

基于改进 YOLOv5 的手机外观缺陷检测算法

潘金晶¹, 曾成¹, 张晶¹, 耿雪娜²

[1. 长园视觉科技(珠海)有限公司, 广东 珠海 519085; 2. 长春理工大学计算机科学技术学院, 长春 130013]

摘要: 提出一种手机外观缺陷检测的改进算法 YOLOv5-CBE。该算法在 YOLOv5 框架的基础上, 在主干网络的 C3 模块中加入坐标注意力 (coordinate attention, CA) 机制, 可同时考虑通道间的关系和位置信息, 使模型更准确地定位并识别到目标区域。借鉴加权双向特征金字塔网络 (Bidirectional feature pyramid network, BiFPN) 的思想, 将 Neck 部分的 concat 模块替换为多尺度特征融合结构, 使不同分辨率的特征更有效地融合。使用 Focal-EIoU 替代原模型中的边界框回归损失函数 CIoU, 使回归过程更专注于高质量的预测框, 提高了定位精度。在工业相机成像的手机外观缺陷数据集上进行测试, 结果表明, 与 YOLOv5 模型相比, 基于 Focal-EIoU 的 YOLOv5 模型召回率 (recall) 和平均精度均值 (mAP50) 分别提升了 4.7%、1.9%; 改进算法的精确率 (precision)、召回率 (recall)、平均精度均值 (mean average precision, mAP50) 均有明显提升, 分别提升了 1.2%、5.6%、5.3%。

关键词: 缺陷检测; 坐标注意力; 多尺度特征融合; Focal-EIoU; YOLOv5

中图分类号: TP391 文献标志码: A 文章编号: 1674-8646(2024)08-0039-05

Mobile Phone Appearance Defect Detection Based on Improved YOLOv5

Pan Jinjing¹, Zeng Cheng¹, Zhang Jing¹, Geng Xuena²

(1. CYG Vision Technology (Zhuhai) Co., Ltd, Zhuhai 519085, China; 2. School of Computer Science and Technology, Changchun University of Science and Technology, Changchun 130013, China)

Abstract: An improved mobile phone appearance defect detection algorithm-YOLOv5-CBE is proposed. Based on the YOLOv5 framework, the algorithm adds coordinate attention (CA) mechanism in C3 module of backbone network, which can consider the relationship between channels and location information at the same time, so that the model can locate and identify the target area more accurately. According to the idea of Bidirectional feature pyramid network (BiFPN), the concat module of Neck part is replaced with multi-scale feature fusion structure, so that features of different resolutions can be fused more effectively. Focal-EIoU replaces the bounding box regression loss function CIoU in the original model, and the regression process focuses more on high-quality samples and improves positioning accuracy. The test is carried out on the data set of mobile phone appearance defects taken by industrial cameras. The results show that compared with the YOLOv5 model, the recall rate and average accuracy mean (mAP50) of YOLOv5 model based on Focal-EIoU increases by 4.7% and 1.9% respectively. The precision, recall and mean average precision (mAP50) of the improved algorithm all be improved significantly by 1.2%, 5.6% and 5.3% respectively.

Key words: Defect detection; Coordinate attention; Multi-scale feature fusion; Focal-EIoU; YOLOv5

0 引言

有效的手机外观缺陷检测手段具有重要的应用及经济价值^[1]。目前, 人工检测仍是主要的检测手段, 但由于评价标准不一、检测效率低下等无法满足企业

快速的生产需求, 因此基于机器视觉的缺陷检测方法逐步被广泛应用。

基于机器视觉的检测技术主要分为传统方法和基于深度学习的检测方法。传统方法是依据缺陷颜色、形状等特征, 利用图像处理方法或结合传统机器学习方法进行检测^[2]。但由于手机表面缺陷种类繁多, 形态多样, 传统方法无法适应不同的生产环境。基于深度学习的检测方法通过卷积运算自动提取图像特征, 克服了传统机器视觉检测方法的缺点, 代表性算法有 Faster R-CNN^[3]、SSD^[4] 和 YOLO 系列^[5-9] 等, 其中 YOLOv5 模型以其准确性、快速性已应用于不同场景下的智能缺陷检测^[10-11], 但在不同检测任务中被证明

收稿日期: 2023-12-29

基金项目: 国家自然科学基金青年科学基金(批准号: 62206028)

作者简介: 潘金晶(1991-), 女, 博士。研究方向: 数据挖掘、推荐系统、计算机视觉研究。

通讯作者: 曾成(1984-), 男, 博士。研究方向: 软件算法及通信协议优化, 深度学习算法在工业检测领域的应用与优化研究。E-mail: hop.zeng@cygvision.com。

检测精度仍需进一步提高。

本研究提出一种适用于手机外观缺陷检测的 YOLO 算法,该算法以 YOLOv5-6.0 模型为基础,从缺陷特征提取、多尺度特征融合、样本平衡等角度设计模块,用于提升模型的泛化能力,在工业相机成像的数据集上证明所提出的算法相较于 YOLOv5 模型改进明显,可以较好地应用于手机外观缺陷检测任务中。

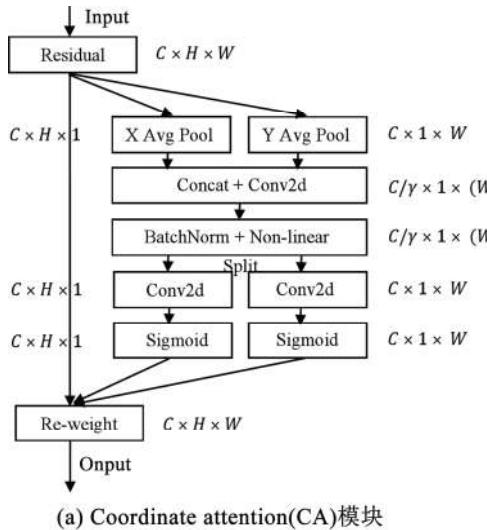
1 改进后的 YOLO 模型

YOLOv5 是单阶段缺陷检测算法的最优模型之一,包括输入端、骨干网络、颈部和头部 4 部分。为了满

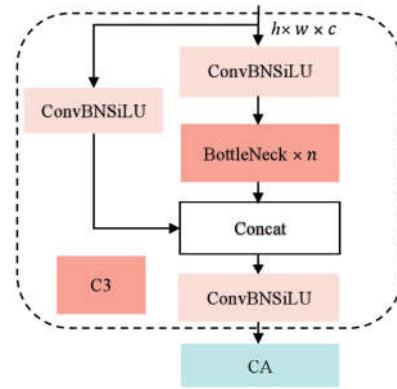
足工业场景下检测速度与精度的需要,选择 YOLOv5s-6.0 为骨干模型。针对 YOLO 模型的改进主要分为骨干网络、颈部和损失函数等方面。

1.1 对骨干网络的改进

注意力机制是计算机视觉中的一个重要概念,可以帮助模型更好地理解输入图像,生产准确特征,从而提高模型性能。坐标注意力^[12](coordinate attention, CA)可以捕获跨通道信息,获取方向感知和位置敏感信息,有助于模型更准确地定位和识别感兴趣的对象^[13]。CA 模块灵活且轻量级,可轻松插入 YOLOv5 网络,增强网络的特征抽取能力。



(a) Coordinate attention(CA)模块



(b) C3CA模块

图 1 CA 模块及 C3CA 模块

Fig. 1 CA module and C3CA module

CA 模块如图 1 (a) 所示。采用两个池化层对输入的每个通道沿着水平与垂直方向进行编码,第 c 个通道在宽度 w 和高度 h 处的输出可表示为:

$$Z_c^w(w) = \frac{1}{H} \sum_{0 \leq i \leq H} x_c(j, w) \quad (1)$$

$$Z_c^h(h) = \frac{1}{W} \sum_{0 \leq i \leq W} x_c(h, i) \quad (2)$$

将得到的特征进行拼接,通过 1×1 卷积 (F_1) 得到:

$$f = \delta(F_1([z^h, z^w])) \quad (3)$$

其中, δ 为非线性激活函数, $f \in R^{C/\gamma \times (H+W)}$ 为在水平和垂直方向上编码空间信息的中间特征图, γ 是控制模块大小的压缩比参数。将 f 沿着空间维度拆分成两个独立张量 $f^h \in R^{C/\gamma \times H}$, $f^w \in R^{C/\gamma \times W}$, 采用两个额外的 1×1 卷积 F_h 和 F_w 将其变换为:

$$g^h = \sigma[F_h(f^h)] \quad (4)$$

$$g^w = \sigma[F_w(f^w)] \quad (5)$$

其中, σ 是 s 型函数。对 g^h 和 g^w 进行扩展并作为注意力权重,CA 模块的输出为:

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

在 YOLOv5 骨干网络中 4 个 C3 模块后均添加 CA 模块 (C3CA 模块),如图 1 (b) 所示,进一步提升 YOLOv5 提取特征的能力。改进后的主干网络如图 2 所示。

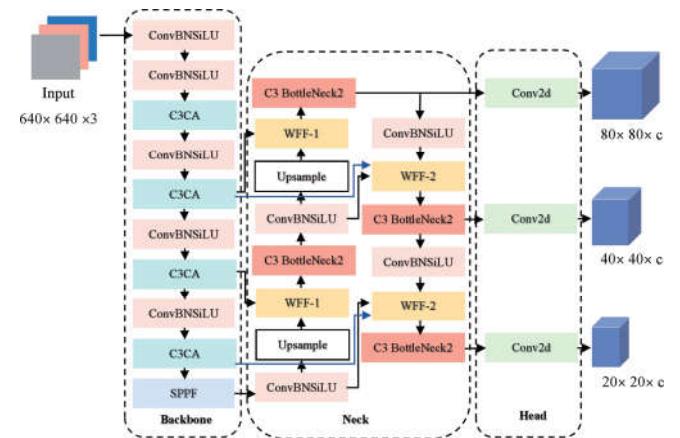


图 2 YOLOv5-CBE 网络结构

Fig. 2 YOLOv5-CBE network structure

1.2 对颈部网络的改进

在原始 YOLOv5 模型中,颈部网络通常采取拼接 (concat) 模块来融合不同特征。但在 YOLOv5 网络中产生的特征通常来源于不同尺度,简单的拼接(即平

等的对待不同特征的贡献)是不合理的。借鉴 BiFPN^[14]中多尺度特征融合方案,将颈部网络的 concat 模块替换为权重特征融合(weighted feature fusion, WFF)模块。该模块在拼接特征时,对于输入的每一个特征引入对应权重,使网络学习各个输入特征的重要性。

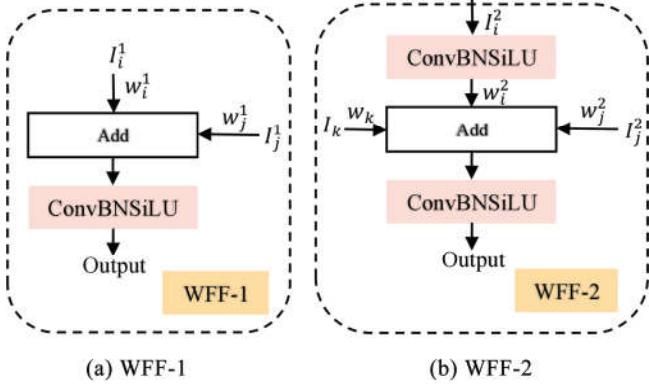


图 3 BiFPN 及 WFF 模块

Fig. 3 BiFPN and WFF module

WFF 模块如图 3 所示,包括 WFF-1 和 WFF-2 模块。WFF-1 将 C3CA 模块及颈部网络的输出特征进行加权融合,融合后的特征通过 1×1 卷积操作将通道数加倍,输出为:

$$Output_{WFF-1} = ConvBNSiLU[Add(w_i^1 \times I_i^1, w_j^1 \times I_j^1)] \quad (7)$$

WFF-2 先通过 1×1 的卷积操作将 C3CA 输出特征图的通道数压缩至原先的一半,再将其与颈部输出特征图加权相加,再通过 1×1 的卷积操作加倍通道数进行输出。

$$Output_{WFF-2} = ConvBNSiLU[Add(w_i^2 \times ConvBNSiLU(I_i^2), w_j^2 \times I_j^2, w_k^2 \times I_k^2)] \quad (8)$$

其中, $w_i^1, w_j^1, w_i^2, w_j^2, w_k^2$ 是可学习的权重因子, 权重因子越大对于特征融合的贡献越大。

1.3 对损失函数的改进

缺陷检测的性能依赖于损失函数的设计。边界框损失函数作为目标检测损失函数的重要组成部分,其良好的定义将为目标检测模型带来显著的性能提升。早期的目标检测工作使用 IoU 作为定位损失。但当目标框与预测框没有交集时, IoU 损失会出现梯度消失现象,导致无法继续优化。这激发了几种改进的基于 IoU 的损失设计,包括 GIoU^[15]、DIoU^[16]、CIoU^[16]。

YOLOv5 模型默认的边界框损失函数为 CIoU, CIoU 损失考虑了 3 个重要的几何因素,即重叠面积、中心点距离和宽高比。但 CIoU 的最后一项 ν 仍没有被很好地定义,将从以下几个方面减慢 CIoU 的精度与收敛速度。^① ν 只反映了宽高比的差异,而不是预测框与目

标框宽高之间的关系。^②CIoU 中, w 和 h 相对于 ν 的梯度,具有相反的符号,即 w 和 h 不能同时增大或减小,这显然也不够合理。^③由于 ν 只反映了宽高比的差异,在训练过程中有可能会增加 CIoU 的损失,从而增加了长宽比的一致性,阻碍了模型有效地减少预测框和目标框之间宽高的真实差异。

为了解决上述问题, EIoU^[17] 损失如式(9)所示,是在 CIoU 损失的基础上得到的,它不仅考虑了中心点距离和宽高比,还考虑了预测框与目标框宽度和高度的真实差异。EIoU 损失函数包含重叠损失、中心距离损失、宽高损失,前两部分损失与 CIoU 相同,但宽高损失直接最小化目标框和预测框的宽高差异,使得回归结果更快、定位结果更精准。

$$Loss_{EIoU} = 1 - (IoU - \frac{\rho^2(b, b^{gt})}{c^2} - \frac{\rho^2(w, w^{gt})}{c_w^2} -$$

$$\frac{\rho^2(h, h^{gt})}{c_h^2}) \quad (9)$$

其中, $\rho^2(\cdot)$ 为欧式距离, b, b^{gt} 分别代表预测框和目标框的中心点, c, c_w 和 c_h 分别为包含两框的最小闭包区域的对角线距离、宽度及高度, w^{gt}, h^{gt} 和 w, h 分别表示目标框和预测框的宽度和高度。

考虑到边界框回归过程中存在训练样本不平衡的问题,即在一张图像中,回归误差小的高质量预测框数量远少于误差大的低质量预测框数量。质量较差的预测框会产生过大的有害梯度,影响训练过程。Focal-EIoU^[17] 损失在 EIoU 损失的基础上从梯度角度出发,用 IoU 值对 EIoU 损失进行加权,平衡高质量样本和低质量样本对损失的贡献,提升高质量样本(IoU 大)的贡献,抑制低质量样本(IoU 小)的贡献,实现更加高效精准的目标检测。Focal-EIoU 公式如下:

$$Loss_{Focal-EIoU} = IoU^\gamma Loss_{EIoU}, \gamma = 0.5 \quad (10)$$

实验将 Focal-EIoU 作为改进 YOLOv5 算法的定位损失函数,将其与 GIoU、DIoU、CIoU 及 EIoU 进行对比,分析不同的定位损失对 YOLOv5 性能的影响。

2 实验分析

2.1 数据集、实验参数及实验环境

实验中使用的数据集来自于工业相机成像,包含张 1126 张含标签图片。样品的缺陷主要包括碰伤、压伤、擦伤,分别对应产品在生产过程中被坚硬物体磕碰受伤、被重物压伤、被尖锐物划伤。按照 8:1:1 的比例将数据集划分为训练集、验证集和测试集。由于原始图像尺寸过大,为 $4096 \times 700 \times 3$,将所有输入图像尺寸统一裁剪为 $960 \times 640 \times 3$ 。模型初始学习率为

0.01, 权重衰减系数为 0.0005, 动量为 0.937, batch size 为 16, epoch 为 450, 参数优化采用随机梯度下降 (stochastic gradient decent, SGD) 优化器, 获取最优的模型权重。模型参数 width_multiple 和 depth_multiple 分别设置为 0.5、0.33, 即在 YOLOv5s 模型的基础上进行改进。基于 Pytorch 深度学习框架, CPU 为 Intel Core i7 6600U, GPU 为 NVIDIA GeForce RTX 3090, 用于加速运算, 开发环境为 Python 3.7.9、Pytorch1.10.2、Cuda 11.6。

2.2 评估指标

采用精确率 (Precision, P)、召回率 (recall, R) 及目标检测领域常用的 mAP50 和 mAP50-95 来评估 YOLOv5 模型及其改进模型的性能。计算公式如下:

$$P = \frac{TP}{TP + FP} \quad (11)$$

$$R = \frac{TP}{TP + FN} \quad (12)$$

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (13)$$

精确率表示模型预测正确物体的能力, 该指标表示在所有预测为缺陷的结果 (TP + FP) 中模型预测为缺陷且正确的目标 (TP) 的占比。召回率表示模型能够找全正确物体的能力, 表示在所有的缺陷样本 (TP +

FN) 中模型预测为缺陷且正确的目标 (TP) 的占比。 AP_i 代表第 i 类的模型平均精度, 如果是在多类别的情况下, 则使用 mAP (mean Average Precision) 来衡量模型性能, 计算方法为将多个 AP 值累加后求平均。mAP50 是指在 IOU 阈值取 0.5 时的 mAP 值, 将 IOU 阈值从 0.5 开始以 0.05 为步长逐渐增加到 0.95, 获得不同阈值下的 mAP 值, 求取它们的平均值即可获得 mAP50-95。

2.3 实验结果分析

表 1 展示了 YOLOv5 模型在不同边界框回归损失函数下在测试集上的性能表现。由表 1 可知, YOLOv5 在 CIoU 上具有最好的精确率, 在 Focal-EIoU 上具有最好的召回率。对于不同模型的 mAP50, Focal-EIoU 表现最好, 其次是 EIoU、CIoU 和 DIoU, GIoU 表现最差。EIoU 在 CIoU 的基础上将宽高损失直接定义为最小化目标框与真实框的宽度和高度之差, 提高了回归精度, mAP50 和召回率较 CIoU 均有所提升。Focal-EIoU 在 EIoU 的基础上平衡高质量样本和低质量样本对损失的贡献, 在实验中获得了最好的 mAP50 和召回率, 较 YOLOv5 原始模型 (YOLOv5 + CIoU) 分别提升了 1.9%、4.7%。因此选用 Focal-EIoU 作为改进 YOLOv5 模型的边界框回归损失函数来提高模型的泛化能力。

表 1 不同边界框回归损失下 YOLOv5 模型在测试集上的实验结果对比

Tab. 1 Comparison of experimental results of YOLOv5 model on test set under different bounding box regression loss

models	AP			mAP50	P	R
	碰伤	压伤	擦伤			
YOLOv5 + GIoU	0.600	0.704	0.814	0.706	0.651	0.695
YOLOv5 + DIoU	0.599	0.703	0.824	0.709	0.753	0.669
YOLOv5 + CIoU	0.633	0.683	0.820	0.712	0.771	0.662
YOLOv5 + EIoU	0.662	0.691	0.811	0.721	0.668	0.706
YOLOv5 + Focal - EIoU	0.659	0.688	0.846	0.731	0.716	0.709

表 2 不同改进在测试集上带来的模型指标变化

Tab. 2 Model metrics change on the test set under different improvements

models	P	R	mAP50	mAP50-95
YOLOv5	0.771	0.662	0.712	0.500
YOLOv5 + C3CA	0.665	0.716	0.725	0.491
YOLOv5 + C3CA + BiFPN-WFF	0.722	0.724	0.737	0.498
YOLOv5-CBE	0.783	0.718	0.765	0.513

改进后的模型 YOLOv5-CBE 在主干网络中添加 C3CA 模块, 将 Neck 部分的 concat 模块替换为 BiFPN-WFF 模块, 将边界框损失函数由 CIoU 替换为 Focal-EIoU, 并应用于整个网络模型。将原模型及改进模型在测试集上进行测试, 观察不同改进对模型性能变化的影响。从表 2 可以看出, 在主干网络的 C3 模块中加

入 CA 后, 相较于 YOLOv5 基础模型, 召回率提升了 5.4%, mAP50 提升了 1.3%, 但精确率和 mAP50-95 有所下降。将 BiFPN-WFF 模块替换 Neck 部分的 concat 模块, 该模型相较于 YOLOv5 模型在召回率上提升了 6.2%, mAP50 提升了 2.5%, mAP50-95 与基础模型相当, 精确率仍低于基础模型。将 CIoU 替换为 Focal-EIoU, YOLOv5-CBE 相较于原模型, 在精确率、召回率、mAP50 及 mAP50-95 上均有明显提升, 分别提升了 1.2%、5.6%、5.3%、0.13%。

图 4 展示了改进前后模型在手机外观缺陷数据集上的检测结果, 对于真实手机外观缺陷数据集, 在第一组对比图中, 由于图像左下角有一微小亮点与缺陷类似, YOLOv5 将其误检为缺陷, YOLOv5-CBE 并未出现

误检。在第二组图像中,由于其中一处碰伤缺陷不明显,YOLOv5 出现漏检现象,YOLOv5-CBE 可以检出,且 YOLOv5 将其中一处碰伤缺陷误检为压伤。在第三四组对比图中,YOLOv5 均将压伤缺陷误检为碰伤缺陷,YOLOv5-CBE 均以较高的置信度正确检出,体现了更好的泛化性能。

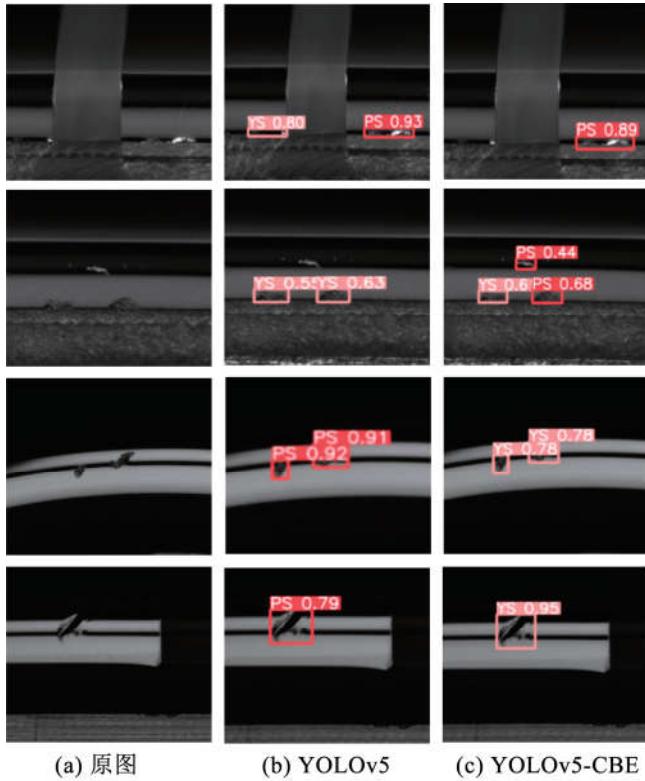


图 4 改进前后模型在手机外观缺陷数据集上的检测结果

Fig. 4 Detection results of the model before and after the improvement on the mobile phone appearance defect dataset

3 结论

基于 YOLOv5 提出了一种改进的手机外观缺陷检测算法 YOLOv5-CBE,在主干网络的 C3 模块中添加坐标注意力机制(CA),增强了模型的特征表达能力,提高了检测精度。借鉴 BiFPN 思想,赋予不同层次不同分辨率的特征可学习的权重因子,使特征融合更加充分有效,减少特征表达能力缺乏而导致的漏检与误检,进而提高缺陷检测精度。针对 ClIoU 的不足,引入 Focal-EIoU 边界框回归损失函数,进一步提高了缺陷的定位精度。实验结果表明,YOLOv5-CBE 算法相较于 YOLOv5 模型在精确率、召回率和平均精度值方面均得到了提升,具有优秀的检测精度,能较好地应用于手机外观缺陷检测任务中。

参考文献:

- [1] Wang T,Zhang C,Ding R,et al. Mobile phone surface defect detection based on improvedfaster R-CNN [C]//2020 25th International Conference on Pattern Recognition (ICPR),2021.
- [2] 罗东亮,蔡雨萱,杨子豪,等.工业缺陷检测深度学习方法综述[J].中国科学:信息科学,2022,52(06):1002–1039.
- [3] Ren SQ,He KM,Girshick R,et al. Faster R-CNN:towards real-time object detection with region proposal networks[J]. Advances in Neural Information Processing Systems,2015,28:91–99.
- [4] Liu W,Anguelov D,Erhan D,et al. SSD:single shot multibox detector [C]//European Conference on Computer Vision. Berlin: Springer,2016,21–37.
- [5] Redmon J,Divvala S,Girshick R,et al. You only look once:unified, real-timeobject detection [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Piscataway, NJ: IEEE,2016:779 –788.
- [6] Redmon J,Farhadi A. YOLO9000:better, faster, stronger [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE,2017:7263 –7271.
- [7] Redmon J,Farhadi A. Yolov3:an incremental improvement [EB/OL]. (2018-04-08)[2021-02-20]. <https://arxiv.org/abs/1804.02767>.
- [8] Bochkovskiy A,Wang CY,Liao HYM. YOLOv4: optimal speed and accuracy ofobject detection [EB/OL]. (2020-04-23)[2022-01-20]. <https://arxiv.org/abs/2009.10934>.
- [9] Jocher G. Yolov5 [EB/OL]. [2022-02-20]. <https://github.com/ultralytics/yolov5>.
- [10] 刘浩翰,孙铖,贺怀清,等. 金属表面缺陷检测方法 YOLOv3I[J]. 吉林大学学报(理学版),2023,61(03):612 –622.
- [11] 左昊,牛晓伟. 基于改进 YOLOv5 的路面缺陷检测算法研究[J]. 信息技术与信息化,2023(01):50 –53.
- [12] Hou QB,Zhou DQ,Feng JS. Coordinate attention for efficient mobile network design[C]//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, NJ: IEEE,2021:13713 –13722.
- [13] 孙伟,常鹏帅,戴亮,等. 基于注意力引导数据增强的车型识别[J]. 计算机工程,2022,48(07):300 –306.
- [14] Tan MX,Pang RM,LE QV. Efficient DET:scalable and efficient object detection [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Seattle, NJ: IEEE,2020:10778 –10787.
- [15] Rezatofighi H,Tsoi N,Gwak JY,et al. Generalized Intersection Over Union: A Metric and a Loss for Bounding Box Regression[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE,2019.
- [16] Zheng ZH,Wang P,Liu W,et al. Distance-IoU loss: faster and better learning for bounding box regression [C]// Proceedings of the 34th AAAI Conference on Artificial Intelligence. PaloAlto: AAAI Press,2020,34(07):12993 –13000.
- [17] Zhang YF,Ren W,Zhang Z,et al. Focal and efficient IOU loss for accurate bounding box regression [J]. Neurocomputing,2022, 506 (09):146 –157.