

本文引用格式: 孟繁,周馨墨,吴烽云,等.基于视觉 SLAM 的水果采摘机器人研究进展[J].自动化与信息工程,2023,44(5):8-13;31.
MENG Fan, ZHOU Xinzhaoh, WU Fengyun, et al. Research progress in fruit picking robots based on visual SLAM [J]. Automation & Information Engineering, 2023,44(5):8-13;31.

基于视觉 SLAM 的水果采摘机器人研究进展*

孟繁^{1,2} 周馨墨² 吴烽云³ 邹天龙²

(1.仲恺农业工程学院, 广东 广州 510225

2.佛山市中科农业机器人与智慧农业创新研究院, 广东 佛山 528010

3.广州商学院, 广东 广州 511363)

摘要: 水果采摘机器人是一种具有较大潜力的农业自动化技术, 不仅需在复杂且不断变化的环境中连续作业, 还面临地形、树木分布、天气变化、环境光照、遮挡等多种挑战。视觉同步定位与地图构建 (SLAM) 作为一种成本低廉且能够提供丰富语义信息的技术, 有望提高水果采摘机器人的效率和自动化程度。近年来, 视觉 SLAM 在水果采摘机器人领域取得了一系列重要进展, 主要包括深度学习优化方法、基于点线特征的优化方法、基于 RGB-D 的视觉 SLAM 优化方法、动态环境中视觉 SLAM 优化方法、回环检测和后端优化方法等。未来, 水果采摘机器人领域的研究将朝着更高的自动化程度和采摘效率方向发展; 可能的发展方向包括更复杂的感知系统、更智能的决策算法、更强大的硬件支持等。此外, 水果采摘机器人在多样化果园中的适应性和鲁棒性研究也将引领这一领域的发展。通过不断推动视觉 SLAM 技术的创新, 水果采摘机器人有望成为现代农业的重要工具, 提高水果产量并减轻农业劳动力短缺的问题。

关键词: 同步定位与地图构建; 水果采摘机器人; 机器视觉; 自动化

中图分类号: TP391.4

文献标志码: A

文章编号: 1674-2605(2023)05-0002-07

DOI: 10.3969/j.issn.1674-2605.2023.05.002

Research Progress in Fruit Picking Robots Based on Visual SLAM

MENG Fan^{1,2} ZHOU Xinzhaoh² WU Fengyun³ ZOU Tianlong²

(1.Zhongkai College of Agricultural Engineering, Guangzhou 510225, China

2.Foshan Zhongke Innovation Research Institute of Intelligent Agriculture, Foshan 528010, China

3.Guangzhou College of Commerce, Guangzhou 511363, China)

Abstract: Fruit picking robots are a promising agricultural automation technology, but they must operate continuously in complex and constantly changing environments, facing various challenges such as terrain, tree distribution, weather changes, environmental lighting, and occlusion. Visual SLAM, as a low-cost technology that can provide rich semantic information, has attracted widespread research interest as it has the potential to significantly improve the efficiency and automation of fruit picking robots. In recent years, visual SLAM has made a series of important progress in the field of fruit picking robots. Mainly including deep learning optimization methods; Optimization method based on point line features; Visual SLAM optimization method based on RGB-D; Visual SLAM optimization methods in dynamic environments; Loop detection and backend optimization methods. In the future, research in the field of fruit picking robots will move towards higher levels of automation and harvesting efficiency. Possible development directions include more complex perception systems, more intelligent decision algorithms, and stronger hardware support. In addition, research on the adaptability and robustness of robots in diverse water orchards will continue to lead the development of this field. By continuously promoting the innovation of visual SLAM technology, fruit picking robots are expected to become an important tool in modern agriculture, increasing production and alleviating the problem of agricultural labor shortage.

Keywords: simultaneous localization and mapping; fruit picking robots; machine vision; automation

0 引言

在非结构化的环境中进行实时、准确和鲁棒性的定位与导航对水果采摘机器人至关重要。随着视觉传感器技术的不断发展,尤其是相机的低成本、微型化和高信息密度等特点,使视觉同步定位与地图创建(simultaneous localization and mapping, SLAM)受到广泛关注。视觉 SLAM 的核心思想是先通过相机连续获取图像帧,再利用图像帧与帧之间的信息来估计相机的运动轨迹并构建周围环境的三维地图。与其他的 SLAM 技术相比,视觉 SLAM 能够提供更加丰富的场景信息^[1]。因此,视觉 SLAM 不仅用于定位和导航^[2]等任务,还用于场景理解、水果识别等任务。

水果采摘机器人通常在室外果园复杂环境中作业,传统的视觉 SLAM 技术在这种动态变化和光照变化的环境中存在挑战^[3-7]。近年来,随着计算机硬件性能的不断提高,深度学习在多个领域成为研究热点^[8-10]。通过深度学习优化的视觉 SLAM 系统具有更强的特征提取能力、动态环境鲁棒性、模型通用性和可迁移性^[11-13]。同时多传感器融合优化 SLAM 方法也是未来发展方向^[14]。本文通过介绍视觉 SLAM 系统的基本原理和关键技术,阐述在复杂环境中水果采摘机器人的视觉 SLAM 优化方法和未来发展方向。

1 视觉 SLAM 系统基本原理

视觉 SLAM 系统的工作流程主要包括视觉传感器数据采集、前端视觉里程计、后端非线性优化、回环检测和建图 5 个模块,如图 1 所示。

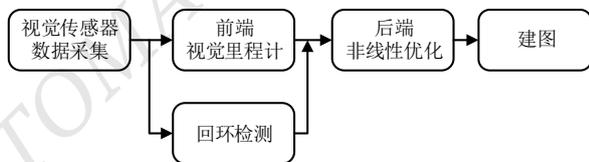


图 1 视觉 SLAM 系统结构图

在视觉传感器数据采集模块,相机或其他视觉传感器采集的图像信息为系统提供原始数据;在前端视觉里程计模块,系统通过在连续的图像帧中提取和匹

配关键点特征,识别场景中的路标点并估算相机的实时位置和姿态,这会随着时间累积误差;为纠正误差,在后端非线性优化模块,参考各图像帧之间的共视信息进行位姿调整和精化,确保长时间运行下视觉 SLAM 系统的准确性和稳定性;若遇到已访问过的区域或场景,回环检测模块通过 2 个图像帧的相似性是否超过预定阈值来判断是否形成回环,一旦检测到回环,对所有路标点和位姿进行优化,以确保地图的一致性和准确性;在建图模块,优化后的路标点被用于构建高质量的三维地图,该地图不仅为水果采摘机器人提供导航信息,还为后续的路径规划和障碍物避让提供基础。

2 视觉 SLAM 优化方法

2.1 深度学习优化方法

文献[15]提出一种融合目标检测网络技术的动态视觉 SLAM 算法,在视觉 SLAM 系统的主跟踪流程中嵌入一个辅助的并行线程,如图 2 所示。

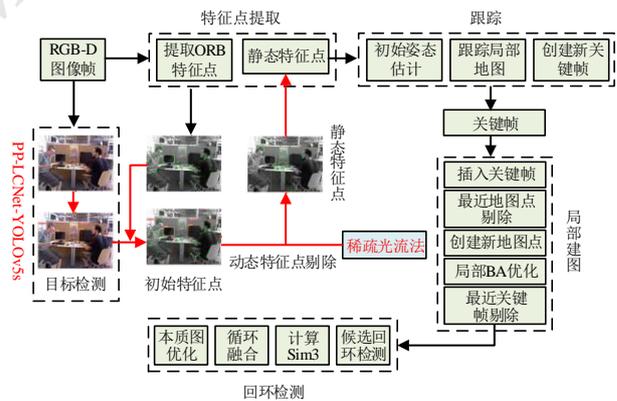


图 2 文献[15]提出的算法流程图

该并行线程结合了优化的目标检测网络与稀疏光流技术,可识别并排除动态物体产生的不稳定特征点,使系统仅使用来源于静态环境的特征点进行特征点匹配和相机位姿估计,从而提高视觉 SLAM 系统的稳定性和定位准确性。

文献[16]提出一种针对回环检测优化的视觉 SLAM 算法。在前端处理阶段,首先,利用定向来自加速段测试的特征(features from accelerated segment

test, FAST) 和旋转二进制鲁棒独立基本特征 (binary robust independent elementary features, BRIEF) 对连续帧图像进行特征提取和匹配; 然后, 通过多点透视成像 (perspective-n-point, PnP) 方法对成功匹配的特征点进行解析, 估计相机的动态轨迹, 并筛选出关键帧。在后端优化阶段, 首先, 采用 SqueezeNet 卷积神经网络对图像进行特征向量提取; 然后, 通过计算图像特征向量的余弦相似度, 判断是否发生了回环事件, 一旦检测到回环, 算法在位姿图中添加相应的约束条件, 并运用图优化理论对系统的位姿进行全局优化, 确保视觉 SLAM 系统具有更高的精度与稳定性。

文献[17]提出一种结合语义信息的 V-SLAM 技术, 主要包括语义地图建构和语义回环检测 2 个核心部分。其中, 在语义地图建构部分, 采用 YOLOv3 深度学习网络对关键帧进行二维图像语义注释, 采用条件随机场模型将二维图像语义注释与场景内部的分割聚类信息结合, 完成语义地图构建; 在语义回环检测部分, 以关键帧的语义标注为基础, 结合去除动态物体特征点的策略, 对传统的词典法进行改进, 有效地去除可能引起误判的动态物体, 从而提高机器人在动态环境中的定位精度。

文献[18]提出利用无监督栈式卷积自编码器 (convolutional autoencoder, CAE) 模型来提取图像特征。经过训练的 CAEs 卷积神经网络可以学习输入图像特征, 并将输入图像特征用于闭环检测。

文献[19]提出一种融合视觉-惯性 SLAM 的非结构化柑橘园 SLAM 方法。通过 BiSeNetV1 对二维 RGB 图像进行语义分割并将其映射到点云, 重建三维语义点云图; 引入统计离群值去除过滤器和 OctoMap 进行后处理, 去除离群值并估计三维空间中的障碍物, 构建更准确、高效和灵活的地图。

2.2 基于点线特征的优化方法

文献[20]提出一种基于点线特征的高速 SLAM 策略。在追踪非关键帧时, 采用描述子来匹配点特征, 同时依赖几何约束来匹配线特征。当插入新的关键帧时, 计算线特征的描述子, 完成关键帧之间的线特征匹配, 并利用线特征三角化算法生成地图线。为确保

视觉 SLAM 系统的实时性, 减少线特征匹配过程的计算量, 利用线特征提供的深度信息构造虚拟的右目线段, 并提出线特征重投影误差的计算策略, 优化视觉 SLAM 系统性能, 特别是在纹理较弱的场景中可以提高视觉 SLAM 系统的定位精度和鲁棒性。

文献[21]提出一种基于点线不变量的线特征匹配策略。该点线不变量是通过对线段及其两侧特征点的局部几何结构进行编码来定义的, 可在现有的特征点上直接进行线匹配, 不仅加快了匹配速度, 还提高了匹配的准确性。此外, 还在点线特征融合阶段加入权重分配机制, 根据场景中的特征丰富程度, 在构建误差函数时为点线特征赋予不同的权重, 进一步优化了算法的性能和鲁棒性。

文献[22]提出一种结合点线特征的视觉 SLAM 方法, 采用模块化的技术策略。首先, 从相机采集的环境图像中提取点与线的特征, 并通过帧间特征进行连续跟踪; 然后, 结合改进的正规迭代最近点 (normal iterative closest point, NICP) 算法和特定的关键帧匹配技术搭建里程计系统; 接着, 为了实现全局一致性的位姿, 融合基于点线特征词典的闭环检测技术和佐治亚理工大学平滑和映射库 (Georgia Tech smoothing and mapping library, GTSAM) 图优化方法, 得到三维点云地图; 最后, 利用机器人技术中间件搭建系统, 不仅提升了系统的实时性能, 还为功能模块提供了更强的可扩展性和灵活性。

文献[23]提出一种并行跟踪和映射 SLAM (parallel tracking and mapping-SLAM, PLP-SLAM) 方法。该方法结合点、线段和平面 3 种特征, 在扩展卡尔曼滤波的框架内工作。首先, 利用点特征来估计机器人的当前位姿; 然后, 构建综合点、线、平面特征的观测模型; 最后, 设计带有平面约束的线段特征数据关联方法, 并构建相应的系统状态更新模型, 进一步增强了视觉 SLAM 系统的稳定性和精度。该方法通过线段和平面特征来描述环境结构更为准确、高效。

2.3 基于 RGB-D 的视觉 SLAM 优化方法

文献[24]提出一种红绿蓝深度 (red green blue-depth, RGB-D) SLAM 方法, 构建基于重投影深度差

值的累积模型，将图像分为动态和静态 2 大区域。考虑到动态区域可能存在过分割的风险，排除与已匹配地图点欧氏距离过大的动态区域特征点，采用 t 分布方法估计剩下特征点的静态概率。在这种情境下，静态区域的特征点和动态区域中可能被认为是静态的特征点都被纳入位姿的优化过程，但被赋予不同的权重。通过这种策略，可以得到经过筛选和优化的位姿，从而在动态环境中实现更为精确和稳定的定位。

文献[25]提出一种基于增强分割技术的 RGB-D SLAM 方法。利用实例分割网络和深度图像聚类来检测当前图像帧中是否存在漏分割的情况。一旦检测到漏分割，结合多帧数据来修复这些分割缺陷。在此基础上，提取当前图像帧中的 Shi-Tomasi 角点，并使用对称转移误差来识别动态角点集合。结合修复后的实例分割结果来确定场景中每个实例对象的运动状态。仅利用静态特征点来追踪相机的位姿，并构建实例级的语义八叉树地图，从而在动态环境中实现高效且准确的定位和地图创建。

文献[26]提出一种融合直接法和特征点法的 RGB-D SLAM 算法。该算法结合了直接法和特征点法的优点，利用后端优化和闭环检测，在复杂环境中提高定位和建图的准确性。在定位环节，利用直接法提供初步估计的相机位姿，通过特征点匹配和重投影

误差最小化进一步优化相机位姿。针对一些特定的情况，如稀疏纹理、光线变化和移动物体，通过优化位姿输出策略和地图点的筛选方法，使算法能够更好地应对这些挑战。在后端优化部分，设计一种关键帧选择策略，同时考虑了直接法和特征点法选取的关键帧，分别维护这两种关键帧，并在滑动窗口和特征地图中进行并行优化。为了进一步增强 SLAM 的全局一致性，对全局关键帧进行闭环检测和优化，为视觉 SLAM 系统提供了一个更加鲁棒和准确的解决方案。

2.4 动态环境中视觉 SLAM 优化方法

文献[27]提出一种融合光流和实例分割技术的 SLAM 方法。针对动态物体与静态背景在光流方向上的差异，提出一种能够实时检测动态区域的掩模算法。该算法在 ORB-SLAM2 的基础上，能够迅速排除动态区域的特征点，去除与相机移动相关的光流，留下与动态物体自身运动相关的光流信息，并通过聚类分析光流大小，准确检测出动态区域。利用动态掩模进一步筛选出与动态物体关联的特征点，结合对极几何约束，去除局部建图过程中的动态路标点，在动态环境中提高了视觉 SLAM 系统的性能，确保更为准确和稳定的地图构建和相机定位。

文献[28]提出一种多焦距动态立体视觉 SLAM 方法，框架图如图 3 所示。

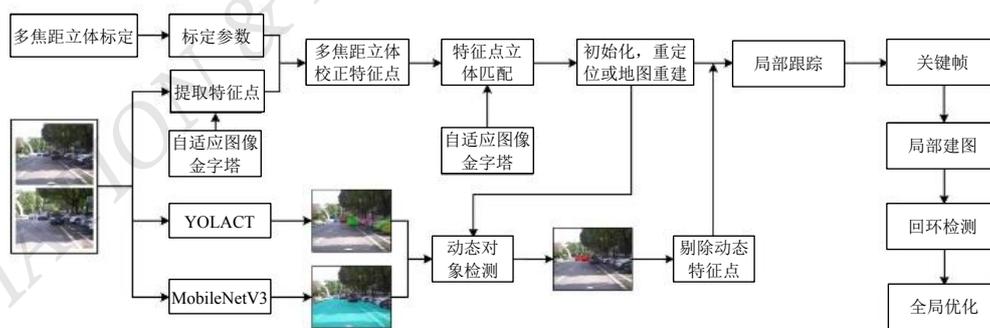


图 3 文献[28]提出的多焦距动态立体视觉 SLAM 方法框架图

与传统立体相机只能固定在一种焦距模式下工作不同，该方法能够同时兼顾远距离的景物和宽广的视场，增强对不同距离和方向物体的感知能力。不同于对整幅图像进行修正，其只调整特征点的位置，因此立体匹配更加准确且计算效率更高。同时设计了一

种自适应的特征提取与匹配算法，针对多焦距图像，增加了特征匹配的数量和准确性。结合多视图几何、区域特征流和相对距离判定来识别并剔除动态物体的特征点，从而确保视觉 SLAM 的准确性和鲁棒性。

文献[29]提出一种具备动态区域剔除能力的双目

视觉 SLAM 方法。该方法的核心思想是检测并剔除受动态目标影响的特征点。利用立体视觉的几何约束来鉴别场景中的动态稀疏特征点，结合场景的深度信息和颜色特征，实现场景区域的高效分割。通过已标识的动态点与场景分割结果，标出场景内的动态区块。针对双目 ORB-SLAM 策略，从算法中移除动态区域中的特征点，减少动态物体对 SLAM 精度的不良影响，为复杂动态环境中的高精度定位和建图提供了有效策略。

文献[30]提出一种在线构建视觉词典的闭环检测方法。该方法优先考虑与已有单词 Surf 描述之间欧式距离最大的特征点。为了使词典更具代表性并降低量化误差，对特征点与单词 Surf 描述之间的最近邻约束条件进行改进，不仅可以更准确地描述和捕获场景中

的关键信息，还能避免因视觉混淆引入的误差。

文献[31]提出面向场景变化的 SLAM 框架—SceneSLAM。该框架支持场景检测与 SLAM 算法的智能切换，实现对环境的动态自适应。依据场景判定，SceneSLAM 自动选择适宜传感器驱动的 SLAM 方法，增强环境的适应性。设计了可以检测室内、室外和低光环境的场景识别模型，并为各种光线变化提供自适应的 SLAM 模型。在算法切换时，系统能自动进行坐标和尺度转换，确保定位与地图构建结果具备全局一致性和尺度一致性。

2.5 回环检测和后端优化方法

文献[32]提出一种结合词袋模型、图像结构校验及跟踪预测模型的算法，算法流程图如图 4 所示。

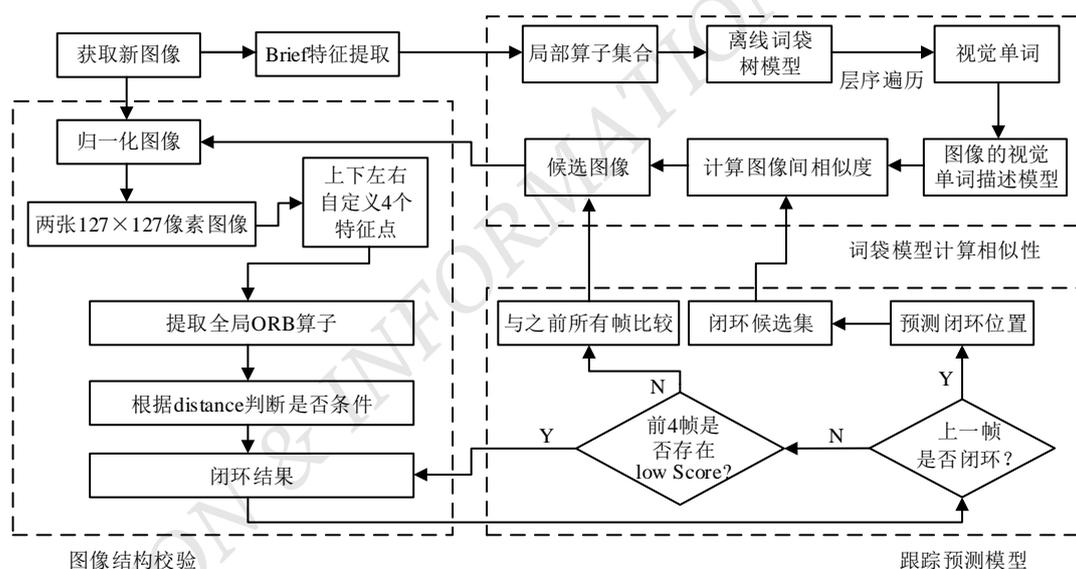


图 4 文献[32]提出的回环检测流程图

首先，通过整合局部和全局特征，构建一个词袋模型，该模型对比不同图像的视觉单词，能有效地筛选出可能形成闭环的候选帧；然后，采用图像结构校验模块对当前图像和闭环候选图像进行灰度化和归一化处理，处理后的图像可作为局部特征的上下文，计算全局描述符；最后，设计跟踪预测模块，避免随着图像数量增加传统的回环检测计算时间大幅上升的问题，加快闭环检测的计算速度，确保回环检测实时性与准确性的平衡。

文献[33]提出一种结合 2 个视觉里程计的后端位姿图优化策略，系统框架如图 5 所示。采用一个运行频率高但精度稍低的端到端视觉里程计和一个运行频率较低但精度更高的视觉里程计同时工作，提供的的数据约束是局部优化的基础，通过高斯-牛顿方法对其进行迭代优化。在进行全局优化时，基于关键帧进行场景匹配，同时与局部优化同步进行，确保高效和准确的轨迹估计。

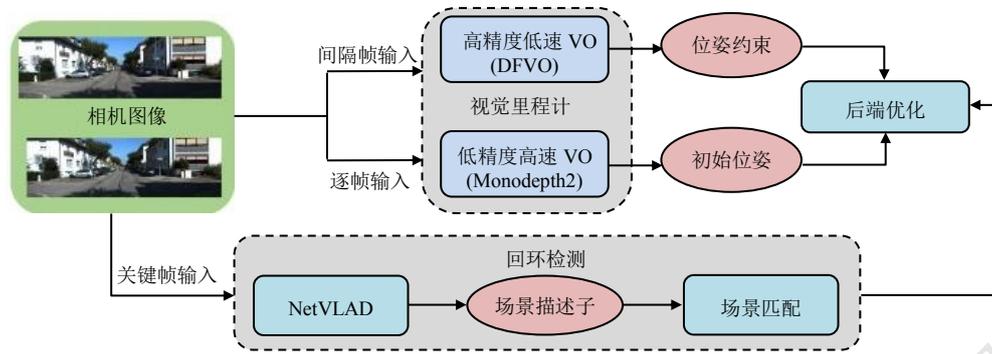


图5 文献[33]提出的系统框架图

3 结论

深度学习在视觉 SLAM 领域已经展现出较大的潜力和价值。尽管这是一个相对较新且不断进化的研究领域,但它已经得到了广大研究者的关注。深度学习与视觉 SLAM 的结合已在视觉里程计、场景识别及全局优化等多个任务上具有突出的表现。深度神经网络以其出色的非线性拟合能力,在处理传统方法难以模拟的复杂非线性关系时,表现出了优越的鲁棒性。此外,将语义信息纳入传统的视觉 SLAM 可以进一步增强对图像特征的理解,为创建语义丰富的高精度地图开辟了新途径。

参考文献

- [1] CHEN M, TANG Y, ZOU X, et al. 3D global mapping of large-scale unstructured orchard integrating eye-in-hand stereo vision and SLAM[J]. Computers and Electronics in Agriculture, 2021,187:106237.
- [2] 曹军,曾碧,何元烈.基于路标观测的改进 EKF-SLAM 算法[J]. 自动化与信息工程,2014,35(1):21-26.
- [3] 李旗,倪江南.一种运用计算机视觉和光学导航技术的采摘机器人[J].电子世界,2021(22):52-53.
- [4] 杨筱慧.猕猴桃采摘机器人视觉导航路径方法研究[J].科技风,2021(15):23-24.
- [5] 倪江楠,石新龙.基于光学导航的采摘机器人系统设计及定位方法研究[J].农机化研究,2021,43(3):221-225.
- [6] 张晓亮.基于图像融合技术的采摘机器人自主导航技术研究[J].农机化研究,2020,42(11):193-196.
- [7] 郭素娜,张丽,刘志刚.一种高精度自主导航定位的葡萄采摘机器人设计[J].农机化研究,2016,38(7):20-24.
- [8] VINCENT P, LAROCHELLE H, LAJOIE I, et al. Stacked

denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion[J]. Journal of machine learning research, 2010, 11(12).

- [9] GU J, WANG Z, KUEN J, et al. Recent advances in convolutional neural networks[J]. Pattern recognition, 2018,77: 354-377.
- [10] ZHANG C, BENGIO S, HARDT M, et al. Understanding deep learning (still) requires rethinking generalization[J]. Communications of the ACM, 2021,64(3):107-115.
- [11] 刘金辉,陈孟元,韩朋朋,等.面向移动机器人大视角运动的图神经网络视觉 SLAM 算法[J/OL].系统仿真学报:1-15. [2023-10-08].DOI:10.16182/j.issn1004731x.joss.23-0053.
- [12] 史蓝兮,颜文旭,倪宏宇,等.基于改进目标检测的动态场景 SLAM 研究[J/OL].系统仿真学报:1-13.[2023-10-08].http://kns.cnki.net/kcms/detail/11.3092.V.20230215.1438.005.html.
- [13] 陈文峰,张学习,蔡述庭,等.基于 YOLO v3 和传感器融合的机器人定位建图系统[J].自动化与信息工程,2021,42(2):34-38;48.
- [14] CREMONA J, CIVERA J, KOFMAN E, et al. GNSS-stereo-inertial SLAM for arable farming[J]. Journal of Field Robotics, 2023.
- [15] 蒋畅江,刘朋,舒鹏.基于改进 YOLOv5s 的动态视觉 SLAM 算法[J/OL].北京航空航天大学学报:1-10. https://doi.org/10.13700/j.bh.1001-5965.2023.0154.
- [16] 郭烈,葛平淑,王肖,等.基于卷积神经网络优化回环检测的视觉 SLAM 算法[J].西南交通大学学报,2021,56(4):706-712;768.
- [17] 郑冰清,刘启汉,赵凡,等.一种融合语义地图与回环检测的视觉 SLAM 方法[J].中国惯性技术学报,2020,28(5):629-637.
- [18] 张云洲,胡航,秦操,等.基于栈式卷积自编码的视觉 SLAM 闭环检测[J].控制与决策,2019,34(5):981-988.

(下转第 31 页)