

DOI: 10.13382/j.jemi.B2104155

注意力机制优化 RetinaNet 的密集工件检测方法研究*

徐健¹ 陆珍¹ 刘秀平¹ 张立昌² 闫焕营³

(1. 西安工程大学电子信息学院 西安 710048; 2. 西安工程大学工程训练中心 西安 710048;

3. 深圳罗博泰尔机器人有限公司 深圳 518109)

摘要:针对密集工件存在相似度高、排列无序的特性导致检测难度大的问题,提出了一种注意力机制优化 RetinaNet 的密集工件检测方法。首先将注意力机制引入到 RetinaNet 主干特征提取网络以减少干扰物对检测效果的影响,提高神经网络的特征提取能力;然后利用 Soft-NMS 构建新的预测框提高重叠定位精度;最后通过迁移学习的方法训练数据集,提高模型训练效率。在密集工件数据集上验证该方法的有效性,实验结果表明,改进后的方法检测精度达到了 98.11%,相较于改进前提高了 2.59%,单张图片检测速度达到了 0.026 s,该方法能够满足实际工业生产过程中精准检测工件的目的,在保证速度的同时降低了漏检率和误检率。

关键词:目标检测;密集工件;注意力机制;RetinaNet

中图分类号: TP391.41 **文献标识码:** A **国家标准学科分类代码:** 520.2

Research on dense workpiece detection method based on attentional mechanism optimization RetinaNet

Xu Jian¹ Lu Zhen¹ Liu Xiuping¹ Zhang Lichang² Yan Huanying³

(1. School of Electronics and Information, Xi'an Polytechnic University, Xi'an 710048, China;

2. School of Engineering Training Centre, Xi'an Polytechnic University, Xi'an 710048, China;

3. Shenzhen Municipal Robotel Robot Technology Co., Ltd., Shenzhen 518109, China)

Abstract: In order to solve the problem of difficult detection due to the existence of dense workpiece with high similarity and disorderly arrangement, an attention mechanism is proposed to optimize RetinaNet's dense workpiece detection method. Firstly, the attention mechanism is introduced into the RetinaNet backbone feature extraction network to reduce the influence of interfering objects on the detection effect and improve the feature extraction ability of the neural network. then a new predictive box is constructed using Soft-NMS to improve the overlap localization accuracy. Finally, the dataset is trained by transfer learning method to improve the model training efficiency. The method effectiveness is verified on the produced dense workpiece dataset; Experimental results show that the detection accuracy of the improved method reaches 98.11%, which is 2.59% higher in comparison with that before the improvement. The detection speed of a single picture is up to 0.026s. The proposed method can meet the purpose of accurate detection of workpiece in actual industrial production process, which can reduce the rate of missed and false detection and assure the speed simultaneously.

Keywords: target detection; dense workpiece; attention mechanism; RetinaNet

0 引言

在“中国制造 2025”国家战略规划的大背景下^[1],智能制造成为当前研究热点。在制造行业中,人工分拣不

同类型工件导致费时费力的问题,使得机器视觉成为最佳解决方案。目标检测作为机器视觉领域的重要分支之一,在工业制造、社会安全、公共交通等领域应用广泛。但是大多数检测算法研究主要集中于稀疏目标,对于密集目标的解决方案并不多,因此密集目标检测有着重要

收稿日期: 2021-04-12 Received Date: 2021-04-12

* 基金项目: 西安市科技局高校人才服务企业项目(GXYD7.5)、陕西省科技厅工业领域一般项目(2018GY-173)资助

的研究意义和价值。

密集工件检测受观测角度、光照强度、外观形状等因素影响,检测难度大,无法满足工业实际需求。传统的目标检测包括区域候选,特征提取和分类回归等 3 个阶段,代表算法有梯度直方图 (histogram of oriented gradient, HOG)^[2]、尺度不变特征变换 (scale-invariant feature transform, SIFT)^[3]、DPM (deformable part model)^[4] 等,该类算法存在特征泛化性差和鲁棒性差等不足。

近年来,由于深度学习具有强大的表征能力和迁移能力,在目标检测方面表现突出^[5]。基于卷积神经网络 (CNN) 的目标检测算法分为单阶段检测算法 (one-stage) 和双阶段检测算法 (two-stage)。Two-stage 通过选择性搜索生成候选区域,对候选区域进行分类和位置回归,从而实现目标检测,代表算法有 R-CNN、Fast R-CNN、Faster R-CNN^[6] 等;One-stage 把特征提取、分类和位置回归的过程整合到同一个卷积神经网络中完成,代表算法有 YOLO^[7-8] (you only look once) 系列、SSD (single shot multiBox detector)、RetinaNet 等。Two-stage 网络结构复杂,检测精度高,但计算量大,检测速度慢;而 One-stage 与 Two-stage 相比检测精度稍有逊色,但检测速度有了大幅度提升。RetinaNet^[9] 是 One-stage 中检测精度较高的算法之一,解决了正负样本分布不均衡导致检测精度低的问题。引入特征金字塔 (feature pyramid networks, FPN)^[10] 结构,对多尺度特征进行融合,在不同尺度上进行目标检测。宋欢欢等^[11] 在 RetinaNet 网络基础之上,使用深度残差网络自主获取图像特征,融合 MobileNet 网络结构进行模型加速,实现车辆目标检测,在保证检测速度的同时 mAP 值也得到了提升。明洪宇等^[12] 提出了一种基于 RetinaNet 的目标检测框架,结合高斯混合模型 (GMM) 和期望最大化 (EM) 算法对钢筋进行计数,并证明了 RetinaNet 网络在密集目标检测方面更具有优势。

针对工业生产过程中密集工件难以检测的问题,本文提出一种注意力机制优化 RetinaNet 的检测方法,将网络的注意力集中于工件目标,滤除背景等无关信息的影响,从而提高工件目标的特征表达能力,同时引入 Soft-NMS 后处理方法获取目标最佳位置。通过对比现有多种检测方法,表明所提方法达到了更高的检测精度,能更好的适应实际工业生产需求。

1 RetinaNet 网络结构

RetinaNet 网络由残差网络、特征金字塔网络、分类和边框回归子网络三部分组成。残差网络 (ResNet50) 作为主干特征提取网络,利用第 a 层的输入和 $a+1$ 层的输出进行叠加作为 $a+2$ 层的输入,对输入图片初步特征提取,代表性结构有 ResNet50, ResNet101, ResNet152 等,网

络结构内部使用残差块进行跳跃连接^[13],缓解了网络深度增加带来的梯度消失或梯度爆炸问题。FPN 在不同尺度上进行目标检测,将包含多尺度目标区域信息的特征图集合作为下一部分的输入,融合残差网络所获取的特征信息。分类和边框回归子网络分别完成目标框类别和位置回归任务得到最终检测结果,网络结构如图 1 所示。

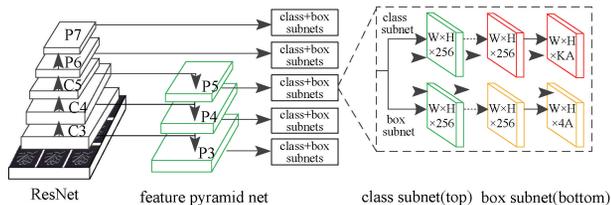


图 1 RetinaNet 网络结构

Fig. 1 RetinaNet network structure

损失函数用来衡量模型预测的好坏,是神经网络中优化的目标函数,神经网络的训练过程是损失函数优化的过程。Focal Loss 主要有两大特点:控制正负样本的权重、控制易分类和难分类样本的权重。通过增加权重与动态因子,解决了正负样本比例失衡的问题,优化了正负样本比例,降低了训练过程负样本所占比重。表达式为:

$$FL(p, y) = \begin{cases} -\alpha(1-p)^\gamma \log_\alpha(p), & y = 1 \\ -(1-\alpha)p^\gamma \log_\alpha(1-p), & y = 0 \end{cases} \quad (1)$$

式中: p 表示模型输出类别概率; y 表示类别标签; α 表示权重; γ 表示超参数。

2 RetinaNet 算法改进

2.1 注意力机制

注意力机制是人类在观察事物的时候,将注意力聚集信息的突出部分,把重要的资源优先分配,从而快速提取到高价值信息的一种有效手段。大量研究^[14-17] 已证明在网络结构中融入注意力机制的有效性。本文将 CBAM (convolutional block attention module) 模块加到主干特征提取网络 ResNet50 之后,残差网络可以通过堆叠多层卷积实现对密集工件的特征提取,CBAM 的加入使得在特征提取时将注意力集中于目标工件,降低形状相似、大小不一、分布密集工件特征带来的检测困难问题;然后将所提取到的信息传入到特征金字塔 FPN,FPN 巧妙地将浅层的细节信息和高层的语义信息进行结合,最终将融合后的信息经过后处理以及相应的分类和回归子网络得到最终检测结果。总体流程如图 2 所示。

输入特征图 $F \in R^{H \times W \times C}$, 经过通道注意力模块得到通道权重系数 M_c , 与输入特征图逐元素相乘得到 F' ; 将 F' 作为空间注意力模块的输入特征,得到空间权重系数 M_s , M_s 与 F' 进行逐元素相乘,最终得到融入多空间

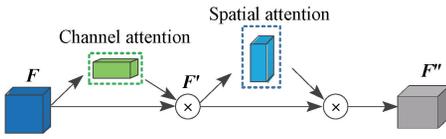


图 2 CBAM 总体流程

Fig. 2 The overall flow of CBAM

关键信息的特征图 F'' 。描述表示为：

$$F' = M_c(F) \otimes F \quad F'' = M_s(F') \otimes F' \quad (2)$$

式中： $M_c(F)$ 表示通道注意力， $M_s(F')$ 表示空间注意力， \otimes 表示逐元素相乘。

1) 通道注意力模块

通道注意力模块借鉴了 Inception^[18] 和 MobileNet^[19] 网络的思想, 根据不同通道特征的重要程度分配不同的权值, 细化特征在图像中所占权重, 如图 3 所示。

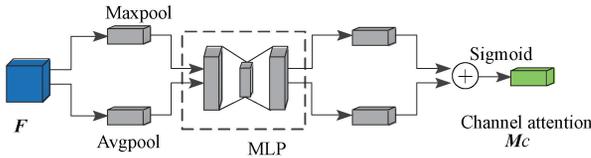


图 3 通道注意力模块结构

Fig. 3 Structure of channel attention module

算法步骤如下：

(1) 对输入特征 F 进行最大池化 (Maxpool) 和平均池化 (Avgpool) 操作, 得到 $F_{avg} \in \mathbf{R}^{1 \times 1 \times C}$ 和 $F_{max} \in \mathbf{R}^{1 \times 1 \times C}$, 并进行融合；

(2) 将结果分别送入多层感知机 MLP 共享网络中, 对特征进行降维和升维操作；

(3) 将维度改变的两个特征逐元素相加, 经过 Sigmoid 激活函数得到通道注意力权重系数 $M_c(F)$ 。描述可表示为^[20]：

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) = \sigma(W_1(W_0(F_{avg})) + W_1(W_0(F_{max}))) \quad (3)$$

式中： σ 表示 Sigmoid 激活函数； W_0 和 W_1 分别表示多层感知机隐层权重和输出层权重； F_{avg} 和 F_{max} 分别表示平均池化和最大池化。

2) 空间注意力模块

空间注意力模块^[21]用于准确定位空间中的目标特征信息, 如图 4 所示。

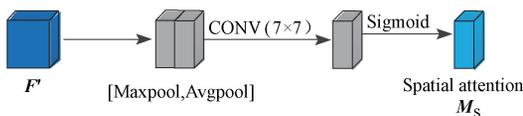


图 4 空间注意力模块结构

Fig. 4 Structure of spatial attention module

算法步骤如下：

(1) 将特征 F' 作为空间注意力模块的输入, 进行最大池化和平均池化操作, 得到 $F'_{avg} \in \mathbf{R}^{H \times W \times 1}$ 和 $F'_{max} \in$

$\mathbf{R}^{H \times W \times 1}$, 并进行融合；

(2) 将通道数为 2 的特征图经过一个 7×7 的卷积层在通道维度上进行压缩；

(3) 经 Sigmoid 激活后, 得到权重系数 $M_s(F')$ 。描述可表示为：

$$M_s(F') = \sigma(f^{7 \times 7}([AvgPool(F'); MaxPool(F')])) = \sigma(f^{7 \times 7}([F'_{avg}; F'_{max}])) \quad (4)$$

式中： σ 表示 Sigmoid 激活函数； $f^{7 \times 7}$ 表示卷积核为 7 的卷积层。

为直观得到特征提取网络所关注的重点信息, 突出加入注意力机制后网络结构的优越性, 本文引用 Grad-CAM^[22]、Guided Backprop、Guided Grad-CAM 三种方法对所检测的密集工件目标区域进行可视化, 结果如图 5 所示。Grad-CAM 的可视化方法很容易定位到目标工件所在区域, 表明图中颜色较深区域的变化会极大影响预测结果, 颜色越深对预测结果的权重贡献越大, 而无关干扰物及背景部分对预测结果影响较小。由图 5(a) 与 (b) 可以看出, 引入注意力机制后, 单类密集工件中聚集了更多深色区域, 涉及到图中的所有目标工件, 而多类密集工件中增大了原有深色区域的面积, 表明网络的注意力集中于目标工件, 忽略了背景等无关因素的影响, 从而能够获取更多有关于目标工件的细节信息。Guided Backprop 的可视化方法可以获得高分辨率以及详细特征的图片, 但是当一张图片分为不同类别时, 较难展现类别差异。Guided Grad-CAM 的可视化方法在图片分类为 5 类目标工件中的某一类时, 能够识别工件的重要特征, 如轮廓, 形状等, 同时也具备当存在类别差异性, 分类为“工件”, 没有展示其他干扰物的特征。

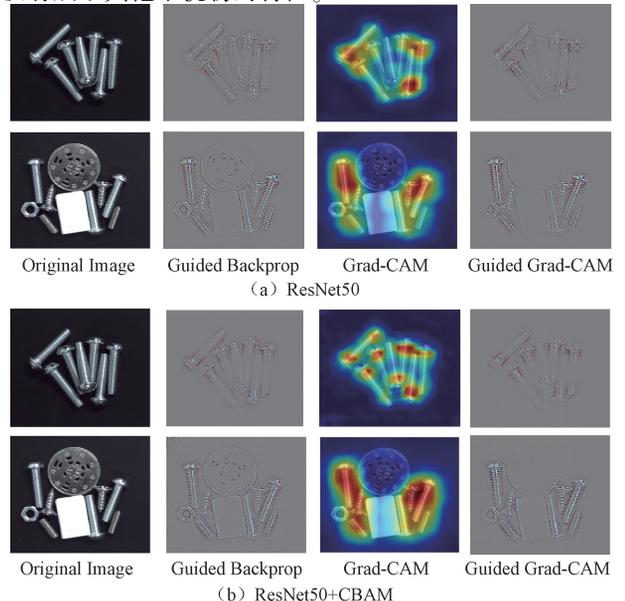


图 5 可视化效果

Fig. 5 Visualization effect

3.1 数据集

数据集的质量很大程度影响目标检测模型的精度。根据实际工业生产情况,本文选取 5 种工件进行数据集制作,如图 8 所示,为使识别效果更加直观,将其用对应的英文字母表示: bolt1 为 M5 十字圆头螺栓, bolt2 为 M2.5 无头内六角螺栓, screw 为 M5 自攻螺丝, nut 为 M5 螺母, gasket 为 M5 垫片。采用工业相机进行不同种类不同数量拍摄,共 784 张原始图片,使用 LabelImg 软件对所采集图片进行类别和位置信息标注,每标注一张就会生成相应的 XML 文件进行保存。将其标注为 PASCAL VOC2007 数据集格式,其中 bolt1 标注约 1 369 个, nut 标注约 2 065 个, gasket 标注约 1 630 个, bolt2 标注约 1 330 个, screw 约 1 710 个。同时为避免过拟合,训练过程中对数据集进行在线数据增强处理。最终选取数据集的 80% 作为训练集, 20% 作为测试集。

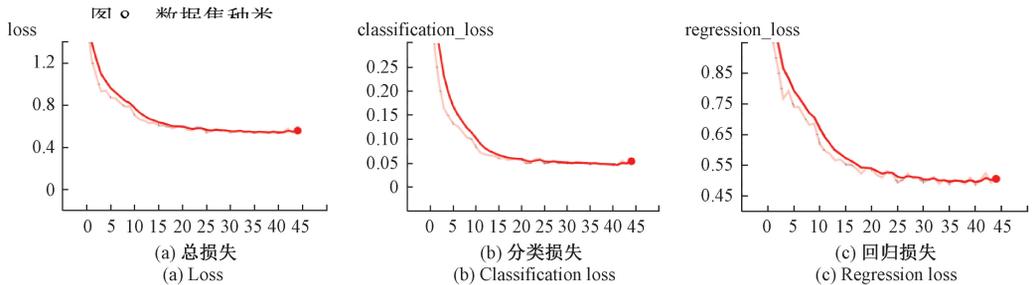


图 9 模型收敛曲线

Fig. 9 Convergence curve of the model

图 10 所示为主干特征提取网络 ResNet50 的部分通道特征图。不同的特征图代表着不同的特征,同一张特征图上所有元素都是对图像不同位置的同一个特征进行检测,特征图中某处数值的大小表示为当前位置对当前特征强弱的反应。以第 1 个卷积层为例,共使用了 64 个 7×7 的卷积核,即卷积操作完成后输出 64 个通道,最终可得到 64 张特征图的输出,相当于这一层共输出了 64 张单通道图片。分别求取 1 通道、2 通道、4 通道、8 通道、16 通道、32 通道以及 64 通道的特征进行可视化。

获得第 1 个卷积层的特征图后,按照此方法提取网络中每层中每个通道的特征图,并将同一层的特征图堆叠到同一个图像的张量中,通道按照顺序并排堆叠。网络中各层的可视化部分特征如图 11 所示。

各层按照如下规则进行命名: res/bn+stage+block+_branch+2a/2b/2c,其中 res 表示卷积层, bn 表示批归一化层, stage 表示第 $n-1$ 个模块, block 表示第几个 resblock 结构, _branch 代表分叉点, 2a/2b/2c 分别代表 1 个模块

3.2 模型训练

本文模型训练主要分为预训练和微调两个阶段。预训练阶段采用迁移学习思想,把在公开的 COCO 数据集上预训练的 ResNet50 模型初始化权重,从而预训练网络模型;将提取特征能力迁移至自制密集工件网络模型中,采用密集工件样本集微调网络。在训练过程中设置了早停机机制,网络性能没有发生提升训练过程就会停止,保存训练得到的权重信息,即为最终训练结果。

采用 AdamOptimizer 优化器进行参数优化,初始学习率为 0.000 1, Batch_size 表示每次输入神经网络的数据大小,每 Batch_size 个数据输入后,进行一次梯度下降法对权重参数进行更新,本文设置为 8, 训练 50 个 Epoch, 最终能够使得损失函数值趋于稳定。

模型训练收敛曲线如图 9 所示,本文运用 Tensorboard 对训练过程进行可视化,其中横坐标代表 Epoch,纵坐标代表损失值,总损失等于分类损失和回归损失之和。从图 9 可以看出,损失函数随着迭代次数的增加逐渐下降,最终总损失值稳定在 0.5 左右。

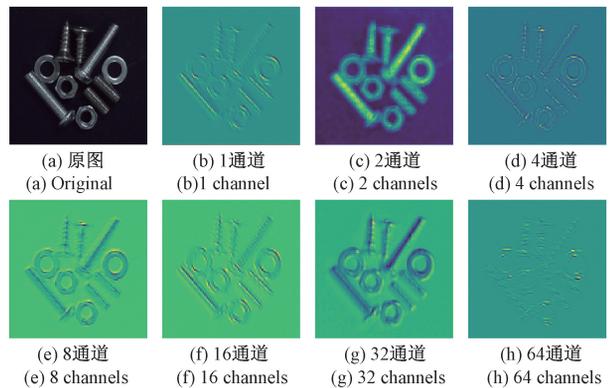


图 10 可视化每个通道特征图

Fig. 10 Visualize each channel feature diagram

中的不同卷积层。conv1 层表示网络中的第 1 个卷积层, res2a_branch2a 和 res2a_branch2b 表示第 1 个模块中第 1 个 resblock 中的第 1 层卷积层与第 2 层卷积层, res2b_branch2a 层表示第 1 个模块中第 2 个 resblock 中的第 1 个卷积层,以此类推。

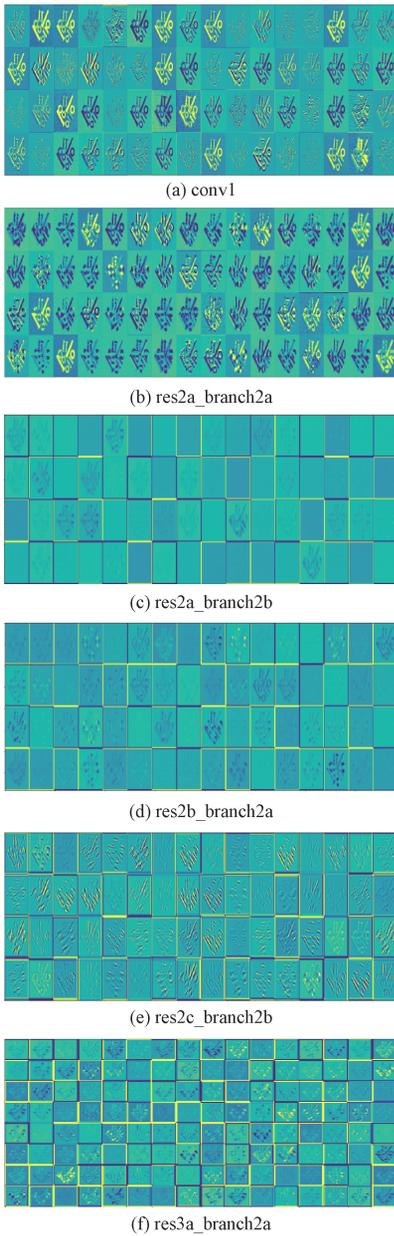


图 11 ResNet 各层卷积特征图

Fig. 11 Convolutional feature maps of each layer in ResNet

从图 11 可以看出,每个卷积层都学习到了不同的特征,在常规卷积神经网络中,随着层数的不断加深,深层得到的特征图不断抽象化,人眼很难识别出目标的纹理、边缘等特征,而残差网络以其独特的跳跃连接,使得上层输出特征可以对下层输出特征进行补充增强,因此在多次跳跃连接过程中,深层特征图能够获得更好的特征表达,同时使得整个网络提取到了不同层次更加丰富的特征。

3.3 工件检测性能分析

为证明本文算法的有效性,分别对单类密集工件和多类密集工件进行对比。图 12(a) 和 (c)、13(a) 和 (c)

分别为改进前检测效果,图 12(b) 和 (d)、图 13(b) 和 (d) 分别为改进后检测效果,对比可知在单类密集工件情况下,改进之后算法效果更好,降低了漏检率,定位更精准;在 multi-class 密集工件情况下,对于误检、漏检的情况都有所改善。

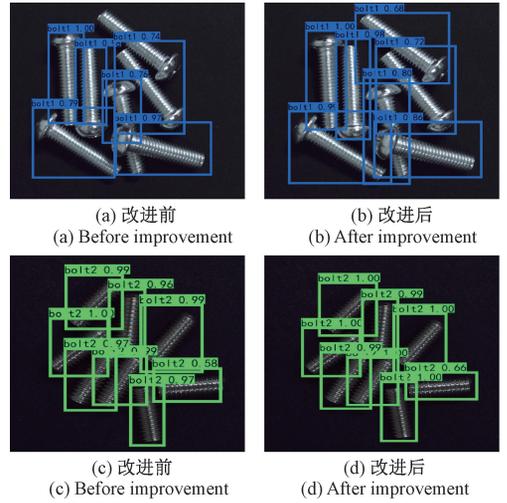


图 12 单类密集工件检测结果

Fig. 12 Detection results of single class dense workpiece

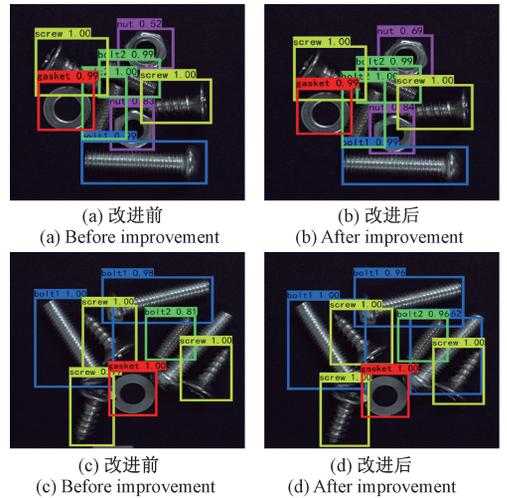


图 13 多类密集工件检测结果

Fig. 13 Detection results of more than one class of dense workpiece

为进一步证明本文所提的注意力机制优化 RetinaNet 方法的有效性,分别对不同光照条件,杂物干扰,变换相机位姿,背景变换等情况进行验证,检测效果如图 14, 15 所示。充分说明了相对复杂条件下,本文改进算法能够精准的对工件目标进行检测,效果显著,并无出现误检、漏检情况。

1) 评价指标

衡量模型分类准确性与定位精度常用的评价指标有

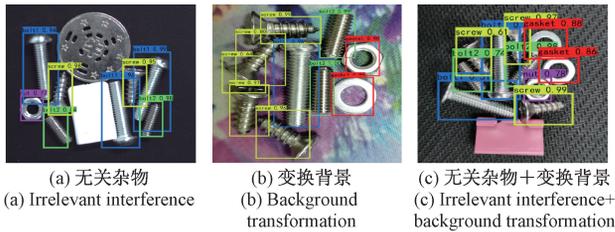


图 14 无关干扰物及背景变换检测效果

Fig. 14 The detection effect of irrelevant interferences and background transformation

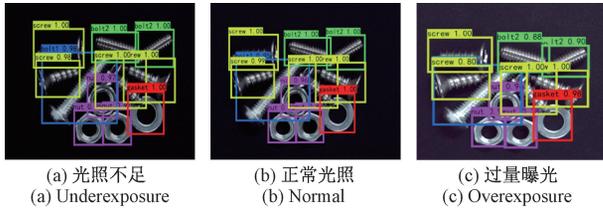


图 15 不同光照条件下检测效果

Fig. 15 Detection effects under different lighting conditions
准确率 (precision)、召回率 (recall)、平均精度值 (AP)、 F_1 分数等, 即:

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

$$AP = \int_0^1 P(R) dR \quad (10)$$

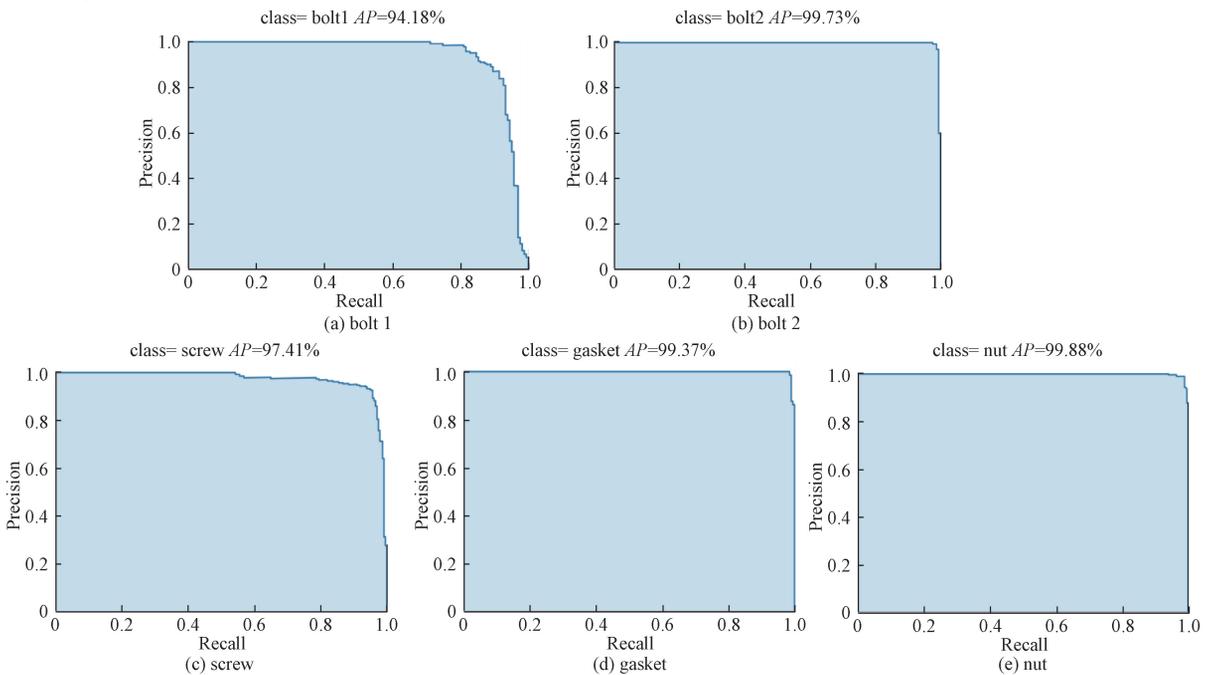


图 16 各类目标 P-R 曲线

Fig. 16 P-R curves of various targets

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (11)$$

式中: TP 表示正样本被预测为正确的样本数目, 即真正例; FP 表示负样本被预测为正确的样本数目, 即伪正例; FN 表示正样本被预测为错误的样本数目, 即伪反例。

表 1 为本文改进算法各类工件的评价指标值。

表 1 各类目标检测效果

Table 1 All kinds of target detection effect (%)

类别	AP	F_1	P	R
bolt1	94.18	89.00	86.13	91.41
bolt2	99.73	96.00	93.18	99.39
screw	97.41	92.00	88.58	96.15
gasket	99.37	97.00	95.65	98.51
nut	99.88	96.00	92.28	99.65

从表 1 可以看出, 本文改进算法在密集工件检测中的平均精度、准确率、召回率和 F_1 分数普遍较高, 效果较好, 更适合密集工件目标检测。

图 16 所示为本文所检测 5 类目标的召回率与准确率 (P-R) 曲线, 是衡量分类器的一个重要性能指标。当曲线出现震荡时, 表明正样本被判断为正例在变多, 但负样本被判断为正例也在变多。PR 曲线、PR 曲线下面积、AP 值, 充分体现了本文算法的良好性能。

2) 算法对比

为客观证明本文改进方法的优越性与可行性,在自制工件数据集上,分别与其他经典算法进行对比,结果如表 2 所示。

表 2 不同模型在本文数据集上检测效果对比

Table 2 Comparison of detection effects of different models on the dataset in this paper

目标检测算法	mAP/%	检测速度/s	FPS
Faster-RCNN	93.31	0.091	11
YOLOv3	90.18	0.014	69
RetinaNet	95.52	0.023	43
本文算法	98.11	0.026	38

从表 2 可以看出,与目标检测经典算法相比,改进算法的平均检测精度达到 98.11%,较改进前检测精度提高了 2.59%,相较于 Faster-RCNN 提高了 4.8%,相较于 YOLOv3 提高了 7.93%。在检测速度方面,本文算法达到了 0.026 s/张,在平均检测精度高出 2.59%的前提下,单张图片检测速度与改进之前算法相比仅降低了 0.003 s,在实际工业生产过程中本文算法所增加的检测时间对工件检测影响较小。实验结果表明,本文改进算法具有较好的检测效果,能够在一定程度上降低工件检测的漏检与误检率,对密集目标检测性能显著。

4 结 论

针对工业生产线上密集工件检测效果不佳的问题,本文提出注意力机制优化 RetinaNet 的密集工件检测方法,在模型中融入注意力机制以提取关于目标更多的细节信息,抑制无关信息;同时改进后处理方法,降低了密集目标工件漏检和误检率。实验结果表明,融合注意力机制并且对后处理算法改进后的模型性能上有了很大的提升,对密集工件达到了较好的检测效果,在非结构性复杂条件下,本文方法识别准确率达到 98.11%,相比原 RetinaNet 网络效果显著提升,并且高于其他对比算法,证明了该网络的优越性,能满足实际生产过程中工件检测需求。后续将对数据集进行扩充,同时提高检测工件的密集度;同时也将从轻量化网络结构出发,在不影响精度的同时提高检测速度,以期达到更好的检测效果。

参考文献

[1] 周济. 智能制造——“中国制造 2025”的主攻方向[J]. 中国机械工程, 2015, 26(17): 2273-2284.
ZHOU J. Intelligent manufacturing——Main direction of “Made in China 2025” [J]. China Mechanical Engineering, 2015, 26(17): 2273-2284.

[2] LUO J H, LIN C H. Pure FPGA implementation of a HOG based real-time pedestrian detection system [J].

Sensors, 2018, 18(4): 1174.

[3] MA X, GRIMSON W. Edge-based rich representation for vehicle classification [C]. Tenth IEEE International Conference on Computer Vision, Beijing: IEEE 2005: 1185-1192.

[4] FELZENSZWALB P F, GIRSHICK R B, MCALLESTER D, et al. Object detection with discriminatively trained part based models [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(9): 1627-1645.

[5] HE Y, ZHANG Z, YU F R, et al. Deep-reinforcement-learning-based optimization for cache-enabled opportunistic interference alignment wireless networks [J]. IEEE Transactions on Vehicular Technology, 2017, 66(11): 10433-10445.

[6] 张晓丽, 董昱. 面向铁路货车车号定位的 Faster R-CNN 卷积神经网络 [J]. 电子测量与仪器学报, 2020, 34(10): 65-73.
ZHANG X L, DONG Y. Faster R-CNN convolutional neural network for the location of freight train number [J]. Journal of Electronic Measurement and Instrumentation, 2020, 34(10): 65-73.

[7] 何静, 余昊宇, 张昌凡, 等. 基于 Canny-YOLOv3 的列车轮对踏面损伤检测 [J]. 电子测量与仪器学报, 2019, 33(12): 25-30.
HE J, YU H Y, ZHANG CH F, et al. Damage detection of train wheelset tread using canny-YOLOv3 [J]. Journal of Electronic Measurement and Instrumentation, 2019, 33(12): 25-30.

[8] 韩航迪, 徐亦睿, 孙博, 等. 基于改进 Tiny-YOLOv3 网络的航天电子焊点缺陷主动红外检测研究 [J]. 仪器仪表学报, 2020, 41(11): 42-49.
HAN H D, XU Y R, SUN B, et al. Using active thermography for defect detection of aerospace electronic solder joint base on the improved Tiny-YOLOv3 network [J]. Chinese Journal of Scientific Instrument, 2020, 41(11): 42-49.

[9] LIN T Y, GOYAL P, GIRSHICK R. Focal loss for dense object detection [C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2980-2988.

[10] LIN T Y, DOLLAR P, GIRSHICK R. Feature pyramid networks for object detection [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117-2125.

[11] 宋欢欢, 惠飞, 景首才, 等. 改进的 RetinaNet 模型的车辆目标检测 [J]. 计算机工程与应用, 2019, 55(13): 225-230.

- SONG H H, HUI F, JING SH C, et al. Improved RetinaNet model for vehicle target detection [J]. Computer Engineering and Applications, 2019, 55(13): 225-230.
- [12] 明洪宇,陈春梅,刘桂华,等. 基于 RetinaNet 的密集型钢筋计数改进算法 [J]. 传感器与微系统, 2020, 39(12): 115-118.
- MING H Y, CHEN CH M, LIU G H, et al. Improved counting algorithm for dense rebars based on RetinaNet [J]. Transducer and Microsystem Technologies, 2020, 39(12): 115-118.
- [13] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [14] MNH V, HEES N, GRAVES A. Recurrent models of visual attention [C]. Advances in Neural Information Processing Systems, 2014: 2204-2212.
- [15] WANG F, JIANG M, QIAN C, et al. Residual attention network for image classification [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017.
- [16] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2020, 42(8): 2011-2023.
- [17] CHU X, YANG W, OUYANG W, et al. Multi-context attention for human pose estimation [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017.
- [18] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 1-9.
- [19] HOWARD A G, ZHU M, CHEN B, et al. Efficient convolutional neural networks for mobile vision applications [J]. arXiv preprint ArXiv: 1704.0486, 2017.
- [20] DENG J, DONG W, SOCHER R, et al. ImageNet: A large-scale hierarchical image database [C]. Computer Vision and Pattern Recognition, 2009: 248-255.
- [21] MAX J, SIMONYAN K, ZISSERMAN A. Spatial transformer networks [C]. Advances in Neural Information Processing Systems, 2015.
- [22] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-Cam: Visual explanations from deep networks via gradient-based localization [C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 618-626.
- [23] JIANG B R, LUO R X, MAO J Y, et al. Acquisition of localization confidence for accurate object detection [C]. Proceedings of the European Conference on Computer Vision, 2018: 816-832.

作者简介



徐健, 1986 年于西安工程大学获得学士学位, 现为西安工程大学教授, 主要研究方向为机器视觉、目标检测、目标跟踪。

E-mail: xu0910@sian.com

Xu Jian received his B. Sc. degree from Xi'an Polytechnic University in 1986. Now he is a professor at Xi'an Polytechnic University. His main research interests include Machine vision target detection and target tracking.