

DOI: 10.3785/j.issn.1008-973X.2022.01.002

旋转框定位的多尺度再生物品目标检测算法

董红召, 方浩杰, 张楠

(浙江工业大学 智能交通系统联合研究所, 浙江 杭州 310014)

摘要: 针对传统目标检测算法未考虑实际分拣场景目标物形态尺度的多样性, 无法获取旋转角度信息的问题, 提出基于 YOLOv5 的改进算法 MR²-YOLOv5. 通过添加角度预测分支, 引入环形平滑标签 (CSL) 角度分类方法, 完成旋转角度精准检测. 增加目标检测层用于提升模型不同尺度检测能力, 在主干网络末端利用 Transformer 注意力机制对各通道赋予不同的权重, 强化特征提取. 利用主干网络提取到的不同层次特征图输入 BiFPN 网络结构中, 开展多尺度特征融合. 实验结果表明, MR²-YOLOv5 在自制数据集上的均值平均精度 (mAP) 为 90.56%, 较仅添加角度预测分支的 YOLOv5s 基础网络提升 5.36%; 对于遮挡、透明、变形等目标物, 均可以识别类别和旋转角度, 图像单帧检测时间为 0.02~0.03 s, 满足分拣场景对目标检测算法的性能需求.

关键词: 再生物品检测; YOLOv5; 旋转框检测; 环形平滑标签; 特征金字塔; 注意力机制

中图分类号: TP 391; X 705 **文献标志码:** A **文章编号:** 1008-973X(2022)01-0016-10

Multi-scale object detection algorithm for recycled objects based on rotating block positioning

DONG Hong-zhao, FANG Hao-jie, ZHANG Nan

(ITS Joint Research Institute, Zhejiang University of Technology, Hangzhou 310014, China)

Abstract: An improved algorithm MR²-YOLOV5 based on YOLOv5 was proposed aiming at the problem that the traditional target detection algorithm did not consider the diversity of the target shape scale in the actual sorting scene and could not obtain the rotation angle information. Precise rotation angle detection was completed by adding angle prediction branches and introducing angle classification method of ring smooth label (CSL). The target detection layer was added to improve the detection ability of different scales of the model. Transformer attention mechanism was used at the end of the backbone network to give different weights to each channel and strengthen feature extraction. The feature graphs of different levels extracted from the backbone network were input into the BiFPN network structure to conduct multi-scale feature fusion. The experimental results showed that the mean average precision (mAP) of MR²-YOLOV5 on the self-made data set was 90.56%, which was 5.36% higher than that of YOLOv5s with only angle prediction branch. Categories and rotation angles can be recognized for objects such as occlusion, transparent and deformation. The detection time of single frame is 0.02-0.03 s, which meets the performance requirements of target detection algorithm for sorting scenes.

Key words: detection of recycled goods; YOLOv5; rotating frame detection; circular smooth label; feature pyramid; attentional mechanism

利用机器视觉与自动化设备协同配合代替人工流水线再生物品分拣成为发展趋势^[1-2]. 瓶罐状物通常由塑料、玻璃、金属等材质制成, 外形均是柱状体, 分拣时易出现误检. 同种材质瓶罐状物

具有多样化形态结构和多尺度的外形尺寸, 易发生变形、遮挡的问题, 透明度较高时易与分拣背景混淆, 目标物外形特征提取难度加大. 统目标检测算法受特定参数约束条件, 降低了分拣对象

收稿日期: 2021-11-25.

网址: www.zjujournals.com/eng/article/2022/1008-973X/202201002.shtml

基金项目: 国家自然科学基金资助项目 (61773347); 浙江公益技术研究项目 (LGF19F030001).

作者简介: 董红召 (1969—), 男, 教授, 从事智能交通、智能机电系统的研究. orcid.org/0000-0001-5905-597X. E-mail: its@zjut.edu.cn

检测的准确性,易出现漏检、误检.利用目标检测算法获取平面坐标、旋转角度等位姿信息,结合激光束、接触传感器等确定执行机构下降高度,在满足机械手精准抓取的前提下,降低传统识别旋转角度设备的成本.

在再生物品图像识别的问题上,诸多学者展开研究. Yang 等^[3]提出轻量级神经网络 WasNet. 陈智超等^[4-5]提出 MobileNet 垃圾图像分类改进算法. Nie 等^[6]结合 Faster R-CNN 和 ResNet50,完成垃圾图像识别、检测. Liang 等^[7]提出整合高、低层次的特征图,设计复合层 CNN 网络. 上述模型无法获取目标物旋转角度信息,缺乏考虑同类物体多形态、多尺度问题,缺少单一图像中多类目标的检测. 周滢慙^[8]改进 Faster-RCNN 目标检测模型,调整原始 VGG16 分类网络,实现瓶子位置和旋转角度的识别. 数据集中瓶子的形态和尺度单一,缺乏考虑目标物之间遮挡、易发生变形等情况. 综上所述,现有再生物品目标检测算法无法同时满足实际分拣场景下位姿信息获取以及多形态、多尺度目标物的检测.

YOLO(you only look once)算法由 Redmon 等^[9]提出, YOLOv5 利用 Mosaic 数据增强、自适应锚框计算和图像缩放等方法,提高模型整体鲁棒性. Liu 等^[10]比较 YOLO 系列模型后表明, YOLOv5 在 mAP 和检测速度方面具有较大优势. YOLOv5 多尺度自适应锚框计算实现多尺度检测,符合再生物品实际分拣场景的性能需求,但缺乏角度预测.

诸多学者提出解决目标物角度预测问题的方法. 朱煜等^[11]提出粗调与细调两阶段旋转框检测网络 R²-FRCNN. Ding 等^[12]提出 ROI-Transformer 检测模型. Xu 等^[13]提出学习水平框转换为旋转框后 4 个点的偏移. Chen 等^[14]引入两阶段 NMS 算法,降低密集对象漏检率. Yang 等提出旋转检测器 SCRDet^[15]、R³Det^[16]. 上述基于回归方法实现旋转框检测,存在角度周期性(PoA)和边缘的交换性(EoE)问题. 针对该问题, Yang 等^[17]提出将物体角度预测视为分类问题,且为限制预测结果,设计环形平滑标签(circular smooth label, CSL),但该方法多用于遥感图像检测,无法直接迁移至分拣场景下的再生物品目标物角度检测.

结合上述问题与方法,提出基于 YOLOv5 旋转框定位的多尺度目标检测算法 MR²-YOLOv5 (multi-scale refined rotated YOLOv5),用于再生物品图像检测. 从以下 4 个方面对模型进行改进:

1)将角度预测作为分类问题,引入环形平滑标签(CSL)角度分类方法,用于对目标物角度的精准预测;2)将检测层增加到 4 层,提升对不同尺度目标尤其是小目标的检测性能;3)在主干网络末端利用 Transformer 模块替换 YOLOv5 的 C3 模块的模型,加强特征提取;4)将原始 PANet 网络结构改进为 BiFPN,进行多尺度的特征融合,以解决再生物品分拣场景下物品结构、尺度多样而导致的高漏检率、误检率的问题. 在形态尺度多样化的瓶罐状物自制数据集下,验证模型的各项改进方法.

1 MR²-YOLOv5 模型

YOLOv5 模型可以通过修改配置文件中 d_m 和 w_m 2 个参数直接影响整体模型通道数以及主干网络瓶颈层模块数量,在不改变整体模型结构的前提下,产生 4 个具有梯度化性能的模型,分别是 YOLOv5s、YOLOv5m、YOLOv5l、YOLOv5x. MR²-YOLOv5 模型的整体结构如图 1 所示,各模块结构如图 2 所示. 图中, m 为输出特征图通道数. MR²-YOLOv5 由主干网络、加强特征提取网络、检测网络 3 部分组成. 其中 Conv 模块中激活函数选择 Swish 函数,将 CIoU 损失函数作为边界框回归损失, CIoU 损失函数定义如下:

$$\left. \begin{aligned} \text{IoU} &= \frac{|A \cap B|}{|A \cup B|}, \\ L_{\text{CIoU}} &= 1 - \text{IoU} + \frac{\rho^2(\mathbf{b}, \mathbf{b}^{\text{gt}})}{c^2} + \alpha v, \\ v &= \frac{4}{\pi^2} \left(\arctan \frac{w^{\text{gt}}}{h^{\text{gt}}} - \arctan \frac{w}{h} \right)^2. \end{aligned} \right\} \quad (1)$$

式中: IoU 为预测框与真实框相交区域与交并区域面积的比值, A 、 B 分别为 2 个边界框的面积, \mathbf{b} 、 \mathbf{b}^{gt} 分别为预测框中心点坐标向量和真实框中心点坐标向量, $\rho(\mathbf{b}, \mathbf{b}^{\text{gt}})$ 为 2 个边界框中心点的欧氏距离, c 为预测框与真实框形成最小矩形区域的对角线长度, α 为权重函数, v 为预测框和真实框长宽比的相似度, w 、 h 为框的宽与高, gt 表示真实框(ground truth).

1.1 增加角度预测分支

针对机器人分拣过程中对于位姿信息的需求,在原模型的基础上添加角度预测分支. 由于长宽比较大的目标对于角度变化十分敏感,目前主流的角度回归方法均存在不连续边界问题,使

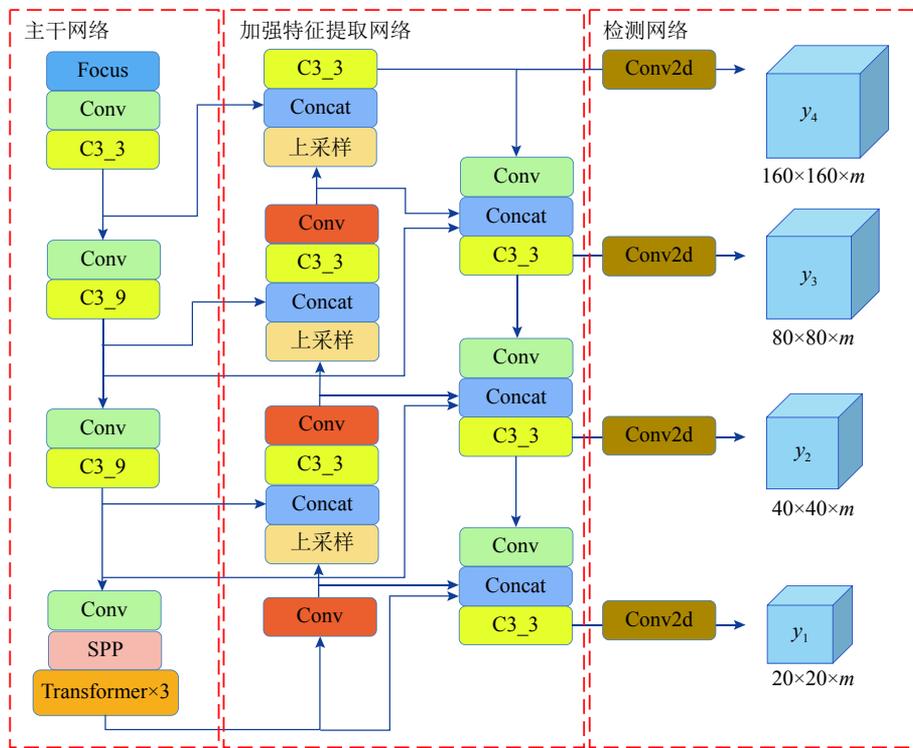


图 1 MR²-YOLOv5 网络结构

Fig.1 MR²-YOLOv5 network structure

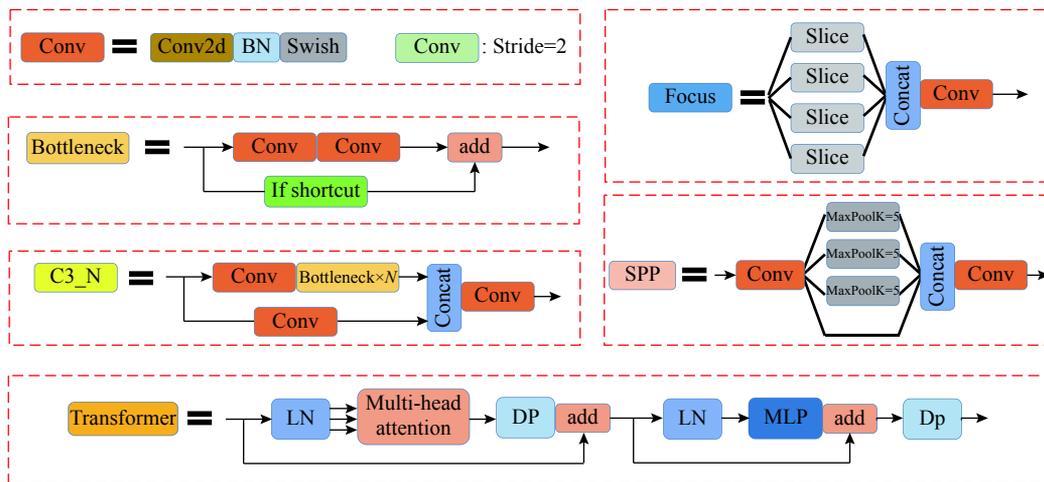


图 2 MR²-YOLOv5 各模块结构

Fig.2 MR²-YOLOv5 structure of each module

模型损失值在处于边界情况下出现突增, 导致该类检测方法在边界条件下检测准确率低^[15]. 将目标物角度预测视为分类问题, 设计符合再生物品分拣物理场景的环形平滑标签(CSL), 增加相邻角度之间的误差容忍度. 修改后的 CSL 可视化表达如图 3 所示, CSL 的具体表达式为

$$CSL(x) = \begin{cases} g(x), & \theta - r < x < \theta + r; \\ 0, & \text{其他.} \end{cases} \quad (2)$$

式中: $g(x)$ 为窗口函数(本次选择高斯函数), r 为

窗函数的半径, θ 为当前边界框的角度.

本次 CSL 角度为 $[0^\circ, 180^\circ)$, 由于角度呈环形分布, 0° 与 180° 结果相同. 旋转框参数表示为 (r_x, r_y, L, S, θ) , 其中 r_x 和 r_y 为目标中心点的横、纵坐标, L 为目标框的较长边, S 为目标框的较短边. 检测图片左上角作为原点, θ 定义为 X 轴顺时针旋转至与目标框长边平行所经过的角度, θ 的定义如图 4 所示. 结合目前 YOLOv5 模型中置信度损失及类别损失的计算方法, 采用二分类交叉熵损失函数 BCEWithLogitsLoss 对角度损失进行计

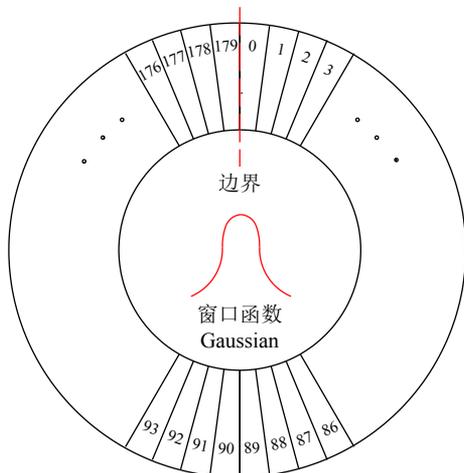


图 3 CSL 示意图

Fig.3 CSL schematic diagram

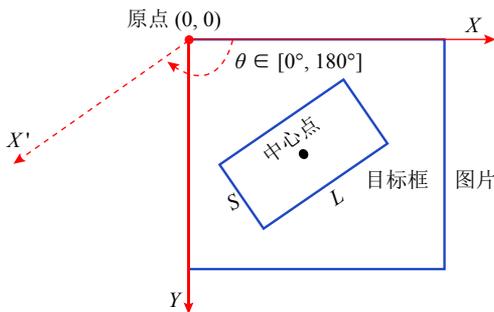


图 4 角度定义示意图

Fig.4 Angle definition diagram

算. 具体定义如下:

$$\left. \begin{aligned} \text{loss}(z, y) &= \text{mean}\{l_0, \dots, l_{N-1}\}, \\ l_n &= \text{sum}\{l_{n,0}, l_{n,1}, \dots, l_{n,179}\}, \\ l_{n,i} &= -[y_{n,i} \ln(\delta(z_{n,i})) + (1 - y_{n,i}) \ln(1 - \delta(z_{n,i}))], \\ y_{n,i} &= \text{CSL}(x). \end{aligned} \right\} \quad (3)$$

式中: N 为样本数量; $i \in [0, 180)$, 共计 180 个类别; δ 为 sigmoid 函数; $z_{n,i}$ 为预测第 n 个样本为第 i 个角度的概率, $z_{n,i}$ 的最大值为 1, 即预测值; $y_{n,i}$ 为第 n 个样本在 $\text{CSL}(x)$ 表达式下第 i 个角度的标签值, 即真实值, $y_{n,i}$ 的最大值为 1. 将第 n 个样本的各个角度预测值与真实值依次代入 $l_{n,i}$ 计算, 第 n 个样本 180 个角度结果求和得到 l_n , 将 N 个样本的计算结果求平均, 即为此次角度损失.

检测网络参数 m 为输出特征图通道数, 具体定义如下:

$$m = (n_c + 5 + \omega)n_a. \quad (4)$$

式中: n_c 为检测类别数; n_a 为锚框数量; 5 表示 5 个参数, 分别指旋转框参数中的 r_x 、 r_y 、 L 、 S 及置信度; ω 为角度分类数量.

此次检测类别数为 3, 锚框数量为 3, 角度 θ 共

分为 180 个类别, 则 $\omega = 180$, 可知输出特征图通道数 $m = (3 + 5 + 180) \times 3 = 564$.

1.2 Transformer 模块

将 Transformer 模块融入 YOLOv5 模型, 可以提升高密度遮挡目标物的检测性能^[18]. Transformer 特征提取结构由 2 部分组成, 分别是编码器 (encoder) 和解码器 (decoder), 由于仅进行图像识别, 只须利用编码器进行特征提取. 利用主干网络对于图像进行特征提取, 结合位置编码将其转换为序列, 作为 Transformer 编码器的输入. Transformer 编码器主要包括 2 个主要模块: 多头注意力机制模块 (multi-head attention) 和前馈神经网络 (MLP). LN (LayerNorm) 层和 DP (Dropout) 层可以防止网络过拟合, 提高特征融合^[19].

Transformer 网络结合多个缩放点积注意力机制, 形成多头注意力机制. 利用缩放点积注意力机制可以批量处理数据, 表达式^[20]为

$$\text{attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V}. \quad (5)$$

式中: \mathbf{Q} 、 \mathbf{K} 、 \mathbf{V} 分别为查询向量、键向量、值向量, d_k 为输入数据维度.

多头注意力机制具有更高的运算效率, 可以从不同的特征空间中学习自适应注意力分布, 表达式为

$$\left. \begin{aligned} \mathbf{h}_i &= \text{attention}(\mathbf{Q}\mathbf{W}_i^Q, \mathbf{K}\mathbf{W}_i^K, \mathbf{V}\mathbf{W}_i^V), \\ \mathbf{H} &= \text{Concat}(\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_n). \end{aligned} \right\} \quad (6)$$

式中: 参数矩阵 $\mathbf{W}_i^Q \in \mathbf{R}^{d_{\text{mod}} \times d_q}$, $\mathbf{W}_i^K \in \mathbf{R}^{d_{\text{mod}} \times d_k}$, $\mathbf{W}_i^V \in \mathbf{R}^{d_{\text{mod}} \times d_v}$, 其中 d_{mod} 为向量维度.

MR²-YOLOv5 主干网络末端利用 3 次 Transformer 网络运算替换 YOLOv5 原始的 C3 模块, 该运算次数与 d_m 参数相关联.

1.3 增加检测层

过多的下采样层会导致图像中的小目标特征消失, 导致网络无法学习到相应的特征, 出现漏检、误检的情况. 原始 YOLOv5 网络的检测层尺度分别为 20×20 、 40×40 、 80×80 , MR²-YOLOv5 保留原始检测尺寸, 不降低原始尺度特征的学习能力, 通过增加 Concat 层、卷积层、C3 模块的方法, 多进行了一次上采样和下采样, 将主干网络中保留较多小目标特征信息的同尺度特征层输入 Concat 层, 进行特征融合操作. 新增检测层对于小目标特征更敏感, 虽然计算成本增加, 但是模型检测性能提高. YOLOv5 具有自适应锚框计算, 会采用 k 均值和遗传算法对自定义数据集进行分析计

算, 获得该数据集各尺度所需的预测锚框. 最终的检测层尺度为 20×20 、 40×40 、 80×80 、 160×160 .

1.4 多尺度特征融合网络

随着网络层数的加深, 每一层网络都会在一定程度上造成特征丢失, 增加检测分支和融合多尺度特征融合网络可以提升检测精度^[21]. 目前已有的特征融合网络 FPN^[22]、PANet^[23]、NAS-FPN^[24]、BiFPN^[25] 等. 为了增强不同尺度的特征信息融合, 将原始加强特征提取网络 PANet 结构改进为 BiFPN 结构, 将主干网络中的多尺度特征图多次输入 BiFPN 结构中, 开展自上而下和自下而上的多尺度特征融合; 对于同尺度大小的特征层, 利用跳跃连接的方式进行特征融合. MR²-YOLOv5 的 BiFPN 自下而上的部分对 2 个同尺度特征进行融合, 自上而下的结构对 3 个同尺度特征进行融合, 在经过 C3 模块以及 Conv2d 模块计算后, 对于 4 个不同的尺度目标进行预测. 利用 BiFPN 网络, 使得预测网络对不同大小的目标更敏感, 提升了整体的模型检测能力, 降低了漏检率与误检率.

2 实验结果与分析

基于 pytorch 深度学习框架构建 MR²-YOLOv5 模型, 硬件配置如下: i7-8700K CPU, GTX1080Ti, 32 GB 内存. 将 YOLOv5s 网络作为整体研究的基础网络 (即 $d_m = 0.33$, $w_m = 0.50$), 可以大幅度减少后续消融实验的训练时间, 提高计算效率. 在添加角度预测分支后, 模型参数量为 7.55×10^6 , 计算量为 1.8×10^{10} FLOPs. 采用 CSL 角度分类方法, θ 为 $[0^\circ, 180^\circ)$.

2.1 实验数据集制作及训练参数设置

现有研究均由自身搭建实验平台利用工业相机进行拍摄, 但未考虑不同材质物体具有相似特征以及同材质物体具有多形态结构和多尺度外形的特点. 在实际的流水线分拣过程中, 分拣物在流水线上的状态多变, 随机性强, 密度较高时易导致复杂背景, 检测难度加大. 复杂背景与单一背景相比而言, 当单个目标物处于单色的输送带或平台时, 可以认为该目标物处于单一背景; 当多目标物处于单色的输送带或平台时, 由于目标物均带有颜色与形状特征, 目标检测时容易相互干扰, 可以认为目标物处于复杂背景.

针对上述情况, 以瓶状物、罐状物为研究对象, 搜集不同材质物品及同材质多种形态结构物品, 模拟实际分拣物的多种随机性状况 (竖放、平

放、遮挡、弯曲变形等), 利用移动端设备进行拍摄. 目标物的拍摄角度尽可能与分拣场景相同 (正上方拍摄), 拍摄场景与拍摄距离多样化, 单张图片内可以包含多个目标物, 使得数据集中的目标角度覆盖 $[0^\circ, 180^\circ)$. 拍摄场景选择为生活场景, 生活场景下的图片背景较输送带或平台背景更复杂. 数据集共计 3 类再生物品, 分别是金属罐状物 (C_s)、塑料制品瓶装物 (P_b)、玻璃制品瓶装物 (G_b). 这 3 类物体的长宽比较大, 对角度检测更敏感; 外形具有相似特征, 较多都是柱状物, 塑料制品和玻璃制品的透明度较高, 易与背景混淆, 更能体现模型的性能. 数据集样本分布如下: C_s 、 P_b 、 G_b 的样本数分别为 1785、2360、2589. 数据集共包含 6064 幅分辨率为 1280×720 至 1920×1080 的可见光图像, 其中 4964 张作为训练集、550 张作为验证集、550 张为测试集, 且较多图片中包含多类物体.

利用 rolabeling 对数据集图片进行旋转框标注, 图片左上角为原点, 将上述 3 类物品的瓶口或开口左侧作为标注起始点, 可视化处理后如图 5 所示. 标签参数定义为 $(C, r_x, r_y, L, S, \theta)$, 即在旋转框参数定义的基础上增加类别 ID 参数 C . 将标注文件 (xml 格式) 转换为所需的标签文件 (txt 格式), 将 r_x 、 r_y 、 L 、 S 基于图像自身宽高进行归一化处理, 多尺度数据集标签的尺寸分布情况如图 6 所示. 可知, 数据集标签中存在大量的小目标物体以及较多的大目标物体, 外形尺度跨度大. 对于传入数据, 采用 Mosaic 数据增强方法, 即对于传入图像进行随机翻转、缩放、裁剪等处理, 将得到的图像进行依次拼接, 大幅度提高了训练时目标背景的复杂度, 使用 Mosaic 数据增强方法后的输入图像如图 7 所示. 模型训练采用 SGD 优化算法, 参数设置如下: 批量大小为 8; 最大迭代次数为



图 5 旋转框标注示意图 (圆点为标注起始点)

Fig.5 Rotation frame annotation schematic diagram (dot is starting point of annotation)

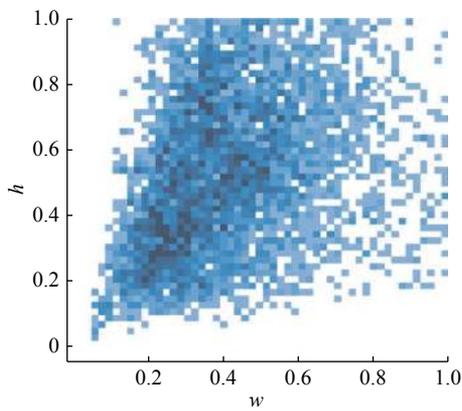


图 6 数据集标签分布情况

Fig.6 Data set label statistics and distribution



图 7 Mosaic 数据增强方法后的输入图像

Fig.7 Input image after Mosaic data enhancement method

200; 图片输入尺寸为 640×640; 采用余弦退火策略动态调整模型训练的学习率, 初始学习率设置为 0.01.

2.2 实验结果分析

基于 YOLOv5 模型添加角度预测分支, 采用 CSL 法对角度进行分类, 其中窗口半径是影响整体模型性能的重要参数之一. 当窗口半径过小时, 窗口函数趋向于脉冲函数, 无法学习方向及比例信息. 当窗口半径过大时, 预测结果的区分度过小. 针对上述问题, 选择 0~8 的合适半径, 如表 1 所示为基于 YOLOv5s 模型在各半径尺寸下的检测性能. 表中, mAP 为均值平均精度. 可知, 基于自制数据集, 当 $r=4$ 时, YOLOv5s 模型的检测性能最佳.

在 YOLOv5s 模型的基础上增加角度预测分支 ($r=4$) 以及其余参数设置相同的情况下, 对当前主流多尺度特征融合网络 FPN、PANet、Bi-FPN 进行检测精度和参数量 M 比较. 从表 2 可知,

传统 FPN 结构由于只进行单方向的特征融合, 预测精度偏低; PANet 结构采用简单的双向融合, 在 FPN 基础上增加了从下而上的融合路径, 明显提升了预测精度; Bi-FPN 双向特征金字塔网络结构更复杂, 将不同的特征层进行反复叠加, 且 Bi-FPN 网络更侧重于多尺度特征融合, 检测精度优于其他网络.

表 1 不同窗口半径下的检测性能比较

Tab.1 Comparison of detection performance under different window radii

模型	mAP/%				
	$r=0$	$r=2$	$r=4$	$r=6$	$r=8$
YOLOv5s	68.79	82.47	84.20	81.20	79.94

表 2 多尺度特征融合网络性能的比较

Tab.2 Performance comparison of multi-scale feature fusion network

多尺度融合网络	mAP/%	$M/10^6$
FPN	81.05	6.52
PANet	84.20	7.55
Bi-FPN	85.36	7.83

为了分析各模块在 MR²-YOLOv5 网络中的作用, 开展消融实验(见表 3), 研究各改进方法对 YOLOv5 模型的影响. 表 3 中, “-”表示不使用, “√”表示使用. 各模型均采用基于 COCO 数据集训练的 YOLOv5s 预训练权重, 可以利用预训练权重进行初始化, 避免过拟合现象, 加速模型收敛.

与未添加角度预测分支网络相比, 添加各改进方法后的模型整体参数量和计算量均有不同程度的提升, 但检测精度大幅提高, mAP 达到 94.46%, 较 YOLOv5s 初始网络提高 5.41%. 可见, 各改进方法可以提升模型检测多形态、多尺度目标物的性能. 其中模型对塑料制品瓶状物的检测性能提升最明显, AP 提升了 8.0%, 金属罐状物与玻璃制品瓶状物的 AP 均提升了 4.5%.

与添加角度预测分支网络相比, 将检测层增加到 4 层后, 增加了不同尺度的预测网络, 与 3 层检测网络相比, mAP 提高了 2.33%, 但计算量增加了 85.6%. 主干网络末端将原来的 C3 模块替换为 Transformer 模块, 利用注意力机制对输入数据进行并行特征提取, 对各通道赋予不同的权重, 强化特征提取, mAP 提高了 1.91%, 参数量提高了 13.5%, 但计算量下降了 2.4%. 将多尺度特征融合

表 3 基于 YOLOv5s 模型的消融实验

Tab.3 Ablation experiment based on YOLOv5s model

角度预测分支(CSL)	检测层	Transformer	融合网络	$M/10^6$	FLOPs/ 10^9	AP			mAP/%
						C_s	G_b	P_b	
-	3	-	PANet	7.28	17.1	91.7	90.1	85.4	89.05
-	4	-	PANet	8.17	28.2	94.5	92.4	87.6	91.34
-	4	√	PANet	8.17	28.0	95.4	94.1	90.2	93.23
-	4	√	BiFPN	9.25	29.6	95.2	94.7	93.4	94.46
√	3	-	PANet	7.55	18.0	90.5	82.5	79.1	85.20
√	4	-	PANet	8.73	33.4	92.1	85.3	82.2	86.53
√	4	√	PANet	9.91	32.6	93.8	87.5	83.9	88.44
√	4	√	BiFPN	10.99	34.2	96.5	91.2	84.2	90.56

网络改进为 BiFPN, 利用跳跃连接的方式, 连接底层、高层特征, 将多个同尺度的特征进行融合, 提高了模型整体的特征学习能力, 虽然参数量和计算量都小幅增大, 但 mAP 提高了 2.12%, 整体模型的 mAP 达到 90.56%, 与仅添加角度预测分支的 YOLOv5s 模型相比, mAP 提高了 5.36%。在添加角度分支后, 利用各改进方法可以提升模型的性能, 且 MR²-YOLOv5 参数量为 10.99×10^6 , 在实际测试中检测单帧图片时间仅需 0.02~0.03 s, 满足分拣场景下对目标检测算法的性能需求。

对比各网络下的 3 种目标物检测精度可知, 塑料制品与玻璃制品的瓶状物的 AP 均较低, 尤其是塑料制品瓶状物的检测精度过低, 影响整体的 mAP。主要原因是两者的透明度均较高, 易与背景色重叠, 且外形相似, 易发生混淆, 塑料制品瓶状物易发生变形, 从而导致检测精度下降。在模型改进后, 对三者的检测性能均有不同程度的提高。

如图 8、9 所示分别为不含角度预测分支与含角度预测分支的各改进网络在瓶状物数据集上的训练结果曲线。图中, E 为训练的迭代次数, L 为训练损失。从训练曲线可知, 在添加角度分支后, 由于角度损失函数的加入, 各个模型训练损失收敛更加缓慢, 但随着模型的改进, 模型的收敛性提高。无角度预测分支的各模型收敛速度更快, 改进后模型的 mAP 与损失函数基本迭代至 40 次左右就趋向于平稳。如图 10 所示为添加角度预测分支后的各模型验证集角度损失曲线。由于窗口函数的设置, 模型可以衡量预测标签和真实标签之间的角度距离, 即在一定范围内越靠近真实值的预测值损失越小。从图 10 可知, 利用各改进方法可以提升角度预测精度。利用各改进方法, 可以加快模型损失收敛, 提升模型精度。如图 11~14 所示为 MR²-YOLOv5 的检测效果图, 显示参数 (从左至右) 分别为该目标物类别、置信度、旋转角度。可知, 该模型对于多尺度目标 (小目标至大

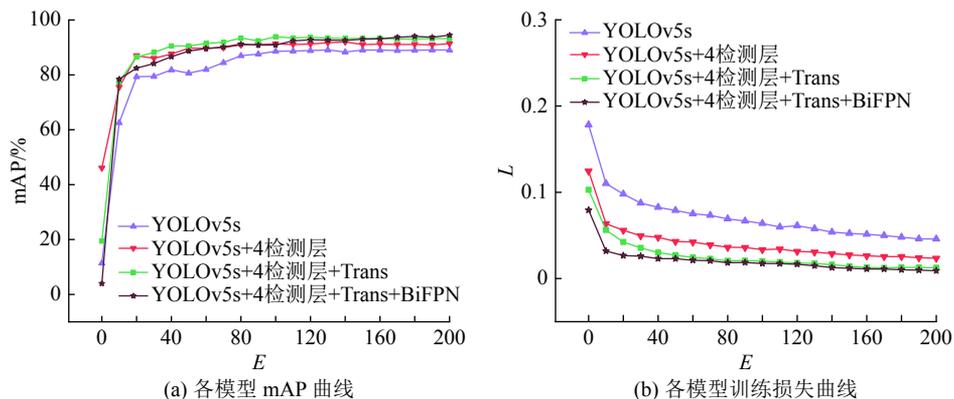


图 8 各模型训练曲线 (无角度预测分支)

Fig.8 Training curves of each model (without angle prediction branch)

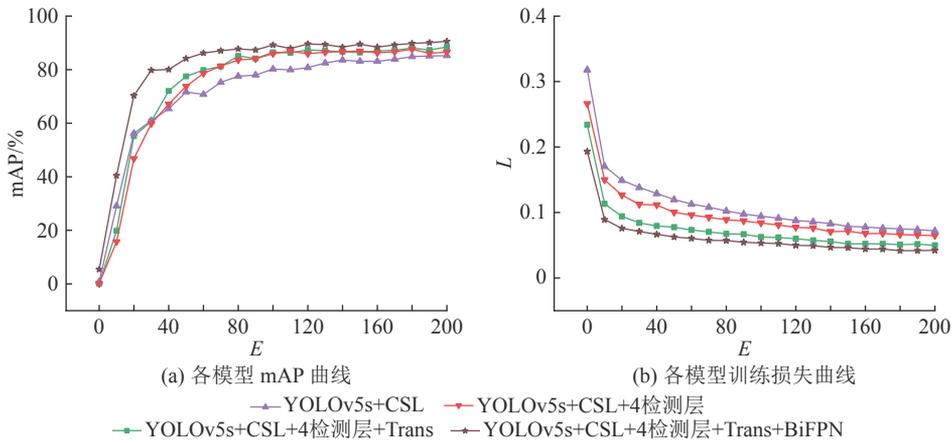


图 9 各模型训练曲线 (含角度预测分支)

Fig.9 Training curves of each model (including angle prediction branch)

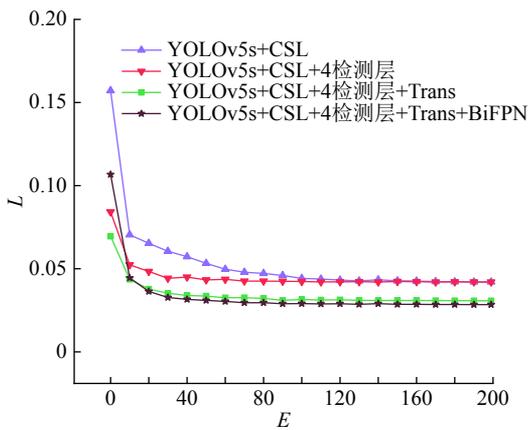


图 10 各模型验证集的角度损失曲线

Fig.10 Angle loss curves of each model validation set

目标), 目标物遮挡、透明度较高、发生变形等情况均有出色的识别效果, 可以检测旋转角度. 目标物发生遮挡、透明度较高、变形等情况, 外形特征不完整或提取难度加大. MR²-YOLOv5 可以提升此类情况下的目标物特征提取能力, 降低漏检率和误检率, 提高检测精度, 可以检测目标物旋转角度.

3 结 论

(1) 通过增加角度检测分支, 引入 CSL 角度分类方法, 完成对角度的精准预测. 通过实验, 比较在自制数据集及 YOLOv5s 模型时不同窗口半径下的模型检测性能. 实验结果表明, 当窗口半径为 4 时, 模型检测性能最佳.

(2) 以添加角度预测分支的 YOLOv5s 为基础, 比较各多尺度特征融合网络在自制数据集下的性能. 实验结果表明, BiFPN 网络结构更侧重于多尺度特征融合, 检测精度优于其他网络结构.

(3) 对添加角度预测分支、未添加角度预测分支的网络进行对比实验. 可知, 利用各改进方法均可以提升模型性能. 添加角度预测分支的 MR²-YOLOv5 在自制数据集下 mAP 达到 90.56%, 整体模型参数数量为 10.99×10⁶, 实际测试中检测单帧图片的时间仅为 0.02~0.03 s, 满足分拣场景下对目标检测算法检测图像的速度要求, 降低传统识别旋转角度设备的成本, 为再生物品分拣提供



图 11 不同尺度目标物检测效果

Fig.11 Detection effects of different scales



图 12 多目标物遮挡情况下的检测效果

Fig.12 Detection effect under occlusion of multiple objects



图 13 目标物透明度较高情况下的检测效果

Fig.13 Detection effect under high target transparency



图 14 目标物发生变形情况下的检测效果

Fig.14 Detection effect in case of deformation of target

新的思路。

(4)MR²-YOLOv5 在复杂场景下有较好的检测性能,但因为是二维图像检测,旋转角度检测仅针对平铺目标物,堆叠物品难以提供准确的位姿信息,可以结合三维图像提高位姿信息检测。当多个目标物重叠度过高时,会出现漏检情况。可以尝试融入边界多边形实例分割,提高模型多目标重叠时的检测性能,使模型更符合再生物品分拣的场景。

参考文献 (References):

[1] 康庄,杨杰,郭濠奇.基于机器视觉的垃圾自动分类系统设计

计[J].浙江大学学报:工学版,2020,54(7):1272-1280.

KANG Zhuang, YANG Jie, GUO Hao-qi. Automatic garbage classification system based on machine vision[J]. **Journal of Zhejiang University: Engineering Science**, 2020, 54(7): 1272-1280.

[2] 谢先武,熊禾根,陶永,等.一种面向机器人分拣的杂乱工件视觉检测识别方法[J].**高技术通讯**,2018,28(4):344-353.

XIE Xian-wu, XIONG He-gen, TAO Yong, et al. A method for visual detection and recognition of clutter workpieces for robot sorting[J]. **Chinese High Technology Letters**, 2018, 28(4): 344-353.

[3] YANG Z, LI D. Wasnet: a neural network-based garbage collection management system[J]. **IEEE Access**, 2020, 8:

- 103984–103993.
- [4] 陈智超, 焦海宁, 杨杰, 等. 基于改进 MobileNet v2 的垃圾图像分类算法 [J]. 浙江大学学报: 工学版, 2021, 55(8): 1490–1499.
CHEN Zhi-chao, JIAO Hai-ning, YANG Jie, et al. Garbage image classification algorithm based on improved MobileNet v2 [J]. *Journal of Zhejiang University: Engineering Science*, 2021, 55(8): 1490–1499.
- [5] 袁建野, 南新元, 蔡鑫, 等. 基于轻量级残差网络的垃圾图片分类方法 [J]. 环境工程, 2021, 39(2): 6.
YUAN Jian-ye, NAN Xin-yuan, CAI Xin, et al. Garbage image classification by lightweight residual network [J]. *Environmental Engineering*, 2021, 39(2): 6.
- [6] NIE Z, DUAN W, LI X. Domestic garbage recognition and detection based on Faster R-CNN [J]. *Journal of Physics: Conference Series*, 2021, 1738(1): 012089.
- [7] LIANG B, WANG Y, WANG Y, et al. Garbage sorting system based on composite layer cnn and multi-robots [J]. *Journal of Physics: Conference Series*, 2020, 1634(1): 012083.
- [8] 周滢懋. 基于机器视觉的生活垃圾智能分拣系统的设计与实现 [D]. 哈尔滨: 哈尔滨工业大学, 2018.
ZHOU Ying-min. Design and implementation of visionbased Sorting system for solid waste [D]. Harbin: Harbin Institute of Technology, 2018.
- [9] 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**. Las Vegas: IEEE, 2016: 779–788.
- [10] LIU K, TANG H, HE S, et al. Performance validation of Yolo variants for object detection [C]// **Proceedings of the 2021 International Conference on Bioinformatics and Intelligent Computing**. Vancouver: [s. n.], 2021: 239–243.
- [11] 朱煜, 方观寿, 郑兵兵, 等. 基于旋转框精细定位的遥感目标检测方法研究 [EB/OL]. [2021-10-01]. <http://www.aas.net.cn/article/doi/10.16383/j.aas.C200261>.
ZHU Yu, FANG Guan-shou, ZHENG Bing-bing, et al. Research on detection method of refined rotated boxes in remote sensing [EB/OL]. [2021-10-01]. <http://www.aas.net.cn/article/doi/10.16383/j.aas.C200261>.
- [12] DING J, XUE N, LONG Y, et al. Learning roi transformer for oriented object detection in aerial images [C]// **2019 IEEE Conference on Computer Vision and Pattern Recognition**. Long Beach: IEEE, 2019: 2849–2858.
- [13] XU Y, FU M, WANG Q, et al. Gliding vertex on the horizontal bounding box for multi-oriented object detection [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 43(4): 1452–1459.
- [14] CHEN Y, DING W, LI H, et al. Arbitrary-oriented dense object detection in remote sensing imagery [C]// **2018 IEEE 9th International Conference on Software Engineering and Service Science**. Beijing: IEEE, 2019: 436–440.
- [15] YANG X, YANG J, YAN J, et al. Scrdet: towards more robust detection for small, cluttered and rotated objects [C]// **2019 IEEE/CVF International Conference on Computer Vision**. South Korea: IEEE, 2019: 8232–8241.
- [16] YANG X, LIU Q, YAN J, et al. R³det: refined single-stage detector with feature refinement for rotating object [EB/OL]. [2021-10-01]. <https://arxiv.org/abs/1908.05612>.
- [17] YANG X, YAN J, HE T. On the arbitrary-oriented object detection: classification based approaches revisited [EB/OL]. [2021-10-01]. <https://arxiv.org/abs/2003.05597v3>.
- [18] ZHU X, LIU S, WANG X, et al. TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios [C]// **Proceedings of the IEEE/CVF International Conference on Computer Vision**. Montreal: IEEE, 2021: 2778–2788.
- [19] FU L, GU W, LI W, et al. Bidirectional parallel multi-branch convolution feature pyramid network for target detection in aerial images of swarm UAVs [J]. *Defence Technology*, 2021, 17(4): 1531–1541.
- [20] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]// **Advances in Neural Information Processing Systems**. Long Beach: [s. n.], 2017: 5998–6008.
- [21] 鲁博, 瞿绍军. 融合 BiFPN 和改进 YOLOv3-tiny 网络的航拍图像车辆检测方法 [J]. *小型微型计算机系统*, 2021, 42(8): 1694–1698.
LU bo, QU Shao-jun. Vehicle detection method in aerial images based on BiFPN and improved YOLOv3-tiny Network [J]. *Journal of Chinese Computer Systems*, 2021, 42(8): 1694–1698.
- [22] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C]// **2017 IEEE Conference on Computer Vision and Pattern Recognition**. Hawaii: IEEE, 2017: 2117–2125.
- [23] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation [C]// **2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition**. Salt Lake City: IEEE, 2018: 8759–8768.
- [24] GHIASI G, LIN T Y, LE Q V. Nas-fpn: Learning scalable feature pyramid architecture for object detection [C]// **2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition**. Long Beach: IEEE, 2019: 7036–7045.
- [25] TAN M, PANG R, LE Q V. EfficientDet: scalable and efficient object detection [C]// **2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition**. Seattle: IEEE, 2020: 10781–10790.