



## 基于实例分割的室内动态场景SLAM

席志红, 温家旭

引用本文:

席志红, 温家旭. 基于实例分割的室内动态场景SLAM[J]. *应用科技*, 2021, 48(6): 18–22.

XI Zhihong, WEN Jiaxu. Indoor dynamic scene SLAM based on instance segmentation[J]. *Applied science and technology*, 2021, 48(6): 18–22.

在线阅读 View online: <https://dx.doi.org/10.11991/ykj.202103022>

## 您可能感兴趣的其他文章

Articles you may be interested in

### 快速室内视觉同步定位与建图研究

Research on fast visual simultaneous indoor localization and mapping

应用科技. 2021, 48(3): 1–6 <https://dx.doi.org/10.11991/ykj.202012024>

### 面向动态环境的机器人同步定位与建图技术

Research on robot simultaneous localization and mapping technology in a dynamic environment

应用科技. 2021, 48(1): 36–41, 47 <https://dx.doi.org/10.11991/ykj.202007016>

### 基于深度特征融合的三维动态手势识别

3D dynamic gesture recognition based on depth feature fusion

应用科技. 2021, 48(1): 18–24 <https://dx.doi.org/10.11991/ykj.202005012>

### 应用于增强现实变电运检三维场景重建的增量八叉树算法

The incremental octree schedule of digital transformer substation scenario reconstructions in augmented–reality environments

应用科技. 2020, 47(5): 58–63 <https://dx.doi.org/10.11991/ykj.202001013>

### 基于滑窗非线性优化的双目视觉SLAM算法

Sliding window based binocular SLAM using nonlinear optimization

应用科技. 2020, 47(1): 55–60 <https://dx.doi.org/10.11991/ykj.201908009>

### 基于点云拼接的植物三维模型重建

Reconstruction of three dimensional model of plant based on point cloud stitching

应用科技. 2019, 46(1): 19–24 <https://dx.doi.org/10.11991/ykj.201806003>



微信公众平台



期刊网址

DOI: 10.11991/yykj.202103022

网络出版地址: <https://kns.cnki.net/kcms/detail/23.1191.U.20210611.1344.012.html>

# 基于实例分割的室内动态场景 SLAM

席志红, 温家旭

哈尔滨工程大学 信息与通信工程学院, 黑龙江 哈尔滨 150001

**摘 要:** 针对动态物体在室内场景中影响定位与建图准确性的问题, 提出一种基于实例分割的室内动态场景同步定位与地图构建(SLAM)系统。首先, 利用 Mask RCNN 网络对输入的图像序列进行实例分割, 结合多视几何方法提高动态对象分割效果; 然后提取图像中的快速提取和描述(ORB)特征点, 将动态物体内的特征点剔除, 利用静态的特征点进行位姿估计; 最后, 利用背景修复和点云拼接技术实现室内场景实例级稠密点云地图和语义八叉树地图的构建。在公开动态场景数据集上进行多次测试的实验结果表明, 相对于 ORB-SLAM2 系统, 该系统相机位姿估计误差明显降低, 对环境信息的理解能力得到提升, 对后续移动机器人的导航工作具有重要的意义。

**关键词:** 实例分割; 动态场景; SLAM; Mask RCNN; 背景修复; 点云拼接; 实例级稠密点云地图; 语义八叉树地图

中图分类号: TP242.6

文献标志码: A

文章编号: 1009-671X(2021)06-0018-05

## Indoor dynamic scene SLAM based on instance segmentation

XI Zhihong, WEN Jiaxu

College of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China

**Abstract:** Aiming at the problem that dynamic objects in indoor scenes affect the accuracy of positioning and mapping, an indoor dynamic scene simultaneous localization and mapping(SLAM) system based on instance segmentation is proposed. First, the Mask RCNN network is used to segment the input image sequence, and combines the multi-view geometric method to improve the dynamic object segmentation effect. Then the oriented fast and rotated brief(ORB) feature points in the image are extracted, the feature points in the dynamic object are eliminated, and the static feature points are used for pose estimation. Finally, the use of background restoration and point cloud registration technologies realize the construction of instance-level dense point cloud maps and semantic octree maps of indoor scenes. The experimental results of multiple tests on the public dynamic scene data set show that compared with the ORB-SLAM2 system, the camera pose estimation error of the system is significantly reduced, and understanding of environmental information is improved, which is of great significance to the subsequent navigation of mobile robots.

**Keywords:** instance segmentation; dynamic scenes; SLAM; Mask RCNN; background restoration; point cloud splicing; instance-level dense point cloud maps; semantic octree maps

同步定位与地图构建 (SLAM) 已广泛应用于服务机器人、无人驾驶、虚拟现实等领域<sup>[1-2]</sup>, 与人们日常生活密切相关。而室内场景中的动态对象会严重影响相机位姿估计的准确性, 从而降低地图创建的效果, 因此室内动态场景 SLAM 成为了一个研究热点<sup>[3]</sup>。

视觉 SLAM 在动态场景下仍然存在很大的挑战, 而且目前 SLAM 通常建立稀疏点云地图<sup>[4]</sup>, 缺少对环境地图信息的理解, 因此基于深度学习的

室内动态场景 SLAM 逐渐受到人们的关注<sup>[5-6]</sup>。

ORB-SLAM2<sup>[7]</sup> 被认为是目前最完整的 SLAM 框架之一, 但是在动态场景中 ORB-SLAM2 系统定位与建图效果并不理想, 并且仅创建稀疏点云地图, 无法用于移动机器人后续导航工作。本文以提高室内动态场景下相机位姿估计准确性以及地图创建效果为目的, 在目前相对完整的 ORB-SLAM2 框架的基础上进行优化, 将深度学习与 SLAM 相结合, 剔除分布在动态物体内的特征点, 减少动态对象对相机位姿估计的影响, 同时提高移动机器人对周围环境的理解能力, 利用背景修复<sup>[8]</sup>和点云拼接技术<sup>[9]</sup>相结合的方法建立无动态

收稿日期: 2021-03-18. 网络出版日期: 2021-06-15.

作者简介: 席志红, 女, 教授, 博士生导师.

通信作者: 席志红, E-mail: xizhihong@hrbeu.edu.cn.

物体干扰的实例级稠密点云地图<sup>[10]</sup>以及语义八叉树地图<sup>[11]</sup>,大大减少了地图存储空间。

## 1 本文 SLAM 系统框架

本文 SLAM 系统采用 ORB-SLAM2 框架来提供 SLAM 方案,并行运行实例分割、特征点提取与跟踪、局部建图、闭环检测和地图构建 5 个线程。本文 SLAM 系统框架如图 1 所示。首先由彩

色和深度 (red, green, blue and depth, RGBD) 相机捕获的彩色 (red, green and blue, RGB) 图像经过实例分割线程处理,得到每个像素实例级别标签;特征点提取与跟踪线程负责提取静态物体内的特征点,并且利用多视几何方法<sup>[8,12]</sup>进一步剔除潜在外点;最后利用筛选后的静态点估计相机位姿,地图构建线程负责建立无动态物体干扰的实例级稠密点云地图以及语义八叉树地图。

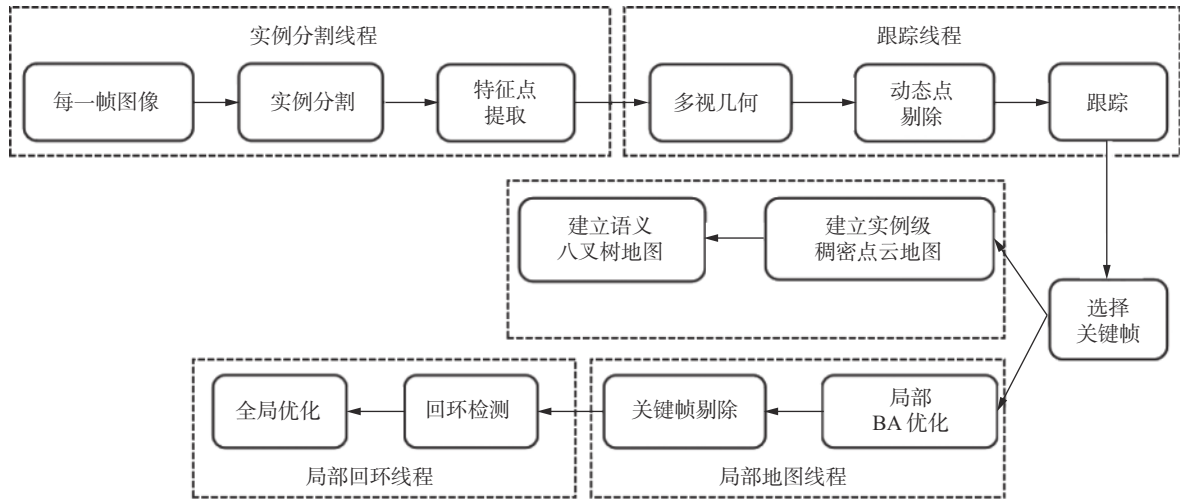


图1 本文 SLAM 系统框架

## 2 动态检测

### 2.1 实例分割

Mask RCNN<sup>[13]</sup>是一个基于深度学习的图像实例分割框架,它利用检测框物体分类、检测框坐标回归和检测框逐像素分割 3 个卷积网络分支来完成实例分割任务,改进了特征金字塔网络的感兴趣区域池化,在区域生成网络 (region proposal network, RPN) 顶部添加了并列的全卷积网络 (fully convolutional networks, FCN) 层来扩展 Faster R-CNN<sup>[14]</sup>,并且计算掩码损失。每个类对应一个掩码可以有效避免类间竞争,通过双线性插值使候选区域和卷积特征对齐,不因量化而损失信息。

Mask RCNN 实例分割网络不仅可以实现对图像中物体语义信息的标注,而且可以准确区分同类物体中的不同个体,这对于帮助移动机器人理解周围环境起着重要的作用。

本文 SLAM 系统采用 Mask RCNN 实例分割网络进行动态物体检测,在实际室内场景中,人、狗、猫被认为是主要动态对象。若实例分割后的结果中含有上述动态对象,则将动态对象内部特征点剔除,这样可以显著降低动态对象的影响。

### 2.2 多视几何

通过使用 Mask RCNN,大多数动态物体可以被分割并且不被用于跟踪和建图。但是,由于一些可移动物体不是先验动态的,因此不能被该方法检测到。本文 SLAM 系统利用多视几何方法进一步检测动态物体,判断动态物体示意如图 2 所示。对于每个输入帧,选择之前具有最大重叠的关键帧,然后计算每个关键点  $n$  从之前关键帧 (key frame, KF) 到当前帧 (current frame, CF) 的投影,获得关键点  $n$  和投影深度  $p$ ;同时获得其对应的 3D 点  $N$ 。计算  $n$  的反投影和关键点  $n'$  之间夹角  $\alpha$ ,如果  $\alpha > 30^\circ$ ,那么静态物体被考虑为动态物体,获得  $n'$  对应的深度值  $z'$ ,若  $z'$  与  $p$  的差值超过设定的阈值  $\tau$ ,则判定关键点  $n$  属于动态物体。

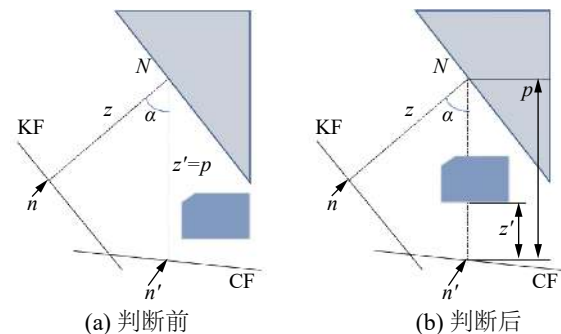


图2 多视几何判断动态物体示意

### 3 位姿计算

经过动态检测并剔除动态特征点,即可仅利用静态特征点进行相机位姿估计,从而提高相机位姿计算精度。假设筛选后的静态特征点 $P$ 的坐标为 $(x,y)$ ,其三维坐标为 $(X,Y,Z)$ ,由几何投影关系可得

$$s \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = K \left( R \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + T \right)$$

式中: $K$ 为内参矩阵, $R$ 为旋转矩阵, $T$ 为平移矩阵(各矩阵均为 $3 \times 3$ 矩阵)。

设对应点 $P$ 的坐标为 $(x',y')$ ,则重投影误差函数为

$$f(R, T) = (P' - 1/sK(RP + T))^2$$

式中 $s$ 为尺度因子。

对相邻关键帧中的所有特征点构建最小二乘问题:

$$(R, T)^* = \arg \min \frac{1}{2} \sum_{l=0}^s f(R, T)_l^2$$

式中 $(R, T)^*$ 即为所求位姿<sup>[15]</sup>。

### 4 背景修复

为了创建无动态物体干扰的地图,需要将图像中的动态对象滤除。本系统采用将之前 20 关键帧彩色图像和深度图像投影到当前帧上的方法实现背景填充,背景修复前后图像如图 3 和图 4 所示。



图3 背景修复前图像



图4 背景修复后图像

### 5 实验测试与结果分析

本文选用 TUM RGB-D 公开数据集<sup>[16]</sup>中的动态序列 f3\_w\_xyz、f3\_w\_halfsphere 和 f3\_w\_static 对本文系统和 ORB-SLAM2 进行对比测试,运行平台为配备 Intel Core i7 处理器、GeForce GTX 1050Ti 型号 GPU、8 GB 内存的台式电脑。

#### 5.1 相机位姿估计误差

本文使用绝对轨迹误差 (absolute trajectory error, ATE) 值 ( $e_{ATE}$ ) 作为评价指标,利用 evo 工具绘制相机的轨迹,并评估估计轨迹与真值的误差。ORB-SLAM2 和本文系统在 3 个公开数据集上的绝对轨迹误差测试结果分别如表 1 和表 2 所示,其中均方根误差 ( $e_{rmse}$ ) 反映估计值与真实值之间的偏差;平均误差 ( $e_{mean}$ ) 反映所有估计误差的平均水平;中值误差 ( $e_{media}$ ) 代表所有误差的中等水平;标准偏差 ( $e_{std}$ ) 反映系统轨迹估计的离散程度。

表 1 ORB-SLAM2 绝对轨迹误差结果 cm

图像序列	$e_{rmse}$	$e_{mean}$	$e_{media}$	$e_{std}$
f3_w_xyz	65.330	55.988	49.180	33.666
f3_w_halfsphere	43.345	39.688	37.231	17.428
f3_w_static	39.779	35.934	28.911	17.063

表 2 本文系统绝对轨迹误差结果 cm

图像序列	$e_{rmse}$	$e_{mean}$	$e_{media}$	$e_{std}$
f3_w_xyz	1.571 8	1.355 7	1.182 3	0.795 3
f3_w_halfsphere	2.823 0	2.428 8	2.040 4	1.438 9
f3_w_static	0.722 9	0.640 7	0.587 9	0.334 8

本文利用 evo 工具分别对 ORB-SLAM2 系统与本文系统绘制相机的轨迹,并评估估计轨迹与真值的误差,在 f3\_w\_xyz 数据集下的实验结果如图 5~8 所示。

由以上实验结果可以看出:在室内动态场景数据集 f3\_w\_xyz、f3\_w\_halfsphere 和 f3\_walking\_static 中,相对 ORB-SLAM2 系统来说,本文系统绝对轨迹均方根误差分别降低了 97.59%、93.49% 和 98.18%,本文 SLAM 系统相机位姿估计误差明显降低,其原因在于本文系统增加了对动态物体的处理,利用动态检测方法筛选后的静态特征点进行位姿估计,从而提升精度。

#### 5.2 地图创建

本文利用实例分割后的图像与相机运动轨迹创建静态语义地图,将二维信息投影到三维地图中,赋予地图物体实例信息,经过实例分割并滤除动态对象后的稠密点云地图如图 9 所示。

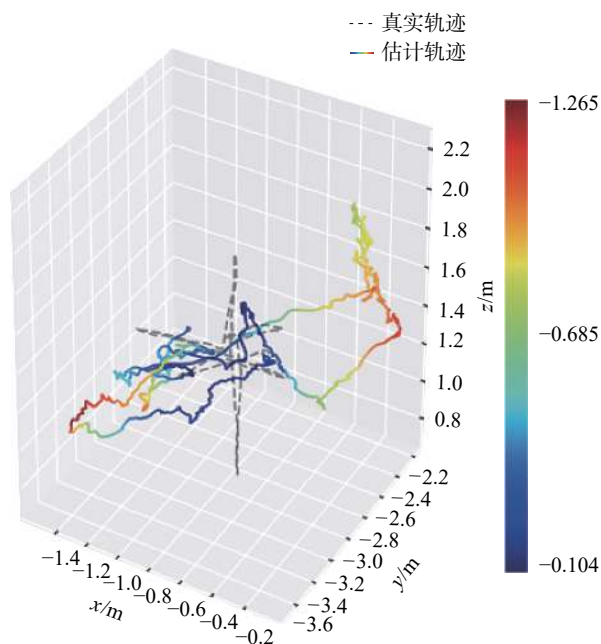


图5 ORB-SLAM2 系统相机估计轨迹与真实轨迹及误差

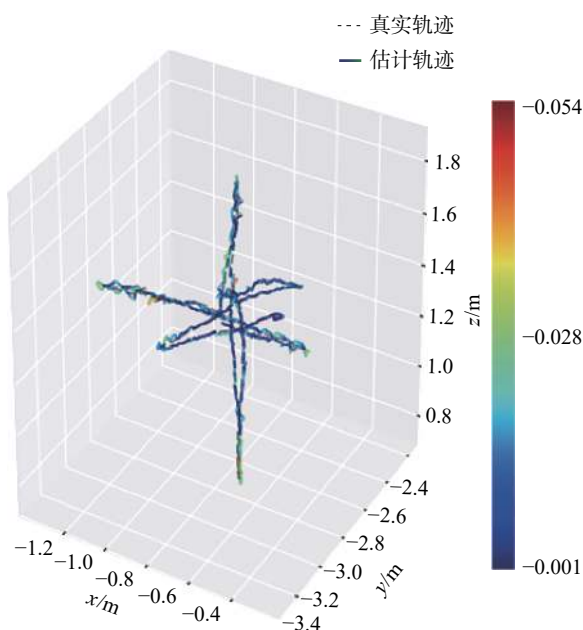


图6 本文系统相机估计轨迹与真实轨迹及误差

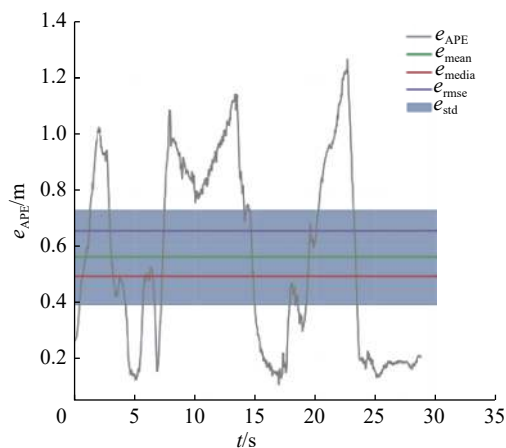


图7 ORB-SLAM2 系统相关误差曲线

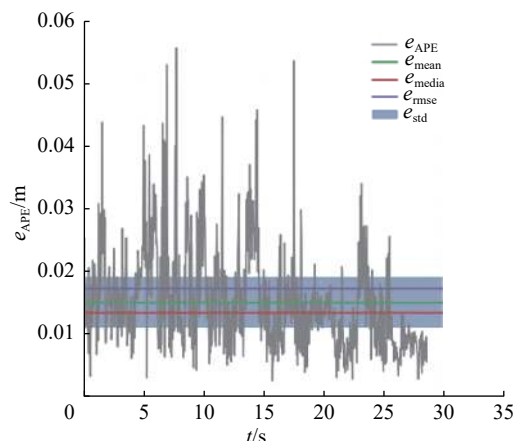


图8 本文系统相关误差曲线

从图 9 可以看出, 对背景修复和实例分割后的图像进行点云拼接, 可以将二维信息投影到三维地图中, 其中不同颜色代表不同物体实例, 然而在背景填充过程中难免存在始终被遮挡区域, 造成图像出现裂痕。

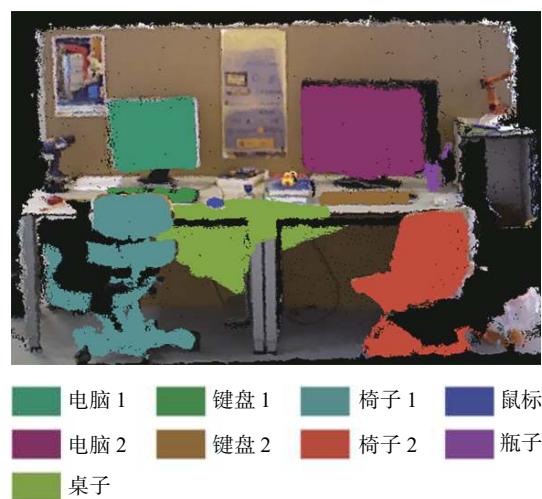


图9 实例级稠密点云地图

相对于实例级稠密点云地图, 语义八叉树地图所占的空间 (3 MB) 约是点云地图 (14.8 MB) 的 20%, 能够节省大量存储空间, 为机器人提供含有环境物体信息的导航地图。本文系统生成的语义八叉树地图如图 10 所示。

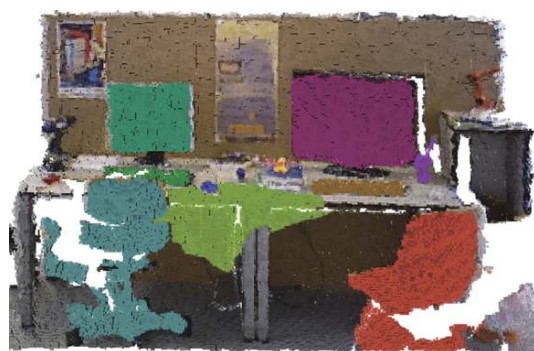


图10 语义八叉树地图

## 6 结 论

本文提出了一种基于实例分割的室内动态场景 SLAM 系统。

1) 该系统实例分割网络不仅可以应用于对环境动静物体的分割, 而且可以在地图中添加物体实例信息, 有效减小动态对象对相机位姿估计的影响, 同时提高地图构建的准确性, 弥补了 ORB-SLAM2 应用于动态场景下的缺陷。

2) 构建的实例级稠密点云地图和语义八叉树地图有助于机器人理解环境, 对后续导航工作有着重要的意义。

然而, 本文 SLAM 系统在减小误差的同时, 也增加了运行时间, 在接下来的工作中, 将进一步提升实例分割网络性能以减小相机位姿估计误差, 并且保证系统的实时性。

## 参考文献:

- [1] LEE T J, KIM C H, CHO D I D. A monocular vision sensor-based efficient SLAM method for indoor service robots[J]. *IEEE transactions on industrial electronics*, 2019, 66(1): 318–328.
- [2] 陈世浪, 吴俊君. 基于 RGB-D 相机的 SLAM 技术研究综述[J]. *计算机工程与应用*, 2019, 55(7): 30–39.
- [3] SAPUTRA M R U, MARKHAM A, TRIGONI N. Visual SLAM and structure from motion in dynamic environments: a survey[J]. *ACM computing surveys*, 2018, 51(2): 37.
- [4] 沈克贤. 基于图像序列的稀疏点云重建[D]. 重庆: 重庆大学, 2018: 49–58.
- [5] 赵洋, 刘国良, 田国会, 等. 基于深度学习的视觉 SLAM 综述[J]. *机器人*, 2017, 39(6): 889–896.
- [6] XIAO Linhui, WANG Jing, QIU Xiaosong, et al. Dynamic-SLAM: semantic monocular visual localization and mapping based on deep learning in dynamic environment[J]. *Robotics and autonomous systems*, 2019, 117: 1–16.
- [7] MUR-ARTAL R, TARDÓS J D. ORB-SLAM2: an open-source SLAM system for monocular, stereo, and RGB-D cameras[J]. *IEEE transactions on robotics*, 2017, 33(5): 1255–1262.
- [8] BESCOS B, FÁCIL J M, CIVERA J, et al. DynaSLAM: tracking, mapping, and inpainting in dynamic scenes[J]. *IEEE robotics and automation letters*, 2018, 3(4): 4076–4083.
- [9] 高翔, 张涛, 刘毅, 等. 视觉 SLAM 十四讲: 从理论到实践[M]. 2 版. 北京: 电子工业出版社, 2019: 152–153.
- [10] 吴皓, 迟金鑫, 田国会. 基于视觉 SLAM 的物体实例识别与语义地图构建[J]. *华中科技大学学报(自然科学版)*, 2019, 47(9): 48–54.
- [11] YU Chao, LIU Zuxin, LIU Xinjun, et al. DS-SLAM: a semantic visual SLAM towards dynamic environments[C]//*Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Madrid, Spain, 2018: 1168–1174.
- [12] 谢理想. 基于多视图几何的无人机稠密点云生成关键技术研究[D]. 郑州: 解放军信息工程大学, 2017: 40–49.
- [13] HE Kaiming, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]//*Proceedings of 2017 IEEE International Conference on Computer Vision*. Venice, Italy, 2017: 2980–2988.
- [14] 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.
- [15] 房立金, 刘博, 万应才. 基于深度学习的动态场景语义 SLAM[J]. *华中科技大学学报(自然科学版)*, 2020, 48(1): 121–126.
- [16] STURM J, ENGELHARD N, ENDRES F, et al. A benchmark for the evaluation of RGB-D SLAM systems[C]//*2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Vilamoura-Algarve, Portugal, 2012: 573–580.

## 本文引用格式:

席志红, 温家旭. 基于实例分割的室内动态场景 SLAM[J]. *应用科技*, 2021, 48(6): 18–22.

XI Zhihong, WEN Jiaxu. Indoor dynamic scene SLAM based on instance segmentation[J]. *Applied science and technology*, 2021, 48(6): 18–22.