

基于改进 DeepLab V3+ 的语义地图构建

李琳^{1*}, 吴怀宇^{1,2}, 张天宇²

¹武汉科技大学信息科学与工程学院, 湖北 武汉 430081;

²武汉科技大学机器人与智能系统研究院, 湖北 武汉 430081

摘要 为了提高移动机器人感知环境、执行高级任务的能力,针对传统视觉同时定位与建图(SLAM)构建的地图缺乏语义信息无法对场景内容进行理解的问题,提出一种基于物体分割的语义地图构建方法。首先,通过改进的语义分割模型 DeepLab V3+ 对二维图像进行分割,获取物体的标签;然后,根据改进的迭代最近点(ICP)点云拼接方法构建稠密地图,并且采用区域生长算法对三维点云进行分割;最后,将二维标签映射到三维稠密地图中,构建出语义地图。实验结果表明:所提改进的 DeepLab V3+ 算法与原方法相比,检测速度提高了约 4 倍;采用改进的 ICP 算法进行点云拼接时,在 TUM 数据集的 fr/360 序列上其相对轨迹误差较 ORB-SLAM 算法减小了约 16.4%;最后,与 ORB+YOLOv3、ORB+MASK-RCNN、ORB+DeepLab V3+ 方法相比,所提方法减少了语义地图的冗余信息,而且建图速度更快,并且占用储存更少。

关键词 机器视觉; 语义地图; 语义分割; 三维稠密地图; 点云拼接; 区域生长

中图分类号 TP391

文献标志码 A

DOI: 10.3788/LOP202259.1015002

Constructing Semantic Map of Mobile Robots Based on Improved DeepLab V3+

Li Lin^{1*}, Wu Huaiyu^{1,2}, Zhang Tianyu²

¹*School of Information Science and Engineering, Wuhan University of Science and Technology, Wuhan 430081, Hubei, China;*

²*Institute of Robotics and Intelligent Systems, Wuhan University of Science and Technology, Wuhan 430081, Hubei, China*

Abstract This study proposes a semantic map construction method based on object segmentation to solve the problem that maps constructed using the traditional visual simultaneous localization and mapping (SLAM) lack semantic information and cannot understand the scene content to improve the ability of mobile robots to perceive the environment and perform advanced tasks. To begin, the improved semantic segmentation model DeepLab V3+ was used to segment a two-dimensional image to obtain the object's label. Further, the dense map was constructed according to the improved iterative closest point (ICP) point cloud splicing method, and the region growth algorithm was used to segment the three-dimensional point cloud. Finally, the semantic map was constructed by mapping the two-dimensional label to the three-dimensional dense map. Experimental results show that the improved DeepLab V3+ detects objects four times faster than the original method; the improved ICP algorithm is used for point cloud splicing, and the relative trajectory error is reduced by 16.4% when compared to the ORB-SLAM algorithm in the fr/360 sequence of the TUM dataset; finally,

收稿日期: 2021-03-31; 修回日期: 2021-05-09; 录用日期: 2021-05-18

基金项目: 国家自然科学基金(62073250, 62003249)、国家重点研发计划(2017YFC0806503)、湖北省科技支撑项目(2015BAA018)

通信作者: *812184225@qq.com

compared with the ORB+YOLOv3, ORB+MASK-RCNN, ORB+DeepLab V3+ methods, the proposed method not only reduces the redundant information of semantic map but also builds faster and occupies less storage.

Key words machine vision; semantic map; semantic segmentation; three-dimensional dense map; point cloud stitching; region growth

1 引言

随着人工智能领域的发展, 移动机器人开始应用于各行各业, 并逐渐成为研究的热点。其中同时定位与建图(SLAM)能利用传感器获取的数据帮助移动机器人进行自定位, 同时构建出周围的环境地图来完成导航和避障, 但是这种传统的SLAM地图无法识别出环境中的物体, 不能执行更高层的语义任务。随着机器人逐步走进人们的日常生活, 如自动驾驶、护理机器人等, 移动机器人对环境的理解能力愈发重要, 因此带有语义信息的语义地图成为一个行之有效的解决方案。

深度学习的发展为解决上述问题提供了新的思路。McCormac等^[1]使用卷积神经网络(CNN)^[2]对RGBD图像中的每个像素进行物体级分类, 然后通过贝叶斯更新把得到的结果和SLAM生成的信息一起关联到语义地图中去。Endo等^[3]提出了一种基于CNN和LSD-SLAM^[4]的单目半稠密三维语义地图构建方法。邹斌等在ORB-SLAM系统^[5-7]的基础上, 使用YOLOv3^[8]对场景进行物体检测, 然后将结果融入到地图中。赵洋^[9]将MASK-RCNN算法引入到视觉SLAM框架中, 然后构建出基于三维体元的以物体个体为单位的语义地图。

以上研究者所提的方法虽然较好地完成语义地图的构建, 但是大多数都是采用目标检测算法获取物体矩形框的, 会包含其他物体的像素, 如果将带有其他边界的语义信息融入到地图中会导致地图包含错误的信息, 会对移动机器人后续工作产生影响。刘雨^[10]使用了DeepLab V3+语义分割算法定位出目标物体在二维图像中的位置, 然后对目标物体进行分割, 最后对分割后的二维图像进行投影得到三维语义地图。虽然这种方法可以实现像素级别的分割, 使得物体边界不包含错误的像素, 但是此模型使用较大的网络模型, 导致整体系统的运行速度变慢, 不满足SLAM系统的实时性需求。基于此, 本文提出基于改进DeepLab V3+的语义地图构建^[11], 采用改进的轻量型DeepLab V3+网络提取语义信息; 使用改进的迭代最近点(ICP)算法进行点云拼接; 最后使用基于区域增长的点云分割算法

分割三维地图并将提取的语义信息融入地图中完成语义地图构建。所提方法在保持较好精度的情况下提高了构建语义地图的速度。

2 基于改进DeepLab V3+的语义地图构建

2.1 构建框架

首先通过Kinect相机采集深度图像和RGB图像。后续处理主要分为两个线程: 第一个线程主要是对图像进行匹配, 然后结合深度数据采用改进的ICP算法进行点云拼接, 构建三维地图并利用区域生长算法对点云进行分割; 第二个子线程主要是对RGB图像进行语义分割, 考虑到传统DeepLab V3+模型较大、耗时长的问题, 采用轻量型MobileNetV2网络代替原来的骨干网络进行分割, 提高实时性。最后将语义信息映射到三维空间, 构建语义地图。所提方法框架如图1所示。

2.2 改进的DeepLab V3+语义分割方法

DeepLab V3+模型是目前语义分割领域较为主流的算法之一。该模型主要采用编码器-解码器的方式。编码部分由骨干网络ResNet-101和空间金字塔池(ASPP)模块组成, 首先通过骨干网络ResNet-101提取图像特征生成高级语义特征图和低级语义特征图, 然后ASPP对骨干网络提取的高级语义特征图进行多尺度采样, 生成多尺度的特征图, 并通过 1×1 的卷积进行通道降维。ASPP模块采用一种多尺度结构, 可以保证目标在图像中所占比例不高同时仍有好的分割效果, ASPP主要由 1×1 卷积、并行空洞率分别为6、12、18的空洞卷积、全局池化并行组成, 对特征图进行不同感受野的特征提取, 获得多尺度的特征。解码部分主要对骨干网络提取的低级特征和ASPP模块输出的多尺度特征进行4倍上采样然后进行融合, 最后再次通过4倍上采样输出语义分割的结果。ASPP模块结构如图2所示, 其中 r 为空洞率, 通过设置 r 的大小可以在不增加参数数量的情况下控制感受野。

DeepLab V3+对目标物体的分割结果不错, 但仍存在一些问题, 由于其骨干网络用的是ResNet-101, 网络分割效果虽然较好, 但模型极大, 占用内

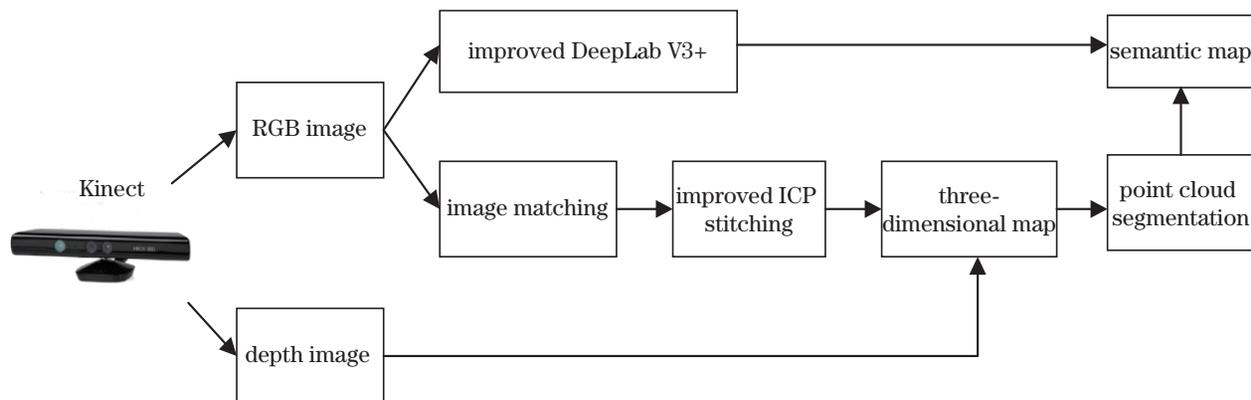


图 1 所提框架

Fig. 1 Proposed framework

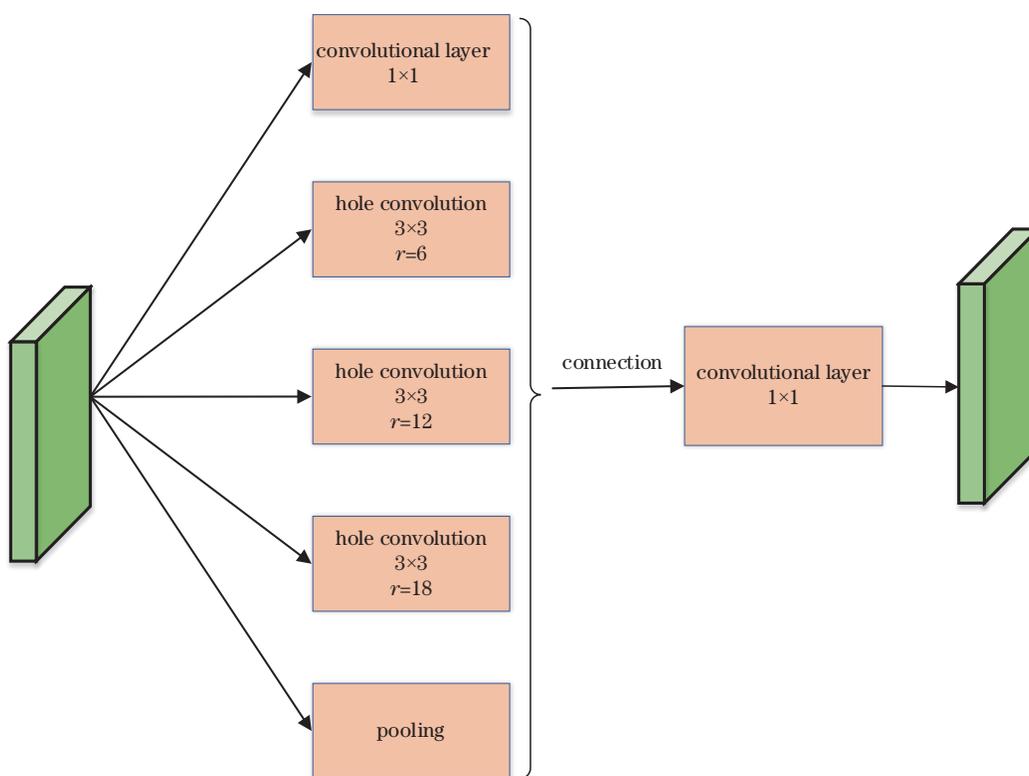


图 2 DeepLab V3+中的 ASPP 模块

Fig. 2 ASPP module in DeepLab V3+

存多,导致整体模型推理时间变长。考虑到 SLAM 系统两个重要指标即实时性和准确性,如果语义分割这一环节消耗内存太多,运行时间太长,那么整体的语义地图构建流程花费时间就太长。考虑到地图构建的实时性,提出一种改进的 DeepLab V3+ 算法,采用 MobileNetV2 替换 ResNet-101 作为骨干网络来提升模型的分割速度。

MobileNetV2 是一种轻量型的网络模型,采用深度可分离卷积形式,可以通过分解卷积层来大幅度降低模型的参数量和计算量,运行速度快,实时

性高,可以为语义地图构建奠定良好的基础。其卷积分解示意如图 3 所示。

设输入的特征图的尺寸为 $D_i \times D_i$, 通道数为 M , 卷积核的尺寸为 $D_K \times D_K$, 输出的特征图的尺寸为 $D_G \times D_G$, 通道数为 N , 则标准卷积的计算公式为

$$\delta = D_K \times D_K \times M \times N \times D_i \times D_i. \quad (1)$$

将标准卷积拆分为深度卷积和逐点卷积^[12]。则深度卷积和逐点卷积的计算量可以表示为

$$D_K \times D_K \times M \times D_i \times D_i + M \times N \times D_i \times D_i. \quad (2)$$

那么总的计算量减少了

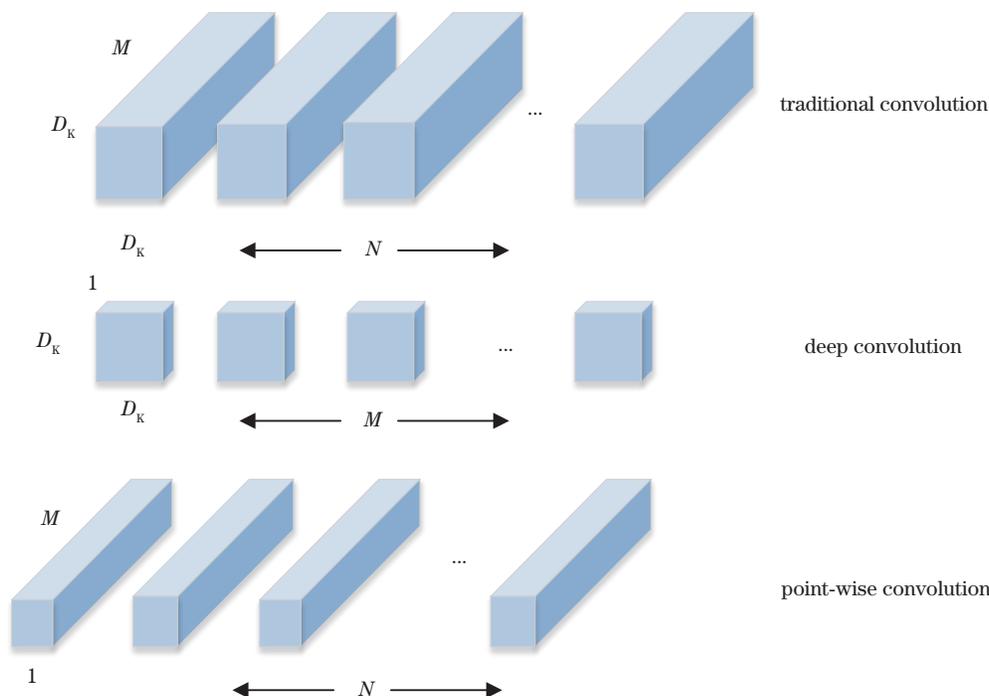


图3 标准卷积分解

Fig. 3 Standard convolution decomposition

$$\frac{D_k \times D_k \times M \times D_f \times D_f + M \times N \times D_f \times D_f}{D_k \times D_k \times M \times N \times D_f \times D_f} = \frac{1}{N} + \frac{1}{D_k^2} \quad (3)$$

从式(3)可以看出,使用深度可分离卷积,参数量可大大减少,例如网络使用 \$3 \times 3\$ 的卷积核时,与标准卷积相比,深度可分离卷积的计算量仅为标准卷积的 \$1/9\$。

2.3 三维地图构建

要想构建三维地图,首先需要知道图像之间的关系,对图像进行特征提取和匹配获取图像间的关系,然后将二维像素点映射到三维空间,最后采用点云拼接算法来构建点云地图。

2.3.1 ORB特征提取和匹配

一般来说,主流的特征点法主要包括 scale-invariant feature transform (SIFT) 算法、speeded up robust feature (SURF) 算法、oriented FAST and rotated BRIEF (ORB) 算法 3 种,综合考虑实时性和鲁棒性,主要选 ORB 算法来进行特征提取和匹配。ORB 特征点主要由 oriented FAST 关键点和 BRIEF 描述子组成,增加了一个旋转不变性特征。Oriented FAST 关键点的检测原理如图 4 所示。

首先选择一个以 \$p\$ 点为中心, \$r\$ 为半径的圆,然后判断圆上是否有连续 \$n\$ 个点的像素值大于 \$I_p + T\$ (其中 \$I_p\$ 为 \$p\$ 点的像素值, \$T\$ 为阈值),若满足要求则

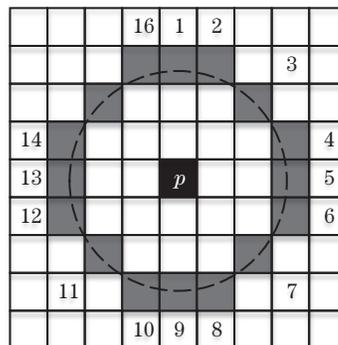


图4 角点提取示意图

Fig. 4 Schematic of corner extraction

认为是一个关键点,否则舍去。BRIEF 是一种二进制描述子,该描述子向量由 0 和 1 组成。取关键点附近两个随机像素(比如 \$p\$ 和 \$q\$),如果 \$p > q\$, 则取 1, 否则取 0。通过计算 128 对这样的 \$p, q\$, 最后得到的是 128 维的向量,也就获取了关键子。

提取到特征后,需要对两张图片进行特征匹配,寻找之间的对应关系,一般来说为了获得更高的匹配精度,采用暴力匹配方法进行匹配,其原理主要是计算所有点的欧氏距离,选择与当前点欧氏距离最小的点进行匹配。利用相机采集了两张实验室场景图片并对其进行暴力匹配,匹配结果如图 5 所示。

但是当图像特征点数目很大时,暴力匹配法的运算量会变得很大,并且误匹配率会增高,为了提



图 5 暴力匹配结果

Fig. 5 Violence matching result

升系统的鲁棒性与准确性,通常会采用 Random Sample Consensus (RANSAC) 算法^[13]对特征进行筛选后获得较为可靠的匹配特征点,优化后的特征结果如图 6 所示。



图 6 筛选后的特征匹配结果

Fig. 6 Feature matching results after screening

2.3.2 点云生成

对于某个像素点 p , 其像素坐标为 (u, v) , 深度值为 d 。该像素点在三维空间的坐标 (X_c, Y_c, Z_c) 为

$$\begin{cases} X_c = \frac{(u - c_x) \cdot Z_c}{f_x} \\ Y_c = \frac{(v - c_y) \cdot Z_c}{f_y} \\ Z_c = d \end{cases} \quad (4)$$

对每个二维像素点进行计算,就可以生成点云。

2.3.3 改进的 ICP 点云拼接方法

生成点云后,需要对点云进行拼接,常见的点云拼接方法有对极几何、perspective- n -point (PnP)、ICP。由于深度信息已知,采用 ICP 算法更为快速和精确,但传统 ICP 算法可能由于初值估计不准确,在迭代过程中容易陷入局部最优解,从而导致点云配准结果不正确,进而影响地图构建的精度。鉴于以上问题,本文采用一种基于离散选取机制的点云拼接方法^[14],首先对初始值进行筛选,然后用筛选后的点进行点云拼接。首先计算特征点集合的最大欧氏距离 d_{dist} (后面用 $d_{\text{dist, max}}$ 表示), 引入阈值参数, 设阈值为 $\alpha \times d_{\text{dist, max}}$, 选取规则为

$$p' = \begin{cases} 1, & d_{\text{dist}}(\mathbf{P}_1, \mathbf{P}_2) > \alpha \times d_{\text{dist, max}} \\ 0, & d_{\text{dist}}(\mathbf{P}_1, \mathbf{P}_2) \leq \alpha \times d_{\text{dist, max}} \end{cases}, \quad (5)$$

式中: $\mathbf{P}_1, \mathbf{P}_2$ 为随机的两个特征点; p' 为特征点被选取的概率。当 $\mathbf{P}_1, \mathbf{P}_2$ 之间的欧氏距离大于阈值时, 将两个特征点加入待配准的点集, 否则舍弃。另一方面, 由于所提方法的效果很依赖 α , 当 α 过小可能无法对特征点进行有效的筛选, 而 α 过大可能会导致特征点过少, 影响点云配准效果, 因此进行反复实验, 得到 $\alpha=0.4$ 时效果最佳。通过筛选后可以获取两组匹配点: $\mathbf{P} = \{p_1, \dots, p_n\}, \mathbf{P}' = \{p'_1, \dots, p'_n\}$ 。

通过最小二乘法使各个点满足

$$\forall i, p_i = \mathbf{R}q_i + \mathbf{T}, \quad (6)$$

即可获取相机的位姿, 同时完成点云的拼接。其中, \mathbf{R} 为旋转矩阵, \mathbf{T} 为平移向量。

2.4 基于区域生长的点云分割

前面已经对 Kinect 相机获取的二维图像进行了分割, 获取了环境物体的类别和位置信息, 但是二维图像只是反映了物体在二维空间的特征, 难以还原出环境原始的三维信息, 如果直接将二维语义信息映射到三维空间中, 可能标注出的物体位置存在偏差。为了避免这种问题, 首先使用三维点云分割算法, 分割出三维空间中属于同一物体的点云, 然后与二维分割出的类别和位置信息进行融合, 获取精度高的语义地图。

基于几何点云分割的算法有很多, 如 Voxel Cloud Connectivity Segmentation (VCCS)、Locally Convex Connected Patches (LCCP) 等。其中 VCCS 算法对三维点云进行超体素分割时, 将点云分割成若干点云块, 容易出现过分割的现象; LCCP 算法在 VCCS 算法分割的基础上, 首先计算不同点云块之间的凹凸性, 然后利用物体凹凸性对过分割的点云进行二次聚类, 但是有时候会产生局部最优的情况, 反而导致分割结果不好, 不具备普适性; 相比较而言, 基于区域生长^[15]的点云分割方法操作起来简单, 分割效果也比较好, 具体步骤如下。

1) 种子选择

对三维点云进行分割前, 首先要确定种子节点, 然后才能根据种子和设置好的生长准则判断一点是否与种子属于同一曲面, 若是, 则将该点确定为新的种子继续生长, 以此类推, 直到所有点云数据被划分完成, 则停止生长。对于种子的选择, 主要根据点云中的一些明显特征如法向量、曲率等进行。

法向量和曲率可以用来描述曲面的变换, 法向量和曲率估算主要是使用主成分分析 (PCA) 方法, 此方法具有易于实现且计算量小的特点。首先将

求解点云的法向量转换为求解点云切平面的法向量,其中切平面可由最小二乘法拟合局部平面的方法确定。设有一平面 S , \mathbf{n}_s 为该平面的法向量,平面中有一点 P , 点 P 的法向量 $\mathbf{n}_s = \mathbf{n}_p$, 则采用 PCA 方法估算法向量和曲率的具体流程如下。

a) 对于点云中任意一点 P_i , 点 P_{ij} 为 P_i 的 k 邻域点, 其中 $P_i = (x_i, y_i, z_i)^T, 1 \leq i \leq n, P_{ij} = (x_{ij}, y_{ij}, z_{ij})^T, 1 \leq j \leq k$, 计算 k 邻域的质心 \mathbf{m}_i

$$\mathbf{m}_i = \frac{1}{k} \sum_{j=1}^k P_{ij} \quad (7)$$

b) 对点 P_i 构建协方差矩阵 N_i

$$N_i = \frac{1}{k} \sum_{j=1}^k (P_{ij} - \mathbf{m}_i)(P_{ij} - \mathbf{m}_i)^T \quad (8)$$

c) 对 N_i 进行特征分解, 可得特征值 r_1, r_2, r_3 对应的特征向量 $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$, 其中 $r_1 < r_2 < r_3$ 。由于 N_i 是半正定矩阵, 因此 $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ 相互正交。此时 P_i 处的最优切平面可由 \mathbf{v}_2 和 \mathbf{v}_3 得出, 由于 \mathbf{v}_1 与切平面正交, 因此可用 \mathbf{v}_1 近似表示点 P_i 的法向量 \mathbf{n}_{P_i} 。 N_i 的特征值 r_1, r_2, r_3 反映了 $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ 方向上的变化程度, 因此点 P_i 处的曲率 c_i 可表示为

$$c_i = \frac{r_1}{r_1 + r_2 + r_3} \quad (9)$$

2) 区域生长

根据步骤 1) 获得的法向量和曲率判断一区域中的点与种子是否属于同一曲面, 若是, 将该点作为新的种子继续生长, 以此类推, 直到所有点检测完毕。将所有相同属性的点云数据划入同一区域, 生长停止, 完成三维点云的分割。

3 实验结果与分析

为了验证所提算法的可行性和准确性, 使用德国慕尼黑工业大学公开的 TUM 标准数据集、ICL-NUIM 数据集和自采数据集进行测试。笔记本电脑的配置: 处理器为 Intel Core i5-7200U 2.5 GHz, 电脑操作系统为 Ubuntu16.04。

3.1 改进的 DeepLab V3+ 语义分割实验

采用像素精度 (PA) 和平均交并比 (mIOU) 作为语义分割预测结果的评价指标。其中, PA 表示被分类正确的像素占总体像素数目的比例, mIOU 是真实标签与预测标签的交集比上它们的并集, 然后计算每个类的交并比, 再取平均。

$$P_A = \frac{\sum_{i_c=0}^c N_{i_c i_c}}{\sum_{i_c=0}^c \sum_{j_c=0}^c N_{i_c j_c}} \quad (10)$$

式中: $N_{i_c j_c}$ 表示本属于类别 i_c 但被预测为类 j_c 的像素数目; $N_{i_c i_c}$ 表示真正的像素数量。

$$C_{mIOU} = \frac{1}{C+1} \sum_{i_c=0}^c \frac{N_{i_c i_c}}{\sum_{i_c=0}^c N_{i_c j_c} + \sum_{i_c=0}^c N_{j_c i_c} - N_{i_c i_c}} \quad (11)$$

式中: $N_{j_c i_c}$ 表示假负的像素数量。

最终超参数选择如表 1 所示。

表 1 网络参数选择

Training parameter	Value
Batch size	4
Learning rate	0.0001
Power	0.9
Epoch	100

分别使用 DeepLab V3+ 和改进的 DeepLab V3+ 算法对 PASCAL VOC2012 数据集进行训练。迭代次数为 100, 算法的总体损失函数曲线如图 7 所示, 其中横坐标为算法迭代次数, 纵坐标为训练总体损失值。

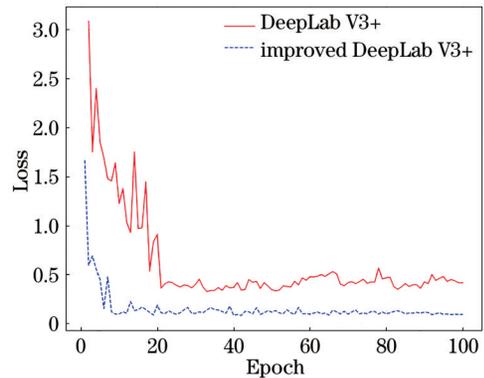


图 7 算法训练总体损失曲线

Fig. 7 Algorithm training overall loss curve

从图 7 中可以看出, 改进前后的 DeepLab V3+ 最终都达到了收敛, 但 DeepLab V3+ 最终收敛到 0.415, 而改进后的 DeepLab V3+ 算法收敛到 0.092, 并且改进后的模型的损失曲线波动幅度更小, 更稳定一些。

表 2 和图 8 是算法对 PASCAL VOC2012 数据集测试的结果, 表 2 是两种算法的评估指标对比, 图 8

表 2 算法评估指标对比

Algorithm	mIOU / %	PA / %	Number of parameters	Time / ms
DeepLab V3+	78.27	89.43	9.03×10^7	318
Improved DeepLab V3+	76.95	87.18	3.53×10^6	75

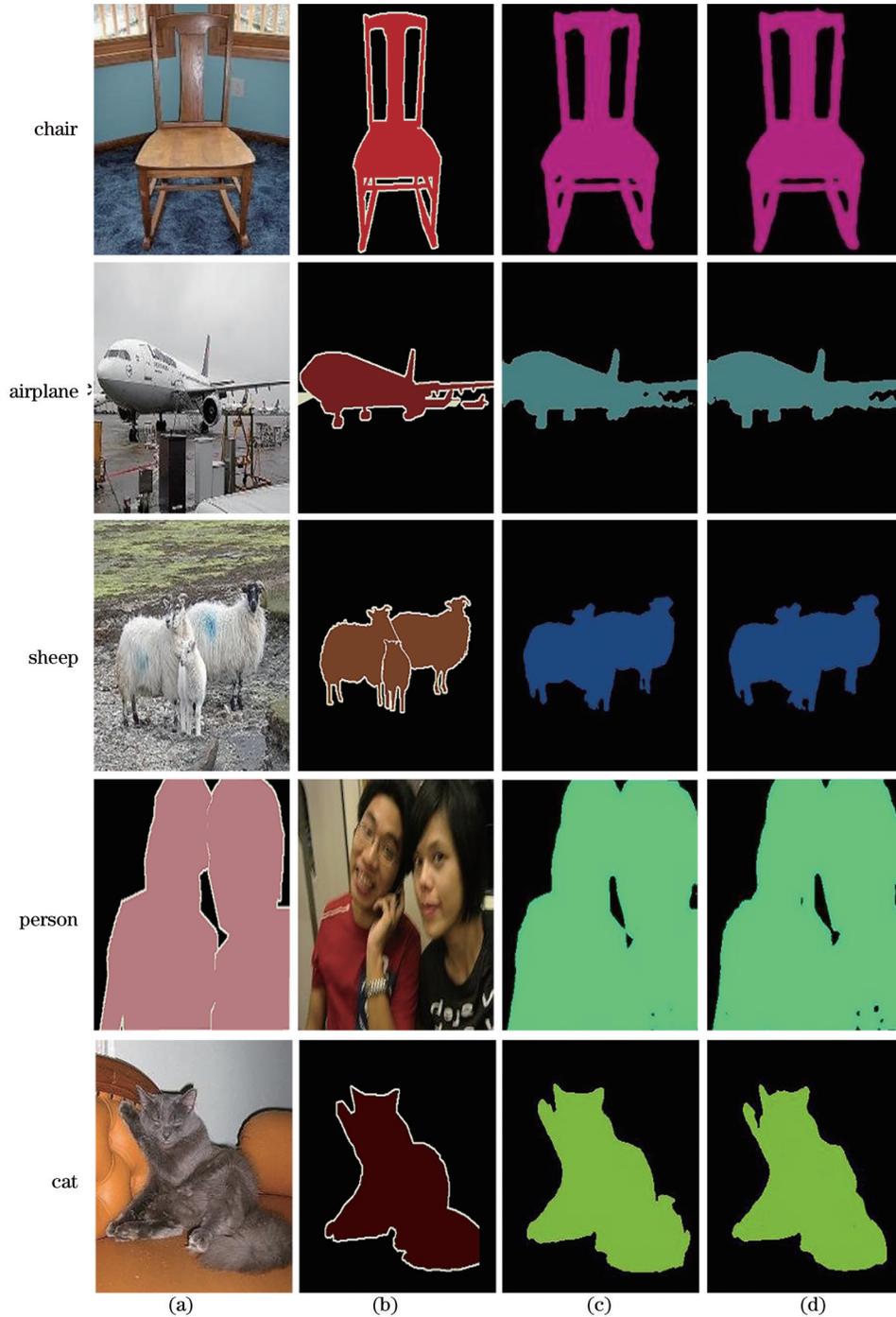


图 8 DeepLab V3+与改进DeepLab V3+的分割结果对比。(a)输入图片;(b)标签;(c) DeepLab V3+的分割结果;(d)改进 DeepLab V3+的分割结果

Fig. 8 Comparison of segmentation results between DeepLab V3+ and improved DeepLab V3+. (a) Input images; (b) ground truth; (c) segmentation results of DeepLab V3+; (d) segmentation results of improved DeepLab V3+

是数据集中的部分图片分割效果。可以看出:虽然改进后的 DeepLab V3+以牺牲 1.32 个百分点的 mIOU 和 2.25 个百分点的 PA 为代价,但从图 8 可以看出微小的分割精度降低对实际分割结果没有多大影响,大致轮廓分割效果相同,且改进后的 DeepLab V3+的参数数量是改进前模型的 0.039,占用内存更

小,处理一张图片的时间仅为 75 ms,实时性更高,完全满足移动机器人构建语义地图的要求。

3.2 语义地图构建

3.2.1 基于改进的 ICP 算法对比

为了体现改进算法的位姿估计效果,选用 TUM 标准数据中 fr1/floor、fr1/xyz、fr2/360 和 fr2/

desk 四个序列进行测试,对相对轨迹误差的均方根误差(RMSE)进行比较,从表 3 中可以看出,在 4 种序列中,所提算法的 RMSE 最小,表明位姿估算精

表 3 相对轨迹的均方根误差
Table 3 RMSE of relative trajectory

Dataset	Number of frames	RMSE of relative trajectory		
		RGBD-SLAM	ORB-SLAM	Proposed algorithm
fr1/floor	1242	0.0044	0.0041	0.0040
fr1/xyz	798	0.0058	0.0059	0.0057
fr2/360	1431	0.035	0.033	0.0276
fr2/desk	2965	0.0037	0.0036	0.0035

表 4 运动估算速度与迭代次数

Table 4 Speed of motion estimation and number of iterations

Dataset	ICP algorithm		Improved ICP algorithm	
	Average time /s	Average number of iterations	Average time /s	Average number of iterations
fr1/floor	0.042	9	0.016	3
fr2/desk	0.033	7	0.012	2

3.2.2 点云地图与区域生长分割

使用 ICL-NUIM 进行地图构建实验,该数据集是一个开源的 RGB-D 数据集,由帝国理工学院制作,该数据集包含两个场景,本文选用 room 场景下

度最高。

另一方面,为了进一步测试改进后的 ICP 算法的速度和迭代次数等性能,从 fr1/floor 序列和 fr2/desk 序列中分别选取连续的 100 帧图片进行位姿估算,以平均每次估算位姿的时间和平均迭代次数作为参数指标。从表 4 中可以看出,改进后的 ICP 算法的平均迭代次数明显降低且平均耗时也大幅降低。这是由于改进 ICP 算法对点云进行了初始的筛选,减少了位姿估算的计算量,同时提供了较为准确的 ICP 计算初值,使得迭代次数大大降低。实验结果表明,改进后的 ICP 算法可以有效提高算法的实时性和准确性,能为三维地图构建提供较好的初始条件。

的数据集。对 ICL-NUIM 数据集的 room 场景构建出三维点云,如图 9 所示。采用区域生长的点云分割算法对三维点云进行分割,分割出点云图中的不同物体,结果如图 10 所示。

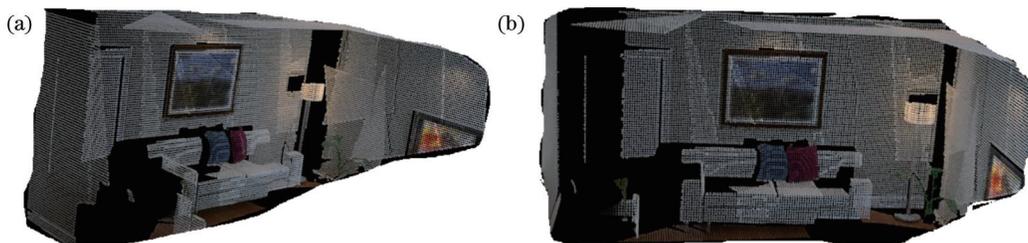


图 9 三维点云地图。(a) 视角 1;(b) 视角 2

Fig. 9 3D point cloud maps. (a) Perspective 1; (b) perspective 2

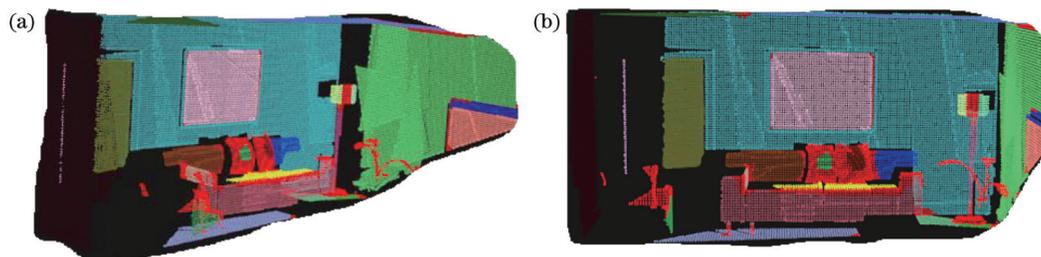


图 10 分割后的三维点云。(a) 视角 1;(b) 视角 2

Fig. 10 3D point cloud after segmentation. (a) Perspective 1; (b) perspective 2

3.2.3 数据集语义地图构建

使用 DeepLab V3+ 对 room 场景的二维图像进行分割,获取物体在图像中的类别和位置,然后将获取的语义信息映射到分割后的三维点云中,使点云

中每个物体所在的区域匹配到指定的语义信息,实现语义地图的构建,结果如图 11 所示。图 11 中红色表示画,黄色表示沙发,蓝色表示抱枕,青色表示门,粉色表示植物,其他空间结构用灰色表示,可以看出所

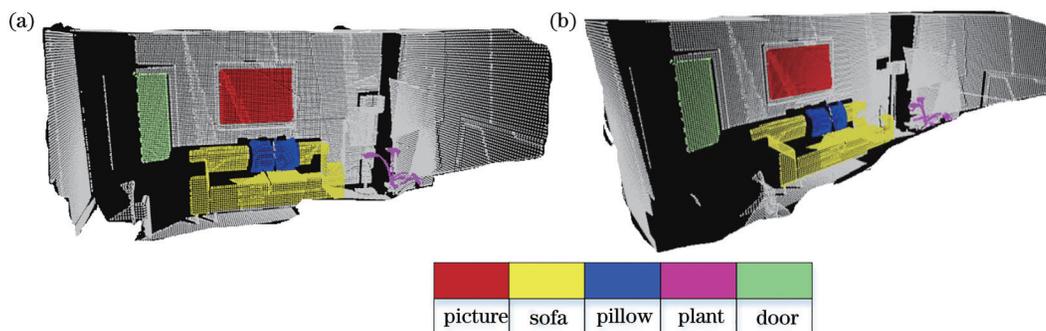


图 11 语义地图。(a) 视角 1;(b) 视角 2

Fig. 11 Semantic maps. (a) Perspective 1; (b) perspective 2

提语义地图构建方法在标准数据集中的表现较好,地图中的物体基本被分割出来,且边界很明显。

为了验证所提方法的有效性,对 ICL-NUIM 数据集构建的语义地图与 ORB+YOLOv3、ORB+

MASK-RCNN以及 ORB+DeepLab V3+方法进行对比,结果如表 5 所示,可以看出所提方法不仅减少了构建地图的冗余信息,而且建图速度更快,并且减少了储存消耗。

表 5 不同方法的建图性能对比

Table 5 Comparison of map construction performance of different methods

Method	Number of point clouds	Total map construction time /ms	Map size /Mbit
ORB + YOLOv3	4132273	508	145.7
ORB+MASK-RCNN	3740152	461	128.2
ORB+DeepLab V3+	3309853	439	110.5
Proposed method	1176592	231	71.2

3.3 实际场景语义地图构建

利用训练好的 DeepLab V3+ 模型对实验场景

景 I 和场景 II 进行分割,分割效果如图 12 所示,分割精度和时间如表 6 所示。从图 12 中可以看出,对

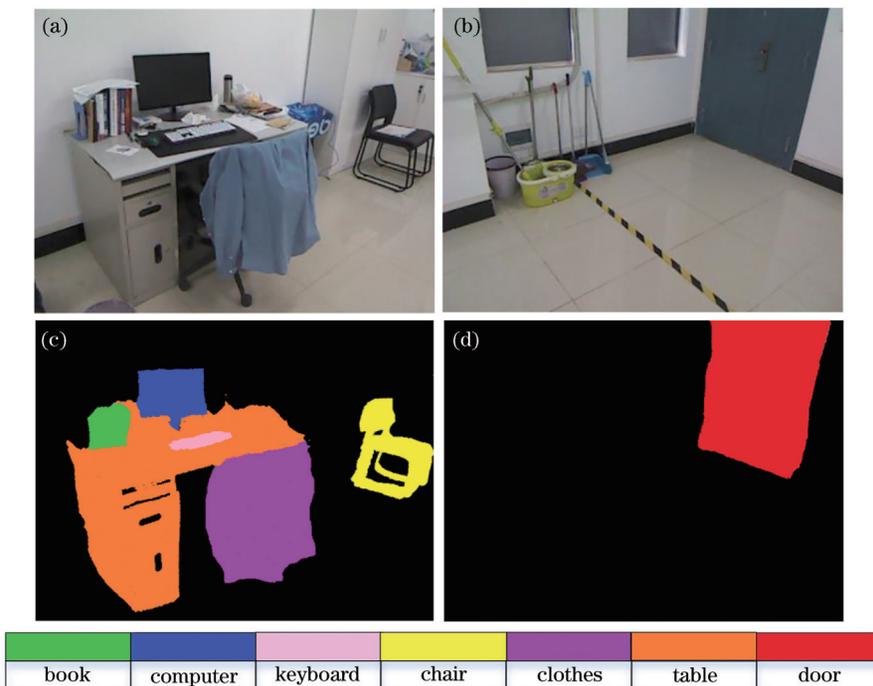


图 12 改进 DeepLab V3+ 的语义分割结果。(a) 场景 I ;(b) 场景 II ;(c) 场景 I 语义分割结果;(d) 场景 II 语义分割结果

Fig. 12 Semantic segmentation results of improved DeepLab V3+. (a) Scene I ; (b) scene II ; (c) semantic segmentation result under scene I; (d) semantic segmentation result under scene II

表 6 不同场景下的分割性能

Table 6 Segmentation performance under different scenes

Scene	mIOU / %	Processing time / ms
I	76.59	86
II	77.92	53

于实验室场景中的物体,所提方法将绝大多数都分割出来了,而且分割效果不错。从表 6 可以看出,所提方法分割准确度较好,实时性高,能够满足室内场景下移动机器人对环境的感知要求。经过测试和验证,所提语义分割方法的准确率和效率满足实际要求。

对实验室场景 I 和场景 II 构建出三维地图,如图 13(a)所示,然后使用区域生长算法对点云进行分割,并将二维语义信息映射到分割后的点云地图中得到语义地图。图 13(b)~(d)是在不同视角下观看的结果,可以看到所提方法可以有效地分割出对室内实验室场景建立的语义地图中的物体,门、桌子、书、衣服、电脑、键盘基本都分割了出来,但是由于椅子在原始地图中重建的效果较差,所以在语义地图中分割效果也不好,但总体来说,二维图像分割的物体在地图中都被标识出来了,且语义一致,验证了所提方法的可行性和准确性。

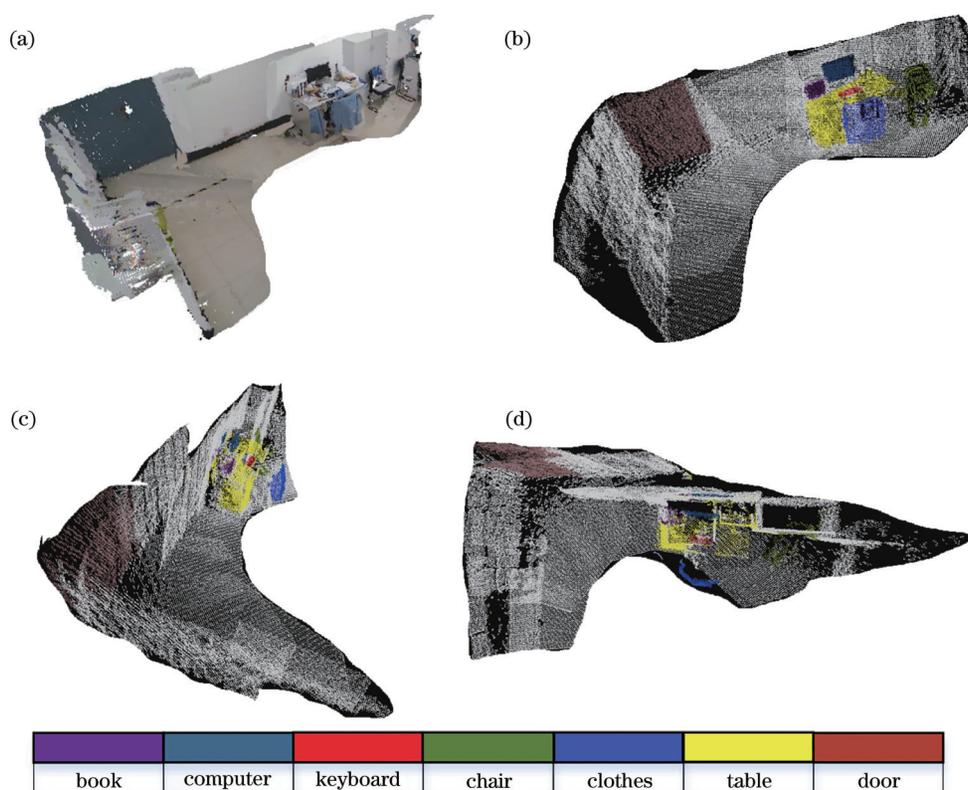


图 13 不同视角下的语义地图。(a) 三维地图;(b) 视角 1 语义地图;(c) 视角 2 语义地图;(d) 视角 3 语义地图

Fig. 13 Semantic maps from different perspectives. (a) 3D map; (b) semantic map from perspective 1; (c) semantic map from perspective 2; (d) semantic map from perspective 3

4 结 论

针对 DeepLab V3+ 网络模型较大影响地图构建的实时性问题,提出一种基于改进 DeepLab V3+ 的语义地图构建方法,采用 MobileNetV2 替换 ResNet-101 作为骨干网络的方法来降低时间的消耗。并且针对地图构建过程中点云拼接算法初值估计不准确,导致点云配准结果不正确等问题,采用一种基于离散选取机制的点云拼接方法,减少了迭代次数且地图构建的精度得到提高。最后将分

割标签映射到三维空间与构建的地图进行融合。实验表明,所提改进 DeepLab V3+ 的分割检测速度较原方法提升了约 4 倍,地图构建的精度较其他方法也获得一定的提升。经实验数据分析,所提方法能够较准确且高效地构建环境语义地图,具有良好的实际应用价值。

参 考 文 献

[1] McCormac J, Clark R, Bloesch M, et al. Fusion++ : volumetric object-level SLAM[C]//

- 2018 International Conference on 3D Vision (3DV), September 5-8, 2018, Verona, Italy. New York: IEEE Press, 2018: 32-41.
- [2] He K M, Gkioxari G, Dollár P, et al. Mask R-CNN [C]//2017 IEEE International Conference on Computer Vision, October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 2980-2988.
- [3] Endo Y, Sato K, Yamashita A, et al. Indoor positioning and obstacle detection for visually impaired navigation system based on LSD-SLAM [C]//2017 International Conference on Biometrics and Kansei Engineering (ICBAKE), September 15-17, 2017, Kyoto, Japan. New York: IEEE Press, 2017: 158-162.
- [4] Engel J, Schöps T, Cremers D. LSD-SLAM: large-scale direct monocular SLAM[M]//Fleet D, Pajdla T, Schiele B, et al. Computer vision-ECCV 2014. Lecture notes in computer science. Cham: Springer, 2014, 8690: 834-849.
- [5] 胡美玉, 张云洲, 秦操, 等. 基于深度卷积神经网络的语义地图构建[J]. 机器人, 2019, 41(4): 452-463.
Hu M Y, Zhang Y Z, Qin C, et al. Semantic map construction based on deep convolutional neural network[J]. Robot, 2019, 41(4): 452-463.
- [6] 邹斌, 林思阳, 尹智帅. 基于 YOLOv3 和视觉 SLAM 的语义地图构建[J]. 激光与光电子学进展, 2020, 57(20): 201012.
Zou B, Lin S Y, Yin Z S. Semantic mapping based on YOLOv3 and visual SLAM[J]. Laser & Optoelectronics Progress, 2020, 57(20): 201012.
- [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] 孙佳, 郭大波, 杨甜甜, 等. 基于改进的 YOLOv3 网络的实时目标检测[J]. 激光与光电子学进展, 2020, 57(22): 221505.
Sun J, Guo D B, Yang T T, et al. Real-time object detection based on improved YOLOv3 network[J]. Laser & Optoelectronics Progress, 2020, 57(22): 221505.
- [9] 赵洋. 基于深度学习的语义同步定位与地图构建[D]. 济南: 山东大学, 2018: 2-10.
Zhao Y. Semantic synchronously localization and mapping based on deep learning[D]. Jinan: Shandong University, 2018: 2-10.
- [10] 刘雨. 服务机器人室内三维环境热场建模及感知技术研究[D]. 哈尔滨: 哈尔滨工业大学, 2019: 43-54.
Liu Y. Research on indoor 3D thermal modeling and perception technology in service robots[D]. Harbin: Harbin Institute of Technology, 2019: 43-54.
- [11] Chen L C, Zhu Y K, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11211: 833-851.
- [12] Sandler M, Howard A, Zhu M L, et al. MobileNetV2: inverted residuals and linear bottlenecks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 4510-4520.
- [13] 王鹏, 朱睿哲, 孙长库. 基于改进的 RANSAC 的场景分类点云粗配准算法[J]. 激光与光电子学进展, 2020, 57(4): 041510.
Wang P, Zhu R Z, Sun C K. Point cloud coarse registration algorithm with scene classification based on improved RANSAC[J]. Laser & Optoelectronics Progress, 2020, 57(4): 041510.
- [14] 张彦铎, 袁博, 李迅. 基于改进 ICP 算法的室内环境三维地图创建研究[J]. 华中师范大学学报(自然科学版), 2017, 51(2): 264-272.
Zhang Y D, Yuan B, Li X. Reconstructing 3D map in indoor environment based on an improved ICP algorithm[J]. Journal of Central China Normal University (Natural Sciences), 2017, 51(2): 264-272.
- [15] 林云森. 面向 3D 无序抓取的目标定位与手眼标定技术研究[D]. 烟台: 烟台大学, 2020: 25-28.
Lin Y S. Research on target location and hand-eye calibration for 3D Bin-picking[D]. Yantai: Yantai University, 2020: 25-28.