

DOI: 10.13382/j.jemi.B2205935

# 基于 YOLOv5-EA-FPNs 的芯片缺陷检测方法研究 \*

张恒 程成 袁彪 赵洪坪 吕雪杭 芹

(重庆邮电大学计算机科学与技术学院 重庆 400065)

**摘要:**针对芯片缺陷检测中,缺陷尺寸跨度大、特征相似、小目标难识别、漏检等问题,本文提出基于 YOLOv5 改进的缺陷检测方法。针对小目标缺陷检测中出现的漏检、误检等问题,提出新增小目标特征检测器(small target feature detector, S-Detector),提升模型对小目标缺陷的学习能力;针对缺陷尺寸跨度大、特征相似等问题,提出具有高效聚焦学习能力的特征金字塔结构 (efficient attention feature pyramid networks, EA-FPNs),提升模型对不同尺寸缺陷的检测能力;针对预测阶段冗余框较多导致时间开销大的问题,提出基于面积的边界框融合算法(bounding box fusion algorithm, BFA),减少冗余框。实验结果表明,本文方法相较于改进前,检测精确度提升 1.2%,小目标缺陷精确度提升 1.6%;采用 BFA 消除冗余框的同时,平均检测时长为 26.8 μs/张,较使用 BFA 前减少了 5.2 μs。本文所提方法具有良好性能,能够提升检测效率。

**关键词:**芯片缺陷检测;深度学习;特征金字塔;多尺度融合;小目标检测;YOLOv5

中图分类号: TP391; TN98 文献标识码: A 国家标准学科分类代码: 520.2060; 51040

## Research on chip defect detection method based on YOLOv5-EA-FPNs

Zhang Heng Cheng Cheng Yuan Biao Zhao Hongping Lyu Xue Hang Qin

(School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

**Abstract:** To address the problems of large defect size span, similar characteristics, difficulty in recognition of small targets, and missed objects in chip defect detection, an improved method based on YOLOv5 is proposed. To solve missed and false detection of small targets, we presented a new small target feature detector (S-Detector) to improve the learning capability of the model. For the large defect size span and similar characteristics, efficient attention feature pyramid networks (EA-FPNs) with highly active focus learning ability are proposed to improve the ability to detect different sizes of defects. The bounding box fusion algorithm (BFA) is developed to reduce the redundant boxes and time overhead in prediction. The experimental results show that the detection accuracy of this method is enhanced by 1.2% and the accuracy of minor target defects is improved by 1.6%; while using BFA to eliminate the redundant boxes, the detection time of a single image is 26.8 μs, which is decreased by 5.2 μs before BFA. The proposed method has good performance and efficiency in chip defect detection.

**Keywords:** chip defect detection; deep learning; FPNs; multi-scale fusion; small object detection; YOLOv5

## 0 引言

半导体芯片在制造流程各个环节都有可能出现表面缺陷,有效的缺陷检测方法是提高芯片制造良率的关键。传统的芯片缺陷检测方法大多是人工或机器视觉检测

法。人工检测法效率、精度相对较低,易受到主观因素影响,检测效率无法满足生产需求;机器视觉检测法基于光学原理,利用自动光学检测设备 (automatic optic inspection, AOI) 完成检测,该方法相较于传统人工检测更加灵活、快速、效率更高。然而,传统机器视觉检测法基于手工特征,需要从业人员具有丰富的专业知识,且该

方法容易受到光线等物理因素的影响,导致检测出现误判、漏检等问题。

随着第一个深度卷积神经网络(convolutional neural networks, CNN)模型的提出,基于深度学习的缺陷检测技术逐渐在工业检测上得到应用,如Park等<sup>[1]</sup>利用CNN实现表面缺陷的端到端检测,避免人工设计特征,降低了人为因素对缺陷特征的影响。目前基于CNN的缺陷检测算法分为两类:一阶段算法,模型训练速度快,实时性强,但精准度欠佳,其代表算法有YOLO(you only look once, YOLO)系列<sup>[2-4]</sup>、SSD<sup>[5]</sup>(single shot detector, SSD)、RetinaNet<sup>[6]</sup>等;二阶段算法,在检测精准度上要强于一阶段算法,但模型实时性较弱,代表算法有区域卷积神经网络<sup>[7]</sup>(region-based convolutional neural network, R-CNN)、Fast R-CNN<sup>[8]</sup>、Faster R-CNN<sup>[9]</sup>、Mask R-CNN<sup>[10]</sup>等。芯片缺陷特征相似、尺寸跨度大、小目标难检测而导致的误检、漏检等一直是困扰技术人员的难题,前人对上述问题进行了研究,如Li等<sup>[11]</sup>提出一种基于卷积神经网络的微小缺陷检测方法,采用从粗到细两阶段的CNN框架去提升模型对微小缺陷的识别能力,但忽略了尺寸跨度大、特征相似等问题;Cao等<sup>[12]</sup>提出的注意力引导结构,能够提升大目标特征图的学习能力,从而降低尺寸跨度大带来的问题,由于该方法只针对大目标,对小目标的识别起到了负面影响。研究人员只专注于解决针对芯片缺陷检测中的单一问题,如小目标缺陷漏检或尺寸跨度大、特征相似而造成的漏检、误检,并未给出较为全面的解决方案,本文针对芯片表面缺陷检测过程中存在的缺陷尺寸差距大、特征相似、小目标检测效果差等问题进行综合考虑,兼顾检测实时性,提出基于YOLOv5的改进方法。本文的主要贡献如下:

1) 针对芯片缺陷尺寸变化大、相似特征等问题,本文提出了EA-FPNs结构,该结构能够聚合多尺度上下文信息,聚焦缺陷特征,以缓解不同尺度缺陷的误检、漏检等问题。进一步的,针对小目标难检测的情况,本文通过新增一层S-Detector,加强EA-FPNs对小目标的识别能力。

2) 针对推理结果中存在较多冗余框导致时间开销大的问题,本文提出了BFA,该算法基于面积权重占比消除冗余框,获得更加准确的检测结果。

## 1 相关工作

随着计算成像、计算机和大数据人工智能等相关领域的快速发展,一些基于统计、滤波器等机器视觉方法被应用到芯片缺陷检测。曹新容等<sup>[13]</sup>利用匹配滤波器克服了背景干扰,消除了噪声影响,使检测图像达到了灰度平衡;朱刘蛊等<sup>[14]</sup>提出基于缺陷纹理匹配Gabor滤波器,该

方法使用滤波器来消除背景纹理的影响,突出缺陷纹理,抑制背景纹理,从而获得良好的精度和检测速度。巢渊<sup>[15]</sup>利用SVM-RFE算法对芯片缺陷特征进行处理,消除无关集冗余特征,获得缺陷样本最优子集;陶志勇等<sup>[16]</sup>利用PSO\_SVM算法对电池板裂缝缺陷检测,这两类方法对形态特征相似的缺陷处理效率较低,处理方式相对繁琐。传统机器视觉技术的准确性并不高,虽有一系列高性能算法不断出现,如Ou等<sup>[17]</sup>采用霍夫线对二值分割芯片缺陷图像中的像素点坐标进行检测,但在实际应用中效果欠佳。

与传统的机器学习缺陷检测方法相比,CNN可以自主学习输入图像的特征,如周天宇等<sup>[18]</sup>利用YOLOv3算法检测载波芯片,提出了4种不同尺度的特征图,并通过增强融合特征的方式来改进多尺度检测;向宽等<sup>[19]</sup>提出了基于Faster R-CNN模型,利用FPNs融合多个不同深度特征图,使高级特征图保留了低维度特征,从而提升检测效果;Wen等<sup>[20]</sup>提出基于CNN的晶圆半导体表面缺陷检测方法,采用了无性卷积特征金字塔结构来提取特征,并生成特征映射,再将生成的特征图传入区域建议网络,生成建议区域,最后在进行分类和回归。上述基于CNN的缺陷检测方法能够有效提取缺陷特征,缓解尺度变化大出现检测效果差的问题,但周天宇等提出的方法只是针对轻量级数据检测,对特征复杂的缺陷检测效果不佳;向宽、Wen等提出的方法在建议网络中需要进行二次学习,在时效性方面存在一定弊端。上述方法虽对芯片检测中遇到的问题提出了解决方案,但多数研究只针对单一问题,如小目标缺陷漏检或尺寸跨度大、特征相似等,无法适用于本文复杂的检测场景。基于以上考虑,本文提出基于改进的YOLOv5芯片表面缺陷检测方法。

## 2 本文工作

本文针对芯片缺陷尺寸跨度大、特征相似、小目标难检测等问题,对YOLOv5网络结构做出了两点改进:1)针对小目标检测存在漏检等问题,新增更大尺寸特征检测器;2)针对缺陷尺寸跨度大、特征相似,提出EA-FPNs结构。图1为模型整体框架,首先对输入网络的图像尺寸进行预处理(Data Preprocessing),包括离线数据增强、调整图像尺寸大小为640×640;将预处理图像输入EA-FPNs结构,通过骨干网络(Backbone)提取网络特征,再将获取的特征图进行特征融合。其中,为了增强模型对小目标的检测能力,本文新增了S-Detector,该检测器拥有更大尺寸的特征图结构,且结合注意力模块(attention module, AM),能够提升模型对小目标的学习能力。在推理阶段,由于非极大抑制(non-maximum suppression,

NMS)<sup>[21]</sup> 算法过滤预测框后仍存在少量冗余框, 影响检测效果, 为此, 本文提出了 BFA, 该算法以预测框面积占

比为权重, 将冗余框进行融合, 最终呈现出最合理的预测效果。下面详细介绍以上改进的内容。

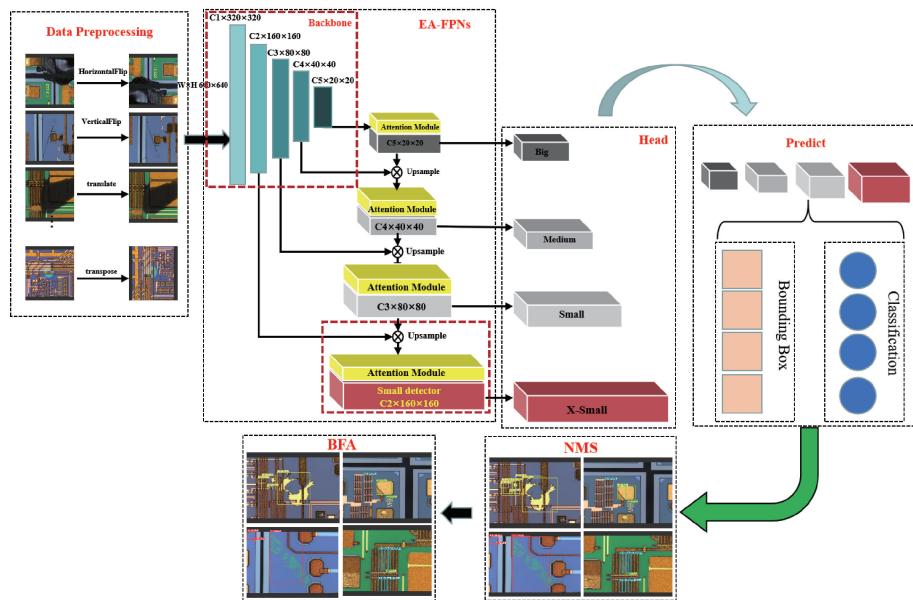


图 1 改进的 YOLOv5 网络结构

Fig. 1 The Improved network structure of YOLOv5

## 2.1 模型结构改进设计

由于在检测过程中, 芯片缺陷存在尺寸跨度大、特征相似、小目标难检测等问题, 本文在 YOLOv5 结构基础上, 提出了 EA-FPNs 结构, 缓解缺陷尺寸差异大而带来的定位不准问题; 在 EA-FPNs 的基础上新增加一层 S-Detector, 该结构拥有更大尺寸的特征检测器, 能够提供更多的空间信息, 从而对小目标缺陷的学习能力。下面介绍二者改进的内容。

### 1) S-Detector 结构介绍

图 2 为芯片缺陷尺寸分布和占比情况, (a) 为芯片尺寸分布情况, (b) 为芯片缺陷尺寸占比情况。以 COCO 标准中对小目标(小于  $32 \times 32$  像素)、中目标( $32 \times 32$  到  $96 \times 96$  像素)、大目标(大于  $96 \times 96$  像素)的定义, 缺陷尺寸主要集中在中、小目标, 共占比 66%, 其中, 小目标占比 16%。

由于在检测过程对小目标缺陷识别效果较差, 本文在 YOLOv5 结构的基础上新增一层 S-Detector, 其结构如图 3 所示。S-Detector 由两个分支构成:一个分支通过常规 CNN 提取特征检测器中小目标的纹理信息, 该分支输入输出的特征检测器通道数保持不变;另一个分支将 AM 融入特征检测器中, 首先将特征检测器通过 1 维卷积进行通道数降维, 再通过平均池化层提取其中的空间、背景信息, 这些信息能够加强模型对小目标缺陷的检测能力, 此时特征检测器会被压缩成 1 维通道, 最后将 1 维检测器进行通道扩充后, 与另外一分支的检测器进行融

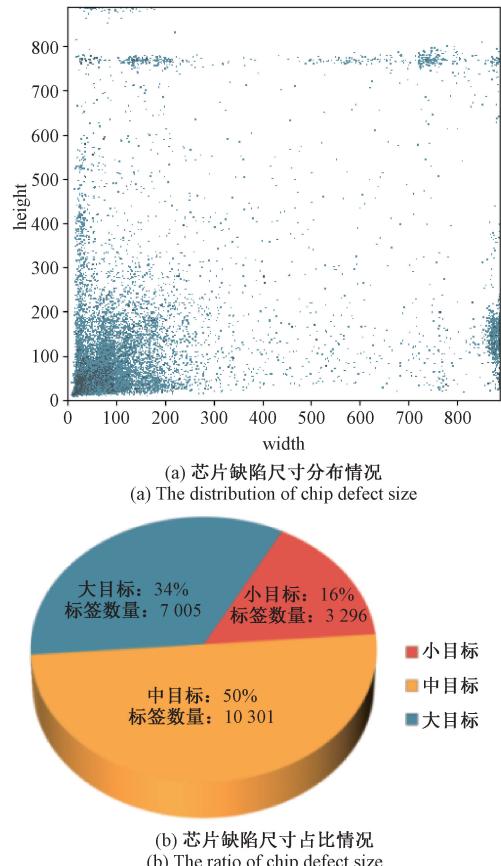


Fig. 2 The distribution and percentage of defect size

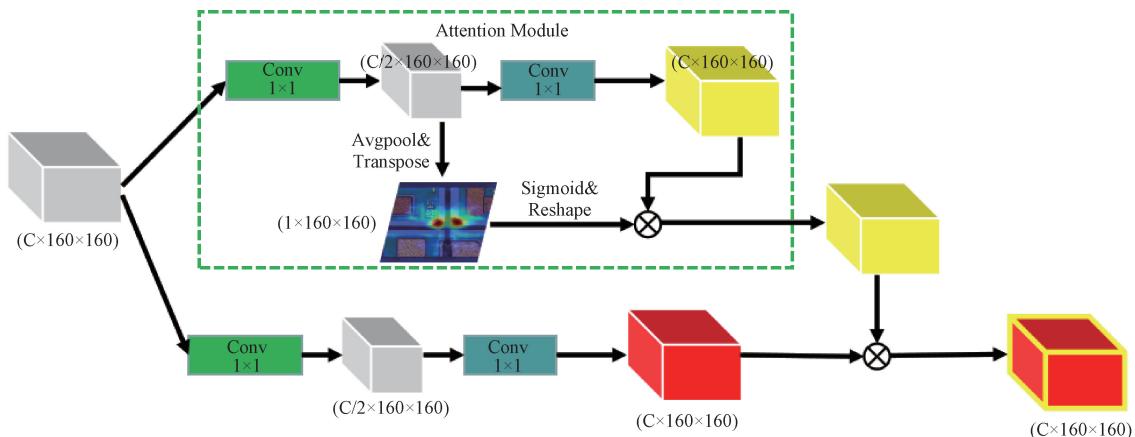


图3 S-Detector 结构

Fig. 3 The structure of the S-Detector

合,生成最终的 S-Detector。通过上述方式可以加强 S-Detector 对小目标缺陷的检测能力。

## 2) EA-FPNs 结构介绍

由于不同网络深度对目标特征的学习能力不同,浅层网络注重学习空间、背景等信息,对小目标更敏感;深层网络注重学习目标特征、纹理等信息,对尺寸较大目标的学习能力更强。FPN 以多尺度融合的方式,融合不同网络深度的特征信息,进而加强模型的对不同尺度目标的学习能力。芯片表面缺陷存在尺寸大小分布较广、跨度较大,在检测过程中易造成定位不准确、漏检等问题,为此,本文在新增 S-Detector 的基础上,对 FPN 结构进行改进,提出 EA-FPNs 结构。图 1 中,以 Darknet53 作为 EA-FPNs 的主干网络,划分为 5 层结构,每层负责不同尺寸缺陷特征的学习。将 5 种不同尺寸特征检测器上采样(Upsample)后进行融合,在融合阶段针对不同特征检测器添加 AM。经过 EA-FPNs 融合后生成 4 种不同尺寸的特征检测器,尺寸分别为  $20 \times 20$ 、 $40 \times 40$ 、 $80 \times 80$ 、 $160 \times 160$ ,前两个检测器负责检测大、中尺寸目标,后两个检测器负责检测小尺寸目标。通过 EA-FPNs 可以聚合不同层次特征图的特征信息,从而减缓缺陷尺寸跨度大带来的检测问题。

## 2.2 BFA 设计

模型在预测阶段会进行 NMS 过滤多余的预测框,经过多次实验发现,在 NMS 后仍然存在部分冗余框,为进一步加强过滤作用,本文提出了 BFA,该算法流程如算法 1 所示。

BFA 是基于面积占比权重的融合算法,首先计算所有冗余框的面积,并获取面积最大的冗余框  $S$ ,同时保留冗余框中置信度最大值;将  $S$  与其余冗余框作交并比<sup>[22]</sup>(intersection over union, IOU)计算;将高于阈值的冗余框与  $S$  进行融合即可获得最终的预测框,融合公式如下:

### 算法 1BFA 流程

开始

输入:冗余框集合  $A_1 = \{b_1, b_2, \dots, b_m\}$ ;

筛选冗余框阈值  $T$ ;

冗余框面积集合  $S_1 = \{s_1, s_2, \dots, s_m\}$ ;

融合框  $B_1$ ;

1. **for**  $i=1, 2, \dots, m$  **in**  $A_1$

计算  $A_1$  中冗余框面积,存放在  $S_1$  中,并按降序排序;获取置信度最大值 Scores;

2. **for**  $j=1, 2, \dots, m$  **in**  $S_1$

计算最大面积预测框  $S_j$  与其余预测框做 IOU 计算,统计大于  $T$  的预测框下标集合  $A_2$

3. **for**  $k=1, 2, \dots, \text{len}(A_2)$

将面积最大预测框  $s_1$  与  $A_{2k}$  进行融合

4. 将 Score 赋值给融合之后的预测框。

结束

$$C_{x_1} = \frac{A_{x_1} \times A_{area} + B_{x_1} \times B_{area}}{A_{area} + B_{area}} \quad (1)$$

$$C_{x_2} = \frac{A_{x_2} \times A_{area} + B_{x_2} \times B_{area}}{A_{area} + B_{area}} \quad (2)$$

$$C_{y_1} = \frac{A_{y_1} \times A_{area} + B_{y_1} \times B_{area}}{A_{area} + B_{area}} \quad (3)$$

$$C_{y_2} = \frac{A_{y_2} \times A_{area} + B_{y_2} \times B_{area}}{A_{area} + B_{area}} \quad (4)$$

其中,  $C_{x_1}, C_{x_2}, C_{y_1}, C_{y_2}$  表示边界框的下标,  $A, B$  表示进行融合的冗余框,  $A_{area}, B_{area}$  分别表示冗余框的面积。相比以置信度为权重进行融合的方法,以面积占比为权重的融合方式避免了出现融合框定位不准的情况。

## 3 实验与结果分析

在本节中详细介绍数据集制作过程、实验过程、以及实验结果和总结,详细说明如下。

### 3.1 数据集准备

本文数据集来自生产线,共 8 818 张,图像大小为  $888 \times 888$ ,训练集、验证集、测试集按 VOC 数据集标准 3 : 1 : 1 进行划分,通过 LabelImg 标准工具进行类别、边

界框标注。数据集共包含 12 个类别,详情如图 4 所示,其中包括:变色、多边少边、裂片、崩边、多金、金疙瘩、图层不全、桥压、划伤、脏污、栅条缺损、钻蚀,数据集类别划分情况如表 1 所示。

表 1 数据集类别划分情况  
Table 1 The classification of data sets

第 1 类	第 2 类	第 3 类	第 4 类	第 5 类	第 6 类	第 7 类	第 8 类	第 9 类	第 10 类	第 11 类	第 12 类
变色	多边少边	裂边	崩边	多金	金疙瘩	图层不全	桥压	划伤	脏污	栅条缺损	钻蚀

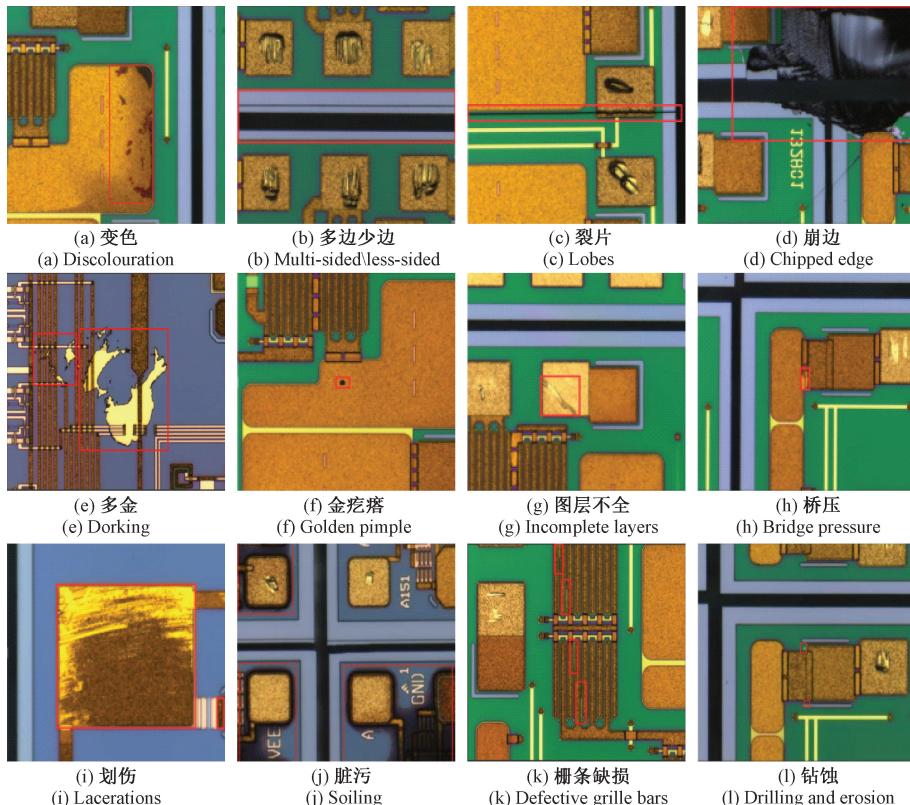


图 4 缺陷类别展示

Fig. 4 The display of defects category

### 3.2 实验准备

1) 实验环境。本文以 PyTorch 作为网络结构和模型训练的框架,实验硬件环境如下:NVIDIA RTX A6000 显卡,显存为 48 GB;CPU 为 Intel (R) Xeon (R) W-2223,64 位操作系统;实验参数如下:训练的图像批处理设置为 bit;共进行 150 次网络训练,学习率为 0.001。

2) 数据增强。缺陷类别分布存在不均匀的情况,将导致实验出现严重过拟合,从而无法进行后续实验。图 5 为数据增强前,各类别标签数量及训练的 AP 情况。由图可知,不同种类缺陷数量严重不平衡,其中第 3、6、12 类缺陷数量太少,第 10 类缺陷数量大多,因此在实验前

采用离线数据增强扩充缺陷数量较少的类别,以此维持训练数据集缺陷种类平衡。增强方式以不改变缺陷原特征,维持数据集比例 3 : 1 : 1 为前提,包括:水平翻转、垂直翻转、平移、转置。图 1 中数据预处理模块为数据增强样例。经过数据增强扩充较少的缺陷数量后,得到训练集 5 005 张、验证集 1 891 张、测试集 1 922 张。

3) 评估指标。实验评估指标采用精确度(precision, P)、召回率(recall, R)、各类别平均精度(average precision, AP)、mAP@0.5、mAP@.5:.95 作为评估指标,小目标缺陷的检测性能使用  $AP^{small}$ 、 $AP^{middle}$ 、 $AP^{large}$  和  $AR^{small}$ 、 $AR^{middle}$ 、 $AR^{large}$ 。

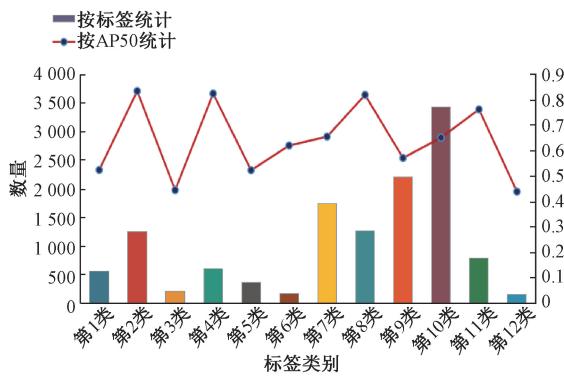


图 5 数据增强前各类别类数量和 AP 情况

Fig. 5 The number of each category and AP before data enhancement

AP、AR 表示某一缺陷的平均 P 值, 平均 R 值, 公式如下:

$$AP = \int_0^1 P(x) dx, P = \frac{TP}{TP + FP} \quad (5)$$

$$AR = \int_0^1 R(x) dx, R = \frac{TP}{TP + FN} \quad (6)$$

其中, TP 表示真实为正样本, 检测为正样本; FP 表示真实为负样本, 检测为正样本; FN 表示真实为正样本, 检测为负样本; P 表示真实为正样本, 且检测为正样本的检测比率; R 表示正确预测缺陷的比率。

mAP@ 0.5 表示 IOU 阈值为 0.5 时, 总类别的平均精度, 公式如下:

$$mAP = \frac{1}{C} \sum_j^c AP_j \quad (7)$$

其中, C 表示总类别数, j 表示类别。

mAP@ . 5 : . 95 表示 IOU 阈值以 0.05 为间隔, 从 0.5~0.95 的平均精度; AP<sup>small</sup>、AP<sup>middle</sup>、AP<sup>large</sup> 和 AR<sup>small</sup>、AR<sup>middle</sup>、AR<sup>large</sup> 表示小、中、大目标的平均 P 值和平均 R 值。

### 3.3 实验结果展示与分析

#### 1) 不同模型结果对比

为进一步验证本文方法的有效性, 将本文模型与常见的缺陷检测模型 Faster R-CNN、YOLOv3, 以及 YOLOv5-M、YOLOv5-L、YOLOv5-X 在相同数据集上进行对比实验, 实验结果如表 2、3 所示, 其中 Layers 表示网络层数, Params 表示网络参数量。

表 2 不同模型结果对比

Table 2 The comparison of results with different models

模型	Backbone	Size	Layers	Params/M	mAP@ 0.5	mAP@ . 5 : . 95
Faster R-CNN	Resnet101	608×608	599	101.2	71.9%	36.4%
YOLOv3	Darknet-53	640×640	261	61.5	75.2%	38.7%
YOLOv5-EA-FPNs	Darknet-53	640×640	576	86.2	79.4%	44.1%

表 3 不同网络深度结果对比

Table 3 The comparison of results with different network depths

模型	Backbone	Size	Layers	Params/M	mAP@ 0.5	mAP@ . 5 : . 95
YOLOv5-M	Darknet-53	640×640	369	20.9	76.3%	40.1%
YOLOv5-L	Darknet-53	640×640	468	41.2	78.3%	36.5%
YOLOv5-X	Darknet-53	640×640	576	86.2	76.5%	39.3%
YOLOv5-EA-FPNs	Darknet-53	640×640	532	68.4	79.4%	44.1%

由实验结果可知, 在相同数据集的情况下, Faster R-CNN 由于采用 resnet101 作为骨干网络, 该网络结构较深, 不适合当前数据集, 所以导致实验结果不佳; YOLOv3 属于轻量级网络结构, 由于当前数据集数量较多, 且缺陷复杂, 所以导致实验结果较差; YOLOv5 三组模型, 主干网络相同, 模型深度不同, 其中 M 网络深度相对较浅, 参数量较少, X 网络相对较深, 参数量较多, 与当前数据集都不契合, 只有 L 模型效果最佳; 本文在 L 模型上进行改进, 提出 YOLOv5-EA-FPNs, 由实验结果可知, 本文改进的模型实验结果最好, mAP@ 0.5、mAP@ . 5 : . 95 分别为 79.4%、44.1%, 相比于 L 模型增长了 1.2%、7.6%, 表明

本文提出的模型更加适合当前芯片数据集。

#### 2) EA-FPNs 实验结果对比

为验证本文提出的 EA-FPNs 结构的有效性, 选取 YOLOv5-L 模型进行消融实验。由表 2 可知, YOLOv5-L 是除本文提出的模型外, 当下实验效果最好的模型, 所以选取 YOLOv5-L、YOLOv5-EA-FPNs 进行消融实验。

图 6 为本次消融实验结果。从图中可知, EA-FPNs 对变色、金疙瘩、划伤、钻蚀提升效果不明显, 对多边少边、崩边、栅条缺损等尺度变化较大的缺陷, 平均精度有明显的提升, 可见 EA-FPNs 能够提升尺度差异较大缺陷的检测效果。

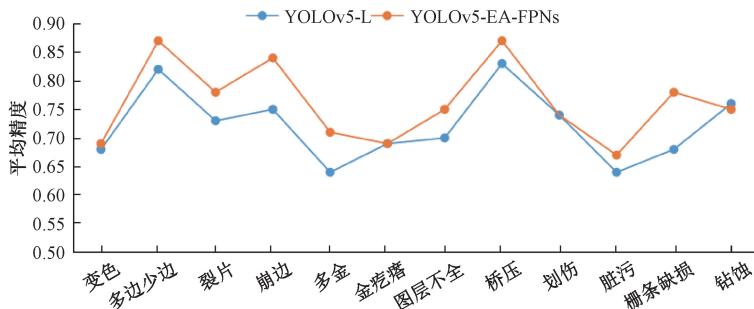


图 6 各类别缺陷平均精度

Fig. 6 The average precision of defects by class

在表 4 中,由小、中、大目标的 AP、AR 实验结果可知,小目标的 AP、AR 分别上升了 1.6% 和 0.8%,且对中、大目标的检测结果没有造成影响。图 9 为 S-Detector 特征图展示结果,其中,前 3 层特征图 (F1、F2、F3) 为

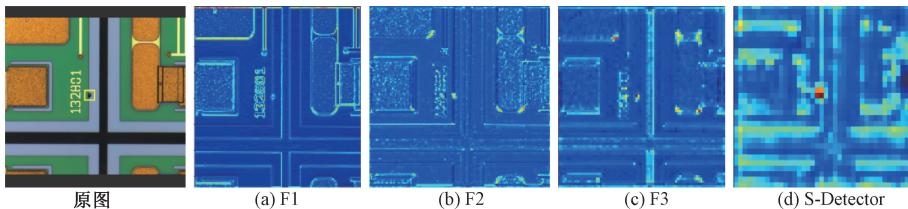


图 7 S-Detector 特征图展示

Fig. 7 The display of S-Detector feature maps

表 4 S-Detector 小目标实验结果对比

Table 4 The comparison of the results with S-Detector small target experiment

	AP <sup>small</sup>	AP <sup>middle</sup>	AP <sup>large</sup>	AR <sup>small</sup>	AR <sup>middle</sup>	AR <sup>large</sup>
YOLOv5-L	12.3%	18.5%	23.3%	17.1%	22.7%	28.4%
YOLOv5-EA-FPNs	13.9% (↑1.6%)	18.4%	23.6% (↑0.3%)	17.9% (↑0.8%)	22.7%	28.4%

图 8 为 YOLOv5-L 与 YOLOv5-EA-FPNs 检测结果对比情况。检测结果表明,本文改进的结构检测精度较高,且漏检率、冗余框、小目标检测率等情况相对较好。

### 3) BFA 检测结果对比

由于 NMS 算法存在没有过滤掉冗余框的情况,本文提出了 BFA,对冗余框进行融合,以获得更好的检测结果。图 9 为 BFA 检测结果对比情况。由图 9 可知,该算法能够在不影响预测框精准度的情况下,有效的消除了冗余框,并且保留了最高置信度分数。

表 5 为使用 BFA 前后 2 500 张数据集测试时长、速度对比情况,没有使用 BFA 前,EA-FPNs 推理时长为 85 s,平均 32  $\mu$ s/张;使用 BFA 的推理时长为 67 s,平均 26.8  $\mu$ s/张。

YOLOv5-L 的小目标缺陷提取效果,S-Detector 为新增特征图的提取效果,从图 7 中可以看出,S-Detector 对小目标特征的提取能力更强,学习效果更好。

表 5 BFA 测试时长对比情况

Table 5 The comparison of BFA test duration

	测试数量/张	总时长/s	平均时长/(张/ $\mu$ s)
Before BFA	2 500	85	32 张/ $\mu$ s
After BFA	2 500	67 (↓18)	26.8 张/ $\mu$ s (↓5.2 $\mu$ s)

## 4 结 论

本文使用改进的 YOLOv5 深度学习网络对芯片表面缺陷进行检测,有如下 2 个方面的改进。

1) 针对芯片缺陷尺寸跨度大,存在类间相似性等问题,本文提出了 EA-FPNs 结构,加强对不同尺度芯片缺陷的学习;针对小目标检测不佳的问题,本文在 YOLOv5 结构基础上,新增 S-Detector,并融合 AM,从而加强模型对小目标缺陷特征的学习能力。

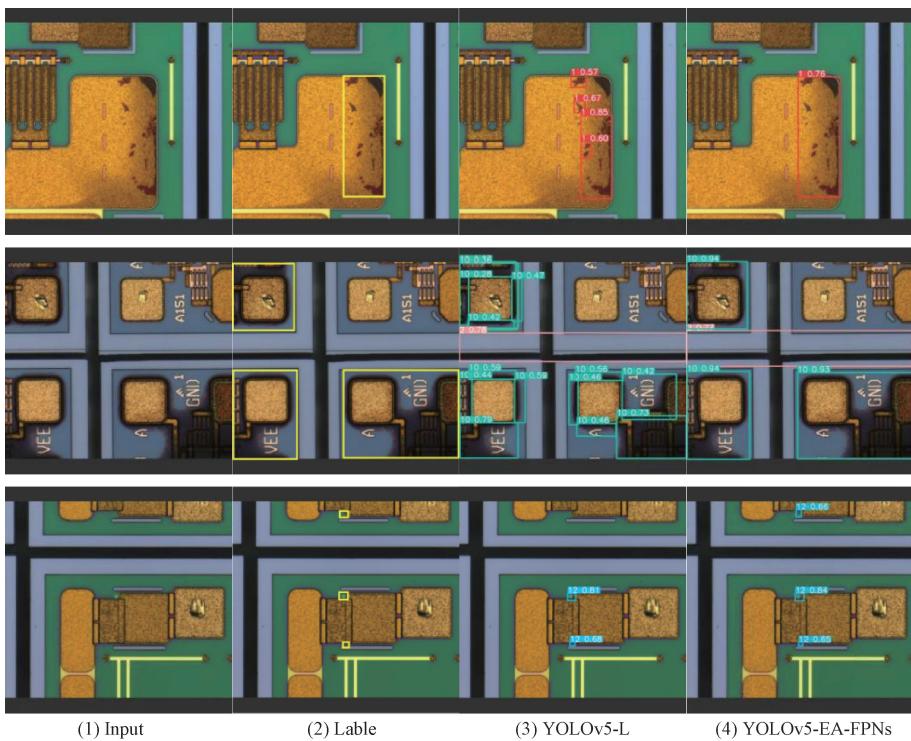


图 8 YOLOv5-L 与 YOLOv5-EA-FPNs 检测结果对比情况

Fig. 8 The comparison of detection results with YOLOv5-L and YOLOv5-EA-FPNs

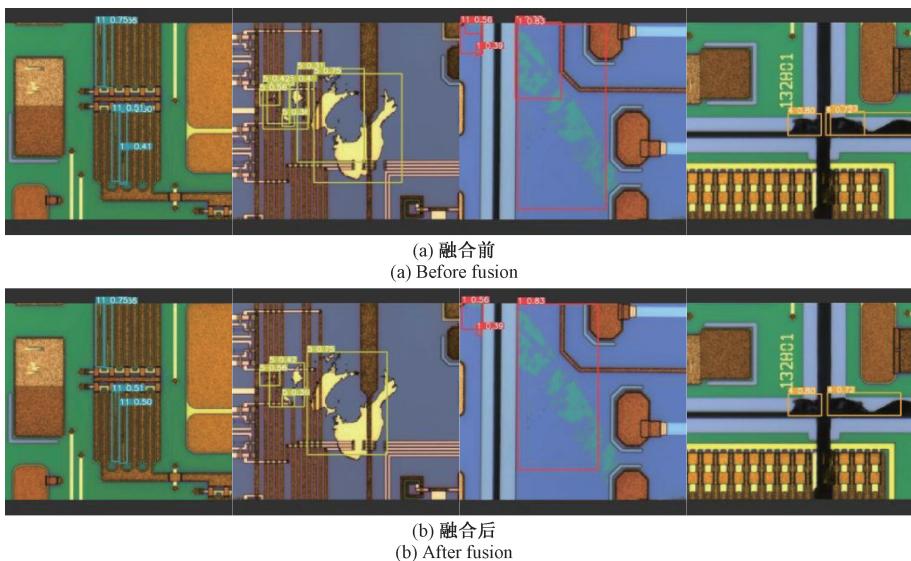


图 9 BFA 检测结果对比情况

Fig. 9 The comparison of bounding BFA detection results

2) 针对后处理阶段,预测框存在冗余情况,造成时间开销大的问题,本文以面积占比为权重,融合多余预测框,并以最高置信度作为融合框的置信度,提出了BFA,该算法在确保最高置信度的同时,避免了预测框定位不准的情况。

本文模型与YOLOv5、Faster R-CNN、YOLOv3等进行实验对比,实验结果表明,YOLOv5-EA-FPNs较上述模型在检测精度上效果更佳,对比YOLOv5-L,mAP@0.5提升了1.2%,小目标检测精度提升了1.6%;使用BFA消除冗余框的同时,平均检测时长为26.8 μs/张,较使用

BFA 前减少了 5.2 μs。为工业芯片表面缺陷检测提供了一种有效的方案。

## 参考文献

- [1] PARK J K, KWON B K, PARK J H, et al. Machine learning-based imaging system for surface defect inspection [J]. International Journal of Precision Engineering and Manufacturing-Green Technology, 2016, 3: 303-310.
- [2] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.
- [3] REDMON J, FARHADI A. YOLOv3: An incremental improvement [J]. arXiv preprint arXiv: 1804.02767, 2018.
- [4] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv:2004.10934, 2020.
- [5] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector [C]. Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part I 14. Springer International Publishing, 2016: 21-37.
- [6] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2980-2988.
- [7] GIRSHICK R, DONAHUE J, DARRELL T, et al. Region-based convolutional networks for accurate object detection and segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 38(1): 142-158.
- [8] GIRSHICK R. Fast R-CNN [C]. Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448.
- [9] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. Advances in Neural Information Processing Systems, 2015, 28.
- [10] HE K, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2961-2969.
- [11] LI D, LI Y, XIE Q, et al. Tiny defect detection in high-resolution aero-engine blade images via a coarse-to-fine framework [J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 1-12.
- [12] CAO J, CHEN Q, GUO J, et al. Attention-guided context feature pyramid network for object detection[J]. arXiv preprint arXiv:2005.11475, 2020.
- [13] 曹新容, 薛岚燕, 林嘉雯, 等. 基于匹配滤波和自动阈值的眼底血管分割方法[J]. 电子测量与仪器学报, 2017, 31(1):51-57.
- [14] CAO X R, XUE L Y, LIN J W, et al. Fundus vascular segmentation method based on matched filtering and automatic threshold [J]. Journal of Electronic Measurement and Instrumentation, 2017, 31(1):51-57.
- [15] 朱刘盅, 李竞择, 吴磊, 等. 基于 Gabor 滤波器的光纤缺陷检测方法[J]. 南京理工大学学报: 自然科学版, 2022, 46(1):119-126.
- [16] ZHU L ZH, LI J Z, WU L, et al. Gabor filter based optical fiber defect detection method [J]. Journal of Nanjing University of Science and Technology: Natural Science Edition, 2022, 46(1):119-126.
- [17] 巢渊. 基于机器视觉的半导体芯片表面缺陷在线检测关键技术研究[D]. 南京:东南大学, 2017.
- [18] CHAO Y. Research on key technologies of online detection of semiconductor chip surface defects based on machine vision [D]. Nanjing: Southeast University, 2017.
- [19] 陶志勇, 于子佳, 林森. PSO\_SVM 算法在太阳能电池板裂缝缺陷检测研究[J]. 电子测量技术, 2021, 44(1):18-25.
- [20] TAO ZH Y, YU Z J, LIN S. Study of PSO\_SVM algorithm in solar panel crack defect detection [J]. Electronic Measurement Technology, 2021, 44 (1): 18-25.
- [21] OU X F, CHEN W, ZHANG M. Design of LQFP chip pin defect detection system based on machine vision[C]. 2021 International Conference on Machine Learning and Intelligent Systems Engineering (MLISE). IEEE, 2021: 69-76.
- [22] 周天宇, 朱启兵, 黄敏, 等. 基于改进 YOLOv3 的载波芯片缺陷检测[J]. 激光与光电子学进展, 2021, 58(12):86-93.
- [23] ZHOU T Y, ZHU Q B, HUANG M, et al. Carrier chip defect detection based on improved YOLOv3 [J]. Advances in Lasers and Optoelectronics, 2021, 58(12): 86-93.
- [24] 向宽, 李松松, 栾明慧, 等. 基于改进 Faster R-CNN 的铝材表面缺陷检测方法[J]. 仪器仪表学报, 2021, 42(1):191-198.
- [25] XIANG K, LI S S, LUAN M H, et al. Aluminum surface defect detection method based on improved faster R-CNN [J]. Chinese Journal of Scientific Instrument, 2021, 42(1):191-198.
- [26] WEN G, GAO Z, CAI Q, et al. A novel method based

- on deep convolutional neural networks for wafer semiconductor surface defect inspection [ J ]. IEEE Transactions on Instrumentation and Measurement, 2020, 69( 12 ): 9668-9680.
- [21] BODLA N, SINGH B, CHELLAPPA R, et al. Improving object detection with one line of code [ J ]. 2017, DOI:10.48550/arXiv.1704.04503.
- [22] HE J, ERFANI S, MA X, et al. Alpha-IoU: A family of power intersection over union losses for bounding box regression [ J ]. Advances in Neural Information Processing Systems, 2021, 34: 20230-20242.

## 作者简介



张恒,2009年于中南大学获得学士学位,2014年于中国科学院大学获得硕士学位,2020年于中国科技大学获得博士学位,现为重庆邮电大学副教授,主要研究方向为人工智能、特种视觉和计算成像。

E-mail: zhangheng@cqupt.edu.cn

**Zhang Heng** received his B. Sc. degree from Central South

University in 2009, M. Sc. degree from University of Chinese Academy of Sciences in 2014 and Ph. D. degree from University of Science and Technology of China in 2020. He is now an associate professor at Chongqing University of Posts and Telecommunications. His main research interests include artificial intelligence, specialty vision and computational imaging.



杭芹(通信作者),2009年于中南大学获得学士学位,2015年于中南大学获得硕士学位,2019年于中国科技大学获得博士学位,现为重庆邮电大学讲师,主要研究方向为人工智能、特种视觉和计算成像。

E-mail: hangqin@cqupt.edu.cn

**Hang Qin** ( Corresponding author ) received her B. S. degree from Central South University in 2009, M. Sc. degree from Central South University in 2015 and Ph. D. degree from University of Science and Technology of China in 2019. She is now a lecturer at Chongqing University of Posts and Telecommunications. Her main research interests include artificial intelligence, specialty vision and computational imaging.