

基于深度学习的多工件抓取点定位方法

安广琳^{1,2}, 李宗刚^{1,2*}, 杜亚江^{1,2}, 康会峰³¹兰州交通大学机电工程学院, 甘肃 兰州 730070;²兰州交通大学机器人研究所, 甘肃 兰州 730070;³北华航天工业学院航空宇航学院, 河北 廊坊 065000

摘要 针对工业生产线中多工件杂乱放置、互相遮挡, 存在漏检、错检以及抓取点定位困难问题, 提出一种基于协同深度学习的多工件抓取点定位方法。首先, 以 YOLOv5 为基础网络, 在输入端增加数据预处理模块用于图像增强时的角度变换, 检测层增加特征细化网络, 通过旋转锚框实现旋转工件的识别定位, 采用轻量化的 Ghost bottleneck 模块代替主干网络中的 bottleneckCSP 模块, 消除旋转锚框二次定位增加的时间成本, 将融合后的特征图分别输入注意力机制模块, 获取工件关键特征; 其次, 根据各工件检测框进行图像裁剪, 将多工件检测近似转化为单工件检测; 最后, 求取工件质心, 并结合旋转检测框的角度值确定抓取点。实验结果表明: 所提方法有效解决了工件距离相近或互相遮挡时抓取点定位问题, 且检测速度和精度均有明显优势, 保证了工业场景中多工件检测的实时性。

关键词 机器视觉; 工件检测; YOLOv5 旋转检测; Ghost bottleneck; 注意力机制

中图分类号 TP399 文献标志码 A

DOI: 10.3788/LOP220857

Multiple Workpiece Grasping Point Localization Method Based on Deep Learning

An Guanglin^{1,2}, Li Zonggang^{1,2*}, Du Yajiang^{1,2}, Kang Huifeng³¹School of Mechanical and Electrical Engineering, Lanzhou Jiaotong University, Lanzhou 730070, Gansu, China;²Robot Research Institute, Lanzhou Jiaotong University, Lanzhou 730070, Gansu, China;³College of Aerospace Engineering, North China Institute of Aerospace Engineering, Langfang 065000, Hebei, China

Abstract In this study, a multi-workpiece grasping point location method based on collaborative depth learning is proposed to solve the problems of disorderly placement and mutual occlusion of multiple workpieces in industrial production lines, such as missing inspection, wrong inspection, and difficult grasping point location. First, YOLOv5 is used as the basic network, and a data preprocessing module is added at the input end for angle transformation during image enhancement. Subsequently, a feature thinning network is added to the detection layer to realize the recognition and positioning of rotating workpieces via rotating anchor frames, and a lightweight Ghost bottleneck module is used to replace the bottleneckCSP module in the backbone network to eliminate the increased time cost due to the secondary positioning of the rotating anchor frames. Additionally, the fused feature maps are inputted into the attention mechanism module to obtain the key features of the workpiece. Subsequently, the image is clipped based on each workpiece detection frame, and the multi-workpiece detection is approximately transformed into single workpiece detection. Finally, the center of mass of the workpiece is obtained, and the grasping point is determined by combining the rotation angles of the detection frame. The experimental results show that the proposed method effectively solves the problem of locating the grab points of multiple workpieces close to or occluding each other. Furthermore, the method has higher detection speed and accuracy, which guarantees the real-time performance of multi-workpiece detection in industrial scenes.

Key words machine vision; workpiece inspection; YOLOv5 rotation inspection; Ghost bottleneck; attention mechanism

收稿日期: 2022-03-02; 修回日期: 2022-04-13; 录用日期: 2022-06-13; 网络首发日期: 2022-06-23

基金项目: 国家自然科学基金(61663020)、甘肃省高等学校科研项目成果转化项目(2018D-10)、兰州交通大学‘百名青年优秀人才培养计划’

通信作者: *lizongg@126.com

1 引言

在工业 4.0 背景下, 机器视觉以便捷、高效的特点得到了广泛应用。智能制造以机械臂为载体, 其结合机器视觉^[1]的实践落地成为研究重点之一。随着人工智能的发展, 工业生产的任务已逐渐从大规模生产转向专业定制和单件小批量灵活生产^[2]。因此, 工业场景中存在遮挡、杂乱放置的工件, 其检测与抓取点的定位成为一种多目标、无规律的任务。

传统的目标检测方法中存在着难以获取目标特征的有效语义信息、检测时间长^[3]、鲁棒性差^[4]等缺陷, 且检测精度和检测速度远未达到实际应用的要求。近年来, 深度学习在目标检测中被广泛应用^[5-6], 其中, 以 R-CNN^[7]为代表的两阶段检测算法和以 YOLO^[8]、SDD^[9]为代表的单阶段检测算法广泛应用于工业领域。王博^[10]将 Faster R-CNN 算法用于复杂工况下汽车小零件的识别分类, 可以精准检测分散的零件, 但对互相遮挡的零件识别成功率低, 存在漏检问题。Gong 等^[11]针对堆叠板类零件使用传统模板匹配难以识别被遮挡零件问题, 将 YOLOv3^[12]算法用于零件识别模型, 提高了检测精度。Li 等^[13]将工件检测定位系统应用于机械臂的抓取系统, 通过 YOLOv3 实现工件的识别定位, 采用双目视觉检测工件的关键点, 完成复杂环境下机械臂的抓取任务。但随着生产任务的需求、复杂度不断增加, 工件结构变得不规则。为了获得更贴合工件轮廓的网络输出形式, 深度学习网络规模不断扩大, 增加了网络推理时间, 导致其难以满足工业视频级的实时检测。因此, 基于深度学习的目标检测方法倾向于结合多种网络结构优点。雷毅^[14]提出一种基于深度学习与图像处理融合的多目标检测算法, 该算法首先采用改进的 YOLOv3 进行工件初步检测定位, 然后经图像裁剪和传统图像处理获取工件轮廓并计算轮廓最小外接矩形和定位矩形的交并比, 最后根据轮廓形状确定抓取点, 两种算法融合使网络轻量化, 对检测速度具有明显的提升。但是对于工业中多工件随机、倾斜放置场景, 使用水平检测算法检测时将会引入大量的背景噪声; 当工件密集堆放时也将造成误检、漏检。而旋转目标检测算法将获得更贴合工件轮廓的网络输出形式, 更适用于遮挡工件的检测。现有的旋转检测算法中, RRPN^[15]在 Faster R-CNN 算法输出的每个检测框基础上各衍生 6 个不同角度的检测框来实现旋转目标的检测, 检测过程中为提高检测精度产生了大量冗余旋转检测框, 影响检测速度。CAD-Net^[16]为两阶段旋转目标检测算法, 进行目标检测时融入全局上下信息, 并使用金字塔局部上下文网络学习对象级的上下文信息, 但是检测速度低于单阶段的旋转检测算法。R³Det^[17]为单阶段旋转目标检测算法, 在生成水平检测框后引入特征细化网络(FRN)实现旋转目标检测。这些方法多用于遥感检测, 适合大纵横比的目标检测, 但仍然存在

较多误检、漏检情况。因此, 针对非结构化工业场景中多工件的检测, 需要更精细化的检测算法, 同时工件抓取点的合理定位也将是工件可靠抓取的前提。

综上所述, 本文从工业生产线中随意放置、互相遮挡的多工件识别和抓取点定位出发, 采用文献^[14]的网络轻量化思想, 将深度学习与传统图像处理融合进行多工件的识别定位。首先, 对实时性好、易于工业生产中部署的 YOLOv5^[18]检测算法增加旋转检测网络, 实现旋转工件的识别定位; 其次, 根据旋转检测框对图像进行裁剪, 将多工件检测问题近似转化为单工件检测问题; 最后, 为保证工件的可靠抓取, 通过传统图像处理求取每个工件质心作为抓取点, 结合旋转检测框的角度值确定工件抓取点。

2 改进的 GB-FRN-YOLOv5 多工件检测算法

图 1 为所提工业场景中多工件检测算法的整体结构, 以 YOLOv5 检测算法为基础, 主要由数据预处理模块、GB-DarkNet53 主干网络、FRN、注意力机制模块等 4 部分构成。

2.1 数据预处理模块

YOLOv5 算法在输入端将 4 张图像随机裁剪、缩放、排列拼接, 以丰富数据集。但是当图像宽高非等比例缩放时, 会造成目标角度信息相较于整个图像发生变化, 从而影响检测精度。目标物体实例角度随图像缩放而改变的示意图如图 2 所示。因此, 在 YOLOv5 网络输入端增加数据预处理模块, 提出角度转换算法, 对缩放前后目标角度进行计算变换, 提高任意方向工件的检测精度。图 2(a) 中缩放前图像矩形长边与 x 轴所成夹角为 θ_1 , 长边端点坐标为 (x_1, y_1) 和 (x_2, y_2) , 宽高为 (w_1, h_1) , 经宽高非等比例缩放后如图 2(b) 所示, 对应角度和端点分别为 $\theta_2, (x_3, y_3), (x_4, y_4)$, 宽高为 (w_2, h_2) 。

由宽高缩放比例可计算对应点间关系:

$$\begin{cases} x_3 = x_1 \times \frac{w_2}{w_1}, y_3 = y_1 \times \frac{h_2}{h_1} \\ x_4 = x_2 \times \frac{w_2}{w_1}, y_4 = y_2 \times \frac{h_2}{h_1} \end{cases} \quad (1)$$

由式(1)可计算角度间换算公式:

$$\begin{cases} \tan \theta_1 = \frac{y_2 - y_1}{x_2 - x_1} \\ \tan \theta_2 = \frac{y_2 - y_1}{x_2 - x_1} \times \frac{h_2 \times w_1}{h_1 \times w_2} \\ \theta_2 = \arctan(\tan \theta_1 \times \frac{h_2 \times w_1}{h_1 \times w_2}) \end{cases} \quad (2)$$

由式(2)可得出, 图像经宽高非等比例缩放时, 缩放后的角度可由缩放前角度及缩放前后图像宽高参数表示, 对数据进行角度换算调整使输入数据更适应训练网络, 提高检测精度。

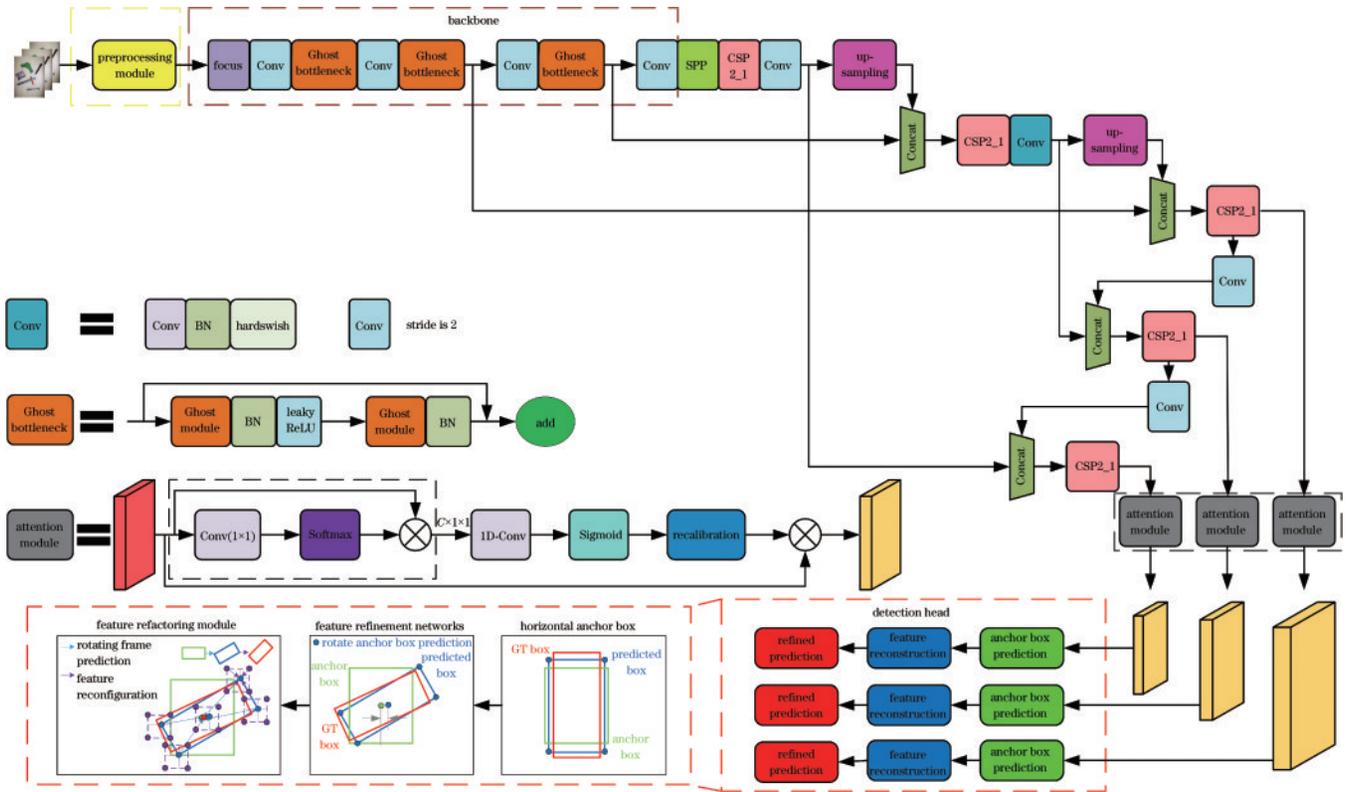


图 1 改进的 GB-FRN-YOLOv5 网络结构图

Fig. 1 Network structure of improved GB-FRN-YOLOv5

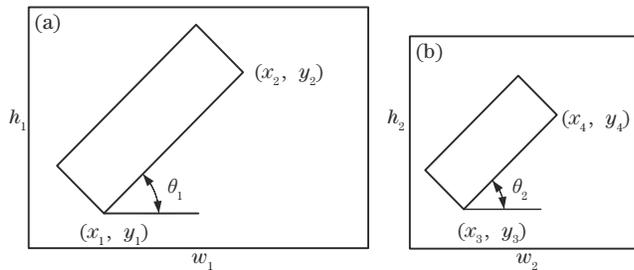


图 2 目标实例角度变化图。(a)缩放前图像；(b)缩放后图像
Fig. 2 Chart of target example angle change. (a) Image before scaling; (b) scaled image

2.2 FRN

借鉴 R³Det 算法生成旋转检测网络思想,为了实现基于 YOLOv5 检测算法的任意方向工件的检测,在检测网络 head 中增加 FRN 得到改进的 FRN-YOLOv5。首先,利用水平 anchor box 高召回率的优势,通过 YOLOv5 检测网络以较快的速度产生大量的水平候选框;然后,在回归子网和分类子网后面添加 FRN,引入角度为 -90° 、 -75° 、 -60° 、 -45° 、 -30° 、 -15° 的多个旋转锚框检测任意方向的工件,FRN 中前景和背景的阈值分别为 0.5 和 0.4。采用五维向量 (x, y, w, h, θ) 定义有方向的矩形检测框,如图 3 所示: 1) 矩形框中以垂直坐标最小的顶点所在的水平方向为参照线 x 轴; 2) θ 表示矩形框旋转到与 x 轴平行的锐角,取值范围为 $[-\pi/2, 0]$, w 为与 x 轴呈锐角的边长, h

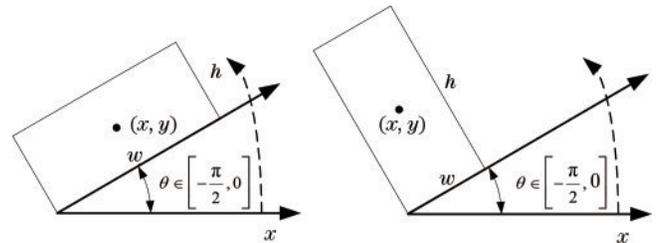


图 3 旋转框定义示意图

Fig. 3 Schematic diagram of definition of rotating frame

为矩形的另一条边长; 3) 矩形框中心点坐标为 (x, y) 。由于引入角度参数 θ , 因此需要在检测框定位时考虑附加角度偏移, 旋转检测框表示为

$$\begin{cases} t_x = (x - x_a) / w_a, t_y = (y - y_a) / h_a \\ t_w = \log(w / w_a), t_h = \log(h / h_a), t_\theta = \theta - \theta_a \\ t'_x = (x' - x_a), t'_y = (y' - y_a) / h_a \\ t'_w = \log(w' / w_a), t'_h = \log(h' / h_a), t'_\theta = \theta' - \theta_a \end{cases}, (3)$$

式中: x, x_a, x' 分别代表 GT box、anchor box、predicted box 坐标, 其他变量类似。

在进行检测框位置确定时引入角度参数使用的是浮点数, 但是图像为离散的像素点, 所以实际目标的位置都为定点, 这导致旋转锚框定位过程中存在特征未对齐现象, 如图 4 所示。以 anchor box 中心点为特征点, 可看出 predicted box 和 anchor box 的中心点存在偏差。

为消除特征偏差, 在增加的 FRN 中设计特征重构

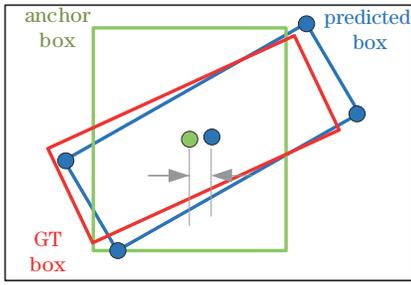


图 4 特征未对齐示意图

Fig. 4 Schematic diagram of feature unaligned

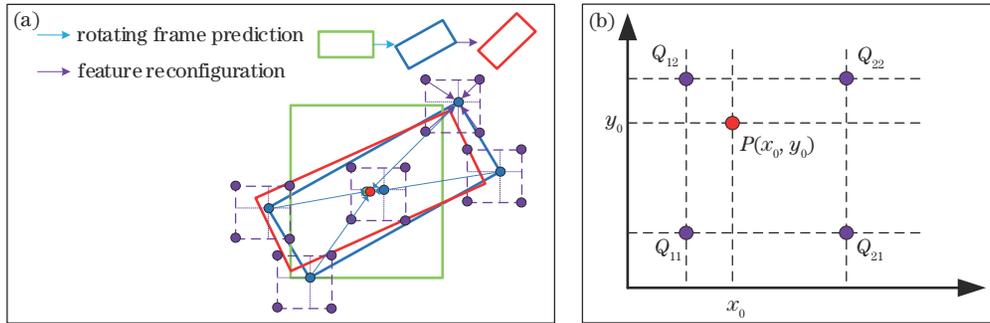


图 5 特征细化阶段示意图。(a)特征重构示意图;(b)双线性插值

Fig. 5 Schematic diagrams of feature refinement stage. (a) Schematic diagram of feature reconstruction; (b) bilinear interpolation

双线性插值原理为通过查询相距特征点 P 最近的 4 个像素点 Q_{11} 、 Q_{12} 、 Q_{21} 、 Q_{22} ，经 2 个方向的单线性插值计算 P 点的特征值，如图 5 (b) 所示。已知 $Q_{11}(x'_0, y'_0)$ 、 $Q_{12}(x'_0, y''_0)$ 、 $Q_{21}(x''_0, y'_0)$ 、 $Q_{22}(x''_0, y''_0)$ ，则 $P(x_0, y_0)$ 的表达式为

$$f(x_0, y_0) = \frac{f(Q_{11})}{(x''_0 - x'_0)(y''_0 - y'_0)} (x''_0 - x_0)(y''_0 - y_0) + \frac{f(Q_{21})}{(x''_0 - x'_0)(y''_0 - y'_0)} (x_0 - x'_0)(y''_0 - y_0) + \frac{f(Q_{12})}{(x''_0 - x'_0)(y''_0 - y'_0)} (x''_0 - x_0)(y_0 - y'_0) + \frac{f(Q_{22})}{(x''_0 - x'_0)(y''_0 - y'_0)} (x_0 - x'_0)(y_0 - y'_0) \quad (4)$$

模块 (FRM)，如图 5(a) 所示。特征细化过程中，在水平框获取高召回率后，引入旋转锚框定位任意方向的工件，定位时对每个特征点只保留置信度得分最高的预测框，过滤其余预测框以提高检测速度，保证工件的实时检测。FRM 采用双线性特征插值方法精确获取细化预测框的位置特征信息，将其重新编码到对应的特征点，并通过逐像素的方式重构整个特征图，实现特征对齐。最后，将重构的特征图添加到原始特征图中，通过目标回归和分类子网重新进行预测得到多工件的检测结果。

2.3 GB-DarkNet53 主干网络

由于增加了 FRN 结构，相当于对工件进行二次预测定位，检测时间成本增加。为保证工件的实时检测，基于轻量化的 Ghost bottleneck^[19] 模块，提出采用 Ghost bottleneck 模块代替 YOLOv5 主干网络中 bottleneckCSP 模块的方法，获得 GB-DarkNet53 主干网络，消除特征映射中普遍存在的特征冗余，在保证网络精度的同时减少网络参数和计算量，得到改进的 GB-FRN-YOLOv5 检测算法。Ghost bottleneck 模块分为 2 条支路，如图 6 所示。2 个 Ghost 模块构成其中的 1 条支路，通过第 1 个 Ghost 模块增加特征通道数，第 2 个 Ghost 模块减少特征通道数，保证输入输出通道数一致的情况下压缩网络模型、降低参数量和计算量，最后通过另一支路连接 2 个 Ghost 模块的输入和输出。

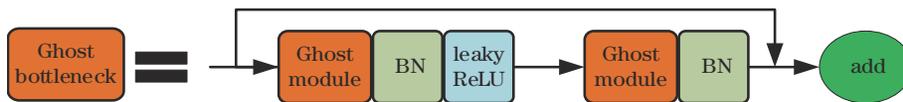


图 6 Ghost bottleneck 模块

Fig. 6 Module of Ghost bottleneck

Ghost 模块首先采用普通卷积操作生成少量的固有特征映射，然后利用线性 ϕ 运算增加特征映射，最后将两部分特征映射组合输出，使特征图的数量与普通卷积层相同。线性 ϕ 运算代替大量的卷积层计算，极大减少计算量，Ghost 模块和普通卷积层计算过程如图 7 所示。

相同的输入特征通过不同方法得到相同的输出特

征，对普通卷积层与 Ghost 模块计算过程量化进行比较。

设输入特征图尺寸为 $H \times W \times e$ ， e 表示输入通道，输出特征图尺寸为 $H' \times W' \times p$ ， p 表示输出通道。则图 7 中普通卷积层计算量为

$$N_{\text{FLOPs}} = e \cdot p \cdot H' \cdot W' \cdot u \cdot u, \quad (5)$$

式中： u 表示卷积核大小。而图 7 中 Ghost 模块计算过程分为两步：

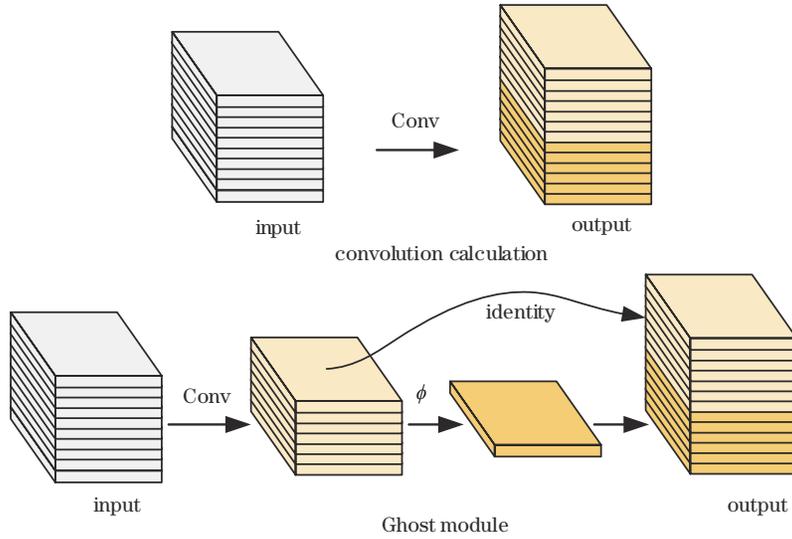


图7 普通卷积层和 Ghost 模块计算示意图

Fig. 7 Schematic diagram of ordinary convolutional layer and Ghost module calculation

1) 输入特征图通过普通卷积操作生成少量中间特征图 $H' \times W' \times r$, r 为第 1 步的输出通道、第 2 步的输入通道。此步骤计算量为

$$N_{FLOPs}'' = e \cdot r \cdot H' \cdot W' \cdot u \cdot u. \quad (6)$$

2) 为得到最终的输出通道数 p , 对中间特征图的 r 个通道通过线性 ϕ 运算进行扩充, 每个通道的扩充倍数为 s , 则有

$$p = r \cdot s. \quad (7)$$

而 Ghost 在第 2 步时每个特征通道有一次恒等变换 (identity), 因此, 实际中每个通道的扩充倍数为 $s - 1$, 则此步骤计算量为

$$N_{FLOPs}''' = r \cdot (s - 1) \cdot H' \cdot W' \cdot v \cdot v, \quad (8)$$

式中: v 表示线性 ϕ 运算卷积核的大小。则联立式(5)~(8)可计算普通卷积与 Ghost 模块的加速比:

$$r_s = \frac{N_{FLOPs}'}{N_{FLOPs}'' + N_{FLOPs}'''} = \frac{e \cdot p \cdot u \cdot u}{e \cdot r \cdot s \cdot H' \cdot W' \cdot u \cdot u + r \cdot (s - 1) \cdot H' \cdot W' \cdot v \cdot v}. \quad (9)$$

设过程中卷积核大小近似相等, 即 $u \approx v$, 同时 $s \ll e$, 则式(9)可改写为

$$r_s \approx \frac{e \cdot s}{e + s - 1} \approx s. \quad (10)$$

即传统卷积的计算量为 Ghost 模块的 s 倍, 普通卷积层计算量明显更大。同样可得普通卷积与 Ghost 模块的压缩比为

$$r_c = \frac{e \cdot p \cdot u \cdot u}{e \cdot r \cdot u \cdot u + r \cdot (s - 1) \cdot v \cdot v} \approx \frac{e \cdot s}{e + s - 1} \approx s. \quad (11)$$

由式(11)可看出, 生成相同的特征图, 普通卷积操作过程中所需参数比 Ghost 模块更多。

2.4 注意力多尺度特征图

工业生产线工件杂乱放置、互相遮挡, 被遮挡工件易出现错检、漏检情况, 为提高工件检测的召回率和精度, 引入选择核心网络 (SENet)^[20]、有效信道注意网络 (ECANet)^[21] 注意力机制思想, 通过改善特征图不同通道的权重系数, 选择有益的特征通道, 有效提取遮挡工件的轮廓特征。将 GB-DarkNet53 主干网络输出的不同特征层间的特征图经上采样并融合后生成的 3 个预测特征图分别输入注意力机制模块中, 通过提高含遮挡工件关键信息的特征通道权重系数, 进行被遮挡工件的自适应特征优化, 最后经过特征图的重标定生成针对密集堆放工件检测任务的特征图, 再输出到 YOLOv5 检测网络中。注意力多尺度特征图如图 8 所示。

引入的注意力机制模块将输入的融合特征信息分为两个分支, 其中一个分支将卷积操作获得的 $C \times 1 \times 1$ 的向量进行特征激活, 然后利用一维卷积计算得到每个通道的权值并采用 Sigmoid 激活函数处理, 最后完成特征图的重标定, 即还原原来的特征图尺寸并通过乘法将其按每个通道加权到另一分支直接传入的初始特征上, 再输入 YOLOv5 检测网络中。其中, 一维卷积的卷积核大小通过一个函数来自适应, 自适应

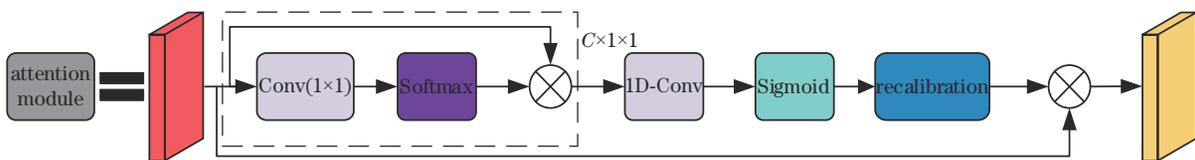


图8 注意力多尺度特征图

Fig. 8 Feature map of attentional multiscale

卷积核大小的计算公式为

$$g = \psi(a) = \left\lfloor \frac{\log_2 a + 1}{\gamma} \right\rfloor, \quad (12)$$

式中: g 为自适应卷积核大小结果; $\psi(a)$ 表示对通道 a 进行非线性映射; a 表示通道数; γ 表示缩放因子, 在这里取 $\gamma = 2$ 。

2.5 损失函数

模型总损失函数由定位损失、角度损失、置信度损失和类别损失构成。由于原始 YOLOv5 的 anchor box 不带角度信息, 相当于 $\theta = 0$, 增加 FRN 后, 使用新定义的旋转锚框, 则将角度损失函数定义为

$$L_\theta = \lambda_2 \sum_{i=0}^{S^2} \sum_{j=0}^B V_{\text{objn}} \left\{ \left[(\theta_i - \theta'_i) \times \frac{\pi}{180} \right]^2 \right\}, \quad (13)$$

式中: λ_2 表示权值系数; θ_i 、 θ'_i 分别表示预测框和真实框的旋转角度。则模型总的损失函数为

$$\begin{aligned} L = L_{\text{loc}} + L_\theta + L_{\text{conf}} + L_{\text{cla}} = & \lambda_1 \sum_{i=0}^{S^2} \sum_{j=0}^B V_{\text{objn}} [(x_i - x'_i)^2 + (y_i - y'_i)^2] + \\ & \lambda_1 \sum_{i=0}^{S^2} \sum_{j=0}^B V_{\text{objn}} [(\sqrt{w_i} - \sqrt{w'_i})^2 + (\sqrt{h_i} - \sqrt{h'_i})^2] + \\ & \lambda_2 \sum_{i=0}^{S^2} \sum_{j=0}^B V_{\text{objn}} \left\{ \left[(\theta_i - \theta'_i) \times \frac{\pi}{180} \right]^2 \right\} + \\ & \sum_{i=0}^{S^2} \sum_{j=0}^B V_{\text{objn}} (C_i - C'_i)^2 + \lambda_3 \sum_{i=0}^{S^2} \sum_{j=0}^B V_{\text{objn}} (C_i - C'_i)^2 + \\ & \sum_{i=0}^{S^2} V_{\text{objn}} \sum_{C \in \text{cla}} [P_i(C) - P'_i(C)]^2, \end{aligned} \quad (14)$$

式中: λ_1 、 λ_3 表示权值系数; B 表示预测框的数量; S 表示特征图边长; (x_i, y_i, w_i, h_i) 表示预测框的中心坐标及宽高; (x'_i, y'_i, w'_i, h'_i) 表示真实框的中心坐标及宽高; C'_i 、 C_i 分别表示真实框和预测框的类别; $P'_i(C)$ 、 $P_i(C)$ 分别表示真实框和预测框的置信度; V_{objn} 表示二进制值, 当 $V_{\text{objn}} = 1$ 时表示预测框目标为前景, $V_{\text{objn}} = 0$ 或无回归值时表示预测框目标为背景。

3 多工件的检测及抓取点定位

将多工件图像输入改进的 GB-FRN-YOLOv5 检测算法中, 获取杂乱放置的工件的矩形检测框, 由于已经检测出工件的类别信息且旋转检测框含有角度信息, 在图像处理中无需分析工件的类别和角度。在图像处理前, 根据矩形检测框对工件图像进行裁剪, 得到多幅单工件图, 将多工件检测问题近似转化为单工件检测问题。

由于异构件质量不均匀, 直接以工件中点作为抓取点, 可能造成抓取不稳掉落情况发生。为保证工件的可靠抓取, 基于带阈值的重心法^[22], 提出一种求工件图像质心的抓取点定位方法。首先对工件图像进行图像预处理: 由于工件图像偏暗, 为突出工件轮廓, 对裁

剪后的单工件图像采用最大值法进行灰度化, 采用中值滤波去除工件图像椒盐噪声和斑块噪声。为提高图像处理效率, 采用全局阈值的方法对工件图像进行二值化, 显示工件轮廓的目标区域。最后求取工件图像的一阶矩用来表示工件图像质心。

设图像灰度为 $f(x_l, y_k)$, 其中 $l = 1, \dots, m$, $k = 1, 2, \dots, n$ 。二值化过程为

$$F(x_l, y_k) = \begin{cases} f(x_l, y_k), & f(x_l, y_k) \geq T \\ 0, & f(x_l, y_k) < T \end{cases}, \quad (15)$$

式中: $F(x_l, y_k)$ 为分割后的图像; T 表示背景阈值, 由于图像中含有工件阴影且整体偏暗, 设定 $T = 80$ 。带阈值的重心法计算公式如下:

$$\begin{cases} x_1 = \frac{\sum_{l=1}^m \sum_{k=1}^n F(x_l, y_k) x_l}{\sum_{l=1}^m \sum_{k=1}^n F(x_l, y_k)} \\ y_k = \frac{\sum_{l=1}^m \sum_{k=1}^n F(x_l, y_k) y_k}{\sum_{l=1}^m \sum_{k=1}^n F(x_l, y_k)} \end{cases}. \quad (16)$$

针对工业生产中的异构件, 以其质心为抓取点, 结合改进的 GB-FRN-YOLOv5 检测算法获得的旋转检测框的角度信息, 确定工件的抓取点。

4 实验结果分析

所提算法的实验过程包括工件数据集的构建, 以及与传统的 YOLOv5、旋转目标检测算法 CAD-Net、R³Det、Gliding vertex^[23] 的实验对比, 验证所提算法在工业生产中多工件的检测与抓取点定位上的有效性与可靠性。

4.1 构建工件数据集

根据工业生产流水线中常见的异构件构建多工件数据集, 包含螺栓 (bolt)、套件 (kit)、支撑板 (support plate)、套管 (bushing)、十字螺栓 (crossbolt) 和卡扣 (buckle) 等 6 类工件共 5600 幅图像, 按照 8:2 的比例划分为训练集和测试集。利用图像标记软件 RolabelImg 对数据集中的样本数据集进行标记, 数据集标签格式与 PASCAL VOC 数据标签格式一致。

4.2 实验参数配置

实验系统为 Ubuntu 18.04, Python 版本为 3.8, PyTorch 版本为 1.6.0, 训练 GB-FRN-YOLOv5 算法时输入的图像大小为 544×544 , 批量大小为 12, 相关参数如表 1 所示。

4.3 评价指标

为了验证所提模型的有效性, 选取平均准确率 (AP) 来评价模型对单个工件类别的检测准确性。AP 从准确率 (P)、召回率 (R) 两方面评价 GB-FRN-YOLOv5 模型对单个工件类别的检测精度, 用平均准

表 1 硬件配置和模型参数

Table 1 Hardware configuration and model parameters

Name	Configuration	Training parameter	Parameter value
GPU	GeForce RTX 2080ti	Warmup_epochs	5.0
CPU	Intel(R) Xeon(R)CPUE52680 v2	Warmup_momentum	0.95
CUDA	10.1	Learning rate	0.01
CuDNN	7.6.5	Weight decay	0.001

准确率(mAP)来评价所提模型对多个工件类别的综合检测性能。通过每秒可处理的工件图像数量(FPS)来衡量模型的检测速度,具体公式如下:

$$\begin{cases} P = \frac{N_{TP}}{N_{TP} + N_{FP}} \\ R = \frac{N_{TP}}{N_{TP} + N_{FN}} \\ R_{AP} = \int_0^1 P(R) dR \end{cases}, \quad (17)$$

式中:以检测工件套件为例, N_{TP} 表示预测的工件为套件、实际为套件的数量; N_{FP} 表示预测的工件为套件、实际不是套件的数量; N_{FN} 表示预测的工件不是套件、实际为套件的数量; $P(R)$ 是以 R 为参数的函数。

4.4 工件检测结果分析

通过对 GB-FRN-YOLOv5 模型性能的评估和实验对比,对所提方法在实际工业生产中的检测性能、消融实验和可行性等 3 方面进行分析。

4.4.1 工件检测性能分析

损失函数在训练中反映真实值与预测值之间的关系,图 9 给出改进的 GB-FRN-YOLOv5 模型训练时总的损失函数。由图 9 可看出,损失函数呈下降趋势,最终趋于稳定,表明所提模型具有良好的检测性能。

将所提算法与 YOLOv5、CAD-Net、R³Det、Gliding vertex 等 4 种检测算法进行任意方向工件的检测实验对比,结果如表 2 所示。从表 2 可看出,水平检

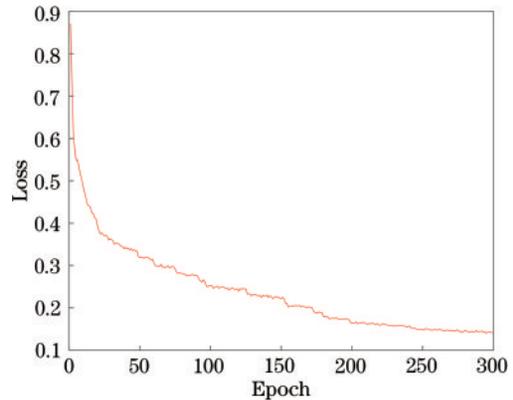


图 9 改进的 GB-FRN-YOLOv5 模型训练损失函数
Fig. 9 Training loss function of improved GB-FRN-YOLOv5 model

测算法 YOLOv5 具有较高的检测速度,但由于其是水平检测,检测图所含背景噪声大,表明以 YOLOv5 为基础网络进行改进的合理性。

所提算法与 3 种旋转检测算法的检测对比结果如图 10 所示。从图 10 可看出,CAD-Net 和 R³Det 旋转检测算法在进行多工件检测时存在漏检问题且定位精度低,虽然 Gliding vertex 算法具有较高的检测精度,但从表 2 可看出,其检测速度低于所提算法。综合对比可得,所提 GB-FRN-YOLOv5 算法相比于其他目标检测算法在检测精度和处理速度上均有着明显优势,满足工业视频级的检测要求。

表 2 GB-FRN-YOLOV5 算法与其他检测算法对比实验结果

Table 2 Experimental result comparison of GB-FRN-YOLOV5 algorithm and other detection algorithms

Detection algorithm	Accuracy of detection for each type of target AP / %						mAP / %	FPS
	bolt	kit	bushing	crossbolt	buckle	supportplate		
YOLOv5	88.21	84.14	83.67	85.23	87.21	90.78	86.54	65.16
CAD-Net	84.41	81.34	82.42	86.65	86.15	91.07	85.34	58.46
R ³ Det	83.65	81.24	80.13	85.25	85.33	89.66	84.21	60.51
Gliding vertex	87.78	85.26	84.02	86.66	88.13	91.59	87.24	63.26
GB-FRN-YOLOv5	90.85	90.55	90.84	90.82	90.79	90.90	90.79	71.43

4.4.2 消融实验

为了更直观地对比不同改进策略对模型性能的影响,以 YOLOv5 为基础网络结构进行消融实验,依次添加不同模块,共形成 5 组网络结构在构建的工件数据集上进行验证。当输入 544×544 的工件图像时,检测结果如表 3 所示。由表 3 可得,所提各改进策略均有

利于任意方向的多工件检测性能的提升。由表 3 中数据可知:模型 B 相较于模型 A, mAP 提高了 1.59 个百分点, FPS 降低了 1.21 frame/s, 表明引入 FRN 增加了工件旋转框定位,对检测速度有所影响,但 FRN 及其 FRM 的引入,提高了检测精度,增强了任意方向工件检测的泛化能力;模型 C 相较于模型 B, mAP 提高了



图 10 不同算法的工件检测结果图。(a)CAD-Net;(b)R³Det;(c)Gliding vertex;(d)GB-FRN-YOLOv5

Fig. 10 Workpiece detection result graphs of different algorithms. (a) CAD-Net; (b) R³Det; (c) Gliding vertex; (d) GB-FRN-YOLOv5

表 3 不同模块对检测性能的影响

Table 3 Effect of different modules on detection performance

Models	A	B	C	D	E
YOLOV5	✓	✓	✓	✓	✓
FRN		✓	✓	✓	✓
Data pre-processing module			✓	✓	✓
Ghost bottleneck				✓	✓
Attention mechanism					✓
mAP / %	86.54	88.13	88.56	88.75	90.79
FPS	65.16	63.95	63.41	72.16	71.43

0.43 个百分点,表明引入数据预处理模块有益于检测精度的提升;模型 D 相较于模型 C, FPS 提高了 8.75 frame/s,表明通过 Ghost bottleneck 模块替换 bottleneckCSP 模块,极大提升了检测速度,该模块很

好弥补了增加的几个模块所消耗的时间成本,保证了工业生产流水线多工件的实时检测;模型 E 相较于模型 D, mAP 提高了 2.04 个百分点,表明引入的注意力机制网络能抑制无用特征,虽然使检测时间有所增加,但提高了检测精度。

图 11 为部分检测结果对比图。图 11(a)为模型 D 检测结果图,可以看出,套件被漏检且 6 类工件置信度普遍偏低;图 11(b)为模型 E 检测结果图,可以看出,6 类工件均被准确检测且置信度均有所提升,表明引入注意力机制有益于避免互相遮挡工件的错检和漏检情况。综合消融实验结果表明:所提各改进模块对工件的旋转检测精度和速度均有所提升,能够很好应用于实际工业生产。

4.4.3 可行性分析

部分检测结果可视化如图 12 所示,图 12(a)、(b)

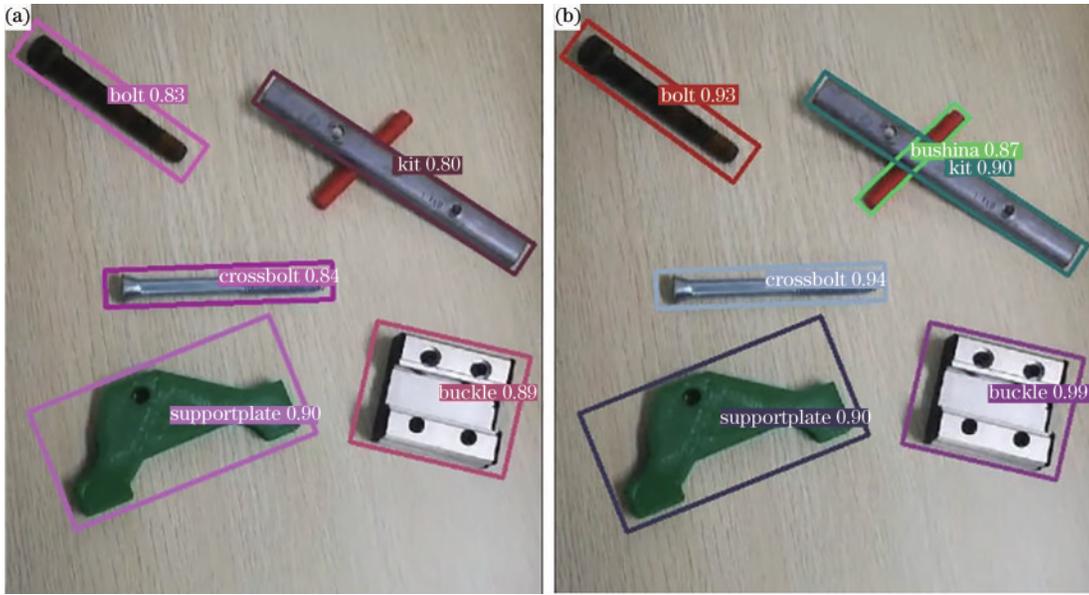


图 11 2 种模型检测对比。(a)模型D检测图;(b)模型E检测图
Fig. 11 Comparison of two models. (a) Test plots of model D; (b) test plots of model E

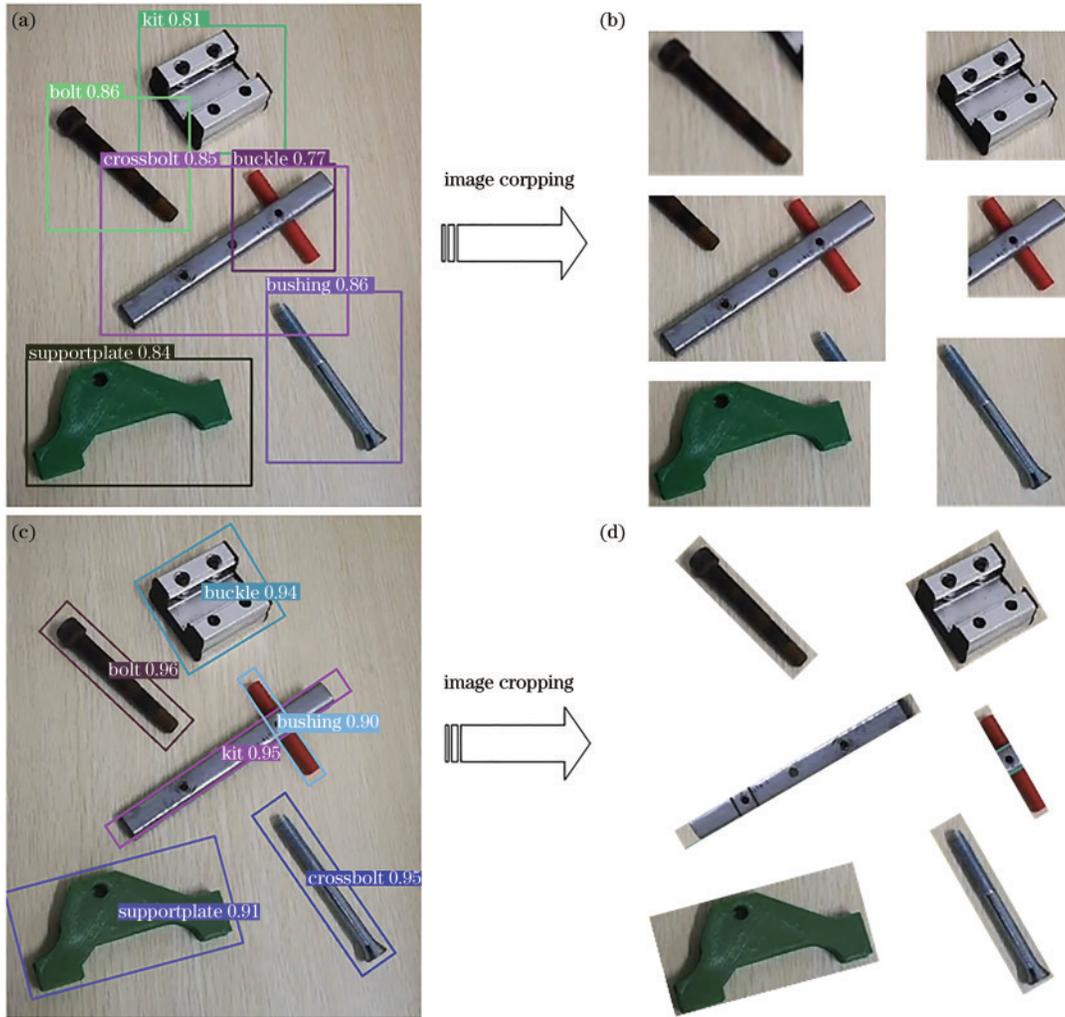


图 12 2 种检测方法裁剪示意图。(a)(b)YOLOv5检测图及图像裁剪;(c)(d) GB-FRN-YOLOv5检测图及图像裁剪
Fig. 12 Cropping schematic of two detection methods. (a)(b)YOLOv5 detection map and image cropping; (c)(d) detection map and image cropping of GB-FRN-YOLOv5

分别为 YOLOv5 检测图和图像裁剪图,图 12(c)、(d) 分别为 GB-FRN-YOLOv5 检测图和图像裁剪图。对于杂乱放置的距离相近或者互相遮挡的工件,图 12(a)中不同工件的水平检测框包含其他工件,且含背景噪声较大,经过裁剪获得的图 12(b)中单个工件图像仍包含其他工件的部分结构,这使得图像处理无法准确求取单个工件的质心,影响工件抓取点定位。而图 12(c)使用旋转检测框,很好地贴合了工件的轮廓,同时经裁剪获得的图 12(d)中每个图像只包

含一个工件,且覆盖像素区域更小,减小了背景环境对图像处理过程中的影响,提高了求取工件质心的效率。

经过检测图像的裁剪,以工件螺栓为例,对单工件图像通过图像处理求取质心过程如图 13 所示。其中,图 13(e)、(f)为二值化取反(背景为黑色)并求取质心图。从图像处理过程可看出,经旋转检测框裁剪后的单幅工件图像背景噪声少,易于图像处理获取工件轮廓并求取质心。同时所求工件质心为工件的质量平衡点,作为工件的抓取点更为可靠。



图 13 图像处理求取工件质心。(a)原图;(b)灰度化;(c)中值滤波;(d)二值化;(e)二值化取反;(f)质心求取

Fig. 13 Image processing to obtain workpiece centroid. (a) Original image; (b) grayscale; (c) median filtering; (d) binarization; (e) inversion of binarization; (f) centroid calculation

5 结 论

从工业生产线随意堆放工件的抓取任务出发,解决存在遮挡、杂乱放置的工件检测与抓取点定位问题。针对工业场景中多工件的检测与抓取任务,提出一种协同深度学习的两阶段多工件检测与抓取点定位算法。首先,对 YOLOv5 算法进行改进,增加旋转检测网络,实现任意方向工件的旋转检测,使以更贴合工件轮廓的网络形式输出,减少背景噪声,并引入轻量化模块和注意力机制等改进策略,提高检测算法的检测速度以及检测精度,对于互相遮挡的工件能够精准检测,相比于现有的旋转检测算法,性能更优;其次,所提 GB-FRN-YOLOv5 算法结合图像处理求取工件质心作为抓取点,两种算法结合能够使工件检测与抓取网络更加轻量化、更适合高速的工业生产;最后,以质心作为工件抓取点,对于结构复杂的工件更能保证可靠抓取,具有实际应用价值。

参 考 文 献

- [1] 解则晓, 陈文柱, 迟书凯, 等. 基于结构光视觉引导的工业机器人定位系统[J]. 光学学报, 2016, 36(10): 1015001. Xie Z X, Chen W Z, Chi S K, et al. Industrial robot positioning system based on the guidance of the structured-light vision[J]. Acta Optica Sinica, 2016, 36(10): 1015001.
- [2] 姚锡凡, 雷毅, 葛动元, 等. 驱动制造业从“互联网+”走向“人工智能+”的大数据之道[J]. 中国机械工程, 2019, 30(2): 134-142. Yao X F, Lei Y, Ge D Y, et al. On big data driving manufacturing from “Internet plus” to “AI plus” [J]. China Mechanical Engineering, 2019, 30(2): 134-142.
- [3] Wang Y, Fu W P, Zheng C S, et al. Image matching for workpiece based on genetic algorithm[C]//2009 International Conference on Artificial Intelligence and Computational Intelligence, November 7-8, 2009, Shanghai, China. New York: IEEE Press, 2009: 152-157.
- [4] Li D, Zhang X, Jiang T, et al. Application of improved wavelet threshold de-noising algorithm in work piece inspection[C]//2013 6th International Congress on Image and Signal Processing (CISP), December 16-18, 2013, Hangzhou, China. New York: IEEE Press, 2013: 256-260.
- [5] 华夏, 王新晴, 王东, 等. 基于改进 SSD 的交通大场景多目标检测[J]. 光学学报, 2018, 38(12): 1215003. Hua X, Wang X Q, Wang D, et al. Multi-objective detection of traffic scenes based on improved SSD[J]. Acta Optica Sinica, 2018, 38(12): 1215003.
- [6] 王文秀, 傅雨田, 董峰, 等. 基于深度卷积神经网络的红外船只目标检测方法[J]. 光学学报, 2018, 38(7): 0712006. Wang W X, Fu Y T, Dong F, et al. Infrared ship target detection method based on deep convolution neural network[J]. Acta Optica Sinica, 2018, 38(7): 0712006.
- [7] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, OH, USA. New York: IEEE Press, 2014: 580-587.
- [8] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-29, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 779-788.

- [9] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9905: 21-37.
- [10] 王博. 基于视觉引导的汽车小零件定位抓取技术研究[D]. 秦皇岛: 燕山大学, 2019: 43-62.
Wang B. Positioning and grasping technology of small parts of automobile based on visual guidance[D]. Qinhuangdao: Yanshan University, 2019: 43-62.
- [11] Gong W, Zhang K B, Yang C W, et al. Adaptive visual inspection method for transparent label defect detection of curved glass bottle[C]//2020 International Conference on Computer Vision, Image and Deep Learning (CVIDL), July 10-12, 2020, Chongqing, China. New York: IEEE Press, 2020: 90-95.
- [12] Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08)[2019-12-05]. <https://arxiv.org/abs/1804.02767>.
- [13] Li Y Z, Yan M D, Liu X. Workpiece intelligent identification and positioning system based on binocular machine vision[C]//2021 IEEE 9th International Conference on Computer Science and Network Technology, October 22-24, 2021, Dalian, China. New York: IEEE Press, 2021: 55-58.
- [14] 雷毅. 基于深度学习的工件检测与双机械臂协作分拣研究[D]. 广州: 华南理工大学, 2020: 13-30.
Lei Y. Research on workpiece detection based on deep learning and collaborative sorting with dual manipulators [D]. Guangzhou: South China University of Technology, 2020: 13-30.
- [15] Ma J Q, Shao W Y, Ye H, et al. Arbitrary-oriented scene text detection via rotation proposals[J]. IEEE Transactions on Multimedia, 2018, 20(11): 3111-3122.
- [16] Zhang G J, Lu S J, Zhang W. CAD-Net: a context-aware detection network for objects in remote sensing imagery[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(12): 10015-10024.
- [17] Yang X, Yan J C, Feng Z M, et al. R³Det: refined single-stage detector with feature refinement for rotating object[EB/OL]. (2019-08-15)[2020-02-21]. <https://arxiv.org/abs/1908.05612v6>.
- [18] 梁霄, 李家炜, 赵小龙, 等. 基于深度学习的红外目标成像液位检测方法[J]. 光学学报, 2021, 41(21): 2110001.
Liang X, Li J W, Zhao X L, et al. Infrared target imaging liquid level detection method based on deep learning[J]. Acta Optica Sinica, 2021, 41(21): 2110001.
- [19] Han K, Wang Y H, Tian Q, et al. GhostNet: more features from cheap operations[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 1577-1586.
- [20] Li X, Wang W H, Hu X L, et al. Selective kernel networks[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 510-519.
- [21] Wang Q L, Wu B G, Zhu P F, et al. ECA-Net: efficient channel attention for deep convolutional neural networks [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 11531-11539.
- [22] 张广军. 机器视觉[M]. 北京: 科学出版社, 2005: 56-58.
Zhang G J. Machine vision[M]. Beijing: Science Press, 2005: 56-58.
- [23] Xu Y C, Fu M T, Wang Q M, et al. Gliding vertex on the horizontal bounding box for multi-oriented object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(4): 1452-1459.