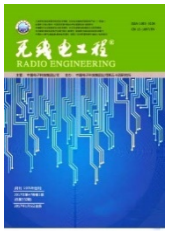


无线电工程
Radio Engineering
ISSN 1003-3106,CN 13-1097/TN

《无线电工程》网络首发论文

题目：动态场景下融合 YOLOv5s 的视觉 SLAM 算法研究
作者：赵燕成，魏天旭，仝棣，赵景波
网络首发日期：2023-06-29
引用格式：赵燕成，魏天旭，仝棣，赵景波．动态场景下融合 YOLOv5s 的视觉 SLAM 算法研究[J/OL]．无线电工程.
https://kns.cnki.net/kcms2/detail/13.1097.TN.20230629.1308.002.html

网络首发时间：2023-06-29 18:07:21
网络首发地址：https://kns.cnki.net/kcms2/detail/13.1097.TN.20230629.1308.002.html

动态场景下融合 YOLOv5s 的视觉 SLAM 算法研究

赵燕成，魏天旭，仝棣，赵景波*

（青岛理工大学信息与控制工程学院，山东 青岛 266520）

摘要：为了解决视觉 SLAM 系统在动态场景下容易受到动态物体干扰，导致算法定位精度和鲁棒性下降的问题，提出了一种融合 YOLOv5s 轻量级目标检测网络的视觉 SLAM 算法。该算法在 ORB-SLAM2 的跟踪线程中添加了目标检测和剔除动态特征点模块，通过剔除图像中的动态特征点，提高 SLAM 系统的定位精度和鲁棒性。首先，改进了 YOLOv5s 的轻量化目标检测算法，来提高了网络在移动设备中的推理速度和检测精度。其次，将轻量化目标检测算法与 ORB 特征点算法结合，以提取图像中的语义信息并剔除先验的动态特征。最后，结合 LK 光流法和对极几何约束来剔除动态特征点，并利用剩余的特征点进行位姿匹配。在 TUM 数据集上的验证表明，提出的算法与原 ORB-SLAM2 相比，在高动态序列下的 ATE 和 RPE 均提高了 95%以上，有效提升了系统的定位精度和鲁棒性。相对当前一些优秀的 SLAM 算法，在精度上也有明显的提升，并且具有更高的实时性，使其在移动设备中拥有更好的应用价值。

关键词：视觉 SLAM；动态场景；轻量级网络；目标检测；LK 光流法

中图分类号 TP39 **文献标识码** A **OSID 码：**

Research on Visual SLAM Algorithm Incorporating YOLOv5s in Dynamic Scenes

ZHAO Yancheng，WEI Tianxu，TONG Di，ZHAO Jingbo*

(School of Information and Control Engineering, Qingdao University of Technology, Qingdao Shandong 266520, China)

ABSTRACT： To address the problem of visual SLAM systems being susceptible to dynamic object interference in dynamic scenes, leading to decreased localization accuracy and robustness, a visual SLAM algorithm incorporating the lightweight YOLOv5s object detection network is proposed. This algorithm introduces a module for target detection and removal of dynamic feature points into the tracking thread of ORB-SLAM2, aiming to improve the localization accuracy and robustness of the SLAM system by eliminating dynamic feature points from the images. Firstly, an enhanced lightweight object detection algorithm based on YOLOv5s is developed to improve the inference speed and detection accuracy of the network on mobile devices. Secondly, the lightweight object detection algorithm is combined with the ORB feature point algorithm to extract semantic information from the images and remove the pre-determined dynamic features. Finally, dynamic feature points are eliminated using the Lucas-Kanade optical flow method and epipolar geometry constraints, and the remaining feature points are utilized for pose estimation. Validation on the TUM dataset demonstrates that the proposed algorithm outperforms the original ORB-SLAM2, achieving over 95% improvement in both ATE and RPE metrics for high dynamic sequences, thus effectively enhancing the localization accuracy and robustness of the system. Moreover, compared to existing state-of-the-art SLAM algorithms, the proposed algorithm exhibits significant improvements in accuracy and real-time performance, making it more valuable for applications on mobile devices.

KEYWORDS： Visual SLAM; Dynamic Scenes; Lightweight Network ； Object Detection; LK Optical Flow

0 引言

同步定位与建图(Simultaneous Localization and Mapping，SLAM)是移动机器人在进入陌生环境后实现自主定位与导航的关键技术[1]，已应用于自动驾驶、生物医疗、无人机等多个领域。当前 SLAM 系统主要借助相机、惯性测量单元(Inertial Measurement Unit，IMU)、激光雷达和超声波雷达等传感器，视觉 SLAM 是使用相机作为外部传感器进行同步定位与建图的技术。得益于计算机视觉技术

基金项目：国家自然科学基金(51475251)；青岛市民生计划（22-3-7-xdny-18-nsh）
Foundation Item: National Natural Science Foundation of China (51475251)；Qingdao People's Livelihood Planning（22-3-7-xdny-18-nsh）

的进步，视觉 SLAM 以其低廉的成本、丰富的环境信息和广泛适应性受到了学者的广泛研究。在 2007 年，Klein 等人[2]提出了一种基于关键帧的 SLAM 解决方案，即并行跟踪和建图(Parallel Tracking And Mapping，PTAM)，是视觉 SLAM 领域的重大突破，使得基于视觉的 SLAM 系统成为研究热点。目前，视觉 SLAM 主要分为两类：特征点法和直接法。其中，PTAM、ORB-SLAM2[3]是基于特征点法的优秀算法，而 LSD-SLAM[4]、DSO[5]则是基于直接法的优秀算法。当前，基于特征点和直接法的视觉 SLAM 算法都是建立在静态环境假设下实现高精确度和鲁棒性，但在现实的生活场景中会出现大量诸如行人、动物和汽车的动态物体，当环境中出现较多此类动态物体时，会使 SLAM 系统的定位精度和鲁棒性严重下降，甚至导致建图失败。

针对上述在动态环境中视觉 SLAM 遇到的问题，国内外学者主要从基于几何、基于光流和基于深度学习的三个方面进行研究。一是基于几何的算法，Kundu 等人[6]提出了一种方法，通过使用多视几何约束来检测物体的静止或运动状态。该方法利用对极线约束和机器人运动知识来估计图像像素沿着对极线的位置界限，以便检测环境中的运动物体。此外，为了准确分类物体的状态，还应用了贝叶斯框架来区分是否为动态物体。Palazzolo 等人[7]提出了一种基于 TSDF 的映射方法，能够在动态环境中跟踪相机的姿态。该算法采用了有效的直接跟踪方法，并利用编码在 TSDF(Truncated Signed Distance Function)中的颜色信息来估计传感器的姿态。同时，该算法还结合了体素哈希表示方法，通过基于配准残差和空闲空间表示的算法来过滤动态特征，从而实现了在动态环境中的稠密建图。二是基于光流的算法，Fang 等人[8]提出了一种基于点匹配技术和均匀采样策略的光流方法有效实现了检测和跟踪移动目标，并引入卡尔曼滤波器改善了检测和跟踪效果，但该算法在提高计算速度的同时损失了一部分精度。Zhang 等人[9]提出了一种基于光流的稠密 RGB-D SL 通过稠密的 RGB-D 点云建立三维地图，使用光流算法来提取当前帧与上一帧之间的运动信息，并计算相应的光流残差提升更准确和高效的动态、静态分割，然后将动态物体进行剔除，在动态和静态环境下都实现了精准和高效的性能。三是基于深度学习的算法，随着深度学习在计算机视觉领域的发展，越来越多的研究人员运用目标检测和语义分割的方法识别并剔除场景中的动态特征点并取得了优秀的效果。清华大学 Yu 等人[10]在 ORB-SLAM2 基础上提出一种名为 DS-SLAM 的方法。该方法加入了语义分割和稠密地图创建线程，并采用 SegNet[11]语义分割网络和运动一致性检测方法相结合的方式，以剔除对系统影响大于设定阈值的特征点，以提高系统在动态环境下的鲁棒性和稳定性。该算法经过验证具有显著的效果改进。同样，在 ORB-SLAM2 基础上 Bescos 等人[12]提出了 DynaSLAM 算法，该算法利用 Mask R-CNN（Region-based Convolutional Neural Network）[13]分割和多视图几何法结合来检测潜在的动态特征并剔除动态元素，从而提升了系统的准确性，但该算法存在着耗时严重和实时性差的问题。在结合目标检测方面，Zhong 等人[14]提出了 Detect-SLAM 系统，将目标检网络 SSD（Single Shot MultiBox Detector）[15]和 SLAM 系统结合，通过预训练好的目标检测网络对图像序列中物体进行检测，然后在 ORB 特征提取阶段将动态特征点剔除，极大地提高了动态环境中 SLAM 的准确性和鲁棒性。Wang 等人[16]提出了一个动态场景下的语义 SLAM 系统，将深度学习方法和基于 LUT SLAM 相结合，利用 YOLOv3 目标检测算法对特定的运动物体进行检测并剔除，生成了剔除移动物体的稠密点云图。

为了减少环境中动态物体对算法的影响，本文针对室内的动态场景提出了一种融合 YOLOv5s 轻量级目标检测网络的视觉 SLAM 算法，运用改进后的轻量化目标检测算法、光流法和结合对极几何约束的方法来剔除场景中的动态特征点，在保证实时性的同时提高视觉 SLAM 系统在动态场景中的定位精度和鲁棒性。

本文有以下两部分的改进和创新：

（1）将 YOLOv5s 的原普通卷积替换为更加轻量级的 Ghost 卷积，以减少网络参数大小；在网络中添加 CA（Coordinate Attention）注意力机制，以增强网络对于重要信息的捕捉能力；同时将损失函数 CIOU 修改为 EIOU，提高模型的稳定性和性能。从而提高算法的推理速度和检测精度。

（2）在 ORB-SLAM2 的框架中添加了目标检测模块和剔除动态特征点模块，将目标检测算法、LK 光流法和对极几何约束相结合，以此剔除环境中的动态特征点。

1 系统框架与流程

ORB-SLAM2 是一种基于特征点的单目/双目/RGB-D 视觉 SLAM 系统，可以通过相机捕捉的图像数据来实现同时定位和地图构建，具有稳定性高、运行速度快、易于实现等优点，是目前视觉 SLAM 领域应用最为广泛的系统，其包含了跟踪线程、局部建图和闭环检测 3 个主要的线程，系统框架如图 1 所示。

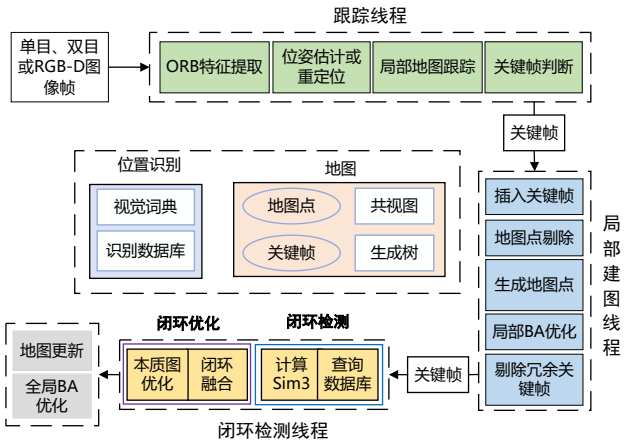


图 1 ORB-SLAM2 算法系统框架

Fig.1 ORB-SLAM2 algorithm system framework

2 基于 YOLOv5s 的轻量化目标检测算法

YOLO 系列作为一阶目标检测算法的杰出代表，相比传统算法，YOLO 算法的结构简单，具有较快的检测速度和较高的检测精度。YOLOv5 在 YOLOv4 的基础上优化了网络结构、训练策略并进行了数据增强，使得其在速度和精度上都有所提升。因 YOLOv5 的轻量化特性，其训练和推理速度比当前最新的 YOLOv7 和 YOLOv8 也要快很多，并且具有较低的内存占用，这使得 YOLOv5 在移动设备或者资源受限的系统应用场景中更具优势，而 Yolov5s 是 Yolov5 系列中模型最小、运行速度最快的网络[17]，对硬件设备要求较低，因此更适合在移动端部署。

考虑到室内动态环境中的检测对象以大目标为主和移动设备算力的限制，为进一步满足动态场景中的目标检测和保证系统能够实时运行的需要，本节以 YOLOv5s 网络为基础，改进了一种基于 YOLOv5s 的轻量化目标检测算法。①将网络的普通卷积替换为更加轻量化的 Ghost 卷积，从而减少网络的计算量，提高运行速度。②在 Backbone 中添加 CA 注意力机制，以增强网络对于重要信息的捕捉能力。③使用新的 EIoU 损失函数替代原 YOLOv5s 使用的 CIoU，提高模型的稳定性和性能。改进后的 YOLOv5s 网络结构如图 2 所示。

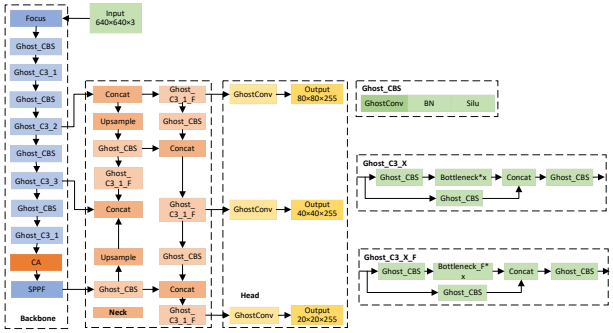


图 2 改进后的 YOLOv5s 网络结构

Fig.2 Improved YOLOv5s network structure

2.1 替换 Ghost 卷积

随着计算机视觉应用场景的不断扩大和多样化，轻量级网络结构的研究和应用成为当前热点之一。在 YOLOv5s 网络中，主干网络是整个模型的核心组成部分，决定了模型的性能和速度。然而，传统的主干网络如 Darknet53 具有较多的参数和计算量，导致模型较大且运行速度较慢。受限于移动设备的硬件条件和环境影响，为提高模型的轻量化和速度，本文将 YOLOv5s 网络原有的普通卷积层替换为更加轻量化的深度可分离卷积 GhostConv，如图 3 所示为 Ghost 模块[18]。

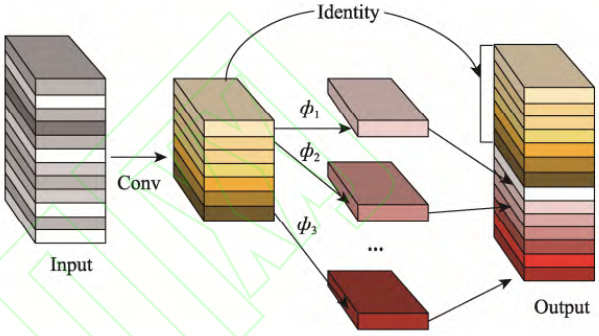


图 3 Ghost 模块

Fig.3 Ghost module

GhostNet 采用轻量化的分组卷积和通道注意力机制，以保持高准确率的同时减少网络的计算量和存储空间。同时，GhostNet 使用反向残差模块来加速模型训练和降低模型的复杂度。因此，将原有的普通卷积层替换为 GhostConv 不仅可以提高 YOLOv5s 的速度，还可以减少网络的参数量和存储空间，更利于在嵌入式和资源受限的移动设备进行实时目标检测。

2.2 添加 CA 注意力机制

在深度学习中，注意力机制已被广泛应用于图像识别、自然语言处理等任务中，取得了良好的效果。注意力机制是一种加强模型特征表达能力的计算单元，可以让模型在处理数据时更关注重要的部分，同时减少不必要的计算。在加入 Ghost 模块后，参数量和计算量都大幅降低，加快了训练和推理速度，但同时网络对全局特征的提取也减少了。因此，为减少冗余信息和增强特征图中重要的特征信息，本文选择在 Backbone 添加 CA 注意力机制[19]。CA 注意力机制模块如图 4 所示。

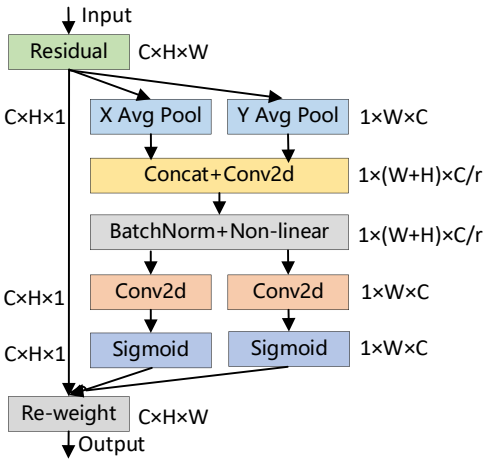


图 4 CA 注意力机制模块

Fig.4 Coordinate attention module

其算法流程主要如下：

（1）输入特征图：特征图输入到 CA 注意力机制中进行处理。(2) 通道权重计算：对于输入的特征图，分别进行全局平均池化操作和全局最大池化操作，得到两个不同的特征向量。将这两个特征向量经过一个全连接层进行变换，得到一个通道权重向量。(3) 特征重要性调整：将通道权重向量乘以输入特征图，得到加权特征图。通道权重的作用是调整每个特征通道的重要性，因此加权特征图中每个通道的特征表示的重要性得到了调整。(4) 通道信息融合：加权特征图通过一个 sigmoid 函数进行激活，得到一个权重矩阵。权重矩阵与输入特征图进行逐元素相乘，得到经过通道信息融合的特征图，这个特征图中的每个像素都包含了整个特征图中所有通道的信息，并且每个通道的重要性已经被调整。

CA 注意力机制作为一种轻量级通道注意力机制，与 SE、CBAM 注意力机制不同，其不涉及空间位置信息，而是关注不同通道间的关系和位置信息，通过自适应地调整每个通道的权重，可有效地提升 YOLOv5s 网络的准确性。

2.3 引入 EIOU 损失函数

在目标检测中，一个检测框与真实框匹配通常采用 IOU（Intersection Over Union）指标来度量。IOU 是通过计算检测框和真实框之间的重叠部分面积与并集面积之比得到的。IoU 的计算公式定义如下：

$$IoU = \frac{P \cap R}{P \cup R} \qquad (1)$$

其中 P 为检测框，R 为真实框，Yolov5s 使用 CIOU[20]作为模型的损失函数，CIOU 同时考虑到回归框宽高比例以及真实框与预测框中心距离。CIoU 的计算公式下：

$$L_{CIoU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \qquad (2)$$

式中 $\rho^2(b, b^{gt})$ 为图预测框和真实框中心点之间的欧氏距离 d，c 是能够同时包含真实框和预测框最小矩形框对角线距离。α 是权重函数，v 可以度量预测框和真实框长宽比的相似性。定义如式下：

$$\alpha = \frac{v}{1 - IoU + v} \qquad (3)$$

$$v = \frac{4}{\pi^2}(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h})^2 \qquad (4)$$

CIoU 损失函数已有效解决了 IOU、DIOU 存在的问题，但由于 v 仅反映纵横比的差异，因此 CIOU 可能会以不恰当的方法优化相似性，即存在当目标框非常小或者存在较大的偏移时，损失函数的值会出现较大偏差的问题。为解决这一问题，本文使用新的 EIoU[21]替换 CIoU 作为 yolov5s 的损失函数，EIoU 函数计算公式如下：

$$\begin{aligned} L_{EIOU} &= L_{IOU} + L_{dis} + L_{asp} \\ &= 1 - IOU + \frac{\rho^2(\mathbf{b}, \mathbf{b}^{gt})}{c^2} + \frac{\rho^2(w, w^{gt})}{C_w^2} + \frac{\rho^2(h, h^{gt})}{C_h^2} \end{aligned} \qquad (5)$$

EIOU 在 CIOU 的基础上计算宽高的差异值代替了纵横比，有效解决了 CIOU 的问题，并且采用 Focal Loss 处理难易样本不平衡的问题。引入 EIOU 损失函数的网络模型训练速度更快、收敛更稳定，能够更好的适应动态场景下目标检测需求。

3 动态特征点剔除算法设计

3.1 基于轻量化目标检测网络的语义信息提取

在计算机视觉领域，语义信息通常指对图像中对象的类别、位置、姿态、形状等高级概念的理解和表达，可以利用语义信息理解场景和动态目标，以此提升在动态场景下 SLAM 系统的鲁棒性。

传统的特征点法 SLAM 系统采用特征点提取和匹配的方法，在位姿初始化时对两帧图像进行处理，接着通过 RANSAC(Random Sample Consensus)等方法去除一些误匹配和动态点。但在动态场景中当环境中的动态数量过多时，传统 SLAM 系统的位姿初始化精度会严重下降。考虑本文针对室内场景进行研究，场景中的动态目标以人或动物为主，因此选择人或动物作为先验的动态目标。本文在 ORB-SLAM2 的跟踪线程上添加目标检测模块，运用改进后的轻量化网络 YOLOv5s 进行目标检测，并提取场景中图像的语义信息，然后将提取的语义信息和 ORB 特征提取相结合获取图像信息，利用目标检测算法预测一些先验的动态区域并剔除其中的动态特征点，将保留下的特征点进入下一环节进行跟踪匹配，从而获得更准确的相机位姿估计。

3.2 基于光流法的特征跟踪和匹配

由于光流法只需要对少量的特征点进行追踪，而不需要处理整张图像，特征点可以通过快速角点检测等方法进行提取，能够快速地计算出相邻两帧图像中运动的点，具有很好的实时性。因此通过前面语义信息滤除先验动态特征点后，使用 LK 光流法[22]对剩余的特征点进行追踪和匹配。LK 光流法示意图如图 5 所示。

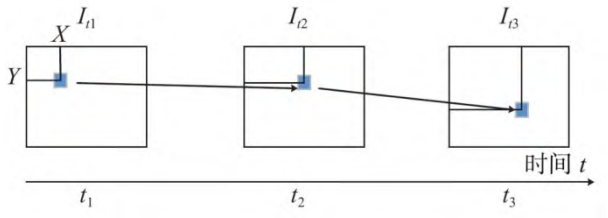


图 5 LK 光流法示意图

Fig.5 Schematic diagram of LK optical flow method

光流法有三个主要假设条件：1.亮度恒定，这是光流法的基本设定。2.小运动，必须满足。3.空间一致性。

在 t 时刻处于 (x,y) 的像素为 $I(x,y,t)$，则 $t+dt$ 时刻处于 $(x+dt,y+dt)$ 的像素点，根据假设 1 有

$$I(x,y,t)=I(x+dx,y+dy,t+dt) \qquad (6)$$

根据假设 2，对式(6)进行泰勒展开并保留一阶项：

$$I(x+dx,y+dy,t+dt)=I(x,y,t)+\frac{\partial I}{\partial x}dx+\frac{\partial I}{\partial y}dy+\frac{\partial I}{\partial t}dt \qquad (7)$$

结合式(6)、(7)可得：

$$\frac{\partial I}{\partial x}dx+\frac{\partial I}{\partial y}dy+\frac{\partial I}{\partial t}dt=0 \qquad (8)$$

式(8)就是计算光流的基本方程，两边同除 dt 有：

$$\frac{\partial I}{\partial x}\frac{dx}{dt}+\frac{\partial I}{\partial y}\frac{dy}{dt}=-\frac{\partial I}{\partial t} \qquad (9)$$

其中 $\frac{\partial I}{\partial x}$、$\frac{\partial I}{\partial y}$ 分别是像素点沿 x 轴、y 轴灰度的梯度，记为 I_x、I_y。$\frac{dx}{dt}$、$\frac{dy}{dt}$ 是像素点沿 x 轴、y 轴的速度。记为 u、v。$\frac{\partial I}{\partial t}$ 记为 I_t，则式(9)可写成：

$$\begin{bmatrix}I_x & I_y\end{bmatrix}\begin{bmatrix}u\\v\end{bmatrix}=-I_t \qquad (10)$$

根据假设 3，利用最小二乘法完成对 u、v 的求解后，可以对某个像素点在图像中的位置进行跟踪估计，然后使用光流法对像素点匹配。

3.3 动态特征点剔除

借助语义信息剔除先验的动态特征点和基于光流法获得基础矩阵后，可以得到每对特征点对应的极线，采用对极几何约束通过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。

假设图 6 中 t_1、t_2 时刻的两个像素特征点 p_1、p_2 是匹配的特征点对，其齐次坐标表示如式(11)。

$$\begin{aligned}p_1&=\begin{bmatrix}x_1 & y_1 & 1\end{bmatrix}\\p_2&=\begin{bmatrix}x_2 & y_2 & 1\end{bmatrix}\end{aligned} \qquad (11)$$

则 p_1 对应的极线 L 为：

$$L=\begin{bmatrix}X\\Y\\Z\end{bmatrix}=Fp_1 \qquad (12)$$

其中矩阵 F 是对应的基础矩阵。可求得 p_2 到极线的距离 D 为：

$$D=\frac{\left|p_2^TFp_1\right|}{\sqrt{\|X\|^2+\|Y\|^2}}>\beta \qquad (13)$$

理想状态下点到极线的距离 $D=0$，但因相机获取的图像受周围环境噪声、光线等影响会产生畸变，所以距离 $D\neq0$，因此通过设置阈值 β 来判断，若 $D>\beta$ 时，则认为是动态特征点，进行剔除，若 $D<\beta$，则认为是静止的点，予以保留。

考虑到单纯地使用一种目标检测或光流法剔除场景中的动态特征点不够全面，本文将改进后的轻量化目标检测算法、光流法以及对极几何约束相结合来剔除场景中的动态特征点。因此在 ORB-SLAM2 框架的跟踪线程中添加了目标检测模块和剔除动态特征点模块，改进后的跟踪线程如图 6 所示。首先，利用改进后的 YOLOv5s 轻量化算法来检测图像中的目标并提取语义信息，将语义信息与 ORB 特征提取相结合，剔除先验的动态特征点；其次，采用光流法将剩余的特征点进行跟踪匹配并计算出基础矩阵；最后，使用对极几何约束设置的阈值进行第二次剔除动态特征点。剩余的静态特征点被用于位姿估计，以减少环境中动态物体的影响，从而提升系统的鲁棒性和定位精度。

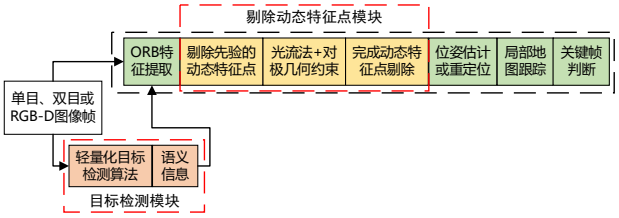


图 6 改进后的跟踪线程

Fig.6 Improved trace threads

4 实验结果分析

为验证本文提出的算法整体性能和有效性，本

节分别对 2、3 节改进后的算法进行了实验验证。

4.1 轻量化目标检测算法验证

4.1.1 数据集

考虑到室内场景中的动态物体以人为主，为验证改进后的 Yolov5s 算法的有效性，本实验选取了 COCO 数据集中“人”类别的图片进行训练和测试，共计 10800 张图片。

4.1.2 性能评估

目标检测算法的性能通常用均值平均精度（mean Average Precision，mAP）评反映了模型在召回率不同的情况下的精度表现，较高的 mAP 值表示模型在高召回率下能保持较高的准确率，因此 mAP 值越高，说明模型的性能越好。计算方式如下：

$$mAP = \frac{1}{n}\sum_{i=1}^{n} AP_i \qquad (14)$$

其中 m 表示平均，AP 表示某类样本的平均精度。本文进行了消融实验，比较了采用不同策略（替换 Ghost 卷积、添加 CA 注意力机制、引入 EIOU 损失函数）对模型性能的影响。在数据集上，通过设置 IoU 阈值为 0.5，得到 mAP@0.5 作为评价指标。

表 1 消融实验对比

Tab.1 Comparison of ablation experiments

Modles	mAP%	mAP (±)%	参数量/MB	FPS
Yolov5s	67.5	0	7.41	83
Ghost	65.6	-1.9	3.35	112
CA	68.8	+1.3	7.56	60
EIOU	70.3	2.8	7.45	95
Ghost+CA+EIOU	71.4	3.9	4.43	102

从表 1 可以看出，相比原版 Yolov5s 算法，使用 Ghost 卷积替换原卷积后，模型参数量降低至 3.35M，检测速度提高了 29 FPS，检测精度略有降低 1.9%。此外，添加 CA 注意力机制和 EIOU 损失函数对检测精度均有一定提升。与原版相比，本文改进的算法在 mAP 上增加了 3.9%，模型大小减小了 40.2%，检测速度提高了 19 FPS。这些改进实现了对轻量级目标检测算法在移动设备上的需求，既提高了检测精度又满足了实时性的要求。

4.2 改进后 ORB-SLAM2 算法验证

4.2.1TMU 数据集

为验证改进后 ORB-SLAM2 算法的有效性，采用由德国慕尼黑工业大学提供的 TUM 数据集[23]，它包含了机器人在室内环境中采集的 RGB-D 彩色图像和深度图像，已成为 SLAM 领域最广泛使用的评估数据集之一，被用于评估和比较不同算法的性能。本文选取了 TUM 数据集中四个不同的图像序列 fr3_walking_xyz 、 fr3_walking_halfsphere 、 fr3_walking_static 和 fr3_sitting_static。其中 walking 序列是高动态场景下的数据集，sitting 序列是低动态场景下的数据集。

4.2.2 动态特征点剔除效果对比

在动态特征点剔除过程中，为了保留存在于动态物体检测框中的静态目标特征点，本文将能检测到的物体分为高动态、中等动态和低动态物体。仅当物体的特征点处于高动态物体检测框并且未处于低动态物体检测框时才将其剔除。图 7 展示了动态特征点剔除前后的效果，其中(a)为剔除前的特征点图像，(b)为运用本文算法剔除动态特征点后的图像。改进后的方法有效检测图像中的物体信息，并剔除了场景中的高动态物体“人”，同时保留了“电脑、键盘、椅子”等低动态物体。

（a）剔除前　　（b）剔除后

图 7 动态特征点剔除效果

Fig.7 Dynamic feature point rejection effect comparison

4.2.3 轨迹误差结果对比

实验结果采用绝对轨迹误差（Absolute Trajectory Error，ATE）和相对轨迹误差（Relative Pose Error，RPE）作为评价指标，其中 RMSE、Mean 和 SD 分别指的是均方根误差、平均误差和标准差。同时将提升效率定义为 Improvement，其公式为：

$$\text{Improvement} = \frac{\text{m} - n}{\text{n}} \times 100\% \quad (1)$$

其中，m、n 分别是本文方法和 ORB_SLAM2 的运结果。表 2、表 3 和表 4 分别为改进前后算法的 ATE、RPE（平移部分）和 RPE（旋转部分）运算结果。

表 2 绝对轨迹误差（ATE）结果对比

Tab.2 Comparison of Absolute Trajectory Error (ATE) results

数据集	ORB-SLAM2			本文算法			Improvement/%		
	RMSE	Mean	SD	RMSE	Mean	SD	RMSE	Mean	SD
walking_xyz	0.7037	0.6059	0.3304	0.0169	0.0148	0.0087	**97.59**	**97.55**	**97.36**
walking_half	0.6375	0.5669	0.2840	0.0304	0.0242	0.0143	**95.23**	**95.73**	**94.96**
walking_static	0.2591	0.2171	0.1415	0.0076	0.0064	0.0035	**97.07**	**97.05**	**97.53**
sitting_static	0.0086	0.0075	0.0043	0.0065	0.0058	0.0032	**24.42**	**22.67**	**25.58**

表 3 相对轨迹误差（RPE）平移部分结果对比

Tab.3 Comparison of results of relative trajectory error (RPE) translation part

数据集	ORB-SLAM2			本文算法			Improvement/%		
	RMSE	Mean	SD	RMSE	Mean	SD	RMSE	Mean	SD
walking_xyz	1.0272	0.8495	0.5774	0.0256	0.0220	0.0128	**97.51**	**97.41**	**97.78**
walking_half	0.9809	0.7943	0.5755	0.0456	0.0395	0.0221	**95.35**	**95.03**	**96.16**
walking_static	0.3663	0.2501	0.2677	0.0115	0.0096	0.0061	**96.86**	**98.50**	**97.72**
sitting_static	0.0128	0.0112	0.0063	0.0122	0.0108	0.0057	**4.69**	**3.57**	**9.52**

表 4 相对轨迹误差（RPE）旋转部分结果对比

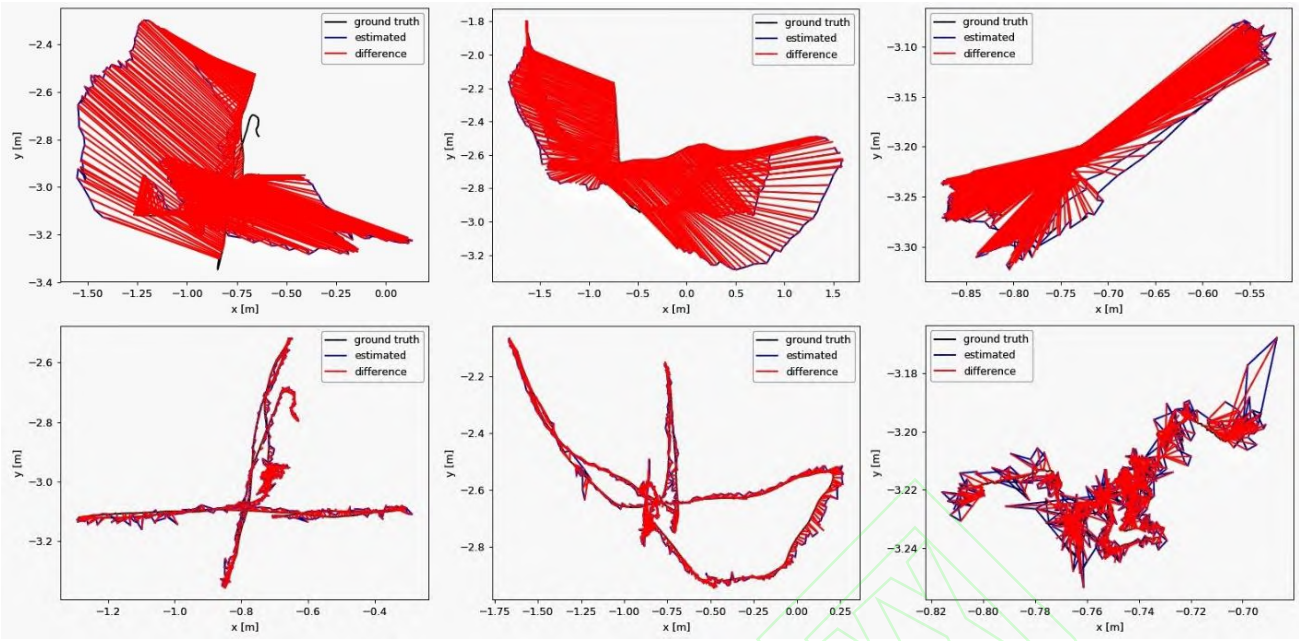
Tab.4 Comparison of relative trajectory error (RPE) rotational part results

数据集	ORB-SLAM2			本文算法			Improvement/%		
	RMSE	Mean	SD	RMSE	Mean	SD	RMSE	Mean	SD
walking_xyz	19.5651	16.1068	11.1068	0.7019	0.5703	0.4090	**96.41**	**96.46**	**96.32**
walking_half	22.5026	18.8807	12.2429	1.0439	0.9292	0.4785	**95.36**	**95.08**	**96.11**
walking_static	10.1236	7.4762	6.8259	0.3397	0.3066	0.1463	**96.64**	**95.90**	**97.86**
sitting_static	0.3608	0.3248	0.1572	0.3435	0.3074	0.1532	**4.79**	**5.36**	**2.54**

在绝对轨迹误差方面，从表 2 中可以看出本文改进后的算法相比于原 ORB-SLAM2 算法在 fr3_walking_xyz 序列中均方根误差、平均误差和标准差分别提升了 97.46%、97.55%和 97.15%，其他三个动态序列中也有明显的提升。该对比实验证明了本文算法在高动态场景中具有较好的性能，可以显著提升定位精度和鲁棒性。在低动态序列 fr3_sitting_static 中均方根误差、平均误差和标准差仅仅提升了 11.63%、9.33%和 9.30%，提升效果相对不明显。在低动态序列中，绝大多数物体的位置、姿态都是相对固定的，因此在序列中很难找到具有显著动态特征的物体或者区域，导致可以用来进行跟踪的特征点非常有限，而 ORB-SLAM2 在低动态环境下具有较好的表现，因此很难在低动态序列中大幅提高其性能。在相对轨迹误差方面，从表 3、4 中可以看出在高动态序列中提升效果明显，同样在低动态序列中提升效果不明显。

为了更加直观体现本文算法与 ORB-SLAM2 算法的效果对比，分别绘制了高动态序列下的 ATE 和 RPE 对比图，其中 ORB-SLAM2 算法（上），本文算法（下）。图 8 为 ATE 对比图，图中的黑色曲线表示相机的真实轨迹，蓝色部分表示估计的轨迹。红色线段则是两者间的误差，误差越小，红色线段就越短，表示系统的精度越高。图 9 为 RPE 对比图，可以看出本文算法相比于 ORB-SLAM2 算法，误差的波动范围很小，其稳定性更好。不难看出，在动态

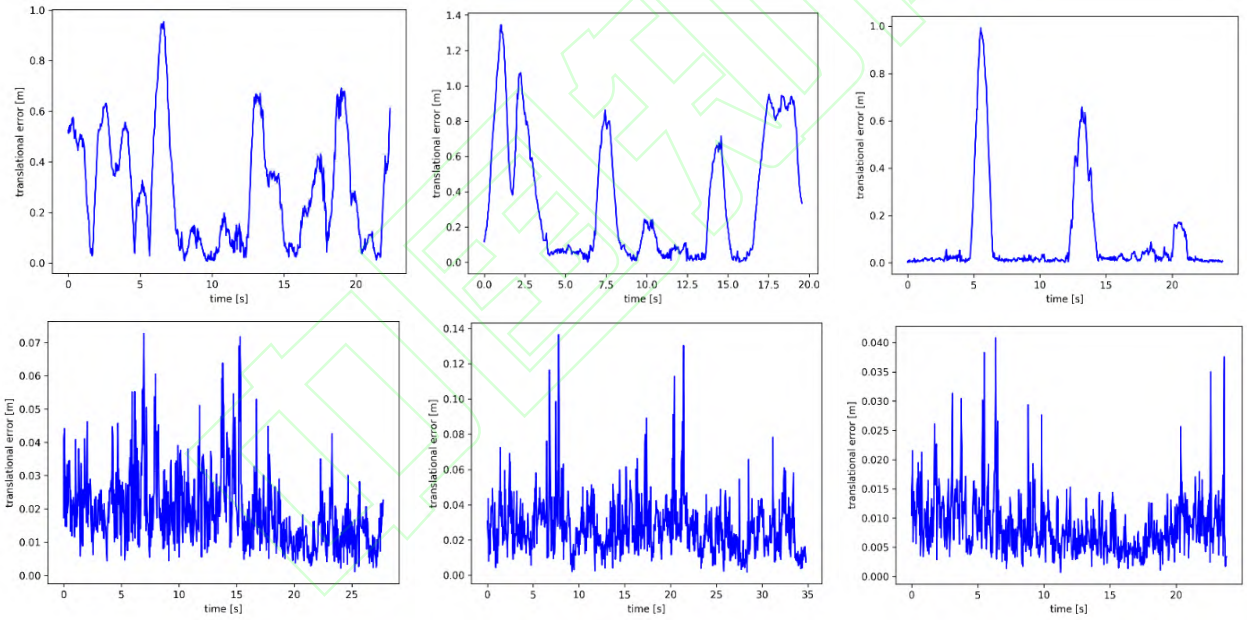
场景下本文算法相比原算法的定位精度和鲁棒性都有显著的提升。



(a) fr3_ walking _xyz　(b) fr3_ walking _half　(c) fr3_ walking _static

图 8 ATE 对比图

Fig.8 ATE comparison graph



(a) fr3_ walking _xyz　(b) fr3_ walking _half　(c) fr3_ walking _static

图 9 RPE 对比图

Fig.9 RPE comparison graph

4.2.4 其他算法对比

为了验证本文算法的先进性，本文对其进行了与近年来比较优秀的 SLAM 算法的比较。其中，DS-SLAM 和 DynaSLAM 是基于 ORB-SLAM2 框架，采用语义分割算法提取动态场景的语义信息；DetectSLAM 则利用 YOLOv3 目标检测算法对特定的动态物体进行检测并剔除；文献[24]是基于几何与运动约束进行特征匹配，来减少动态特征点和错误匹配点的影响。表 5 和 6 展示了本文算法与其他算法误差对比和部分算法平均运行时间对比。

表 5 本文算法与其他算法误差对比

Tab.5 Comparison of the error of this algorithm with other algorithms

数据集	DS-SLAM	Dyna SLAM	Detect SLAM	文献[24]	本文
	RMSE	RMSE	**RMSE**	**RMSE**	**RMSE**
walking_xyz	0.0247	0.0184	0.0241	0.0354	**0.0169**
walking_half	0.0303	0.0250	0.0514	**0.0285**	0.0304
walking_static	0.0081	**0.0064**	-	0.0131	0.0076
sitting_static	0.0065	0.0085	-	0.0446	**0.0065**

表 6 部分算法平均运行时间对比

Tab.6 Comparison of average running time of some algorithms

算法	平均运行时间
本文算法	46.5 ms
DS-SLAM	76.44 ms
DynaSLAM	335.4 ms

综合比较表 5 和表 6 可以看出，与 DS-SLAM 算法相比，本文算法在高动态场景下的定位精度和运行时间效率均得到了显著提升，在 walking_xyz 和 walking_static 序列上绝对轨迹误差的 RMSE 分别降低了 31.6%、6.2%，平均运行时间降低了 39.2%；与 DynaSLAM 算法相比，本文算法在定位精度方面的表现相近，但在运行时间方面快了 7 倍多，运行速度更快，这是因为 DynaSLAM 采用了 Mask R-CNN 实例分割算法，处理图像时较为耗时；同时，与 DetectSLAM 算法和文献[24]中的算法进行对比，本文算法在定位精度也有着不同程度的提升。通过以上的比较，进一步验证了本文算法的先进性。

5 结论

本文提出了一种融合 YOLOv5s 轻量级目标检测网络的视觉 SLAM 算法，旨在解决动态场景下视觉 SLAM 系统受到动态物体影响导致定位精度和鲁棒性下降的问题。该算法采用了基于 YOLOv5s 的轻量化目标检测算法来实时检测动态物体，再结合 ORB-SLAM2 算法提取图像中的语义信息并剔除先验的动态特征，最后通过 LK 光流法和对极几何约束来剔除动态特征点。实验结果表明，相比原 ORB-SLAM2 算法，该算法在高动态序列下的 ATE 和 RPE 均提高了 95%以上，并且在保证实时性的同时，提高了定位精度和鲁棒性，相比当前一些优秀的 SLAM 算法，在精度和实时性上有着显著的提升。因此，本文提出的融合 YOLOv5s 目标检测的视觉 SLAM 算法具有较好的实际应用前景。下一步，考虑采用多传感器融合和使用新的算法框架（如 ORB-SLAM3）进行优化改进，使算法适应更多场景的需要。

参考文献

[1] 赵燕成,房桐,杜保帅,赵景波.移动机器人视觉 SLAM 回环检测现状研究[J].无线电工程,2023,53(1):129-139.

[2] G. Klein and D. Murray, “Parallel tracking and mapping for small ar workspaces,” in 2007 6th IEEE and ACM international symposium on mixed and augmented reality. IEEE, 2007, pp. 225–234.

[3] MUR-ARTAL R，TARDOS J D.ORB-SLAM2：an opensource SLAM system for monocular，stereo，and RGB- D cameras[J].IEEE Transactions on Robotics，2017，33（5）： 1255-1262

[4] ENGEL J，SCHOPS T，CREMERS D.LSD-SLAM:largescale direct monocular SLAM [C]//Computer Vision-ECCV，2014：834-849.

[5] Matsuki H, Von Stumberg L, Usenko V, et al. Omnidirectional DSO: Direct sparse odometry with fisheye cameras [J]. IEEE Robotics and Automation Letters, 2018, 3 (4): 3693-3700.

[6] Kundu A, Krishna K M, Sivaswamy J. Moving object detection by multi-view geometric techniques from a single camera mounted robot[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, USA: IEEE, 2009: 4306-4312.

[7] Palazzolo E, Behley J, Lottes P, et al. ReFusion: 3D reconstruction in dynamic environments for RGB-D cameras exploiting residuals[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, USA: IEEE, 2019: 7855-7862

[8] Fang Y Q, Dai B. An improved moving target detecting and tracking based on optical flow technique and Kalman filter[C]// 4th International Conference on Computer Science & Education. Piscataway, USA: IEEE, 2009: 1197-1202.

[9] Zhang T W, Zhang H Y, Li Y, et al. FlowFusion: Dynamic dense RGB-D SLAM based on optical flow[C]//IEEE International Conference on Robotics and Automation. Piscataway, USA: IEEE, 2020: 7322-7328.

[10] Yu C, Liu Z X, Liu X J, et al. DS-SLAM: A semantic visual SLAM towards dynamic environments[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, USA: IEEE, 2018: 1168-1174

[11] 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

[12] Bescos B, Facil 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
[13] HE K，GKIOXARIG，DOLLARP，et al．Mask R-CNN ［C］// Proceedings of the IEEE International Conference on Computer Vision． Piscataway: IEEE，2017: 2961- 2969
[14] Zhong F W, Wang S, Zhang Z Q, et al. Detect-SLAM: Making object detection and SLAM mutually beneficial[C]//IEEE Winter Conference on Applications of Computer Vision. Piscataway, USA: IEEE, 2018: 1001-1010
[15] LIU W，ANGUELOV D，ERHAN D，et al． SSD: single shot multibox detector［C］// European Conference on Computer Vision． Berlin: Springer，2016: 21-37
[16] Wang Z M, Zhang Q, Li J S, et al. A computationally efficient semantic SLAM solution for dynamic scenes[J]. Remote Sensing, 2019, 11(11). DOI: 10.3390/rs11111363.
[17] 伍子嘉,陈航,彭勇,宋威.动态环境下融合轻量级YOLOv5s 的视觉 SLAM[J]. 计算机工程,2022,48(8):187-195+205.
[18] Han K, Wang Y, Tian Q, et al. GhostNet: More Features From Cheap Operations [C]//Proceedings of the 2020 IEEE/ CVF Conference on Computer Vision and Pattern Recognition，2020：1580-1589.
[19] HOU Q，ZHOU D，FENG J.Coordinate attention for efficient mobile network design [C]//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition，2021：13713-13722
[20] Zheng Z, Wang P, Ren D, et al. Enhancing Geometric Factors in Model Learning and Inference for Object Detection and Instance Segmentation[J]. 2020：8574-8586
[21] Zhang Y F, Ren W, Zhang Z, et al. Focal and Efficient IOU Loss for Accurate Bounding Box Regression[J]. 2021.
[22] Lucas B D. An iterative image registration technique with an application to stereo vision[C]// Proc. of the 7th International Conference on Artificial Intelligence, 1981.
[23] STURM J, ENGELHARD N, ENDRES F, et al. A benchmark for the evaluation of RGB-D SLAM systems[C]. IEEE/ RSJ International Conference on Intelligent Robots and Systems, Piscataway (IROS), USA: IEEE, 2012: 573- 580.
[24] 艾青林,刘刚江,徐巧宁.动态环境下基于改进几何与运动约束的机器人 RGB-D SLAM 算法[J]. 机器人,2021,43(2):167-176.

作者简介：

赵燕成，（1999-）男、硕士，就读于青岛理工大学电子信息专业，主要研究方向：视觉 SLAM、深度学习。

魏天旭，（1997-）男、硕士，就读于青岛理工大学控制科学与工程专业，主要研究方向：深度学习、网络控制系统。

仝棣，（1996-）男、硕士，就读于青岛理工大学电子信息专业，主要研究方向：目标检测、神经网络。

（*通讯作者）赵景波，（1971-）男、博士，教授，主要从事机器人工程、计算机控制领域的研究和教学。