

面向多尺度缺陷的轻量级双分支检测网络研究

余铭琪¹ 岳洪伟^{2*}

YU Mingqi YUE Hongwei

摘要

针对感光元件缺陷人工检测效率低、小目标漏检率高等问题,文章提出了一种基于YOLOv4的双分支轻量级缺陷检测方法。首先采用MobileNetV2替换CSPDarknet53主干网络构建轻量基线,利用K-means++聚类生成适配缺陷特征的锚框;引入Focal Loss缓解样本不平衡,结合余弦退火策略优化训练过程;最终设计双分支网络TBNNet,通过分尺度处理增强多尺度缺陷检测能力。实验表明,所提方法在自建数据集上平均精度达95.45%,其中坏点检测精度97.64%,污点检测精度93.26%,在保持轻量化的同时显著提升小目标检出率,为高分辨率工业图像缺陷检测提供有效解决方案。

关键词

感光元件;YOLOv4;轻量级;缺陷检测;MobileNetV2

doi: 10.3969/j.issn.1672-9528.2026.05.026

0 引言

摄像头模组需要在无尘车间生产组装,其中目检、感光元件清洁、镜头点胶等是必须的生产工艺流程。由于生产工艺和环境的制约,摄像头模组在生产过程中难免会引入工业缺陷,比如感光元件长时间在高温状态下就会产生坏点缺陷。同时,摄像头模组的生产组装必须在无尘环境中进行,如果组装前对感光元件和镜头除尘不彻底,则灰尘会导致产生污点缺陷。其中坏点缺陷往往较小,在可视化下通常表现为黑色斑点;污点缺陷较大,又可以分为明显污点和不明显污点。而且灰点的面积较小,污点呈现透明形态。这些特性使得检测具有一定的困难。

1 研究现状

传统人工检测方法受人工经验和主观因素的影响大,存在劳动强度高,分拣质量不稳定等问题。基于机器视觉的检测方法可以很大程度上克服人工检测弊端^[1-2]。由于本文的透明污点缺陷和背景接近,同时人工设定的缺陷识别特征缺乏一定的鲁棒性,当背景环境、拍摄角度、缺陷位置发生改变时,机器视觉方法无法得到较好的识别效果。

随着人工智能技术的发展,基于深度学习的目标识别方法在工业器件的缺陷识别中得到广泛应用,并取得较好的效果^[3]。EfficientDet^[4]、Faster R-CNN^[5]和CenterNet^[6]等经典目标检测模型,经过研究者在工业缺陷特定场景下的

优化,持续展现出强大的性能和实用价值。麻越等^[7]利用YOLOv3针对热轧板卷表面缺陷增加检测分支扩大感受野,提高模型对小目标缺陷检测的能力,并且在模型训练上,优化了学习率设置、损失函数及最优模型的选择策略,提高了模型的检测速度以及准确率。王立琦等^[8]基于YOLOv4针对锥桶目标,从锚框、损失函数、轻量化都进行了对应的优化,其中利用K-means++优化锚框,用SIoU替换了原始的CIoU损失函数,用MobileNetV1的深度可分离卷积达到轻量化的效果。

由于感光元件缺陷样本较少,而且坏点缺陷和污点缺陷往往会同时出现,受上述文献的启发,因此本文提出了一种基于YOLOv4的轻量级双分支联合检测网络(two branch joint detection network, TBNNet)完成感光元件缺陷检测方法。首先利用K-means++优化网络的锚框;然后利用Focal Loss^[9]来减少易分样本的权重,从而使模型更加关注难分样本;接着在余弦退火调度器^[10]的基础上,使用固定学习率进行微调达到优化网络的检测性能的目的;最后利用TBNNet来完成污点和坏点的同时检测。结果表明,该方法能够有效处理感光元件的缺陷检测。

2 高精度轻量YOLOv4缺陷检测

在本节中,首先详细介绍了YOLOv4算法,考虑到YOLOv4算法参数量大,检测精度存在提升空间,本文把YOLOv4的主干特征提取网络替换成轻量级网络MobileNetV2^[11],这大大降低了计算成本;然后提出了一系列的改进措施:一是锚框的优化;二是损失函数的优化;三是针对小目标难检问题而提出的TBNNet网络。

1. 广州工商学院 广东广州 510850

2. 广东第二师范学院 广东广州 510303

[基金项目] 广东高等学校科研平台和项目(2023ZDZX4039)

2.1 YOLOv4 基线模型概述

作为本文工作的基础，YOLOv4 检测框架主要包含三部分：采用 CSPDarknet53 作为主干网络以提取丰富特征；使用融合了空间金字塔池化与路径聚合网络的颈部网络进行多尺度特征融合；其检测头则沿用 YOLOv3 的设计，负责最终的目标分类与定位。

2.2 轻量级主干网络替换

虽然 CSPDarknet53 具有强大的特征提取能力，但是其网络结构复杂，导致模型参数数量和计算成本较高，不利于实时检测任务。针对此问题，本文将其替换为轻量化的 MobileNetV2 网络。该网络通过深度可分离卷积实现高效计算，理论分析表明，其计算成本仅为标准卷积的约 1/9。基于这一显著优势，本文构建了 YOLOv4-MobileNetV2 基础轻量模型，其结构如图 1 所示。

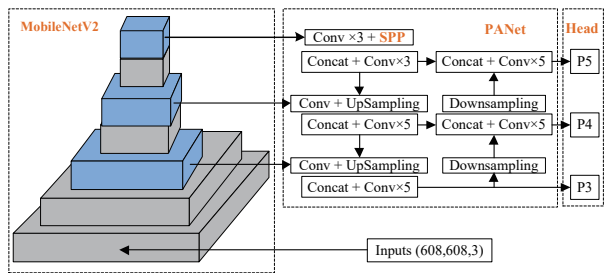


图 1 YOLOv4-MobileNetV2 网络结构

2.3 检测性能优化策略

2.3.1 基于 K-means++ 的锚框初始化设计

YOLOv4 使用锚框作为先验知识，以提升模型收敛速度与检测性能。然而，其默认锚框尺寸基于 VOC/COCO 等通用数据集设计，与本文所采用的感光元件缺陷数据集中目标尺寸分布存在显著差异。为增强锚框对缺陷特征的适应性，本文采用 K-means++ 聚类方法，重新生成适用于本任务的锚框尺寸。

具体而言，本文以训练集中所有真实边界框的宽度与高度作为聚类对象，使用 K-means++ 算法进行聚类。该算法通过概率加权方式选取初始聚类中心，有效缓解了传统 K-means 对初始值敏感的问题，从而获得更具代表性的锚框分布。在聚类过程中，采用 1-IoU 作为距离度量，以直接优化锚框与真实框之间的重叠程度。参考 YOLOv4 的典型设置，聚类数目设为 9，并将其结果按尺度划分为三个层级，分别对应不同大小的目标，具体锚框尺寸如表 1 所示。

表 1 K-means++ 聚类生成的锚框尺寸

尺度	目标类别	锚框尺寸
76×76	小目标	(7,7)(16,17)(25,24)
38×38	中目标	(35,34)(40,53)(55,44)
19×19	大目标	(61,67)(83,84)(134,127)

2.3.2 基于 Focal Loss 的损失函数优化

单阶段目标检测器（如 YOLOv4）在训练过程中常受到正负样本不平衡以及难易样本不平衡问题的困扰。大量的简单背景样本会导致损失函数被其主导，使得模型难以专注于学习稀疏的前景目标和困难样本。

为解决上述问题，本文采用 Focal Loss 以替代 YOLOv4 原目标置信度损失函数。Focal Loss 在标准交叉熵损失的基础上引入了可调节的缩放因子，用公式表示为：

$$FL(p_i) = -\alpha_i(1 - p_i)^\gamma \log(p_i)$$
(1)

式中： p_i 为模型对目标类别的预测概率， α_i 为用于平衡正负样本数量的权重因子， γ 为聚焦参数，用于降低易分类样本的损失贡献，使模型训练更专注于困难样本。

在本文实验中，设定 $\alpha_i=0.25$, $\gamma=2$ 。最终，将 YOLOv4 总体损失函数中的原始目标置信度损失函数 L_{conf} 替换为 $FL(p_i)$ ，形成优化后的总损失函数：

$$Loss_{Ours} = L_{loc} + L_{cls} + FL(p_i)$$
(2)

式中： L_{loc} 与 L_{cls} 分别为边界框回归损失与分类损失，其定义与原 YOLOv4 保持一致。

2.3.3 双分支检测网络

YOLOv4 虽然具备多尺度特征提取能力，但其输入尺寸固定为 608 px×608 px。当将本文所使用的高分辨率原始图像（3 280 px×2 464 px）下采样至该尺寸时，图像中的微小坏点缺陷特征会严重损失，从而导致检测性能下降。

为解决该问题，本文提出一种专为高分辨率图像中小目标检测设计的双分支网络。该网络基于前述 YOLOv4-MobileNetV2 结构构建，整体架构如图 2 所示。其核心思想是通过两个分支分别处理不同尺度的输入，从而兼顾大目标的定位精度与小目标的特征保留：分支 1 将图像缩放至 608 px×608 px，利用感受野较大的深层特征（输出尺度为 38×38 和 19×19，对应 P5 与 P4）进行预测，主要检测污点等尺度较大的缺陷；分支 2 则输入尺寸为 1 568 px×1 568 px 的较大分辨率图像，借助高分辨率浅层特征（输出尺度为 196×196，对应 P3）进行预测，专门负责识别坏点等微小缺陷，有效避免了因图像过度下采样而引起的特征丢失。

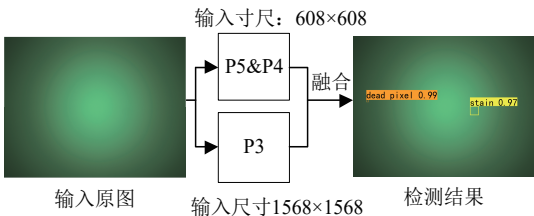


图 2 双分支联合检测网络

两个分支共享同一个预训练的主干网络权重，通过利用 YOLOv4 固有的多尺度预测头，并为其分配不同尺度的输入

与特征层，实现了高效的双分支协同检测。最终，融合两个分支的预测结果，当缺陷置信度大于 0.5 时判定为有效目标。在图 2 中，污点与坏点缺陷分别用 stain 与 dead pixel 表示。

3 实验结果及分析

3.1 实验参数设置

本实验采用 Adam 优化器与余弦退火学习率调度器，训练过程分为三个阶段。第一阶段进行冻结训练，共 100 个 Epoch，初始学习率为 $1e-3$ ，下限学习率为 $1e-5$ ，批处理大小设为 16。第二阶段为解冻训练，共 200 个 Epoch，初始学习率设置为 $1e-4$ ，下限学习率为 $1e-5$ ，批处理大小调整为 8。第三阶段在前一阶段基础上使用固定学习率 $1e-6$ 微调 50 个 Epoch，以稳定模型参数并促进最终收敛。

在模型训练过程中，本文采用了分阶段的学习率调度策略。如图 3 所示，训练过程可分为两个阶段：在前阶段使用余弦退火调度器，使学习率周期性下降，这有助于模型跳出局部最优点，损失曲线呈现稳步下降趋势；在后阶段切换为固定学习率进行微调，损失值最终收敛并在 0.028 附近小幅波动，表明模型已充分训练，达到稳定状态。

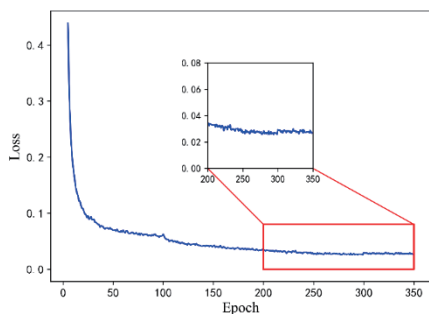


图 3 Loss 曲线

3.2 消融实验

为系统验证各项改进策略的有效性，本文进行了消融实验，结果如表 2 所示。实验设置了不同配置的模型，其中“M”代表基础轻量模型（MobileNetV2 主干）。

表 2 消融实验结果

编号	M	K	F	C	T	AP20 (坏点) /%	AP50 (污点) /%	mAP /%
1						81.08	84.81	82.95
2	√					77.74	85.18	81.46
3	√	√				78.23	90.93	84.58
4	√	√	√			79.58	92.68	86.13
5	√	√	√	√		80.54	93.26	86.38
6	√	√	√	√	√	97.64	93.26	95.45

分析可得出以下结论：一是基础轻量模型“M”在参数

量大幅减少的情况下，其 mAP 仅比原 YOLOv4 基准模型低 1.49 个百分点，证明了轻量化设计的有效性；二是在“M”基础上引入 K-means++ 的锚框优化（记为“M+K”）后，模型对污点的检测精度（AP50）显著提升 5.75 个百分点，这得益于定制的锚框尺寸更贴合本数据集中缺陷的实际分布；三是进一步引入 Focal Loss（记为“M+K+F”）以缓解正负样本不平衡问题，使模型更聚焦于困难样本，从而将整体 mAP 提升了 1.55 个百分点。随后，结合余弦退火与固定学习率调度策略（记为“M+K+F+C”）带来了性能的稳定微幅提升。最后，引入专为小目标设计的双分支网络（记为“M+K+F+C+T”，即完整 TBNNet）对坏点检测产生了决定性改善，使其 AP50 指标大幅提升 17.1 个百分点。集成所有改进的完整模型在所有缺陷类别上均达到了最优性能。

为直观展示本文改进策略的渐进式效果，对 YOLOv4（基线模型）、YOLOv4-M（主干网络替换为 MobileNetV2 的模型）及 TBNNet（本文最终模型）的检测结果进行对比，如表 3 所示。对比结果显示，相较于 YOLOv4-M，本文最终模型 TBNNet 在目标定位精度与不明显污点的检出能力方面均有显著改善。尤其对于分布于图像边缘或特征微弱的坏点，其在常规下采样过程中极易丢失有效特征，导致漏检；而 TBNNet 凭借其双分支结构中的高分辨率输入，能够有效保留此类目标的细节信息，从而实现对边缘坏点与低对比度缺陷的稳定检测。

表 3 不同模型检测效果对比

缺陷	方法	YOLOv4	YOLOv4-M	TBNNet
不明显污点				
边缘坏点				

3.3 方法对比

为全面评估 TBNNet 算法的有效性，本文将其与多种主流目标检测方法进行对比。鉴于本文核心在于通过双分支架构解决高分辨率下的小目标检测难题，而非追求极致的理论计算效率，因此对比实验聚焦于更具实际意义的指标：检测精度（AP、mAP）与模型部署成本（权重文件大小）。传统的参数量（Params）与计算量（FLOPs）指标难以准确衡量双分支网络中特征共享与异构路径带来的综合效能，而所选指

标能更直接地反映模型在真实工业场景下的价值。对比实验结果如表 4 所示。

表 4 不同模型结果对比

模型	AP20（坏点）/%	AP50（污点）/%	mAP /%	权重文件大小 /MB
EfficientDet-D1	—	85.57	42.79	25.7
Faster R-CNN (ResNet50)	8.83	79.42	44.13	108
CenterNet (ResNet50)	76.47	89.07	82.77	124
YOLOv3	75.79	89.97	82.88	235
YOLOv4	81.08	84.81	82.95	244
YOLOv4-M	77.74	85.18	81.46	46.5
TBNet	97.64	93.26	95.45	46.5

分析可知，EfficientDet-D1 由于深层特征图经多次下采样导致小目标特征严重丢失，未能检测到坏点；Faster R-CNN 对浅层位置信息利用不足，而该信息对小目标检测至关重要，因此检测效果欠佳；CenterNet 虽通过 Focal Loss 降低了背景权重，但其依赖上采样还原特征，导致浅层细节受损，影响小目标检测精度；YOLOv3 通过特征金字塔实现自上而下的多尺度融合与独立预测，但缺乏自下而上的特征增强路径，限制了浅层特征保留能力，故在坏点检测上略逊于 YOLOv4。

本文所提 TBNet 在各项指标上均表现优异。在坏点检测方面，其 AP20 达到 97.64%，较次优的 YOLOv4 提升 16.56 个百分点，显著领先其他算法，证明双分支结构能有效缓解小目标特征丢失问题；在污点检测方面，AP50 为 93.26%，优于 YOLOv3 约 3.29 个百分点，这得益于 K-means++ 初始化锚框提升定位准确性、Focal Loss 缓解样本不平衡，以及学习率调度策略对模型收敛的进一步优化。模型权重仅 46.5MB，具备良好轻量化特性。最终，TBNet 在 mAP 上取得 95.45% 的综合性能，显著优于所有对比模型。

4 结论

本文针对高分辨率工业图像中小目标缺陷检测的精度与效率平衡问题，提出了一种基于改进 YOLOv4 的高效解决方案。该方法在轻量化的 YOLOv4-MobileNetV2 框架基础上，通过集成双分支检测结构、引入余弦退火学习率调度机制以及采用 K-means++ 进行锚框优化，有效提升了模型对微小缺陷的检测能力，同时保持了模型的轻量化特性。

实验结果表明，优化后的模型在坏点检测（AP20 为 97.64%）、污点检测（AP50 为 93.26%）和综合检测精度（mAP 为 95.45%）方面均有显著提升，而模型权重仅为 46.5 MB，在检测精度与模型效率之间取得了良好平衡。本

研究为高分辨率图像下的小目标实时检测提供了一种有效的轻量化解决方案，后续工作将重点围绕生成式数据增强与精细化缺陷分析展开，以进一步提升系统性能。

参考文献：

[1] 陈烨强. 基于机器视觉的表面缺陷检测技术研究 [J]. 现代制造技术与装备, 2024,60(12):167-169.

[2] 付山, 谢晖, 易建业. 基于机器视觉的金属表面缺陷检测 [J]. 电镀与精饰, 2024, 46(12):145-146.

[3] 黄克腾, 王玉奇, 王清, 等. 基于深度学习的表面缺陷检测算法文献研究 [J]. 青岛大学学报 (工程技术版), 2025, 40(2): 106-113.

[4] Mao Fuqi, LI Jing, Yang Jian, et al. A grain boundary defects detection algorithm with improved localization accuracy based on efficientdet[J]. Automatic Control and Computer Sciences, 2023, 57:81-92.

[5] 姜菲菲, 李宁, 邱翠翠, 等. 基于改进 Faster RCNN 的金属丝网缺陷检测方法 [J]. 中国科技论文, 2024,19(2):153-159.

[6] Chen Zhihong, Huang Xuhong, Kang Ronghao, et al. Aluminum product surface defect detection method based on improved centernet[J]. IEEE Transactions on Electrical and Electronic Engineering, 2024,20(3):415-421.

[7] 麻越, 李月林, 张琳, 等. 基于改进 YOLOv3 模型的热轧板卷表面缺陷检测技术研究 [J]. 冶金自动化, 2024, 48(3): 83-91.

[8] 王立琦, 张汝波. 基于轻量化 YOLOv4 的 FSAC 目标检测方法研究 [J]. 计算机仿真, 2025,42(2):350-355.

[9] 曾维国, 刘晓群, 郝娟. 多注意力机制结合的绝缘子缺陷检测模型 [J]. 现代信息科技, 2025,9(9):37-42.

[10] 刘帅, 嵇小辅. 基于改进 ResNet34 卷积神经网络的硅棒缺陷检测 [J]. 计算机测量与控制, 2025,33(2):37-43.

[11] 赖肖强, 胡泓. 基于机器视觉算法和改进 MobileNetV2 的晶圆缺陷检测研究 [J]. 机械与电子, 2025,43(1):34-39.

【作者简介】

余铭琪（1997—），男，广东英德人，硕士，研究方向：人工智能。

岳洪伟（1979—），通信作者（email:springfall@foxmail.com），男，安徽阜阳人，博士，副教授，研究方向：人工智能。

（收稿日期：2025-11-07 修回日期：2026-05-13）