
- 12 Selvaraju R R, Cogswell M, Das A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization. In: *Proceedings of the IEEE International Conference on Computer Vision*, 2017. 618–626
- 13 Smilkov D, Thorat N, Kim B, et al. Smoothgrad: removing noise by adding noise. 2017. ArXiv:1706.03825
- 14 Zuo B, Wang F L. Surface cutting defect detection of magnet using Fourier image reconstruct. *Comput Eng Appl*, 2016, 52: 256–260
- 15 Chetverikov D, Hanbury A. Finding defects in texture using regularity and local orientation. *Pattern Recogn*, 2002, 35: 2165–2180
- 16 Hou Z, Parker J M. Texture defect detection using support vector machines with adaptive Gabor wavelet features. In: *Proceedings of the 7th IEEE Workshops on Applications of Computer Vision*, 2005. 1: 275–280
- 17 Peng X, Chen Y, Yu W, et al. An online defects inspection method for float glass fabrication based on machine vision. *Int J Adv Manuf Technol*, 2008, 39: 1180–1189
- 18 Murino V, Bicego M, Rossi I A. Statistical classification of raw textile defects. In: *Proceedings of the 17th International Conference on Pattern Recognition*, 2004. 4: 311–314
- 19 Shi Y, Cui L, Qi Z, et al. Automatic road crack detection using random structured forests. *IEEE Trans Intell Transp Syst*, 2016, 17: 3434–3445
- 20 Deng Y S, Luo A C, Dai M J. Building an automatic defect verification system using deep neural network for PCB defect classification. In: *Proceedings of the 4th International Conference on Frontiers of Signal Processing (ICFSP)*, 2018. 145–149
- 21 Liang Q, Zhu W, Sun W, et al. In-line inspection solution for codes on complex backgrounds for the plastic container industry. *Measurement*, 2019, 148: 106965
- 22 Deitsch S, Christlein V, Berger S, et al. Automatic classification of defective photovoltaic module cells in electroluminescence images. *Sol Energy*, 2019, 185: 455–468
- 23 Zhang Z, Wen G, Chen S. Weld image deep learning-based on-line defects detection using convolutional neural networks for Al alloy in robotic arc welding. *J Manuf Processes*, 2019, 45: 208–216
- 24 Ma L, Xie W, Zhang Y. Blister defect detection based on convolutional neural network for polymer lithium-ion battery. *Appl Sci*, 2019, 9: 1085
- 25 Cha Y J, Choi W, Büyüköztürk O. Deep learning-based crack damage detection using convolutional neural networks. *Comput-Aided Civil Infrastruct Eng*, 2017, 32: 361–378
- 26 Xu X, Zheng H, Guo Z, et al. SDD-CNN: small data-driven convolution neural networks for subtle roller defect inspection. *Appl Sci*, 2019, 9: 1364
- 27 Cha Y J, Choi W, Suh G, et al. Autonomous structural visual inspection using region-based deep learning for

- detecting multiple damage types. *Comput-Aided Civil Infrastruct Eng*, 2018, 33: 731–747
- 28 Ren S, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks. In: *Proceedings of Advances in Neural Information Processing Systems*, 2015. 28: 91–99
- 29 Tao X, Zhang D, Wang Z, et al. Detection of power line insulator defects using aerial images analyzed with convolutional neural networks. *IEEE Trans Syst Man Cybern Syst*, 2020, 50: 1486–1498
- 30 He Y, Song K, Meng Q, et al. An end-to-end steel surface defect detection approach via fusing multiple hierarchical features. *IEEE Trans Instrum Meas*, 2020, 69: 1493–1504
- 31 Li J, Su Z, Geng J, et al. Real-time detection of steel strip surface defects based on improved YOLO detection network. *IFAC-PapersOnLine*, 2018, 51: 76–81
- 32 Zhang C, Chang C, Jamshidi M. Concrete bridge surface damage detection using a single-stage detector. *Comput-Aided Civil Infrastruct Eng*, 2020, 35: 389–409
- 33 Redmon J, Farhadi A. Yolov3: an incremental improvement. 2018. ArXiv:1804.02767
- 34 Chen S H, Tsai C C. SMD LED chips defect detection using a YOLOv3-dense model. *Adv Eng Inf*, 2021, 47: 101255
- 35 Huang G, Liu Z, van der Maaten L, et al. Densely connected convolutional networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017. 4700–4708
- 36 Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015. 3431–3440
- 37 Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: *Proceedings of International Conference on Medical Image Computing and Computer-assisted Intervention*. Cham: Springer, 2015. 234–241
- 38 Qiu L, Wu X, Yu Z. A high-efficiency fully convolutional networks for pixel-wise surface defect detection. *IEEE Access*, 2019, 7: 15884–15893
- 39 Tabernik D, Šelaj S, Skvarč J, et al. Segmentation-based deep-learning approach for surface-defect detection. *J Intell Manuf*, 2020, 31: 759–776
- 40 Huang Y, Qiu C, Yuan K. Surface defect saliency of magnetic tile. *Vis Comput*, 2020, 36: 85–96
- 41 Xie Y, Zhu F, Fu Y. Main-secondary network for defect segmentation of textured surface images. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2020. 3531–3540
- 42 Valente A, Wada C, Neves D, et al. Print defect mapping with semantic segmentation. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2020. 3551–3559
- 43 Zhang G, Cui K, Hung T Y, et al. Defect-GAN: high-fidelity defect synthesis for automated defect inspection. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021. 2524–2534
- 44 DeVries T, Taylor G W. Improved regularization of convolutional neural networks with cutout. 2017. ArXiv:1708.04552
- 45 Li C L, Sohn K, Yoon J, et al. CutPaste: self-supervised learning for anomaly detection and localization. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021. 9664–9674
- 46 Harris C, Stephens M. A combined corner and edge detector. In: *Proceedings of Alvey Vision Conference*, 1988. 147–151
- 47 Lowe D G. Distinctive image features from scale-invariant keypoints. *Int J Comput Vision*, 2004, 60: 91–110
- 48 Wang Z, Wu F, Hu Z. MSLD: a robust descriptor for line matching. *Pattern Recogn*, 2009, 42: 941–953
- 49 Zhang L L, Koch R. An efficient and robust line segment matching approach based on LBD descriptor and pairwise geometric consistency. *J Visual Commun Image Represent*, 2013, 24: 794–805
- 50 Nistér D, Stewénius H. Linear time maximally stable extremal regions. In: *Proceedings of the 10th European Conference on Computer Vision*. Marseille: Springer, 2008. 183–196
- 51 Xu Y C, Monasse P, Géraud T, et al. Tree-based morse regions: a topological approach to local feature detection. *IEEE Trans Image Process*, 2014, 23: 5612–5625
- 52 Movafeghi A, Mohammadzadeh N, Yahaghi E, et al. Defect detection of industrial radiography images of ammonia pipes by a sparse coding model. *J Nondestruct Eval*, 2018, 37: 3
- 53 Tong L, Wong W K, Kwong C K. Fabric defect detection for apparel industry: a nonlocal sparse representation approach. *IEEE Access*, 2017, 5: 5947–5964
- 54 Liang L Q, Li D, Fu X, et al. Touch screen defect inspection based on sparse representation in low resolution images.

- Multimed Tools Appl, 2016, 75: 2655–2666
- 55 Zhou W, Fei M, Zhou H, et al. A sparse representation based fast detection method for surface defect detection of bottle caps. *Neurocomputing*, 2014, 123: 406–414
 - 56 Hou Y, Cuyt A, Lee W S, et al. Decomposing textures using exponential analysis. In: *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, 2021
 - 57 Hocenski Z, Vasilic S, Hocenski V. Improved Canny edge detector in ceramic tiles defect detection. In: *Proceedings of the 32nd Annual Conference on IEEE Industrial Electronics*, 2006. 3328–3331
 - 58 Mak K L, Peng P, Yiu K F C. Fabric defect detection using morphological filters. *Image Vision Comput*, 2009, 27: 1585–1592
 - 59 Yang E J, Liao Y H, Liu A D, et al. Detection for fabric defects based on low-rank decomposition. *J Textile Res*, 2020, 41: 72–78
 - 60 Li C, Liu C, Gao G, et al. Robust low-rank decomposition of multi-channel feature matrices for fabric defect detection. *Multimed Tools Appl*, 2019, 78: 7321–7339
 - 61 Shi B, Liang J, Di L, et al. Fabric defect detection via low-rank decomposition with gradient information. *IEEE Access*, 2019, 7: 130423
 - 62 Li C, Gao G, Liu Z, et al. Defect detection for patterned fabric images based on GHOG and low-rank decomposition. *IEEE Access*, 2019, 7: 83962–83973
 - 63 Wang J, Xu G, Li C, et al. Surface defects detection using non-convex total variation regularized RPCA with kernelization. *IEEE Trans Instrum Meas*, 2021, 70: 1–13
 - 64 Schölkopf B, Williamson R C, Smola A J, et al. Support vector method for novelty detection. In: *Proceedings of Advances in Neural Information Processing Systems*. Breckenridge: MIT Press, 2000. 582–588
 - 65 Tax D M J, Duin R P W. Support vector data description. *Mach Learn*, 2004, 54: 45–66
 - 66 Lee K Y, Kim D W, Lee K H, et al. Density-induced support vector data description. *IEEE Trans Neural Netw*, 2007, 18: 284–289
 - 67 Huang G, Chen H, Zhou Z, et al. Two-class support vector data description. *Pattern Recogn*, 2011, 44: 320–329
 - 68 Juszczak P, Tax D, Duin R P W. Feature scaling in support vector data description. In: *Proceedings of the 8th Annual Conference of the Advanced School for Computing and Imaging*, 2002. 95–102
 - 69 Hinton G E, Zemel R S. Autoencoders, minimum description length, and Helmholtz free energy. In: *Proceedings of the 6th International Conference on Neural Information Processing Systems*, 1994. 6: 3–10
 - 70 Kingma D P, Welling M. Auto-encoding variational Bayes. 2013. ArXiv:1312.6114
 - 71 Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. In: *Proceedings of the 27th International Conference on Neural Information Processing Systems*, 2014. 2672–2680
 - 72 Yang H, Chen Y, Song K, et al. Multiscale feature-clustering-based fully convolutional autoencoder for fast accurate visual inspection of texture surface defects. *IEEE Trans Automat Sci Eng*, 2019, 16: 1450–1467
 - 73 Yang J, Shi Y, Qi Z. DFR: deep feature reconstruction for unsupervised anomaly segmentation. 2020. ArXiv:2012.07122
 - 74 Chung H, Park J, Keum J, et al. Unsupervised anomaly detection using style distillation. *IEEE Access*, 2020, 8: 221494
 - 75 Dehaene D, Frigo O, Combexelle S, et al. Iterative energy-based projection on a normal data manifold for anomaly localization. 2020. ArXiv:2002.03734
 - 76 Zhou Q, Mei J, Zhang Q, et al. Semi-supervised fabric defect detection based on image reconstruction and density estimation. *Textile Res J*, 2021, 91: 962–972
 - 77 Bergmann P, Löwe S, Fauser M, et al. Improving unsupervised defect segmentation by applying structural similarity to autoencoders. 2018. ArXiv:1807.02011
 - 78 Abati D, Porrello A, Calderara S, et al. Latent space autoregression for novelty detection. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019. 481–490
 - 79 Schlegl T, Seeböck P, Waldstein S M, et al. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In: *Proceedings of International Conference on Information Processing in Medical Imaging*. Cham: Springer, 2017. 146–157
 - 80 Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial

- networks. 2015. ArXiv:1511.06434
- 81 Zenati H, Foo C S, Lecouat B, et al. Efficient GAN-based anomaly detection. 2018. ArXiv:1802.06222
- 82 Schlegl T, Seeböck P, Waldstein S M, et al. f-AnoGAN: fast unsupervised anomaly detection with generative adversarial networks. *Med Image Anal*, 2019, 54: 30–44
- 83 Gulrajani I, Ahmed F, Arjovsky M, et al. Improved training of wasserstein GANs. 2017. ArXiv:1704.00028
- 84 Hu G, Huang J, Wang Q, et al. Unsupervised fabric defect detection based on a deep convolutional generative adversarial network. *Textile Res J*, 2020, 90: 247–270
- 85 Komoto K, Aizawa H, Kato K. Consistency ensured bi-directional GAN for anomaly detection. In: *Proceedings of International Workshop on Frontiers of Computer Vision*. Singapore: Springer, 2020. 236–247
- 86 Donahue J, Krähenbühl P, Darrell T. Adversarial feature learning. 2016. ArXiv:1605.09782
- 87 Sabokrou M, Khalooei M, Fathy M, et al. Adversarially learned one-class classifier for novelty detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 3379–3388
- 88 Baur C, Wiestler B, Albarqouni S, et al. Deep autoencoding models for unsupervised anomaly segmentation in brain MR images. In: *Proceedings of International MICCAI Brainlesion Workshop*. Cham: Springer, 2018. 161–169
- 89 Akcay S, Atapour-Abarghouei A, Breckon T P. Ganomaly: semi-supervised anomaly detection via adversarial training. In: *Proceedings of Asian Conference on Computer Vision*. Cham: Springer, 2018. 622–637
- 90 Akçay S, Atapour-Abarghouei A, Breckon T P. Skip-ganomaly: skip connected and adversarially trained encoder-decoder anomaly detection. In: *Proceedings of International Joint Conference on Neural Networks (IJCNN)*, 2019. 1–8
- 91 Tang T W, Kuo W H, Lan J H, et al. Anomaly detection neural network with dual auto-encoders GAN and its industrial inspection applications. *Sensors*, 2020, 20: 3336
- 92 Zhou K, Xiao Y, Yang J, et al. Encoding structure-texture relation with P-Net for anomaly detection in retinal images. In: *Proceedings of the 16th European Conference on Computer Vision*, Glasgow, 2020. 360–377
- 93 Berthelot D, Schumm T, Metz L. BEGAN: boundary equilibrium generative adversarial networks. 2017. ArXiv:1703.10717
- 94 Liu W, Li R, Zheng M, et al. Towards visually explaining variational autoencoders. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020. 8642–8651
- 95 Venkataramanan S, Peng K C, Singh R V, et al. Attention guided anomaly localization in images. In: *Proceedings of European Conference on Computer Vision*. Cham: Springer, 2020. 485–503
- 96 Kimura D, Chaudhury S, Narita M, et al. Adversarial discriminative attention for robust anomaly detection. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2020. 2172–2181
- 97 Pidhorskyi S, Almohsen R, Adjero D A, et al. Generative probabilistic novelty detection with adversarial autoencoders. In: *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, 2018. 6823–6834
- 98 Perera P, Nallapati R, Xiang B. OCGAN: one-class novelty detection using GANs with constrained latent representations. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019. 2898–2906
- 99 Hong E, Choe Y. Latent feature decentralization loss for one-class anomaly detection. *IEEE Access*, 2020, 8: 165658
- 100 Gong D, Liu L, Le V, et al. Memorizing normality to detect anomaly: memory-augmented deep autoencoder for unsupervised anomaly detection. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019. 1705–1714
- 101 Yang Y, Xiang S, Zhang R. Improving unsupervised anomaly localization by applying multi-scale memories to autoencoders. 2020. ArXiv:2012.11113
- 102 Hou J, Zhang Y, Zhong Q, et al. Divide-and-assemble: learning block-wise memory for unsupervised anomaly detection. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021. 8791–8800
- 103 Park H, Noh J, Ham B. Learning memory-guided normality for anomaly detection. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020. 14372–14381
- 104 Tan D S, Chen Y C, Chen T P C, et al. TrustMAE: a noise-resilient defect classification framework using memory-augmented auto-encoders with trust regions. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021. 276–285
- 105 Ye F, Huang C, Cao J K, et al. Attribute restoration framework for anomaly detection. *IEEE Trans Multimedia*,

- 2022, 24: 116–127
- 106 Salehi M, Eftekhari A, Sadjadi N, et al. Puzzle-AE: novelty detection in images through solving puzzles. 2020. ArXiv:2008.12959
 - 107 Noroozi M, Favaro P. Unsupervised learning of visual representations by solving Jigsaw puzzles. In: Proceedings of European Conference on Computer Vision. Cham: Springer, 2016. 69–84
 - 108 Haselmann M, Gruber D P, Tabatabai P. Anomaly detection using deep learning based image completion. In: Proceedings of the 17th IEEE International Conference on Machine Learning and Applications (ICMLA), 2018. 1237–1242
 - 109 Li Z, Li N, Jiang K, et al. Superpixel masking and inpainting for self-supervised anomaly detection. In: Proceedings of the 31st British Machine Vision Conference, 2020. 7–10
 - 110 Collin A S, de Vleeschouwer C. Improved anomaly detection by training an autoencoder with skip connections on images corrupted with stain-shaped noise. In: Proceedings of the 25th International Conference on Pattern Recognition (ICPR), 2021. 7915–7922
 - 111 Gal Y, Ghahramani Z. Dropout as a Bayesian approximation: representing model uncertainty in deep learning. In: Proceedings of International Conference on Machine Learning, 2016. 1050–1059
 - 112 Yang H, Zhou Q, Song K, et al. An anomaly feature-editing-based adversarial network for texture defect visual inspection. IEEE Trans Ind Inf, 2021, 17: 2220–2230
 - 113 Mei S, Yang H, Yin Z. An unsupervised-learning-based approach for automated defect inspection on textured surfaces. IEEE Trans Instrum Meas, 2018, 67: 1266–1277
 - 114 Zavrtanik V, Kristan M, Skočaj D. Reconstruction by inpainting for visual anomaly detection. Pattern Recogn, 2021, 112: 107706
 - 115 Yan X, Zhang H, Xu X, et al. Learning semantic context from normal samples for unsupervised anomaly detection. In: Proceedings of the AAAI Conference on Artificial Intelligence, 2021. 35: 3110–3118
 - 116 Erfani S M, Rajasegarar S, Karunasekera S, et al. High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning. Pattern Recogn, 2016, 58: 121–134
 - 117 Chalapathy R, Menon A K, Chawla S. Anomaly detection using one-class neural networks. 2018. ArXiv:1802.06360
 - 118 Ruff L, Vandermeulen R, Goernitz N, et al. Deep one-class classification. In: Proceedings of International Conference on Machine Learning. Stockholm: ACM, 2018. 4393–4402
 - 119 Kitamura S, Nonaka Y. Explainable anomaly detection via feature-based localization. In: Proceedings of International Conference on Artificial Neural Networks. Munich: Springer, 2019. 408–419
 - 120 Perera P, Patel V M. Learning deep features for one-class classification. IEEE Trans Image Process, 2019, 28: 5450–5463
 - 121 Yi J, Yoon S. Patch SVDD: patch-level SVDD for anomaly detection and segmentation. In: Proceedings of the Asian Conference on Computer Vision, 2020
 - 122 Liznerski P, Ruff L, Vandermeulen R A, et al. Explainable deep one-class classification. In: Proceedings of International Conference on Learning Representations, 2020
 - 123 Wu P, Liu J, Shen F. A deep one-class neural network for anomalous event detection in complex scenes. IEEE Trans Neural Netw Learning Syst, 2019, 31: 2609–2622
 - 124 Massoli F V, Falchi F, Kantarci A, et al. MOCCA: multi-layer one-class classification for anomaly detection. 2020. ArXiv:2012.12111
 - 125 Golan I, El-Yaniv R. Deep anomaly detection using geometric transformations. In: Proceedings of the 32nd International Conference on Neural Information Processing Systems, 2018. 9781–9791
 - 126 Gidaris S, Singh P, Komodakis N. Unsupervised representation learning by predicting image rotations. In: Proceedings of International Conference on Learning Representations, 2018
 - 127 Bergman L, Hoshen Y. Classification-based anomaly detection for general data. In: Proceedings of International Conference on Learning Representations, 2019
 - 128 Pourreza M, Mohammadi B, Khaki M, et al. G2d: generate to detect anomaly. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2021. 2003–2012
 - 129 Napoletano P, Piccoli F, Schettini R. Anomaly detection in nanofibrous materials by CNN-based self-similarity. Sensors, 2018, 18: 209

- 130 Cohen N, Hoshen Y. Sub-image anomaly detection with deep pyramid correspondences. 2020. ArXiv:2005.02357
- 131 Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks. In: Proceedings of Advances in Neural Information Processing Systems, 2012. 25: 1097–1105
- 132 Rippel O, Mertens P, Merhof D. Modeling the distribution of normal data in pre-trained deep features for anomaly detection. In: Proceedings of the 25th International Conference on Pattern Recognition (ICPR), 2021. 6726–6733
- 133 Defard T, Setkov A, Loesch A, et al. PaDiM: a patch distribution modeling framework for anomaly detection and localization. In: Proceedings of International Conference on Pattern Recognition. Cham: Springer, 2021. 475–489
- 134 Rudolph M, Wandt B, Rosenhahn B. Same same but DifferNet: semi-supervised defect detection with normalizing flows. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2021. 1907–1916
- 135 Rezende D, Mohamed S. Variational inference with normalizing flows. In: Proceedings of International Conference on Machine Learning, 2015. 1530–1538
- 136 Kirichenko P, Izmailov P, Wilson A G. Why normalizing flows fail to detect out-of-distribution data. In: Proceedings of Advances in Neural Information Processing Systems, 2020. 33
- 137 Bergmann P, Fauser M, Sattlegger D, et al. Uninformed students: student-teacher anomaly detection with discriminative latent embeddings. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020. 4183–4192
- 138 Xiao Q, Wang J, Lin Y, et al. Unsupervised anomaly detection with distilled teacher-student network ensemble. Entropy, 2021, 23: 201
- 139 Salehi M, Sadjadi N, Baselizadeh S, et al. Multiresolution knowledge distillation for anomaly detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021. 14902–14912
- 140 Li T, Wang Z, Liu S, et al. Deep unsupervised anomaly detection. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2021. 3636–3645
- 141 Tao X, Wang Z, Zhang Z, et al. Wire defect recognition of spring-wire socket using multitask convolutional neural networks. IEEE Trans Compon Packag Manufact Technol, 2018, 8: 689–698
- 142 Lin D, Cao Y, Zhu W, et al. Few-shot defect segmentation leveraging abundant normal training samples through normal background regularization and crop-and-paste operation. 2020. ArXiv:2007.09438
- 143 Gong Y, Shao H, Luo J, et al. A deep transfer learning model for inclusion defect detection of aeronautics composite materials. Composite Struct, 2020, 252: 112681
- 144 Imoto K, Nakai T, Ike T, et al. A CNN-based transfer learning method for defect classification in semiconductor manufacturing. In: Proceedings of International Symposium on Semiconductor Manufacturing (ISSM), 2018. 1–3
- 145 Jing J F, Ma H, Zhang H H. Automatic fabric defect detection using a deep convolutional neural network. Coloration Technol, 2019, 135: 213–223
- 146 Lu Y, Yu F, Reddy M K K, et al. Few-shot scene-adaptive anomaly detection. In: Proceedings of European Conference on Computer Vision. Cham: Springer, 2020. 125–141
- 147 Finn C, Abbeel P, Levine S. Model-agnostic meta-learning for fast adaptation of deep networks. In: Proceedings of International Conference on Machine Learning, 2017. 1126–1135
- 148 Di H, Ke X, Peng Z, et al. Surface defect classification of steels with a new semi-supervised learning method. Optics Lasers Eng, 2019, 117: 40–48
- 149 Odena A. Semi-supervised learning with generative adversarial networks. 2016. ArXiv:1606.01583
- 150 Gao Y, Gao L, Li X, et al. A semi-supervised convolutional neural network-based method for steel surface defect recognition. Robotics Comput-Integrated Manufact, 2020, 61: 101825
- 151 He Y, Song K, Dong H, et al. Semi-supervised defect classification of steel surface based on multi-training and generative adversarial network. Optics Lasers Eng, 2019, 122: 294–302
- 152 Chu W H, Kitani K M. Neural batch sampling with reinforcement learning for semi-supervised anomaly detection. In: Proceedings of the 16th European Conference on Computer Vision, Glasgow, 2020. 751–766
- 153 Lin D, Li Y, Prasad S, et al. CAM-UNET: class activation MAP guided UNET with feedback refinement for defect segmentation. In: Proceedings of 2020 IEEE International Conference on Image Processing (ICIP), 2020. 2131–2135
- 154 Niu S, Lin H, Niu T, et al. DefectGAN: weakly-supervised defect detection using generative adversarial network. In: Proceedings of IEEE 15th International Conference on Automation Science and Engineering (CASE), 2019. 127–132
- 155 Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks.

- In: Proceedings of the IEEE International Conference on Computer Vision, 2017. 2223–2232
- 156 Mayr M, Hoffmann M, Maier A, et al. Weakly supervised segmentation of cracks on solar cells using normalized L_p norm. In: Proceedings of IEEE International Conference on Image Processing (ICIP), 2019. 1885–1889
 - 157 Pirnay J, Chai K. Inpainting transformer for anomaly detection. 2021. ArXiv:2104.13897
 - 158 Ristea N C, Madan N, Ionescu R T, et al. Self-supervised predictive convolutional attentive block for anomaly detection. 2021. ArXiv:2111.09099 2021
 - 159 Tayeh T, Aburakhia S, Myers R, et al. Distance-based anomaly detection for industrial surfaces using triplet networks. In: Proceedings of the 11th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 2020. 372–377
 - 160 Schlüter H M, Tan J, Hou B, et al. Self-supervised out-of-distribution detection and localization with natural synthetic anomalies (NSA). 2021. ArXiv:2109.15222
 - 161 Zavrtanik V, Kristan M, Skocaj D. DRAEM-A discriminatively trained reconstruction embedding for surface anomaly detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021. 8330–8339
 - 162 Song J, Kong K, Park Y I, et al. AnoSeg: anomaly segmentation network using self-supervised learning. 2021. ArXiv:2110.03396
 - 163 Xia X, Pan X, He X, et al. Discriminative-generative representation learning for one-class anomaly detection. 2021. ArXiv:2107.12753
 - 164 Jezequel L, Vu N S, Beaudet J, et al. Efficient anomaly detection using self-supervised multi-cue tasks. 2021. ArXiv:2111.12379
 - 165 Sheynin S, Benaim S, Wolf L. A hierarchical transformation-discriminating generative model for few shot anomaly detection. 2021. ArXiv:2104.14535
 - 166 Shaham T R, Dekel T, Michaeli T. SinGAN: learning a generative model from a single natural image. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019. 4570–4580
 - 167 Howard A G, Zhu M, Chen B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications. 2017. ArXiv:1704.04861
 - 168 Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016. 2818–2826
 - 169 Jaderberg M, Vedaldi A, Zisserman A. Speeding up convolutional neural networks with low rank expansions. 2014. ArXiv:1405.3866
 - 170 Han S, Pool J, Tran J, et al. Learning both weights and connections for efficient neural network. In: Proceedings of Advances in Neural Information Processing Systems (NIPS). Montreal: MIT Press, 2015. 1135–1143
 - 171 Bailer C, Habtegebrial T A, Varanasi K, et al. Fast dense feature extraction with CNNs that have pooling or striding layers. In: Proceedings of British Machine Vision Conference (BMVC), 2017
 - 172 Bergmann P, Batzner K, Fauser M, et al. The MVTec anomaly detection dataset: a comprehensive real-world dataset for unsupervised anomaly detection. Int J Comput Vis, 2021, 129: 1038–1059
 - 173 Buerhop-Lutz C, Deutsch S, Maier A, et al. A benchmark for visual identification of defective solar cells in electroluminescence imagery. In: Proceedings of the 35th European PV Solar Energy Conference and Exhibition, 2018. 1287–1289
 - 174 Cai K, Ma L. User behavior data analysis of Taobao online based on flink-based K-means algorithm. In: Proceedings of International Conference on Applications and Techniques in Cyber Security and Intelligence. Cham: Springer, 2020. 852–859
 - 175 Hao Z, Notghi B, Vandenbossche J M, et al. Optimal wave propagation-based nondestructive test design for quantitative damage characterization. J Nondestruct Eval, 2015, 34: 12
 - 176 Tang S, He F, Huang X, et al. Online PCB defect detector on a new PCB defect dataset. 2019. ArXiv:1902.06197
 - 177 Gan J, Li Q, Wang J, et al. A hierarchical extractor-based visual rail surface inspection system. IEEE Sens J, 2017, 17: 7935–7944
 - 178 Bergmann P, Fauser M, Sattlegger D, et al. MVTec AD—a comprehensive real-world dataset for unsupervised anomaly detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019. 9592–9600
 - 179 Silvestre-Blanes J, Albero-Albero T, Miralles I, et al. A public fabric database for defect detection methods and

- results. *Autex Res J*, 2019, 19: 363–374
- 180 Mishra P, Verk R, Fornasier D, et al. VT-ADL: a vision transformer network for image anomaly detection and localization. 2021. ArXiv:2104.10036
- 181 Carrera D, Manganini F, Boracchi G, et al. Defect detection in SEM images of nanofibrous materials. *IEEE Trans Ind Inf*, 2017, 13: 551–561
- 182 Wang T, Chen Y, Qiao M, et al. A fast and robust convolutional neural network-based defect detection model in product quality control. *Int J Adv Manuf Technol*, 2018, 94: 3465–3471
- 183 Krizhevsky A, Hinton G. Learning multiple layers of features from tiny images. In: *Handbook of Systemic Autoimmune Diseases*, 2009
- 184 Liu W, Luo W, Lian D, et al. Future frame prediction for anomaly detection-a new baseline. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 6536–6545
- 185 Mahadevan V, Li W, Bhalodia V, et al. Anomaly detection in crowded scenes. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010. 1975–1981
- 186 Lu C, Shi J, Jia J. Abnormal event detection at 150 FPS in MATLAB. In: *Proceedings of the IEEE International Conference on Computer Vision*, 2013. 2720–2727
- 187 Tan M, Le Q. EfficientNet: rethinking model scaling for convolutional neural networks. In: *Proceedings of International Conference on Machine Learning*, 2019. 6105–6114
- 188 He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016. 770–778
- 189 Li Y, Chen Z, Zha D, et al. AutoOD: automated outlier detection via curiosity-guided search and self-imitation learning. 2020. ArXiv:2006.11321
- 190 Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. In: *Proceedings of Advances in Neural Information Processing Systems*, 2017. 5998–6008
- 191 Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16×16 words: transformers for image recognition at scale. 2020. ArXiv:2010.11929
- 192 Carion N, Massa F, Synnaeve G, et al. End-to-end object detection with transformers. In: *Proceedings of European Conference on Computer Vision*. Cham: Springer, 2020. 213–229
- 193 Zheng S, Lu J, Zhao H, et al. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021. 6881–6890
- 194 Liu Z, Lin Y, Cao Y, et al. Swin transformer: hierarchical vision transformer using shifted windows. 2021. ArXiv:2103.14030
- 195 Choromanski K, Likhoshesterov V, Dohan D, et al. Rethinking attention with performers. 2020. ArXiv:2009.14794
- 196 Xie X, Liu H, Na Z, et al. DPiT: detecting defects of photovoltaic solar cells with image transformers. *IEEE Access*, 2021, 9: 154292
- 197 Pinaya W H L, Tudosi P D, Gray R, et al. Unsupervised brain anomaly detection and segmentation with transformers. 2021. ArXiv:2102.11650
- 198 Oord A, Vinyals O, Kavukcuoglu K. Neural discrete representation learning. 2017. ArXiv:1711.00937
- 199 Matsoukas C, Haslum J F, Söderberg M, et al. Is it time to replace CNNs with transformers for medical images? 2021. ArXiv:2108.09038
- 200 Xia Y, Zhang Y, Liu F, et al. Synthesize then compare: detecting failures and anomalies for semantic segmentation. In: *Proceedings of European Conference on Computer Vision*. Cham: Springer, 2020. 145–161

Survey on industrial defect detection with deep learning

Dongliang LUO¹, Yuxuan CAI¹, Zihao YANG¹, Zheyang ZHANG¹, Yu ZHOU^{1*} & Xiang BAI^{2*}

1. *School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan 430074, China;*

2. *School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 430074, China*

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

Abstract Deep-learning-based industrial defect detection methods play an increasingly important role in intelligent manufacturing, as they provide compelling benefits in reducing the cost spent on manual product inspection, while improving inspection accuracy and efficiency. They have been widely used in various manufacturing, operation and maintenance (O&M) applications such as automated inspection, smart patrol, and quality control. This survey aims to make a comprehensive introduction of industrial defect detection, which mainly spans its definition, difficulties, challenges, mainstream methods, open datasets, and evaluation protocols, so as to help researchers gain rapid and comprehensive knowledge. Specifically, we firstly introduce some background knowledge. Secondly, based on the difference of the provided annotations of different datasets in practical scenarios, we categorize most methods into three task settings: known defects, unknown defects, and few-shot defects. We go over these methods in further depth and illustrate a detailed analysis. We expound the connections between different algorithms and actual demands to present a clear picture of how different algorithms evolve. Thirdly, this paper summarizes some useful strategies that can effectively improve defect detection performance. Finally, based on our understanding of this area, we conclude several limitations of existing methods in practical applications as well as several directions of future research that embrace further efforts.

Keywords defect detection, anomaly detection, computer vision, industrial vision, deep learning