Modeling and Simulation 建模与仿真, 2025, 14(3), 179-185
Published Online March 2025 in Hans. https://www.hanspub.org/journal/mos
https://doi.org/10.12677/mos.2025.143213

Hans汉斯

动态场景下基于YOLO的场景重建

沈康鹏

上海理工大学机械工程学院，上海

收稿日期：2025年2月13日；录用日期：2025年3月6日；发布日期：2025年3月14日

摘　要

在实际动态环境中，深度传感器在获取环境信息时不可避免地会受到运动物体的干扰。如何有效处理动态物体、使机器人准确理解周围环境并完成复杂任务，仍然是一个亟待解决的难题。本文提出了一种基于改进型ORB-SLAM3与改进的YOLOv5相结合的语义分割方法。该方法通过识别并剔除动态特征，同时最大程度地保留静态环境的有效特征，结合ORB-SLAM3算法实现了高精度的场景重建，成功生成稠密点云地图。实验结果表明，在TUM-RGB-D数据集上，本文提出的方法相比原始ORB-SLAM3算法，在高动态场景中的RMSE平均降低了92.04%，在低动态场景中的RMSE平均降低了19.48%。特别是在动态物体比例较高的场景中，系统表现出优异的鲁棒性和准确性。此外，本文还对系统的实时性进行了优化，通过轻量化的目标检测网络和高效的特征筛选策略，确保了系统在普通硬件平台上的实时运行能力。研究结果为解决动态环境下的视觉SLAM问题提供了一种高效可靠的解决方案。

关键词

同时定位和建图，动态场景，ORBSLAM3，语义分割

Scene Reconstruction Based on YOLO in Dynamic Environments

Kangpeng Shen

School of Mechanical Engineering, University of Shanghai for Science and Technology, Shanghai

Received: Feb. 13th, 2025; accepted: Mar. 6th, 2025; published: Mar. 14th, 2025

Abstract

In real-world dynamic environments, depth sensors inevitably encounter interference from moving objects while acquiring environmental information. How to effectively process dynamic objects, enable robots to accurately understand their surroundings, and accomplish complex tasks remains a challenging problem. This paper proposes a semantic segmentation method that combines an

文章引用: 沈康鹏. 动态场景下基于 YOLO 的场景重建[J]. 建模与仿真, 2025, 14(3): 179-185.
DOI: 10.12677/mos.2025.143213

improved ORB-SLAM3 with an enhanced YOLOv5. The method identifies and eliminates dynamic features while maximally preserving effective features of the static environment. By integrating the ORB-SLAM3 algorithm, it achieves high-precision scene reconstruction and successfully generates dense point cloud maps. Experimental results on the TUM-RGB-D dataset show that compared to the original ORB-SLAM3 algorithm, our proposed method reduces RMSE by an average of 92.04% in highly dynamic scenes and 19.48% in low dynamic scenes. The system demonstrates excellent robustness and accuracy, particularly in scenarios with a high proportion of dynamic objects. Additionally, we optimized the system's real-time performance through a lightweight object detection network and efficient feature filtering strategy, ensuring real-time operation on standard hardware platforms. The research provides an efficient and reliable solution for visual SLAM problems in dynamic environments.

Keywords

Simultaneous Localization and Mapping (SLAM), Dynamic Scenes, ORB-SLAM3, Semantic Segmentation

Copyright © 2025 by author(s) and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

1. 引言

SLAM 技术能让机器人在未知环境中同步完成定位和地图构建。系统通过处理传感器数据来实现这一目标，常用的传感器包括不同类型的摄像头和激光雷达。当使用摄像头作为主要传感器时，这类系统被称为视觉 SLAM。它在服务机器人、增强现实和无人驾驶等领域有广泛应用，可以准确估计相机位置并构建环境的三维模型。

视觉 SLAM 在静态环境下已取得较为成熟的成果，例如 MonoSLAM、PTAM、ORB-SLAM、ORB-SLAM2、LSD-SLAM、SVO 和 DSO 等经典方法[1]-[7]。然而，在动态环境中，由于行人、车辆等动态目标的干扰，静态世界假设常常失效，导致系统性能下降甚至失败。例如，在高动态环境中，如果大部分特征点落在动态目标上(如图 1 所示)，这些动态特征点会直接影响系统的定位与地图构建精度。常规的鲁棒优化方法(如 RANSAC)无法有效过滤出动态点，最终导致轨迹估计偏离实际轨迹。

为了解决动态环境中的挑战，检测和剔除动态目标是关键步骤。早期的方法主要依赖于几何约束，如通过对极几何分析光流场的变化检测移动目标[8]。这类方法计算复杂度低，但在动态目标复杂或光照变化剧烈的场景下容易失效。随着深度学习的发展，基于语义分割的动态目标检测方法得到了广泛应用，例如 DS-SLAM 结合 SegNet 网络进行语义分割，利用动态目标的位置信息辅助 SLAM 定位[9]；DynaSLAM 结合 Mask R-CNN 和多视图几何检测动态目标，从而提升了系统的鲁棒性[10]。然而，这些方法通常计算开销较大，且依赖于强大的硬件设备。

针对这一问题，我们提出了一种结合 YOLO 检测与对极几何的 SLAM 改进方案。该方法首先使用 YOLO 快速检测动态目标(如行人、车辆)，并通过检测框剔除动态目标相应的特征点。然后，基于对极几何验证特征点的静态性，进一步去除动态干扰。与传统的语义分割方法相比，该方案不仅提高了实时性能，还有效降低了计算负担。我们仅在关键帧执行 YOLO 检测，而在其他帧中通过对极几何更新特征点，从而实现高效的动态目标剔除。

与其他现有方法相比，本方法在多个数据集上的测试结果表明，它能够显著提升定位精度，同时保

持实时性能。例如，文献[11]提出的稀疏运动去除模型(SMR)虽然能够过滤动态特征，但在帧间差异过大时容易导致跟踪丢失。文献[12]通过相邻帧间差异检测运动目标，然而在复杂动态环境下仍难以达到理想效果。基于 ORB-SLAM3 的研究[13]则通过语义分割来去除动态目标，尽管精度有所提高，但仍依赖于较大的计算资源。此外，YOLO-SLAM [14]利用 YOLOv3 检测动态目标，但当动态目标过多时，定位精度仍然下降。相比之下，我们的方案不仅能在实时性上保持优势，还能在较为复杂的动态环境中提高定位精度，成为解决动态场景 SLAM 的一种有效方案。

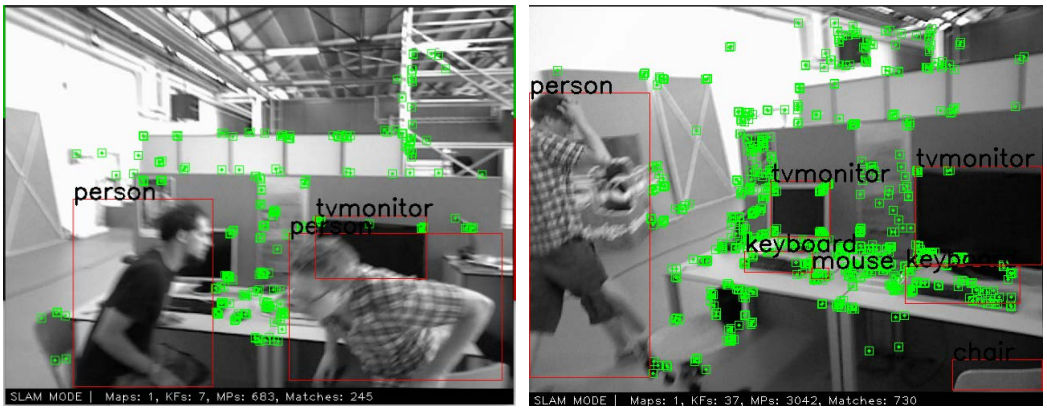

Figure 1. Dynamic objects
图 1. 动态物体

2. 系统框架

我们基于 ORB-SLAM3 设计了一个新的动态物体检测模块，通过独立线程运行。如图 2 所示，系统由多个并行线程构成：跟踪线程负责提取 ORB 特征，检测线程用 YOLOv5 识别动态物体。系统先通过动态掩膜去除相应特征点，再用多视图几何约束过滤剩余动态点。最终，处理后的关键帧被用于定位和建图。

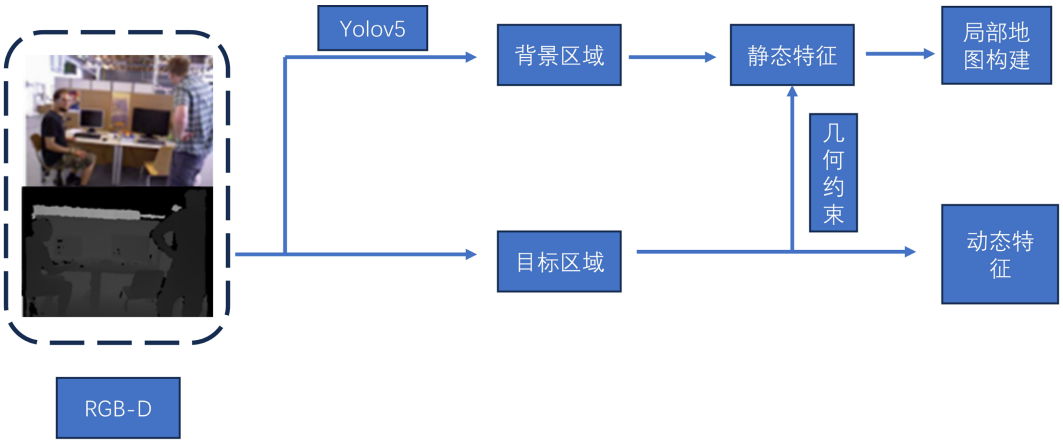

Figure 2. System framework
图 2. 系统框架

2.1. 动态目标检测模块

首先，他们提取图像中的潜在对象位置，然后对这些对象进行分类并细化其位置属性。该多步骤方

法被设计为在识别和精确定位感兴趣的对象时提供增强的准确性。相比之下，一级网络被设计成直接预测图像内的对象类别和位置。虽然两级目标检测网络倾向于提供更高的识别精度，但与一级网络相比，它们以较慢的训练和推理速度为代价。针对视觉 SLAM 对实时性的强烈要求，本文采用了经济高效的单级目标检测网络 YOLO-V5 s。在本研究中，YOLO-V5 s 经过了进一步的微调和轻量化，以满足视觉 SLAM 的特定要求。

传统的卷积神经网络模型往往具有大量参数的主要原因之一是卷积过程生成了大量相似的特征图。特征图的这种丰富性增加了模型的参数和计算复杂度，从而影响检测速度。为了解决这一问题，我们采用了轻量级模块来优化网络主干。在本文中，Ghost 模块和 Ghost Bottleneck 取代了 YOLO-V5 的传统卷积和 BottleneckCSP 组件。Ghost 模块能够通过线性变换和组合操作生成更多种类的特征图，从而实现更全面的特征信息表示。与传统的卷积网络相比，传统的卷积网络都是利用卷积来生成特征图，网络模型参数的压缩比由(1)给出。

$$r_c = \frac{n \cdot c \cdot k \cdot k}{\frac{n}{s} \cdot c \cdot k \cdot k + (s-1) \cdot \frac{n}{s} \cdot d \cdot d} \approx \frac{s \cdot c}{s + c - 1} \approx s \tag{1}$$

2.2. 动态感知 SLAM

我们对 ORB-SLAM 3 进行了动态感知改进。ORB-SLAM 3 是顶级的特征点 SLAM 系统，在静态场景中表现出色。改进版增加了两个功能：用 Yolo-v5 检测动态目标并标记边界框，以及利用几何约束筛选特征点。系统先对 RGB 图像进行目标检测和特征提取，再通过几何约束过滤动态特征点，最后用剩余的静态特征点估计位姿

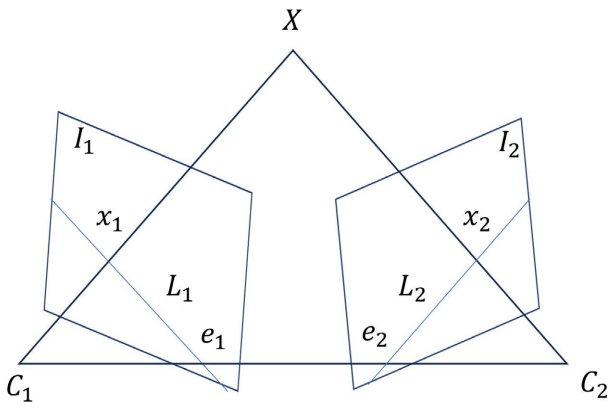

Figure 3. Epipolar geometry constraints
图 3. 对极几何约束

2.3. 多视图几何约束

基于深度学习的目标检测算法虽然可以识别预设的运动目标，但对于体积较小的物体或人为移动的物体的检测效果仍有不足。为解决这一问题，本研究采用对极几何原理进行动态特征点的筛除。对极几何是研究相机运动关系的基础理论模型，如图 3 所示：空间点 X 经过投影，在不同视角下的成像平面上形成对应点 X_1 和 X_2。齐次坐标如(2)所示。

$$\begin{cases} X_1 = [u_1 \ v_1 \ 1] \\ X_2 = [u_2 \ v_2 \ 1] \end{cases} \tag{2}$$

其中 u 和 v 是像素的相应水平和垂直坐标。然后，对应于点 E1 的核线 E1 由(3)给出。

$$L_1 = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = FP_1 \tag{3}$$

其中 F 表示相应的基本矩阵。X、Y 和 Z 是极线的向量 1。根据从点到直线的距离公式，从点 X_1 到极线 X_1 的距离 d 如(4)所示。

$$d = \frac{\left|P_2 F P_1\right|}{\sqrt{X^2 + Y^2}} \tag{4}$$

实际应用中，静态点的对应点对由于特征匹配和基础矩阵计算存在误差，可能会偏离极线约束。为此，本文在获取基本运动矩阵后，采用 RANSAC 方法筛选相邻帧间的稳定特征点作为内点集。基于基本矩阵计算两帧间的极线，得到投影点在相邻帧上的坐标及其到极线的距离 d_1 和 d_2。将两个距离之和定义为特征点的总偏差 D。通过比较 D 值与预设阈值 τ 的大小，可判定特征点的属性：当 D 值超过阈值时，将其归类为动态特征点并剔除；反之则视为静态特征点，用于后续位姿估计。

3. 实验和分析

本研究在搭载 Intel i7-12650H 处理器、16GB 内存和 RTX 4060Ti GPU 的 Ubuntu 18.04 平台上进行实验。系统核心采用 C++开发并由 cmake 构建，分割模块使用 Python 实现。实验数据选取 TUM-RGB-D 数据集中的四组动态序列，包括表征低动态环境的 sitting 系列和代表高动态场景的 walking 系列。

系统评估采用绝对轨迹误差(ATE)指标，通过 RMSE、平均值和中值三个维度进行精度分析。为确保结果可靠性，本文对比了改进算法与原始 ORB-SLAM3 在各数据集上的表现，每组实验重复执行 3 次并取平均值。

表 1 展示了改进算法与 ORB-SLAM3 算法的 ATE 对比结果，其中的提升率使用公式(5)计算。

$$\rho = \frac{e_{orb} - e_{plyo}}{e_{orb}} \times 100\% \tag{5}$$

式中：e_{orb} 为 ORB-SLAM3 算法运行结果，e_{plyo} 为本文提出的算法运行结果。

Table 1. Experimental data
表 1. 实验数据

图像序列	ORB_SLAM3/m			Our-Method/m		
	RMSE	Mean	Median	RMSE	Mean	Median
Sitting_halfsphere	0.0578	0.0545	0.0491	0.0503	0.0505	0.0431
Sitting_xyz	0.0154	0.0138	0.0123	0.0124	0.0107	0.0095
Waliking_halfphere	0.2649	0.2308	0.1724	0.0210	0.0170	0.0141
Waliking_static	0.0267	0.0231	0.0205	0.0072	0.0064	0.0060

如表 1 所示，Our-Method 在四组数据集的精度均优于 ORB-SLAM3，RMSE 平均下降 49.39%。其中高动态场景 RMSE 平均下降 92.04%，低动态场景则下降 19.48%，表明去除动态物体后定位精度显著提升，在高动态场景中尤为明显。视觉 SLAM 的最终目的是建图，通过计算轨迹误差与真实轨迹之间的差距来评估算法性能。在轨迹误差对比中，本文使用 EVO 工具对不同算法生成的轨迹文件和真实轨迹进行

评估，并进行可视化展示。

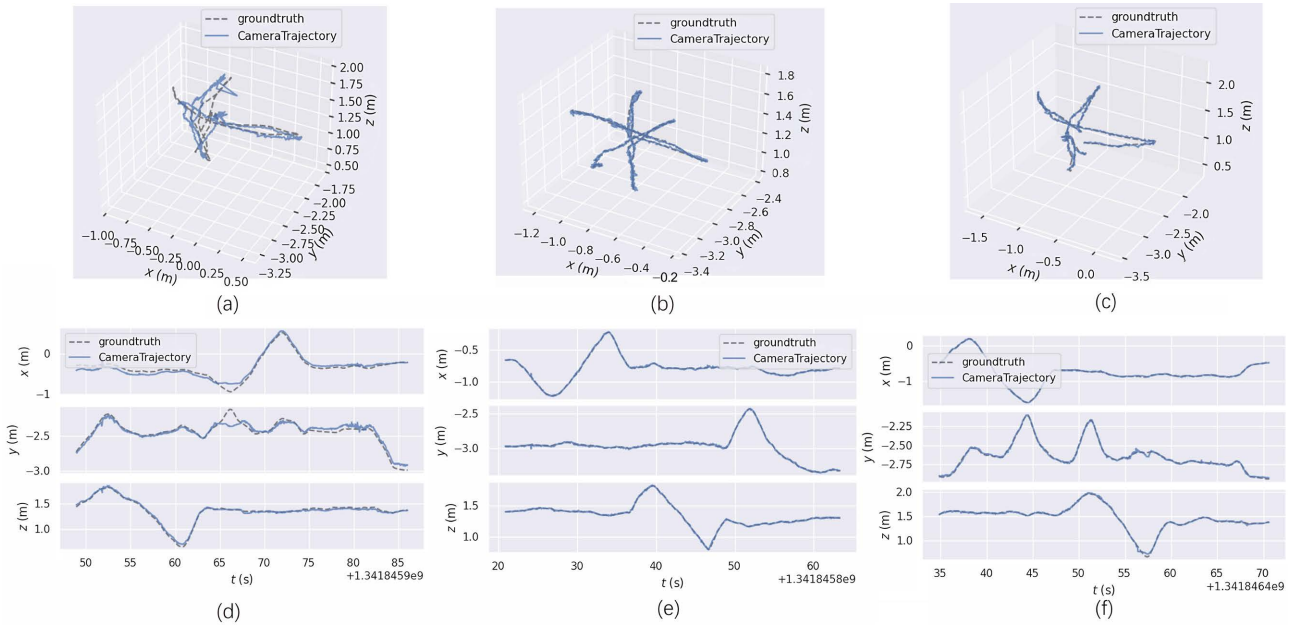

Figure 4. Trajectories of sitting and walking
图 4. Sitting，walking 轨迹图

walking_half 序列中，相机运动状态由位置和姿态两部分表示：XYZ 坐标反映相机在三个方向的位移，RPY 角度描述相机围绕各轴的旋转。

本文设计的 SLAM 系统结合了语义和几何分析方法。图 4 展示了系统在动态环境下的轨迹估计结果，将实际轨迹与参考真值进行对比。在动态目标较多时，语义模块通过筛除动态特征提高精度，几何模块则保留静态特征维持跟踪；而原始 ORB-SLAM3 易受动态对象影响导致跟踪失效。当场景中动态成分较少时，由于语义处理的影响减弱，两种方法表现接近。测试结果表明该系统能有效抑制轨迹漂移。

4. 结论

本研究提出了一种结合语义和几何信息的改进型 ORB-SLAM3 系统，以解决动态环境下的视觉 SLAM 问题。通过引入 YOLOv5 目标检测器来识别动态物体，并利用对极几何约束进行动态特征点筛选，系统有效剔除动态物体的干扰，显著提升了高动态环境中的定位精度。实验结果表明，本文方法在 TUM-RGB-D 数据集上的 RMSE 相比原始 ORB-SLAM3 算法，尤其在高动态场景中，平均降低了 92.04%，在低动态场景中也有显著改善，达到 19.48%。此外，通过优化实时性和轻量化的网络设计，系统在标准硬件平台上也能高效运行。

本文的研究成果为动态场景下的 SLAM 提供了一种高效可靠的解决方案，不仅增强了系统的鲁棒性，还能有效减少动态物体对定位精度的影响。未来的研究将集中在进一步优化实时性、探索更稳健的特征检测方法，以及结合多传感器融合技术，以提升系统在更复杂环境中的适应性和精度。

参考文献

[1] Davison, A.J., Reid, I.D., Molton, N.D. and Stasse, O. (2007) MonoSLAM: Real-Time Single Camera SLAM. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **29**, 1052-1067. https://doi.org/10.1109/tpami.2007.1049

[2] Engel, J., Koltun, V. and Cremers, D. (2018) Direct Sparse Odometry. *IEEE Transactions on Pattern Analysis and*

Machine Intelligence, **40**, 611-625. https://doi.org/10.1109/tpami.2017.2658577

[3] Mur-Artal, R., Montiel, J.M.M. and Tardos, J.D. (2015) ORB-SLAM: A Versatile and Accurate Monocular SLAM System. *IEEE Transactions on Robotics*, **31**, 1147-1163. https://doi.org/10.1109/tro.2015.2463671

[4] Mur-Artal, R. and Tardos, J.D. (2017) ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras. *IEEE Transactions on Robotics*, **33**, 1255-1262. https://doi.org/10.1109/tro.2017.2705103

[5] Engel, J., Schöps, T. and Cremers, D. (2014) LSD-SLAM: Large-Scale Direct Monocular SLAM. In: Fleet, D., Pajdla, T., Schiele, B. and Tuytelaars, T., Eds., *Computer Vision—ECCV* 2014, Springer, 834-849. https://doi.org/10.1007/978-3-319-10605-2_54

[6] Forster, C., Pizzoli, M. and Scaramuzza, D. (2014) SVO: Fast Semi-Direct Monocular Visual Odometry. 2014 *IEEE International Conference on Robotics and Automation* (*ICRA*), Hong Kong, 31 May-7 June 2014, 15-22. https://doi.org/10.1109/icra.2014.6906584

[7] Klappstein, J., Barth, A., Franke, U. and Maurer, M. (2006) Detecting Moving Objects in Car Environment by Motion Analysis and Ego-Motion Compensation. *IEEE Intelligent Vehicles Symposium*, Tokyo, 13-15 June 2006, 682-687.

[8] Yu, C., Liu, Z., Liu, X., Xie, F., Yang, Y., Wei, Q., *et al.* (2018) DS-SLAM: A Semantic Visual SLAM Towards Dynamic Environments. 2018 *IEEE/RSJ International Conference on Intelligent Robots and Systems* (*IROS*), Madrid, 1-5 October 2018, 1168-1174. https://doi.org/10.1109/iros.2018.8593691

[9] Bescos, B., Facil, J.M., Civera, J. and Neira, J. (2018) DynaSLAM: Tracking, Mapping, and Inpainting in Dynamic Scenes. *IEEE Robotics and Automation Letters*, **3**, 4076-4083. https://doi.org/10.1109/lra.2018.2860039

[10] Redmon, J. and Farhadi, A. (2018) YOLOv3: An Incremental Improvement. arXiv: 1804.02767.

[11] Cheng, J., Wang, C. and Meng, M.Q. (2020) Robust Visual Localization in Dynamic Environments Based on Sparse Motion Removal. *IEEE Transactions on Automation Science and Engineering*, **17**, 658-669. https://doi.org/10.1109/tase.2019.2940543

[12] Sun, Y., Liu, M. and Meng, M.Q. (2017) Improving RGB-D SLAM in Dynamic Environments: A Motion Removal Approach. *Robotics and Autonomous Systems*, **89**, 110-122. https://doi.org/10.1016/j.robot.2016.11.012

[13] Jin, J., Jiang, X., Yu, C., Zhao, L. and Tang, Z. (2023) Dynamic Visual Simultaneous Localization and Mapping Based on Semantic Segmentation Module. *Applied Intelligence*, **53**, 19418-19432. https://doi.org/10.1007/s10489-023-04531-6

[14] Wu, W., Guo, L., Gao, H., You, Z., Liu, Y. and Chen, Z. (2022) YOLO-SLAM: A Semantic SLAM System towards Dynamic Environment with Geometric Constraint. *Neural Computing and Applications*, **34**, 6011-6026. https://doi.org/10.1007/s00521-021-06764-3