

中图法分类号: TP391.41; TN305 文献标识码: A 文章编号: 1006-8961(2025)01-0025-26

论文引用格式: Hu Z Q and Wu Y Q. 2025. Survey of semiconductor wafer defect detection method based on machine vision. Journal of Image and Graphics, 30(01):0025-0050(胡志强, 吴一全. 2025. 基于机器视觉的半导体晶圆缺陷检测方法综述. 中国图象图形学报, 30(01):0025-0050)[DOI:10.11834/jig.240053]

基于机器视觉的半导体晶圆缺陷检测方法综述

胡志强, 吴一全*

南京航空航天大学电子信息工程学院, 南京 211106

摘要: 当今许多新兴科技领域的发展需要集成电路(integrated circuit, IC)技术的支撑。半导体晶圆作为集成电路芯片中的关键角色, 因制作工艺复杂, 容易产生各种缺陷, 其故障会极大地影响芯片的最终工作性能并增加成本。因此半导体晶圆的缺陷检测是保证其良品率和生产率的重要手段。结合机器视觉算法的晶圆缺陷检测方法普适性强、速度快, 能更好地满足工业检测的相关需求。鉴于此, 综述了近十几年来半导体晶圆缺陷检测方法的研究进展, 介绍了晶圆制造的复杂工艺、表面缺陷检测的流程, 根据不同的标准将晶圆缺陷进行分类。重点阐述了基于学习的方法, 包括有监督机器学习、无监督机器学习、混合学习、半监督学习以及迁移学习5大类。对晶圆缺陷检测的深度神经网络分类为检测网络、分类网络、分割网络与组合网络。随后梳理了在晶圆缺陷检测领域常用的数据集以及性能评价指标。最后总结了晶圆缺陷检测当前存在的问题, 对未来的研究方向进行了展望。

关键词: 半导体晶圆; 缺陷检测; 机器视觉; 深度神经网络(DNN); 机器学习; 晶圆缺陷数据集

Survey of semiconductor wafer defect detection method based on machine vision

Hu Zhiqiang, Wu Yiquan*

College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China

Abstract: The integrated circuit chip is developed on a semiconductor wafer substrate and serves as a fundamental electronic component in many electronic devices. The semiconductor wafer, which is the foundation for manufacturing integrated circuits, contains billions of tiny electronic components. Production is gradually moving toward high-performance and small-sized directions because of the semiconductor industry's demand for high-quality wafers. In the manufacturing process of wafers, unexpected structures may remain due to environmental, operational, and process-related reasons, leading to a decrease in the operational performance of the wafer chip. These unforeseen structures are referred to as wafer defects. The causes of wafer defects are diverse, with common types including scratches, pits, chemical stains, and dust contamination. The occurrence of unpredictable, mixed, and complex defects can result in increased production costs, decreased product yield, and a decline in manufacturing process stability. Therefore, defect detection on semiconductor wafers is crucial to ensure their performance and improve the yield of high-quality products. Currently, methods for detecting surface defects on semiconductor wafers include manual visual inspection, optical inspection, ultrasonic testing, and machine vision detection. Manual visual inspection involves operators using microscopes or the naked eye to check for obvi-

收稿日期: 2024-02-01; 修回日期: 2024-04-29; 预印本日期: 2024-05-06

* 通信作者: 吴一全 nuaaimage@163.com

基金项目: 国家自然科学基金项目(61573183)

Supported by: National Natural Science Foundation of China (61573183)

ous defects on the wafer surface, but this method has low efficiency and accuracy for detecting small defects. Optical inspection uses light sources to illuminate the wafer surface and determines the presence of defects by detecting the intensity and shape of reflected and transmitted light. This non-contact method is suitable for large-area defect detection but is less effective for detecting small and shallow defects and cannot identify internal defects within the wafer. Ultrasonic testing utilizes the propagation speed of ultrasonic waves in the wafer, along with the reflection and scattering characteristics of defects such as cracks and voids, but it is sensitive to material density and sound speed, leading to reduced detection effectiveness in specific material. In addition to these three methods, machine vision detection uses image processing and computer vision technology to detect defects on the wafer surface. It is a non-contact, high-speed, automated, cost-effective, and scalable approach, making machine vision-based detection the primary method in the wafer defect detection field today. With the development of computer vision, machine learning, and especially deep learning, the performance of defect identification and classification algorithms for semiconductor wafers has been further enhanced. This not only improves detection efficiency but also reduces production costs and enhances product quality. Machine vision-based detection methods generally fall into three categories: traditional image processing methods, machine learning methods, and deep learning methods. The development of many emerging technology fields today requires the support of integrated circuit technology. As a key role in integrated circuit chips, semiconductor wafers are prone to various defects due to complex manufacturing processes, and their failures will greatly affect the final performance of the chip and increase costs. Therefore, defect detection of semiconductor wafers is an important means to ensure its yield and productivity. Wafer defect detection combined with machine vision algorithms has strong universality and high speed and can better meet the relevant needs of industrial inspection. To have a deeper understanding of the research status of wafer defect detection, this paper conducts a comprehensive and thorough research on the defect detection method based on machine vision. Wafer manufacturing, its surface defect detection process, and the classification of wafer defects are introduced. The detection methods based on traditional image processing, machine learning, and deep learning algorithms are fully explored. Detection methods based on traditional image processing are divided according to the spatial domain and the transform domain. Detection methods based on the spatial domain include template matching, edge gradient, optical flow method, and spatial filtering method and are divided according to the correlation between the defect pattern and the template or the dissimilarity of the background. Detection methods based on the transform domain include wavelet transform, frequency filtering, and Fourier transform. Further research is being conducted on methods based on Fourier transform in the wavelet transform domain. It focuses on the application of machine learning and deep neural network in wafer defect detection. According to the learning method, the learning techniques utilized for wafer defect identification are divided into supervised machine learning, unsupervised machine learning, hybrid learning, semi-supervised learning, and transfer learning; according to the network category, deep neural networks for wafer defect detection is divided into detection network, classification network, segmentation network, and combination network. The detection performance and advantages and disadvantages of each method are compared in depth in each section. From the collected literature, it indicates that research on wafer defect detection based on transformer is increasing year by year. The large datasets shared in this field mainly include WM-811K and MixedWM38. Most of the datasets are not open to the public and are maintained and updated by teams or companies. Representative evaluation indicators mainly include single-label indicators of supervised learning/semi-supervised learning (accuracy, precision, recall, F1 score) and multi-label indicators (micro-precision, micro-recall, exact match ratio, Hamming loss). Unsupervised learning indicators include Rand index, adjusted Rand index, normalized mutual information, adjusted mutual information, and purity. Finally, the existing problems of the current vision-based wafer defect detection are identified, namely, low data availability, serious class imbalance problem, high computational complexity, and limited research on mixed-type defects. Future development trends include multimodal and non-destructive detection techniques and more efficient feature representation learning.

Key words: semiconductor wafer; defect inspection; machine vision; deep neural network (DNN); machine learning; wafer defect dataset

0 引言

集成电路(integrated circuit, IC)芯片在半导体晶圆基板上开发,是许多电子设备的基本电子部件。半导体晶圆是制造集成电路的基础,其中包含了数亿个微小的电子元件,因半导体行业对高质量晶圆的需求,其制作逐渐朝着高性能小尺寸方向发展。在晶圆的制造流程中,由于环境、操作和工艺等原因在其中留下预想外的结构,并可能因此导致晶圆芯片电路工作性能降低,这类预想外的结构称为晶圆缺陷,其聚集形成的空间模式称为总失效区域(gross failure area, GFA)。晶圆缺陷的成因不尽相同、种类繁多,常见的缺陷类型有划痕、凹坑、化学污渍、灰尘污染等。不可预测、混合和复杂的缺陷的出现将导致生产成本增加、产品产量下降和制造工艺稳定性下滑等问题,因此对半导体晶圆进行缺陷检测,是保证其性能和提高其良品率的重要手段。

目前,关于半导体晶圆表面缺陷检测的方法主要包括:人工目视法、光学检测法、超声波检测法以及机器视觉检测法。人工目视法由操作人员使用显微镜或者裸眼检查晶圆表面是否有明显的缺陷,但这种检测方法的效率和针对微小缺陷的检测精度较低。光学检测法使用光源照射晶圆表面,通过检测反射光与透射光的强度、形状来判断缺陷是否存在,这种方法属于非接触式检测,适用于大面积的缺陷检测,但对微小和较浅缺陷的检测效果差,无法检测晶圆内部缺陷(Shon 等, 2021)。超声波检测法利用超声波在晶圆中的传播速度和缺陷反射、散射特性来检测其内部缺陷,如裂纹、空洞等,但容易受材料密度和声速的影响,在特定材料中检测效果变差。除了上述3种方法外,机器视觉检测法利用图像处理和计算机视觉技术来检测晶圆表面的缺陷,其具有检测过程非接触式、高速、自动化、成本低以及可扩展性等特点,因此基于机器视觉的检测方法成为当今晶圆缺陷检测领域中最主要的应用方法(汤勃等, 2017)。随着计算机视觉、机器学习、尤其是深度学习的发展,半导体晶圆的缺陷识别和分类算法的性能得到进一步增强,在提高检测效率的同时,减少了制作成本,提高了产品质量。一般基于机器视觉的检测方法可分为:传统图像处理的方法、机器学习的方法和深度学习的方法。

以机器视觉进行半导体晶圆缺陷检测的研究越来越受到关注。目前的相关综述文献情况如下: Batool 等人(2021)对晶圆图(wafer map, WM)缺陷检测中深度学习技术的应用进行系统性的综述,介绍了用于半导体晶圆的故障评估和缺陷检测的深度学习模型,但深度学习网络的文献搜集不全面,且未涉及经典的机器学习算法,基于卷积神经网络(convolutional neural network, CNN)的缺陷检测方法分类较混乱。Kim 等人(2022)介绍了晶圆的制作与检测过程、数据集的晶圆仓图(wafer binary map, WBM)组成与相关的评价指标,并根据检测的缺陷特征、学习算法对其中有代表性的检测方法进行比较。选取的文献时间跨度大、质量高,包含的方法较多,但较多检测方法存在分类重叠的问题,性能评估不够全面,且未涉及传统图像算法的介绍。Theodosiou 等人(2023)从 Scopus 数据库中按一系列限定条件检索2017年1月至2022年5月期间发表的晶圆缺陷检测论文,并按照机器学习、深度学习和混合学习进行分类。搜集的论文质量高,统计了文献、数据集的占比与出版物的逐年变化,但分类较粗略,深入探讨的文献过少,仅29篇,未以列表的形式对所搜集的方法进行整体的性能评估。本文系统全面地综述了近十几年来基于传统图像处理、基于机器学习和基于深度学习的半导体晶圆缺陷检测方法,深入分析比较了该领域的不同算法与模型的检测性能。

本文专注于晶圆缺陷识别和分类的各种机器学习和深度学习的研究进展,目的在于帮助半导体晶圆缺陷检测的研究人员能快速、系统地了解当前相关技术的发展状况。本文第1节介绍制造工艺的背景、表面检测的流程以及WM缺陷的分类情况。第2节将基于学习的多种模型和方法分类为有监督机器学习、无监督机器学习、混合学习、半监督学习以及迁移学习等5类,其中有监督学习模块(2.1节)介绍一系列监督式的机器学习算法;在无监督学习小节(2.2节)中阐述用于WM缺陷聚类、自编码器等几种算法;在混合学习中(2.3节)介绍不同种类的机器学习集成算法;在半监督学习中(2.4节)对低标签依赖性和高分类有效性的算法展开介绍;在迁移学习中(2.5节)比较几种大型的网络转移至晶圆缺陷检测领域上的性能。第3节重点论述用于晶圆缺陷检测的深度学习网络架构,主要分为检测网络、分类网络、分割网络和组合网络。第4节梳理半导体领

域内缺陷的数据集以及性能评价指标。第5节分析目前晶圆缺陷检测存在的问题,并展望进一步的发展趋势,最后对本文的整理工作进行总结。期望能为半导体晶圆缺陷检测领域的研究人员提供一些思路和启示。

1 晶圆缺陷与检测流程

晶圆是指在制作IC中所使用的一种结构复杂的圆形半导体硅晶圆,其上可承载各种电子器件与电路,是半导体电路制造业中重要的中间产物。晶圆制备工艺和制作标准会影响其生产产量、产品质量与缺陷复杂性的变化。而随着机器学习、深度学习和计算机视觉的发展,在晶圆图缺陷检测中逐渐应用学习算法与检测网络,自动检测技术可以提高模型的准确性、降低计算负载、减少运行时间和提升学习能力。本节将主要介绍不同种类的半导体晶圆缺陷模式、表面缺陷检测过程以及半导体晶圆的制造流程。

1.1 晶圆缺陷的分类

表1概述了半导体晶圆表面常见的缺陷,以及

表1 半导体晶圆缺陷与对应成因			
Table 1 Semiconductor wafer defects and their causes			
缺陷	过程	成因	图像
随机	环境清理	工作间的环境条件较差,颗粒和碎屑进入硅片表面	
局部	研磨、抛光	晶圆表面不均匀,清洁不到位	
边缘	研磨、抛光	晶圆表面不均匀,清洁不到位	
中心	抛光	晶圆表面不均匀	
边环	研磨、光刻	晶圆表面不均匀,因化学蚀刻失误导致的层间错位	
划痕	组装抛光、环境清理	抛光垫硬化、粒子积聚、机械操作不当	
近满	-	随机性缺陷与多种系统性缺陷的集聚	
圆环	研磨、抛光	晶圆表面不均匀、设备处理不当、抛光垫硬化	

注:“-”表示暂无数据。

这些缺陷所涉及的过程和导致它们的原因。晶圆缺陷检测的目标是分析系统缺陷的空间模式,而随机缺陷在检测过程中增加了分析系统缺陷模式的挑战,通常视为噪声。对系统性的局部缺陷进行模式判别和成因追溯,不仅可以降低生产成本,还能有效提高良品率和晶圆质量。

晶圆缺陷尺度非常小,同样以纳米(nm)、微米(μm)为单位衡量。根据不同分类标准,晶圆缺陷有不同的分类方式。

根据 Kaempf 标准(Kaempf, 1995)可将晶圆缺陷分为以下3类:

1)随机缺陷。此缺陷是均匀随机的,具有稳定的平均密度。这类缺陷没有具体的聚类现象。造成此种类型缺陷的原因复杂,往往很难找到其成因。但可以通过提高制作工艺的稳定性和精准度,减少由此类缺陷导致的未合格的产品数量。

2)系统缺陷。此缺陷是系统性的,在WM中具有重复性、周期性和规律性。此类缺陷具有明显的聚类现象。这些缺陷可能由制造工艺、设备问题或材料质量等因素引起,可通过WM上的缺陷分布模式找到其成因。

3)组合缺陷。这种缺陷是半导体制造中最常见的缺陷分布,因WM而异。组合缺陷是随机缺陷和系统缺陷的组合。对于此类型缺陷,往往采用消除随机缺陷、保留系统缺陷的手段找到故障产生的原因。

依据缺陷表征与特性作为分类标准,可将晶圆缺陷分为以下3类:

1)形貌缺陷。晶圆的形貌缺陷包括平整度不良、边缘裂纹和表面缺陷等,其会影响IC图案的形成,通常是在抛光或光刻过程中,工人或机械操作不当引起的。

2)污染物。常见污染物如颗粒,通常由环境中的脏物污染晶圆所致。在分子层面上包括有机层和无机层等污染物;在原子层面上包括离子、重金属缺陷等。

3)晶体缺陷(Bollmann, 2012)。通常由晶体生长不均导致,硅晶片层面存在的晶体缺陷会影响构建在这些晶圆上的器件的电性能,甚至由晶体缺陷导致的极端退化情况可能引起设备故障。常见晶体缺陷有:硅原子失位、硅原子错位以及非硅原子掺杂。

1.2 表面缺陷检测的流程与晶圆的制造

晶圆表面缺陷检测的整体流程如图1所示,共包含5个步骤。

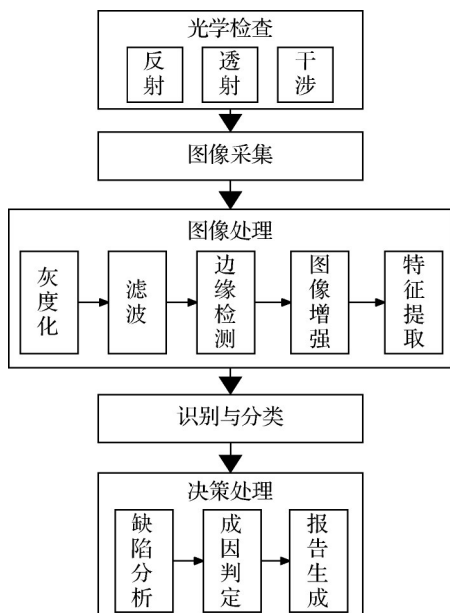


图1 晶圆表面缺陷的检测流程

Fig. 1 Workflow for inspecting surface defects on wafers

第1步主要是使用光学显微镜检查表面显著缺陷。第2步通过高分辨率的光学显微镜采集晶圆表面的图像。其中WM是晶圆经过电性测试后的空间结果,WM是二进制的WM图像,由自动视觉或电子束检查系统获得(Tilli等,2020)。缺陷图像是故障模式的集群,每种模式具有不同的特征。第3步是对所采集WM进行灰度化、滤波、边缘检测、图像增强等处理。第4步通过检测系统对晶圆表面的缺陷进行自动识别和分类,对于每个缺陷,提供其尺寸、位置和严重程度等评估参数。最后一步根据缺陷的评估参数,生成缺陷报告,对其进行缺陷成因判定。随后利用光刻技术和化学气相沉积(chemical vapor deposition, CVD)等方法,在晶圆表面逐步形成各种层和结构,包括导线、晶体管、电容器等。对结构化的晶圆进行二次清洗后,进入半导体的后端制造工艺。通过电性测试生成表面的晶圆图,验证于晶圆上形成的电子器件的性能和质量,确保其符合规格和要求。在晶圆接受测试中采集的晶圆图通常由多个图层组成,包括衬底层、绝缘层、传感层、金属层、圆中划片槽等。图2显示了划片槽的部分结构,左侧是完整的WM,为了将芯片进行包装,通常

使用划片刀将晶圆切割成芯片,其中的凹槽称为划片槽,每一个小格代表一颗芯片,图2右侧展示了中间的划片槽在显微镜放大下的图像。

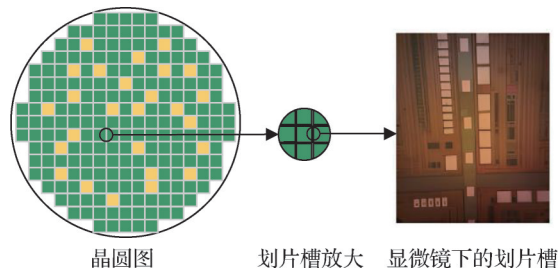


图2 晶圆中划片槽的部分结构

Fig. 2 Part of the structure of the dicing slot in a wafer

晶圆的完整流程与IC的制造过程涉及许多复杂流程,如提取硅锭、切割、研磨、蚀刻、抛光、清洗、掺杂、氧化。各制备步骤如表2所示。

2 基于学习的晶圆缺陷检测方法

基于学习的晶圆缺陷检测方法利用传统机器学习技术和深度学习算法来检测和分类晶圆缺陷,采用的学习方法主要包含有监督学习、无监督学习、混合学习、半监督学习和迁移学习等。大部分研究主要包含有监督学习和无监督学习,基于混合学习、半监督学习以及迁移学习的检测方法则用于提升算法的学习性能、突破模型的数据依赖性(雷林建等,2020)。有监督学习与无监督学习包含许多算法,根据具体场景和数据特征选择合适的方法能提高缺陷检测的准确性和效率。

2.1 基于有监督机器学习的方法

有监督学习通过利用带有标签的训练数据学习类别特征,从而能够完成分类或预测任务。为决策提供有力的支持,监督学习在自动解决许多复杂问题方面表现出色。在本节中重点介绍应用于晶圆缺陷检测的监督传统机器学习算法。包括支持向量机(support vector machines, SVM)、决策树、K近邻(K-nearest neighbor, KNN)、多层感知器(multilayer perceptron, MLP)、线性判别分析(linear discriminant analysis, LDA)、自适应共振理论(adaptive resonance theory, ART)以及集成学习算法等。

SVM是一种用于模式分类和回归分析的监督学习算法,通过将样本映射到高维特征空间,寻找最

表2 晶圆的制造流程
Table 2 Wafer manufacturing process

制备流程	流程描述
硅的生长与提取	晶圆通常由单晶硅制成。制造晶圆的第1步是准备硅单晶圆。原始硅通常以石英砂的形式存在,经过多个步骤的冶炼、提纯和晶体生长过程,最终获得高纯度的硅单晶。在实践中,硅锭通常使用直拉法或浮动区法生长。
切割	硅锭生长后,将其提取并裁剪去除非圆柱体端部,通过金刚石线切割将硅锭切成薄片,即晶圆,然后使用异性金刚石砂轮,将晶圆边缘切割成定制的边缘轮廓,调整直径,降低滑动和碎裂的风险。常见直径尺寸包括200 mm或300 mm。
研磨	切割得到的晶圆表面不够光滑和平坦,需要进行研磨。通过机械和化学方法,去除晶圆表面的不均匀性和缺陷,使其达到高度平坦和光滑的状态。在研磨操作后,晶圆的表面或边缘可能出现残余的机械损伤。
化学蚀刻	晶圆经过研磨后,会残留有颗粒和杂质。为了确保晶圆的纯净度,需要进行化学蚀刻。清洗过程使用碱性或酸性化学溶液和超纯水,以去除表面的杂质。
抛光	晶圆通过化学蚀刻后,进行抛光以达到所需的厚度和平坦度,控制总厚度变化在合理区间中。
掺杂	晶圆上的某些区域需要引入特定的杂质,以改变其电学性质,这个过程称为掺杂。通常使用离子注入或扩散等技术,在晶圆表面或内部中引入特定的杂质原子。
氧化	在晶圆上形成氧化硅层,以便作为绝缘层或隔离层。通常通过将晶圆暴露在氧气或蒸汽中进行热氧化来实现。

优的超平面,将不同类别的样本尽可能分类,具有良好的泛化能力和较强的鲁棒性。Wang和Ni(2019)通过主成分分析(principal component analysis, PCA)提取晶圆主仓值,用于训练非线性SVM模型。研究结果表明,结合PCA和SVM的方法能有效提高WBM分类的准确性。Piao等人(2018)提取Radon变换特征,应用于集成决策树模型,对晶圆缺陷进行分类。Kim等人(2020a)从缺陷形状、位置信息中提取WM的不确定特征,提出了广义非确定决策树模型KNN的核心思想是根据样本之间的距离来进行分类。

LDA是一种经典的模式分类和特征提取方法。它通过线性变换将原始数据投影到低维空间,以实现数据的分类和可视化。余建波等人(2016)基于对数正态泊松点(log-normal Poisson point, LNPP)、局部和非局部的LDA,提出了新的流形学习算法,即局部与非局部线性判别分析法(local and nonlocal linear discriminant analysis, LNLDA),该框架利用流形学习来提取高度判别的特征。对于晶圆缺陷检测,进一步提出了基于LNLDA的Fisher判别式,用于计算晶圆缺陷的判别函数值。Abdullah等人(2021)对于模拟数据集的缺陷模式识别,比较了3种不同集成方法,即bagging、voting和AdaBoost。所有模型都实现了相似的分类性能,AdaBoost和bagging的精度略有提高。Liao等人(2021)基于真实WM的自定义数据集提出了一种识别划痕缺陷的方法,使用自适应增强(adaptive boosting, AdaBoost)进行分类。

结果在划痕图案中获得了超过89%的识别率,在常见缺陷图案中获得超过94%的识别率。陈晋贤等人(2020)提出了一种综合利用混合特征选择和基于序列模型优化(sequential model-based optimization, SMBO)的方法,用于预测缺陷。该方法以适用于高维、多噪声、多模态和线性不可分数据的随机森林(random forest, RF)和SVM这两种机器学习算法为基础,采用序列前向搜索策略来寻找降序排列的特征。在搜索过程中,利用SMBO方法进行优化,最终从中选择性能表现最佳且具有最少特征数量的模型,用于缺陷预测。Liao等人(2021)在真实WM数据集上进行对网格型缺陷模式的识别。前4个步骤侧重于增强WM,最后阶段旨在利用机器学习集成模型识别缺陷图案,包括AdaBoost、决策树、RF、极致梯度增强(extreme gradient boosting, XGBoost)、bagging。Saqlain等人(2019)提出了一种用于晶圆缺陷识别和分类的软投票分类器。集成了4种机器学习分类器:逻辑回归(logistic regression, LR)、梯度提升机(gradient boosting machine, GBM)、人工神经网络和RF。结果证明了与单个基本分类器相比,集成分类器的性能有所提高。尽管LNLDA实现了较高的识别率,但其性能取决于手动生成的特征,容易形成性能瓶颈。

表3总结了用于晶圆缺陷检测的几种有监督学习算法的分类性能与优缺点。

表 3 主要的有监督机器学习方法对比
Table 3 Comparison of major supervised machine learning methods

类别	方法	数据	准确率	贡献/局限性
SVM	PCA+SVM(Wang 和 Ni, 2019)	真实和模拟数据	0.895 3	主仓提取法性能良好,鲁棒性待提升。
	基于 Radon 的聚合决策树 (Piao 等, 2018)	WM-811K	0.905 0	基于 random 变换做多类型特征描述,集成决策树对分支预测做汇总,聚合性能好。
	广义不确定决策树(Kim 等, 2020a)	真实数据	0.956 0	局限于集群、圆形、重复和随机类型缺陷识别。
LDA	LNLDA(余建波 等, 2016)	WM-811K	0.905 0	分类准确性高,验证了模型有效性与实用性。
集成学习	bagging, voting, boosting (Abdullah 等, 2021)	模拟数据	0.998 0	对环形、斑点、月型等不同缺陷形状保持高分类精度,耗时短。
	AdaBoost 预测(Liao 等, 2021)	真实和模拟数据	0.940 0	降低测试成本,精度与效率提升。
	SMBO(陈晋贤 等, 2020)	真实数据	0.813 0	单特征选择稳定性更好,识别准确率优于 KNN、NN 模型,仅针对蚀刻制程中的缺陷进行预测。
	AdaBoost, bagging, RF, GBM (Liao 等, 2021)	真实数据	0.964 5 (XGBoost)	有效提取隐藏网格模式的 WM,显著减低晶圆的测试成本与不可靠性。
	LR, GB, RF(Saqlain 等, 2019)	WM-811K	0.958 7	减少过拟合、欠拟合,提高分类性能。

2.2 基于无监督机器学习的方法

无监督学习的目的是从无标签的数据中发现数据的潜在结构、特征,无需预先给定目标标签。相比监督学习,无监督学习不需要依赖人工标记的数据进行训练,而是依靠算法自主地对数据进行学习和探索。本节主要针对聚类、自编码器(autoencoder, AE)、深度信念网络(deep belief network, DBN)、原型网络(prototype network, PN) 等无监督机器学习算法展开介绍。

1)概率聚类。Hwang 和 Kim(2020)将狄利克雷过程高斯混合模型(Dirichlet process Gaussian mixture model, DPGMM)与变分自动编码器(variational auto-encoder, VAE)相结合。狄利克雷过程用于自动更新簇的数量, GMM 用于先验分布,以学习不同 WM 的细微差别。Ezzat 等人(2021)提出了一种识别晶圆缺陷空间模式的框架,该框架集成了用于空间滤波的邻接聚类(adjacency-clustering, AC)模型和用于空间聚类的混合模型 iWMM。作者比较了 AC 和连通域滤波算法,证明了 AC 能更有效地过滤高复杂度的缺陷。

2)密度聚类。Jin 等人(2019)提出一种新的基于密度的聚类方法(density-based spatial clustering of applications with noise, DBSCAN),提取边缘缺陷

的笛卡尔坐标和极坐标用于聚类。

3)划分聚类。Taha 等人(2018)提出一种晶圆聚类算法,即显性缺陷模式查找器(dominant defect pattern finder, DDPfinder),考虑了所有仓图的空间依赖性。Tulala 等人(2018)使用 VAE 从 WM 图案中提取特征,应用 K-means 算法将 WM 划分为不相交的簇。

4)层次聚类。卢笑蕾和余建波(2018)采用基于流形调节的局部连续高斯混合模型(Gaussian mixture model with local consistency, LCGMM)将晶圆表面的局部缺陷划分为线性、曲线形和椭圆形缺陷簇,并进一步提出集成 LCGMM 和主曲线模型(principal curve, PC)的混合模型,提高了缺陷模式识别的准确性。

AE 用于晶圆缺陷识别中的各项任务:特征学习、缺陷分割、聚类和分类。AE 的基本组成是编码器和解码器。编码器将输入转化为潜在空间表示,解码器则用潜在空间表示重构输入。对于数据集的干扰噪声, Lee 等人(2017)提出使用堆叠式去噪自动编码器(stacked denoising auto-encoders, SDAE)进行特征提取与分类,能识别晶圆样本的全局特征和不变特征,对测试噪声具有较强的鲁棒性。刘国梁和余建波(2022)提出一种基于知识的堆叠降噪自编

码器 (knowledge-based stacked denoising auto-encoder, KBSDAE)。首先对SDAE的网络特点设计一套符号规则系统,提高其与神经网络的有效融合;其次研究了一种完整的知识提取和插入算法,实现深度网络和规则之间的知识转换。实验结果表明,该模型取得了较好的检测率,综合性能甚至优于经典的深度网络。而对于数据集的类不平衡问题,Shawon 等人(2019)使用卷积自动编码器(convolutional auto-encoder, CAE)进行数据合成,在每种类型中添加了2 000幅图像,采用了数据增强解决类不平衡问题。

DBN是一种由多层受限玻尔兹曼机(restricted Boltzmann machine, RBM)组成的无监督学习模型,可以学习数据的分布和高级特征表示,适用于数据降维、特征提取、生成模型等任务。Zhuang 等人(2020)使用DBN对WM的6种缺陷模式进行分类。针对真实WM数据中的6类缺陷分别训练了6种DBN,在具有单类型和混合类型缺陷的数据集上进

行了测试。Yu 等人(2021a)基于混合自注意(mixture of self-attention, MSA)和PN 提出少样本学习。PN 不同于传统的机器学习方法,其采用元学习训练范式来处理数据稀缺性学习。使用少量数据,该方法也能对多数缺陷模式表现较高的分类性能。表4归纳比较了基于无监督机器学习的缺陷检测方法的性能、贡献和局限性。

2.3 基于混合学习的方法

混合学习结合不同类型的学习算法、模型或策略,如有监督学习、无监督学习、强化学习、遗传算法等,以及它们的组合形式,利用不同方法的优势来解决复杂的问题,以提高学习系统的鲁棒性、泛化能力和适应性。Shon 等人(2021)使用WM-811K的未标记数据,训练卷积变分自动编码器(convolutional variational auto-encoders, CVAE)。CVAE的判别网络具有较高的分类性能,但难以区分混合缺陷。针对上文对混合型缺陷的局限性,Jin 等人(2020)提出一种组合CNN、纠错输出码(error correction output

表4 主要的无监督机器学习方法对比
Table 4 Comparison of major unsupervised machine learning methods

类别	方法	数据	准确率	贡献/局限性
概率聚类	DPGMM(Hwang 和 Kim, 2020)	真实数据	-	在SVAE中结合狄利克雷过程实现两次聚类,聚类精度高,获取多种确定簇。
	AC+iWMM(Ezzat 等, 2021)	真实数据	-	对复杂混合型缺陷分类性能优越,AC滤除噪声可靠、快速。
密度聚类	DBSCAN(Jin 等, 2019)	WM-811K	-	适用于检测不同尺寸晶圆的缺陷,单类型缺陷检测精度较低。
划分聚类	DDPfinder(Taha 等, 2018)	真实与模拟数据	0.998 0	精度随运行时间和样本的增加而增加,但计算复杂度也增加。
	VAE(Tulala 等, 2018)	真实数据	-	有助于提高上游生产效率,提高发现早期生产问题的效率。
层次聚类	LCGMM+PC (卢笑蕾和余建波, 2018)	真实与模拟数据	0.997 5	有效识别晶圆表面缺陷簇,提高晶圆质量,但识别的形状较简单。
AE	SDAE(Lee 等, 2017)	模拟数据	0.985 0	对噪声有鲁棒性,精度高,只考虑光刻工艺造成的缺陷。
	KBSDAE (刘国梁和余建波, 2022)	WM-811K	0.911 4	知识转化算法提高模型的识别性能,复杂模型的可解释性待提升。
	CAE(Shawon 等, 2019)	WM-811K	0.992 9	使用数据增强技术解决类不平衡问题,显著提高识别准确率。
DBN	DBN(Zhuang 等, 2020)	真实数据	0.912 0	简化WBM模式提取,提高模式诊断的效率和准确性。
PN	MSA+PN(Yu 等, 2021a)	WM-811K MixedWM38	0.923 2	分类精度在小样本方法中最高,但只能识别有限数量的缺陷类别。

注:“-”表示暂无数据。

code, ECOC)和SVM的框架,用于识别WM的4种常见的缺陷模式:划痕、半圆、团簇和环,解决了因图像旋转和平移引起的误判问题,其准确率为98.43%。Park(2021)提出一种无监督的晶圆缺陷聚类方法,该方法对Siamese网络进行判别性特征学习,并在学习到的特征空间中对基于高斯均值聚类的重构类标签进行重复交叉学习。该方法在训练数据集的分类标签被破坏的情况下,可以由被破坏的样本找到其原始的分类标签来提高测试数据集的分类精度。结果表明,该方法的准确率比CNN提高了7.8%。此外,作者发现一个新的未知类型缺陷,没有用于学习,新的缺陷类识别的准确率超过73%。类标签重构的结果受孤立噪声样本和初始嵌入空间的样本分布的影响,因此需要研究更严格的模式细分和初步隔离噪声的方法,其网络结构如图3所示。

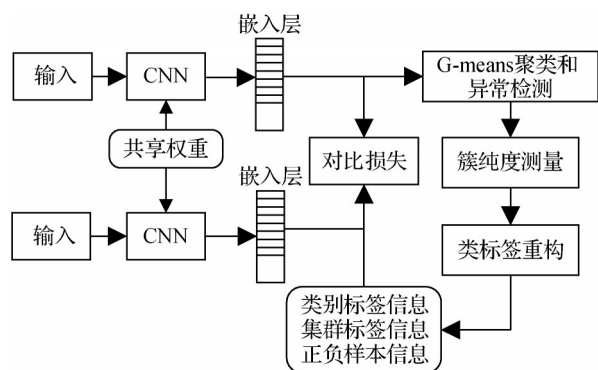


图3 结合类标签重构的孪生网络

Fig. 3 Siamese network combined with class label reconstruction

Wang等人(2021)提出一种VAE和CNN结合的框架,用于数据增强解决WM-811K的数据不平衡问题。其中VAE负责生成WM,深度卷积神经网络(deep convolutional neural network, DCNN)负责分类操作。结果显示准确度达到99.19%。Yang(2019)以及Kang和Kang(2021)均采用了数据增强的手段以解决数据的类不平衡问题。Yang(2019)检查了基于CNN、SVM、AdaBoost和XGBoost的分类器。CNN模型在台积电300mm晶圆厂的真实WM数据集中实现了99.2%的最高精度。Kang和Kang(2021)提出了一种组合基本分类器和元分类器堆叠的方法,基本分类器基于手动特征提取(manual feature extraction, MFE)和CNN创建,能够识别WM缺陷类型,并自动提取其特征。对于参数复杂化与模

型难调整的问题,Tsai和Lee(2020b)根据WM-811K和21-defect两个真实数据集,分别提出了轻量级的CNN和残差网络(residual network, ResNet),并使用CAE对每类图像进行过采样。表5对比了晶圆缺陷检测中混合学习方法的性能与优缺点。

2.4 基于半监督学习的方法

受限于可用标签数量的有监督学习以及无监督学习对未标记缺陷的分类性能可能的限制,半监督学习应运而生。这种方法克服了有监督学习和无监督学习的局限性,通过使用标记的WM学习每个缺陷图案的相关特征,同时利用未标记的WM来进一步细化特征表示。Yu等人(2019a)提出了一种半监督学习模型,堆叠卷积稀疏去噪自动编码器(stacked convolutional sparse denoising auto-encoder, SCSDAE)。SCSDAE体系结构由两个具有卷积和池化层的CSDAE和一个具有SVM分类器的分类层组成。Yu(2019)提出了一种具有流形正则化的增强堆叠去噪自动编码器(enhanced stacked denoising auto-encoder, ESDAE)用于特征学习。ESDAE实现了97.03%的最佳缺陷识别准确率,且与Shon等人(2021)训练的CVAE类似,ESDAE对单类型缺陷表现良好,但对混合类型缺陷表现欠佳。Yu和Liu(2021)提出了一种将半监督二维主成分分析(conditional two dimensional principal component analysis, C2DPCA)和卷积自动编码器(principal component analysis-based convolutional auto-encoder, PCACAE)相结合的方法。C2DPCA进行特征提取,然后通过自动编码器级联特征。该方法优于流型的CNN分类器,如局部敏感哈希—卷积神经网络、孪生网络、多尺度卷积网络和基于度量学习的卷积神经网络,在小类别上表现良好。

Chen等人(2021a)提出的框架共包括5个阶段,即数据清理、DBSCAN去噪、特征提取、基于bagging算法分类、基于SOM聚类,实验结果产生了94.37%的总体准确率。Li等人(2022)提出了一个半监督框架。将集成学习算法用于监督学习与识别可能错误标记的样本,无监督学习用于提取未定义的缺陷类型。Kong和Ni(2018)以半监督的方式训练了基于CNN的Ladder网络,以检测和分类WM缺陷。半监督梯形网络由两个编码器和一个解码器组成。编码器负责学习WM的潜在特征,解码器负责重建WM。由于缺陷分类是有限的,Ladder对混合

类型缺陷的检测需要进行修改和重训练。Kong 和 Ni (2020b) 提出了另一个半监督变分自动编码器 (semi-supervised variational autoencoder, SVAE) 用于 WM 缺陷分类。使用信息熵对半监督模型训练, 选择具有最大信息熵的 WM 进行标记, 将置信度高的 WM 用于模型微调。与标准 CNN 的比较表明, 梯形网络具有优越性能, 但 SVAE 优于梯形网络和标准 CNN, 尤其当标记数据百分比较低时。Lee 和 Kim (2020) 使用了熵模糊 c 均值 (entropy fuzzy c means, EFCM) 与谱聚类结合的方案, 来分类不同类型的晶圆缺陷模式, 并构建了一棵基于凸性和特征值比两个聚类特征的决策树来识别具体的缺陷类型。

Lee 和 Kim (2020) 提出了一种半监督深度卷积生成模型 (semi-supervised convolutional deep generative model with multiple discriminative networks, SS-CDGMM)。该模型利用标记数据和非标记数据, 将混合类型缺陷模式的分类问题转化为多标签分类问题。为了表示不同的单一模式, 研究团队采用了多个潜在类变量。整体而言, SS-CDGMM 的架构如图 4 所示, 展现了其复杂而高效的设计。

这项研究的贡献在于将半监督学习与深度生成模型相结合, 为处理混合类型缺陷模式提供了一种新颖的方法。通过引入多个潜在类变量, 该模型能够更精确地捕捉不同模式之间的差异, 为分类问题提供了更准确的表示。这一方法的应用有望在晶圆

缺陷检测领域以及其他复杂模式分类问题中取得显著的成果。所提出的生成模型具有生成标记 WBM 数据的优势, 为后续需要大量的 WBM 数据的研究提供了帮助。

Hu 等人 (2021) 提出一个针对单类型缺陷模式的对比学习框架, 然后对分类器进行监督微调。尽管性能低于其他算法, 但其检测率与最先进的对比方法 SimCLR 不相上下, 具有巨大的对比学习潜力。针对混合型缺陷模式的分类与检测, Lee 等人 (2023) 提出了一种半监督学习方法 (semi-supervised attend-infer-repeat, SS-AIR)。使用不同 WBM 数据集的测试结果表明, SS-AIR 在混合型缺陷模式的分类性能优于其他算法, 并能自动检测位置。针对相似缺陷模式难识别与部分缺陷模式未被定义等问题, 在表 5 中比较了晶圆缺陷检测中主要的半监督学习方法。

2.5 基于迁移学习的方法

迁移学习是一种机器学习方法, 目的在于利用在一个任务上学到的知识和经验来改善在另一个相关任务上的学习性能。传统的机器学习方法通常独立地从头开始学习每个任务, 而迁移学习则通过将已学习的知识迁移到新任务上, 加速学习并提高性能。沈宗礼和余建波 (2020) 提出了 T-DenseNet 框架。首先在 ImageNet 数据集上对预训练的 DenseNet 模型进行微调, 然后将细化的特征表示作为在线测

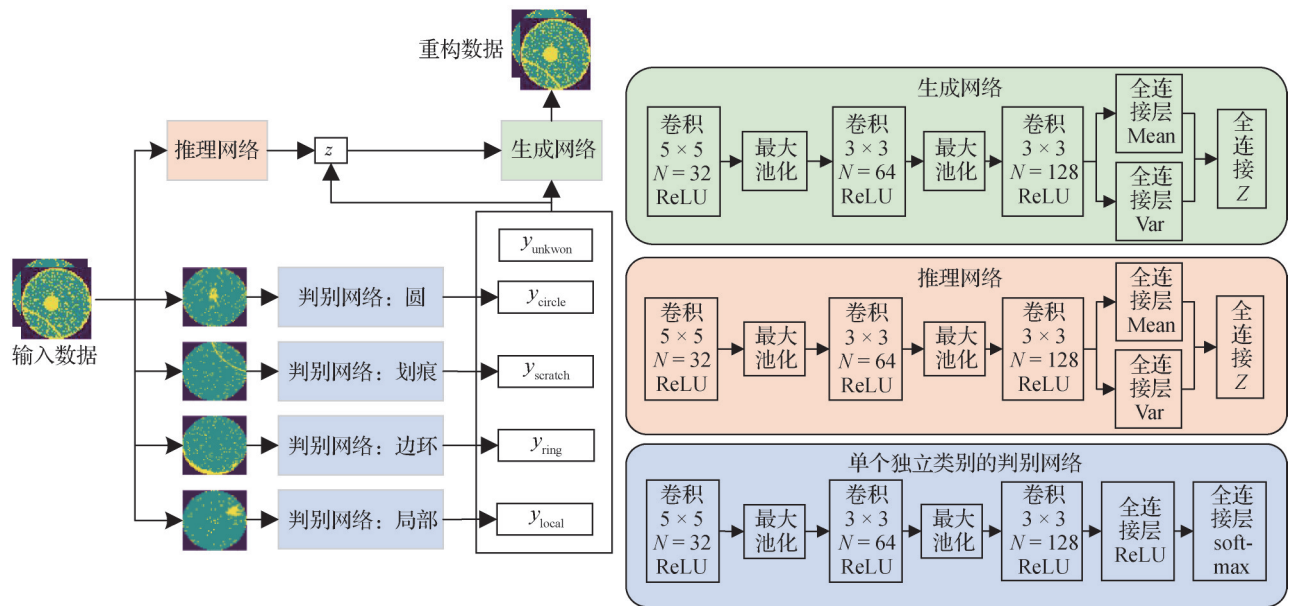


图 4 SS-CDGMM 的总体结构
Fig. 4 Overall structure of SS-CDGMM

表 5 主要的半监督学习方法对比
Table 5 Comparison of major semi-supervised learning methods

方法	准确率	贡献/局限性
SCSDAE(Yu等,2019a)	0.988 3	鲁棒性和特征提取表现良好,存在数据稀缺、类不平衡和过拟合问题。
ESDAE(Yu,2019)	0.970 3	应用流形正则化技术,算法有效性良好。
PCACAE(Yu和Liu,2021)	0.937 7	网络收敛与训练成本降低,但适用性差,只能识别单缺陷模式,不能处理非线性数据。
bagging+SOM+CNN(Chen等,2021a)	0.943 7	分类准确性与效率高,但存在类不平衡问题。
DBSCAN+SOM(Li等,2022)	0.943 7	解决标记稀缺与出错的问题,提高预测模型有效性,获得更高模式识别率。
Ladder网络(Kong和Ni,2018)	0.902 0	通过总失效区域GFA识别缺陷的根本原因,对GFA模式有高精度分类。
SVAE(Kong和Ni,2020b)	0.903 0	用梯形网络对具有总失效区域GFA模式的WBM进行高精度分类。
SS-CDGMM(Lee和Kim,2020)	0.948 8	分类准确性与效率高,但存在类不平衡问题。
SS-AIR(Lee等,2023)	0.939 5	扩展了AIR,对随机缺陷有更强的鲁棒性,对噪声滤波算法有较强要求,无法自主学习新的缺陷模式。

试系统的输入。它们的结果表明,迁移学习是一种更快、更准确的WM缺陷分类方法。Ishida等人(2019)提出一种基于CNN的WM模式识别框架VGGNet(Visual Geometry Group network)。在数据准备阶段使用一系列变换操作,对晶圆灰度图像的随机缺陷进行去噪。该框架的准确性和宏观F1分数分别为94.7%和73.67%。Chien等人(2020)利用DCNN对中心缺陷、局部缺陷、随机缺陷和划痕缺陷进行模型训练,并利用迁移学习微调预先训练的CNN,分别比较了在COCO(common objects in context)和KITTI(Karlsruhe Institute of Technology and Toyota Technological Institute)数据集上预训练的Faster R-CNN的性能,发现KITTI数据集上预训练的模型表现更佳。预训练的VGGNet模型和R-CNN模型被用于WM缺陷识别和分类。表6对比了几种迁移模型对不同缺陷类型的分类性能,反映了预训练的模型选择将影响下游任务的性能,体现了深度迁

移学习的有效性以及不同模型经迁移后的学习能力。目前已有的基于迁移学习的晶圆缺陷检测方法较少,其中预训练的T-DenseNet和VGGNet都被转移至WM-811K数据集中训练,能对大部分的缺陷模型分类,但不同种类的分类性能相差过大,包括中心、边缘、边环、划痕、随机等模式的晶圆缺陷。在KITTI和COCO数据集上预训练的Faster R-CNN网络,分类精度相比前两者更高,但支持识别的缺陷模式有限,仅集中于中心、局部、划痕和随机等的晶圆缺陷模式。

3 晶圆缺陷检测方法的深度神经网络

深度学习模型是一种基于人工神经网络的机器学习方法,通过多层次的非线性变换来学习样本数据的特征。其通常由多个神经网络层组成,通过学习数据的不同特征层次来逐渐提高模型的性

表 6 主要的迁移学习方法在不同缺陷类型上的分类准确率对比
Table 6 Comparison of classification accuracy of main transfer learning methods on different defect types

网络	中心	圆环	边缘	边环	局部	划痕	近满	随机	无图案
T-DenseNet	64.52	91.18	81.46	66.79	100.00	72.60	99.31	66.52	85.45
VGGNet	88.80	89.00	86.90	83.90	72.60	77.20	90.50	73.50	95.60
Faster R-CNN KITTI	98.92	-	-	-	97.27	98.32	-	99.26	-
Faster R-CNN COCO	97.81	-	-	-	98.50	96.14	-	99.06	-

注:“-”表示暂无数据。

能。由于深度学习模型具有自动特征提取能力,许多研究将深度学习模型用于混合类型的WM缺陷识别。

3.1 晶圆缺陷检测的检测网络

检测网络基于深度学习方法,通过训练大量的样本数据来学习缺陷的特征表示,并能够实现高效准确地缺陷检测和定位。在晶圆缺陷检测中,常用的检测网络包括以下几种:单阶段检测网络,如单点检测器(single shot multibox detector, SSD)、YOLO(you only look once);双阶段检测网络,如Mask R-CNN(mask region-based convolutional neural network);还有的缺陷检测方法基于Transformer(Vaswani等, 2023)模型进行研究。

3.1.1 单阶段检测网络

传统的单阶段检测网络通常是直接在输入图像上进行密集预测,预测图像中各个位置是否存在目标以及目标的位置和类别。这种方法在速度上具有优势,但可能存在目标定位不准确、漏检或误检的问题。SSD是一种常见的单阶段检测网络,通过预测图像在多个尺度上的边界框和类别信息,以检测不同尺度的目标。Kim和Behdinin(2023)将SSD应用于半导体晶圆缺陷检测中,利用Python中的PIL库和空间滤波器来获得边界框。SSD算法同时解决了对象分类和定位问题,有效识别、分割和分类混合型缺陷样本中的多个缺陷模式,减少了检测时间。与CNN模型相比,所提出的SSD模型对单类型和混合类型的缺陷实现了更高的精度。

YOLO是另一种非常流行的单阶段检测网络。它将检测任务转化为回归问题,通过将图像划分为网格单元,并在每个单元格中预测边界框和类别信息来实现实时目标检测。郭毅强(2018)利用Zhang-Suen细化算法提取骨架,并基于YOLOv3深度学习网络来检测缺陷。史浩琛等人(2022)提出了改进的YOLOv4算法,结合定制的CCD(charge-coupled device)工业相机能很好地识别各种复杂条件下的不同类型的缺陷,具有良好的鲁棒性。Shinde等人(2022)评估了YOLOv3、YOLOv4等检测网络在晶圆缺陷定位与分类上的性能,其中YOLO系列能得到94%的分类准确度,而ResNet50和DenseNet121的晶圆缺陷分类率只有89%与92%,且没有缺陷定位的能力。

3.1.2 双阶段检测网络

与单阶段检测器相比,双阶段检测网络通常具有更高的准确性和更低的漏检率。Mask R-CNN是一种基于Faster R-CNN的双阶段目标检测模型,它在目标检测的基础上进一步添加了实例分割的功能。Mask R-CNN可以准确地检测目标物体的位置并生成每个目标物体的类别信息。Li和Wang(2021)使用增强的Mask R-CNN对混合型缺陷模式进行了缺陷检测,使用ResNet架构作为骨干特征提取网络,并利用特征金字塔网络和非极大值抑制(non-maximum suppression, NMS)算法进一步改进了Mask R-CNN。在MixedWM38上的实验结果表明,与Mask R-CNN相比,改进后的模型具有更高的检测精度。针对缺陷间的类别相关性与一般对象间的类别相关性不同,Wang等人(2022)提出了一种类别相关非极大值抑制方法(category-related non-maximum suppression, CR-NMS)和一种双阶段边界框回归算法以去除重复的边界框。基于该算法,分别构建了Faster R-CNN、RetinaNet、Cascade R-CNN和Mask R-CNN缺陷检测器,并在公共数据集上训练以及测试性能。

3.1.3 基于Transformer网络

Wei和Wang(2022)提出了一种新的多尺度信息融合Transformer框架MSF-Trans,该架构使用多头注意力机制对WM全局特征进行编码,对缺陷图案之间的内部关系进行建模。MSF-Trans在细节特征与全局特征学习中充分融合CNN与Transformer的优势,对混合型缺陷展现出优异的识别能力。Song等人(2022)介绍了一种基于Transformer的DETR算法用于晶圆的缺陷检测。在现有数据集的基础上对网络进行再训练,调整参数。实验结果表明,其与改进的Mask R-CNN网络相比,精度提高了8.4%。Shukla(2023)提出了一种新型的可变性卷积Transformer(deformable convolutional Transformer, DC Transformer),可变形卷积头能帮助整个网络关注WM中全局特征的学习,忽略非重要特征,并能对WM与缺陷间的内部关系进行简单建模。表7整理了晶圆缺陷检测中主要的几种检测网络的性能与优缺点。

3.2 晶圆缺陷检测的分类网络

分类网络将输入数据分为不同的预定义类别。模型通过学习样本的特征和类别之间的潜在联系进

表 7 主要的检测网络对比
Table 7 Comparison of major detection networks

模型	数据	准确率	贡献/局限性
SSD	WM-811K	0.920 0	解决混合型缺陷分类与定位问题,但未用真实数据验证。
YOLOv3	模拟数据	0.851 1	只能检测垂直、平行直线区域的缺陷,大尺寸污染物会被识别为两个污染物。
改进的 YOLOv4	真实数据	0.995 3	增强算法的特征提取能力,提高精度,但模型的泛化性较差,识别种类较少。
YOLOv3 YOLOv4	WM-811K	0.944 0 0.957 0	启发基于深度学习模型的缺陷定位与分类的研究,提高检测效率与精度,实现缺陷定位。
Mask R-CNN	MixedWM38	0.807 0	融合残差网络和特征金字塔,提高了特征提取的效率,对微小缺陷识别率更高。
CR-NMS+ Mask R-CNN	真实数据	0.923 0	有效抑制冗余边界框,降低不确定判断类的比例,对未来神经网络的优化有一定参考意义。
MSF-Trans	真实数据	0.971 7	混合型缺陷识别效果良好,缺陷特征的学习具备选择性。
DETR	真实数据	-	使用迁移学习权值和微调参数处理晶圆图,显著提高检测精度与检测性能的稳定性。
DC Transformer	MixedWM38	0.973 3	混合型缺陷识别效果良好,更有效地学习全局特征。

注:“-”表示暂无数据。

行分类预测,对不同的缺陷模式标注对应的类别标签。本节按照标签分类数量、晶圆缺陷检测的分类网络分为单标签缺陷分类、多标签缺陷分类展开介绍。

3.2.1 单标签缺陷分类

单标签缺陷分类旨在将输入的 WM 样本区分为好坏晶圆,描述其存在的单种缺陷模式。Shih 等人(2020)提出基于深度学习的再生晶圆缺陷分类方法,采用 MLP、CNN 和 ResNet 三种神经网络进行分类比较,CNN 和 ResNet 在晶圆缺陷的识别中表现出优异的性能,大规模的 MLP 在合理的网络结构下也获得了良好的效果。Di Bella 等人(2019)采用子流形稀疏卷积网络(submanifold sparse convolutional network, SSCN)作为稀疏数据的二元分类器。数据集从意大利 Agrate Brianza 的 STMicroelectronics 获得,包含 12 个类别的 29 746 个 WM。图像分辨率为 20 000 × 20 000 像素。所有缺陷类通过旋转、水平翻转、平移和噪声注入等图像变换进行扩展。Nakata 等人(2017)结合数据挖掘技术和机器学习技术,提出了可再训练的 CNN 分类器,具有 5 层结构,但其只能对单类型缺陷进行识别。Du 和 Shi(2020)提出一种易于训练且具有高识别率的 DCNN 分类模型,其替换了平坦层和完全连接层,以减少参数集,提高了训练效率。

Kim 等人(2020b)提出了一种使用带有残差单

元的 CNN 分类网络,并应用分布外图像检测技术来执行 WM 分类,其训练数据集包含有 13 种缺陷模式的真实二进制 WM。Batool 等人(2020)介绍了用于晶圆缺陷分类的 8 层 CNN 结构,与现有的机器学习检测方法相比,准确率提高了 12%。Alawieh 等人(2020)提出一种基于 CAE 模型的数据增强框架,用于合成样本的生成。通过深度选择性学习来构建具有预测和选择输出的 CNN,在成功检测新缺陷类型的同时,Tsai 和 Lee(2020a)提出一种基于深度可分离卷积的减权分类器,通过结合深度可分离卷积(depthwise separable convolution, DSC),提高运行时间并减少了过拟合,并且模型参数比标准卷积神经网络少。为了增加稀缺训练数据的数量和多样性,Yang 和 Sun(2022)提出基于量子电路的混合经典量子深度学习框架(hybrid classical-quantum deep learning, HCQDL),用于 WM 缺陷模式分类、热点检测等,探索了具有不同可表达性和纠缠能力的参数化量子电路。表 8 总结了主要的单标签晶圆缺陷分类网络。

3.2.2 多标签缺陷分类

1)DCNN。Maksim 等人(2019)提出一种新的学习 DCNN 模型的策略,使用随机梯度下降算法对学习参数进行自适应评估。选取 VGG-19、ResNet-34、ResNet50 和 MobileNetV2(Howard 等,2017),比较它们对晶圆缺陷分类的能力,其中 ResNet50 分类准确

表 8 主要的单标签缺陷分类网络
Table 8 Main single-label defect classification networks

模型	数据	准确率	贡献/局限性
MLP,CNN,ResNet	真实数据	0.998 2	鲁棒性好,解决过拟合问题,学习速度快, CNN 与其他网络共享权值和偏差特征,具有局部接受野特征。
SSCN	ST数据集 WM-811K	0.990 2	解决样本稀缺性问题,对 WBM 实现有效的数据增强。
CNN	真实数据	0.934 0	由大数据分析实现长期监控自动化,分类精度高,应用前景广。
WMDP-DCNN	WM-811K	0.946 8	分类准确度,优于基于人工设计特征和多参数网络。
CNN	真实数据	0.971 8	使用欠采样法来解决类不平衡问题,但根本原因分析的准确性低。
CNN	WM-811K	0.904 4	欠采样解决类不平衡问题,但会丢失部分样本数据。
CNN+CAE	WM-811K	0.940 0	解决类不平衡问题,分类精度高,能高选择性学习新缺陷类型。
CNN+DSC	真实和模拟数据	0.966 3	通过分层学习生产强特征,但训练成本高,参数调整复杂。
HCQDL	WM-811K	0.981 0	体现参数化量子电路的潜在用途与量子学习的新颖性,启发半导体晶圆缺陷检测的量子深度学习方面的技术路线。

度最佳。Hsu 和 Chien (2022) 提出基于 LeNet、AlexNet (Krizhevsky 等, 2017) 和 GoogLeNet 的端到端的集成卷积神经网络 (ensemble convolutional neural network, ECNN) 模型, 采用加权多数函数预测输出, 为基础模型赋予更多权重因子, 以提高性能。Yu 等人 (2021c) 提出一种基于 DenseNet (Huang 等, 2018) 模型的 WBM 缺陷模式检测方法, 根据 WBM 的特点, 改进模型的结构与其训练损失函数, 并提出一种约束均值滤波算法来滤除噪声颗粒。

2) 结合其他算法的 CNN。Geng 等人 (2021) 提出另一种用于 WM-811K 的 CNN 分类器, 其使用了少样本学习。所提出的分类框架分别包含标记数据和未标记数据的少样本损失和自监督损失。Nakazawa 和 Kulkarni (2018) 提出一种基于 CNN 的模式分类方法, 但模型难以区分相似的缺陷模式。Kyeong 和 Kim (2018) 开发了多个 CNN 模型进行混合型缺陷模式分类, 为每个单模式缺陷建立单独的模型。Saqlain 等人 (2020) 采用随机旋转、水平翻转、宽度偏移、高度偏移、剪切和缩放等变化操作, 使数据多样化。实验结果表明, 在分类准确率方面, 所提出的用于晶圆缺陷自动识别的卷积神经网络 (convolutional neural network for automatic wafer defect identification, CNN-WDI) 模型优于其他模型。Zhang 等人 (2022) 提出基于二值化神经网络的 WDP-BNN 框架, 并使用随机欠采样方法与数据增强技术, 解决因二值化引起的不平衡问题与性能损失。Devika 和

George (2019) 使用 CNN 对 4 种基本单类型缺陷模式进行训练, 其可以从真实数据集中检测基本缺陷类型的组合。

Cheon 等人 (2019) 提出一种结合 KNN 与 CNN 模型的方法, 可对未知缺陷类分类。Byun 和 Baek (2020) 使用 CAE 初始化 CNN 权重。该模型对单类型缺陷的分类性能良好, 但对混合类型缺陷分类效果有待提升。Shim 等人 (2020) 提出一种通过不确定性采样进行主动学习的 CNN, 用于 WM 缺陷分类。结果表明通过贝叶斯不一致主动学习 (Bayesian active learning by disagreement, BALD) 的 CNN 错误率低, 精度高, 参数较少, 训练和验证时间都较短。Jang 等人 (2020) 应用旋转和水平翻转操作创建多样化、旋转不变的 WM。在训练数据上通过高斯过程调整 CNN 超参数, 以防止高复杂度模型导致的过拟合出现。Chen 等人 (2020) 研究了一种基于 CNN 快速准确的晶圆缺陷检测方法 WDD-Net (wafer defect detection network), 相比 VGG-16 和 MobileNet-v2, 检测速度更快且模型尺寸减小, 该方法的检测精度大于 99%。Hyun 和 Kim (2020) 针对包含了高度不平衡的混合型缺陷模式的 WBM 数据, 提出了记忆增强 CNN。

3) 其他网络。Wang 等人 (2020) 使用了 5 块深度可变性卷积网络 (deep deformable convolutional networks, DCN) 进行多标签分类, 并介绍了一个基于真实和模拟的 WM 的数据集 MixedWM38, 其有

38种缺陷模式。Yu等人(2021b)从WM提取的Radon特征作为自组织增量神经网络(self-organizing incremental neural network, SOINN)模型的输入。该模型能判断是否需要增加神经元,并识别新的未知缺陷模式。Cha等人(2020)使用Xception(extreme inception)模型对晶圆缺陷进行分类,并使用CAE进行数据扩充。Yu等人(2023)提出一种基于PeeleNet模块的轻量级网络WM-PeeleNet,通过在CAE的隐藏层中添加随机高斯噪声实现晶圆数据增强,使少数的缺陷类别的识别精度得到显著提高。

生成对抗网络(generative adversarial network, GAN)也被应用于晶圆多标签缺陷分类任务,其由两个CNN组成:生成器和鉴别器。生成器学习真实数据和模拟数据,鉴别器负责区分真实数据和模拟数据。用于晶圆缺陷的研究较少,只有3项研究探讨了用于晶圆缺陷分类和数据生成的GAN。李阳和蒋三新(2023)提出了一种改进的GAN无监督晶圆表面缺陷检测模型,使用带有两层跳跃连接和记忆模块的编码器—解码器卷积神经网络来搭建生成器,跳跃连接用于捕获多尺度的图像特征,记忆模块对潜在特征实施约束,扩大真实缺陷样本与重构样本间距离。

Wang等人(2019)提出了一种改进的自适应平衡生成对抗性网络(adaptive balancing generative adversarial network, AdaBalGAN)。该框架由两部分组成:分类生成对抗网络和自适应生成控制器。该控制器通过考虑缺陷类的大小以及每个缺陷类与多数缺陷类的识别准确率差异来识别少数缺陷类。通过识别不平衡的类分布,自适应生成控制器自动调整每种缺陷类型的合成晶圆图的数量,该方法的分类准确率达到96%。Ji和Lee(2020)开发了一种使用GAN做数据增强的深度卷积分类器,它复合了多个卷积层的图像处理能力。与经典扩增相比,该模型的性能从97.0%显著提高至98.3%。

表9归纳了主要的多标签晶圆缺陷分类网络。

3.3 晶圆缺陷检测的分割网络

分割网络负责从复杂背景中分割出晶圆缺陷,其通常采用编码器—解码器结构。编码器负责提取晶圆缺陷的纹理、颜色等特征,解码器则将特征映射回原始图像尺寸中,并产生像素级的预测结果。目前在晶圆缺陷检测中,主要的分割网络包括Seg-

Net、U-Net、Mask R-CNN和DeepLeb等模型。本节主要介绍上述分割模型在晶圆缺陷检测中的应用与改进。

Nakazawa和Kulkarni(2019)研究基于SegNet、U-Net和全卷积网络的CNN自动编码器架构,可以检测和分割WM缺陷图案。研究结果表明,SegNet和U-Net表现出相似的训练精度,而基于全卷积网络的精度较低。Cha和Jeong(2022)提出了一种改进的U-Net模型。该模型将注意力机制与残差块相结合来分割混合缺陷,通过抑制不相关特征来提取改进后的特征图,有效分割单类型与混合型缺陷,但模型的自学习能力较弱。Han等人(2020)提出了一种基于深度卷积网络的多晶硅晶圆缺陷分割方法。首先应用区域生成网络(region proposal network, RPN)生成潜在缺陷区域,然后将其反馈至基于U-Net和膨胀卷积的改进的分割网络。针对上文无法有效检测微小缺陷的问题,Wen等人(2020)开发了一种具有空洞卷积的特征金字塔网络(feature pyramid networks with atrous convolution, FPNAC),来提取特征并生成特征图,然后将特征图输入至RPN中以生成候选区域,作为深度多分支神经网络(deep multi-branches neural network, DMBNN)的输入。该框架能精确地对缺陷进行分类和分割,但需要大量带标签数据的支撑。对于其数据依赖性,Chiu和Chen(2021)将数据增强与Mask R-CNN实例分割模型相结合,在有限训练数据下,能在WM上对缺陷模式进行精确地分类与定位,并进一步追溯缺陷成因,但所提出的模型泛化性弱。因此Chen等人(2019)提出了一种具有强泛化性的模型,可用于其他领域,即鲁棒多尺度特征显著性图(multi-scale feature saliency map, MFSM)来精确分割位错缺陷,采用参数优化大气散射模型(parameter-optimized atmospheric scattering Model, PASM)增强图像对比度,保留缺陷区域信息。

表10总结了用于晶圆缺陷检测的分割网络性能特点,其性能与样本集质量密切相关。

3.4 晶圆缺陷检测的组合网络

组合网络是一种集成学习方法,通过将多个独立的模型组合,来提高整体的预测性能。组合网络通过结合多个模型的预测结果,可以获得更准确、稳定和鲁棒的预测结果。

Yang和Sun(2020)提出基于VGG16、Inception

表 9 主要的多标签缺陷分类网络
Table 9 Main multi-label defect classification networks

类别	模型	数据	准确率	贡献/局限性
DCNN	ResNet50	WM-811K	0.878 0	准确性高,但存在数据稀缺问题。
	ECNN	WM-811K	0.985 7	解决了类不平衡问题,模型的泛化性较差。
	DenseNet	WM-811K	0.887 0	计算 WBM 识别模型的不确定性,降低分类错误率,提高模型的自适应力。
结合其他算法的 CNN	CNN	WM-811K	0.966 9	解决类不平衡问题,支持检测复杂缺陷。
	CNN	真实、模拟数据	0.982 0	噪声鲁棒性好,检测性能良好。
	CNN	真实、模拟数据	0.974 0	混合缺陷模式分类精度高,对噪声有鲁棒性。
	CNN-WDI	WM-811K	0.962 0	克服类不平衡与过拟合问题,分类准确度高。
	WDP-BNN	WM-811K	0.948 3	解决类不平衡问题,由对比实验和 PCA 验证数据增强方法的有效性。
	CNN	真实、模拟数据	0.941 4	对噪声有鲁棒性,分类精度高。
	CNN	真实数据	0.962 0	分类精度高,但计算负担大。
	CNN	WM-811K	-	单缺陷检测表现良好,混合缺陷表现不佳。
	OVO-ECNN	真实数据	0.948 0	有更高的扩展性,未知缺陷模式检测率提高。
	WDD-Net	真实数据	0.994 4	提高检测速度与精度,存在类不平衡问题,需要大量标注数据。
其他类型	记忆增强 CNN	模拟数据	-	解决类不平衡问题,分类性能受类数量限制。
	DCN	MixedWM38	0.932 0	解决了类不平衡问题,多缺陷识别表现良好,但未实现缺陷成因追溯。
	SOINN	WM-811K	0.933 3	提高检测效率,降低测试成本。
	RGRN	模拟数据	0.998 0	提高模型泛化能力与分类精度,应用随机子空间学习和自举种子技术。
	CAE+Xception	WM-811K	0.999 0	解决了类不平衡问题,分类准确。
	WM-PeleeNet	WM-811K	0.954 0	部分缺陷模式存在错误分类,训练成本大,无法对小样本 WM 分类。
		MixedWM38		
	GAN	真实数据	0.934 0	提高无监督模型的检测性能,难以应用于缺陷精确定位的场景中。
	AdaBalGAN	WM-811K	0.960 0	准确性和稳定性提升,存在类不平衡问题。
	CNN+GAN	WM-811K	0.983 0	提高分类精度,解决类不平衡与数据稀缺问题。

表 10 主要的分割网络对比
Table 10 Main segmentation networks comparison

模型	数据	准确率	贡献/局限性
SegNet, U-Net, FCN	模拟数据	0.943 0 (随机缺陷)	检测真实晶圆缺陷的能力有限,训练数据少,可作为模式提取的上游步骤。
改进的 U-Net	WM-811K MixedWM38	0.980 0	有效检测单类型与混合型缺陷,应用自动缺陷屏蔽技术提高效率,模型学习能力待提高。
RPN+膨胀卷积+U-Net	真实数据	0.834 0	解决类不平衡和样本稀缺问题,训练少量粗标记缺陷图像,能准确分割各类缺陷。
FPNAC+RPN+DMBNN	真实数据	0.939 7	对微小缺陷精确分类和分割,有良好的综合性能,但需要大量带标签的训练数据。
Mask R-CNN	WM-811K	0.977 0	提高产量,及时判断缺陷成因,模型的泛化性较弱,特征提取算法待优化。
PASM-MFSM	真实数据	0.820 0	复杂背景下具有良好的分割精度与鲁棒性,有良好的位错缺陷分割性能,可应用于其他领域。

和 ResNet 组合的网络 RadonNet。该模型使用 Radon 变换进行初次特征提取,然后将预处理数据和

原始数据组合,作为 CNN 分类器的输入,进行第 2 次特征提取。在真实数据集上的实验验证了所提方法

具有较高的缺陷分类性能,识别准确率达到 98.5%。邴鑫和史峥(2018)利用扫描电子显微镜(scanning electron microscope, SEM)图像,采用一种新的ZFNet网络和改进式的Faster R-CNN进行分类。Tello 等人(2018)将随机广义回归网络(random generalized regression neural network, RGRN)与 CNN 模型相结合。RGRN 和 CNN 分别对单类型和复合类型的缺陷进行分类,总体准确率达到 86.17%。Yu 等人(2019b)在框架中使用了两种神经网络:一种用于缺陷模式

的检测,另一种用于分类。8层检测模型将缺陷图像划分为模式化和非模式化,再通过 13 层分类器对缺陷模式进一步分类,最后应用 PCA 降低其维度后,对缺陷进行根本原因分析。Kong 和 Ni(2020a)使用 U-Net 和 CNN 来定位和分割图案化的晶圆,U-Net 用于缺陷边界分割,CNN 用于分类。所提出的两个模型实现了与其他高性能模型相当的分类性能。

表 11 比较了主要的晶圆缺陷组合网络性能、贡献与局限性。

表 11 主要的组合网络对比
Table 11 Major combined networks comparison

目标	模型	数据	准确率	贡献/局限性
分类	RadonNet	真实数据	0.985 0	对晶圆表面洁净度要求高,两次特征提取,识别准确率高。
	ZFNet, Faster R-CNN, Patch-base CNN	模拟数据	F1:0.971 0 (ZFNet)	数据适应能力较强,可准确定位缺陷位置,从扫描电镜图像中缺陷检测多种类型缺陷,耗时短。
	空间滤波+RGRN+ CNN	真实和模拟数据	0.861 7	单类型缺陷分类精度高,高效分类混合型缺陷,但不能实时识别,计算复杂度高,模型调参复杂。
分割 检测	13 层 CNN 分类模型 8 层 CNN 检测模型	WM-811K	0.999 8(检测) 0.931 3(分类)	对特征图进行 PCA 降维和相似度排序,适合检测大规模 WM,有效检测常见缺陷模式,实现缺陷根源分析。
	U-Net, CNN	真实数据	0.852 4(分割) 0.969 0(分类)	边界检测法实现定位和分割,分离、分割重复晶圆缺陷图案,能对混合 WBM 展现良好的定量和定性分析。
分割 分类	WaferSegClassNet	MixedWM38	0.982 0	率先在 MixedWM38 上实现分割,分类精度高但模型泛化性弱。

Nag 等人(2022)提出一种轻量级深度学习架构的新型分割分类组合网络 WSCN(WaferSegClass-Net),如图 5 所示。WSCN 由编码器、解码器组成。

编码器执行下采样,采用可分离卷积,减少内存带宽和计算需求的同时提高了特征表示效率;解码器执行上采样,包含分割分支和分类分支,分割分支中的

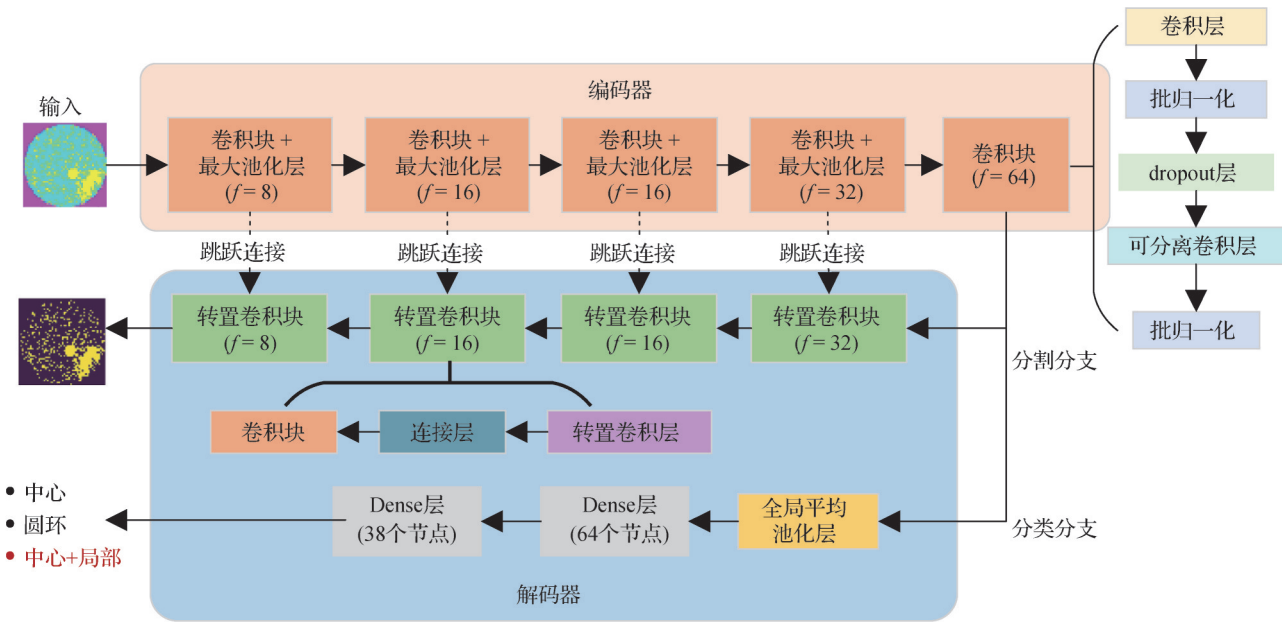


Fig. 5 WaferSegClassNet model structure

转置卷积可补充上采样细节。其使用 N 对对比损失对编码器做预训练,基于BCE-Dice损失进行分割,基于分类交叉熵进行分类。相比轻量级网络,WSCN有更高的分类精确度,虽然精度略小于ResNet50和DenseNet121等大型网络,但WSCN只有0.009 M参数,模型大小也仅为0.51 MB,约为ResNet50模型尺寸的1/200。因其在训练阶段没有引入模型的真实缺陷类,在不可见类上的识别性能有待提升,并且为了缩短训练时间,可以考虑将迁移学习运用在该网络中。

4 晶圆图缺陷数据集及评价方法

4.1 数据集

数据是用于训练、验证和测试模型的输入。晶

圆缺陷数据集来源于真实世界或通过建模获得。晶圆图缺陷数据集通常有两个限制:类严重不平衡、缺乏标签。由于缺陷晶圆出现的频率低于正常晶圆,数据集会出现严重的类别不平衡。而因人为错误,WM数据集会存在不正确的标签(Northcutt等,2021),导致未标记或错误标记的WM出现。随着强大的深度生成模型的出现,许多研究已经将生成模型应用于WM数据生成中(Lee等,2023;Ji和Lee,2020)。本文搜集了所有目前已公开的数据集和部分私有数据集,如表12所示。部分数据集样例如图6所示。

WM-811K数据集(Wu等,2015)是已知的最大的公用晶圆数据集,由从现实世界中的46 393个制造批次中收集的811 457个WM和其他制作工艺数据组成,在该领域研究工作中占有重要地位,广泛用

表 12 晶圆表面缺陷检测的常用数据集

Table 12 Common datasets for surface defect detection in wafer inspection

数据集	规模	类别	标记情况	分辨率/像素	GFA分布	网络模型	年份	下载链接
WM-811K	811 457	9	已标记:0.031 未标记:0.969	15~200	单GFA	SVM	2015	http://mirlab.org/dataSet/public/
MixedWM38	38 015	38	-	52 × 52	混合GFA	DC-Net	2020	https://github.com/Junliangwangdhu/WaferMap
21-Defect	16 388	21	已标记:0.638 未标记:0.362	64 × 64	单GFA	DMC	2020	未公开
MBWBM	1 050	9	-	35 × 19	混合GFA	3D U-Net	2022	https://ieee-dataport.org/documents/synthesized-multi-gfa-multi-bin-wafer-bin-map-dataset
ST	29 746	12	-	-	单GFA	SSCN	2019	未公开

注:“-”表示暂无数据。

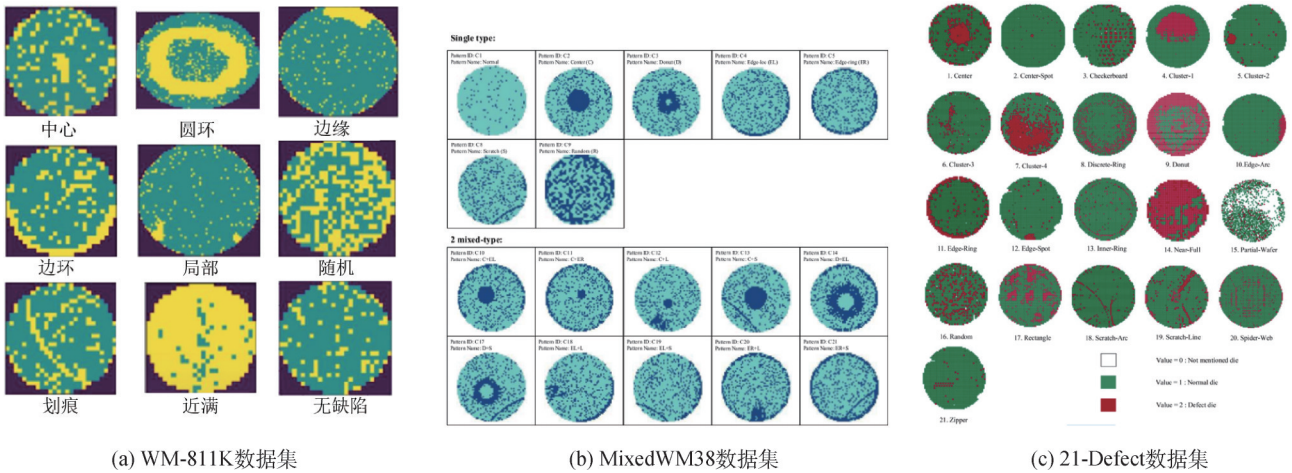


图6 常用的数据集示例

Fig. 6 Commonly used dataset examples ((a) WM-811K; (b) MixedWM38; (c) 21-Defect)

于训练和测试模型任务中。其组成主要包含训练集与测试集,规模分别为54 356幅和118 595幅。其中仅有172 950幅图像带有手动标签,前8种缺陷模式的数据分布分别是:中心(Center)4 294幅,圆环(Donut)555幅,边缘(EdgeLoc)5 189幅,边环(EdgeRing)9 680幅,位置(Loc)3 593幅,近满(Near-full)149幅,随机(Random)866幅,划痕(Scratch)1 193幅。如图6(a)所示,图像分辨率的范围为15×15像素到200×200像素。虽然WM-811K数据量大但是图像质量不好,识别类别单一,所以使用该数据集来进行实验,效果可能不尽人意。

MixedWM38数据集(Wang等,2020)是另一个公开可用的数据集,如图6(b)所示,它由38 015个混合型缺陷的WBM组成,该数据集有38种缺陷类型,其中有29种2~4个模式混合的混合型缺陷,8种单模式缺陷和1个无缺陷。由于一些混合型晶圆缺陷图案的样本不足,采用了模拟方法来生成样本,以保证各种类型的缺陷样本间的平衡。

21-Defect数据集(Tsai和Lee,2020b)的数据质量与WM-811K相同,如图6(c)所示,由Tsai团队从真实WM中提取获得,共有16 388个WM,其中63.8%的WM(10 460个晶圆)为有标签缺陷,36.2%的晶圆(5 928个晶圆)没有标签。

MBWM数据集(Wang和Ni,2023)由浙江大学的倪东和王毅通过使用训练过的pix2pix模型,对WM-811K的WBM数据集和真实世界的多种不同的二进制WBM进行合成而得到。该数据集包含1 050个多模式多bin的WBM,pix2pix是一种GAN模型,其中鉴别器负责对真假MBWM的分类,WM-811K的WBM样本馈送到训练好的生成器中,生成单模式的MBWM,然后将其叠加获得合成的MBWM。

ST(STMicroelectronics)数据集(Di Bella等,2019)由意法半导体在意大利Abrate Brianza的生产基地获得的29 746个WM组成,包括了12种缺陷类型与标准晶圆图像,但目前未对外界公开。

4.2 评价方法

机器学习和深度学习的评估指标取决于数据和学习方法。在WM缺陷检测文献中,已确定并总结了常见的性能评估指标(Li和Wang,2021;Saqlain等,2019;Hwang和Kim,2020;Lee和Kim,2020)。变量TP(true positive)、TN(true negative)、FP(false positive)和FN(false negative)分别表示真阳性、真阴性、

假阳性和假阴性。有监督学习算法的评估方法通过比较正确和不正确分类的样本数来表示模型的学习效果,主要的单标签缺陷分类评价指标包括:准确率(accuracy, Acc)、精确度(precision)、召回率(recall)、F1分数。对于多标签缺陷分类问题,可以选取精确匹配率(exact match ratio, EMR)、微精度(micro-precision, MPre)、微召回(micro-recall, MRe)和汉明损失(Hamming loss)等度量进行评价(Lee和Kim,2020;Wang等,2020)。

由于半监督算法在训练过程中利用未标记数据进行标签插补和特征表示学习,因此评估方法与有监督算法相同,都是根据标记的数据进行计算的,如Acc、EMR等。对于无监督的晶圆缺陷检测方法,性能指标通常评估缺陷聚类的结果。以下可以作为无监督缺陷检测算法的常用评估指标:随机指数(rand index, RI)、调整后的随机指数(adjusted rand index, ARI)、归一化互信息(normalized mutual information, NMI)、调整后互信息(adjusted mutual information, AMI)、纯度(purity),侧重于通过相似性和信息共享程度来比较聚类。表13整理了在晶圆缺陷检测中常用的评价指标。

5 现存问题与未来展望

5.1 现存问题

尽管将机器学习、深度学习技术应用于晶圆缺陷检测中,能够有效地提高精度与效率,使检测系统变得更智能,但在将理论方法应用于实际中时,仍遇到了诸多问题与挑战。当前半导体晶圆缺陷的机器视觉方法中存在的问题主要有:

1)数据可用性问题。由于晶圆缺陷数据集采集成本、时间、难度等原因,导致算法在训练过程中缺乏足够的样本来学习缺陷的特征和变化模式。经过梳理后,可将晶圆缺陷数据集的可用性问题整理为以下几点:(1)不平衡的类别分布。晶圆缺陷数据集中不同类别的缺陷可能存在不平衡的情况。(2)缺少标注和真实标签。晶圆缺陷数据集中可能存在缺少标注或者标注不准确的问题,使得训练算法的难度增加。(3)缺乏时间序列数据,晶圆缺陷数据集缺乏针对同一晶圆在不同时间点采集的数据。

2)混合型缺陷检测问题。目前大多数现有文献都集中于单一缺陷检测的研究,虽然存在混合类型

表 13 晶圆缺陷检测方法的主要评价指标
Table 13 Main evaluation indicators of wafer defect detection methods

分类	评价指标	计算公式	含义
有监督、 半监督 学习	准确率 Accuracy	$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$	正确识别的 WM 缺陷样本总数, TP 表示正确预测为缺陷晶圆的样本数, TN 表示正确预测为正常晶圆的样本数, FP 表示错误预测为缺陷晶圆的样本数, FN 表示错误预测为正常晶圆的样本数。
	精确率 Precision	$Precision = \frac{TP}{TP + FP}, Precision_i = \frac{TP_i}{TP_i + FP_i}$	正确识别的 WM 缺陷样本数与所有已识别的 WM 缺陷样本数的比值, i 表示某种缺陷模式。
	召回率 Recall	$Recall = \frac{TP}{TP + FN}, Recall_i = \frac{TP_i}{TP_i + FN_i}$	正确识别的 WM 缺陷样本数与实际 WM 缺陷样本数的比值, i 表示某种缺陷模式。
	F1 分数	$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$	精确率与召回率的加权平均值, 表示预测值和基本真实值的接近程度, 用于评估多类别分类性能。
	微精度 MPre	$MPre = \frac{\sum_{i=1}^N \sum_{j=1}^C y_i^j \hat{y}_i^j}{\sum_{i=1}^N \sum_{j=1}^C \hat{y}_i^j} = \frac{\sum_j TP^j}{\sum_j TP^j + \sum_j FP^j}$	表示某种缺陷模式精确率的平均值。 N 和 C 分别表示样本的总数和已知类的总数, y_i 和 $\hat{y}_i$ 分别表示真实标签和分类标签, y_i^j 和 $\hat{y}_i^j$ 表示第 j 类下的真实标签和分类标签。
	微召回 MRe	$MRe = \frac{\sum_{i=1}^N \sum_{j=1}^C y_i^j \hat{y}_i^j}{\sum_{i=1}^N \sum_{j=1}^C \hat{y}_i^j} = \frac{\sum_j TP^j}{\sum_j TP^j + \sum_j FN^j}$	表示某种缺陷模式召回率的平均值。
	精确匹配率 EMR	$EMR = \frac{1}{N} \sum_{i=1}^N I(y_i = \hat{y}_i)$	计算方式与精确率相似, 是所有正确识别类的度量。
无监督 学习	汉明损失	$Hamming\ loss = \frac{1}{NC} \sum_{i=1}^N \sum_{j=1}^C I(y_i \neq \hat{y}_i)$	反映在单标签缺陷中, 错误预测的标签与标签总数的比例, 当预测标签与实际标签不匹配时, 汉明损失函数评估值为 1。
	随机指数 RI	$RI = \frac{a + b}{\frac{n}{2}} = \frac{TP + TN}{TP + FP + FN + TN}$	校正相似对 a 和校正不相似对 b 的数量与所有可能的组合对的比率, n 表示样本数量。
	调整后随机指数 ARI	$ARI = \frac{RI - E[RI]}{1 - E[RI]}$	经过调整的 RI, 以消除聚类、样本的数量的影响。随机样本时值接近 0, 高相似度的样本值靠近 1。
	归一化互信息 NMI	$NMI = \frac{I(X; Y)}{\sqrt{H(X)H(Y)}}$	$I(X; Y)$, $H(X)$ 和 $H(Y)$ 分别表示 X 和 Y 之间的互信息以及 X 和 Y 的熵, 衡量两种聚类之间的相似性。
	调整后互信息 AMI	$AMI = \frac{MI - E[I(X; Y)]}{\max(H(X), H(Y)) - E[I(X; Y)]}$	经过调整后的 MI, 以消除类标签对度量分数的影响。
	纯度 Purity	$Purity = \frac{1}{N} \sum_{i=1}^k \max_j c_j \cap t_j $	累积每个簇中分配至频数最高的簇中的样本数除以样本总数 N , 以衡量聚类算法的分类准确性。

的缺陷检测算法, 但大多数算法在标记数据的可用性、缺陷模式类型的范围和计算的负载等方面受限。其中常见的问题包括以下几点: (1) 缺陷定位与分类。混合型晶圆缺陷由不同类型的缺陷组成, 需要对缺陷进行分类和定位。(2) 缺陷稳定性分析。在检测到混合型缺陷后, 需要长期监测缺陷是否会随着时间的推移而增加或变化。(3) 缺陷原因追溯。对于混合型缺陷, 可以通过分析材料、回顾工艺参数等手

段, 追溯其来源和形成机制。

3) 有效特征表示学习。晶圆缺陷的有效特征表示是指将 WM 中的缺陷信息转化为具有区分性和表达能力的特征向量或特征描述子, 以便进行缺陷检测、分类和识别。以下是晶圆缺陷的有效特征表示可能面临的问题: (1) 多尺度问题。晶圆缺陷可能具有不同的尺度和尺寸。对于同一尺度的缺陷, 存在形状和纹理上的变化。(2) 高维数据问题。WM 数

据通常具有高维度的特征表示。高维数据可能引起计算量和存储量的增加,并可能影响算法的性能和效率。(3)非确定性和噪声问题。WM数据可能受到噪声、光照变化和图像失真等因素的影响。这些不确定性因素会影响缺陷的特征表示。

5.2 未来展望

基于机器视觉的晶圆缺陷检测技术能很好地检测出微小且罕见的缺陷,与人工检测方法相比整体效率更高。基于学习的方法检测精度更好、分类效果更佳,但相比传统的图像处理,需要考虑数据集与训练成本的影响。虽然机器视觉技术在晶圆缺陷检测领域已经得到广泛应用,但还是存在特征有效表示学习困难、可用数据集少、混合缺陷不稳定以及训练成本高等问题。期望未来能从以下工作展开:

1)更有效的特征提取。基于形态学、边缘检测和纹理分析等图像处理技术,有效地提取与缺陷相关的特征。通过结合多种特征表示方法,如形状特征、颜色特征等,以获取更全面和多样化的缺陷信息。结合领域知识和先验信息,对特定模式的晶圆缺陷设计特征表示法,以提高最终的检测和分类的准确性。

2)多模态与无损检测。多模态检测可以提供更多的表征参数和多角度观察,从而提高缺陷检测的可靠性。未来的晶圆缺陷检测可能会结合多种检测技术,例如光学、电子显微镜、红外等,以获取更全面和准确的缺陷信息。目前的晶圆缺陷检测主要依赖于表面显微镜和断层分析等技术,这些方法可能对晶圆造成一定的损伤。未来的发展将致力于开发更多的无损检测技术,以减少对晶圆的影响,并提高检测效率和可重复性。

3)计算复杂度的有效降低。数据降维技术将越来越多地应用于未来的晶圆缺陷检测中,此外,通过采用在线学习和增量学习的方法,系统能够灵活地更新模型和算法,以适应不断变化的晶圆制造环境和缺陷类型,从而降低重新计算的需求。同时,考虑运用知识蒸馏技术对模型进行压缩,可以构建精度和性能不受损的轻量化小模型。这一整合方法为晶圆缺陷检测提供了更为高效和灵活的解决方案。

4)晶圆缺陷数据集的扩充。能够反映多种晶圆缺陷模式的大型公有数据集有限,其中最典型的WM-811K数据集,在晶圆尺寸、IC尺寸等数据上已经不能满足现有的研究工作,并存在大量未标记的

晶圆缺陷模式。私有数据能很好地反映现行的设计标准,但由于访问的限制,减缓了WMDD的创新和研究。另外,模拟的合成晶圆图也与真实缺陷模式存在差距,提高合成数据与真实缺陷模式相似度的同时,人工晶圆应该能反映更真实的类大小和缺陷形态,因此希望有更多的大型晶圆缺陷数据集得到发布。

6 结 语

本文搜集了关于晶圆缺陷检测方面的相关文献,并对不同的研究方法进行了比较。晶圆的缺陷模式主要可以分为随机、局部、边缘、中心、边环和划痕等类型,集中在研磨、抛光和化学蚀刻等过程产生。在晶圆缺陷的图像处理的检测方法中,大部分方法基于图像的空间域进行Hough变换、分割与检测,基于图像变换域进行滤波与特征提取。但随着缺陷种类的增加与制作工艺的复杂化,传统图像处理应对缺陷提取与分类越加困难。基于机器学习和深度学习的算法在不断发展,涉及的学习算法主要包括有监督学习、无监督学习、混合学习以及其他学习算法,主要解决了分类、聚类、类不平衡、实例标记、数据合成和特征提取等具体的问题,网络结构包括CNN、GAN、GRN、DBN和AE等。学习模型的训练需要数据集的支撑,目前公用的大型数据集较少,数据扩充与数据校正慢,而私有数据维护更好。现阶段的问题主要集中于可用数据集量少、高计算量、类不平衡与模型的优化等。最后希望本文能为后续的研究提供有价值的参考。

参考文献(References)

- Abdullah M, Rahman M H and Akhter S. 2021. Pattern recognition in analog wafermaps with multiple ensemble approaches//Proceedings of the 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST). DHAKA, Bangladesh: IEEE: 587-591 [DOI: 10.1109/icrest51555.2021.9331084]
- Alawieh M B, Boning D and Pan D Z. 2020. Wafer map defect patterns classification using deep selective learning//Proceedings of the 57th ACM/IEEE Design Automation Conference (DAC). San Francisco, USA: IEEE: 1-6 [DOI: 10.1109/dac18072.2020.9218580]
- Batool U, Shapiai M I, Fauzi H and Fong J X. 2020. Convolutional neural network for imbalanced data classification of silicon wafer

- defects//Proceedings of the 16th IEEE International Colloquium on Signal Processing and Its Applications (CSPA). Langkawi, Malaysia: IEEE: 230-235 [DOI: 10.1109/cspa48992.2020.9068669]
- Batool U, Shapiai M I, Tahir M, Ismail Z H, Zakaria N J and Elakharany A. 2021. A systematic review of deep learning for silicon wafer defect recognition. *IEEE Access*, 9: 116572-116593 [DOI: 10.1109/ACCESS.2021.3106171]
- Bollmann W. 2012. *Crystal Defects and Crystalline Interfaces*. Berlin, Heidelberg, Germany: Springer [DOI: 10.1007/978-3-642-49173-3]
- Byun Y and Baek J G. 2020. Mixed pattern recognition methodology on wafer maps with pre-trained convolutional neural networks//Proceedings of the 12th International Conference on Agents and Artificial Intelligence. Valletta, Malta: SciTePress: 974-979 [DOI: 10.5220/0009177909740979]
- Cha J and Jeong J. 2022. Improved U-Net with residual attention block for mixed-defect wafer maps. *Applied Sciences*, 12 (4): #2209 [DOI: 10.3390/app12042209]
- Cha J, Oh S, Kim D and Jeong J. 2020. A defect detection model for imbalanced wafer image data using CAE and Xception//Proceedings of 2020 International Conference on Intelligent Data Science Technologies and Applications (IDSTA). Valencia, Spain: IEEE: 28-33 [DOI: 10.1109/idsta50958.2020.9264135]
- Chen H Y, Liu J L, Wang S and Liu K. 2019. Robust dislocation defects region segmentation for polysilicon wafer image with random texture background. *IEEE Access*, 7: 134318-134329 [DOI: 10.1109/access.2019.2942218]
- Chen J X, Ji Y D, Lin Y Z and Zhu D H. 2020. Wafer etch defect prediction based on hybrid feature selection and hyper-parameter optimization method. *Computer Integrated Manufacturing Systems*, 26(9): 2396-2403 (陈晋贤, 季颖娣, 林义征, 朱定海. 2020. 基于混合特征选择和超参优化的晶圆蚀刻缺陷预测方法. *计算机集成制造系统*, 26(9): 2396-2403) [DOI: 10.13196/j.cims.2020.09.009]
- Chen L L Y, Li K S M, Jiang X H, Wang S J, Huang A Y A, Chen J E, Liang H C and Hsu C L. 2021a. Semi-supervised framework for wafer defect pattern recognition with enhanced labeling//Proceedings of 2021 IEEE International Test Conference (ITC). Anaheim, USA: IEEE: 208-212 [DOI: 10.1109/itc50571.2021.00029]
- Chen S H, Zhang Y X, Yi M L, Ma J and Hou X N. 2021b. Wafer maps defect recognition based on transfer learning of handwritten pre-training network//International Symposium on Computer Technology and Information Science (ISCTIS). Guilin, China: IEEE: 280-283 [DOI: 10.1109/isctis51085.2021.00064]
- Chen X Y, Chen J Y, Han X G, Zhao C D, Zhang D Y, Zhu K F and Su Y J. 2020. A light-weighted CNN model for wafer structural defect detection. *IEEE Access*, 8: 24006-24018 [DOI: 10.1109/access.2020.2970461]
- Cheon S, Lee H, Kim C O and Lee S H. 2019. Convolutional neural network for wafer surface defect classification and the detection of unknown defect class. *IEEE Transactions on Semiconductor Manufacturing*, 32(2): 163-170 [DOI: 10.1109/tsm.2019.2902657]
- Chien J C, Wu M T and Lee J D. 2020. Inspection and classification of semiconductor wafer surface defects using CNN deep learning networks. *Applied Sciences*, 10(15): #5340 [DOI: 10.3390/app10155340]
- Chiu M C and Chen T M. 2021. Applying data augmentation and Mask R-CNN-based instance segmentation method for mixed-type wafer maps defect patterns classification. *IEEE Transactions on Semiconductor Manufacturing*, 34(4): 455-463 [DOI: 10.1109/tsm.2021.3118922]
- Devika B and George N. 2019. Convolutional neural network for semiconductor wafer defect detection//Proceedings of the 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT). Kanpur, India: IEEE: 1-6 [DOI: 10.1109/iccncnt45670.2019.8944584]
- Di Bella R, Carrera D, Rossi B, Fragneto P and Boracchi G. 2019. Wafer defect map classification using sparse convolutional networks//Proceedings of the 20th International Conference on Image Analysis and Processing. Trento, Italy: Springer: 125-136 [DOI: 10.1007/978-3-030-30645-8_12]
- Du D Y and Shi Z. 2020. A wafer map defect pattern classification model based on deep convolutional neural network//Proceedings of the 15th IEEE International Conference on Solid-State and Integrated Circuit Technology (ICSICT). Kunming, China: IEEE: 1-3 [DOI: 10.1109/icsict49897.2020.9278021]
- Ezzat A A, Liu S, Hochbaum D S and Ding Y. 2021. A graph-theoretic approach for spatial filtering and its impact on mixed-type spatial pattern recognition in wafer bin maps. *IEEE Transactions on Semiconductor Manufacturing*, 34(2): 194-206 [DOI: 10.1109/tsm.2021.3062943]
- Fang X and Shi Z. 2018. Wafer defect detection and classification algorithms based on convolutional neural network. *Computer Engineering*, 44(8): 218-223 (邢鑫, 史峥. 2018. 基于卷积神经网络的晶圆缺陷检测与分类算法. *计算机工程*, 44(8): 218-223) [DOI: 10.19678/j.issn.1000-3428.0047107]
- Geng H, Yang F, Zeng X and Yu B. 2021. When wafer failure pattern classification meets few-shot learning and self-supervised learning//Proceedings of 2021 IEEE/ACM International Conference on Computer Aided Design (ICCAD). Munich, Germany: IEEE: 1-8 [DOI: 10.1109/iccad51958.2021.9643518]
- Guo Y Q. 2018. *Research on Visual Inspection of Wafer Surface Defects*. Harbin: Harbin Institute of Technology (郭毅强. 2018. 晶圆表面缺陷视觉检测研究. 哈尔滨: 哈尔滨工业大学)
- Han H, Gao C Q, Zhao Y, Liao S S, Tang L and Li X D. 2020. Polycrystalline silicon wafer defect segmentation based on deep convolutional neural networks. *Pattern Recognition Letters*, 130: 234-241 [DOI: 10.1016/j.patrec.2018.12.013]
- Howard A G, Zhu M L, Chen B, Kalenichenko D, Wang W J, Weyand

- T, Andreetto M and Adam H. 2017. MobileNets: efficient convolutional neural networks for mobile vision applications [EB/OL]. [2024-02-01]. <http://arxiv.org/pdf/1704.04861.pdf>
- Hsu C Y and Chien J C. 2022. Ensemble convolutional neural networks with weighted majority for wafer bin map pattern classification. *Journal of Intelligent Manufacturing*, 33(3): 831-844 [DOI: 10.1007/s10845-020-01687-7]
- Hu H B, He C and Li P. 2021. Semi-supervised wafer map pattern recognition using domain-specific data augmentation and contrastive learning//*Proceedings of 2021 IEEE International Test Conference (ITC)*. Anaheim, USA: IEEE: 113-122 [DOI: 10.1109/itc50571.2021.00019]
- Huang G, Liu Z, Van Der Maaten L and Weinberger K Q. 2018. Densely connected convolutional networks [EB/OL]. [2024-02-01]. <http://arxiv.org/pdf/1608.06993.pdf>
- Hwang J and Kim H. 2020. Variational deep clustering of wafer map patterns. *IEEE Transactions on Semiconductor Manufacturing*, 33(3): 466-475 [DOI: 10.1109/tsm.2020.3004483]
- Hyun Y and Kim H. 2020. Memory-augmented convolutional neural networks with triplet loss for imbalanced wafer defect pattern classification. *IEEE Transactions on Semiconductor Manufacturing*, 33(4): 622-634 [DOI: 10.1109/tsm.2020.3010984]
- Ishida T, Nitta I, Fukuda D and Kanazawa Y. 2019. Deep learning-based wafer-map failure pattern recognition framework//*Proceedings of the 20th International Symposium on Quality Electronic Design (ISQED)*. Santa Clara, USA: IEEE: 291-297 [DOI: 10.1109/isqed.2019.8697407]
- Jang J, Seo M and Kim C O. 2020. Support weighted ensemble model for open set recognition of wafer map defects. *IEEE Transactions on Semiconductor Manufacturing*, 33(4): 635-643 [DOI: 10.1109/tsm.2020.3012183]
- Ji Y S and Lee J H. 2020. Using GAN to improve CNN performance of wafer map defect type classification; yield enhancement//*Proceedings of the 31st Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC)*. Saratoga Springs, USA: IEEE: 1-6 [DOI: 10.1109/asmc49169.2020.9185193]
- Jin C H, Kim H J, Piao Y, Li M J and Piao M. 2020. Wafer map defect pattern classification based on convolutional neural network features and error-correcting output codes. *Journal of Intelligent Manufacturing*, 31(8): 1861-1875 [DOI: 10.1007/s10845-020-01540-x]
- Jin C H, Na H J, Piao M, Pok G and Ryu K H. 2019. A novel DBSCAN-based defect pattern detection and classification framework for wafer bin map. *IEEE Transactions on Semiconductor Manufacturing*, 32(3): 286-292 [DOI: 10.1109/tsm.2019.2916835]
- Kaempf U. 1995. The binomial test: a simple tool to identify process problems. *IEEE Transactions on Semiconductor Manufacturing*, 8(2): 160-166 [DOI: 10.1109/66.382280]
- Kang H and Kang S. 2021. A stacking ensemble classifier with hand-crafted and convolutional features for wafer map pattern classification. *Computers in Industry*, 129: #103450 [DOI: 10.1016/j.comp-ind.2021.103450]
- Kim B, Jeong Y S, Tong S H and Jeong M K. 2020a. A generalised uncertain decision tree for defect classification of multiple wafer maps. *International Journal of Production Research*, 58(9): 2805-2821 [DOI: 10.1080/00207543.2019.1637035]
- Kim T and Behdinan K. 2023. Advances in machine learning and deep learning applications towards wafer map defect recognition and classification: a review. *Journal of Intelligent Manufacturing*, 34(8): 3215-3247 [DOI: 10.1007/s10845-022-01994-1]
- Kim T S, Lee J W, Lee W K and Sohn S Y. 2022. Novel method for detection of mixed-type defect patterns in wafer maps based on a single shot detector algorithm. *Journal of Intelligent Manufacturing*, 33(6): 1715-1724 [DOI: 10.1007/s10845-021-01755-6]
- Kim Y, Cho D and Lee J H. 2020b. Wafer map classifier using deep learning for detecting out-of-distribution failure patterns//*Proceedings of 2020 IEEE International Symposium on the Physical and Failure Analysis of Integrated Circuits (IPFA)*. Singapore, Singapore: IEEE: 1-5 [DOI: 10.1109/ipfa49335.2020.9260877]
- Kong Y T and Ni D. 2018. Semi-supervised classification of wafer map based on ladder network//*Proceedings of the 14th IEEE International Conference on Solid-State and Integrated Circuit Technology (ICSICT)*. Qingdao, China: IEEE: 1-4 [DOI: 10.1109/icsict.2018.8564982]
- Kong Y T and Ni D. 2020a. Qualitative and quantitative analysis of multi-pattern wafer bin maps. *IEEE Transactions on Semiconductor Manufacturing*, 33(4): 578-586 [DOI: 10.1109/tsm.2020.3022431]
- Kong Y T and Ni D. 2020b. A semi-supervised and incremental modeling framework for wafer map classification. *IEEE Transactions on Semiconductor Manufacturing*, 33(1): 62-71 [DOI: 10.1109/tsm.2020.2964581]
- Krizhevsky A, Sutskever I and Hinton G E. 2017. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6): 84-90 [DOI: 10.1145/3065386]
- Kyeong K and Kim H. 2018. Classification of mixed-type defect patterns in wafer bin maps using convolutional neural networks. *IEEE Transactions on Semiconductor Manufacturing*, 31(3): 395-402 [DOI: 10.1109/tsm.2018.2841416]
- Lee H and Kim H. 2020. Semi-supervised multi-label learning for classification of wafer bin maps with mixed-type defect patterns. *IEEE Transactions on Semiconductor Manufacturing*, 33(4): 653-662 [DOI: 10.1109/tsm.2020.3027431]
- Lee H, Kim Y and Kim C O. 2017. A deep learning model for robust wafer fault monitoring with sensor measurement noise. *IEEE Transactions on Semiconductor Manufacturing*, 30(1): 23-31 [DOI: 10.1109/tsm.2016.2628865]
- Lee H, Lee J and Kim H. 2023. Semi-supervised learning for simultaneous location detection and classification of mixed-type defect patterns in wafer bin maps. *IEEE Transactions on Semiconductor*

- Manufacturing, 36(2): 220-230 [DOI: 10.1109/tsm.2023.3264279]
- Lei L J, Sun S L, Xiang Y K, Zhang Y and Liu H K. 2020. Bottleneck issues of computer vision in intelligent manufacturing. *Journal of Image and Graphics*, 25(7): 1330-1343 (雷林建, 孙胜利, 向玉开, 张悦, 刘会凯. 2020. 智能制造中的计算机视觉应用瓶颈问题. *中国图象图形学报*, 25(7): 1330-1343) [DOI: 10.11834/jig.190446]
- Li K S M, Jiang X H, Chen L L Y, Wang S J, Huang A Y A, Chen J E, Liang H C and Hsu C L. 2022. Wafer defect pattern labeling and recognition using semi-supervised learning. *IEEE Transactions on Semiconductor Manufacturing*, 35(2): 291-299 [DOI: 10.1109/tsm.2022.3159246]
- Li Y and Jiang S X. 2023. Unsupervised wafer defect detection based on improved generative adversarial network. *Electronic Measurement Technology*, 46(6): 91-99 (李阳, 蒋三新. 2023. 基于改进生成对抗网络的无监督晶圆缺陷检测. *电子测量技术*, 46(6): 91-99) [DOI: 10.19651/j.cnki.emt.2210828]
- Li Y and Wang J G. 2021. A defect detection method based on improved mask R-CNN for wafer maps//*Proceedings of 2021 International Conference on Computer Network, Electronic and Automation (ICCNEA)*. Xi'an, China: IEEE: 133-137 [DOI: 10.1109/iccnea53019.2021.00038]
- Liao P Y Y, Li K S M, Chen L L Y, Wang S J, Huang A Y A, Cheng K C C, Tsai N C Y and Chou L. 2021. WGrid: wafermap grid pattern recognition with machine learning techniques//*Proceedings of 2021 IEEE International Test Conference (ITC)*. Anaheim, USA: IEEE: 309-313 [DOI: 10.1109/ite50571.2021.00043]
- Liu G L and Yu J B. 2022. Application of neural-symbol model based on stacked denoising auto-encoders in wafer map defect recognition. *Acta Automatica Sinica*, 48(11): 2688-2702 (刘国梁, 余建波. 2022. 基于堆叠降噪自编码器的神经—符号模型及在晶圆表面缺陷识别. *自动化学报*, 48(11): 2688-2702) [DOI: 10.16383/j.aas.c190857]
- Lu X L and Yu J B. 2018. Recognition of wafer defects based on hybrid models and manifold regulation. *Computer Integrated Manufacturing Systems*, 24(2): 302-308 (卢笑蕾, 余建波. 2018. 基于混合模型与流形调节的晶圆表面缺陷识别. *计算机集成制造系统*, 24(2): 302-308) [DOI: 10.13196/j.cims.2018.02.003]
- Maksim K, Kirill B, Eduard Z, Nikita G, Aleksandr B, Arina L, Vladislav S, Daniil M and Nikolay K. 2019. Classification of wafer maps defect based on deep learning methods with small amount of data//*Proceedings of 2019 International Conference on Engineering and Telecommunication (EnT)*. Dolgoprudny, Russia: IEEE: 1-5 [DOI: 10.1109/ent47717.2019.9030550]
- Nag S, Makwana D, R S C T, Mittal S and Mohan C K. 2022. WaferSeg-*ClassNet*: a light-weight network for classification and segmentation of semiconductor wafer defects. *Computers in Industry*, 142: #103720 [DOI: 10.1016/j.compind.2022.103720]
- Nakata K, Orihara R, Mizuoka Y and Takagi K. 2017. A comprehensive big-data-based monitoring system for yield enhancement in semiconductor manufacturing. *IEEE Transactions on Semiconductor Manufacturing*, 30(4): 339-344 [DOI: 10.1109/TSM.2017.2753251]
- Nakazawa T and Kulkarni D V. 2018. Wafer map defect pattern classification and image retrieval using convolutional neural network. *IEEE Transactions on Semiconductor Manufacturing*, 31(2): 309-314 [DOI: 10.1109/tsm.2018.2795466]
- Nakazawa T and Kulkarni D V. 2019. Anomaly detection and segmentation for wafer defect patterns using deep convolutional encoder-decoder neural network architectures in semiconductor manufacturing. *IEEE Transactions on Semiconductor Manufacturing*, 32(2): 250-256 [DOI: 10.1109/tsm.2019.2897690]
- Northcutt C, Jiang L and Chuang I. 2021. Confident learning: estimating uncertainty in dataset labels. *Journal of Artificial Intelligence Research*, 70: 1373-1411 [DOI: 10.1613/jair.1.12125]
- Park S, Jang J and Kim C O. 2021. Discriminative feature learning and cluster-based defect label reconstruction for reducing uncertainty in wafer bin map labels. *Journal of Intelligent Manufacturing*, 32(1): 251-263 [DOI: 10.1007/s10845-020-01571-4]
- Piao M, Jin C H, Lee J Y and Byun J Y. 2018. Decision tree ensemble-based wafer map failure pattern recognition based on radon transform-based features. *IEEE Transactions on Semiconductor Manufacturing*, 31(2): 250-257 [DOI: 10.1109/tsm.2018.2806931]
- Saqlain M, Abbas Q and Lee J Y. 2020. A deep convolutional neural network for wafer defect identification on an imbalanced dataset in semiconductor manufacturing processes. *IEEE Transactions on Semiconductor Manufacturing*, 33(3): 436-444 [DOI: 10.1109/tsm.2020.2994357]
- Saqlain M, Jargalsaikhan B and Lee J Y. 2019. A voting ensemble classifier for wafer map defect patterns identification in semiconductor manufacturing. *IEEE Transactions on Semiconductor Manufacturing*, 32(2): 171-182 [DOI: 10.1109/tsm.2019.2904306]
- Shawon A, Faruk M O, Habib M B and Khan A M. 2019. Silicon wafer map defect classification using deep convolutional neural network with data augmentation//*Proceedings of the 5th IEEE International Conference on Computer and Communications (ICCC)*. Chengdu, China: IEEE: 1995-1999 [DOI: 10.1109/icc47050.2019.9064029]
- Shen Z L and Yu J B. 2020. Wafer map defect recognition based on transfer learning and deep forest. *Journal of Zhejiang University (Engineering Science)*, 54(6): 1228-1239 (沈宗礼, 余建波. 2020. 基于迁移学习与深度森林的晶圆图缺陷识别. *浙江大学学报(工学版)*, 54(6): 1228-1239) [DOI: 10.3785/j.issn.1008-973X.2020.06.021]
- Shi H C, Jin Z Y, Tang W J, Wang J, Jiang K and Xia W. 2022. Research on high precision wafer defect detection based on deep learning. *Journal of Electronic Measurement and Instrumentation*, 36(11): 79-90 (史浩琛, 金致远, 唐文婧, 王静, 蒋楷, 夏伟. 2022. 基于深度学习的高精度晶圆缺陷检测方法研究. *电子测*

- 量与仪器学报, 36(11): 79-90 [DOI: 10.13382/j. jemi. B2205613]
- Shih P C, Hsu C C and Tien F C. 2020. Automatic reclaimed wafer classification using deep learning neural networks. *Symmetry*, 12(5): #705 [DOI: 10.3390/sym12050705]
- Shim J, Kang S and Cho S. 2020. Active learning of convolutional neural network for cost-effective wafer map pattern classification. *IEEE Transactions on Semiconductor Manufacturing*, 33(2): 258-266 [DOI: 10.1109/tsm.2020.2974867]
- Shinde P P, Pai P P and Adiga S P. 2022. Wafer defect localization and classification using deep learning techniques. *IEEE Access*, 10: 39969-39974 [DOI: 10.1109/access.2022.3166512]
- Shon H S, Batbaatar E, Cho W S and Choi S G. 2021. Unsupervised pre-training of imbalanced data for identification of wafer map defect patterns. *IEEE Access*, 9: 52352-52363 [DOI: 10.1109/ACCESS.2021.3068378]
- Shukla N. 2023. Efficient mixed-type wafer defect pattern recognition using compact deformable convolutional transformers. [EB/OL]. [2023-10-17]. <http://arxiv.org/pdf/2303.13827.pdf>
- Song D N, Liu B L and Li Y. 2022. Based on end-to-end object detection algorithm with transformers for detecting wafer maps//Proceedings of 2022 International Conference on Computer Network, Electronic and Automation (ICCNEA). Xi'an, China: IEEE: 297-302 [DOI: 10.1109/iccnea57056.2022.00072]
- Taha K, Salah K and Yoo P D. 2018. Clustering the dominant defective patterns in semiconductor wafer maps. *IEEE Transactions on Semiconductor Manufacturing*, 31(1): 156-165 [DOI: 10.1109/tsm.2017.2768323]
- Tang B, Kong J Y and Wu S Q. 2017. Review of surface defect detection based on machine vision. *Journal of Image and Graphics*, 22(12): 1640-1663 (汤勃, 孔建益, 伍世虔. 2017. 机器视觉表面缺陷检测综述. *中国图象图形学报*, 22(12): 1640-1663) [DOI: 10.11834/jig.160623]
- Tello G, Al-Jarrah O Y, Yoo P D, Al-Hammadi Y, Muhaidat S and Lee U. 2018. Deep-structured machine learning model for the recognition of mixed-defect patterns in semiconductor fabrication processes. *IEEE Transactions on Semiconductor Manufacturing*, 31(2): 315-322 [DOI: 10.1109/tsm.2018.2825482]
- Theodosiou T, Rapti A, Papageorgiou K, Tziolas T, Papageorgiou E, Dimitriou N, Margetis G and Tzovaras D. 2023. A review study on ML-based methods for defect-pattern recognition in wafer maps. *Procedia Computer Science*, 217: 570-583 [DOI: 10.1016/j.procs.2022.12.253]
- Tilli M, Paulasto-Kröckel M, Petzold M, Theuss H, Motooka T and Lindroos V. 2020. Handbook of Silicon based MEMS Materials and Technologies. 3rd ed. Oxford, UK: Elsevier [DOI: 10.1016/c2018-0-01845-9]
- Tsai T H and Lee Y C. 2020a. Wafer map defect classification with depthwise separable convolutions//Proceedings of 2020 IEEE International Conference on Consumer Electronics (ICCE). Las Vegas, USA: IEEE: 1-3 [DOI: 10.1109/icce46568.2020.9043041]
- Tsai T H and Lee Y C. 2020b. A light-weight neural network for wafer map classification based on data augmentation. *IEEE Transactions on Semiconductor Manufacturing*, 33(4): 663-672 [DOI: 10.1109/tsm.2020.3013004]
- Tulala P, Mahyar H, Ghalebi E and Grosu R. 2018. Unsupervised wafermap patterns clustering via variational autoencoders//Proceedings of 2018 International Joint Conference on Neural Networks (IJCNN). Rio de Janeiro, Brazil: IEEE: 1-8 [DOI: 10.1109/ijcnn.2018.8489422]
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser L and Polosukhin I. 2023. Attention is all you need [EB/OL]. [2023-08-02]. <http://arxiv.org/pdf/1706.03762.pdf>
- Wang J L, Yang Z L, Zhang J, Zhang Q H and Chien W T K. 2019. AdaBalGAN: an improved generative adversarial network with imbalanced learning for wafer defective pattern recognition. *IEEE Transactions on Semiconductor Manufacturing*, 32(3): 310-319 [DOI: 10.1109/tsm.2019.2925361]
- Wang J L, Xu C Q, Yang Z L, Zhang J and Li X O. 2020. Deformable convolutional networks for efficient mixed-type wafer defect pattern recognition. *IEEE Transactions on Semiconductor Manufacturing*, 33(4): 587-596 [DOI: 10.1109/tsm.2020.3020985]
- Wang S Y, Zhong Z T, Zhao Y L and Zuo L. 2021. A variational autoencoder enhanced deep learning model for wafer defect imbalanced classification. *IEEE Transactions on Components, Packaging and Manufacturing Technology*, 11(12): 2055-2060 [DOI: 10.1109/tpmt.2021.3126083]
- Wang X Y, Jia X L, Jiang C Y and Jiang S X. 2022. A wafer surface defect detection method built on generic object detection network. *Digital Signal Processing*, 130: #103718 [DOI: 10.1016/j.dsp.2022.103718]
- Wang Y and Ni D. 2019. Multi-bin wafer maps defect patterns classification//Proceedings of 2019 IEEE International Conference on Smart Manufacturing, Industrial and Logistics Engineering (SMILE). Hangzhou, China: IEEE: 48-52 [DOI: 10.1109/smile45626.2019.8965299]
- Wang Y and Ni D. 2023. A deep learning analysis framework for complex wafer bin map classification. *IEEE Transactions on Semiconductor Manufacturing*, 36(3): 367-377 [DOI: 10.1109/tsm.2023.3269230]
- Wei Y X and Wang H. 2022. Mixed-type wafer defect recognition with multi-scale information fusion transformer. *IEEE Transactions on Semiconductor Manufacturing*, 35(2): 341-352 [DOI: 10.1109/tsm.2022.3156583]
- Wen G J, Gao Z J, Cai Q, Wang Y D and Mei S. 2020. A novel method based on deep convolutional neural networks for wafer semiconductor surface defect inspection. *IEEE Transactions on Instrumentation and Measurement*, 69(12): 9668-9680 [DOI: 10.1109/tim.2020.

- 3007292]
- Wu M J, Jang J S R and Chen J L. 2015. Wafer map failure pattern recognition and similarity ranking for large-scale data sets. *IEEE Transactions on Semiconductor Manufacturing*, 28 (1) : 1-12 [DOI: 10.1109/tsm.2014.2364237]
- Yang Y F. 2019. A deep learning model for identification of defect patterns in semiconductor wafer map//*Proceedings of the 30th Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC)*. Saratoga Springs, USA: IEEE: 1-6 [DOI: 10.1109/asmc.2019.8791815]
- Yang Y F and Sun M. 2020. Double feature extraction method for wafer map classification based on convolution neural network//*Proceedings of the 31st Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC)*. Saratoga Springs, USA: IEEE: 1-6 [DOI: 10.1109/asmc49169.2020.9185393]
- Yang Y F and Sun M. 2022. Semiconductor defect detection by hybrid classical-quantum deep learning//*Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New Orleans, USA: IEEE: 2313-2322 [DOI: 10.1109/cvpr52688.2022.00236]
- Yu J B. 2019. Enhanced stacked denoising autoencoder-based feature learning for recognition of wafer map defects. *IEEE Transactions on Semiconductor Manufacturing*, 32 (4) : 613-624 [DOI: 10.1109/tsm.2019.2940334]
- Yu J B and Liu J T. 2021. Two-dimensional principal component analysis-based convolutional autoencoder for wafer map defect detection. *IEEE Transactions on Industrial Electronics*, 68 (9) : 8789-8797 [DOI: 10.1109/tie.2020.3013492]
- Yu J B, Lu X L and Zong W Z. 2016. Wafer defect detection and recognition based on local and nonlocal linear discriminant analysis and dynamic ensemble of Gaussian mixture models. *Acta Automatica Sinica*, 42(1): 47-59 (余建波, 卢笑蕾, 宗卫周. 2016. 基于局部与非局部线性判别分析和高斯混合模型动态集成的晶圆表面缺陷探测与识别. *自动化学报*, 42(1): 47-59) [DOI: 10.16383/j.aas.2016.c150311]
- Yu J B, Zheng X Y and Liu J T. 2019a. Stacked convolutional sparse denoising auto-encoder for identification of defect patterns in semiconductor wafer map. *Computers in Industry*, 109: 121-133 [DOI: 10.1016/j.compind.2019.04.015]
- Yu N G, Chen H S, Xu Q and Hasan M M. 2021a. Wafer map defect recognition with few shot learning based on hybrid self-attention mechanism and prototype network//*Proceedings of 2021 China Automation Congress (CAC)*. Beijing, China: 4128-4134 [DOI: 10.1109/cac53003.2021.9727346]
- Yu N G, Chen H S, Xu Q, Hasan M M and Sie O. 2023. Wafer map defect patterns classification based on a lightweight network and data augmentation. *CAAI Transactions on Intelligence Technology*, 8(3): 1029-1042 [DOI: 10.1049/cit2.12126]
- Yu N G, Xu Q, Hasan M M and Jiang K. 2021b. Wafer map defect pattern recognition based on self-organizing incremental neural network//*Proceedings of 2021 China Automation Congress (CAC)*. Beijing, China: IEEE: 5617-5622 [DOI: 10.1109/cac53003.2021.9728184]
- Yu N G, Xu Q and Wang H L. 2019b. Wafer defect pattern recognition and analysis based on convolutional neural network. *IEEE Transactions on Semiconductor Manufacturing*, 32(4): 566-573 [DOI: 10.1109/tsm.2019.2937793]
- Yu N G, Xu Q, Wang H L and Lin J. 2021c. Wafer bin map inspection based on DenseNet. *Journal of Central South University*, 28(8): 2436-2450 [DOI: 10.1007/s11771-021-4778-7]
- Zhang Q, Zhang Y H, Li J Z and Li Y F. 2022. WDP-BNN: efficient wafer defect pattern classification via binarized neural network. *Integration*, 85: 76-86 [DOI: 10.1016/j.vlsi.2022.04.003]
- Zhuang J J, Mao G Y, Wang Y, Chen X and Wei Z Y. 2020. A neural-network approach to better diagnosis of defect pattern in wafer bin map//*Proceedings of 2020 China Semiconductor Technology International Conference (CSTIC)*. Shanghai, China: IEEE: 1-3 [DOI: 10.1109/estic49141.2020.9282438]

作者简介

胡志强,男,博士研究生,主要研究方向为图像处理与机器视觉。E-mail: 157751723@qq.com

吴一全,通信作者,男,教授,博士生导师,主要研究方向为遥感图像处理与理解、红外目标检测与识别、视觉检测与图像测量、视频处理与智能分析。E-mail: nuaaimage@163.com