

文章编号 1004-924X(2024)21-3231-13

面向机械臂抓取的航天器位姿检测

黄 成*, 苏俊飞, 许家忠

(哈尔滨理工大学 自动化学院, 黑龙江 哈尔滨 150080)

摘要:针对机械臂非合作抓取任务中目标检测效率低的问题,提出了一种基于改进 YOLOv8n 的目标位姿检测算法。将大型可分离核注意力(Large Separable Kernel Attention, LSKA)引入金字塔池化层(Spatial Pyramid Pooling Fusion, SPPF),以增强模型的多尺度聚合特征能力;设计一种全新的轻量化模块 RGCSPELAN,以减少模型的运行成本与计算量;然后,用 inner 思想改造平均成对距离交并比(Mean Pairwise Distance Intersection over Union, MPDIoU),进而改造加权交并比(Wise-IoU),形成新的损失函数 Wise-MPDIoU^{inner},以提高模型的训练效果与检测性能;最后,基于目标位置检测信息和深度信息,通过构建目标实时坐标系得到其空间三维姿态信息并完成机械臂抓取任务。实验结果表明:本文算法的精确率达到了 96.5%,mAP@0.5 指标达到了 96.7%,参数量降低了 16%,推理速度提升了 33%,实现了模型精度与轻量化之间的平衡,且满足 UR5 机械臂在非合作抓取任务中的实时性要求。

关键词:机械臂;抓取检测;轻量化;YOLOv8n;位姿检测;损失函数

中图分类号:TP18;TP391.41 **文献标识码:**A **doi:**10.37188/OPE.20243221.3231

Spacecraft posture detection for robotic arm grabbing

HUANG Cheng*, SU Junfei, XU Jiazhong

(College of Automation, Harbin University of Science and Technology, Harbin 150080, China)

* Corresponding author, E-mail: huangchengsunxi@163.com

Abstract: An improved YOLOv8n-based target position detection algorithm is proposed to address the inefficiency of target detection in non-cooperative robotic arm gripping tasks. First, the Large Separable Kernel Attention (LSKA) mechanism is integrated into the Spatial Pyramid Pooling Fusion (SPPF) layer to enhance the model's multiscale feature aggregation capability. Second, a novel lightweight module, RGCSPELAN, is designed to reduce the computational cost and runtime of the model. Additionally, the average pairwise distance intersection over union (MPDIoU) is restructured with an inner transformation concept, which is further combined with the weighted intersection over union (Wise-IoU) to develop a new loss function, Wise-MPDIoU^{inner}. This loss function enhances both the training efficiency and detection performance of the model. Finally, using target position detection and depth information, a real-time coordinate system is constructed to determine the target's 3D spatial attitude, enabling the completion of robotic arm grasping tasks. Experimental results demonstrate that the proposed algorithm achieves an accuracy of 96.5%, an mAP@0.5 of 96.7%, a 16% reduction in parameters, and a 33% improvement in inference speed. The algorithm effectively balances model accuracy and computational efficiency, meeting the

收稿日期:2024-07-12;修订日期:2024-08-16.

基金项目:国家自然科学基金资助项目(No. 52102455)

real-time requirements of the UR5 robot for non-cooperative grasping tasks.

Key words: robotic arm; grabbing detection; lightweighting; YOLOv8n; posture detection; loss function

1 引言

近年来,随着机械臂技术的高速发展,机械臂抓取作业在工业生产中替代了大量的人力资源。根据被抓取目标能否提供位姿信息,可以将其分为合作与非合作两类。相比合作式目标,非合作抓取任务在航天、生产制造、物流分拣等领域具有更广阔的应用前景。为了执行复杂多样的非合作抓取任务,目标位姿检测成为该研究领域的一项关键技术^[1-3]。目前,大多数视觉检测技术能够保证检测精度,但实时性一般,尤其在目标姿态估计时,难以及时应对目标形状复杂的时敏抓取任务,因此,研究面向机械臂抓取的高效位姿检测方法势在必行。

空间非合作目标具有形状多样且结构复杂的特点,通常是不规则物体。除此之外,其表面材料的异质性和结构的独特性,进一步增加了检测难度。空间非合作目标的检测与常规目标有显著区别,尤其涉及精密操作或抓取任务时,对系统的稳定性和反应速度提出了更高的要求。

近年来,针对机械臂抓取任务的位姿检测算法研究取得了显著进展。其中,基于深度学习的机械臂抓取位姿估计得到了广泛的关注和应用^[4-6]。Yang 等^[7]以 RGB 图像为输入,结合深度神经网络模型与 Deeplabv3 姿态估计网络,实现了在像素级对物体抓取姿态的预测。Zhang 等^[6]为解决由于物体表面形状突出以及物体尺寸而导致的物体姿态预测偏差,提出通过对象分类辅助任务的形式将对象类别信息合并到姿态估计网络中的基于多任务参数共享(PMP)框架的六维姿态估计。Meng 等^[8]针对多目标环境下机械臂抓取精度低、网络结构复杂等问题,提出了一种基于无锚机制的多任务网络模型来检测物体的类别和抓取姿势。该模型通过残差网络实现浅层特征的提取,引入基于关键点的抓取策略以完成抓取姿态的准确估计,同时引入混合注意力提高抓取预测的准确性。阮国强等^[9]针对部分遮挡物体抓取姿态效率低的问题,基于 Point-

Net++ 网络设计了一种抓取姿态估计算法。该算法从目标点云中生成有效的抓取姿态,简化了学习过程并加快了学习速度。李树春等^[10]针对机械臂抓取过程中场景的复杂性和遮挡问题,提出了一种基于深度相机的物体位姿估计方法,该方法提高了抓取的稳定性和精确性。此外,还有许多学者将 YOLO(You Only Look Once)算法与位姿估计相结合,提高机械臂的抓取性能^[11]。Li 等^[12]针对传统收割机器人收割性能差等问题,设计了一种根据农作物表型和姿态估计进行收割的新型机器人。该机器人基于 YOLOv5 与自适应 DBScan 聚类算法进行果实检测和位姿估计,从而提高机器人的采摘效率。巩大康^[13]为解决 2D 图像只能识别物体位置,无法获得物体的深度信息,且传统的 3D 点云视觉检测算法复杂且运算时间较长等问题,提出了 2D 彩色图与 3D 深度图结合的物体识别和位姿估计方法,通过改进 YOLOv7-tiny 算法与 PCA 位姿估计算法,实现了复杂环境下的抓取任务,提高了服务机器人的智能化程度,但是该算法的位姿估计实时性较差。李启煌^[14]针对复杂场景中不同物体因遮挡导致的位姿估计精度不高、抓取效率低的问题,基于 YOLO-6D 算法提出了 YOLO-SA 和 YOLO-SCA 目标位姿估计算法。该算法提高了机器人的抓取效率和位姿估计的检测精度。然而,上述抓取检测算法均不能很好地兼顾检测精度、检测效率以及实时性。YOLO 算法自问世以来,一直在不断地更新迭代,YOLOv8 算法在保持高精度和适度计算复杂度的前提下,实现了较高的实时性,但存在训练成本高、参数量较大的问题。

因此,本文选择模型较小、实时性较高的 YOLOv8n 作为基线模型进行改进,结合目标实时坐标系构建方法,提出一种基于 SRW-YOLOv8n 算法的机械臂抓取位姿检测方法。该方法将 LSKA 注意力与金字塔池化层 SPPF 融合,增强模型的多尺度聚合特征能力;结合 GhostNet 思想,聚合网络结构(Efficient Layer Aggregation Network, ELAN)与重参数化模块(RepConv)设

计一种全新的轻量化模块 RGCSPELAN,在增强特征提取和梯度流通能力的基础上实现了模型的轻量化;将 inner 思想引入平均成对距离交并比 MPDIoU 中形成 $\text{MPDIoU}^{\text{inner}}$,然后替换加权交并比 Wise-IoU 中的交并比 (Intersection over Union, IoU),形成新的损失函数 Wise-MPDIoU^{inner},以提高模型的训练效果与检测性能;为了利用 SRW-YOLOv8n 算法实现对目标位姿的实时检测,将目标 4 个区域上中心点的二维位置信息与深度信息结合,通过构建目标实时坐标系获取空间位姿信息,从而提高航天器位姿的检测速度。

2 YOLOv8n 基本结构

YOLOv8 算法是 Ultralytics 公司于 2023 年 1 月发布的一款 YOLO 模型。作为一种高效而强大的目标检测算法,它具备优秀的实时性能和高准确性。根据模型大小,YOLOv8 可分为 n, s, m, l 和 x 五个版本。随着模型网络结构的加深,检测精度一般会提高,但检测速度会降低。由于 YOLOv8n 模型最小,具备快速检测速度和较低的训练成本,因此本文选择 YOLOv8n 作为基准模型。

YOLOv8 算法主要由主干网络 (Backbone), 颈部 (Neck) 和检测头 (Head) 三部分构成,其网络结构与 YOLOv5 类似。Backbone 部分在 YOLOv8 中的作用是对输入图像进行特征提取,并为后续目标检测任务提供重要的信息基础。Neck 部分在 YOLOv8 中的作用是进行特征融合,为后续的任务提供更加丰富、准确的特征表达,从而提高模型的性能和效果。除此之外,YOLOv8 仍采用 FPN-PAN 思想来构建和优化多尺度特征表示,以提高网络对于不同尺度目标的检测能力,从而使得模型在复杂场景中具有更好的性能表现。Head 部分在 YOLOv8 中的作用是将经过 Backbone 和 Neck 处理得到的特征图转换为目标检测的结果。在损失函数方面,YOLOv8 抛弃了以往的 IoU 匹配或单边比例的分配方式,采用 Task-Aligned Assigner 正负样本匹配的方式。分类分支仍然采用二值交叉熵损失,而回归分支则采用分布聚焦损失 (Distribution Focal Loss, DFL-Loss) 和完全交并比 (Complete

Intersection over Union, CIoU) 损失函数。该损失函数通过优化模型的分类与定位能力,增强了它在复杂场景中的性能表现。

3 原 理

3.1 SRW-YOLOv8n 模型

为了进一步提升模型的性能,本文对其主干网络和颈部网络进行优化,整体框架如图 1 所示。其中, Detect 模块为网络的检测头,主要功能是从模型提取的特征中生成具体的目标检测结果。3 个 detect 模块可以实现多尺度特征图检测,分别负责小、中、大尺度的目标物体。因此可以对其进行改进,进一步提高目标检测模型的检测精度,增强稳定性。

首先,将 LSKA 与 SPPF 融合,进一步提高不同特征层之间的特征信息融合能力。其次,基于 GhostNet 思想,ELAN 结构与重参数化思想设计一种轻量化模块 RGCSPELAN,以此减少模型的参数量与计算量,同时降低训练成本。最后,用 inner 思想改造 MPDIoU,进而以 $\text{MPDIoU}^{\text{inner}}$ 替换 Wise-IoU 中的 IoU,形成新的损失函数 Wise-MPDIoU^{inner}。

3.1.1 LSKA 机制

为了提高对航天器模型关键特征的提取与融合能力,在 YOLOv8n 网络结构中的 SPPF 模块中引入 LSKA 机制^[15-16]。LSKA 机制能够引导模型更精确地聚焦于航天器模型中的关键特征,融合后的模块进一步增强了特征融合效果,有效提升了模型的整体性能。与以往的注意力机制不同,LSKA 具有长距离依赖性、空间适应性、通道适应性和极大型核的可扩展性 4 个关键特性。因此,LSKA 能够显著提高模型对于不同空间尺度、形状和位置的输入数据,以及输入数据中不同通道特征的适应能力,从而提高模型的性能、稳定性和泛化能力,在处理大尺度数据和复杂任务时表现出色。改进后的 SPPF-LSKA 模块结构示意图如图 2 所示。经卷积层 Conv 与池化层 MaxPool2d 处理后的结果输入到 LSKA 模块中,随后通过 LSKA 模块以更大的感受野进行特征提取,帮助模型更好地理解数据,最后再经过卷积层处理后输出。改进后的 SPPF-LSKA 模块极大提高了模型在多个尺度上的特征聚合能力。

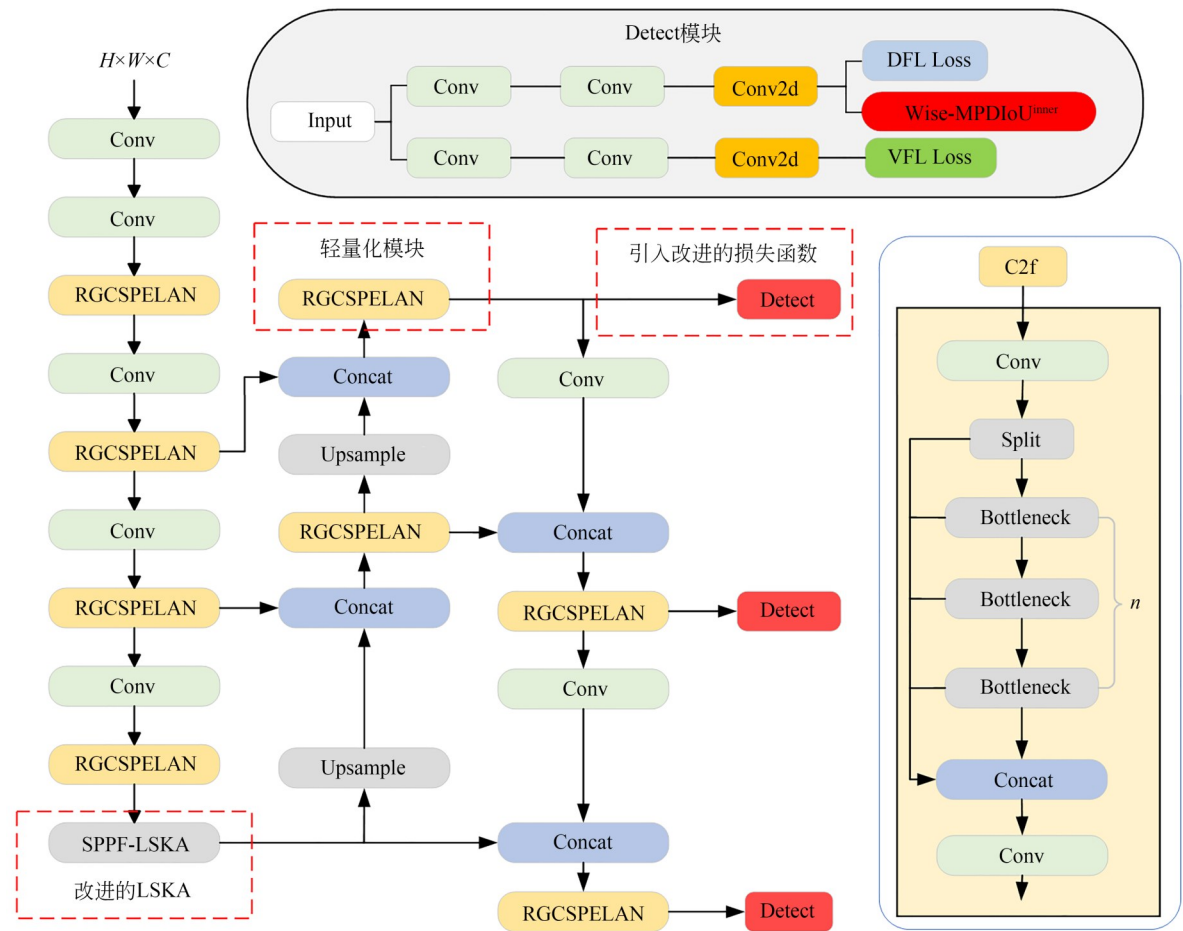


图1 SRW-YOLOv8n网络结构

Fig. 1 Structure of SRW-YOLOv8n network

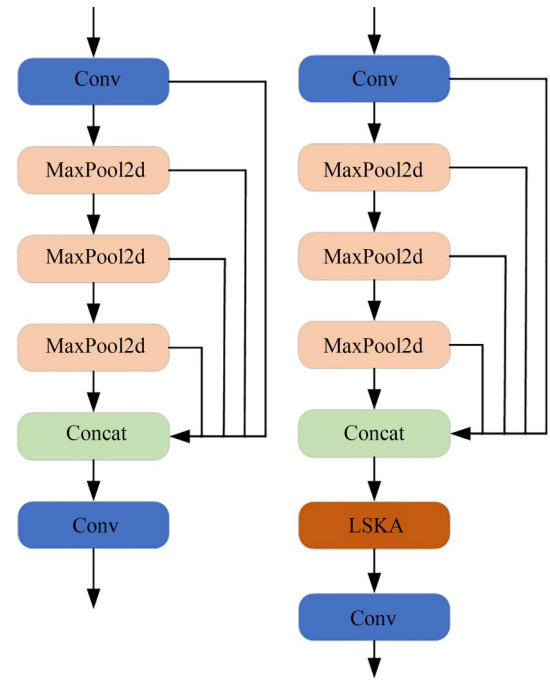


图2 SPPF与SPPF-LSKA结构示意图

Fig. 2 Schematic diagram of SPPF and SPPF-LSKA

3. 1. 2 RGCSPELAN 轻量化模块

为了提高模型的训练和推理效率并减少计算成本,本文结合 GhostNet 网络与 ELAN 结构的设计思想以及重参数化思想设计了一种轻量化模块 RGCSPELAN。RGCSPELAN 模块结构示意图如图 3 所示。

GhostNet 网络是一种轻量级神经网络架构。与传统的 CNN 模型相比, GhostNet 网络通过生成少量的基特征图,再经一系列廉价的线性变换处理后获得更多的特征图,进而提高网络的表达能力。该网络由于采用了廉价操作,计算成本显著降低,同时减少了参数量,提高了模型的泛化能力。ELAN 结构在设计思想上与 C3 模块和 C2f 模块类似,三者都采用特征融合的方式丰富特征表示,并且它们都采用了多分支结构来捕捉特征信息。而重参数化思想则是能够实现更高的计算效率和更快的推理速度。Ding 等基于 VGG 风格提出了 RepVGG 架构。RepConv 模块

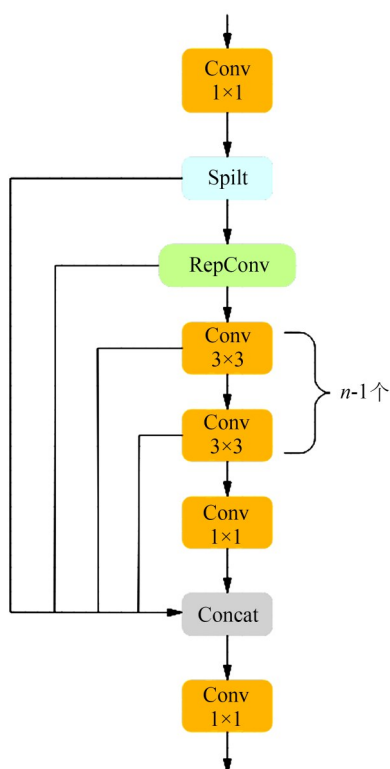


图3 RGCSPELAN结构示意图

Fig. 3 Schematic diagram of RGCSPELAN

是RepVGG架构的基本组成部分,采用了卷积重参数化方法。在模型训练阶段,通过多个分支来训练,以提高模型的精度。在推理阶段,通过结构重参数化,将训练时的多分支结构等效转换为单分支结构,以加快推断速度,降低模型复杂度和计算成本。卷积层、BN层以及合并简化后的公式为:

$$\text{Conv}(x) = W(x) + b, \quad (1)$$

$$\text{BN}(x) = \gamma \times \frac{x - \text{mean}}{\sqrt{\text{var}}} + \beta, \quad (2)$$

$$\text{BN}(\text{Conv}(x)) = \gamma \times \frac{\text{Conv}(x) - \text{mean}}{\sqrt{\text{var}}} + \beta = \frac{\gamma \times W(x)}{\sqrt{\text{var}}} + \left(\frac{\gamma \times (b - \text{mean})}{\sqrt{\text{var}}} + \beta \right), \quad (3)$$

式中: x 为输入值, W 为卷积层权重, b 为卷积层偏置, γ 为BN权重, β 为BN偏置, $\gamma \times W(x)/\sqrt{\text{var}}$ 为融合后卷积的权重, $\gamma \times (b - \text{mean})/\sqrt{\text{var}} + \beta$ 为融合后卷积的偏置。RepConv结构如图4所示。

RGCSPELAN模块借鉴了GhostNet网络中

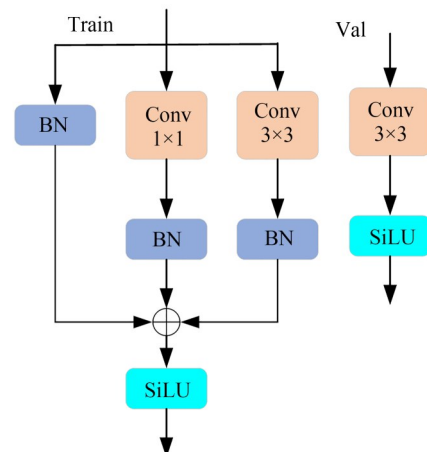


图4 RepConv结构

Fig. 4 Structure of RepConv

的结构,通过第一个卷积生成初始特征图,然后经过多级特征提取。另外还借鉴了ELAN结构的设计思想,将初始特征图由split模块处理后形成两个分支,一支继续进行多级特征提取;另一支直接与多级处理后的特征进行拼接,再经卷积操作后输出。除此之外,舍弃了YOLOv8中常用的BottleNeck。为了弥补舍弃残差块所导致的性能缺失,在梯度流通分支上使用重参数化模块RepConv,以此增强特征提取与梯度流通的能力,并且在推理时进行重参数化,加快推理速度,降低模型复杂度。

3.1.3 损失函数

损失函数在机器学习和深度学习中的作用不可或缺。它不仅用于量化模型预测与真实值之间的差异,还在模型参数优化、正则化以及反向传播等关键任务中发挥核心作用。IoU自提出后就成为对象检测中常用的评价指标,其公式为:

$$\text{IoU} = \frac{|B^{\text{pred}} \cap B^{\text{gt}}|}{|B^{\text{pred}} \cup B^{\text{gt}}|}, \quad (4)$$

其中 B^{pred} 与 B^{gt} 分别为预测框与真实框。在YOLOv8n网络模型中,损失函数采用CIOU损失函数。CIOU损失函数综合考虑了目标框之间的位置、尺寸和形状等因素,具有较好的目标定位精度和检测准确度。然而,CIOU也存在一些缺点,比如计算复杂度较高,需要更多的计算资源;在一些极端情况下,可能对目标框的评价不够准确。

IoU^{inner} 损失函数通过引入尺度因子比 ratio 控制生成不同尺度的辅助边框来计算损失,能够很好地根据不同检测任务进行自我调整。具体计算公式如下:

$$\begin{cases} b_l^{gt} = x_c^{gt} - \frac{w^{gt} \cdot \text{ratio}}{2}, b_r^{gt} = x_c^{gt} + \frac{w^{gt} \cdot \text{ratio}}{2} \\ b_t^{gt} = y_c^{gt} - \frac{h^{gt} \cdot \text{ratio}}{2}, b_b^{gt} = y_c^{gt} + \frac{h^{gt} \cdot \text{ratio}}{2} \\ b_l = x_c - \frac{w \cdot \text{ratio}}{2}, b_r = x_c + \frac{w \cdot \text{ratio}}{2} \\ b_t = y_c - \frac{h \cdot \text{ratio}}{2}, b_b = y_c + \frac{h \cdot \text{ratio}}{2} \end{cases}, \quad (5)$$

$$\text{inter} = (\min(b_r^{gt}, b_r) - \max(b_l^{gt}, b_l)) \cdot (\min(b_b^{gt}, b_b) - \max(b_t^{gt}, b_t)), \quad (6)$$

$$\text{union} = (w^{gt} \cdot h^{gt}) \cdot (\text{ratio})^2 + (w \cdot h) \cdot (\text{ratio})^2 - \text{inter}, \quad (7)$$

$$\text{IoU}^{\text{inner}} = \frac{\text{inter}}{\text{union}}, \quad (8)$$

其中: b_l^{gt}, b_r^{gt} 分别表示真实框、预测框的角顶点计算结果, (x_c^{gt}, y_c^{gt}) 为真实框的中心点, w^{gt} 和 h^{gt} 分别表示宽度和高度, (x_c, y_c) 为预测框的中心点, w 和 h 分别表示宽度和高度。由式(5)可得真实框与辅助检测框的角顶点,并计算出预测框与真实框之间重叠的面积。由式(8)可计算出重叠区域与预测框之间的比值。ratio $\in [0.5, 1.5]$, ratio < 1 时,辅助边框面积小于实际边框面积,其回归的有效范围小于 IoU 损失的有效范围,但是其梯度绝对值大于 IoU 损失的梯度,可以加速高 IoU 样本的收敛。而当 ratio > 1 时,辅助边框有助于低 IoU 的回归。

MPDIoU 是一种基于最小点距离的边界框相似度的 IoU 度量方法,能很好地兼顾重叠、非重叠区域,以及中心点距离等因素。其表达式为:

$$\text{MPDIoU} = \text{IoU} - \frac{\rho_{ul}^2(P, G)}{w^2 + h^2} - \frac{\rho_{br}^2(P, G)}{w^2 + h^2}, \quad (9)$$

其中: $\rho_{ul}^2(P, G)$ 计算预测框与真实框的左上角点的距离平方和。将 IoU^{inner} 替换 MPDIoU 中的 IoU 即可得到 MPDIoU^{inner} 损失函数,其公式为:

$$\text{MPDIoU}^{\text{inner}} = \text{IoU}^{\text{inner}} - \frac{\rho_{ul}^2(P, G)}{w^2 + h^2} - \frac{\rho_{br}^2(P, G)}{w^2 + h^2}. \quad (10)$$

Wise-IoU v3 是一种具有动态聚焦机制的损失函数,它在 Wise-IoU v1 的基础上,使用“离群度”来描述锚框的质量。Wise-IoU v1 的计算公式如下:

$$L_{\text{WIoUv1}} = R_{\text{WIoU}} L_{\text{IoU}}, \quad (11)$$

$$R_{\text{WIoU}} = \exp\left(\frac{(x_c - x_c^{gt})^2 + (y_c - y_c^{gt})^2}{(W_g^2 + H_g^2)^*}\right), \quad (12)$$

式中: L_{WIoUv1} 为 Wise-IoU v1 损失, L_{IoU} 为边界框损失 IoU, R_{WIoU} 为距离注意力, W_g, H_g 分别为预测框与真实框最小外接矩形的宽和高, * 表示从计算图中分离操作,使其变为不带梯度的常量。

离群度的定义如下:

$$\beta = \frac{L_{\text{IoU}}^*}{L_{\text{IoU}}} \in [0, +\infty), \quad (13)$$

其中 $\overline{L_{\text{IoU}}}$ 为归一化因子。

利用离群度 β 构造非单调聚焦系数 k , 进而得到边界框损失 Wise-IoU v3, 其公式为:

$$L_{\text{WIoUv3}} = k L_{\text{WIoUv1}}, \quad (14)$$

式中非单调聚焦系数 $k = \beta / \delta \alpha^{\beta - \delta}$ 。本文中,超参数 α, δ 分别设置为 1.9 和 3^[16]。将 MPDIoU^{inner} 替换 Wise-IoU v3 中的 IoU, 即可构造出 Wise-MPDIoU^{inner} 损失函数。

3.2 位姿检测方法

近年来,端对端的位姿估计得到了广泛的关注,并随着深度学习技术的不断进步取得了显著进展。该技术能够对规则目标进行高效的特征学习,提升了估计精度和模型的稳定性。然而,本文的研究对象是不规则目标,尽管端对端方法具有自动特征学习的优势,但其训练过程复杂,需要大量计算资源,并且稳定性较差,容易受到输入数据质量的影响。因此,这里采用建立实时坐标系的方法来获取物体的三维姿态,不仅节省了计算资源,也显著提高了位姿估计的实时性。

基于 SRW-YOLOv8n 算法的位姿检测流程如下:首先,通过深度相机,利用 SRW-YOLOv8n 算法实时获取目标 4 个区域中心点的位置与深度信息;然后,基于所获取的 4 个空间位置信息,通

过计算平面法向量构造目标实时坐标系;最后,通过目标物坐标系与机器人末端坐标系之间的转换,实现对目标抓取位姿的检测。实时坐标系的构建主要通过识别目标物上 4 个有效点,如图 5 所示,然后构建两个平面向量。通过这两个向量构造一个平面,随后根据右手定则确定平面的法向量,该法向量为坐标系的 Z 轴,最后由以上 3 个向量即可构成目标物的实时坐标系。

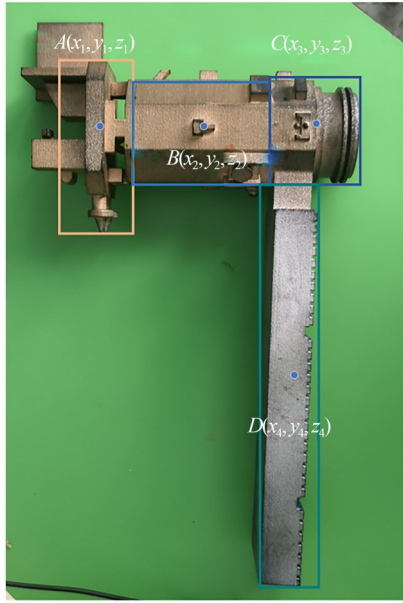


图 5 坐标系的建立

Fig. 5 Establishment of coordinate system

A, B, C 和 D 为航天器模型上 4 个部分的有效点,接下来做向量 $BA(x_1 - x_2, y_1 - y_2, z_1 - z_2)$, $CD(x_4 - x_3, y_4 - y_3, z_4 - z_3)$, 平面法线与两向量垂直,故而法向量 n 为:

$$n = BA \times CD = \begin{vmatrix} i & j & k \\ x_1 - x_2 & y_1 - y_2 & z_1 - z_2 \\ x_4 - x_3 & y_4 - y_3 & z_4 - z_3 \end{vmatrix} = (ai + bj + ck = (a, b, c)). \quad (15)$$

本文采用 RPY 角描述目标物坐标系和机器人末端坐标系之间的姿态转换,即通过旋转矩阵表示目标物体坐标系与机器人坐标系的转换,再将旋转矩阵转换为 RPY 角度组合,进而实现目标物体与机器人末端姿态表述的统一。首先,将机器人末端坐标系与目标物上的实时坐标系进行比对,若存在角度为 0 的轴,如两坐标

系的 Z 轴平行,则只需绕 Z 轴旋转即可使得机器人末端坐标系得到期望姿态,若不存在,则需进行绕轴旋转;其次,将机器人末端坐标系绕 Z 轴旋转,使得两坐标系 X 轴呈二维夹角关系,即两 X 轴在同一平面内;然后,将机器人末端坐标系绕 Y 轴旋转,使得两坐标系 X 轴平行;最后,将机器人末端坐标系绕 X 轴旋转,使得两坐标系的 Y 轴, Z 轴重合。此时,目标物坐标系和机器人末端坐标系姿态统一,可以完成对航天器模型的抓取任务。RPY 角姿态转换过程如图 6 所示。

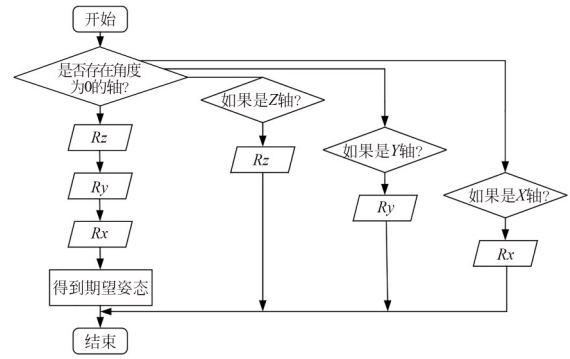


图 6 RPY 角姿态转换过程

Fig. 6 RPY angle attitude conversion process

4 实验及结果分析

4.1 实验环境

本文以航天器模型为研究对象,采用实验室的自采图像制作数据集。图像采集设备为 vivoX100 手机,日期为 2024 年 1 月。在数据采集过程中,综合考虑环境的复杂度与光照强度,最终采集了 3 000 张图片。按背景复杂度可划分为:单一背景 1 594 张,复杂背景 1 406 张;按光照强度可划分为:光亮 2 683 张,昏暗 317 张。随后,利用 LabelImg 工具对每张图像进行标注。航天器模型按照抓取的可行性划分为 4 个部分,即 grabs, ungrabs1, ungrabs2 和 ungrabs3。其中,grabs 为可被机械臂夹爪抓取的部分,其余则为不可被抓取的部分。最后,根据 7:2:1 的比例划分训练集、测试集和验证集。图 7 为标注图像的不同类别。

本文使用精确率 P (Precision), 召回率 R (Recall), 平均精度均值 mAP (Mean Average

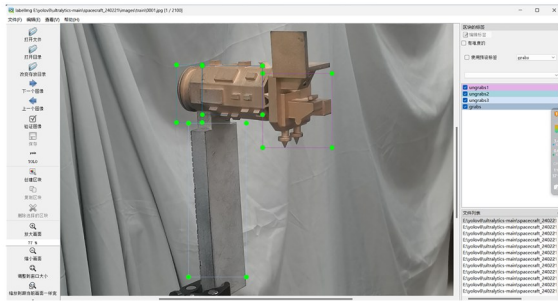


图 7 数据集标注

Fig. 7 Dataset annotation

Precision), 推理速度 FPS (Frames Per Second), 权重大小, 参数量 Param (Parameters), 浮点计算量 FLOPs (Floating Point of Operations) 以及权重来评估航天器模型位置检测的性能。各项指标的计算公式如下:

$$P = \frac{TP}{TP + FP} \times 100\%, \quad (16)$$

$$R = \frac{TP}{TP + FN} \times 100\%, \quad (17)$$

其中: TP 为预测正确的正样本数量, FN 为预测错误的负样本数, FP 为预测错误的正样本数。

$$AP = \int_0^1 p(r) dr, \quad (18)$$

$$mAP = \frac{1}{m} \sum_{i=1}^m AP_i, \quad (19)$$

$$FPS = \frac{1000}{t}, \quad (20)$$

其中: AP 为平均精度, $p(r)$ 指 PR 曲线上各点的数值, m 指数据集中类别的个数, t 表示推理一张图片需要的时间, 单位是毫秒。

模型训练的操作系统为 Windows 10 系统, CPU 型号为 Intel Core i7-12700KF, GPU 型号为 NVIDIA GeForce RTX3080Ti, 采用 Pytorch1.13.1 深度学习框架构建模型, 并通过 CUDA 11.7 加速训练, 编程环境为 python3.8。实验的迭代次数设置为 300, 初始学习率为 0.01, 训练轮数设置为 32, 输入图片尺寸为 640×640 , 线程数为 4, 动量参数为 0.937, 权重衰减率设置为 0.0005。

4.2 不同尺度因子对比实验

为测试不同尺度因子比下 Wise-MPDIoU^{inner} 损失函数对基线模型的影响, 从而选取合适的辅助边框以提高模型的泛化能力, 本文以 0.2 为步

长选取尺度因子比, 进行实验对比。

不同尺度因子比的实验结果对比如表 1 所示。当尺度因子比 ratio=0.6 时, 精确率与 mAP@0.5 以及推理速度 FPS 最低; 当尺度因子比 ratio=1.4 时, 模型的改进效果最佳, 精确率为 96.0%, mAP@0.5 为 96.5%, 推理速度 FPS 则达到 91.7 帧每秒, 相较于最低的 70.9 帧每秒, 提高了 29.3%; 当尺度因子比 ratio=1.6 时, 精确率与 mAP@0.5 以及推理速度 FPS 相较于 ratio=1.4 时均有所降低。除此之外, 将不同尺度因子比下的边界框损失值进行对比得到损失曲线, 如图 8 所示。其中, 尺度因子比 ratio=1.4 时的损失函数表现最好, 能够兼顾模型的检测速率与精度。因此, 为了进一步提高航天器模型检测的准确性与实时性, 本实验选择尺度因子比 ratio=1.4 以改进基线模型。

表 1 尺度因子对比实验

Tab. 1 Scale factor ratio comparison experiment

尺度因子比	精确率/%	mAP@0.5/%	FPS
0.6	95.1	95.9	70.9
0.8	96.0	96.1	86.2
1.0	95.8	96.3	84.2
1.2	95.1	96.4	80.6
1.4	96.0	96.5	91.7
1.6	95.5	96.3	84.6

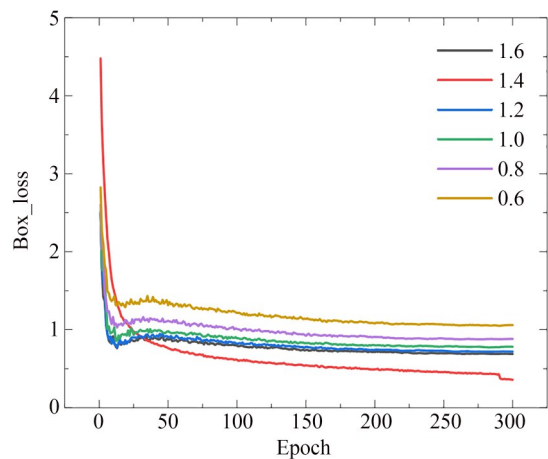


图 8 边界损失

Fig. 8 Bounding Box Loss

4.3 注意力对比实验

为了验证 SPPF-LSKA 模块的有效性, 将它

与 BiFormer^[17], SimAM, CPCA (Channel Prior Convolutional Attention), BiFPN, SE (Squeeze-and Excitation Networks)和 CBAM(Convolutional Block Attention Module)等注意力机制进行对比,结果如表 2 所示。

表 2 不同注意力机制对比

Tab. 2 Comparison of different attention mechanisms

模 型	精确率/%	mAP@0.5/%	FPS
YOLOv8n	96.3	96.5	80.0
SPPF-LSKA	96.4	96.3	97.1
BiFormer	95.9	96.4	91.7
SimAM	96.3	96.6	90.9
CPCA	96.7	96.3	71.4
BiFPN	95.5	95.8	72.5
SE	94.6	95.9	66.2
CBAM	95.0	96.2	68.5

由表 2 可知,不同的注意力机制对模型产生的作用不同。将 CBAM,SE 和 BiFPN 注意力加入 YOLOv8n 模型中,使得原模型的精确率分别

降低了 1.3%,1.7% 和 0.8%,mAP@0.5 分别降低 0.3%,0.6% 和 0.7%,而推理速度则分别降低 14.4%,17.3% 和 9.4%;将 CPCA 注意力加入 YOLOv8n 模型中,使得原模型的精确率提高了 0.4%,但 mAP@0.5 降低了 0.2%,推理速度降低了 10.8%;将 SimAM 注意力加入 YOLOv8n 模型中,精确率与 mAP@0.5 的值基本不变,但推理速度提高了 13.6%;将 BiFormer 注意力加入 YOLOv8n 模型中,精确率与 mAP@0.5 的值略微降低,但推理速度提高了 14.6%;将 SPPF-LSKA 注意力加入 YOLOv8n 模型中,精确率提高了 0.1%,mAP@0.5 降低了 0.2%,但推理速度提升了 21.4%。综上所述,与其他注意力机制相比,SPPF-LSKA 注意力在保持模型检测精度变化较小的情况下,显著提升了推理速度。因此,选择 SPPF-LSKA 注意力来改进基线模型是可行的。

4.4 消融实验

为了评估本文提出的不同改进模块对模型检测性能的影响,在基于 YOLOv8n 网络的基础上进行了消融实验,实验结果如表 3 所示。

表 3 消融实验

Tab. 3 Ablation experiment

模型	精确率%	mAP@0.5/%	参数量/M	浮点计算量/G	FPS	权重大小/MB
A	96.3	96.5	3.00	8.1	80.0	5.96
B	96.4	96.3	3.28	8.3	97.1	6.48
C	96.1	96.0	2.20	6.1	91.7	4.48
D	96.0	96.5	3.00	8.1	91.7	5.96
E	96.1	96.8	2.67	6.8	94.7	5.41
F	95.3	95.6	3.00	8.1	87.9	5.96
G	96.5	96.7	2.67	6.8	106.4	5.40

本文设计了 7 组消融实验,以评估模型的实时性和有效性。模型 A 代表基线模型 YOLOv8n,模型 B 代表引入 SPPF-LSKA 模块,模型 C 代表引入 RGCSPELAN 模块,模型 D 代表引入 Wise-MPDIoU^{inner} 损失函数,模型 E 代表同时引入 SPPF-LSKA 模块与 RGCSPELAN 模块,模型 F 代表引入 MPDIoU 损失函数,模型 G 代表本文所提出的 SRW-YOLOv8n 模型。

从表 3 可以看出,与基线模型 A 相比,模型 B 的精确率与 mAP@0.5 变化不大,参数量与计算

量略有提升,但它的推理速度提升了 21.4%。模型 C 的精确率与 mAP@0.5 比基线模型 A 略低,参数量降低了 26.7%,计算量降低了 24.7%,模型权重为 4.48MB,推理速度提升了 14.6%。这说明 RGCSPELAN 模块的引入使得基线模型在牺牲较小检测精度的同时,大幅度降低了模型的大小与计算成本,增强了模型的实时性。模型 D 相较于基线模型 A,其检测精度与模型大小几乎不变,但推理速度提升了 14.6%。模型 E 相较于基线模型 A,其 mAP@0.5 提升了 0.3%,参数量

降低了11%，计算量降低了16.0%。说明SPPF-LSKA模块与RGCSPELAN模块的引入，可以在略微提升精度的同时，对模型的计算成本与大小进行降低。模型F相较于其他模型，其精确率与mAP@0.5较低，但推理速度比基线模型A高了9.9%。模型G相较于基线模型A，其精确率提升了0.2%，mAP@0.5提升了0.2%，参数量降低了11%，计算量降低了16.0%，权重为5.40MB，推理速度提升最大，提升了33%。这说明SRW-YOLOv8n模型可以兼顾检测精度与推理速度，可以很好地满足实际应用，进一步证明了本文提出的轻量化模型在提升航天器模型检测效率和精度方面具有显著优势。

4.5 对比实验

为验证本文算法的有效性，将其与其他主流目标检测算法进行对比，实验结果如表4所示。由表可知，YOLOv8n模型的mAP@0.5与其他主流目标检测算法相差不大，但参数量与计算量

均小于其他算法，在推理速度上仅逊色于YOLOv7-tiny，故本文选择YOLOv8n作为基线模型。

与基线模型YOLOv8n相比，SRW-YOLOv8n模型的mAP@0.5提升了0.2%，参数量降低了11.0%，计算量降低了16.0%，推理速度提高了33.0%。与目前常用YOLO系列算法YOLOv5s，YOLOv7-tiny，YOLOv8s，YOLOv9-S相比，SRW-YOLOv8n模型的mAP@0.5值相差不大，参数量分别降低了62.4%，55.5%，75.9%，72.2%，计算量分别降低了58.8%，48.5%，76.1%，82.4%，推理速度提升了36.2%，22%，63.9%，55.3%。与两阶段模型Faster-RCNN，Cascade-RCNN相比，SRW-YOLOv8n模型的mAP@0.5值提高了1.1%，0.9%，推理速度提高了2.3倍，3.2倍，同时参数量与计算量远小于两阶段算法。与同为一阶段算法Tood相比，SRW-YOLOv8n模型的mAP@0.5值提高了1.7%，推理速度提高了3.1倍。综上所述，本文提出的轻量化模型SRW-YOLOv8n能更好地完成对航天器模型的位置检测任务。

4.6 检测效果对比

YOLOv8n与SRW-YOLOv8n的热力图对比如图9所示。图9(a)与图9(b)是复杂环境下的输出热力图。在图9(a)中，YOLOv8n网络关注区域分散，航天器区域关注度不高。在图9(b)中，SRW-YOLOv8n网络关注区域集中在航天器上，其他背景均无关注。图9(c)与图9(d)是单一背景环境下的小目标输出热力图。在图9(c)中，YOLOv8n网络关注重点集中到背景中的黄色纸板上，说明在小目标即航天器距离较远时，YO-

表 4 对比实验

Tab. 4 Comparative experiment

模 型	mAP@0.5/%	参数量/M	计算量/G	FPS
Tood	95.0	32.02	123.0	25.9
Faster-RCNN	95.6	41.36	134.0	32.5
Cascade-RCNN	95.8	69.20	162.0	25.6
YOLOv5s	96.3	7.10	16.5	78.1
YOLOv7-tiny	96.8	6.00	13.2	87.2
YOLOv8s	96.9	11.10	28.4	64.9
YOLOv9-S	96.8	9.60	38.7	68.5
YOLOv8n	96.5	3.00	8.1	80.0
SRW-YOLOv8n	96.7	2.67	6.8	106.4

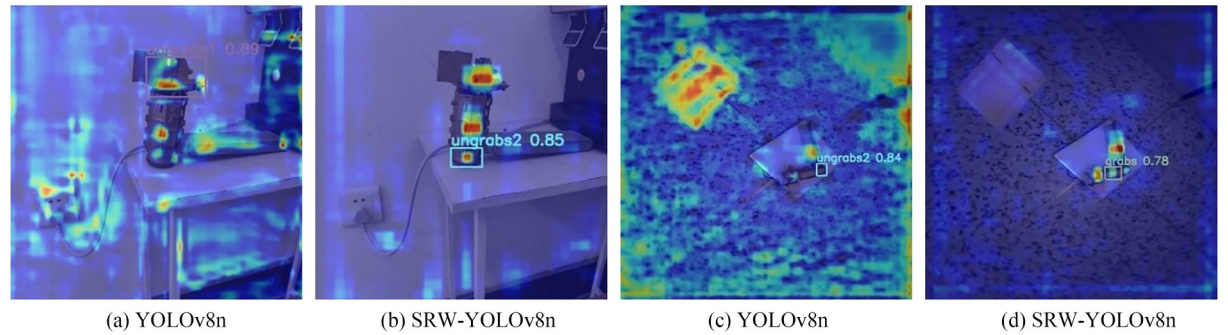


图9 YOLOv8n与SRW-YOLOv8n的热力图对比

Fig. 9 Heatmap comparison between YOLOv8n and SRW-YOLOv8n

LOv8n 网络会将其他黄色物品错认为是航天器黄色主体。在图 9(d)中,SRW-YOLOv8n 网络的注意力集中在航天器上,并未对其他颜色相近物体产生关注,说明 SRW-YOLOv8n 算法在航天器距离较远时有明显作用。

4.7 实际抓取实验

实物抓取实验平台由 UR5 机械臂、D435i 深度相机以及 Robotiq2F-140 二指夹爪组成。首先,基于 SRW-YOLOv8n 算法对航天器模型进行训练,以获取模型上 4 个区域的中心点坐标,并将中心点坐标与深度信息结合形成三维坐标;其次,以模型上的 4 点坐标构成两个平面向量,随后

由右手定则确定模型平面的法向量以构造目标物体的实时坐标系;然后,不断调整机器人末端坐标系,直至目标物体坐标系与机器人坐标系重合,即通过 RPY 角实现坐标系之间的姿态转换,同时调整机械臂位置从而使得两坐标系重合;最后,控制机械臂完成航天器模型的抓取任务。实际抓取结果如图 10 所示,第一行为明亮环境下的抓取,第二行为模拟昏暗环境下的抓取。根据航天器模型的实时位置数据不断调整机械臂,直至机械臂夹爪正对航天器模型为止,即目标物体坐标系与机器人坐标系重合,继而完成对航天器模型的抓取任务。

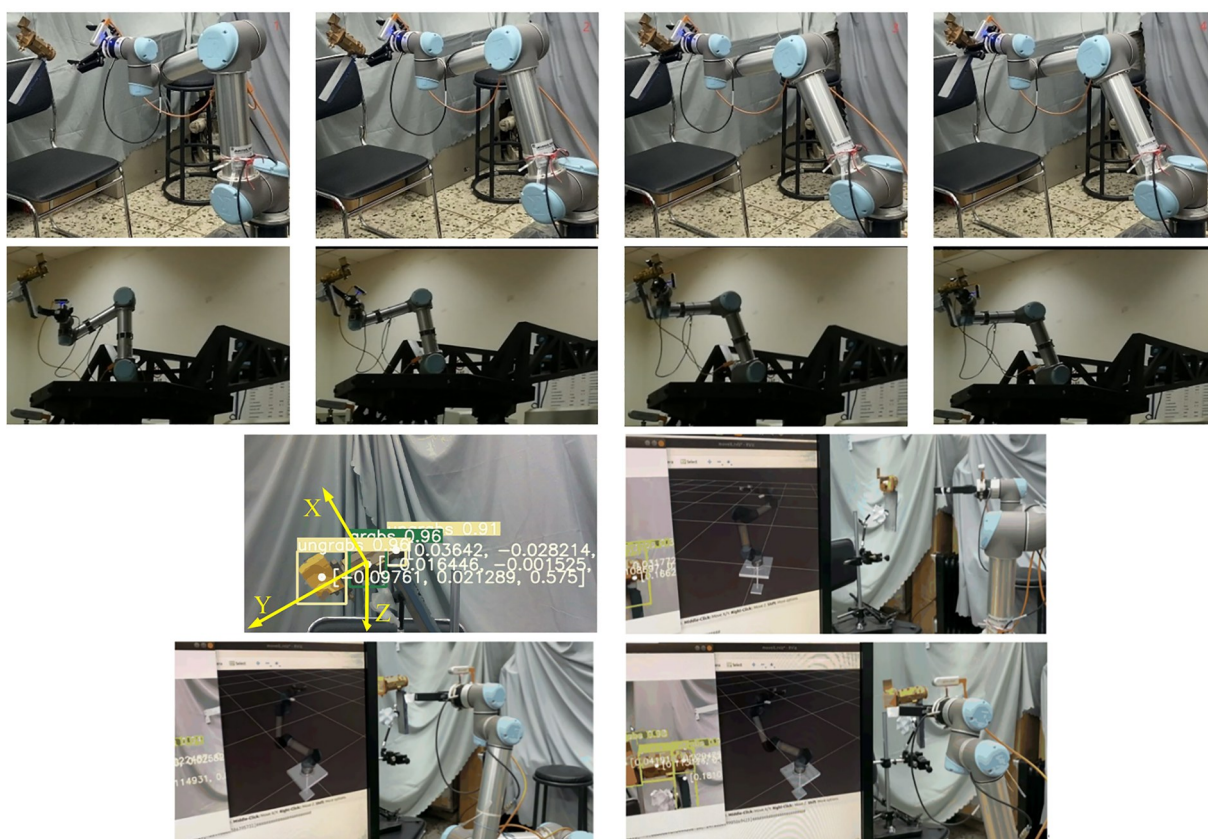


图 10 实际抓取结果

Fig. 10 Actual grabbing results

5 结 论

为提高机械臂非合作抓取任务中目标的检测效率,本文提出了一种基于改进 YOLOv8n 的目标位姿检测方法。通过引入 LSKA 注意力改造 SPPF 模块,设计轻量化模块 RGCSPELAN 以及改造损失函数 Wise-MPDIoU^{tinner} 以增强模型的特

征提取能力,提高模型的训练效果和检测性能以及实现模型的轻量化。最后,通过深度相机与 SRW-YOLOv8n 算法结合获取目标 4 个区域中心点的位置与深度信息,计算平面法向量构造目标实时坐标系,并完成抓取任务。实验结果表明,与 YOLOv8n 相比,本文算法的精确率提升了 0.2%,mAP@0.5 提升了 0.2%,参数量降低了

11%, 计算量降低了 16.0%, 推理速度提升了 33%, 实现了模型精度与轻量化之间的平衡, 证明了该方法的有效性和优越性。在执行 UR5 机械臂抓取任务时, 该算法能够在 2 s 完成航天器模型的位姿识别及抓取, 而原始算法的识别与抓取时

间为 2.5 s, 这表明该算法具备出色的实时性能。

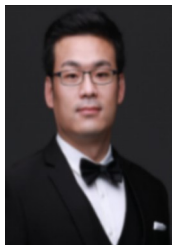
本文用 SRW-YOLOv8n 算法进行目标位姿检测, 其中只对模块进行改进, 未来的工作重点是通过改进模型架构来提升模型的精度与检测效率。

参考文献:

- [1] 陈敏佳, 盖绍彦, 达飞鹏, 等. 采用辅助学习的物体六自由度位姿估计[J]. 光学精密工程, 2024, 32(6): 901-914.
CHEN M J, GAI S H Y, DA F P, *et al.* Object 6-DoF pose estimation using auxiliary learning [J]. *Opt. Precision Eng.*, 2024, 32(6): 901-914. (in Chinese)
- [2] 周苒, 刘延芳, 齐乃明, 等. 面向空间应用的视觉位姿估计技术综述[J]. 光学精密工程, 2022, 30(20): 2538-2553.
ZHOU R, LIU Y F, QI N M, *et al.* Overview of visual pose estimation methods for space missions [J]. *Opt. Precision Eng.*, 2022, 30(20): 2538-2553. (in Chinese)
- [3] WANG C X, FANG H S, GOU M H, *et al.* Graspness discovery in clutters for fast and accurate grasp detection[C]. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV)*. October 10-17, 2021. Montreal, QC, Canada. IEEE, 2021: 15944-15953.
- [4] JABALAMELI A, BEHAL A. From single 2D depth image to gripper 6D pose estimation: a fast and robust algorithm for grabbing objects in cluttered scenes[J]. *Robotics*, 2019, 8(3): 63.
- [5] GE J Y, SHI J L, ZHOU Z Q, *et al.* A grasping posture estimation method based on 3D detection network [J]. *Computers and Electrical Engineering*, 2022, 100: 107896.
- [6] ZHANG L M, ZHOU X, LIU J Q, *et al.* Instance-level 6D pose estimation based on multi-task parameter sharing for robotic grasping [J]. *Scientific Reports*, 2024, 14(1): 7801.
- [7] YANG X Q, XIANG Y, ZHANG J Q, *et al.* Research on pixel-level grasp configuration prediction method based on deep neural network[C]. *IECON 2023- 49th Annual Conference of the IEEE Industrial Electronics Society*. October 16-19, 2023. Singapore, Singapore. IEEE, 2023: 1-5.
- [8] MENG M M, SUN X T, CHEN W H, *et al.* Multi-object recognition and grasping detection based on the anchor-free network[C]. 2023 *16th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*. October 28-30, 2023. Taizhou, China. IEEE, 2023: 1-6.
- [9] 阮国强, 曹雏清. 基于 PointNet++ 的机器人抓取姿态估计[J]. 仪表技术与传感器, 2023(5): 44-48.
RUAN G Q, CAO C Q. Robot grasping attitude estimation based on PointNet++ [J]. *Instrument Technique and Sensor*, 2023(5): 44-48. (in Chinese)
- [10] 李树春, 张静, 张华, 等. 面向机器人抓取过程中目标位姿估计方法[J]. 传感器与微系统, 2019, 38(7): 32-34, 38.
LI S C, ZHANG J, ZHANG H, *et al.* Object pose estimation method for robot grasping process [J]. *Transducer and Microsystem Technologies*, 2019, 38(7): 32-34, 38. (in Chinese)
- [11] JIANG P Y, ERGU D J, LIU F Y, *et al.* A review of yolo algorithm developments[J]. *Procedia Computer Science*, 2022, 199: 1066-1073.
- [12] LI X X, MA N, HAN Y H, *et al.* AHPPEBot: autonomous robot for tomato harvesting based on phenotyping and pose estimation[EB/OL]. 2024: 2405.06959. <https://arxiv.org/abs/2405.06959v1>
- [13] 巩大康. 服务机器人的抓取物体识别与位姿估计[D]. 南京: 南京信息工程大学, 2023: 1-7.
GONG D K. *Grabbing Object Recognition and Pose Estimation of Service Robot* [D]. Nanjing: Nanjing University of Information Science & Technology, 2023: 1-7. (in Chinese)
- [14] 李启煌. 基于注意力机制与 YOLO 相融合的 6D 位姿估计算法研究[D]. 桂林: 桂林电子科技大学, 2023: 2-8.
LI Q H. *Research on 6D Pose Estimation Algorithm Based on the Fusion of Attention Mechanism and YOLO* [D]. Guilin: Guilin University of Elec-

- tronic Technology, 2023; 2-8. (in Chinese)
- [15] LAU K W, PO L M, REHMAN Y A U. Large separable kernel attention: rethinking the large kernel attention design in CNN[J]. *Expert Systems with Applications*, 2024, 236: 121352.
- [16] 马超凡, 李翔, 王晓霞, 等. 基于重参数化的轻量化非机动车目标检测[J/OL]. 计算机工程与应用, 2024 (2024-05-23). <https://kns.cnki.net/kcms/detail/11.2127.tp.20240522.1323.006.html>. MA CH F, LI X, WANG X X, *et al.* 基于重参数化的轻量化非机动车目标检测[J/OL]. 计算机工程与应用, 2024 (2024-05-23). <https://kns.cnki.net/kcms/detail/11.2127.tp.20240522.1323.006.html>. (in Chinese)
- [17] ZHU L, WANG X J, KE Z H, *et al.* BiFormer: vision transformer with bi-level routing attention [C]. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 17-24, 2023. *Vancouver, BC, Canada*. IEEE, 2023; 10323-10333.

作者简介:



黄 成(1986—),男,黑龙江哈尔滨人,博士,副教授,博士生导师,2009年于哈尔滨理工大学获得学士学位,2012年于哈尔滨工程大学获得硕士学位,2018年于哈尔滨工业大学获得博士学位,主要研究方向为空间非合作目标捕获、目标检测和图像处理。E-mail: huangchengsunxi@163.com



苏俊飞(1997—),男,山西大同人,硕士研究生,2021年于哈尔滨理工大学获得学士学位,主要从事目标检测、图像处理的研究。E-mail: 563093788@qq.com