Modeling and Simulation 建模与仿真, 2024, 13(4), 4180-4194
Published Online July 2024 in Hans. https://www.hanspub.org/journal/mos
https://doi.org/10.12677/mos.2024.134379

Hans汉斯

基于大场景下识别动态物体的视觉SLAM研究

李兴州[1]，何　锋[1*]，余国宽[2]

[1]贵州大学机械工程学院，贵州　贵阳
[2]贵州师范大学机械与电气工程学院，贵州　贵阳

收稿日期：2024年6月10日；录用日期：2024年7月3日；发布日期：2024年7月10日

摘　要

目前轮式机器人基于YOLOv5的视觉SLAM算法并没有融合IMU解决大场景下鲁棒性较差的问题。利用目前较为先进的YOLOv7动态目标检测算法，将其与ORB_SLAM3算法的IMU数据融合，根据其检测的特征点判断是否为动态目标，保证其在动态场景下稳定工作。通过仿真分析，在TUM-VI数据集的动态场景下剔除动态目标特征点稳定，在Mono-IMU和Stereo-IMU两种模式下绝对误差精度平均提高30%以上，相对误差精度平均提升20%以上。因此，本文所采用的方法在大场景且存在动态物体的情况下提高了定位与建图精度。

关键词

视觉SLAM，动态场景，YOLOv7，动态目标点

Visual SLAM Research Based on Recognizing Dynamic Objects in Large Scenes

Xingzhou Li[1], Feng He[1*], Guokuan Yu[2]

[1]School of Mechanical Engineering, Guizhou Universitv, Guiyang Guizhou
[2]School of Mechanical and Electrical Engineering, Guizhou Normal University, Guiyang Guizhou

Received: Jun. 10th, 2024; accepted: Jul. 3rd, 2024; published: Jul. 10th, 2024

Abstract

The current YOLOv5 vision-based SLAM algorithm for wheeled robots does not fuse IMU to solve the problem of poor robustness in large scenes. Using the more advanced YOLOv7 dynamic target detection algorithm, it fuses the data with the IMU data of the ORB_SLAM3 algorithm to determine

*通讯作者。

文章引用: 李兴州, 何锋, 余国宽. 基于大场景下识别动态物体的视觉 SLAM 研究[J]. 建模与仿真, 2024, 13(4): 4180-4194.
DOI: 10.12677/mos.2024.134379

whether it is a dynamic target or not according to its detected feature points, which ensures that it can work stably in dynamic scenes. Through simulation analysis, the rejection of dynamic target feature points is stabilized under the dynamic scene of the TUM-VI dataset, and the absolute error accuracy is improved by more than 30% on average and the relative error accuracy is improved by more than 20% on average in both Mono-IMU and Stereo-IMU modes. Therefore, the method adopted in this paper improves the localization and map building accuracy in large scenes with the presence of dynamic objects.

Keywords

Visual SLAM, Dynamic Scenes, YOLOv7, Dynamic Target Points

1. 引言

移动机器人在未知环境下利用传感器检测周边数据，完成对周围环境的定位与构建称为即时定位与地图构建(Simultaneous Localization And Mapping, SLAM) [1]。

在传统的 SLAM 方案中，诸如 RANSAC [2] (Random sample consensus)、鲁棒核函数等用于处理外点的算法在应对动态目标时可发挥一定作用，但是此类方法对于当有动态目标占据图像大部分区域时或大部分纹理特征位于动态目标上时效果较差。现有大多数动态场景 SLAM 算法仅依据视觉信息进行工作，其方案大致可分为基于几何、基于光流和基于语义三种。

(1) 几何约束：Abhijit Kundu [3]等基于多视图几何对特征进行静止或动态判断，该方法定义了两个约束并使用贝叶斯滤波器得出特征为静态或动态的概率，进一步提升判断的鲁棒性。Wang [4]等首先利用对极几何进行匹配外点的排除，再根据深度相机获取深度信息进行聚类，从而对运动目标进行识别。然而，这些方法无法保证实时性的要求。

(2) 光流约束：Y. Fang、Wang 和 Sun Y [5]-[7]等人均采用稠稠密光流及其改进方法对场景进行运动分割，再基于静态部分进行相机位姿求解。Xu [8]等人提出的 MID-Fusion 可以实现多实例分割，通过对图像的每一帧用 Mask R-CNN 进行实例分割和边框的精修，最后完成相机的位姿估计。Long [9]等人提出的 PSPNet-SLAM 使用一个金字塔结构的 PSPNet 和误差补偿矩阵来实现动态点的检测。Ji [10]等人提出的系统能够检测已知和未知的移动物体，此外还提出了一个有效几何模块，通过将深度图像聚类到几个区域并通过它们的重投影误差识别动态区域来检测未知的运动对象。Zhang [11]等人提出了一种基于学习的快速的人体对象检测器，用于隔离动态人体对象的同时并实现稠密建图。Zhang [12]等人提出了一种新的视听融合方法，将声源方向融合到 RGB-D 图像中，从而消除动态障碍物对多机器人 SLAM 系统的影响。Hu [13]等人提出了一种称为 DYP-SLAM 的视觉 SLAM 算法，该算法使用目标检测框架检测环境中的移动对象，并给予一定的概率来判断是否为动态目标，同时结合语义约束和几何约束计算物体、关键点和地图点的静态概率，并将它们作为权重参与相机位姿估计。虽然上述方法提高了 SLAM 精度，但是由于光流法基于光度不变性假设过于严格，在现实中无法保证。

(3) 语义约束：随着深度学习的研究不断深入，部分学者开始考虑通过利用语义信息来提升 SLAM 算法在动态场景下的表现。Xiao [14]提出一种结合语义信息的单目视觉 SLAM 算法 Dynamic-SLAM，该

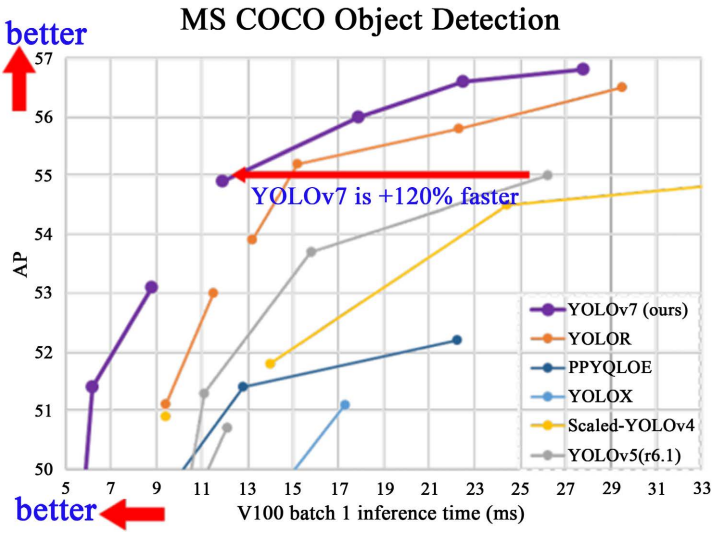
方法基于 ORB-SLAM2 框架，利用 SSD (Single Shot Multi Box Detector) [15]得出目标的类型与位置，使用基于相邻帧速度不变性的漏检补偿算法及选择跟踪算法，改善了原有框架在动态环境下的表现。Ruidong Zhang [16]利用几何约束融合改进的 YOLOv5 提高了精度，Mseddi [17]通过改进 YOLO5 的网格来提高检测精度。

由于 YOLOv7 目标网络检测模块使用了更快的卷积操作和更小的模型[18]，因此在相同的计算资源下可以达到更高的检测速度，此外，YOLOv7 还提供了较高的识别精度，能够检测更多的细粒度对象。因此本文选择 YOLOv7 目标检测网络融合到 ORB_SLAM3 [19]中，以满足对动态场景下的动态物体检测及剔除，提高其在动态场景下定位与建图的性能。

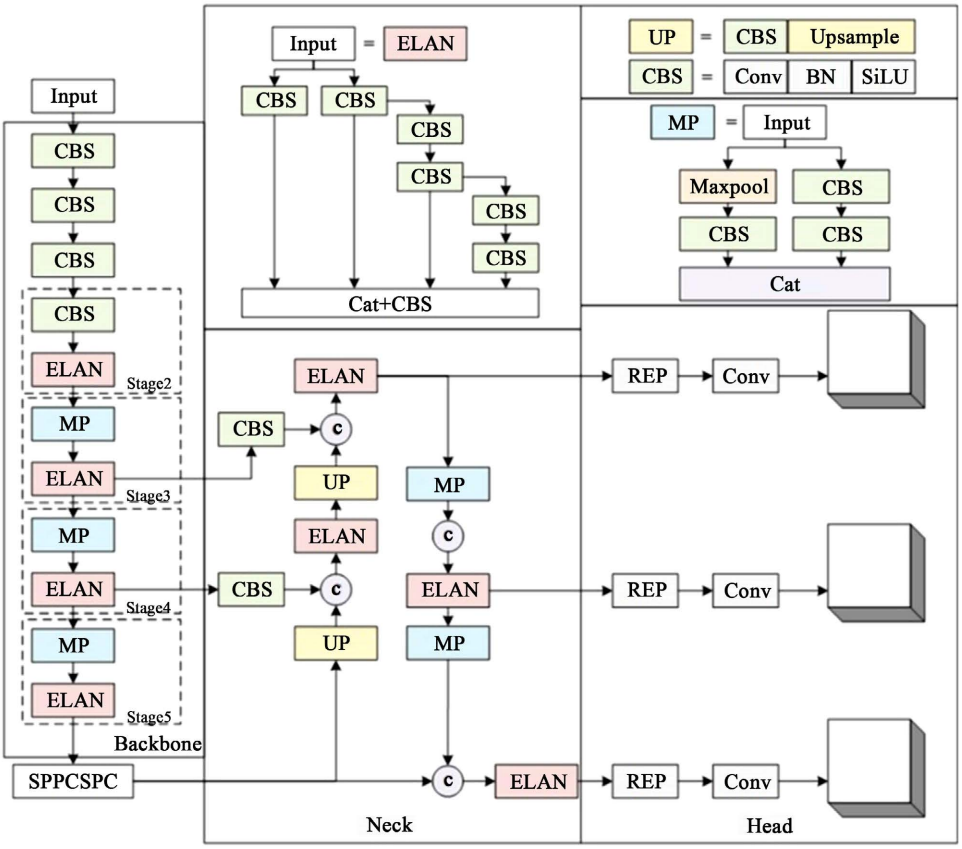
本文针对动态场景中进行研究，通过在 ORB-SLAM3 的跟踪线程上加入 YOLOv7 目标检测模块进行检测，并提取图像的语义信息与 ORB 特征相结合，通过目标检测算法预测动态区域并剔除动态特征点，保留的特征点用于跟踪匹配，以获得更准确的相机位姿估计。

2. YOLOv7 目标检测算法

YOLOv7 是一种目标检测算法，用于对象识别和定位，其模型处理速度和精度相较于其他算法有明显的优势(如图 1(a)所示)。它由 Input、Backbone、Head 和 Neck 组成(如图 1(b)所示)。Input 部分对输入的图片使用 Mosaic 数据增强、自适应锚框计算等方法，选取 4 张图片通过缩放、裁减、随机排布的方式对图片进行拼接、缩放等操作组合成新的图片，以达到增加图片中小目标数量、丰富目标检测数据集的目的，使网络鲁棒性更好；Backbone 部分是由 CSB 模块、ELAN 模块、MP 模块组合而成，其中 CBS 模块由 Conv 层、BN 层、与 SLU 激活函数构成。其目的主要是对图片进行特征提取；ELAN 模块由多个 CBS 模块分支堆暨构成，通过分支得到的特征层堆叠之后会再经过 CBS 模块进行特征整合，以便充分提取图片的特征信息；MP 模块由 CBS 模块与 Maxpool 块构成，它分为两个分支：左分支是一个步长为 2 × 2 的 Maxpool 和一个卷积大小为 1 × 1 的 CBS 模块，右分支是一个卷积大小为 1 × 1 的 CBS 模块和一个卷积大小为 3.3，步长为 2 × 2 的 CBS 模块，两个分支的结果输出后再进行一次堆叠，对提取的特征进行特征融合；Neck 模块由特征金字塔网络构成，将具有丰富物体信息的浅层特征与具有丰富语义信息的深层特征进行双向特征融合；Head 部分由 REP 块和 CBS 模块构成，将通过特征金字塔网格输出的特征图进行通道数调整，输出三个不同尺寸的结果。



(a) YOLOv7 模型处理速度与精度对比图



(b) YOLOv7 检测算法框架

Figure 1. YOLOv7 target detection algorithm
图 1. YOLOv7 目标检测算法

3. 动态特征点剔除算法设计

在计算机视觉中，语义信息指的是对图像中对象的高级理解，如类别、位置、姿态和形状等，它有助于理解场景和动态目标。在动态场景下，SLAM 可以利用语义信息提高系统的稳定性和建图定位的精度。SLAM 系统通过提取和匹配特征点来初始化位姿，然后使用 RANSAC 方法来排除误匹配和动态点。但是在动态目标过多的情况下，位姿初始化的精度会降低。

3.1. ORB 特征提取原理

ORB-SLAM3 算法[19]由提取特征点和计算描述子两部分组成，分别是 FAST 角点检测法和 BRIEF 二进制特征描述子法。

如图 2 所示，FAST 角点检测法是基于像素灰度差原理完成的。取一像素点 p 为圆心，半径为 R 的圆上按顺时针顺序选择 16 个像素点，并计算 p 与这些像素点的像素灰度差。如果存在 N 个连续像素点与 p 的像素灰度值的差值都大于阈值 t，则确定 p 为一个角点。一般地，N 取值为 9，阈值 t 为 p 点灰度值的 20%。

由于 FAST 检测法提取的角点会因为图像旋转和尺度不一致而导致后续特征点提取无法正确匹配，为解决旋转不变性和尺度不变性的问题，ORB 算法采用灰度质心法和构建图像金字塔法来解决上述问题。

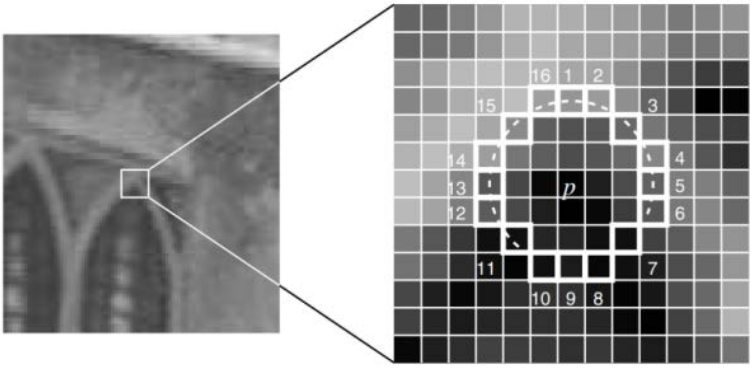


Figure 2. FAST corner extraction map
图 2. FAST 角点提取图

灰度质心法是通过计算图像的质心与形心之间的向量作为 FAST 特征点的主方向，其步骤为：

(1) 定义特征点局部图像块的矩为：

$$m_{pq}=\sum_{x,y}x^{p}y^{q}I(x,y),p,q=\{0,1\} \tag{1}$$

其中，p,q 取 0 或者 1；$I(x,y)$ 表示在该点图像灰度值；m_{pq} 表示图像的矩。在半径为 R 的范围内，沿 x，y 坐标轴方向的图像矩分别为：

$$m_{10}=\sum_{x=-R}^{R}\sum_{y=-R}^{R}xI(x,y) \tag{2}$$

$$m_{01}=\sum_{x=-R}^{R}\sum_{y=-R}^{R}yI(x,y) \tag{3}$$

整个圆形范围内灰度值总和为：

$$m_{00}=\sum_{x=-R}^{R}\sum_{y=-R}^{R}I(x,y) \tag{4}$$

(2) 图像的质心 C 为：

$$C=(c_x,c_y)=\left(\frac{m_{10}}{m_{00}},\frac{m_{01}}{m_{00}}\right) \tag{5}$$

(3) 图像形心 O 指向质心 C 的方向向量 $\overrightarrow{OC}$ 表示关键点的“主方向”，且旋转角度记为：

$$\theta=\arctan 2(c_y,c_x)=\arctan 2(m_{01},m_{10}) \tag{6}$$

然后利用二进制描述子快速找到角点。

3.2. 动态特征点剔除策略

利用上述特征点投影得到像素点，利用特征点法获得基础矩阵 F 后，得到每对特征点对应的极线关系，通过对极几何约束计算每个特征点到其对应极线的距离，判断该点是否为动态特征点。假设两时刻的两个像素特征点 p_1、p_2 是匹配的特征点对，其齐次坐标表示如下式：

$$\begin{cases}p_1=[x_1 \quad y_1 \quad 1]\\ p_2=[x_2 \quad y_2 \quad 1]\end{cases} \tag{7}$$

则 p_1 对应的极线 L 为：

$$L=\begin{bmatrix}X\\Y\\Z\end{bmatrix}=Fp_1 \tag{8}$$

进一步计算 p_2 到极线的距离 D 为：

$$D=\frac{\left|p_2^T F p_1\right|}{\sqrt{\|x\|^2+\|y\|^2}}>\beta \tag{9}$$

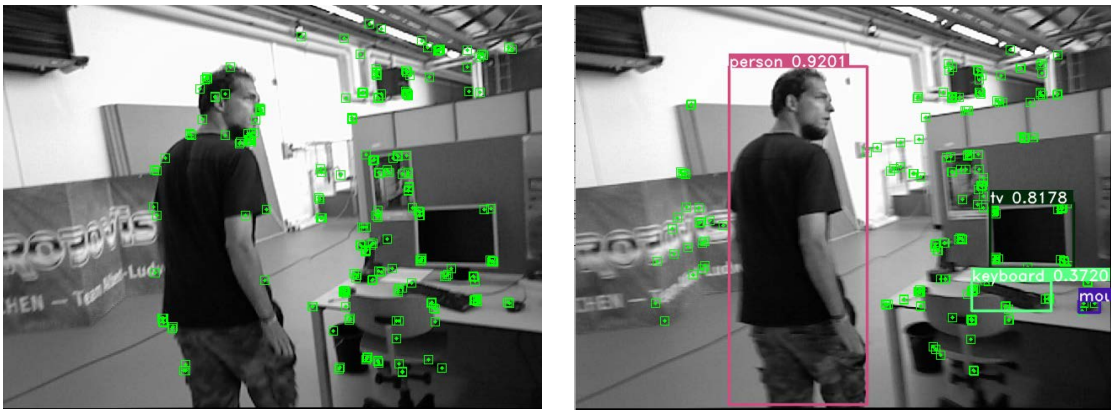
理想状态下点到极线的距离 $D=0$，然而，由于相机获取的图像受周围环境噪声、光线等的影响会产生畸变，故距离 $D\neq 0$。因此可通过设置阈值 β 来进行辅助判断，若 $D>\beta$ 时，则认为该像素点是动态特征点，需进行剔除；反之则认为其是静止的点，予以保留。

4. 仿真结果分析

本研究的仿真基于 Linux 平台，其中系统版本为 Ubuntu18.04，CPU 为 i5-12490F 的处理器、内存为 16GB。在 TUM-VI 数据集上分别进行 ORB_SLAM3 和 ORB_SLAM3_YOLOv7 仿真测试，并以绝对轨迹误差(Absolute Trajectory Error, ATE)和相对误差轨迹(Relative Pose Error, RPE) [20]作为评价两者精度的标准，分别针对大、小场景进行仿真分析。

4.1. 动态物体特征点识别

为确保定位与建图效果，剔除动态特征点的过程中需保留静态特征点。因此，只有当检测到的动态特征点位于高动态物体检测框时，才会进行剔除。在 waiking_rpy、walking_xyz 和室外动态数据集 outdoor4 下的仿真结果显示，图 3(a)展示动态特征点识别前的效果，而图 3(b)展示了识别动态目标点后的效果。由图可知，ORB_SLAM3 在存在大量动态物体的场景中表现受限，因为工作人员的持续行走导致动态特征点对 SLAM 系统的定位效果产生持续影响。引入了 ORB_SLAM3_YOLOv7 后，通过结合 YOLOv7 目标检测算法，系统更有效地应对动态物体的干扰。ORB_SLAM3_YOLOv7 在处理动态场景时能准确识别多个动态目标的特征点，如 person 和 chair，并同时保留一些静态特征点如 tv、mouse、keyboard 等。当在室外动态数据集 outdoor4 下仿真中，在室外场景下多纹理的动态场景下，且由于本次数据集天气阴暗，光线不足导致画面较暗，但是由图观察得到，在室外动态场景且光线不足的情况下，本章提出的 ORB_SLAM3_YOLOv7 依旧能高效的识别动态物体并剔除，同时对于场景中的静态物体提起效果也不错(绿色点为静态特征点，红色框为动态剔除特征)，可见 ORB_SLAM3_YOLOv7 显著提升了 SLAM 系统在动态场景下的定位效果。



waiking_rpy 数据集对比

walking_xyz 数据集对比

outdoor4 数据集对比

(a) 动态特征点识别前　　(b) 动态特征点识别后

Figure 3. Dynamic feature point recognition effect
图 3. 动态特征点识别效果

4.2. 轨迹误差测试

为了进行针对动态场景的识别的仿真测试，选择使用融合 YOLOv7 的 ORB_SLAM3_YOLOv7 算法，并与 ORB_SLAM3 算法进行比较。为保证仿真的通用性，本文采用 TUM-VI 数据集作为测试序列，并使用 ATE 和 RPE 作为评价标准来衡量算法性能。为了评估精度，通过时间戳将真实值和估计值进行对齐，计算图像位姿之间的误差值。最终，ATE 和 RPE 以图表形式输出，以便更加直观地呈现评估结果。

(1) 对其真实轨迹与估计轨迹，得到两者轨迹之间的最小二乘解变换矩阵 S。绝对轨迹误差定义如下：

$$F_i = Q_i^{-1} S P_i \tag{10}$$

其中 P_i 表示 SLAM 系统估计位姿，Q_i 表示轨迹真实位姿。

(2) 完整时间戳内位姿的均方根误差 RMSE 为：

$$\mathrm{RMSE}\left(F_{1:n}\right)=\left(\frac{1}{n}\sum_{i=1}^{n}\left\|\mathrm{trans}\left(F_i\right)\right\|^2\right)^{\frac{1}{2}} \tag{11}$$

其中 trans 为 SLAM 系统得到的平移误差。

(3) 相对位姿误差指的是相隔固定时间差Δ两帧位姿差的精度，意在直接测量里程计的误差。因此第 i 帧的 RPE 定义如下：

$$E_{\mathrm{i}}=\left(Q_i^{-1}Q_{i+\Delta}\right)^{-1}\left(P_i^{-1}P_{i+\Delta}\right) \tag{12}$$

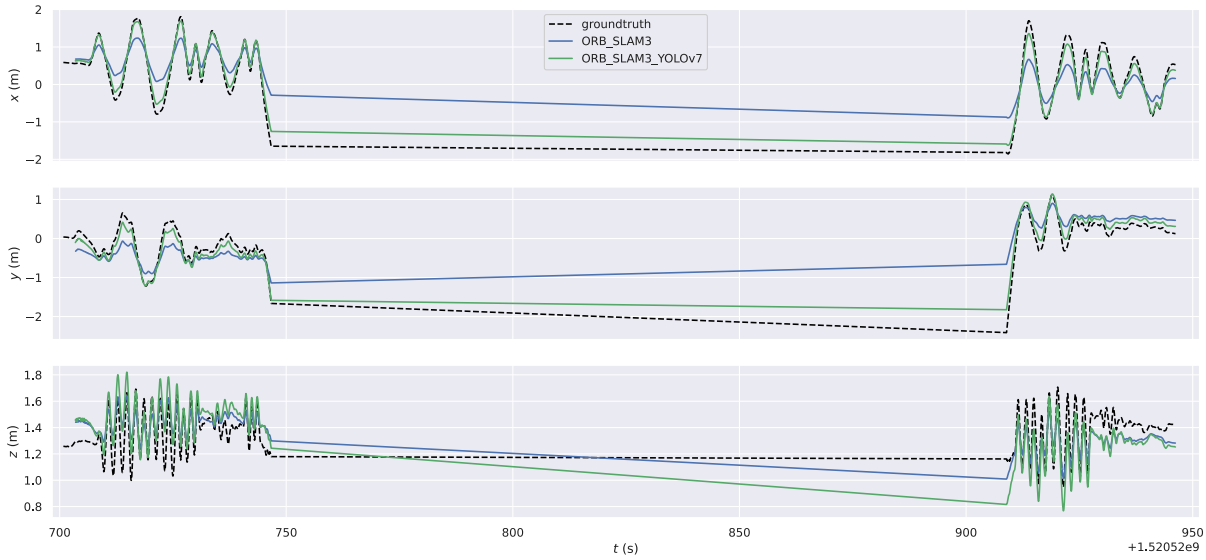
(4) 完整时间戳内位姿的均方根误差 RMSE 为：

$$\mathrm{RMSE}\left(E_{1:m}\right)=\left(\frac{1}{m}\sum_{i=1}^{m}\left\|\mathrm{trans}\left(E_i\right)\right\|^2\right)^{\frac{1}{2}} \tag{13}$$

面对当前主流 ORB_SLAM3_YOLOv5 算法未针对大场景地形融合 IMU 数据，以及大场景中视觉 SLAM 算法在存在动态物体时建图鲁棒性不足的问题，本研究在 ORB_SLAM3 的 IMU 模式基础上融合 YOLOv7，旨在确保大场景下视觉 SLAM 的鲁棒性，同时有效识别动态目标，从而提升定位与建图精度。研究通过对 Mono-IMU 和 Stereo-IMU 两种模式在 TUM-VI 大场景数据集下的仿真分析。

4.2.1. Mono-IMU 模式仿真

由于单目相机会缺乏尺度信息，因此在整个 SLAM 工作过程中，单目模式下的 SLAM 会受到环境的较大影响，尤其是当面对动态物体时这种劣势会被放大，导致整个定位与建图效果明显变差；利用单目相机和 IMU 融合数据虽然可以改善在大场景下受到的影响，但是面对动态物体时依旧没办法很好的完成定位与建图效果。以 slides2 数据集为例，对 ORB_SLAM3 和 ORB_SLAM3_YOLOv7 在 Mono-IMU 模式下分别进行仿真测试得到轨迹误差曲线如图 4 所示，由图观察可得在 Mono-IMU 模式下 ORB_SLAM3 在 x、y、z 三轴方向的误差偏离明显，整体未出现与真实轨迹吻合的部分，且在 750~920 s 时间段 x 轴方向上出现约 1.2 m 的平均偏差，y 轴上出现约 1.1 m 的平均偏差，而 ORB_SLAM3_YOLOv7 整个 SLAM 工作过程中更加靠近真实值，且在 750~920 s 时间段出现约 0.3 m 的平均偏差，相较于 ORB_SLAM3 精度提高了 75%，在 y 轴上出现了约 0.4 m 的平均偏差，相较于 ORB_SLAM3 精度提高了 64%。



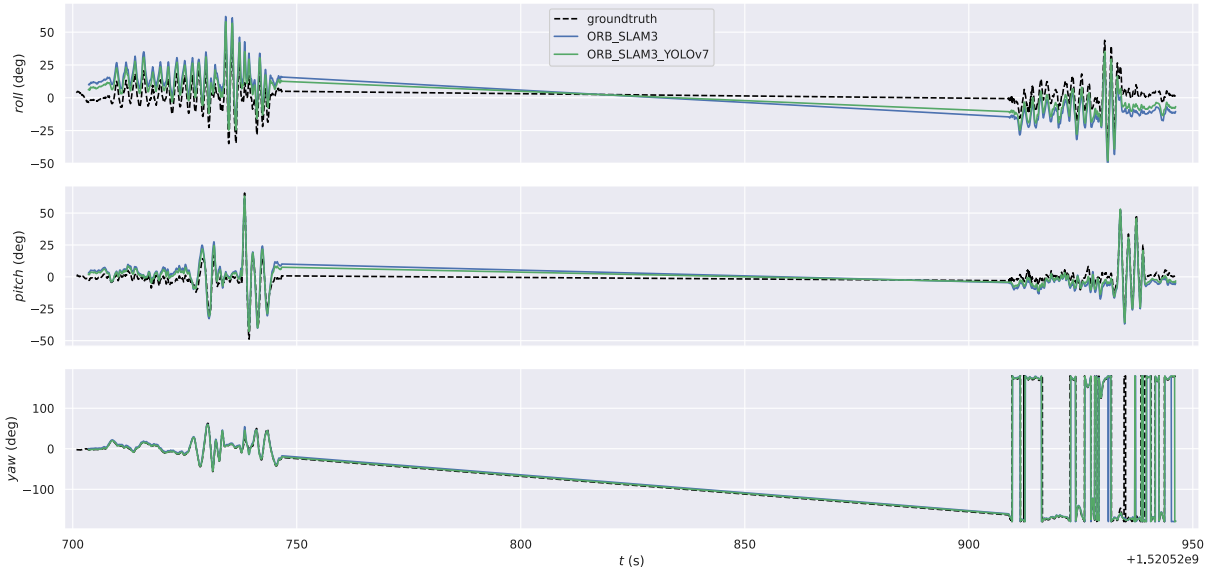
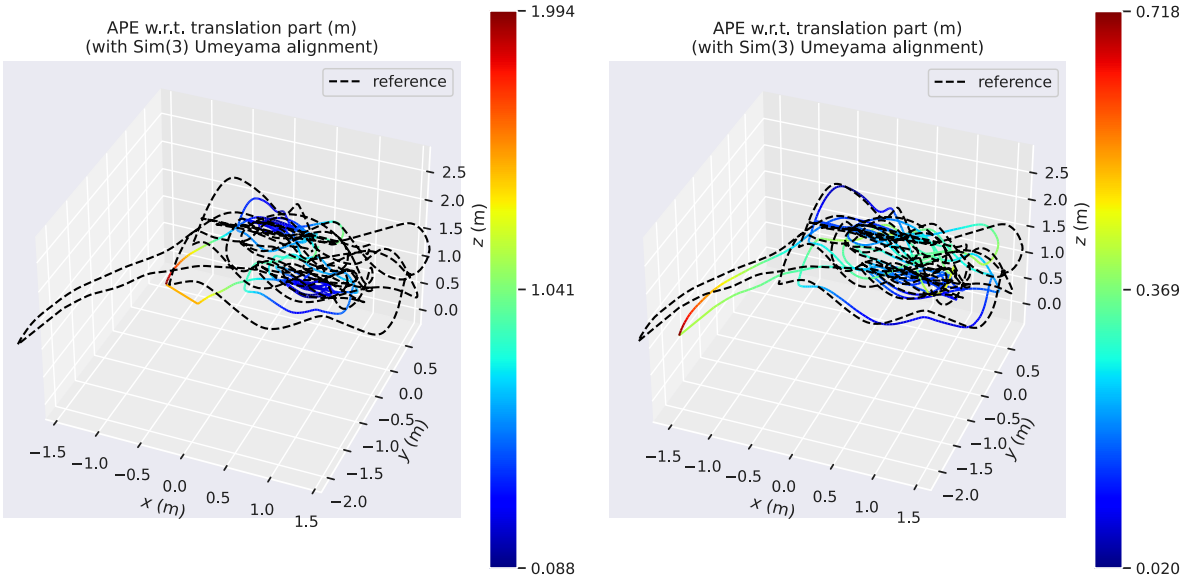


Figure 4. Trajectory error curve
图 4. 轨迹误差曲线

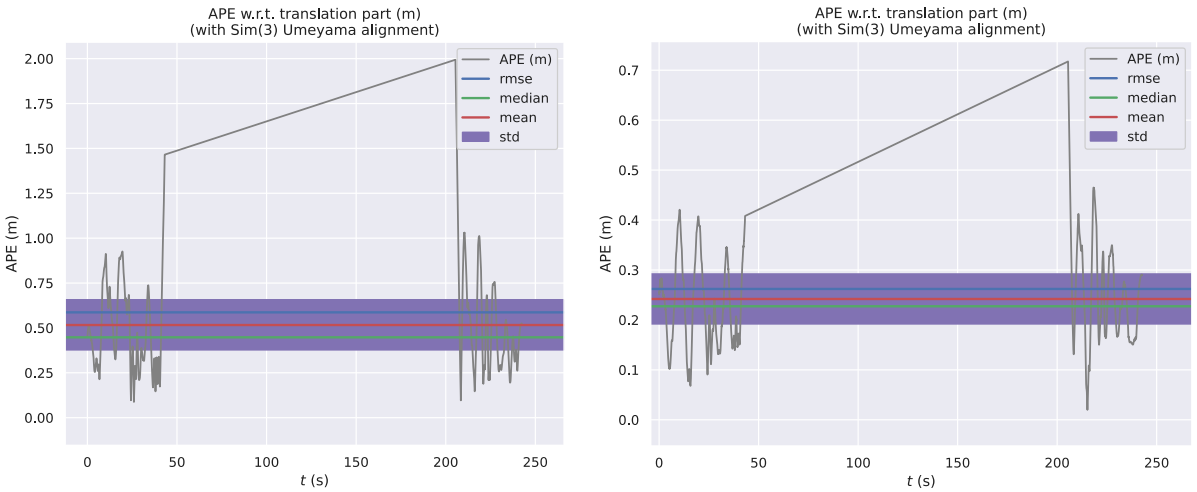
将两种方法的算法对大场景 slides2 动态数据集进行全局仿真得到轨迹曲线图如图 5 所示，图 5(a)为 ORB_SLAM3 算法的轨迹曲线图，图中可以观察出在整个 SLAM 工作过程由于无法剔除动态特征点导致整体的定位精度受到这些动态特征点的严重影响，轨迹曲线与真实值偏离较大，出现较大的误差，最大达到 1.994 m；图 5(b)为 ORB_SLAM3_YOLOv7 算法的轨迹曲线图，整个 SLAM 工作过程中紧跟真实轨迹，且最大偏差为 0.718 m，相较于 ORB_SLAM3 算法精度提高 64%，并且由于单目且动态特征点造成缺乏尺度影响了 ORB_SLAM3 算法轨迹的尺度信息，但是 ORB_SLAM3_YOLOv7 算法利用剔除动态特征点的优势，保留静态特征点信息，使其具备与真实值近似的尺度信息。



(a) (b)

Figure 5. Trajectory curve graph
图 5. 轨迹曲线图

在图 6 中，本章节对比了两种算法在整个 SLAM 工作过程中的相对误差分布情况，图 6(a)为 ORB_SLAM3 算法的全局误差分布情况，整体的误差波动区间为 0.00~2.00 m，且误差分布出现明显的波动，因此其定位精度较差；图 6(b)为 ORB_SLAM3_YOLOv7 算法的全局误差分布情况，整体误差波动区间为 0.0~0.7 m，相对于 ORB_SLAM3 算法误差分布更小且密集，rmse 用于衡量预测模型在连续性数据上的预测精度的指标，其面积越小说明 SLAM 的定位精度越高，由图观察得 ORB_SLAM3_YOLOv7 算法的 rmse 相较于 ORB_SLAM3 算法约提高 60%，说明 ORB_SLAM3_YOLOv7 算法的定位精度相较于 ORB_SLAM3 算法提高了 60%。



(a) (b)

Figure 6. Error distribution graph
图 6. 误差分布图

为了减少仿真随机性引发的误差，对三组数据集 room1、slide2 和 corridor4 进行了 10 次的 Mono-IMU 模式仿真测试，并取得平均值作为最终的仿真结果。同时对比了 ORB-SLAM3 和 ORB-SLAM3_YOLOv7 算法在 Mono-IMU 模式下的定位与建图效果，将结果整理在表 1 (绝对轨迹误差结果对比)和表 2 (相对轨迹误差结果对比)。从表中提升效率的数据来看，在 APE 这两种误差对比中，可以明显观察到，在 Mono-IMU 模式下，ORB_SLAM3_YOLOv7 算法在 RMSE 和 STD 的精度方面均取得了显著的提升，尤其在场景变化时，精度提高幅度超过 55%；在 RPE 这两种误差对比中，可以明显观察到，在 Mono-IMU 模式下，ORB_SLAM3_YOLOv7 算法在 RMSE 和 STD 的精度方面均取得了显著的提升，精度提高幅度均超过 55%，当在多个场景变化下如 corridor4 数据集下，精度更是提高了 87%。这结果明确表明，ORB_SLAM3_YOLOv7 算法在大场景下具有更高的精度，提供更卓越的建图与定位效果，更加符合实际应用的需求。

Table 1. Comparison of Mono-IMU absolute trajectory error (ATE) results (mm)
表 1. Mono-IMU 绝对轨迹误差(ATE)结果对比(mm)

数据集	ORB_SLAM3		ORB_SLAM3_YOLOv7		提升效率(%)	
	RMSE	STD	RMSE	STD	RMSE	STD
room1	29.860	28.269	11.573	9.118	61.24	67.75
slide2	586.603	278.475	261.933	100.360	55.35	63.96
corridor4	100.607	68.857	72.750	18.378	27.69	73.31

Table 2. Comparison of Mono-IMU relative trajectory error (RTE) results (mm)
表 2. Mono-IMU 相对轨迹误差(RTE)结果对比(mm)

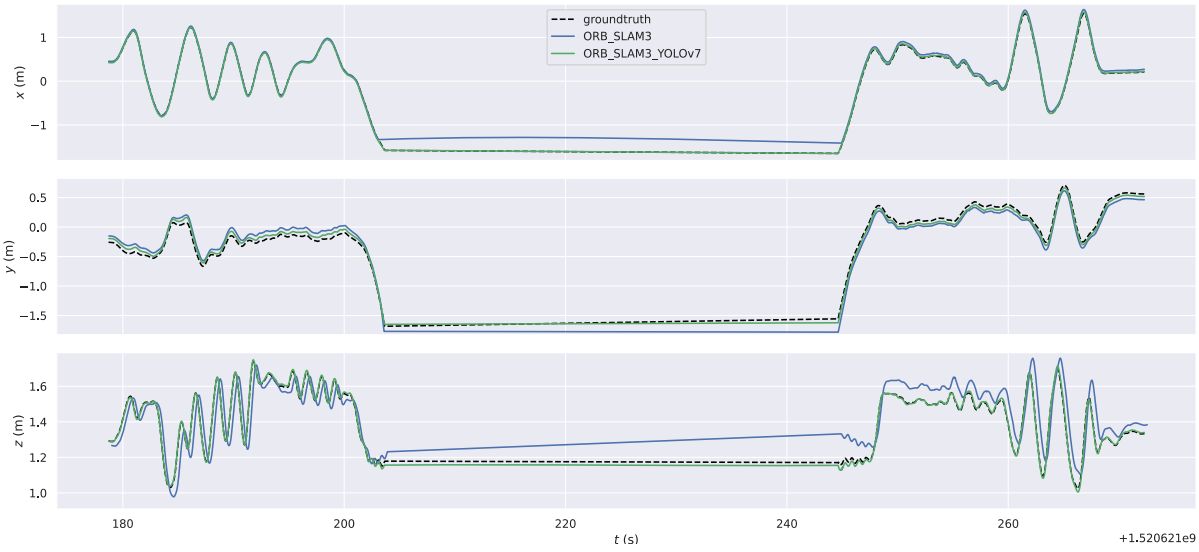
数据集	ORB_SLAM3		ORB_SLAM3_YOLOv7		提升效率(%)	
	RMSE	STD	RMSE	STD	RMSE	STD
room1	21.761	21.252	9.688	8.988	55.48	57.71
slide2	40.275	33.404	16.653	15.250	68.65	54.35
corridor4	87.659	86.886	11.256	10.797	87.19	87.57

4.2.2. Stereo-IMU 模式仿真评估

由于双目相机可以通过左右两幅图像得到尺度信息，因此在面对大多数静态场景时定位与建图精度可以得到保证，但是面对动态场景时，依旧会被动态特征点影响定位的问题，而且大场景下缺乏鲁棒性，在 SLAM 工作中容易丢失前期建立的地图信息导致后续建图效果下降，因此利用双目相机和 IMU 融合的信息可以提高在大场景下的鲁棒性。

以 corridor4 数据集为例，对 ORB_SLAM3 和 ORB_SLAM3_YOLOv7 在 Stereo-IMU 模式下分别进行仿真测试得到轨迹误差曲线如图 7 所示，由图观察可得 Stereo-IMU 模式下 ORB_SLAM3 在 x、y、z 三轴方向的误差偏离明显，整体基本与真实轨迹吻合程度较差，在 210~250 s 时间段 x 轴方向上出现约 0.3 m 的平均偏差，y 轴上基本与真实值偏离，且在同一时间段出现约 0.1 m 的平均偏差，z 轴上更是达到约 0.4 m 的平均偏差，而 ORB_SLAM3_YOLOv7 整个 SLAM 工作过程中更加靠近真实值，在任何时间段几乎无偏差，相较于 ORB_SLAM3 精度提高了 20%；在 210~250 s 时间段 pitch 航向角 ORB_SLAM3 出现约 8 度的偏差，而 ORB_SLAM3_YOLOv7 在 r、p、y 三种航向角几乎与真实值一致。

将两种方法的算法对大场景 corridor4 动态数据集进行全局仿真得到轨迹曲线图如图 8 所示，图 8(a)为 ORB_SLAM3 算法的轨迹曲线图，图中可以观察出在整个 SLAM 工作过程由于无法剔除动态特征点导致整体的定位精度受到这些动态特征点的严重影响，轨迹曲线与真实值偏离较大，出现较大的误差，最大达到 0.212 m；图 8(b)为 ORB_SLAM3_YOLOv7 算法的轨迹曲线图，整个 SLAM 工作过程中紧跟真实轨迹，且最大偏差为 0.094 m，相较于 ORB_SLAM3 算法精度提高 56%。



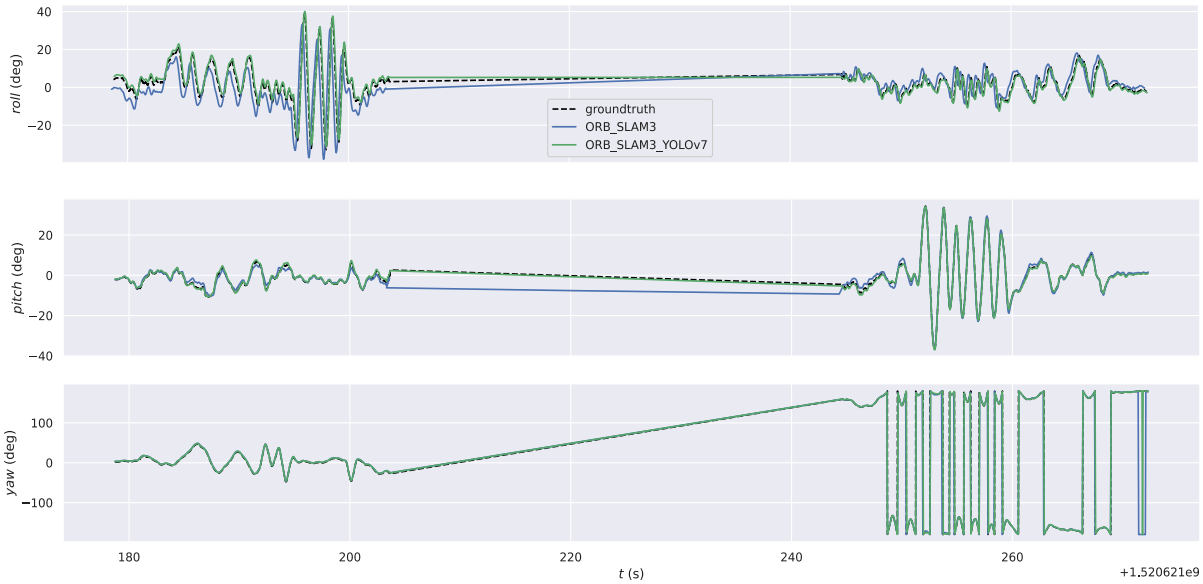
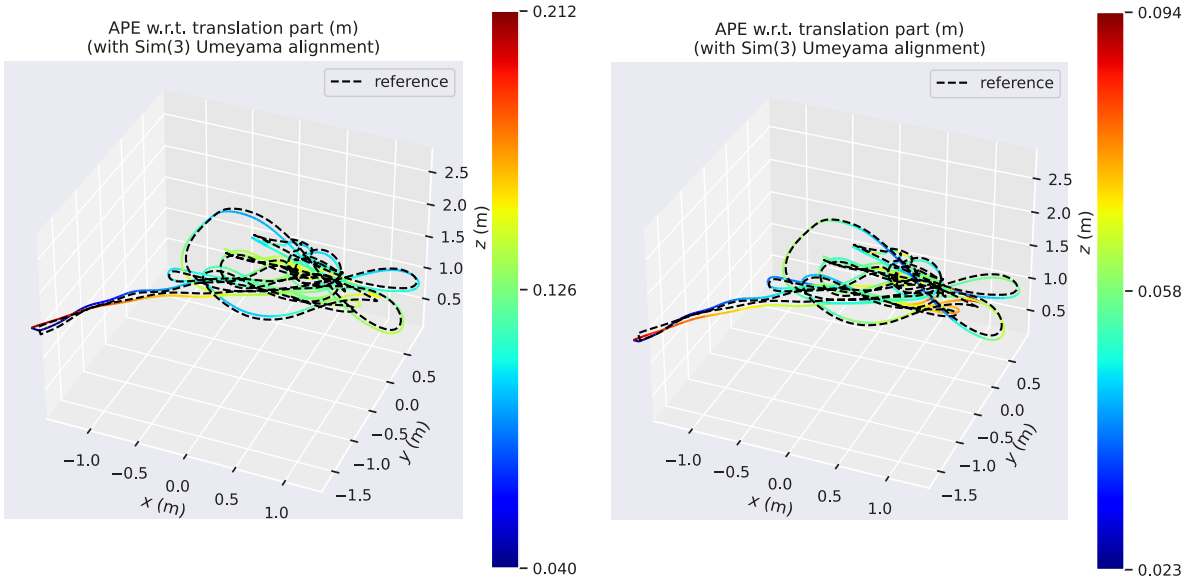


Figure 7. Trajectory error curve
图 7. 轨迹误差曲线



(a) (b)

Figure 8. Trajectory curve graph
图 8. 轨迹曲线图

在图 9 中，本章节对比了两种算法在整个 SLAM 工作过程中的绝对误差分布情况，图 9(a)为 ORB_SLAM3 算法的全局误差分布情况，整体的误差波动区间为 0.04~0.2 m，图 9(b)为 ORB_SLAM3_YOLOv7 算法的全局误差分布情况，整体误差波动区间为 0.02~0.09 m，相对于 ORB_SLAM3 算法误差分布更小且密集；rmse 用于衡量预测模型在连续性数据上的预测精度的指标，其面积越小说明 SLAM 的定位精度越高，由图观察得 ORB_SLAM3_YOLOv7 算法的 rmse 相较于 ORB_SLAM3 算法约提高 90%，说明 ORB_SLAM3_YOLOv7 算法的定位精度相较于 ORB_SLAM3 算法提高了 90%。

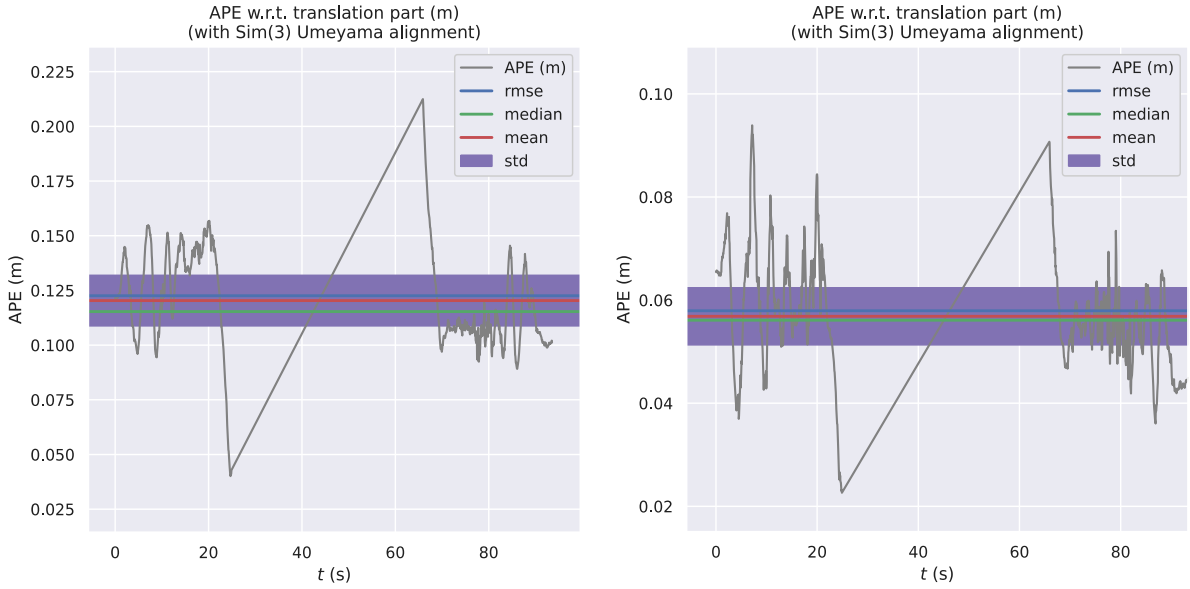


Figure 9. Error distribution graph
图 9. 误差分布图

为了减少仿真随机性引发的误差，对三组数据集 room1、slide2 和 corridor4 进行了 10 次的 Stereo-IMU 模式仿真测试，并取得平均值作为最终的仿真结果。同时对比了 ORB-SLAM3 和 ORB-SLAM3_YOLOv7 算法在 Stereo-IMU 模式下的定位与建图效果，将结果整理在表 3 (绝对轨迹误差结果对比)和表 4 (相对轨迹误差结果对比。表中提升效率的数据来看，在 APE，明显可以观察到，Stereo-IMU 模式下，ORB-SLAM3-YOLOv7 算法在 RMSE 和 STD 的精度方面都取得了显著提升，尤其在室外场景(corridor4)时，

Table 3. Comparison of Stereo-IMU absolute trajectory error (ATE) results (mm)
表 3. Stereo-IMU 绝对轨迹误差(ATE)结果对比(mm)

数据集	ORB_SLAM3		ORB_SLAM3_YOLOv7		提升效率(%)	
	RMSE	STD	RMSE	STD	RMSE	STD
room1	7.331	3.615	6.702	3.063	8.58	15.23
slide2	269.299	60.971	178.406	59.174	33.75	2.95
corridor4	122.559	23.109	57.928	11.089	52.73	52.01

Table 4. Comparison of Stereo-IMU relative trajectory error (RTE) results (mm)
表 4. Stereo-IMU 相对轨迹误差(RTE)结果对比(mm)

数据集	ORB_SLAM3		ORB_SLAM3_YOLOv7		提升效率(%)	
	RMSE	STD	RMSE	STD	RMSE	STD
room1	3.666	2.908	3.119	2.353	14.92	19.09
slide2	14.949	14.613	12.218	11.434	18.27	21.75
corridor4	7.217	6.936	3.838	3.299	46.82	52.44

精度提高幅度超过 50%；在 RPE 明显可以观察到，Stereo-IMU 模式下，ORB-SLAM3-YOLOv7 算法在 RMSE 和 STD 的精度方面都取得了显著提升，尤其在室外场景(corridor4)时，精度提高幅度超过 46%。表明 Stereo-IMU 模式下的 ORB-SLAM3-YOLOv7 算法在大场景中具有更高的精度，提供更好的建图与定位效果，更符合实际应用的需求。

5. 结论与分析

针对视觉 SLAM 算法在动态场景下定位与建图效果较差的情况，本研究选择 YOLOv7 网络检测算法融合得到的 ORB_SLAM3_YOLOv7 算法可提高视觉 SLAM 在动态场景下的定位与地图构建效果以及大场景下的鲁棒性，在 TUM 和 TUM-VI 数据集上进行仿真分析，得到以下结论：

(1) 在动态而小尺度的场景中，ORB_SLAM3_YOLOv7 算法在处理动态物体方面表现突出，明显减少了特征点受动态物体干扰的情况。相对于当前主流的 ORB_SLAM3_YOLOv5 算法，其精度提高超过 25%，与 ORB_SLAM3 算法相比，精度提高超过 80%，并且生成的轨迹更接近真实轨迹。

(2) 在具有动态目标的场景中，本本提出的 ORB_SLAM3_YOLOv7 算法融合 imu 数据，表现出更强大的鲁棒性。相对于传统 ORB_SLAM3 算法，它实现了显著的定位精度提升，特别是在开放室外环境中，其精度改善可超过 50%。

基金项目

国家自然科学基金(52262044)，黔科合支撑[2023]一般 400。

参考文献

[1] 程俊廷, 郭博洋, 田宽. 改进的 LK 光流法在 SLAM 中的应用[J]. 黑龙江科技大学学报, 2019, 29(6): 736-740.

[2] 杨永刚, 武楚健, 杨正全. 基于融合改进 RANSAC 光流法的无人机视觉 SLAM 研究[J]. 半导体光电, 2023, 44(2): 277-283.

[3] Kundu, A., Krishna, K.M. and Sivaswamy, J. (2009) Moving Object Detection by Multi-View Geometric Techniques from a Single Camera Mounted Robot. 2009 *IEEE/RSJ International Conference on Intelligent Robots and Systems*, St. Louis, 10-15 October 2009, 4306-4312. https://doi.org/10.1109/iros.2009.5354227

[4] Wang, R., Wan, W., Wang, Y. and Di, K. (2019) A New RGB-D SLAM Method with Moving Object Detection for Dynamic Indoor Scenes. *Remote Sensing*, **11**, Article 1143. https://doi.org/10.3390/rs11101143

[5] Fang, Y., Dai, B., *et al*. (2009) An Improved Moving Target Detecting and Tracking Based on Optical Flow Technique and Kalman Filter. 2009 4*th International Conference on Computer Science & Education*, Nanning, 25-28 July 2009, 1197-1202. https://doi.org/10.1109/ICCSE.2009.5228464

[6] Wang, Y. and Huang, S. (2014) Towards Dense Moving Object Segmentation Based Robust Dense RGB-D SLAM in Dynamic Scenarios. 2014 13*th International Conference on Control Automation Robotics & Vision* (*ICARCV*), Singapore, 10-12 December 2014, 1841-1846. https://doi.org/10.1109/icarcv.2014.7064596

[7] Sun, Y., Liu, M. and Meng, M.Q.-H. (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

[8] Xu, B., Li, W., Tzoumanikas, D., Bloesch, M., Davison, A. and Leutenegger, S. (2019) MID-Fusion: Octree-Based Object-Level Multi-Instance Dynamic SLAM. 2019 *International Conference on Robotics and Automation* (*ICRA*), Montreal, 20-24 May 2019, 5231-5237. https://doi.org/10.1109/icra.2019.8794371

[9] Long, X., Zhang, W. and Zhao, B. (2020) PSPNet-SLAM: A Semantic SLAM Detect Dynamic Object by Pyramid Scene Parsing Network. *IEEE Access*, **8**, 214685-214695. https://doi.org/10.1109/access.2020.3041038

[10] Ji, T., Wang, C. and Xie, L. (2021) Towards Real-Time Semantic RGB-D SLAM in Dynamic Environments. 2021 *IEEE International Conference on Robotics and Automation* (*ICRA*), Xi'an, 30 May-05 June 2021, 11175-11181. https://doi.org/10.1109/icra48506.2021.9561743

[11] Zhang, H., Zhang, T., Lam, T.L. and Vijayakumar, S. (2021) PoseFusion2: Simultaneous Background Reconstruction and Human Shape Recovery in Real-time. 2021 *IEEE/RSJ International Conference on Intelligent Robots and Systems* (*IROS*), Prague, 27 September-01 October 2021, 7631-7638. https://doi.org/10.1109/iros51168.2021.9636658

[12] Zhang, T., Zhang, H., Li, X., Chen, J., Lam, T.L. and Vijayakumar, S. (2021) Acousticfusion: Fusing Sound Source Localization to Visual SLAM in Dynamic Environments. 2021 *IEEE/RSJ International Conference on Intelligent Robots and Systems* (*IROS*), Prague, 27 September-1 October 2021, 6868-6875. https://doi.org/10.1109/iros51168.2021.9636585

[13] Hu, X., Zhang, Y., Cao, Z., Ma, R., Wu, Y., Deng, Z., *et al*. (2022) CFP-SLAM: A Real-Time Visual SLAM Based on Coarse-to-Fine Probability in Dynamic Environments. 2022 *IEEE/RSJ International Conference on Intelligent Robots and Systems* (*IROS*), Kyoto, 23-27 October 2022, 4399-4406. https://doi.org/10.1109/iros47612.2022.9981826

[14] Xiao, L., Wang, J., Qiu, X., Rong, Z. and Zou, X. (2019) Dynamic-SLAM: Semantic Monocular Visual Localization and Mapping Based on Deep Learning in Dynamic Environment. *Robotics and Autonomous Systems*, **117**, 1-16. https://doi.org/10.1016/j.robot.2019.03.012

[15] 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

[16] Zhang, R. and Zhang, X. (2023) Geometric Constraint-Based and Improved YOLOv5 Semantic SLAM for Dynamic Scenes. *ISPRS International Journal of Geo-Information*, **12**, Article 211. https://doi.org/10.3390/ijgi12060211

[17] Mseddi, W.S., Sedrine, M.A. and Attia, R. (2021) YOLOv5 Based Visual Localization for Autonomous Vehicles. 2021 29*th European Signal Processing Conference* (*EUSIPCO*), Dublin, 23-27 August 2021, 746-750. https://doi.org/10.23919/EUSIPCO54536.2021.9616354

[18] Wang, C., Bochkovskiy, A. and Liao, H.M. (2023) YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Vancouver, 17-24 June 2023, 7464-7475. https://doi.org/10.1109/cvpr52729.2023.00721

[19] Campos, C., Elvira, R., Rodriguez, J.J.G., M. Montiel, J.M. and D. Tardos, J. (2021) ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial, and Multimap SLAM. *IEEE Transactions on Robotics*, **37**, 1874-1890. https://doi.org/10.1109/tro.2021.3075644

[20] Zhang, Z. and Scaramuzza, D. (2018) A Tutorial on Quantitative Trajectory Evaluation for Visual (-Inertial) Odometry. 2018 *IEEE/RSJ International Conference on Intelligent Robots and Systems* (*IROS*), Madrid, 1-5 October 2018, 7244-7251. https://doi.org/10.1109/iros.2018.8593941