

无序地标视觉导航

Hao Ren¹, Junzhe Zhu¹, Yihan Li¹, Zetong Bi¹, Le Zheng¹, Zhi Li¹,
Yiqing Yuan¹, Zhaoliang Wan², Dizhe Zhang², Lu Qi², Hui Cheng^{1*}

¹ 中山大学, 中国广州

² Insta360 研究院, 中国深圳

<https://hren20.github.io/ulvn-website>

摘要 图像目标导航 (Image-Goal Navigation) 是具身智能的一项基础能力, 但其实际部署受到强先验假设的制约。现有方法大多依赖时序有序的视频流或辅助传感器 (如深度、激光雷达) 来维持空间一致性。这类时序依赖与多模态依赖严重限制了可扩展性, 尤其是在使用众包或预先录制的无序图像集合部署机器人时。一旦去除时序先验, 现有方法便会遭遇严重的感知混淆 (Perceptual Aliasing)、噪声关联以及灾难性的建图失败。针对这一尚未充分探索的挑战, 本文提出无序地标视觉导航 (Unordered Landmark Visual Navigation, ULVN) ——一个不依赖时序与里程计先验的纯 RGB 统一框架。ULVN 通过整合建图、定位与规划, 系统性抑制误差累积。具体而言, 它借助标定式几何验证与最大生成森林 (Maximum Spanning Forest, MSF) 精化, 直接从非结构化图像构建鲁棒的二维拓扑地图。在闭环执行阶段, ULVN 摒弃时序启发式, 采用基于图的置信度传播滤波器, 配合熵自适应融合 (Entropy-Adaptive Fusion) 完成全局定位与动态子目标规划。仿真与真机部署的大量实验表明, ULVN 显著优于现有最先进方法。

关键词: 视觉导航 · 拓扑建图 · 视觉定位

1 引言

视觉导航 [6, 16, 21, 78] 是自主智能体的基础能力, 支撑着从运动规划 [17] 到家用机器人 [35] 的广泛应用。从认知科学的视角看, 人类导航主要依靠视觉观测形成对环境的拓扑理解, 无需精确的度量坐标, 也无需激光雷达这类昂贵的主动传感器 [52, 57]。因此, 在具身智能中复现这一能力已成为极具价值的目标 [1, 57, 83]。理想情况下, 具身智能体应主要基于 RGB 观测在预先构建的拓扑地图 [7, 10, 11, 52, 59, 65, 67] 中运行, 形成整体空间感知与可靠的目标导向行为。

与这一理想相比, 许多现有视觉导航系统隐式依赖在真实场景中很少成立的简化条件。它们往往要求时序有序的视频流, 以为图构建和策略学习提供强时序先验 [65, 80]。一旦缺失这种时序结构, 例如使用房产列表中的无序照片集合时, 学到的转移关系往往失效。没有时序先验, 区分相似地点变得困难, 拓扑连接充满噪声、动作选择不稳定, 进而频繁引发振荡行为。此外, 许多系统依赖深度或激光雷达等辅助传感器稳定定位 [19, 26]。仅依赖外观会使几何验证变得复

* 通讯作者: Hui Cheng (chengh9@mail.sysu.edu.cn)。

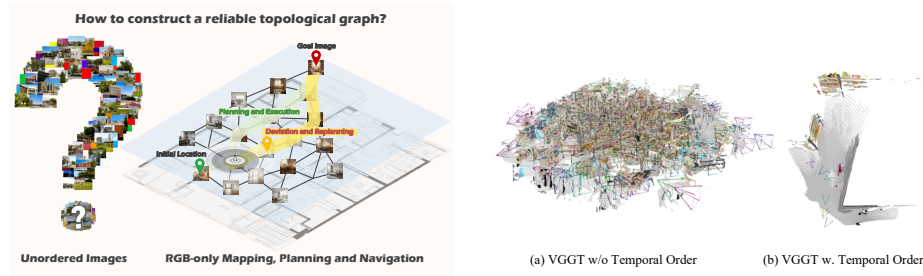


图 1: 从无序图像到导航。本文系统 ULVN 仅从非结构化的 RGB 图像集合自动构建拓扑表示, 从而支持闭环规划、执行与重规划。相比之下, VGGT 等基础模型难以处理无序图像。

杂, 即便先进的重建模型也难以从稀疏、不连通的观测中估计位姿 [61, 66, 71]。因此, 当前方法的成功严重依赖时序连续性、辅助传感, 或两者兼有。

消除这些假设带来两个递进的挑战。第一, 去除时序先验后, 拓扑建图只能仅凭外观推断连通性, 这使其易受感知混淆与虚假边影响, 规划与定位随之脆弱——缺少时序更新抑制歧义, 误差会不断累积 [7, 11, 39, 73]。第二, 在去除时序先验的同时去除多模态传感会进一步加剧难度。没有深度、激光雷达或里程计, 尺度与视角歧义的消解显著变难, 累积漂移增大, 仅凭 RGB 证据进行闭环校正也更加复杂 [6, 8, 74, 78]。这些挑战引出本文的核心问题: 能否构建一个纯 RGB 的导航框架, 仅使用无序图像即可实现可靠的闭环导航, 而不依赖里程计、深度或时序先验?

为回答该问题, 本文提出 ULVN——一个从统一系统优化视角设计的无序地标导航框架, 确保建图、定位与规划以一致性和可靠性为目标协同设计。在拓扑建图方面, ULVN 采用视觉位置识别 (Visual Place Recognition, VPR) 与几何验证相结合的方案, 并配备一次性 (One-Shot) 数据驱动阈值标定, 以提升边验证的鲁棒性。进一步通过提取最大生成森林骨架并重新插入强回环闭合来精化图结构, 从无序图像中得到可靠拓扑。在定位方面, 本文提出全局置信度传播方案, 通过贝叶斯估计与基于熵的自适应融合更新空间概率, 有效缓解视觉歧义下的定位漂移。最后, ULVN 利用该全局置信度状态在图上规划路径, 并在出现偏离时动态重规划。本文在仿真与真实环境中对 ULVN 进行了全面验证, 主要贡献如下:

- **无序地标视觉导航的统一框架。** 本文研究不含时序与多模态先验的纯 RGB 导航——一个重要却仍未充分探索的设定, 并提供整合建图、定位与导航的统一框架, 以抑制误差累积。
- **从无序 RGB 图像构建更可靠的拓扑地图。** 为获得高质量拓扑地图, ULVN 采用两阶段方案: 先用一次性阈值估计标定基于 VPR 的候选检索与几何验证, 再通过带回环重插入的 MSF 骨架精化图结构。
- **二维图上无里程计的全局定位与导航。** ULVN 通过贝叶斯估计与基于熵的自适应融合更新全局置信度状态, 抑制视觉歧义下的漂移, 并利用该置信度在拓扑图上规划与重规划路径。
- **基准与全面实验验证。** 本文公开发布仿真与真实环境下的无序地标视觉导航数据集, 同时提供数据采集流程与评测指标, 以支持可复现研究。本文对各组件的误差累积给出消融分析, 并表明 ULVN 性能强于现有方法。

2 相关工作

拓扑地图构建。早期基于外观的视觉导航与 SLAM 出于效率和鲁棒性考虑使用拓扑地图，将视觉位置识别 (VPR) 与概率推断耦合以缓解感知混淆 [11, 39, 81]。这条研究路线也表明，在许多场景中仅用图像的表示可以成为稠密度量重建 [15, 27, 51] 的可扩展替代方案。近期大量纯 RGB 系统假设采集过程时序有序 [47, 58, 63, 64]，实际上得到的是一维地标链 [20, 49, 62, 65]，并聚焦于局部规划器 [37, 48, 50, 69, 78]。这类设计虽实用，却依赖时序结构，与不存在时序信号的图像袋 (Bag-of-Images) 设定不兼容。当序列缺失时，二维拓扑图通常通过基于 VPR 的候选检索加几何验证构建 [18, 38]。经典流程用局部特征检索邻居，并通过极或同伦 (Homography) 检验验证边 [54, 55]，但对超参数 (如 RANSAC 内点阈值) 敏感，需在边的精确率与召回率之间权衡，最终影响连通性。学习得到的全局描述符提升了检索的不变性 (视角、光照、季节)，但描述符相似并不意味着可通行，验证仍不可或缺。近期的鲁棒估计器 [2, 46, 60, 72] 缓解但并未消除跨环境对阈值和上界的依赖 [14]。

拓扑地图上的定位与规划。给定地图后，智能体必须完成定位与路径规划。许多规划器假设可获得精确的全局信息 (如深度或度量地图) [5, 41]，而这在纯 RGB 场景中往往不可用。一种流行的替代方案是估计当前视图与目标视图之间的“时序距离” [28, 36, 42, 52, 57, 59]。然而，这一代理量与测地距离的相关性可能很差 [4, 32, 43]，通常预设近似恒定速度 [30, 40, 82]，在大量候选上评估代价高昂 [65]，且训练需要特定机器人的轨迹数据 [57, 65, 76]。近期工作将子目标选择重新表述为基于 VPR 的定位 [9, 65]，提升了效率与鲁棒性，但许多模型仍隐式保留一维拓扑假设 [26, 65]，在含交叉口、分支与回环的通用二维图上会失效。与这类时序假设不同，本文方法利用图邻接定义的拓扑转移模型，在无里程计条件下于二维图上定位。

3 方法

为严格仅从无序 RGB 图像实现图像目标导航，本文框架 ULVN 采用纯拓扑路线，摒弃易出错的度量重建与时序启发式。在所给图像库具备充分视觉重叠的前提下，ULVN 通过三个紧密集成的模块系统性抑制误差累积：鲁棒拓扑图构建 (RAVEL, 第 3.1 节)、全局状态估计 (BPL, 第 3.2 节) 与闭环视觉规划 (BASS, 第 3.3 节)。

问题定义与记号。设 $\mathcal{L} = \{I_i\}_{i=1}^N$ 为采集到的无序图像集合， I_g 为目标图像， I_t 为时刻 t 的实时观测，且不提供里程计或时序先验。全局图像编码器 $f(\cdot)$ 产生经 L2 归一化的嵌入 $\mathbf{z}_i^r = f(I_i)$ 、 $\mathbf{z}_g = f(I_g)$ 与 $\mathbf{z}_t = f(I_t)$ 。ULVN 构建一个有向加权图：

$$\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{W}), \quad \mathcal{V} \subseteq \{1, \dots, N\}, \quad \mathbf{W} = \{W_{ij}\}_{(i,j) \in \mathcal{E}} \quad (1)$$

由于度量距离不可用，本文以视觉重叠作为空间邻近性的代理量。每条边 $(i, j) \in \mathcal{E}$ 具有整数权重 $W_{ij} \in \mathbb{Z}_{>0}$ ，表示已验证的局部特征内点数量。

3.1 视觉地标图构建

本文提出 RAVEL (Robust Augmentation and VERification of Landmarks, 地标的鲁棒增强与验证)，一条将 $\mathcal{L}$ 转换为加权有向视觉图的流程。无序集合天然

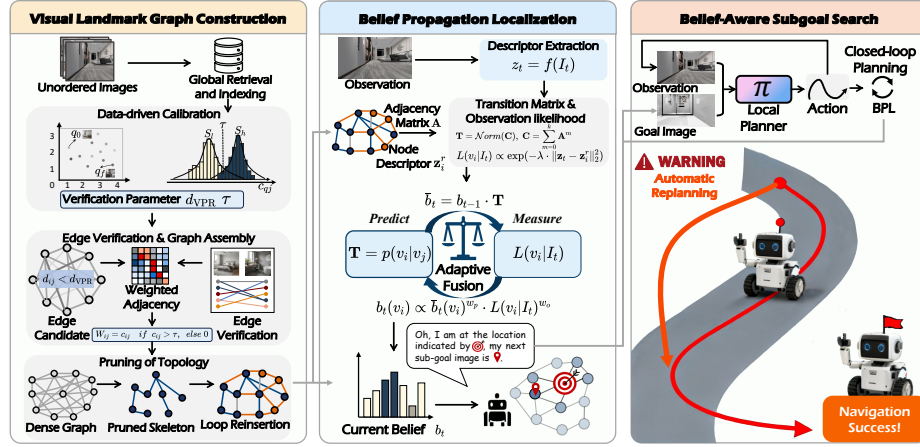


图 2: 无序地标视觉导航总览。(左) 在无序图像上通过特征检索构建经标定与剪枝的拓扑结构。(中) 嵌入当前观测, 并借助自适应融合, 用观测似然与转移矩阵 T 更新节点置信度。(右) 依据置信度选取下一个地标并闭环规划; 置信度发生偏移时自动触发向目标的重规划。

产生稠密冗余的候选对, 会造成内存膨胀并降低下游规划质量。RAVEL 先用全局检索实现快速召回, 再以几何验证作为高精确率瓶颈, 最终将图剪枝为紧凑且具结构意义的骨架。

全局检索与索引。为避免穷举式两两匹配, 本文先用图像编码器 $f(\cdot)$ 为每张图像提取经 L2 归一化的全局描述符 $z_i^r = f(I_i)$ 。虽然 $f(\cdot)$ 可以是通用骨干网络 (如 ResNet、DINOv2), 但实验表明专门针对视觉位置识别 (VPR) 训练的表示能给出最鲁棒的检索候选。所有全局描述符均使用 FAISS [13, 29] 建立索引。

为自动标定验证阈值, 本文选取首张图像 q_0 , 并检索其最远的有效邻居 $q_f = \arg \max_{j \in \mathcal{N}_{\text{FAISS}}(q_0)} \|z_{q_0} - z_j\|_2$, 同时显式排除无效索引输出。对每个锚点 $q \in \{q_0, q_f\}$, 与所有被检索图像执行完整特征匹配, 为每一对给出内点数 c_{qj} 与锚点描述符距离 d_{qj} 。对汇集的内点数应用 k 均值聚类, 得到最高置信簇 S_h 与次高置信簇 S_l 。内点阈值 τ 定义为:

$$\tau = \frac{1}{2} (\min(S_h) + \max(S_l)). \quad (2)$$

全局距离阈值由 S_h 导出: $d_{VPR} = \max_{(q,j) \in S_h} \|z_q - z_j\|_2$ 。

局部几何验证与图剪枝。对每个全局描述符距离满足 $s_{ij} \leq d_{VPR}$ 的候选对, 使用深度匹配器 (如 LightGlue [34]) 执行局部几何验证, 得到精确的内点匹配数 c_{ij} 。加权邻接矩阵更新为:

$$W_{ij} = W_{ji} = \begin{cases} c_{ij}, & c_{ij} > \tau \\ 0, & \text{其他.} \end{cases} \quad (3)$$

由无序集合构建的原始图本身稠密, 且包含弱边与冗余边, 这不仅造成内存膨胀, 还会把概率质量扩散到噪声路径中, 从而削弱后续的置信度传播。为提取紧凑可靠的骨架, 本文计算最大生成森林 (MSF) $\mathcal{F} = (\mathcal{V}, \mathcal{E}_{\text{MSF}})$ 。通过最大化

总边权 $\sum_{(i,j) \in \mathcal{E}_{\text{MSF}}} W_{ij}$, MSF 确保导航骨干严格由最强的几何转移构成, 最小化局部执行失败。

MSF 虽能有效消除视觉混淆, 却会破坏对备选路径与拓扑纠错至关重要的环结构。为恢复这些环, 本文引入数据驱动的回环重插入 (Strong-Loop Reinsertion)。不采用脆弱的启发式规则, 而是通过 k 均值将 $\mathcal{G}$ 中所有边权聚为 $k = 10$ 个簇, 并将动态阈值 τ_{add} 定义为两个最高质心簇内的最小权重, 以保证对不同场景的适应性。重新插入满足 $W_{ij} > \tau_{\text{add}}$ 的被排除边, 即得最终鲁棒图 $\mathcal{G}_{\text{pruned}}$, 在结构可靠性与关键拓扑冗余之间取得平衡。

3.2 置信度传播定位

在经典贝叶斯滤波 [68] 的基础上, 本文将先前工作 [65, 73] 使用的一维时序定位滤波器扩展到具有复杂分支与回环的任意二维空间图上, 提出置信度传播定位 (Belief Propagation Localization, BPL) 滤波器, 通过递归的预测—校正循环在所有节点 $v_i \in \mathcal{V}$ 上维护置信度分布 b_t 。

拓扑预测步。 设 $\mathbf{A}$ 为由 $\mathcal{G}_{\text{pruned}}$ 导出的二值邻接矩阵。为在无里程计条件下建模机器人多步可能的运动, 本文定义聚合长度至多为 $K = 3$ 的路径的累积可达矩阵:

$$\mathbf{C} = \sum_{m=0}^K \mathbf{A}^m, \quad (4)$$

其中 $\mathbf{A}^0 = \mathbf{I}$ 。对 $\mathbf{C}$ 按行归一化得到随机转移矩阵 $\mathbf{T}_{ij} = \mathbf{C}_{ij} / \sum_{j'} \mathbf{C}_{ij'}$ 。预测置信度为 $\bar{b}_t = b_{t-1} \cdot \mathbf{T}$ 。

自适应观测融合。 观测似然定义为 $L(v_i | I_t) = \exp(-\lambda \cdot \|\mathbf{z}_t - \mathbf{z}_i^r\|_2^2)$, 其中 $\lambda = 5$ 。本文引入熵自适应融合机制, 用先验置信度的归一化香农熵量化不确定性:

$$H_n(b_{t-1}) = -\frac{1}{\log |\mathcal{V}|} \sum_{v_i \in \mathcal{V}} b_{t-1}(v_i) \log b_{t-1}(v_i). \quad (5)$$

为动态平衡预测与观测, 融合权重直接由熵给出: $w_p = 1 - H_n(b_{t-1})$ 、 $w_o = H_n(b_{t-1})$, 确保两者严格和为 1。后验置信度通过归一化的加权几何平均计算:

$$b_t(v_i) \propto \bar{b}_t(v_i)^{w_p} \cdot L(v_i | I_t)^{w_o}. \quad (6)$$

估计位置取最大后验 (Maximum A Posteriori, MAP) 节点:
 $\hat{v}_t = \arg \max_{v_i} b_t(v_i)$ 。

3.3 面向导航的置信度感知子目标搜索

在图构建完成、定位滤波器初始化后, 置信度感知子目标搜索 (Belief-Aware Subgoal Search, BASS) 完全基于图像完成路径规划与执行。

任务初始化。 导航任务由目标图像 I_g 启动。规划前需先将起点与目标状态锚定到图 $\mathcal{G}$ 上。起点节点 v_s 由初始全局置信度 b_0 的最大后验估计确定: $v_s = \arg \max_{v_i \in \mathcal{V}} b_0(v_i)$ 。目标节点 v_g 取描述符相似度最大者: $v_g = \arg \max_{v_i \in \mathcal{V}} (\mathbf{z}_g^\top \mathbf{z}_i^r)$ 。

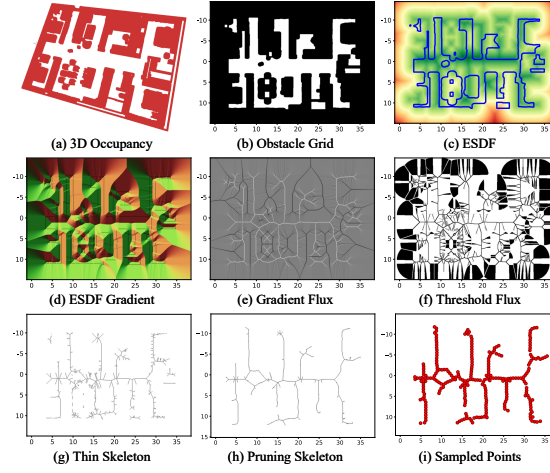


图3: 数据采集流程中的拓扑骨架提取。流程始于 (a) 三维占据地图，将其投影为 (b) 二维二值障碍栅格；(c) 欧氏符号距离函数 (ESDF) 与 (d) 其梯度；(e) 随后计算 ESDF 梯度通量，(f) 经阈值化分离中轴，(g) 细化，(h) 剪枝得到干净的拓扑骨架，最后 (i) 标出采集图像所在的节点。

拓扑规划。将 $\mathcal{G}$ 视为边权 W_{ij} 反映视觉重叠的加权图。由于视觉导航路径的鲁棒性取决于其最弱的视觉连接，本文把图 $\mathcal{G}$ 上从 v_s 到 v_g 的候选路径 $\mathcal{P}$ 的置信度定义为其最小边权：

$$\text{conf}(\mathcal{P}) = \min_{(v_i, v_j) \in \mathcal{P}} W_{ij}. \quad (7)$$

BASS 用 Dijkstra 算法提取最宽路径 $\mathcal{P}^* = \arg \max_{\mathcal{P}} \text{conf}(\mathcal{P})$ 。这一最大化最小值 (Max-Min) 目标保证所选视觉子目标序列 $(v_s, v_1, \dots, v_g)$ 具有最高的图像到图像局部控制成功概率。

纯视觉闭环执行。本文框架严格不使用里程计。机器人依次将规划出的子目标传递给基于图像的局部规划器（如 ViNT [59] 这类视觉基础模型，或通用视觉伺服策略），从而沿 $\mathcal{P}^*$ 行进。每一步，局部规划器以观测 I_t 和当前拓扑子目标图像 I_{s_n} 为输入，直接输出底层速度指令驱动机器人趋向目标。

执行过程中，BPL 持续更新置信度 b_t 。若 MAP 节点 v_t 落在规划路径 $\mathcal{P}^*$ 之外，或 v_t 到当前子目标的图距离超过 $D_{\text{thres}} = 3$ ，则判定发生偏离。一旦偏离，BASS 从 v_t 到 v_g 动态重规划新的最宽路径。当置信度在目标节点上收敛且概率超过 0.5 时，导航成功。

4 实验

4.1 实验设置

评测场景。本文在 NVIDIA Isaac Sim 仿真器中，使用面向通用机器人的高保真基准 GRScenes 数据集 [70] 评测 ULVN。共使用 10 个不同场景，在家居 (Home) 与商业 (Commercial) 环境间均分，所有场景均具备真实的物理与材质。本文还

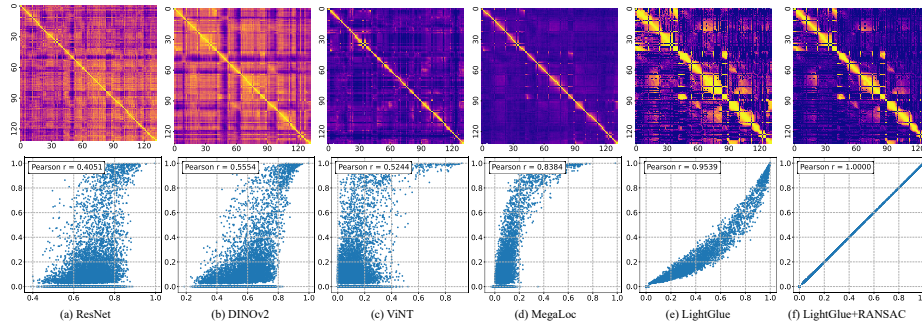


图 4: 全局描述符作为几何验证的高效代理量。本文比较多种全局描述符给出的成对相似度分数——(a) ResNet、(b) DINOv2、(c) ViNT (时序距离) 与 (d) MegaLoc (经 VPR 训练)——与 (f) LightGlue+RANSAC 给出的真值几何一致性分数。上排: 无序图像集合的成对亲度矩阵。下排: 各描述符分数 (纵轴) 与真值 (横轴) 的相关性图, 标注皮尔逊 r 。经 VPR 训练的 MegaLoc (d) 呈现强线性相关, 验证其可作为几何验证的有效且高效代理量。

使用 CARLA [12] 仿真器, 在不同感知条件下单独隔离并验证 RAVEL 与 BPL 模块。

任务与指标。本文围绕三项核心能力评测 ULVN: 1) 拓扑地图构建: 报告所生成图的边相对于真值空间连通性 (由第 4.2 节导出) 的精确率 (P)、召回率 (R)、F1 分数 ($F1$) 与准确率 ($Acc.$)。同时分析全局描述符相似度与真实几何内点之间的皮尔逊相关系数 (r), 以论证检索设计的合理性。2) 定位: 报告将连续的噪声图像流对预先构建的非序列图进行定位时的定位准确率 (MAP 估计正确识别最近拓扑节点的成功率)。3) 导航: 采用标准 Habitat ImageNav 指标 [53]: 成功率 (Success Rate, SR)、路径长度加权成功率 (Success weighted by Path Length, SPL) 与平均碰撞次数。

基线方法。本文与强有力的最先进表示与系统进行对比。拓扑地图构建方面, 评测 ResNet-50 [23]、DINOv2 [45]、MegaLoc [3]、ViNT [59] 与 PlaceNav [65]。由于 ViNT 与 PlaceNav 传统上依赖时序有序视频, 本文将其时序距离预测或 VPR 检索分数直接作为边权代理量, 以适配无序建图。定位方面, 与单帧 VPR (MegaLoc、PlaceNav)、时序距离 (ViNT) 与序列平滑 (JIST [4]) 对比。导航方面, 评测本文框架与视觉局部规划器 (ViNT、NoMaD) 的集成效果, 并与其他最先进方法对比。

实现细节。全局描述符由 MegaLoc [3] 提取并作 L2 归一化, 几何验证通过 LightGlue [34] 完成。RAVEL 标定阶段, 若高置信簇 S_h 有效 ($k=2$), 则 τ 与 d_{VPR} 由其统计量动态设定; 否则退化为默认值 $\tau_{\text{default}} = 15$ 、 $d_{VPR, \text{default}} = 1.7$ 。MSF 使用 Kruskal 算法计算, 强回环重插入以 $\tau_{\text{add}} = 1.5\tau$ 为界。BPL 的观测似然缩放因子取 $\lambda = 10$ 。

4.2 数据采集与真值流程

为严格评测拓扑地图指标, 需要真值空间连通性。如图 3 所示, 本文构建了一条自动化的仿真内流程: 导出三维场景占据信息, 并依据机器人尺寸转换为二维可通行栅格。参照 [33], 使用欧氏符号距离场 (Euclidean Signed Distance Field, ESDF) [44] 提取可通行区域的拓扑骨架, 并沿骨架施加非极大值抑制, 采样

表 1: GRScenes 上不同检索基线与本文方法的对比。

方法	P	R	F1	Acc.
VGGT [71]	0.1599	0.1659	0.1629	0.9931
PlaceNav top 2 [65]	0.5262	0.7113	0.6043	0.9948
PlaceNav top 5 [65]	0.2699	0.7655	0.3731	0.9853
ViNT top 2 [59]	0.5201	0.6775	0.5815	0.9946
ViNT top 5 [59]	0.2660	0.8246	0.3806	0.9816
RAVEL (本文)	0.7104	0.7656	0.7365	0.9970

表 2: GRScenes 上拓扑图构建的定量评测。消融实验将 RAVEL 与 Top- k ANN 检索基线以及本文方法各变体对比。

方法	P	R	F1	Acc.
Top- k ANN	0.1245	0.9121	0.2156	0.9584
RAVEL w/o (MSF, AVP τ)	0.3676	0.7177	0.4496	0.9898
RAVEL w/o MSF	0.6910	0.4220	0.5157	0.9956
RAVEL (本文)	0.7104	0.7656	0.7365	0.9970

出稀疏且空间均匀的节点。底层骨架结构提供绝对真值连通性矩阵，用于评测 RAVEL 生成图的精确率与召回率。更多细节见附录。

4.3 RAVEL 评测

全局描述符作为几何验证的代理量。本文首先评估以全局描述符作为高代价几何验证的快速召回代理量的可行性。利用 CARLA 数据，分析描述符相似度与真值几何分数 (LightGlue + RANSAC 内点) 之间的皮尔逊相关系数 (r)。如图 4 所示，ResNet (a) 与 DINOv2 (b) 等通用特征表现出弱且非线性的相关性，导致严重的感知混淆与误检索；ViNT 的时序距离 (c) 同样无法在无序设定下可靠映射到几何重叠。相反，经 VPR 训练的 MegaLoc (d) 与几何真值呈现强且紧密聚集的线性相关。这证实在初始检索阶段必须使用 VPR，才能在不丢弃有效拓扑邻居的前提下有效筛选候选。

建图性能评测。表 1 与表 2 给出 10 个 GRScenes 场景上的图构建性能。仅检索的基线 (CosPlace、ViNT) 暴露出一个根本矛盾：提高 Top- k 阈值可提升召回率，却使精确率大幅下降，F1 分数很差。这说明原始外观相似度无法替代结构验证。

消融实验 (表 2) 揭示了本文方案的作用机理。自适应阈值变体 (*Ours w/o MSF*) 通过剪除弱视觉匹配获得高精确率，但纯局部的边拒绝会使图碎片化，因而召回率偏低。引入 MSF 可解决该问题：把目标从孤立的边评估转为全局结构一致性，MSF 重建出统一骨架，使 RAVEL 能激进过滤误检的同时保持高召回率，最终取得最高 F1 分数，并在多样场景中展现稳健的泛化能力。

对视觉扰动的鲁棒性。为评估图构建对真实传感器缺陷的鲁棒性，本文引入 *+noise* 设定 (表 3)，在建图前对图像施加一系列随机视觉扰动，包括光照变化 (随机亮度缩放 $\in [0.75, 1.25]$ 、RGB 颜色偏移 $\pm 5\%$ 、对比度调整 $\in [0.85, 1.15]$ ，以及亮度低于 0.9 时触发的低光高斯噪声 $\sigma \in [3, 10]$) 与方向性运动模糊 (随机核长 $\in \{3, 5, 7, 9, 11\}$ ，角度 $\in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$)。

如表 3 所示，采用 VPR 的纯检索方法 (CosPlace) 与时序距离方法 (ViNT) 对这些像素级扰动高度敏感，精确率与 F1 分数大幅下滑，说明仅依赖外观距离

表 3: RAVEL 的鲁棒性

方法	精确率	召回率	F1 分数	准确率
RAVEL+ 噪声	0.6703 (-5.64%)	0.7072 (-7.63%)	0.6882 (-6.56%)	0.9963 (-0.07%)
CosPlace top 2+ 噪声	0.4592 (-12.73%)	0.6615 (-7.00%)	0.5420 (-10.29%)	0.9936 (-0.12%)
CosPlace top 5+ 噪声	0.2192 (-18.78%)	0.7739 (+1.10%)	0.3416 (-8.44%)	0.9826 (-0.27%)
ViNT top 2+ 噪声	0.4197 (-19.30%)	0.6064 (-10.49%)	0.4961 (-15.70%)	0.9913 (-0.33%)
ViNT top 5+ 噪声	0.2065 (-22.36%)	0.7976 (-3.27%)	0.3281 (-18.43%)	0.9769 (-0.48%)

表 4: 全局场景与困难路径场景下全局定位的定量对比。

场景 指标	本文	MegaLoc [3]	ViNT [59]	JIST [4]
全部	节点数	976	976	976
	成功	932	889	845
	失败	44	87	131
	Acc.(%)	95.49	91.09	86.58
困难 路径	节点数	383	383	383
	成功	360	341	270
	失败	23	42	113
	Acc.(%)	93.99	89.03	70.50

在环境噪声下会迅速退化。相比之下, RAVEL 表现出显著韧性, F1 分数仅下降 6.56%。尽管噪声不可避免地减少了局部特征内点的原始数量, 但 RAVEL 的几何验证与 MSF 的全局结构约束成功阻止了受污染图像形成虚假边, 从而保持了高准确率与结构完整性。

4.4 BPL 评测

为评测定位性能, 本文在 GRScenes 中沿随机轨迹采集连续图像流以模拟机器人运动。关键在于, 这些测试图像拍摄于与拓扑图参考节点不同的新位姿。本文还引入困难路径 (Difficult Path) 子集, 其特点是路径长、转角大、起点与目标视觉重叠低。

与其他范式的对比。如表 4 所示, BPL 取得最高准确率。基线方法在复杂路径上明显退化, 而 BPL 性能保持稳定。ViNT 这类一维时序距离方法在含交叉口或回环的困难路径上失效严重。BPL 的鲁棒性源自其由图导出的转移矩阵: 置信度可在多跳邻域上自然扩散, 使概率质量在交叉口分叉、并在回环闭合时重新汇聚。

BPL 消融。表 6 分析 BPL 的内部机制。改变传播深度 (k) 可见性能相对稳定, 在 2-3 跳附近达到峰值; 过度扩散会过平滑置信度并略微增加误差。更重要的是, 禁用熵自适应融合会使准确率从 0.9549 严重下降至 0.9057。这证明结构扩散虽有帮助, 但依据当前不确定性动态加权预测与观测才是消解视觉歧义的关键因素。

对感知退化的鲁棒性。为评估 BPL 在实际执行中对真实传感器噪声与视角偏移的韧性, 本文在严重图像扰动下测试定位准确率 (表 5), 使用来自 RECON [56]、SCAND [31]、GoStanford [25] 与 SACSOn [24] 四个多样数据集的连续轨迹。实时查询图像被施加四种退化条件: 纯角度视角偏移 (Rotation), 以及旋转分别叠加高斯噪声、泊松噪声或随机结构性裁剪。

如表 5 所示, BPL 在所有退化条件下取得最高且最稳定的定位准确率 (平均 0.930 ± 0.040)。纯 VPR (MegaLoc)、序列平滑 (JIST) 与时序模型 (ViNT)

表 5: 不同方法在不同条件下的定位准确率对比。结果以四个数据集上的均值 $\pm$ 标准差报告。

条件	定位准确率			
	MegaLoc [3]	JIST [4]	ViNT [59]	BPL (本文)
旋转	0.913 $\pm$ 0.104	0.722 $\pm$ 0.066	0.895 $\pm$ 0.044	0.966 $\pm$ 0.020
旋转 + 高斯	0.743 $\pm$ 0.088	0.452 $\pm$ 0.177	0.885 $\pm$ 0.045	0.914 $\pm$ 0.038
旋转 + 泊松	0.717 $\pm$ 0.114	0.445 $\pm$ 0.176	0.888 $\pm$ 0.037	0.896 $\pm$ 0.036
旋转 + 裁剪	0.855 $\pm$ 0.151	0.525 $\pm$ 0.076	0.640 $\pm$ 0.096	0.945 $\pm$ 0.027
平均	0.807 $\pm$ 0.133	0.536 $\pm$ 0.167	0.827 $\pm$ 0.124	0.930 $\pm$ 0.040

表 6: 置信度传播定位 (BPL) 的消融实验

k	1	2	3	4	5	1
自适应融合	✓	✓	✓	✓	✓	✗
Acc. (%)	95.49	94.47	93.65	92.42	92.73	90.57

在复合像素噪声或结构性遮挡下表现出严重脆弱性，而 BPL 始终保持 0.89 以上的准确率。这一韧性验证了熵自适应融合的设计：当严重感知退化使观测似然趋于平坦时，滤波器会动态转向依赖拓扑预测，借助多跳空间先验平稳越过暂时的视觉中断而不丢失跟踪。

4.5 BASS 评测

本文在 GRScenes 的随机轨迹上评测 BASS。图 5 与图 7(a-d) 展示了闭环执行过程：在简单布局中机器人平滑跟踪子目标；在复杂拓扑中，当 BPL 检测到物理偏离超过 D_{thres} 时，系统动态重规划恢复路径。本文还将 ULVN 与最先进的端到端模型 Uni-Navid 对比 (图 7e)。Uni-Navid 纯反应式的自我中心导航存在严重的轨迹振荡，而 ULVN 在拓扑图引导下执行平滑高效的轨迹。这证实轻量级结构记忆对于缓解纯反应式方法固有的短视规划至关重要。

表 7 定量比较了本文框架与端到端基线，并消融了局部规划器的集成方式。ULVN 显著超越 Uni-Navid、UniGoal 等纯反应式模型，配合 NoMaD 取得 71.9% 的峰值成功率 (SR)，并保持明显更高的 SPL。本文还消融了局部规划器训练中的时序距离损失 (记为 $-A$)：移除该损失会严重降低性能，对基于回归的 ViNT 影响尤为明显，而生成式的 NoMaD 相对更鲁棒。这表明时序距离是一项关键的辅助训练信号。然而，在执行阶段用基于时序距离的定位替换 BPL 滤波器 ($ULVN + ViNT w. d_{temp}$) 会把 SR 降至 59.6%，说明尽管时序距离有利于局部控制的训练，但基于 VPR 的 BPL 对可靠的全局定位是必需的。

本文观察到定位准确率 ($\sim 95\%$) 与导航成功率 ($\sim 71\%$) 之间存在明显差距，主要源于局部规划器的物理局限：障碍感知有限与轨迹生成振荡，会导致机器人在拥有正确拓扑子目标的情况下偶发失败。

4.6 真机评测

为验证仿真到真实 (Sim-to-Real) 的迁移能力，本文将 ULVN 部署在配备 Azure Kinect 相机与 NVIDIA Jetson Orin 的 Diablo 轮式机器人上。图 6 展示了一次典型试验：初始规划路线需要右转，但物理扰动使机器人偏离路径。由于 BPL

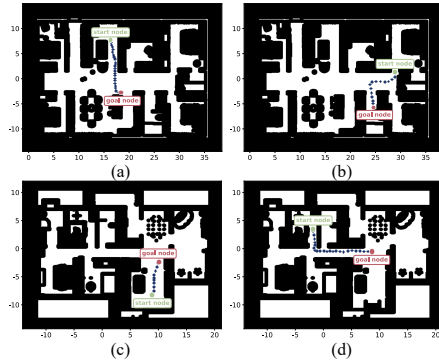


图 5: BASS 在简单任务 (a,c) 与困难任务 (b,d) 中的导航。绿色: 起点; 红色: 目标; 蓝色: 实际执行轨迹。

表 7: GRScenes 上不同算法的对比。“-A”表示仅用动作头损失函数重新训练; “w. d_{temp} ”表示用时序距离而非 BPL 进行定位。

方法	SR (%)	平均碰撞	平均 SPL
Uni-Navid [77]	32.0	1.96	0.2391
UniGoal [75]	61.6	0.88	0.3176
ULVN+VINT-A [59]	31.0	1.13	0.812
ULVN+NoMaD-A [62]	54.3	0.94	0.7458
ULVN+ViNT w. d_{temp} [59]	59.6	0.88	0.7752
ULVN+ViNT [59]	68.1	0.76	0.8398
ULVN+NoMaD [62]	71.9	0.42	0.7978

持续将观测与图进行匹配, 熵自适应滤波器把置信度集中到一个离线路节点上; 当 MAP 节点与目标子目标之间的拓扑距离超过偏离阈值后, BASS 成功触发实时重规划, 给出新的子目标并安全引导机器人到达目的地。这证实了该框架在无里程计物理部署中的实用可行性。

5 结论

本文提出 ULVN——一个完全基于无序 RGB 图像集合运行的图像目标导航框架, 无需里程计或时序先验。为消解严重的感知混淆, RAVEL 流程通过协调局部几何验证与全局结构一致性构建鲁棒拓扑骨架。在该图之上, 置信度传播定位 (BPL) 滤波器借助熵自适应融合, 在严重视觉噪声、结构性遮挡与复杂回环下维持可靠跟踪。以此状态估计为支撑, 置信度感知子目标搜索 (BASS) 实现了具备动态漂移恢复能力的一致闭环执行。大量仿真到真实的评测表明, 本文“先建图再导航”的范式对感知退化具有高度韧性, 并克服了纯反应式端到端模型固有的短视轨迹振荡。最终, ULVN 验证了轻量级拓扑记忆是在非结构化环境中实现可泛化导航的高可扩展、鲁棒方案。

致谢

本工作受国家自然科学基金 (U22A2095) 资助。

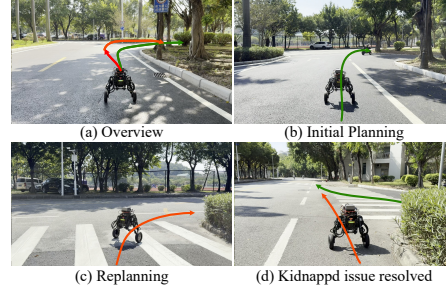


图 6: 真机案例。(a) 总览。(b) 初始规划。(c) 偏离后的重规划。(d) 通过重定位与重规划解决的绑架机器人 (Kidnapped Robot) 案例。绿色: 规划路径; 红色: 观测到的偏离; 橙色: 重规划路径。

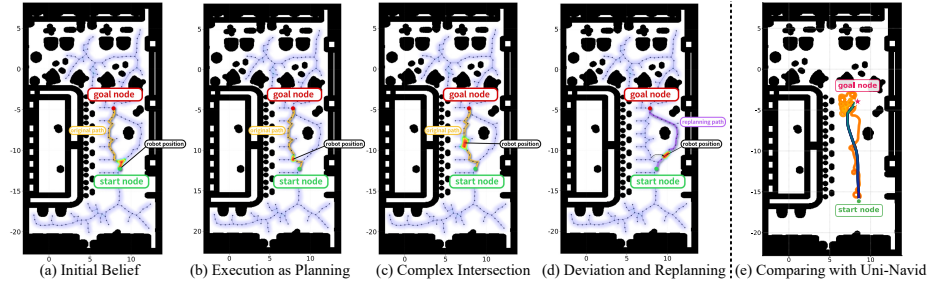


图 7: 重规划过程的可视化以及与 Uni-Navid 的对比。

参考文献

1. Anderson, P., Chang, A., Chaplot, D.S., Dosovitskiy, A., Gupta, S., Koltun, V., Kosecka, J., Malik, J., Mottaghi, R., Savva, M., et al.: On evaluation of embodied navigation agents. arXiv preprint arXiv:1807.06757 (2018)
2. Barath, D., Mishkin, D., Cavalli, L., Sarlin, P.E., Hruby, P., Pollefeys, M.: Stereoglue: Robust estimation with single-point solvers. In: European Conference on Computer Vision. pp. 421–441. Springer (2024)
3. Berton, G., Masone, C.: Megaloc: One retrieval to place them all. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 2861–2867 (2025)
4. Berton, G., Trivigno, G., Caputo, B., Masone, C.: Jist: Joint image and sequence training for sequential visual place recognition. IEEE Robotics and Automation Letters **9**(2), 1310–1317 (2023)
5. Blochliger, F., Fehr, M., Dymczyk, M., Schneider, T., Siegwart, R.: Topomap: Topological mapping and navigation based on visual slam maps. In: 2018 IEEE International Conference on Robotics and Automation (ICRA). pp. 3818–3825. IEEE (2018)
6. Bonin-Font, F., Ortiz, A., Oliver, G.: Visual navigation for mobile robots: A survey. Journal of intelligent and robotic systems **53**, 263–296 (2008)
7. Booij, O., Terwijn, B., Zivkovic, Z., Krose, B.: Navigation using an appearance based topological map. In: Proceedings 2007 IEEE International Conference on Robotics and Automation. pp. 3927–3932. IEEE (2007)
8. Cadena, C., Carlone, L., Carrillo, H., Latif, Y., Scaramuzza, D., Neira, J., Reid, I., Leonard, J.J.: Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age. IEEE Transactions on robotics **32**(6), 1309–1332 (2016)
9. Claxton, O., Malone, C., Carson, H., Ford, J.J., Bolton, G., Shames, I., Milford, M.: Improving visual place recognition based robot navigation by verifying localization estimates. IEEE Robotics and Automation Letters (2024)
10. Cui, X., Liu, Q., Liu, Z., Wang, H.: Frontier-enhanced topological memory with improved exploration awareness for embodied visual navigation. In: European Conference on Computer Vision. pp. 296–313. Springer (2024)
11. Cummins, M., Newman, P.: Fab-map: Probabilistic localization and mapping in the space of appearance. The International journal of robotics research **27**(6), 647–665 (2008)
12. Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., Koltun, V.: Carla: An open urban driving simulator. In: Conference on robot learning. pp. 1–16. PMLR (2017)

13. Douze, M., Guzhva, A., Deng, C., Johnson, J., Szilvasy, G., Mazaré, P.E., Lomeli, M., Hosseini, L., Jégou, H.: The faiss library. *IEEE Transactions on Big Data* (2025)
14. Edstedt, J.: Less biased noise scale estimation for threshold-robust ransac. In: *Proceedings of the Computer Vision and Pattern Recognition Conference*. pp. 2811–2820 (2025)
15. Fan, S., Liu, R., Wang, W., Yang, Y.: Navigation instruction generation with bev perception and large language models. In: *European Conference on Computer Vision*. pp. 368–387. Springer (2024)
16. Furgale, P., Barfoot, T.D.: Visual teach and repeat for long-range rover autonomy. *Journal of field robotics* **27**(5), 534–560 (2010)
17. Gao, C., Liu, S., Chen, J., Wang, L., Wu, Q., Li, B., Tian, Q.: Room-object entity prompting and reasoning for embodied referring expression. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **46**(2), 994–1010 (2023)
18. Garg, S., Fischer, T., Milford, M.: Where is your place, visual place recognition? In: *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI-21)*. pp. 4416–4425. International Joint Conferences on Artificial Intelligence (2021)
19. Garg, S., Rana, K., Hosseinzadeh, M., Mares, L., Sünderhauf, N., Dayoub, F., Reid, I.: Robohop: Segment-based topological map representation for open-world visual navigation. In: *2024 IEEE International Conference on Robotics and Automation (ICRA)*. pp. 4090–4097. IEEE (2024)
20. Gode, S., et al.: Flownav: Learning efficient navigation policies via conditional flow matching. *arXiv* (2024)
21. Gornet, J., Thomson, M.: Automated construction of cognitive maps with visual predictive coding. *Nature Machine Intelligence* **6**(7), 820–833 (2024)
22. Guo, Z., Hall, R.W.: Parallel thinning with two-subiteration algorithms. *Communications of the ACM* **32**(3), 359–373 (1989)
23. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 770–778 (2016)
24. Hirose, N., Shah, D., Sridhar, A., Levine, S.: Sacson: Scalable autonomous control for social navigation. *IEEE Robotics and Automation Letters* (2023)
25. Hirose, N., Xia, F., Martín-Martín, R., Sadeghian, A., Savarese, S.: Deep visual mpc-policy learning for navigation. *IEEE Robotics and Automation Letters* **4**(4), 3184–3191 (2019)
26. Hossain, J., Faridee, A.Z., Roy, N., Freeman, J., Gregory, T., Trout, T.: Toponav: Topological navigation for efficient exploration in sparse reward environments. In: *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. pp. 693–700. IEEE (2024)
27. Irschara, A., Zach, C., Frahm, J.M., Bischof, H.: From structure-from-motion point clouds to fast location recognition. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. pp. 2599–2606. IEEE (2009)
28. Jiang, Y., Liu, Q., Yang, Y., Ma, X., Zhong, D., Hu, H., Yang, J., Liang, B., XU, B., Zhang, C., et al.: Episodic novelty through temporal distance. In: *The Thirteenth International Conference on Learning Representations* (2025)
29. Johnson, J., Douze, M., Jégou, H.: Billion-scale similarity search with gpus. *IEEE Transactions on Big Data* **7**(3), 535–547 (2019)
30. Kamila, S., Hasanuzzaman, M., Ekbal, A., Bhattacharyya, P.: Measuring temporal distance focus from tweets and investigating its association with psycho-

- demographic attributes. *IEEE Transactions on Affective Computing* **13**(2), 1086–1097 (2020)
31. Karnan, H., Nair, A., Xiao, X., Warnell, G., Pirk, S., Toshev, A., Hart, J., Biswas, J., Stone, P.: Socially compliant navigation dataset (scand): A large-scale dataset of demonstrations for social navigation. *IEEE Robotics and Automation Letters* (2022)
 32. Li, H., Peng, G., Zhang, J., Wen, M., Ma, Y., Wang, D.: Casevpr: Correlation-aware sequential embedding for sequence-to-frame visual place recognition. *IEEE Robotics and Automation Letters* (2025)
 33. Li, Z., Zheng, K., Yuan, Y., Huang, J., Zhang, X., Wu, J., Cheng, H.: Learning to explore efficiently: Heterogeneous topological graphs and lightweight global reasoning for robotic exploration. *IEEE Robotics and Automation Letters* (2025)
 34. Lindenberger, P., Sarlin, P.E., Pollefeys, M.: Lightglue: Local feature matching at light speed. In: *Proceedings of the IEEE/CVF international conference on computer vision*. pp. 17627–17638 (2023)
 35. Liu, P., Orru, Y., Paxton, C., Shafiullah, N.M.M., Pinto, L.: Ok-robot: What really matters in integrating open-knowledge models for robotics. *Proceedings of Robotics: Science and Systems (RSS)* (2024)
 36. Ma, Y.J., Sodhani, S., Jayaraman, D., Bastani, O., Kumar, V., Zhang, A.: Vip: Towards universal visual reward and representation via value-implicit pre-training. *arXiv preprint arXiv:2210.00030* (2022)
 37. Majumdar, A., Aggarwal, G., Devnani, B., Hoffman, J., Batra, D.: Zson: Zero-shot object-goal navigation using multimodal goal embeddings. *Advances in Neural Information Processing Systems* **35**, 32340–32352 (2022)
 38. Miao, J., Jiang, K., Wen, T., Wang, Y., Jia, P., Wijaya, B., Zhao, X., Cheng, Q., Xiao, Z., Huang, J., Zhong, Z., Yang, D.: A survey on monocular re-localization: From the perspective of scene map representation. *IEEE Transactions on Intelligent Vehicles* **10**(4), 2519–2550 (2025)
 39. Milford, M.J., Wyeth, G.F.: Seqslam: Visual route-based navigation for sunny summer days and stormy winter nights. In: *2012 IEEE international conference on robotics and automation*. pp. 1643–1649. IEEE (2012)
 40. Montano-Oliván, L., Placed, J.A., Montano, L., Lázaro, M.T.: G-loc: Tightly-coupled graph localization with prior topo-metric information. *IEEE Robotics and Automation Letters* (2024)
 41. Muravyev, K., Melekhin, A., Yudin, D., Yakovlev, K.: Prism-topomap: online topological mapping with place recognition and scan matching. *IEEE Robotics and Automation Letters* (2025)
 42. Myers, V., Zheng, C., Dragan, A., Levine, S., Eysenbach, B.: Learning temporal distances: Contrastive successor features can provide a metric structure for decision-making. In: *International Conference on Machine Learning*. pp. 37076–37096. PMLR (2024)
 43. Neubert, P., Schubert, S., Protzel, P.: A neurologically inspired sequence processing model for mobile robot place recognition. *IEEE Robotics and Automation Letters* **4**(4), 3200–3207 (2019)
 44. Noël, T., Lehuger, A., Marchand, E., Chaumette, F.: Skeleton disk-graph roadmap: A sparse deterministic roadmap for safe 2d navigation and exploration. *IEEE Robotics and Automation Letters* **9**(1), 555–562 (2023)
 45. Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., et al.: Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193* (2023)

46. Piedade, V., Miraldo, P.: Bansac: A dynamic bayesian network for adaptive sample consensus. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. pp. 3738–3747 (2023)
47. Ren, H., Bi, Z., Zeng, Y., Wan, Z., Qi, L., Cheng, H.: Strnet: Visual navigation with spatio-temporal representation through dynamic graph aggregation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 42464–42473 (2026)
48. Ren, H., Bi, Z., Zeng, Y., Zheng, L., Li, Z., Wan, Z., Qi, L., Cheng, H.: Fisher-preserving guidance: Training-free manifold constraints for safe diffusion control. *arXiv preprint arXiv:2605.29937* (2026)
49. Ren, H., Zeng, Y., Bi, Z., Wan, Z., Huang, J., Cheng, H.: Prior does matter: Visual navigation via denoising diffusion bridge models. In: *Proceedings of the Computer Vision and Pattern Recognition Conference*. pp. 12100–12110 (2025)
50. Roth, P., Nubert, J., Yang, F., Mittal, M., Hutter, M.: Viplanner: Visual semantic imperative learning for local navigation. In: *2024 IEEE International Conference on Robotics and Automation (ICRA)*. pp. 5243–5249. IEEE (2024)
51. Sattler, T., Leibe, B., Kobbelt, L.: Efficient & effective prioritized matching for large-scale image-based localization. *IEEE transactions on pattern analysis and machine intelligence* **39**(9), 1744–1756 (2016)
52. Savinov, N., Dosovitskiy, A., Koltun, V.: Semi-parametric topological memory for navigation. In: *International Conference on Learning Representations* (2018)
53. Savva, M., Kadian, A., Maksymets, O., Zhao, Y., Wijmans, E., Jain, B., Straub, J., Liu, J., Koltun, V., Malik, J., et al.: Habitat: A platform for embodied ai research. In: *Proceedings of the IEEE/CVF international conference on computer vision*. pp. 9339–9347 (2019)
54. Schonberger, J.L., Frahm, J.M.: Structure-from-motion revisited. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 4104–4113 (2016)
55. Schönberger, J.L., Zheng, E., Frahm, J.M., Pollefeys, M.: Pixelwise view selection for unstructured multi-view stereo. In: *European conference on computer vision*. pp. 501–518. Springer (2016)
56. Shah, D., Eysenbach, B., Kahn, G., Rhinehart, N., Levine, S.: Rapid exploration for open-world navigation with latent goal models. *arXiv preprint arXiv:2104.05859* (2021)
57. Shah, D., Eysenbach, B., Kahn, G., Rhinehart, N., Levine, S.: Ving: Learning open-world navigation with visual goals. In: *2021 IEEE International Conference on Robotics and Automation (ICRA)*. pp. 13215–13222. IEEE (2021)
58. Shah, D., Sridhar, A., Bhorkar, A., Hirose, N., Levine, S.: Gnm: A general navigation model to drive any robot. In: *2023 IEEE International Conference on Robotics and Automation (ICRA)*. pp. 7226–7233. IEEE (2023)
59. Shah, D., Sridhar, A., Dashora, N., Stachowicz, K., Black, K., Hirose, N., Levine, S.: Vint: A foundation model for visual navigation. *arXiv preprint arXiv:2306.14846* (2023)
60. Shi, P., Yan, S., Xiao, Y., Liu, X., Zhang, Y., Li, J.: Ransac back to sota: A two-stage consensus filtering for real-time 3d registration. *IEEE Robotics and Automation Letters* **9**(12), 11881–11888 (2024)
61. Song, M., Zhang, D., Ren, H., Zhang, R., Du, B., Yang, M.H., Qi, L.: Unisharp: Universal sharp monocular view synthesis. *arXiv preprint arXiv:2606.07514* (2026)
62. Sridhar, A., Shah, D., Glossop, C., Levine, S.: Nomad: Goal masked diffusion policies for navigation and exploration. In: *2024 IEEE International Conference on Robotics and Automation (ICRA)*. pp. 63–70. IEEE (2024)

63. Sun, X., Liu, L., Zhi, H., Qiu, R., Liang, J.: Prioritized semantic learning for zero-shot instance navigation. In: European Conference on Computer Vision. pp. 161–178. Springer (2024)
64. Suomela, L., Arachchige, S.K., Torres, G.F., Edelman, H., Kämäräinen, J.K.: Synthetic vs. real training data for visual navigation. arXiv preprint arXiv:2509.11791 (2025)
65. Suomela, L., Kalliola, J., Edelman, H., Kämäräinen, J.K.: Placenv: Topological navigation through place recognition. In: 2024 IEEE International Conference on Robotics and Automation (ICRA). pp. 5205–5213. IEEE (2024)
66. Thoma, J., Paudel, D.P., Chhatkuli, A., Probst, T., Gool, L.V.: Mapping, localization and path planning for image-based navigation using visual features and map. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7383–7391 (2019)
67. Thrun, S.: Learning metric-topological maps for indoor mobile robot navigation. *Artificial intelligence* **99**(1), 21–71 (1998)
68. Thrun, S.: Probabilistic robotics. *Communications of the ACM* **45**(3), 52–57 (2002)
69. Wan, Z., Bi, Z., Zhou, Z., Ren, H., Zeng, Y., Li, Y., Qi, L., Yang, X., Yang, M.H., Cheng, H.: Rapid hand: A robust, affordable, perception-integrated, dexterous manipulation platform for generalist robot autonomy. arXiv preprint arXiv:2506.07490 (2025)
70. Wang, H., Chen, J., Huang, W., Ben, Q., Wang, T., Mi, B., Huang, T., Zhao, S., Chen, Y., Yang, S., et al.: Grutopia: Dream general robots in a city at scale. arXiv preprint arXiv:2407.10943 (2024)
71. Wang, J., Chen, M., Karaev, N., Vedaldi, A., Rupprecht, C., Novotny, D.: Vggt: Visual geometry grounded transformer. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 5294–5306 (2025)
72. Wei, T., Patel, Y., Shekhovtsov, A., Matas, J., Barath, D.: Generalized differentiable ransac. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 17649–17660 (2023)
73. Xu, M., Fischer, T., Sünderhauf, N., Milford, M.: Probabilistic appearance-invariant topometric localization with new place awareness. *IEEE Robotics and Automation Letters* **6**(4), 6985–6992 (2021)
74. Yasuda, Y.D., Martins, L.E.G., Cappabianco, F.A.: Autonomous visual navigation for mobile robots: A systematic literature review. *ACM Computing Surveys (CSUR)* **53**(1), 1–34 (2020)
75. Yin, H., Xu, X., Zhao, L., Wang, Z., Zhou, J., Lu, J.: Unigoal: Towards universal zero-shot goal-oriented navigation. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 19057–19066 (2025)
76. Zeng, Y., Ren, H., Wang, S., Huang, J., Cheng, H.: Navidiffusor: Cost-guided diffusion model for visual navigation. arXiv preprint (2025)
77. Zhang, J., Wang, K., Wang, S., Li, M., Liu, H., Wei, S., Wang, Z., Zhang, Z., Wang, H.: Uni-navid: A video-based vision-language-action model for unifying embodied navigation tasks. arXiv preprint arXiv:2412.06224 (2024)
78. Zhang, T., Hu, X., Xiao, J., Zhang, G.: A survey of visual navigation: From geometry to embodied ai. *Engineering Applications of Artificial Intelligence* **114**, 105036 (2022)
79. Zhang, T.Y., Suen, C.Y.: A fast parallel algorithm for thinning digital patterns. *Communications of the ACM* **27**(3), 236–239 (1984)
80. Zhang, W., Hara, Y., Nakamura, S.: Topological mapping and navigation using a monocular camera based on anyloc. In: 2025 IEEE 21st International Conference on Automation Science and Engineering (CASE). pp. 3077–3083. IEEE (2025)

81. Zhang, X., Wang, L., Su, Y.: Visual place recognition: A survey from deep learning perspective. *Pattern Recognition* **113**, 107760 (2021)
82. Zhao, J., Zhang, F., Cai, Y., Tian, G., Mu, W., Ye, C., Feng, T.: Learning sequence descriptor based on spatio-temporal attention for visual place recognition. *IEEE Robotics and Automation Letters* **9**(3), 2351–2358 (2024)
83. Zhu, Y., Mottaghi, R., Kolve, E., Lim, J.J., Gupta, A., Fei-Fei, L., Farhadi, A.: Target-driven visual navigation in indoor scenes using deep reinforcement learning. In: 2017 IEEE international conference on robotics and automation (ICRA). pp. 3357–3364. IEEE (2017)

A 附录概览

本补充材料提供全面的实现细节、数学推导与扩展实验分析，以支撑正文内容，组织如下：

- **第 B 节：数据采集细节**。详细说明用于生成真值拓扑图以及从仿真环境获取无序图像数据集的流程。
- **第 C 节：RAVEL 扩展算法细节**。深入解释地标的鲁棒增强与验证（Robust Augmentation and VERification of Landmarks, RAVEL）框架，包括数据驱动的阈值标定与结构剪枝机制。
- **第 D 节：置信度传播定位（BPL）推导**。给出定位模块背后的数学形式化，详述基于图的转移模型与基于熵的自适应融合策略。
- **第 E 节：置信度感知子目标搜索（BASS）细节**。描述本文导航策略的完整工作流程，重点是动态子目标选择与偏离触发的重规划逻辑。
- **第 F 节：扩展实验结果**。给出图构建与定位鲁棒性的补充定量消融实验，以及复杂场景中导航轨迹的定性可视化。

B 数据采集细节

本节对所采集数据集以及用于构建 ULVN 无序地标库与相应真值拓扑结构的数据采集流程给出更详尽的描述。正文以简洁方式介绍了各步骤，这里则希望更清晰地阐明每一次变换为何必要、中间表示如何构建，以及它们最终如何支撑建图、定位与导航的可靠评测。整条流程把原始仿真场景转换为稀疏骨架图与一组相机观测，二者共同构成参考拓扑。

B.1 所采集数据集信息

数据集构建采用基于环境骨架结构的空间离散化策略。采样节点由骨架地图系统性生成，并存储为带朝向的位姿向量 $[x, y, \text{yaw}]$ ，据此引导图像采集过程。数据采集覆盖十个不同环境，均分为五个家居（Home）与五个商业（Commercial）场景，以确保环境多样性。具体而言，住宅场景 Home 0、1、2、3、4 的图像数量分别为 540、226、332、216 与 342 帧；商业场景 Commercial 0、1、2、3、4 分别贡献 474、398、298、388 与 380 帧。数据集共包含 3,594 张分辨率为 1920×1080 像素的高分辨率图像，为拓扑图构建提供稠密且高质量的视觉表示。

B.2 数据采集流程细节

本文流程大致包含以下阶段：提取三维占据栅格、压缩为二维可通行地图、计算 ESDF、检测中轴区域、生成细化骨架、采样拓扑节点、构建连通性图，最后在节点位置采集 RGB 图像。下面依次描述这些阶段。

三维占据提取。首先从仿真器导出体素化的三维占据栅格，其中每个体素记为

$$O(x, y, z) \in \{0, 1\}, \quad (8)$$

被占据的体素对应场景中任意静态物体或结构元素。由于机器人在地平面上运动，且其传感器只能感知有限的垂直范围，本文仅关注高度切片

$$z \in [z_{\min}, z_{\max}], \quad (9)$$

取值由机器人设计确定：

$$z_{\min} = 0.2 \text{ m}, \quad z_{\max} = 1.2 \text{ m}. \quad (10)$$

体素分辨率（如 0.05 m）决定该表示的保真度。此三维占据地图是所有下游拓扑结构的几何基础。

二维可通行性投影。为得到与平面导航兼容的表示，本文把所选高度切片压平为二维栅格：

$$G(x, y) = \max_{z \in [z_{\min}, z_{\max}]} O(x, y, z). \quad (11)$$

只要垂直区间内任一体素被占据，该单元即视为阻塞，从而对悬挑障碍作保守处理。所得二维离散地图分辨率为 0.05 m。该表示编码了地板平面上每个位置是否可通行，并成为距离计算与骨架提取的基础。

ESDF 计算。为以连续且利于导航的方式刻画自由空间几何，本文在障碍地图上计算欧氏符号距离场（ESDF）。对每个单元：

$$\text{ESDF}(x, y) = \begin{cases} +d(x, y), & G(x, y) = 0, \\ -d(x, y), & G(x, y) = 1, \end{cases} \quad (12)$$

其中 $d(x, y)$ 是到最近障碍单元的欧氏距离。该计算使用欧氏距离变换算法完成（由 `scipy.ndimage` 实现）。ESDF 不仅支持鲁棒的自由空间查询，还能揭示指导中轴曲线提取的几何结构。

面向中轴检测的梯度与通量。为识别位于走廊或开阔区域中心附近的点，本文计算 ESDF 梯度：

$$\nabla \text{ESDF}(x, y) = \left(\frac{\partial \text{ESDF}}{\partial x}, \frac{\partial \text{ESDF}}{\partial y} \right), \quad (13)$$

使用有限差分以保证数值稳定。随后计算梯度通量

$$F(x, y) = -\nabla \cdot \left(\frac{\nabla \text{ESDF}(x, y)}{\|\nabla \text{ESDF}(x, y)\|} \right), \quad (14)$$

它会突出距离梯度汇聚的位置——这是拓扑“中心线”的特征。施加阈值

$$F(x, y) > \tau_{\text{flux}}, \quad \tau_{\text{flux}} = -0.01, \quad (15)$$

即得到粗略的中轴掩码，大致刻画可通行空间的连通性。

骨架提取与细化。原始中轴掩码往往偏厚且含噪。为得到干净的单像素宽拓扑骨架，本文施加若干精化步骤：连通域过滤、形态学细化（如 Zhang-Suen [79] 或 Guo-Hall [22]），以及无意义分支的剪枝。具体地，任何短于

$$L_{\min} = 15 \text{ px} \quad (16)$$

算法 1 数据采集流程：骨架提取**Require:** 三维仿真环境，高度阈值 $[z_{\min}, z_{\max}]$ **Ensure:** 骨架地图 $\mathcal{S}_{\text{GT}}$

```

1: 阶段 1：环境体素化
2: 初始化三维占据栅格  $\mathcal{O}$ 
3: for 环境中每个体素  $v = (x, y, z)$  do
4:   若为障碍则  $\mathcal{O}[x, y, z] \leftarrow 1$ ，否则为 0
5: 阶段 2：二维投影
6: 初始化二维栅格  $\mathcal{G}_{2D}$ 
7: for 每个栅格单元  $(x, y)$  do
8:    $\mathcal{G}_{2D}[x, y] \leftarrow \max_{z \in [z_{\min}, z_{\max}]} \mathcal{O}[x, y, z]$  ▷ 压平三维地图
9: 阶段 3：ESDF 计算
10: 为  $\mathcal{G}_{2D}$  计算欧氏距离场  $d(x, y)$ 
11: for 每个栅格单元  $(x, y)$  do
12:   if  $\mathcal{G}_{2D}[x, y] = 1$  then
13:      $\text{ESDF}[x, y] \leftarrow -d(x, y)$  ▷ 障碍内部取负
14:   else
15:      $\text{ESDF}[x, y] \leftarrow d(x, y)$  ▷ 自由空间取正
16: 阶段 4：中轴检测
17: for ESDF 中每个栅格单元  $(x, y)$  do
18:   计算梯度  $\mathbf{g} \leftarrow \nabla \text{ESDF}[x, y]$ 
19:   基于  $\mathbf{g}$  的散度计算通量  $F(x, y)$ 
20:   if  $F(x, y) > \tau_{\text{flux}}$  then
21:      $\mathcal{S}_{\text{raw}}[x, y] \leftarrow 1$  ▷ 标记为候选骨架
22: 阶段 5：剪枝与细化
23:  $\mathcal{S} \leftarrow \text{MorphologicalThinning}(\mathcal{S}_{\text{raw}})$  ▷ 细化至单像素宽
24: for 每条分支  $b \in \mathcal{S}$  do
25:   if  $\text{Length}(b) < L_{\min}$  then
26:     从  $\mathcal{S}$  中移除分支  $b$  ▷ 剪除短伪影
27: return  $\mathcal{S}$ 

```

的分支都会被移除，以避免产生虚假节点和无法有效反映环境结构的微小分支。细化后的骨架是一个紧凑的曲线网络，表示可通行区域的核心拓扑。

拓扑节点采样。为把骨架转换为图表示，本文沿其曲线采样节点。以步长

$$s = 0.5 \text{ m} \quad (17)$$

均匀采样可提供一致的空间覆盖。由于细化仍可能产生局部聚簇，本文施加半径为

$$r_{\text{NMS}} = 0.5 \text{ m} \quad (18)$$

的非极大值抑制（NMS），确保节点分布良好且不冗余。在交叉口（骨架上度数 ≥ 3 的点）附近提高采样密度，因为这些位置对表示环境的分支结构至关重要。

图连通性恢复。确定节点位置 $\{v_i\}$ 后，通过检查沿骨架的局部连通性推断边。若两个节点位于同一骨架分支上且测地间距小于阈值，则将其相连：

$$d(v_i, v_j) < d_{\max}, \quad d_{\max} = 1 \text{ m}. \quad (19)$$

算法 2 数据采集流程：图像采集与节点采样

Require: 骨架地图 $\mathcal{S}$, 节点间距 s , 最大边距离 $d_{\max}$
Ensure: 拓扑图 $\mathcal{G}_{\text{GT}}$, 图像数据集 $\mathcal{I}$

- 1: **阶段 1: 骨架上的节点采样**
- 2: $V_{\text{sampled}} \leftarrow \emptyset$
- 3: **for** $\mathcal{S}$ 中每段路径 **do**
- 4: 以固定间隔 s 采样节点
- 5: $V_{\text{sampled}} \leftarrow V_{\text{sampled}} \cup \{\text{采样得到的节点}\}$
- 6: $V \leftarrow \text{NonMaxSuppression}(V_{\text{sampled}}, r_{\text{nms}})$ ▷ 移除重复节点
- 7: **阶段 2: 图连通性**
- 8: $E \leftarrow \emptyset$
- 9: **for** 每一对 $v_i, v_j \in V$ **do**
- 10: **if** $\text{dist}(v_i, v_j) < d_{\max}$ **and** $\text{LineOfSight}(v_i, v_j)$ **then**
- 11: $E \leftarrow E \cup \{(v_i, v_j)\}$ ▷ 连接可通行节点
- 12: $\mathcal{G}_{\text{GT}} \leftarrow (V, E)$
- 13: **阶段 3: 图像采集与朝向**
- 14: 初始化数据集 $\mathcal{I} \leftarrow \emptyset$
- 15: **for** 每个节点 $v_i \in V$ **do**
- 16: // 通过 PCA 或梯度确定朝向
- 17: $\mathbf{g}_i \leftarrow \nabla \text{ESDF}(v_i)$
- 18: $\theta_i \leftarrow \text{ComputePrincipalDirection}(\mathbf{g}_i)$
- 19: 将智能体移动到位置 v_i 并设朝向为 θ_i
- 20: 采集图像 I_i
- 21: 将 (I_i, v_i, θ_i) 加入 $\mathcal{I}$
- 22: **return** 带位姿的图像集合 $\mathcal{I}$

该过程得到真值拓扑图

$$\mathcal{G}_{\text{GT}} = (V, E), \quad (20)$$

其中编码了核心导航结构。边可选地存储几何长度或评测时使用的其他元数据。

RGB 图像采集与相机朝向。为把拓扑位置与视觉观测配对，本文将机器人虚拟相机置于每个节点 v_i 处采集 RGB 图像。除节点位置外，流程还利用 ESDF 及其中轴骨架的局部几何，在每个节点计算一组规范观察朝向。

对骨架上每个采样节点位置 $p_i = (x_i, y_i)$ ，先将其映射回底层 ESDF 栅格并查询局部 ESDF 梯度

$$\nabla \text{ESDF}(x_i, y_i) = (g_x(x_i, y_i), g_y(x_i, y_i)). \quad (21)$$

梯度方向

$$\theta_{\text{grad}} = \text{atan2}(g_y(x_i, y_i), g_x(x_i, y_i)) \quad (22)$$

由骨架中心线指向邻近障碍或自由空间边界。为使相机与走廊方向对齐，本文将局部切向朝向定义为该梯度旋转 90° ：

$$\theta_{\text{tan}} = \theta_{\text{grad}} + \frac{\pi}{2}, \quad (23)$$

它沿中轴方向，因而近似为行进方向。

围绕该切向方向，通过叠加固定角度偏移生成一组离散航向：

$$\theta_{i,k} = \theta_{\text{tan}} + k \Delta\theta, \quad k = 0, \dots, K-1, \quad (24)$$

其中 $\Delta\theta$ 为角度间隔（例如 $\Delta\theta = 60^\circ$ ，则 $K = 6$ 个视角在 360° 上均匀分布）。实际实现中，位置及其对应航向存储为

$$[x_i, y_i, \theta_{i,0}, \theta_{i,1}, \dots, \theta_{i,K-1}], \quad (25)$$

仿真器把相机置于 (x_i, y_i) 并依次偏转到各 $\theta_{i,k}$ ，以在每个节点渲染多个视角。相机模型、图像分辨率 1920×1080 以及视场 $121.04^\circ \times 35.20^\circ$ （水平 $\times$ 垂直）均沿用实验所用的机器人配置。最终得到的无序图像集合

$$\mathcal{L} = \{I_{i,k}\}_{i=1,\dots,N}^{k=0,\dots,K-1} \quad (26)$$

构成 ULVN 唯一的视觉输入，用于评测建图与定位组件。实际实现中，每个节点采样 2 张图像。

小结。总体而言，该流程把原始仿真数据转换为具有拓扑意义的图和相应的相机观测集合。需要强调的是，这些真值结构仅用于评测：ULVN 本身只接收无序 RGB 图像，不含度量先验、里程计或时序信息。通过标准化拓扑生成过程，本文数据集确保了跨多样环境的公平比较，并为评估拓扑建图与导航系统提供了可靠基础。所有参数在全部环境中保持固定，并按机器人足迹/走廊宽度尺度选取；第 F 节报告了对关键图构建超参数的鲁棒性。

C RAVEL：扩展算法细节

本节对 RAVEL (**R**obust **A**ugmentation and **V**erification of **L**andmarks，地标的鲁棒增强与验证) 流程给出扩展说明。正文在概念层面介绍了 RAVEL，这里进一步展开其动机、设计原则与算法步骤，旨在阐明 RAVEL 如何把一组含噪的无序 RGB 图像转换为结构可靠的视觉拓扑图。

在没有时序或度量先验的条件下仅凭 RGB 图像构建图本身极具挑战。直接基于外观的检索在视觉相似但空间相距很远的位置发生匹配时，常引入误检；反之，过严的几何阈值可能剔除有效边，使拓扑碎片化。RAVEL 通过把验证过程标定到环境自身统计特性，再借助有原则的剪枝与回环重插入施加全局结构一致性，来应对这些问题。以下各小节描述这些组件。

全局检索与候选生成。给定无序图像库 $L = \{I_1, \dots, I_N\}$ ，先将每张图像映射为全局描述符：

$$z_i = f(I_i), \quad z_i \in \mathbb{R}^D. \quad (27)$$

描述符经 L2 归一化并用 FAISS 建立索引。其作用并非对图连通性作最终判定，而是提供一种高效方式来筛出可能相关的图像。对每张图像，检索描述符距离在阈值 d_{VPR} 内的所有邻居，该阈值将在标定阶段自动确定。这种检索仅在描述符相似度支持的范围内扩展搜索空间，从而减少昂贵几何验证的次数。

数据驱动标定的动机。几何验证的性能关键取决于两个量：判定两图是否具有几何有效关系的内点阈值 τ ，以及候选检索的描述符截断 d_{VPR} 。在经典流程中，

这些阈值是固定常数，必须针对每个环境手工调参。然而光照、视角、纹理分布与场景布局在不同环境间差异极大，使通用固定阈值难以稳定。

因此 RAVEL 通过探测场景自身的匹配统计特性执行一次性标定（One-Shot Calibration），得到能自动适应环境难度与纹理丰富度的阈值。

基于锚点的探测与分数采集。为估计有意义的阈值，本文选取两个锚点图像：基准锚点 q_0 及其描述符最远邻居

$$q_f = \arg \max_j \|z_{q_0} - z_j\|_2. \quad (28)$$

这两个锚点共同覆盖环境中视觉相似与视觉差异较大的区域。对每个锚点，与 L 中所有图像穷举匹配，得到成对测量：

$$(c_{qj}, d_{qj}) = (\text{通过 LightGlue+RANSAC 得到的内点数}, \|z_q - z_j\|_2). \quad (29)$$

这些样本反映了场景自身的匹配质量分布。

基于聚类的阈值导出。RAVEL 汇集两个锚点的内点数并执行二均值聚类。高置信簇 $\mathcal{S}_h$ 代表强几何匹配， $\mathcal{S}_l$ 则包含较弱或虚假的匹配。内点阈值定义为最低强匹配与最高弱匹配的中点：

$$\tau = \frac{1}{2} (\min(\mathcal{S}_h) + \max(\mathcal{S}_l)). \quad (30)$$

类似地，描述符阈值取强匹配中的最大描述符距离：

$$d_{\text{VPR}} = \max_{(q,j) \in \mathcal{S}_h} d_{qj}. \quad (31)$$

该标定技术无需任何环境专属超参数调优，即可可靠地把有意义的关系与噪声分开。

几何验证与图装配。对经标定描述符阈值检索出的每个候选对 (i, j) 执行几何验证。LightGlue 提取局部特征并产生对应关系，随后进行基于 RANSAC 的检验（如同伦或本质矩阵一致性）。所得内点数 c_{ij} 决定是否形成边：

$$W_{ij} = W_{ji} = \begin{cases} c_{ij}, & c_{ij} > \tau, \\ 0, & \text{其他}. \end{cases} \quad (32)$$

这会得到一个初始稠密图，其中可能含有环、捷径与偶发误检——当几何信号含噪或存在重复纹理时，这是典型结果。

结构剪枝的必要性。仅靠局部验证无法保证全局结构一致。例如纹理丰富的区域可能产生大量“强而虚假”的连接，形成不对应真实空间可通行性的环；反之，过严的局部裁剪可能遗漏对维持连通性至关重要的边。为化解这些矛盾，RAVEL 使用最大生成森林（MSF）构建全局结构骨干。

最大生成森林（MSF）。本文用 Kruskal 算法计算 MSF。MSF 具有以下性质：

- 保留最强的已验证边，

算法 3 RAVEL 图构建

Require: 无序图像集合 $\mathcal{I} = \{I_1, \dots, I_N\}$
Ensure: 拓扑图 $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$

- 1: **阶段 1: 初始化与特征提取**
- 2: 计算全局描述符 $\mathcal{Z} = \{z_i \mid z_i = f(I_i)\}$
- 3: 用 $\mathcal{Z}$ 建立 FAISS 索引以实现高效检索
- 4: **阶段 2: 参数估计**
- 5: 选取锚点 q_0 (随机) 与 q_f (q_0 的最远邻居)
- 6: **for** $q \in \{q_0, q_f\}$ **do**
- 7: 对所有 j 计算几何内点 c_{qj} ▷ 使用 LightGlue + RANSAC
- 8: 记录描述符距离 d_{qj}
- 9: 将内点统计量聚类为集合 $\mathcal{S}_{high}$ 与 $\mathcal{S}_{low}$
- 10: 导出验证阈值 τ 与检索半径 d_{VPR}
- 11: **阶段 3: 稠密图构建**
- 12: 初始化邻接矩阵 $W \leftarrow \mathbf{0}_{N \times N}$
- 13: **for** 每张图像 $i \in \{1, \dots, N\}$ **do**
- 14: Retrieve candidates $\mathcal{C}_i \leftarrow \{j \mid \|z_i - z_j\| \leq d_{VPR}\}$ ▷ FAISS 半径搜索
- 15: **for** 每个 $j \in \mathcal{C}_i$ **do**
- 16: $c_{ij} \leftarrow \text{GeometricVerify}(I_i, I_j)$
- 17: **if** $c_{ij} > \tau$ **then**
- 18: $W_{ij} \leftarrow c_{ij}$ ▷ 边权即内点数
- 19: **阶段 4: 拓扑精化**
- 20: $\mathcal{T}_{MSF} \leftarrow \text{MaximumSpanningForest}(W)$ ▷ 保证全局连通性
- 21: $E_{final} \leftarrow \mathcal{T}_{MSF} \cup \{(i, j) \mid W_{ij} > \tau_{add}\}$ ▷ 重新插入强回环闭合
- 22: 构建最终图 $\mathcal{G} = (\mathcal{V}, E_{final}, W)$
- 23: **return** $\mathcal{G}$

- 消除所有环,
- 确保各连通分量内部连通,
- 促进全局结构一致性。

直观上, 该过程在仅保留最可靠关系的前提下, 找到连接图所需的最少边数。

回环重插入。真实环境常包含真实回环——例如矩形走廊或环形路径。由于 MSF 移除了所有环, RAVEL 仅重新引入几何支持格外强的边。当满足

$$W_{ij} > \tau_{add}, \quad \tau_{add} = 1.5 \tau \quad (33)$$

时, 边 (i, j) 被恢复。仅重插入高置信回环, 使 RAVEL 在拓扑稀疏性与表示完整性之间取得平衡。

讨论与遗留挑战。相比仅检索或朴素验证的流程, RAVEL 显著提升了边精确率与全局一致性。不过, 纹理极度贫乏的环境或大面积开阔区域仍可能引入歧义。这类情形反映的是单目图像的本质局限, 而非算法本身的缺陷。尽管存在这些挑战, RAVEL 始终为正文所述的定位与导航模块提供稳定基础。

算法 4 置信度传播定位的更新过程**Require:** 当前观测 I_t , 上一时刻置信度 $\mathbf{b}_{t-1}$, 转移矩阵 $\mathbf{T}$, 节点描述符 $\mathcal{Z} = \{z_1, \dots, z_N\}$ **Ensure:** 更新后的置信度状态 $\mathbf{b}_t$

- 1: **1. 置信度预测 (转移)**
- 2: $\bar{\mathbf{b}}_t \leftarrow \mathbf{b}_{t-1} \cdot \mathbf{T}$ ▷ 沿图拓扑传播置信度
- 3: **2. 观测似然估计**
- 4: 提取特征向量 $z_t \leftarrow f(I_t)$
- 5: **for** 每个节点 $v_i \in V$ **do**
- 6: $d_i \leftarrow \|z_t - z_i\|_2^2$ ▷ 欧氏距离
- 7: $L(v_i|I_t) \leftarrow \exp(-\lambda \cdot d_i)$ ▷ 计算似然
- 8: **3. 自适应权重计算**
- 9: $\eta_t \leftarrow \frac{-\sum_i \bar{\mathbf{b}}_t(v_i) \log \bar{\mathbf{b}}_t(v_i)}{\log N}$ ▷ 归一化熵
- 10: 依据式 (32) 计算 $w_p(\eta_t)$
- 11: $w_o(\eta_t) \leftarrow 1 - w_p(\eta_t)$
- 12: **4. 融合与更新**
- 13: **for** 每个节点 $v_i \in V$ **do**
- 14: $\mathbf{b}_t(v_i) \leftarrow \bar{\mathbf{b}}_t(v_i)^{w_p} \cdot L(v_i|I_t)^{w_o}$ ▷ 自适应融合
- 15: **5. 归一化**
- 16: $\eta \leftarrow \sum_j \mathbf{b}_t(v_j)$
- 17: $\mathbf{b}_t \leftarrow \mathbf{b}_t / \eta$ ▷ 保证 $\sum \mathbf{b}_t = 1$
- 18: **return** $\mathbf{b}_t$

D 置信度传播定位：补充推导

本节对 ULVN 所用的置信度传播 (Belief Propagation, BP) 定位方法给出更深入的说明。正文概述了 BP 定位的基本原理, 这里通过讨论底层数学框架、解释拓扑转移模型并阐明置信度传播的执行方式来扩展推导, 同时描述将预测置信度与新观测结合的自适应融合过程。

置信度传播定位的动机。定位的目标是估计机器人在拓扑图中的位置, 做法是维护一个覆盖所有可能节点的置信度分布, 表示机器人处于各位置的概率。传统定位方法常依赖高度依赖时序先验的序列滤波 (如卡尔曼滤波)。然而本文方法没有关于机器人运动的直接时序或里程计信息, 因此必须依靠环境拓扑, 并根据新观测自适应更新置信度。

本文提出的置信度传播方法通过在图结构上传播置信度、并以迭代方式融入新观测来解决该问题。它不假设数据具有时序性、也不依赖里程计, 而是利用拓扑图中编码的空间关系逐步更新置信度。

定位问题的图表示。本文定位方法作用于拓扑图 $G = (V, E)$, 其中 V 是节点集合 (环境中的位置), E 是边集合 (位置之间的连接)。每个节点 $v_i \in V$ 对应一个唯一地标 (或一组地标), 边 E 表示地标之间的可通行性。

设置信度状态 $\mathbf{b}_t \in \mathbb{R}^N$ 表示时刻 t 在 N 个节点上的概率分布。该置信度初始为均匀分布, 或依据首次检索初始化, 随后通过结合图的先验知识与机器人传感器的新观测进行更新。

通过多步转移进行置信度预测。置信度预测步使用由图邻接矩阵 $\mathbf{A}$ 导出的转移矩阵 $\mathbf{T}$ 。与仅考虑直接邻居的简单马尔可夫模型不同，本文实现考虑至多 K 步的可达性（实验中 $K = 3$ ），以建模机器人在两次更新之间可能的运动。

通过累加邻接矩阵的幂构建累积可达矩阵：

$$\mathbf{C} = \sum_{k=0}^K \mathbf{A}^k, \quad (34)$$

其中 $\mathbf{A}^0 = \mathbf{I}$ （单位矩阵）。对 $\mathbf{C}$ 按行归一化以保证随机性，得到转移矩阵 $\mathbf{T}$ ：

$$\mathbf{T}_{ij} = \frac{\mathbf{C}_{ij}}{\sum_j \mathbf{C}_{ij}}. \quad (35)$$

为预测时刻 t 的置信度状态，将上一时刻的后验置信度 b_{t-1} 通过该转移模型传播：

$$\bar{b}_t = b_{t-1} \cdot \mathbf{T}. \quad (36)$$

这里 $\bar{b}_t$ 表示尚未纳入当前观测的时刻 t 先验置信度。该形式有效地把概率质量扩散到上一估计的拓扑邻域。

置信度与观测的自适应融合。置信度传播完成后，将其与当前视觉观测 I_t 融合。

1. 观测似然：本文依据查询嵌入 z_t 与节点嵌入 z_i 之间的欧氏距离 $d(z_t, z_i)$ 计算似然 $L(v_i|I_t)$ 。为应对特征空间中动态范围的变化，计算自适应缩放因子 λ ：

$$\lambda = \frac{\ln(\delta)}{q_{0.975}(\mathbf{d}) - q_{0.025}(\mathbf{d})}, \quad (37)$$

其中 $\mathbf{d}$ 是 z_t 与所有节点之间的距离向量， q 表示分位数函数。似然随后建模为：

$$L(v_i|I_t) \propto \exp(-\lambda \cdot d(z_t, z_i)). \quad (38)$$

2. 基于熵的加权：本文方法的一项核心贡献是依据当前置信度的不确定性自适应加权预测与观测。本文用置信度分布的归一化香农熵 η 量化不确定性：

$$\eta(b) = \frac{-\sum_i b(v_i) \ln(b(v_i))}{\ln(N)}. \quad (39)$$

基于 η 用分段线性函数定义预测权重 w_p 与观测权重 w_o 。熵低（置信度高）时更依赖运动预测；熵高（不确定性大）时更依赖当前观测以重定位。

具体地，预测权重 w_p 计算为：

$$w_p(\eta) = \begin{cases} 0.6 & \text{若 } \eta < 0.3 \\ 0.6 - 0.75(\eta - 0.3) & \text{若 } 0.3 \leq \eta \leq 0.7 \\ 0.3 & \text{若 } \eta > 0.7 \end{cases} \quad (40)$$

观测权重取互补形式 $w_o(\eta) = 1.0 - w_p(\eta)$ 。

3. 几何融合更新：最后，更新后的后验置信度 b_t 通过先验置信度与似然的加权几何平均计算：

$$b_t(v_i) \propto \bar{b}_t(v_i)^{w_p} \cdot L(v_i|I_t)^{w_o}. \quad (41)$$

算法 5 置信度感知子目标搜索 (BASS)**Require:** 拓扑地图 $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, 目标图像 I_{goal} , 前视窗口 Δ **Ensure:** 通向目标的机器人轨迹

```

1: 阶段 1: 初始化
2: 给定初始观测, 用 BPL 初始化置信度  $b_0$ 
3:  $v_{goal} \leftarrow \operatorname{argmax}_{v \in \mathcal{V}} P(v | I_{goal})$  ▷ 目标识别
4:  $v_{curr} \leftarrow \operatorname{argmax}_{v \in \mathcal{V}} b_0(v)$ 
5:  $P \leftarrow \text{GlobalPlan}(\mathcal{G}, v_{curr}, v_{goal})$  ▷ 计算全局路径  $P = \{v_0, \dots, v_{goal}\}$ 
6: 阶段 2: 导航循环
7: while 机器人尚未到达  $v_{goal}$  do
8:   获取当前观测  $o_t$ 
9:    $b_t \leftarrow \text{BPL}(b_{t-1}, o_t)$  ▷ 更新置信度状态
10:  // 偏离检测与重规划
11:   $p_{on\_path} \leftarrow \max_{v \in P} b_t(v)$ 
12:  if  $p_{on\_path} < \tau_{threshold}$  or  $\text{IsDeviated}(b_t, P)$  then
13:     $v_{curr} \leftarrow \operatorname{argmax}_{v \in \mathcal{V}} b_t(v)$ 
14:     $P \leftarrow \text{GlobalPlan}(\mathcal{G}, v_{curr}, v_{goal})$  ▷ 重规划路径
15:  // 依赖路径的子目标选择
16:   $k^* \leftarrow \operatorname{argmax}_k \{b_t(v_k) | v_k \in P\}$  ▷ 定位在路径上的索引
17:   $v_{sub} \leftarrow P_{\min(k^* + \Delta, |P|)}$  ▷ 前视选择
18:  // 闭环执行
19:   $a_t \leftarrow \text{LocalPlanner}(v_{sub}, b_t)$ 
20:  执行动作  $a_t$ 

```

这种乘性融合（等价于对数空间中的加权求和）比加性融合更尖锐，能有效抑制那些同时不被转移模型与视觉证据支持的节点。

小结。置信度传播定位为在拓扑地图中估计机器人位置提供了简洁方案，无需依赖时序或里程计数据。通过把多步拓扑传播与熵感知自适应融合结合，系统能有效应对各类不确定性。动态权重 w_p 与 w_o 使系统在置信度高时保持既有判断，在位置歧义时仍对新视觉证据保持响应。

E 置信度感知子目标搜索细节

本节详细说明置信度感知子目标搜索 (BASS) 算法，它在 ULVN 的视觉导航中起关键作用。BASS 把置信度传播与路径规划整合起来，使机器人在不确定环境中高效导航，其置信度状态随移动持续更新。正文在较高层面介绍了 BASS，本节深入算法细节，逐步解释其运行方式与背后动机。

BASS 的动机。当机器人必须依据视觉线索和不确定的置信度状态导航时，简单的路径规划方案并不够。传统方法通常在已知地图与精确位置的假设下，规划从当前状态到目标的路径。然而在视觉导航中，机器人对自身位置的置信度是不确定且持续演化的，因此需要一种更鲁棒的方案，在拓扑导航语境下同时考虑置信度状态与目标。

BASS 通过引入**置信度感知子目标搜索**应对该挑战。它不是僵化地依次跟随节点序列，而是依据机器人当前置信度从全局规划中动态确定即时子目标，从

而确保机器人瞄准路径上可达且具前瞻性的路点，或在偏离足够大时检测出需要重规划。

BASS 的规划与执行过程。 BASS 算法作用于 ULVN 构建的拓扑地图，过程分为全局路径规划与动态子目标选择：

1. 初始定位：先用置信度传播定位（BPL）模块更新机器人置信度，确定其在拓扑图上的初始概率分布。
2. 目标识别：目标由一张目标图像表示，机器人的置信度状态被更新以反映最可能对应目标的节点，该目标节点即为导航任务的终点。
3. 全局路径规划：在选择即时子目标之前，先从最可能的起始节点到目标节点计算全局拓扑路径 $P = \{v_0, v_1, \dots, v_{goal}\}$ ，该路径由可通行边连接的一串视觉地标（节点）组成。
4. 动态子目标选择：机器人导航过程中，BASS 依据当前置信度状态判断其在路径 P 上的进展，并选取路径上一个“前视”节点作为即时子目标，以驱动高效前进。
5. 偏离检测与重规划：关键在于，BASS 持续核验机器人当前置信度状态是否与全局路径 P 所含节点一致。若机器人把自身定位到路径之外，系统即触发重规划。

依赖路径的子目标选择机制。 BASS 的核心是在给定全局规划 P 与当前置信度 b_t 时选择即时目标图像（子目标）的机制。与信息寻求式探索不同，该过程聚焦于高效的路径遍历。

1. 路径上定位：设全局路径为节点序列 P 。在时刻 t ，机器人估计自身在路径上的位置索引 k ，使置信度 $b_t(v_k)$ 在 P 的所有节点中最大：

$$k^* = \operatorname{argmax}_k b_t(v_k), \quad \text{其中 } v_k \in P \quad (42)$$

2. 前视选择：确定机器人在路径上的当前位置索引 k^* 后，BASS 选取一个下游节点作为即时子目标以鼓励前进。子目标 v_{sub} 取为：

$$v_{sub} = v_{k^* + \Delta}, \quad (43)$$

其中 Δ 表示前视窗口。这确保机器人朝未来路点前进，而非试图收敛到一个已在经过的位置。

闭环导航与重规划。 子目标选定后，机器人借助局部规划器朝其导航。然而由于视觉导航在定位与运动两方面都存在不确定性，机器人可能偏离规划路径。此时 BASS 通过持续更新置信度状态并调整路径来实现闭环导航。

1. 路径执行：机器人从当前位置沿路径行至所选子目标。该路径由同时考虑置信度状态与拓扑图的局部规划器规划。机器人移动过程中，其位置置信度由 BPL 模块更新。
2. 偏离检测：若机器人显著偏离规划路径，BASS 触发重规划机制。偏离通过比较机器人当前置信度状态与依据规划路径的期望置信度来检测。若置信度状态显示机器人已不在正确路径上，则选择新的子目标并将机器人重新导向该目标。
3. 重规划与自适应：一旦发生偏离，或获得新信息（如通过视觉观测），置信度状态即被更新，并重新计算通向目标的路径。该迭代过程持续进行，直到机器人到达最终目标或置信度状态收敛。

计算复杂度与性能。 BASS 算法的计算复杂度受拓扑图规模与所选子目标数量影响。选择子目标需要评估所有候选节点，为每个节点计算期望信息增益与置信度熵，该过程为 $O(N)$ ，其中 N 是图中节点数。

子目标选定后，路径规划步使用图搜索算法（如 Dijkstra 或 A*），时间复杂度为 $O(E \log N)$ ，其中 E 是图中边数。总体而言，BASS 的复杂度主要由置信度传播与路径搜索两步主导。

小结。 置信度感知子目标搜索 (BASS) 算法通过纳入置信度更新与自适应路径规划，为视觉导航提供了鲁棒框架。依据机器人置信度状态选择中间子目标，BASS 确保机器人始终朝环境中信息量最大的位置前进。该方案在机器人位置不确定的场景中尤为有用，能在复杂动态环境中实现高效导航。BASS 的闭环能力使机器人可从偏离中恢复并继续趋向目标，成为真实世界视觉导航部署的有力工具。

F 扩展实验结果

表 8: GRScenes 数据集上不同拓扑图构建方法的详细对比评测。

方法	精确率	召回率	F1	准确率
商业				
PlaceNav top 2 [65]	0.5078 $\pm$ 0.0453	0.6803 $\pm$ 0.0604	0.5682 $\pm$ 0.0477	0.9943 $\pm$ 0.0011
PlaceNav top 5 [65]	0.2324 $\pm$ 0.0262	0.7878 $\pm$ 0.0479	0.3587 $\pm$ 0.0265	0.9857 $\pm$ 0.0022
ViNT top 2 [59]	0.5291 $\pm$ 0.0724	0.6839 $\pm$ 0.1021	0.5962 $\pm$ 0.0701	0.9951 $\pm$ 0.0013
ViNT top 5 [59]	0.2823 $\pm$ 0.1686	0.8239 $\pm$ 0.0410	0.3882 $\pm$ 0.1522	0.9862 $\pm$ 0.0059
RAVEL (本文)	0.7012 $\pm$ 0.0886	0.7687 $\pm$ 0.0599	0.7328 $\pm$ 0.0756	0.9972 $\pm$ 0.0011
家居				
PlaceNav top 2 [65]	0.5446 $\pm$ 0.0390	0.7224 $\pm$ 0.0690	0.6203 $\pm$ 0.0478	0.9946 $\pm$ 0.0018
PlaceNav top 5 [65]	0.3075 $\pm$ 0.1192	0.8032 $\pm$ 0.0590	0.4149 $\pm$ 0.0491	0.9850 $\pm$ 0.0060
ViNT top 2 [59]	0.4752 $\pm$ 0.1361	0.6910 $\pm$ 0.1406	0.6062 $\pm$ 0.0857	0.9935 $\pm$ 0.0022
ViNT top 5 [59]	0.1502 $\pm$ 0.0714	0.8281 $\pm$ 0.0672	0.3289 $\pm$ 0.1095	0.9761 $\pm$ 0.0094
RAVEL (本文)	0.7196 $\pm$ 0.1291	0.7625 $\pm$ 0.1217	0.7402 $\pm$ 0.1252	0.9967 $\pm$ 0.0016
全部 (家居 + 商业)				
PlaceNav top 2 [65]	0.5262 $\pm$ 0.0503	0.7113 $\pm$ 0.0805	0.6043 $\pm$ 0.0606	0.9948 $\pm$ 0.0014
PlaceNav top 5 [65]	0.2699 $\pm$ 0.0915	0.7655 $\pm$ 0.1584	0.3731 $\pm$ 0.0244	0.9853 $\pm$ 0.0052
ViNT top 2 [59]	0.5201 $\pm$ 0.0718	0.6775 $\pm$ 0.1252	0.5815 $\pm$ 0.0724	0.9946 $\pm$ 0.0015
ViNT top 5 [59]	0.2660 $\pm$ 0.1540	0.8246 $\pm$ 0.0560	0.3806 $\pm$ 0.1329	0.9816 $\pm$ 0.0108
RAVEL (本文)	0.7104 $\pm$ 0.1048	0.7656 $\pm$ 0.0905	0.7365 $\pm$ 0.0976	0.9970 $\pm$ 0.0013

本节给出补充实验结果，以充实正文中的结论。

本文提供更详细的结果以补充正文实验，并对 RAVEL 中的自适应参数 τ 与 d_{VPR} 做敏感性分析，以论证锚点选择策略的鲁棒性与方法整体的有效性。本文还通过报告连通分量数、平均最短路径长度与图直径等指标，分析所构建拓扑图的结构性质，并给出不同方法所得拓扑图的可视化对比。考虑到多数现有方法依赖时序序列，本文在若干包含大量时序序列的数据集上验证全局定位算法

表 9: GRScenes 数据集上拓扑图构建的详细定量评测。

场景	指标	Top-k ANN	本文 w/o (MSF, AVP τ)	本文 w/o MSF	本文
家居	精确率	0.1537 ± 0.0551	0.4625 ± 0.1459	0.7198 ± 0.1075	0.7196 ± 0.1291
	召回率	0.8898 ± 0.1095	0.6177 ± 0.1604	0.3521 ± 0.0492	0.7625 ± 0.1217
	F1 分数	0.2598 ± 0.0835	0.5070 ± 0.0754	0.4705 ± 0.0564	0.7402 ± 0.1252
	准确率	0.9673 ± 0.0106	0.9929 ± 0.0020	0.9951 ± 0.0016	0.9967 ± 0.0016
商业	精确率	0.0953 ± 0.0341	0.2728 ± 0.1145	0.6621 ± 0.1202	0.7012 ± 0.0886
	召回率	0.9343 ± 0.0274	0.8177 ± 0.0978	0.4920 ± 0.0981	0.7687 ± 0.0599
	F1 分数	0.1713 ± 0.0554	0.3922 ± 0.0967	0.5609 ± 0.0984	0.7328 ± 0.0756
	准确率	0.9496 ± 0.0220	0.9868 ± 0.0050	0.9962 ± 0.0011	0.9972 ± 0.0011
全部	精确率	0.1245 ± 0.0530	0.3676 ± 0.1590	0.6910 ± 0.1117	0.7104 ± 0.1048
	召回率	0.9121 ± 0.0788	0.7177 ± 0.1637	0.4220 ± 0.1039	0.7656 ± 0.0905
	F1 分数	0.2156 ± 0.0815	0.4496 ± 0.1017	0.5157 ± 0.0894	0.7365 ± 0.0976
	准确率	0.9584 ± 0.0188	0.9898 ± 0.0048	0.9956 ± 0.0014	0.9970 ± 0.0013

BPL。为模拟机器人运动过程中出现的位置、朝向与运动模糊变化，本文施加图像裁剪、视角修改与噪声注入等处理。这些测试进一步验证了本文方法的有效性与泛化能力。最后，本文给出部分导航轨迹的可视化以说明系统性能，重点展示重规划机制的效用。

F.1 实验细节

为严格评测系统性能，本文对全局定位与自主导航分别开展实验，在已建图环境中使用随机选取的轨迹。定位实验的测试协议是在所有场景中遍历随机路径以评估位置估计准确率；为确保统计显著性，每个环境生成五条不同路径。导航评测采用类似的随机路径生成策略，并根据场景具体复杂度略作调整。

F.2 扩展实验

评测指标与实验设置。在扩展实验中，本文跨多个任务评测 ULVN 以全面评估其性能，核心任务包括：

1. 拓扑建图：

本文通过敏感性分析验证 RAVEL 中一次性标定机制的鲁棒性，并从连通分量数、平均最短路径长度与图直径等多个维度对所构建拓扑图的性质做统计评估。

2. 定位：

本文在时序导航数据集上开展仿真全局定位实验。对每个片段中的图像，施加裁剪、视角修改与噪声注入等图像处理，以模拟机器人运动中遇到的位置、朝向与运动模糊变化。处理后的图像作为机器人观测，原始图像作为拓扑图中的节点图像。本文对每张图像执行全局定位并报告准确率。这些实验在 RECON [56]、SCAND [31]、GoStanford [25] 与 SACSOn [24] 数据集上进行。

3. 导航：

本文给出部分场景中导航路径的可视化，并提供导航方法的消融实验，包括重规划机制与全局定位的实验。

表 10: 不同方法在不同数据集上的定位准确率对比。

数据集	条件	定位准确率			
		MegaLoc [3]	JIST [4]	ViNT [59]	BPL (本文)
GoStanford [25]	旋转	0.9930	0.8047	0.9302	0.9891
	旋转 + 高斯	0.8674	0.3395	0.9295	0.9559
	旋转 + 泊松	0.8488	0.3209	0.9226	0.9412
	旋转 + 裁剪	0.9884	0.6209	0.6140	0.9853
RECON [56]	旋转	0.7597	0.7394	0.9306	0.9429
	旋转 + 高斯	0.7022	0.7039	0.9069	0.9375
	旋转 + 泊松	0.6362	0.6937	0.9120	0.9062
	旋转 + 裁剪	0.6379	0.5262	0.7733	0.9301
SACSoN [24]	旋转	0.9507	0.6494	0.8410	0.9565
	旋转 + 高斯	0.6669	0.3172	0.8267	0.8862
	旋转 + 泊松	0.6089	0.3188	0.8410	0.8556
	旋转 + 裁剪	0.9006	0.4356	0.5437	0.9394
SCAND [31]	旋转	0.9471	0.6960	0.8784	0.9753
	旋转 + 高斯	0.7364	0.4486	0.8753	0.8772
	旋转 + 泊松	0.7746	0.4481	0.8768	0.8827
	旋转 + 裁剪	0.8920	0.5183	0.6305	0.9259
平均	总体	0.8069	0.5363	0.8270	0.9304

拓扑建图性能の詳細结果。不同拓扑图构建方法对比的详细数据见表 8。细分类别的考察表明, RAVEL 在多样环境语境下保持了更优的性能稳定性。在商业场景中, RAVEL 取得 0.7012 的精确率与 0.7328 的 F1 分数, 显著领先最接近的竞争者; 在家居场景中这一优势更为明显, RAVEL 达到 0.7196 的精确率与 0.7402 的 F1 分数。值得关注的是, ViNT top 5 等基线方法从商业 (0.2823) 转到家居 (0.1502) 环境时精确率急剧下降, 而 RAVEL 表现出显著的一致性。这说明本文方法对区分住宅布局与商业空间的结构变异性与视觉杂乱程度差异远为鲁棒。就统计方差而言, 表中报告的标准差反映了图构建的可靠性。ViNT top 5 等方法虽取得最高召回率, 但其精确率标准差很高 (商业场景中高达 ± 0.1686) 且均值精确率极低, 表明其依赖产生大量误检, 导致拓扑图不稳定。相反, RAVEL 展现出均衡的权衡。尽管其在家居场景中的精确率标准差高于商业场景, 反映住宅环境固有的不规则性, 但该方法始终以很窄的误差范围取得最高准确率。总体准确率上的低方差证实 RAVEL 是拓扑建图中最可靠、最具泛化性的方案, 有效抑制了启发式基线中观察到的剧烈波动。

基于表 9 给出的详细消融结果, 本文分析各模块在不同场景类别中的具体贡献以及方法的统计稳定性。场景细分结果显示, 最大生成森林 (MSF) 策略的作用在非结构化环境中尤为突出。在家居场景中, 移除 MSF 组件会使召回率从 0.7625 急剧降至 0.3521。这一大幅下降说明在常含不规则通路与视觉遮挡的住宅布局中, MSF 机制对维持图连通性、防止拓扑地图碎片化至关重要。相反, 在商业环境中, 同时缺失 MSF 与自适应验证参数 (Adaptive Verification Parameter, AVP) 会使精确率显著恶化至 0.2728。这说明在纹理重复、视觉特征相似的商业空间中, 自适应剪枝与结构约束对过滤简单最近邻方法无法区分的误检连接是必需的。就实验稳定性而言, 标准差反映了方法对环境差异的鲁棒性。数据显示家居场景通常引发更高的性能波动, 完整方法在家居场景中的精确率标准差 (± 0.1291) 高于商业场景 (± 0.0886), 这一差异反映了住宅内部固有的多样性与复杂性。尽管面临这一挑战, 本文方法在两个子类别中的准确

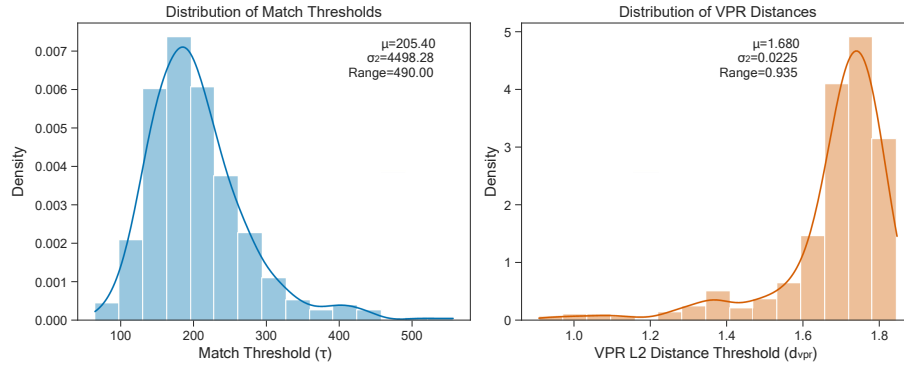


图 8: GRScenes 数据集上 RAVEL 自适应阈值的分布。

率标准差都极低, 介于 ± 0.0011 与 ± 0.0016 之间。这种一致性表明完整模型有效抑制了消融变体与 Top-k ANN 基线中的不稳定性, 能在不同场景特性下提供可靠的拓扑构建。

自适应参数敏感性分析。基于前述方法描述与 GRScenes 数据集上的统计可视化, 以下分析验证本文数据驱动标定策略的合理性与有效性。

自适应阈值必要性的验证。图 8 (左) 给出的匹配阈值 (τ) 分布以及图 10 中的敏感性分析, 为采用自适应方案而非固定参数提供了有力实证。匹配阈值 τ 呈现显著离散: 极差为 490.00, 方差约 4498, 形成以 $\mu = 205.40$ 为中心的宽泛类高斯分布。这一高方差说明内点与外点之间的最优分割点会随场景视觉辨识度剧烈波动。此外, 图 10 表明 τ 并非静态, 而是随查询索引动态变化, 对锚点对 (q_0, q_f) 的具体外观属性作出响应。固定的手工阈值必然无法覆盖这种异质性, 会在纹理贫乏区域造成漏检、在重复环境中造成误检。本文方法通过从汇集的匹配数聚类中导出 τ , 为每个具体语境有效地把阈值居中, 从而保证鲁棒的特征匹配性能。

基于距离的候选选择的有效性。图 8 (右) 中 VPR 距离 (d_{VPR}) 的分布展示了基于聚类的估计在定义有效搜索半径上的效力。该分布严重偏向较大距离值 ($\mu = 1.68$), 并在取值范围上界附近 (0.935) 紧密集中。这一统计行为与方法设计一致: d_{VPR} 由最高置信簇 (S_h) 内的最大距离确定。通过自适应该边界, 系统在不施加任意 “top-k” 限制的情况下最大化回环闭合候选的召回。分布中清晰的峰值说明该方法能一致地为验证确定一条 “几何视界”, 成功区分有效邻居与无关位置的嵌入空间。

图连通性的鲁棒性。标定策略的最终验证体现在图 9 所示的边权分布上。所得邻接矩阵权重服从形态良好的正态分布, 均值较高 ($\mu = 684.19$), 且在 500–900 区间密度显著。这说明经标定的阈值 (τ 与 d_{VPR}) 成功滤除了低置信噪声, 同时保留了强连接。分布中不存在明显的低权重尾部或双峰式碎裂, 说明 K-Means 聚类逻辑有效地把高置信簇 (S_h) 与噪声 (S_l) 分离。因此, 生成的拓扑图保持了高结构完整性, 其特征是强且可验证的边, 而非弱而歧义的连接。这证实一次性标定机制起到了有效过滤器的作用, 把原始视觉数据转化为高质量的拓扑表示。

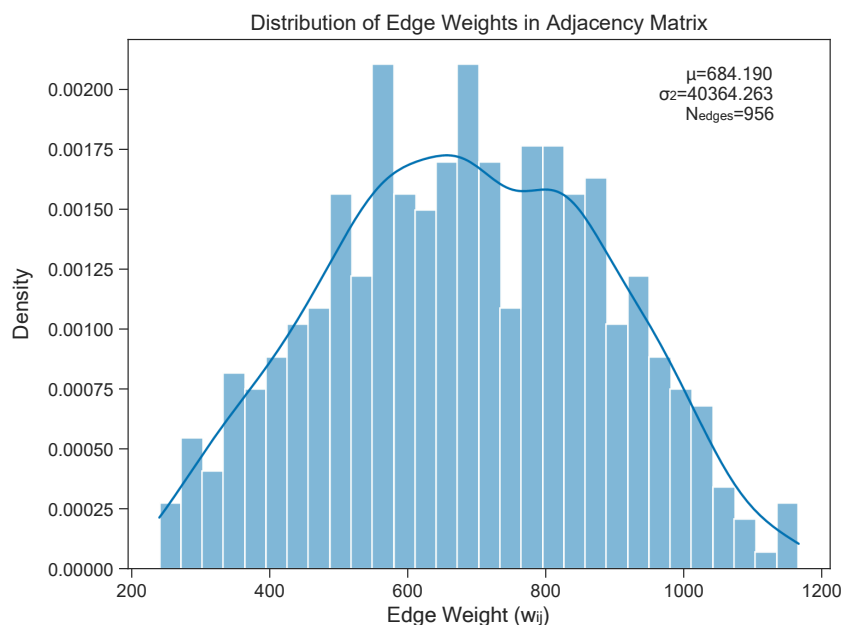


图 9: GRScenes 数据集上的边权分布。

拓扑图构建的定性分析。为定性评估所生成地图的结构完整性与可导航性，本文可视化了不同方法在三个不同场景中构建的拓扑图。图 13 给出对比视图，将 RAVEL 框架与基于特征、基于时序学习、基于全局检索的基线对照，并以人工核验的图作为参考真值（Reference）。

视觉对比揭示出图质量的显著差异，直接凸显启发式边过滤的局限。仅依赖 LightGlue 特征匹配密度的 Top-k ANN 方法表现出严重过连接：如第 (a) 行所示，它生成混乱的边网络，频繁穿透障碍与墙体，在纹理重复的场景 1 中尤为明显。这证实了在缺乏严格几何验证的情况下，仅凭原始匹配计数无法把相邻节点与视觉相似但空间相距很远的位置区分开。类似地，ViNT 与 PlaceNav 方法虽结构性稍好，仍存在大量噪声：在场景 2 的开阔布局与场景 3 的多房间布局中，这些基线频繁在互不相连的区域间建立虚假连接，形成违反环境物理约束的“蛛网”结构。

相比之下，RAVEL 产生的拓扑结构与参考真值高度吻合，这一优势可归因于流程中的两个具体组件。第一，数据驱动标定有效确定了自适应内点阈值 τ ，严格滤除了在 Top-k ANN 基线中造成“穿墙”边的弱几何匹配。第二，也是最关键的，两阶段剪枝策略确保了结构洁净：通过提取最大生成森林（MSF），RAVEL 构建出强健的无环骨架骨干，消除了场景 2 中所见的冗余噪声；随后强回环重插入机制仅选择性地恢复维持连通性所必需的高置信环。这一组合使 RAVEL 在场景 3 中既尊重房间分隔、又保持通过门口的有效可通行性，最终得到稀疏、几何一致且拓扑忠实的图。

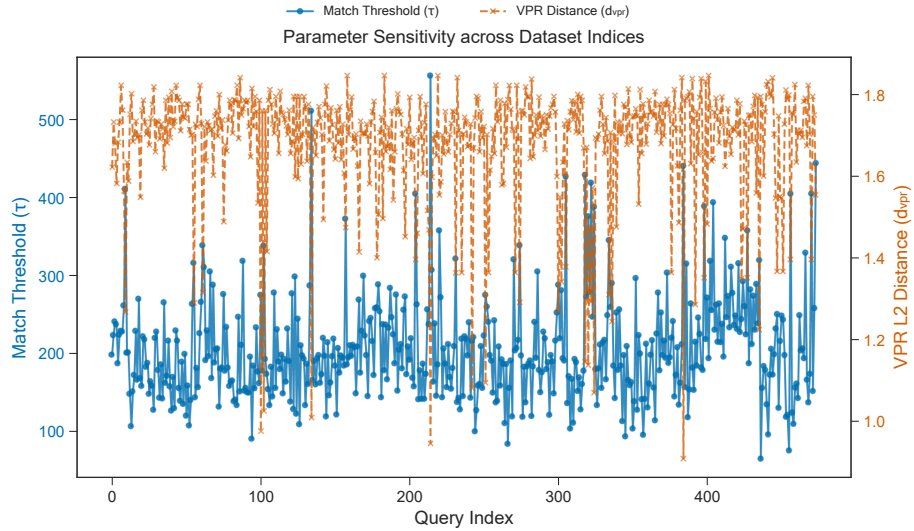


图 10: 锚点选择在不同数据集索引上的敏感性分析。



图 11: 家居场景中的导航可视化示例。

定位性能。为评估本文系统在真实导航挑战下的鲁棒性，本文在多个时序数据集上开展仿真全局定位实验，通过系统性施加视角修改与噪声注入等图像处理扰动，模拟机器人主动运动中遭遇的感知缺陷。表 10 给出这些不同条件下定位准确率的定量对比，反映各拓扑定位方法的稳定性。

结果表明，BPL 在所有评测基准上取得更优的泛化能力，总体平均准确率最高，达 0.9304，显著优于竞争基线，在涉及复杂视觉退化的场景中尤为突出。基线方法（尤其是 MegaLoc）暴露出的一个关键局限是对高频噪声明显敏感：MegaLoc 在 GoStanford 数据集的纯旋转场景中表现具有竞争力，但引入高斯或泊松噪声后准确率迅速恶化，在 SACSoN 与 SCAND 等数据集上下降近 30%。相比之下，BPL 对这类信号污染表现出极强韧性，在“旋转 + 高斯”与“旋转 + 泊松”条件下仍保持一致的高准确率，说明其特征表示能在抑制传感器噪声的同时保留结构相关的拓扑信息。

此外，“旋转 + 裁剪”条件是对部分可观测性与剧烈视角偏移的压力测试。在该设定下，ViNT 与 JIST 等方法性能大幅退化，在 SACSoN 与 SCAND 数据集上 ViNT 准确率显著下滑；而 BPL 框架在这些遮挡严重的设定中始终取得超过 0.92 的准确率。这说明即便视场中大部分内容缺失或被改变，BPL 仍能成功



图 12: 商业场景中的导航可视化示例。

利用全局上下文对智能体定位。综上，这些发现证实 BPL 不仅在理想条件下准确，还能在真实机器人导航固有的不可预测视觉变化下保持鲁棒。

导航性能。为定性评估所构建拓扑图的实用性，本文在真实感仿真环境中可视化了导航过程。该评测涵盖完整导航流程：从用无序图像构建图，到起点一目标定位，再到局部规划器执行路径规划。重点关注系统对执行误差的动态响应——置信度传播定位（BPL）模块识别出对最优轨迹的偏离，并触发拓扑重规划以确保导航成功。

图 11 与图 12 分别展示家居与商业场景中的导航性能。两类场景中，蓝色轨迹表示在拓扑图上计算的初始最优路径，橙色轨迹表示智能体因控制噪声或局部避障而偏离预定路线的情形。“偏离与重规划”行中的视觉数据表明，在这些偏离过程中机器人遇到的视角与初始规划显著不同。尽管存在这些视觉差异，BPL 模块仍成功检测出期望状态与实际状态的不匹配，促使系统生成新的拓扑路径。

在两个不同环境中都能成功恢复，说明本文框架具有鲁棒性。在杂物密集的家居场景中，机器人偏离后能有效绕过中央障碍，证明所构建的图在狭窄不规则空间中仍保持足够连通性；在纹理重复的商业场景中，系统纠正了显著的横向漂移，说明定位模块对大型开阔厅堂中常见的感知混淆具有韧性。这些结果证实，由无序图像构建的拓扑图提供了可靠的空间表示，配合 BPL 后可在执行条件不完美的情况下实现自主自我纠正与一致的目标达成。

实验结果小结。本节给出扩展实验结果，评测了 ULVN 在建图、定位与导航任务上的性能。ULVN 在所有任务上均优于基线方法，表明其在构建可靠拓扑地图、在地图中定位以及在复杂环境中导航方面兼具鲁棒性与有效性。仿真与真实环境的结果共同展示了 ULVN 在非结构化空间中实现自主导航的实用潜力。

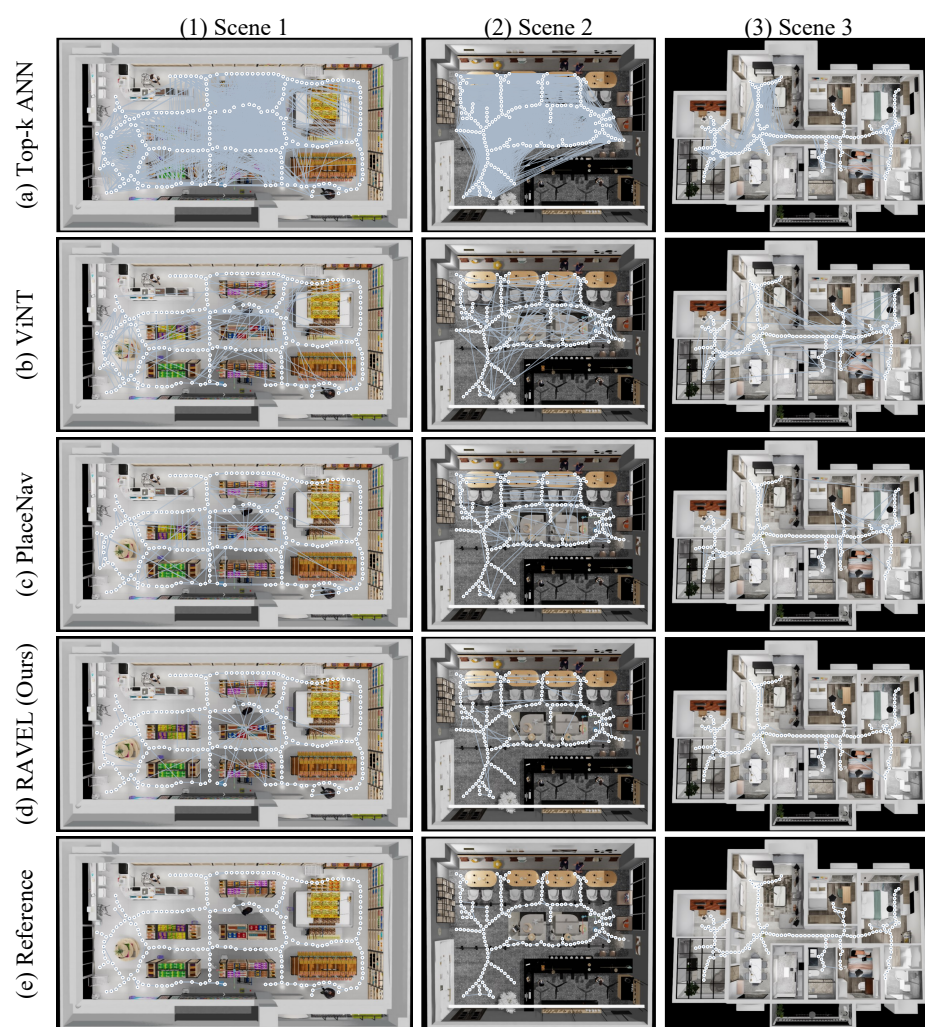


图 13: 由无序图像构建的拓扑图可视化示意。