

基于SGV-YOLOv8模型的机械零件智能识别与抓取方法

罗杭 杨晔* 陈本永

浙江理工大学机械工程学院,杭州,310018

摘要:针对工业机器人抓取机械零件过程中零件识别速度慢、抓取成功率低等问题,提出了一种基于SGV-YOLOv8模型的机械零件智能识别与抓取方法。采用单目相机和激光测距模块构建深度视觉检测装置,实现机械零件三维定位;将YOLOv8模型作为基本架构,在骨干网络使用StarNet网络替换原有结构,并在颈部引入GSConv模块和VoV-GSCSP结构,实现了降低模型复杂程度的同时提高检测速度和抓取率。实验结果表明,与原模型相比,设计的SGV-YOLOv8模型(StarNet-GSConv-VoV YOLOv8)的模型参数量和浮点运算数(GFLOPs)分别下降了51.9%和51%,而每秒检测帧数(FPS)提高了37.6%;构建的工业机器人抓取装置的零件抓取成功率为80%。

关键词:机械臂抓取;机器视觉;激光测距模块;YOLOv8模型;零件识别

中图分类号:TP241.2

DOI:10.3969/j.issn.1004-132X.2026.02.019

开放科学(资源服务)标识码(OSID):



Intelligent Part Identification and Grabbing Method Based on SGV-YOLOv8 Model

LUO Hang YANG Ye* CHEN Benyong

School of Mechanical Engineering, Zhejiang Sci-Tech University, Hangzhou, 310018

Abstract: To solve the problems of slow part identification and low success rate in grabbing mechanical parts by industrial robots, an intelligent part identification and grabbing method was proposed based on SGV-YOLOv8 model. The monocular camera and laser ranging module were used to build a depth vision detection device to realize the three-dimensional positioning of mechanical parts; Taking the YOLOv8 model as the basic architecture, StarNet network was used in the backbone network to replace the original structure, and GSConv module and VoV-GSCSP structure were introduced in the neck, so as to reduce the complexity of the model and improve the detection speed and capture rate. The experimental results show that compared with the original model, the number of model parameters and the number of floating point operations (GFLOPs) of the designed SGV-YOLOv8 increases 51.9% and 51% respectively, while the number of detection frames per second (FPS) increases 37.6%; The success rate of part grasping in the constructed industrial robot grasping devices is 80%.

Key words: mechanical arm grabbing; machine vision; laser ranging module; YOLOv8 model; part identification

0 引言

采用工业机器人实现自动抓取在工业生产中具有重要的应用。机器视觉是工业机器人零件识别与抓取的关键技术之一^[1-2]。以往一般采用双目立体视觉或激光雷达的方式,但是双目视觉需要根据立体视差计算每个像素的深度^[3],以双目视觉为原理的立体深度相机的目标定位取决于对应匹配的水平、图像质量、重复场景和光照变化^[4]。激光雷达虽然能够提供精确的距离测量信息,但

其硬件成本较高,不适合大规模推广。因此,在最近的研究和工业应用中,单点激光测距仪和相机的集成越来越多地被投入使用^[5]。PATEL等^[6]使用基于激光的测距仪、方向传感器和摄像头获取物体的绝对三维位置。WITHER等^[7]将小型单点激光测距仪集成到可穿戴系统中,移动设备也可以方便地配备单点激光测距仪。本文提出了一种低成本、相对简单且精度更高的解决方案,即单目相机与激光测距模块相结合。该方案尤其适合于应用于六轴串联机械臂的工业抓取任务。

在机械臂抓取任务中,零件目标识别是重要的环节之一。在实际生产过程中,工业零件具有

收稿日期:2025-01-15

基金项目:浙江省科技计划(2024C01174)

多尺度、形状复杂、相互遮盖的特点,并且对识别速度有较高要求。传统的人为特征提取图像处理算法受外界环境的影响,外部光照、摆放位置、拍摄角度等原因都会造成机器人抓取的失败^[8]。HINTON等^[9]提出了神经网络的概念,从此神经网络在目标检测领域得到大量应用。目标检测算法主要分为一阶段算法和二阶段算法,二阶段算法以R-CNN^[10]为代表,它为后来出现的SPP-net^[11]、Fast R-CNN^[12]、Faster R-CNN^[13]、R-FCN^[14]开创了一种基于候选区域的深度学习目标检测思路^[15]。二阶段算法的问题是训练速度慢,不利于工业零件检测。为解决该问题,REDMON等^[16]提出了一阶段算法YOLO(you only look once),直接在图像上回归物体的边界框和类别概率而无需分为两步。为了改进YOLO算法对密集目标识别准确度不高的问题,LIU等^[17]提出SSD目标检测算法,引入锚框(anchor box)和交并比的概念对标记框进行检验,但该算法对小目标的识别能力不够高。此后,YOLOv2、YOLOv3算法相继被提出。然而,深度学习模型通常需要较大的计算资源和存储空间,这对设备硬件的要求较高。

针对上述存在问题,本文提出了一种SGV-YOLOv8轻量化网络模型对目标零件进行识别,该模型在骨干部分使用StarNet网络代替原有结构,并使用GSCONV模块和VoV-GSCSP模块细化颈部。使用改进后的SGV-YOLOv8网络对零件进行实时检测,再结合激光测距模块,最终实现机械臂的零件抓取。

1 基于SGV-YOLOv8模型的机械零件智能识别与抓取系统设计

机械零件智能识别与抓取系统主要由三部分组成:视觉系统、机械臂和SGV-YOLOv8识别网络。视觉系统包括单目相机和激光测距模块,机械臂末端安装机械爪抓取零件,SGV-YOLOv8网络识别零件种类及抓取点。将机械臂基座设立为世界坐标系 $\{O_w X_w Y_w Z_w\}$,机械爪末端设立工具坐标系 $\{O_h X_h Y_h Z_h\}$,相机坐标系设为 $\{O_c X_c Y_c Z_c\}$,激光器坐标系设为 $\{O_l X_l Y_l Z_l\}$ 。首先,零件的位置信息经过2D相机获取,再利用SGV-YOLOv8网络确定零件种类与抓取点,获得抓取点在相机内的像素坐标。根据相机和激光器的联合标定结果,利用激光测距模块获得距离,最终得到抓取点在世界坐标系下的坐标,并将该坐标发送给机械臂。机械臂按照计算的轨迹路径运动,最终抓取目标零件。系统示意图见图1。机

械臂对零件进行抓取操作至少需要三个坐标: X 、 Y 、 Z (深度值)^[18]。其中 X 和 Y 坐标可以通过单目相机获得, Z 坐标则通过激光测距模块的结果得到。

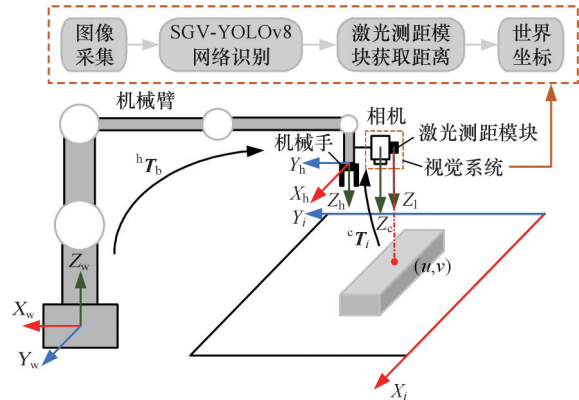


图1 零件智能识别与抓取系统示意图

Fig.1 Schematic diagram of intelligent part identification and grabbing system

在相机坐标系下检测激光光斑,得到光斑位于相机坐标系下的表达式:

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = \begin{bmatrix} L \cos \theta_x + t_x \\ L \cos \theta_y + t_y \\ L \cos \theta_z + t_z \end{bmatrix} \quad (1)$$

其中, t_x 、 t_y 、 t_z 为相机中心与激光测距模块之间的偏移量; L 是激光测距模块测出的距离; θ_x 、 θ_y 、 θ_z 是激光与相机坐标系夹角,如图2所示。根据几何关系,可得 $\cos^2 \theta_x + \cos^2 \theta_y + \cos^2 \theta_z = 1$ 。

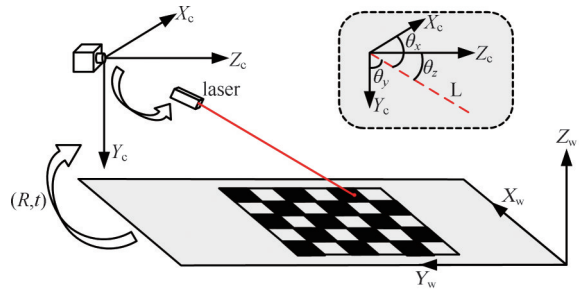


图2 激光测距模块和相机的标定示意图

Fig.2 Calibration diagram of laser ranging module and camera

根据针孔模型,可得相机坐标 $P = (X_c, Y_c, Z_c)^T$ 和图像坐标 $p = (u, v)^T$ 的关系:

$$p \sim KP \quad (2)$$

其中, K 为相机内参矩阵。又因像素坐标系的实际坐标在图像中心而非图像左上角,结合式(1)可得

$$\frac{f_x(L \cos \theta_x + t_x)}{L \sqrt{1 - \cos^2 \theta_x - \cos^2 \theta_y} + t_z} + u_0 = u \quad (3)$$

$$\frac{f_y(L \cos \theta_y + t_y)}{L \sqrt{1 - \cos^2 \theta_x - \cos^2 \theta_y} + t_z} + v_0 = v \quad (4)$$

其中, f_x 、 f_y 是图像轴上的焦距; u_0 、 v_0 是图像投影中心。最后利用最小二乘法求解 θ_x 、 θ_y 、 t_x 、 t_y 、 t_z 。

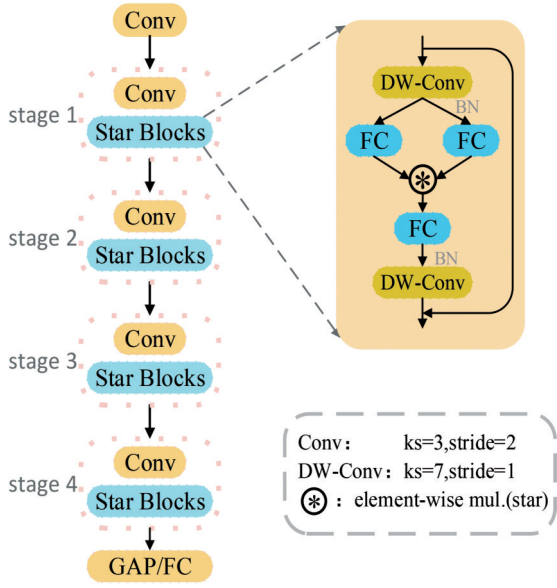


图4 StarNet网络结构

Fig.4 StarNet network architecture

其中,将权重和偏差分别表示为 $W=[W \ B]^T$, $X=[X \ 1]^T$. 定义 $w_1, w_2, x \in \mathbb{R}^{(d+1) \times 1}$, d 为输入通道号。

单个通道可以扩展成多个输出通道和多个特征元素, $W_1, W_2 \in \mathbb{R}^{(d+1) \times (d+1)}$, $X \in \mathbb{R}^{(d+1) \times n}$, 即

$$w_1^T x * w_2^T x = \left(\sum_{d+1}^{i=1} w_1^i x^i \right) * \left(\sum_{d+1}^{j=1} w_2^j x^j \right) = \sum_{i=1}^{d+1} \sum_{j=1}^{d+1} w_1^i w_2^j x^i x^j = \underbrace{\alpha_{(1,1)} x^1 x^1 + \dots + \alpha_{(4,5)} x^4 x^5 + \dots + \alpha_{(d+1,d+1)} x^{d+1} x^{d+1}}_{(d+2)(d+1)/2} \quad (7)$$

$$\alpha_{(i,j)} = \begin{cases} w_1^i w_2^j & i=j \\ w_1^i w_2^j + w_1^j w_2^i & i \neq j \end{cases} \quad (8)$$

其中, i, j 为索引通道; α 是每一项的系数。从式(8)可以看出,除了第 $\alpha_{(d+1,d+1)} x^{d+1} x^{d+1}$ 项外,其余项均与 x 项成非线性关系,即独立的维度。因此,在 d 维空间中进行星形运算,可得

$$\frac{(d+2)(d+1)}{2} \approx \left(\frac{d}{\sqrt{2}} \right)^2$$

维空间的表达,显著放大了特征维度。

2.3 YOLOv8 颈部网络的改进

YOLOv8 网络的颈部采用 4 次 CSP (cross stage partial) 的思想,这一设计虽然提高了特征融合能力,但也引入了大量的卷积操作。由于标准卷积参数量较大,导致模型的计算负担显著增加,故本文采用 GSConv 卷积代替传统卷积,其结构如图 5 的右侧部分所示。

GSConv 模块主要由普通卷积与深度可分离卷积 DW-Conv (depthwise separable convolution)

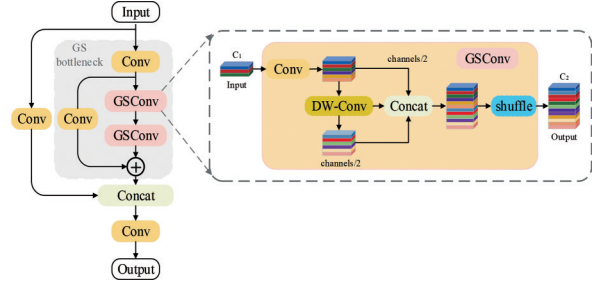


图5 VoV-GSCSP模块及GSConv模块结构

Fig.5 Architecture of VoV-GSCSP module and GSConv module

构成。其中,深度可分离卷积能够有效减少参数量和浮点运算量,但其设计会导致通道信息分离,降低特征提取的完整性;而标准卷积虽然特征提取能力更强,却存在计算量较大的问题。GSConv 卷积将两者结合,将标准卷积生成的信息渗透至深度可分离卷积的每个部分中^[19],并在洗牌任务 (shuffle) 中使用线性操作,进一步促进信息的高效融合,从而显著增强了特征的表达能力。

通常,卷积计算的时间复杂度是由 GFLOPs 来定义的,标准卷积 (SC) 和 GSConv 的时间复杂度为

$$Time_{SC} \sim O(W \cdot H \cdot K_1 \cdot K_2 \cdot C_1 \cdot C_2) \quad (9)$$

$$Time_{GSConv} \sim O \left[W \cdot H \cdot K_1 \cdot K_2 \cdot \frac{C_2}{2} (C_1 + 1) \right] \quad (10)$$

其中, W, H 为输出特征图的高度和宽度; K_1, K_2 为卷积核大小; C_1, C_2 为输入和输出特征图的通道数。从式(9)和式(10)可看出, C_1 值越大, GSConv 卷积的计算成本越接近 SC 卷积的 50%, 显著降低了模型的时间复杂度。

C2f 模块是 YOLOv8 网络提取图像特征最重要的模块之一,其中包含大量卷积过程。为了实现网络轻量化,并在保持模型速度的同时捕捉更丰富的特征信息,在 GSConv 模块的基础上使用 VoV-GSCSP 模块代替 C2f 模块,有助于模型保持速度的同时捕捉更丰富的特征信息。

VoV-GSCSP 模块结构如图 5 所示。该模块在 GSConv 的基础上引入不同尺度的卷积核捕捉图像中的多种特征,使用单次聚合方法设计跨阶段聚合网络模型,在不同层上进行特征聚合。VoV-GSCSP 模块将具有 C_1 个通道的输入分为两部分,每个分支通过一个卷积层处理,将通道数减半。其中进入 GS bottleneck 模块的分支经过两次 GSConv 层,再与原始输入特征图相加,得到输出特征图。随后,两个分支的特征图进行拼接操作,并通过一次最终卷积层进一步融合和提炼

特征,从而提高模型的适应性和鲁棒性。

3 实验设计及结果分析

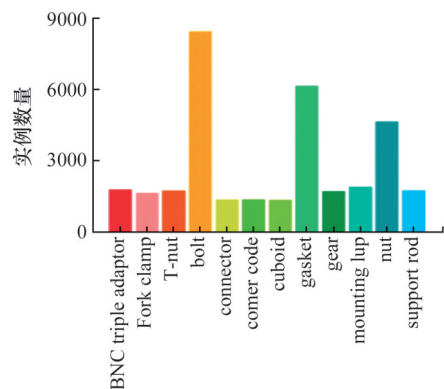
3.1 图像采集和数据集处理

为验证模型的泛化能力,本文选择在自制的数据集 Self Parts 以及公开数据集 Industrial Tool 上分别进行实验。Self Parts 数据集为本次实验的主要数据集,在该数据集的基础上进行基础框架的验证以及消融实验。

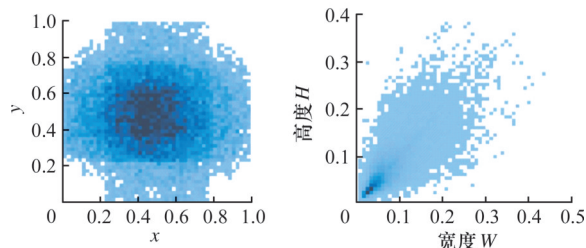
Industrial Tool 数据集包含 2100 张带标签的图像。该数据集所涵盖的对象为工业领域中常见的 13 类零件,具体包括双六角柱(double hexagonal column)、法兰螺母(flange nut)、六角螺母(hexagon nut)、六角柱(hexagon pillar)、六角螺丝(hexagon screw)、六角钢柱(hexagonal steel column)、水平气泡仪(horizontal bubble)、键条(key-bar)、塑料缓冲柱(plastic cushion pillar)、方形螺母(rectangular nut)、圆头螺丝(round head screw)、弹簧垫圈(spring washer)以及 T 形螺丝(T-shaped screw)。数据集中大部分零件目标尺寸较小,这为评估模型对小目标的检测能力提供了理想的测试场景。本实验采取 7:2:1 的比例进行随机抽取并组成训练集、验证集和测试集。

Self Parts 数据集中的图像均使用分辨率为 1920 像素 \times 1080 像素的普通工业相机进行拍摄,拍摄的图像以 PNG 格式进行保存。因为工业零件抓取一般在室内进行,故拍摄场景选择室内,设置了强光和弱光两种可能,进行多角度拍摄。另外考虑到实际生产过程中零件摆放方式各有不同,故在图像采集过程中特意设置了密集零件相互重叠和遮挡的情况。零件类型包括三通转接头(BNC triple adaptor),三叉压板(fork clamp),垫片(gasket),螺栓(bolt),螺母(nut),T 型螺母(T-nut),齿轮(gear),安装耳(mounting lug),光学支撑杆(support rod)在内共计 12 种零件,其中包含大目标零件和小目标零件。对可用的零件图像使用 Labelme 进行标注后采用翻转、随机噪声、模糊、剪影方法对图像进行数据增强。最终保存了 4902 张图片,并按照 7:2:1 的比例进行随机抽取,组成训练集、验证集和测试集。

图 6 所示为整个 Self Parts 数据集情况的可视化。分析可知,在该自建数据集中,虽然垫片和螺栓螺母的数量较多,但是符合工业抓取的实际情况,尺寸特征呈现一定的线性关系,该数据集适用于工业零件识别场景。

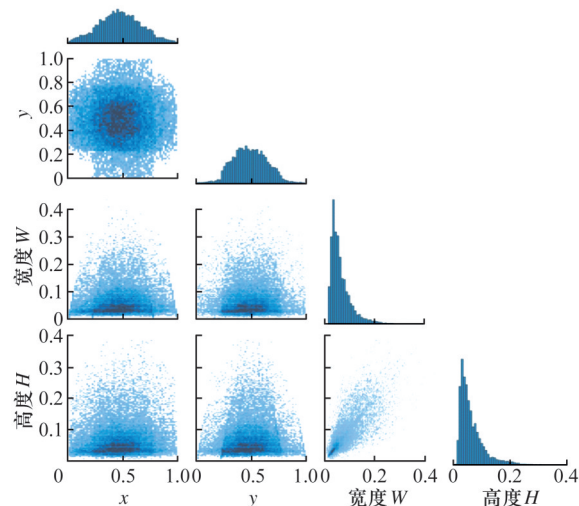


(a) 数据集类别数量



(b) 检测目标中心点分布情况

(c) 标签宽度与高度分布情况



(d) 原始数据集中分布的细节

图 6 数据集的可视化

Fig.6 Visualization of data sets

3.2 模型训练环境与评估指标

3.2.1 实验环境

本实验进行网络训练时使用的硬件配置为 Windows Server 2022 的操作系统和 Intel(R) Xeon(R) Gold 5118 CPU @ 2.30 GHz 处理器,内存为 384 GB, GPU 为 NVIDIA RTX 3090。网络的编译语言使用 Python 3.8,编译平台为 Visual Studio 2022,并且使用 PyTorch2.1.0 作为网络的深度学习框架,安装 CUDA12.1 进行加速。模型的超参数配置见表 1。

3.2.2 评价指标

本研究旨在几乎不降低检测精度的前提下减少模型的计算量和检测时间,从而减少对硬件的

表 1 模型超参数设置

Tab.1 Model super parameter setting

网络配置项	结构参数
Epochs	150
lr0	0.01
Optimizer	SGD
Momentum	0.937
Weight decay	0.0005
Batch size	64

需求,故本实验选择了以下6个指标对模型进行评估:参数规模、浮点运算数(GFLOPs)、精度(precision)、召回率(recall)、平均精度(mAP@0.5)、每秒帧数(frames per second, FPS),具体公式略。FPS值、参数规模和浮点运算是衡量网络大小的重要评价指标。FPS值越高、参数规模和浮点运算越小,则网络对硬件的要求越低,更适合实际应用,且能及时检测。

3.3 实验结果与分析

3.3.1 基准框架的验证

为全面评估不同基准框架的目标检测性能,本实验选择YOLO系列中的其他模型以及二阶段算法Faster-R-CNN作为参照模型,在Self Parts零件数据集上进行实验验证,从参数规模、GFLOPs、推理速度(FPS)和精度(mAP@0.5)等多个关键维度进行综合考量,旨在深入对比不同基准框架的模型大小和推理能力。对比实验结果见表2。

表 2 Self Parts 零件数据集在不同算法上的性能比较

Tab.2 Performance comparison of part datasets on different algorithms

模型	参数规模/MB	GFLOPs/G	推理速度/(帧·s ⁻¹)	mAP@0.5/%
Faster R-CNN	140.8	406.6	10	83.1
SSD	50.2	360.9	107	75.7
YOLOv5	13.8	15.9	11	98.8
YOLOv6s	8.3	11.8	263	98.0
YOLOv8m	49.6	79.1	208	99.2
YOLOv8n	6.0	8.9	164	99.0
YOLOv8s	21.4	28.8	303	98.9

Self Parts 零件数据集包括大小范围在11~100 mm的工业零件。该数据集具有目标尺度变化明显、零件间存在相互遮挡等特点,这会导致图像背景的复杂度增加。此外,大多数零件的尺寸较小,这给目标检测带来了一些障碍。在实验设置方面,为确保对比的公平性和一致性,所有参与对比试验的模型均将输入图像统一设置为640像素×640像素,批大小均设置为16,epoch均设置为100。

实验结果表明,尽管YOLOv8m在检测性能上表现最优,但它的参数规模大小是YOLOv8s的2.32倍,GFLOPs是YOLOv8s的2.75倍。在推理速度上,YOLOv8s每秒检测的图像帧数比YOLOv8m多95帧。虽然YOLOv8s的mAP@0.5略低于YOLOv8m等框架,但它在参数规模、GFLOPs和推理速度方面展现出显著优势。综合考虑模型的综合性能和实际应用需求,最终选定YOLOv8s作为基线网络。

3.3.2 YOLOv8的骨干网络优化结果与分析

为验证2.2节中针对YOLOv8模型骨干网络改进的有效性,选择几种主流轻量化网络与StarNet网络进行对比,分别是MobileNet, ShuffleNet, GhostNet, FasterNet,并在Self Parts数据集上进行骨干网络的优化实验。实验结果见表3。

表 3 不同骨干网络的比较

Tab.3 Comparison of different backbone networks

模型	参数规模/MB	GFLOPs/G	mAP@0.5/%
MobileNet	11.2	22.6	98.5
ShuffleNet	12.4	17.4	98.7
GhostNet	12.4	17.3	98.6
FasterNet	16.7	21.7	99.0
StarNet	11.1	17.3	98.7

从检测精度层面分析,针对Self Parts数据集,以下5种轻量化网络均展现出较高的检测精度。其中,FasterNet取得了最高的检测精度,其mAP@0.5为99.0%,而StarNet网络的mAP@0.5为98.7%,FasterNet仅比StarNet高出0.3%。然而,在参数规模和计算复杂度方面,两者存在显著差异。FasterNet的参数规模达16.7 MB,相比StarNet的11.1 MB高出50.5%;其GFLOPs为21.7 G,相较于StarNet的17.3 G高出25.4%。

从轻量化网络的角度考虑,本文选择StarNet网络作为YOLOv8的骨干网络,它可以在模型的参数量大小和检测精度之间找到平衡,保证轻量化的同时兼顾检测精度。

3.3.3 消融实验

为验证SGV-YOLOv8模型的有效性及其轻量化结果,本文在Self Parts零件数据集上进行了一组消融实验,以评估SGV-YOLOv8网络中每个模块的单独影响,实验结果见表4。

本研究的基准网络为YOLOv8s网络模型,通过使用StarNet网络替换YOLOv8s的骨干部分,使得模型参数规模从21.4 MB减至11.1 MB,减少约50.9%,同时GFLOPs从28.8 G降至17.3

表4 YOLOv8的消融实验结果

Tab.4 Ablation test results of YOLOv8

	原始 YOLOv8 网络	YOLOv8 网络+ StarNet	YOLOv8 网络+ GSConv	YOLOv8 网络+ VoV-GSCSP	YOLOv8 网络+ StarNet+ GSConv	YOLOv8 网络+ StarNet+ VoV-GSCSP	YOLOv8 网络+ GSConv+ VoV-GSCSP	原始 YOLOv8 网络+ StarNet+ GSConv+ VoV-GSCSP
YOLOv8	✓	✓	✓	✓	✓	✓	✓	✓
StarNet		✓			✓	✓		✓
GSConv			✓		✓		✓	✓
VoV-GSCSP				✓		✓	✓	✓
参数规模/MB	21.4	11.1	5.81	19.3	12.0	12.7	19.9	11.1
GFLOPs/G	28.8	17.3	26.2	21.3	16.9	17.3	25.1	14.1
推理速度/(帧·s ⁻¹)	303.4	384.6	277.5	286	323.3	344.9	293.7	417.2
mAP@0.5/%	98.9	98.7	99.0	98.9	98.5	98.7	99.2	98.9

G,降低约40%。尽管如此,推理速度显著提高至384.6帧/s,提高了26.7%。引入GSConv模块在颈部代替普通卷积后,模型参数规模进一步减少至5.81MB,相较于基准模型减少约72.8%,同时GFLOPs降至26.2G,这一改进使mAP@0.5提高至99.0%。此外,在颈部结构中,采用更加精简的VoV-GSCSP模块替代C2f模块后,参数规模减少至19.3MB,相较于基准模型减少约9.8%,同时GFLOPs降至21.3G。这一改进使网络在mAP@0.5不变的前提下实现了参数和计算量的进一步减少。

综合上述策略对模型进行改进后,最终得到的SGV-YOLOv8模型与YOLOv8s相比,在mAP仍然保持98.9%高精度的情况下,参数规模仅为原模型的51.9%,GFLOPs降低了51%,并且推理速度显著提高至417.2帧/s,相较于基准模型提高约114帧/s。上述实验结果充分证明了各个改进模块对网络轻量化的有效性,以及SGV-YOLOv8模型在实现高效目标检测任务中的优越性能。

为直观展示SGV-YOLOv8模型在消融实验过程中结合各个模块所实现的性能提升,将其转化为散点图形式,如图7所示。图中A表示基线网络YOLOv8s,在A的基础上分别加入StarNet、GSConv、VoVGSCSP、StarNet+GSConv、StarNet+VoVGSCSP、GSConv+VoVGSCSP、StarNet+GSConv+VoVGSCSP,分别记为B、C、D、E、F、G、H。通过在YOLOv8s中逐步增加各个模块评估SGV-YOLOv8的性能。可以直观看出,SGV-YOLOv8的表现明显优于YOLOv8s,在有效实现模型轻量化的同时保持了较高的检测精度。此外,一系列的改进将模型的推理速度从

303.4提高到417.2,使其更适用于小目标、更快地抓取,更有利于模型部署在设备上。

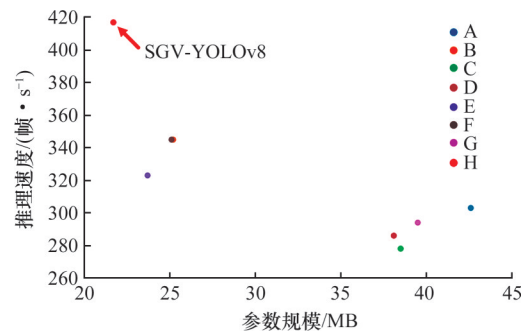


图7 消融实验结果散点图

Fig.7 Visualization of data sets

图8所示为实验室环境中各种零件目标检测结果的对比,表明改进后的YOLOv8网络提高了检测效率并降低了误检率。图9所示为YOLOv8网络和改进后的SGV-YOLOv8网络在Self Parts零件数据集的实际检测效果对比。综上,SGV-YOLOv8有更好的性能。

3.3.4 模型泛化性实验

为验证模型的泛化性,本文在公开数据集Industrial Tool上进行试验,并与其他经典算法进行对比。实验结果见表5。可以看出,在Industrial Tool公共数据集上,改进的SGV-YOLOv8网络在mAP@0.5和mAP@0.5:0.95上相较于YOLOv8s分别提高0.3%和1.8%,推理速度则从312.5帧/s提高至344.8帧/s,提高10.3%,模型参数则减小了48.1%。与其他模型相比,SGV-YOLOv8不仅在精度上表现出色,拥有最高的mAP@0.5和mAP@0.5:0.95,而且在推理速度上也达到了最快,为344.8帧/s。尽管SGV-YOLOv8的参数规模和浮点运算数不是最小的,但其综合性能显著优于其他模型,在保持较高精度

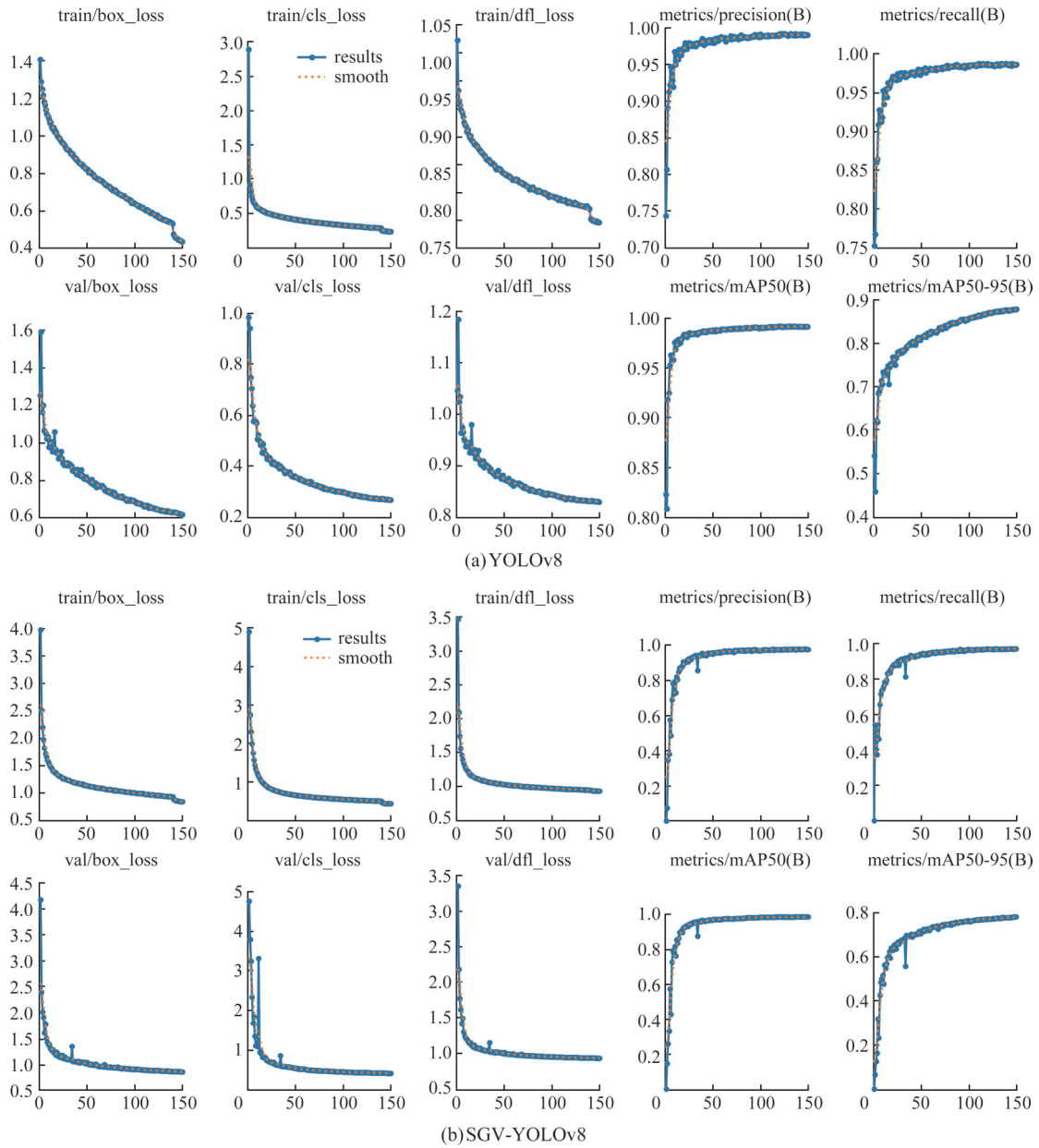


图8 YOLOv8和SGV-YOLOv8网络模型的结果对比

Fig.8 Comparison of results between YOLOv8 and SGV-YOLOv8 network

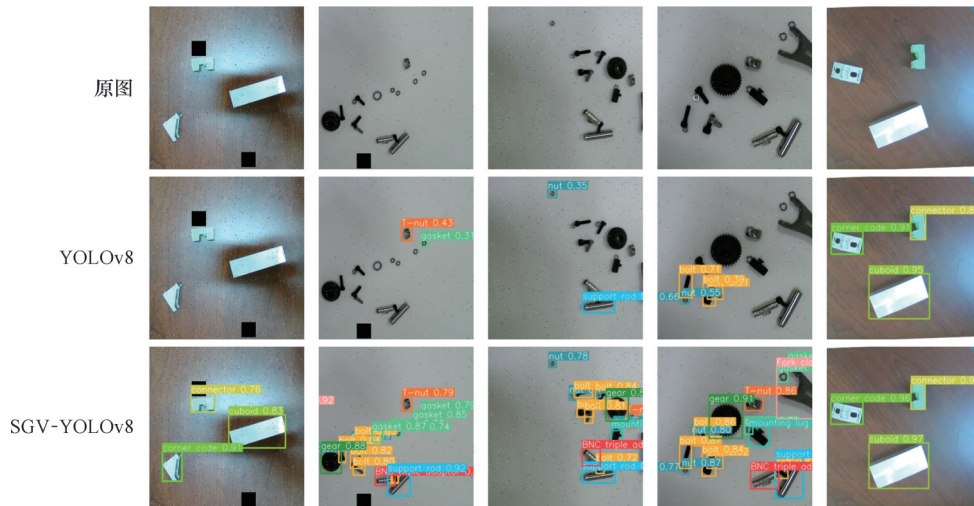


图9 不同模型对所选测试集检测对比

Fig.9 Comparison of different models for detection of selected test set

表5 Industrial Tool数据集上的泛化实验

Tab.5 Generalization experiments on industrial tool datasets

模型	参数规模/MB	GFLOPs/G	推理速度FPS/(帧·s ⁻¹)	mAP@0.5/%	mAP@0.5:0.95/%
YOLOv5	13.8	15.9	21.2	99.2	76.9
YOLOv6s	8.3	11.8	266.0	98.7	74.4
YOLOv8n	6.0	8.9	128.2	99.0	77.9
YOLOv8s	21.4	28.8	312.5	98.9	76.8
YOLOv8m	49.6	79.1	288.9	99.1	77.4
SGV-YOLOv8	11.1	14.1	344.8	99.2	78.6

的同时实现了更快的推理速度和更小的模型参数量。综上,SGV-YOLOv8在公开数据集上同样有优异表现,证明其具有泛化性。

3.3.5 机械臂抓取零件实验结果分析

为验证改进后的SGV-YOLOv8算法在抓取机器人视觉系统中的性能表现,本研究将该算法通过香橙派部署于相机端进行实时目标识别,并在自制的六自由度串联机械臂平台上开展实验测试,如图10所示。



图10 机械臂实物图

Fig.10 Physical image of mechanical arm

本次实验在实验室环境下开展,机器人抓取零件流程主要包括视觉识别、激光测距模块测量及抓取执行三个关键环节。首先通过单目相机获取目标零件的图像信息,利用目标检测算法对零件进行种类和抓取点的识别;识别完成后,移动激光测量模块获取距离值,并将抓取点的坐标赋值给机械臂进行工作;最后机械臂根据规划路径运动至目标位置,由末端执行器完成精确抓取操作。

在本次实验中,让采用不同识别模型的机械臂均进行30次的零件抓取,结果见表6。使用YOLOv8网络进行零件识别时,抓取成功率为70%,其中7次由于定位失败,2次因为网络未能正确识别零件而失败。而使用改进后的SGV-YOLOv8网络进行零件识别时,零件均可以被正

确识别,抓取成功率提高至80%。结果表明,改进后的SGV-YOLOv8算法可以更好地应用于机械臂,帮助实现零件的抓取。

表6 基于改进YOLOv8的机械臂零件抓取结果

Tab.6 Robot arm part grab results based on improved YOLO8

模型	试验次数	定位失败的零件数量	识别错误的零件数量	成功抓取次数	成功率/%
YOLOv8	30	7	2	21	70
SGV-YOLOv8	30	6	0	24	80

4 结语

本文提出了一种轻量化SGV-YOLOv8目标识别算法和激光测距模块相结合的方法,用于零件的识别和定位,解决了零件目标检测中计算量大、模型尺寸大、检测速度慢的问题。SGV-YOLOv8使用StarNet网络对YOLOv8s的骨干网络部分进行重构,降低模型复杂度,同时增加网络的特征提取能力。此外,改进后的模型在颈部使用GSCConv模块代替普通卷积,使用VoV-GSCSP替换C2f模块细化颈部。通过消融实验,改进后的SGV-YOLOv8网络在自制的工业零件数据集上优于原有的YOLOv8网络,在mAP@0.5保持98.9%的高精度的情况下,网络的参数量从21.4 MB下降至11.1 MB,GFLOPs从28.8 G下降至14.1 G,分别下降了48.1%和51%。而检测速度提高37.5%,从303.4帧/s增加至417.2帧/s。实验结果验证了改进后的网络对零件数据集的有效性和适用性。本文将改进后的SGV-YOLOv8目标识别算法与激光测距模块相结合,提出一种零件定位方法,并应用于自制的零件抓取机械臂上,抓取成功率从70%提高至80%,验证了该方法的有效性。

本研究为实现在机械臂上对工业零件目标进行更快速、准确的检验提供了方法支持,但SGV-YOLOv8也存在不足,如对小目标零件识别的准确度有待提高。未来将致力于增强其对不同场景和环境条件的适应性,通过探索更复杂的深度学习技术,并结合该领域的最新进展,算法的推理速度将得到进一步提高。此外,本实验考虑到实际抓取情况导致样本的不平衡问题,未来将对数量少的类别进行相应的数据扩充。由于本实验的机械臂为自制的机械臂,定位精度也有待提高。

参考文献:

- [1] 谢丰隆,韩建海,李向攀.一种快速的机器人固定视觉标

- 定方法[J]. 机械设计与制造, 2018(11): 237-240.
- XIE Fenglong, HAN Jianhai, LI Xiangpan. A Fast Way of Stable Camera Calibration with Robot[J]. Machinery Design & Manufacture, 2018(11): 237-240.
- [2] 那一鸣, 胡超, 邱业余, 等. 基于机器视觉的汽车车门三维定位引导[J]. 中国机械工程, 2024, 35(9): 1677-1687.
- NA Yiming, HU Chao, QIU Yeyu, et al. Three-dimensional Positioning Guidance of Automobile Doors Based on Machine Vision [J]. China Mechanical Engineering, 2024, 35(9): 1677-1687.
- [3] NAKAGUCHI V M, LIU Zifu, et al. 3D Camera and Single-point Laser Sensor Integration for Apple Localization in Spindle-type Orchard Systems[J]. Sensors, 2024, 24(12): 3753.
- [4] LUHMANN T, FRASER C, MAAS H G. Sensor Modelling and Camera Calibration for Close-range Photogrammetry[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2016, 115: 37-46.
- [5] LIU Zewei, LU Dongming, QIAN Weixian, et al. Calibration of a Single-point Laser Range Finder and a Camera [J]. Optical and Quantum Electronics, 2018, 50(12): 447.
- [6] PATEL S N, REKIMOTO J, ABOWD G D. ICam: Precise At-a-distance Interaction in the Physical Environment [C]//Pervasive Computing. Berlin, 2006: 272-287.
- [7] WITHER J, COFFIN C, VENTURA J, et al. Fast Annotation and Modeling with a Single-point Laser Range Finder [C]//2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality. Cambridge, 2008: 65-68.
- [8] 吕张成, 张建业, 陈哲钥, 等. 基于深度学习的工业零件识别与抓取实时检测算法[J]. 机床与液压, 2023, 51(24): 33-38.
- LYU Zhangcheng, ZHANG Jianye, CHEN Zheyao, et al. Real-time Detection Algorithm for Industrial Parts Recognition and Grabbing Based on Deep Learning [J]. Machine Tool & Hydraulics, 2023, 51(24): 33-38.
- [9] HINTON G E, SALAKHUTDINOV R R. Reducing the Dimensionality of Data with Neural Networks[J]. Science, 2006, 313(5786): 504-507.
- [10] 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. Columbus, 2014: 580-587.
- [11] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition [C]//Computer Vision - ECCV 2014. Cham, 2014: 346-361.
- [12] GIRSHICK R. Fast R-CNN [C]//2015 IEEE International Conference on Computer Vision (ICCV). Santiago, 2015: 1440-1448.
- [13] REN Shaoqing, HE Kaiming, GIRSHICK R, et al. Faster R-CNN: Towards Real-time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [14] DAI J, LI Y, HE K, et al. R-FCN: Object Detection via Region-based Fully Convolutional Network[C]//30th Conference on Neural Information Processing Systems. Barcelona, 2016: 379-387.
- [15] 黎洲, 黄妙华. 基于YOLO_v2模型的车辆实时检测[J]. 中国机械工程, 2018, 29(15): 1869-1874.
- LI Zhou, HUANG Miaohua. Vehicle Detections Based on YOLO_v2 in Real-time [J]. China Mechanical Engineering, 2018, 29(15): 1869-1874.
- [16] 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 (CVPR). Las Vegas, 2016: 779-788.
- [17] LIU Wei, ANGUOLOV D, ERHAN D, et al. SSD: Single Shot MultiBox Detector [C]//Computer Vision - ECCV 2016. Cham, 2016: 21-37.
- [18] MA Xu, DAI Xiyang, BAI Yue, et al. Rewrite the Stars [C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, 2024: 5694-5703.
- [19] LI H, LI J, WEI H, et al. Slim-neck by GSConv: a Better Design Paradigm of Detector Architectures for Autonomous Vehicles [J]. arXiv Preprint arXiv: 2206.02424, 2022.

(编辑 陈勇)

作者简介: 罗杭, 女, 2000年生, 硕士研究生。研究方向为机器人控制。杨晔*(通信作者), 男, 1977年生, 实验师。研究方向为工业机器人的设计与控制。E-mail: hooohoo9999@zstu.edu.cn。

本文引用格式:

罗杭, 杨晔, 陈本永. 基于SGV-YOLOv8模型的机械零件智能识别与抓取方法[J]. 中国机械工程, 2026, 37(2): 442-451.

LUO Hang, YANG Ye, CHEN Benyong. Intelligent Part Identification and Grabbing Method Based on SGV-YOLOv8 Model[J]. China Mechanical Engineering, 2026, 37(2): 442-451.