

融合离散小波变换的YOLOv11晶圆缺陷检测算法

刘廷强¹, 陈明淑¹, 费赞强², 闫思安³

¹西京学院电子信息学院, 陕西 西安

²双林镇综合服务中心, 浙江 湖州

³智立方自动化设备股份有限公司, 深圳

收稿日期: 2026年4月15日; 录用日期: 2026年5月9日; 发布日期: 2026年5月18日

摘要

针对半导体制造过程中晶圆表面微小缺陷尺度小、特征弱且易受复杂背景干扰, 导致传统目标检测方法存在漏检率高、定位精度不足等问题, 本文提出了一种融合离散小波变换(Discrete Wavelet Transform, DWT)特征保留机制的YOLOv11晶圆缺陷检测算法。该方法在主干网络下采样阶段引入WaveletConv模块, 以二维离散小波变换替代传统步长卷积, 实现特征图的多频分解与结构化表征。具体而言, 输入特征被分解为低频全局分量以及水平、垂直和对角方向的高频细节分量, 在完成空间降采样的同时, 有效保留了与微小缺陷识别密切相关的边缘、纹理和局部突变信息。进一步地, 通过通道维融合策略将多频信息联合传递至深层网络, 增强了模型对微弱缺陷特征的表征能力与判别能力, 从而缓解了传统卷积下采样过程中高频信息衰减带来的性能损失。实验结果表明, 所提出的方法在晶圆缺陷数据集上取得了优于基线模型的检测性能, 在微小缺陷检出率和定位精度方面表现出明显优势。研究表明, 将小波域多尺度、多频特征建模引入目标检测框架, 能够有效提升模型对复杂工业场景中微观缺陷的感知能力, 为半导体制造过程中的高精度视觉检测提供了新的解决思路。

关键词

离散小波变换, YOLO, 晶圆缺陷检测, 半导体制造

Wafer Defect Detection Algorithm Based on YOLOv11 with Wavelet Transform Integration

Tingqiang Liu¹, Mingshu Chen¹, Zanqiang Fei², Si'an Yan³

¹School of Electronic Information, Xijing University, Xi'an Shaanxi

²Shuanglin Town Comprehensive Service Center, Huzhou Zhejiang

文章引用: 刘廷强, 陈明淑, 费赞强, 闫思安. 融合离散小波变换的YOLOv11晶圆缺陷检测算法[J]. 传感器技术与应用, 2026, 14(3): 439-449. DOI: 10.12677/jsta.2026.143045

Abstract

In semiconductor manufacturing, wafer surface defects are characterized by small scales, weak features, and susceptibility to complex background interference, which leads to high miss-detection rates and insufficient localization accuracy in traditional object detection methods. To address these challenges, this paper proposes a YOLOv11-based wafer defect detection algorithm integrated with a Discrete Wavelet Transform (DWT) feature preservation mechanism. This method introduces a WaveletConv module at the downsampling stages of the backbone network, replacing conventional strided convolutions with two-dimensional discrete wavelet transform to achieve multi-frequency decomposition and structured representation of feature maps. Specifically, the input features are decomposed into low-frequency global components and high-frequency detail components in horizontal, vertical, and diagonal orientations. This approach accomplishes spatial downsampling while effectively preserving edge, texture, and local mutation information crucial for tiny defect recognition. Furthermore, through a channel-wise fusion strategy, the multi-frequency information is jointly propagated to deeper network layers, enhancing the model's representation and discriminative capabilities for subtle defect features, thereby mitigating performance degradation caused by high-frequency information attenuation during traditional convolutional downsampling. Experimental results demonstrate that the proposed method achieves superior detection performance compared to baseline models on wafer defect datasets, exhibiting significant advantages in tiny defect detection rates and localization precision. The study reveals that introducing wavelet-domain multi-scale, multi-frequency feature modeling into object detection frameworks can effectively improve the model's perception capability for microscopic defects in complex industrial scenarios, providing novel insights for high-precision visual inspection in semiconductor manufacturing processes.

Keywords

Discrete Wavelet Transform, YOLO, Wafer Defect Detection, Semiconductor Manufacturing

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着集成电路制造工艺持续向高集成度、高精度和微纳尺度演进，半导体制造过程对检测精度、响应速度和质量控制能力提出了更高要求。晶圆作为芯片制造的核心载体，其表面质量直接影响器件性能、成品良率以及后续封装测试成本[1]。在晶圆加工、传输与检测过程中，受机械应力、化学残留、环境微尘及工艺波动等因素影响，晶圆表面容易产生划痕、颗粒污染、边缘破损及局部图案异常等缺陷。这类缺陷不仅可能导致局部功能失效，还可能在后续工艺中被进一步放大，造成批量性质量风险[2]。因此，实现晶圆表面缺陷的快速、准确检测，是半导体生产检测环节中的关键技术问题之一。

近年来，基于深度学习的目标检测方法在工业缺陷检测领域取得了较好效果，其中 YOLO 系列算法因检测速度快、部署效率高而得到广泛关注[3]。然而，晶圆表面缺陷检测仍面临若干突出困难：其一，

缺陷目标通常尺寸较小、形态不规则,且边缘与纹理特征较弱,容易在网络下采样过程中发生信息丢失;其二,晶圆缺陷类型复杂,局部细节差异显著,常规卷积下采样更倾向于保留低频语义信息,而对高频边缘和纹理响应不足;其三,在复杂背景和多尺度缺陷并存条件下,检测网络需要同时兼顾全局定位能力与局部细节辨识能力,这对特征提取与信息传递机制提出了更高要求[4]。由此可见,如何在保证检测效率的同时尽可能保留浅层细节信息,是提升晶圆缺陷检测性能需要重点解决的问题。

针对上述问题,本文提出一种基于离散小波变换(DWT)的 WaveletConv 下采样模块,并在此基础上构建融合该模块的改进 YOLOv11 网络,在保持与原有 neck/head 结构无缝衔接的同时增强对微小缺陷高频细节信息的保留能力,最终通过在 WM-811K 数据集上开展对比实验与消融实验验证了所提方法的有效性。

2. 算法介绍

2.1. YOLOv11 基础网络结构及其在晶圆缺陷检测中的局限性

YOLOv11 作为单阶段目标检测网络,整体上仍由主干网络(Backbone)、颈部网络(Neck)和检测头(Head)三部分构成。其中,主干网络负责对输入图像进行分层特征提取,通过多级卷积与下采样操作逐步扩大感受野并形成不同尺度的特征表示;颈部网络结合特征金字塔与路径聚合机制,对浅层细节特征与深层语义特征进行跨尺度融合[4];检测头则基于融合后的多尺度特征图完成目标类别判别与边界框回归。该结构能够实现端到端的目标检测,为后续网络改进提供了清晰的模块化基础。

然而,在晶圆缺陷检测场景中,缺陷目标通常具有尺度小、形态不规则、边界模糊及局部灰度变化微弱等特点,其判别依据更多依赖于浅层特征中的边缘、纹理及局部突变信息[5]。YOLOv11 主干网络中的传统步长卷积虽然能够有效完成空间下采样并增强高层语义表征,但在特征压缩过程中往往会削弱高频细节信息的保留能力,使微小缺陷相关的局部结构响应随网络加深而逐渐衰减。对于晶圆图像这类细粒度异常目标而言,若前期下采样阶段对高频信息保留不足,容易导致小尺度缺陷漏检、边界定位偏移以及相邻异常区域区分不充分等问题[6]。因此,如何在主干网络下采样过程中兼顾空间尺度压缩与高频细节保留,成为提升晶圆微小缺陷检测性能的关键。

2.2. 离散小波变换的特征分解机制及其适用性分析

离散小波变换(Discrete Wavelet Transform, DWT)是一种典型的多分辨率分析方法,能够同时表征信号的空间局部信息与频率分布特征。对于二维图像而言,一次二维 DWT 可将输入特征分解为 1 个低频子带和 3 个高频子带,其中低频子带主要保留图像的整体轮廓与主体结构信息,高频子带则分别刻画水平方向、垂直方向和对角方向上的边缘细节与局部变化特征。与仅能描述全局频域分布的传统傅里叶分析不同,DWT 具有更强的空间-频率局部化能力,因此更适合用于表征具有局部突变特征的图像结构[7]。

对于晶圆缺陷检测任务而言,微小缺陷通常具有尺度小、边界弱、纹理不明显以及局部灰度变化细微等特点,其有效判别信息更多体现在边缘扰动、局部异常响应和弱纹理变化之中。相比之下,传统卷积下采样在压缩空间分辨率的过程中更倾向于保留低频语义信息,而与微小缺陷识别密切相关的高频细节容易被削弱甚至丢失。离散小波变换能够在实现空间降采样的同时,将低频结构信息与高频细节信息显式分离,从而为后续网络提供更具针对性的特征表达。因此,将 DWT 引入晶圆缺陷检测网络的下采样过程,具有明确的理论合理性与任务适配性。

在具体实现上,本文选用二维 Haar 小波作为特征分解工具。其原因在于,Haar 小波结构简单、计算过程清晰,既便于嵌入卷积神经网络,又能够在下采样过程中同步完成频带分解,额外计算代价较低。对于本文所关注的晶圆微小缺陷检测任务,采用 Haar 小波不仅能够满足网络轻量化设计需求,还能够为

后续高频细节保留与多频信息融合提供有效支撑[8]。

设输入特征图为 $X \in \mathbb{R}^{H \times W}$ ，对于其中以 $(2i, 2j)$ 为左上角的 2×2 局部区域，其二维 Haar 小波分解结果可表示为。

$$LL_{i,j} = \frac{1}{2}(x_{2i,2j} + x_{2i,2j+1} + x_{2i+1,2j} + x_{2i+1,2j+1}) \quad (1)$$

$$LH_{i,j} = \frac{1}{2}(x_{2i,2j} + x_{2i+1,2j} - x_{2i,2j+1} - x_{2i+1,2j+1}) \quad (2)$$

$$HL_{i,j} = \frac{1}{2}(x_{2i,2j} - x_{2i+1,2j} + x_{2i,2j+1} - x_{2i+1,2j+1}) \quad (3)$$

$$HH_{i,j} = \frac{1}{2}(x_{2i,2j} - x_{2i+1,2j} - x_{2i,2j+1} + x_{2i+1,2j+1}) \quad (4)$$

其中， (i, j) 表示降采样后特征图的像素坐标。

其中， (i, j) 表示降采样后子带特征图中的空间坐标， $x_{m,n}$ 表示输入特征图在对应位置的特征响应。经过一次二维小波分解后，输入特征图被映射为四个子带： LL 子带主要保留低频轮廓与整体结构信息， LH 、 HL 和 HH 子带则分别表征不同方向上的高频边缘与局部细节信息[9]。

对于晶圆缺陷检测任务，缺陷区域通常具有尺度小、边界弱、局部突变明显等特点，相关判别线索往往集中在高频成分中。相较于传统步长卷积仅通过局部聚合完成降采样，离散小波变换能够在压缩空间分辨率的同时显式分离低频结构信息与高频细节信息，从而为后续特征提取提供更充分的边缘与纹理表达[10]。因此，DWT 为本文后续构建基于 WaveletConv 的下采样算子提供了理论基础。

2.3. 基于 DWT 的 WaveletConv 下采样算子设计

针对传统步长卷积在空间降采样过程中易造成高频细节信息衰减的问题，本文基于离散小波变换设计了一种可嵌入卷积神经网络的下采样算子，记为 WaveletConv。该模块用于替代 YOLOv11 主干网络中的传统 $\text{stride} = 2$ 卷积单元，其核心思想是在实现特征图空间分辨率降低的同时，将输入特征中的低频结构信息与高频细节信息显式分离，并通过后续通道融合生成新的尺度特征表示。整体上，WaveletConv 由频域分解、通道重组和特征聚合三个步骤构成。

设输入特征张量为

$$X \in \mathbb{R}^{C \times H \times W} \quad (12)$$

其中， C 、 H 和 W 分别表示输入特征图的通道数、高度和宽度。首先，对输入特征图的每个通道独立施加二维离散小波变换，将原始特征分解为四个频带子特征，分别记为低频子带 LL 以及高频子带 LH 、 HL 和 HH 。经过一次二维小波分解后，各子带特征张量的空间尺寸均变为原来的二分之一，即

$$LL, LH, HL, HH \in \mathbb{R}^{C \times \frac{H}{2} \times \frac{W}{2}} \quad (13)$$

其中， LL 主要保留输入特征中的整体轮廓和主体结构信息，而 LH 、 HL 与 HH 分别表征不同方向上的边缘响应、局部纹理以及突变细节信息。该过程本质上在完成空间下采样的同时，将传统卷积隐式压缩的信息显式分解到不同频带中，从而增强了特征表达的结构可分性。

为了使后续网络能够同时利用低频语义信息与高频细节信息，本文将四个子带沿通道维进行拼接，得到重组后的多频特征表示：

$$X_{cat} = \text{Concat}(LL, LH, HL, HH) \quad (14)$$

拼接后特征张量的尺寸可表示为

$$X_{cat} \in \mathbb{R}^{4C \times \frac{H}{2} \times \frac{W}{2}} \quad (15)$$

相较于传统 $\text{stride} = 2$ 卷积直接在空间域完成采样与聚合, 通道拼接操作能够在保持空间分辨率降采样结果不变的前提下, 将不同频带中的响应统一映射至扩展后的通道维度, 为后续跨频信息融合提供更充分的表达基础。

然而, 直接将四个子带拼接后的特征送入后续网络会导致通道数显著增加, 从而带来不必要的参数冗余与计算负担。为此, 本文在通道拼接后引入 1×1 卷积, 并结合批归一化(Batch Normalization, BN)与 SiLU 激活函数, 对多频特征进行跨通道融合与维度压缩, 生成最终输出特征:

$$Y = \text{SiLU}\left(\text{BN}\left(\text{Conv}_{1 \times 1}\left(X_{cat}\right)\right)\right) \quad (16)$$

其中, $\text{Conv}_{1 \times 1}$ 表示卷积核大小为 1×1 、步长为 1 的卷积操作, 输出通道数设定为与对应下采样阶段原始卷积层一致的目标维度 C_{out} 。因此, WaveletConv 的最终输出尺寸为

$$Y \in \mathbb{R}^{C_{out} \times \frac{H}{2} \times \frac{W}{2}} \quad (17)$$

可见, 该模块在输出空间尺度上与传统 $\text{stride} = 2$ 卷积保持一致, 因而能够在不改变后续网络尺度对应关系的前提下, 直接替换原有下采样层。

从信息流角度看, 传统步长卷积主要依赖局部感受野内的采样与加权聚合完成下采样, 容易在特征压缩过程中丢失与微小缺陷识别密切相关的高频边缘和弱纹理信息。相比之下, WaveletConv 先通过 DWT 将特征映射到低频与高频子带, 再利用通道拼接和 1×1 卷积实现多频信息重组, 使网络在完成尺度变换的同时仍能够保留较丰富的局部异常响应。对于晶圆缺陷检测任务而言, 该设计尤其有利于增强模型对小尺度缺陷、边界模糊缺陷以及弱对比度异常区域的表征能力。

为保证模块设计具备明确的实现路径, 本文对 WaveletConv 的具体实现方式作如下说明: 首先, 采用二维 Haar 小波对输入特征图进行逐通道分解, 即每个输入通道独立生成 LL 、 LH 、 HL 和 HH 四个子带, 而非采用分组变换或跨通道联合分解方式; 其次, 四个子带信息全部保留, 并沿通道维拼接形成多频特征张量, 以避免在下采样初期提前丢弃高频细节响应; 再次, 拼接后的特征通过 $1 \times 1 \text{ Conv} + \text{BN} + \text{SiLU}$ 进行通道压缩与非线性融合, 其中输出通道数设置为与原 $\text{stride} = 2$ 卷积层的输出维度一致; 最后, 模块输出特征的空间尺寸与原始下采样层保持相同, 均为输入特征图的一半, 因此可直接嵌入 YOLOv11 主干网络的下采样位置, 而无需对后续 neck 和 head 的结构进行额外调整。

2.4. 融合 WaveletConv 的改进 YOLOv11 网络架构

在原始 YOLOv11 中, 主干网络通常通过多级 $\text{stride} = 2$ 卷积逐步降低特征图分辨率, 并同步提升通道维度, 以实现感受野扩展和深层语义信息提取。然而, 对于晶圆微小缺陷检测任务而言, 浅层特征中包含大量与缺陷识别密切相关的边缘、纹理及局部突变信息, 如图 1 所示基线网络结构前期阶段采用传统步长卷积进行连续下采样, 易导致高频细节响应快速衰减, 从而影响后续网络对微小异常区域的表征能力。为此, 本文将主干网络前两级 $\text{stride} = 2$ 下采样卷积替换为 WaveletConv, 以增强浅层高频细节保留能力, 同时控制额外计算开销。

设第 l 个下采样阶段的输入特征为

$$X_l \in \mathbb{R}^{C_l \times H_l \times W_l} \quad (18)$$

在原始 YOLOv11 中, 该阶段通常通过 $\text{stride} = 2$ 卷积得到下一尺度特征; 而在本文的改进网络中,

该过程被表示为

$$X_{I+1} = \text{WaveletConv}(X_I) \tag{19}$$

其中，WaveletConv(·)表示由二维离散小波分解、通道拼接和1×1卷积融合组成的下采样映射。经过该模块后，输出特征尺寸变为

$$X_{I+1} \in \mathbb{R}^{C_{I+1} \times \frac{H_I \times W_I}{2}} \tag{20}$$

可见，WaveletConv在空间降采样倍率上与原 stride = 2 卷积保持一致，因此能够在不改变特征层级尺度关系的前提下，完成对原有下采样单元的直接替换。

如图2所示在完成主干网络特后，不同尺度的特征图仍按照原YOLOv11的连接方式送入颈部网络进行多尺度融合。本文保留原有neck中的特征金字塔与路径聚合结构，使浅层定位信息与深层语义信息继续通过自顶向下和自底向上的路径进行交互。由于浅层下采样阶段引入了WaveletConv，输入到后续特征融合模块中的特征图能够保留更充分的高频边缘与局部纹理信息，从而提升多尺度特征表达的完整性，增强模型对小尺度、弱对比度及边界模糊缺陷的检测能力。

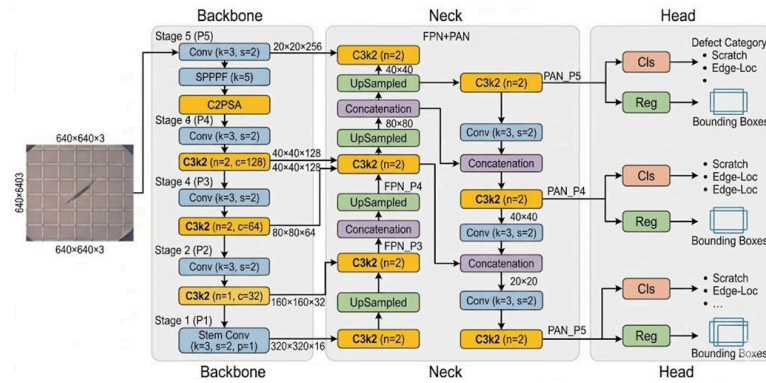


Figure 1. Baseline architecture of YOLOv11
图 1. YOLOv11 基线网络结构

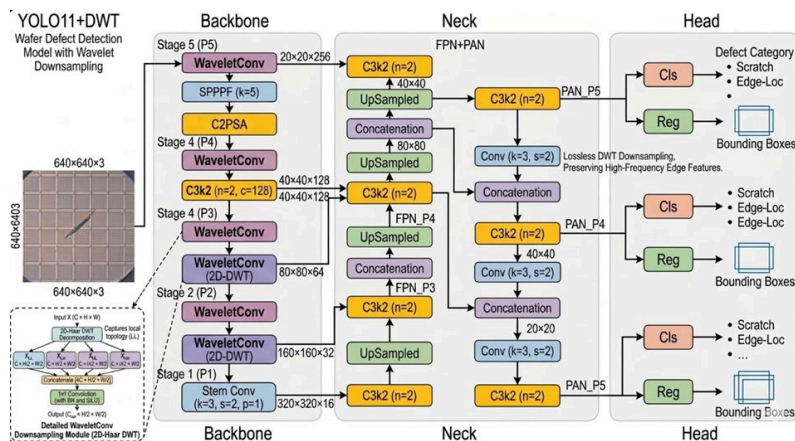


Figure 2. Architecture of the proposed YOLOv11 with WaveletConv
图 2. 融合 WaveletConv 的改进 YOLOv11 网络结构

在预测阶段，改进网络仍采用原YOLOv11检测头对融合后的多尺度特征进行类别预测与边界框回归。由于本文的改进主要集中在主干网络前端的下采样机制，因此检测头的输出形式、损失函数构成以

及训练优化方式均与基线模型保持一致。这样的设计有助于将性能提升更直接地归因于 WaveletConv 对高频细节保留能力的增强，而非其他结构性因素的干扰。

需要指出的是，WaveletConv 与 $\text{stride} = 2$ 卷积具有相同的空间降采样倍率，因此改进后的 backbone 不改变 neck 与 detection head 的输入尺度对应关系，能够与原 YOLOv11 网络结构实现无缝衔接。这一设计保证了方法改进具有良好的结构兼容性和实现可行性，也为后续实验中公平比较基线模型与改进模型提供了基础。


3. 实验验证与分析

3.1. 数据集构建

本文实验采用公开的 WM-811K 晶圆缺陷数据集开展验证。该数据集包含多种典型晶圆缺陷模式，能够较好反映晶圆制造过程中常见异常分布形态。考虑到无缺陷样本对目标检测任务的贡献有限，本文剔除类别标签为 None 的样本，仅保留中心缺陷(Center)、圆环缺陷(Donut)、局部边缘缺陷(Edge-Loc)、边缘环形缺陷(Edge-Ring)、局部缺陷(Loc)、随机缺陷(Random)、划痕缺陷(Scratch)以及近全屏缺陷(Near-full) 8 类典型异常样本，用于构建晶圆缺陷检测实验数据集。各类缺陷的形态特征及样本统计信息见表 1。

Table 1. Common defect types on wafer surfaces

表 1. 晶圆表面常见缺陷类型

图像	缺陷	过程	成因
	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、清洁不到位
	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	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
	随机	研磨、抛光	晶圆表面不均匀、清洁不到位

为保证实验评估的规范性, 本文按照分层抽样原则将数据集划分为训练集、验证集和测试集, 比例设置为 7:2:1。在划分过程中, 尽可能保持各类别样本在不同子集中的分布一致, 以减小类别分布偏移对实验结果的影响。鉴于 WM-811K 数据集中不同缺陷类别样本数量存在一定不均衡性, 本文在训练阶段采用数据增强与少数类样本重采样相结合的策略缓解类别不平衡问题, 从而提升模型对少样本缺陷类型的学习能力。为适配 YOLOv11 网络的输入要求并兼顾检测精度与计算开销, 所有样本在输入网络前统一缩放至 640×640 。

3.2. 实验环境与超参数设置

本文实验基于深度学习框架 PyTorch 2.0 搭建, 编程语言环境为 Python 3.9, CUDA 版本为 12.0。硬件平台采用 Intel Core i9-13900K 处理器和单张 NVIDIA RTX 4090 GPU。为与轻量化检测任务相匹配, 本文选取 YOLOv11 作为基线模型, 并在此基础上构建融合 WaveletConv 的改进网络。为保证实验结果具有可比性, 基线模型与改进模型均在相同软硬件环境下完成训练与测试。

在训练策略方面, 输入图像尺寸统一设为 640×640 , 训练集、验证集和测试集按照 7:2:1 的比例进行划分。训练阶段采用 SGD 优化器进行参数更新, 初始学习率设为 0.01, 动量系数设为 0.937, 权重衰减系数设为 0.0005, 并采用余弦退火策略对学习率进行动态衰减。模型总训练轮数设为 300 Epoch, 批大小(Batch Size)设为 32。为增强模型对缺陷尺度变化和局部形态扰动的适应能力, 训练过程中采用 Mosaic 数据增强、随机翻转以及随机缩放等策略; 同时, 针对 WM-811K 数据集中部分类别样本分布不均衡的问题, 结合少数类样本重采样方式缓解类别不平衡对模型训练的影响。测试阶段置信度阈值设为 0.25, 非极大值抑制(NMS)的 IoU 阈值设为 0.45。此外, 为降低随机初始化对实验结果的干扰, 本文将随机种子固定为 42, 并在相同设置下独立重复实验 3 次, 最终结果取平均值。模型性能评价指标采用精确率(Precision, P)、召回率(Recall, R)和 $mAP@0.5$, 以综合衡量模型的检测准确性与定位能力。

3.3. 实验与结果分析

晶圆缺陷检测结果

为验证所提方法在晶圆缺陷检测任务中的有效性, 本文在 WM-811K 数据集上对基线模型 YOLOv11 与融合 WaveletConv 的改进模型进行了对比实验, 整体检测结果如表 2~4 及图 3 所示。由总体评价指标可以看出, 引入 WaveletConv 后, 模型的 $mAP@0.5$ 由 0.897 提升至 0.935, 提高了 3.8 个百分点, 表明所提方法能够在统一实验设置下有效提升晶圆缺陷检测性能。与此同时, 改进模型的 Precision、Recall 及参数量指标见表 2。由于本文方法仅对主干网络下采样模块进行改进, 因此其性能增益能够较直接地归因于高频细节保留能力的增强。

从不同缺陷类别的检测结果来看, 本文方法对细粒度异常和边界复杂缺陷的提升更为明显。具体而言, Edge-Loc 类别的 AP 由 0.739 提升至 0.850, 增幅为 0.111; Scratch 类别由 0.759 提升至 0.860, 增幅为 0.101; 此外, Random、Loc 和 Edge-Ring 三类缺陷的 AP 也分别提升了 0.047、0.043 和 0.044。这说明, WaveletConv 模块能够在主干网络下采样过程中更有效地保留与缺陷识别密切相关的高频边缘、局部纹理及弱异常响应信息, 从而增强模型对微小缺陷和复杂边界缺陷的表征能力。

对于 Center、Donut 和 Near-full 三类形态较为显著、区域分布相对稳定的缺陷, 改进前后 AP 基本保持一致, 说明本文方法在提升困难类别检测性能的同时, 并未削弱模型对易识别类别的检测稳定性。综合来看, 融合 WaveletConv 的改进 YOLOv11 网络在整体检测性能和困难缺陷识别能力方面均优于基线模型, 验证了在下采样阶段引入离散小波分解机制的合理性与有效性。

Table 2. Ablation study results of the wavelet transform module
表 2. 小波变换模块消融实验结果

模型	P (%)	R (%)	mAP0.5 (%)	参数量(M)
Baseline (YOLOv11)	89.4	86.2	89.7	3.2
Ours (YOLOv11 + DWT)	92.1	90.5	93.5	3.85

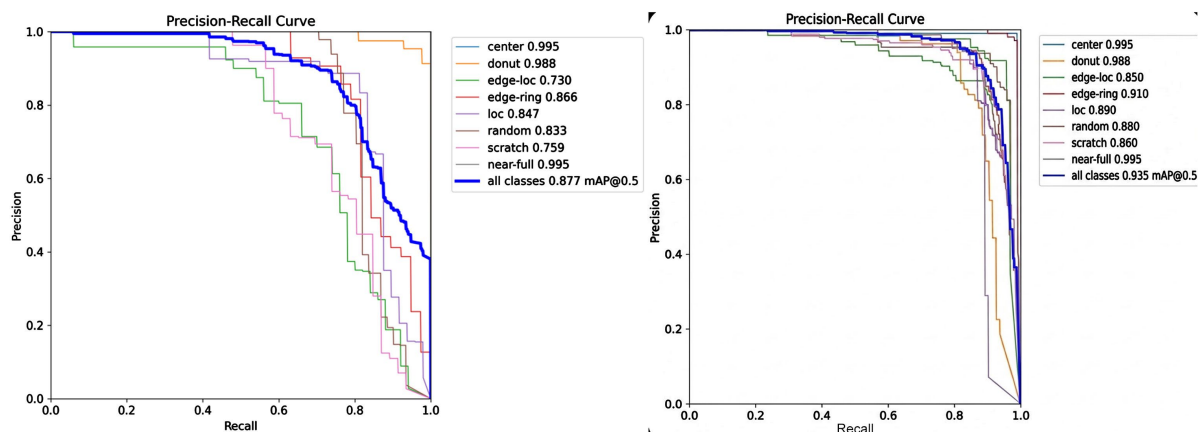


Figure 3. Precision-Recall curve comparison of YOLOv11 and the proposed method on the WM-811K dataset
图 3. YOLOv11 与本文方法在 WM-811K 数据集上的 PR 曲线对比

Table 3. WM811K experimental verification
表 3. WM811K 实验验证

缺陷类型	未定位原图	YOLO11 定位图	DWT-YOLO11 定位图
中心			
圆环			
边缘			
边环			

续表

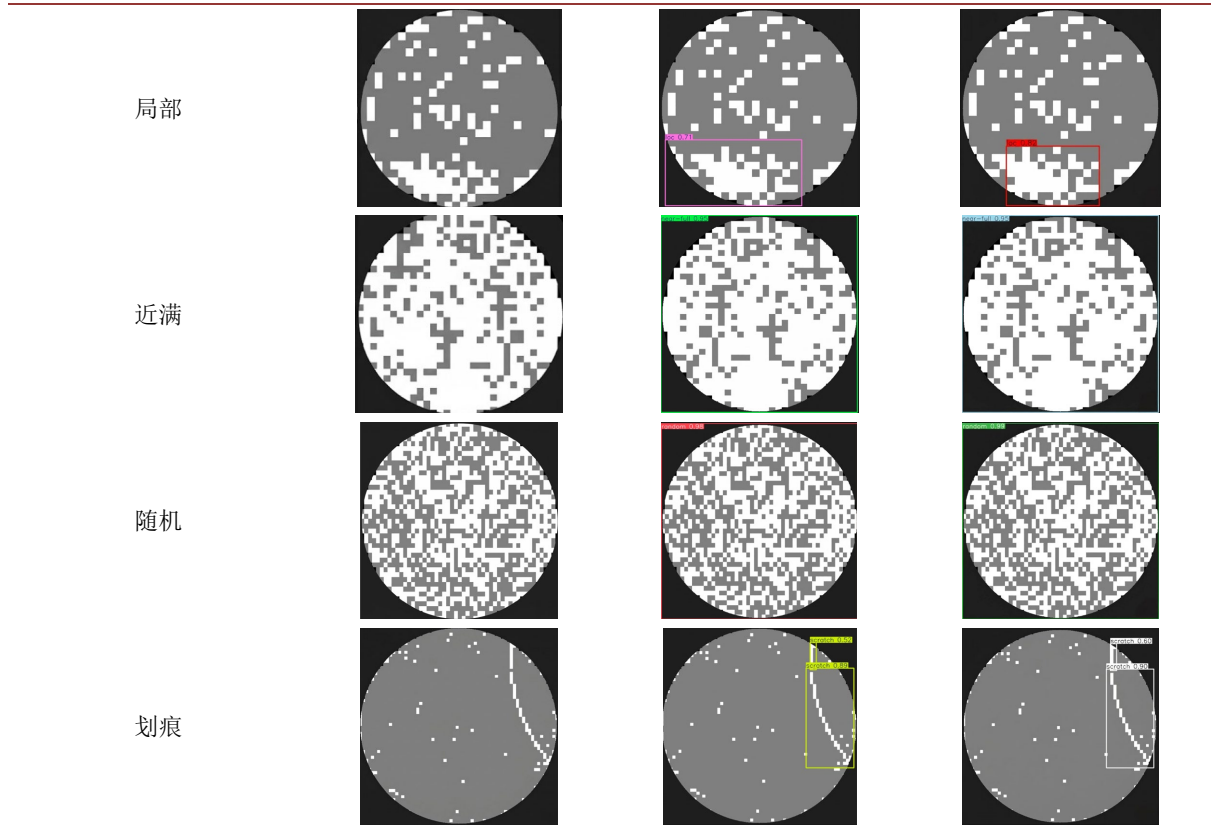


Table 4. Ablation study
表 4. 消融实验

缺陷类型	YOLOv11	YOLOv11 + WaveletConv	提升幅度
Center	0.995	0.995	0.000
Donut	0.988	0.988	0.000
Edge-Loc	0.739	0.850	+0.111
Edge-Ring	0.866	0.910	+0.044
Loc	0.847	0.890	+0.043
Random	0.833	0.880	+0.047
Scratch	0.759	0.860	+0.101
Near-full	0.995	0.995	0.000
mAP@0.5	0.897	0.935	+0.038

4. 总结

本文针对晶圆微小缺陷在下采样过程中易发生高频细节信息损失的问题，提出了一种融合离散小波变换的改进 YOLOv11 检测方法。通过在主干网络中引入 WaveletConv 模块，实现了低频结构信息与高频细节信息的协同建模，从而增强了模型对边缘模糊、纹理微弱及局部异常缺陷的表征能力。基于 WM-811K 数据集的实验结果表明，所提方法在整体检测性能及困难类别识别能力方面均优于基线模型，验证

了将离散小波分解机制引入下采样阶段的有效性。未来将进一步结合真实工业场景数据，围绕模型轻量化、复杂噪声鲁棒性及工程部署适应性开展深入研究。

参考文献

- [1] Ma, J., Zhang, T., Yang, C., Cao, Y., *et al.* (2023) Review of Wafer Surface Defect Detection Methods. *Electronics*, **12**, Article 1787. <https://doi.org/10.3390/electronics12081787>
- [2] Zou, J., Wang, H. and Zhong, J. (2025) Wafer Defect Image Generation Method Based on Improved Styleganv3 Network. *Micromachines*, **16**, Article 844. <https://doi.org/10.3390/mi16080844>
- [3] Han, C., Chun, H., Lee, J., *et al.* (2024) Hybrid Semiconductor Wafer Inspection Framework Minimizing Loads for Feature Engineering. *Journal of Manufacturing Science and Engineering*, **146**, Article 070906. <https://doi.org/10.1115/1.4065276>
- [4] Singh, J.P., Ghosh, D., Kumar, A., *et al.* (2026) Hybrid DenseNet Architectures and Keras Tuner-Based Optimization for Rice Leaf Disease Detection. *IEEE Access*, **14**, 26845-26868. <https://doi.org/10.1109/ACCESS.2026.3664467>
- [5] Yu, J. and Lu, X. (2016) Wafer Map Defect Detection and Recognition Using Joint Local and Nonlocal Linear Discriminant Analysis. *IEEE Transactions on Semiconductor Manufacturing*, **29**, 33-43. <https://doi.org/10.1109/TSM.2015.2497264>
- [6] Liu, S., Qi, L., Qin, H., *et al.* (2018) Path Aggregation Network for Instance Segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 8759-8768. <https://doi.org/10.1109/CVPR.2018.00913>
- [7] 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. <https://doi.org/10.1109/ACCESS.2022.3166512>
- [8] Swaroop, P. and Sharma N. (2016) An Overview of Various Template Matching Methodologies in Image Processing. *International Journal of Computer Applications*, **153**, 8-14. <https://doi.org/10.5120/ijca2016912165>
- [9] 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**, 3215-3247. <https://doi.org/10.1007/s10845-022-01994-1>
- [10] Wu, M.J., Jang, J.S.R. and Chen, J.L. (2014) Wafer Map Failure Pattern Recognition and Similarity Ranking for Large-Scale Data Sets. *IEEE Transactions on Semiconductor Manufacturing*, **28**, 1-12. <https://doi.org/10.1109/TSM.2014.2364237>