

DOI:10.19651/j.cnki.emt.2211120

融合坐标注意力和自适应特征的 YOLOv5 陶瓷膜缺陷检测方法*

雷震霆 朱兴龙 孙进 马昊天 梁立 游志刚

(扬州大学机械工程学院 扬州 225127)

摘要: 针对平板陶瓷膜表面缺陷实时检测时存在检测准确率较低的问题,本文提出了一种融合坐标注意力和自适应特征的 YOLOv5 陶瓷膜缺陷检测方法。通过在原有 YOLOv5 模型的主干网络中加入坐标注意力机制,建立位置信息和通道之间的关系,从而更准确地获取感兴趣区域。在原始网络的预测网络中融入自适应特征融合机制,提高模型对多尺度缺陷的检测能力。将空洞空间卷积池化金字塔模块替换原始网络中的空间金字塔池化模块,提高卷积核视野获取更多的有用信息。实验结果表明:本文模型平均精度为 97.8%,检测帧数为 32 FPS,平均精度与原始 YOLOv5 模型相比提高了 5.5%。本文提出的模型在满足平板陶瓷膜缺陷的实时检测条件下,提高了模型的检测准确率,对推动平板陶瓷膜缺陷检测的发展具有一定的参考价值。

关键词: YOLOv5s;平板陶瓷膜;目标检测;坐标注意力;自适应特征融合

中图分类号: TP242.6 **文献标识码:** A **国家标准学科分类代码:** 510.1050

YOLOv5 ceramic film defect detection method incorporating coordinate attention and adaptive features

Lei Zhenting Zhu Xinglong Sun Jin Ma Haotian Liang Li You Zhigang

(School of Mechanical Engineering, Yangzhou University, Yangzhou 225127, China)

Abstract: To address the problem of low detection accuracy in real-time detection of defects on the surface of flat ceramic films, this paper proposes a YOLOv5 ceramic film defect detection method that incorporates coordinate attention and adaptive features. By adding a coordinate attention mechanism to the backbone network of the original YOLOv5 model, the relationship between location information and channels is established to obtain the region of interest more accurately. The adaptive feature fusion mechanism is incorporated into the prediction network of the original network to improve the detection capability of the model for multi-scale defects. Replace the spatial pyramid pooling module in the original network with the spatial pyramid pooling module of the null space convolution pooling module to improve the convolutional kernel field of view to obtain more useful information. The experimental results show that the average accuracy of this model is 97.8%, the number of detection frames is 32 FPS, and the average accuracy is improved by 5.5% compared with the original YOLOv5 model. The model proposed in this paper improves the detection accuracy of the model under the condition of satisfying the real-time detection of flat ceramic film defects, which has certain reference value for promoting the development of flat ceramic film defect detection.

Keywords: YOLOv5s; flat ceramic film; object detection; coordinate attention; adaptive feature fusion

0 引言

平板陶瓷膜作为一种新型无机膜,因其可回收利用^[1]、使用寿命长^[2]和过滤效果好^[3],被广泛的应用在海水淡

化^[4],污水处理等领域^[5]。平板陶瓷膜在实际生产过程中会发生表面损伤等情况,如果生产阶段不能及时发现,在实际使用中会造成严重的后果。

目前,针对表面缺陷检测方法主要分为两大类。一类

收稿日期:2022-08-19

* 基金项目:国家自然科学基金(51775484,51475409)、2022年扬州市科技计划项目(YZ2022184)、2021年扬州市产业前瞻与共性关键技术项目(YZ2021020)、2020年江苏省产学研合作项日(BY2020663)、2020年扬州大学市校企合作专项(YZ2020166)资助

是两阶段的检测方法;Zhang 等^[6]针对复杂纹理干扰瓷砖表面缺陷检测问题,将采集的图像进行分割和校正后使用显著性检测方法对表面缺陷进行初检,再对包含缺陷区域进行二次检测,使用缺陷定位方法实现对瓷砖表面缺陷的检测。赵楚等^[7]针对瓷砖表面较小缺陷的检测问题,提出了一种改进的 Faster-RCNN 检测算法,在主干提取网络后面加入可变形卷积,使用卷积神经网络学习小特征。同时,改进锚点的相应参数,使其定位更加准确。最后实现了对瓷砖小目标的检测。两阶段的检测方法虽然具有检测准确率高,但是实时检测性较差,不能应用在平板陶瓷膜缺陷的实时检测。

另一类为一阶段的检测方法,Huang 等^[8]针对氧化锆陶瓷手机背板的缺陷,提出了一种基于数字图像处理的缺陷检测系统,通过对获取陶瓷手机背板图像样本进行提取和预处理,放入 YOLO3 中进行训练和分类,实现了对陶瓷手机背板表面缺陷的检测。孙乐杨等^[9]针对遥感影像中的输电塔的检测问题,提出了一种基于改进 YOLOv5 的检测方法,通过增大其大尺度检测层,提高了对图像小目标的识别率。Yang^[10]针对苹果的实时检测问题,提出了一种改进 YOLOv5s 模型,通过对瓶颈网络 CSP 模块进行改进,实现了对遮挡的苹果的实时检测。刘素行等^[11]、徐晓光等^[12]通过 YOLO 算法的改进,实现相应的检测要求。虽然上述对 YOLO 网络进行改进,可以实现缺陷的实时检测,但是针对平板陶瓷膜的缺陷检测,仍存在检测准确率较低的问题。

本文为了实现平板陶瓷膜实时缺陷检测过程中具有较高的准确率,采用工业相机获取平板陶瓷膜的缺陷图像,制作成平板陶瓷膜检测的数据集。通过在原有 YOLOv5 模型的主干网络中加入坐标注意力机制、融入自适应特征融合机制、将空洞空间卷积池化金字塔模块替换原始网络中的空间金字塔池化模块来提高模型对平板陶瓷膜缺陷的检测能力,进而增强模型的泛化能力和鲁棒性。

1 YOLOv5 原理

1.1 YOLOv5 网络

YOLOv5 模型主要包括主干提取网络(Backbone)、瓶颈网络(Neck)和预测网络(Prediction)等 3 部分网络结构。其中,Backbone:在不同图像细粒度上聚合形成图像特征的卷积神经网络;Neck:一系列混合和组合图像特征的网络层,负责将图像特征传递到预测层;Prediction:对特征图进行预测合并分类处理。YOLOv5 网络主要有 4 个版本:YOLOv5s、YOLOv5m、YOLOv5l 和 YOLOv5x。其中 YOLOv5s 模型最小,检测速率最快,但是检测准确率较低^[13]。本文的网络模型是基于 YOLOv5s 中的 v5.0 版本上进行改进的。

1.2 Focus 结构

Focus 层是通过对输入的图像进行切片处理,可以得到 4 张互补的图像,如图 1 所示。Focus 将输入的 RGB 图

像变成了 12 通道,提高了特征的提取能力。通过对切片结束的图像进行张量拼接,然后再进行卷积处理即可得到没有信息丢失的二倍下采样特征图。

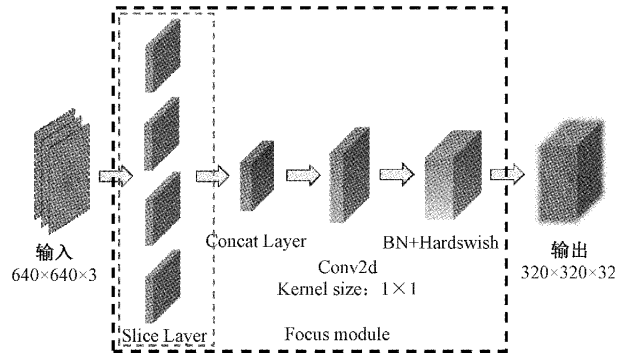


图 1 Focus 模块示意图

1.3 空间金字塔池化结构

空间金字塔池化(spatial pyramid pooling,SPP),能将任意大小的特征图转换成固定大小的特征向量,如图 2 所示。首先,将特征图进行卷积操作;然后,完成特征图网格的划分并进行最大池化;最后,将池化后的特征图进行张量拼接后再进行卷积,输出特征图。

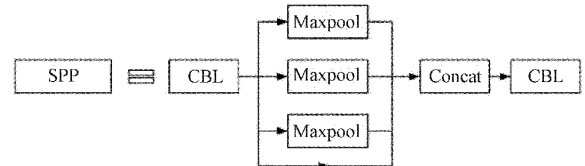


图 2 SPP 模块示意图

2 改进的 YOLOv5 模型

2.1 空洞空间卷积池化金字塔结构

空洞空间卷积池化金字塔(atrous spatial pyramid pooling,ASPP),如图 3 所示。其中,空洞卷积与普通卷积不同之处在于空洞卷积在相邻卷积之间进行插值处理,在保证不降低图像大小情况下,增加了卷积核的感受野^[14]。同时避免了传统方法因增加滤波器大小和计算时间来增加感受野导致检测速度降低的问题。

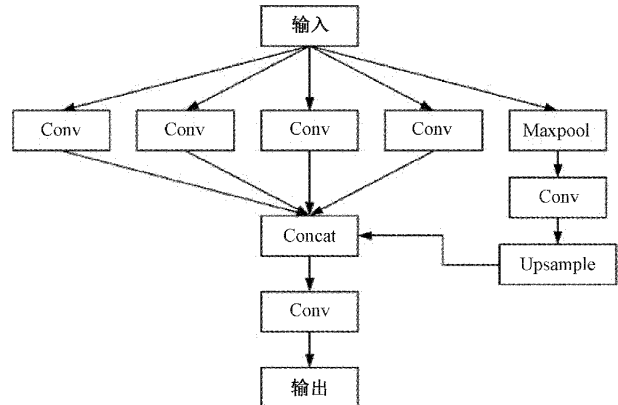


图 3 ASPP 模块示意图

通过卷积代替SPP结构中的最大池化,并分别进行1个 1×1 、3个 3×3 (rate分别为6、12、18)的卷积操作和1个最大池化。在最大池化后再进行卷积操作,通过上采样来扩充张量大小。最后将所有张量进行拼接后再进行卷积,输出特征图,增加了网络整体的特征提取能力。

2.2 坐标注意力机制

注意力机制源于人类视觉观察机制,指人们关注多个事物时会集中关注较为重要的事物。神经网络通过增加注意力机制主要使得网络更加倾向于关注特征图中有用的信息,忽略无用信息,进而提高模型特征提取关键信息的能力^[13]。

坐标注意力(coordinate attention, CA)^[16]如图4所示,主要分为2步:1)完成坐标信息的嵌入;2)生成坐标注意力机制。下面对这两个方面进行阐述。

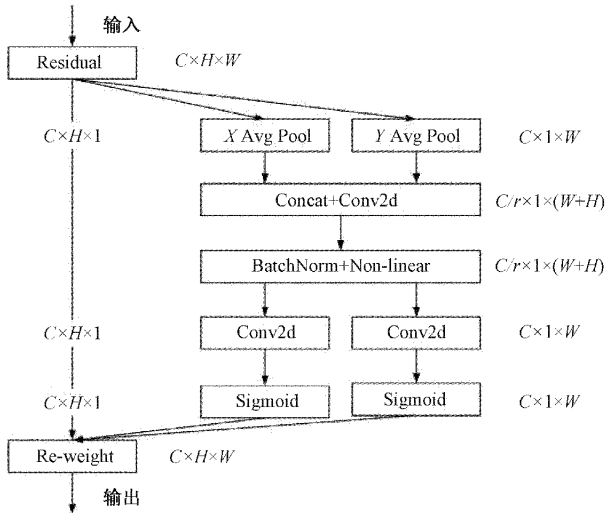


图4 坐标注意力机制结构

1)当给定一个特征图进行全局池化操作,分解成沿着特征图水平方向的 $(H, 1)$ 和沿着特征图垂直方向 $(1, W)$,在两个方向上分别进行平均池化,获取特征图 X, Y 方向的坐标信息。在水平方向获取特征的计算公式如式(1)所示。

$$z_c^h(h) = \frac{1}{W} \sum_{i=0}^{W-1} x_c(h, i) \quad (1)$$

在垂直方向获取特征的计算公式如式(2)所示。

$$z_c^w(w) = \frac{1}{H} \sum_{j=0}^{H-1} x_c(j, w) \quad (2)$$

2)根据式(1)和(2),模型可以准确地获得全局感受野并编码精确的位置信息,然后将 X 方向和 Y 方向平均池化后的特征图,进行张量的拼接,最后通过共享的 1×1 卷积进行 F_1 变换,具体如式(3)所示。

$$f = \zeta(F_1([\cdot^h, \cdot^w])) \quad (3)$$

其中, f 为水平方向的和垂直方向的进行编码的中间特征映射, $[\cdot, \cdot]$ 表示沿空间维度的串联操作, F_1 表示将水平和垂直的特征图进行张量的拼接, ζ 为非线性激活函数。

经过非线性和归一化处理,沿着空间维度将 f 切分

为两个独立的张量 $f^h \in R^{\frac{C}{r} \times H}$ 和 $f^w \in R^{\frac{C}{r} \times W}$,再利用两个 1×1 的卷积 F_h 和 F_w 将特征图 f^h 和 f^w 变换到和 X 输入相同的通道,具体计算公式如式(4)和(5)所示。

$$g^h = \text{sigmoid}(F_h(f^h)) \quad (4)$$

$$g^w = \text{sigmoid}(F_w(f^w)) \quad (5)$$

为了降低模型的复杂度和计算量,通常会适当的缩减通道数。最后采用矩阵相乘的方法来获取模型的注意力权重矩阵,计算公式如式在(6)所示。

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \quad (6)$$

2.3 自适应空间特征融合

本文在Neck网络后面增加自适应空间特征融合机制(adaptively spatial feature fusion, ASFF),如图5所示。

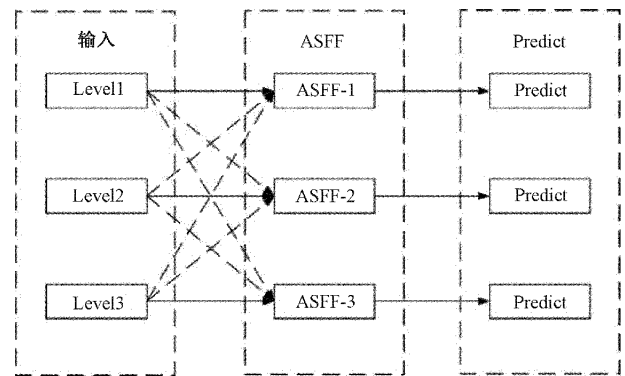


图5 ASFF网络结构示意图

将Neck中输出的Level1、Level2和Level3特征图进行加权融合,是为了对底层边缘轮廓、形状、颜色信息以及图像高层的语义信息进行充分融合^[17]。自适应空间特征融合机制的公式如式(7)所示。

$$W_{ij}^1 = Y_{ij}^{1-1} \alpha_{ij}^1 + Y_{ij}^{2-1} \beta_{ij}^1 + Y_{ij}^{3-1} \gamma_{ij}^1 \quad (7)$$

其中, $Y_{ij}^{1-1}, Y_{ij}^{2-1}, Y_{ij}^{3-1}$ 为不同特征图的输出, $\alpha_{ij}^1, \beta_{ij}^1, \gamma_{ij}^1$ 为来自不同层的权重。例如,将Level1、Level2与Level3融合,首先采用 1×1 大小的卷积核将Level1和Level2压缩到相同的通道数,然后通过上采样的方式将Level1和Level2调整到与Level3相同的维度,进行相应的融合。

3 实验结果与分析

3.1 平板陶瓷膜的数据集

从公司获取含有缺陷的平板陶瓷膜样品,通过工业相机采集企业实际生产过程中的缺陷图像和人为因素产生的缺陷图像,将上述图像全部用于模型的训练。将平板陶瓷膜缺陷图像中的80%作为训练集,20%作为测试集。其中数据数量以及类型详见表1所示,缺陷图像如图6所示。

3.2 训练图像数量对模型性能的影响

本文针对数据集的数量设置梯度为1 000、2 000、3 000、4 000、5 000、6 000;依次将上述梯度数量的平板陶瓷膜缺陷图像作为YOLOv5s网络模型的输入,可以获得不同数量集与模型性能的关系,如图7所示。

表 1 平板陶瓷膜缺陷集

样本类型	缺陷名称	数量	总数
平板陶瓷膜训练集	孔洞	1 635	6 783
	凹坑	1 721	
	划痕	1 668	
	落渣	1 759	

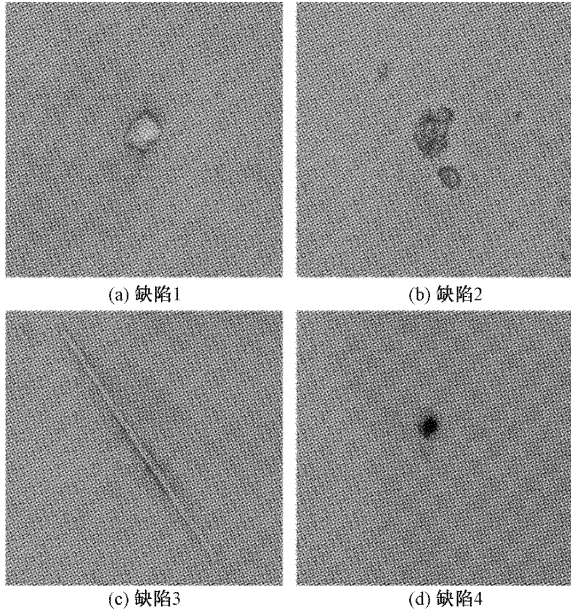


图 6 平板陶瓷膜缺陷图像

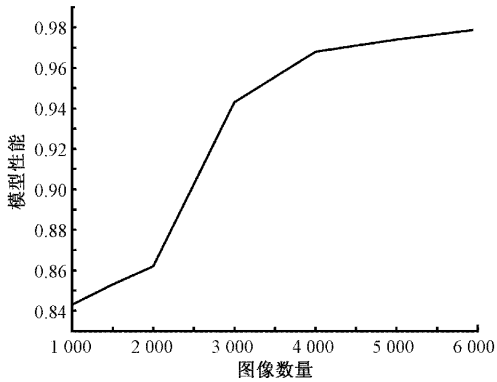


图 7 不同数量缺陷图像对 YOLOv5s 模型性能的影响

YOLOv5s 网络模型整体性能随着训练集数量的增加而提升。训练集数量在 2 000~3 000 之间时模型性能增加最快;训练集数量在 3 000~4 000 之间时模型性能增加较快;训练集数量在 4 000~6 000 之间时模型性能增加缓慢,对 YOLOv5 模型性能的影响较小。因此,本文用于训练的 5 426 幅(总样本数据集的 80%)平板陶瓷膜缺陷图像,可以满足 YOLOv5s 模型对数据集的需求并且可以得到一个较好的检测训练权重。

3.3 消融实验

消融对比实验是为了验证各个改进模块的优化作用,

消融实验结果如表 2 所示。其中,改进模型 1 表示在主干网络中加入坐标注意力机制模块,改进模型 2 表示将 ASPP 模块替换原始网络的 SPP 模块,改进模型 3 表示融合自适应特征融合机制。加入坐标注意力机制模型的识别准确率提高了 2.4%,速度降低了 2 ms;加入 ASPP 模块模型的准确率增加了 1.6%,速度降低了 3 ms;加入自适应特征融合机制模型的准确率提高 1.1%,速度降低了 1 ms。将这 3 个改进到一个模型中,模型的准确率提高了 5.5%,速度降低了 5 ms。改进的模型虽然降低了检测的速率,但是却较大地提高了模型的准确率。

表 2 消融实验结果

对比模型	加入 CA	加入 ASPP	加入 ASFF	mAP@0.5/ %	时间/ s
YOLOv5s	×	×	×	92.3	0.026
改进模型 1	✓	×	×	94.7	0.028
改进模型 2	×	✓	×	93.9	0.029
改进模型 3	×	×	✓	93.4	0.027
本文	✓	✓	✓	97.8	0.031

3.4 不同模型之间的性能的比较

为了更好地评估本文 YOLOv5 模型的性能,分别在相同数据集上对 SSD、Faster-RCNN、YOLOv3、YOLOv4、YOLOv5 和本文的模型进行相关预测实验,结果如表 3 所示。本文模型在检测帧数相比于 Faster-RCNN 高出 29 PFS,mAP@0.5 数值上相比于 SSD 高出 12.7%,本文在保证平板陶瓷膜缺陷实时检测的条件下,提高了模型的检测精度。

表 3 各类算法的比较

序号	模型	mAP@0.5/ %	Precision/ %	Recall/ %	时间/ s	帧数/ FPS
1	SSD	85.1	86.2	84.7	0.125	8
2	Faster-RCNN	94.7	95.8	92.6	0.333	3
3	YOLOv3	90.6	91.3	88.7	0.045	22
4	YOLOv4	91.5	92.1	89.1	0.029	35
5	YOLOv5	92.3	94.4	91.8	0.026	38
6	YOLOX	91.7	92.9	90.8	0.029	34
7	本文	97.8	98.2	97.4	0.031	32

3.5 不同模型的检测效果图

本文采集图像训练集时采用分时采集方法,获取不同光照时间的缺陷数据集,使得训练集更加贴近实际生产条件。本文随机选取不同缺陷图像测试不同模型检测性能,检测效果如图 8 所示。

由不同模型的检测效果图可得:采用 SSD 和 YOLOv3 模型对平板陶瓷膜缺陷图像进行检测时,会出现漏检和置

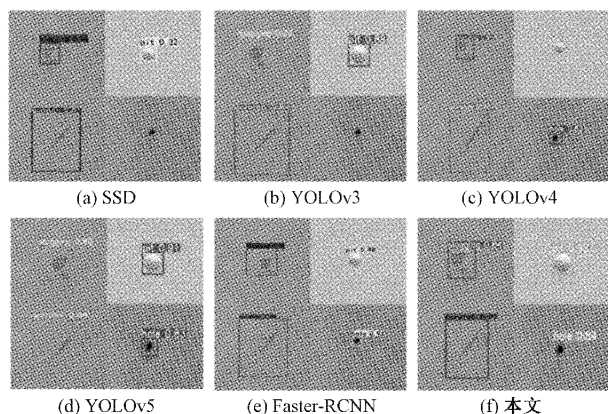


图8 不同模型的检测效果图

信度偏低的现象,说明模型检测效果较差。采用 Faster-RCNN 模型对平板陶瓷膜缺陷图像进行检测时,虽然检测置信度最高,但是预测框与实际大小的缺陷相差较大,不利于检测后期的精准定位。本文提出的模型预测框与实际大小的缺陷相近,没有出现过大框选缺陷的现象,缺陷检测的置信度较高,说明模型的检测效果较好。

4 结 论

针对平板陶瓷膜表面缺陷实时检测时存在检测准确率较低的问题,本文提出了一种融合坐标注意力和自适应特征的 YOLOv5 陶瓷膜缺陷检测方法。通过在原有 YOLOv5 模型的主干网络中加入坐标注意力机制、融入自适应特征融合机制、使用 ASPP 模块替换原始网络中的 SPP 模块等改进,来提高模型对平板陶瓷膜缺陷的检测能力,进而增强模型的泛化能力和鲁棒性。实验结果表明:改进的 YOLOv5 模型比原有的 YOLOv5 模型,在小目标和多尺度目标检测效果上优于原始 YOLOv5 网络,并且还能获得更高的检测精度。该模型具有较好的实时检测性能和较好的泛化能力,不仅可以实时检测平板陶瓷膜的表面缺陷,还可以适应于其他类型目标的检测。

参考文献

- [1] 黄文靓,付宛宜,张锡辉. 氧化/陶瓷膜过滤组合工艺处理石化废水研究[J]. 中国给水排水, 2022, 38(8): 65-72.
- [2] HE Z, LYU Z, GU Q, et al. Ceramic-based membranes for water and wastewater treatment[J]. Colloids and Surfaces A: Physicochemical and Engineering Aspects, 2019, 578:123513-123513.
- [3] 赵成根,李根利,张程,等. 陶瓷平板膜 MBR 技术在印钞废水中的中试应用[J]. 资源信息与工程, 2022, 37(2):125-128.
- [4] ASIF M B, ZHANG Z H. Ceramic membrane technology for water and wastewater treatment: A critical review of performance, full-scale applications,

membrane fouling and prospects [J]. Chemical Engineering Journal, 2021(418): 129481-129499.

- [5] 徐振峰,陈文松,黎钱波,等. 陶瓷膜工艺在浙江省农村饮用水处理中的应用[J]. 净水技术, 2022, 41(3): 39-45.
- [6] ZHANG H, PENG L, YU S, et al. Detection of surface defects in ceramic tiles with complex texture[J]. IEEE Access, 2021, 9: 92788-92797.
- [7] 赵楚,段先华,苏俊楷. 改进 Faster RCNN 的瓷砖表面瑕疵检测研究[J]. 计算机工程与应用, 2022, DOI: 10.3778/j.issn.1002-8331.2203-0414.
- [8] HUANG W Z, ZHANG C, WU X, et al. The detection of defects in ceramic cell phone backplane with embedded system[J]. Measurement, 2021, 181: 109598-109605.
- [9] 孙乐杨,凌振宝,王永志. 基于改进的 YOLOv5 识别遥感影像中输电塔的方法[J]. 实验技术与管理, 2022, 39(4):19-24.
- [10] YANG F. A real-time apple targets detection method for picking robot based on improved YOLOv5 [J]. Remote Sensing, 2021, 13(9):1619-1642.
- [11] 刘素行,吴媛,张军军. 基于 YOLOv3 的交通场景目标检测方法[J]. 国外电子测量技术, 2021, 40(2): 116-120.
- [12] 徐晓光,李海. 多尺度特征在 YOLO 算法中的应用研究[J]. 电子测量与仪器学报, 2021, 35(6): 96-101.
- [13] 郎松,曹选,张艳微,等. 融合改进 YOLOv5 算法的图像全站仪全自动测量方法[J]. 仪器仪表学报, 2022, 43(5):120-127.
- [14] 孙进,张道周,张洋,等. 基于双通道空洞卷积神经网络的手势识别[J]. 传感器与微系统, 2022, 41(3): 126-128.
- [15] 邓佳桐,程志江,叶浩勛. 改进 YOLOv3 的多模态融合行人检测算法[J]. 中国测试, 2022, 48(5): 108-115.
- [16] HOU Q B, ZHOU D Q, FENG J S. Coordinate attention for efficient mobile network design [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13713-13722.
- [17] 罗禹杰,张剑,陈亮,等. 基于自适应空间特征融合的轻量化目标检测算法[J]. 激光与光电子学进展, 2022, 59(4): 310-320.

作者简介

雷震霆,硕士研究生,主要研究方向为机器视觉智能检测。

E-mail: leizhenting2021@163.com

朱兴龙(通信作者),教授,博士,主要研究方向为机器视觉。

E-mail: xlzhu@yzu.edu.cn