FASTER, FEWER, & FARTHER: GRANULARITY-ALIGNED HIERARCHICAL LIDAR RELOCALIZATION IN A SHARED SPECTRAL GEOMETRY

Anonymous authors
Paper under double-blind review

ABSTRACT

Global LiDAR relocalization requires transferable association across growing maps, where geographically distant places can share similar structure. Exhaustive flat frame retrieval preserves a simple comparison interface but incurs map-linear search and increasing exposure to distant aliases. Hierarchical retrieval reduces this search, yet makes coarse routing an irreversible geographic hypothesis-pruning operation: a discarded location cannot be recovered by fine matching. We identify spatial, granularity, and embedding-consistency mismatches: coarse candidates can be appearance-coherent yet geographically distant, routing may compare evidence at unequal granularities, and separately fitted stage representations may disagree under domain shift. F^{3}Loc aligns support and granularity in a simple hierarchy with no routing-specific representation learning. Query context routes to spatially coherent map regions, after which the current frame localizes among the retained frames. The proposed shared-basis spectral calibration establishes a fixed comparison geometry before aggregation, and granularity-aligned routing preserves the resulting frame-neighborhood semantics from regional routing to final localization. Across Oxford, NCLT, and MCD, F^{3}Loc discards 96.1–99.5% of map frames, yielding 2.35–8.90× speedups over same-encoder exhaustive retrieval. Mean translation error falls by 54–56% on Oxford and 31–52% on NCLT and MCD. Under the same no-target-training protocol, F^{3}Loc achieves substantially lower localization error than industrial ANN and hierarchical retrieval controls, while attaining translation accuracy competitive with, and on several routes better than, in-domain regressors. Code will be released.

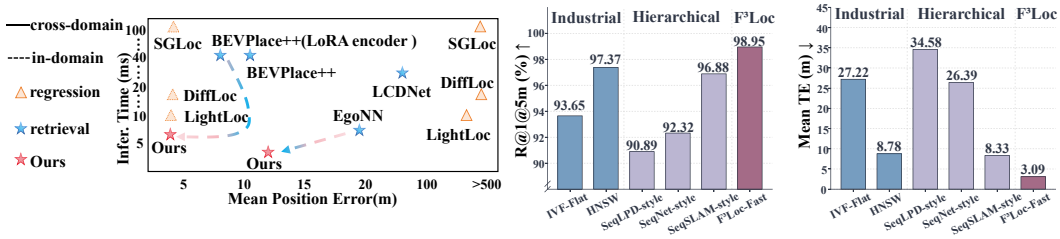


Figure 1: **Positioning F^{3}Loc among global localization and retrieval alternatives.** Left: representative localization paradigms under in-domain and cross-domain evaluation. Right: a controlled Oxford 11-13→17-14 comparison of industrial ANN, hierarchical retrieval baselines, and F^{3}Loc. Under a controlled Oxford protocol, F^{3}Loc achieves substantially lower localization error.

1 INTRODUCTION

Global LiDAR relocalization seeks a sensor pose in a previously mapped environment when external positioning is unavailable. Its central challenge is to preserve the correct geographic hypothesis as the map grows and operating conditions shift. Descriptor retrieval supports this transfer without

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

scene-specific refitting, as illustrated by EgoNN, BEVPlace++, and recent robust retrieval systems (Komorowski et al., 2022; Luo et al., 2025; Cao et al., 2026), but a larger global search space increases the opportunity for geographically distant structures with similar appearance to enter the candidate set. Pose regression avoids map search but remains tied to a learned coordinate frame (Li et al., 2023; 2025; Xing et al., 2026), while registration provides local precision only after a valid global association (Ao et al., 2023; Fung et al., 2026). Figure 1 places F^{3}Loc across the localization landscape and contrasts it with industrial ANN and hierarchical retrieval baselines.

Hierarchical retrieval is well established in vector search, sequence recognition, and coarse-to-fine localization (Malkov & Yashunin, 2020; Johnson et al., 2021; Milford & Wyeth, 2012; Garg & Milford, 2021; Fu et al., 2024). In relocalization, however, routing does more than reduce computation. It determines which geographic hypotheses remain reachable. Coarse routing is therefore an irreversible hypothesis-pruning operation whose quality depends not only on shortlist compactness, but on whether the retained support preserves the correct location. This motivates our central question of whether a relocalizer can change the granularity at which evidence is accumulated without changing the geometry by which that evidence is compared.

We identify three mismatches that make this difficult (Figure 2). First, sequence and appearance hierarchies build coarse hypotheses from frame-wise similarities. Distant aliases shared across neighboring query frames can form a coherent false trajectory, thereby allowing temporal aggregation to reinforce the wrong region (Milford & Wyeth, 2012; Liu et al., 2019). Second, a frame descriptor summarizes one observation, whereas a coarse prototype or centroid summarizes many map descriptors, so direct frame-to-prototype comparison can distort coarse ranking relative to the frame-level decision that follows (Johnson et al., 2021). Third, our diagnostics reveal an embedding-consistency mismatch in which a learned sequence descriptor or separately fitted coarse projection can reshape a transferable target-domain neighborhood under domain shift. On Oxford, a source-trained SeqNet-style mapping reverses an initially correct positive–negative ordering for 10.54% of eligible queries, causing coarse and fine stages to disagree on which candidates remain plausible (Garg & Milford, 2021). These mismatches concern different interfaces but share the same consequence. The hierarchy can prune the correct location before final localization is allowed to resolve it. Appendix C further analyzes the embedding-consistency mismatch and the associated neighborhood reshaping.

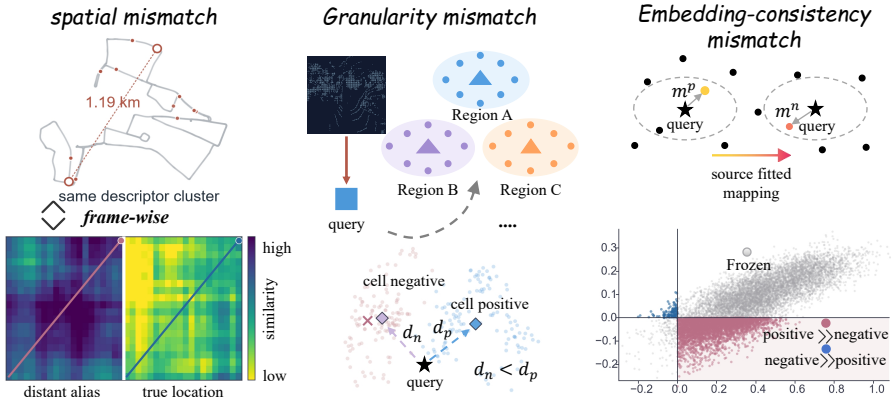


Figure 2: **Three mismatches in hierarchical relocalization.** Spatial mismatch arises when frame-wise distant aliases form a coherent false trajectory; granularity mismatch compares a frame observation with a region summary; embedding-consistency mismatch allows a separately fitted stage mapping to reshape transferable neighborhoods. Each can remove the correct geographic hypothesis before final frame matching.

Over fixed deployment descriptors, F^{3}Loc uses a deliberately simple, non-learned hierarchy to align physical support and evidence granularity in one comparison geometry. Map traversal defines spatially coherent regions; query context supplies routing evidence, and the current frame resolves the stored pose within selected regions. All frame descriptors are first placed in the shared spectral geometry, which then remains fixed as temporal and regional evidence is formed. Query context changes routing support; final localization reuses the calibrated frame kernel. Thus, F^{3}Loc changes evidence granularity without adding a stage-specific geometry.

Our contributions are as follows.
1. We identify spatial, granularity, and embedding-consistency mismatches as coupled failure modes under domain shift. They motivate a single principle. F^{3}Loc changes evidence support and granularity without changing comparison geometry, establishing that geometry at the frame level and preserving the same comparison rule through regional routing and final matching.
2. Across Oxford, NCLT, and MCD, and with both BEVPlace++ and EgoNN descriptors, our experiments isolate candidate formation, evidence granularity, and cross-stage geometry. Without target-scene training, F^{3}Loc substantially narrows the accuracy–efficiency–generalization trade-off. Mean TE falls by 31–56% relative to same-encoder exhaustive retrieval, 96.1–99.5% of frame comparisons are removed, and same-encoder speedups reach 2.35–8.90×.

2 Related Work

2.1 Global LiDAR Relocalization

LiDAR relocalization is commonly formulated as regression, registration, or retrieval. Regression predicts pose efficiently but binds its output to a learned coordinate frame (Chen et al., 2025; Yang et al., 2025; Wu et al., 2026); related image-based methods regress scene coordinates from sparse mapping features (Barroso-Laguna et al., 2026). Registration gives precise local alignment when observations overlap, but needs a valid upstream association (Qin et al., 2023; Ao et al., 2023; Fung et al., 2026). Retrieval provides global association without target-domain fitting across varied rotations, scenes, viewpoints, and weather (Komorowski et al., 2022; Luo et al., 2025; Xu et al., 2023; Lu et al., 2025; Shen et al., 2025; Griffiths et al., 2025; Shi et al., 2025; Cao et al., 2026), but flat search remains map-dependent and exposed to distant aliases.

These advances improve pose readout, local refinement, or the frame descriptor itself. F^{3}Loc instead asks what changes when transferable frame retrieval is embedded in an irreversible coarse-to-fine routing process. The coarse stage no longer merely accelerates search; it determines which geographic hypotheses remain available to final localization.

2.2 Hierarchical and Sequence Retrieval

Industrial ANN systems partition or connect descriptor space through inverted files and navigable graphs (Johnson et al., 2021; Malkov & Yashunin, 2020). Sequence methods organize observations through path consistency in SeqSLAM or appearance-defined super keyframes in SeqLPD (Milford & Wyeth, 2012; Liu et al., 2019). These mechanisms reduce candidate search, but their cells, graph neighborhoods, clusters, and paths need not coincide with a geographically local neighborhood. For relocalization, appearance or sequence coherence therefore does not guarantee spatial coherence. An internally consistent coarse hypothesis can still represent a distant structural alias.

Hierarchical methods also summarize coarse evidence through compact keys, learned sequence descriptors, or overlap verification (Kim et al., 2022; Ma et al., 2022; Vidanapathirana et al., 2021; Garg & Milford, 2021; Fu et al., 2024). SeqNet, for example, learns a sequence descriptor for shortlisting and then aggregates frame scores for reranking. When routing compares a single-frame observation with an aggregated candidate, the two sides encode unequal evidence support, and their similarity need not preserve the ordering required by final frame-level localization. F^{3}Loc instead aligns granularity by comparing ordered query context with spatially coherent regions during routing and the current frame with retained frames during localization.

2.3 Transferable Metric Geometry and Cross-Stage Consistency

Classical PCA and whitening reshape retrieval distances from aggregate feature statistics (Jégou & Chum, 2012). Multi-group models provide a closer precedent. CPC shares principal directions but retains group-specific spectra (Flury, 1984), while group-structured PCA estimates common components from aggregate covariances (Krzanowski, 1984). These characterize covariance structure, whereas training-time DG learns representations through RobustNet, MixStyle, and CP-CANet, which respectively suppress domain-sensitive covariance, mix source statistics, and learn a shared source-domain subspace by unfolding CPCA (Choi et al., 2021; Zhou et al., 2021; Chen & Seghouane, 2026). These approaches improve descriptor geometry or representation robustness, but do not specify how one comparison rule should survive irreversible coarse-to-fine aggregation.

F^{3}Loc uses this perspective for a different objective. In each source-only setting, pooled descriptors establish a reference basis, and source-wise variation measured there is aggregated into one retrieval geometry rather than retained as group-specific spectra. F^{3}Loc does not fit a CPC model or another learned representation; it instead constructs a fixed comparison interface for hierarchical retrieval. Fixed before evidence aggregation and reused by coarse routing and fine localization, this geometry lets the hierarchy change evidence granularity without redefining the comparison space. The 10.54% inversion under a source-trained mapping shows why transferable frame features alone do not ensure this cross-stage consistency.

3 METHOD

Hierarchical relocalization must decide which part of the map remains searchable and which stored frame supplies the pose. The three mismatches in Figure 2 imply that these decisions should respect physical support, align evidence granularity, and preserve comparison geometry. F^{3}Loc therefore makes three coupled commitments. Coarse hypotheses use spatially coherent support, routing compares query context with region-level evidence, and one source-calibrated geometry is established upstream of the hierarchy and reused through final frame matching. It fits no separate coarse representation. The hierarchy changes evidence support and granularity, not the underlying similarity rule.

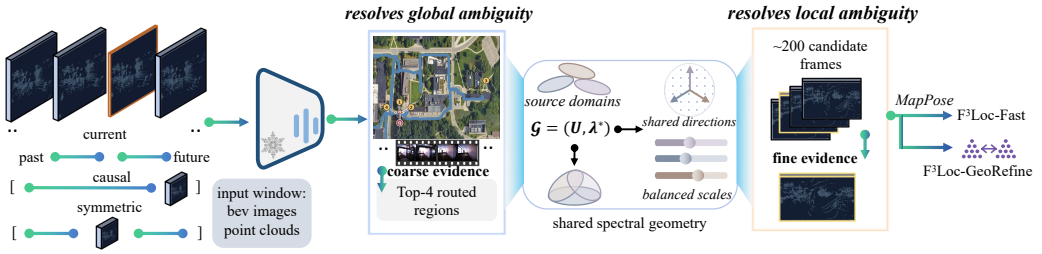


Figure 3: *Note: Shared spectral geometry is applied first to frame descriptors.* **F^{3}Loc builds a non-learned routing hierarchy over descriptors fixed after source-side adaptation and calibration.** Query context routes to spatially coherent map regions, and the current frame localizes among roughly 200 unique frame candidates. The shared geometry $\mathcal{G} = (\mathbf{U}, \boldsymbol{\lambda}^*)$ is applied to frame descriptors first, so both regional routing and final localization are built from the same frame kernel. F^{3}Loc-Fast returns MapPose; F^{3}Loc-GeoRefine optionally aligns local geometry.

3.1 PROBLEM FORMULATION AND DESIGN PRINCIPLES

Let the map be an ordered sequence $\mathcal{M} = \{(\mathbf{P}_i, \mathbf{T}_i)\}_{i=1}^{M}$ of LiDAR scans and stored poses $\mathbf{T}_i \in \mathrm{SE}(3)$, and let $\mathcal{Q} = \{\mathbf{Q}_t\}_{t=1}^{T}$ be an ordered query sequence from an unseen traversal of the mapped environment, available either in full or progressively over time. A frozen encoder g_θ maps each scan to a descriptor $\mathbf{f} \in \mathbb{R}^d$. At time t, relocalization selects a map index $\hat{i}_t$ and returns the corresponding MapPose $\widehat{\mathbf{T}}_t = \mathbf{T}_{\hat{i}_t}$. For a given source-only transfer setting, the spectral geometry is fitted once from its source domains $\{\mathcal{D}_e\}_{e=1}^{E}$ and remains frozen while processing the held-out target; no target statistics are used in fitting. We call this direct map-pose readout F^{3}Loc-Fast. F^{3}Loc-GeoRefine uses the retrieved MapPose to initialize local geometric alignment and returns a refined 6DoF pose. Coarse units have spatially coherent rather than appearance-defined support; contiguous traversal intervals provide this support in our datasets, although other region primitives are admissible. Both stages use the same transformed descriptors and distance (Figure 3). Appendix H specifies GeoRefine initialization, pose composition, and fallback behavior.

3.2 SHARED SPECTRAL COMPARISON GEOMETRY

The embedding-consistency mismatch in Figure 2 motivates a geometry fixed before aggregation and preserved across stages, which classical PCA/whitening and training-time DG do not explicitly target. CPC and group-structured PCA provide a useful view of shared directions and group-specific

variation (Flury, 1984; Krzanowski, 1984). F^{3}Loc departs from these formulations by using pooled descriptors to establish the reference basis and domain-level source statistics to determine a shared scale vector for hierarchical retrieval. Thus pooled samples set directions, while domains set scales.

Let $\bar{\mathbf{f}}_{e,i} = \mathbf{f}_{e,i}/\|\mathbf{f}_{e,i}\|_2$ denote the i-th normalized descriptor from source e, let $n_e = |\mathcal{D}_e|$, and let $N = \sum_e n_e$. The pooled mean and covariance are

$$\begin{aligned} \boldsymbol{\mu} &= \frac{1}{N}\sum_{e=1}^{E}\sum_{i=1}^{n_e}\bar{\mathbf{f}}_{e,i}, \\ \mathbf{C}_{\mathrm{src}} &= \frac{1}{N-1}\sum_{e,i}(\bar{\mathbf{f}}_{e,i}-\boldsymbol{\mu})(\bar{\mathbf{f}}_{e,i}-\boldsymbol{\mu})^\top. \end{aligned} \tag{1}$$

We eigendecompose the regularized pooled covariance as

$$\mathbf{C}_{\mathrm{src}} + \epsilon\mathbf{I} = \mathbf{U}\mathbf{D}\mathbf{U}^\top, \tag{2}$$

where $\epsilon > 0$ is a regularizer, $\mathbf{I}$ is the identity, and diagonal $\mathbf{D}$ contains the eigenvalues. The pooled basis gives all source domains one reference coordinate system before any domain-specific scaling is introduced. For each source, we measure its directional scale around the shared center along these common reference directions,

$$\mathbf{S}_e = \frac{1}{n_e-1}\sum_{i=1}^{n_e}(\bar{\mathbf{f}}_{e,i}-\boldsymbol{\mu})(\bar{\mathbf{f}}_{e,i}-\boldsymbol{\mu})^\top, \qquad \boldsymbol{\lambda}_e = \mathrm{diag}(\mathbf{U}^\top\mathbf{S}_e\mathbf{U}). \tag{3}$$

Here $\mathbf{S}_e$ is the source-wise second-order matrix about shared mean $\boldsymbol{\mu}$; measuring every source in the same $\mathbf{U}$ retains source-specific directional variation without allowing each source to redefine the coordinate system. We use their mean as the shared directional scale,

$$\lambda_j^* = \frac{1}{E}\sum_{e=1}^{E}\lambda_{e,j}. \tag{4}$$

This domain-level average lets each source contribute once to scaling after pooled-cache directions are fixed, rather than weighting scale only by sample count. Together, $\mathbf{U}$ and $\boldsymbol{\lambda}^*$ define $\mathcal{G} = (\mathbf{U}, \boldsymbol{\lambda}^*)$. For any descriptor $\mathbf{f}$, let $\bar{\mathbf{f}} = \mathbf{f}/\|\mathbf{f}\|_2$. The resulting geometry maps it to

$$\mathbf{A} = \mathbf{U}\,\mathrm{diag}\Big((\boldsymbol{\lambda}^*+\epsilon)^{-1/2}\Big)\,\mathbf{U}^\top, \qquad \mathbf{z}(\mathbf{f}) = \mathrm{norm}_2\big[\mathbf{A}(\bar{\mathbf{f}}-\boldsymbol{\mu})\big]. \tag{5}$$

Here $\mathbf{A}$ is the source-fitted transform and $\mathrm{norm}_2(\mathbf{v}) = \mathbf{v}/\|\mathbf{v}\|_2$. Crucially, $\mathbf{A}$ is applied before any temporal or regional aggregation, so Equation 11 expresses regional routing as an aggregate of the same calibrated frame kernel used for final localization.

3.3 Granularity-Aligned Retrieval over Physical Support

With $\mathcal{G}$ fixed, spatially coherent map regions address the spatial mismatch, while context-to-region routing followed by frame-to-frame localization aligns evidence granularity without changing the comparison rule. Let $\mathbf{z}_i^{\mathcal{M}} = \mathbf{z}(g_\theta(\mathbf{P}_i))$ and $\mathbf{z}_t^{\mathcal{Q}} = \mathbf{z}(g_\theta(\mathbf{Q}_t))$ be map and query frame descriptors in the shared geometry. In the current implementation, we approximate physical support by J intervals along the ordered mapping trajectory. Each interval has a 50-frame core and 10 frames of boundary context on either side, clipped at the map boundary; $\mathcal{C}_j \subseteq \{1,\dots,M\}$ denotes this expanded support. The cores advance by 40 frames, so adjacent supports overlap. More general map topologies can use spatially local graph partitions. Each region is represented by the normalized consensus of its frames,

$$\mathbf{r}_j = \mathrm{norm}_2\left(\frac{1}{|\mathcal{C}_j|}\sum_{i\in\mathcal{C}_j}\mathbf{z}_i^{\mathcal{M}}\right). \tag{6}$$

For query time t, let $\mathcal{I}_T = \{1,\dots,T\}$ denote the valid query indices and W the temporal extent. Temporal consensus can use either a symmetric or causal neighborhood,

$$\mathcal{W}_t^{\mathrm{sym}}(W) = \{\tau\in\mathcal{I}_T : |\tau - t| \le W\}, \qquad \mathcal{W}_t^{\mathrm{cau}}(W) = \{\tau\in\mathcal{I}_T : 0 \le t-\tau \le W\}. \tag{7}$$

Writing $\mathcal{W}_t$ for either regime, the analogous query consensus is

$$\mathbf{c}_t = \mathrm{norm}_2\left(\frac{1}{|\mathcal{W}_t|}\sum_{\tau\in\mathcal{W}_t}\mathbf{z}_\tau^{\mathcal{Q}}\right). \tag{8}$$

The symmetric neighborhood uses context around the query; the causal neighborhood is restricted to the observed prefix. Both sides summarize regional evidence; we rank with squared Euclidean distance, equivalent to cosine for unit-normalized descriptors. Regional routing selects the K closest regions,

$$\mathcal{R}_t = \operatorname*{arg\,min}_{\substack{\mathcal{R}\subseteq\{1,\dots,J\}\\|\mathcal{R}|=K}}\sum_{j\in\mathcal{R}}\|\mathbf{c}_t-\mathbf{r}_j\|_2^2, \qquad \mathcal{P}_t = \bigcup_{j\in\mathcal{R}_t}\mathcal{C}_j. \tag{9}$$

The final decision returns to unpooled frame evidence,

$$\hat{i}_t = \arg\min_{i\in\mathcal{P}_t}\|\mathbf{z}_t^{\mathcal{Q}}-\mathbf{z}_i^{\mathcal{M}}\|_2^2, \qquad \widehat{\mathbf{T}}_t = \mathbf{T}_{\hat{i}_t}. \tag{10}$$

Define the frame similarity $k_{\mathcal{G}}(\mathbf{f},\mathbf{f}') = \mathbf{z}(\mathbf{f})^\top\mathbf{z}(\mathbf{f}')$. Writing $\mathbf{s}_t = \sum_{\tau\in\mathcal{W}_t}\mathbf{z}_\tau^{\mathcal{Q}}$ and $\mathbf{s}_j = \sum_{i\in\mathcal{C}_j}\mathbf{z}_i^{\mathcal{M}}$, regional similarity satisfies

$$\mathbf{c}_t^\top\mathbf{r}_j = \frac{\displaystyle\sum_{\tau\in\mathcal{W}_t}\sum_{i\in\mathcal{C}_j}k_{\mathcal{G}}(g_\theta(\mathbf{Q}_\tau),g_\theta(\mathbf{P}_i))}{\|\mathbf{s}_t\|_2\,\|\mathbf{s}_j\|_2}. \tag{11}$$

Equation 11 links the stages. Its numerator sums the same frame-kernel values later used for localization, while its denominator normalizes the aggregates. This formalizes the central design principle that routing broadens evidence support without redefining the frame-level comparison geometry. Figure 5 tests this relation by changing the geometry at either stage.

Map-frame and region descriptors are computed once. After query projection, consensus, routing, and localization require $O((|\mathcal{W}_t|+J+|\mathcal{P}_t|)d)$ operations, versus $O(Md)$ for exhaustive retrieval; frame matching is restricted to $|\mathcal{P}_t|$ temporally supported candidates. The context, region, and routing budgets are fixed across targets and specified in Section 4.1.

4 Experiments

Our experiments ask four questions. Can source-only hierarchical retrieval match strong localization baselines? Is shared geometry sufficient by itself, or must the hierarchy preserve it across stages? How do the three alignment mechanisms recover the accuracy lost by naive routing? Do the gains persist in challenging cross-domain scenes, under long-tail aliases, and with another descriptor encoder?

4.1 Setup

We insert rank-4 LoRA adapters (α=8) into the 1×1 convolutions of the pretrained BEVPlace++ encoder and NetVLAD pooling layer. The frozen base leaves 1,536 trainable parameters, optimized for 20 epochs with source triplets. Oxford fitting uses KITTI 00/02/05/06/08 (17,135 frames) and MulRan DCC01/02 (11,440); NCLT and MCD additionally use 2,323 sampled Oxford frames. Appendix B gives sequence-level counts, adaptation controls, and the target-exposure comparison. Oxford 11-13 and NCLT 02-18 form the maps; Table 1 lists the queries, while Appendix A extends evaluation to long-horizon NCLT 05-26 and MCD day01→day10.

Table 1 compares retrieval systems EgoNN (RA-L 2022), LCDNet (T-RO 2022), HOTFormerLoc (CVPR 2025), and BEVPlace++ (T-RO 2025) (Komorowski et al., 2022; Cattaneo et al., 2022; Griffiths et al., 2025; Luo et al., 2025), together with regressors SGLoc, DiffLoc, and LightLoc (CVPR 2023–2025) (Li et al., 2023; 2024; 2025). Table 1 uses outdoor LiDAR methods with aligned localization protocols; Section 2.1 covers recent systems in adjacent modalities and settings. Under one frozen encoder and a no-target-training protocol, Table 2 compares IVF-Flat, HNSW, SeqLPD-style, SeqNet-style, and SeqSLAM-style controls; HNSW and SeqNet-style are additionally evaluated with $\mathcal{G}$ (Johnson et al., 2021; Malkov & Yashunin, 2020; Liu et al., 2019; Garg & Milford, 2021;

Table 1: Localization on Oxford and NCLT (mean TE, m / RE, °; end-to-end inference at batch 16). Regression is evaluated in-domain and by direct transfer; retrieval follows Section 4.1. Lower is better; best/second-best values are **bold**/underlined.

Method	Training regime	Oxford: TE / RE			NCLT: TE / RE			Infer. (ms, B=16)
		15-13	17-13	17-14	02-12	02-19	03-31	
Retrieval								
EgoNN	Cross-domain	26.93/5.62	22.80/5.47	19.88/4.99	15.53/28.08	12.25/37.00	10.94/22.44	7.4/6.9
F^{3}Loc-Fast (EgoNN, cau.)	Cross-domain	16.03/4.12	15.14/3.92	12.04/3.79	7.11/27.55	4.50/37.74	4.81/21.78	**4.2/4.9**
LCDNet	Cross-domain	105.91/22.65	94.73/22.42	93.80/25.35	16.97/32.26	18.98/39.44	15.85/27.57	27.2/52.5
HOTFormerLoc	Cross-domain	–	–	–	33.51/25.44	45.64/29.69	32.39/20.32	–/24.6
BEVPlace++	Cross-domain	9.58/4.03	16.95/4.42	11.66/4.46	3.00/25.40	2.71/36.50	3.28/20.71	41.4/29.5
Regression (in-domain training / cross-domain transfer)								
LightLoc	In-domain	**2.33**/1.21	3.19/1.34	3.11/1.24	0.98/2.76	0.89/2.51	0.86/2.67	9.9/75.0
	Cross-domain	481.66/97.41	478.22/100.30	472.57/98.03	528.67/95.58	528.75/95.98	535.02/96.24	
DiffLoc	In-domain	3.57/0.88	3.65/0.68	4.03/0.70	0.99/2.40	0.92/**2.14**	0.98/2.27	17.5/32.1
	Cross-domain	>500/90.28	>500/93.64	>500/95.35	417.05/89.54	417.10/92.93	409.95/92.80	
SGLoc	In-domain	3.01/1.91	4.07/2.07	3.37/1.89	1.20/3.08	1.20/3.05	1.12/3.28	113.7/210.6
	Cross-domain	>500/93.67	>500/88.90	>500/87.03	418.10/91.71	418.04/92.47	410.99/91.64	
Ours								
BEVPlace++ (LoRA encoder)	Cross-domain	7.58/3.59	8.97/3.72	8.08/3.32	2.59/25.23	2.04/36.26	2.48/20.36	41.4/29.5
F^{3}Loc-Fast (sym.)	Cross-domain	5.53/2.26	2.15/2.21	3.11/2.31	1.51/24.02	1.42/36.17	1.46/19.59	7.2/6.7
F^{3}Loc-Fast (cau.)	Cross-domain	4.78/2.39	3.43/2.33	3.09/2.35	1.51/24.17	1.83/36.31	1.53/19.68	7.2/6.7
F^{3}Loc-GeoRefine (sym.)	Cross-domain	5.44/0.82	**1.95/0.57**	2.96/**0.67**	0.47/**2.15**	**0.47**/2.17	**0.49/2.20**	224.4/284.7
F^{3}Loc-GeoRefine (cau.)	Cross-domain	4.70/**0.81**	3.25/0.69	**2.94**/0.71	**0.46/2.15**	0.88/2.27	0.56/2.25	224.4/284.7

Retrieval baselines use frozen checkpoints; latency is Oxford/NCLT query time; sym./cau. denotes symmetric/causal context; TE >500 m denotes coordinate collapse. All retrieval methods use MapPose, consistent with F^{3}Loc-Fast.

Milford & Wyeth, 2012). Figure 5 tests whether the same geometry must be preserved across stages (a–b) and whether the proposed source calibration improves over alternative ways of combining source statistics (c).

All targets use 50-frame region cores, 10-frame boundary context, and Top-4 routing. We use $W=50$ for symmetric context and $W=40$ for causal context, fixed across all datasets; Appendix E traces TE around this operating point as context, routing budget, and region-core length vary. Table 1 uses the full map and every map-positive query, defined by at least one map frame within 5 m. R@1 uses this radius, FAR counts errors above 50 m, and latency is end-to-end at batch 16 excluding offline map encoding. Appendix L consolidates metric definitions, candidate accounting, and fixed parameters.

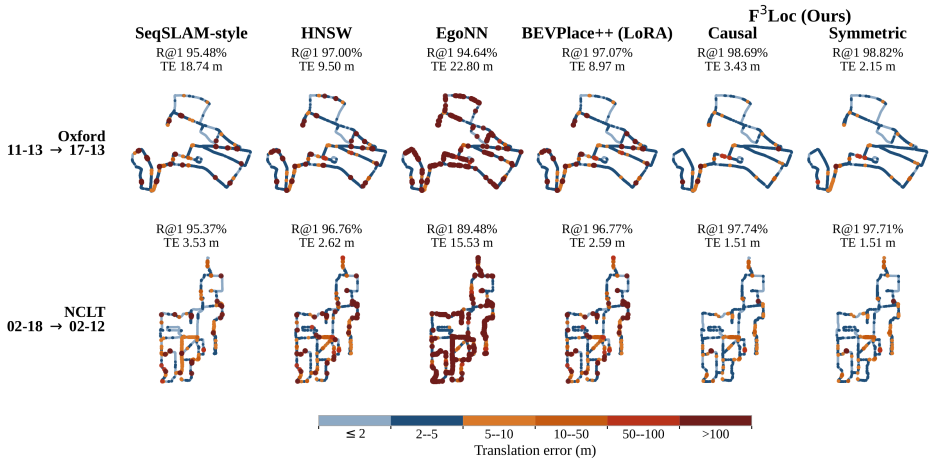


Figure 4: **Spatial distribution of retrieval error.** Oxford 11-13→17-13 and NCLT 02-18→02-12. Darker colors denote larger TE. Baseline methods exhibit geographically clustered large errors, whereas F^{3}Loc substantially reduces this tail.

4.2 Main Results

Table 1 separates in-domain fit from cross-domain association. Regression collapses under direct transfer, whereas retrieval remains independent of a learned target coordinate frame. Before geo-

metric refinement, causal F^{3}Loc-Fast reaches 3.09 m on Oxford 11-13→17-14, matching the 3.11 m result of the best in-domain regressor while using no target fitting. With the same LoRA-adapted BEVPlace++ encoder, hierarchical routing reduces Oxford/NCLT inference from 41.4/29.5 ms for exhaustive retrieval to 7.2/6.7 ms. Across the three Oxford routes, its mean TE is 3.77 m, compared with 8.21 m for exhaustive BEVPlace++ retrieval.

Without target fitting, GeoRefine surpasses the best in-domain regressor on Oxford 17-13/17-14 (1.95/2.94 versus 3.19/3.11 m) and all NCLT routes (0.46–0.88 versus 0.86–0.98 m), with NCLT RE of 2.15°–2.27°. Figure 4 shows that F^{3}Loc removes geographically clustered distant errors. Fixed-EgoNN results below test encoder generality.

4.3 Industrial and Hierarchical Retrieval Comparisons

HNSW stays near exhaustive retrieval (9.50 versus 8.97 m), while $\mathcal{G}$ lowers its TE to 4.26 m. Placing the same calibration before unchanged source-trained SeqNet-style mappings instead raises TE to 54.56 m at a matched 220-frame budget; input calibration alone cannot preserve consistency through separately learned stage functions. F^{3}Loc combines calibrated geometry, coherent support, and aligned granularity, reaching 3.43 m, 84 FAR, and 222 candidates. ANN indices accelerate descriptor lookup, whereas F^{3}Loc structures the localization hypotheses that survive for frame matching; Table 2 therefore reports both accuracy and total query inference. Appendix F gives the execution breakdown.

Table 2: Candidate formation on Oxford 11-13→17-13 with a fixed cross-domain encoder. $\mathcal{G}$ denotes our shared geometry. Inference follows the Table 1 B=16 query-level scope: query encoding and the complete online retrieval path are included, while map and index construction are offline. ANN lookup runs on CPU; F^{3}Loc routing and matching run on GPU. ↑ R@1; ↓ TE/FAR.

Method	Candidate formation	R@1	TE	FAR	Cand.	Infer. (ms)
IVF-Flat (Johnson et al., 2021)	Single-cell inverted search	93.36	33.82	2106	281	5.7
HNSW (Malkov & Yashunin, 2020)	Frame-level graph search	97.00	9.50	702	–	4.4
HNSW+$\mathcal{G}$	Frame-level graph search	<u>98.57</u>	<u>4.26</u>	<u>157</u>	–	4.6
SeqLPD-style (Liu et al., 2019)	Super-keyframe routing	90.29	44.80	3055	635	13.5
SeqNet-style (Garg & Milford, 2021)	Sequence Top-44 → frame rerank	93.35	23.38	1875	220	4.6
SeqNet-style+$\mathcal{G}$	Sequence Top-44 → frame rerank	86.38	54.56	4656	220	4.5
SeqSLAM-style (Milford & Wyeth, 2012)	Ordered sequence, Top-128	95.48	18.74	1001	128	129.4
F^{3}Loc-Fast (cau.)	Context → region → frame	**98.69**	**3.43**	**84**	222	7.2

4.4 Dissecting the Three Alignment Mechanisms

Table 3 adds the three mismatch remedies step by step. Current-frame-to-region routing cuts 40,565 frames to 233 candidates, but worsens Oxford/NCLT TE from 8.97/2.59 to 10.26/3.31 m, showing that a smaller search can hurt when frame evidence is compared with a regional summary. Matching query context at the same granularity recovers 8.04/1.93 m, and shared geometry reaches 2.15/1.51 m. At similar budgets, coherent support controls where routing searches, granularity alignment controls what evidence is compared, and shared geometry controls how it is compared. With fixed EgoNN descriptors, raw hierarchy supplies most of the gain over flat retrieval and shared geometry improves every route further, confirming complementary roles whose balance varies with the encoder (Appendix J). Appendix D visualizes one routed NCLT query.

Table 3: Alignment ablation over fixed contiguous map support.

Variant	Physical support	Query context	Shared geometry	TE / RE		Cand.	B1 (ms)
				Oxford 17-13	NCLT 02-12		
Exhaustive raw retrieval	–	–	–	8.97 / 3.72	2.59 / 25.23	40565	476.4
+ shared spectral geometry	–	–	✓	3.71 / 2.37	1.80 / 24.08	40565	475.9
Hierarchy only (frame→region)	✓	–	–	10.26 / 4.19	3.31 / 25.53	233	53.5
+ granularity alignment	✓	✓	–	8.04 / 4.05	1.93 / 25.30	225	53.5
F^{3}Loc-Fast (sym.)	✓	✓	✓	2.15 / 2.21	1.51 / 24.02	216	53.5

Query context uses symmetric aggregation; Cand. is the number of unique frames exposed to final matching, and Cand./B1 is measured on Oxford 17-13 at batch 1.

4.5 Controlled Analysis of Shared Geometry

Figure 5 tests whether routing stages should preserve one comparison geometry and how that geometry should be estimated from source statistics. Mixed raw/spectral stages reach 12.12 m TE and 601 FAR on Oxford, versus 3.43 m and 84 FAR when both use k_G. The pooled-reference construction also gives the lowest mean TE on Oxford and MCD; Appendix B.1 details these controls.

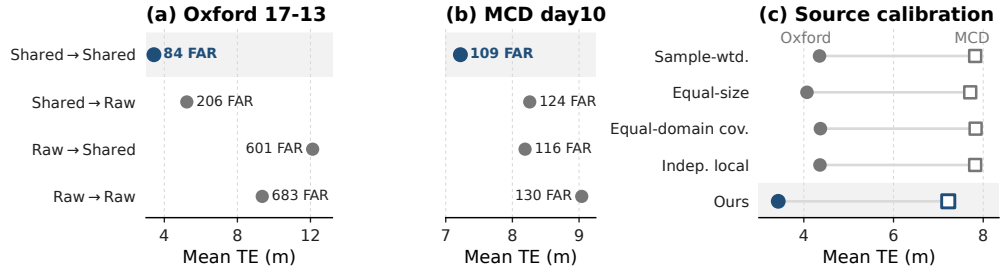


Figure 5: **Shared geometry across stages and source calibration.** (a–b) Mixed geometries raise error; Shared→Shared preserves the Eq. 11 frame kernel. (c) The proposed calibration gives the lowest mean TE on Oxford and MCD.

4.6 Efficiency, Challenging Scenes, and Long-Tail Suppression

Roughly 200 final candidates remove 96.1–99.5% of comparisons. Batch-1 speedups are 2.35–8.90×; at batch 16, they remain $5.77/4.43/1.57\times$ on Oxford/NCLT/MCD. Appendix I decomposes latency, storage, and batch scaling.

On the challenging long-horizon NCLT 05-26 and MCD day01→day10 settings, causal/symmetric TE falls from 9.14 to 5.52/4.99 m on NCLT and from 12.62 to 7.22/6.04 m on MCD, while FAR drops by 31.9–59.5%. Appendix A reports the full challenging-scene evaluation. With fixed EgoNN descriptors, the hierarchy lowers TE on every route and Oxford/NCLT latency from 7.4/6.9 to 4.2/4.9 ms (Appendix J).

Figure 6 shows that mean TE and failures above 50 m decrease together from raw retrieval through calibration to the full hierarchy. FAR falls from 1.01% to 0.30% on Oxford and from 6.35% to 4.33% on MCD. Alongside Figure 4, this indicates that F^{3}Loc mainly suppresses geographically distant failures; Appendix K decomposes the residual errors.

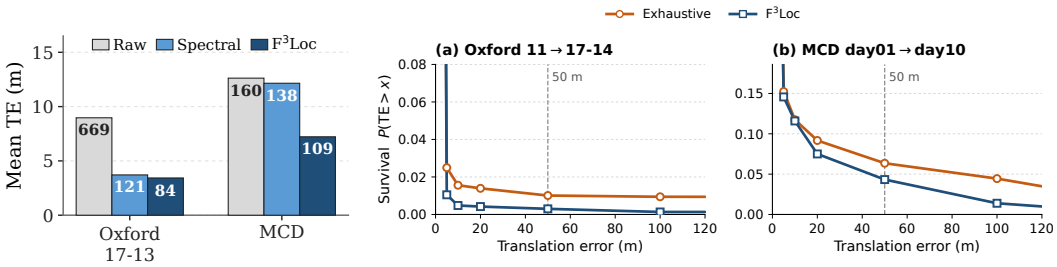


Figure 6: **Long-tail suppression.** Left: mean TE; numbers above bars report failures with TE > 50 m. Right: TE survival on Oxford and MCD.

5 Conclusion

Appendices A–L cover challenging routes, mechanisms, scaling, transfer, failures, and reproducibility. F^{3}Loc narrows the accuracy–efficiency–generalization trade-off by fixing a source-calibrated geometry before aggregation and preserving its frame kernel through coherent, granularity-aligned routing, while removing over 96% of comparisons without target-scene training.

AI USE STATEMENT

Generative AI tools were used in a limited advisory capacity for language refinement and presentation-oriented brainstorming. All technical content, implementation, experiments, analysis, and conclusions were independently developed, verified, and finalized by the authors.

REFERENCES

Sheng Ao, Qingyong Hu, Hanyun Wang, Kai Xu, and Yulan Guo. BUFFER: Balancing accuracy, efficiency, and generalizability in point cloud registration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1255–1264, 2023.

Axel Barroso-Laguna, Tommaso Cavallari, Victor Adrian Prisacariu, and Eric Brachmann. A scene is worth a thousand features: Feed-forward camera localization from a collection of image features. In *International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=rmDA02o8MV.

Xuewei Cao, Jiayue Yang, Zhiwen Zeng, Yanyong Zhang, and Yan Xia. C-LaV: Conditional latent velocity field denoising for weather-robust LiDAR place recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2928–2937, 2026.

Daniele Cattaneo, Matteo Vaghi, and Abhinav Valada. LCDNet: Deep loop closure detection and point cloud registration for LiDAR SLAM. *IEEE Transactions on Robotics*, 38(4):2074–2093, 2022. doi: 10.1109/TRO.2022.3150683.

Yidong Chen, Qi Li, Yuyang Yang, Wen Li, Sheng Ao, and Cheng Wang. Unleashing the power of data generation in one-pass outdoor lidar localization. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pp. 199–208, 2025.

Yu-Hsi Chen and Abd-Krim Seghouane. CPCANet: Deep unfolding common principal component analysis for domain generalization. *arXiv preprint arXiv:2605.05136*, 2026.

Sungha Choi, Sanghun Jung, Huiwon Yun, Joanne T. Kim, Seungryong Kim, and Jaegul Choo. RobustNet: Improving domain generalization in urban-scene segmentation via instance selective whitening. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11580–11590, 2021.

Bernhard N. Flury. Common principal components in k groups. *Journal of the American Statistical Association*, 79(388):892–898, 1984. doi: 10.1080/01621459.1984.10477108.

Chencan Fu, Lin Li, Jianbiao Mei, Yukai Ma, Linpeng Peng, Xiangrui Zhao, and Yong Liu. A coarse-to-fine place recognition approach using attention-guided descriptors and overlap estimation. In *IEEE International Conference on Robotics and Automation*, 2024.

Sheldon Fung, Wei Pan, Ling Cao, Fei Hou, Ling Chen, Shasha Mao, Hongdong Li, and Xuequan Lu. SuP: Sub-cloud driven point cloud registration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 24185–24194, 2026.

Sourav Garg and Michael Milford. Seqnet: Learning descriptors for sequence-based hierarchical place recognition. *IEEE Robotics and Automation Letters*, 6(3):4305–4312, 2021.

Ethan Griffiths, Maryam Haghighat, Simon Denman, Clinton Fookes, and Milad Ramezani. HOTFormerLoc: Hierarchical octree transformer for versatile lidar place recognition across ground and aerial views. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6648–6658, 2025.

Hervé Jégou and Ondřej Chum. Negative evidences and co-occurences in image retrieval: The benefit of PCA and whitening. In *European Conference on Computer Vision*, pp. 774–787. Springer, 2012. doi: 10.1007/978-3-642-33709-3_55.

Jeff Johnson, Matthijs Douze, and Hervé Jégou. Billion-scale similarity search with GPUs. *IEEE Transactions on Big Data*, 7(3):535–547, 2021. doi: 10.1109/TBDATA.2019.2921572.

Giseop Kim, Sunwook Choi, and Ayoung Kim. Scan context++: Structural place recognition robust to rotation and lateral variations in urban environments. *IEEE Transactions on Robotics*, 2022.

Jacek Komorowski, Monika Wysoczanska, and Tomasz Trzcinski. Egonn: Egocentric neural network for point cloud based 6dof relocalization at the city scale. *IEEE Robotics and Automation Letters*, 7(2):722–729, 2022. doi: 10.1109/LRA.2021.3133593.

W. J. Krzanowski. Principal component analysis in the presence of group structure. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 33(2):164–168, 1984. doi: 10.2307/2347442.

Wen Li, Shangshu Yu, Cheng Wang, Guosheng Hu, Siqi Shen, and Chenglu Wen. Sgloc: Scene geometry encoding for outdoor lidar localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9286–9295, 2023.

Wen Li, Yuyang Yang, Shangshu Yu, Guosheng Hu, Chenglu Wen, Ming Cheng, and Cheng Wang. Diffloc: Diffusion model for outdoor lidar localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15045–15054, 2024.

Wen Li, Chen Liu, Shangshu Yu, Dunqiang Liu, Yin Zhou, Siqi Shen, Chenglu Wen, and Cheng Wang. Lightloc: Learning outdoor lidar localization at light speed. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 6680–6689, 2025.

Zhe Liu, Chuanzhe Suo, Shunbo Zhou, Huanshu Wei, Yingtian Liu, Hesheng Wang, and Yun-Hui Liu. SeqLPD: Sequence matching enhanced loop-closure detection based on large-scale point cloud description for self-driving vehicles. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1218–1223, 2019.

Sha Lu, Xuecheng Xu, Dongkun Zhang, Yuxuan Wu, Haojian Lu, Xieyuanli Chen, Rong Xiong, and Yue Wang. RING#: PR-by-PE global localization with roto-translation equivariant gram learning. *IEEE Transactions on Robotics*, 41:1861–1881, 2025.

Lun Luo, Si-Yuan Cao, Xiaorui Li, Jintao Xu, Rui Ai, Zhu Yu, and Xieyuanli Chen. BEVPlace++: Fast, robust, and lightweight LiDAR global localization. *IEEE Transactions on Robotics*, 2025.

Junyi Ma, Xieyuanli Chen, Jingyi Xu, and Guangming Xiong. SeqOT: A spatial-temporal transformer network for place recognition using sequential LiDAR data. *IEEE Transactions on Industrial Electronics*, 2022.

Yu. A. Malkov and D. A. Yashunin. Efficient and robust approximate nearest neighbor search using hierarchical navigable small world graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(4):824–836, 2020. doi: 10.1109/TPAMI.2018.2889473.

Michael J. Milford and Gordon F. Wyeth. SeqSLAM: Visual route-based navigation for sunny summer days and stormy winter nights. In *IEEE International Conference on Robotics and Automation*, pp. 1643–1649, 2012.

Zheng Qin, Hao Yu, Changjian Wang, Yulan Guo, Yuxing Peng, Slobodan Ilic, Dewen Hu, and Kai Xu. GeoTransformer: Fast and robust point cloud registration with geometric transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(8):9806–9821, 2023.

Yanqing Shen, Turcan Tuna, Marco Hutter, Cesar Cadena, and Nanning Zheng. ForestLPR: LiDAR place recognition in forests attentioning multiple BEV density images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6659–6669, 2025.

Ziwei Shi, Xiaoran Zhang, Yan Xia, Yu Zang, Siqi Shen, and Cheng Wang. L2RSI: Cross-view LiDAR-based place recognition for large-scale urban scenes via remote sensing imagery. In *Advances in Neural Information Processing Systems*, volume 38, 2025.

Kavisha Vidanapathirana, Peyman Moghadam, Ben Harwood, Muming Zhao, Sridha Sridharan, and Clinton Fookes. Locus: LiDAR-based place recognition using spatiotemporal higher-order pooling. In *IEEE International Conference on Robotics and Automation*, pp. 5075–5081, 2021.

Jianshi Wu, Minghang Zhu, Dunqiang Liu, Wen Li, Sheng Ao, Siqi Shen, Chenglu Wen, and Cheng Wang. LEADER: Learning reliable local-to-global correspondences for LiDAR relocalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9932–9942, 2026.

Leyuan Xing, Huanjia Zhang, Dongyu Pan, Hai Wu, Qiming Xia, Kezheng Xiong, Wen Li, Chenglu Wen, and Cheng Wang. TACO: Task-aware contrastive learning for joint LiDAR localization and 3d object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9965–9975, 2026.

Xuecheng Xu, Sha Lu, Jun Wu, Haojian Lu, Qiuguo Zhu, Yiyi Liao, Rong Xiong, and Yue Wang. Ring++: Roto-translation invariant gram for global localization on a sparse scan map. *IEEE Transactions on Robotics*, 39(6):4616–4635, 2023.

Yuyang Yang, Wen Li, Sheng Ao, Qingshan Xu, Shangshu Yu, Yu Guo, Yin Zhou, Siqi Shen, and Cheng Wang. Raloc: Enhancing outdoor lidar localization via rotation awareness. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3304–3313, 2025.

Kaiyang Zhou, Yongxin Yang, Yu Qiao, and Tao Xiang. Domain generalization with mixstyle. In *International Conference on Learning Representations*, 2021.

SUPPLEMENTARY MATERIAL OVERVIEW

The supplement follows the same argument as the main paper and expands the evidence behind each decision in the hierarchy. Appendix A evaluates F^{3}Loc in challenging cross-domain scenes where distant aliases are most visible. Appendices B and C then examine source adaptation and the geometry inherited by individual descriptors. Appendices D and E turn to evidence support and routing. Appendices F–J document the controlled comparisons, cross-stage behavior, refinement, efficiency, and encoder transfer. Appendix K separates the residual errors by the stage at which they arise, and Appendix L closes with the protocol needed to reproduce the reported system.

Part	Question	Evidence
A	How does F^{3}Loc perform in challenging cross-domain scenes?	Long-horizon NCLT and MCD results, complementing the error-survival curves and spatial maps in the main paper.
B	What is fitted on the source side?	LoRA placement, source composition, target exposure, and calibration definitions.
C	How does descriptor geometry change?	Positive–negative margins and neighborhood retention under a learned mapping, followed by fixed-pool frame-ranking evidence for shared geometry.
D	Why align evidence granularity?	Candidate survival, a real routed sequence, and the frame-kernel decomposition of the regional score.
E	Is the operating point fragile?	Context extent, routing budget, and region-core length under fixed boundary context.
F	How do alternative retrieval structures form candidates?	IVF, HNSW, SeqLPD-, SeqNet-, and SeqSLAM-style controls under a common deployment protocol.
G	Must geometry be shared across stages?	All four coarse/fine geometry assignments on Oxford and MCD.
H	What does GeoRefine add?	Initialization, local alignment, pose composition, and fallback behavior.
I	How do cost and storage scale?	Batch latency, resident map arrays, candidate counts, and complexity.
J	Does the hierarchy transfer across encoders?	Route-wise EgoNN results with its checkpoint kept fixed.
K	Which failures remain?	Routing and fine-matching failure rates, followed by pose errors and limits of contiguous map support.
L	How is the system reproduced?	Dataset roles, hyperparameters, metrics, and offline/symmetric/causal procedures.

This organization separates source calibration, candidate formation, and local refinement.

A Performance in Challenging Cross-Domain Scenes

Appendix A evaluates F^{3}Loc on long-horizon NCLT 05-26 and MCD day01→day10, two challenging cross-domain settings with widely separated traverses. On MCD, the dominant difficulty is a long tail of geographically distant confusions. Exhaustive retrieval retains 84.76% R@1 but reaches 12.62 m mean TE and 160 FAR. Cross-day appearance variation, a complex urban layout, and frequent changes in travel direction are consistent with this long-error tail. Figure 7 summarizes recall, mean TE, and search size.

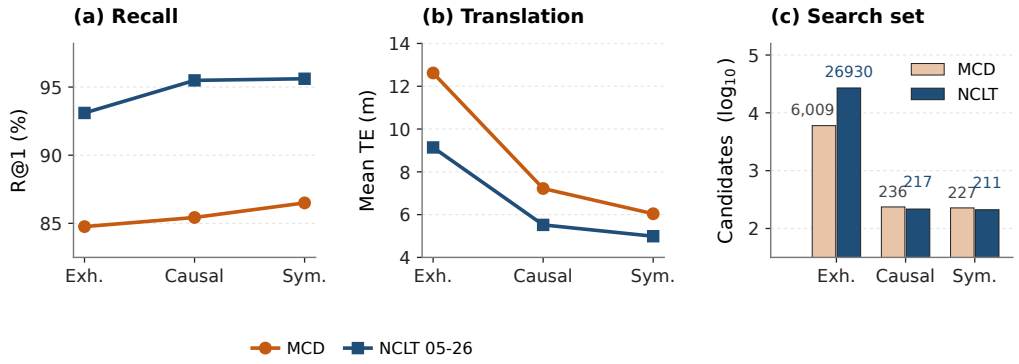


Figure 7: **Retrieval in challenging cross-domain settings.** Recall, mean translation error, and routed candidates on MCD day01→day10 and long-horizon NCLT 05-26. Candidate counts are shown on a logarithmic axis with raw values annotated. The hierarchy retains a small frame pool while reducing the large-error tail.

On NCLT 05-26, exhaustive retrieval gives 9.14 m mean TE. Causal and symmetric consensus reduce it to 5.52 and 4.99 m. On MCD, the corresponding values are 12.62, 7.22, and 6.04 m. These reductions extend beyond the center of the error distribution. FAR falls by 31.9–59.5%, which explains why the mean changes more strongly than the typical successful estimate.

The survival curves in main-paper Figure 6 show the large-error tail. Figure 4 shows spatially clustered errors, consistent with location-dependent ambiguity. Appendix K separates coarse hypothesis loss from errors after positive support survives.

B SOURCE-SIDE ADAPTATION AND SPECTRAL CALIBRATION

All trainable components used by the BEVPlace++ route are fitted before deployment. Rank-4 LoRA adapters with α=8 are inserted into the encoder's 1×1 convolutions and NetVLAD pooling layer. The base network stays frozen, leaving 1,536 trainable parameters. AdamW uses a learning rate of 5×10^{-5}, weight decay 10^{-4}, and batches of eight for 20 epochs with 20,000 source samples per epoch. Positives lie within 5 m and negatives beyond 20 m; the random seed is 42. The resulting descriptor encoder is fixed when the map and query traverses are processed. The routing hierarchy itself introduces no learned representation.

The source pool changes with the held-out target. Oxford uses KITTI and MulRan, giving 28,575 source descriptors. NCLT and MCD use KITTI, sampled Oxford, and MulRan, giving 30,898. Table 4 gives the sequence-level composition used for both LoRA adaptation and spectral fitting. The sampled Oxford entries are deliberately much smaller than the full evaluation traverses.

Table 4: **Source descriptors used for adaptation and spectral fitting.** Columns specify the held-out target; Oxford source traverses are sampled.

Domain	Sequence	Oxford target	NCLT / MCD target
KITTI	00	4,541	4,541
	02	4,661	4,661
	05	2,761	2,761
	06	1,101	1,101
	08	4,071	4,071
Oxford, sampled	11-13	–	1,893
	15-13	–	430
MulRan	DCC01	5,419	5,419
	DCC02	6,021	6,021
Total		28,575	30,898

LoRA training draws 20,000 triplets per epoch from the corresponding source pool. Spectral fitting uses the complete descriptor cache listed in Table 4; it does not resample the sources to the training epoch length. Our shared-basis spectral calibration is fitted once from these normalized descriptors and then reused without target statistics. Its basis $\mathbf{U}$ comes from the regularized pooled source covariance. Each source is measured in this basis, and the directional scales are averaged at the domain level to obtain $\boldsymbol{\lambda}^*$. This produces the single geometry $\mathcal{G} = (\mathbf{U}, \boldsymbol{\lambda}^*)$ used by both retrieval stages. Accordingly, $\mathcal{G}$ is fixed within each held-out-target protocol; different source pools are fitted separately.

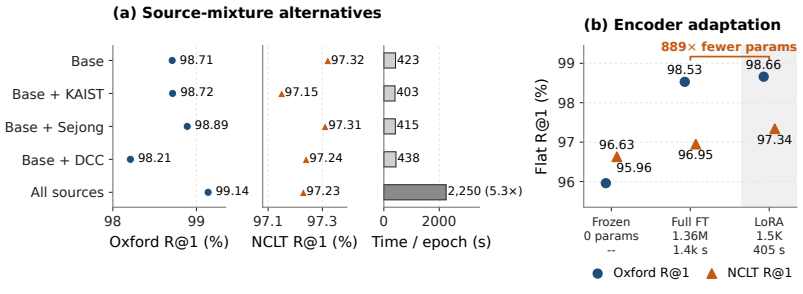


Figure 8: **Source composition and adaptation capacity.** Source mixtures change cross-domain recall non-monotonically. Under the fixed protocol, low-rank adaptation gives the strongest transfer among the frozen, full-tuning, and LoRA controls while changing only 1,536 parameters.

Figure 8 separates source diversity from adaptation capacity. Adding a source is not uniformly beneficial, while full fine-tuning does not improve the frozen representation as reliably as the low-rank update. The LoRA-adapted encoder is then frozen for all routing experiments.

B.1 CALIBRATION ALTERNATIVES

The calibration controls in Figure 5 differ in how source statistics define the comparison geometry. *Sample-weighted* pooling estimates one covariance from the union of all source descriptors, so each domain contributes in proportion to its sample count. *Equal-size* first subsamples the same number of descriptors from each source and then estimates a pooled covariance. *Equal-domain covariance* averages the full per-domain covariance matrices in the original descriptor coordinates before eigendecomposition. *Independent local* fits one whitening transform per source and then averages the transforms. Our design instead obtains a reference basis from the pooled covariance, projects every source-wise second-order matrix into that fixed basis, and averages only the resulting directional statistics. It therefore separates shared-direction construction from domain-level scale aggregation and never averages transforms expressed in independently determined bases.

Together, these controls examine how the reference basis is estimated and how source domains contribute to directional scaling. Sample-weighted pooling lets source sample count influence both choices; equal-size pooling equalizes source contributions by subsampling each domain to the same descriptor count; equal-domain covariance derives both directions and scales from one domain-averaged covariance; and independent local whitening estimates transforms in separate eigenbases before combining them. F^{3}Loc instead uses the full pooled cache for the basis and lets each source contribute once to scaling. This construction uses all source descriptors to define the reference basis, while domain-level scale averaging gives each source one explicit contribution to scaling. Measuring all source statistics in the same basis also avoids combining transforms defined in incompatible local coordinates. Figure 5(c) shows the lowest mean TE on both Oxford and MCD.

B.2 TARGET EXPOSURE

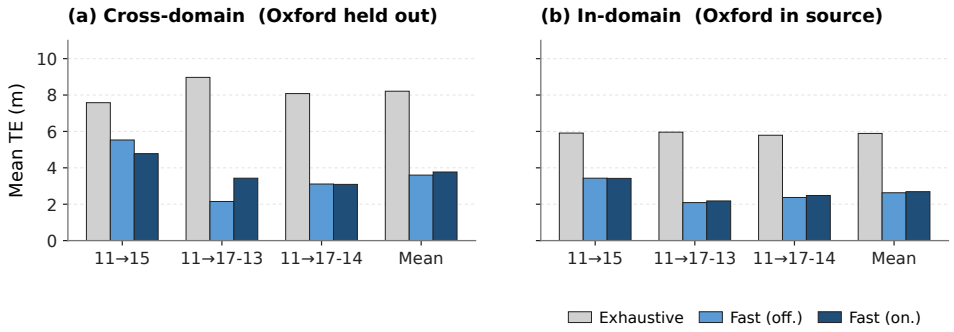


Figure 9: **Oxford MapPose with and without target exposure.** Cross-domain fitting uses KITTI and MulRan; the in-domain comparison adds Oxford. The retrieval stack is held fixed within each column.

Figure 9 distinguishes the hierarchy gain from target exposure. With Oxford excluded, routing reduces mean TE from 8.21 m for exhaustive retrieval to 3.60/3.77 m with symmetric/causal context; adding Oxford further improves all settings.

C DIAGNOSING THE EMBEDDING-CONSISTENCY MISMATCH

C.1 LEARNED STAGE MAPPING

This section examines whether a source-trained stage mapping can overturn a target-domain neighborhood that is correct at its input. This diagnostic uses Oxford 11-13 as the map and 17-13 as the query traverse. For query q, let $\mathcal{P}(q)$ contain map frames within 5 m and let $\mathcal{N}(q)$ contain the remaining frames. For representation g, define

$$\Delta_g(q) = \min_{m \in \mathcal{N}(q)} \|g(q) - g(m)\|_2^2 - \min_{m \in \mathcal{P}(q)} \|g(q) - g(m)\|_2^2. \tag{12}$$

A positive margin places a valid map frame ahead of every negative. We compare the fixed S1 input representation, g_0, with the source-trained frame mapping, g_θ, on exactly the same query–map pairs. A harmful inversion is the event $\Delta_{g_0}(q) > 0$ and $\Delta_{g_\theta}(q) < 0$.

Among the queries ordered correctly at the mapping input, 10.54% undergo this harmful inversion. Since the samples and granularity are unchanged, the inversion isolates the learned mapping from sequence pooling and candidate construction. It shows that source-trained parameters may alter a useful target-domain ordering. This observation motivates keeping routing in the same fixed comparison geometry as final localization.

C.2 Neighborhood retention

The margin test asks whether the nearest positive–negative ordering becomes incorrect. Neighborhood retention measures how broadly the local candidate set changes. If $\mathcal{T}_K^0(q)$ and $\mathcal{T}_K^\theta(q)$ are the Top-K map identities before and after the learned mapping, respectively, then

$$\text{Retention@}\,K(q) = \frac{|\mathcal{T}_K^0(q) \cap \mathcal{T}_K^\theta(q)|}{K}. \tag{13}$$

The statistic records whether the same map hypotheses survive without conflating identity changes with absolute descriptor distances.

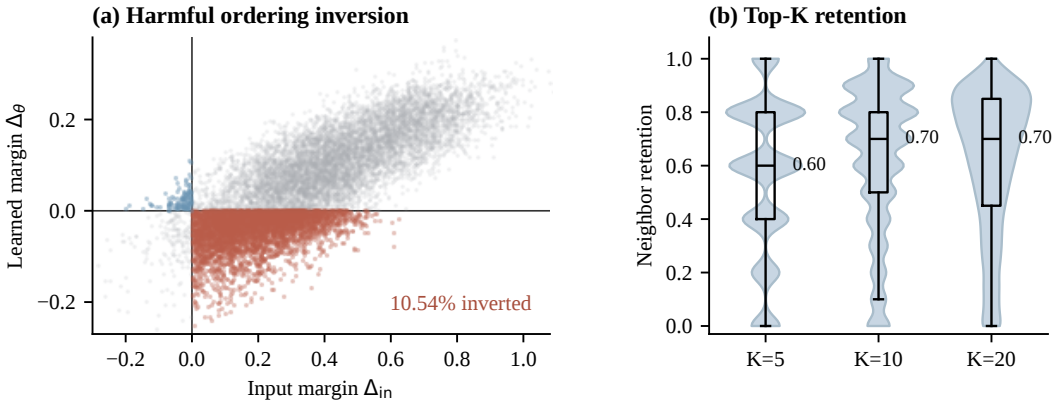


Figure 10: **Effect of a source-trained stage mapping on target-domain neighborhoods.** (a) Harmful ordering reversals affect 10.54% of initially correct queries. (b) Top-K retention measures neighborhood replacement; median overlap is 0.60, 0.70, and 0.70 for $K = 5, 10, 20$.

The mean retained fractions are 0.56, 0.63, and 0.62 for $K = 5, 10, 20$. A changed neighborhood is not necessarily worse, but the margin inversions in Figure 10 identify changes that reverse a correct local decision. The diagnostics are complementary. Margin inversion measures lost ordering, whereas retention measures the breadth of candidate replacement. Together with the end-to-end SeqNet-style result in Table 2, they connect local geometric reshaping to the candidates available after an irreversible coarse decision.

C.3 Frame-level effect of shared geometry

To isolate final frame ranking from candidate survival, Table 5 conditions evaluation on queries whose routed pool already contains a 5 m positive. Within each row, raw and shared fine matching receive the same candidate pool; only the frame-level comparison geometry changes.

Shared fine geometry reduces both conditional mean TE and distant failures under either fixed routing pool. This complements the stage-wise coverage result in Appendix G, where the coarse geometry keeps the correct region reachable while the fine geometry improves frame ordering after that region has survived.

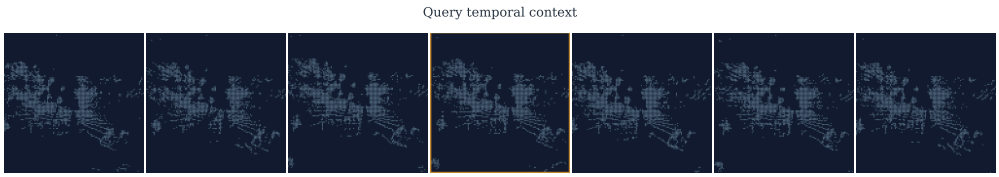
Table 5: **Fine-stage ranking on fixed positive candidate pools, Oxford 17-13.** Shared geometry improves the ordering of frames already exposed to final matching. FAR counts errors with TE > 50 m.

	Raw fine geometry		Shared fine geometry	
Routing pool	TE	FAR	TE	FAR
Raw routing	3.39	110	**2.46**	**32**
Shared routing	4.39	150	**2.42**	**30**

D Granularity Alignment and Candidate Survival

The hierarchy must decide where to search before it decides which frame is nearest. A current frame and a region summary carry different evidence support; matching them directly reduces computation but worsens Oxford/NCLT TE from 8.97/2.59 to 10.26/3.31 m in Table 3. Query context restores matched support and recovers 8.04/1.93 m even before shared spectral geometry is added.

Equation 11 makes this relation explicit. The regional score is a normalized sum of the same frame-kernel values used by the fine stage. Routing changes the set over which evidence is accumulated, but does not introduce a region-specific metric. Thus temporal consensus can suppress an isolated alias while preserving the frame-level distinction needed after routing.



(a) Query context



(b) Correct map support

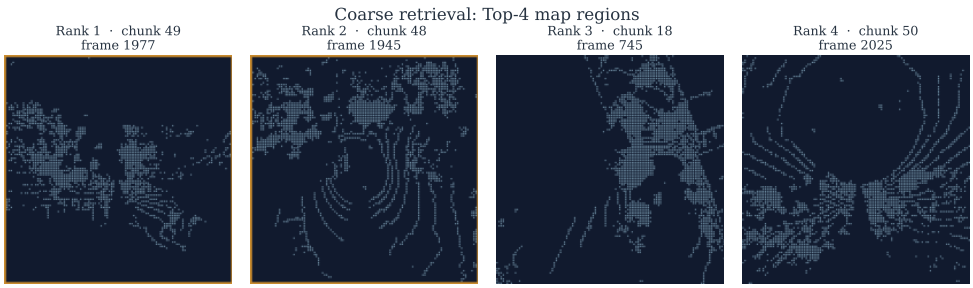
Figure 11: **A successful NCLT 02-12→02-18 retrieval.** The query and correct map sequences share extended structural support. Routing scores this support through frame similarities; final localization returns to the current query frame and unpooled map frames.

D.1 What routing retains and rejects

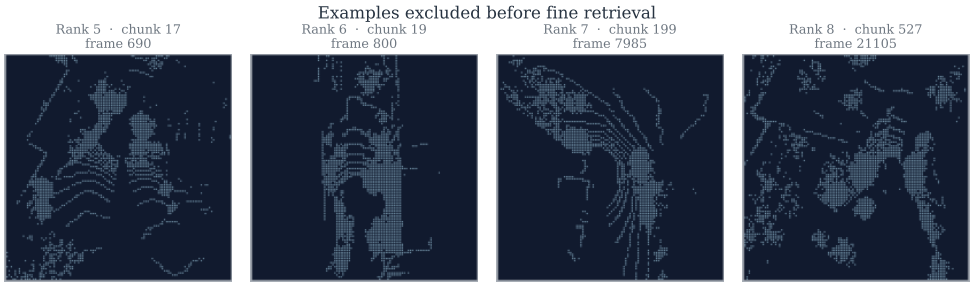
Figure 12 follows the same query through the coarse decision. The Top-4 regions contain the candidate support exposed to final matching. Frames outside these regions are never evaluated by the fine stage. The design objective is therefore not merely a small pool, but a pool that keeps the geographically correct support while removing distant alternatives.

As reported directly in Appendix G, shared coarse geometry retains a positive region on Oxford for 99.85% of valid queries, compared with 98.61% for raw coarse routing. This 1.24-point gain keeps 524 additional queries reachable and, together with the fine geometry, reduces failures above 50 m from 601–683 to 84.

Symmetric and causal inference use the same kernel and region construction. Symmetric context has access to a centered neighborhood and gives the strongest offline results; causal context uses only the observed prefix and remains the online variant. Their difference is evidence availability, not a change in comparison rule.



(a) Frames retained by Top-4 routing



(b) Examples excluded before fine matching

Figure 12: **Candidate survival in a real hierarchical query.** Coarse routing retains contiguous map support rather than a collection of unrelated frame hits. The fine stage compares the current descriptor only with frames inside the retained regions.

E Routing Sensitivity

The fixed operating point uses 50-frame region cores with 10 frames of boundary context on each side, Top-4 routing, symmetric extent $W = 50$, and causal extent $W = 40$. Figure 13 varies the parameters that most directly control the evidence support and candidate pool.

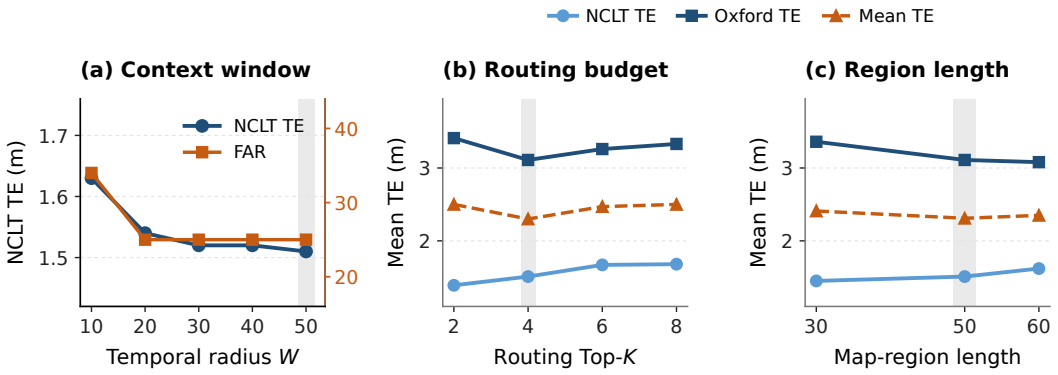


Figure 13: **Post-hoc sensitivity on NCLT 02-12 and Oxford 11-13→17-14.** (a) Symmetric context extent. (b) Routing Top-K; the dashed curve is cross-dataset mean TE. (c) Region-core length. Gray bands mark the operating point used in the paper.

Context aggregation saturates quickly. Once $W \geq 20$, positive-pool coverage reaches 100% in the reported NCLT sweep, R@1 remains within 0.1 percentage point, and mean TE changes from 1.54 to 1.51 m. The narrow $W = 10$ setting is the exception, with lower coverage and more large errors.

The causal extents are similarly close. Across the six Table 1 routes, $W = 40$ and $W = 50$ give 2.695 and 2.720 m macro TE.

Routing budget and region size expose the expected cross-target tradeoff. Top-$K = 2$ gives the highest Oxford R@1 in the plotted sweep but raises Oxford TE to 3.41 m; Top-$K = 4$ lowers cross-target mean TE to 2.30 m. Thirty-frame regions improve NCLT while raising Oxford TE to 3.36 m. Fifty-frame regions give the lower cross-target mean of 2.31 m.

The reported curves cover the parameters that directly change the amount of evidence or the final candidate budget. All F^{3}Loc variants and target datasets use the same 10-frame core overlap and boundary context, holding the physical map partition constant across these ablations.

F Controlled Retrieval Comparisons

Baselines suffixed with "-*style*" implement the candidate-formation mechanisms of the corresponding methods under a common deployment descriptor bank, target protocol, and candidate accounting. The original systems use different descriptors, training data, and inference interfaces; these controls isolate their hierarchy mechanisms. All use the frozen BEVPlace++ bank and map-positive Oxford 17-13 queries without target training. Non-learned controls use the common frame-distance matcher; SeqNet-style retains its source-trained frame reranking branch as part of the learned hierarchy.

Table 6: **Retrieval controls under the common deployment protocol.** "Train" denotes parameters learned beyond the fixed encoder; "Target use" distinguishes fixed target-map index construction from target-domain parameter fitting.

Method	Encoder	Train	Target use	Coarse evidence	Final matcher	Budget
IVF-Flat	BEVPlace++	no	map index	descriptor centroid	frame distance	281
HNSW	BEVPlace++	no	map index	graph neighborhood	frame distance	dynamic
HNSW+$\mathcal{G}$	BEVPlace++	no	map index	graph neighborhood	frame distance in $\mathcal{G}$	dynamic
SeqLPD-style	BEVPlace++	no	no	super keyframe	frame distance	635
SeqSLAM-style	BEVPlace++	no	no	ordered frame scores	frame distance	128
SeqNet-style	BEVPlace++	yes, source	no	learned sequence key	learned frame branch	220
SeqNet-style+$\mathcal{G}$	BEVPlace++	yes, source	no	learned sequence key	learned frame branch	220
F^{3}Loc	BEVPlace++	no routing fit	no	context and region	frame distance in $\mathcal{G}$	222

IVF-Flat and HNSW. IVF-Flat fits a coarse quantizer on the map, probes one cell, and exhaustively reranks the frames assigned to that cell. HNSW uses $M = 32$ and $efSearch = 128$ for all comparisons. HNSW+$\mathcal{G}$ changes only the descriptor geometry before index construction; its graph procedure is unchanged. Its candidate count is omitted because graph visits do not define an explicit shortlist comparable to region routing.

Sequence controls. SeqLPD-style groups appearance-similar super keyframes and expands selected groups to frame candidates. Its support can be geographically disconnected because membership is defined in descriptor space. SeqSLAM-style begins with the common frame-wise similarity matrix, scores length-20 causal paths, and retains 128 endpoints. Consecutive high similarities can therefore support either the true route or a coherent distant alias.

Learned hierarchical control. The SeqNet-style implementation learns a 256-D projection and causal convolution for length-5 sequence descriptors, together with a linear frame reranking branch. Both branches are trained on KITTI and MulRan for 30 epochs. The Top-44 sequence endpoints expand to 220 frame comparisons. The $+\mathcal{G}$ row transforms the input descriptors but leaves these source-trained stage mappings intact; this is a test of whether a shared input calibration remains sufficient after the hierarchy learns distinct comparison functions.

These controls separate the source of each retrieval decision. Industrial indices alter how frame neighbors are explored, sequence controls alter how evidence is accumulated, and the learned

SeqNet-style control introduces distinct coarse and fine mappings. The comparison thus isolates retrieval structure under a common deployment protocol while retaining each control's decision rule.

Table 7: **Online latency under the Table 1/Table 2 protocol.** Total inference includes query encoding and the complete online retrieval path at B=16. The middle column isolates descriptor-resident lookup or routing; map and index construction are offline.

Method	Lookup/routing (ms/query)	Total inference (ms/query)	Execution
IVF-Flat	1.55	5.7	CPU lookup
HNSW	0.24	4.4	CPU lookup
HNSW+$\mathcal{G}$	0.46	4.6	CPU lookup
SeqLPD-style	9.40	13.5	GPU routing/matching
SeqNet-style	0.46	4.6	GPU routing/matching
SeqNet-style+$\mathcal{G}$	0.39	4.5	GPU routing/matching
SeqSLAM-style	125.28	129.4	GPU path scoring
F^{3}Loc-Fast	3.06	7.2	GPU routing/matching

The query encoder and its data path are shared across rows. Total inference combines this common path with the measured method-specific online stage. ANN indexing and hierarchical localization address different bottlenecks. The former accelerates descriptor-space lookup, while F^{3}Loc structures the evidence support retained for final localization.

G Cross-Stage Geometry Consistency

Shared geometry plays two roles that can be separated experimentally. At the coarse stage, it determines which physical regions survive. At the fine stage, it determines the ranking of frames inside the surviving pool. Table 8 evaluates all four raw/shared assignments without changing context, region construction, or Top-K.

Table 8: **Geometry assigned to coarse routing and fine matching.** Pool coverage is the fraction of valid queries whose routed pool contains a 5 m positive. TE is mean translation error. All entries use causal query context.

		Oxford 17-13				MCD day01→day10			
Coarse	Fine	Cover.	R@1	TE	FAR	Cover.	R@1	TE	FAR
Shared	Shared	**99.85**	**98.69**	**3.43**	**84**	**95.83**	**85.43**	**7.22**	**109**
Raw	Shared	98.61	97.46	12.12	601	94.52	83.80	8.19	116
Shared	Raw	**99.85**	98.06	5.23	206	**95.83**	84.36	8.26	124
Raw	Raw	98.61	96.95	9.36	683	94.52	82.69	9.04	130

Oxford makes the two-stage effect particularly clear. Replacing raw coarse routing with shared geometry raises pool coverage from 98.61 to 99.85%, independent of the fine metric. Once the pool is fixed, shared fine matching lowers TE from 5.23 to 3.43 m. Using shared geometry only at the fine stage cannot recover a positive already removed by raw routing and instead gives 12.12 m. MCD shows the same ordering with a smaller margin.

The regional score in Eq. 11 aggregates the frame kernel used by final localization. Mixed assignments change this relationship and can remove hypotheses that remain plausible to the fine matcher.

H GeoRefine

F^{3}Loc-Fast returns the stored pose of the top-ranked map frame. GeoRefine is an optional local stage initialized by the global association. It refines local translation and orientation after a valid match; distant incorrect regions remain outside its capture range.

Frozen BEV features are sampled with FAST keypoints and matched by L2 distance at 0.4 m per pixel. A planar transform is estimated with 1,000 RANSAC iterations, a 0.5 m inlier threshold, and a minimum of eight inliers. The planar estimate is lifted to $\mathrm{SE}(3)$ and composed with the retrieved MapPose. Where scans are available, 0.5 m voxelized point-to-plane ICP then uses at most 30 iterations, a 2 m correspondence threshold, and minimum fitness 0.1. Failed validity checks retain the preceding estimate.

Fast measures global association and direct MapPose readout, whereas GeoRefine measures the local precision available after that association.

I Efficiency and Map Scaling

The computational distinction between flat and hierarchical retrieval is concentrated in the frame comparison. Exhaustive matching evaluates all M map descriptors. F^{3}Loc evaluates J region descriptors and then the union $|\mathcal{P}_t|$ of frames in the selected regions. With a fixed 50-frame core, 10-frame boundary context, and Top-K, the fine-stage pool remains near 200 frames across the evaluated maps, while the full map ranges from 6,009 to more than 40,000 frames.

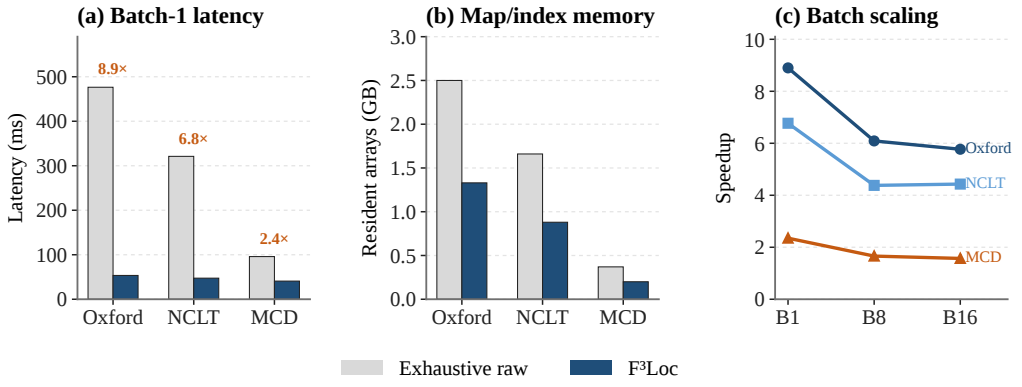


Figure 14: **Efficiency and map scaling.** The panels report batch-1 latency, resident map/index storage, and speedup across batch sizes. Within each dataset, exhaustive retrieval and F^{3}Loc use the same encoded map; the F^{3}Loc routing budget is fixed.

At batch 1, measured speedups range from 2.35× on the smaller MCD map to 8.90× on Oxford. At batch 16, they remain 1.57×, 5.77×, and 4.43× on MCD, Oxford, and NCLT. The reduced comparison set therefore remains visible after batching, although exhaustive matrix multiplication benefits more strongly from large batches.

The resident map and index arrays fall by approximately 47% because the deployment cache need not retain every auxiliary bank used by exhaustive evaluation. Region descriptors add only a small index. Peak GPU allocation rises by roughly 302 MiB in the measured implementation because query context and temporary routing tensors coexist during inference; this is a workspace cost rather than persistent map storage.

Offline costs are map encoding, projection into $\mathcal{G}$, and construction of overlapping region summaries. They are paid once per map and excluded from query latency; query encoding is included. The resulting online complexity is $O((|\mathcal{W}_t| + J + |\mathcal{P}_t|)d)$ rather than $O(Md)$. Region search remains linear in J in the present implementation; an approximate index over region descriptors is compatible with the formulation but is not used in the reported numbers.

J GENERALITY ACROSS DESCRIPTOR ENCODERS

The BEVPlace++ experiments combine source-side adaptation with the proposed retrieval hierarchy. To test whether the routing principle depends on that encoder, we also apply it to a fixed EgoNN checkpoint. No EgoNN parameters are retrained. The same causal region construction and Top-4 routing are used, and the source-fitted geometry is recomputed in the EgoNN descriptor space.

Table 9: **EgoNN ablation with a fixed checkpoint.** Entries are mean TE in meters. Raw hierarchy adds physical regions and causal context; the full row additionally uses the shared spectral geometry fitted in the EgoNN descriptor space.

Variant	Oxford 15-13	Oxford 17-13	Oxford 17-14	NCLT 02-12	NCLT 02-19	NCLT 03-31	Cand. (Ox./NCLT)
Flat raw EgoNN	26.93	22.80	19.88	15.53	12.25	10.94	40,565 / 26,930
+ temporal hierarchy (raw)	18.91	17.93	13.65	9.00	4.71	5.22	235–239 / 222–226
+ shared spectral geometry (F^3Loc)	**16.03**	**15.14**	**12.04**	**7.11**	**4.50**	**4.81**	232–235 / 222–224

The flat row follows the finalized baseline in Table 1. Temporal hierarchy reduces TE while shrinking the frame pool by two orders of magnitude. Shared geometry further improves each route under the fixed routing budget. Oxford 17-13 gives 22.80→17.93→15.14 m, and NCLT 02-12 gives 15.53→9.00→7.11 m. These results extend the comparison of coherent support and shared geometry to a second descriptor encoder.

The absolute errors remain encoder-dependent. The consistent ablation trend across both encoders indicates that the routing mechanism does not require a BEVPlace++-specific routing head.

K STAGE-WISE FAILURE ANALYSIS

Where does the hierarchy first fail? A routing failure occurs when the candidate pool contains no 5 m positive. A fine-matching failure occurs when a positive survives but the returned frame is farther than 5 m. These disjoint events have rates $1 - \mathrm{Coverage}$ and $\mathrm{Coverage} - \mathrm{R@1}$, respectively.

Table 10: **Residual association failures by stage.** Percentage points of all valid queries, derived from the rounded Shared→Shared entries in Table 8.

Target	Routing failure	Fine-matching failure	Total R@1 failure
Oxford 17-13	0.15	1.16	1.31
MCD day10	4.17	10.40	14.57

Most residual association failures occur after positive support has survived. Oxford routing loses only 0.15% of valid queries, while fine ranking accounts for 1.16%. MCD exposes both a larger routing loss and a larger fine-ranking error. This separates the need for stronger regional evidence from the need for better frame discrimination.

Pose readout and local geometry. Even a frame within 5 m can retain translation offsets and heading error because Fast returns a discrete stored MapPose. GeoRefine addresses local alignment after association. The Fast/GeoRefine results in Table 1 illustrate this distinction; they report aggregate metrics across valid queries.

Structural support. Contiguous trajectory intervals are an operational proxy for spatially coherent support on the evaluated routes. Distance- or graph-based partitions could extend the formulation to branched, multi-pass, or otherwise complex map topologies. This structural issue is distinct from an individual query's routing or fine-ranking failure.

L REPRODUCIBILITY DETAILS

This section records the dataset roles, fixed operating points, and execution order needed to reproduce the reported protocol. Source fitting, map construction, and target inference are kept distinct throughout.

Table 11: **Dataset roles in the principal cross-domain evaluations.**

Target	Source fitting	Map	Query
Oxford	KITTI + MulRan	11-13	15-13, 17-13, 17-14
NCLT	KITTI + Oxford + MulRan	02-18	02-12, 02-19, 03-31
NCLT long horizon	same as NCLT	02-18	05-26
MCD	KITTI + Oxford + MulRan	day01	day10

Table 12: **Fixed retrieval and refinement parameters.**

Quantity	Value	Quantity	Value
Positive radius	5 m	FAR threshold	50 m
Region core	50 frames	Boundary context	10 frames/side
Routed regions	4	Symmetric extent	$W = 50$
Causal extent	$W = 40$	Spectral regularizer	10^{-3}
LoRA rank / α	4 / 8	LoRA epochs	20
RANSAC iterations	1,000	RANSAC threshold	0.5 m
ICP voxel / threshold	0.5 / 2 m	ICP iterations	30

R@1 is the fraction of map-positive queries whose returned frame lies within 5 m. TE is the Euclidean distance between the predicted and ground-truth 3-D translations, and RE is the geodesic angle between their rotations. FAR counts queries with TE above 50 m. Mean TE is always computed over the complete map-positive query set, including large errors. Latency includes query encoding, projection, temporal consensus, routing, and final matching; map encoding and index construction are offline.

L.1 EVALUATION AND CANDIDATE ACCOUNTING

Map scans are encoded once and never enter the online timing path. Query scans are encoded online; the current descriptor contributes both to its temporal context and to final frame matching. Symmetric inference is evaluated after a centered context is available, whereas causal inference uses only the observed prefix. They share the map representation, region partition, and final frame rule, so their difference is confined to the query evidence available to routing.

Candidate counts refer to unique map frames exposed to final matching, not to region scores or repeated comparisons. If $\mathcal{R}_t$ is the set of four routed regions, the fine pool is $\mathcal{C}_t = \bigcup_{r \in \mathcal{R}_t} r$. Each region starts from a 50-frame core and adds 10 boundary frames on both sides before clipping; duplicate frame indices are removed after the four supports are merged. A region therefore contains up to 70 frames, while overlap among selected regions and map boundaries reduce the union. This construction explains the observed averages above 200 and their small variation across routing regimes and encoders.

All aggregate localization metrics use the same precomputed map-positive query mask. Every eligible query contributes to mean TE and RE, and errors beyond 50 m additionally contribute to FAR. R@1, TE, RE, FAR, and candidate count are computed from the same returned frame before optional local refinement unless a GeoRefine result is explicitly named.

L.2 EXECUTION SEQUENCE

Source calibration. Normalize the cached source descriptors and estimate the regularized pooled covariance that defines $\mathbf{U}$. Rotate each source covariance into this basis, average its directional scales, and form $\boldsymbol{\lambda}^*$. The resulting $\mathcal{G} = (\mathbf{U}, \boldsymbol{\lambda}^*)$ is fixed before target processing; Table 4 lists the cache used for each held-out target.

Algorithm 1: offline map construction. Encode each map scan once, transform its descriptor with $\mathcal{G}$, partition the trajectory into 50-frame cores with 10 frames of boundary context on either side, and store each normalized regional mean with the transformed frame descriptors and MapPoses. One frame bank supports both region construction and final matching.

Algorithm 2: symmetric inference. For query time t, transform every available descriptor in $\mathcal{W}_t^{\mathrm{sym}}(50)$ and normalize their mean. Retrieve the four nearest regions, form the union of their frame indices, and match the current transformed query descriptor against this union. Return the MapPose of the nearest frame, optionally followed by GeoRefine.

Algorithm 3: causal inference. Replace the centered context by the observed prefix $\mathcal{W}_t^{\mathrm{cau}}(40)$. Region scoring, candidate union, frame matching, and pose readout remain unchanged. This is the online configuration reported as F^{3}Loc-Fast (cau.).

All routes use fixed budgets and a precomputed map-positive mask. Target descriptors never select the geometry, map caches remain immutable, and fixed-cache retrieval is deterministic; only source training and RANSAC are stochastic.