DOI: 10.13382/j.jemi.B2407825

融合注意力机制的室外场景视觉 SLAM 算法研究

马金睿 梁 浩 林剑辉

(北京林业大学工学院 林业装备与自动化国家林业和草原局重点实验室 北京 100083)

摘 要:室外场景中特征点丰富且具有多样的几何形状和尺度,但因光照变化明显、纹理重复性较高,导致传统的视觉同步定位与建图算法在进行场景的三维重建过程时,存在特征点提取与匹配精度低的问题。为了提升系统在复杂环境中的建图精度和鲁棒性,提出一种融合注意力机制的视觉同步定位与建图(SLAM)算法,对 SLAM 系统中的特征提取和匹配方式进行改进。首先,将通道-空间融合的卷积注意力模块融合到 SuperPoint 网络编码器的卷积层中,以增强模型的特征提取和匹配能力;然后,将改进后的 SuperPoint 网络与 ORB-SLAM2 算法的后端相结合,实现在复杂场景中更准确的位姿估计和地图构建;最后,在 KITTI 数据集上进行验证。结果表明,融合通道-空间卷积注意力模块的 SuperPoint 网络在保持特征点稳定性和描述子判别性的基础上,有效提升了图像间特征匹配的精度,所提出的 SLAM 算法与 ORB-SLAM2 算法相比,绝对轨迹误差减少了 30.05%,相对位姿误差减少了 14.49%,实验结果表明,方法在光照变化明显和纹理重复性高的室外环境中具有更强的鲁棒性和稳定性,有效地提升了 SLAM 系统在室外复杂环境中的建图精度。

关键词: 视觉 SLAM;室外场景;特征提取;卷积注意力机制

中图分类号: TN911.73;TP391.41 **文献标识码**: A **国家标准学科分类代码**: 510.40;510.80

Research on visual SLAM algorithm for outdoor scenes with integrated attention mechanism

Ma Jinrui Liang Hao Lin Jianhui

(Key Laboratory of National Forestry and Grassland Administration for Forestry Equipment and Automation, School of Technology, Beijing Forestry University, Beijing 100083, China)

Abstract: Outdoor scenes are rich in feature points with diverse geometric shapes and scales; however, significant illumination variations and high texture repetitiveness often lead to low feature extraction and matching accuracy in conventional visual simultaneous localization and mapping (SLAM) algorithms during 3D reconstruction. To improve mapping accuracy and robustness in complex environments, this paper proposes a visual SLAM algorithm integrated with an attention mechanism, aiming to enhance the feature extraction and matching strategies within SLAM systems. Specifically, a channel-spatial convolutional attention module is embedded into the convolutional layers of the SuperPoint encoder to strengthen the model's feature detection and matching capabilities. The improved SuperPoint network is then integrated with the backend of the ORB-SLAM2 algorithm, enabling more accurate pose estimation and map construction in complex scenarios. The proposed approach is validated on the KITTI dataset. Experimental results demonstrate that the SuperPoint network integrated with the channel-spatial convolutional attention module significantly improves feature matching accuracy between images while maintaining the stability of keypoints and the discriminability of descriptors. Compared with the original ORB-SLAM2 algorithm, the proposed method achieves a 30.05% reduction in absolute trajectory error (ATE) and a 14.49% reduction in relative pose error (RPE). These results confirm that the proposed SLAM algorithm exhibits stronger robustness and stability in outdoor environments characterized by significant illumination changes and repetitive textures, effectively enhancing the mapping accuracy of SLAM systems in complex outdoor scenes.

Keywords: visual SLAM; outdoor scenes; feature extraction; convolutional attention mechanism

收稿日期: 2024-09-13 Received Date: 2024-09-13

0　引　言

同步定位与建图(simultaneous localization and mapping,SLAM)技术主要用于解决移动机器人在运动过程中的定位与地图构建问题[1]。SLAM 通过构建环境地图,使机器人能够感知障碍物、识别可通行区域,并基于优化后的位姿估计进行路径规划[2]。在视觉 SLAM 系统中,前端数据关联通常分为特征点法和直接法,图像中特征点之间的对应关系通常通过特征描述子计算匹配度来估计[3]。传统的视觉 SLAM 方法利用前端视觉里程计来提取和匹配特征点,进行相机位姿估计,使用后端优化消除帧间累计误差,并运用闭环检测修正全局轨迹[4]。在特征提取方面,经典的 SIFT 特征[5]和 ORB 特征[6],均依赖于手工设计的描述符算子,在纹理丰富的环境中能够实现稳定匹配,但这些传统特征存在计算复杂度高、鲁棒性不足的问题,尤其在光照变化剧烈或纹理重复区域,易出现特征匹配错误。

ORB-SLAM2 是目前主流的视觉 SLAM 算法[7],采用 FAST 角点检测算法和 BRIEF 描述子构建适用于实时应用的 ORB 特性。通过多尺度特征检测和方向计算,但 ORB 特征依赖于图像的灰度信息,在光照条件变化明显的场景中,特征提取和匹配的效果会显著下降,而 BRIEF 描述子的表达能力有限,在更多具有重复性纹理的场景中易出现错配的情况,导致定位和建图的误差增大,出现轨迹漂移和丢失的问题。近年来,具有图像匹配功能的深度学习算法在特征检测、鲁棒性、泛化性和复杂任务处理方面展现出了明显的优势[8],在三维重建、特征提取和目标检测等领域得到了广泛应用,如 Key. net 和 GCNv2[9-10],均同时进行特征点检测和描述子提取,实现了端到端优化特征点检测和描述子提取,显示了较强的特征表达能力和适应能力。

此外,注意力机制的引用提升了神经网络对图像中关键区域的关注能力[11]。SuperGlue[12]结合了图神经网络和注意力机制,有效提升了神经网络算法在图像特征提取与匹配任务中的准确率。Zhao 等[13]提出了一种结合多尺度特征融合和双注意力机制的孪生网络算法,用于提高模板匹配的准确性。刘冬等[14]提出了一种基于深度学习和注意力机制的稳定、实时图像特征提取与匹配方法,提高了 SLAM 系统在动态环境下的精度、稳定性和实时性。端到端的特征学习网络不仅降低了工程复杂度,也减少了算法运行过程中的误差累积[15]。

SuperPoint[16]是一种在无监督方法下训练用于提取图像特征点和描述符的网络。SuperPoint 兴趣点检测器和描述符网络可以在一定程度上应对光照和视点的变化,能够改善提取特征点的局部奇异性和累积漂移引起的误差问题。与 ORB 相比,它的定位精度和稳定性更具优势,此外,深度学习比传统方法具有更好的泛化能力。尽管 SuperPoint 在多个基准测试中表现良好,其在处理复杂公园场景时的表现仍有提升空间。注意力机制的应用可以增强卷积神经网络的特征表示能力。卷积注意力模块(convolutional block attention module,CBAM)是一种用于提升卷积神经网络性能的注意力机制,增强卷积神经网络对于特征图的表达能力[17]。CBAM 通过引入空间注意力和通道注意力两个模块来提升网络对关键特征的关注能力,在目标检测[18]、图像分割[19]、图像分类[20]等其他计算机视觉任务中均有广泛的应用。在特征点检测与描述任务中,引入 CBAM 能够提升特征匹配的准确性。

尽管当前基于深度学习的特征提取方法已有显著进展,但现有研究多集中于特征提取精度的提升,而较少结合具体 SLAM 系统结构进行整体优化。特别是对于室外自然场景中存在的光照变化明显、尺度形状多样与纹理重复性高等问题,现有 SLAM 系统在特征匹配与闭环检测方面仍面临挑战。针对上述问题,本文提出一种融合通道-空间注意力机制的视觉 SLAM 方法,将改进后的 SuperPoint 卷积神经网络与 ORB-SLAM2 系统的后端相结合,构建能够在特征点较多、尺度形状多样和光照强度变化明显的室外复杂场景运动过程中能更好的进行跟踪和闭环检测的高精度建图视觉 SLAM 系统。

1　算法结构

本文算法共有两个部分,分别为融合注意力机制的特征提取和匹配网络以及在包含室外场景的公共数据集上进行轨迹地图构建。将通道和空间融合的注意力机制与卷积神经网络结合,使用简单的几何形状进行无监督训练。在常见的室外场景中,环境复杂,光照变化明显,纹理重复性较高、植被遮挡部分较多这些因素均会影响算法的特征点检测性能,因此,利用单应性训练来提升模型的泛化性能。

传统的特征匹配算法主要依赖于手工设计的特征描述符,这些描述符在极端光照条件下可能无法稳定提取和匹配特征点,导致匹配准确度下降。融合注意力机制的神经网络特征匹配算法在光照变化明显的室外场景中,比传统手工设计的特征匹配算法具有更好的适应性和鲁棒性。利用匹配后的特征点进行相机位姿估计,对地图点的位置和优化进行稳健估算,从而对摄像机位置进行估算。估计结果插入关键帧,本地地图生成更新模块,然后调整本地捆绑以优化地图结构。当系统检测到回环时,回环检测模块会执行全局优化,通过全局捆绑调整消除累积误差,确保地图的精度,避免漂移现象。该方

法结合了深度学习的特征增强能力与传统 SLAM 的优化策略,提高了系统在复杂环境下的鲁棒性。本文算法框架如图 1 所示。

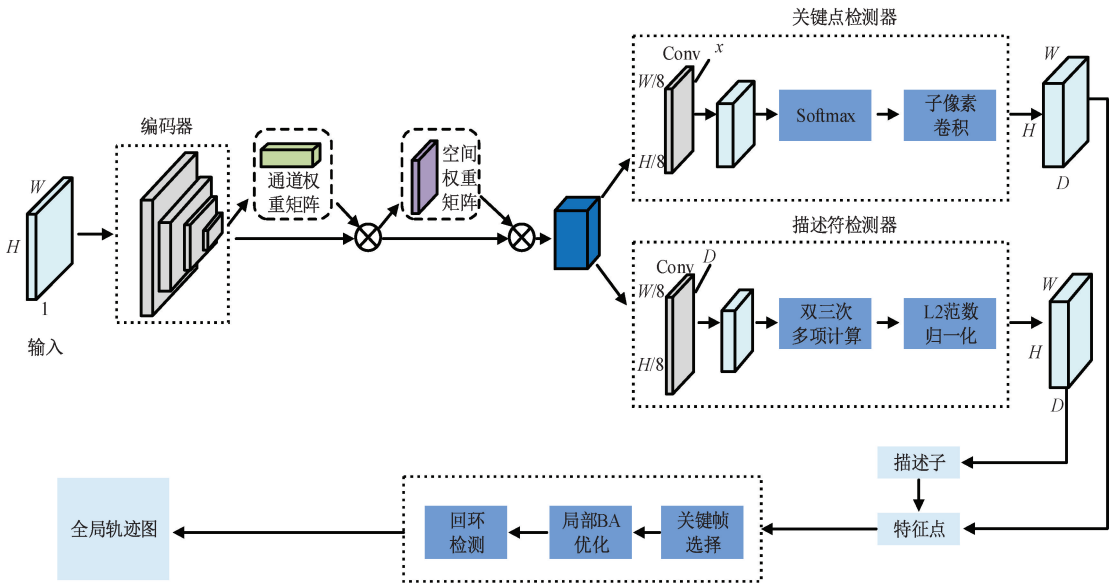


图 1　算法框架图

Fig. 1　Algorithm framework diagram

2　融合注意力机制的特征提取和匹配网络

本文针对传统的视觉 SLAM 算法在进行场景三维重建过程时存在特征点提取与匹配精度低的问题,提出了融合注意力机制的特征提取和匹配网络。网络框架分别由编码器、卷积注意力机制模块和解码器 3 个主要部分构成。其中编码器的卷积层提取图像的多层次关键信息,之后通过 CBAM 模块来进行通道注意力和空间注意力调整,以此得到增强特征图,并将特征点传递给热图生成器和描述子生成器,特征点热图生成器生成特征点的概率热图,通过 Softmax 函数转换为概率分布。特征描述子生成器生成特征点的描述子,并进行 L2 归一化以便匹配时计算欧氏距离。

在室外场景下,移动机器人使用传统特征匹配算法时面临诸多困难。首先,在早晨、黄昏或阴影区域,传统算法难以应对光照差异,导致特征提取和匹配的不稳定。此外,移动机器人运动时的视角变化也会导致同一场景中的特征点在不同视角下难以匹配,而重复纹理区域的出现会进一步增加误匹配的可能性。本文提出的融合注意力机制的特征提取和匹配网络中的关键点检测是基于 Softmax 归一化的概率输出,选择概率最高的位置作为特征点,提取这些位置对应的特征描述子。使用特征描述子进行特征点匹配,常用的方法包括暴力匹配和基于树的匹配[21]。根据特征描述子之间的距离进行匹配,找到相似度最高的特征点对。通过匹配的特征点对进行姿态估计,计算相机的相对位置和姿态,将姿态信息和特征点用于构建稀疏地图。相比传统算法复杂的特征匹配方式而言,本文算法可以在光照变化明显、特征点种类较多的复杂环境中捕捉更多的全局图像特征信息,使 SLAM 系统在面对自然公园等复杂环境时具有更加精准的位姿跟踪能力。

2.1　融合注意力机制的特征提取网络模型

本文所使用的网络编码器由 3 部分组成,分别是卷积层、池化空间下采样层和非线性激活函数层。其中编码器使用 3 个最大池化层,将低维输出的像素作为一个单元,编码器的 3 个下采样池化操作过程中会产生 8×8 像素单元,作为输入图像中相应的特征映射集合,将输入图像映射到空间维度更小和通道深度更大的中间张量中。编码器的浅层通常负责提取图像中的边缘、纹理、棱角等低级特征,这些特征对于整体图像的表征作用较弱[22]。将 CBAM 模块插入编码器浅层时,其中的的注意力机制虽然可以增强特征的重要性,但是由于浅层网络的感受野较小,输入的特征本身较为简单且表达能力有限,经过增强后的特征依然较为粗糙,难以有效提升特征匹配的精度。因此在考虑将 CBAM 模块插入 superpoint 编码器的卷积层时,主要考虑深层插入和跨层插入这两种方式。

CBAM 通过通道注意力模块(channel attention module, CAM)和空间注意力模块(spatial attention module, SAM)分别在通道和空间维度上应用注意力机

制。对于输入特征图,通道注意力模块在通道维度上对其进行加权,强调重要的通道特征,抑制不重要的通道特征。空间注意力模块通过在特征图的空间维度上分配不同的权重来关注重要的特征位置。通过添加此模块,可以增强模型特征提取的能力[17],网络结构如图 2 所示。

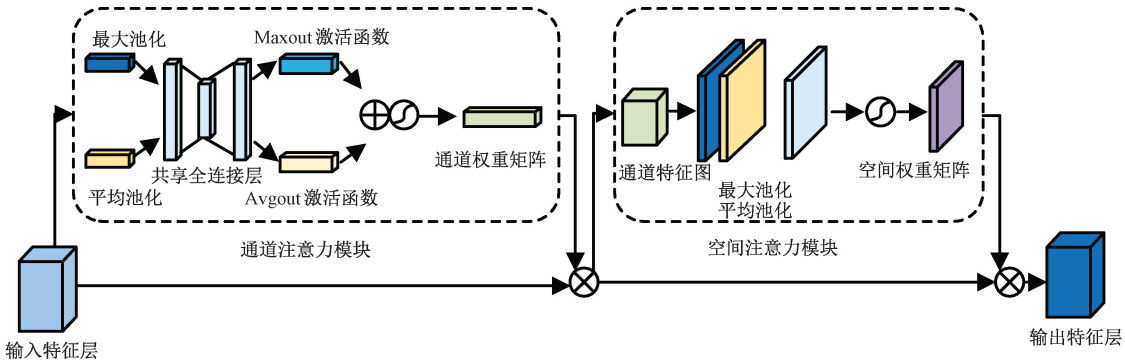


图 2　通道-空间注意力机制网络结构

Fig. 2　Structure of the network of channel-spatial attention mechanisms

其中,通道注意力模块由全局平均池化、全局最大池化、共享全连接层和 Sigmoid 激活函数组成。空间注意力模块接收通道注意力模块输出的特征图,进行通道维度上的全局平均池化和全局最大池化,生成两个单通道的特征图,将这两个单通道的特征图在通道维度上进行拼接。将拼接后的特征图输入到一个卷积层中,得到一个单通道的空间注意力图。CBAM 模块通过全局池化、全连接和卷积操作,实现了对特征图的注意力调整,适用于各种卷积神经网络。通过通道和空间维度的注意力机制,CBAM 模块提升了卷积神经网络在图像特征提取与匹配任务中的算法性能。

2.2　损失函数构建

在构建融合 CBAM 模块的 SuperPoint 网络的损失函数时,主要的损失函数包括关键点检测损失和描述子匹配损失两部分。生成每个图像的真实关键点热图 H_{true}。由 SuperPoint 网络输出的预测关键点热图 H_{pred},表示网络对每个位置是否为关键点的预测概率。关键点检测损失用于衡量预测的关键点热图与真实热图之间的差异,使用交叉熵损失函数计算 H_{true} 和 H_{pred} 之间的差异,交叉熵损失函数定义如下:

$$Loss_{detection} = -\sum_{i=1}^{N}[H_{true,i}\log(H_{pred,i}) + (1 - H_{true,i})\log(1 - H_{pred,i})] \tag{1}$$

描述子匹配损失用于衡量生成的特征描述子与真实特征描述子之间的相似性。对每个关键点生成特征描述子 D_{pred},生成成对的特征描述子,包括匹配对和不匹配对。匹配对表示在两张图像中对应同一特征点的描述子,不匹配对表示对应不同特征点的描述子,计算每对描述子之间的欧氏距离 d,对于匹配点对,希望欧氏距离 d 尽可能小;对于不匹配对,希望欧氏距离 d 尽可能大。对比损失函数定义如下:

$$Loss_{descriptor} = \frac{1}{2N}\sum_{i=1}^{N}[y_i d_i^2 + (1 - y_i)\max(margin - d_{i,0})^2] \tag{2}$$

式中:N 是描述子对的数量;y_i 表示第 i 对描述子的匹配标签(1 表示匹配对,0 表示不匹配对);d_i 表示第 i 对描述子的欧氏距离;$margin$ 是设定的阈值,用于区分匹配对和不匹配对。总损失函数是关键点检测损失和描述子匹配损失的加权和,通过调整权重系数来平衡两部分的损失,从而提升模型的整体性能。总损失函数定义如下:

$$Loss = aLoss_{detection} + bLoss_{descriptor} \tag{3}$$

式中:a 和 b 是权重系数。交叉熵损失与对比损失在目标上是互补的,前者确保关键点的可检测性,后者增强特征点的匹配能力。然而,在联合训练时,二者可能会在优化过程中产生一定的竞争关系。例如,关键点检测可能会在高对比度区域生成特征点,而这些区域的局部纹理可能不稳定,导致后续匹配困难。因此,在定义总损失函数时设置了权重系数以确保关键点检测和匹配的有效性,避免了关键点检测和描述子匹配联合训练时可能出现的矛盾问题。通过以上步骤构建融合 CBAM 模块的 SuperPoint 网络的损失函数。关键点检测损失用于衡量预测的关键点热图与真实热图之间的差异,而描述子匹配损失用于衡量生成的特征描述子与真实描述子之间的相似性。总损失函数是这两部分损失的加权和,通过调整权重系数,可以平衡特征点检测和描述子匹配的效果,提升模型的整体性能。

2.3　模型与训练

本文实验设备配置情况为显卡 RTX 3080Ti,显存 12 G,内存 80 G,使用 PyTorch 框架进行网络模型的搭建和训练,Python 版本为 3.8,训练数据集选择 MS-COCO2014[23],训练过程中,首先生成一个由简单的 2D

几何图形组成的合成形状数据集,再将单应性变换应用于每个图像以增加训练数据的数量。模型在合成数据集上进行训练可以增加在真实图像上的特征点提取效果。由于室外场景中大部分区域都是平面区域,应用这种训练方式可以提升模型在视角不断变化的真实场景中进行特征提取和特征匹配的准确性。

本文算法训练过程中,通过前向传播将输入图像经过编码器和 CBAM 模块生成特征图,再经过特征点检测模块和特征描述子生成模块生成特征点热图和描述子;然后根据真实热图和生成的热图计算关键点检测损失,根据真实描述子和生成的描述子计算描述子匹配损失,最后通过反向传播算法计算梯度,使用优化器更新网络参数。在特征点匹配环节,首先根据预测的特征点热图选择概率最高的位置作为特征点,提取对应的特征描述子,再通过描述子之间的欧氏距离进行特征点匹配,找到相似度最高的特征点对,完成特征匹配任务。此外,为验证 CBAM 模块不同插入位置对匹配精度造成的影响,进行了不同插入位置(仅深层插入、跨层插入)的对比实验,将模块分别插入到了编码器的第 5 层卷积层和第 1、3、5 层卷积层。在模型训练过程中,发现跨层插入会导致过拟合的现象,而仅深层插入得到了良好的收敛,并能够在保证稳定性的基础上,增强特征提取效果,从而提高模型的特征提取和匹配性能。训练过程中具体的损失函数曲线。如图 3 所示,蓝色曲线表示仅深层插入时的损失函数值,可以看到训练损失和验证损失均平稳下降,而跨层插入时的验证损失在一定的训练轮次后上升,出现了过拟合现象。因此,本文模型将 CBAM 模块插入到了 SuperPoint 的第 5 层卷积层中,以保证能够有效提升模型的特征提取和匹配性能。训练过程中具体的损失函数曲线图如图 3 所示。

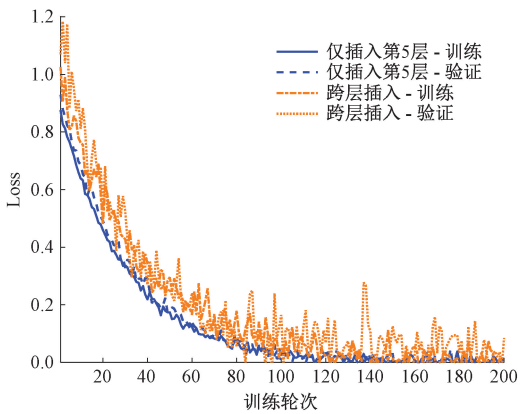


图 3　仅深层插入和跨层插入的损失函数曲线

Fig. 3　Loss function curves for deep insertion and cross layer insertion only

3　实验结果与分析

本文主要从 3 个部分来验证算法性能的有效性,分别是特征提取效果、特征匹配效果和在公开数据集上的建图轨迹精度,所用系统版本为 Ubuntu20.04,选取合适的室外场景数据集分别验证本文算法的特征提取和匹配能力并使用公开数据集 KITTI 对本文算法的建图轨迹精度进行验证。

3.1　特征提取效果对比实验

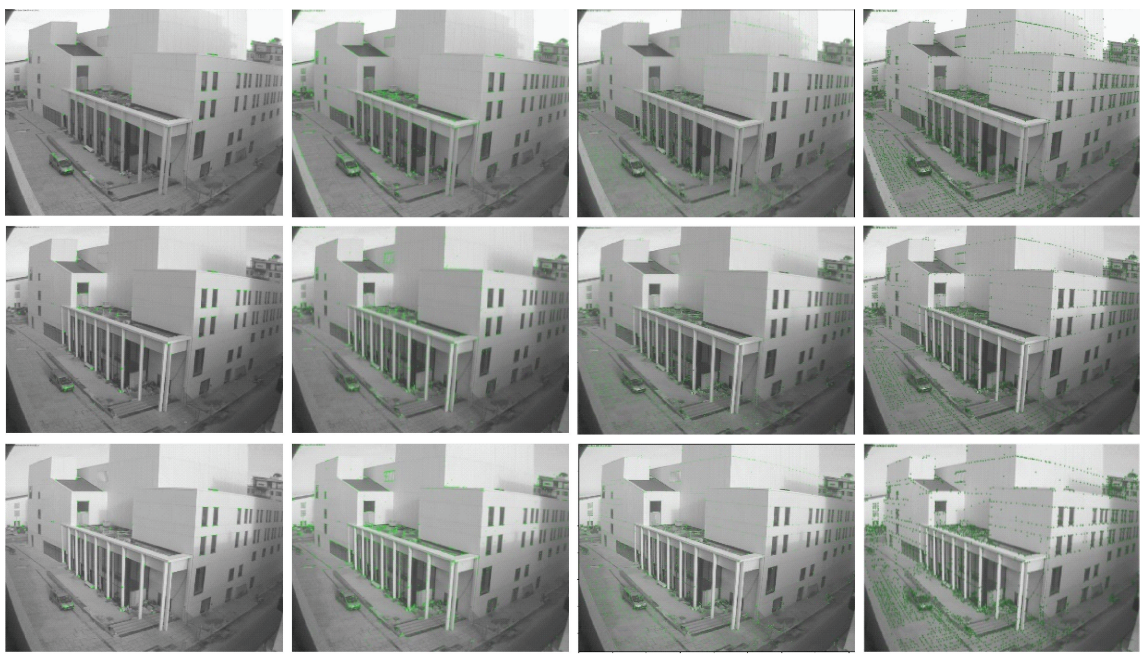
HPatches 数据集[24]中包含了本文所提到的光照强度变化较大和特征纹理重复性较高的室外真实场景图像序列,HPatches 数据集用于评估特征点检测的鲁棒性,特别是在光照变化明显和视角变化条件下的稳定性。本文采用数据集中多个场景的图像序列,并针对 CBAM 改进后的 SuperPoint 进行关键点检测和匹配准确率评估。以其中的 i_castle 序列和 i_whitebuilding 序列为例进行特征提取的效果展示,并将结果分别与 ORB、SIFT、SuperPoint 算法特征提取结果进行对比。4 种算法在两个序列上的特征提取效果对比结果如图 4 和 5 所示。从图 4(a)可以看出,在同一场景不同光照强度环境下,ORB 算法的提取的特征点分布不均匀,且局部较为密集。这是由于在具有丰富纹理或强烈对比的图像区域,FAST 算法会检测到大量角点[22],导致特征点在这些区域呈现聚集分布。由图 4(b)可知,SIFT 算法虽然在光照变化的条件下具有一定的鲁棒性,但特征点分布仍不均匀。由图 4(c)可知,Superpoint 算法在光照变化明显的图像中提取的特征点数量变化较大。由图 4(d)可知,本文算法使用通道空间注意力机制融合了空间位置信息,经过增强的特征图在描述子生成时能够更准确地捕捉重要的局部特征,进一步提升了特征提取的准确性和鲁棒性,通过选择性地关注重要的特征通道和空间位置,增强了对关键特征的响应。在光照变化明显的场景中也能保持特征点数量稳定性。

由图 5(a)可以看出,在灰度均匀和纹理特征重复性高的场景中,ORB 算法的特征点检测不足、特征点描述易混淆。这是由于 FAST 检测器很难找到足够的角点,即使检测到了角点,BRIEF 描述符由于基于像素强度比较的方法,也难以生成具有强辨别性的特征描述[25]。由图 5(c)可知,SuperPoint 算法的深度学习特性虽然可以在一定程度上缓解上述情况,但在完全灰度均匀的区域,它仍然可能无法提取到有效的特征点。此外也会存在特征点提取困难或不准确的问题,在光照强度变化明显的图像中,特征点数量稳定性较低。由图 5(d)可知,本文算法能够更有效地关注到图像中的重要信息,忽略重复的特征,增强了在纹理重复性较高的场景下的特征提取

(a) ORB　(b) SIFT　(c) SuperPoint　(d) 本文算法
(d) The proposed algorithm

图 4　4 种算法在光照变化明显的场景中的特征提取效果

Fig. 4　Feature extraction performance of four algorithms in scenes with significant illumination changes



(a) ORB　(b) SIFT　(c) SuperPoint　(d) 本文算法
(d) The proposed algorithm

图 5　4 种算法在纹理重复性高的场景中的特征提取效果

Fig. 5　Feature extraction performance of four algorithms in scenes with high texture repetition

能力。

实验结果表明,本文所提出的算法相较于其他 3 种算法,在光照变化明显、纹理重复性高的场景下均展示出了较强的特征提取能力。

3.2　特征匹配效果对比实验

为验证本文算法的特征匹配能力,选用 HPatches 数

据集中纹理重复性高的多视角变化图像序列进行特征匹配对比实验,选取单应性估计(*Epsilon* = 1, 3, 5)、检测器指标(重复性和最大似然估计),以及描述符指标(最近邻平均精度和匹配分数)等作为算法匹配性能的评价指标[26]。单应性估计是衡量算法对图像之间透视变换的鲁棒性的指标,Epsilon 数值越小,表示对估计误差的要求越高[27]。重复性表示在不同图像之间重复检测到同一特征点的能力,数值越高,说明检测器的稳定性越好[28]。最大似然估计用于衡量模型对噪声的适应性,值越小表示检测器的预测越稳定。噪声影响主要发生在光照变化、运动模糊、纹理重复性高等情况下,使得特征点检测不稳定。最近邻平均精度和匹配分数分别用来衡量描述符在特征匹配中的精度和效果,数值越高越好。为验证 2.1 节所提出的 CBAM 模块两种插入方式对匹配精度的影响,特征匹配对比实验分别选取了 ORB、SIFT、SuperPoint 和 CBAM 模块跨层插入和仅深层插入 SuperPoint 5 种算法进行对比,实验结果如表 1 所示。

表 1　Hpatches 数据集特征匹配性能对比结果

Table 1　Comparison of feature matching performance on the HPatches dataset

算法	Epsilon			检测器指标		描述符指标	
	1	3	5	重复性	最大似然估计	最近邻平均精度	匹配分数
ORB	0.150	0.395	0.538	**0.641**	1.157	0.735	0.266
SIFT	**0.424**	0.676	0.759	0.495	**0.833**	0.694	0.313
SuperPoint	0.310	0.684	0.829	0.581	1.158	0.821	0.470
跨层插入	0.380	0.700	0.850	0.600	1.095	0.825	0.480
仅深层插入(本文)	0.401	**0.708**	**0.875**	0.612	1.104	**0.836**	**0.503**

由表 1 可以看出,SIFT 对于 *Epsilon* = 1 的情况表现良好,并且具有最低的最大似然估计值。ORB 的重复性最高,但它的检测方式往往会在整个图像中形成稀疏簇,因此在最终的单应性估计任务中得分较低。SuperPoint 在以描述符为中心的指标中得分高于前两种算法。本文算法在单应性估计中表现良好,在检测器指标方面,通道注意力和空间注意力机制可以有效地过滤掉冗余信息,突出显示与特征匹配任务相关的区域。因此,本文算法在特征检测时能够更稳定地检测到同一特征点。本文在室外与纹理密集、重复性高的环境中进行对比,使用 HPatches 数据集中的视角变换图像序列验证本文算法的图像匹配性能,选取了 Hpatches 数据集中的 v_gardens、v_home 和 v_dogman 3 个序列不同的视角变化角度进行特征匹配,不同视角变化下的特征匹配情况如图 6 所示。

实验结果表明,不同位置插入 CBAM 模块对特征提取与匹配精度有明显影响。特别是将 CBAM 模块插入编码器的第 5 层卷积层,即“仅深层插入”方案,在训练稳定性和验证损失上均表现优于“跨层插入”方案,有效避免了过拟合现象,验证了注意力模块在该层级的嵌入最为合理。该消融实验验证了注意力机制对视觉 SLAM 系统性能提升的有效性和关键性。

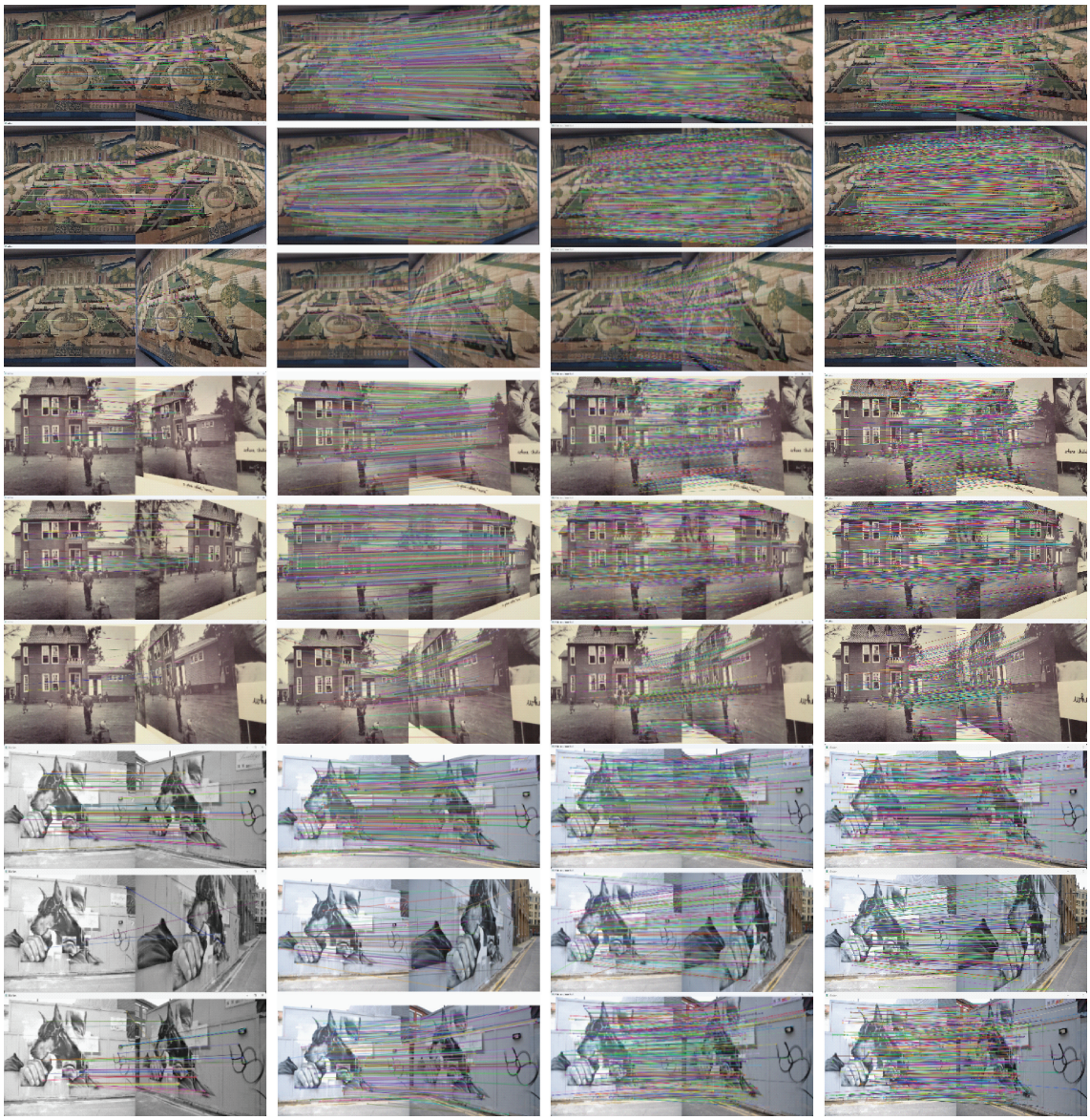
在纹理重复性高的场景中,ORB 算法基于简单的二值化特征比较,描述符指标中的匹配分数较低,SIFT 算法的最大似然估计值较低,但在多视角变化的图像序列中误匹配率较高。SuperPoint 描述符能捕获更复杂的特征模式和语义信息,在尺度变化和旋转不变性上表现较为优越,但也出现了误匹配的现象。本文算法在纹理重复性高的场景下特征提取与匹配能力较强。

3.3　公开数据集轨迹对比实验

使用 KITTI 数据集验证室外复杂特征点环境情况下本文算法建图的性能。选择了 00~10 序列,这些序列包含了符合本文研究场景的环境条件,如公园场景、城市街道、乡村道路等。本文对数据集进行了归一化预处理,以适配 SuperPoint-SLAM 和本文算法的输入格式要求。为了验证本文提出算法的有效性,将本文算法与 ORB-SLAM2 算法和 SuperPoint-SLAM 算法相对比,3 种算法的建图轨迹(图 7)并计算绝对轨迹误差(absolute trajectory error, ATE)与相对位姿误差(relative pose error, RPE)的均方根(root mean square error, RMSE)作为评价指标,建图轨迹对比结果。

表 2 为 3 种视觉 SLAM 算法在 KITTI 数据集上的建图精度对比结果。

从整体结果来看,本文算法在 KITTI 数据集的 6 个序列的建图精度都高于其他两种对比算法。由图 7 和表 2 可以看出,3 种算法在 KITTI 数据集的 09 序列上的建图效果差异较为明显,且 ORB-SLAM2 算法和 SuperPoint-SLAM 算法均出现了未检测到闭环的现象。这是由于 09 序列中主要包含一些光照变化较大、速度较快的场景,导致传统的 ORB-SLAM2 算法在这些区域上难以提取和匹配到图像中足够的特征点,故 ORB-SLAM2 在 09 序列的运行过程中未能检测到闭环,且由于初始的位姿估计不准确,导致建图过程中轨迹地图的累积误差逐渐增大,最终导致系统跟丢。SuperPoint-SLAM 算法未出现大幅跟踪丢失的情况,但由于 SuperPoint-SLAM 算法生成的特征点可能在视觉上相似但属于不同的物理位置,生成的特征描述符可能在某些情况下不够独特或稳定,在序列的



(a) ORB　(b) SIFT　(c) SuperPoint　(d) 本文算法
(d) The proposed algorithm

图 6　HPatches 数据集视角渐变下特征匹配对比实验

Fig. 6　Feature matching comparison experiment under viewpoint variations in the HPatches dataset

表 2　KITTI 数据集轨迹误差对比

Table 2　Comparison of trajectory errors on the KITTI dataset

序列	ORB-SLAM2		SuperPoint-SLAM		本文	
	RPE	ATE	RPE	ATE	RPE	ATE
00	**0. 305**	8. 149	0. 944	7. 841	0. 777	**6. 319**
02	**0. 903**	32. 086	—	—	—	—
03	0. 073	1. 016	0. 068	1. 954	**0. 065**	**0. 942**
04	0. 104	1. 386	0. 090	1. 337	**0. 017**	**0. 175**
05	0. 401	6. 859	0. 401	7. 656	**0. 353**	**6. 816**
06	**0. 659**	**12. 770**	0. 954	17. 214	0. 732	15. 908
07	0. 190	**2. 218**	**0. 152**	2. 779	0. 174	4. 305
08	0. 727	**47. 194**	**0. 565**	60. 982	0. 687	55. 641
09	12. 056	199. 787	0. 407	49. 899	**0. 122**	**6. 638**
10	0. 203	8. 849	0. 144	7. 617	**0. 131**	**6. 548**

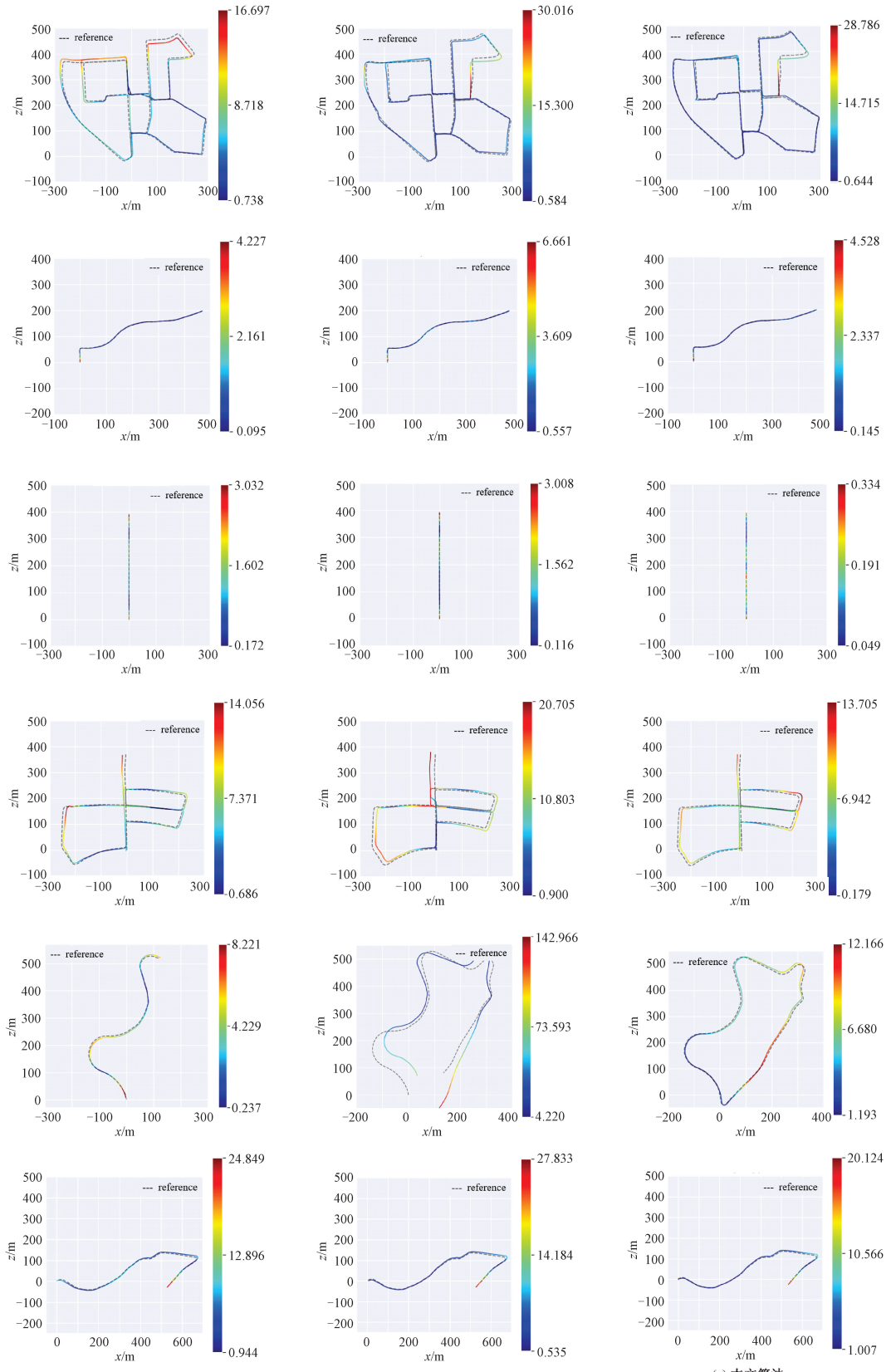


图 7　KITTI 数据集轨迹对比

Fig. 7　Trajectory comparison on the KITTI dataset

最后部分难以正确识别已访问过的区域,导致了错误的匹配,从而在闭环检测时失败,未能生成完整的轨迹地图。本文算法在对网络的编码层的改进过程中加入了通道-空间注意力机制,增强了网络对重要特征的感知能力,同时抑制了不重要的背景信息,这种机制有助于在复杂场景中提取更鲁棒的特征。通过聚焦于图像中的关键区域,增强了特征描述符的区分度,从而提高了匹配的准确性。故本文算法在处理具有复杂特征点的情况下表现更好。

由表 2 可知,从整体轨迹误差对比情况来看,SuperPoint-SLAM 算法在 KITTI 数据集上的绝对轨迹误差和相对位姿误差略低于 ORB-SLAM2 算法,本文算法在 KITTI 数据集的 10 个序列上相较 ORB-SLAM2 算法的平均绝对轨迹误差的 RMSE 值减小了 30.05%,平均相对位姿误差的 RMSE 值减小了 14.49%,相比 SuperPoint-SLAM 算法的平均绝对轨迹误差 RMSE 值减小了 28.62%,平均相对位姿误差 RMSE 值减小了 16.49%,表明本文算法在光照强度变化较大和具有复杂特征点的室外环境中性能优异。

4　结　论

本文围绕复杂室外场景中视觉 SLAM 系统在特征提取与匹配方面存在的精度不足和鲁棒性欠缺的问题,提出了一种融合通道-空间注意力机制的视觉 SLAM 算法,旨在提升系统在光照变化明显、纹理重复性高以及尺度几何形状多样等挑战性环境下的建图精度与定位稳定性。该算法通过将 CBAM 模块与 SuperPoint 网络的编码器融合,显著增强了图像特征的表达能力,实现了更稳健的特征点检测与描述,并将改进后的 SuperPoint 网络与 ORB-SLAM2 的后端相结合,构建了兼具深度学习表达能力与传统优化稳定性的视觉 SLAM 系统。实验结果表明本文算法在室外光照变化剧烈和特征点形状尺度多样的情况下表现出了更强的建图能力。未来的工作将聚焦于模型轻量化与部署优化,在保障匹配精度的前提下提升系统的实时性与适应性,进一步增强视觉 SLAM 系统在复杂环境下的应用能力与泛化性能。

参考文献

[1] 张耀,吴一全,陈慧娴. 基于深度学习的视觉同时定位与建图研究进展[J]. 仪器仪表学报,2023,44(7):214-241.
ZHANG Y, WU Y Q, CHEN H X. Research progress of visual simultaneous localization and mapping based on deep learning [J]. Chinese Journal of Scientific Instrument, 2023, 44(7):214-241.

[2] 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.
HUANG Z X, SHAO CH L. Survey of visual SLAM based on deep learning [J]. Robot, 2023, 45(6): 756-768.

[3] MAO X R, LIU K M, HANG Y F. Feature extraction and matching of SLAM image based on improved SIFT algorithm [C]. International Academy of Computer Technology (IACT), 2020:72-77.

[4] 王朋,郝伟龙,倪翠,等. 视觉 SLAM 方法综述[J]. 北京航空航天大学学报,2024,50(2):359-367.
WANG P, HAO W L, NI C, et al. An overview of visual SLAM methods [J]. Journal of Beijing University of Aeronautics and Astronautics, 2024, 50(2):359-367.

[5] CHEUNG W, HAMARNEH G. n-SIFT: n-dimensional scale invariant feature transform[J]. IEEE Transactions on Image Processing, 2009, 18(9):2012-2021.

[6] 行芳仪,徐成,高宏伟. 高效高精度光照自适应的 ORB 特征匹配算法[J]. 电子测量与仪器学报,2023,37(7): 140-147.
XING F Y, XU CH, GAO H W. Efficient and high-precision illumination adaptive ORB feature matching algorithm[J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(7): 140-147.

[7] SUN H C, WEI L F, JIANG Y X. Research on mobile robot localization algorithm with improved ORB extraction matching[C]. 2024 IEEE International Conference on Mechatronics and Automation (ICMA), 2024: 351-356.

[8] 梁继然,陈壮,董国军, 等. 结合注意力机制和密集连接网络的车辆检测方法[J]. 电子测量与仪器学报,2022,36(3):210-216.
LIANG J R, CHEN ZH, DONG G J, et al. Vehicle detection method combining attention mechanism and dense connection network [J]. Journal of Electronic Measurement and Instrumentation, 2022, 36(3): 210-216.

[9] BAROSSO-LAGUNA A, MIKOLAJCZYK K. Key. Net: Keypoint detection by handcrafted and learned CNN filters revisited [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(1): 698-711.

[10] TANG J, ERIKSSON L, FOLKESSON J, et al. GCNv2: Efficient correspondence prediction for real-time SLAM[J]. IEEE Robotics and Automation Letters, 2019, 4(4): 3505-3512.

[11] 杨锋,丁之桐,邢蒙蒙, 等. 深度学习的目标检测算法改进综述[J]. 计算机工程与应用,2023,59(11): 1-15.
YANG F, DING ZH T, XING M M, et al. Review of object

detection algorithm improvement in deep learning [J]. Computer Engineering and Applications, 2023, 59(11): 1-15.

[12] SARLIN P E, DETONE D, MALISIEWICZ T, et al. SuperGlue: Learning feature matching with graph neural networks[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020: 4937-4946.

[13] ZHAO K, HE B, PAN S, et al. Siamese network with multi-scale feature fusion and dual attention mechanism for template matching[C]. 2022 41st Chinese Control Conference (CCC),2022: 6588-6592.

[14] 刘冬,于涛,丛明, 等. 基于深度学习图像特征的动态环境视觉SLAM方法[J]. 华中科技大学学报(自然科学版),2024,52(6):156-163.
LIU D, YU T, CONG M, et al. Visual SLAM method for dynamic environment based on deep learning image features[J]. Journal of Huazhong University of Science and Technology (Natural Science Edition), 2024, 52(6): 156-163.

[15] HOU Y. An end-to-end convolutional neural network model for autonomous driving[C]. 2023 Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC), 2023: 360-365.

[16] DETONE D, MALISIEWICZ T, RABINOWICH A. SuperPoint: Self-supervised interest point detection and description [C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2018.

[17] FENG J, ZHAO X, ZHU T, et al. Detection mature bud for daylily based on Faster R-CNN integrated with CBAM[J]. IEEE Access, 2023, 11: 81646-81655.

[18] ZHAO J, YEUNG A, ALI M, et al. CBAM-SwinT-BL: Small rail surface defect detection method based on swin transformer with block level CBAM enhancement [J]. IEEE Access, 2024,12:181997-182009.

[19] DONG J, WANG N, FANG H, et al. CBAM-optimized automatic segmentation and reconstruction system for monocular images with asphalt pavement potholes[J]. IEEE Transactions on Intelligent Transportation Systems, 2024, 25(8): 10313-10330.

[20] JIN Y, YE X, FENG N, et al. Lesion classification of coronary artery CTA images based on CBAM and transfer learning[J]. IEEE Transactions on Instrumentation and Measurement, 2024,DOI: 10.1109/TIM.2024.3385035.

[21] CHANG J, DONG N, LI D. A real-time dynamic object segmentation framework for SLAM system in dynamic scenes[J]. IEEE Transactions on Instrumentation and Measurement, 2021, DOI: 10.1109/TIM.2021.3109718.

[22] LUO Y, RAO Z, WU R. FD-SLAM: A semantic SLAM based on enhanced Fast-SCNN dynamic region detection and DeepFillv2-Driven background inpainting[J]. IEEE Access, 2023, 11: 110615-110626.

[23] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: Common objects in context [C]. Computer VisionECCV 2014. Cham: Springer International Publishing, 2014: 740-755.

[24] BALNTAS V, LENC K, VEDALDI A, et al. HPatches: A benchmark and evaluation of handcrafted and learned local descriptors [C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 3852-3861.

[25] 董蕊芳,王宇鹏,阚江明. 基于改进ORB_SLAM2的机器人视觉导航方法[J]. 农业机械学报,2022,53(10): 306- 317.
DONG R F, WANG Y P, KAN J M. Visual navigation method for robot based on improved ORB_SLAM2[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(10):306-317.

[26] 徐冰冰,岑科廷,黄俊杰, 等. 图卷积神经网络综述[J]. 计算机学报,2020,43(5):755-780.
XU B B, CEN K T, HUANG J J, et al. A survey on graph convolutional neural network[J]. Chinese Journal of Computers, 2020, 43(5):755-780.

[27] 盖绍彦,黄妍妍,达飞鹏. 基于通道注意力和特征切片的图像快速匹配算法[J]. 光学学报,2023,43(22): 158-166.
GAI SH Y, HUANG Y Y, DA F P. Fast image matching based on channel attention and feature slicing[J]. Acta Optica Sinica, 2023, 43(22):158-166.

[28] 周非,陈帅,吴凯, 等. 快速跟踪分割辅助的动态SLAM[J]. 仪器仪表学报,2023,44(5):313-321.
ZHOU F, CHEN SH, WU K, et al. Dynamic SLAM assisted by fast tracking segmentation [J]. Chinese Journal of Scientific Instrument, 2023, 44(5):313-321.

作者简介

马金睿,2022年于河北工业大学获得学士学位,现为北京林业大学硕士研究生,主要研究方向为视觉SLAM和机器人路径规划。
E-mail: 13653167091@163.com

Ma Jinrui received her B. Sc. degree from Hebei University of Technology in 2022. Now she is a M. Sc. candidate at Beijing Forestry University. Her main research interests include visual SLAM and robot path planning.

梁浩(通信作者),2012 年于东北林业大学获得学士学位,2014 年于东北林业大学获得硕士学位,2017 年于东北林业大学获得博士学位,现为北京林业大学工学院副教授,主要研究方向为机器视觉、智慧林业、人工智能和无损检测。

Liang Hao (Corresponding author), received his B. Sc. degree from Northeast Forestry University in 2012, M. Sc. degree from Northeast Forestry University in 2014, and Ph. D. degree from Northeast Forestry University in 2017. Now he is an associate professor at Beijing Forestry University. His main research interests include machine vision, smart forestry, artificial intelligence and non-destructive testing.

林剑辉,2000 年于中国农业大学获得学士学位,2007 年于中国农业大学获得博士学位,现为北京林业大学工学院副教授,主要研究方向为智慧林业、智能检测与信号处理和精准灌溉控制技术。

E-mail: swiq_lin@ 163. com

Lin Jianhui received his B. Sc. degree from China Agricultural University in 2000 and Ph. D. degree from China Agricultural University in 2007. Now he is an associate professor at Beijing Forestry University. His main research interests include smart forestry, intelligent detection and signal processing, precision irrigation control technology.