激光与光电子学进展 先进成像

动态场景下基于加权静态的视觉SLAM算法

李勇[1,2]，吴海波[1,2*]，李万[1,2]，李东泽[1,2]

[1]昆明理工大学机电工程学院，云南 昆明 650500；

[2]云南省先进装备智能制造技术重点实验室，云南 昆明 650500

摘要 针对传统视觉同步定位和映射(SLAM)系统在动态环境中鲁棒性和定位精度低等问题，基于ORB-SLAM2算法框架，提出一种在室内动态环境中运行稳健的视觉SLAM算法。首先，语义分割线程采用改进的轻量化语义分割网络YOLOv5获得动态对象的语义掩码，并通过语义掩码选择ORB特征点，同时，几何线程通过加权几何约束的方法检测动态对象的运动状态信息。然后，提出一种给语义静态特征点赋予权值，并对相机的位姿和特征点的权值进行局部光束平差法(BA)联合优化的算法，有效地减少动态特征点的影响。最后，在TUM数据集和真实的室内动态场景中进行实验，结果表明，与改进之前的ORB-SLAM2算法相比，所提算法有效地提高了系统在高动态数据集中的定位精度，并且绝对轨迹误差和相对轨迹误差的均方根误差(RMSE)分别提升了96.10%和92.06%以上。

关键词 视觉SLAM；动态环境；加权几何约束；语义掩码；BA联合优化

中图分类号 TP242.6 **文献标志码** A **DOI**：10.3788/LOP231254

Visual SLAM Algorithm Based on Weighted Static in Dynamic Environment

Li Yong[1,2], Wu Haibo[1,2*], Li Wan[1,2], Li Dongze[1,2]

[1]*Faculty of Mechanical and Electrical Engineering, Kunming University of Science and Technology, Kunming* 650500, *Yunnan, China*;

[2]*Key Laboratory of Intelligent Manufacturing Technology for Advanced Equipment in Yunnan Province, Kunming* 650500, *Yunnan, China*

Abstract To address the low robustness and positioning accuracy of the traditional visual simultaneous localization and mapping (SLAM) system in a dynamic environment, this study proposed a robust visual SLAM algorithm in an indoor dynamic environment based on the ORB-SLAM2 algorithm framework. First, a semantic segmentation thread uses the improved lightweight semantic segmentation network YOLOv5 to obtain the semantic mask of the dynamic object and selects the ORB feature points through the semantic mask. Simultaneously, the geometric thread detects the motion-state information of the dynamic objects using weighted geometric constraints. Then, an algorithm is proposed to assign weights to semantic static feature points and local bundle adjustment (BA) joint optimization is performed on camera pose and feature point weights, effectively reducing the influence of the dynamic feature points. Finally, experiments are conducted on a TUM dataset and a genuine indoor dynamic environment. Compared with the ORB-SLAM2 algorithm before improvement, the proposed algorithm effectively improves the positioning accuracy of the system on highly dynamic datasets, showing improvements of root mean square error (RMSE) of the absolute and relative trajectory errors by more than 96.10% and 92.06%, respectively.

Key words visual simultaneous localization and mapping (SLAM); dynamic environment; weighted geometric constraint; semantic mask; jointly optimized by BA

收稿日期：2023-05-08；**修回日期**：2023-06-05；**录用日期**：2023-07-24；**网络首发日期**：2023-08-15

基金项目：云南省人才培养基金(KKSY201701001)、国家自然科学基金(52065035)

通信作者：*whb_kust@kust.edu.cn

1　引　　言

同步定位和映射(SLAM)是应用某种传感器来估计系统当前位姿的技术。近年来,视觉SLAM系统得到了广泛的研究,一些经典的SLAM有ORB-SLAM2[1]、ORB-SLAM3[2]、LSD-SLAM[3]等。然而,这些算法进行位姿状态估计时都假设在静态条件下运行的,当机器人在具有动态对象(如人、自行车等)的环境下移动时,动态对象的出现会改变图像的信息,导致不匹配的数据关联,使SLAM系统定位精度产生较大误差。目前,大多数传统SLAM系统采用随机采样一致性(RANSAC)[4]剔除动态点,但是动态对象在图像区域占据较大面积时,RANSAC算法易出现误匹配现象。近些年,国内外有较多学者对静/动态环境下的视觉SLAM进行研究。Sun等[5]利用自我补偿图像差异的方法来区分动态对象的运动,然后使用粒子滤波器和最大后验估计来精确优化运动分割,从而剔除运动对象对SLAM系统的干扰,但是该方法要求动态背景不能占比太大。

随着深度学习的发展,越来越多的研究将深度学习引入到SLAM领域中,以提高SLAM系统的鲁棒性。邹斌等[6]利用YOLOv3获取的语义信息结合ORB-SLAM2算法进行语义地图构建。Dyna-SLAM是一种动态鲁棒的SLAM算法,该算法利用Mask R-CNN的语义分割结果和多视图几何来检测运动对象[7],提高了位姿估计的精度,但Mask R-CNN运行速度缓慢,导致SLAM系统的实时性能差。王梦瑶等[8]通过语义分割与分配的等级传递获取移动特征点的运动状态。Li等[9]为基于深度不连续点获得的深度边缘点分配了静态权值,提出了一种针对关键帧中深度边缘点的静态加权方法,以减少动态信息造成的影响。DP-SLAM算法结合了几何约束和语义分割,通过利用动态概率和光流分割等方法来处理动态环境下的SLAM问题,有效地减少了动态物体对SLAM系统的干扰,从而提高系统的鲁棒性和精度[10]。

针对室内动态场景,本文提出了一种多线程并行的视觉SLAM系统,在ORB-SLAM2算法框架中引入了语义分割线程和几何线程。在语义分割线程中,采用改进的轻量化语义分割网络YOLOv5模型,以提高算法的运行效率,更准确地分割对象边界。为了减少网络的参数量,提高网络的运行效率和训练速度,用高效的特征提取模块GhostV2C3来替换YOLOv5模型中的C3模块,以减少冗余的卷积计算。在几何线程中,为了突破极约束判断动态特征点的局限性,采用加权几何约束的方法来检测对象的运动状态。语义分割线程与几何线程相结合能够分割出动态物体轮廓,并能根据目标检测结果为语义掩码选择的语义静态特征点赋予不同的权值,然后进行局部光束平差法(BA)联合优化。

2　系统框架

所提系统主要是基于ORB-SLAM2算法[1]框架的,如图1所示。在改进的框架中增加了语义分割线程和几何线程来检测动态目标,并根据检测结果给语义静态特征点赋予权重,然后对语义静态特征点权值和相机位姿估计进行BA联合优化。语义分割线程中,将YOLOv5语义分割网络中的C3模块替换为改进的特征提取模块GhostV2C3,有助于保持网络分割精度的同时最小化网络的复杂性,并根据语义掩码来提取oriented FAST and rotated BRIEF (ORB)特征点,获得语义静态特征点。几何线程通过加权几何约束的方法区分图像中的动态特征点和静态特征点。最后,将跟踪线程优化后的位姿和关键帧传递到后续的线程,进行位姿状态估计。

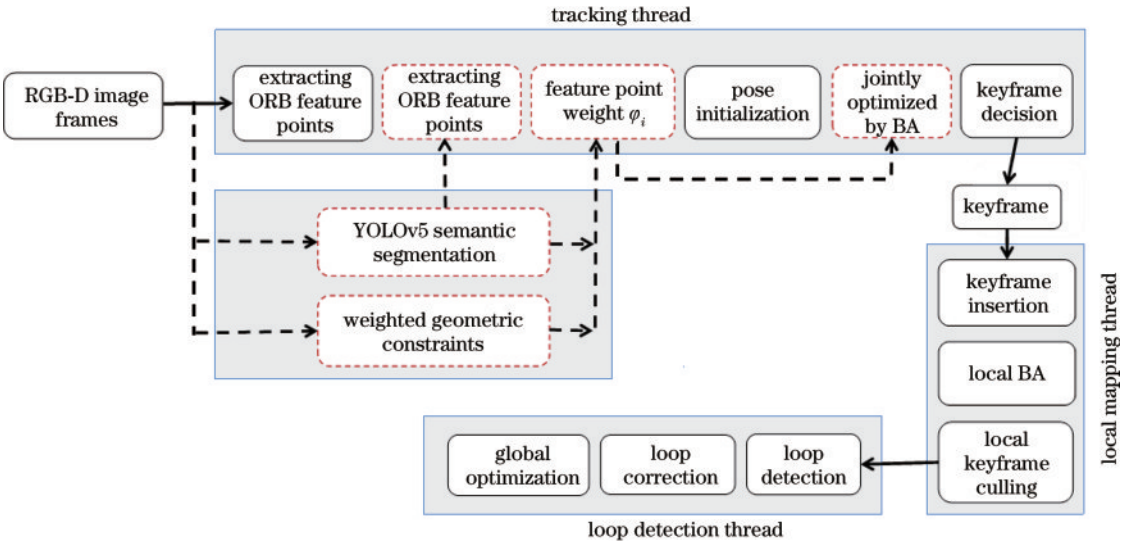

图1　系统框架图

Fig. 1　System framework diagram

3　动态目标检测

3.1　轻量化 YOLOv5 语义分割

传统的语义 SLAM 系统很难兼顾语义分割网络的高精度和实时性。Mask R-CNN[11]作为一种两阶段实例分割方法，会导致网络对图像的处理时间延迟较高，难以满足实时性需求。一些卷积神经网络，如全卷积神经网络(FCN)[12]和编码器-解码器结构的 SegNet 算法[13]，已经被应用到 SLAM 系统中，然而它们缺乏上下文信息，并且仅使用单个尺度的特征图进行分割，缺乏多尺度的特征融合，难以准确地分割不同尺度的目标。

为了提高动态 SLAM 的运行效率，需要选择一种轻量级的语义分割网络。YOLOv5-seg 系列存在 n、s、m、l 和 x 5 个不同规模的模型，在这些模型中，YOLOv5s-seg 在精度和复杂度方面能够取得较好的平衡。因此本文采用 YOLOv5s-seg 进行改进。YOLOv5s-seg 的分割原理与 YOLACT[14]相似，利用两个并行任务来执行分割，在 YOLOv5s 检测网络的基础上增加了原型掩码输出分支，所以图像数据的预处理部分与 YOLOv5s 相同。为了实现精度和速度的平衡，本文将 YOLOv5s-seg 网络中所有 C3 模块替换为用于特征提取的改进模块 GhostV2C3。如图 2 所示，改进的 YOLOv5s-seg 网络主要由特征提取 Backbone、特征融合 Neck、检测头分支和原型掩码分支组成。特征提取结构主要由 CBS、GhostV2C3、空间金字塔池-快速(SPPF)模块组成。特征融合结构采用特征金字塔网络(FPN)与路径聚合网络(PAN)[15]融合。首先，特征提取网络对输入图像进行特征提取；然后，FPN 结构将可靠的语义信息从底层传递到顶层进行上采样，PAN 结构将可靠的定位信息从顶层传递到底层进行下采样，这两个特征融合后为预测头分支和原型掩码分支网络生成提供了基础；接着，将 FPN 和 PAN 融合的多尺度特征信息作为检测头分支的输入并生成各候选框的类别置信度、锚框和原型掩码系数，这些参数经非极大值抑制(NMS)滤波后得到新的输出；同时将 PAN 结构自底向上进行上采样的底层送入原型掩码分支(protonet)，生成原型掩码；最后将原型掩码与掩码系数线性组合，通过 Sigmoid 非线性函数来生成最终的实例掩码。

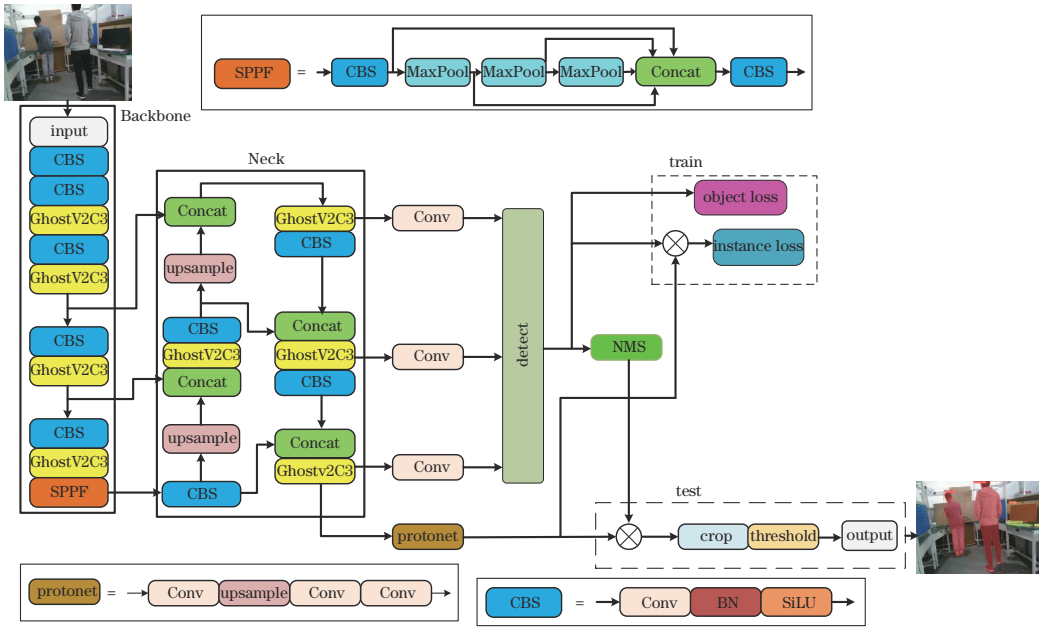

图 2　改进的 YOLOv5s-seg 网络结构

Fig. 2　Improved YOLOv5s-seg network structure

GhostNetV2[16]在 GhostNet 架构[17]上增加解耦全连接(DFC)注意力机制，DFC 注意力机制将全连接(FC)层分解为水平 FC 和垂直 FC，表达式为

$$a'_{hw}=\sum_{h'=1}^{H}F_{h,h'w}^{H}\odot z_{h'w},\ h=1,2,\cdots,H,\ w=1,2,\cdots,W,\quad(1)$$

$$a_{hw}=\sum_{w'=1}^{W}F_{w,hw'}^{W}\odot a'_{hw'},\ h=1,2,\cdots,H,\ w=1,2,\cdots,W,\quad(2)$$

式中：⊙为按元素乘法；F^H 和 F^W 为变换权重；输入是原始特征 $\boldsymbol{Z}$，$\boldsymbol{Z}\in\mathbf{R}^{H\times W\times C}$，$\boldsymbol{Z}=\{z_{11},z_{12},\cdots,z_{HW}\}$；$\boldsymbol{A}=\{a_{11},a_{12},\cdots,a_{HW}\}$ 是得到的注意图。式(1)和式(2)顺序作用在特征上，在卷积神经网络的二维特征图中聚合像素。这两个 FC 层包含沿各自方向较长范围内的像素，将它们叠加，产生一个全局的接受域。DFC 注意力机制捕捉了不同空间位置的像素之间的长期依赖关系，增强了模型的表达性，如图 3(a)所示。

GhostNetV2Bottleneck 模块如图 3(b)所示，其由 1 个 DFC 注意力机制、2 个核为 1×1 的 Ghost 卷积、残

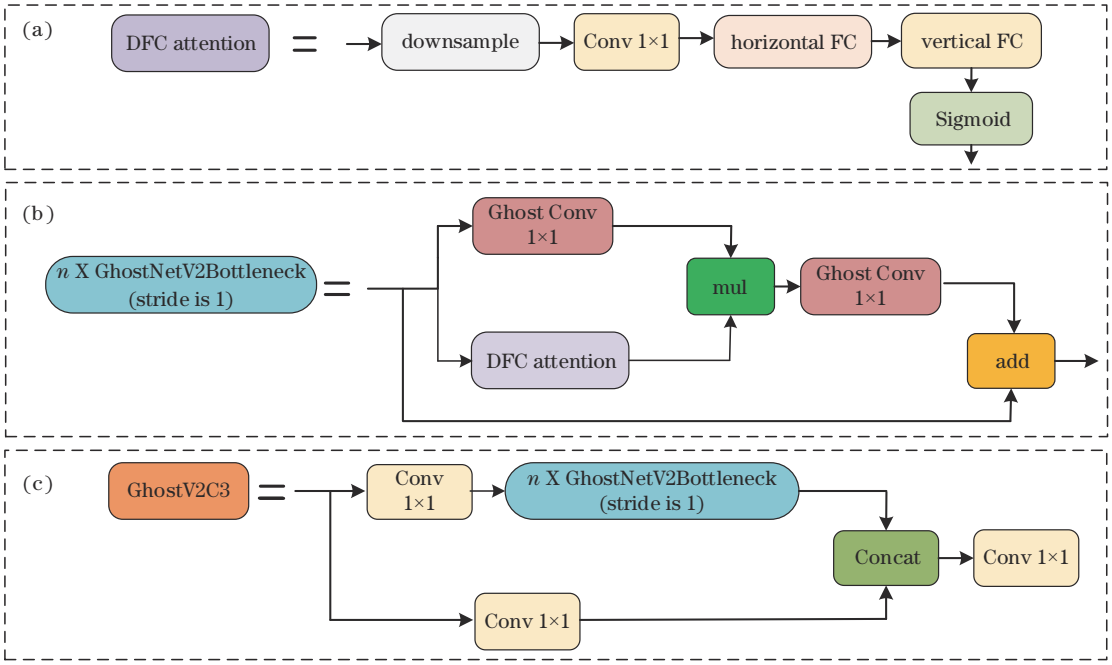

图 3　两种 Ghost 模块。(a)DFC 注意力;(b)GhostNetV2Bottleneck 模块;(c)GhostV2C3 模块

Fig. 3　Two Ghost modules. (a) DFC attention; (b) GhostNetV2Bottleneck module; (c) GhostV2C3 module

差连接构成。GhostNetV2Bottleneck 层将输入的特征图输入第一个 1×1 的 Ghost 卷积进行特征通道数扩展,在较大的通道数中使用深度可分离卷积提取大感受野特征,从而更好地捕捉输入图像的局部特征。同时 Ghost 卷积与 DFC 注意力机制分支并行,以增强扩展的特性。然后,增强后的特性传输到第二个 1×1 的 Ghost 卷积,以生成输出特性。

YOLOv5s-seg 使用 C3 模块在 Backbone 中的各个阶段进行特征提取和在 PANet 中进行多尺度特征融合。其中多个 Bottleneck 层的堆叠带来较大的参数量,造成检测效率低,考虑上述问题,设计了特征提取效率更高的 GhostV2C3 模块以减少网络时间开销,如图 3(c)所示。GhostV2C3 模块通过将 C3 模块的 Bottleneck 结构替换成 GhostNetV2Bottleneck 结构,可以有效地减少 YOLOv5 网络中 C3 模块使用的 3×3 卷积带来的额外计算开销,并使网络从原特征图中学习有用的信息,丢弃冗余的特征信息,从而在不降低特征提取能力的条件下有效地减少参数量。同时,该模块中的残差连接结构能够促进网络的浅层特征向深层特征传递,有效地解决普通卷积带来的梯度消失问题。通过用 GhostV2C3 模块替换掉 YOLOv5 网络中的 C3 模块,进一步降低模型的复杂性,增强分割网络轻量化,使 YOLOv5 分割网络在所提算法语义分割线程中降低了图像处理的速度,并保持一定的准确性。

3.2　加权几何约束

利用语义分割只消除了先验动态特征点,一些语义上静态而实际上是运动的潜在特征点无法被语义分割识别。利用极约束[18]来判断某点是否为动态特征点,即静态特征点满足极约束,动态特征点违反该约束,但是极约束检验动态点的方法并不能找到所有的动态点,如图 4(b)所示,空间点 P_{Q1} 沿着 A_1P_{Q1} 的方向移动到 P_{Q2} 点上,P_{Q2} 在当前帧的投影同样落在极线 I_2 上且满足极约束,但是该空间点为动态特征点,所以通过极约束来检测潜在的动态特征点的方法存在一定的局限性,本文通过一种加权几何约束方法来检测动态特征点。如图 4(a)所示,当空间点 P 在当前帧的投影落在极线 I_2 上时,满足

$$\overrightarrow{A_2p_2}\times\boldsymbol{n}=\boldsymbol{0},\tag{3}$$

式中:$\boldsymbol{n}$ 为由 A_1、P、A_2 三点所确定的极平面的法向量。

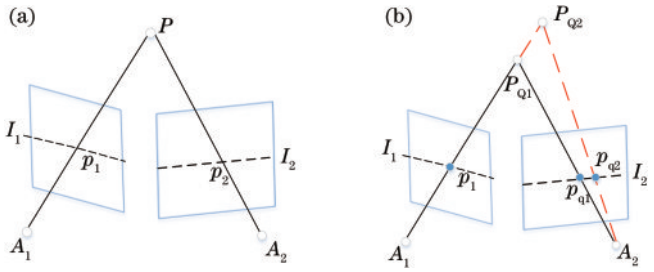

图 4　极约束示意图。(a)满足极约束的情况;(b)不满足极约束的情况

Fig. 4　Schematic of the epipolar constraint. (a) Case of meeting the epipolar constraint; (b) case where the epipolar constraint is not met

对于每个输入帧,选择与当前帧重合度最高的 5 个关键帧作为参考帧,利用深度信息差 $\Delta z=z-z'$ 和极平面的法向量 $\boldsymbol{n}$ 与 $\overrightarrow{A_2p_2}$ 的余弦值 $\cos(\boldsymbol{n},\overrightarrow{A_2p_2})$ 来检测动态特征点,表达式为

$$\delta=\alpha|\cos(\boldsymbol{n},\overrightarrow{A_2p_2})|+\beta|z-z'|,\tag{4}$$

式中：α和β分别是极约束影响因子和深度影响因子。当余弦值越接近90°时，说明投影点p_2离极线l_2的距离越近，则该点为静态特征点的概率越高。设定一个阈值σ_1，如果余弦值的绝对值大于σ_1时，令$\alpha=1$，$\beta=1$，即表示该投影点离极线的距离偏远，不需要计算深度差就可以确定它为动态点，否则令$\alpha=0$，$\beta=1$。最后设定阈值σ_2，若$\delta\geqslant\sigma_2$，则确定为动态点。图5是TUM数据集fr3/walking_xyz序列中包含潜在动态特征点的某一帧。图5(b)为只通过语义分割来去除先验的动态特征点的情况，从图5(b)可以看出右侧人椅子上存在大量潜在动态特征点。利用极约束的方法，从图5(c)可以看出右侧人椅子上仍然有少量潜在动态特征点未去除。从图5(d)可以明显地看出，加权几何约束方法去除了人椅子上的潜在动态特征点。

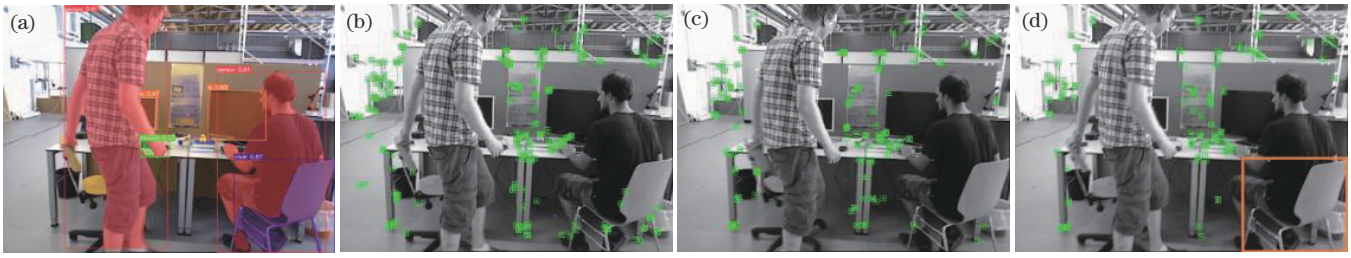

图5　去除动态特征点。(a) 语义分割；(b) 语义分割去除；(c) 极约束去除；(d) 加权几何约束去除

Fig. 5　Removing dynamic feature points. (a) Semantic segmentation; (b) semantic segmentation removal; (c) pole constraint removal; (d) weighted geometric constraint removal

4　基于加权静态的BA联合优化

由于大多数可用的SLAM系统都依赖于静态场景，不适用于动态场景，从而影响了它们在现实场景中的部署。ORB-SLAM2在进行线程跟踪的过程中，利用所有的特征点来进行局部BA优化[19]，估计相机位姿，有

$$\{\boldsymbol{R}^*,\boldsymbol{t}^*\}=\arg\min_{\boldsymbol{R},\boldsymbol{t}}\frac{1}{2}\sum_{i=1}^{n}\left\|\boldsymbol{p}_i-\pi(\boldsymbol{R}\boldsymbol{P}_i+\boldsymbol{t})\right\|_2^2,\tag{5}$$

式中：$\{\boldsymbol{R}^*,\boldsymbol{t}^*\}$表示相机位姿；$\boldsymbol{p}_i$为投影的像素坐标；$\boldsymbol{P}_i$为空间点坐标；$\pi$表示从三维坐标系到像素坐标系的投影。在动态场景下，动态特征点参与最小化重投影误差将会导致相机位姿估计产生较大误差。本文根据语义掩码提取ORB特征点，获得语义静态特征点。语义静态特征点包含静态特征点和潜在动态特征点，所以通过定义一种静态加权约束，给语义静态特征点赋予权值φ_i，表达式为

$$\varphi_i=\eta\mathrm{mask}_{\mathrm{seg}}(u,v)+\omega\mathrm{mask}_{\mathrm{seg}}(u,v),\tag{6}$$

式中：$\mathrm{mask}_{\mathrm{seg}}(u,v)$为动态目标检测结果；$\eta$和$\omega$分别为语义掩码和几何掩码对特征点的贡献比例，且满足$\eta+\omega=1$。φ_i表示语义静态特征点的可靠性，权值越大越可靠，说明某点是静态特征点的可靠性越大，对BA优化过程中目标函数的贡献越大；反之，值越小，说明某点是潜在动态特征点的可能性越大，对目标函数的贡献越少。通过语义掩码提取ORB特征点，并通过式(6)初始化语义静态特征点的权值。η和ω的初始值都为0.5。然后通过BA联合优化估计相机的位姿$\{\boldsymbol{R}^*,\boldsymbol{t}^*\}$和语义静态特征点的权值$\varphi_i$来更新相机的位姿和特征点的权值。

采用静态加权约束方法可以检测环境中潜在动态特征点的运动信息，减少参与BA优化过程的特征点对数，提高位姿优化的精度。本文将重投影误差乘以语义静态特征点的初始化权值，并将结果相加，通过与权值的乘积，可以根据特征点的可靠性，合理地分配特征点对目标函数的贡献。权值越大的，特征点越可靠，说明某点是静态特征点的可能性越大，对BA优化目标函数的贡献越大。最后对特征点的权值和相机位姿进行优化，表达式为

$$\{\boldsymbol{R}^*,\boldsymbol{t}^*,\varphi^*\}=\arg\min_{\boldsymbol{R},\boldsymbol{t},\varphi_i}\frac{1}{2}\varphi_i\sum_{i=1}^{n}\left\|\boldsymbol{p}_i-\pi(\boldsymbol{R}\boldsymbol{P}_i+\boldsymbol{t})\right\|_2^2,\tag{7}$$

本文在图优化库(g2o)中采用Levenburg-Marquardt方法进行BA联合优化，需要手工计算误差项对优化变量的偏导数，所以特征点权值的偏导数为

$$\frac{\partial e}{\partial\varphi_i}=\begin{bmatrix}u_i-\left(\frac{X}{Z}f_x+c_x\right)\\ v_i-\left(\frac{Y}{Z}f_y+c_y\right)\\ u_i-\left(\frac{X}{Z}f_x+c_x-\frac{b_f}{Z}\right)\end{bmatrix},\tag{8}$$

式中：f_x、f_y、c_x、c_y、b_f为相机的参数，均已知；$\boldsymbol{P}=(X\ Y\ Z)^{\mathrm{T}}$为相机坐标系下的空间点坐标，$(u_i,v_i)$为其对应的像素坐标点。

5　实验与分析

5.1　与ORB-SLAM2及其他先进SLAM系统进行比较

为了评估所提系统的性能，对ORB-SLAM2和所提算法在TUM RGB数据集[20]下进行比较。图6(a)～

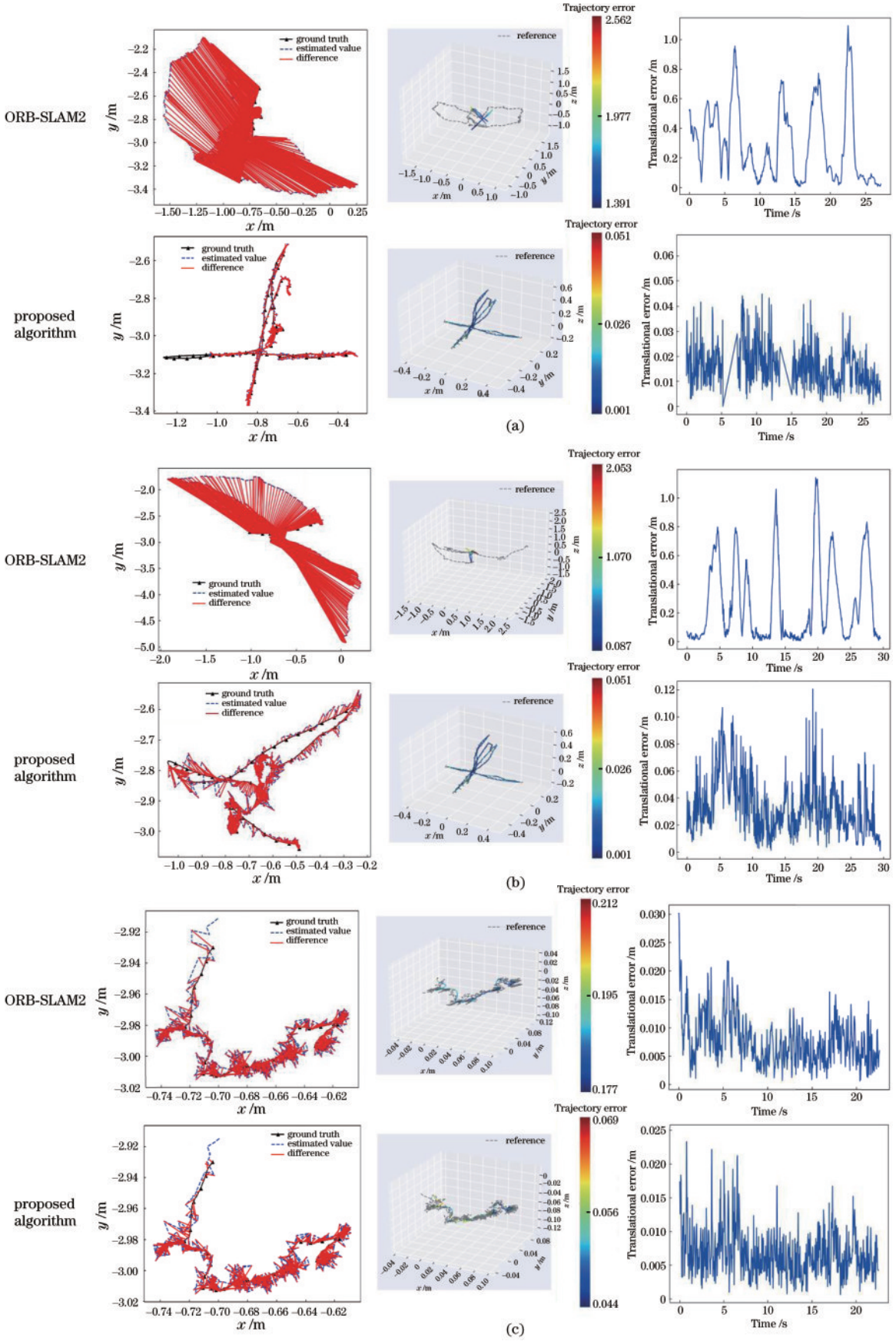

图 6　绝对轨迹误差图、三维轨迹误差热力图和相对位姿误差图。(a) fr3/walking_xyz 序列；(b) fr3/walking_rpt 序列；(c) fr3/sitting_static 序列

Fig. 6　Absolute trajectory error map, 3D trajectory error heat map, and relative pose error map. (a) fr3/walking_xyz sequence; (b) fr3/walking_rpt sequence; (c) fr3/sitting_static sequence

(c)分别表示ORB-SLAM2和所提算法在3个高/低动态数据集上的绝对轨迹误差(ATE)图、三维轨迹误差热力图和相对位姿误差(RPE)图。如图6所示,在高动态环境下,ORB-SLAM2系统估计的运动轨迹与真实轨迹存在较大的差异,甚至在某些区域产生错误的轨迹。相反,所提算法估计的运动轨迹和真实轨迹高度重叠。

表1和表2分别展示了所提算法与ORB-SLAM2运行结果中ATE和RPE的均方根误差(RMSE)、误差中值(Median)和标准差(S. D.)。与ORB-SLAM2相比,所提算法在ATE和PRE方面有显著提高;ATE方面,在典型的高动态序列(fr3/walking_xyz)中,ORB-SLAM2的RMSE从0.7323下降到所提算法的0.0133,提升了98.18%,对于其他的高动态序列,所提算法的RMSE与ORB-SLAM2相比提高了96.10%~97.98%;但是在低动态序列下,改进的程度不是很显著,RMSE值只提升了24.71%。这主要是因为ORB-SLAM2本身是为低动态环境而设计的,能够很好地处理低动态场景,能取得良好的效果。为了进一步体现所提算法的有效性,对所提算法与Dyna-SLAM[7]、DS-SLAM[21]和RDS-SLAM[22]在TUM RGB-D数据集中选择的5组序列上进行比较,如表3所示。比较结果表明,所提算法相较于大多数处理动态场景的先进SLAM系统具有更好的性能。

表1 绝对轨迹误差的对比

Table 1 Comparison of the absolute trajectory error unit: m

Sequence	ORB-SLAM2			Proposed algorithm			Improvement percentage /%		
	RMSE	Median	S. D.	RMSE	Median	S. D.	RMSE	Median	S. D.
w_half	0.6173	0.5074	0.2495	0.0241	0.0184	0.0112	96.10	96.37	95.51
w_rpy	0.8154	0.6542	0.4236	0.0275	0.0204	0.0158	96.63	96.88	96.27
w_static	0.3518	0.2639	0.1607	0.0071	0.0054	0.0030	97.98	97.95	98.13
w_xyz	0.7323	0.6673	0.2564	0.0133	0.0113	0.0068	98.18	98.31	97.34
s_staic	0.0085	0.0074	0.0042	0.0064	0.057	0.0033	24.71	22.97	21.42

表2 相对位姿误差的对比

Table 2 Comparison of the relative pose error unit: m

Sequence	ORB-SLAM2			Proposed algorithm			Improvement percentage /%		
	RMSE	Median	S. D.	RMSE	Median	S. D.	RMSE	Median	S. D.
w_half	0.4458	0.2334	0.2859	0.0251	0.0195	0.0122	94.37	91.65	95.73
w_rpy	0.4803	0.121	0.3416	0.0381	0.0274	0.0262	92.06	77.36	92.33
w_static	0.2214	0.017	0.1463	0.0092	0.0092	0.0048	95.84	95.06	96.70
w_xyz	0.4512	0.1653	0.2923	0.0179	0.0139	0.0090	96.03	91.59	96.23
s_staic	0.0083	0.0072	0.0043	0.0076	0.0058	0.0038	8.43	19.44	11.63

表3 一些先进的SLAM算法的绝对轨迹误差的对比

Table 3 Comparison of the absolute trajectory error of some advanced SLAM algorithms unit: m

Sequence	Dyna-SLAM		DS-SLAM		RDS-SLAM		Proposed algorithm	
	RMSE	S. D.	RMSE	S. D.	RMSE	S. D.	RMSE	S. D.
w_half	0.0296	0.0157	0.0303	0.0159	0.0807	0.0454	0.0241	0.0112
w_rpy	0.0415	0.0271	0.4442	0.2350	0.1604	0.0873	0.0275	0.0158
w_static	0.0068	0.0034	0.0081	0.0036	0.0206	0.012	0.0071	0.0030
w_xyz	0.0157	0.0083	0.0247	0.0161	0.0571	0.0229	0.0133	0.0068
s_staic	0.0067	0.0046	0.0065	0.0033	0.0084	0.0043	0.0064	0.0033

5.2 真实环境下的评估

为了证明所提系统的有效性和实时性,采集了两个人在实验室里站立和行走的真实动态场景的实验数据。图7(b)第1行和第2行分别表示相机捕获的真实的实验室动态场景原始RGB图像和语义分割图像,第3行和第4行表示原始ORB-SLAM2算法和所提算法的ORB特征提取情况。由图7(b)第3行可知,ORB-SLAM2提取了大量的特征,而在图7(b)第4行中,移动的人身上的特征点基本上被所提算法去除,几乎所有的特征点都是在静态背景下提取的。如图7(a)所示,使用奥比中光(Astra Pro)相机在真实的动态场景中采集一个沿着圆形路径移动的高动态数据集和一个沿着凸形路径移动的较高动态数据集。图8展示了ORB-SLAM2、Dyna-SLAM和所提算法在两个现实场景下的相机轨迹比较。由于实验室动态特征点的存在,ORB-SLAM2估计的虚线轨迹存在多处偏移现

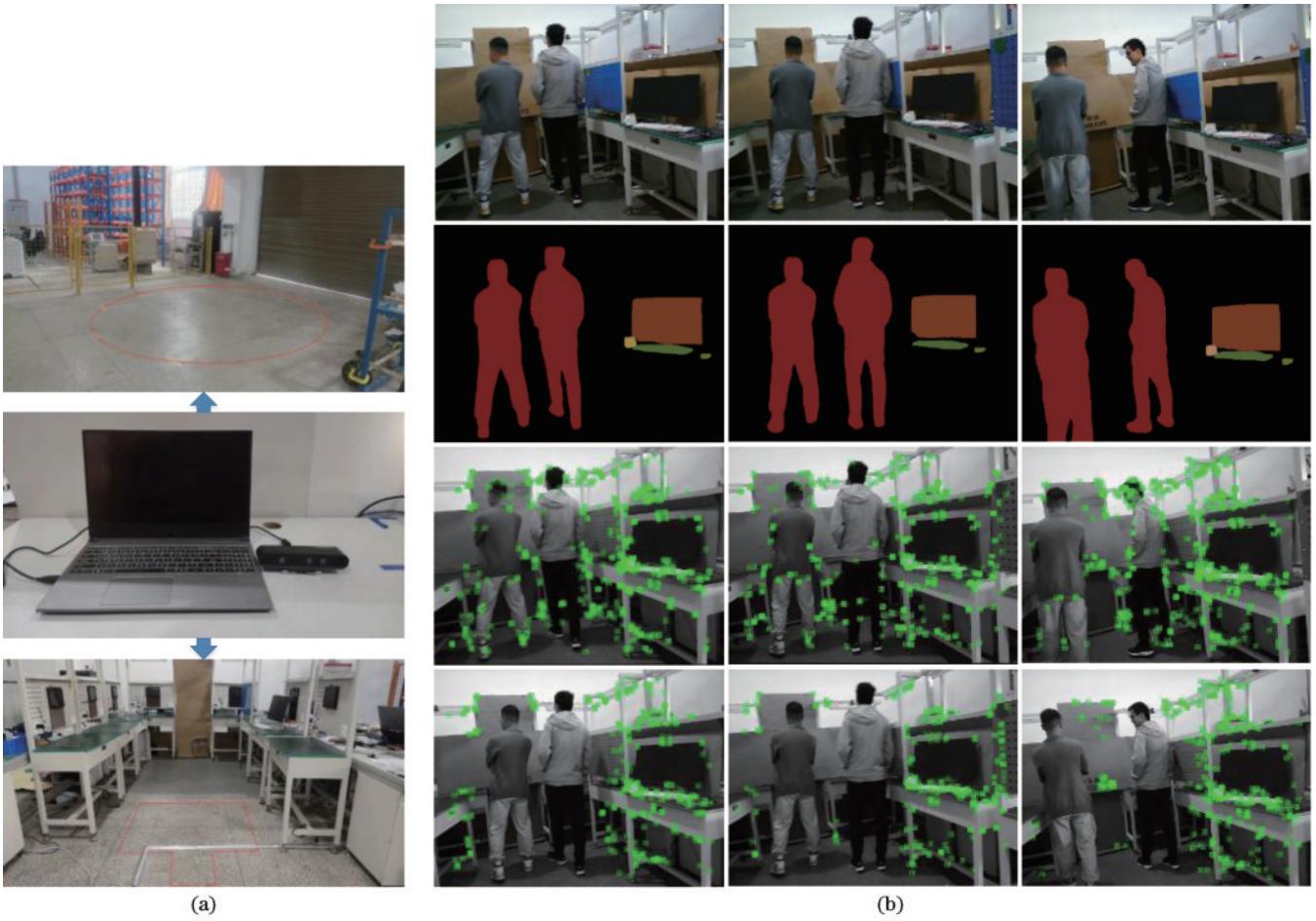

图 7　真实的动态场景中使用 RGB-D 相机进行的实验。(a)实验场景;(b)实验效果

Fig. 7　Experiments using RGB-D cameras in a real dynamic scene. (a) Experiment scene; (b) experimental result

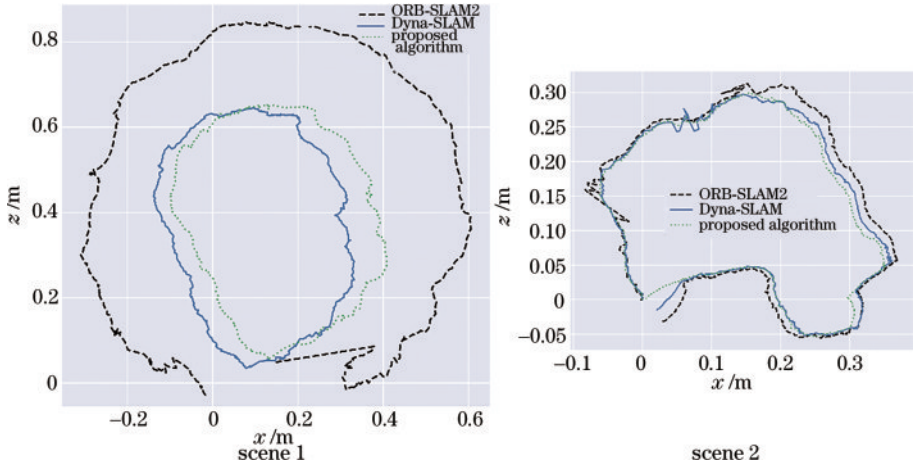

图 8　不同算法在真实的动态场景中的比较

Fig. 8　Comparison of different algorithms in a real dynamic scene

象,且最终无法形成一个完整的闭环。与 ORB-SLAM2 相比,所提算法估计的轨迹很好地形成了一个闭环,这定性地反映了所提算法的准确性。与 Dyna-SLAM 相比,所提算法估计的相机轨迹也有较好的定位精度。

5.3　实时性评估

在实际的应用中,实时性是评估 SLAM 的重要性能指标之一,跟踪时间指在跟踪线程完成对一帧图像的提取、匹配和位姿估计等所需的时间,不同的 SLAM 系统的平均跟踪时间如表 4 所示。所提算法在语义分割部分耗时比较低,在联合优化部分耗时比较严重,但是整体的跟踪线程耗时远优于 Dyna-SLAM,与 RDS-SLAM 跟踪时间相近,但定位精度高于 RDS-SLAM。

表 4 跟踪阶段的耗时对比

Table 4 Time-consuming comparison of the tracking phase

Method	Semantic segmentation/detection time /ms	Time of other modules related to tracking /ms	Track each frame time /ms	GPU
ORB-SLAM2	–	–	38.23	GeForceRTX1650
Dyna-SLAM	198.63	Geometry module: 239.52	672.38	GeForceRTX1650
RDS-SLAM	200	Mask generation: 5.31	85.43	GeForceRTX1650
Proposed algorithm	20.48	Joint optimization: 132.44	90.56	GeForceRTX1650

6 结 论

为了消除动态目标对系统定位精度的影响，提出了一种动态场景下基于加权静态的视觉 SLAM 算法，在 ORB-SLAM2 算法中引入了语义分割线程和几何线程。采用改进的轻量化 YOLOv5 语义分割网络来提高所提算法的实时性，同时利用几何加权约束在几何线程中确定特征点的真实运动状态。将语义分割线程的分割结果与几何线程的检测信息相结合，获得动态目标检测结果，基于检测结果给语义静态特征点赋予权值，并对特征点权值和相机位姿估计进行 BA 联合优化，从而有效地提升了 SLAM 系统的定位精度。最后为了评估所提算法的可行性，在 TUM 数据集和真实的室内动态场景下，与 ORB-SLAM2 进行对比，所提算法在高动态环境数据集上的定位精度和稳定性有较高的提升，与 Dyna-SLAM 和 RDS-SLAM 等先进的视觉 SLAM 算法相比，所提算法的定位精度和实时性也得到一定的提升。尽管在定位精度和实时性能方面取得了进展，但仍然存在以下问题：一方面，系统的实时性能有待提高，需要进一步的改进；另一方面，需要对语义分割网络进行不断的优化，从而提高动态目标检测的分割精度。

参 考 文 献

[1] 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.

[2] Campos C, Elvira R, Rodríguez J J G, et al. ORB-SLAM3: an accurate open-source library for visual, visual-inertial, and multimap SLAM[J]. IEEE Transactions on Robotics, 2021, 37(6): 1874-1890.

[3] Engel J, Schöps T, Cremers D. LSD-SLAM: large-scale direct monocular SLAM[M]//Cremers D, Reid I, Saito H, et al. Computer vision-ECCV 2014. Lecture notes in computer science. Cham: Springer, 2014, 8690: 834-849.

[4] Fischler M A, Bolles R C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography[J]. Communications of the ACM, 1981, 24(6): 381-395.

[5] Sun Y X, Liu M, Meng M Q H. Improving RGB-D SLAM in dynamic environments: a motion removal approach[J]. Robotics and Autonomous Systems, 2017, 89: 110-122.

[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] 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.

[8] 王梦瑶, 宋薇. 动态场景下基于自适应语义分割的RGB-D SLAM算法[J]. 机器人, 2023, 45(1): 16-27.
Wang M Y, Song W. RGB-D SLAM algorithm based on adaptive semantic segmentation in dynamic scene[J]. Robot, 2023, 45(1): 16-27.

[9] Li S L, Lee D. RGB-D SLAM in dynamic environments using static point weighting[J]. IEEE Robotics and Automation Letters, 2017, 2(4): 2263-2270.

[10] Li A, Wang J K, Xu M, et al. DP-SLAM: a visual SLAM with moving probability towards dynamic environments[J]. Information Sciences, 2021, 556: 128-142.

[11] He K M, Gkioxari G, Dollár P, et al. Mask R-CNN[C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 2980-2988.

[12] Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 39(4): 640-651.

[13] Badrinarayanan V, Kendall A, Cipolla R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481-2495.

[14] Bolya D, Zhou C, Xiao F Y, et al. YOLACT: real-time instance segmentation[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Republic of Korea. New York: IEEE Press, 2020: 9156-9165.

[15] Kasper-Eulaers M, Hahn N, Berger S, et al. Short communication: detecting heavy goods vehicles in rest areas in winter conditions using YOLOv5[J]. Algorithms, 2021, 14(4): 114.

[16] Tang Y H, Han K, Guo J Y, et al. GhostNetV2:

enhance cheap operation with long-range attention[EB/OL]. (2022-11-23)[2023-02-03]. https://arxiv.org/abs/2211.12905.

[17] Han K, Wang Y H, Tian Q, et al. GhostNet: more features from cheap operations[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 1577-1586.

[18] Hartley R, Zisserman A. Multiple view geometry in computer vision[M]. 2nd ed. Cambridge: Cambridge University Press, 2003.

[19] Engels C, Stewénius H, Nistér D. Bundle adjustment rules[EB/OL]. [2023-02-03]. https://www.isprs.org/proceedings/XXXVI/part3/singlepapers/O_24.pdf.

[20] 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, October 7-12, 2012, Vilamoura-Algarve, Portugal. New York: IEEE Press, 2012: 573-580.

[21] Yu C, Liu Z X, Liu X J, et al. DS-SLAM: a semantic visual SLAM towards dynamic environments[C]//2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), October 1-5, 2018, Madrid, Spain. New York: IEEE Press, 2019: 1168-1174.

[22] Liu Y B, Miura J. RDS-SLAM: real-time dynamic SLAM using semantic segmentation methods[J]. IEEE Access, 2021, 9: 23772-23785.