ORCHNet：一种用于果园中基于三维激光雷达的地点识别的鲁棒全局特征聚合方法

T. Barros、L. Garrote、P. Conde、M.J. Coombes、C. Liu、C. Premebida、U.J. Nunes

arXiv:2303.00477v2 [cs.CV] 6 Feb 2024

摘要—在农业环境中，实现鲁棒且可靠的地点识别和回环检测仍然是一个尚未解决的问题。特别是，由于整个田地具有结构相似性，果园是一个较为困难的案例研究对象。在本研究中，我们利用三维激光雷达数据解决果园中的地点识别问题，因为三维激光雷达被认为是提升鲁棒性的关键模态。因此，我们提出了 ORCHNet，这是一种基于深度学习的方法，可将三维激光雷达扫描映射为全局描述符。具体而言，本文提出了一种新的全局特征聚合方法，将多种聚合方法融合为一个鲁棒的全局描述符。ORCHNet 在果园中采集的真实世界数据上进行了评估，其中包括夏季和秋季的数据。为评估其鲁棒性，我们将 ORCHNet 与最先进的聚合方法在同季节数据和跨季节数据上的表现进行了比较。此外，我们还将所提出的方法作为定位框架的一部分进行评估，其中 ORCHNet 被用作回环检测器。实证结果表明，在地点识别任务中，ORCHNet 的性能优于其余方法，并且在跨季节场景下也更加鲁棒。至于定位任务，当将 ORCHNet 集成为回环检测器后，路径穿过树木的边缘情况得以解决，展示了所提方法在该任务中的潜在应用价值。代码将在以下网址公开提供：https://github.com/Cybonic/ORCHNet.git

索引词—**定位、地点识别、同步定位与建图、农业机器人。**

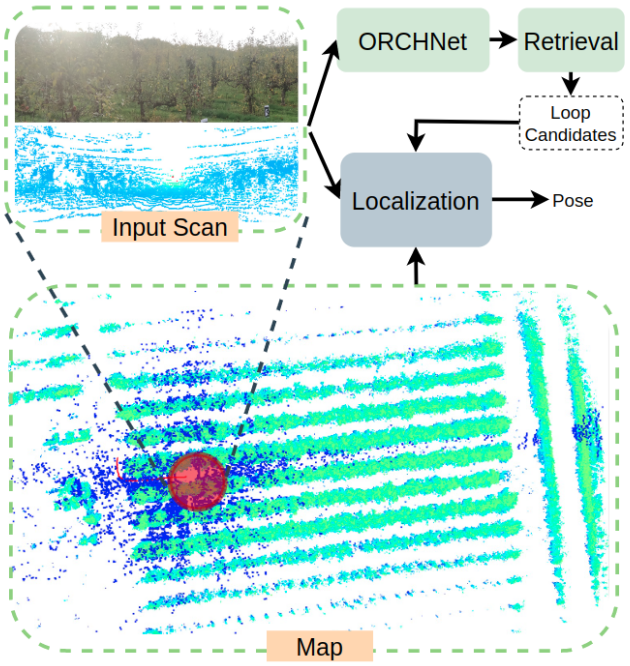


图 1：所提出的 ORCHNet 在定位框架中的集成示意图。

一、引言

地点识别可理解为一种基于感知的全局定位方法，它利用视觉、结构和／或语义线索识别先前到访过的地点，并据此生成描述符。在典型应用中，会将当前描述符与先前输入数据中的描述符进行比较，以识别重访地点。近年来，地点识别已被用作同步定位与建图或定位方法中的高效回环检测器[1], [2]。

在研究领域，自动驾驶（AD）社群最为活跃，利用地点识别实现长期定位，并将三维激光雷达数据视为提升鲁棒性的关键模态 [3]。目前，基于三维激光雷达的方法大多采用深度学习（DL）方法进行地点建模 [4]。随着深度学习近期取得进展，网络得以直接处理点云，这一技术已得到广泛应用；PoinNet[5]便是此类网络的一个例子。

在地点识别和回环检测方面，野外机器人学是一个研究较少的领域，尤其是农业机器人技术。在这一领域，由于环境条件恶劣且不断变化，基于感知的定位难以长期实现[6]。在农业应用领域中，果园是一个极具挑战性的案例，因为与类城市环境相比，果园环境缺乏“相关”的几何特征；类城市环境中的几何特征通常从车辆、建筑物或其他静态城市“设施”中提取。在果园中，树冠密度较低且可能随季节变化，因此会产生稀疏扫描，导致描述性特征较差。此外，树木以规则间距排列成平行行列，导致激光雷达扫描数据的几何信息相似，使得连续扫描以及来自相邻行列的扫描几乎无法区分。这对于地点识别等方法而言是一个不利特征，因为此类方法旨在学习独特且具有描述性的特征来识别地点。

本研究针对果园中的地点识别和闭环检测问题，提出了 ORCHNet——一种将三维激光雷达扫描映射到描述子空间的地点建模方法。ORCHNet 的主要贡献在于全局特征聚合方法，该方法将多种聚合方法融合为稳健的全局描述子。

T.Barros、L.Garrote、P.Conde、C.Premebida 和 U.J.Nunes 任职于葡萄牙科英布拉大学系统与机器人研究所电气与计算机工程系。电子邮箱：{tiagobarros, garrote, pedro.conde, cpremebida, urbano}@isr.uc.ptM.J.Coombes 和 C.Liu 任职于英国拉夫堡大学 LUCAS实验室航空与汽车工程系。电子邮箱：{M.J.Coombes, C.Liu5}@lboro.ac.uk

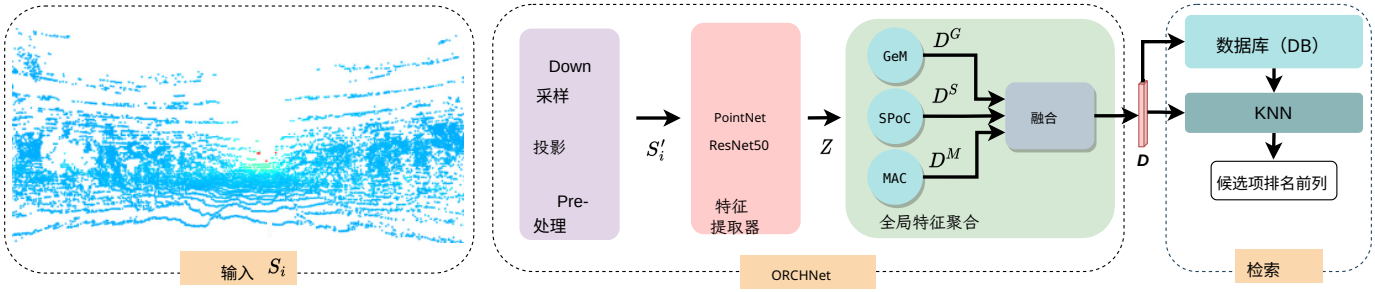


图 2：ORCHNet 作为基于检索的地点识别任务一部分的架构。ORCHNet 接收点云S_i作为输入，对其进行下采样和预处理，返回 S_i'。从 S_i'中使用特征提取器提取局部特征 Z。局部特征被输入全局特征聚合模块，该模块将 GeM、SPoC 和 MAC 的输出融合为全局描述子 D。利用全局描述子查询数据库，数据库根据相似性度量返回前 N 个闭环候选项。

针对这一应用，研究人员使用一台配备三维激光雷达和 RTK-GPS 等传感器的移动机器人，在英国的果园中采集了一个自建数据集，并在该数据集上对 ORCHNet 进行了评估。该数据集包含夏季和秋季的果园数据，可用于在果园中实际出现的变化条件下测试所提出的方法。

为评估 ORCHNet 的性能，尤其是其在地点识别任务中跨季节的鲁棒性，开展了两类实验：同季节实验，即训练数据和测试数据均来自秋季序列；以及跨季节实验，即训练数据来自秋季序列，测试数据来自夏季序列。此外，还将 ORCHNet 作为定位框架的一部分进行了评估（见图 1），在该框架中，ORCHNet 用于寻找闭环，并据此进行定性评估。

实证结果表明，与其他全局特征聚合方法相比，所提出的方法更适用于果园。此外，结果还表明，该方法在跨季节实验中的性能不受影响。对于定位任务，ORCHNet 能够解决一些边缘情况，例如路径穿过树木的情况。

简而言之，本研究的主要贡献如下：

- 一种适用于果园/树木密集环境的新型稳健基于三维激光雷达的地点识别方法。

II. 相关工作

过去十年来，地点识别一直是备受研究的课题，其中基于三维激光雷达的方法尤为活跃 [4]。三维激光雷达传感器数据被认为是实现稳健地点识别的一种关键模态，因为它不受复杂视觉变化条件的影响；这些条件可能在长期运行期间出现，尤其是在不同季节和光照条件下重访地点时。

三维激光雷达地点识别方法是应对基于视觉的方法所面临挑战的自然选择。在三维激光雷达方法中，深度学习方法是地点建模领域最常见的方法。基于点云的深度学习方法可分为两个主要分支：一类直接从点云中提取特征，例如 PointNetVLAD [7]或 Minkloc3d[8],；另一类则间接从点云中提取特征，即将点云投影为代理表示，例如体素 [9], 、极坐标 [10]或深度范围图像 [11]。在这两类方法中，这些输入都会送入特征提取模块，以提取局部特征，随后将其聚合为全局描述子。

尽管基于三维激光雷达的方法天然能够抵御外观变化条件，但仍存在一些局限性，例如难以生成具有旋转不变性的全局描述子，而在从相反方向重访某个地点时，这一性质至关重要。因此，模型必须从数据中学习如何在生成具有旋转不变性的全局描述子与确保其具备足够的描述能力以识别地点之间取得平衡。在这一过程中，特征聚合模块至关重要，因为局部特征正是在聚合步骤中转换为全局描述子的，而网络则从全局角度确定哪些特征最为重要。为此，已有研究提出了多种针对三维激光雷达数据的聚合方法，其中 NetVLAD [7], [12] 是最受欢迎的方法之一。NetVLAD 将特征空间划分为多个聚类，以生成全局描述子。其他方法包括注意力 [13], 广义均值池化（GeM）[8]，全局最大池化（MAC）[14], 平均池化（SPoC）[15]。

本研究采用 GeM、SpoC 和 MAC 聚合方法，这些方法效率高，且已被证明适用于点云，并将各自的输出融合为一个全局描述子。

III. 所提方法

本节详细介绍了所提出的 ORCHNet 在基于检索的地点识别框架中的应用。图 2 展示了 ORCHNet 及该框架的概览。该地点识别框架包含两个主要模块：ORCHNet，通过将点云映射到描述子空间来建模地点；以及检索模块。建模主要通过以下步骤实现：输入点云 S_i 经

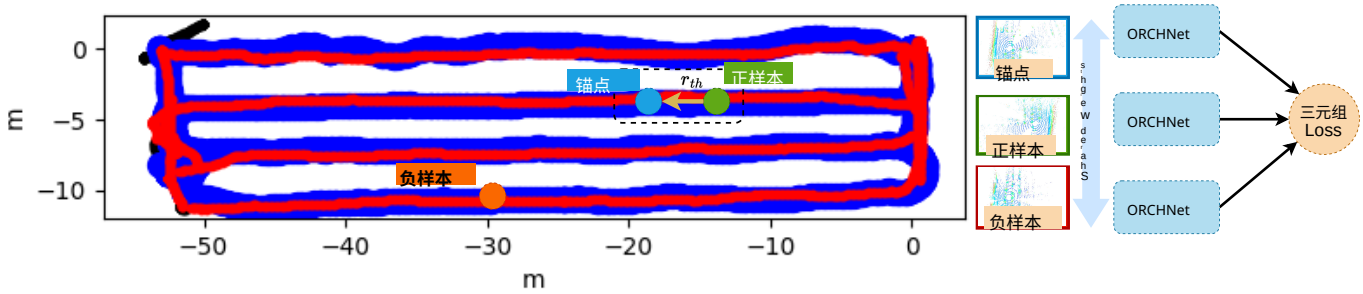


图3：使用LazyTriplet损失的训练方法。锚点-正样本对必须来自同一物理位置（即在同一行内彼此接近），而锚点-负样本对必须来自不同的位置。LazyTriplet损失通过测量锚点-正样本对与锚点-负样本对之间的相似度距离来计算。

预处理后输入特征提取器，得到局部特征 Z。利用所提出的全局特征聚合模块，将局部特征聚合为全局描述子 D，该模块是本文的主要贡献。

注意：以下各小节中定义的集合是有序集（即，集合元素之间存在序关系），且该顺序由元素的索引定义。

A. 预处理与特征提取器

预处理和特征提取模块旨在通过将扫描数据映射到中间特征空间，从中提取足够的特征。因此，给定一个扫描数据$S_i \in \mathbb{R}^{n_i \times 3}$，其中 $i \in [1, ..., N]$， N表示序列中的扫描数据数量， n_i 表示 i^{th} 点云中的点数，特征提取器会将 S_i 映射为固定的特征向量 $Z_i \in \mathbb{R}^{n \times F}$，其中 $F \gg 3$。为了比较不同的特征提取方法，本文采用了一种可直接处理点云的特征提取器（PoinNet），以及另一种从类图像代理表示中提取特征的特征提取器（ResNet50）。ResNet50最初用于图像，但通过适当投影到类图像表示，也已用于其他模态。我们对这两种网络都进行了少量修改，以适用于本任务。

B. 全局特征聚合

所提出的全局特征聚合方法利用三种不同的聚合方法：MAC[14]、GeM[16], [8]和SPoC[15], ，并通过可训练加权和将它们的输出融合为一个全局描述子。所提出的方法如图2所示。该方法接收一个特征向量$Z \in \mathbb{R}^{N \times F}$ 并输出一个描述符 $D = [d_k] \in \mathbb{R}^K$，其函数为 μ： $\mathbb{R}^{N \times F} \to \mathbb{R}^K$。

1) MAC： MAC模块是一种全局最大池化算子，用于从输入特征 $Z = [z_{i,j}] \in \mathbb{R}^{N \times F}$中计算 i^{th} 维度上的最大激活值：

$$z_i^M = \max_j z_{i,j}, \ \forall i \in [1, ..., N] \tag{1}$$

与 $Z^M = [z_i^M] \in \mathbb{R}^N$。然后，将 Z^M 输入全连接函数 h^M： $\mathbb{R}^N \to \mathbb{R}^K$，该函数输出 MAC 的输出描述子 $D^M = [d_k^M]$，其维度为 K。

2) SPoC：SPoC 模块是一个平均池化算子，它根据输入特征 $Z = [z_{i,j}] \in \mathbb{R}^{N \times F}$计算 i^{th} 维度的平均激活值：

$$z_i^S = \frac{1}{F} \sum_{j=1}^{F} z_{i,j}, \ \forall i \in [1, ..., N] \tag{2}$$

其中 $Z^S = [z_i^S] \in \mathbb{R}^N$。然后，将 Z^S 输入全连接函数 h^S： $\mathbb{R}^N \to \mathbb{R}^K$，输出SPoC的输出描述子 $D^S = [d_k^S]$，其维度为 K。

3) GeM： GeM模块是一种广义均值池化算子，作用于输入特征向量 $Z = [z_{i,j}] \in \mathbb{R}^{N \times F}$：

$$z_i^G = \left(\frac{1}{F} \sum_{j=1}^{F} z_{i,j}^p \right)^{1/p}, \ \forall i \in [1, ..., N] \tag{3}$$

其中 $p \in \mathbb{R}$ 是一个可训练参数，且 $Z^G = [z_i^G] \in \mathbb{R}^N$。然后，将 Z^G 输入全连接函数 $h^G: \mathbb{R}^N \to \mathbb{R}^K$，输出GeM的输出描述子 $D^G = [d_k^G]$，其维度为 K。

4) 融合： 融合模块使用$A = [a_1, a_2, a_3] \in \mathbb{R}^3$将三个描述符（$D^M$、$D^S$ 和 D^G）合并为一个全局描述子 D；这些可训练参数用于学习最佳激活组合。首先，将描述符拼接为 $D' = [D^M, D^S, D^G]^T = [d'_{i,k}] \in \mathbb{R}^{3 \times K}$，然后计算加权和：

$$d_k = \sum_{i=1}^{3} a_i d'_{i,k}, \ \forall k \in [1, ..., K] \tag{4}$$

其中， $D = [d_k] \in \mathbb{R}^K$是所提出的 ORCHNet 的 K维输出描述子。

C. 网络训练

在本研究中，使用文献 [7] 提出的 LazyTriplet损失对全局描述符进行训练。LazyTriplet损失迫使网络学习一个描述符空间，使来自同一物理位置的描述符（相对于锚点的正样本 ）彼此接近，而来自不同位置的描述符（相对于锚点的负样本 ）彼此远离。

因此，给定一组扫描数据 $\mathcal{S} = \{S_i | S_i \in \mathbb{R}^{n_i \times 3}\}$及其对应的位姿 $\mathcal{P} = \{p_i | p_i \in \mathbb{R}^3\}$ ，其中 $i =$

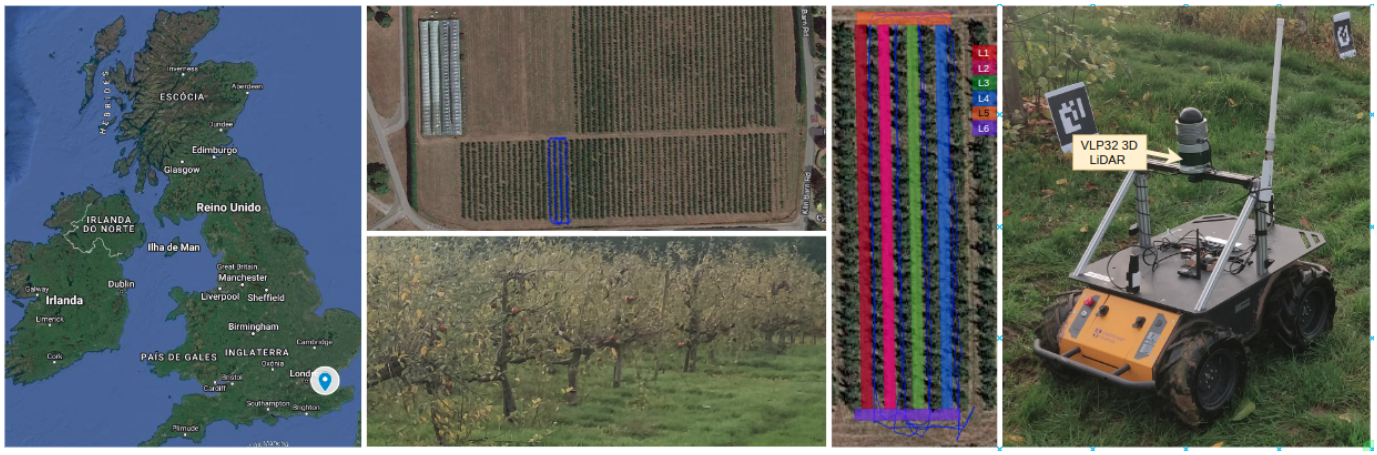


图4：用于采集数据的果园和机器人平台示意图。这些果园位于英国南部。果园按行划分，此处展示的是秋季序列，该序列被划分为6行。

$[1, ..., N]$ 其中 N 是序列中的样本数，n_i 是 i^{th} 扫描中的点数。考虑集合 $\Gamma = \{(S_i, p_i) | S_i \in \mathcal{S}, p_i \in \mathcal{P}\} \subseteq \mathcal{S} \times \mathcal{P}$，它定义了每个扫描 S_i 都存在一个对应的位姿 p_i。

定义一个集合 $\Gamma^A \subseteq \Gamma$，其中每个元素 $(S_j^A, p_j^A) \in \Gamma^A$ 都是一个锚点（即存在与之对应的正回环），其中 $j \in [1, ..., J]$，且 $J < N$ 是序列 $\mathcal{S}$ 中的锚点数量。

我们还定义（对于某个参数 γ）一个正样本集合 $\Gamma_\gamma^{P_j} \subseteq \Gamma$（$\forall j \in [1, ..., J]$），相对于锚点 j^{th}（S_j^A, p_j^A）。首先，定义函数 $\omega : \Gamma \to \mathbb{N}$，它将每一对 Γ 映射到其行号。现在，我们可以正式定义

$$\Gamma_\gamma^{P_j} = \Big\{ (S_l^{P_j}, p_l^{P_j}) \mid \|p_l^{P_j} - p_j^A\|_2 < r_{th} \wedge \omega(S_l^{P_j}, p_l^{P_j}) = \omega(S_l^A, p_l^A) \wedge l \notin [j-\gamma, j] \Big\} \quad (5)$$

其中每个元素 $(S_l^{P_j}, p_l^{P_j}) \in \Gamma_\gamma^{P_j}$，都是相对于锚点 j^{th} 的正样本。总之，由 $\Gamma_\gamma^{P_j}$ 定义的正样本满足三个不同条件，按以下顺序在公式（5）中描述：属于锚点 j^{th} 的邻域，该邻域由半径 r_{th} 定义；与锚点 j^{th} 位于同一行；不属于 Γ 中相对于锚点 j^{th} 的前 γ 个元素集合。

最后，我们还定义一组相对于 $\Gamma^{N_j} \subseteq \Gamma$ 锚点（S_j^A, p_j^A）的负样本。这些负样本是位于由 r_{th} 定义的感兴趣区域之外的样本，即

$$\Gamma^{N_j} = \{ (S_m^{N_j}, p_m^{N_j}) \mid \|p_m^{N_j} - p_j^A\|_2 \geq r_{th} \} \quad (6)$$

其中每个元素 $(S_m^{N_A}, p_m^{N_A}) \in \Gamma^{N_j}$ 都是相对于 j^{th} 锚点的负样本。实际上，负样本集是通过从原始 Γ^{N_j} 中随机抽取样本生成的。

现在，为了计算给定锚点（S_j^A, p_j^A）的损失，LazyTriplet损失通过以下表达式计算：

$$\mathcal{L}_T = \max(\|d_j^A - d_{c_1}^P\|_2 - \|d_j^A - d_{c_2}^N\|_2 + m, 0) \quad (7)$$

其中，d_j^A 是 j^{th} 锚点扫描 S_j^A 的描述符；d_c^P 是物理空间中距离最近的正样本的描述符

$(S_{c_1}^P, p_{c_1}^P) \in \Gamma^P$ S_j^A, p_j^A $c_1 = \arg\min_l \|p_l^P - p_j^A\|_2$ d_c^N 相对于 (，)，即 argmin；$S_{c_2}^P$, $p_{c_2}^P$ 是描述子空间中距离最近的负样本的描述符，$\in \Gamma^P$ 相对于 (S_j^A, p_j^A)，即 $c_2 = \arg\min_m \|d_m^P - d_j^A\|_2$。最后，$m$ 是一个间隔值。该过程如图3所示。

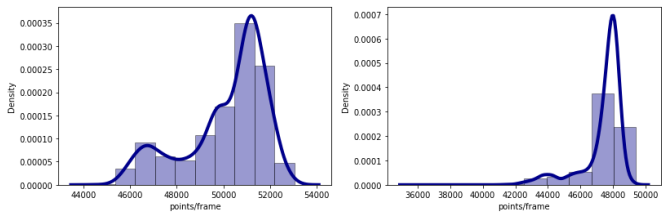
IV. 实验评估

本工作的目标是评估果园中地点识别的鲁棒性，即目标之一是评估所提出方法在不同季节下的鲁棒性。因此，为支持所提出的贡献，本节介绍了一个自行构建的果园数据集、实验与评估协议以及实现细节，最后讨论了在检索任务中以及作为定位框架一部分所获得的结果。

A. 果园数据集

该数据集包含两个序列，均使用配备 Velodyne VLP32 三维激光雷达（10Hz）和 ZED-F9P RTK-GPS（5Hz）的 Clearpath Husky 移动机器人，在英国肯特郡同一座布拉姆利苹果园中采集，但采集季节不同（即夏季和秋季）。这两个序列记录了处于不同开花状态的树木，因此树木及环境状况的变化带来了额外挑战。两次采集均在晴朗干燥、风速较低的日子进行。图 4 展示了记录期间配备传感器的机器人。对于本研究，当来自不同重复访问的扫描数据之间距离在 10m 以内时，即视为存在闭环；当来自不同重复访问的两份扫描数据之间距离在 10m 以内时，即构成锚点-正样本对。

在这两个序列中，机器人均沿果园行间行驶，形成了包含同向和反向重复访问路段的路径。夏季序列于7月采集，包含3244次扫描及其同步位姿。其中，954次扫描为锚点。夏季序列中的路段仅被重复访问一次，而秋季序列中的部分路段则被重复访问多次。秋季序列于11月采集，包含3674次扫描及其同步



(a) (b)

图5：a）summer序列和 b）autumn序列的每帧点数分布示意图。

表I：数据集中帧和回环的数量。回环数量对应于一个 $r_{th} \leq 10m$。

序列	长度 [帧]	回环 [帧]	点数/帧
秋季	3674	2311	47k±1k
夏季	3244	954	50k±2k

位姿，其中2311次扫描为锚点。图5展示了每个序列中每次扫描的点分布，而真实回环如图6(a)和图6(c)所示。表I总结了每个序列中扫描数据及现有回环的相关信息。图6(b)和图6(d)分别展示了对于给定锚点，在夏季序列和秋季序列中10米范围内对应的正样本。

B. 地点识别评估

采用标准检索指标召回率对所提出方法进行评估，其定义如下：

$$\text{Recall} = \frac{\text{TP (\# Retrieved Loops)}}{\text{TP + FN (\# Loops)}}, \tag{8}$$

即报告检索到的最相似候选项的召回率（即 recall@1），以及最相似的 1% 候选项的召回率（即 recall@1%）。

因此，为评估所提出方法的性能，我们采用了以下检索协议。首先，考虑扫描数据及其对应的位姿$\Gamma = \{(S_i, p_i) | S_i \in \mathcal{S}, p_i \in \mathcal{P}\}$。对于每个元素 $(S_i, p_i) \in \Gamma$，我们将模型生成的 (S_i, p_i) 描述符记为 D_i，由此定义描述符集合 $\mathcal{D} = \{D_i | D_i \in \mathbb{R}^K\}$，其维度为 K，且 $|D| = |\Gamma|$。我们还定义数据库 $\Theta_t = \{D_m^{\Theta_t} | D_m^{\Theta_t} \in \mathbb{R}^K\} \subseteq \mathcal{D}$，其中包含 $m \in [1, ..., M_t]$，M_t 表示迭代 t时数据库中的描述符数量。

此外，我们还将考虑先前定义的锚点对集合 Γ^A，以定义 $\mathcal{D}^A = \{D_j^A \in D | (S_j^A, p_j^A) \Gamma^A\} \subseteq \mathcal{D}$，即与锚点 $(S_j^A, p_j^A) \in \Gamma^A$相关联的描述符集合；请注意 $|D^A| = |\Gamma^A|$。

给定锚点的描述符 $D_j^A \in \mathcal{D}^A$ 和数据库 Θ_t，使用K近邻（KNN）方法从 Θ_t 中 $\mathcal{D}_{N\,t}^j = \{d_n \in \Theta_t | \|d_n - d_j^A\|_2 \leq \|d_{n+1} - d_j^A\|_2 \wedge n \in [1, ..., N]\}$ $|\mathcal{D}_{N\,t}^j| = N$ 检索出锚点 j^{th} 的 N 候选项排名前列$\mathcal{D}_{N,\,t}^j$ 相对于锚点的集合，并定义如下：，（其中，）即

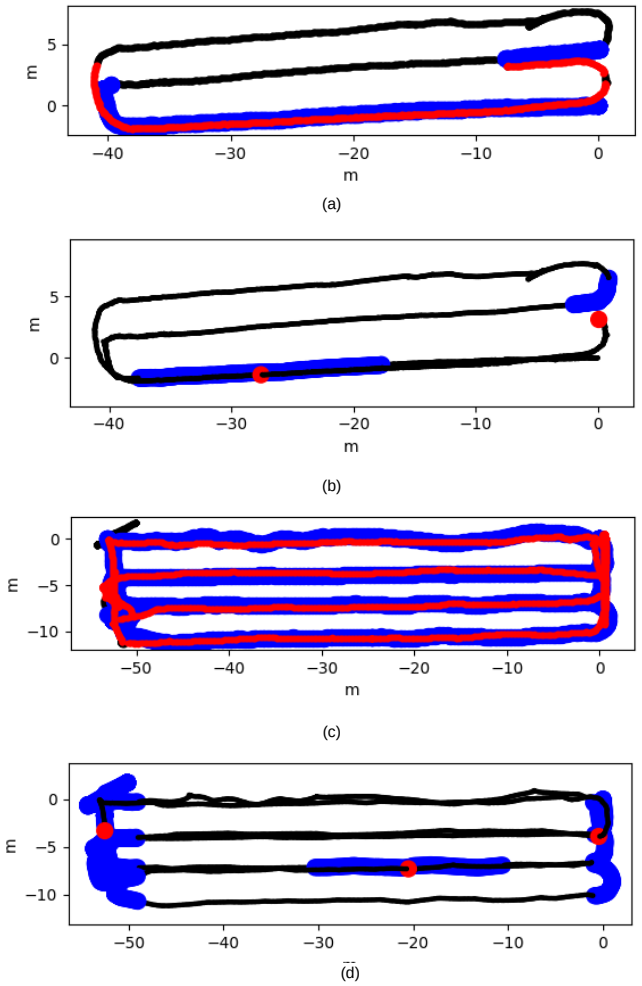


图6：真实路径示意图。图 a) 和 c) 分别展示了夏季和秋季序列中的回环，其中锚点以红色标出，正样本以蓝色标出，其余点以黑色标出。图 b) 和 d) 分别展示了夏季和秋季序列中距给定锚点 10 米以内的正样本，其中锚点以红色标出，正样本以蓝色标出。

N 中与 j^{th} 锚点的描述符在描述子空间中距离更近的 Θ_t 描述符集合。

因此，为了计算给定模型的 recall@N（其中 N 表示前 N 个候选项），给定一组排名靠前的候选描述符 $\mathcal{D}_{N\,t}^j$ $\Gamma_{N\,t}^j = \{S_n\ p_n \in \Gamma | d_n \in \mathcal{D}_{N\,t}^j\}$ $|\Gamma_{N\,t}^j| = |\mathcal{D}_{N\,t}^j| = N$，我们首先定义，（，）（其中，，），即与排名靠前的候选描述符集合相关联的（扫描数据和位姿）对集合。然后，相对于 j^{th} 锚点，真正例的数量（见公式 (8)）定义为$|\Gamma_{N,\,t}^j \cap \Gamma_\gamma^{P_j}|$，即先前定义的候选项排名前列集合 $(\Gamma_{N,\,t}^j)$ 与正样本集合 $(\Gamma_\gamma^{P_j})$ 中的共同元素数量。

本文提出的方法与最先进的全局特征聚合方法进行比较，例如 NetVLAD [7],GeM[8]MAC [8], 和 SPoC[15]。因此，为了公平地比较各种方法并评估每种方法的优劣，所有全局特征聚合方法都使用相同的特征提取器（即骨干网络）

网络。本文采用两种特征提取器：PointNet[5] 和 ResNet50[17]。这两种网络处理点云的方式各不相同：PointNet 是一种直接从点云中学习特征的神经网络；而 ResNet50 是一种基于卷积神经网络的网络，最初是为图像提出的，因此需要将点云投影到球面或鸟瞰图（BEV）等代理表示中。在基于 ResNet50 的实验中，点云被投影为 BEV表示。比较这两种特征提取器，有助于更深入地评估所提出的方法在接收不同输入表示时的表现。

C. 实现与训练细节

所有模型均在 NVIDIA GeForce RTX 3090 图形处理器上训练了 100 个 epoch，使用最近的正样本和 20 个负样本。间隔值 m 设置为 0.5，并使用 AdamW优化器优化模型参数，学习率（Lr）为 0.0001，权重衰减（Wd）为 0.0005。此外，所有提出的实验均使用 Python 3.8 和 CUDA 11.6 版 PyTorch 进行网络训练。

在所有实验中，除非另有说明，点云分别沿 x 轴和 y 轴在 $\pm 15m$ 处进行裁剪，然后下采样至 20k 个点用于 PointNet，或下采样至 512 个点用于 ResNet50，采样点均为随机选取。此外，在每个训练步骤中，点云都会通过施加最大旋转角度为 $\pm 180°$ 的随机刚体变换进行增强。

此外，如 [7], 所述，NetVLAD 的聚类数量设为 64。可训练参数 $A = [a_1, a_2, a_3]^T$ 根据正态分布进行初始化，其中$\mu = 0$ 和 $\sigma = 0.1$。此外，还调整了每个网络的配置，以获得最佳性能。例如，ResNet50 在全局描述子维度为 512 时表现最佳，而 PointNet 则需要 2048 维。

1）基于 *PointNet*： PointNet 是一种直接从点云中学习特征的神经网络。在本研究中，按照 [7]的建议，在最大池化层之前截取 PointNet。尽管 PointNet 能够直接从点云中学习，但仍需要固定大小的输入，因此不直接使用原始扫描数据作为输入，而是将 S_i 下采样至 S_i'。

PointNet 的输入为张量 $S' \in \mathbb{R}^{b\times 1\times n'\times F}$，其中$b$ 表示批大小， n' 表示点的数量（已下采样），F 表示特征维度。在本研究中，测试时批大小设为 20，训练时设为 1；点的数量设为 $20k$， F 设为 3。网络的第一层将这个三维张量映射到 1024 维，并在 PointNet 的输出端将其降至 512 维（即 $Z \in \mathbb{R}^{b\times 20k\times 512}$）。全局特征聚合模块接收这个 Z 张量，并输出一个描述符 $D \in \mathbb{R}^{b\times 2048}$。

2）基于 *ResNet50*： ResNet50 是一种基于卷积神经网络的网络，可从类似图像的表示中提取特征。在本研究中，使用 ResNet50 作为特征提取器时，会将点云投影为 BEV表示，该表示通过将点云离散化为二维网格得到，其中

表 II：数据集中帧和闭环的数量。

实验	序列	训练集	测试集
同季节	秋季	60%（1386）	40%（925）
跨季节	秋季/夏季	2311	954

表 III：同季节实验中获得的实证结果，其中秋季序列被划分为60%的训练集和40%的测试集。

	Recall@1		Recall@1%	
	PointNet	ResNet	PointNet	ResNet
VLAD	0.11	0.41	0.68	0.92
SPoC	0.50	0.44	0.85	0.93
GeM	0.49	0.41	0.84	0.93
MAC	0.49	0.44	0.84	0.92
ORCHNet（我们的模型）	**0.52**	**0.45**	**0.92**	**0.94**

各单元格填充落入相应单元格的高度（z 坐标）、点密度和强度。因此，使用 ResNet50 从点云中提取特征的过程如下：首先，将扫描 $S_i \in \mathbb{R}^{n\times 3}$下采样为 $S_i' \in \mathbb{R}^{500\times 3}$，再将其投影为图像 $B_i = [R_h, R_d, R_i] \in \mathbb{R}^{b\times 3\times h\times w}$。该图像由扫描投影到 BEV表示后的高度（$z$ 坐标） $R_h \in \mathbb{R}_h^{h\times w}$ 、点密度 $R_d \in \mathbb{R}^{h\times w}$ 和强度 $R_i \in \mathbb{R}^{h\times w}$ 拼接而成，其中 b为批大小， h 和 w 分别为图像的高度和宽度。在本研究中， h 和 w 均设为 256，测试期间批大小设为 20，训练期间设为 1。ResNet50 以 B_i 作为输入，并输出张量$Z' \in \mathbb{R}^{b\times 2048\times 16\times 16}$，随后将其重塑为 $Z \in \mathbb{R}^{b\times 2048\times 256}$。然后，将 Z 输入全局特征聚合模块，输出全局描述子 $D \in \mathbb{R}^{b\times 512}$。

D. 检索性能

本节介绍并讨论使用 ORCHNet 获得的实证结果。我们进行了两项主要实验：同季节实验和跨季节实验。对于同季节数据实验，秋季序列被划分为 60% 用于训练，40% 用于测试。对于跨季节实验，模型在秋季序列上训练，并在夏季序列上测试。表 II 汇总了各项实验中训练样本和测试样本的划分情况。

同季节数据上的结果如表 III 所示，结果表明，所有方法（包括 ORCHNet）在仅检索一个候选项时（即 recall@1）表现都不佳。表现不佳的原因可能是果园中的几何差异较小，导致网络无法提取描述性特征，从而区分不同行中的扫描数据。另一方面，将数据库中排名前 1% 的条目作为闭环候选项（本实验中为 31 个候选项）进行检索时，性能显著提升，召回率几乎达到完美水平。这意味着，尽管

表 IV：跨季节实验的实证结果，其中秋季序列用于训练，夏季序列用于测试。

	Recall@1		Recall@1%	
	PointNet	ResNet	PointNet	ResNet
VLAD	0.061	0.35	0.58	0.92
SPoC	0.43	0.41	0.76	0.93
GeM	0.43	0.38	0.79	0.93
MAC	0.40	0.40	0.79	**0.96**
ORCHNet（我们的）	**0.50**	**0.42**	**0.85**	0.95

排名第一的候选项存在歧义，但在检索排名前 1% 的条目时，所有网络都能取得尚可的性能，其中 ORCHNet 优于其他方法。

跨季节实验的结果见表 IV。结果表明，尽管所有模型的性能都出现了预期中的下降，基于PointNet的方法在top@1检索方面仍表现更好，而基于ResNet50的方法在top 1%检索方面表现更佳。结果还表明，ORCHNet能够在不同季节保持性能，由此展现出其对季节变化的鲁棒性。

E. 定位框架评估

我们还在自适应蒙特卡洛定位（AMCL）框架中对ORCHNet进行了评估，并报告了秋季序列的定性结果。评估内容是比较AMCL（仅AMCL）、AMCL-ORCHANet、同步定位与建图（SLAM）和RTK-GPS生成的路径。首先，采用[18] 提出的基于HMaps的SLAM方法，为秋季序列创建环境表示（见图7）。随后，对AMCL方法进行调整，使其纳入ORCHNet模型提供的一组闭环候选。在AMCL的重采样阶段进行了调整，采用 [19] 提出的分层方法来选择：（i）从当前粒子集合中采样给定粒子，或（ii）从闭环候选项集合中采样该粒子。闭环候选的采样取决于每个闭环候选项的相似度得分，这意味着相似的地点候选具有更高的采样概率。该过程通过多项式重采样实现。此外，还采用基于扫描匹配的LiDAR里程计方法来预测粒子状态。关于权重更新，需要注意的是，计算每个粒子的状态时仅使用少量点（每50个点取1个）。

为补充实验结果，图7展示了对所提出的AMCL-ORCHNet框架的定性评估，并将其与一种AMCL实现（即 不包含回环检测）、基于HMaps的同步定位与建图方法以及通过RTK-GPS获取的真实路径进行了比较。尽管秋季序列所在的果园空间范围较小，且各次扫描的重叠度很高，但点云较为稀疏（如图1所示），这给将扫描数据配准到地图中带来了挑战。尽管存在这些挑战，

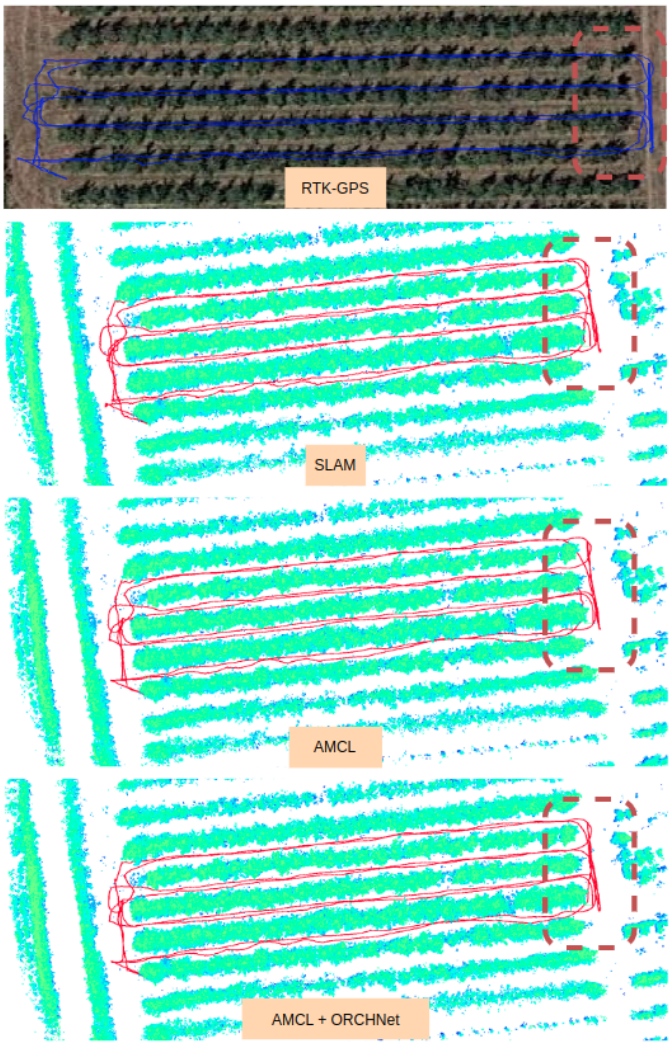


图7：秋季序列和定位框架的结果。从上到下依次为：RTK-GPS 数据、基于HMaps的SLAM、AMCL，以及采用 ORCHNet 提供的地点候选的 AMCL。

基于HMaps的同步定位与建图方法仍生成了良好的地图表示。图7显示，基线AMCL整体上取得了合理的定位性能；然而，在几何路径最右侧部分（以红色标出，见图8），AMCL有几处穿过了树木，而AMCL-ORCHNet方法纠正了这一问题。另一方面，在该区域，AMCL-ORCHNet方法计算出的路径与基于HMaps的同步定位与建图方法和RTK-GPS得到的路径相近。这些结果凸显了所提出的ORCHNet作为定位框架的一部分，在真实环境条件下的适用性和可行性。

V. 结论

在这项工作中，我们提出了一种基于三维激光雷达的地点识别方法（称为 ORCHNet），通过结合多种全局特征聚合方法的描述符，获得稳健的全局描述符。我们在

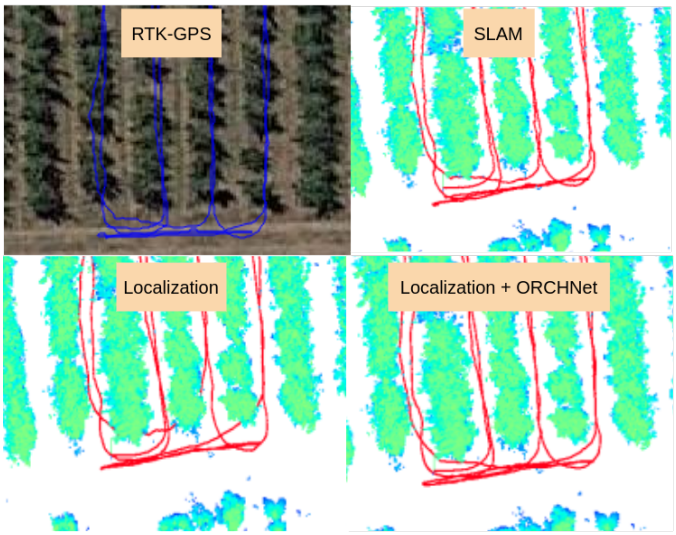


图8：秋季序列定性评估结果概览。从上到下、从左到右：RTK-GPS 数据、基于HMaps的SLAM、经典 AMCL 和基于 ORCHNet 的 AMCL。

地点识别任务和基于 AMCL 的定位框架中，将其用作闭环检测器。

ORCHNet 的实验评估使用了在夏季和秋季期间于果园采集的真实场景数据，数据由配备 32通道激光雷达的移动机器人平台获取。研究开展了两项不同任务领域的实验：第一项聚焦于基于三维激光雷达的地点识别，第二项则探索了将 ORCHNet 集成到定位框架中。在基于激光雷达的地点识别方面，研究将 ORCHNet 的鲁棒性与其他最先进方法进行了比较，并在具有挑战性的果园数据集上进行了展示，结果表明，所提出的方法在不同季节均具有可靠性。作为定位框架的一部分，AMCL-ORCHNet 方法修正了路径，而 AMCL 基线方法无法正确预测该路径。

致谢

本研究得到 GreenBotics 项目（编号 PTDC/EEI-ROB/2459/2021）的支持，该项目由葡萄牙科学技术基金会（FCT）资助。此外，本研究还部分获得了 FCT 的 UIDB/00048/2020 项目资助，以及编号为 2021.06492.BD 的博士奖学金资助。作者还要感谢 NIAB East Malling 的 Charles Whitfield博士 为果园数据采集工作提供便利。

参考文献

[1] D. Cattaneo、M. Vaghi 和 A. Valada，“Lcdnet：用于激光雷达SLAM的深度回环检测与点云配准”，*IEEE*机器人学汇刊，第38卷，第4期，第2074–2093页，2022年。[2] T. Cieslewski、S. Choudhary 和 D. Scaramuzza，“数据高效的去中- 心化视觉SLAM”，载于*2018 IEEE*机器人与自动化国际会议（*ICRA*）。IEEE，2018年，第2466–2473页。[3] C. Cadena、L. Carlone、H. Carrillo、Y. Latif、D. Scaramuzza、J. Neira、I. Reid 和 J. J. Leonard，“同时定位与建图的过去、现在与未来：迈向鲁棒感知时代”，*IEEE*机器人学汇刊，第32卷，第6期，第1309–1332页，2016年。

[4] T. Barros、R. Pereira、L. Garrote、C. Premebida 和 U. J. Nunes，“地点识别综述：深度学习方法的最新进展”，*arXiv*预印本 *arXiv:2106.10458*，2021年。[5] C. R. Qi、H. Su、K. Mo 和 L. J. Guibas，“Pointnet：用于三维分类和分割的点集深度学习”，载于*IEEE*计算机视觉与模式识别会议（*CVPR*）论文集，2017年7月。[6] F. Ou、Y. Li 和 Z. Miao，“利用注意力得分图进行大规模非结构化果园的地点识别”，*IEEE*机器人与自动化快报，2023年。[7] M. Angelina Uy 和 G. Hee Lee，“PointNetVLAD：基于深度点云的大规模地点识别检索”，载于*IEEE*计算机视觉与模式识别会议论文集，2018年，第4470–4479页。[8] J. Komorowski，“Minkloc3d：基于点云的大规模地点识- 别”，载于*IEEE/CVF*冬季计算机视觉应用会议论文集，2021年，第1790–1799页。[9] S. Siva、Z. Nahman 和 H. Zhang，“用于基于三维点云地点识别的基于体素的表示学习”，载于*2020 IEEE/RSJ*智能机器人与系统国际会议（*IROS*）。IEEE，2020年，第8351–8357页。[10] L. Li、X. Kong、X. Zhao、T. Huang、W. Li、F. Wen、H. Zhang 和 Y. Liu，“Rinet：利用旋转不变神经网络的高效基于三维激光雷达的地点识别”，*IEEE*机器人与自动化快报，第7卷，第2期，第4321–4328页，2022年。[11] T. Barros、L. Garrote、R. Pereira、C. Premebida 和 U. J. Nunes，“Attdl- net：用于三维激光雷达地点识别的基于注意力的深度网络”，载于*ROBOT2022*：第五届伊比利亚机器人会议：机器人技术进展，第*1*卷。Springer，2022年，第309–320页。[12] R. Arandjelovic、P. Gronat、A. Torii、T. Pajdla 和 J. Sivic，“Netvlad：用于弱监督地点识别的卷积神经网络架构”，载于*IEEE*计算机视觉与模式识别会议论文集，2016年，第5297–5307页。[13] C. R. Qi、H. Su、K. Mo 和 L. J. Guibas，“PointNet：用于三维分类和分割的点集深度学习”，载于*IEEE*计算机视觉与模式识别会议论文集，2017年，第652–660页。[14] G. Tolias、R. Sicre 和 H´egou，“通过卷积神经网络激活的积分最大池化进行特定物体检索”，*arXiv*预印本 *arXiv:1511 05879*，2015年。[15] A. B. Yandex 和 V. Lempitsky，“聚合局部深度特征用于图像检索”，载于 *2015 IEEE* 国际计算机视觉会议（*ICCV*），2015年，第1269–1277页。[16] F. Radenovi´c、G. Tolias 和 O. Chum，“无需人工标注的卷积神经网络图像检索微调”，*IEEE*模式分析与机器智能汇刊，第41卷，第7期，第1655–1668页，2018年。[17] K. He、X. Zhang、S. Ren 和 J. Sun，“用于图像识别的深度残差学习”，载于*IEEE*计算机视觉与模式识别会议论文集，2016年，第770–778页。[18] L. Garrote、C. Premebida、D. Silva 和 U. J. Nunes，“Hmaps——用于环境表示的混合高度体素地图”，载于*2018 IEEE/RSJ*智能机器人与系统国际会议（*IROS*）。IEEE，2018年，第1197–1203页。[19] L. Garrote、M. Torres、T. Barros、J. Perdiz、C. Premebida 和 U. J. Nunes，“由绝对室内定位系统辅助、采用强化学习地图更新决策的移动机器人定位”，载于*2019 IEEE/RSJ*智能机器人与系统国际会议（*IROS*），2019年，第1620–1626页。