E-mail: jig@aircas.ac.cn
Website: www.cjig.cn
Tel: 010-58887035
中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS

中图法分类号:TP391.4　　**文献标识码**:A　　**文章编号**:1006-8961(2025)07-2484-15

论文引用格式:Hao W, Lyu Y, Jin H Y and Shi Z H. 2025. Large-scale point cloud place recognition method based on attention mechanism and rotation invariant features. Journal of Image and Graphics, 30(7):2484-2498(郝雯, 吕炎, 金海燕, 石争浩. 2025. 融合注意力机制与旋转不变特征的大规模点云地点识别. 中国图象图形学报, 30(7):2484-2498)[DOI:10.11834/jig.240593]

融合注意力机制与旋转不变特征的大规模点云地点识别

郝雯[1,2*],吕炎[1],金海燕[1,2],石争浩[1,2]

1. 西安理工大学计算机科学与工程学院,西安　710048; 2. 陕西省网络计算与安全技术重点实验室,西安　710048

摘　要:**目的**　地点识别是机器人利用实时扫描到的点云数据进行定位和自主导航的核心。现有的针对大规模点云的地点识别方法往往忽略了真实驾驶中存在的旋转问题。当查询场景发生旋转时,这些方法的识别性能会显著下降,这严重阻碍了它们在复杂现实场景中的应用。因此,本文提出一种有效的面向三维点云的具有旋转感知的地点识别网络(efficient rotation-aware network for point cloud based place recognition,ERA-Net)。**方法**　首先,利用自注意机制与邻域注意力机制,在捕获点与点之间全局依赖关系的同时,捕获每个点与其邻域点之间的局部依赖关系,充分提取点间的语义特征。同时,利用点与其k邻近点的坐标信息,计算距离、角度以及角度差等低维几何特征,并设计基于特征距离的注意力池化模块,通过在高维空间分析特征之间的相关性,提取具有较强区分性且具有旋转特性的几何特征。最后,将提取的语义特征以及几何特征进行有效融合,通过NetVLAD模块,产生更具判别性的全局描述符。**结果**　将提出的ERA-Net在公共数据集Oxford Robotcar上进行验证并与当前先进的方法(state-of-the-art method,SOTA)进行比较。在Oxford数据集中,ERA-Net的前1%平均召回率(average recall@1%,AR@1%)指标可以达到96.48%,在U.S.(university sector)、R.A.(residential area)以及B.D.(business district)数据集上的识别效果均优于其他方法。特别地,当查询场景进行旋转时,ERA-Net的识别效果优于已有方法。**结论**　ERA-Net能够充分考虑点间的上下文信息,以及特征间的相关性,提取具有较强独特性的场景特征,在面对旋转问题时能够展现出较好的鲁棒性,具有较强的泛化能力。

关键词:点云场景;地点识别;旋转感知;注意力机制;特征距离

Large-scale point cloud place recognition method based on attention mechanism and rotation invariant features

Hao Wen[1,2*], Lyu Yan[1], Jin Haiyan[1,2], Shi Zhenghao[1,2]

1. Department of Computer Science, Xi'an University of Technology, Xi'an　710048, China;

2. Shaanxi Key Laboratory for Network Computing and Security Technology, Xi'an　710048, China

Abstract: **Objective** Point cloud-based place recognition involves identifying and matching a specific location within a pre-built map by comparing global descriptors generated from a scene. This process is crucial for robot localization and autonomous navigation, wherein real-time point clouds are continuously scanned. For example, autonomous vehicles and robots

收稿日期:2024-11-13;**修回日期**:2025-03-03;**预印本日期**:2025-03-10

*通信作者:郝雯　haowensxsf@163.com

基金项目:国家自然科学基金项目(61602373);陕西省教育厅青年创新团队项目(24JP116);中国国家留学基金资助

Supported by: National Natural Science Foundation of China (61602373); Shaanxi Provincial Department of Education Youth Innovation Team Project (24JP116); China Scholarship Council Program

rely on robust place recognition to navigate safely and efficiently, ensuring that they can recognize and reidentify locations within a scene. The primary goal of place recognition is to enable a system to determine accurately whether it has previously encountered a location. This ability is essential for certain tasks, such as loop closure in simultaneous localization and mapping and re-localization in dynamic environments. Place recognition methods can be categorized into two approaches in accordance with data type: image-based and LiDAR-based methods. Image-based methods have achieved promising results by using 2D images in the place recognition task. However, the performance of image-based methods is easily degraded in real-world environments due to illumination changes and viewpoint variation. Point clouds can be captured easily by 3D sensors, and they provide high-precision geometry information of objects. Hence, misleading caused by lighting change and viewpoint variation can be avoided to a certain extent. Although many methods have been proposed for 3D place recognition, these methods typically perform well in prescribed travel routes. However, a significant challenge that is frequently overlooked is the effect of data rotation. In real-world scenarios, vehicles do not travel on fixed routes. When a vehicle deviates from a prescribed route in the same place, the acquired point clouds are rotated relative to the submaps. Existing location recognition methods for large-scale point clouds often ignore the rotation problem in real driving. When a query scene is rotated, the recognition performance of these methods degrades significantly, seriously hindering their application in complex real-world scenarios. **Method** In this study, we propose an efficient rotation-aware network for point cloud-based place recognition (ERA-Net) for point cloud-based place recognition that leverages semantic and geometric features to achieve better discriminative ability. ERA-Net employs an attention mechanism that thoroughly considers global dependencies among points and local correlations between a point and its neighboring points to extract semantic features. In addition, it utilizes the coordinate information of a point and its k-nearest neighbors to derive low-dimensional geometric features that exhibit rotational invariance, such as distance, angle, and angular difference. By calculating the associations of these features in high-dimensional space, ERA-Net extracts geometric characteristics that demonstrate strong uniqueness and rotational invariance. First, global dependencies among points are captured using a self-attention mechanism, while local dependencies between each point and its k-nearest neighbor points are modeled through neighboring attention mechanisms. These mechanisms ensure the comprehensive extraction of attention features among points. The self-attention mechanism computes self-coefficients by considering the geometric properties of individual points, while the neighboring attention mechanism focuses on local coefficients by considering the neighborhood structure. Simultaneously, several low-dimensional geometric features are calculated using the coordinates of points. The sampled points, their local center of mass, and k-nearest neighbors form a stable local triangular structure. Low-dimensional geometric features, such as distances, angles, and angular differences, are calculated. These geometric features remain invariant to point cloud rotations. Subsequently, a novel feature distance-based attention pooling module is designed to extract geometric features with high discrimination and rotational invariance by analyzing correlations among features in high-dimensional space. Feature distance measures the similarity among different points in high-dimensional feature space. Points may be geometrically distant from one another in terms of their spatial positions. However, they can share similar features semantically. The feature distance-based pooling method captures the relationships among these geometrically distant but semantically similar points. It can also effectively distinguish points with similar spatial coordinates but considerable semantic differences. Finally, the extracted semantic and geometric features are fused through the NetVLAD module, generating more discriminative global descriptors for point clouds. **Result** The proposed ERA-Net is validated on the public dataset Oxford Robotcar and compared with state-of-the-art (SOTA) methods. In the Oxford Robotcar dataset, the average recall (AR)@1% of ERA-Net reaches 96.48% and the AR@1 reaches 90.47%. ERA-Net outperforms other SOTA methods on the U. S. (university sector), R. A. (residential area), and B. D. (business district) datasets. We select three representative queries of the Oxford dataset to verify the retrieval capability of ERA-Net. Although the scene contains structurally unstable trees, slender poles, or a large number of planar points, ERA-Net can retrieve the correct place. In particular, ERA-Net outperforms existing methods in terms of recognition when the query scene is rotated. The recognition performance of most existing methods decreases significantly when the query scene is rotated. The results of AR@1% and AR@1 of ERA-Net and other SOTA methods under different rotation angles are tested separately. The tested rotation angles include 30°, 60°, 90°, 120°, 150°, and 180°, and these angles represent the maximum rotation range. The query point cloud is randomly rotated

from 0° to the maximum rotation angle. The performance of ERA-Net is better than those of the other methods when the rotation angle of the query scene is increasing. The experimental results show that ERA-Net can still extract features with strong uniqueness and rotation invariance from the scene and correctly retrieve the corresponding scene even when facing the rotation of point cloud data. **Conclusion** In this study, we propose a rotation-aware place recognition network, i. e., ERA-Net, which leverages semantic and geometric features. The experimental results show that the recognition performance of ERA-Net is better than those of SOTA methods. ERA-Net performs significantly better than other methods on the three in-house datasets. For AR@1%, ERA-Net reaches 91. 04%, 87. 37%, and 87. 12% on the U. S., R. A., and B. D. datasets, respectively. Simultaneously, ERA-Net achieves better results in the rotation test. The extraction of low-dimensional geometric features can better improve the recognition effect of ERA-Net. The attention pooling module based on feature distance can capture the relationship among distant but semantically similar points and effectively distinguish points with similar coordinates but considerable semantic differences. ERA-Net can extract scene features with strong uniqueness by fully considering the contextual information among points and the correlation among features. It exhibits good robustness and strong generalization ability when facing the rotation problem.

Key words: scene point clouds; place recognition; rotation-aware; attention mechanism; feature distance

0 引 言

地点识别是机器人学和计算机视觉领域的一项基本任务,其目标是通过分析环境的几何特征,检索与查询地图中最相似的地点,已广泛应用于自动驾驶、机器人导航以及同步定位与绘图(simultaneous localization and mapping,SLAM)等多个领域。在自动驾驶中,准确的地点识别能够帮助车辆更好地理解周围环境,从而提高导航的安全性和准确性。目前很多学者通过图像特征的提取与匹配,完成地点识别(Ali-Bey等,2023;刘熠晨 等,2024;Izquierdo和Civera,2024)。然而,由于光照变化和视点变化,基于图像的方法在现实环境中的性能很容易下降。与图像相比,点云能够捕捉到更加细致和全面的空间信息,可以在一定程度上避免光照变化和视点变化造成的信息误导。此外,三维扫描技术的快速发展使得人们能够快速便捷地获取大规模场景的点云数据。因此,基于点云的地点识别已成为一个热门研究课题。

早期的基于点云的地点识别方法往往基于人工设计提取点的几何特征,并对特征进行统计分析,形成独特的描述符,然后通过相似性度量等手段完成场景识别或闭环检测(Dubé等,2017;He等,2016;Röhling等,2015;Rizzini,2017)。这些方法提取特征过程繁复耗时,易受到环境干扰,且局限于特定的模型与特定的应用。随着深度学习技术的快速发展,越来越多的学者通过设计不同的卷积神经网络(convolutional neural network,CNN)学习点的深层次特征,然后将学习到的特征编码为特定维度的全局描述符向量,最后通过全局描述符比对,从场景数据库中检索得到结构上最接近查询点云的场景(郝雯等,2022)。根据数据表示形式不同,基于深度学习的地点识别方法可分为基于点的方法、基于体素的方法以及基于投影的方法(Zhang等,2024)。PointNetVLAD(Uy和Lee,2018)是第1个用于大规模三维点云地点识别的卷积神经网络,它将PointNet(Qi等,2017)和NetVLAD(Arandjelovic等,2016)结合起来,利用PointNet提取点的全局特征,通过NetVLAD网络对特征进行聚合。由于PointNet对点云局部特征的提取能力不足,这对生成的全局描述符的独特性有一定的影响。同时,该网络并未考虑局部邻域的空间分布关系。继PointNetVLAD之后,PCAN(point contextual attention network)(Zhang和Xiao,2019)将注意力机制整合到NetVLAD中,将注意力加权的局部特征汇总为全局描述符。EPC-Net(Hui等,2022)提出一种轻量级的边卷积模块ProxyConv,它利用空间相邻矩阵和代理点简化了原始边缘卷积,从而减少内存消耗。通过ProxyConv模块,构造代理点卷积神经网络,聚合多尺度局部几何特征,完成地点的识别。

在真实的驾驶中,汽车的行驶路线不是固定的,当车辆在同一地点不按规定路线行驶时,获取的点云相当于相对预构建地图进行了旋转。图1(a)为Oxford RobotCar数据集(Maddern等,2017)中部分行驶路线,图1(b)为该路线扫描得到的点云场景数

据。图 1(c)为图 1(a)场景中(紫色方框)一个直线行驶路线的实景图。红色箭头为汽车行驶方向，如果汽车不按原定路线行驶，按照绿色箭头方向行驶，那么扫描得到的点云数据相对于预先构建的地图旋转了 180°。图 1(d)为一个 T 型路口的实景图(图 1(a)中橙色方框)，红色箭头为汽车行驶方向，如果汽车不按原定路线行驶，按照绿色箭头方向行驶，扫描得到的点云数据相对于预先构建的地图旋转了 90°。目前已有的地点识别方法并未过多地考虑数据旋转的情况，当查询点云发生旋转时，其识别性能会严重下降。因此，针对已有方法不能较好地处理地点识别中数据旋转的问题，本文提出一种新的基于点云的高效旋转感知场景识别网络(efficient rotation-aware network for point cloud based pace recognition，ERA-Net)，主要工作如下：1)分别利用自注意机制捕获点与点之间的全局依赖关系，利用邻域注意力机制分析点与邻域点间的局部相关性，动态获取注意力权重信息，从而加强关键点的影响，并利用多头机制，将提取的多个语义特征进行拼接，保证上下文注意力特征提取的鲁棒性。2)根据采样点、k 近邻点以及局部中心点，计算具有旋转不变特性的低维几何特征，并设计一个新的基于特征距离的注意力池化模块，将这些低维几何特征作为输入，通过分析高维空间中特征间的距离信息，充分挖掘低维几何特征间的相关性，提取具有较强不变性与旋转鲁棒性的几何特征，提高网络的识别性能。3)将本文提出的 ERA-Net 在公共数据集 Oxford RobotCar 上进行验证，实验结果表明，ERA-Net 在 Oxford 数据集中，前 1% 平均召回率达到 96.48%。对于未知的内部数据集，ERA-Net 的识别结果优于其他最先进的方法(state of the art，SOTA)。特别地，当查询点云进行旋转时，ERA-Net 的识别效果也优于其他方法。

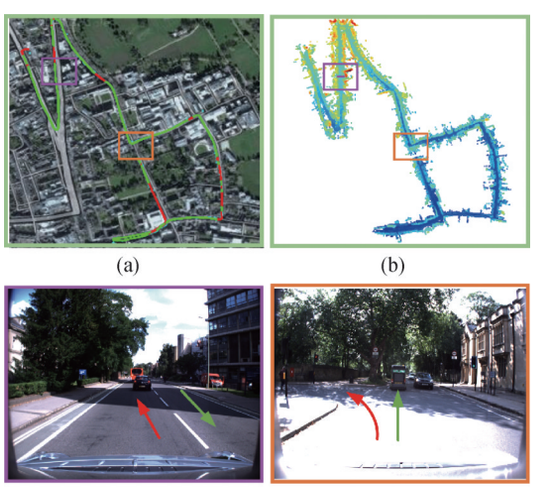

图 1　Oxford RobotCar 数据集中数据旋转示意图

Fig. 1　Schematic diagram of data rotation in the Oxford RobotCar dataset

1　相关工作

目前基于点的地点识别方法可以归为两类：基于注意力机制的点云地点识别方法以及基于图的点云地点识别方法。

1.1　基于注意力机制的点云地点识别方法

基于注意力机制的地点识别方法利用注意力机制，学习点云数据中不同点的重要性以增强对特定特征的关注，从而对关键点的局部特征进行自适应加权。

SOE-Net (self-attention and orientation encoding network)(Xia 等，2021)提出一个点方向编码模块，对 8 个方向的邻域信息进行编码，同时利用注意力机制捕获局部描述符之间的远程特征依赖关系。然而，SOE-Net 只能覆盖有限的空间方向差异，这也限制了其对未知环境的泛化能力。SphereVLAD++(Zhao 等，2023)将点云投影到每个独特区域的球面透视图上，并通过注意力机制捕捉局部特征之间的上下文联系及其与全局三维几何分布之间的依赖关系。Retriever(Wiesmann 等，2022)在对三维点云地图压缩的基础上，利用 PointNet 提取特征，然后利用基于感知器的注意力聚合模块 Perceiver(Jaegle 等，2021)将特征聚合为全局描述符。PTNet-PLT(point transformer network with pyramid learnable tokens)(Wen 等，2023)提出一个移动立方体注意力模块，该模块由自我注意力模块和交叉注意力模块组成，分别用于提取局部特征和聚合区域特征。为了获取不同分辨率的区域特征，PTNet-PLT 在每一层都加入了可学习标记，并逐层减少标记的数量。最后，将每一层的所有标记区域连接起来，得到用于场景识别的最终全局描述符。所提出的注意力模块可以学习局部特征之间的内在联系以及局部到全局的联系。SAGE-Net (spatial attention and geometric encoding network)(Hao 等，2024)利用空间注意力机制学习点间的语义信息，并通过基于动态图的邻域聚合模块动态获取特征间的空间分布信息。

同时，还有许多工作利用Transformer框架的自注意力机制来有效捕捉点云中的全局上下文信息。Hui等人(2021)设计了一个金字塔点变换器模块，通过利用分组的Transformer自注意力机制提取特征，自适应地学习不同尺度下点云不同区域之间的空间关系。Hou等人(2022)提出一种基于分层Transformer的点云地点识别方法。首先通过降采样和近邻搜索，将点云初步划分为一系列小单元，接着，通过短程Transformer(short-range Transformer，SRT)提取局部特征，通过长程Transformer(long-range transformer，LRT)提取全局上下文特征，最后，通过NetVLAD完成特征的聚合。Chen等人(2024)设计了一种基于变换器编码器的多层融合特征提取网络(multi-layer fusion feature extraction network，MFFEN)，利用动态图卷积网络提取具有局部关系的局部语义特征。此外，设计了一个语义引导的局部注意力网络(semantics-guided local attention network，SLAN)，识别场景的突出局部特征，从而增强全局描述符的描述能力。

基于注意力机制的方法比较关注点间的上下文特征的提取，而忽略了点间具有旋转不变性几何特征的提取。

1.2 基于图的点云地点识别方法

点云是非结构化的稀疏数据，点与点之间没有固定的关系。基于图卷积的点云处理方法将点云建模为图结构，通过卷积操作捕捉点与其邻域之间的关系。每个点视做图中的一个节点，点与点之间的连线视为边，通过图卷积操作对这些邻域点进行特征聚合。

LPD-Net(large-scale place description network)(Liu等，2019)首先计算每点的曲率、高度差、点密度等10个几何特征，并构建特征转换模块，将计算得到的局部特征映射到高维特征空间。然后，利用图卷积神经网络提取点间的空间分布信息。最后，通过NetVLAD网络对各种特征进行聚合，生成场景的全局描述符。DAGC(dual attention and graph convolution for point cloud based place recognition)(Sun等，2020)提出将图卷积与双注意力机制结合，对每个点的多层邻域进行卷积运算，进一步提取有效的局部特征。SRNet(scene recognition network)(Fan等，2020)使用静态图卷积神经网络提取点的局部几何特征，提出密集语义融合策略，通过重用浅层特征层补偿信息丢失。该类方法着重提取点间关系，对抽象语义信息提取不足，并且图卷积会消耗大量的内存。Kong等人(2020)提出一种基于语义图的大规模场景识别方法。首先对点云场景进行语义分割、实例分割以获取场景物体的语义类别，并进一步收集语义和拓扑信息形成的节点语义图。然后，原始点云场景被转换成拓扑语义图，对场景的识别转化为图匹配问题。该方法需要事先对场景中包含物体的语义类别进行定义，并且不能区分相同语义类别中的不同物体。Li等人(2024)首先对点云场景中的物体进行分割，将每个物体用结点表示，将点云转换为场景图结构。然后将图结构输入到变换器—注意力模块，提取整个场景的相对位置信息和全局图特征。最后利用多层感知机(multilayer perceptron，MLP)模块计算图相似度。P-GAT(pose-graph attentional graph neural network)(Ramezani等，2024)提出一种基于姿态图(pose-graph)和图神经网络(graph neural network，GNN)的点云场景识别方法。P-GAT利用姿态图SLAM，生成相邻云描述符之间的最大空间和时间信息，并采用多头注意力图神经网络对点云的上下文和视点信息进行聚合，形成每个点的描述符。Shu和Kwon(2024)提出一种分层双向图卷积网络(hierarchical bidirected graph convolution network，HiBi-GCN)，该模型使用条件概率作为有向边的权重，按从细到粗的方向构建分层图，并按从粗到细的方向获得几何特征。

基于图的方法往往依赖点间的几何邻近关系，并未考虑在高维空间特征间的关系。一些研究人员通过学习旋转不变的全局描述符来解决旋转问题，用于地点识别任务(Fan等，2022；Weng等，2022)。RPR-Net(rotation-aware large scale place recognition network)(Fan等，2022)以ARIConv为基本单元提取旋转不变特征，能够在保持旋转不变性的同时学习几何和语义特征。ERINet(Weng等，2022)设计了一个核心旋转不变性模块，增强了提取三维点云旋转不变性特征的能力。然而，上述这些方法并没有挖掘高维空间中特征间的距离对几何特征提取的促进作用。

2 ERA-Net网络模型

2.1 网络整体结构

图2为ERA-Net的网络结构图。ERA-Net包含语义特征提取分支(上层)和几何特征提取分支(下

层)。对于语义特征提取分支,首先利用T-Net(Qi等,2017)对点云数据进行对齐操作,然后经过语义特征提取模块后,将得到的语义特征与三维坐标信息进行拼接,送入两层MLP,得到语义特征;对于几何特征提取分支,通过某个点与k近邻点的坐标信息,计算点间的距离、角度与角度差等低维几何特征,送入两层MLP,将得到的特征作为几何特征提取模块的输入,得到几何特征F_{geo}。最后,将语义特征与几何特征F_{geo}拼接,送入NetVLAD模块,生成256维全局描述符。

2.2 语义特征提取模块

受Chen等人(2021)方法的启发,本文利用自注意机制与邻域注意力机制,提取点间的语义特征,如图3(a)所示。自注意力机制通过考虑每个点的几何信息学习自系数,而邻域注意力机制则通过考虑点的k近邻信息得到局部系数。然后通过非线性激活函数leakyReLU(leaky rectified linear unit)将两个系数融合在一起,获得注意力系数,并通过softmax函数对注意力系数进一步规范化。最后,将近邻特征与注意力系数相乘,得到点间的语义特征。

首先,将三维点云送入两个连续的MLP获得自系数$coef_{self}$,具体为

$$\begin{aligned} p_i' &= h(p_i,\theta) \\ coef_{self} &= h(p_i',\theta) \end{aligned} \tag{1}$$

式中,$h(\cdot)$是一个参数非线性函数,p_i是检索点的坐标,θ是可学习的参数。

同时,对于点p_i,利用K最近邻(K-nearest neighbor, KNN)查找其k近邻点$\{p_i^1, p_i^2, \cdots, p_i^j, \cdots, p_i^k\}$,计算点与邻近点的边信息$e_{ij} = p_i - p_i^j$,将边信息送入两个连续的MLP,从而获得邻域系数$coef_{neig}$,具体为

$$\begin{aligned} e_{ij}' &= h(e_{ij},\theta) \\ coef_{neig} &= h(e_{ij}',\theta) \end{aligned} \tag{2}$$

然后,将学习到的自系数与邻域系数相加,通过非线性激活函数leakyReLU将两个系数融合在一起,并利用softmax函数对系数进行归一化,具体为

$$f_{ij} = LeakyReLU(coef_{self} + coef_{neig}) \tag{3}$$

$$s_{ij} = \frac{\exp(f_{ij})}{\sum_k \exp(f_{ik})} \tag{4}$$

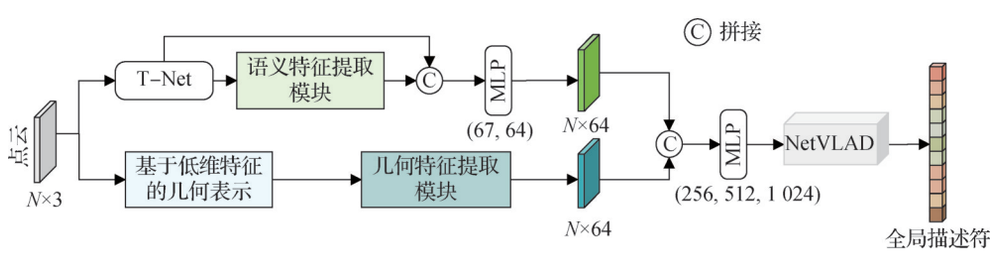

图2 ERA-Net网络结构

Fig. 2 The architecture of ERA-Net

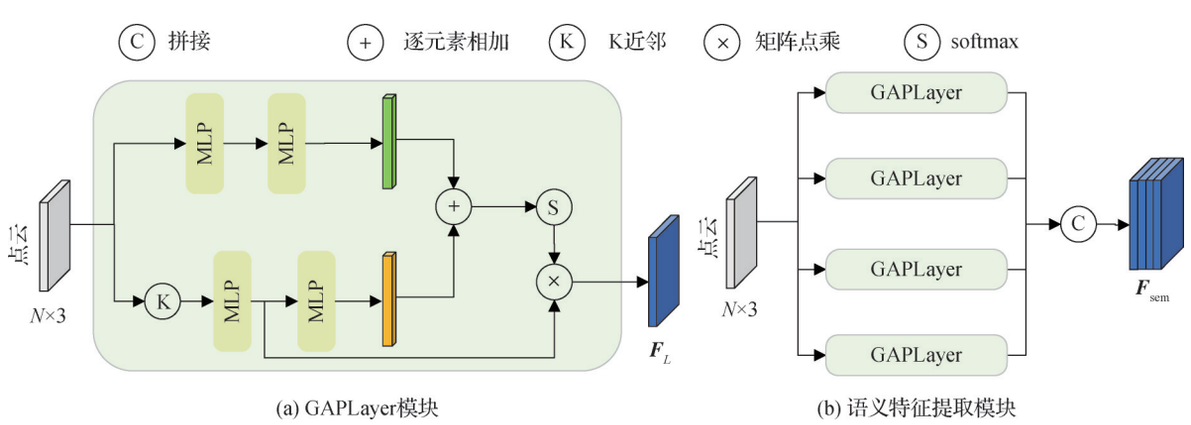

图3 语义特征提取模块网络结构

Fig. 3 The structure of the semantic feature extraction module ((a) GAPLayer module; (b) semantic feature extraction module)

最后，将近邻特征与权重矩阵相乘，获得点间的语义信息，具体为

$$\boldsymbol{F}_L = g\left(\sum s_{ij} e'_{ij}\right) \tag{5}$$

式中，$g(\cdot)$是非线性激活函数。将4个GAPLayer模块的输出特征进行拼接，得到语义特征$\boldsymbol{F}_{\text{sem}}$，如图3(b)所示。

2.3 基于低维特征的几何表示模块

本文拟通过点云场景中一点与其k邻近点构建一个局部稳定的三角形结构，计算边长、角度和角度差等低维几何特征。无论点云场景如何旋转，这些特征都保持不变(Fan等，2021；Zhang等，2022)。

图4为具有旋转不变性的低维几何特征计算示意图，点p_i(红色点)为点云场景中任意一点，灰色点为p_i的k近邻点，其中p_i^k是k近邻点中的一个点，该k近邻点的质心为p_i^c(黄色点)。p_i，p_i^k，p_i^c三点组成一个三角形T。在图4(b)(c)中，$p_i^{k'}$与$p_i^{c'}$是点p_i^k、p_i^c在xOy平面上的投影点。

首先，根据点p_i与其k邻近点构建三角形，计算边长与角度(图4(a))。计算组成三角形T的两条边$p_ip_i^k$与$p_i^kp_i^c$的边长d_1、d_2。将得到的欧氏距离取以e为底的负指数，得到最终的距离值dis_1和dis_2，具体为

$$\begin{cases} dis_1 = \exp(-d_1) \\ dis_2 = \exp(-d_2) \end{cases} \tag{6}$$

此外，计算三角形T中边$p_ip_i^k$与$p_ip_i^c$构成的夹角θ_1，$p_i^cp_i^k$与$p_i^cp_i$构成的夹角θ_2。

接着，计算角度差(图4(b))。将p_i^k和p_i^c投影至xOy平面，得到投影点$p_i^{k'}$和$p_i^{c'}$，利用式(7)分别计算$p_i^{k'}p_i$与x轴之间的夹角α_i^k以及$p_i^{c'}p_i$与x轴之间的夹角α_i^c，具体为

$$\alpha_i^k = \arctan\frac{\left|y_i - y_i^k\right|}{\left|x_i - x_i^k\right|} \tag{7}$$

同时，利用式(8)计算$p_i^kp_i$与$p_ip_i^{k'}$之间的角度β_i^k以及$p_i^cp_i$与$p_ip_i^{c'}$之间的角度β_i^c，具体为

$$\beta_i^k = \arctan\frac{\left|z_i - z_i^k\right|}{\sqrt{(x_i - x_i^k)^2 + (y_i - y_i^k)^2}} \tag{8}$$

利用式(9)计算两个角度α_i^k与α_i^c之间的差值以及β_i^k与β_i^c之间的角度差值。

$$\begin{cases} \alpha = \alpha_i^k - \alpha_i^c \\ \beta = \beta_i^k - \beta_i^c \end{cases} \tag{9}$$

最后，分别将计算得到的距离差、角度以及角度差进行拼接，得到具有旋转不变特性的低维几何特征$\boldsymbol{RI}(p_i) = \{dis_1, dis_2, \theta_1, \theta_2, \alpha, \beta\}$。同时，为了保留原始点的位置信息，将检索点坐标$p_i$与近邻点坐标$p_i^k$也纳入特征组合中，得到最终的几何表示$\boldsymbol{GR}$，即

$$\boldsymbol{GR}(p_i) = \{p_i, p_i^k, dis_1, dis_2, \theta_1, \theta_2, \alpha, \beta\} \tag{10}$$

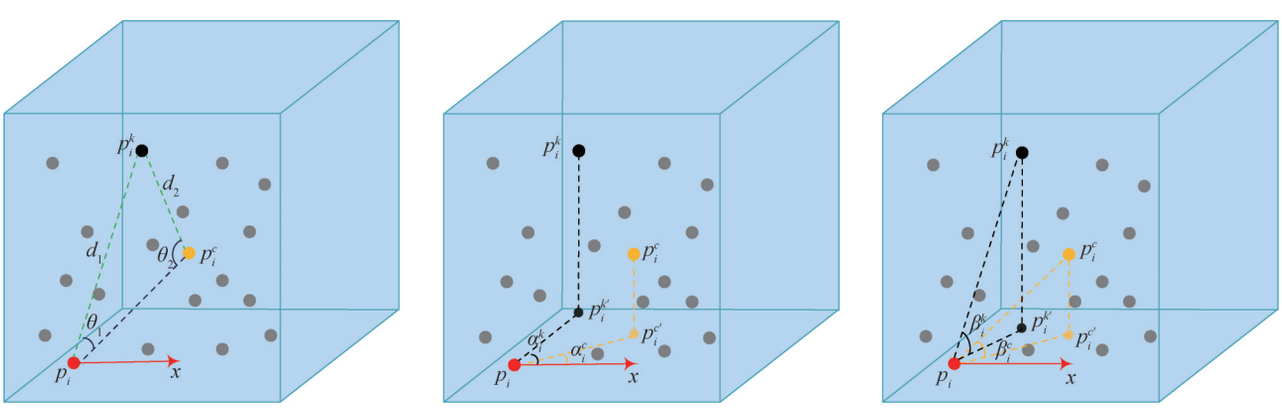

图4 低维旋转不变几何特征示意图

Fig. 4 The diagram of low-dimensional rotation-invariant geometric features

2.4 几何特征提取模块

特征距离表示点云中不同点在高维特征空间中的相似性。点云场景中的点可能在几何位置上相距较远，但在语义上却具有相似的特征。为了能够捕捉到这些远距离但语义相近的点的相互关系，同时有效区分坐标相近但是语义差异较大的点，本文设计了一个基于特征距离注意力池化模块(feature distance based attention pooling module，FDAP)，将低维特征作为输入，根据特征距离捕获较为重要的邻域特征，提取具有独特性且具有旋转不变性的特征。

1)特征距离计算。特征距离计算模块如图5所示，将点云三维坐标作为输入，经过两个MLP，得到特征$\boldsymbol{F}_1$，同时，利用KNN查找点云的k近邻，将边信息送入两个MLP，得到边特征$\boldsymbol{F}_2$，将特征$\boldsymbol{F}_1$与$\boldsymbol{F}_2$拼接得到特征$\boldsymbol{F}_m$。然后对特征$\boldsymbol{F}_m$进行扩展操作($tile(\cdot)$)，同时，利用KNN查找特征$\boldsymbol{F}_m$的k近邻($K(\cdot)$)。将两者作差并取绝对值。最后求均值得到特征距离Dis_{F_m}，具体为

$$Dis_{F_m} = \mathrm{mean}\left(\left|tile(\boldsymbol{F}_m) - K(\boldsymbol{F}_m)\right|\right) \tag{11}$$

式中，$|\cdot|$表示取绝对值，最终的特征距离FD要对Dis_{F_m}取以e为底的负指数，具体为

$$FD = \exp(-Dis_{F_m}) \tag{12}$$

2)基于特征距离的注意力池化模块。图6为基于特征距离的注意力池化模块网络图。首先，将几何表示$\boldsymbol{GR}$送入多层感知机，得到新特征$\boldsymbol{f}_i$，将$\boldsymbol{f}_i$与特征距离FD进行拼接，得到特征$\boldsymbol{F}_i$，具体为

$$\boldsymbol{f}_i = \mathrm{MLP}(\boldsymbol{GR}(p_i)) \tag{13}$$

$$\boldsymbol{F}_i = concat(\boldsymbol{f}_i, FD) \tag{14}$$

然后，对$\boldsymbol{F}_i$进行softmax操作，得到权重矩阵$\boldsymbol{W}$，具体为

$$\boldsymbol{W} = \mathrm{softmax}(FC(\boldsymbol{F}_i)) \tag{15}$$

最后，将特征$\boldsymbol{f}_i$与权重矩阵W相乘，得到最终的几何特征$\boldsymbol{F}_{\mathrm{geo}}$，具体为

$$\boldsymbol{F}_{\mathrm{geo}} = \mathrm{MLP}\left(\sum_k (\boldsymbol{W} \times \boldsymbol{f}_i)\right) \tag{16}$$

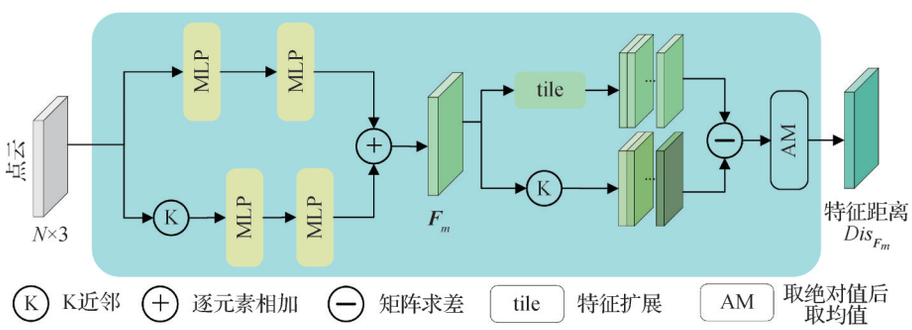

图5　特征距离计算模块结构

Fig. 5　The structure of the feature distance calculation module

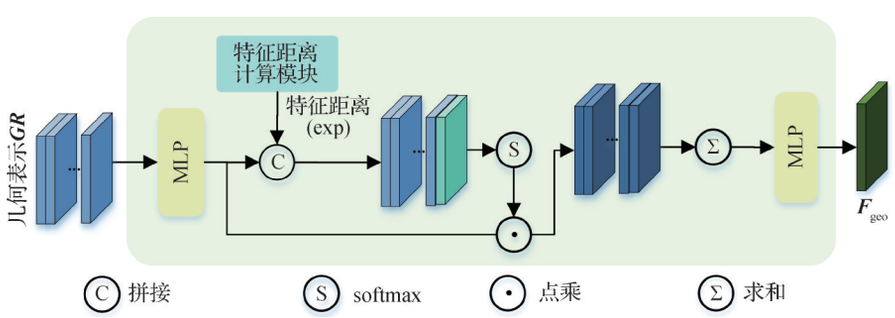

图6　基于特征距离的注意力池化模块结构

Fig. 6　The structure of the feature distance based attention pooling module

3　实验与分析

3.1　数据集与评估指标

本文使用Oxford RobotCar数据集验证ERA-Net。该数据集包括4个不同的场景：Oxford数据集(Maddern等，2017)和3个内部数据集(Uy和Lee，2018)：大学区(U. S.)、居民区(R. A.)和商业区(B. D.)。该数据集中所有的点云场景均去除地面点，下采样到4 096个点，并归一化到[-1，1]范围内。对于Oxford数据集中所有子地图，其中21 711个子地图用于训练，另外3 030个子地图用于测试。在训练过程中，如果点云场景之间的距离不超过10 m，则视为正确匹配；如果点云场景之间的距离大于50 m，则视为错误匹配。在测试过程中，如果检索到的点云子图与查询点云之间的距离在25 m以内，则认为检索得到的点云与查询点云匹配正确，即是同一个地点。本文只使用Oxford数据集中的训练数据集进

行训练，在Oxford测试数据集，U. S. ，R. A. 和B. D. 数据集上进行评估。所有的实验均在NVIDIA RTX 2080Ti GPU上进行训练和测试。本文将ERA-Net训练20个epoch，batchsize设置为1，初始学习率为0.000 5，在特征提取模块中输入点数为4 096，*KNN*算法中k设置为20，NetVLAD中聚类中心个数设置为K = 64，最终输出全局描述符向量的维数设置为256。

与Uy和Lee（2018）、Hui等人（2022）、Xia等人（2021）的评价指标相同，本文也采用前1平均召回率（average recall @1，AR@1）和前1%平均召回率（average recall @1%，AR@1%），用以评估地点识别网络的准确性。召回率是信息检索和机器学习领域中的一个关键性能指标，衡量模型正确识别正类的能力，即被正确识别为正类样本占所有正类样本的比例。

3.2 实验结果对比

将ERA-Net与当前先进的方法（SOTA）进行比较，包括PointNetVLAD（Uy和Lee，2018）、PCAN（Zhang和Xiao，2019）、LPD-Net（Liu等，2019）、MinkLoc3D（Komorowski，2021）、SOE-Net（Xia等，2021）、EPC-Net（Hui等，2022）、RPR-Net（Fan等，2022）、HiTPR-F8（Hou等，2022）、Retriever（Wiesmann等，2022）和HiBi-Net（Shu和Kwon，2024）。

表1展示了ERA-Net与其他方法在4个基准数据集上的AR@1%和AR@1。其中，MinkLoc3D在训练过程中，对输入数据进行随机平移、抖动等增强操作，旨在提高模型的检索性能。MinkLoc3D*则表示在训练过程中不使用特征增强的召回率结果。ERA-Net在Oxford数据集中的实验结果低于进行数据增强的MinkLoc3D。但是如果在训练过程中，去除数据增强操作，ERA-Net在AR@1%与AR@1两个指标中均高于MinkLoc3D*。实验结果表明，ERA-Net在3个内部数据集上性能明显优于其他方法。对于指标AR@1%，ERA-Net在U. S. 、R. A. 和B. D. 数据集上分别达到96.97%、94.69%和91.28%。

对于指标AR@1，在U. S. 数据集中，ERA-Net比HiBi-Net高3.23%，在R. A. 数据中，ERA-Net比HiBi-Net高1.61%，在B. D. 数据集中，比SOE-Net高出3.79%。这表明ERA-Net能够有效地生成具有判别性的全局描述符，对于未知的环境有较好的泛化能力。此外，图7展示了ERA-Net、PN_VLAD、PCAN、SOE-Net、EPC-Net在4个数据集上前25检索结果的召回率曲线。可以看出，在Oxford、U. S、

表1 Oxford数据集上不同网络的AR@1%和AR@1比较

Table 1 Comparison of AR@1% and AR@1 of different networks on the Oxford dataset

/%

方法	Oxford		U.S.		R.A.		B.D.	
	AR@1%	AR@1	AR@1%	AR@1	AR@1%	AR@1	AR@1%	AR@1
PN_VLAD	81.01	62.76	77.82	63.01	69.75	56.18	65.30	57.21
PCAN	83.90	69.05	79.13	62.49	71.15	56.99	66.90	58.14
LPD-Net	94.53	85.62	<u>95.93</u>	85.68	90.72	83.39	87.50	80.89
MinkLoc3D	**97.90**	**93.70**	95.00	86.00	91.20	81.10	<u>88.50</u>	82.60
MinkLoc3D*	95.90	88.20	93.60	83.20	86.00	74.70	82.20	74.00
SOE-Net	96.40	89.37	93.17	82.46	<u>91.47</u>	82.92	88.45	<u>83.33</u>
EPC-Net	95.20	86.86	95.42	86.33	88.10	79.18	84.24	77.79
RPR-Net	92.20	81.00	94.50	83.20	91.30	83.30	86.40	80.40
HiTPR-F8	94.64	87.77	94.01	86.00	89.11	81.32	88.31	81.80
HiBi-Net	–	87.46	–	<u>87.81</u>	–	<u>85.76</u>	–	83.03
Retriever	91.93	–	91.88	–	87.44	–	85.53	–
ERA-Net（本文）	<u>96.48</u>	<u>90.47</u>	**96.97**	**91.04**	**94.69**	**87.37**	**91.28**	**87.12**

注：加粗、下划线字体分别表示各列最优、次优结果。“–”表示无数据 。

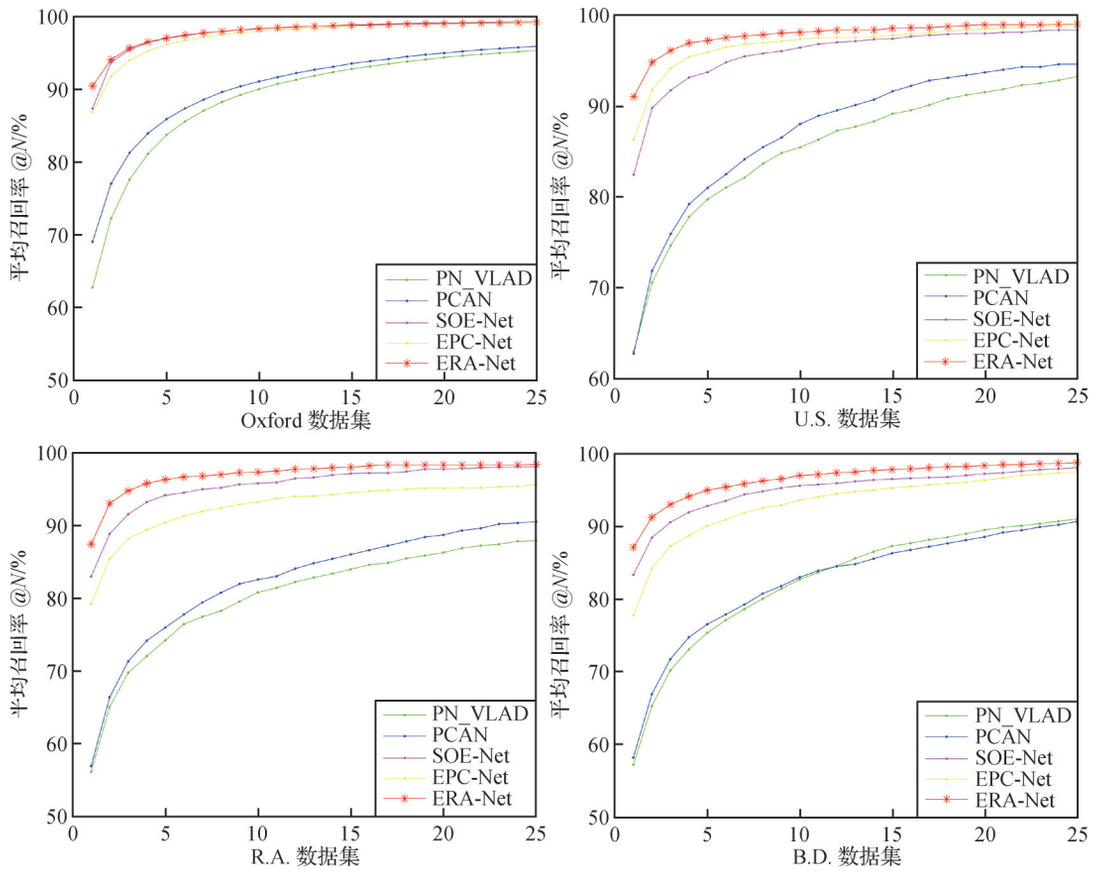

图 7　不同方法在数据集 Oxford、U. S. 、R. A. 和 B. D. 上排名前 25 检索结果的召回率曲线

Fig. 7　Average recall @25 of different methods on the Oxford RobotCar, U. S. , R. A. , and B. D. datasets

R. A. 、B. D. 4 个数据集上，ERA-Net 的结果明显优于其他几种方法。

当查询场景的数据发生旋转时，大多数现有方法的识别性能会出现显著下降。本文分别测试 ERA-Net、PN_VLAD（Uy 和 Lee，2018）、PCAN（Zhang 和 Xiao，2019）、MinkLoc3D（Komorowski，2021）、EPC-Net（Hui 等，2022）、ERINet（Weng 等，2022）与 LPS-Net（Liu 等，2024）在不同旋转角度下的 AR@1% 以及 AR@1，识别结果如表 2 所示。其中，测试的旋转角度包括：30°、60°、90°、120°、150°以及 180°，将这些角度作为最大旋转角度。将查询测试点云随机旋转，旋转范围从 0°到最大旋转角度。结果表明，当查询测试点云进行轻微的旋转时，如旋转 30°、60°，PN_VLAD、PCAN 和 MinkLoc3D 的识别效果下降明显。随着旋转角度的增大，ERA-Net 的识别性能优于其他方法。在指标 AR@1 中，虽然在旋转 30°时 ERA-Net 会在指标 AR@1% 上低于 LPS-Net 0. 47%，在指标 AR@1 上低于 LPS-Net 1. 43%，但是在其他旋转角度下，ERA-Net 都优于其他方法。当查询场景旋转 90°时，ERA-Net 在 AR@1% 与 AR@1 两个指标中比 LPS-Net 的 75. 32%、60. 14% 高出 7. 57% 和 7. 91%。当查询场景旋转 120°时，ERA-Net 在 AR@1% 与 AR@1 两个指标中比 ERINet 高出 2. 98% 和 4. 44%。当查询场景旋转角度不断增大时，本文提出的 ERA-Net 识别性能优于其他方法。实验结果表明，ERA-Net 即使在面对点云数据旋转的情况下，仍然能从散乱的点中提取出具有较强独特性以及旋转不变性的特征，正确检索出对应的场景。表 2 中倒数第 2 行罗列了 ERA-Net 去除 6 组低维几何特征 RI 后的旋转测试结果，可以看出，当查询点云旋转角度不断增大时，ERA-Net[-RI]在 AR@1% 与 AR@1 两个指标中下降明显。实验结果表明，本文提出的距离、角度与角度差 6 组低维几何特征在提高网络的旋转不变性上发挥了重要的作用。

3. 3　结果可视化

图 8 展示了 ERA-Net 在 Oxford 数据集中不同场景的检索结果。第 1 列为查询点云，第 2、3、4 列分别是检索得到的召回率前 3 名的场景。绿色框为正确

表2　Oxford数据集不同方法旋转测试结果

Table 2　Rotation performance on the Oxford dataset for different methods

/%

方法	旋转30°		旋转60°		旋转90°		旋转120°		旋转150°		旋转180°	
	AR@1%	AR@1	AR@1%	AR@1	AR@1%	AR@1	AR@1%	AR@1	AR@1%	AR@1	AR@1%	AR@1
PN_VLAD	58.93	37.99	40.04	24.43	31.75	19.48	33.01	20.02	29.80	18.00	28.34	17.03
PCAN	58.80	39.28	38.02	23.06	27.95	16.08	32.83	19.55	28.68	16.77	27.10	15.57
MinkLoc3D	72.37	54.54	60.62	42.31	53.89	36.03	50.36	32.69	48.14	30.79	46.32	28.81
EPC-Net	92.16	80.79	79.39	62.75	68.86	51.29	60.48	43.11	55.80	38.66	52.39	35.18
ERINet	95.31	85.30	82.09	64.91	70.96	52.31	71.66	54.30	66.55	48.73	64.34	46.29
LPS-Net	**95.70**	**88.56**	87.23	74.81	75.32	60.14	65.96	50.34	60.46	44.31	56.69	40.68
ERA-Net [-RI]	76.63	60.58	59.24	43.11	47.26	32.56	40.78	27.28	37.04	24.89	34.15	22.33
ERA-Net(本文)	95.23	87.13	**90.66**	**78.46**	**82.89**	**68.05**	**74.64**	**58.74**	**69.63**	**53.31**	**65.96**	**49.51**

注：加粗字体表示各列最优结果。

检索的场景，红色框为未能正确检索的场景。可以看出，对于不同的检索场景，top1的检索结果与查询点云成功匹配。场景1主要包括建筑物与树木，树木扫描得到的点云数据会随着季节的变化而变化，例如在夏季，扫描得到的树木点云包括较多的叶点，而在冬季，扫描得到的树木点云主要包括枝杆。尽管树木点云数据变化很大，ERA-Net仍能检索得到正确的场景。场景2中包括杆状物，杆状物点云比

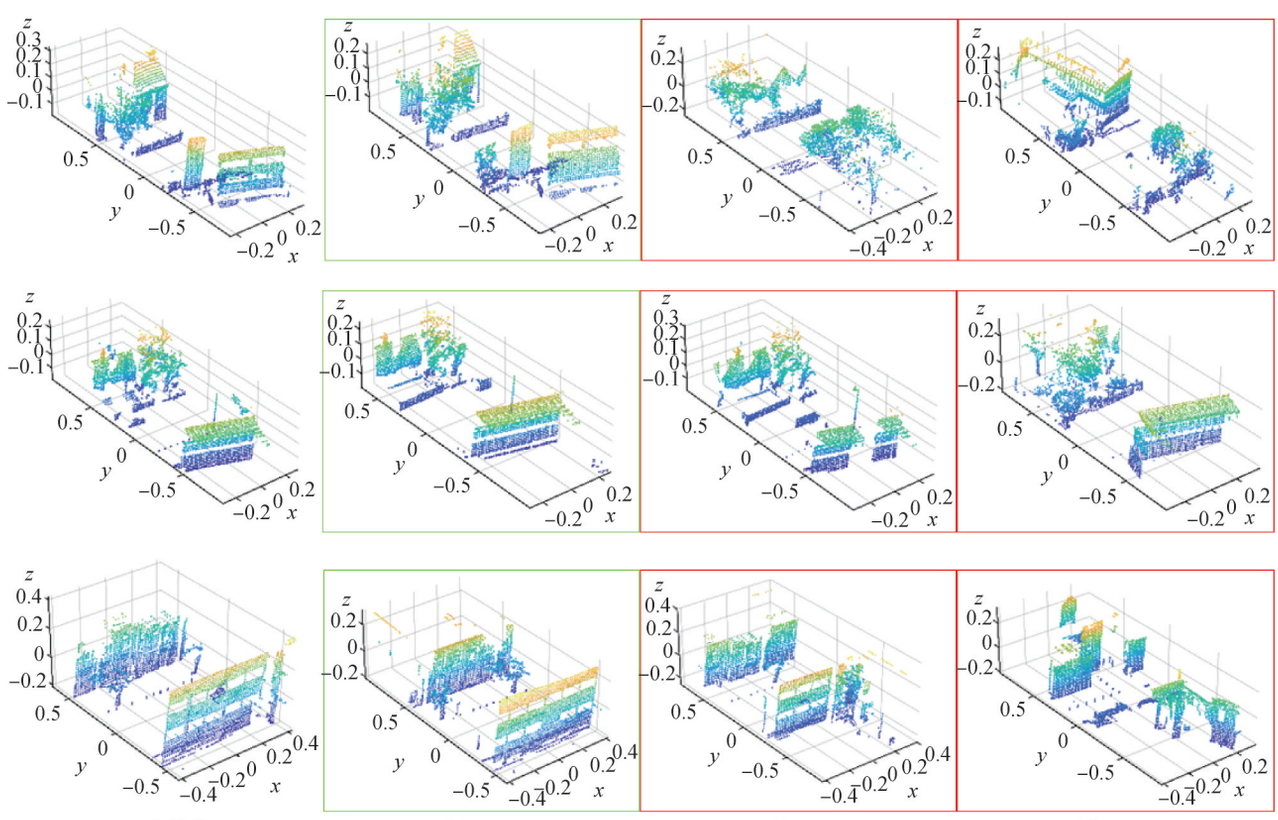

图8　ERA-Net在Oxford数据集的检索结果

Fig. 8　Retrival results of ERA-Net on the Oxford dataset((a)query point cloud;(b)top1;(c)top2;(d)top3)

较细，ERA-Net能较好地提取其几何特征，完成场景的正确匹配。场景3中建筑物占主要部分，尽管场景中包括较多的平面点，ERA-Net仍能提取具有较强独特性与不变性的特征。

为验证ERA-Net对旋转问题的有效性，本文在对B. D. 数据集进行测试时，将查询点云子图进行旋转，模拟现实车辆驾驶过程中不按规定路径行驶的情况。图9展示了利用EPC-Net和ERA-Net两个模型对旋转90°的点云子图进行检索的结果。其中，第1列为检索场景，第2列为将检索子图旋转90°得到的点云数据，第3列为当查询子图旋转90°时，EPC-Net检索得到的top1子图，第4列为当查询子图旋转90°时，ERA-Net检索得到的top1子图。可以看出，EPC-Net并未检索出正确的场景，本文提出的ERA-Net在查询子图旋转90°的情况下，仍然能检索出正确的场景。

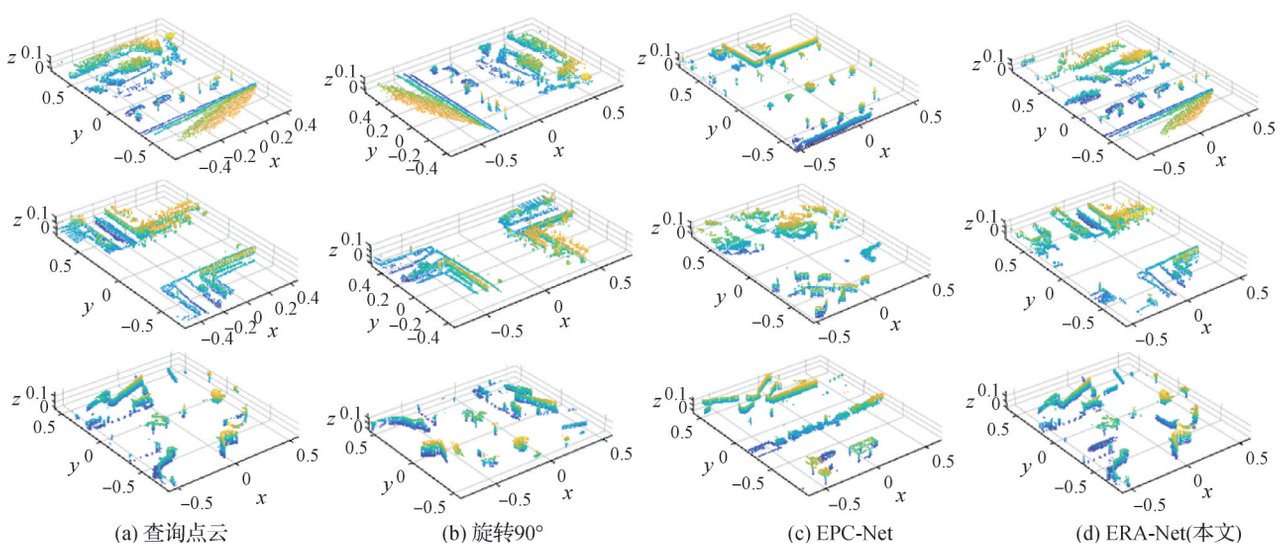

图9　B. D. 数据集场景旋转测试

Fig. 9　Rotation test on B. D. dataset((a)query point cloud;(b)rotating 90°;(c)EPC-Net;(d)ERA-Net(ours))

3.4　消融实验

为进一步验证本文提出的具有旋转不变性的低维几何特征模块以及基于特征距离的注意力池化模块的有效性，本文在Oxford数据集上进行消融实验，表3展示了ERA-Net在Oxford数据集上的AR@1%和AR@1的测试结果。

1）去除低维几何特征模块ERA-Net[-RI]。根据点与k近邻点之间的坐标信息，计算6组具有旋转不变性的低维几何特征（距离、角度与角度差），本文通过去除6组低维几何特征，将三维点云坐标与每点的k近邻直接输入到基于特征距离的注意力池化模块中，以此验证6组具有旋转不变性的低维几何特征对网络结构的重要性。ERA-Net[-RI]在Oxford数据集中AR@1%比ERA-Net低了11. 13%，在AR@1指标中，ERA-Net[-RI]比ERA-Net低17. 43%。实验结果表明，ERA-Net中具有旋转不变性的低维特征能够有效地捕获点间的几何信息，提取具有较强不变性和鲁棒性的特征，极大提升了ERA-Net的场景识别性能。

2）去除基于特征距离的注意力池化模块ERA-Net[-FDAP]。将ERA-Net中基于特征距离的注意力池化模块去除，在Oxford数据集中AR@1%比ERA-Net降低0. 36%，在AR@1指标中，ERA-Net[-FDAP]比ERA-Net降低1. 03%。可以看出，特征间欧氏距离可以有效地衡量特征之间的相关性和相互影响，基于特征距离的注意力池化模块能够有效

表3　ERA-Net在Oxford数据集上不同模块消融实验

Table 3　ERA-Net ablation experiments of different modules on the Oxford dataset

/%

不同模块	Oxford	
	AR@1%	AR@1
ERA-Net [-RI]	85.35	73.04
ERA-Net [-FDAP]	96.12	89.44
ERA-Net	96.48	90.47

地提取关键局部特征。

3.5 计算资源耗费分析

表4列出了PN_VLAD、LPD-Net、ERA-Net的参数量、所需的GPU显存以及每帧运行时间。可以看出，ERA-Net每帧运行时间高于PN_VLAD和LPD-Net，这主要是由于ERA-Net需要计算场景中每点具有旋转不变性的低维几何特征，同时引入4层GAPLayer模块，需要进行大量的矩阵运算。ERA-Net的参数量稍低于PN_VLAD和LPD-Net。ERA-Net所需的显存与LPD-Net相当，但是ERA-Net能获得更好的识别效果。

表4 不同网络的参数量、所需显存与每帧运行时间比较

Table 4 Comparison results of the parameters, GPU memory required and runtime per frame for different networks

方法	参数/M	GPU 显存/G	每帧运行时间/ms
PN_VLAD	19.78	5.28	40.3
LPD-Net	19.81	10.66	102.6
ERA-Net	18.55	10.83	440.0

4 结 论

本文提出一种具有旋转感知的点云场景地点识别方法，该方法利用注意力机制，充分考虑点间的全局依赖关系以及点与邻域点间的局部相关性，提取点的语义特征。同时，利用点与k邻近点的坐标信息，提取距离、角度和角度差等具有旋转不变性的低维几何特征。并通过计算特征在高维空间的关联性，提取具有较强独特性以及旋转不变性的几何特征。最后，通过NetVLAD模块将提取的语义特征与几何特征聚合，完成地点的准确识别。实验结果表明，ERA-Net的识别性能优于目前SOTA方法，低维几何特征的提取能够较好地提升网络的识别效果，在旋转测试中，也能取得比较好的效果。基于特征距离的注意力池化模块能够捕捉远距离但语义相近的点的相互关系，同时能够有效区分坐标相近但是语义差异较大的点。

参考文献(References)

Ali-Bey A, Chaib-Draa B and Giguere P. 2023. MixVPR: feature mixing for visual place recognition//Proceedings of 2023 IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa, USA: IEEE: 2997-3006 [DOI: 10.1109/wacv56688.2023.00301]

Arandjelovic R, Gronat P, Torii A, Pajdla T and Sivic J. 2016. NetVLAD: CNN architecture for weakly supervised place recognition//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 5297-5307 [DOI: 10.1109/cvpr.2016.572]

Chen C, Fragonara L Z and Tsourdos A. 2021. GAPointNet: graph attention based point neural network for exploiting local feature of point cloud. Neurocomputing, 438: 122-132 [DOI: 10.1016/j.neucom.2021.01.095]

Chen Y W, Zhuang Y, Huai J Z, Li Q P, Wang B L, El-Bendary N and Yilmaz A. 2024. Semantics-enhanced discriminative descriptor learning for LiDAR-based place recognition. ISPRS Journal of Photogrammetry and Remote Sensing, 210: 97-109 [DOI: 10.1016/j.isprsjprs.2024.03.002]

Dubé R, Dugas D, Stumm E, Nieto J, Siegwart R and Cadena C. 2017. SegMatch: segment based place recognition in 3D point clouds//Proceedings of 2017 IEEE International Conference on Robotics and Automation (ICRA). Singapore, Singapore: IEEE: 5266-5272 [DOI: 10.1109/icra.2017.7989618]

Fan S Q, Dong Q L, Zhu F H, Lyu Y S, Ye P J and Wang F Y. 2021. SCF-Net: learning spatial contextual features for large-scale point cloud segmentation//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 14499-14508 [DOI: 10.1109/cvpr46437.2021.01427]

Fan Z X, Liu H Y, He J, Sun Q and Du X Y. 2020. SRNet: a 3D scene recognition network using static graph and dense semantic fusion. Computer Graphics Forum, 39(7): 301-311 [DOI: 10.1111/cgf.14146]

Fan Z X, Song Z B, Zhang W P, Liu H Y, He J and Du X Y. 2022. RPR-Net: a point cloud-based rotation-aware large scale place recognition network//Proceedings of 2022 European Conference on Computer Vision. Tel Aviv, Israel: IEEE: 709-725 [DOI: 10.1007/978-3-031-25056-9_45]

Hao W, Zhang W J and Jin H Y. 2024. SAGE-Net: employing spatial attention and geometric encoding for point cloud based place recognition. IEEE Robotics and Automation Letters, 9(6): 4958-4965 [DOI: 10.1109/LRA.2024.3387112]

Hao W, Zhang W J, Liang W, Xiao Z L and Jin H Y. 2022. Scene recognition for 3D point clouds: a review. Optics and Precision Engineering, 30(16): 1988-2005 (郝雯, 张雯静, 梁玮, 肖照林, 金海燕. 2022. 面向三维点云的场景识别方法综述, 光学 精密工程, 30(16): 1988-2005) [DOI: 10.37188/OPE.20223016.1988]

He L, Wang X L and Zhang H. 2016. M2DP: a novel 3D point cloud descriptor and its application in loop closure detection//Proceedings of 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Daejeon, Korea (South): IEEE: 231-237 [DOI: 10.1109/iros.2016.7759060]

Hou Z X, Yan Y, Xu C Z and Kong H. 2022. HiTPR: hierarchical transformer for place recognition in point cloud//Proceedings of 2022 International Conference on Robotics and Automation (ICRA). Philadelphia, USA: IEEE: 2612-2618 [DOI: 10.1109/icra46639.2022.9811737]

Hui L, Cheng M M, Xie J, Yang J and Cheng M M. 2022. Efficient 3D point cloud feature learning for large-scale place recognition. IEEE Transactions on Image Processing, 31: 1258-1270 [DOI: 10.1109/tip.2021.3136714]

Hui L, Yang H, Cheng M M, Xie J and Yang J. 2021. Pyramid point cloud transformer for large-scale place recognition//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 6078-6087 [DOI: 10.1109/iccv48922.2021.00604]

Izquierdo S and Civera J. 2024. Optimal transport aggregation for visual place recognition//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 17658-17668 [DOI: 10.1109/cvpr52733.2024.01672]

Jaegle A, Gimeno F, Brock A, Vinyals O, Zisserman A and Carreira J. 2021. Perceiver: general perception with iterative attention//Proceedings of the 38th International Conference on Machine Learning. [s.l.]: ICML: 4651-4664

Komorowski J. 2021. MinkLoc3D: point cloud based large-scale place recognition//Proceedings of 2021 IEEE Winter Conference on Applications of Computer Vision. Waikoloa, USA: IEEE: 1789-1798 [DOI: 10.1109/wacv48630.2021.00183]

Kong X, Yang X M, Zhai G Y, Zhao X R, Zeng X F, Wang M M, Liu Y, Li W L and Wen F. 2020. Semantic graph based place recognition for 3D point clouds//Proceedings of 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Las Vegas, USA: IEEE: 8216-8223 [DOI: 10.1109/iros45743.2020.9341060]

Li Q P, Zhuang Y, Huai J Z, Chen Y W and Yilmaz A. 2024. An efficient point cloud place recognition approach based on transformer in dynamic environment. ISPRS Journal of Photogrammetry and Remote Sensing, 207: 14-26 [DOI: 10.1016/j. isprsjprs. 2023.11.013]

Liu C X, Chen G Y and Song R. 2024. LPS-Net: lightweight parameter-shared network for point cloud-based place recognition//Proceedings of 2024 IEEE International Conference on Robotics and Automation (ICRA). Yokohama, Japan: IEEE: 448-454 [DOI: 10.1109/ICRA57147.2024.10610758]

Liu Y C, Yu L, Yu H and Yang W. 2024. Visual place recognition with fusion event cameras. Journal of Image and Graphics, 29(4): 1018-1029 (刘熠晨, 余磊, 余淮, 杨文. 融合事件相机的视觉场景识别. 中国图象图形学报, 2024, 29(4): 1018-1029) [DOI: 10.11834/jig.230003]

Liu Z, Zhou S B, Suo C Z, Yin P, Chen W, Wang H S, Li H A and Liu Y H. 2019. LPD-Net: 3D point cloud learning for large-scale place recognition and environment analysis//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 2831-2840 [DOI: 10.1109/iccv.2019.00292]

Maddern W, Pascoe G, Linegar C and Newman P. 2017. 1 year, 1000 km: the Oxford RobotCar dataset. The International Journal of Robotics Research, 36(1): 3-15 [DOI: 10.1177/0278364916679498]

Qi C R, Su H, Kaichun M and Guibas L J. 2017. PointNet: deep learning on point sets for 3D classification and segmentation//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 77-85 [DOI: 10.1109/cvpr.2017.16]

Ramezani M, Wang L, Knights J, Li Z B, Pounds P and Moghadam P. 2024. Pose-Graph attentional graph neural network for Lidar place recognition. IEEE Robotics and Automation Letters, 9(2): 1182-1189 [DOI: 10.1109/lra.2023.3341766]

Rizzini D L. 2017. Place recognition of 3D landmarks based on geometric relations//Proceedings of 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Vancouver, Canada: IEEE: 648-654 [DOI: 10.1109/iros.2017.8202220]

Röhling T, Mack J and Schulz D. 2015. A fast histogram-based similarity measure for detecting loop closures in 3-D LIDAR data//Proceedings of 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Hamburg, Germany: IEEE: 736-741 [DOI: 10.1109/iros.2015.7353454]

Shu D W and Kwon J. 2024. Hierarchical bidirected graph convolutions for large-scale 3-D point cloud place recognition. IEEE Transactions on Neural Networks and Learning Systems, 35(7): 9651-9662 [DOI: 10.1109/tnnls.2023.3236313]

Sun Q, Liu H Y, He J, Fan Z X and Du X Y. 2020. DAGC: employing dual attention and graph convolution for point cloud based place recognition//Proceedings of 2020 International Conference on Multimedia Retrieval. Dublin, Ireland: ACM: 224-232 [DOI: 10.1145/3372278.3390693]

Uy M A and Lee G H. 2018. PointNetVLAD: deep point cloud based retrieval for large-scale place recognition//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 4470-4479 [DOI: 10.1109/cvpr.2018.00470]

Wen C C, Huang H, Liu Y S and Fang Y. 2023. Pyramid learnable tokens for 3D LiDAR place recognition//Proceedings of 2023 IEEE International Conference on Robotics and Automation (ICRA). London, United Kingdom: IEEE: 4143-4149 [DOI: 10.1109/icra48891.2023.10161523]

Weng S C, Zhang R N and Li G. 2022. ERINet: effective rotation invariant network for point cloud based place recognition//Proceedings of 2022 IEEE International Conference on Visual Communications and Image Processing. Suzhou, China: IEEE: 1-5 [DOI: 10.1109/vcip56404.2022.10008830]

Wiesmann L, Marcuzzi R, Stachniss C and Behley J. 2022. Retriever: point cloud retrieval in compressed 3D maps//Proceedings of 2022 International Conference on Robotics and Automation (ICRA). Philadelphia, USA: IEEE: 10925-10932 [DOI: 10.1109/icra46639.2022.9811785]

Xia Y, Xu Y S, Li S, Wang R, Du J, Cremers D and Stilla U. 2021. SOE-Net: a self-attention and orientation encoding network for point cloud based place recognition//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 11343-11352 [DOI: 10.1109/cvpr46437.2021.01119]

Zhang W X and Xiao C X. 2019. PCAN: 3D attention map learning using contextual information for point cloud based retrieval//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 12428-12437 [DOI: 10.1109/cvpr.2019.01272]

Zhang Y J, Shi P C and Li J Y. 2024. LiDAR-based place recognition for autonomous driving: a survey. ACM Computing Surveys, 57(4): #106 [DOI: 10.1145/3707446]

Zhang Z Y, Hua B S and Yeung S K. 2022. RIConv++: effective rotation invariant convolutions for 3D point clouds deep learning. International Journal of Computer Vision, 130(5): 1228-1243 [DOI: 10.1007/s11263-022-01601-z]

Zhao S Q, Yin P, Yi G and Scherer S. 2023. SphereVLAD++: attention-based and signal-enhanced viewpoint invariant descriptor. IEEE Robotics and Automation Letters, 8(1): 256-263 [DOI: 10.1109/lra.2022.3223555]

作者简介

郝雯,女,副教授,主要研究方向为点云地点识别、点云场景分割。E-mail: haowensxsf@163.com

吕炎,男,硕士研究生,主要研究方向为点云地点识别。E-mail:2231221145@stu.xaut.edu.cn

金海燕,女,教授,主要研究方向为计算机视觉、图像处理和智能信息处理。E-mail: jinhaiyan@xaut.edu.cn

石争浩,男,教授,主要研究方向为机器视觉、医学图像处理及机器学习。E-mail: ylshi@xaut.edu.cn