

羊森海, 陈丹. 基于改进 YOLOv7 的抓取图像小目标检测算法[J]. 智能计算机与应用, 2025, 15(7): 99–103. DOI: 10.20169/j. issn. 2095–2163. 250714

基于改进 YOLOv7 的抓取图像小目标检测算法

羊森海, 陈 丹

(福州大学 电气工程与自动化学院, 福州 350000)

摘 要: 针对工业化环境下用于抓取的小目标物体存在难以识别、检测精度低等问题, 本文提出一种基于改进 YOLOv7 的小目标检测算法。在 YOLOv7 模型的特征融合部分引入动态稀疏注意力机制 (BiFormer), 有效增加待检测小目标物体的细节信息, 提升网络模型的检测精度; 在检测部分引进递归特征金字塔模块 (RFP), 将特征金字塔结构输出的特征与自下而上的骨架层融合。实验结果表明, 改进后的 YOLOv7 模型性能得到了一定提升, 检测精确度提升 3.6 个百分点, 对于小目标物体的检测效果更佳。

关键词: 小目标检测; YOLOv7; 动态稀疏注意力机制; 递归特征金字塔

中图分类号: TP389.1

文献标志码: A

文章编号: 2095–2163(2025)07–0099–05

Small target detection algorithm based on improved YOLOv7 in grasping environment

YANG Miaohai, CHEN Dan

(College of Electrical Engineering and Automation, Fuzhou University, Fuzhou 350000, China)

Abstract: Aiming at the problems of difficult recognition and low detection accuracy for small target objects in grasping tasks under industrial environments, this paper proposes an improved YOLOv7-based small target detection algorithm. The BiFormer dynamic sparse attention mechanism is introduced into the feature fusion part of the YOLOv7 model to effectively enhance the detail information of small target objects and improve the detection accuracy. Additionally, a recursive feature pyramid module is incorporated into the detection section to fuse the output features from the feature pyramid structure with the bottom-up backbone layers. Experimental results demonstrate that the enhanced YOLOv7 model achieves improved performance, showing a 3.6 percentage point increase in detection precision and superior detection performance for small target objects.

Key words: small target detection; YOLOv7; bi-level routing attention; recursive feature pyramid

0 引 言

如今, 机械臂抓取在国家工业自动化生产中应用广泛, 其中基于视觉的机械臂抓取扮演着非常重要的角色, 可以提升生产效率、降低人力成本^[1-2]。但在实际的抓取任务中, 机器人往往处于非理想化的复杂环境, 检测更加艰难, 尤其是小目标物体的检测, 会出现小目标物体难以识别、检测精度低等情况^[3]。由于目标物体尺寸小、特征少, 导致检测结果准确度低、识别速度慢, 因而研究基于图像的小目标检测算法具有重要的意义^[4]。

小目标检测是目标检测的一个分支, 指在图像中

检测和定位尺寸较小的目标, 而现今大多数检测都是基于目标检测的算法模型进一步改进。在注意力机制改进方面, 徐诚极等^[5]在 YOLOv3 算法基础上针对小目标的识别进行改进, 通过引入注意力机制来定位物体的位置, 解决了小目标的边界框位置漂移以及重叠物体难以区分等方面的问题。在瞄框方面, Ge 等^[6]提出高性能的 YOLOX 算法, 采用无瞄框机制, 直接预测边界框的位置信息; Li 等^[7]的 YOLOv6 沿用 YOLOX 的无瞄机制, 并提出一个基于 RepVGG 的新骨架 EfficientRep, 在分类和回归任务使用自蒸馏策略, 动态调整教师模型和标签, 便于学生模型的训练。高分辨率方面, Wang 等^[8]为实时、精准探测海面物

基金项目: 国家自然科学基金面上项目 (61973085); 福建省自然科学基金面上项目 (2022J01114)。

作者简介: 羊森海 (1999—), 男, 硕士研究生, 主要研究方向: 计算机图像处理, 深度学习, 基于网络的机器人控制。

通信作者: 陈 丹 (1977—), 女, 博士, 副教授, 硕士生导师, 主要研究方向: 基于网络的机器人控制, 机械臂轨迹规划, 机器视觉研究等。

Email: nzpovkr04@foxmail.com。

收稿日期: 2023–11–16

哈尔滨工业大学主办 ◆ 系统开发与应用

2 YOLOv7 算法的改进

为提高工业环境下机械臂抓取的小目标检测精度,本文在原 YOLOv7 算法的基础上融入递归特征金字塔结构与动态稀疏注意力机制,能够有效增加小目标物体细节特征,提升算法检测精度。

2.1 递归特征金字塔

递归特征金字塔在金字塔网络结构的基础上,将 FPN 网络的特征输出连接到主干网络层中,将输入图像特征进行反复的提纯,丰富金字塔网络结构的表达能力,如图 2 所示。RFP 结构中的递归能够更新主干网络的参数,提高算法鲁棒性。

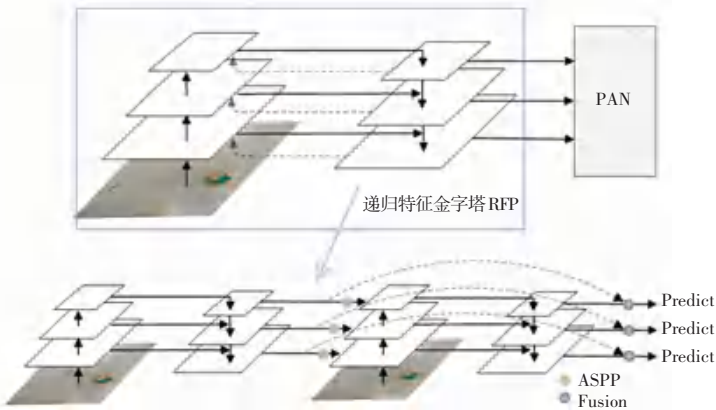


图 2 递归特征金字塔 RFP 结构

Fig. 2 Recursive feature pyramid RFP structure

递归特征金字塔将带有空洞卷积的空间金字塔池化模块 (ASPP) 作为两个递归特征金字塔的级联连接模块^[12]。ASPP 金字塔池化结构如图 3 所示。

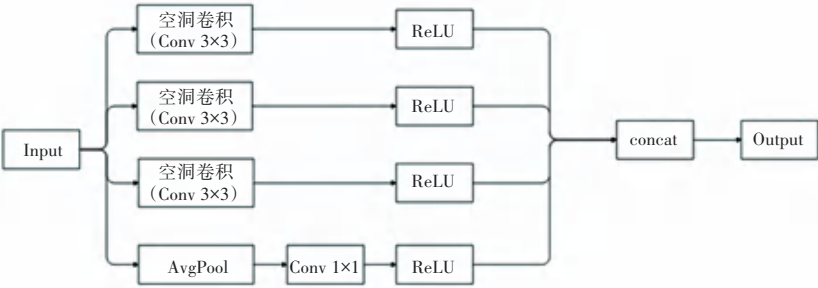


图 3 ASPP 金字塔池化结构

Fig. 3 ASPP pyramid pool structure

金字塔网络的 Fusion 融合模块通过卷积与 Sigmoid 操作计算注意力,将两个相邻阶段的特征按照学习到的权重重新更新,再进行融合,如图 4 所示。 f_i^t 表示第 t 阶段第 i 层的输出, f_i^{t+1} 表示第 t 阶段第 $i+1$ 层的输出, 权重 α 由 f_i^{t+1} 后接卷积层和 Sigmoid 函数输出实现。

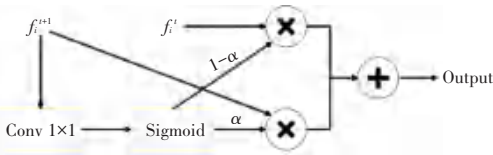


图 4 Fushion 金字塔融合模块

Fig. 4 Fushion pyramid fusion module

2.2 动态稀疏注意力机制

Zhou 等^[13] 提出注意力机制具有稀疏性,呈现长

尾分布,只有少数点积对的注意力有贡献。动态稀疏注意力机制基于 BRA (BiLevel Routing Attention) 模块构建而成,BRA 模块是一种动态的、查询感知的稀疏注意力机制,能够在粗糙区域级别过滤掉大部分不相关的键值对,提升模型鲁棒性。组成 BRA 模块包括 3 个步骤,首先将给定特征图 $X(W \times H \times C)$ 分为若干个不重叠区域 $(S \times S)$ 并进行线性映射,获得查询、键、值的张量 Q, K, V :

$$\begin{aligned} Q &= X^r W^q \\ K &= X^r W^k \\ V &= X^r W^v \end{aligned} \tag{1}$$

其中, W^q, W^k, W^v 分别是查询、键、值的投影。通过 Q 和转置 K 的矩阵乘法推导出相邻矩阵 A :

$$\mathbf{A}^r = \mathbf{Q}^r(\mathbf{K}^r)^T \tag{2}$$

利用邻接矩阵 $\mathbf{A}^r$ 构建有向图来查找不同键值对对应的关系,采用逐行 topk 运算符推导出一个路由索引矩阵 $\mathbf{I}^r$:

$$\mathbf{I}^r = \text{topkIndex}(\mathbf{A}^r) \tag{3}$$

$\mathbf{I}^r$ 的第 i 行包含了 i 区域内最相关的键索引,利用区域到区域的路由索引矩阵 $\mathbf{I}^r$ 来构建双层路由注意力机制 O , O 就是得到的稀疏注意力机制模块 BRA,公式如下:

$$O = \text{Attention}(\mathbf{Q}, \mathbf{K}^g, \mathbf{V}^g) + \text{LCE}(\mathbf{V}) \tag{4}$$

其中, $\mathbf{K}^g, \mathbf{V}^g$ 是聚集起来的键和值, LCE 是引入的上下文增强机制,公式如下:

$$\begin{cases} \mathbf{K}^g = \text{gather}(\mathbf{K}, \mathbf{I}^r) \\ \mathbf{V}^g = \text{gather}(\mathbf{V}, \mathbf{I}^r) \end{cases} \tag{5}$$

本文将 BiFormer 中的 BRA 模块引入 YOLOv7 改进的递归特征金字塔部分,改进的递归特征金字塔结构如图 5 所示。

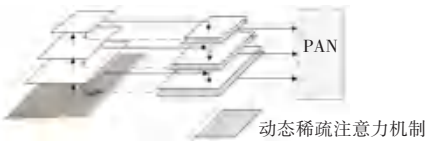


图 5 改进的递归特征金字塔结构

Fig. 5 Improved recursive feature pyramid structure

3 实验结果与分析

3.1 实验设置及评价指标

实验环境:Windows10 操作系统,显卡为英伟达的 GTX3080Ti, 12 GB 显存。训练 300 轮次,使用 Adam 优化器,学习率为 0.001,批处理大小为 16。

为了检验算法在机械臂抓取的小目标物体检测图像上的有效性,本文使用平均精度均值(mAP)、

平均准确率(Average Precision, AP)和算法检测速率FPS作为算法的评价指标。

3.2 数据集

本文对不同光照、遮挡等复杂环境下的小目标物体进行拍照,获取小目标物体图像数据集。为提高算法泛化性能,对所有图像数据进行旋转、平移和缩放等数据增强操作。

选用 LabelImg 软件作为标注工具,对采集和处理得到的 997 张图像数据进行标注,标注好的数据集按照 8 : 2 的比例随机分为训练集和测试集。

3.3 实验与结果分析

利用训练集和测试集对改进后的 YOLOv7 目标检测算法与 YOLOv5s 和 YOLOv7 目标检测算法分别进行训练和测试,将图像尺寸设置为 640×640,实验结果及对比见表 1。

表 1 多个目标检测算法的实验结果对比

Table 1 Comparison of experimental results of multiple target detection algorithms

算法	AP/%	FPS	mAP@ 0. 5
YOLOv5s	86. 3	68. 5	0. 718
YOLOv7	92. 7	64. 7	0. 784
本文算法	96. 3	64. 3	0. 827

由表 1 可知,本文算法准确度最高、平均精度均值(mAP)最大。由于算法增加了递归特征金字塔,在提升模型精度的同时检测速率会有略微下降,但也能满足实时性检测的需求,能够适用于工业环境下用于机械臂抓取的小目标物体检测。

本文将 YOLOv7 网络算法与改进后的网络算法进行对比实验,检测结果如图 6、7 所示,可见改进后的网络算法识别准确度有所提升,且能识别更小的物体,证明了算法的有效性。



图 6 YOLOv7 检测结果图

Fig. 6 Detection results of YOLOv7



图 7 改进 YOLOv7 检测结果图

Fig. 7 Test results of improvement YOLOv7

4 结束语

针对工业环境下小目标检测困难的问题,本文提出一种改进的 YOLOv7 检测算法。在 YOLOv7 网络算法的特征融合部分引入动态稀疏注意力机制,将原有的 FPN 网络结构改进为递归的 RPN 结构,丰富 FPN 的表达能力,可以有效增加待检测小目标物体的细节信息。实验结果表明,改进后的 YOLOv7 网络算法比原网络算法检测性能更佳。

本文提出的网络算法在工业环境下对小目标物体的检测精度有较大提高,且速度上也满足实时检测的要求。后续实验中,将对数据集进行扩充,提高算法鲁棒性;针对算法进行剪枝、量化等研究操作,以实现进一步应用。

参考文献

[1] 王田苗,陶永. 我国工业机器人技术现状与产业化发展战略[J]. 机械工程学报, 2014, 50(9): 1-13.

[2] 刘建廷. 浅析智能化技术在电气工程自动化中的应用[J]. 科技致富向导, 2014 (12): 188-188.

[3] 刘亚,艾海舟,徐光佑. 一种基于背景模型的运动目标检测与跟踪算法[J]. 信息与控制, 2002, 31(4): 315-319.

[4] 潘晓英,贾凝心,穆元震,等. 小目标检测研究综述[J]. 中国图象图形学报,2023,28(9):2587-2615.

[5] 徐诚极,王晓峰,杨亚东. Attention-YOLO:引入注意力机制的 YOLO 检测算法[J]. 计算机工程与应用, 2019,55(6):13-23.

[6] GE Z, LIU S, WANG F, et al. Yolox: Exceeding yolo series in 2021[J]. arXiv preprint arXiv,2107.08430, 2021.

[7] LI C, LI L, JIANG H, et al. YOLOv6: A single-stage object detection framework for industrial applications[J]. arXiv preprint arXiv,2209.02976, 2022.

[8] WANG B, JIANG P, LIU Z, et al. An adaptive lightweight small object detection method for incremental few-shot scenarios of unmanned surface vehicles [J]. Engineering Applications of Artificial Intelligence,2024,133(PA):107989.

[9] 于永军,徐锦法,张梁,等. 惯导/双目视觉位姿估计算法研究[J]. 仪器仪表学报, 2014, 35(10): 2170-2176.

[10]王静,金玉楚,郭苹,等. 基于深度学习的相机位姿估计方法综述[J]. 计算机工程与应用,2023,59(7):1-14.

[11] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2023: 7464-7475.

[12] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 2117-2125.

[13]ZHOU H, ZHANG S, PENG J, et al. Informer: Beyond efficient transformer for long sequence time-series forecasting [C]// Proceedings of the AAAI Conference on Artificial Intelligence. AAAI, 2021:11106-11115.