

DOI: 10.19926/j.cnki.issn.1674-232X.2023.02.111

文献引用:石兰娟,张梦斯,刘文浩,等. G-YOLO:基于改进 YOLOv5 的嵌入式小目标缺陷检测算法[J]. 杭州师范大学学报(自然科学版), 2024,23(2):201-208.

SHI Lanjuan, ZHANG Mengsi, LIU Wenhao, et al. G-YOLO: an embedded small target defect detection algorithm based on improved YOLOv5[J]. Journal of Hangzhou Normal University(Natural Science Edition), 2024,23(2):201-208.

G-YOLO:基于改进 YOLOv5 的嵌入式小目标缺陷检测算法

石兰娟,张梦斯,刘文浩,周迪斌

(杭州师范大学信息科学与技术学院,浙江 杭州 311121)

摘 要:针对人工检测缺陷模式或 YOLOv5 等深度学习算法对工业产品的缺陷检测存在识别准确率低、模型参数规模大等问题,提出一种对微小缺陷端到端检测的嵌入式算法 G-YOLO. 该算法使用卷积核为 3 和卷积核为 1 的双层卷积 FConv 模块,改善了原单层卷积带来的参数量较大的问题;改进的轻量级跨阶段 GSP 模块融合坐标注意力机制用于主干网络中能够利用冗余信息实现廉价的线性操作和聚焦缺陷信息来增强特征,以提高网络对缺陷特征的提取能力;去除原 YOLOv5 的颈部模块,减少网络的参数量和提升网络检测速度. 结果表明, G-YOLO 嵌入式算法减少了模型大小,改善了缺陷检测的效果,较好地满足轻量化嵌入式模型的要求.

关键词: 嵌入式;特征融合;注意力机制;冗余信息;特征提取

中图分类号: TP399

文献标志码: A

文章编号: 1674-232X(2024)02-0201-08

工业缺陷检测旨在发现各种工业制品的外观瑕疵,是保障产品质量、维持生产稳定的重要技术. 以往的缺陷检测需要人工筛查,成本高、效率低,难以覆盖大规模的质检需求. 随着工业成像、计算机视觉和深度学习等领域新技术的出现,基于视觉的工业缺陷检测技术得到了长足发展. 工业缺陷检测不仅可以用于检测各种工业制品,如金属、纺织物、半导体等,而且具有优秀的检测精度和效率,还能提供简便、安全的操作环境. 因此,工业缺陷检测已成为智能制造领域重要的基础研究方向.

对图像中缺陷部位检测方法的研究中,Chen 等^[1]将 Dense Net^[2]与 YOLOv3^[3]结合来检测 LED 缺陷;He 等^[4]通过融合多尺度特征提升检测精度;Tao 等^[5]将 Faster R-CNN^[6]级联,实现电力巡检中绝缘子的缺陷检测,第一阶段从自然场景检测绝缘子以排除其他背景干扰,第二阶段从裁剪的绝缘子区域检测缺陷,整个网络采用端到端训练方式;Zhang 等^[7]采用 YOLOv3 检测桥梁缺陷,并使用迁移学习、批再规范化与 Focal Loss 提升检测性能;Duan 等^[8]引入双密度卷积层结构,增加模型预测尺度,改进 YOLOv3,得到更好的铸件缺陷检测网络模型;Guo 等^[9]在 YOLOv5 中引入了 TRANS 模块,提升对钢铁表面缺陷的检测性能;Jin 等^[10]改进了 YOLOv5 对织物缺陷图像特征的检测,改善对织物缺陷的检测.

上述方法在检测精度上有一定的进步,但在实际部署时,需要考虑到计算时间和存储资源的开销. 同

收稿日期:2023-02-11

修回日期:2023-02-27

通信作者:周迪斌(1978—),男,讲师,博士,主要从事数字图像处理方面的研究. E-mail:dibinz@zju.edu.cn

时工业零件表面缺陷检测和识别算法需要考虑模型在移动设备上的部署和满足生产线的实时性要求,目前的移动设备和嵌入式设备的计算能力太低,现有的深度卷积模型不适合直接用于嵌入式设备.因此,本文重点改进 YOLOv5 的体系结构设计,提出一种轻量级的神经网络模型(G-YOLO)来解决当前的问题.

1 网络结构

1.1 YOLOv5 网络结构

YOLOv5 模型的网络结构根据处理阶段分为 4 个部分,包括输入、主干、颈部和预测^[11].在主干网络部分,为了避免细节丢失,通过执行 Focus 模块将输入图像的空间信息传递到通道维度. YOLOv5 的原始骨干部分是 CSPDarknet53,通过跨阶段部分(CSP)模块用于特征提取,CSP 模块由 1 个瓶颈结构和 3 个卷积组成. CSP 模块有 2 种,其中一种由 1 个残差单元(CSP1_1)或 3 个残差单元(CSP1_3)组成,用于主干网络中;另一种是颈部中的 CSP(CSP2_1),用 CBS 模块替换残差单元.此外,空间金字塔池(SPP)模块提供不同的感受野来丰富特征的表达能力.在颈部,采用了特征金字塔网络(FPN)和路径聚合网络(PAN)的结构,FPN 模块自顶向下,通过上采样将高层特征图与低层特征图融合,实现语义特征增强;同时,PAN 模块是自下而上的,在 FPN 基础上,PAN 将定位信息从浅层传输到深层以获得增强的空间特征.输出处理后的特征作为预测层的输入,最后对输入特征进行 head 预测,分类器得到类别对象并生成边界框的最终坐标.

1.2 G-YOLO 网络结构

为了解决 YOLOv5 模型的数量和检测速度不适用于嵌入式设备,以及检测工业表面缺陷数据集的小目标缺陷准确度不高的问题,本文提出的 G-YOLO 算法做了相关改进. G-YOLO 网络主要由输入、主干和预测 3 部分组成.在本文轻量级网络的设计中,将 GSP 模块注入骨干网络,用来减轻网络和特征提取;在下采样处理中,将第 4 层和第 6 层的 Conv 替换为 FConv 结构,其目的是降低维数,减少网络中学习的参数数量;去掉颈部模块,防止信息丢失,尽可能压缩网络参数,减少网络计算量,提高模型推理速度;最后将该特征送入检测分支进行目标类别和目标边框位置的预测. G-YOLO 算法的结构示意图如图 1 所示.

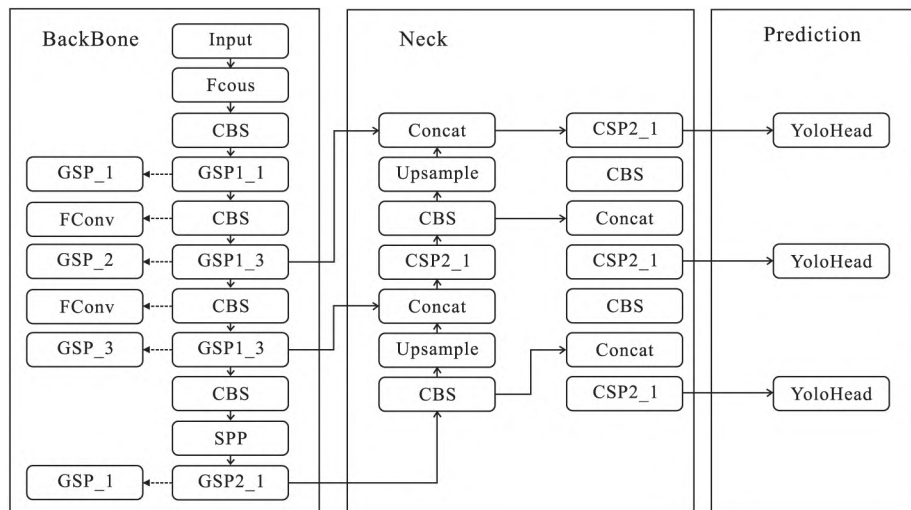


图 1 G-YOLO 算法结构示意图

Fig. 1 Schematic diagram of G-YOLO algorithm structure

1.2.1 双层卷积 FConv

双层卷积 FConv 模块主要由 2 个卷积组成.为了获得轻量级主干网络,该模块采用 CBS 单元(卷积核为 3 的 Conv 层、BN 层、SiLU 激活层和卷积核为 1 的 Conv 层、BN 层、SiLU 激活层),使用 1×1 卷积控制卷积核数量实现升维,以最少的参数拓宽网络通道,进行跨通道信息交互,解决原始网络的卷积带来的较

大参数量,提高网络的检测速度和满足实时检测的高要求.该模块的详细结构如图 2 所示.

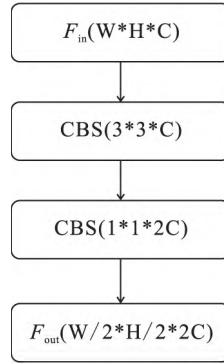


图 2 FConv 结构示意图

Fig. 2 Schematic diagram of FConv structure

1.2.2 GSP

原始 CSP 结构用于主干网络中,通过不同的卷积通道提取物体的不同特征,会导致太多的训练资源被用于无缺陷区域、网络的训练效率低下及参数量过大.为了更好地提取特征,将 GSP 模块作为主干部分,该模块主要由 1 个瓶颈结构、1 个注意力机制和 3 个卷积组成.当特征映射 F_{in} 输入时,传输路径被分成 2 个平行分支,其中 F_{in} 的通道减少一半,以生成 2 个新的特征映射;然后,将 2 个平行分支的特征图连接为一个完整的特征图(输出特征地图 F_{out}).GSP 模块的详细结构如图 3 所示.

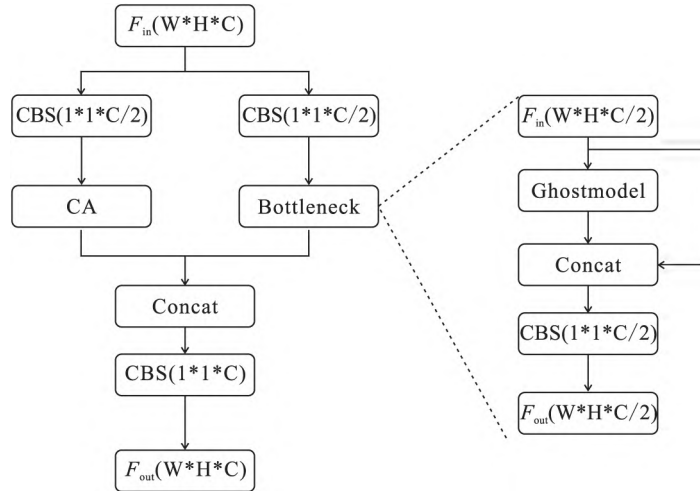


图 3 GSP 结构示意图

Fig. 3 Schematic diagram of GSP structure

为了更全面地提取特征,普通卷积采用大量卷积生成特征图.但较大的卷积核数量与通道数会产生冗余信息,导致计算量较大.相比之下,在 Bottleneck 模块中使用 Ghost 卷积,在不增加网络参数的情况下减小模型的总体大小,使网络更好更快地理解冗余信息,提高模型的准确性.假设卷积核数量为 n ,输入特征图的大小为 $h \times w \times c$,输出特征图大小为 $h' \times w' \times n$,卷积核大小为 $k \times k$,则普通卷积所需的计算量可表示为 $n \times h \times w \times c \times k \times k$.通道数 c 为 256 时,卷积计算量将超过 10^5 ,Ghost 卷积所需的计算量可表示为 $\frac{n}{s} \times h' \times w' \times c \times k \times k + (s-1) \times \frac{n}{s} \times h' \times w' \times d \times d$,其中 $d \times d$ 为线性运算核的大小,与 $k \times k$ 类似, s 表示线性运算的数量,且 s 远小于 c .

比较二者的计算量,Ghost 模块的理论加速比为

$$r = \frac{n \times h' \times w' \times c \times k \times k}{\frac{n}{s} \times h \times w \times c \times k \times k + (s-1) \times \frac{n}{s} \times h' \times w' \times d \times d} =$$

$$\frac{c \times k \times k}{\frac{1}{s} \times c \times k \times k + \frac{s-1}{s} \times d \times d} \approx \frac{s \times c}{s + c - 1} \approx s.$$

从上式可知,普通卷积的计算量约为 Ghost 卷积的 s 倍,同样参数量的计算量也约为 s 倍,这充分展示了 Ghost 卷积在计算量方面的优势.因此,在模型中使用 Ghost 卷积可以有效地降低计算成本,压缩模型参数量.

本文在 GSP 模型中引入坐标注意力机制(coordinate attention,CA)^[12],坐标注意力机制能够将横向和纵向的位置信息编码到通道注意力中,使得网络能够关注大范围的位置信息又不会带来过多的计算量,并且使模型的定位和目标识别更加准确.注意力机制通过关注输入对象的重要区域来获取更多关键信息.空间注意力机制(bottleneck attention module,BAM)^[13]、基于卷积块的注意力机制(convlutional block attention,CBAM)^[14]、通道注意力机制(squeeze-and-excitation networks,SE-Net)^[15]和坐标注意力机制等主流注意力机制已经被验证可以提高检测模型的性能.实验也证明,坐标注意力机制可以通过增加复杂度和添加少量参数进一步提升性能.

本文将 GSP 模块添加到 YOLOv5 目标检测模型的骨干特征提取中,通过集成坐标注意力机制,为不同的卷积通道分配不同的权重,以突出工业零件的缺陷特征.同时,通过适当的跨通道交互,模型的复杂性明显降低,避免了降维对学习通道的影响,添加少量的参数实现了客观的性能提升.

2 实验方法

2.1 实验平台

使用 GPU 型号为 NVIDIA GeForce RTX3060、CPU 型号为 E5-2650 v3 的 Windows 10 操作系统进行实验的训练部分,采用基于 Python3.8 语言的 PyTorch1.7.0 作为测试框架.

2.2 实验参数

采用适应性矩估计(Adam)算法训练网络,网络输入大小为 640×640 ,批量大小为 4.训练期间,对网络进行了 300 轮的训练,学习率设置为 0.0005,动量设置为 0.937,其他未提及的超参数与 YOLOv5 保持相同.

2.3 实验数据

本文使用的数据集 NEU-DET 是东北大学发布的一个表面缺陷数据库,收集了 6 种典型的热轧钢带表面缺陷,包括碎裂、包体、补丁、坑状表面、滚入式刻度线和划痕.该数据库包括 6 种不同类型典型表面缺陷的 1800 张灰度图像,每种类型的缺陷包含 300 个样本,每个样本图像包含多个缺陷.印刷电路板(printed circuit board,PCB)数据集由北京大学人机交互开放实验室发布,本文使用其中的 445 张图像,包含 6 种缺陷(missing hole, mouse bite, open circuit, short, spur, spurious copper).纺织物图像数据集原样本来自文献[16],随机抽样了 2352 张图片,根据缺陷类型将样本数据集分为 3 种类型,分别是破损(broken)、污渍(stain)和油渍(oil stain)缺陷.在整个实验过程中均采用 8:2 的比例划分训练集和验证集.

3 结果和分析

本文设计了 2 组实验,第 1 组在 PCB、热轧钢带和纺织物数据集上进行消融实验,分析本文不同改进部分在 YOLOv5 架构下对网络性能的影响;第 2 组将本文模型与主流目标检测模型进行性能比较,综合比较分析 G-YOLO 与主流检测网络 YOLOv4^[17]、CenterNet^[18]、YOLOv5 和 Faster R-CNN 的性能.

3.1 消融实验

采用消融实验验证不同改进对原始 YOLOv5 算法性能的影响,评估 3 种数据集去除颈部部分、添加 FConv 和添加 GSP 模块的效果,结果如表 1 所示.使用初始数据集进行 YOLOv5 网络测试的均值平均精度(mAP)分别为 85.8%、71.9%和 86.5%.去除颈部对网络压缩后,mAP 有所降低.网络主干使用 GSP

结构后提高了网络提取重要特征的能力,模型尺寸减小,mAP 提高. 通过上述 FConv 模块的通道融合和模型修剪,模型尺寸减小,mAP 提高,提高了检出率. 本文提出的算法(G-YOLO)最终 mAP 值分别为 87.0%、73.0%和 90.8%,有效提高了准确率.

表 1 不同改进方法对原始 YOLOv5 算法性能的影响

组别	纺织物数据集		热轧钢带数据集		PCB 数据集	
	均值平均精度/%	参数量/MB	均值平均精度/%	参数量/MB	均值平均精度/%	参数量/MB
未处理	85.8	7.2	71.9	7.2	86.5	7.2
去除颈部	76.6	4.2	64.5	4.2	54.5	4.2
去除颈部+添加 GSP	86.7	1.6	72.9	1.6	92.5	1.6
去除颈部+添加 GSP+ 添加 FConv	87.0	1.5	73.0	1.5	90.8	1.5

图 4 展示了 G-YOLO 网络对 PCB、热轧钢带和纺织物数据缺陷的检测结果. 结果显示,原始 YOLOv5 算法在 3 种数据集上没有检测到缺陷的位置,G-YOLO 通过算法改进未降低原始算法对缺陷的置信水平,并做出了正确的预测.

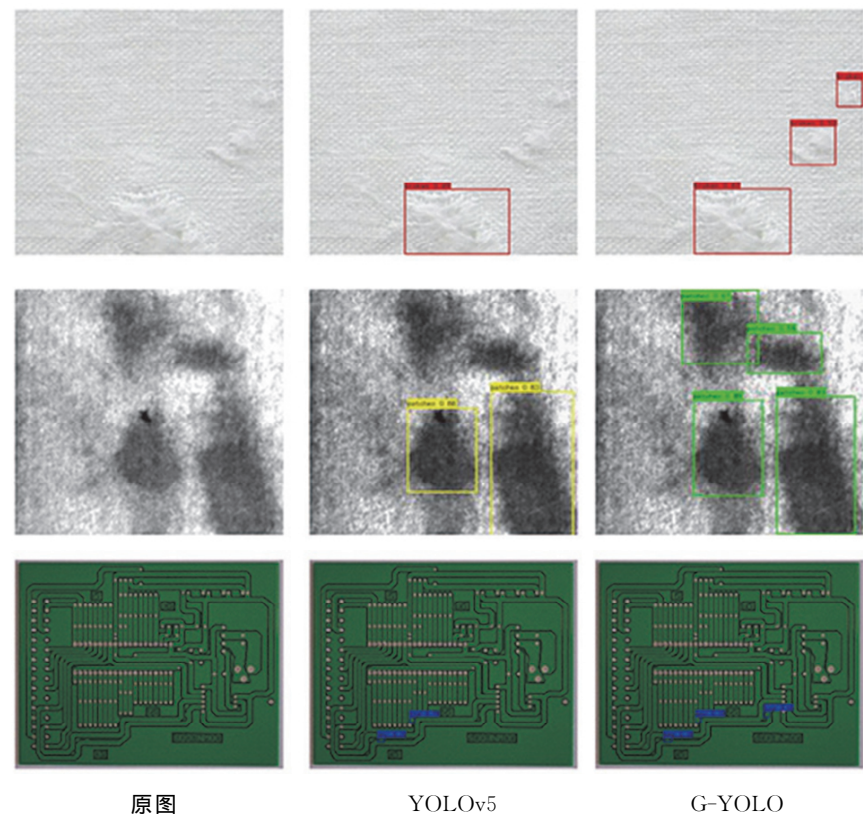


图 4 G-YOLO 与 YOLOv5 算法的检测结果图

Fig. 4 Detection results of G-YOLO and YOLOv5

本文提取了 YOLOv5 算法和 G-YOLO 算法对比主干网络训练后的特征图视觉,结果如图 5 所示. 对比 YOLOv5 算法,G-YOLO 算法在主干网络能够通过 GSP 模块聚合更多的语义特征,能够在一定程度上学会利用目标对象区域信息,更清晰地识别表面缺陷的具体位置,提高检测效果.

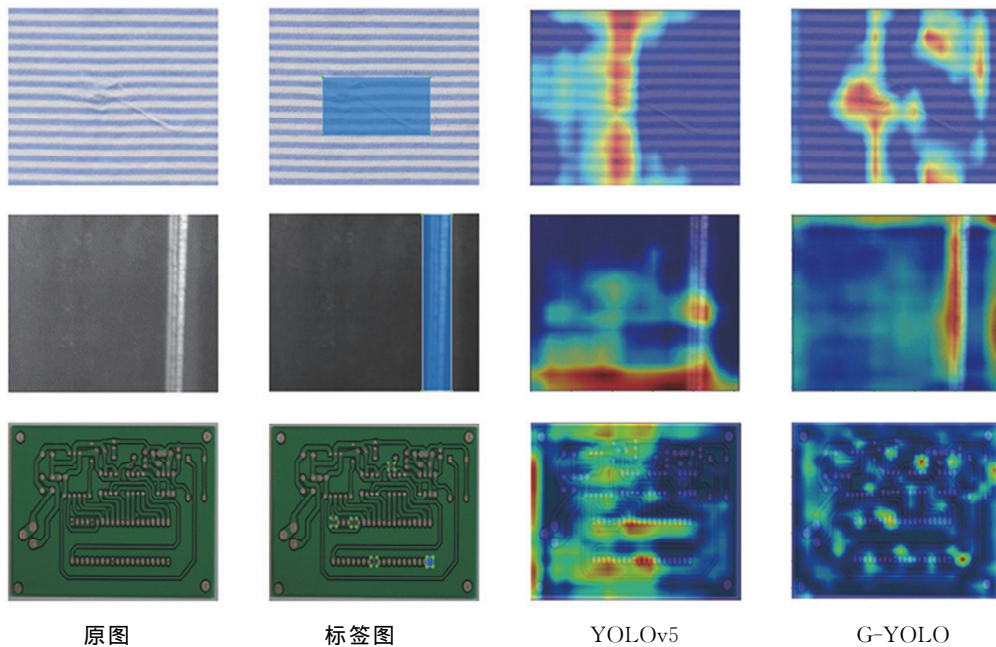


图 5 YOLOv5 和 G-YOLO 算法特征图

Fig. 5 Feature maps of YOLOv5 and G-YOLO

3.2 对比实验

将 G-YOLO 算法、Faster R-CNN、CenterNet、YOLOv4 和 YOLOv5 分别在 3 种数据集上进行训练和测试,结果如表 2 所示. Faster R-CNN、CenterNet、YOLOv4 和 YOLOv5 在模型大小和检测速度上并不理想. G-YOLO 算法的模型大小和在 CPU 上测试单张图片的每秒传输帧数(FPS)领先于其他网络模型,经过模型修剪后,减少了模型的训练时间,提高了对缺陷的检测率和准确率,证明了网络的有效性.

表 2 不同网络在 PCB、热轧钢带和纺织物数据集上的实验结果

Tab. 2 Results of different networks on PCB, hot-rolled steel strip and textile data sets

数据集	模型	均值平均精度/%	参数量/MB	每秒传输帧数/(帧/s)
纺织物	Faster R-CNN	83.0	28.3	0.04
	YOLOv4	83.2	63.9	0.13
	CenterNet	85.8	32.6	0.62
	YOLOv5	85.8	7.2	3.34
	G-YOLO	87.0	1.5	4.73
热轧钢带	Faster R-CNN	75.1	28.3	0.04
	YOLOv4	73.0	63.9	0.13
	CenterNet	71.3	32.6	0.62
	YOLOv5	71.9	7.2	3.34
	G-YOLO	73.0	1.5	4.73
PCB	Faster R-CNN	73.3	28.3	0.04
	YOLOv4	92.8	63.9	0.13
	CenterNet	89.3	32.6	0.62
	YOLOv5	86.5	7.2	3.34
	G-YOLO	90.8	1.5	4.73

比较 YOLOv5、YOLOv4、Faster R-CNN、CenterNet 和 G-YOLO 这 5 种算法对 PCB、热轧钢带和纺织物数据集上缺陷的检测效果,结果如图 6 所示. YOLOv5、YOLOv4、CenterNet 和 Faster R-CNN 算法对小目标的检测效果不佳,会出现漏检和误检的情况. G-YOLO 采用 Ghost 模块可以有效解决漏检和误

检的问题,将坐标注意力机制融合到GSP层,可以突出复杂背景下表面缺陷的表达性,提高小目标缺陷的检测,改善检测效果,目标缺陷标记准确,置信度高。

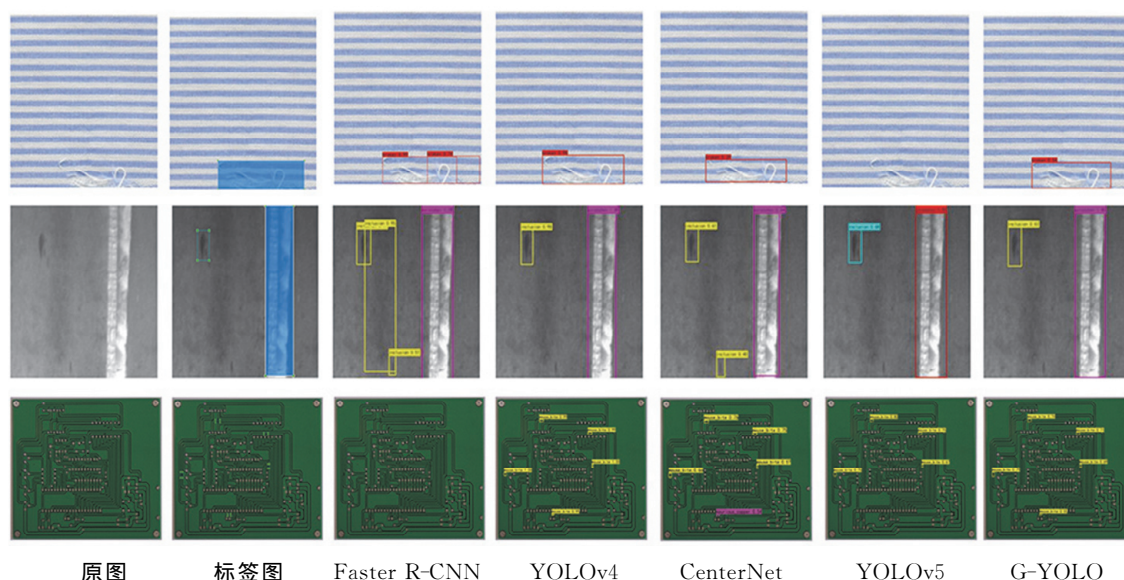


图6 不同网络在PCB、热轧钢带和纺织物数据集上的测试结果图

Fig. 6 Test maps of different networks on PCB, hot-rolled steel strip and textile data sets

4 结论

本文提出一个用于工业表面缺陷检测的轻量级算法G-YOLO.该算法采用端到端方法进行目标检测和识别,无需复杂的设计过程,直接选择整个特征图进行模型训练,可以更好地区分目标和背景区域,以解决传统工业表面缺陷检测精度低、模型推理时间长、对小缺陷的检测能力不足等问题.该模型通过修剪颈部,在不损失精度的情况下,减少模型的训练时间.网络主干主要通过GSP模块中引入Ghost卷积和在残差边集成坐标注意力机制,实现特征增强和突出复杂背景下表面缺陷的信息,有效提高网络的特征提取能力.最后,为了提高检测效率,采用双层卷积FConv结构,降低网络的参数量.PCB、热轧钢带和纺织物数据实验结果表明,与YOLOv5相比,G-YOLO算法减少了模型大小,提高了mAP,能够有效地检测出小的目标缺陷.本文仍需进一步研究与完善:

1)通过更多的图像处理与计算机视觉技术对缺陷漏检情况进行更全面的分析.本文仅对实验结果进行了对比展示与分析,后续需深入探究导致缺陷漏检的具体原因,调整实验方案、实验步骤及相关参数设置等.

2)通过先验知识对输入网络的预选框进行调整,提升缺陷识别正确率,以便提出更优的工业表面缺陷检测方案.

参考文献:

- [1] CHEN S H, TSAI C C. SMD LED chips defect detection using a YOLOv3-dense model[J]. Adv Eng Inform, 2021, 47: 101255.
- [2] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017: 2261-2269.
- [3] REDMON J, FARHADI A. YOLOv3: an incremental improvement[EB/OL]. (2018-08-08)[2023-01-15]. <http://arxiv.org/abs/1804.02767.pdf>.
- [4] HE Y, SONG K C, MENG Q G, et al. An end-to-end steel surface defect detection approach via fusing multiple hierarchical features[J]. IEEE Trans Instrum Meas, 2020, 69(4): 1493-1504.

- [5] TAO X, ZHANG D P, WANG Z H, et al. Detection of power line insulator defects using aerial images analyzed with convolutional neural networks[J]. *IEEE Trans Syst Man Cybern Syst*, 2020, 50(4):1486-1498.
- [6] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. *IEEE Trans Pattern Anal Mach Intell*, 2017, 39(6):1137-1149.
- [7] ZHANG C B, CHANG C, JAMSHIDI M. Concrete bridge surface damage detection using a single-stage detector[J]. *Comput Aided Civ Infrastruct Eng*, 2020, 35(4):389-409.
- [8] DUAN L M, YANG K, RUAN L. Research on automatic recognition of casting defects based on deep learning[J]. *IEEE Access*, 2020, 9:12209-12216.
- [9] GUO Z X, WANG C S, YANG G, et al. MSFT-YOLO: improved YOLOv5 based on transformer for detecting defects of steel surface [J]. *Sensors*, 2022, 22(9):3467.
- [10] JIN R, NIU Q. Automatic fabric defect detection based on an improved YOLOv5[J]. *Math Probl Eng*, 2021, 2021:7321394.
- [11] 柳进元, 张明锋. 基于 YOLOv5 算法的图像水深自动提取[J]. *福建师范大学学报(自然科学版)*, 2023, 39(1):86-92.
- [12] HOU Q B, ZHOU D Q, FENG J S. Coordinate attention for efficient mobile network design[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville: IEEE, 2021:13708-13717.
- [13] PARK J, WOO S, LEE J Y, et al. BAM: bottleneck attention module[EB/OL]. (2018-07-17)[2023-01-15]. <http://arxiv.org/abs/1807.06514>. pdf.
- [14] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//European Conference on Computer Vision. Cham: Springer, 2018:3-19.
- [15] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018:7132-7141.
- [16] ZHANG C K, FENG S Z, WANG X, et al. ZJU-leaper: a benchmark dataset for fabric defect detection and a comparative study[J]. *IEEE Trans Artif Intell*, 2020, 1(3):219-232.
- [17] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. (2020-04-23) [2023-01-15]. <http://arxiv.org/abs/2004.10934>. pdf.
- [18] ZHOU X Y, WANG D Q, KRÄHENBÜHL P. Objects as points[EB/OL]. (2019-04-16)[2023-01-15]. <http://arxiv.org/abs/1904.07850>. pdf.

G-YOLO: An Embedded Small Target Defect Detection Algorithm Based on Improved YOLOv5

SHI Lanjuan, ZHANG Mengsi, LIU Wenhao, ZHOU Dibin

(School of Information Science and Technology, Hangzhou Normal University, Hangzhou 311121, China)

Abstract: An embedded algorithm G-YOLO for end-to-end detection of small defect was proposed to solving problems of low recognition accuracy and large model parameter scale during industrial product defect detection using manual defect detection or deep learning algorithms such as YOLOv5. Firstly, G-YOLO used a double-layer convolution FConv module with a convolution kernel of three and a convolution kernel of one, which improved the problem of large parameter quantities caused by the original single-layer convolution. Secondly, the improved lightweight cross-stage GSP module fused with coordinate attention (CA) mechanism used in the backbone network could enhance features by utilizing redundant information for inexpensive linear operations and concentrating the defect information. Consequently, the network's ability to extract defect features was improved. Finally, the neck of the original YOLOv5 module was removed, the amount of network parameters was reduced and the speed of network detection was improved. The results showed that the embedded algorithm G-YOLO compresses the size of models and improved the effectiveness of defect detection, which better met the requirements of lightweight embedded models.

Key words: embedded; feature fusion; attention mechanism; redundant information; feature extraction