

文章编号 1004-924X(2025)19-3150-12

基于单目视觉的集装箱角件自动对位

张永雷, 胡 涛*, 申立群

(哈尔滨工业大学 仪器科学与工程学院, 黑龙江 哈尔滨 150001)

摘要: 为了提高集装箱装卸效率, 实现装卸自动化, 基于单目视觉, 结合深度学习和图像处理算法, 提出了一种吊具与集装箱的自动对位技术。以单目相机拍摄的集装箱吊装工况图像为研究对象, 分析角件区域特征。针对角件在高分辨率图像中像素占比低的问题, 提出“粗定位-精分割”的处理方法。然后, 基于分割结果检测关键特征点, 构建 2D-3D 点对, 通过 L-M 算法完成位姿估计。最后, 在实验室 AUBO-i10 机械臂对位平台上开展自动对位实验, 验证所提算法的有效性。实验结果表明, 在实验室环境和真实场景下, 集装箱角件的检测精度 mAP 值均达到 95% 以上, 角件分割精度 mIoU 值分别达到 98.15% 和 93.89%, 相比原模型 SegFormer-B0 分别提高 1.24% 和 1.64%, 模型计算量下降约 23.2%; 在相机距离集装箱角件 2 m 左右的情况下, 对位位置瞄准误差小于 1.0 mm, 在 X, Y, Z 3 个轴上自动对位的对位平移绝对误差小于 5.0 mm, 旋转绝对误差小于 0.5°。结果证明, 本文所提算法精度可靠, 满足单角件自动对位需求。

关键词: 单目视觉; 集装箱; 自动对位; 目标识别; 位姿估计

中图分类号: TH247; TP29 **文献标识码:** A

doi: 10.37188/OPE.20253319.3150

CSTR: 32169.14.OPE.20253319.3150

Automatic alignment of container corner based on monocular vision

ZHANG Yonglei, HU Tao*, SHEN Liqun

(School of Instrumentation Science and Engineering, Harbin Institute of Technology,
Harbin 150001, China)

* Corresponding author, E-mail: hutao@hit.edu.cn

Abstract: To enhance the efficiency of container loading and unloading and realize automated operations, this study investigates monocular-vision-based automatic alignment between spreader and container, integrating deep learning and image-processing techniques. Monocular images of container hoisting conditions were analyzed with emphasis on the regional characteristics of corner components. To address the low pixel proportion of corner regions in high-resolution images, a two-stage "coarse positioning-fine segmentation" strategy was proposed. Based on the segmentation results, key feature points were detected, 2D-3D point correspondences were established, and pose estimation was performed using the Levenberg-Marquardt algorithm. Validation was conducted on an AUBO-i10 manipulator alignment platform in laboratory settings. Experimental results demonstrate that the mean average precision (mAP) for detection of container corner components exceeds 95% in both laboratory and real-scene environments. Mean intersection-

收稿日期: 2025-06-16; 修订日期: 2025-07-20.

基金项目: 航天联合基金资助项目 (No. Y22144)

over-union (mIoU) for corner segmentation reached 98.15% and 93.89%, respectively-improvements of 1.24% and 1.64% over the baseline SegFormer-B0-while model computational cost was reduced by approximately 23.2%. At a camera-to-corner distance of about 2 m, the aiming error of the alignment position was below 1.0 mm. Absolute translation errors on the X , Y , and Z axes were all below 5.0 mm, and absolute rotation error was below 0.5° . These results indicate that the proposed method achieves reliable accuracy and satisfies the requirements for automatic alignment of single-angle components.

Key words: monocular vision; containers; automatic alignment; target recognition; pose estimation

1 引 言

近年来,随着全球经济的快速增长,集装箱运输业得到了快速发展,集装箱运输通过标准化的装载和运输方式,大大提高了货物的运输效率和安全性,降低了物流成本,促进了国际贸易的繁荣^[1]。集装箱正面起重机采用集装箱吊具来装卸集装箱,其操作流程需依赖人工控制完成。在进行集装箱装卸作业中,操作人员通过驾驶室人机交互界面操控吊具三维运动,在吊具与集装箱形成空间对位后,需要将吊具旋锁机构与集装箱顶的角件孔位精准对接。然而,通常进行这项作业时,操作人员观察视角非常有限,往往是靠感觉和运气进行操作,导致锁闭机构对接失败率增加,且频繁的定位误差易引发吊具结构件与集装箱顶角件的机械碰撞,造成设备损耗以及集装箱滑移等安全隐患,因此,亟需提高集装箱装卸的自动化和智能化程度以增强装卸效率降低作业成本。

目前,激光雷达(Light Detection and Ranging, LiDAR)技术因其高精度、快速响应及良好的环境适应性,在集装箱自动化对位中的应用逐渐趋于成熟,尤其在港口自动化领域中。国内外研究主要集中在激光雷达的三维建模、定位精度的提高以及与其他传感器的融合应用等方面。李文锋等^[2]采用激光雷达扫描集装箱堆场环境,结合GPS\INS模块采集的数据和先验集装箱尺寸信息确定集装箱的空间位置。Zhang等^[3]提出一种基于三维点云数据处理的集装箱锁销自动卸货方法。虽然基于激光雷达的对位技术在定位精度和环境适应性方面优于基于导向板的辅助对位方法,但由于需要对集装箱多个表面进行点云数据采集与处理,算法复杂度较高,难以满足吊具与集装箱快速精准对位的实时性需求。

随着计算机视觉的不断发展,图像处理技术广泛应用于各行各业,如医学影像^[4]、自动驾驶^[5-6]、工业检测与质量控制^[7-8]等领域。基于图像处理的集装箱对位方案分为基于传统机器视觉的方法和基于深度学习的方法两大类。在基于传统机器视觉技术方面,日本三菱重工公司设计了一种集装箱检测设备。Li等^[9]使用双目立体相机测量集装箱的深度信息,利用霍夫直线变换检测出集装箱区域。在基于深度学习技术方面,唐鑫^[10]基于YOLOv4-tiny网络检测锁孔边框,然后利用EGNet网络分割角件,最后通过双目测距原理计算出角件中心的三维世界坐标,但该方法视场角受限,且算法复杂度增加,实时性提升较为困难。Zhang等^[11]基于改进的SSD模型,结合模糊图像形态匹配算法实现对集装箱角件的精确检测和跟踪。李源鑫等^[12]提出基于轻量化YOLOv5s的集装箱角件识别算法,提高集装箱角件的检测效率,该方法仅仅完成集装箱角件的检测任务,并无法完成角件的自动对位任务。

将图像处理技术结合深度学习应用到吊具与集装箱对位中,既能克服传统人工对位的缺点,降低工作强度,又能提高工作效率,保证工作质量。针对目前研究方法存在的问题,本文基于深度学习的图像方法,采用单目视觉手段研究吊具与集装箱的对位技术。首先,基于轻量级YOLOv8的角件检测网络,提取集装箱角件图像中的角件区域,实现快速定位的目的。其次,通过改进的SegFormer-B0实现集装箱角件的精细分割。然后,基于LM-PNP算法完成集装箱角件的位姿估计。最后,搭建实验环境下单目视觉自动对位平台,系统验证所提算法的有效性。该算法能够精确识别集装箱角件,完成集装箱角件的位姿估计,进而驱动旋锁插进集装箱角件内部。

2 算法原理

2.1 集装箱角件识别算法

吊具与集装箱自动对位需要使吊具旋锁插入到集装箱角件中,该过程的关键在于集装箱角件的识别和集装箱角件的位姿获取。

针对集装箱角件的识别部分,以集装箱角件图像为研究对象,对其进行特征分析,在整幅高分辨率的图像中,集装箱角件部分像素占比很低,直接在集装箱图像上分割集装箱角件

将浪费大量计算资源,降低对位实时性,因此本文采用粗略定位和精细分割相结合的方法完成集装箱角件的识别任务。首先,构建基于YOLOv8s的角件检测网络,提取集装箱角件图像中的角件区域,实现快速定位的目的。然后,基于改进的SegFormer-B0算法实现集装箱角件的精细分割。保证在复杂的集装箱装卸工况下和低算力硬件设备上完成精确识别任务。图1给出了集装箱角件识别的技术方案。

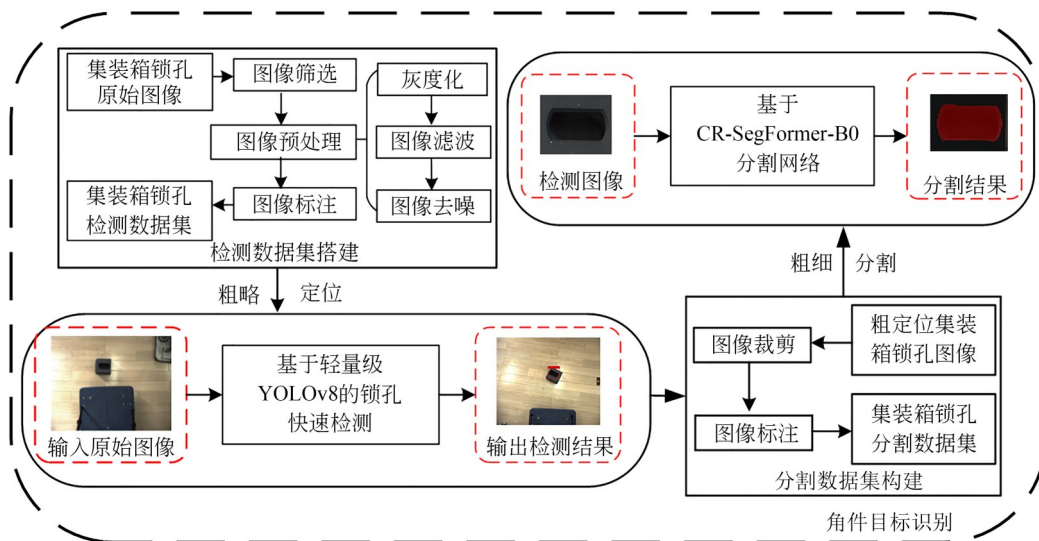


图1 集装箱角件识别算法流程

Fig. 1 Flow chart of container corner recognition algorithm

2.1.1 CR-SegFormer-B0网络搭建

为了进一步提高网络的特征表达能力和更高的分割精度,本文在SegFormer-B0的基础上进行改进,提出了CR-SegFormer-B0网络,在保持较低模型复杂度的同时,提高模型的分割效果。

SegFormer是2021年建立在Transformer结构基础上的新型语义分割模型,保留了原有CNN模型的编解码网络结构^[13]。在编码器部分,SegFormer网络设计了Mix Transformer Encoders (MiT)作为主干特征提取网络(Backbone),分别使用4个MiT提取输入图像的不同尺度特征,输出多层的Feature Map,包含高分辨率的粗特征和低分辨率的精细特征,4个MiT的输出分辨率分别为原始集装箱角件图像分辨率的1/4,1/8,1/16和1/32。根据Backbone的层数不同,分为MiT-

B0~MiT-B5, MiT-B0是轻量级的预测模型, MiT-B5是性能最好的也是最大的模型。

为了降低网络整体的计算复杂度, SegFormer对Transformer模块中的多头自注意力机制(Multi-head Self-Attention, MSA)进行改进,在原始自注意力机制基础上,添加缩放因子 R ,提出了高效的多头自主力机制(Efficient Multi-head Self Attention, EMSA),降低网络整体的计算复杂度。自注意力机制可以表示为:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V}, \quad (1)$$

其中: $\mathbf{Q}, \mathbf{K}$ 和 $\mathbf{V}$ 的维度都是 $N \times C, N = H \times W$ 。

原始自注意力机制的计算复杂度为 $O(N^2)$, 引入缩放因子的复杂度为 $O(N^2/R)$ 。在解码器

部分,引入轻量级的MLP解码器,相较于传统复杂的解码结构更加高效。其关键在于所采用的多级Transformer编码器比CNN具有更大的感受野,从不同层次提取高分辨率的粗特征与低分辨率的精细特征,并在解码阶段进行特征融合,从而在保持模型轻量化的同时显著提升了语义分割的精度和稳定性。

CR-SegFormer-B0网络在SegFormer-B0的基础上融合RFB模块和CBAM注意力机制。在网络的解码器部分,每一层MLP上采样之后添加RFB模块和CBAM注意力机制,改进后的网络结构如图2所示。CR-SegFormer-B0融合了RFB的多尺度特征提取模块,增强了感受野,在解码器部分,将高分率低层次的语义信息和低分辨率高层次的语义信息进行融合,完成上采样。为了更好地恢复语义信息,通过RFB模块增强感受野之后,加入CBAM注意力机制,增强网络表达能力。为了缓和融合之后带来的复杂度的提升,将解码器MLP每一层的输出通道数减半,将原来的256降至128。

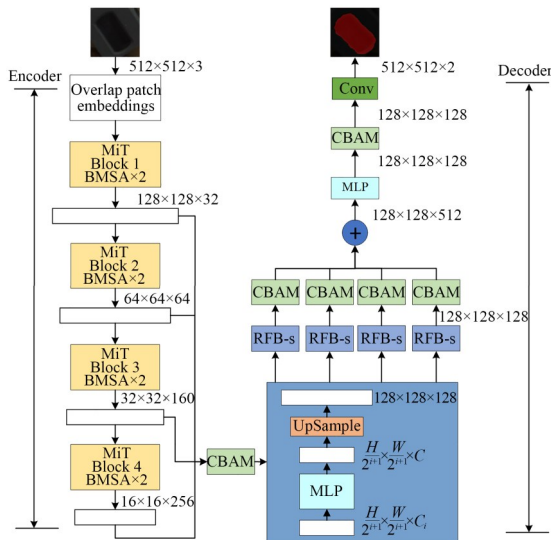


图2 CR-SegFormer-B0网络结构

Fig. 2 Structure of CR-SegFormer-B0 network

2.1.2 多尺度特征融合模块

RFB模块^[14]在Inception思想的基础上加入了空洞卷积,有效增大了感受野,提高特征提取能力。RFB-s在RFB的基础上进行改进,其结构如图3所示。

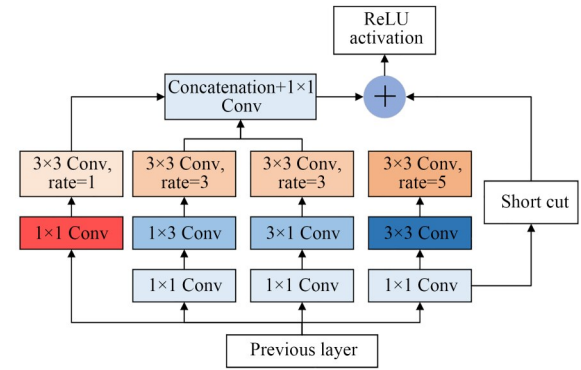


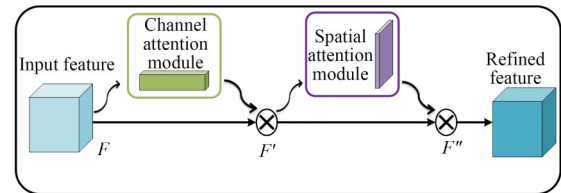
图3 RFB-s模块结构

Fig. 3 Structure of RFB-s module

RFB-s一方面使用 3×3 卷积层代替 5×5 卷积层,另一方面使用 1×3 和 3×1 卷积层代替 3×3 卷积层,降低了模型计算量。其结构主要有两个特点:(1)不同尺寸卷积核的卷积层构成了多分支结构;(2)通过三通道并行式结构,采用不同空洞系数的空洞卷积进行特征提取,增加感受野。

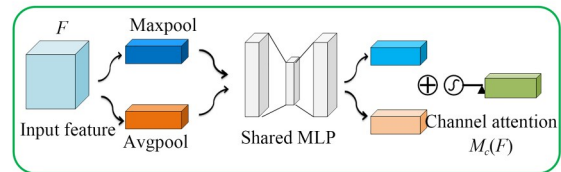
2.1.3 轻量级注意力机制

CBAM注意力机制^[15]的结构如图4所示。



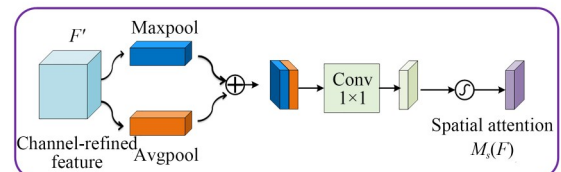
(a) CBAM结构

(a) CBAM structure



(b) 通道注意力

(b) Channel attention



(c) 空间注意力

(c) Spatial attention

图4 CBAM及两种注意力模块结构

Fig. 4 Structure of CBAM and two attention modules

该注意力机制属于轻量级模块,对模型的计算量影响较小。

CBAM模块结合了通道注意力与空间注意力。输入特征首先经过通道注意力模块,利用全局平均池化与最大池化生成通道注意力图,突出重要通道特征。然后,输出特征再输入空间注意力模块,通过对通道维度的池化操作生成空间注意力图,进一步强调关键空间区域,从而实现特征的精细增强。最终的输出特征是输入特征与两个注意力图的逐元素相乘结果,通过上述操作提高网络特征表达能力。该模块可以轻松地插入网络中,并不会对网络的性能产生负面影响。

2.2 集装箱角件位姿估计算法

通过深度学习的手段来识别和定位 2D 目标物体,但实际情况下,目标物体都有其自身的三维几何和姿态信息,只获得目标物体的二维坐标位置,缺乏其在三维空间中的位置和姿态,难以进行对位操作。针对 2D 目标检测在位姿估计上的局限性,本文基于 2D-3D 对应关系,设计集装箱角件的位姿估计算法。具体算法流程如图 5 所示。

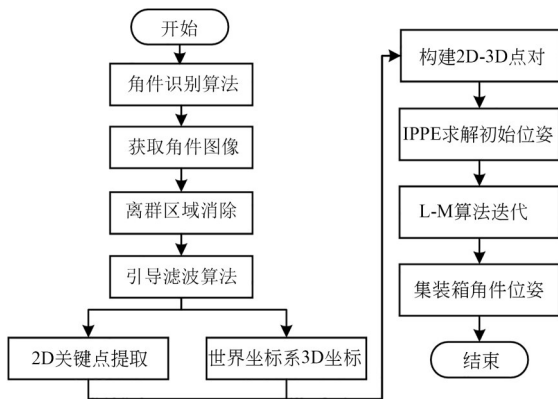


图 5 角件位姿估计算法流程

Fig. 5 Flow chart of corner pose estimation algorithm

2.2.1 LM-PnP 算法

透视 n 点问题 (Perspective-n-Points, PnP), 是指 3D 到 2D 点对的物体运动定位问题, 已知物体在世界坐标系中的三维坐标和在相机图像平面上对应的像素坐标, 求解物体相对于相机坐标系的六自由度位姿 (6D 位姿), 包括三维空间中的位置 (X, Y, Z) 和朝向 (绕 X, Y, Z 轴的旋转角)。该问题是计算机视觉中的核心任务, 广泛应用于目标识别、姿态估计和机器人操作等

领域。

迭代式算法 (Levenberg-Marquardt, L-M) 利用已知的 2D 特征点坐标与对应的 3D 特征点坐标, 结合相机的内参, 通过最小化 2D 特征点的重投影误差的平方和迭代求解。该方法收敛性较好, 对于特征点较少情况下, 求解效果较好, 但受初始估计值的影响较大^[16]。为了防止在迭代求解过程中, 在局部最小值附近收敛, 本文结合 IPPE 和 L-M 算法对角件进行位姿估计, 使用直接法 IPPE 代替弱透视投影模型进行初始位姿值的求解, 然后将初值解代入 L-M 算法, 以重投影误差的平方和作为目标函数, 迭代求取使得重投影误差平方和最小的解, 从而获取集装箱角件的位姿信息。

设世界坐标系下 n 个空间特征点的坐标为 $P_{wi} = (X_{wi}, Y_{wi}, Z_{wi}) (i = 1, 2, 3, \dots, n)$, 相机坐标系下的坐标为 (X_{ci}, Y_{ci}, Z_{ci}) 在相机图像平面对应的投影点坐标为 $p_i = (u_i, v_i) (i = 1, 2, 3, \dots, n)$, 由相机成像原理得:

$$Z_{ci} \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{f}{dx} & 0 & u_0 & 0 \\ 0 & \frac{f}{dy} & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & T \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} X_{wi} \\ Y_{wi} \\ Z_{wi} \\ 1 \end{bmatrix} = M_1 M_2 \begin{bmatrix} X_{wi} \\ Y_{wi} \\ Z_{wi} \\ 1 \end{bmatrix}, \quad (2)$$

其中: f 为相机焦距; dx, dy 为单位像素在物理坐标系下的实际物理尺寸; (u_0, v_0) 为相机光学中心; M_1, M_2 分别为相机的内参矩阵和外参矩阵。

对于旋转矩阵 R , 用四元数 $q = (q_w, q_x, q_y, q_z)$ 形式表示为:

$$R = \begin{bmatrix} 1 - 2(q_y^2 + q_z^2) & 2(q_x q_y - q_z q_w) & 2(q_x q_z + q_y q_w) \\ 2(q_x q_y + q_z q_w) & 1 - 2(q_x^2 + q_z^2) & 2(q_y q_z - q_x q_w) \\ 2(q_x q_z - q_y q_w) & 2(q_y q_z + q_x q_w) & 1 - 2(q_x^2 + q_y^2) \end{bmatrix}, \quad (3)$$

其中: q_x, q_y 和 q_z 表示向量部分, q_w 表示标量部分, 且满足 $q_w^2 + q_x^2 + q_y^2 + q_z^2 = 1$ 。

令 $q' = [q_w, q_x, q_y, q_z, t_x, t_y, t_z]^T$, 则空间特征点在像素坐标系下的投影坐标可以表示为关于

变量 q' 的函数:

$$\hat{p}_i = T(q'), \quad (4)$$

其中: $\hat{p}_i = (\hat{u}_i, \hat{v}_i) (i = 1, 2, 3 \cdots n)$ 表示投影点坐标。

特征点的重投影误差可以表示为:

$$e_i(q') = \|p_i - \hat{p}_i\|. \quad (5)$$

则所有空间特征点的重投影误差平方和可以表示为:

$$E_i(q') = \sum_{i=1}^n \|p_i - \hat{p}_i\|^2. \quad (6)$$

将式(6)构建为具有约束条件下的非线性最小二乘法的目标函数,通过 L-M 迭代求解,使得 $E_i(q')$ 值最小的解 q' , 即 $q_w, q_x, q_z, t_x, t_y, t_z$, 获得相机外参, 即集装箱角件在相机坐标系下的位姿信息。

2.2.2 角件位姿估计算法

由于在集装箱装卸过程中,集装箱正面起重机的吊具位于集装箱的上方,相机拍摄装卸场景时,采集的图像为集装箱上表面图像,而集装箱角件属于标准件,其性能和规格在国际上有严格要求。图6给出了普通集装箱角件的尺寸规格说明,该图仅为尺寸展示图,并非真实工况图。以集装箱角件平面为 XOY 平面,垂直于角件平面为 Z 轴,建立角件坐标系,其中外接矩形的四个角点以及中心点的世界坐标可以表示为: $(0, 63.5, 0), (0, 0, 0), (124.0, 0, 0), (124.0, 63.5, 0)$ 和 $(62.0, 31.75, 0)$, 进而将求取集装箱角件与相机的相对位姿转化为共面参考点的 PnP 问题。

针对集装箱角件的位姿估计,本文基于 2D-3D 对应的迭代式位姿估计算法进行集装箱角件位姿估计。以单目视觉为基础,采集的 RGB 图

像作为输入,首先进行角件识别,通过粗略定位以确定角件的大致位置,然后通过精细分割进一步提取角件区域。在此基础上,使用膨胀和腐蚀去除小的毛刺,应用图像处理技术对角件图像进行离群区域消除和引导滤波算法处理,然后获取角件的最小外接矩形。该过程能够有效地简化角件的形状,并为后续的特征点提取奠定基础。

在获取到最小外接矩形后,提取角件的关键特征点,以最小外接矩形的四个角点和矩形的中心点作为关键特征点,通过准确的 2D-3D 特征点匹配,利用非直接法 IPPE 求出初始位姿解,代入 L-M 算法中,再精确计算出集装箱角件在相机坐标系下的位姿。该位姿信息是实现吊具自动对位的关键,直接决定了吊具旋锁能否准确插入角件,从而完成集装箱的装卸操作。由于目标检测进行图像截取,RGB 图像当中目标的尺寸大小和坐标图上的尺寸大小不一样,为了能够建立原始 RGB 图像像素与空间坐标的 2D-3D 对应关系,把像素从坐标图映射到了原始的 RGB 图像上面。

3 实验结果与分析

为了验证所提算法的有效性,开展实验验证和结果分析。通过搭建集装箱角件数据集,包括目标检测数据集和语义分割数据集,适应深度学习网络的训练。通过对比实验和消融研究,验证各个模块的有效性,并搭建实验平台,系统验证算法的可靠性。

3.1 数据集构建

针对集装箱角件的粗略定位任务和目标检测任务,搭建了共 590 张实验室环境下的数据集和 550 张真实场景下的数据集。通过针对集装箱角件的精细分割任务,搭建了共 176 张实验室环境下的数据集和 250 张真实场景下的数据集。为了提高深度神经网络的性能和泛化能力,对于集装箱角件的精细分割任务,采用图像增强的方式对数据集进行扩充,通过尺度变换、水平镜像、图像旋转和添加噪声的方式增强数据集。将原有集装箱角件图像分别扩充至 1 410 张和 1 500 张。对扩充后的数据集按照 4:1 的比例划分为训练集和测试集,划分后的实验室环境下的数据集包含 1 128 张训练图像和 282 张测试图

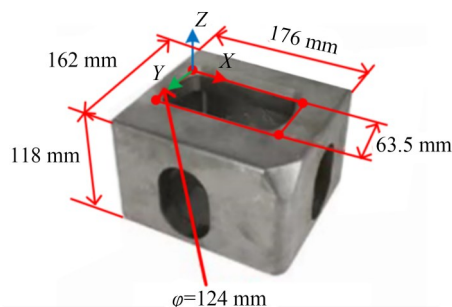


图6 集装箱角件规范

Fig. 6 Container corner specification

像,真实场景下的数据集包含 1 200 张训练图像和 300 张测试图像。其中,验证集和测试集共用一类图像。

3.2 实验配置及评价指标

为了完成集装箱角件的识别任务,对深度学习实验平台进行合理化配置。实验平台选择 Dell Precision 5820 工作站,配置 Intel Core i9-10900X 处理器,同时搭载 NVIDIA GeForce RTX 3080Ti 显卡,内存 32 GB ECC,满足训练要求。网络采用迁移学习策略进行训练,即训练过程从主干网络的预训练权重开始训练。选择在 PASCAL VOC 数据集上训练好的模型作为网络预训练模型,在不改变原有模型复杂度的基础上,进一步提高模型精度。

针对粗略定位任务,本文使用准确率(Precision)、召回率(Recall)、平均精度均值(mAP)、浮点运算次数(FLOPs)和参数量(Params)5 个指标进行评估。其中 Precision,Recall 和 mAP 作为评价模型检测精度的指标,FLOPs 和 Params 作为评价模型复杂度的指标。针对精细分割任务,采用 FLOPs 和 Params 作为评价模型复杂度的指标,采用均交并比(mIoU)作为评价模型分割精度的指标。

3.3 角件粗略定位实验及结果分析

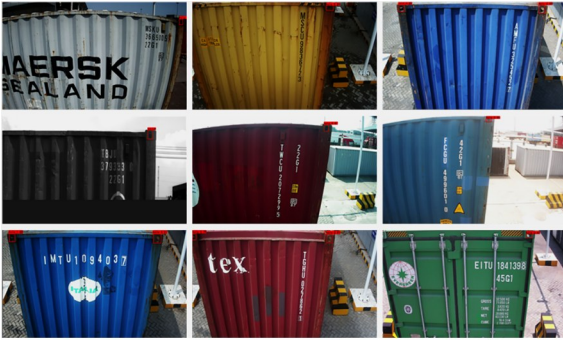
在目标检测数据集上训练 YOLOv8s 网络,从优化模型性能和节省存储空间的角度出发,将模型的数据类型从 32 位浮点数转换为 16 位浮点数,采用混合精度的方式训练。由于角件检测任务属于定位任务,故选用 BCE-Loss(交叉熵损失函数)损失函数,在反向传播中采用 CIoU-Loss(边界框回归损失函数)、BCE-Loss 和 DFL-Loss(离散焦点损失函数)三者的加权和,分别设置为 7.5,0.5 和 1.5。

获得相应训练权重后,将测试集图像输入到模型之中,统计模型推理结果并计算模型的精确度和速度。可视化结果如图 7 所示,模型检测的量化结果如表 1 所示。由于本文中检测目标只有集装箱角件一个类别,故 mAP 值和 AP 值相同。

由表 1 可知,在 75% 的 IoU 阈值下,轻量级 YOLOv8s 模型在实验室环境下检测的平均精度达到 96.40%,在真实场景下检测的平均精度达



(a) 实验环境
(a) Experimental environment



(b) 真实场景
(b) Real scene

图 7 角件定位可视化结果
Fig. 7 Visualization results of corner positioning

表 1 网络模型评估结果					
Tab. 1 Network model evaluation results					
指标	Precision	Recall	mAP	Params	FLOPs
实验环境下	100%	97.74%	96.40%	11.2	28.6
真实场景下	100%	97.12%	95.61%	11.2	28.6

到 95.61%。因此,本文选用 YOLOv8s 模型,其采用的轻量级模型架构和多尺度训练技术,能够实现复杂背景、低算力场景下的实时检测。

3.4 角件精细分割实验及结果分析

3.4.1 对比实验

针对本课题的集装箱角件的精细分割任务,选取 U-Net, PSPNet, DeepLabv3+, SegFormer-B0, SegFormer-B4 以及 SenFormer(Swin_T) 模型进行对比实验。量化结果如表 2 所示,可视化

结果如图 8 所示。其中,D-V3、S-B0 和 S-B4 分别表示 DeepLabv3+,SegFormer-B0 和 SegFormer-B4 网络。

表 2 各网络模型的评估结果

Tab. 2 Evaluation results of each network model

Model	mIoU	Params	FLOPs
Unet(VGG)	92.52/86.95	24.89	451.67
PSPnet(Resnet50)	86.41/81.48	46.71	118.43
DeepLabv3+(MobileNetv2)	93.94/87.48	5.81	52.64
SenFormer(Swin_T)	97.45/94.96	59.02	179.02
SegFormer-B0	96.91/92.25	3.72	13.70
SegFormer-B4	97.83/95.85	171.33	64.01

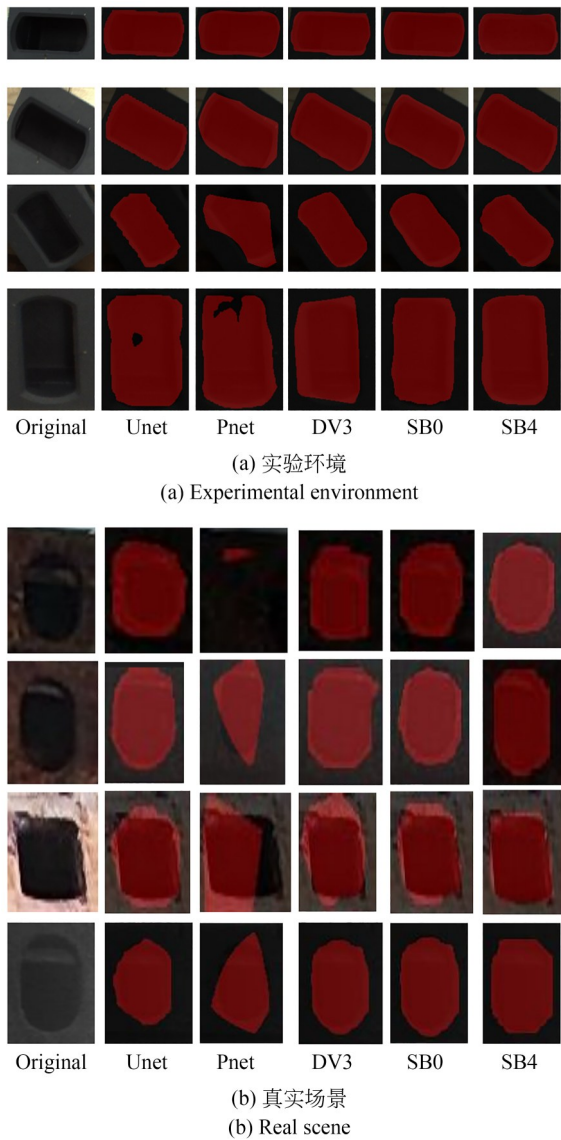


图 8 角件分割可视化结果

Fig. 8 Visualization results of corner segmentation

为了满足既要保证模型具有较高的分割精度又要保证较低的模型复杂度这一条件,通过对表 2 的分割结果分析,SegFormer-B0 模型优势明显,模型复杂度最低,且分割精度较高。

3.4.2 消融实验

为了验证改进网络 CR-SegFormer-B0 的有效性,现分别采取融合 CBAM 模块方式、融合 RFB 模块方式、融合两者的方式(不改变通道数)以及融合两者的方式(改变通道数)进行消融实验。表 3 给出不同改进方式在实验环境下测试集的评价指标数据。

表 3 消融实验评估结果

Tab. 3 Ablation experiment evaluation results

Model	mIOU	Param	FLOPs
SegFormer-B0	96.91	3.72	13.70
SegFormer-B0+CBAM	97.32	3.77	13.79
SegFormer-B0+RFB	97.79	5.01	27.78
SegFormer-B0+CBAM+RFB	98.18	5.05	27.80
CR-SegFormer-B0	98.15	3.79	10.52

由表 3 可知,无论是单独融合 RFB 特征提取模块,还是单独融合 CBAM 注意力机制模块,都可以对 SegFormer-B0 模型实现精度的提升,但提升都不大。第三种方式,两者结合使用可以很好地帮助 SegFormer-B0 模型实现更高分割精度的提升,与此同时,也造成模型复杂度的提高。第四种方式的改进,有效缓和了模型复杂度的提高,模型参数量相比原模型仅增加了 0.09 M,提高了约 2.4%;但计算量减少了 3.18 GFLOPs,下降了约 23.2%,分割精度相比于原模型提高了 1.24%。因此,本文选择使用融合 RFB 和 CBAM 的 SegFormer-B0 模型,并将解码器的每层输出通道数降低为原来的一半,最后构成了 CR-SegFormer-B0 网络,作为集装箱角件图像的分割模型。

3.5 角件位姿估计实验及结果分析

为了验证 L-M 算法的有效性,本文分别采用 E-PnP,IPPE,SQPnP 和 L-M 算法对标定板的位姿进行估计,对比重投影误差(Reprojection error, RE)和对应的实际尺寸误差(Actual Size Error, AE)。通过计算像素当量(Per-pixel, P),获

取实际尺寸误差,计算方法如下:

$$\begin{cases} AE = RE \times Perp \\ P = \frac{S}{H} \times \frac{L}{f} \end{cases}, \quad (7)$$

其中: S 为传感器尺寸, H 为分辨率大小, L 为相机距离目标物体的实际距离, f 为相机焦距。

表 4 给出了不同距离和角度上的集装箱角件

位姿估计实验结果。基于 L-M 算法的重投影误差与实际误差相对其它几种方法最小,在相机距离集装箱角件 2 m 左右的情况下,求解的重投影误差约为 0.6 个像素,经过式(7)转换,对位位置瞄准误差小于 1 mm。结果证明,该算法能够在少量特征点情况下获得高精度的位姿估计结果。

表 4 角件位姿估计算法结果对比

Tab. 4 Comparison of corner pose estimation algorithm results

距离/mm	RE/pix				AE/mm			
	E-PNP	IPPE	SQPnP	L-M	E-PNP	IPPE	SQPnP	L-M
855	10.113	1.530	1.515	1.513	3.744	0.566	0.561	0.560
875	6.598	1.406	1.378	1.372	2.450	0.533	0.522	0.520
1 120	5.121	1.338	1.304	1.295	2.483	0.649	0.632	0.628
1 175	3.325	1.281	1.242	1.235	1.692	0.652	0.632	0.628
2 075	2.088	0.847	0.648	0.647	1.876	0.761	0.582	0.581

3.6 自动对位实验验证

为了进一步系统验证本文设计方案的正确性,在实验室条件下搭建单目视觉对位实验平台,开展自动对位实验,如图 9 所示。选用六自由度 AUBO-i10 机械臂作为执行机构,其末端通过定制转接件搭载海康 MV-CS050-90GC 工业相机(500 万像素 CMOS 传感器,千兆网口面阵)及真实旋锁。为保证角件在视场内,选用 8 mm 焦距的 MVL-MF0828M-8MP 镜头。

实验条件为光照,近似户外日照光,工况场

景为相机位于角件上方,相机初始视场包含角件,且对准过程中角件不离开相机视场。根据实际应用需求及我国相关部门发布的角件技术要求,本文开展单角件自动对位的位置误差应小于 5 mm,姿态误差应小于 0.5°。

在进行自动对位实验前,需标定系统。对选型的相机进行标定,使用张正友标定法标定获得的相机焦距为 8.276 mm。由于旋锁是垂直于转接件表面安装在机械臂末端,且转接件平面和末端平面处于平行,故旋锁的姿态保持和末端一致,因此只需要进行位置标定即可。通过四点法对旋锁末端的位置标定结果为(2.31, -60.43, 83.92),单位为 mm。在机器人系统中,通过手眼标定获得相机坐标系在机器人系统基坐标系下的位姿。采用 TSAI 方法对标定板进行多次标定,获得的标定结果为(-1.48, 170.53, 24.58, -0.90, 0.24, -178.60),前 3 个为位置结果,单位为 mm;后 3 个为姿态结果,单位为°。

在完成实验条件下的硬件平台搭建、相机标定、末端执行器标定和手眼标定实验之后,结合软件开发平台开展集装箱角件的自动对位实验并进行误差分析。设计的自动对位的实验流程如图 10 所示,识别集装箱角件的 6D 位姿,完成整

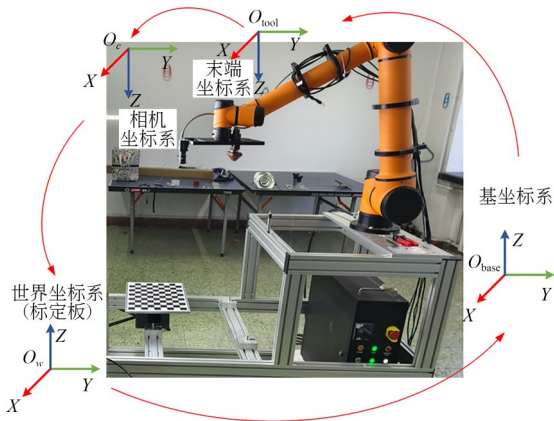


图 9 单目视觉对位实验平台

Fig. 9 Monocular visual alignment experimental platform

体对位实验验证。多次对位操作的结果如表 5 所示。表 6 给出了集装箱角件在系统基坐标系的真实位姿(示教器读取)。表 7 给出了位姿误差信息。

通过开展批量对位实验,得到自动对位的误差信息,对集装箱角件的 6D 位姿进行精度分析,位置误差小于 5 mm,姿态误差小于 0.5°,实验所得误差相对较小。

为了对设计算法的可靠性进行验证,在不同光照环境和不同工况下开展 10 组实验,每组进行 20 次对位实验,对位实验结果如表 8 所示。在进行的 10 组实验均识别出角件的 6D 位姿,对位成功率高达 98.5%,验证了本文所设计算法的可靠性。其中,对位不成功的原因主要有 3 点:(1)机器人系统本身的误差,对单目对位系统的精度产生影响;(2)相机标定及手眼标定误差。由于标定是通过计算得到的结果,所以在计算中产生的

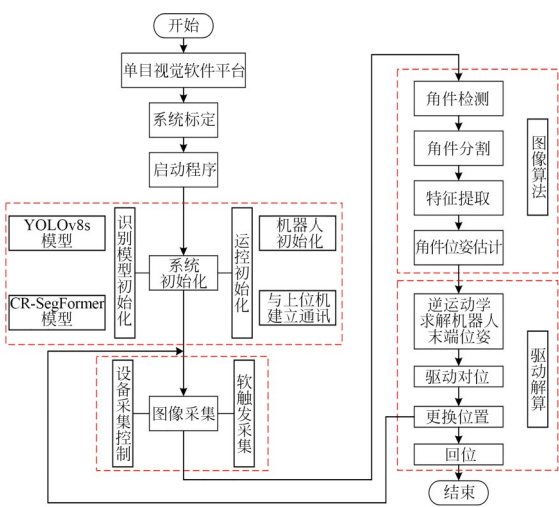


图 10 自动对位整体流程

Fig. 10 Overall process of automatic alignment

误差也会影响整个系统的精度;(3)图像算法的误差。由于图像算法会受到光照的影响,当光照条件

表 5 单目视觉系统角件计算位姿

Tab. 5 Corner calculation pose of monocular vision system

实验组别	X/m	Y/m	Z/m	$R_x/(^{\circ})$	$R_y/(^{\circ})$	$R_z/(^{\circ})$
1	-705.18	86.42	-212.67	179.45	0.43	79.12
2	-669.12	134.88	-215.13	179.57	-0.37	82.40
3	-735.90	190.48	-104.81	179.63	-0.33	90.19
4	-747.12	164.60	-107.22	179.65	0.31	91.22
5	-646.52	75.10	-215.71	179.50	0.45	101.21
6	-728.03	-74.49	-212.12	179.48	0.39	87.52
7	-721.78	79.81	-279.40	176.93	2.65	60.04
8	-694.02	-39.01	-283.07	176.87	2.59	78.59
9	-619.83	-239.27	-311.02	174.56	4.54	83.75
10	-624.81	-251.57	-313.93	173.79	4.61	96.80

表 6 角件的真实位姿

Tab. 6 Real pose of corner piece

实验组别	X/m	Y/m	Z/m	$R_x/(^{\circ})$	$R_y/(^{\circ})$	$R_z/(^{\circ})$
1	-702.12	84.88	-216.51	180.00	0.00	79.44
2	-666.90	131.12	-219.48	180.00	0.00	81.99
3	-731.53	193.22	-107.87	180.00	0.00	90.45
4	-750.19	167.32	-105.01	180.00	0.00	90.83
5	-642.83	72.02	-217.86	180.00	0.00	100.94
6	-724.22	-70.87	-218.01	180.00	0.00	87.36
7	-725.60	76.22	-277.02	177.32	2.97	60.30
8	-690.11	-36.52	-279.48	177.32	2.97	78.28
9	-615.41	-236.68	-314.92	174.15	4.89	83.40
10	-622.13	-256.02	-309.18	174.15	4.89	96.34

较差,目标识别的结果就会产生一定的误差,而该误差也会影响系统对位精度。

表 7 位姿误差信息表
Tab. 7 Pose error information

误 差	X/m	Y/m	Z/m	$R_x/(^{\circ})$	$R_y/(^{\circ})$	$R_z/(^{\circ})$
平均绝对误差	3.21	3.17	3.23	0.43	0.36	0.32
均方根误差	3.98	4.02	4.05	0.44	0.37	0.33

表 8 自动对位实验结果
Tab. 8 Experimental results of automatic alignment

实验组别	实验次数	对位成功次数	对位成功率/%
1	20	20	100
2	20	20	100
3	20	19	95
4	20	20	100
5	20	20	100
6	20	19	95
7	20	20	100
8	20	20	100
9	20	20	100
10	20	19	95

为了保证单角件对位成功率,图像算法采用多帧计算动态修正的策略。在旋锁运动至角件的过程中,距离角件不同高度拍摄集装箱角件图像计算角件位姿,舍弃差异较大的结果,对结果进行修正,进而保证驱动解算结果的正确性,提

高算法稳定性。

4 结 论

本文基于单目视觉,提出结合深度学习和图像处理技术的集装箱自动对位算法。该算法能够精确识别集装箱角件,并完成位姿估计。基于 AUBO-i10 机械人平台开展算法验证实验,结果表明,本文所提算法满足单角件对位的精度需求。在实际的集装箱对位中,为了提升对位精度,保证对位的成功率,采用对角线安装单目相机的方式实现对位,通过采集集装箱对角线上的角件图像,进行图像处理,协同计算集装箱与吊具的相对位姿关系,实现了集装箱与吊具的自动对位。

作者贡献声明:

张永雷:论文撰写,算法设计与实验分析;
胡涛:论文审核与编辑,提供资源;
申立群:项目管理与论文审核。

参考文献:

[1] 程国政. 我国港口集装箱运输发展现状与趋势[J]. 中国航务周刊, 2024(30): 45-47.
CHENG G ZH. Present situation and trend of container transportation in China's ports[J]. *China Shipping Gazette*, 2024(30): 45-47. (in Chinese)
[2] 南通润邦重机有限公司,无锡卡尔曼导航技术有限公司. 一种结合 GPS\INS 的集装箱识别定位方法: 201910905943.1[P]. 2023-02-21.
Nantong Runbang Heavy Machinery Co., Ltd., Wuxi Kalman Navigation Technology Co., Ltd. A container identification and positioning method combined with GPS\INS: 201910905943.1 [P]. 2023-

02-21. (in Chinese)
[3] ZHANG L G, ZHANG W, LI Y C, *et al.* Acquisition method of container lock pin model based on point cloud data[C]. 2022 *China Automation Congress (CAC)*. November 25-27, 2022, Xiamen, China. IEEE, 2022: 490-494.
[4] 周涛,张祥祥,陆惠玲,等. 基于 LL-GG-LG Net 的 CT 和 PET 医学图像融合[J]. 光学精密工程, 2023, 31(20): 3050-3064.
ZHOU T, ZHANG X X, LU H L, *et al.* CT and PET medical image fusion based on LL-GG-LG Net [J]. *Opt. Precision Eng.*, 2023, 31(20): 3050-3064. (in Chinese)
[5] WANG L, ZHANG X Y, SONG Z Y, *et al.*

- Multi-modal 3D object detection in autonomous driving: a survey and taxonomy[J]. *IEEE Transactions on Intelligent Vehicles*, 2023, 8(7): 3781-3798.
- [6] 蒋志豪, 张美香, 薛卫涛, 等. 面向点云分类和分割的形状自适应特征聚合网络[J]. *光学精密工程*, 2025, 33(5): 777-788.
- JIANG ZH H, ZHANG M X, XUE W T, *et al.* Shape adaptive feature aggregation network for point cloud classification and segmentation[J]. *Opt. Precision Eng.*, 2025, 33(5): 777-788. (in Chinese)
- [7] QI D L, HAN Y F, ZHOU Z Q, *et al.* Review of defect detection technology of power equipment based on video images[J]. *Journal of Electronics & Information Technology*, 2022, 44 (11): 3709-3720.
- [8] 刘玉淇, 吴一全. 基于机器视觉的太阳能电池片缺陷检测算法综述[J]. *光学精密工程*, 2024, 32 (6): 868-900.
- LIU Y Q, WU Y Q. Review of defect detection algorithms for solar cells based on machine vision[J]. *Opt. Precision Eng.*, 2024, 32 (6): 868-900. (in Chinese)
- [9] WEI L, LEE E. Real-time container shape and range recognition for implementation of container auto-landing system[J]. *Journal of Korea Multimedia Society*, 2009, 12: 794-803.
- [10] 唐鑫. 基于深度学习的吊具与集装箱视觉对位技术研究[D]. 成都: 西南交通大学, 2022.
- TANG X. *Research on Visual Alignment Technology of Spreader and Container Based on Deep Learning*[D]. Chengdu: Southwest Jiaotong University, 2022. (in Chinese)
- [11] ZHANG Y J, MI CH. *A Fast Vision-based Algorithm for Automated Container Pose Measurement System*[M]. The 8th International Conference on Advances in Construction Machinery and Vehicle Engineering. Singapore: Springer Nature Singapore, 2024: 817-825.
- [12] 李源鑫, 郭忠峰, 杨钧麟. 基于轻量化YOLOv5s的集装箱锁孔识别算法[J]. *计算机科学*, 2024, 51(S1): 524-529.
- LI Y X, GUO ZH F, YANG J L. Container lock hole recognition algorithm based on lightweight YOLOv5s[J]. *Computer Science*, 2024, 51(S1): 524-529. (in Chinese)
- [13] XIE E Z, WANG W H, YU ZH D, *et al.*, SegFormer: simple and efficient design for semantic segmentation with transformers [C]. 2021: 35th Annual Conference on Neural Information Processing Systems (NeurIPS). December 6-14, 2021, Canada.
- [14] LIU S T, HUANG D, WANG Y H. *Receptive Field Block Net for Accurate and Fast Object Detection* [M]. Computer Vision-ECCV 2018. Cham: Springer International Publishing, 2018: 404-419.
- [15] ZHANG Y Y, ZHANG M H, ZANG Y X, *et al.* Efficient video deblurring guided by Motion magnitude and Convolutional block attention module [C]. *Artificial Intelligence in China*. Singapore: Springer, 2024: 259-267.
- [16] 陈津平, 吴晓亮. 参考点共面条件下的稳健相机位姿估计方法[J]. *激光与光电子学进展*, 2024, 61(4): 147-155.
- CHEN J P, WU X L. Robust camera pose estimation method under coplanar reference points condition[J]. *Laser & Optoelectronics Progress*, 2024, 61(4): 147-155. (in Chinese)

作者简介:



张永雷(2001—),男,山东菏泽人,硕士研究生,2025年于哈尔滨工业大学获得硕士学位,主要从事视觉检测 and 智能图像处理方面的研究。E-mail: 23S101104@stu.hit.edu.cn

通讯作者:



胡涛(1976—),男,河南许昌人,副教授,硕士生导师,2001年、2006年于哈尔滨工业大学分别获得硕士和博士学位,主要从事图像处理和图像测量方面的算法研究。E-mail: hutao@hit.edu.cn