

DOI:10.16356/j.1005-2615.2022.01.003

## 面向飞行机械臂的实时目标检测与定位算法

张睿<sup>1</sup>, 王尧尧<sup>1,2</sup>, 段雅琦<sup>3</sup>, 陈柏<sup>1</sup>

(1. 南京航空航天大学机电学院, 南京 210016; 2. 浙江大学流体动力与机电系统国家重点实验室, 杭州 310027;  
3. 中国兵器工业集团航空弹药研究院有限公司, 哈尔滨 150046)

**摘要:** 飞行机械臂要完成自主抓取的任务, 对目标物的识别与定位尤为关键。当前飞行机械臂视觉识别算法多采用传统的特征提取等方法。为提升目标物识别及定位的精度和效率, 本文设计了一种基于 YOLOv5 深度学习目标检测算法和 RGB-D 传感器的视觉识别与定位算法, 该算法可以实时检测目标物并对其位姿进行估计, 为飞行机械臂的抓取工作服务。同时针对深度学习算法计算量庞大, 在嵌入式端无法实现高性能实时检测的问题, 引入了模型量化技术优化算法, 大幅提升算法推理速度。本文介绍了算法的整体框架及实现过程, 利用 COCO 数据集和动作捕捉系统分别验证了目标检测和位姿估计部分算法的有效性。

**关键词:** 飞行机械臂; YOLOv5; RGB-D; 目标检测; 模型量化; 位姿估计

**中图分类号:** TP391      **文献标志码:** A      **文章编号:** 1005-2615(2022)01-0027-07

### Real-Time Object Detection and Location Algorithm for Aerial Manipulator

ZHANG Rui<sup>1</sup>, WANG Yaoyao<sup>1,2</sup>, DUAN Yaqi<sup>3</sup>, CHEN Bai<sup>1</sup>

(1. College of Mechanical & Electrical Engineering, Nanjing University of Aeronautics & Astronautics, Nanjing 210016, China; 2. The State Key Lab of Fluid Power Transmission and Control, Zhejiang University, Hangzhou 310027, China;  
3. Norinco Group Air Ammunition Research Institute Co., Ltd., Harbin 150046, China)

**Abstract:** To accomplish the task of aerial manipulator's autonomous grasping, it is very important to recognize and locate the object. At present, most of the recognition algorithms of aerial manipulators use traditional feature extraction methods. In order to improve the accuracy and efficiency of object recognition and positioning, this paper designs a visual recognition and positioning algorithm based on YOLOv5 deep learning object detection algorithm and RGB-D sensor. The algorithm can detect the object in real-time and estimate its pose, which serves for the grasping task of aerial manipulator. At the same time, aiming at the problems that the deep learning algorithm has a huge amount of calculation and cannot achieve high-performance real-time detection in the embedded system, model quantization is introduced to optimize the algorithm, which can greatly improve the processing speed of the algorithm. This paper introduces the overall framework and implementation process of the algorithm and verifies the effectiveness of some algorithms of object detection and pose estimation by using the COCO dataset and motion capture system.

**Key words:** aerial manipulator; YOLOv5; RGB-D; object detection; model quantification; pose estimation

近年来, 无人机在各领域的应用得到了迅猛的发展。然而现有的无人机大多只能完成监视、探测等任务, 而不能与目标进行交互, 无法满足当前对

无人机空中作业的迫切需求。为了让无人机代替人在许多难以接近的地点和危险的环境中工作, 研究人员开始对搭载多关节机械臂的旋翼飞行机械

**基金项目:** 国家自然科学基金 (52175097); 南京航空航天大学基本科研业务费 (NS2021033)。

**收稿日期:** 2021-05-31; **修订日期:** 2021-11-16

**通信作者:** 王尧尧, 男, 博士, 副教授, E-mail: yywang\_cmeec@nuaa.edu.cn。

**引用格式:** 张睿, 王尧尧, 段雅琦, 等. 面向飞行机械臂的实时目标检测与定位算法[J]. 南京航空航天大学学报, 2022, 54(1): 27-33. ZHANG Rui, WANG Yaoyao, DUAN Yaqi, et al. Real-time object detection and location algorithm for aerial manipulator[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2022, 54(1): 27-33.

臂展开相关研究<sup>[1]</sup>。

飞行机械臂在大多数环境下执行任务时,需要具有一定的独立获取和处理信息的能力,其中非常重要的环节是目标物体的识别和定位,因此其视觉系统的构建是一个关键问题。针对飞行机械臂的视觉伺服控制,国内外研究者已取得了许多重要成果<sup>[2-4]</sup>。然而,受无人机机载系统计算性能的限制,当前视觉识别算法主要采用传统的基于特征提取的方法,如颜色、形状等<sup>[5-6]</sup>。近年来随着人工智能研究的不断深入,在目标检测领域越来越多地开始采用基于深度学习的方法。与传统的基于特征提取的检测方法相比,基于深度学习的目标检测算法具有较高的识别精度,在各种场景中得到了广泛的应用,主要有将目标检测分步进行的双阶段方法,包括RCNN<sup>[7]</sup>、F-RCNN<sup>[8]</sup>等;单阶段算法是将目标检测视为一个回归问题,可以直接输出目标物的边界框和类别,包括SSD(Single shot multibox detector)<sup>[9]</sup>、YOLO(You only look once)<sup>[10]</sup>等。然而基于深度学习的目标检测算法由于计算复杂、计算参数多,往往需要强大的GPU来保证其实时检测性能。因此,如何将深度学习算法应用于飞行机械臂等嵌入式系统中,实现目标的实时检测,仍然是一个难点。为使深度学习算法更好地适配计算性能较为有限的系统,研究者们也提出了许多模型压缩和加速方法,主要包括卷积通道裁剪<sup>[11]</sup>、模型量化<sup>[12]</sup>和模型蒸馏<sup>[13]</sup>等方法。2017年提出的深度可分离卷积,可以在深度学习训练阶段压缩模型<sup>[14]</sup>。为进一步提升目标检测的实时性,在嵌入式端运行或者面向检测高速移动对象时,有学者引入视频运动自适应推理策略减少对相似视频帧的重复检测<sup>[15]</sup>。

在得到目标物在二维图像中的位置后,需要获取其三维信息。传统单、双目摄像头一般通过计算图像视差来感知图像中各像素点的距离,但其点云计算需要占用计算资源且对图像质量有一定要求。RGB-D传感器(例如Microsoft Kinect、Intel RealSense、ASUS Xtion Pro Live等)可以通过物理手段直接获取目标深度信息,不需要进行额外的视差计算且精度较高,近年来得到了广泛的应用。在传感器得到深度信息后,有多种算法可以实现对目标物姿态的估计,常用的有迭代最近点(Iterative closest point, ICP)算法。由于ICP算法计算量庞大,尤其是在处理大型点云数据时难以保证良好的实时性,研究者们提出了多种改进方法,如基于点线匹配的PLICP<sup>[16]</sup>、基于K-D树的优化ICP<sup>[17]</sup>等,也有学者通过深度神经网络直接估计包括目标物位置姿态在内的六维信息<sup>[18-20]</sup>。此类姿态估计算法多针对合作目标,需要预先给定待检测

目标的三维模型或需要提供大量数据集。近年来有学者针对非合作目标的姿态估计算法展开研究,如通过区域生长算法进行点云分割,利用点云聚类配准实现非合作目标姿态估计及一系列相关算法的改进<sup>[21]</sup>。

本文根据飞行机械臂工作特点,设计了一种基于YOLOv5目标检测算法和RGB-D传感器的实时目标检测与定位算法,在NVIDIA Xavier NX嵌入式开发板上构建了整体的视觉系统。该算法通过对彩色图像的识别得到目标物边界框,并根据传感器得到的深度图像来提取目标物区域的点云。算法利用模型量化方法优化深度学习推理过程,在识别精度基本不变的情况下大幅提升了算法处理速度。本文选用COCO数据集对算法目标检测部分进行测试,该数据集是目标检测领域广泛使用的开源数据集,可以较好地检验目标检测算法模型的综合性能。优化后的算法在COCO数据集的测试显示在处理时间降低到33 ms的同时平均精确率均值(Mean average precision, mAP)的下降仅在0.001以内。算法通过RGB-D传感器结合点云滤波及改进ICP算法对目标物位姿进行估计,最终利用NOKOV动作捕捉系统(Motion capture, MO-CAP)验证了该算法的有效性。算法可以为飞行机械臂的抓取工作提供实时且精确的目标物位置和姿态信息,帮助完成复杂和危险环境下对目标物的自主飞行和抓取。

## 1 关键技术及原理

### 1.1 YOLOv5目标检测算法

YOLO作为经典的基于回归的单阶段目标检测算法,其最早提出于2016年<sup>[22]</sup>。为满足飞行机械臂空中作业对视觉算法的实时性、高精度等要求,本文采用了最新的YOLOv5目标检测算法来完成目标物在二维图像中的定位。YOLOv5在整体网络结构上与YOLOv4相似,新版本使用易于部署的Pytorch深度学习框架代替Darknet框架,在输入端采用了Mosaic数据增强技术,其原理为通过随机缩放、裁剪和排列图像,增强了模型对小目标的检测效果。同时算法在骨干网中加入跨阶段模块(Cross stage partial, CSP)模块,将底层的特征映射分为两部分,然后通过跨阶段的层次结构进行合并,减少了计算量,保证了算法的准确性。在损失函数的计算部分,先前版本所采用的IOU\_Loss可以表示为

$$\text{IOU} = \frac{|A \cap B|}{|A \cup B|} \quad (1)$$

式中A、B分别表示目标物实际边界框和候选边界

框的面积。而在 YOLOv5 中采用了 GIOU\_Loss, 其表达式为

$$\text{GIOU} = \text{IOU} - \frac{|C - (A \cup B)|}{|C|} \quad (2)$$

式中  $C$  表示可以将边界框  $A$  和  $B$  覆盖在内的最小矩形框的面积。传统的 IOU\_Loss 在候选边界框与实际边界框没有相交时始终为 0, 其值并不能体现两者的远近关系, 无法进行优化, 同时存在两者对齐方式不同而 IOU\_Loss 相同的可能性。而 GIOU\_Loss 引入了变量  $C$ , 将两框以外的空间信息也考虑在内, 较好地解决了传统损失函数存在的问题。

通过这一系列的优化与改进, YOLOv5 整体的性能有着较大的提升。然而算法主要的改进方向为提升自身的识别精度, 尽管在运行效率提升部分也做了较多的工作, 但要在 GPU 性能不足的嵌入式系统上实时运行仍有改进的空间。

## 1.2 基于量化的模型推理加速

在目标检测部分, 考虑到飞行机械臂飞行及抓取过程中速度变化较大, 对检测算法的实时性要求较高。同时需要考虑机载嵌入式系统有限的计算资源和后续目标位姿估计算法所消耗的时间, 都对算法的处理速度有很高的要求。目前多数主流的深度学习目标检测算法更倾向于提高识别的精度, 而在飞行机械臂上使用则需要算法的效率越高越好。考虑到本文选用的 RGB-D 传感器获取深度图像的帧率在 30 f/s, 因此期望在目标检测部分能达到相应的输出帧率。本文利用量化方法来优化和加速模型推理过程, 在识别精度基本不变的情况下, 使目标检测速度显著提高。

基于深度学习的目标检测算法的实现可以分为训练和推理两个阶段。训练阶段的主要工作是构造神经网络, 准备数据集, 使用各种深度学习框架进行训练。这一阶段的优化主要是通过调整网络结构、进行数据扩充和调整训练超参数来实现的。YOLOv5 主要优化了这个阶段。在此基础上, 对网络推理阶段进行定量加速, 使模型得到更快的推理速度。

由于目前主流的神经网络算法是以计算机为运行环境, 为了保证计算精度, 一般采用单精度浮点数据类型 (FP32) 来进行数据计算。在神经网络的训练阶段, 由于权值参数的微小变化就能影响反向传播的梯度下降过程, 会对训练结果有非常大的影响, 因此在神经网络训练过程中使用单精度数据类型是必要的。然而, 大多数神经网络中的权值在以 FP32 数据类型存储时是存在冗余的, 实际

运行过程中半精度类型 (FP16) 甚至 INT8 就能完成绝大部分参数的计算。理论上, 如果在嵌入式系统使用半精度数据类型 (FP16) 或 INT8 数据类型, 推理速度可以提高一倍甚至更多且精度基本不会受到影响。表 1 中显示了 FP32、FP16 和 INT8 数据类型的比较。

表 1 各数据类型比较

Table 1 Comparison of data types

数据类型	数值范围	精度
FP32	$-3.4 \times 10^{38} \sim 3.4 \times 10^{38}$	$1.4 \times 10^{-45}$
FP16	$-65\,504 \sim 65\,504$	$5.96 \times 10^{-8}$
INT8	$-128 \sim 127$	1

在本文实际模型训练过程中得到的数据显示, 整体的网络权值数据各项参数均在  $\pm 30$  间浮动, 即使 INT8 数据类型, 也能较好地满足要求。同时为进一步提升推理速度, 算法通过 NVIDIA 显卡支持的 TensorRT 完成网络的垂直和水平结构整合, 将输入为相同张量和执行相同操作的层融合在一起, 减少数据运算和传输过程中的效率损耗。

本文针对 YOLOv5 算法在飞行机械臂这样的嵌入式系统部署后实时性差的问题, 采用量化加速的方法对算法进行优化, 并在 COCO 数据集进行测试。实验结果表明, 识别精度基本不变的情况下, 改进后算法在机载嵌入式系统中达到了 30 f/s 的帧率, 确保能够实现实时、准确的目标检测。

## 2 算法设计及实现

### 2.1 硬件结构

本文构建的飞行机械臂视觉系统主要包括目标检测和位姿估计两大部分, 考虑到深度学习算法的 GPU 计算资源需求, 选用 NVIDIA Xavier NX 嵌入式开发板作为机载系统, 开发板质量 200 g, 功耗 15 W, 不会为飞行机械臂带来较大的负担。在目标物位姿估计部分选用 ASUS Xtion Pro Live 传感器, 可以直接获取深度信息。图 1 展示了本文所设计的整体飞行机械臂视觉硬件系统。

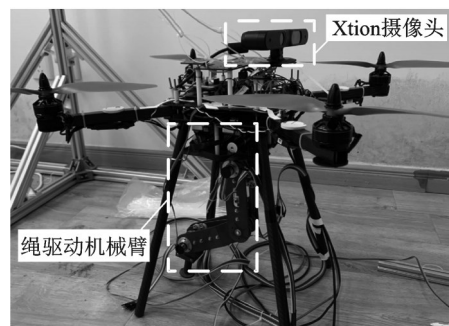


图 1 搭载视觉系统的旋翼飞行机械臂

Fig.1 Rotor aerial manipulator equipped with vision system



## 2.2 算法实现

系统的软件部分主要分为目标检测和定位以及姿态估计两个环节。首先目标物在图像中的边界框可以通过量化加速后的YOLOv5算法得到。在得到目标物二维图像区域后,飞行机械臂需要获得物体的位置和姿态,以规划抓取动作。本文采用RGB-D传感器直接获取点云信息的方式,采用直通滤波截取目标物区域点云信息并过滤去除离群点,之后对点云数据拟合其中心,并利用标定好的相机内参计算目标物在相机坐标系下的三维空间位置,最终通过机载惯导传感器获取的自身姿态换算得到目标物在世界坐标系下的位置。

飞行机械臂在抓取过程中姿态数据主要用于调整手爪角度,若目标物没有移动则在世界坐标系下其姿态是确定的,其姿态可以利用算法得到的相机坐标系下目标物姿态信息,结合机载惯导传感器提供的飞行器当前姿态信息换算得到。因此,姿态估计部分算法对实时性的要求相较位置检测略有降低。考虑到姿态估计算法对计算资源的占用,算法在各方向上设置了抓取范围,只有在飞行器与目标物距离进入该范围时才会执行姿态估计算法。在姿态估计部分采用K-D树优化的ICP算法将得到的目标物点云与设定的原始点云进行配准,计算得到相机坐标系下目标物姿态与初始姿态间的旋转矩阵,实现为飞行机械臂的抓取提供实时的目标物位置及姿态数据的功能。整体的目标检测与位姿估计算法流程图如图2所示。

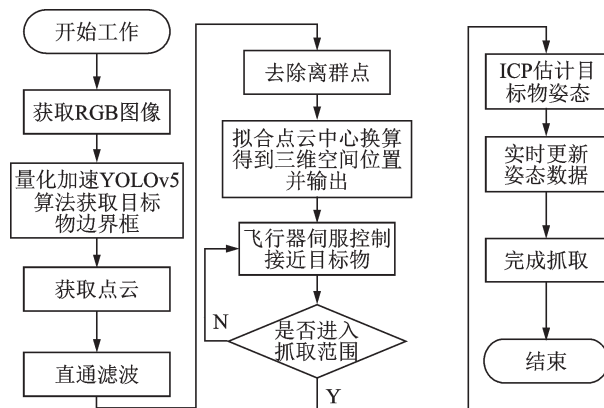


图2 目标检测与位姿估计算法流程图

Fig.2 Flow chart of object detection and pose estimation algorithm

在目标检测环节,并非所有的嵌入式开发板都支持模型量化的半精度或INT8数据类型的处理,为此本文采用了最新的NVIDIA Xavier NX嵌入式开发板,运算数据类型为半精度数据且搭载深度学习加速引擎。本文算法的具体实现过程如下:

(1)利用OpenNI2读取RGB-D传感器的输出数据,分别得到彩色图像和深度图像,同时利用

PCL库拟合点云。

(2)通过NVIDIA的TensorRT接口对YOLOv5网络进行重新构建,将计算参数的数据类型调整为FP32,同时利用TensorRT针对GPU特性对网络结构进行优化,压缩模型,实现高速、高精度的实时推理。

(3)将RGB图像输入量化加速后的YOLO算法网络,得到目标物边界框坐标信息,并利用机器人操作系统(Robot operating system, ROS)进行发布。

(4)利用在ROS中订阅得到边界框坐标信息,对RGB-D相机获取的点云数据通过PCL库进行直通滤波并去除离群点得到目标物点云。拟合点云中心点,结合飞行器自身位姿数据换算得到目标物当前在世界坐标系下的三维空间位置。

(5)利用K-D树优化后的ICP算法将当前目标物点云与设定的初始点云进行迭代,计算得到目标物的姿态信息。

算法实现过程中需要在得到目标物在相机坐标系下的坐标后,将其转换为世界坐标系下的坐标。相机自身的姿态一般由机载惯导系统给出,在实验时也可由MOCAP系统给出。可将得到的相机姿态四元数数据设为 $q = w + xi + yj + zk$ ,其对应的旋转矩阵可表示为

$$R = \begin{pmatrix} 1 - 2y^2 - 2z^2 & 2xy + 2\tau wz & 2xz - 2wy \\ 2xy - 2\tau wz & 1 - 2x^2 - 2z^2 & 2yz + 2\tau wx \\ 2xz + 2\tau wy & 2yz - 2\tau wx & 1 - 2x^2 - 2y^2 \end{pmatrix} \quad (3)$$

假设目标物在世界坐标系和相机坐标系下的坐标分别为 $a_w = (x_w \ y_w \ z_w)^T$ 和 $a_c = (x_c \ y_c \ z_c)^T$ 。平移向量 $t$ 可以通过MOCAP系统给出的相机坐标得到,则算法最终得到的目标物在世界坐标系下的坐标为

$$a_w = R^{-1}a_c + t \quad (4)$$

## 3 实验设计及结果分析

本文所有算法的实验及评估均在NVIDIA Xavier NX嵌入式开发板上进行,处理器运行模式为15W6core;软件运行环境为Ubuntu18.04, CUDA10.1, CUDNN8.0。实验采用普通水瓶作为待检测目标,目标检测算法训练集采用COCO数据集,主要从复杂的日常场景中截取,包含91类目标,33万图像和250万标签。本文直接使用已完成训练的模型,主要关注对现有模型在嵌入式开发板上推理过程的量化加速实验评估。在姿态估计算法部分,本文预先采集了目标物静置于地面的点云作为其初始姿态用于后续配准计算。

由于点云数据由传感器直接给出,目标检测的准确性直接影响了整体算法的准确性和有效性。YOLOv5 提供了多种训练得到的模型,根据系统整体的需求,实验在目标检测部分模型采用了较为轻量化的 YOLOv5s 并对其进行了量化加速处理,对改进后的算法采用 COCO 数据集的测试集在开发板上进行测试,结果如表 2 所示。

表 2 YOLO 模型量化对比实验

Table 2 Comparison experiment of YOLO model quantization

网络类型	推理时间/ ms	帧率/ (f·s <sup>-1</sup> )	mAP/%
YOLOv4	120	8.3	49.2
YOLOv5s (FP32)	55	18.2	43.1
YOLOv5s (FP16)	33	30.3	43.0

由表 2 结果可以看出,相较单精度推理模式的 YOLOv5s,半精度推理模式下算法有明显的加速效果,帧率提升明显,而最终的 mAP 损失只有 0.1%,基本不影响识别精度。实验结果表明,通过模型量化手段加速模型推理可以在保证识别精度的情况下大幅减少推理时间,在嵌入式端应用有良好的实时检测效果。

考虑到团队现有的飞行机械臂尚不支持完成完整的飞行抓取控制实验,为验证算法的有效性,对目标物位姿估计算法的测试选择在室内可控环境进行。图 3 展示了目标物位姿估计的实验环境。

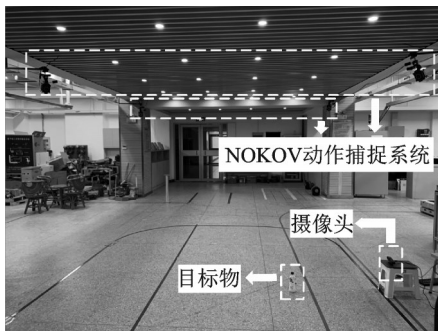
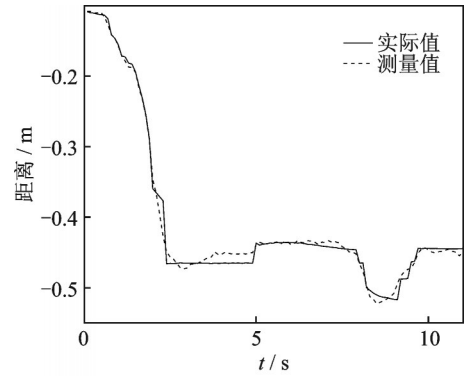


图 3 目标物位姿估计实验平台

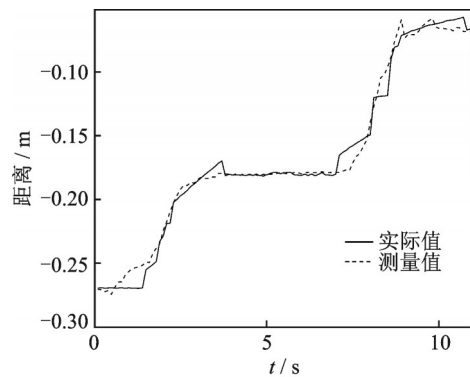
Fig.3 Object pose estimation experimental platform

实验主要利用 MOCAP 对算法进行验证。通过在摄像头和目标物上放置多个反光标志物分别建立刚体,系统可以实时输出两者在世界坐标系下的位姿数据。本文实验中摄像头自身姿态数据的获取采用 MOCAP,而相机实际搭载于飞行机械臂系统时,自身的姿态数据由机载惯导系统获取,得到的姿态信息可用于飞行器控制和相机坐标转换。实验通过向各方向移动摄像头模拟飞行机械臂飞行过程中相机的运动,测试算法的整体性能。由于 MOCAP 系统的精度极高,实验设置系统输出坐标为放置在目标物中心位置的反光标志物坐

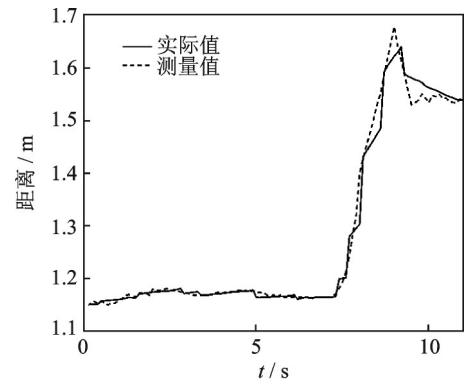
标,根据预先测量的目标物尺寸数据换算得到其几何中心坐标并将其视作实际值,对比视觉系统经目标检测,拟合点云中心以及坐标转换得到的坐标信息,拟合曲线后对比结果如图 4 所示。



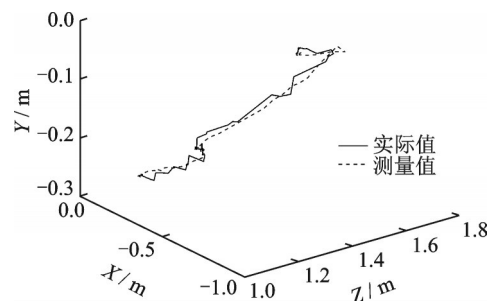
(a) X-axis direction



(b) Y-axis direction



(c) Z-axis direction



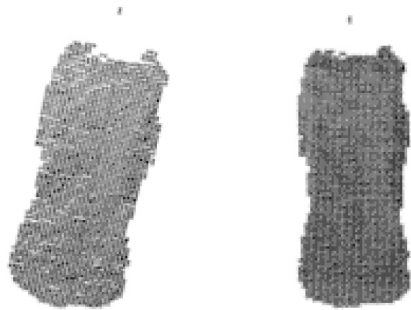
(d) Three-dimensional trajectory

图 4 算法得到目标物位置与实际值的对比

Fig.4 Comparison of object position obtained by algorithm and actual value

实验结果显示,算法在各方向上的检测和跟踪效果较好,对目标物三维空间坐标估计在各方向上的实时数据误差均方根平均值小于0.02 m,可以较好地为飞行机械臂抓取工作提供目标物三维空间位置信息。同时根据实验中各方向上检测的结果曲线对比显示,Z轴方向位置的检测相比X与Y轴效果略有不佳,这可能与待检测的目标物形状不规则以及现场光照对传感器的影响有关。在检测速率方面由于加入量化推理加速,算法目标检测与定位部分整体检测速率降低至35 ms,与RGB-D传感器自身视频流30 f/s的速率基本匹配,验证了算法实时检测与定位的有效性。

在姿态估计实验部分,本文采用的是基于K-D树优化的ICP算法,在GPU计算加速下通过调整目标物姿态得到多组检测数据,检测时间平均耗时为587 ms,若适当降低精度要求,则检测速率也能得到进一步提高。图5展示了一组姿态估计实验效果图,该组数据共有有点云数目3 261个,姿态估计耗时542 ms。考虑到飞行机械臂实际抓取过程中对姿态估计实时性要求低于对位置检测实时性的要求,结果基本满足算法要求。



(a) Point cloud after filtered (b) Setting template point cloud

图5 目标物ICP姿态估计效果图

Fig.5 ICP object pose estimation

## 4 结 论

本文提出了一种面向旋翼飞行机械臂自主抓取的基于YOLOv5和RGB-D传感器的目标物识别与定位算法。为解决传统目标检测算法精度较低而深度学习算法难以在嵌入式端保证良好的实时工作性能的问题,对YOLOv5算法进行模型量化加速,采用半精度推理模式优化神经网络推理过程,实现了识别精度基本不变的情况下算法处理速度的大幅提高;采用RGB-D传感器代替传统单双目摄像头获取点云,利用滤波算法得到目标物附近点云并去除离群点,之后利用K-D树优化ICP算法估计目标物姿态信息。该算法实现了对单一目标物体的识别定位与姿态估计,各部分分别在CO-

CO数据集和动作捕捉系统下进行了实验测试,有效性均得到了验证。下一步将会继续研究相应的视觉伺服控制策略,结合当前视觉算法部署在旋翼飞行机械臂上,最终实现飞行机械臂对目标物的自主飞行抓取。同时也将在视觉算法方面有针对性地根据不同应用场景进行相应优化,帮助飞行机械臂更好地在各种危险环境下替代人工作业。

## 参考文献:

- [1] 杨斌,何玉庆,韩建达,等. 作业型飞行机器人研究现状与展望[J]. 机器人, 2015, 37(5): 628-640.  
YANG Bin, HE Yuqing, HAN Jianda, et al. Survey on aerial manipulator systems[J]. Robot, 2015, 37(5): 628-640.
- [2] RAMON-SORIA P, ARRUE B C, OLLERO A. Grasp planning and visual servoing for an outdoors aerial dual manipulator[J]. Engineering, 2020, 6(1): 77-88.
- [3] LUO B, CHEN H, QUAN F, et al. Natural feature-based visual servoing for grasping target with an aerial manipulator[J]. Journal of Bionic Engineering, 2020, 17: 215-228.
- [4] 孙敬陶,钟杭,王耀南,等. 旋翼飞行机械臂的混合视觉伺服和分层控制方法[J]. 仪器仪表学报, 2018, 39(7): 56-65.  
SUN Jingtao, ZHONG Hang, WANG Yaonan, et al. Hybrid visual servoing with hierarchical task control for unmanned aerial manipulation[J]. Chinese Journal of Scientific Instrument, 2018, 39(7): 56-65.
- [5] BRUCE J, BALCH T, VELOSO M. Fast and inexpensive color image segmentation for interactive robots [C]//Proceedings of 2000 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2000) (Cat. No. 00CH37113). Takamatsu, Japan: IEEE, 2000, 3: 2061-2066.
- [6] ZHOU H, YUAN Y, SHI C. Object tracking using SIFT features and mean shift[J]. Computer Vision and Image Understanding, 2009, 113(3): 345-352.
- [7] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA: IEEE, 2014: 580-587.
- [8] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137-1149.
- [9] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector[C]//Proceedings of Eu-

- ropean Conference on Computer Vision. Springer, Cham: Springer, 2016: 21-37.
- [10] REDMON J, FARHADI A. YOLOv3: An incremental improvement[EB/OL]. (2018-04-08) [2021-05-31]. <https://arxiv.org/abs/1804.02767.pdf>.
- [11] HE Y, ZHANG X, SUN J. Channel pruning for accelerating very deep neural networks[C]//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017: 1389-1397.
- [12] HAN S, MAO H, DALLY W J. Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding[EB/OL]. (2016-02-15) [2021-05-31]. <https://arxiv.org/abs/1510.00149.pdf>.
- [13] HINTON G, VINYALS O, DEAN J. Distilling the knowledge in a neural network[J]. Computer Science, 2015, 14(7): 38-39.
- [14] CHOLLET F. Xception: Deep learning with depth-wise separable convolutions[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE, 2017.
- [15] 尹彦卿, 龚华军, 王新华. 基于YOLOv3的嵌入式实时视频目标检测算法[J]. 计算机工程, 2020, 46(2): 230-234.
- YIN Yanqin, GONG Huajun, WANG Xinhua. Embedded real-time video object detection algorithm based on YOLOv3[J]. Computer Engineering, 2020, 46(2): 230-234.
- [16] CENSI A. An ICP variant using a point-to-line metric [C]//Proceedings of 2008 IEEE International Conference on Robotics and Automation. [S. l.]: IEEE, 2008: 19-25.
- [17] 刘江, 张旭, 朱继文. 一种基于K-D树优化的ICP三维点云配准方法[J]. 测绘工程, 2016, 25(6): 15-18.
- LIU Jiang, ZHANG Xu, ZHU Jiwen. ICP three-dimensional point cloud registration based on KD tree optimization[J]. Engineering of Surveying and Mapping, 2016, 25(6): 15-18.
- [18] XIANG Y, SCHMIDT T, NARAYANAN V, et al. Posecnn: A convolutional neural network for 6D object pose estimation in cluttered scenes[EB/OL]. (2018-05-26) [2021-05-31]. <https://arxiv.org/abs/1711.00199.pdf>.
- [19] GUALTIERI M, TENPAS A, SAENKO K, et al. High precision grasp pose detection in dense clutter [C]//Proceedings of 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Daejeon, Korea (South): IEEE, 2016: 598-605.
- [20] GAO G, LAURI M, WANG Y, et al. 6D object pose regression via supervised learning on point clouds [EB/OL]. (2020-01-24) [2021-05-31]. <https://arxiv.org/abs/2001.08942.pdf>.
- [21] 卢祺, 林婷婷, 李程鹏, 等. 空间非合作目标点云聚类配准方法[J]. 红外与激光工程, 2021, 50(9): 362-371.
- LU Qi, LIN Tingting, LI Chengpeng, et al. Improved registration algorithm for spatial non-cooperative target point cloud clustering[J]. Infrared and Laser Engineering, 2021, 50(9): 362-371.
- [22] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 2016: 779-788.

(编辑: 胥橙庭)