Journal of Sensor Technology and Application 传感器技术与应用, 2025, 13(6), 827-837
Published Online November 2025 in Hans. https://www.hanspub.org/journal/jsta
https://doi.org/10.12677/jsta.2025.136081

Hans汉斯

融合深度与语义信息的双目视觉SLAM

潘长征*，杨　艺

江苏理工学院机械工程学院，江苏　常州

收稿日期：2025年8月29日；录用日期：2025年9月23日；发布日期：2025年10月17日

摘　要

针对动态场景下传统视觉SLAM系统因动态物体干扰导致定位精度下降的问题，本文提出一种融合深度特征与语义信息的双目视觉SLAM优化方法。首先，采用轻量化改进的DeepLabV3+语义分割网络，在保证实时性的同时提升分割精度；其次，设计语义与深度一致性融合的动态掩码构建机制：通过语义分割初筛潜在动态区域，结合双目深度图分析时序深度变化，并引入置信度加权评分及时序一致性判断策略，精准区分动态与静态特征点；最后，在ORB-SLAM3框架中集成上述模块，构建包含跟踪、局部建图与回环检测的完整系统。在Euroc和KITTI数据集上的实验表明：与ORB-SLAM3相比，本文方法在室内场景绝对轨迹误差(ATE)降低8%~15%，室外场景降低7%~35%，显著提升了动态环境下的鲁棒性与定位精度。

关键词

视觉SLAM，语义分割，深度特征，特征提取，动态环境

Binocular Vision SLAM Integrating Depth and Semantic Information

Changzheng Pan*, Yi Yang

School of Mechanical Engineering, Jiangsu University of Technology, Changzhou Jiangsu

Received: August 29, 2025; accepted: September 23, 2025; published: October 17, 2025

Abstract

This paper proposes a binocular visual SLAM optimization method that integrates depth features and semantic information, addressing the issue of reduced positioning accuracy in traditional visual SLAM systems due to interference from dynamic objects in dynamic scenes. Firstly, a lightweight

*通讯作者。

文章引用: 潘长征, 杨艺. 融合深度与语义信息的双目视觉 SLAM [J]. 传感器技术与应用, 2025, 13(6): 827-837.
DOI: 10.12677/jsta.2025.136081

improved DeepLabV3 semantic segmentation network is adopted to enhance segmentation accuracy while ensuring real-time performance; secondly, a dynamic mask construction mechanism based on semantic and depth consistency is designed: it initially filters potential dynamic areas through semantic segmentation, analyzes temporal depth changes using binocular depth maps, and introduces a confidence-weighted scoring and temporal consistency judgment strategy to accurately distinguish between dynamic and static feature points; finally, the above modules are integrated into the ORB-SLAM3 framework to construct a complete system that includes tracking, local mapping, and loop detection. Experiments on the Euroc and KITTI datasets demonstrate that compared to ORB-SLAM3, the proposed method reduces absolute trajectory error (ATE) by 8%~15% in indoor scenes and by 7%~35% in outdoor scenes, significantly enhancing robustness and positioning accuracy in dynamic environments.

Keywords

Visual SLAM, Semantic Segmentation, Depth Features, Feature Extraction, Dynamic Environment

1. 引言

视觉同步定位与建图技术(Visual Simultaneous Localization and Mapping, VSLAM)通过环境感知实现机器人位姿的精准估计。该技术凭借低成本硬件架构与高精度定位特性，已成为自动驾驶、机器人导航等领域的核心感知方案[1]。经典视觉 SLAM 方法(如基于直接法的 LSD-SLAM 和基于特征点的 ORB-SLAM 系列[2])普遍依赖静态环境假设，其在静态场景中表现出优异的鲁棒性，但在动态干扰环境下存在明显局限[3]：特征提取过程对运动目标响应迟滞[4]，且动态物体引发的特征误匹配将导致位姿估计精度急剧恶化。

为提升动态及弱纹理场景的适应性，深度学习驱动的 SLAM 系统通过增强特征表达能力优化定位性能。典型案例如下：DS-SLAM [5]融合 SegNet [6]实现动态目标分割，但仅完成传统 ORB 特征的语义筛选，未挖掘深度特征潜力；DynaSLAM [7]结合 Mask R-CNN [8]与多视图几何检测动态区域，其固定阈值剔除机制造成弱纹理区域静态特征大量丢失；Detect-SLAM [9]采用 YOLO 检测模型与动态概率评估，因缺失语义–几何联合约束导致筛选效能受限；SG-SLAM [10]基于 MobileNetV3 构建动态特征滤除器，虽融合几何约束提升动态点识别率，却在剧烈运动时因位姿漂移引发跟踪失效；RDS-SLAM [11]虽在 ORB-SLAM3 中集成多速率语义分割线程，但纯语义驱动的动态点剔除策略降低了系统准确性。现有方法的核心缺陷在于过度依赖传统手工特征，未能有效协同深度特征与语义信息[12]，造成特征过度剔除与跟踪稳定性下降。

针对上述问题，本文提出基于深度特征点提取与动态语义约束的协同优化框架。通过引入深度特征，结合轻量级目标检测网络提供语义先验剔除动态特征点，实现精准定位和更鲁棒的算法。本文的核心贡献包含三方面：

1、提出轻量化深度语义分割网络架构，将 DeepLabV3+的主干网络替换为 MobileNetV3，显著降低计算负载；在 ASPP 模块后集成 ECA-Net 注意力机制，增强多尺度特征融合能力。该设计在保持较高分割精度的同时，提升推理速度，满足动态 SLAM 实时性需求。

2、建立语义–深度协同的动态掩码生成机制，创新性融合语义先验与深度一致性约束：首先通过语

义分割初筛潜在动态区域，继而基于双目深度图计算时序深度差异，最后提出置信度加权评分模型及时序一致性判决策略，有效解决静态误判问题。

3、构建深度特征增强的双目 SLAM 框架，在 ORB-SLAM3 基础上集成上述模块：跟踪线程采用改进 DeepLabV3+实时输出动态掩码，结合多尺度特征提取剔除动态特征点；局部建图引入深度约束优化位姿估计，回环检测增加语义一致性验证。在 Euroc 数据集和 KITTI 数据集测试，鲁棒性和定位精度有显著的提升。

2. 系统框架

如图 1 所示，本研究构建了深度－语义融合的双目视觉 SLAM 架构，其核心由并行运作的跟踪线程、局部建图线程与闭环修正线程构成。双目输入图像同步馈入语义分割与跟踪流程：在语义分支中，基于视差计算生成深度图以增强场景理解，采用轻量化 DeepLabv3+ (MobileNetV3 主干 ＋ECA-Net 模块)输出语义先验，进而融合深度信息生成动态目标掩码；在跟踪分支中，通过多尺度特征提取与立体匹配获取鲁棒性特征点集，继而依据动态掩码过滤潜在运动特征，最终保留的静态特征点用于位姿估计、后端优化及闭环检测。

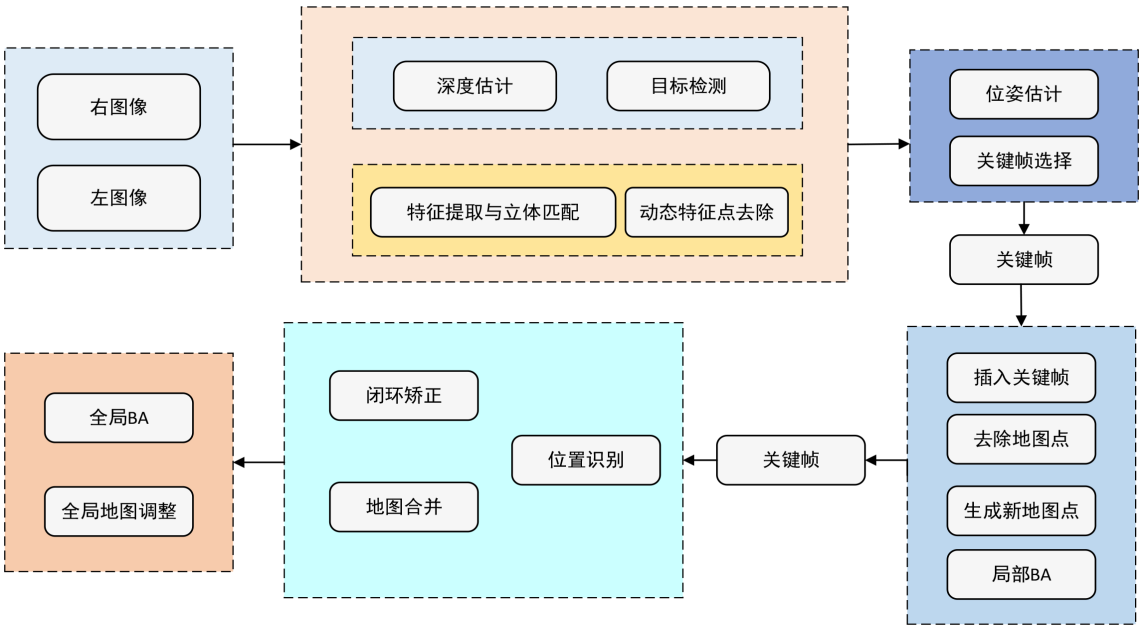


Figure 1. Overall diagram of the algorithm
图 1. 算法整体框图

3. 改进的 Deeplabv3+的语义分割算法

DeepLabV3+采用编码－解码架构[13]，其核心 Xception 主干网络通过深层卷积堆叠实现高精度分割，但高参数量级导致计算负担显著。该模型的关键创新在于空洞空间金字塔池化(ASPP)模块——利用多膨胀率卷积捕获多尺度上下文信息，使输出特征图具备大感受野特性，相较于常规卷积可获得更密集的特征表达。解码器部分衔接 ASPP 输出，采用 1 × 1 卷积进行通道压缩与分辨率恢复[14]，特别强化了复杂场景下的边缘细节重建能力，有效提升分割边界一致性。

如图 2 所示，本研究对 DeepLabV3+进行了轻量化与特征增强双重优化。针对原 Xception 主干网络参数量过大导致的实时性瓶颈，采用基于深度可分离卷积的 MobileNetV3 替代其主干结构[15]，在维持分

割精度的同时显著提升推理速度。此外，针对 ASPP 模块在高膨胀率卷积下引发的细节丢失问题(表现为边缘特征弱化与小目标漏检率上升)，在解码器嵌入 ECA-Net 通道注意力机制。该模块通过自适应通道加权强化多尺度特征融合，有效提升边界分割一致性，且只增加较小的计算开销。

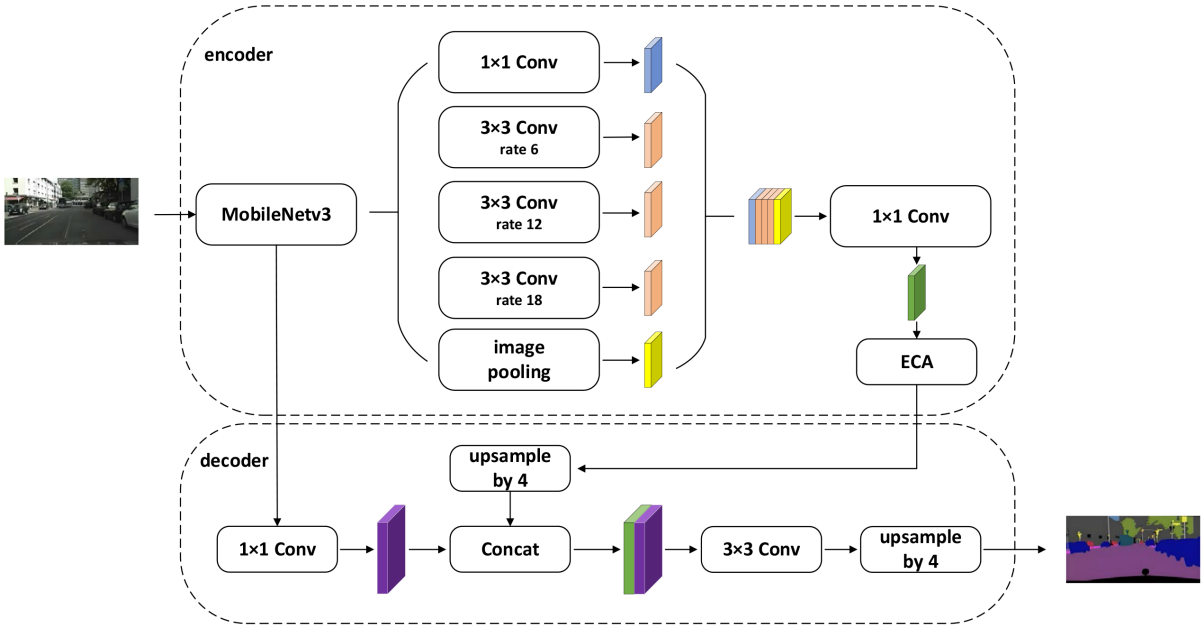


Figure 2. Improved DeepLabV3 model network structure
图 2. 改进的 DeepLabV3+模型网络结构

3.1. MobileNetV3 主干网络

MobileNetV3 作为该系列最新的轻量化主干网络，专为资源受限的边缘设备设计，在特征提取效率与精度间实现优化平衡。其延续 MobileNet 系列[16]的深度可分离卷积架构，显著降低计算复杂度同时保持特征表征能力。核心改进包括：重构 MobileNetV2 的线性逆残差结构，引入 SE 通道注意力机制[17]与 h-swish 激活函数[18]，实现特征通道权重的自适应调整；特征处理流程通过 1 × 1 逐点卷积对输入张量升维，经 3 × 3 深度卷积结合 SE 模块提取特征，再通过 1 × 1 卷积降维并与残差输入叠加，最终由 h-swish 激活函数输出，该设计强化了特征间相关性建模能力。此外，采用神经架构搜索(NAS)优化网络拓扑[19]：以 3 × 3 卷积层初始化特征提取，堆叠改进型逆残差模块构建多层次特征表达，并通过输出层精简降低计算开销。本文将其作为 DeepLabV3+的主干替代方案，有效缩减模型复杂度并提升特征提取效率。

3.2. ECA-Net 模块

ECA-Net [20]作为 SE 模块的演进形式，构建了高效通道注意力机制。其核心创新在于规避 SE 模块的全连接降维操作，通过保留原始通道维度优化注意力权重计算。该设计显著抑制非关键区域响应，强化对判别性特征的聚焦能力，从而提升特征选择精度与模型识别效能。本文将其集成于 ASPP 输出层之后，通过增强多尺度特征的通道交互效率，显著改善分割边界一致性(结构详见图 3)。

4. 语义与深度一致性的动态掩码构建方法

在动态环境下，传统基于静态假设的视觉 SLAM 系统往往由于动态物体的干扰导致特征点匹配失败、位姿估计误差增大[21]，严重影响系统鲁棒性与精度。为有效解决这一问题，本文提出一种融合语义信息

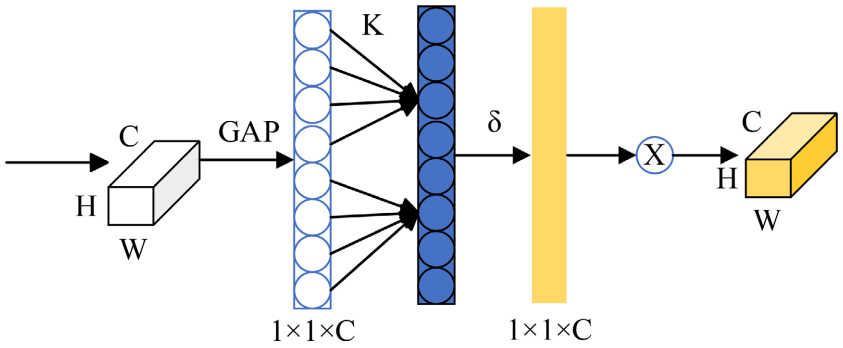


Figure 3. ECA-Net module structure
图 3. ECA-Net 模块结构

与深度一致性的动态掩码构建方法，真实的动态物体不仅在语义上具有“可疑”标签，还会在多帧间呈现出明显的深度不连续性或深度运动不一致性。因此，本文将语义分割结果作为潜在动态候选区域的初筛依据，再结合双目视觉估计的深度图，在时序帧之间分析深度变化趋势，以实现对动态区域的精确判断。

4.1. 语义掩码提取

首先，利用改进的 DeepLabv3+网络对输入图像进行语义分割，输出每个像素的语义类别标签。对于具备潜在动态属性的类别(如人、车、自行车、动物等)，将其对应像素位置标记为语义候选动态区域 M_{sem}，构建语义掩码，该语义掩码提供了初步的动态区域候选信息，但由于静态的人体雕像、停放的车辆等也可能被判为“动态类”，存在误判静态为动态的问题，因此需要进一步融合几何信息进行剔除。

4.2. 基于深度一致性的动态检测

为进一步确认像素点是否属于真实动态区域，本文基于深度一致性原理，判断连续帧中某点的几何变化情况。设定左目图像中某点为 $p=(u,v)$，视差为 d，那么该点在相机坐标系中的三维坐标为：

$$Z=\frac{f\cdot B}{d},\ X=\frac{(u-c_x)\cdot Z}{f},\ Y=\frac{(v-c_y)\cdot Z}{f}$$

其中 f 为焦距，B 为双目基线，(c_x,c_y) 为主点坐标。

设当前帧点 $P_t=(X_t,Y_t,Z_t)$，前一帧的相机相对位姿为 $[R|t]$，则前一帧中的对应三维点为：

$$P_{t-1}=R^{-1}(P_t-t)$$

将 P_{t-1} 投影回图像平面得：

$$u'=\frac{f\cdot X_{t-1}}{Z_{t-1}}+c_x,\ v'=\frac{f\cdot Y_{t-1}}{Z_{t-1}}+c_y$$

由此可以找到该点在前一帧图像上的对应位置，用于计算深度差：

$$\Delta D(p)=\left|Z_t(p)-Z_{t-1}(u',v')\right|$$

设当前帧和前一帧的深度图分别为 D_t 和 D_{t-1}，像素点坐标 (u,v)，前一帧像素坐标为 (u',v')，计算该点在连续帧中的深度差异 ΔD：

$$\Delta D(u,v)=\left|D_t(u,v)-D_{t-1}(u',v')\right|$$

若该差值超过一定的深度变化阈值 δ_d，则认为该点的深度变化异常，标记为动态点；

$$M_{depth}(u,v)=\begin{cases}1, & \Delta D(u,v)>\delta_d\\ 0, & \Delta D(u,v)<\delta_d\end{cases}$$

4.3. 动态掩码融合策略

在完成语义掩码与深度一致性检测后，本文进一步提出一种改进的动态掩码融合策略，以提高动态区域判定的准确性和鲁棒性。传统的掩码融合方法通常采用简单的逻辑“或”运算，虽然实现简单，但在实际应用中存在误判率高、可信度不区分、边界不连续等问题，影响后续特征点剔除的精度。因此，本文在此基础上引入基于置信度加权的动态评分融合方法，并结合时序一致性判断机制，共同构建更加鲁棒的动态区域掩码。首先，将语义信息和深度变化信息分别转化为归一化的动态评分图，设像素点(u,v)的语义动态置信度为$P_{sem}(u,v)$，由语义分割网络输出获得；同时，定义深度动态评分为：

$$P_{depth}(u,v)=\min\left(\frac{\Delta D(u,v)}{\delta_{\max}},1\right)$$

其中$\delta_{\max}$为深度差归一化上限，用于避免极端值干扰评分分布。两者融合后构建最终的动态评分图$P_{dyn}(u,v)$，其加权计算如下：

$$P_{dyn}(u,v)=\alpha\cdot P_{sem}(u,v)+(1-\alpha)\cdot P_{depth}(u,v)$$

其中α表示语义信息在融合中的权重，随后设定一个动态判定阈值T_{dyn}，将高于阈值的像素标记为动态区域：

$$M_{dyn}(u,v)=\begin{cases}1, & P_{dyn}(u,v)>T_{dyn}\\ 0, & P_{dyn}(u,v)\le T_{dyn}\end{cases}$$

为增强动态掩码的时间稳定性，本文引入时序一致性判断策略。设该像素在前N帧中每一帧的动态判断结果为$M_{dyn}^{(t-i)}(u,v)$，统计其连续动态次数$C(u,v)$，若满足以下条件，则认为其为真实动态像素：

$$C(u,v)=\sum_{i=0}^{N-1}M_{dyn}^{(t-i)}(u,v)\ge\tau$$

其中τ为一致性判断阈值，最终构建稳定的动态掩码$M_{final}(u,v)$：

$$M_{final}(u,v)=\begin{cases}1, & C(u,v)\ge\tau\\ 0, & C(u,v)<\tau\end{cases}$$

该融合策略充分结合语义信息、深度变化趋势及时间一致性，提升了动态掩码的检测精度与鲁棒性。

5. 实验与分析

本文实验采用的计算机处理器为Intel i5 12500H，操作系统为Ubuntu 20.04，所使用的数据集为Euroc数据集和KITTI数据集部分序列，Euroc数据集主要是模拟室内日常环境，KITTI数据集包含市区、乡村和高速公路等场景。采用开源工具EVO来评估SLAM算法在开源数据集中运行所得到的轨迹，以绝对轨迹误差(ATE)的均方根误差(RMSE)作为评价标准。

5.1. 基于Euroc数据集的性能评估

分别选择其中的MH_01_easy、MH_04_difficult、V1_03_difficult、V2_02_medium等序列进行实验对比。在Euroc数据集测试中，本文算法展现出显著的精度提升。

如表1所示，MH_04_difficult序列的ATE从ORB-SLAM3的0.085 m降至0.072 m (提升15%)；

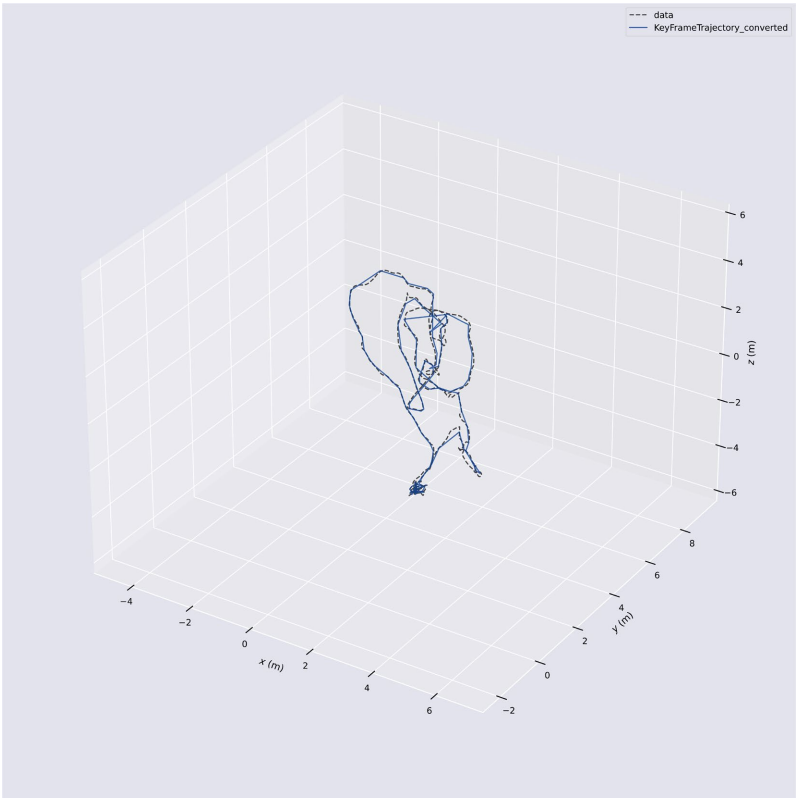
V2_02_medium 序列在剧烈旋转场景下 ATE 由 0.102 m 优化至 0.093 m (提升 8%)。轨迹对比图(图 4)进一步表明：在 M01e、M04d 等序列中，改进后轨迹几乎与真值重合；而 V102m、V202m 序列仅在飞行器姿态突变时出现轻微漂移(<0.015 m)，常态运动下稳定性优于基线系统。

5.2. 基于 KITTI 数据集的性能评估

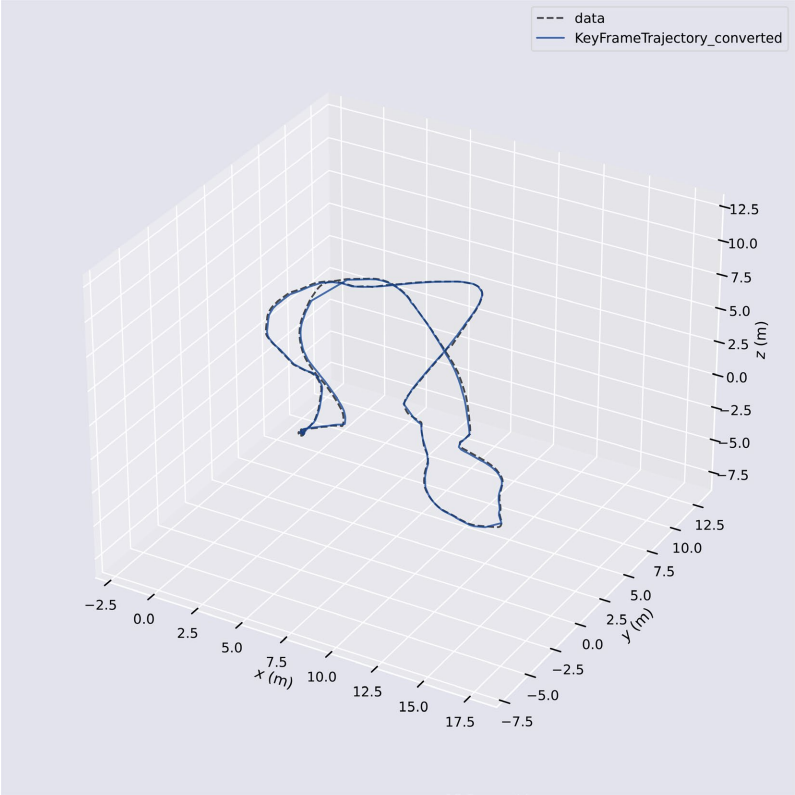
为进一步验证本文算法在室外场景的定位表现，本文在 KITTI 数据集上进行算法的定位精度评估。针对 KITTI 数据集的动态交通环境，本文方法有效克服了车辆与行人的干扰。如表 2 所示，序列 03 (高密度行人区域)的 ATE 从 0.81 m 大幅降至 0.52 m (提升 35%)，序列 09 (高速公路场景)从 3.62 m 优化至 2.90 m (提升 19%)。实时性方面，轻量化 DeepLabV3+实现 30 fps 语义分割，系统整体帧率达 25 fps，满足动态 SLAM 实时需求，结果显示本文算法能够较好的去除车辆、行人等动态物体上的特征点，本文算法在室外数据集上有较好的动态点去除效果，不会被动态物体所限制。

Table 1. Comparison results of Absolute Trajectory Error (ATE) under the Euroc dataset
表 1. 在 Euroc 数据集下绝对轨迹误差(ATE)对比结果

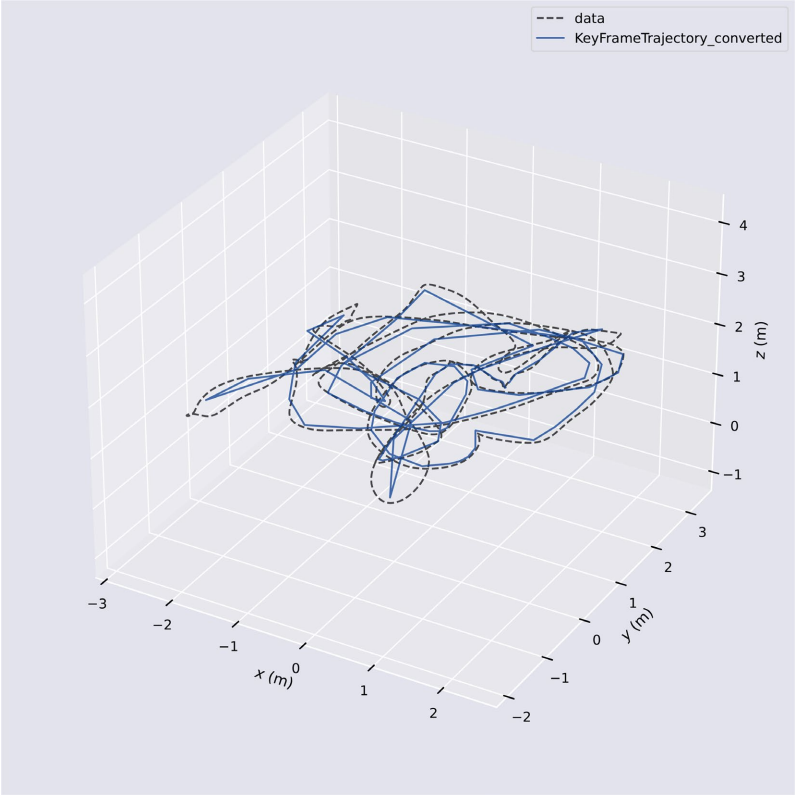
序列	ORB-SLAM3 ATE/m	本文算法 ATE/m	提升率 ATE/%
MH_01	0.048	0.041	14
MH_04	0.085	0.072	15
V1_02	0.074	0.065	12
V2_02	0.102	0.093	8



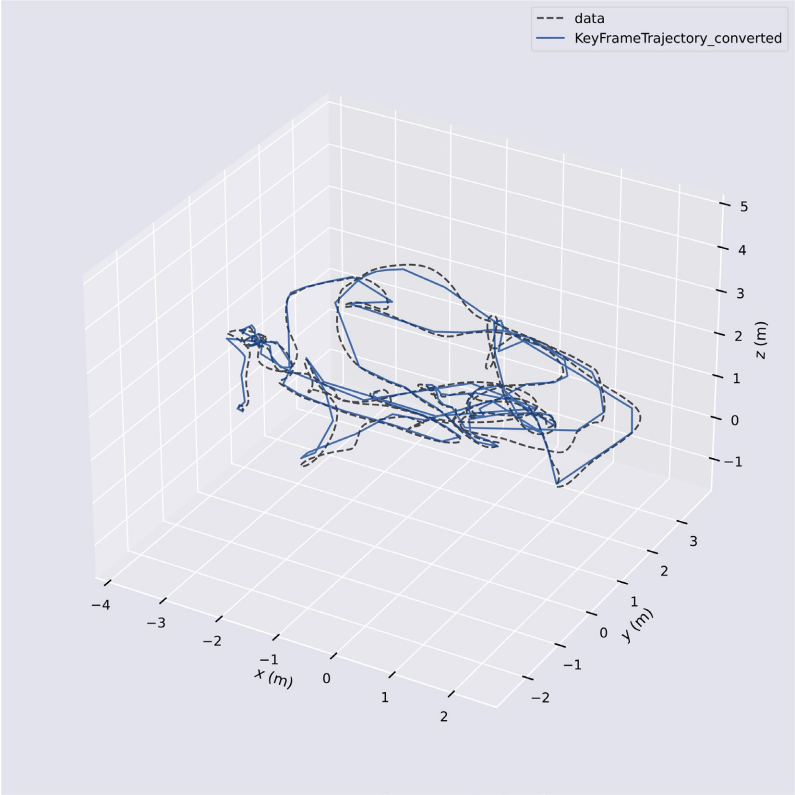
(a) MH_01


data
KeyFrameTrajectory_converted
x (m)
y (m)
z (m)

(b) MH_04


data
KeyFrameTrajectory_converted
x (m)
y (m)
z (m)

(c) V1_02



(d) V2_02

Figure 4. Trajectory comparison chart
图 4. 轨迹对比图

Table 2. Comparison results of Absolute Trajectory Error (ATE) on the KITTI dataset
表 2. 在 KITTI 数据集下绝对轨迹误差(ATE)对比结果

序列	ORB-SLAM3 ATE/m	本文 ATE/m	提升率 ATE/%
00	1.30	1.16	10
03	0.81	0.52	35
04	1.54	1.37	11
05	1.27	1.16	13
06	1.04	0.88	15
07	2.80	2.58	7
09	3.62	2.90	19

5.3. 消融实验

为验证各子模块对系统性能的影响，本文设计了消融实验，如表 3 所示，结果表明，单独使用语义分割可在大多数序列中降低 10%左右的 ATE，但在复杂动态环境中存在静态目标误判的问题；加入深度一致性检测后，进一步过滤误判动态点，ATE 显著下降，进一步减少轨迹误差，验证本文掩码融合策略的有效性。

Table 3. Comparison results of the absolute trajectory error (ATE) (m)
表 3. 绝对轨迹误差(ATE)对比结果(米)

序列	ORB-SLAM3	ORB-SLAM3 + 语义分割	ORB-SLAM3 + 语义分割 + 深度一致性检测
MH_04	0.085	0.079	0.072
V1_02	0.074	0.070	0.065
03	0.81	0.61	0.52
09	3.62	3.01	2.90

5.4. 时间分析

为验证系统在实际部署中的实时性，本文在 Intel i5-12500H 处理器平台上对关键模块进行了耗时分析。测试使用 Euroc 数据集的 MH_01 序列(分辨率 752 × 480)。各核心模块平均处理耗时如表 4 所示。

Table 4. Analysis of average processing time of key modules (ms)
表 4. 关键模块平均处理耗时分析(毫秒)

序列	语义分割	深度图计算	ORB 特征提取	跟踪与位姿估计
MH_01	33.2	12.5	15.3	18.6

分析可知，语义分割模块耗时最高(33.2 ms)，成为系统主要计算瓶颈，约占单帧总处理时间的 42%。ORB 特征提取与跟踪及位姿估计耗时相当，分别为 15.3 ms 与 18.6 ms。深度图计算耗时相对较低，为 12.5 ms。各模块耗时分布表明，深度融合语义信息虽引入一定计算开销，但仍在可接受范围内，系统整体仍能满足实时性要求。后续工作可着重于进一步优化语义分割网络的计算效率。

6. 结论

本文针对动态环境下传统视觉 SLAM 系统因动态物体干扰导致的定位漂移问题，提出一种融合深度特征与语义约束的双目视觉 SLAM 优化框架。通过设计轻量化改进的 DeepLabV3+语义分割网络，在保证分割精度的同时提升推理速度，显著优于原 Xception 网络。进一步创新性地构建语义－深度协同的动态掩码机制：基于语义先验初筛潜在动态区域后，结合双目深度图分析时序深度差异，并引入置信度加权评分模型及时序一致性判决策略，有效解决静态误判问题，提高了动态点识别率。

在 ORB-SLAM3 框架中集成上述模块，形成深度特征增强的双目 SLAM 系统。Euroc 数据集实验表明，在 MH_04_difficult 序列中，绝对轨迹误差(ATE)从 0.085 m 降至 0.072 m (提升 15%)，且轨迹漂移量减少。KITTI 数据集验证显示：高动态场景(序列 03)的 ATE 从 0.81 m 优化至 0.52 m (提升 35%)，移动车辆特征点识别率提高。系统整体帧率满足基本实时要求。这些结果充分证明，深度特征与语义约束的协同优化能显著提升动态环境下的定位鲁棒性。

尽管本文所提方法取得了良好效果，但仍存在一些局限性。首先，动态目标的识别依赖于预定义的语义类别，无法泛化至未知类别。其次，动态判断易受双目深度估计噪声影响，在弱纹理、远距离等区域可能发生误判。未来工作将致力于提升系统对未知类别动态物体的泛化能力，并增强在深度退化场景下的鲁棒性。

参考文献

[1] 黄泽霞, 邵春莉. 深度学习下的视觉 SLAM 综述[J]. 机器人, 2023, 45(6): 756-768.

[2] Ai, Y., Rui, T., Yang, X., He, J., Fu, L., Li, J., *et al.* (2021) Visual SLAM in Dynamic Environments Based on Object Detection. *Defence Technology*, **17**, 1712-1721. https://doi.org/10.1016/j.dt.2020.09.012

[3] Liu, J., Meng, Z. and You, Z. (2020) A Robust Visual SLAM System in Dynamic Man-Made Environments. *Science China Technological Sciences*, **63**, 1628-1636. https://doi.org/10.1007/s11431-020-1602-3

[4] Liu, C., Qin, J., Wang, S., Yu, L. and Wang, Y. (2022) Accurate RGB-D SLAM in Dynamic Environments Based on Dynamic Visual Feature Removal. *Science China Information Sciences*, **65**, Article No. 202206. https://doi.org/10.1007/s11432-021-3425-8

[5] 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

[6] Jonnalagadda, V.A. and Hashim, A.H. (2024) SegNet: A Segmented Deep Learning Based Convolutional Neural Network Approach for Drones Wildfire Detection. *Remote Sensing Applications*: *Society and Environment*, **34**, Article ID: 101181.

[7] 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

[8] Zhang, X., Wang, X. and Zhang, R. (2022) Dynamic Semantics SLAM Based on Improved Mask R-CNN. *IEEE Access*, **10**, 126525-126535. https://doi.org/10.1109/access.2022.3226212

[9] Zhong, F., Wang, S., Zhang, Z., Chen, C. and Wang, Y. (2018) Detect-SLAM: Making Object Detection and SLAM Mutually Beneficial. 2018 *IEEE Winter Conference on Applications of Computer Vision* (*WACV*), Lake Tahoe 12-15 March 2018, 1001-1010. https://doi.org/10.1109/wacv.2018.00115

[10] Cheng, S., Sun, C., Zhang, S. and Zhang, D. (2023) SG-SLAM: A Real-Time RGB-D Visual SLAM toward Dynamic Scenes with Semantic and Geometric Information. *IEEE Transactions on Instrumentation and Measurement*, **72**, 1-12. https://doi.org/10.1109/tim.2022.3228006

[11] Liu, Y. and Miura, J. (2021) RDS-SLAM: Real-Time Dynamic SLAM Using Semantic Segmentation Methods. *IEEE Access*, **9**, 23772-23785. https://doi.org/10.1109/access.2021.3050617

[12] Hu, Z., Zhao, J., Luo, Y. and Ou, J. (2022) Semantic SLAM Based on Improved DeepLabv3^{+} in Dynamic Scenarios. *IEEE Access*, **10**, 21160-21168. https://doi.org/10.1109/access.2022.3154086

[13] Niu, J., Chen, Z., Zhang, T. and Zheng, S. (2024) Improved Visual SLAM Algorithm Based on Dynamic Scenes. *Applied Sciences*, **14**, Article 10727. https://doi.org/10.3390/app142210727

[14] 于明洋, 徐海青, 张文焯, 等. 融合 ASPP 与双注意力机制的建筑物提取模型[J]. 航天返回与遥感, 2024, 45(1): 136-146.

[15] 夏晓华, 苏建功, 王耀耀, 等. 基于 DeepLabv3+的轻量化路面裂缝检测模型[J]. 激光与光电子学进展, 2024, 61(8): 182-191.

[16] Howard, A., Sandler, M., Chen, B., Wang, W., Chen, L., Tan, M., *et al.* (2019) Searching for MobileNetV3. 2019 *IEEE/CVF International Conference on Computer Vision* (*ICCV*), Seoul, 27 October-2 November 2019, 1314-1324. https://doi.org/10.1109/iccv.2019.00140

[17] Yao, Y. (2023) SE-CNN: Convolution Neural Network Acceleration via Symbolic Value Prediction. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, **13**, 73-85. https://doi.org/10.1109/jetcas.2023.3244767

[18] Zhao, D., Zhang, W. and Wang, Y. (2024) Research on Personnel Image Segmentation Based on Mobilenetv2 H-Swish CBAM PSPNet in Search and Rescue Scenarios. *Applied Sciences*, **14**, Article 10675. https://doi.org/10.3390/app142210675

[19] Luo, J., Fang, H., Shao, F., Zhong, Y. and Hua, X. (2021) Multi-Scale Traffic Vehicle Detection Based on Faster R-CNN with NAS Optimization and Feature Enrichment. *Defence Technology*, **17**, 1542-1554. https://doi.org/10.1016/j.dt.2020.10.006

[20] Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W. and Hu, Q. (2020) ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Seattle, 13-19 June 2020, 11531-11539. https://doi.org/10.1109/cvpr42600.2020.01155

[21] 王梦瑶, 宋薇. 动态场景下基于自适应语义分割的 RGB-DSLAM 算法[J]. 机器人, 2023, 45(1): 16-27.